INSTRUCTOR'S SOLUTIONS MANUAL

WILLIAM CRAINE III

STATS: DATA AND MODELS

FOURTH EDITION

Richard De Veaux
Williams College

Paul Velleman
Cornell University

David Bock
Cornell University

PEARSON

Boston Columbus Hoboken Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Reproduced by Pearson from electronic files supplied by the author.

Publishing as Pearson, 501 Boylston Street, Boston, MA 02116.

ISBN-13: 978-0-321-98994-9
ISBN-10: 0-321-98994-5

www.pearsonhighered.com

PEARSON

Contents

Chapter 1	Stats Starts Here	*1*
Chapter 2	Displaying and Describing Categorical Data	*6*
Chapter 3	Displaying and Summarizing Quantitative Data	*23*
Chapter 4	Understanding and Comparing Distributions	*40*
Chapter 5	The Standard Deviation as a Ruler and the Normal Model	*57*
Review of Part I	Exploring and Understanding Data	*79*
Chapter 6	Scatterplots, Association, and Correlation	*97*
Chapter 7	Linear Regression	*112*
Chapter 8	Regression Wisdom	*144*
Chapter 9	Re-expressing Data: Get It Straight!	*162*
Review of Part II	Exploring Relationships Between Variables	*180*
Chapter 10	Understanding Randomness	*203*
Chapter 11	Sample Surveys	*213*
Chapter 12	Experiments and Observational Studies	*223*
Review of Part III	Gathering Data	*241*
Chapter 13	From Randomness to Probability	*255*
Chapter 14	Probability Rules!	*267*
Chapter 15	Random Variables	*289*
Chapter 16	Probability Models	*309*
Review of Part IV	Randomness and Probability	*340*
Chapter 17	Sampling Distribution Models	*360*
Chapter 18	Confidence Intervals for Proportions	*390*
Chapter 19	Testing Hypotheses About Proportions	*407*
Chapter 20	Inferences About Means	*428*
Chapter 21	More About Tests and Intervals	*449*
Review of Part V	From the Data at Hand to the World at Large	*467*
Chapter 22	Comparing Groups	*491*
Chapter 23	Paired Samples and Blocks	*536*
Chapter 24	Comparing Counts	*556*
Chapter 25	Inferences for Regression	*582*
Review of Part VI	Accessing Associations Between Variables	*609*
Chapter 26	Analysis of Variance	*652*
Chapter 27	Multifactor Analysis of Variance	*664*
Chapter 28	Multiple Regression	*675*
Review of Part VII	Inferences When Variables Are Related	*684*
Chapter 29	Multiple Regression Wisdom	*708*

Chapter 1 – Stats Starts Here

Section 1.1

1. **Grocery shopping.** Discount cards at grocery stores allow the stores to collect information about the products that the customer purchases, what other products are purchased at the same time, whether or not the customer uses coupons, and the date and time that the products are purchased. This information can be linked to demographic information about the customer that was volunteered when applying for the card, such as the customer's name, address, sex, age, income level, and other variables. The grocery store chain will use that information to better market their products. This includes everything from printing out coupons at the checkout that are targeted to specific customers to deciding what television, print, or Internet advertisements to use.

2. **Online shopping.** Amazon hopes to gain all sorts of information about customer behavior, such as how long they spend looking at a page, whether or not they read reviews by other customers, what items they ultimately buy, and what items are bought together. They can then use this information to determine which other products to suggest to customers who buy similar items, to determine which advertisements to run in the margins, and to determine which items are the most popular so these items come up first in a search.

Section 1.2

3. **Super Bowl.** When collecting data about the Super Bowl, the games themselves are the *who.*

4. **Nobel laureates.** Each year is a case, holding all of the information about that specific year. Therefore, the year is the *who.*

Section 1.3

5. **Grade level.**

   a) If we are, for example, comparing the percentage of first-graders who can tie their own shoes to the percentage of second-graders who can tie their own shoes, grade-level is treated as categorical. It is just a way to group the students. We would use the same methods if we were comparing boys to girls or brown-eyed kids to blue-eyed kids.

   b) If we were studying the relationship between grade-level and height, we would be treating grade level as quantitative.

6. **ZIP codes.**
   a) ZIP codes are categorical in the sense that they correspond to a location. The ZIP code 14850 is a standardized way of referring to Ithaca, NY.
   b) ZIP codes generally increase as the location gets further from the east coast of the United States. For example, one of the ZIP codes for the city of Boston, MA is 02101. Kansas City, MO has a ZIP code of 64101, and Seattle, WA has a ZIP code of 98101.
7. **Voters.** The response is a categorical variable.
8. **Job hunting.** The answer is a categorical variable.
9. **Medicine.** The company is studying a quantitative variable.
10. **Stress.** The researcher is studying a quantitative variable.

Chapter Exercises

11. **The News.** Answers will vary.
12. **The Internet.** Answers will vary.
13. **Gaydar.** *Who* – 40 undergraduate women. *What* – Whether or not the women could identify the sexual orientation of men based on a picture. *Population of interest* – All women.
14. **Hula-hoops.** *Who* – An unknown number of participants. *What* – Heart rate, oxygen consumption, and rating of perceived exertion. *Population of interest* – All people.
15. **Bicycle Safety.** *Who* – 2,500 cars. *What* – Distance from the bicycle to the passing car (in inches). *Population of interest* – All cars passing bicyclists.
16. **Investments.** *Who* – 30 similar companies. *What* – 401(k) employee participation rates (in percent). *Population of interest* – All similar companies.
17. **Honesty.** *Who* – Workers who buy coffee in an office. *What* – amount of money contributed to the collection tray. *Population of interest* – All people in honor system payment situations.
18. **Blindness.** *Who* – 24 patients. *What* – Whether the patient had Stargardt's disease or dry age-related macular degeneration, and whether or not the stem cell therapy was effective in treating the condition. *Population of interest* – All people with these eye conditions.
19. **Not-so-diet soda.** *Who* – 474 participants. *What* – whether or not the participant drank two or more diet sodas per day, waist size at the beginning of the study, and waist size at the end of the study. *Population of interest* – All people.

20. Molten iron. *Who* – 10 crankshafts at Cleveland Casting. *What* – The pouring temperature (in degrees Fahrenheit) of molten iron. *Population of interest* – All crankshafts at Cleveland Casting.

21. Weighing bears. *Who* – 54 bears. *What* – Weight, neck size, length (no specified units), and sex. *When* – Not specified. *Where* – Not specified. *Why* - Since bears are difficult to weigh, the researchers hope to use the relationships between weight, neck size, length, and sex of bears to estimate the weight of bears, given the other, more observable features of the bear.
How – Researchers collected data on 54 bears they were able to catch. *Variables* – There are 4 variables; weight, neck size, and length are quantitative variables, and sex is a categorical variable. No units are specified for the quantitative variables. *Concerns* – The researchers are (obviously!) only able to collect data from bears they were able to catch. This method is a good one, as long as the researchers believe the bears caught are representative of all bears, in regard to the relationships between weight, neck size, length, and sex.

22. Schools. *Who* – Students. *What* – Age (probably in years, though perhaps in years and months), race or ethnicity, number of absences, grade level, reading score, math score, and disabilities/special needs. *When* – This information must be kept current. *Where* – Not specified. *Why* – Keeping this information is a state requirement. *How* – The information is collected and stored as part of school records. *Variables* – There are seven variables. Race or ethnicity, grade level, and disabilities/special needs are categorical variables. Number of absences, age, reading test score, and math test score are quantitative variables. *Concerns* – What tests are used to measure reading and math ability, and what are the units of measure for the tests?

23. Arby's menu. *Who* – Arby's sandwiches. *What* – type of meat, number of calories (in calories), and serving size (in ounces). *When* – Not specified. *Where* – Arby's restaurants. *Why* – These data might be used to assess the nutritional value of the different sandwiches. *How* – Information was gathered from each of the sandwiches on the menu at Arby's, resulting in a census. *Variables* – There are three variables. Number of calories and serving size are quantitative variables, and type of meat is a categorical variable.

24. Age and party. *Who* – 1180 Americans. *What* – Region, age (in years), political affiliation, and whether or not the person voted in the 2006 midterm Congressional election. *When* – First quarter of 2007. *Where* – United States. *Why* – The information was gathered for presentation in a Gallup public opinion poll. *How* – Phone Survey. *Variables* – There are four variables. Region, political affiliation, and whether or not the person voted in 1998 are categorical variables, and age is a quantitative variable.

25. Babies. *Who* – 882 births. *What* – Mother's age (in years), length of pregnancy (in weeks), type of birth (caesarean, induced, or natural), level of prenatal care (none, minimal, or adequate), birth weight of baby (unit of measurement not specified, but probably pounds and ounces), gender of baby (male or female), and baby's health problems (none, minor, major).
When – 1998-2000. *Where* – Large city hospital. *Why* – Researchers were investigating the impact of prenatal care on newborn health. *How* – It appears that they kept track of all births in the form of hospital records, although it is not specifically stated. *Variables* – There are three quantitative variables: mother's age, length of pregnancy, and birth weight of baby. There are four categorical variables: type of birth, level of prenatal care, gender of baby, and baby's health problems.

26. Flowers. *Who* – 385 species of flowers. *What* – Date of first flowering (in days). *When* – Not specified. *Where* – Southern England. *Why* – The researchers believe that this indicates a warming of the overall climate. *How* – Not specified. *Variables* – Date of first flowering is a quantitative variable. *Concerns* - Hopefully, date of first flowering was measured in days from January 1, or some other convention, to avoid problems with leap years.

27. Herbal medicine. *Who* – experiment volunteers. *What* – herbal cold remedy or sugar solution, and cold severity. *When* – Not specified. *Where* – Major pharmaceutical firm. *Why* – Scientists were testing the efficacy of an herbal compound on the severity of the common cold.
How – The scientists set up a controlled experiment. *Variables* – There are two variables. Type of treatment (herbal or sugar solution) is categorical, and severity rating is quantitative. *Concerns* – The severity of a cold seems subjective and difficult to quantify. Also, the scientists may feel pressure to report negative findings about the herbal product.

28. Vineyards. *Who* – American Vineyards. *What* – Size of vineyard (in acres), number of years in existence, state, varieties of grapes grown, average case price (in dollars), gross sales (probably in dollars), and percent profit. *When* – Not specified. *Where* – United States. *Why* – Business analysts hoped to provide information that would be helpful to producers of American wines. *How* – Not specified. *Variables* – There are five quantitative variables and two categorical variables. Size of vineyard, number of years in existence, average case price, gross sales, and percent profit are quantitative variables. State and variety of grapes grown are categorical variables.

29. Streams. *Who* – Streams. *What* – Name of stream, substrate of the stream (limestone, shale, or mixed), acidity of the water (measured in pH), temperature (in degrees Celsius), and BCI (unknown units). *When* – Not specified. *Where* – Upstate New York. *Why* – Research was conducted for an Ecology class. *How* – Not specified. *Variables* – There are five variables. Name and substrate of the stream are categorical variables, and acidity, temperature, and BCI are quantitative variables.

30. Fuel economy. *Who* – Every model of automobile in the United States. *What* – Vehicle manufacturer, vehicle type, weight (probably in pounds), horsepower (in horsepower), and gas mileage (in miles per gallon) for city and highway driving. *When* – This information is collected currently. *Where* – United States. *Why* – The Environmental Protection Agency uses the information to track fuel economy of vehicles. *How* – The data is collected from the manufacturer of each model. *Variables* – There are six variables. City mileage, highway mileage, weight, and horsepower are quantitative variables. Manufacturer and type of car are categorical variables.

31. Refrigerators. *Who* – 353 refrigerators. *What* – Brand, cost (probably in dollars), size (in cu. ft.), type, estimated annual energy cost (probably in dollars), overall rating, and repair history (in percent requiring repair over the past five years). *When* – 2013. *Where* – United States.
Why – The information was compiled to provide information to the readers of *Consumer Reports. How* – Not specified. *Variables* – There are 7 variables. Brand, type, and overall rating are categorical variables. Cost, size, estimated energy cost, and repair history are quantitative variables.

32. Walking in circles. *Who* – 32 volunteers. *What* – Sex, height, handedness, the number of yards walked before going out of bounds, and the side of the field on which the person walked out of bounds. *When* – Not specified. *Where* – Not specified. *Why* – The researcher was interested in whether people walk in circles when lost. *How* – Data were collected by observing the people on the field, as well as by measuring and asking the participants. *Variables* – There are 5 variables. Sex, handedness, and side of the field are categorical variables. Height and number of yards walked are quantitative variables.

33. Kentucky Derby 2014. *Who* – Kentucky Derby races. *What* – Year, winner, jockey, trainer, owner, and time (in minutes, seconds, and hundredths of a second. *When* – 1875 – 2013. *Where* – Churchill Downs, Louisville, Kentucky. *Why* – It is interesting to examine the trends in the Kentucky Derby. *How* – Official statistics are kept for the race each year. *Variables* – There are 6 variables. Winner, jockey, trainer and owner are categorical variables. Date and duration are quantitative variables.

34. Indianapolis 500 . *Who* – Indy 500 races. *What* – Year, driver, time (in minutes, seconds, and hundredths of a second), and speed (in miles per hour). *When* – 1911 – 2013.
Where – Indianapolis, Indiana. *Why* – It is interesting to examine the trends in Indy 500 races.
How – Official statistics are kept for the race every year. *Variables* – There are 4 variables. Driver is a categorical variable. Year, time, and speed are quantitative variables.

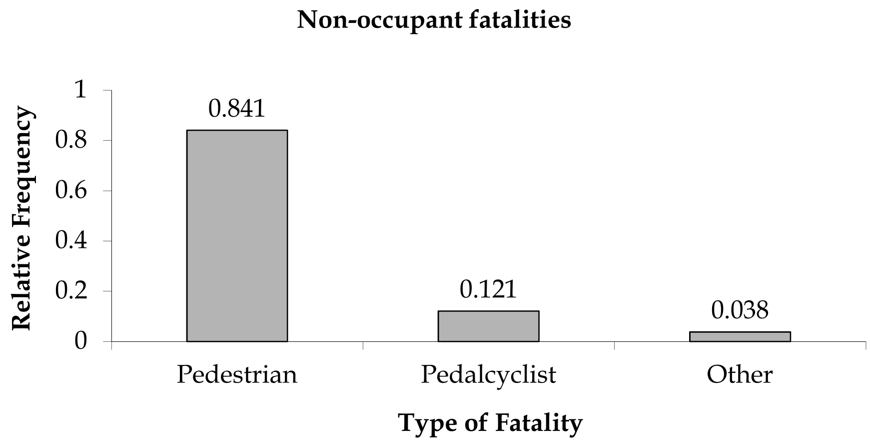
Chapter 2 – Displaying and Describing Categorical Data

Section 2.1

1. Automobile fatalities.

Subcompact and Mini	0.1128
Compact	0.3163
Intermediate	0.3380
Full	0.2193
Unknown	0.0137

2. Non-occupant fatalities.



3. Movie genres.

a) 2008 **b)** 1996 **c)** 2006 **d)** 2012

4. Marriage in decline.

a) People Living Together Without Being Married (ii)

b) Gay/Lesbian Couples Raising Children (iv)

c) Unmarried Couples Raising Children (iii)

d) Single Women Having Children (i)

Section 2.2

5. Movies again.

a) 170/348 ≈ 48.9% of these films were rated R.

b) 41/348 ≈ 11.8% of these films were R-rated comedies.

c) 41/170 ≈ 24.1% of the R-rated films were comedies.

d) 41/90 ≈ 45.6% of the comedies were R-rated.

6. Labor force.

a) 14,824/237,828 ≈ 6.2% of the population was unemployed.

b) 8858/237,828 ≈ 3.7% of the population was unemployed and between 25 and 54.

c) 12,699/21,047 ≈ 60.3% of those 20 to 24 years old were employed.

d) 4378/139,063 ≈ 3.1% of employed people were between 16 and 19.

Chapter Exercises

7. Graphs in the news. Answers will vary.

8. Graphs in the news II. Answers will vary.

9. Tables in the news. Answers will vary.

10. Tables in the news II. Answers will vary.

11. Movie genres.

a) A pie chart seems appropriate from the movie genre data. Each movie has only one genre, and the 193 movies constitute a "whole".

b) "Other" is the least common genre. It has the smallest region in the chart.

12. Movie ratings.

a) A pie chart seems appropriate for the movie rating data. Each movie has only one rating, and the 20 movies constitute a "whole". The percentages of each rating are different enough that the pie chart is easy to read.

b) The most common rating is PG-13. It has the largest region on the chart.

13. Genres, again.

a) SciFi/Fantasy has a higher bar than Action/Adventure, so it is the more common genre.

b) This is easier to see on the bar chart. The percentages are so close that the difference is nearly indistinguishable in the pie chart.

14. Ratings, again.

a) The least common rating was G. It has the shortest bar.

b) The bar chart does not support this claim. These data are for a single year only. We have no idea if the percentages of G and PG-13 movies changed from year to year.

15. Magnet Schools.

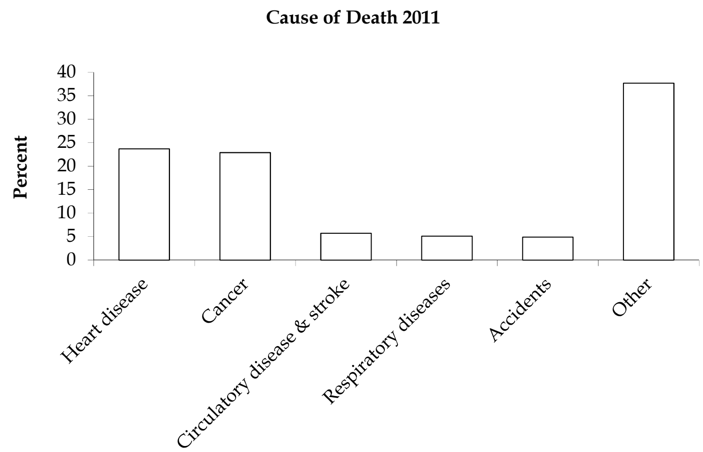
There were 1755 qualified applicants for the Houston Independent School District's magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away for lack of space.

16. Magnet schools again.

There were 1755 qualified applicants for the Houston Independent School District's magnet schools program. 29.5% were Black or Hispanic, 16.6% were Asian, and 53.9% were white.

17. Causes of death 2011.

a) Yes, it is reasonable to assume that heart and respiratory disease caused approximately 29.4% of U.S. deaths in 2007, since there is no possibility for overlap. Each person could only have one cause of death.

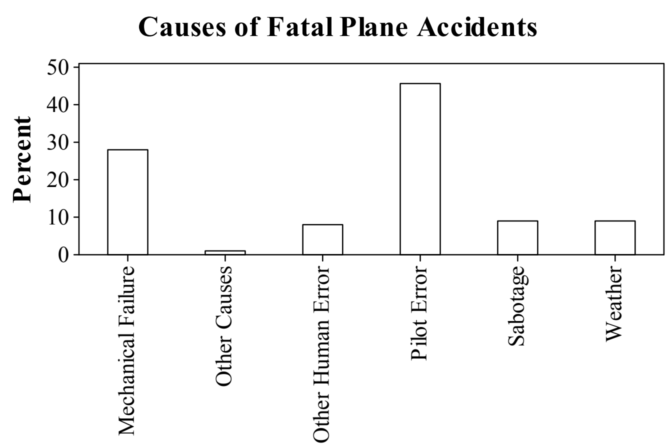


b) Since the percentages listed add up to 62.3%, other causes must account for 37.7% of US deaths.

c) A bar chart is a good choice (with the inclusion of the "Other" category). Since causes of US deaths represent parts of a whole, a pie chart would also be a good display.

18. Plane crashes.

a) As long as each plane crash had only one cause, it would be reasonable to assume that weather or mechanical failures were the causes of about 37% of crashes.



b) It is likely that the numbers in the table add up to 101% due to rounding.

c) A relative frequency bar chart is a good choice. A pie chart would also be a good display, as long as each plane crash has only one cause.

19. Oil spills as of 2013.

a) Grounding, accounting for approximately 150 spills, is the most frequent cause of oil spillage for these 459 spills. A substantial number of spills, approximately 140, were caused by Collision. Less prevalent causes of oil spillage in descending order of frequency were Hull or equipment failures, Fire & Explosions, and Other/Unknown causes.

b) A pie chart is an appropriate display of the data, since there is only a single cause attributed to each spill, and all spills are represented in some category.

c) There were more spills due to Grounding than Collisions. This is much easier to see on the bar chart.

20. Winter Olympics 2010.

a) There are too many categories to construct an appropriate display. In a bar chart, there are too many bars. In a pie chart, there are too many slices. In each case, we run into difficulty trying to display those countries that didn't win many medals.

b) Perhaps we are primarily interested in countries that won many medals. We might choose to combine all countries that won fewer than 6 medals into a single category. This will make our chart easier to read. We are probably interested in number of medals won, rather than percentage of total medals won, so we'll use a bar chart. A bar chart is also better for comparisons.

21. Global warming.

Perhaps the most obvious error is that the percentages in the pie chart only add up to 93%, when they should, of course, add up to 100%. Furthermore, the three-dimensional perspective view distorts the regions in the graph, violating the area principle. The regions corresponding to No Solid Evidence and Due to Human Activity should be roughly the same size, at 32% and 34% of respondents, respectively. However, the angle for the 32% region looks much bigger. Always use simple, two-dimensional graphs. Additionally, the graph does not include a title.

22. Modalities.

a) The bars have false depth, which can be misleading. This is a bar chart, so the bars should have space between them. Running the labels on the bars from top to bottom and the vertical axis labels from bottom to top is confusing.

b) The percentages sum to 100%. Normally, we would take this as a sign that all of the observations had been correctly accounted for. But in this case, it is extremely unlikely. Each of the respondents was asked to list *three* modalities. For example, it would be possible for 80% of respondents to say they use ice to treat an injury, and 75% to use electric stimulation. The fact that the percentages total greater than 100% is not odd. In fact, in this case, it seems wrong that the percentages add up to 100%, rather than correct.

23. Teen smokers.

According to the Monitoring the Future study, teen smoking brand preferences differ somewhat by region. Although Marlboro is the most popular brand in each region, with about 58% of teen smokers preferring this brand in each region, teen smokers from the South prefer Newports at a higher percentage than teen smokers from the West, 22.5% to approximately 10%, respectively. Camels are more popular in the West, with 9.5% of teen smokers preferring this brand, compared to only 3.3% in the South. Teen smokers in the West are also more likely to have no particular brand than teen smokers in the South. 12.9% of teen smokers in the West have no particular brand, compared to only 6.7% in the South. Both regions have about 9% of teen smokers that prefer one of over 20 other brands.

24. Handguns.

76.4% of handguns involved in Milwaukee buyback programs are small caliber, while only 20.3% of homicides are committed with small caliber handguns. Along the same lines, only 19.3% of buyback handguns are of medium caliber, while 54.7% of homicides involve medium caliber handguns. A similar disparity is seen in large caliber handguns. Only 2.1% of buyback handguns are large caliber, but this caliber is used in 10.8% of homicides. Finally, 2.2% of buyback handguns are of other calibers, while 14.2% of homicides are committed with handguns of other calibers. Generally, the handguns that are involved in buyback programs are not the same caliber as handguns used in homicides in Milwaukee.

25. Movies by genre and rating.

a) The table uses column percents, since each column adds to 100%, while the rows do not.

b) 25.86% of these movies are comedies.

c) 28.57% of the PG-rated movies were comedies.

d) **i)** 27.36% of the PG-13 movies were comedies.

ii) You cannot determine this from the table.

iii) None (0%) of the dramas were G-rated.

iv) You cannot determine this from the table.

26. The last picture show.

a) Since neither the columns nor the rows total 100%, but the table itself totals 100%, these are table percentages.

b) The most common genre/rating combination was the R-rated drama. 18.68% of the 348 movies had this combination.

c) 5.17% of the 348 movies, or 18 movies, were PG-rated comedies.

d) A total of 2.59% of the 348 movies, or 9 movies, were rated G.

e) 2.59% of the movies were rated G, and 18.10% of them were rated PG. So patrons under 13 can see only 20.69% of these movies. This supports the assertion that approximately three-quarters of movies can only be seen by patrons 13 years old or older.

27. Seniors.

a) A table with marginal totals is to the right. There are 268 White graduates and 325 total graduates. 268/325 ≈ 82.5% of the graduates are white.

Plans	White	Minority	TOTAL
4-year college	198	44	242
2-year college	36	6	42
Military	4	1	5
Employment	14	3	17
Other	16	3	19
TOTAL	268	57	325

b) There are 42 graduates planning to attend 2-year colleges. 42/325 ≈ 12.9%

c) 36 white graduates are planning to attend 2-year colleges. 36/325 ≈ 11.1%

d) 36 white graduates are planning to attend 2-year colleges and there are 268 whites graduates. 36/268 ≈ 13.4%

e) There are 42 graduates planning to attend 2-year colleges, and 36 of them are white.
36/42 ≈ 85.7%

28. Politics.

a) There are 192 students taking Intro Stats. Of those, 115, or about 59.9%, are male.

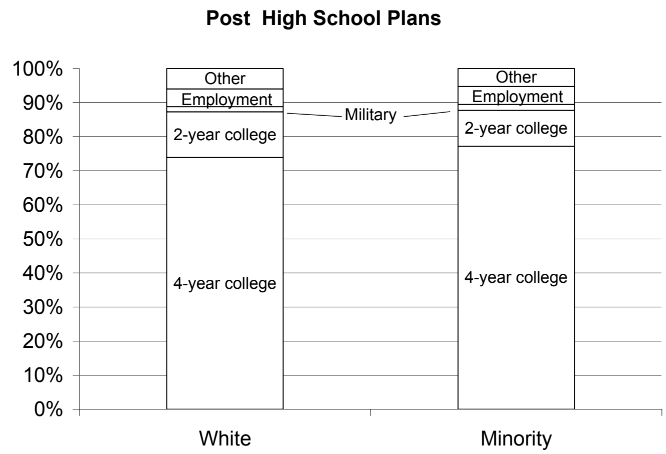
b) There are 192 students taking Intro Stats. Of those, 27, or about 14.1%, consider themselves to be "Conservative".

c) There are 115 males taking Intro Stats. Of those, 21, or about 18.3%, consider themselves to be "Conservative".

d) There are 192 students taking Intro Stats. Of those, 21, or about 10.9%, are males who consider themselves to be "Conservative".

29. More about seniors.

a) For white students, 73.9% plan to attend a 4-year college, 13.4% plan to attend a 2-year college, 1.5% plan on the military, 5.2% plan to be employed, and 6.0% have other plans.



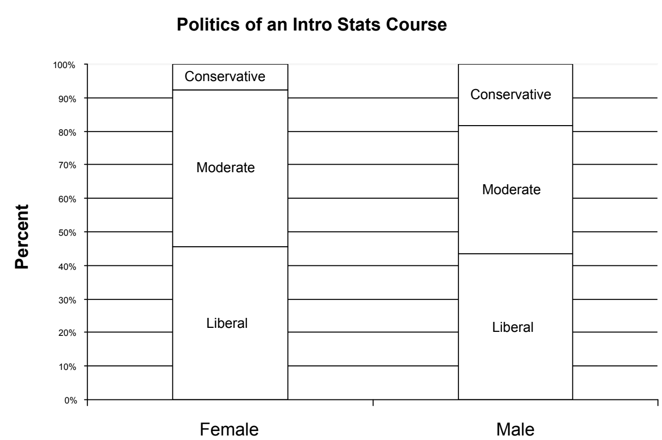
b) For minority students, 77.2% plan to attend a 4-year college, 10.5% plan to attend a 2-year college, 1.8% plan on the military, 5.3% plan to be employed, and 5.3% have other plans.

c) A segmented bar chart is a good display of these data.

d) The conditional distributions of plans for Whites and Minorities are similar:
White – 74% 4-year college, 13% 2-year college, 2% military, 5% employment, 6% other.
Minority – 77% 4-year college, 11% 2-year college, 2% military, 5% employment, 5% other.
Caution should be used with the percentages for Minority graduates, because the total is so small. Each graduate is almost 2%. Still, the conditional distributions of plans are essentially the same for the two groups. There is little evidence of an association between race and plans for after graduation.

30. Politics revisited.

a) The females in this course were 45.5% Liberal, 46.8% Moderate, and 7.8% Conservative.



b) The males in this course were 43.5% Liberal, 38.3% Moderate, and 18.3% Conservative.

c) A segmented bar chart comparing the distributions is at the right.

d) Politics and sex do not appear to be independent in this course. Although the percentage of liberals was roughly the same for each sex, females had a greater percentage of moderates and a lower percentage of conservatives than males.

31. Magnet schools revisited.

a) There were 1755 qualified applicants to the Houston Independent School District's magnet schools program. Of those, 292, or about 16.6% were Asian.

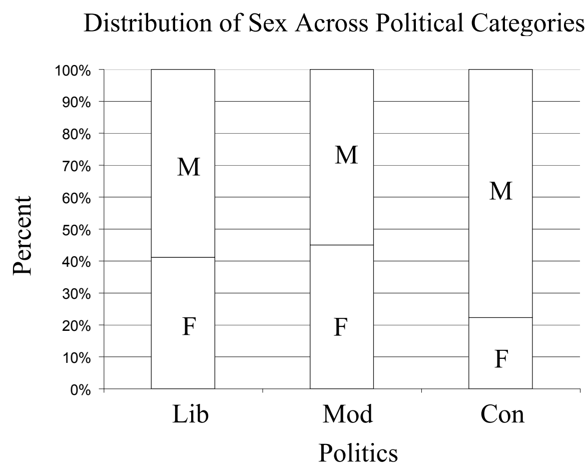
b) There were 931 students accepted to the magnet schools program. Of those, 110, or about 11.8% were Asian.

c) There were 292 Asian applicants. Of those, 110, or about 37.7%, were accepted.

d) There were 1755 total applicants. Of those, 931, or about 53%, were accepted.

32. More politics.

a)



b) The percentage of males and females varies across political categories. The percentage of self-identified Liberals and Moderates who are female is about twice the percentage of Conservatives who are female. This suggests that *sex* and *politics* are not independent.

33. Back to school.

There were 1,755 qualified applicants for admission to the magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away. While the overall acceptance rate was 53%, 93.8% of Blacks and Hispanics were accepted, compared to only 37.7% of Asians, and 35.5% of whites. Overall, 29.5% of applicants were Black or Hispanics, but only 6% of those turned away were Black or Hispanic. Asians accounted for 16.6% of applicants, but 25.3% of those turned away. It appears that the admissions decisions were not independent of the applicant's ethnicity.

34. Parking lots.

a) In order to get percentages, first we need totals. Here is the same table, with row and column totals. Foreign cars are defined as non-American. There are 45+102=147 non-American cars or 147/359 ≈ 40.95%.

	Driver		
Origin	Student	Staff	Total
American	107	105	212
European	33	12	45
Asian	55	47	102
Total	195	164	359

b) There are 212 American cars of which 107 or 107/212 ≈ 50.47% were owned by students.

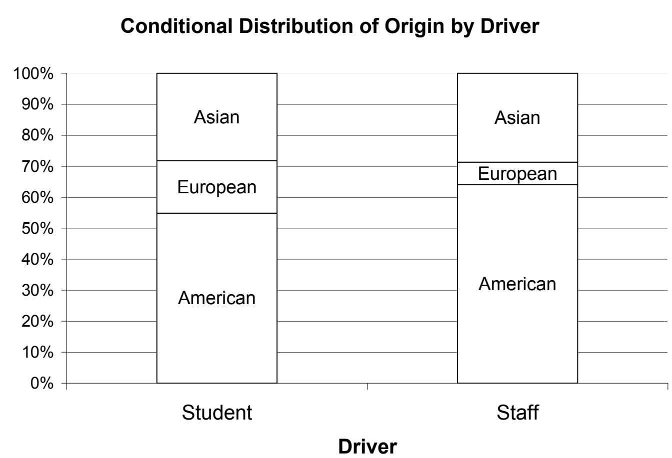
c) There are 195 students of whom 107 or 107/195 ≈ 54.87% owned American cars.

d) The marginal distribution of Origin is displayed in the third column of the table at the right: 59% American, 13% European, and 28% Asian.

Origin	Totals
American	212 (59%)
European	45 (13%)
Asian	102 (28%)
Total	359

e) The conditional distribution of Origin for Students is: 55% (107 of 195) American, 17% (33 of 195) European, and 28% (55 of 195) Asian.
The conditional distribution of Origin for Staff is:
64.0% (105 of 164) American, 7.3% (12 of 164) European, and 28.7% (47 of 164) Asian.

f) The percentages in the conditional distributions of Origin by Driver (students and staff) seem slightly different. Let's look at a segmented bar chart of Origin by Driver, to compare the conditional distributions graphically.



The conditional distributions of Origin by Driver have similarities and differences. Although students appear to own a higher percentage of European cars and a smaller percentage of American cars than the staff, the two groups own nearly the same percentage of Asian cars. However, because of the differences, there is evidence of an association between Driver and Origin of the car.

35. Weather forecasts.

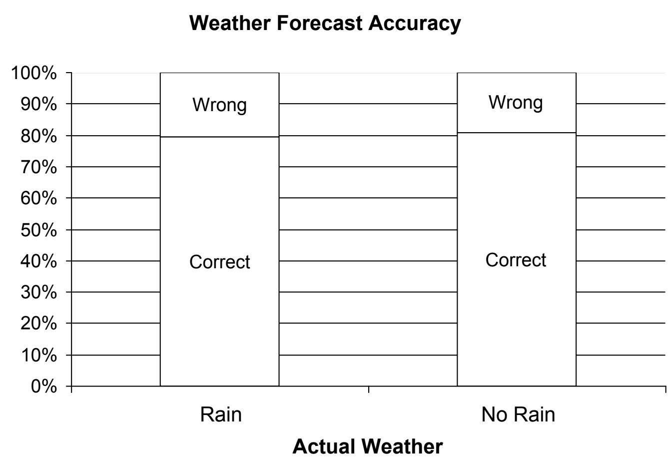
a) The table shows the marginal totals. It rained on 34 of 365 days, or 9.3% of the days.

		Actual Weather		
		Rain	No Rain	**Total**
Forecast	Rain	27	63	90
	No Rain	7	268	275
	Total	34	331	365

b) Rain was predicted on 90 of 365 days. 90/365 ≈ 24.7% of the days.

c) The forecast of Rain was correct on 27 of the days it actually rained and the forecast of No Rain was correct on 268 of the days it didn't rain. So, the forecast was correct a total of 295 times. 295/365 ≈ 80.8% of the days.

d) On rainy days, rain had been predicted 27 out of 34 times (79.4%). On days when it did not rain, forecasters were correct in their predictions 268 out of 331 times (81.0%). These two percentages are very close. There is no evidence of an association between the type of weather and the ability of the forecasters to make an accurate prediction.



36. Twin births.

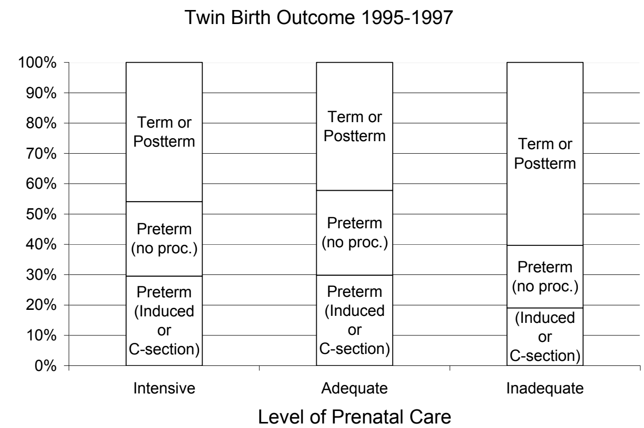
a) Of the 278,000 mothers who had twins in 1995-1997, 63,000 had inadequate health care during their pregnancies. 63,000/278,000 = 22.7%

Twin Births 1995-97 (in thousands)				
Level of Prenatal Care	**Preterm (Induced or Caesarean)**	**Preterm (without procedures)**	**Term or Postterm**	**Total**
Intensive	18	15	28	61
Adequate	46	43	65	154
Inadequate	12	13	38	63
Total	76	71	131	278

b) There were 76,000 induced or Caesarean births and 71,000 preterm births without these procedures. (76,000 + 71,000)/278,000 = 52.9%

c) Among the mothers who did not receive adequate medical care, there were 12,000 induced or Caesarean births and 13,000 preterm births without these procedures. 63,000 mothers of twins did not receive adequate medical care. (12,000 + 13,000)/63,000 = 39.7%

d)



e) 52.9% of all twin births were preterm, while only 39.7% of births in which inadequate medical care was received were preterm. This is evidence of an association between level of prenatal care and twin birth outcome. If these variables were independent, we would expect the percentages to be roughly the same. Generally, those mothers who received adequate medical care were more likely to have preterm births than mothers who received intensive medical care, who were in turn more likely to have preterm births than mothers who received inadequate health care. This does *not* imply that mothers should receive inadequate health care do decrease their chances of having a preterm birth, since it is likely that women that have some complication *during* their pregnancy (that might lead to a preterm birth), would seek intensive or adequate prenatal care.

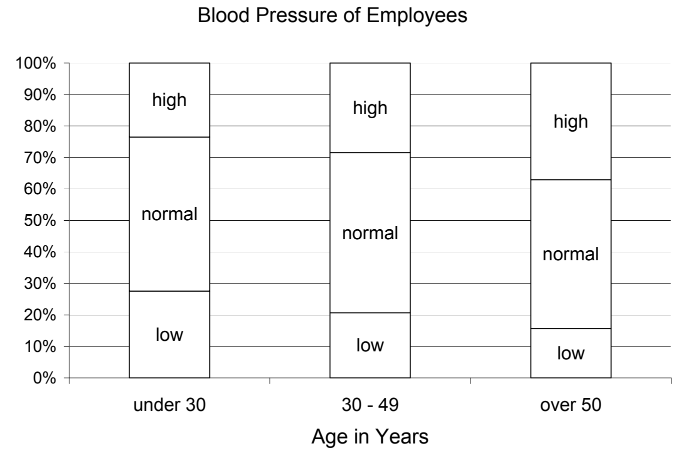
37. Blood pressure.

Blood pressure	under 30	30 - 49	over 50	Total
low	27	37	31	95
normal	48	91	93	232
high	23	51	73	147
Total	98	179	197	474

a) The marginal distribution of blood pressure for the employees of the company is the total column of the table, converted to percentages. 20% low, 49% normal and 31% high blood pressure.

b) The conditional distribution of blood pressure within each age category is:
Under 30 : 28% low, 49% normal, 23% high
30 – 49 : 21% low, 51% normal, 28% high
Over 50 : 16% low, 47% normal, 37% high

c) A segmented bar chart of the conditional distributions of blood pressure by age category is at the right.

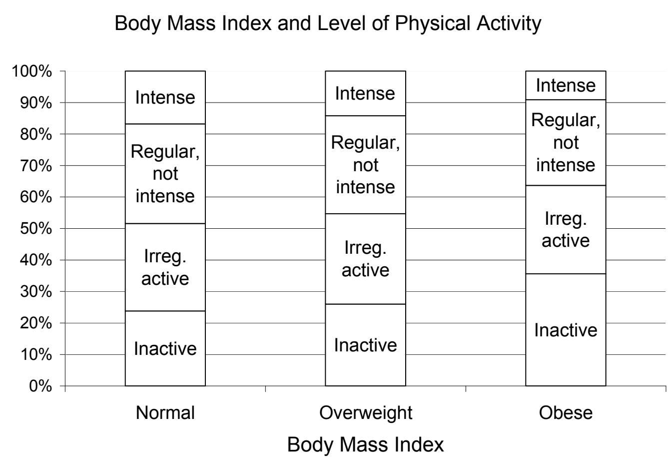


d) In this company, as age increases, the percentage of employees with low blood pressure decreases, and the percentage of employees with high blood pressure increases.

e) No, this does not prove that people's blood pressure increases as they age. Generally, an association between two variables does not imply a cause-and-effect relationship. Specifically, these data come from only one company and cannot be applied to all people. Furthermore, there may be some other variable that is linked to both age and blood pressure. Only a controlled experiment can isolate the relationship between age and blood pressure.

38. Obesity and exercise.

a) Participants were categorized as Normal, Overweight or Obese, according to their Body Mass Index. Within each classification of BMI (column), participants self reported exercise levels. Therefore, these are column percentages. The percentages sum to 100% in each column, *not* across each row.



b) A segmented bar chart of the conditional distributions of level of physical activity by Body Mass Index category is at the right.

c) No, even though the graphical displays provide strong evidence that lack of exercise and BMI are not independent. All three BMI categories have nearly the same percentage of subjects who report "Regular, not intense" or "Irregularly active", but as we move from Normal to Overweight to Obese we see a decrease in the percentage of subjects who report "Regular, intense" physical activity (16.8% to 14.2% to 9.1%), while the percentage of subjects who report themselves as "Inactive" increases. While it may seem logical that lack of exercise causes obesity, association between variables does not imply a cause-and-effect relationship. A lurking variable (for example, overall health) might influence both BMI and level of physical activity, or perhaps lack of exercise is *caused by* obesity. Only a controlled experiment could isolate the relationship between BMI and level of physically activity.

39. Anorexia.

These data provide no evidence that Prozac might be helpful in treating anorexia. About 71% of the patients who took Prozac were diagnosed as "Healthy", while about 73% of the patients who took a placebo were diagnosed as "Healthy". Even though the percentage was higher for the placebo patients, this does not mean that Prozac is hurting patients. The difference between 71% and 73% is not likely to be statistically significant.

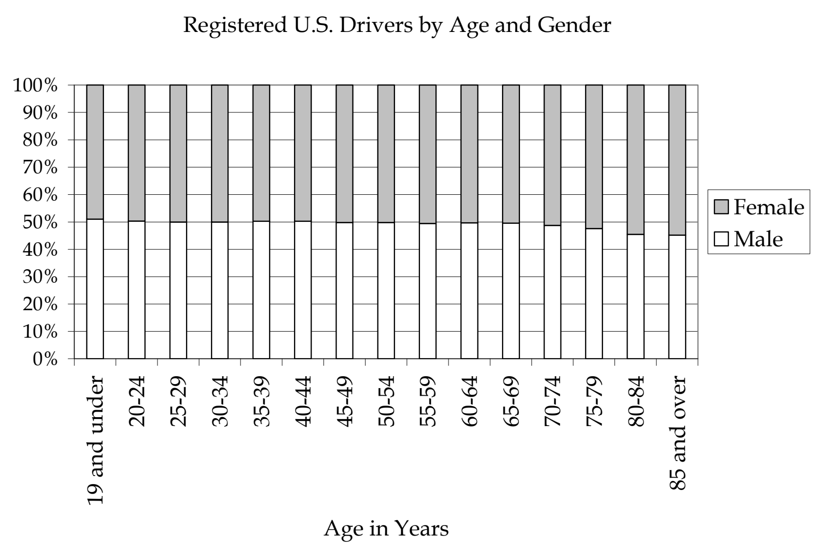
40. Antidepressants and bone fractures.

These data provide evidence that taking a certain class of antidepressants (SSRI) might be associated with a greater risk of bone fractures. Approximately 10% of the patients taking this class of antidepressants experience bone fractures. This is compared to only approximately 5% in the group that were not taking the antidepressants.

41. Driver's licenses 2011.

a) There are 10.0 million drivers under 20 and a total of 208.3 million drivers in the U.S. That's about 4.8% of U.S. drivers under 20.

b) There are 103.5 million males out of 208.4 million total U.S. drivers, or about 49.7%.

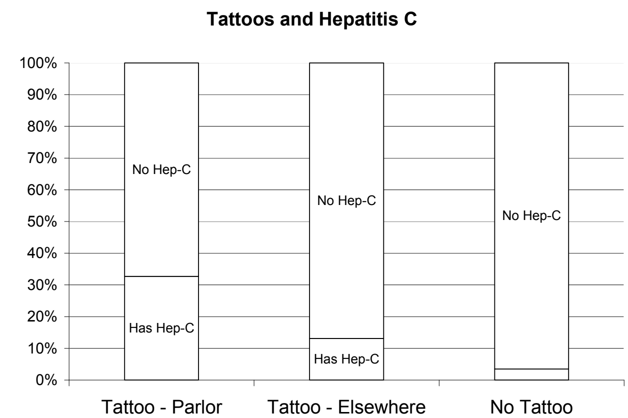


c) Each age category appears to have about 50% male and 50% female drivers. The segmented bar chart shows a pattern in the deviations from 50%. At younger ages, males form the slight majority of drivers. This percentage shrinks until the percentages are 50% male and 50% for middle aged drivers. The percentage of male drivers continues to shrink until, at around age 45, female drivers hold a slight majority. This continues into the 85 and over category.

d) There appears to be a slight association between age and gender of U.S. drivers. Younger drivers are slightly more likely to be male, and older drivers are slightly more likely to be female.

42. Tattoos.

The study by the University of Texas Southwestern Medical Center provides evidence of an association between having a tattoo and contracting hepatitis C. Around 33% of the subjects who were tattooed in a commercial parlor had hepatitis C, compared with 13% of those tattooed elsewhere, and only 3.5% of those with no tattoo. If having a tattoo and having hepatitis C were independent, we would have expected these percentages to be roughly the same.



43. Hospitals.

a) The marginal totals have been added to the table:

Procedure		Discharge delayed: Large Hospital	Small Hospital	Total
	Major surgery	120 of 800	10 of 50	130 of 850
	Minor surgery	10 of 200	20 of 250	30 of 450
	Total	130 of 1000	30 of 300	160 of 1300

160 of 1300, or about 12.3% of the patients had a delayed discharge.

b) Yes. Major surgery patients were delayed 130 of 850 times, or about 15.3% of the time.
Minor Surgery patients were delayed 30 of 450 times, or about 6.7% of the time.

c) Large Hospital had a delay rate of 130 of 1000, or 13%.
Small Hospital had a delay rate of 30 of 300, or 10%.
The small hospital has the lower overall rate of delayed discharge.

d) Large Hospital: Major Surgery 15% delayed and Minor Surgery 5% delayed.
Small Hospital: Major Surgery 20% delayed and Minor Surgery 8% delayed.
Even though small hospital had the lower overall rate of delayed discharge, the large hospital had a lower rate of delayed discharge for each type of surgery.

e) No. While the overall rate of delayed discharge is lower for the small hospital, the large hospital did better with *both* major surgery and minor surgery.

f) The small hospital performs a higher percentage of minor surgeries than major surgeries. 250 of 300 surgeries at the small hospital were minor (83%). Only 200 of the large hospital's 1000 surgeries were minor (20%). Minor surgery had a lower delay rate than major surgery (6.7% to 15.3%), so the small hospital's overall rate was artificially inflated. Simply put, it is a mistake to look at the overall percentages. The real truth is found by looking at the rates after the information is broken down by type of surgery, since the delay rates for each type of surgery are so different. The larger hospital is the better hospital when comparing discharge delay rates.

44. Delivery service.

a) Pack Rats has delivered a total of 28 late packages (12 Regular + 16 Overnight), out of a total of 500 deliveries (400 Regular + 100 Overnight). 28/500 = 5.6% of the packages are late. Boxes R Us has delivered a total of 30 late packages (2 Regular + 28 Overnight) out of a total of 500 deliveries (100 Regular + 400 Overnight). 30/500 = 6% of the packages are late.

b) The company should have hired Boxes R Us instead of Pack Rats. Boxes R Us only delivers 2% (2 out of 100) of its Regular packages late, compared to Pack Rats, who deliver 3% (12 out of 400) of its Regular packages late. Additionally, Boxes R Us only delivers 7% (28 out of 400) of its Overnight packages late, compared to Pack Rats, who delivers 16% of its Overnight packages late. Boxes R Us is better at delivering Regular and Overnight packages.

c) This is an instance of Simpson's Paradox, because the overall late delivery rates are unfair averages. Boxes R Us delivers a greater percentage of its packages Overnight, where it is comparatively harder to deliver on time. Pack Rats delivers many Regular packages, where it is easier to make an on-time delivery.

45. Graduate admissions.

a) 1284 applicants were admitted out of a total of 3014 applicants. 1284/3014 = 42.6%

Program	Males Accepted (of applicants)	Females Accepted (of applicants)	Total
1	511 of 825	89 of 108	600 of 933
2	352 of 560	17 of 25	369 of 585
3	137 of 407	132 of 375	269 of 782
4	22 of 373	24 of 341	46 of 714
Total	**1022 of 2165**	**262 of 849**	**1284 of 3014**

b) 1022 of 2165 (47.2%) of males were admitted. 262 of 849 (30.9%) of females were admitted.

c) Since there are four comparisons to make, the table at the right organizes the percentages of males and females accepted in each program. Females are accepted at a higher rate in every program.

Program	Males	Females
1	61.9%	82.4%
2	62.9%	68.0%
3	33.7%	35.2%
4	5.9%	7%

d) The comparison of acceptance rate within each program is most valid. The overall percentage is an unfair average. It fails to take the different numbers of applicants and different acceptance rates of each program. Women tended to apply to the programs in which gaining acceptance was difficult for everyone. This is an example of Simpson's Paradox.

46. Be a Simpson!

Answers will vary. The three-way table below shows one possibility. The number of local hires out of new hires is shown in each cell.

	Company A	**Company B**
Full-time New Employees	40 of 100 = 40%	90 of 200 = 45%
Part-time New Employees	170 of 200 = 85%	90 of 100 = 90%
Total	210 of 300 = 70%	180 of 300 = 60%

Chapter 3 – Displaying and Summarizing Quantitative Data

Section 3.1

1. Details.

Boxplots don't tell us much about the shape of a distribution, beyond a basic idea of symmetry or skewness. The given boxplot could be displaying a distribution that has multiple modes (the first histogram), is reasonably unimodal (the second histogram), has gaps and clusters (the third histogram), or has outliers (the fourth histogram). We simply can't determine shape from a boxplot.

2. Opposites.

a) The tallest bars on the histogram are where the vertical lines on the boxplot are closest together.

b) The boxplot indicates a skewed distribution when the vertical lines on one side of the median are closer together than the vertical lines on the other side. The histogram indicates a skewed distribution when the bars on one side of the histogram are generally taller than the bars on the other side of the histogram.

c) The histogram shows a second mode in the distribution, as well as clusters and gaps. The boxplot does not show this.

d) The boxplot shows the quartiles and the median, as well as showing outliers. These cannot be determined from the histogram.

Section 3.2

3. Outliers.

IQR = Q3 – Q1 = 116 – 98 = 18.
Using the Outlier Rule (1.5 IQRs beyond quartiles):

Upper Fence:
$$\begin{aligned} Q3+1.5(IQR) &= 116+1.5(18) \\ &= 116+27 \\ &= 143 \end{aligned}$$

Since the maximum, 160 minutes, is above the upper fence, there is at least one high outlier.

Lower Fence:
$$\begin{aligned} Q1-1.5(IQR) &= 98-1.5(18) \\ &= 98-27 \\ &= 71 \end{aligned}$$

Since the minimum, 43 minutes, is below the lower fence, there is at least one low outlier.

4. Adoptions.

Since the mean age at adoption is higher than the median age at adoption, the distribution of adoption ages is likely to be skewed to the right, with many adoptions happening when children are relatively young, with fewer adoptions of older children.

Section 3.3

5. Adoptions II.

The mean number of adoptions is expected to by higher than the median number of adoptions, since the distribution of the number of adoptions is skewed to the right.

6. Test score centers.

The median test score will not be affected, but the mean test score will increase by 0.4 points.

Section 3.4

7. Test score spreads.

The IQR of the test scores will not be affected, but the standard deviation of the test scores will increase.

8. Fuel economy.

If the outlier is removed, the standard deviation will decrease. The IQR will not change substantially. (It may change slightly, since removing one value from the data set may change the quartiles, which would change the IQR.)

Chapter Exercises

9. Histogram. Answers will vary.

10. Not a histogram. Answers will vary.

11. In the news. Answers will vary.

12. In the news II. Answers will vary.

13. Thinking about shape.

a) The distribution of the number of speeding tickets each student in the senior class of a college has ever had is likely to be unimodal and skewed to the right. Most students will have very few speeding tickets (maybe 0 or 1), but a small percentage of students will likely have comparatively many (3 or more?) tickets.

b) The distribution of player's scores at the U.S. Open Golf Tournament would most likely be unimodal and slightly skewed to the right. The best golf players in the game will likely have around the same average score, but some golfers might be off their game and score 15 strokes above the mean. (Remember that high scores are undesirable in the game of golf!)

c) The weights of female babies in a particular hospital over the course of a year will likely have a distribution that is unimodal and symmetric. Most newborns have about the same weight, with some babies weighing more and less than this average. There may be slight skew to the left, since there seems to be a greater likelihood of premature birth (and low birth weight) than post-term birth (and high birth weight).

d) The distribution of the length of the average hair on the heads of students in a large class would likely be bimodal and skewed to the right. The average hair length of the males would be at one mode, and the average hair length of the females would be at the other mode, since women typically have longer hair than men. The distribution would be skewed to the right, since it is not possible to have hair length less than zero, but it is possible to have a variety of lengths of longer hair.

14. More shapes.

a) The distribution of the ages of people at a Little League game would likely be bimodal and skewed to the right. The average age of the players would be at one mode and the average age of the spectators (probably mostly parents) would be at the other mode. The distribution would be skewed to the right, since it is possible to have a greater variety of ages among the older people, while there is a natural left endpoint to the distribution at zero years of age.

b) The distribution of the number of siblings of people in your class is likely to be unimodal and skewed to the right. Most people would have 0, 1, or 2 siblings, with some people having more siblings.

c) The distribution of pulse rate of college-age males would likely be unimodal and symmetric. Most males' pulse rates would be around the average pulse rate for college-age males, with some males having lower and higher pulse rates.

d) The distribution of the number of times each face of a die shows in 100 tosses would likely be uniform, with around 16 or 17 occurrences of each face (assuming the die had six sides).

15. Sugar in cereals.

a) The distribution of the sugar content of breakfast cereals is bimodal, with a cluster of cereals with sugar content around 6% sugar and another cluster of cereals around 48% sugar. The lower cluster shows a bit of skew to the right. Most cereals in the lower cluster have between 0% and 10% sugar. The upper cluster is symmetric, with center around 45% sugar.

b) There are two different types of breakfast cereals, those for children and those for adults. The children's cereals are likely to have higher sugar contents, to make them taste better (to kids, anyway!). Adult cereals often advertise low sugar content.

16. Singers.

a) The distribution of the heights of singers in the chorus is bimodal, with a mode at around 65 inches and another mode around 71 inches. No chorus member has height below 60 inches or above 76 inches.

b) The two modes probably represent the mean heights of the male and female members of the chorus.

17. Vineyards.

a) There is information displayed about 36 vineyards and it appears that 28 of the vineyards are smaller than 90 acres. That's around 78% of the vineyards. (75% would be a good estimate!)

b) The distribution of the size of 36 Finger Lakes vineyards is skewed to the right. Most vineyards are smaller than 90 acres, with a few larger ones, from 90 to 160 acres. One vineyard was larger than all the rest, over 240 acres. The mode of the distribution is between 0 and 30 acres.

18. Run times.

The distribution of runtimes is skewed to the right. The shortest runtime was around 28.5 minutes and the longest runtime was around 35.5 minutes. A typical run time was between 30 and 31 minutes, and the majority of runtimes were between 29 and 32 minutes. It is easier to run slightly slower than usual and end up with a longer runtime than it is to run slightly faster than usual and end up with a shorter runtime. This could account for the skew to the right seen in the distribution.

19. Heart attack stays.

a) The distribution of length of stays is skewed to the right, so the mean is larger than the median.

b) The distribution of the length of hospital stays of female heart attack patients is bimodal and skewed to the right, with stays ranging from 1 day to 36 days. The distribution is centered around 8 days, with the majority of the hospital stays lasting between 1 and 15 days. There are a relatively few hospital stays longer than 27 days. Many patients have a stay of only one day, possibly because the patient died.

c) The median and IQR would be used to summarize the distribution of hospital stays, since the distribution is strongly skewed.

20. Emails.

a) The distribution of the number of emails sent is skewed to the right, so the mean is larger than the median.

b) The distribution of the number of emails received from each student by a professor in a large introductory statistics class during an entire term is skewed to the right, with the number of emails ranging from 1 to 21 emails. The distribution is centered at about 2 emails, with many students only sending 1 email. There is one outlier in the distribution, a student who sent 21 emails. The next highest number of emails sent was only 8.

c) The median and IQR would be used to summarize the distribution of the number of emails received, since the distribution is strongly skewed.

21. Super Bowl points 2013.

a) The median number of points scored in the first 48 Super Bowl games is 45 points.

b) The first quartile of the number of points scored in the first 48 Super Bowl games is 35 points. The third quartile is 54.5 (or 55) points.

c) In the first 48 Super Bowl games, the lowest number of points scored was 21, and the highest number of points scored was 75. The median number of points scored was 45, and the middle 50% of Super Bowls has between 35 and 55 points scored.

22. Super Bowl wins 2013.

a) The median winning margin in the first 48 Super Bowl games is 12 points.

b) The first quartile of the winning margin in the first 48 Super Bowl games is 4.5 points. The third quartile is 19 points.

c) In the first 48 Super Bowl games the lowest winning margin was 1 point and the highest winning margin was 45 points, which was an outlier. The second highest winning margin was only 36 points. The median winning margin was 12 points, with the middle 50% of winning margins between 4.5 and 19 points.

23. Summaries.

a) The mean cost of the compact refrigerators is \$144.44.

b) The median cost of the compact refrigerators is \$150. The first quartile is \$130, and the third quartile is \$150.

c) The range of the cost of the compact refrigerators is \$180 – \$120 = \$60. The IQR is \$150 – \$130 = \$20.

24. Tornadoes 2013.

a) The mean number of annual deaths from tornadoes in the United States from 1998 through 2013 is 125.1.

b) The median number of deaths is 60.5. The first quartile is 40 deaths and the third quartile is 109.5 deaths.

c) The range is 555 – 21 = 534 deaths. The IQR is 109.5 – 40 = 69.5 deaths.

25. Mistake.

a) As long as the boss's true salary of $200,000 is still above the median, the median will be correct. The mean will be too large, since the total of all the salaries will decrease by $2,000,000 - $200,000 = $1,800,000, once the mistake is corrected.

b) The range will likely be too large. The boss's salary is probably the maximum, and a lower maximum would lead to a smaller range. The IQR will likely be unaffected, since the new maximum has no effect on the quartiles. The standard deviation will be too large, because the $2,000,000 salary will have a large squared deviation from the mean.

26. Sick days.

The company probably uses the mean, while the union uses the median number of sick days. The mean will likely be higher, since it is affected by probable right skew. Some employees may have many sick days, while most have relatively few.

27. Standard deviation I.

a) Set 2 has the greater standard deviation. Both sets have the same mean (6) but set two has values that are generally farther away from the mean.
SD(Set 1) = 2.24 SD(Set 2) = 3.16

b) Set 2 has the greater standard deviation. Both sets have the same mean (15), maximum (20), and minimum (10), but 11 and 19 are farther from the mean than 14 and 16.
SD(Set 1) = 3.61 SD(Set 2) = 4.53

c) The standard deviations are the same. Set 2 is simply Set 1 + 80. Although the measures of center and position change, the spread is exactly the same.
SD(Set 1) = 4.24 SD(Set 2) = 4.24

28. Standard deviation II.

a) Set 2 has the greater standard deviation. Both sets have the same mean (7), maximum (10), and minimum (4), but 6 and 8 are farther from the mean than 7.
SD(Set 1) = 2.12 SD(Set 2) = 2.24

b) The standard deviations are the same. Set 1 is simply Set 2 + 90. Although the measures of center and position are different, the spread is exactly the same.
SD(Set 1) = 36.06 SD(Set 2) = 36.06

c) Set 2 has the greater standard deviation. The central 4 values of Set 2 are simply the central 4 values of Set 1 + 40, but the maximum and minimum of Set 2 are farther away from the mean than the maximum and minimum of Set 1. Range(Set 1) = 18 and Range(Set 2) = 22. Since the Range of Set 2 is greater than the Range of Set 1, the standard deviation is also larger.
SD(Set 1) = 6.03 SD(Set 2) = 7.24

29. Pizza prices.

The mean and standard deviation would be used to summarize the distribution of pizza prices, since the distribution is unimodal and symmetric.

30. Neck size.

The mean and standard deviation would be used to summarize the distribution of neck sizes, since the distribution is unimodal and symmetric.

31. Pizza prices again.

a) The mean pizza price is closest to $2.60. That's the balancing point of the histogram.

b) The standard deviation in pizza prices is closest to $0.15, since that is the typical distance to the mean. There are no pizza prices as far as $0.50 or $1.00.

32. Neck sizes again.

a) The mean neck size is closest to 15 inches. That's the balancing point of the histogram.

b) The standard deviation in neck sizes is closest to 1 inch, because a typical value lies about 1 inch from the mean. There are a few points as far away as 3 inches from the mean, and none as far away as 5 inches. Those are too large to be the standard deviation.

33. Movie lengths 2010.

a) A typical movie would be a little over 100 minutes long. This is near the center of the unimodal and slightly skewed histogram, with the outlier set aside.

b) You would be surprised to find that your movie ran for 150 minutes. Only 3 movies ran that long.

c) It's difficult to say which would be higher. While the distribution of movie lengths is generally skewed to the right, which would raise the mean, there is a low outlier, which would lower the mean. (The actual mean of 107.07 minutes is a bit higher than the median of 104.50 minutes.)

34. Golf drives 2013.

a) The distribution of golf drives is roughly unimodal and symmetric, with a typical drive of a little over 290 yards. Professional golfers on the men's PGA tour had drives that were as short as about 255 yards, and as long as about 320 yards.

b) Approximately 25% of professional male golfers drive less than 280 yards.

c) According to the graph, the mean drive is between 285 and 295 yards.

d) The distribution of golf drives is approximately symmetric, so the mean and the median should be relatively close.

35. Movie lengths II 2010.

a) i) The distribution of movie running times is fairly consistent, with the middle 50% of running times between 98 and 116 minutes. The interquartile range is 18 minutes.

ii) The standard deviation of the distribution of movie running times is 16.6 minutes, which indicates that movies typically have running times fairly close to the mean running time.

b) Since the distribution of movie running times is generally skewed to the right and contains an outlier, the standard deviation is a poor choice of numerical summary for the spread. The interquartile range is better, since it is resistant to outliers.

36. Golf drives II 2013.

a) i) The distribution of PGA golf drives is fairly consistent, with the middle 50% of the drives having distances between 282.5 and 295.6 yards. The interquartile range is 13.1 yards.

ii) The standard deviation of the distribution of PGA golf drives is 11.2 yards, which indicates that golf drives are typically within 11.2 yards of the mean gold drive.

b) Since the distribution of golf drives is reasonably symmetric, both the standard deviation and the interquartile range are reasonable measures of spread.

37. Movie earnings 2013.

The industry publication is using the median, while the watchdog group is using the mean. It is likely that the mean is pulled higher by a few high earning movies.

38. Cold weather.

a) The mean temperature will be lower. The median temperature will not change, since the incorrect temperature is still the lowest temperature, and the median is based only on position.

b) The range and standard deviation in temperature will both increase, since the incorrect temperature is more extreme than the correct temperature. The IQR will not change, since the both the correct and incorrect scores are below the first quartile, and the IQR measures the distance between the first and third quartiles.

39. Payroll.

a) The mean salary is $\frac{1200+700+6(400)+4(500)}{12} = \$525$.

The median salary is the middle of the ordered list:

400 400 400 400 400 400 500 500 500 500 700 1200

The median is $450.

b) Only two employees, the supervisor and the inventory manager, earn more than the mean wage.

c) The median better describes the wage of the typical worker. The mean is affected by the two higher salaries.

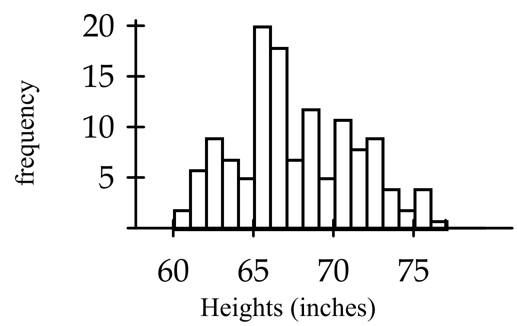
d) The IQR is the better measure of spread for the payroll distribution. The standard deviation and the range are both affected by the two higher salaries.

40. Singers full choir.

a) 5-number summary: 60, 65, 66, 70, 76, so the median is 66 inches and the IQR is 70 – 65 = 5 inches.

b) The mean height of the singers is 67.12 inches, and the standard deviation of the heights is 3.79 inches.

c) The histogram of heights of the choir members is at the right.



d) The distribution of the heights of the choir members is bimodal (probably due to differences in height of men and women) and skewed slightly to the right. The median is 66 inches. The distribution is fairly spread out, with the middle 50% of the heights falling between 65 and 70 inches. There are no gaps or outliers in the distribution.

41. Gasoline 2014.

a) Gasoline Prices

31	1
31	5
32	1233
32	6678
33	
33	9
34	23
34	556

Key : 32 | 1 = $3.21/gal

b) The distribution of gas prices is bimodal, with two clusters, one centered around \$3.45 per gallon, and another centered around \$3.25 per gallon. The lowest and highest prices were \$3.11 and \$3.46 per gallon.

c) There is a gap in the distribution of gasoline prices. There were no stations that charged between \$3.28 and \$3.39.

42. The Great One.

a)

Wayne Gretzky –
Games played per season

```
8 | 000000122
7 | 8899
7 | 0344
6 |
6 | 4          Key:
5 |            7 | 8 = 78
5 |                    games
4 | 58
4 |
```

b) The distribution of the number of games played by Wayne Gretzky is skewed to the left.

c) Typically, Wayne Gretzky played about 80 games per season. The number of games played is tightly clustered in the upper 70s and low 80s.

d) Two seasons are low outliers, when Gretzky played fewer than 50 games. He may have been injured during those seasons. Regardless of any possible reasons, these seasons were unusual compared to Gretzky's other seasons.

43. States.

a) The distribution of state populations is skewed heavily to the right. Therefore, the median and IQR are the appropriate measures of center and spread.

b) The mean population must be larger than the median population. The extreme values on the right affect the mean greatly and have no effect on the median.

c) There are 50 entries in the stemplot, so the median must be between the 25th and 26th population values. Counting in the ordered stemplot gives median = 4.5 million people. The middle of the lower 50% of the list (25 state populations) is the 13th population, or 2 million people. The middle of the upper half of the list (25 state populations) is the 13th population from the top, or 7 million people. The IQR = Q3 – Q1 = 7 – 2 = 5 million people.

d) The distribution of population for the 50 U.S. States is unimodal and skewed heavily to the right. The median population is 4.5 million people, with 50% of states having populations between 2 and 7 million people. There are two outliers, a state with 37 million people, and a state with 25 million people. The next highest population is only 19 million.

44. Wayne Gretzky.

a) The distribution of the number of games played per season by Wayne Gretzky is skewed to the left, and has low outliers. The median is more resistant to the skewness and outliers than the mean.

b) The median, or middle of the ordered list, is 79 games. Both the 10^{th} and 11^{th} values are 79, so the median is the average of these two, also 79.

c) The mean should be lower. There are two seasons when Gretzky played an unusually low number of games. Those seasons will pull the mean down.

45. A-Rod 2013.

The distribution of the number of homeruns hit by Alex Rodriguez during the 1994 – 2013 seasons is reasonably symmetric, with the exception of a second mode around 10 homeruns. A typical number of homeruns per season was in the high 30s to low 40s. With the exception of 3 seasons in which A-Rod hit 0 , 5, and 7 homeruns, his total number of homeruns per season was between 16 and the maximum of 57.

46. Bird species 2013.

a) The results of the 2013 Laboratory of Ornithology Christmas Bird Count are displayed in the stem and leaf display at the right.

b) The distribution of the number of birds spotted by participants in the 2013 Laboratory of Ornithology Christmas Bird Count is skewed right, with a median of 117 birds. There are three high potential outliers, with participants spotting 150, 166, and 184 birds. With the exception of these outliers, most participants saw between 82 and 136 birds.

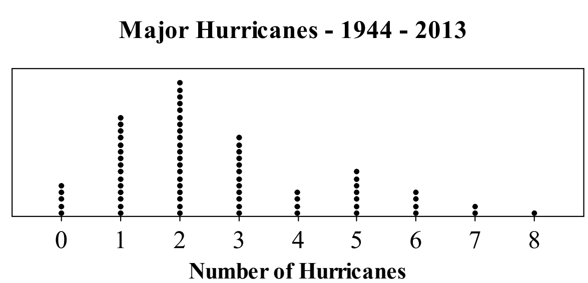
Number of Birds

8	2368
9	78
10	1156
11	8
12	468
13	136
14	
15	0
16	6
17	
18	4

Key : 15 | 0 = 150 birds

47. Major Hurricanes 2013.

a) A dotplot of the number of hurricanes each year from 1944 through 2013 is displayed. Each dot represents a year in which there were that many hurricanes.



b) The distribution of the number of hurricanes per year is unimodal and skewed to the right, with center around 2 hurricanes per year. The number of hurricanes per year ranges from 0 to 8. There are no outliers. There may be a second mode at 5 hurricanes per year, but since there were only 6 years in which 5 hurricanes occurred, this may simply be natural variability.

48. Horsepower.

The distribution of horsepower of cars reviewed by *Consumer Reports* is nearly uniform. The lowest horsepower was 65 and the highest was 155. The center of the distribution was around 105 horsepower.

Horsepower

15	05
14	2
13	0358
12	0559
11	00555
10	359
9	00577
8	058
7	01158
6	55889

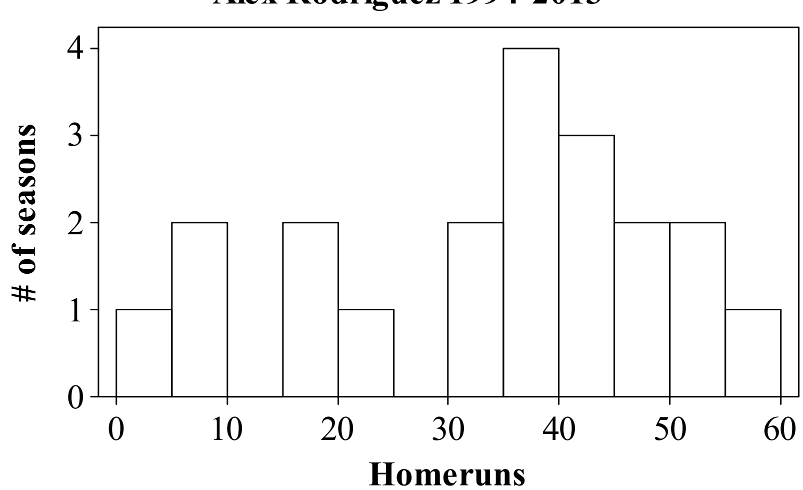
Key : 15 | 0 = 150 hp

49. A-Rod again 2013.

a) This is not a histogram. The horizontal axis should the number of home runs per year, split into bins of a convenient width. The vertical axis should show the frequency; that is, the number of years in which A-Rod hit a number of home runs within the interval of each bin. The display shown is a bar chart/time plot hybrid that simply displays the data table visually. It is of no use in describing the shape, center, spread, or unusual features of the distribution of home runs hit per year by A-Rod.

b) The histogram is at the right.

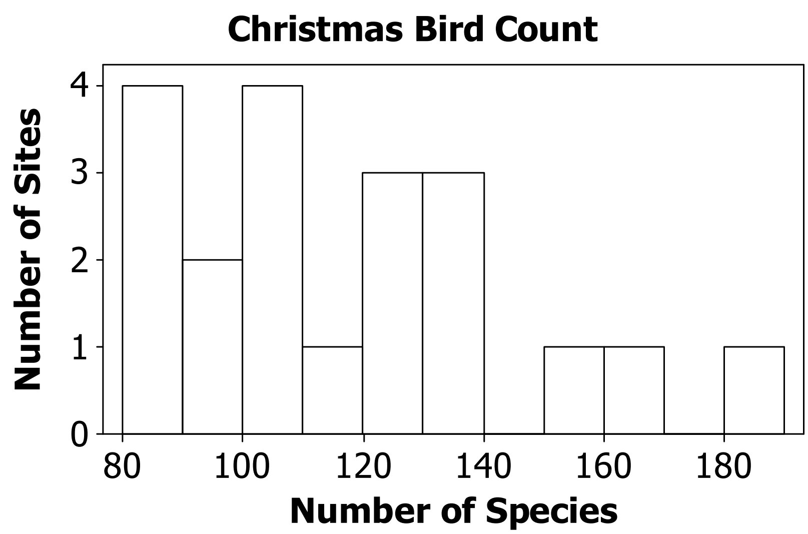
Alex Rodriguez 1994-2013



50. Return of the birds 2013.

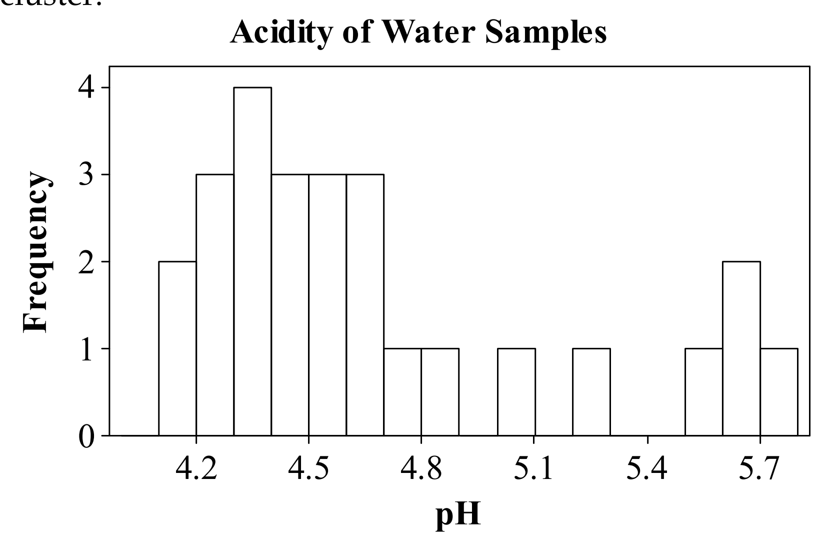
a) This is not a histogram. The horizontal axis should split the number of counts from each site into bins. The vertical axis should show the number of sites in each bin. The given graph is nothing more than a bar chart, showing the bird count from each site as its own bar. It is of absolutely no use for describing the shape, center, spread, or unusual features of the distribution of bird counts.

b) The histogram is below.



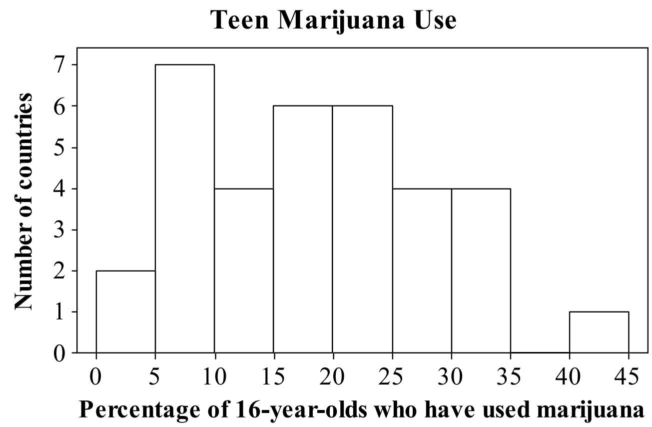
51. Acid rain.

The distribution of the pH readings of water samples in Allegheny County, Penn. is bimodal. A roughly uniform cluster is centered around a pH of 4.4. This cluster ranges from pH of 4.1 to 4.9. Another smaller, tightly packed cluster is centered around a pH of 5.6. Two readings in the middle seem to belong to neither cluster.



52. Marijuana 2007.

The distribution of the percentage of 16-year-olds in 34 countries who have used marijuana is somewhat bimodal, with 9 countries having between 3% and 10% of 16-year-olds having used marijuana. Another group of 12 countries has between 15% and 25% of teens who have used marijuana. Armenia, at 3%, had the lowest percentage of 16-year-olds who have tried marijuana. Czech Republic had the highest percentage, at 45%. A typical country might have a percentage of approximately 20%.



53. Final grades.

The width of the bars is much too wide to be of much use. The distribution of grades is skewed to the left, but not much more information can be gathered.

54. Final grades revisited.

a) This display has a bar width that is much too narrow. As it is, the histogram is only slightly more useful than a list of scores. It does little to summarize the distribution of final exam scores.

b) The distribution of test scores is skewed to the left, with center at approximately 170 points. There are several low outliers below 100 points, but other than that, the distribution of scores is fairly tightly clustered.

55. Zip codes.

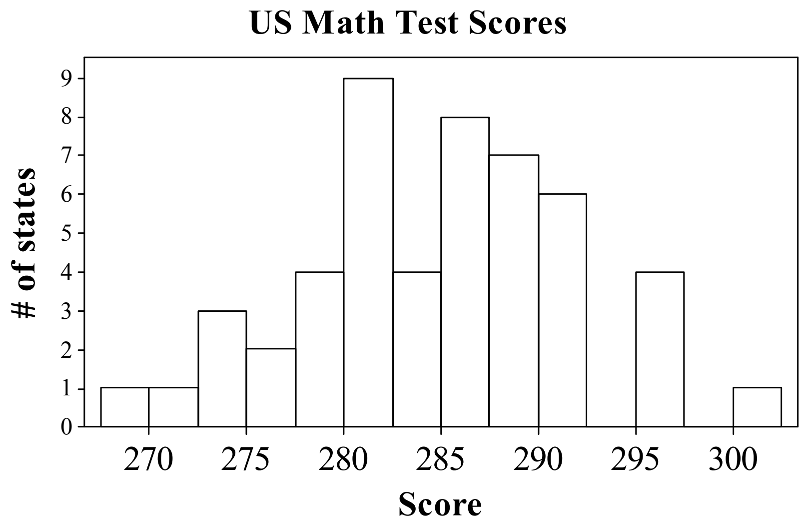
Even though zip codes are numbers, they are not quantitative in nature. Zip codes are categories. A histogram is not an appropriate display for categorical data. The histogram the Holes R Us staff member displayed doesn't take into account that some 5-digit numbers do not correspond to zip codes or that zip codes falling into the same classes may not even represent similar cities or towns. The employee could design a better display by constructing a bar chart that groups together zip codes representing areas with similar demographics and geographic locations.

56. Zip codes revisited

The statistics cannot tell us very much since zip codes are categorical. However, there is *some* information in the first digit of zip codes. They indicate a general East (0-1) to West (8-9) direction. So, the distribution shows that a large portion of their sales occurs in the West and another in the 32000 area. But a bar chart of the first digits would be the appropriate display to show this information.

57. Math scores 2013.

a) Median: 285
IQR: 9
Mean: 284.36
Standard deviation: 6.84

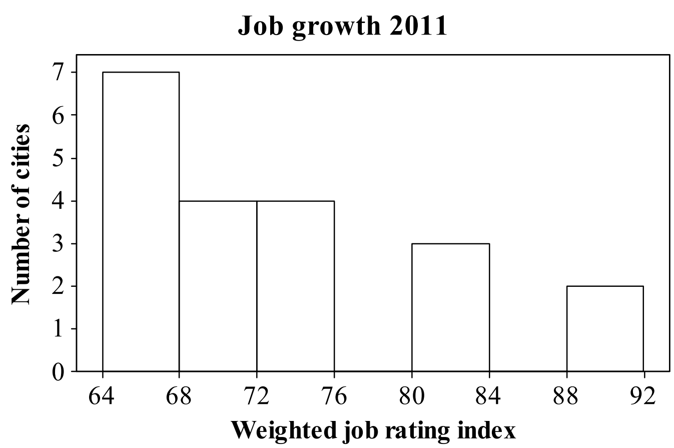


b) Since the distribution of Math scores is skewed to the left, it is probably better to report the median and IQR.

c) The distribution of average math achievement scores for eighth graders in the United States is skewed slightly to the left, and roughly unimodal. The distribution is centered at 285. Scores range from 269 to 301, with the middle 50% of the scores falling between 280 and 289.

58. Boomtowns 2011.

a) A histogram of the job growth rates of NewGeography.com's best cities for job growth is at the right. A boxplot, stemplot, or dotplot would also have been an acceptable display.



b) The mean weighted job rating index is 73.03% and the median weighted job rating index is 71.80%. The mean is higher because distribution is skewed to the right.

c) The median would be the appropriate measure of center of the distribution of weighted job rating indices, since the distribution is skewed to the right.

d) The standard deviation of the distribution of weighted job rating indices is 7.61% and the IQR is 10.10%.

e) The IQR is the appropriate measure of spread, because the skewness influences the standard deviation.

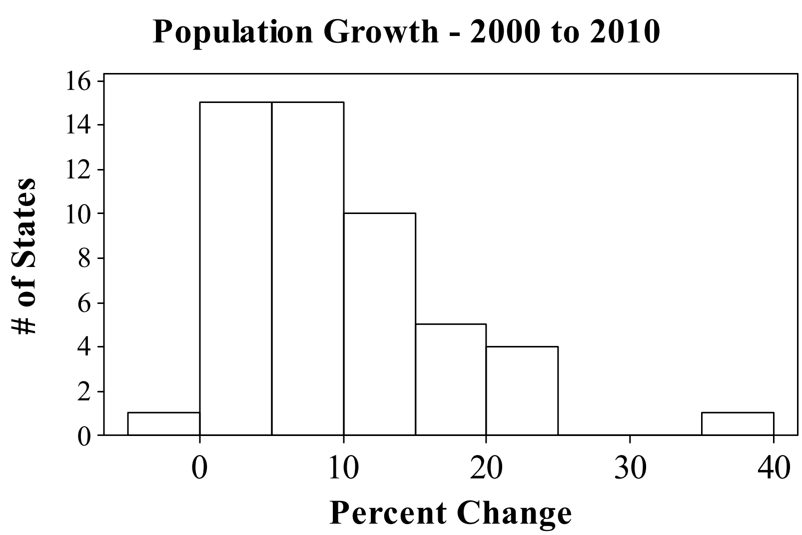
f) If 49.23% were subtracted from each of the weighted job rating indices, the mean and median would each decrease by 49.23%. The standard deviation and the IQR would not change.

g) If we were to set aside Austin-Round Rock-San Marcos, the highest weighted job rating index, the mean would decrease. The skewness was pulling it up. The standard deviation would decrease, since the skewness gave the impression of more spread. The median and IQR would be relatively unaffected, since those measures are resistant to the presence of skewness, although they would change slightly, since they are each based upon relative position. With the highest rating removed, there would only be 19 rating indices, instead of 20. This would cause the median and the quartiles to shift down slightly.

h) The distribution of weighted job rating indices is roughly unimodal and skewed to the right. The median weighted job rating index for these cities is 71.80%. The middle 50% of the cities had weighted job rating indices between 67.25% and 77.35%, for an interquartile range of 10.10%. The median and IQR are the best measures of spread, since the distribution is skewed.

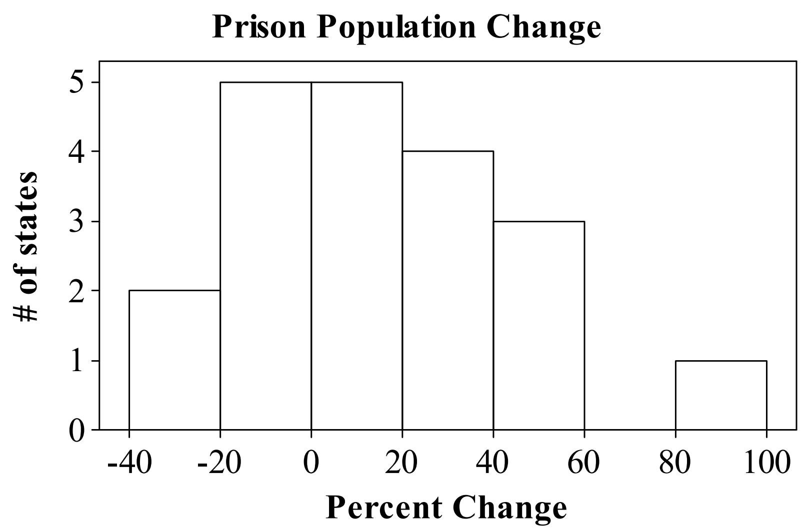
59. Population growth 2010.

The distribution of population growth among the 50 United States and the District of Columbia is unimodal and skewed to the right. Most states experienced modest growth, as measured by percent change in population between 2000 and 2010. Nearly every state experienced positive growth, with the exception of Michigan. The median population growth was 7.8%, with the middle 50% of states experiencing between 4.30% and 14.10% growth, for an IQR of 9.80. The distribution contains one high outlier. Nevada experienced population growth of 35.1%.



60. Prisons 2013.

The median increase in federal prison populations from 2000 to 2012 in 20 northeastern and midwestern states was 10.4% with 7 of the 20 states showing a decrease. The distribution is unimodal and skewed to the right. The large IQR of 35.3% indicates much variability from state to state, with half of these states experiencing prison population increases in excess of 10%.



Chapter 4 – Understanding and Comparing Distributions

Section 4.1

1. Load factors, 2013.

The distribution of domestic load factors and the distribution of international load factors are both unimodal and skewed to the left. The distribution of international load factors may contain a low outlier. Because the distributions are skewed, the median and IQR are the appropriate measures of center and spread. The medians are very close, which tell us that typical international and domestic load factors are about the same. The IQRs show a bit more variability in the domestic load factors.

2. Load factor, 2013 by season.

The distribution of Spring/Summer load factors and the distribution of Fall/Winter load factors are both unimodal and skewed to the left. Load factors in the Fall/Winter period vary less than load factors in the Spring/Summer period, but are generally higher. The center of the distribution of Fall/Winter load factors is around 82, while the center of the distribution of Spring/Summer load factors is around 77.

Section 4.2

3. Load factors 2013 by month.

Load factors are generally higher and less variable in the summer months (June – August). They are lower and more variable in the winter and spring.

4. Load factors 2013 by year.

Load factors have generally increased steadily since 2001. They may have become less variable in recent years.

Section 4.3

5. Extraordinary months.

Air travel immediately after the events of 9/11 was not typical of air travel in general. If we want to analyze monthly patterns, it might be best to set these months aside.

6. Extraordinary months again.

Outliers are dependent on context. The low outlier evident in the single boxplot must be the lowest value from 2001, but load factors were generally lower in 2001 than they were overall. That value wasn't an outlier when compared to the other low values of 2001, but it stood out overall, as load factors increased.

Section 4.4

7. Load factors 2013 over time.

a) After a period of little change in 2000-2001, load factors have been increasing steadily.

b) We would never assume that a pattern like this would continue. This case illustrates one of the reasons why we wouldn't assume this. Since load factors are percentages, they cannot exceed 100%. At the very least, the load factors would have to level out in the future.

8. Load factors 2013 over time, a second look.

a) With the median smoother, the seasonal pattern that was witnessed in Exercise 3 becomes evident. Higher load factors are expected in the summer months.

b) Yes, we can expect this pattern to persist, because it reflects seasonal effects, such as summer vacation time, that will probably continue.

Section 4.5

9. Exoplanets.

It is difficult to summarize data with a distribution this skewed. The extremely large values will dominate any summary or description.

10. Exoplanets re-expressed.

a) Yes, this re-expressed scale is better for understanding these distances. The log scale provides a nearly symmetric distribution, and points out that the sun was included in the data, probably accidentally.

b) The sun should not be included in data about extra-solar planets.

Chapter Exercises

11. In the news. Answers will vary.

12. In the news. Answers will vary.

13. Time on the Internet. Answers will vary.

14. Groups on the Internet. Answers will vary.

15. Pizza prices.

a) Pizza prices appear to be both higher on average, and more variable, in Baltimore than in the other three cities. Prices in Chicago may be slightly higher on average than in Dallas and Denver, but the difference is small.

b) There are low outliers in the distribution of pizza prices in Baltimore and Chicago. There is one high outlier in the distribution of pizza prices in Dallas. These outliers do not affect the overall conclusions reached in the previous part.

16. Costs.

a) Coffee is the most expensive commodity, on average.

b) Newspapers are generally more expensive than a ride on public transportation, but there are cities in which a ride on public transportation is more expensive than a newspaper in other cities.

c) Each distribution has a high outlier, but the presence of the outlier doesn't affect any of the previous conclusions.

17. Rock concert accidents.

a) The histogram and boxplot of the distribution of "crowd crush" victims' ages both show that a typical crowd crush victim was approximately 18 - 20 years of age, that the range of ages is 36 years, that there are two outliers, one victim at age 36 - 38 and another victim at age 46 – 48.

b) This histogram shows that there may have been two modes in the distribution of ages of "crowd crush" victims, one at 18 – 20 years of age and another at 22 – 24 years of age. Boxplots, in general, can show symmetry and skewness, but not features of shape like bimodality or uniformity. Most victims were between 16 and 24 years of age.

c) The median is the better measure of center, since the distribution of ages has outliers. The median is more resistant to outliers than the mean.

d) The IQR is a better measure of spread, since the distribution of ages has outliers. The IQR is more resistant to outliers than the standard deviation.

18. Slalom times 2010.

a) The histogram and boxplot of the distribution of Men's Giant Slalom times both show that a typical time of around 165 seconds, that the range of slalom times is about 55 seconds. Both distributions also show that the distribution of slalom times is skewed to the high end.

b) Since the distribution of slalom times is skewed, and contains possible outliers, the median is the better summary of center.

d) In the presence of skewness and possible outliers, we'd prefer the IQR to the standard deviation as a measure of spread.

19. Cereals.

a) The maximum sugar content is approximately 60% and the minimum sugar content is approximately 1%, so the range of sugar contents is about 60 – 1 = 59%.

b) The distribution of sugar content of cereals is bimodal, with modes centered around 5% and 45% sugar by weight.

c) Some cereals are healthy, low-sugar brands, and others are very sugary.

d) Yes. The minimum sugar content in the children's cereals is about 35% and the maximum sugar content of adult cereals is only 34%.

e) The range of sugar contents is about the same for the two types of cereals, approximately 28%, but the IQR is larger for the adult cereals. This is an indication of more variability in the sugar content of the middle 50% of adult cereals.

20. Tendon transfers.

a) The distribution of pushing strength scores is unimodal and symmetric.

b) The maximum pushing strength score is 4 and the minimum is 1, so the range is a score of 4 – 1 = 3.

c) The histogram does not show that the results of the two procedures typically resulted in different strengths.

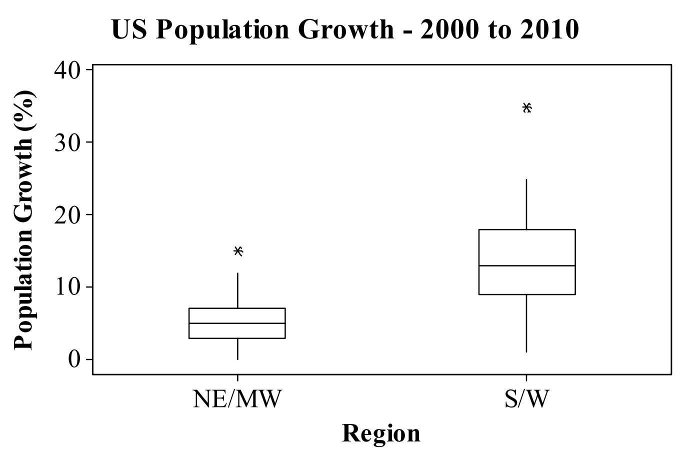
d) The distribution of biceps transfer strength scores had a higher median than the distribution of deltoid transfer strength scores.

e) The biceps transfer was not always the best. The highest strength score in the deltoid transfer was higher than the lowest 25% of biceps transfer strength scores.

f) The deltoid transfer produced more consistent strength scores. The IQR is much smaller for this group, even though the ranges are approximately the same.

21. Population growth 2010 by region.

a) Comparative boxplots are at the right.



b) The distribution of population growth in NE/MW states is unimodal, symmetric and tightly clustered around 5% growth. The distribution of population growth in S/W states is much more spread out, with most states having population growth between 5% and 25%. A typical state had about 15% growth. There was an outliers, with 35% growth. Generally, the growth rates in the S/W states were higher and more variable than the rates in the NE/MW states.

22. Camp sites.

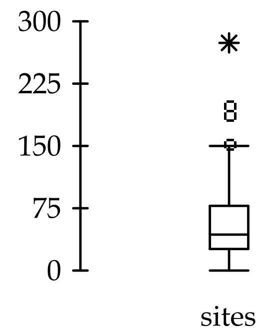
a) The distribution of the number of campsites in public parks in Vermont is strongly skewed to the right, so median and IQR are appropriate measures of center and spread.

b) IQR = Q3 – Q1 = 78 – 28 = 50.
Using the Outlier Rule (1.5 IQRs beyond quartiles):

Upper Fence: Q3+1.5(IQR)=78+1.5(50)
=78+75
=153

Lower Fence: Well below 0 campsites.

There are 3 parks with greater than 180 campsites. These are definitely outliers. There are 2 parks with between 150 and 160 campsites each. These may be outliers as well.



c) A boxplot of the distribution of number of campsites is at the right.

d) The distribution of the number of campsites at public parks in Vermont is unimodal and skewed to the right. The center of the distribution is approximately 44 campsites. The distribution of campsites is quite spread out, with several high outliers. These parks have in excess of 153 campsites each.

23. Hospital stays.

a) The histograms of male and female hospital stay durations would be easier to compare if the were constructed with the same scale, perhaps from 0 to 20 days.

b) The distribution of hospital stays for men is skewed to the right, with many men having very short stays of about 1 or 2 days. The distribution tapers off to a maximum stay of approximately 25 days. The distribution of hospital stays for women is skewed to the right, with a mode at approximately 5 days, and tapering off to a maximum stay of approximately 22 days. Typically, hospital stays for women are longer than those for men.

c) The peak in the distribution of women’s hospital stays can be explained by childbirth. This time in the hospital increases the length of a typical stay for women, and not for men.

24. Deaths 2011.

a) The distributions of ages at death of black Americans and white Americans are both unimodal and skewed to the left, towards the lower ages of death.

b) The distributions of death ages differ mainly in their spreads and centers. The age at death for black Americans is more variable than the age at death for white Americans. A greater proportion of black Americans die between the ages of 25 and 74 than do white Americans, while a greater proportion of white Americans dies at ages older than 74 than do black Americans.

c) The interval widths are not consistent. Most of the bars are 10 years wide, but the first bar is only 1 year wide, the second bar is 5 years wide, while the last bar has width "85 and older".

25. Women's basketball.

a) Both girls have a median score of about 17 points per game, but Scyrine is much more consistent. Her IQR is about 2 points, while Alexandra's is over 10.

b) If the coach wants a consistent performer, she should take Scyrine. She'll almost certainly deliver somewhere between 15 and 20 points. But, if she wants to take a chance and needs a "big game", she should take Alexandra. Alex scores over 24 points about a quarter of the time. On the other hand, she scores under 11 points about as often.

26. Gas prices 2013.

a) Gas prices generally increased yearly for 2009 through 2011, then stablilized. Gas prices in 2010 were much more consistent than the prices in other years.

b) The distribution of gas prices in 2009 shows the greatest range and the biggest IQR, so the prices varied a great deal. 2010 had very consistent gas prices. Gas prices in 2011-2013 had similar variability, more consistent than 2009, but less consistent than 2010.

27. Marriage age.

The distribution of marriage age of U.S. men is skewed right, with a typical man (as measured by the median) first marrying at around 24 years old. The middle 50% of male marriage ages is between about 23 and 26 years. For U.S. women, the distribution of marriage age is also skewed right, with median of around 21 years. The middle 50% of female marriage age is between about 20 and 23 years. When comparing the two distributions, the most striking feature is that the distributions are nearly identical in spread, but have different centers. Females typically seem to marry earlier than males. In fact, between 50% and 75% of the women marry at a younger age than *any* man.

28. Fuel economy by cylinders.

Cars with 4 cylinders generally get better gas mileage than cars with 6 cylinders, which generally get better gas mileage than cars with 8 cylinders. Additionally, the greater the number of cylinders, the more consistent the mileage becomes. 4 cylinder cars typically get between 27 – 33 mpg, 6 cylinder cars typically get between 18 – 22 mpg, and 8 cylinder cars typically get between 16 – 19 mpg. We don't have enough data to compare 5-cylinder cars.

29. Fuel economy 2012.

(Note: numerical details may vary.) In general, fuel economy is highest in mid-size cars, lower in SUVs, and lowest in pickup trucks. There are numerous outliers on both ends for cars, one high outlier for pickups and two high outliers for SUVs. The top 50% of cars get higher fuel economy than 75% of SUVs and all pickups. The distributions of fuel economy for cars and SUVs have approximately the same IQR, about 5 miles per gallon, but cars generally have higher combined fuel economy. Furthermore, there are several cars with much higher fuel economy. Pickups trucks have consistently low fuel economy.

30. Ozone.

a) April had the highest recorded ozone level, approximately 440.

b) March had the largest IQR of ozone level, approximately 50.

c) August had the smallest range of ozone levels, approximately 50.

d) January had a lower median ozone level than June, 340 and 350, respectively, but June's ozone levels were much more consistent.

e) Generally, ozone levels rose through the winter and were highest in the spring, then fell through the summer and were lowest in the fall. Additionally, ozone levels were very consistent in the summer, became more variable in the fall, were most variable in the winter, and became more consistent through the spring.

31. Test scores.

Class A is Class 1. The median is 65, but has less spread than Class B, which is Class 2. Class C is Class 3, since it's median is higher, which corresponds to the skew to the left.

32. Eye and hair color.

The graph is not appropriate. Boxplots are for quantitative data, and these are categorical data, although coded as numbers. The numbers used for hair color and eye color are arbitrary, so the boxplot and any accompanying statistics for eye color make no sense.

33. Graduation?

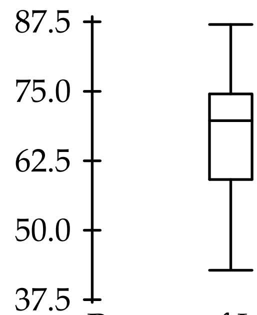
a) The distribution of the percent of incoming college freshman who graduate on time is roughly symmetric. The mean and the median are reasonably close to one another and the quartiles are approximately the same distance from the median.

b) Upper Fence: Q3+1.5(IQR)=74.75+1.5(74.75-59.15)
=74.75+23.4
=98.15

Lower Fence: Q1-1.5(IQR)=59.15-1.5(74.75-59.15)
=59.15-23.4
=35.75

Since the maximum value of the distribution of the percent of incoming freshmen who graduate on time is 87.4% and the upper fence is 98.15%, there are no high outliers. Likewise, since the minimum is 43.2% and the lower fence is 35.75%, there are no low outliers. Since the minimum and maximum percentages are within the fences, all percentages must be within the fences.

c) A boxplot of the distribution of the percent of incoming freshmen who graduate on time is at the right.



Percent of Incoming Freshman Graduating in 4 Years

d) The distribution of the percent of incoming freshmen who graduate on time is roughly symmetric, with mean of approximately 68% of freshmen graduating on time. Universities surveyed had between 43.2% and 87.4% of students graduating on time, with the middle 50% of universities reporting between 59.15% and 74.75% graduating on time.

34. Vineyards.

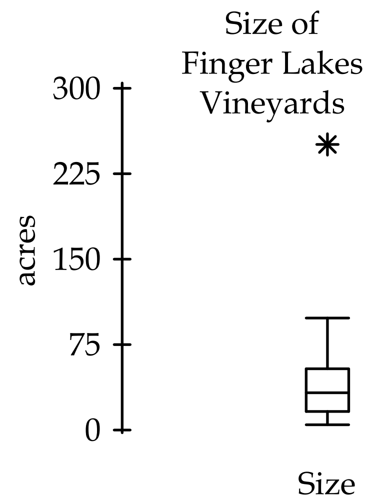
a) The distribution of size of Finger Lakes vineyards is skewed heavily to the right. The mean size is a great deal higher than the median size.

b) Upper Fence: Q3+1.5(IQR)=55+1.5(55-18.5)
=55+54.75
=109.75

Lower Fence: Q1-1.5(IQR)=18.5-1.5(55-18.5)
=18.5-54.75
=-36.25

The maximum of 250 acres is well above the upper fence of 109.75 acres. Therefore, there is at least one high outlier, 250 acres. Since the lower fence is negative, there are no low outliers, since it is certainly impossible to have a vineyard with negative size.

c) The boxplot of the distribution of sizes of Finger Lakes vineyards is at the right. There may be additional outliers, but we are sure that there is at least one, the maximium.



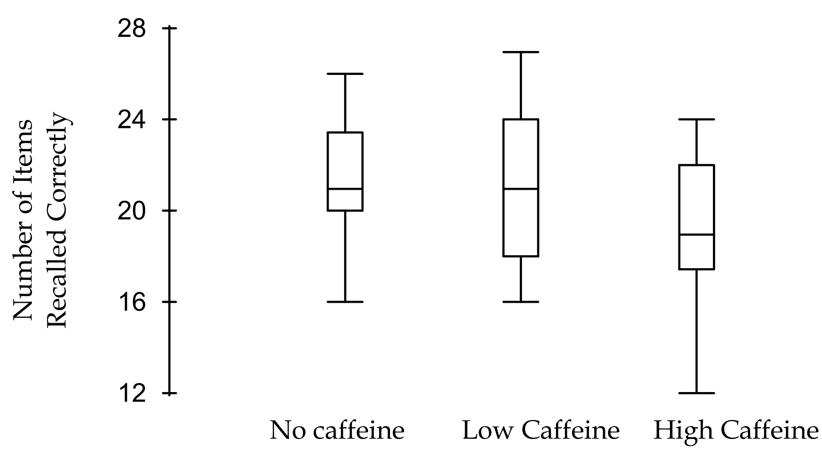
d) The distribution of sizes of Finger Lakes vineyards is skewed to the right. Many vineyards have moderate sizes, with the middle 50% of vineyards consisting of 18.5 to 55 acres. The smallest vineyard is 6 acres. At least one vineyard is comparatively bigger, at 250 acres. The median acreage is 33.5.

35. Caffeine.

a) *Who* – 45 student volunteeers. *What* – Level of caffeine consumption and memory test score. *When* – Not specified. *Where* – Not specified. *Why* – The student researchers want to see the possible effects of caffeine on memory. *How* – It appears that the researchers imposed the treatment of level of caffeine consumption in an experiment. However, this point is not clear. Perhaps they allowed the subjects to choose their own level of caffeine.

b) *Variables* – Caffeine level is a categorical variable with three levels: no caffeine, low caffeine, and high caffeine. Test score is a quantitative variable, measured in number of items recalled correctly.

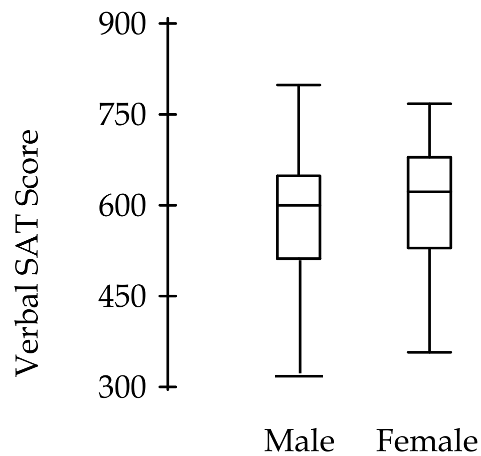
c)



d) The groups consuming no caffeine and low caffeine had comparable memory test scores. A typical score from these groups was around 21. However, the scores of the group consuming no caffeine were more consistent, with a smaller range and smaller interquartile range than the scores of the group consuming low caffeine. The group consuming high caffeine had lower memory scores in general, with a median score of about 19. No one in the high caffeine group scored above 24, but 25% of each of the other groups scored above 24.

36. SAT scores.

a) Parallel boxplots comparing the scores of boys and girls SAT scores are at the right.



b) Females in this graduating class scored slightly higher on the Verbal SAT, with a median of 625, compared to the median of 600 for the males. Additionally, the females had higher first and third quartiles. The IQR of the males' scores was slightly smaller, than the IQR for the females' scores, indicating a bit more consistency in male scores. However, the overall spread of male scores was greater than that of female scores, with males having both the minimum and maximum score. Both distributions of scores were slightly skewed to the left.

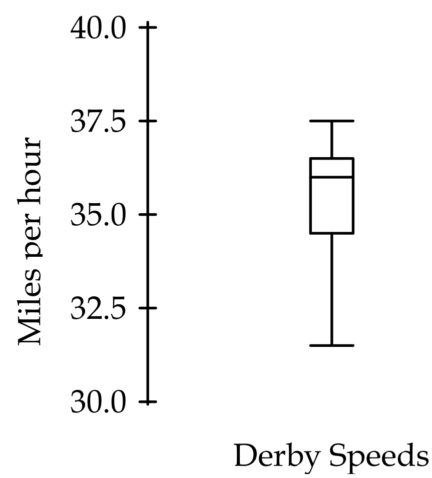
37. Derby speeds 2014.

a) The median speed is the speed at which 50% of the winning horses ran slower. Find 50% on the left, move straight over to the graph and down to a speed of about 36.25 mph.

b) Quartile 1 is at 25% on the left, and Quartile 3 is at 75% on the left. Matching these to the ogive, Q1 = 35 mph and Q3 = 36.8 mph, approximately.

c) Range = Max – Min = 37.5 – 31.4 = 6.1 mph
IQR = Q3 – Q1 = 36.8 – 35 = 1.8 mph

d) An approximate boxplot of winning Kentucky Derby Speeds is at the right.

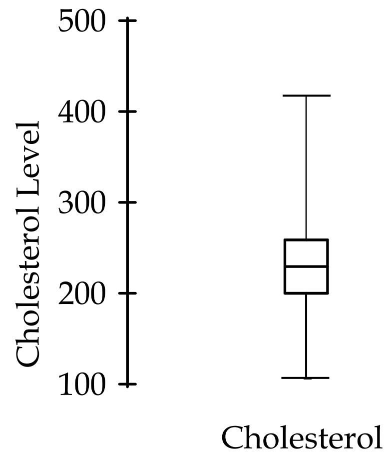


e) The distribution of winning speeds in the Kentucky Derby is skewed to the left. The lowest winning speed is just under 31 mph and the fastest speed is about 38 mph. The median speed is approximately 36 mph, and 75% of winning speeds are above 35 mph. Only a few percent of winners have had speeds below 33 mph. The middle 50% of winning speeds are between 35 and 37 mph.

38. Cholesterol.

A boxplot for the distribution of cholesterol levels of 1400 men is at the right. The five number summary is estimated from the ogive to be: 100, 200, 230, 260, 425.

The distribution of cholesterol levels is skewed to the right, and tightly clustered around the median. The median cholesterol level is approximately 230, with the middle 50% of cholesterol levels between 200 and 260. A very small percentage of men had cholesterol below 150 or above 325.



39. Reading scores.

a) The highest score for boys was 6, which is higher than the highest score for girls, 5.9.

b) The range of scores for boys is greater than the range of scores for girls.
Range = Max – Min Range(Boys) = 4 Range(Girls) = 3.1

c) The girls had the greater IQR.
IQR = Q3 – Q1 IQR(Boys) = 4.9 – 3.9 = 1 IQR(Girls) = 5.2 – 3.8 = 1.4

d) The distribution of boys' scores is more skewed. The quartiles are not the same distance from the median. In the distribution of girls' scores, Q1 is 0.7 units below the median, while Q3 is 0.7 units above the median.

e) Overall, the girls did better on the reading test. The median, 4.5, was higher than the median for the boys, 4.3. Additionally, the upper quartile score was higher for girls than boys, 5.2 compared to 4.9. The girls' lower quartile score was slightly lower than the boys' lower quartile score, 3.8 compared to 3.9.

f) The overall mean is calculated by weighting each mean by the number of students. $\dfrac{14(4.2)+11(4.6)}{25} = 4.38$

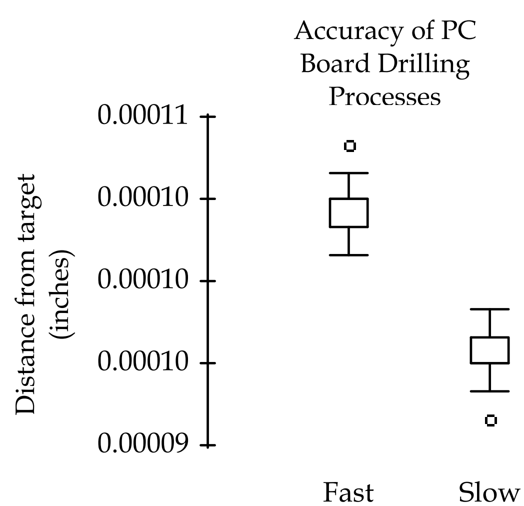
40. Cloud seeding.

a) Median and IQR, as well as the quartiles are the appropriate summary statistics, since the distribution of amount of rain produced is either skewed to the right or has high outliers. Indication of skewness and outliers can be seen in the comparison of median and mean. The mean amount of rain produced is significantly higher than the median for both seeded and unseeded clouds. Skewness or outliers pulled up the sensitive mean.

b) There is evidence that that the seeded clouds produced more rain. The median and both quartiles are higher than the corresponding statistics for unseeded clouds. In fact, the median amount of rainfall for seeded clouds is 221.60 acre-feet, about 5 times the median amount for unseeded clouds.

41. Industrial experiment.

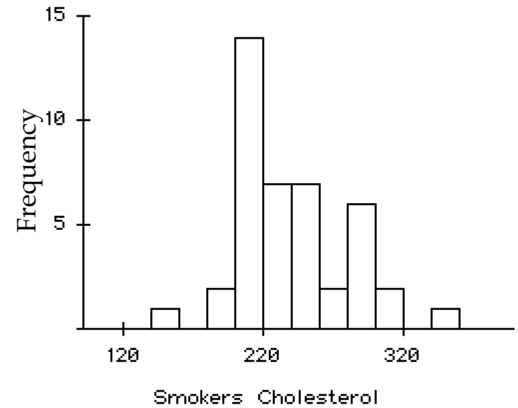
First of all, there is an extreme outlier in the distribution of distances for the slow speed drilling. One hole was drilled almost an inch away from the center of the target! If that distance is correct, the engineers at the computer production plant should investigate the slow speed drilling process closely. It may be plagued by extreme, intermittent inaccuracy. The outlier in the slow speed drilling process is so extreme that no graphical display can display the distribution in a meaningful way while including that outlier. That distance should be removed before looking at a plot of the drilling distances.



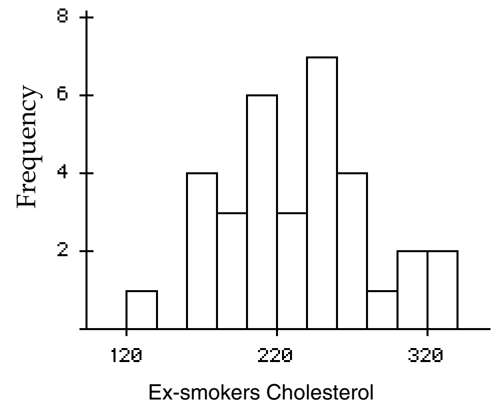
With the outlier set aside, we can determine that the slow drilling process is more accurate. The greatest distance from the target for the slow drilling process, 0.000098 inches, is still more accurate than the smallest distance for the fast drilling process, 0.000100 inches.

42. Cholesterol.

The distribution of cholesterol levels for smokers is unimodal and skewed slightly to the right, with a mode around 210. Cholesterol levels vary from approximately 140 to 350, but are generally clustered between 200 and 300. There is one low cholesterol level and one high cholesterol level, but these don't depart from the overall pattern.



The distribution of cholesterol levels for ex-smokers is unimodal and roughly symmetric, with a center around 240. Cholesterol levels vary from approximately 120 to 340, and seem spread out. There is one value, but not unusually low.

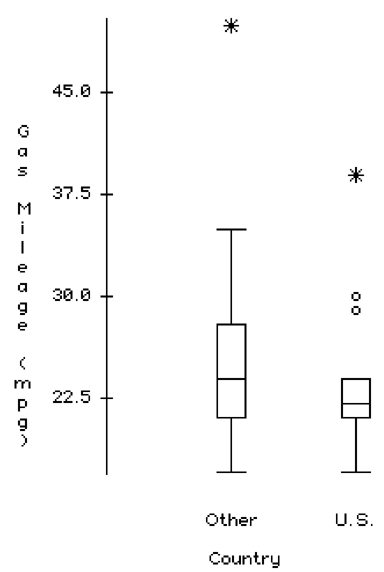


In general, the cholesterol levels of smokers seem to be slightly lower than the cholesterol levels of ex-smokers. Additionally, the cholesterol levels of smokers appear more consistent than cholesterol levels of ex-smokers.

43. MPG.

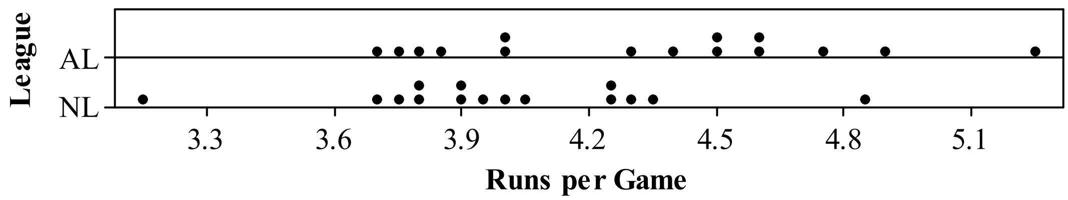
a) Comparative boxplots are at the right. Comparative dotplots, histograms, or stemplots would also be acceptable.

b) In general, the Other cars got better gas mileage than the US Cars. The median of the distribution of mileage for US cars was approximately 22 miles per gallon, while the median for Other cars was approximately 24 miles per gallon. U.S. cars had a distribution that was less variable than the distribution of Other cars, with 75% of US cars having gas mileage that was lower than the median of cars from Other countries. Both groups had some high outliers, most likely hybrid models.



44. Baseball 2013.

a) A comparative dotplot appears below. Back-to-back stem-and-leaf displays, comparative histograms, and boxplots on the same scale would also be acceptable.



b) The distributions of average number of runs per game for American League teams and National League teams are both reasonably symmetric. American League teams typically score slightly more runs per game on average, with a median of about 4.4 runs. The median number of runs scored for National League teams is about 3.95 runs. The American League has a distribution that is slightly more spread out than the National League, with IQRs of approximately 0.75 runs and 0.44 runs, respectively.

c) The 5.27 runs per game scored by the Boston Red Sox and the 4.83 runs are not outliers in their respective leagues for the 2013 season, although they are both the highest average number of runs scored for each league.

45. Fruit Flies.

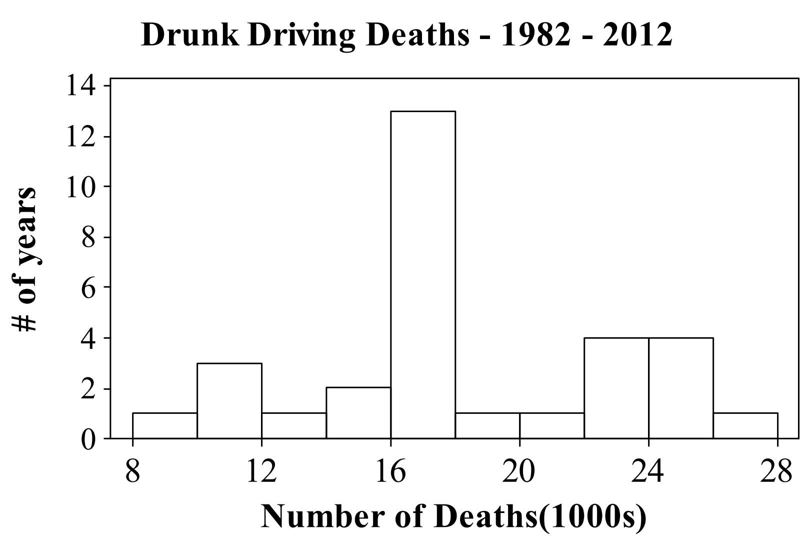
a) The most fruit flies died around day 16. Over 60,000 fruit flies died that day.

b) The largest proportion of fruit flies died around day 65. About 20% of that previously surviving population died that day.

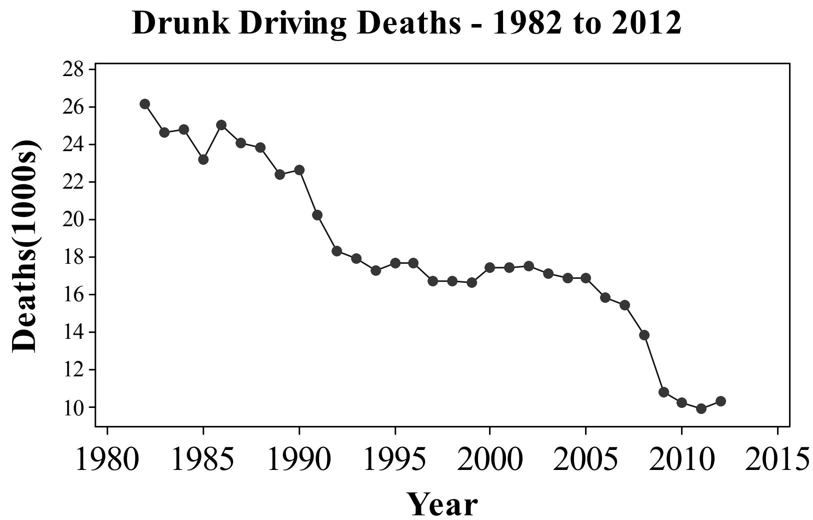
c) At around day 50, the number of fruit flies wasn't changing by very much day to day.

46. Drunk driving 2012.

a) The histogram shows the distribution of drunk driving deaths.



b) The timeplot shows the change in drunk driving deaths over time.



c) The distribution of the number of drunk driving deaths is bimodal, with a cluster between 22 and 25 thousand deaths and another cluster between 16 and 17 thousand deaths. The timeplot shows that this corresponds to a rapid decrease in the drunk driving deaths in the early nineties. The number of deaths was high, then decreased dramatically. Starting at about 1995, the number of drunk driving deaths leveled off, but recently began to decrease again.

47. Assets.

a) The distribution of assets of 79 companies chosen from the *Forbes* list of the nation's top corporations is skewed so heavily to the right that the vast majority of companies have assets represented in the first bar of the histogram, 0 to 10 billion dollars. This makes meaningful discussion of center and spread impossible.

b) Re-expressing these data by, for example, logs or square roots might help make the distribution more nearly symmetric. Then a meaningful discussion of center might be possible.

48. Music library.

a) The distribution of the number of songs students had in their digital libraries is extremely skewed to the right. That makes it difficult to determine a center. The typical number of songs in a library is probably in the first bar of the histogram.

b) Re-expressing these data by, for example, logs or square roots might help make the distribution more nearly symmetric. Then a meaningful discussion of center might be possible.

49. Assets again.

a) The distribution of logarithm of assets is preferable, because it is roughly unimodal and symmetric. The distribution of the square root of assets is still skewed right, with outliers.

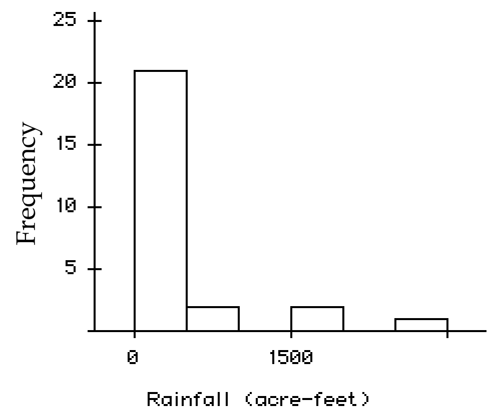
b) If $\sqrt{Assets} = 50$, then the companies assets are approximately $50^2 = 2500$ million dollars.

c) If $\log(Assets) = 3$, then the companies assets are approximately $10^3 = 1000$ million dollars.

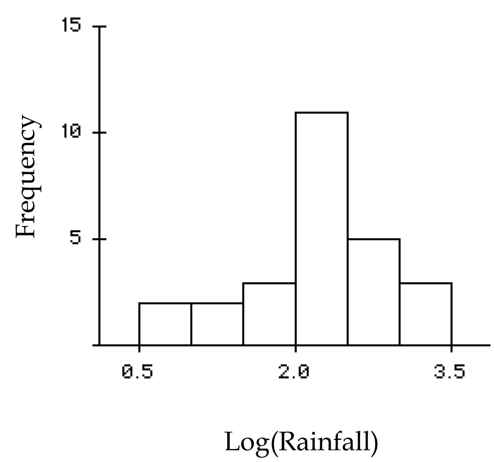
50. Rainmakers.

a) Since one acre-foot is about 320,000 gallons, these numbers are more manageable than gallons.

b) The distribution of rainfall from 26 clouds seeded with silver iodide is skewed heavily to the right, with the vast majority of clouds producing less than 500 acre-feet of rain. Several clouds produced more, with a maximum of 2745 acre-feet.



c) The distribution of log (base 10) of rainfall is much more symmetric than the distribution of rainfall. We can see that the center of the distribution is around log 2 – log 2.5 acre-feet.



d) Since the reexpressed scale is measured in log (acre-feet), we need to raise 10 to the power of the number on our scale to convert back to acre feet. For example, if a cloud in the new scale has a log (rainfall) of 2.3, we convert back to rainfall as follows:

$$\log(rainfall) = 2.3$$
$$rainfall = 10^{2.3}$$
$$rainfall = 199.5$$

The cloud produced 199.5 acre-feet of rain.

51. Stereograms.

a) The two variables discussed in the description are fusion time and treatment group.

b) Fusion time is a quantitative variable, measured in seconds. Treatment group is a categorical variable, with subjects either receiving verbal clues only, or visual and verbal clues.

c) Both groups have distributions that are skewed to the right. Generally, the Visual/Verbal group had shorter fusion times than the No/Verbal group. The median for the Visual/Verbal group was approximately the same as the lower quartile for the No/Verbal group. The No/Verbal Group also had an extreme outlier, with at least one subject whose fusion time was approximately 50 seconds. There is evidence that visual information may reduce fusion time.

52. Stereograms, revisited.

The re-expression using logarithms has a distribution that is more symmetric than the original distribution of fusion times, and the re-expression has no outliers. This symmetry makes it easier to compare the two groups.

Chapter 5 – The Standard Deviation as a Ruler and the Normal Model

Section 5.1

1. Stats test.

Gregor scored 65 points on the test.

(Or, 75 – 2(5) = 65)

$$z = \frac{y-\mu}{\sigma}$$
$$-2 = \frac{y-75}{5}$$
$$y = 65$$

2. Mensa.

According to this scale, persons with an IQ of 140 or higher are considered geniuses.

(100 + 2.5(16) = 140)

$$z = \frac{y-\mu}{\sigma}$$
$$2.5 = \frac{y-100}{16}$$
$$y = 140$$

3. Temperatures.

In January, with mean temperature 36° and standard deviation in temperature 10°, a high temperature of 55° is almost 2 standard deviations above the mean. In July, with mean temperature 74° and standard deviation 8°, a high temperature of 55° is more than two standard deviations below the mean. A high temperature of 55° is less likely to happen in July, when 55° is farther away from the mean.

4. Placement Exams.

On the French exam, the mean was 72 and the standard deviation was 8. The student's score of 82 was 10 points, or 1.25 standard deviations, above the mean. On the math exam, the mean was 68 and the standard deviation was 12. The student's score of 86 was 18 points or 1.5 standard deviations above the mean. The student did better on the math exam.

Section 5.2

5. Shipments.

a) Adding 4 ounces will affect only the median. The new median will be 68 +4 = 72 ounces, and the IQR will remain at 40 ounces.

b) Changing the units will affect both the median and IQR. The median will be 72/16 = 4.5 pounds and the IQR will be 40/16 = 2.5 pounds.

6. Hotline.

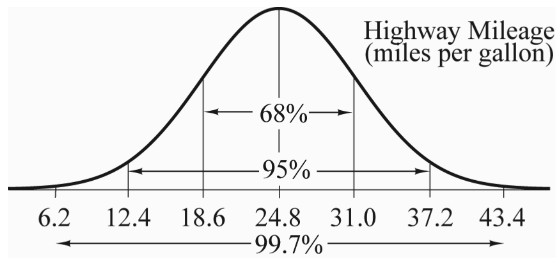
a) Changing the units will affect both the median and IQR. The median will be 4.4(60) = 264 seconds and the IQR will be 2.3(60) = 138 seconds.

b) Subtracting 24 seconds will affect only the median. The new median will be 264 – 24 = 240 seconds and the new IQR will remain 138 seconds.

Section 5.3

7. Guzzlers?

a) The Normal model for auto fuel economy is at the right.



b) Approximately 68% of the cars are expected to have highway fuel economy between 18.6 mpg and 31.0 mpg.

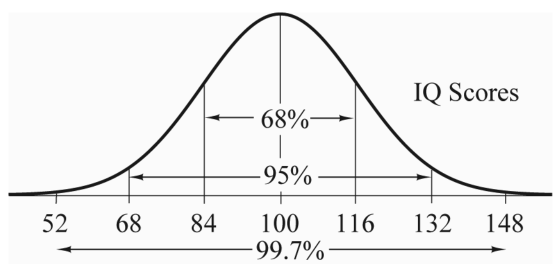
c) Approximately 16% of the cars are expected to have highway fuel economy above 31 mpg.

d) Approximately 13.5% of the cars are expected to have highway fuel economy between 31 mpg and 37 mpg.

e) The worst 2.5% of cars are expected to have fuel economy below approximately 12.4 mpg.

8. IQ.

a) The Normal model for IQ scores is at the right.



b) Approximately 95% of the IQ scores are expected to be within the interval 68 to 132 IQ points.

c) Approximately 16% of IQ scores are expected to be above 116 IQ points.

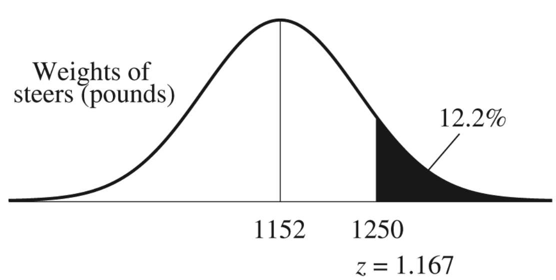
d) Approximately 13.5% of IQ scores are expected to be between 68 and 84 IQ points.

e) Approximately 2.5% of the IQ scores are expected to be above 132.

Section 5.4

9. Normal cattle.

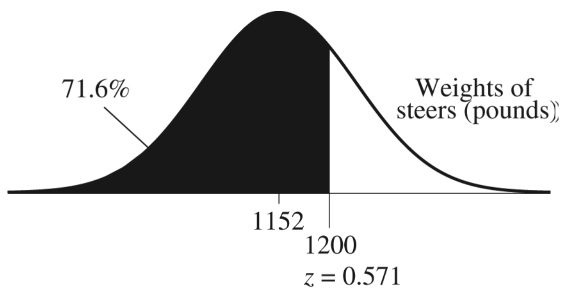
a)



$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{1250 - 1152}{84}$$
$$z \approx 1.167$$

According to the Normal model, we expect 12.2% of steers to weigh over 1250 pounds.

b)



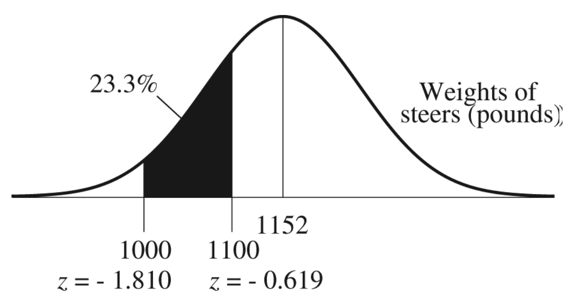
$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{1200 - 1152}{84}$$
$$z \approx 0.571$$

According to the Normal model, 71.6% of steers are expected to weigh under 1200 pounds.

c)

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{1000 - 1152}{84}$$
$$z \approx -1.810$$

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{1100 - 1152}{84}$$
$$z \approx -0.619$$

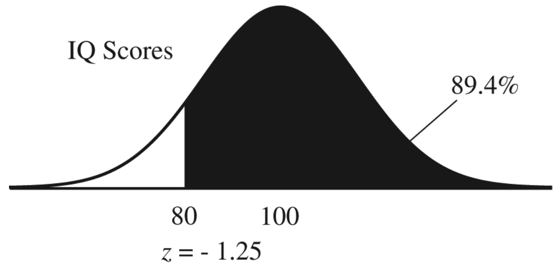


According to the Normal model, 23.3% of steers are expected to weigh between 1000 and 1100 pounds.

10. IQs revisited.

a)

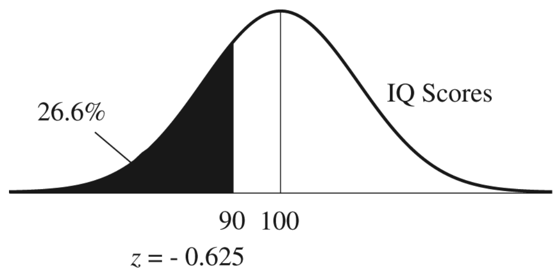
$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{80 - 100}{16}$$
$$z = -1.25$$



According to the Normal model, 89.4% of IQ scores are expected to be over 80.

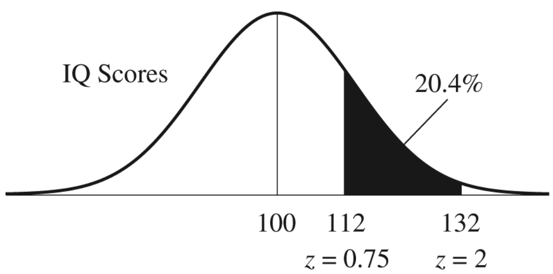
b)

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{90 - 100}{16}$$
$$z = -0.625$$



According to the Normal model, 26.6% of IQ scores are expected to be under 90.

c)



$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{112 - 100}{16}$$
$$z = 0.75$$

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{132 - 100}{16}$$
$$z = 2$$

According to the Normal model, about 20.4% of IQ scores are between 112 and 132.

Section 5.5

11. Music library.

a) The Normal probability plot is not straight, so there is evidence that the distribution of the lengths of songs in Corey's music library is not Normal.

b) The distribution of the lengths of songs in Corey's music library appears to be skewed to the right. The Normal probability plot show that the longer songs in Corey's library are much longer than the lengths predicted by the Normal model. The song lengths are much longer than their quantile scores would predict for a Normal model.

12. Wisconsin ACT math.

a) The distribution of mean ACT math scores is bimodal, so it is not approximately Normal, and 78.8% of scores are within one standard deviation of the mean. If a Normal model is useful, we would need approximately 68% of the scores within one standard deviation of the mean.

b) With Milwaukee area schools removed, the distribution of mean ACT math scores is slightly skewed, but the Normal probability plot is reasonably straight, so the Normal model is appropriate.

Chapter Exercises.

13. Payroll.

a) The distribution of salaries in the company's weekly payroll is skewed to the right. The mean salary, \$700, is higher than the median, \$500.

b) The IQR, \$600, measures the spread of the middle 50% of the distribution of salaries.

$$\begin{aligned} Q3 - Q1 &= IQR \\ Q3 &= Q1 + IQR \\ Q3 &= \$350 + \$600 \\ Q3 &= \$950 \end{aligned}$$

50% of the salaries are found between \$350 and \$950.

c) If a \$50 raise were given to each employee, all measures of center or position would increase by \$50. The minimum would change to \$350, the mean would change to \$750, the median would change to \$550, and the first quartile would change to \$400. Measures of spread would not change. The entire distribution is simply shifted up \$50. The range would remain at \$1200, the IQR would remain at \$600, and the standard deviation would remain at \$400.

d) If a 10% raise were given to each employee, all measures of center, position, and spread would increase by 10%.

Minimum = \$330	Mean = \$770	Median = \$550	Range = \$1320
IQR = \$660	First Quartile = \$385	St. Dev. = \$440	

14. Hams.

a) Range = Maximum – Minimum = 7.45 – 4.15 = 3.30 pounds
IQR = Q3 – Q1 = 6.55 – 5.6 = 0.95 pounds

b) The distribution of weights of hams is slightly skewed to the left because the mean is lower than the median and the first quartile is farther from the median than the third quartile.

c) All of the statistics are multiplied by 16 in the conversion from pounds to ounces.

Mean = 96 oz.	St. Dev. = 10.4 oz.	First Quartile = 89.6 oz.
Third Quartile = 104.8 oz.	Median = 99.2 oz.	IQR = 15.2 oz.
Range = 52.8 oz.		

d) Measures of position increase by 30 ounces. Measures of spread remain the same.

Mean = 126 oz.	St. Dev. = 10.4 oz.	First Quartile = 119.6 oz.
Third Quartile = 134.8 oz.	Median = 129.2 oz.	IQR = 15.2 oz.
Range = 52.8 oz.		

e) If a 10-pound ham were added to the distribution, the mean would change, since the total weight of all the hams would increase. The standard deviation would also increase, since 10 pounds is far away from the mean. The overall spread of the distribution would increase. The range would increase, since 10 pounds would be the new maximum. The median, quartiles, and IQR may not change. These measures are summaries of the middle 50% of the distribution, and are resistant to the presence of outliers, like the 10-pound ham.

15. SAT or ACT?

Measures of center and position (lowest score, top 25% above, mean, and median) will be multiplied by 40 and increased by 150 in the conversion from ACT to SAT by the rule of thumb. Measures of spread (standard deviation and IQR) will only be affected by the multiplication.

Lowest score = 910	Mean = 1230	Standard deviation = 120
Top 25% above = 1350	Median = 1270	IQR = 240

16. Cold U?

Measures of center and position (maximum, median, and mean) will be multiplied by $\frac{9}{5}$ and increased by 32 in the conversion from Fahrenheit to Celsius. Measures of spread (range, standard deviation, IQR) will only be affected by the multiplication.

Maximum temperature = 51.8°F	Range = 59.4°F
Mean = 33.8°F	Standard deviation = 12.6°F
Median = 35.6°F	IQR = 28.8°F

17. Stats test, part II.

A z-score of 2.20 means that your score was 2.20 standard deviations above the mean.

18. Checkup.

A z-score of –1.88 means that the boy's height was 1.88 standard deviations below the mean height of American children his age.

19. Music library again.

On the Nickel, by Tom Waits has a z-score of 1.20

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{380 - 242.4}{114.51}$$
$$z = 1.20$$

20. Windy.

a) February: 2.79

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{6.73 - 2.324}{1.577}$$
$$z = 2.79$$

June: 3.865

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{3.93 - 0.857}{0.795}$$
$$z = 3.865$$

August 3.18

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{2.53 - 0.63}{0.597}$$
$$z = 3.18$$

Since it has the highest z-score, June was the most extraordinary month.

21. Combining test scores.

The z-scores, which account for the difference in the distributions of the two tests, are 1.5 and 0 for Derrick and 0.5 and 2 for Julie. Derrick's total is 1.5 which is less than Julie's 2.5.

22. Combining scores again.

The z-scores, which account for the difference in the distributions of the two tests, are 0 and 1 for Reginald, for a total of 1. For Sara, they are 2.0 and –0.33 for a total of 1.67. While her raw score is lower, her z-score is higher.

23. Final Exams.

a) Anna's average is $\frac{83 + 83}{2} = 83$. Megan's average is $\frac{77 + 95}{2} = 86$.

Only Megan qualifies for language honors, with an average higher than 85.

b) On the French exam, the mean was 81 and the standard deviation was 5. Anna's score of 83 was 2 points, or 0.4 standard deviations, above the mean. Megan's score of 77 was 4 points, or 0.8 standard deviations below the mean.

On the Spanish exam, the mean was 74 and the standard deviation was 15. Anna's score of 83 was 9 points, or 0.6 standard deviations, above the mean. Megan's score of 95 was 21 points, or 1.4 standard deviations, above the mean.

Measuring their performance in standard deviations is the only fair way in which to compare the performance of the two women on the test.

Anna scored 0.4 standard deviations above the mean in French and 0.6 standard deviations above the mean in Spanish, for a total of 1.0 standard deviation above the mean.

Megan scored 0.8 standard deviations below the mean in French and 1.4 standard deviations above the mean in Spanish, for a total of only 0.6 standard deviations above the mean.

Anna did better overall, but Megan had the higher average. This is because Megan did very well on the test with the higher standard deviation, where it was comparatively easy to do well.

24. MP3s.

a) Standard deviation measures variability, which translates to consistency in everyday use. Batteries with a small standard deviation would be more likely to have lifespans close to their mean lifespan than batteries with a larger standard deviation.

b) RockReady batteries have a higher mean lifespan and smaller standard deviation, so they are the better battery. 8 hours is $2\frac{2}{3}$ standard deviations below the mean lifespan of RockReady and $1\frac{1}{2}$ standard deviations below the mean lifespan of DuraTunes. DuraTunes batteries are more likely to *fail* before the 8 hours have passed.

c) 16 hours is $2\frac{1}{2}$ standard deviations higher than the mean lifespan of DuraTunes, and $2\frac{2}{3}$ standard deviations higher than the mean lifespan of RockReady. Neither battery has a good chance of lasting 16 hours, but DuraTunes batteries have a greater chance than RockReady batteries.

25. Cattle.

a) A steer weighing 1000 pounds would be about 1.81 standard deviations below the mean weight.

$$z = \frac{y - \mu}{\sigma} = \frac{1000 - 1152}{84} \approx -1.81$$

b) A steer weighing 1000 pounds is more unusual. Its z-score of –1.81 is further from 0 than the 1250 pound steer's z-score of 1.17.

26. Car speeds.

a) A car going the speed limit of 20 mph would be about 1.08 standard deviations below the mean speed.

$$z = \frac{y - \mu}{\sigma} = \frac{20 - 23.84}{3.56} \approx -1.08$$

b) A car going 10 mph would be more unusual. Its z-score of –3.89 is further from 0 than the 34 mph car's z-score of 2.85.

27. More cattle.

a) The new mean would be 1152 – 1000 = 152 pounds. The standard deviation would not be affected by subtracting 1000 pounds from each weight. It would still be 84 pounds.

b) The mean selling price of the cattle would be 0.40(1152) = $460.80. The standard deviation of the selling prices would be 0.40(84) = $33.60.

28. Car speeds again.

a) The new mean would be 23.84 – 20 = 3.84 mph over the speed limit. The standard deviation would not be affected by subtracting 20 mph from each speed. It would still be 3.56 miles per hour.

b) The mean speed would be 1.609(23.84) = 38.359 kph. The speed limit would convert to 1.609(20) = 32.18 kph. The standard deviation would be 1.609(3.56) = 5.728 kph.

29. Cattle, part III.

Generally, the minimum and the median would be affected by the multiplication and subtraction. The standard deviation and the IQR would only be affected by the multiplication.

Minimum = 0.40(980) – 20 = $372.00
Median = 0.40(1140) – 20 = $436
Standard deviation = 0.40(84) = $33.60
IQR = 0.40(102) = $40.80

30. Caught speeding.

Generally, the mean and the maximum would be affected by the multiplication and addition. The standard deviation and the IQR would only be affected by the multiplication.

Mean = 100 + 10(28 – 20) = $180
Maximum = 100 +10(33 – 20) = $230
Standard deviation = 10(2.4) = $24
IQR = 10(3.2) = $32

31. Professors.

The standard deviation of the distribution of years of teaching experience for college professors must be 6 years. College professors can have between 0 and 40 (or possibly 50) years of experience. A workable standard deviation would cover most of that range of values with ±3 standard deviations around the mean. If the standard deviation were 6 months ($\frac{1}{2}$ year), some professors would have years of experience 10 or 20 standard deviations away from the mean, whatever it is. That isn't possible. If the standard deviation were 16 years, ±2 standard deviations would be a range of 64 years. That's way too high. The only reasonable choice is a standard deviation of 6 years in the distribution of years of experience.

32. Rock concerts.

The standard deviation of the distribution of the number of fans at the rock concerts would most likely be 2000. A standard deviation of 200 fans seems much too consistent. With this standard deviation, the band would be very unlikely to draw more than a 1000 fans (5 standard deviations!) above or below the mean of 21,359 fans. It seems like rock concert attendance could vary by much more than that. If a standard deviation of 200 fans is too small, then so is a standard deviation of 20 fans. 20,000 fans is too large for a likely standard deviation in attendance, unless they played several huge venues. Zero attendance is only a bit more than 1 standard deviation below the mean, although it seems very unlikely. 2000 fans is the most reasonable standard deviation in the distribution of number of fans at the concerts.

33. Small steer.

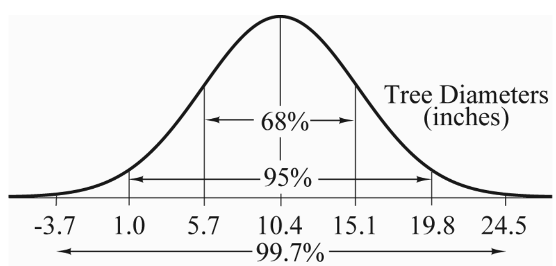
Any weight more than 2 standard deviations below the mean, or less than 1152 – 2(84) = 984 pounds might be considered unusually low. We would expect to see a steer below 1152 – 3(84) = 900 very rarely.

34. High IQ.

Any IQ more than 2 standard deviations above the mean, or more than 100 + 2 (16) = 132 might be considered unusually high. We would expect to find someone with an IQ over 100 + 3(16) = 148 very rarely.

35. Trees.

a) The Normal model for the distribution of tree diameters is at the right.



b) Approximately 95% of the trees are expected to have diameters between 1.0 inch and 19.8 inches.

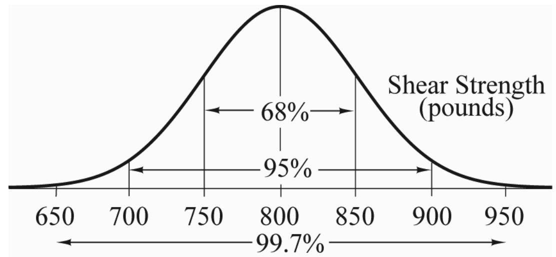
c) Approximately 2.5% of the trees are expected to have diameters less than an inch.

d) Approximately 34% of the trees are expected to have diameters between 5.7 inches and 10.4 inches.

e) Approximately 16% of the trees are expected to have diameters over 15 inches.

36. Rivets.

a) The Normal model for the distribution of shear strength of rivets is at the right.



b) 750 pounds is 1 standard deviation below the mean, meaning that the Normal model predicts that approximately 16% of the rivets are expected to have a shear strength of less than 750 pounds. These rivets are a poor choice for a situation that requires a shear strength of 750 pounds, because 16% of the rivets would be expected to fail. That's too high a percentage.

c) Approximately 97.5% of the rivets are expected to have shear strengths below 900 pounds.

d) In order to make the probability of failure very small, these rivets should only be used for applications that require shear strength several standard deviations below the mean, probably farther than 3 standard deviations. (The chance of failure for a required shear strength 3 standard deviations below the mean is still approximately 3 in 2000.) For example, if the required shear strength is 500 pounds (6 standard deviations below the mean), the chance of one of these bolts failing is approximately 1 in 1,000,000.

37. Trees, part II.

The use of the Normal model requires a distribution that is unimodal and symmetric. The distribution of tree diameters is neither unimodal nor symmetric, so use of the Normal model is not appropriate.

38. Car speeds, the picture.

The distribution of cars speeds shown in the histogram is unimodal and roughly symmetric, and the normal probability plot looks quite straight., so a normal model is appropriate.

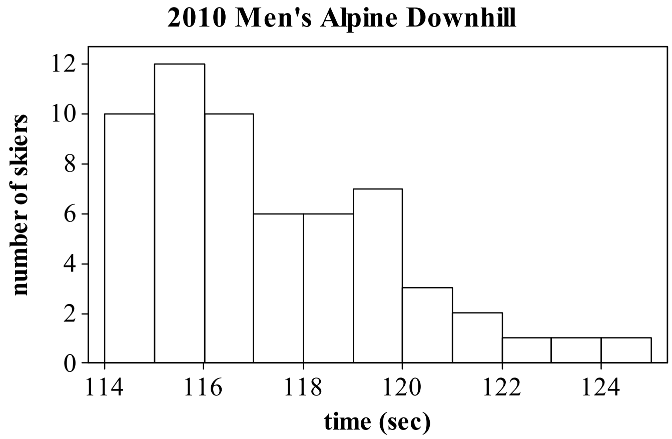
39. Winter Olympics 2010 downhill.

a) The 2010 Winter Olympics downhill times have mean of 117.34 seconds and standard deviation 2.456 seconds. 114.875 seconds is 1 standard deviation below the mean. If the Normal model is appropriate, 16% of the times should be below 99.7 seconds.

b) 8 out of 59 times (13.56%) are below 114.875 seconds.

c) The percentages in parts a and b do not agree because the Normal model is not appropriate in this situation.

d) The histogram of 2010 Winter Olympic Downhill times is skewed to the right. The Normal model is not appropriate for the distribution of times, because the distribution is not symmetric.



40. Check the model.

a) We know that 95% of the observations from a Normal model fall within 2 standard deviations of the mean. That corresponds to 23.84 – 2(3.56) = 16.72 mph and 23.84 + 2(3.56) = 30.96 mph. These are the 2.5 percentile and 97.5 percentile, respectively. According to the Normal model, we expect only 2.5% of the speeds to be below 16.72 mph, and 97.5% of the speeds to be below 30.96 mph.

b) The actual 2.5 percentile and 97.5 percentile are 16.638 and 30.976 mph, respectively. These are very close to the predicted values from the Normal model. The histogram in Exercise 38 is unimodal and roughly symmetric. It is very slightly skewed to the right and there is one outlier, but the Normal probability plot is quite straight. We should not be surprised that the approximation from the Normal model is a good one.

41. Receivers 2013.

a) Approximately 2.5% of the receivers are expected to gain more yards than 2 standard deviations above the mean number of yards gained.

b) The distribution of the number of yards gained has mean 426.28 yards and standard deviation 408.34 yards. According to the Normal model, we expect 2.5% of the receivers, or 4.525 of them, to gain more than 2 standard deviations above the mean number of yards. This means more than 426.98 + 2(408.34) = 1244 yards. In 2013, 12 receivers ran for more than 1244 yards.

c) The distribution of the number of yards run by wide receivers is skewed heavily to the right. Use of the Normal model is not appropriate for this distribution, since it is not symmetric.

42. Customer database.

a) The median of 93% is the better measure of center for the distribution of the percentage of white residents in the neighborhoods, since the distribution is skewed to the left. The median is a better summary for skewed distributions since the median is resistant to effects of the skewness, while the mean is pulled toward the tail.

b) The IQR of 17% is the better measure of spread for the distribution of the percentage of white residents in the neighborhoods, since the distribution is skewed to the left. IQR is a better summary for skewed distributions since the IQR is resistant to effects of the skewness, and the standard deviation is not.

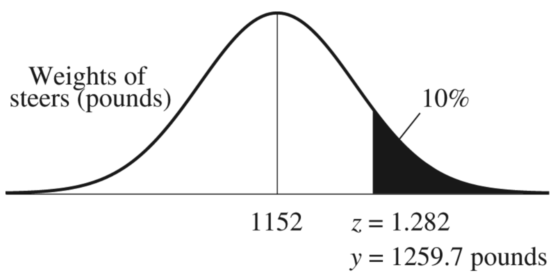
c) According to the Normal model, approximately 68% of neighborhoods are expected to have a percentage of whites within 1 standard deviation of the mean.

d) The mean percentage of whites in a neighborhood is 83.59%, and the standard deviation is 22.26%. 83.59%± 22.26% = 61.33% to 105.85%. Estimating from the graph, more than 80% of the neighborhoods have a percentage of whites greater than 61.33%.

e) The distribution of the percentage of whites in the neighborhoods is strongly skewed to the left. The Normal model is not appropriate for this distribution. There is a discrepancy between c) and d) because c) is wrong!

43. More cattle.

a)

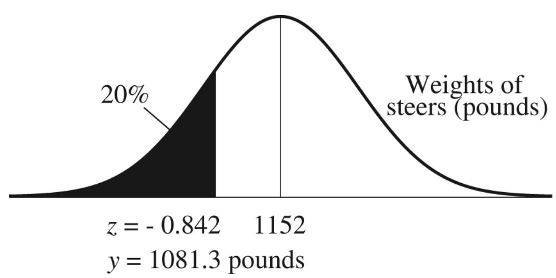


$$z = \frac{y-\mu}{\sigma}$$
$$1.282 = \frac{y-1152}{84}$$
$$y \approx 1259.7$$

According to the Normal model, the highest 10% of steer weights are expected to be above approximately 1259.7 pounds.

b)

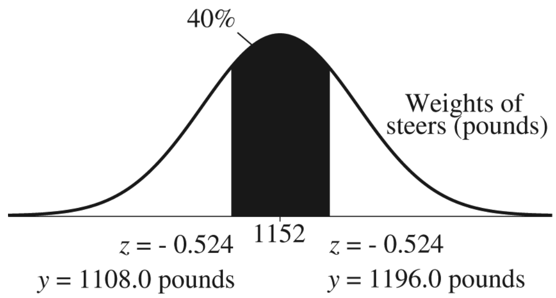
$$z = \frac{y-\mu}{\sigma}$$
$$-0.842 = \frac{y-1152}{84}$$
$$y \approx 1081.3$$



According to the Normal model, the lowest 20% of weights of steers are expected to be below approximately 1081.3 pounds.

c)

$$z = \frac{y-\mu}{\sigma}$$
$$-0.524 = \frac{y-1152}{84}$$
$$y \approx 1108.0$$

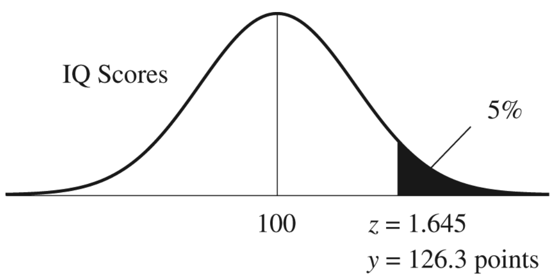


$$z = \frac{y-\mu}{\sigma}$$
$$0.524 = \frac{y-1152}{84}$$
$$y \approx 1196.0$$

According to the Normal model, the middle 40% of steer weights is expected to be between about 1108.0 pounds and 1196.0 pounds.

44. More IQs.

a)

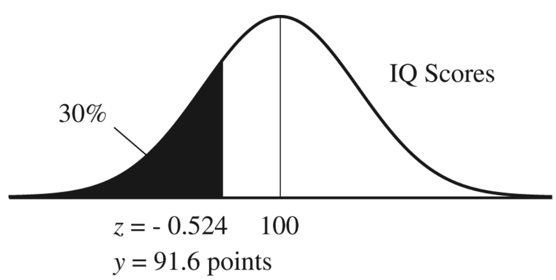


$$z = \frac{y-\mu}{\sigma}$$
$$1.645 = \frac{y-100}{16}$$
$$y \approx 126.3$$

According to the Normal model, the highest 5% of IQ scores are above about 126.3 points.

b)

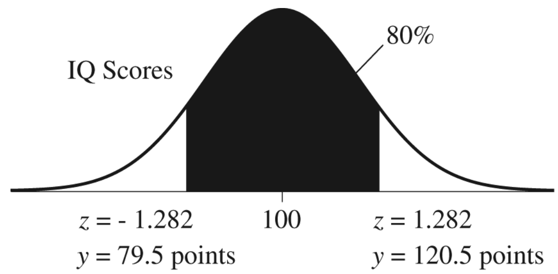
$$z = \frac{y-\mu}{\sigma}$$
$$-0.524 = \frac{y-100}{16}$$
$$y \approx 91.6$$



According to the Normal model, the lowest 30% of IQ scores are expected to be below about 91.6 points.

c)

$$z = \frac{y-\mu}{\sigma}$$
$$-1.282 = \frac{y-100}{16}$$
$$y \approx 79.5$$



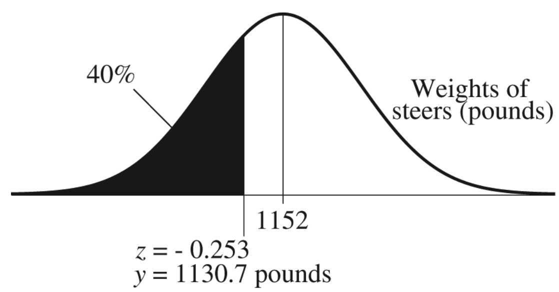
$$z = \frac{y-\mu}{\sigma}$$
$$1.282 = \frac{y-100}{16}$$
$$y \approx 120.5$$

According to the Normal model, the middle 80% of IQ scores is expected to be between 79.5 points and 120.5 points.

45. Cattle, finis.

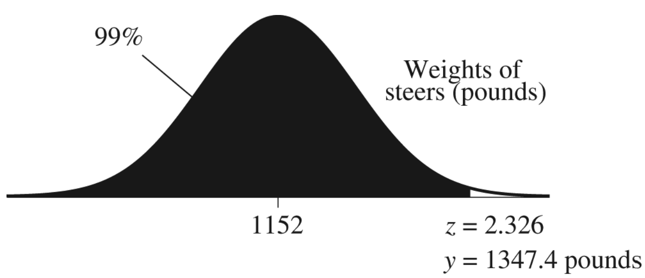
a)

$$z = \frac{y-\mu}{\sigma}$$
$$-0.253 = \frac{y-1152}{84}$$
$$y \approx 1130.7$$



According to the Normal model, the weight at the 40th percentile is 1130.7 pounds. This means that 40% of steers are expected to weigh less than 1130.7 pounds.

b)

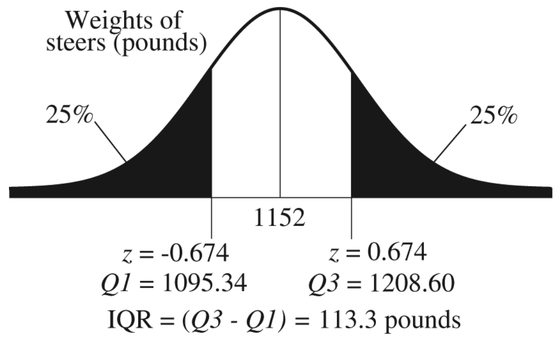


$$z = \frac{y-\mu}{\sigma}$$
$$2.326 = \frac{y-1152}{84}$$
$$y \approx 1347.4$$

According to the Normal model, the weight at the 99th percentile is 1347.4 pounds. This means that 99% of steers are expected to weigh less than 1347.4 pounds.

c)

$$z = \frac{y - \mu}{\sigma}$$
$$-0.674 = \frac{Q1 - 1152}{84}$$
$$Q1 \approx 1095.34$$



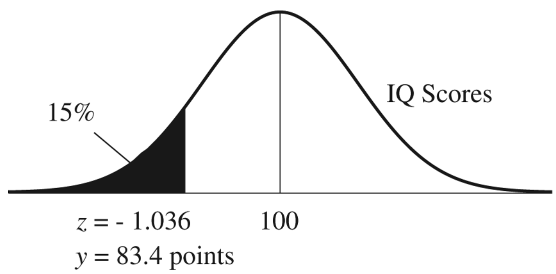
$$z = \frac{y - \mu}{\sigma}$$
$$0.674 = \frac{Q3 - 1152}{84}$$
$$Q3 \approx 1208.60$$

According to the Normal model, the IQR of the distribution of weights of Angus steers is about 113.3 pounds.

46. IQ, finis.

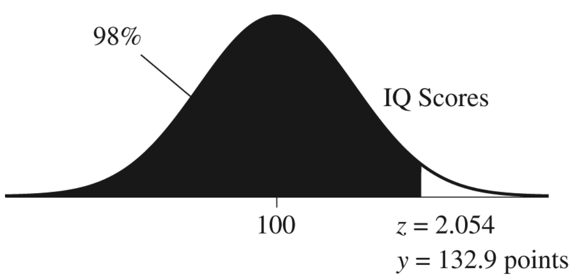
a)

$$z = \frac{y - \mu}{\sigma}$$
$$-1.036 = \frac{y - 100}{16}$$
$$y \approx 83.4$$



According to the Normal model, the 15th percentile of IQ scores is about 83.4 points. This means that we expect 15% of IQ scores to be lower than 83.4 points.

b)

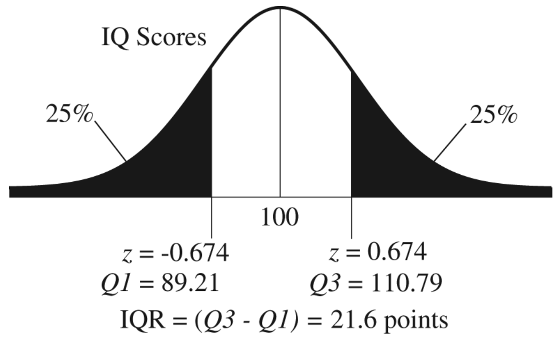


$$z = \frac{y - \mu}{\sigma}$$
$$2.054 = \frac{y - 100}{16}$$
$$y \approx 132.9$$

According to the Normal model, the 98th percentile of IQ scores is about 132.9 points. This means that we expect 98% of IQ scores to be lower than 132.9 points.

c)

$$z = \frac{y - \mu}{\sigma}$$
$$-0.674 = \frac{Q1 - 100}{16}$$
$$Q1 \approx 89.21$$

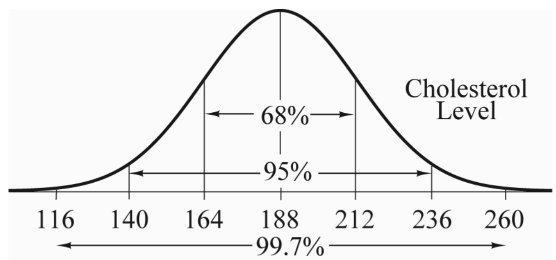


$$z = \frac{y - \mu}{\sigma}$$
$$0.674 = \frac{Q3 - 100}{16}$$
$$Q3 \approx 110.79$$

According to the Normal model, the IQR of the distribution of IQ scores is 21.6 points.

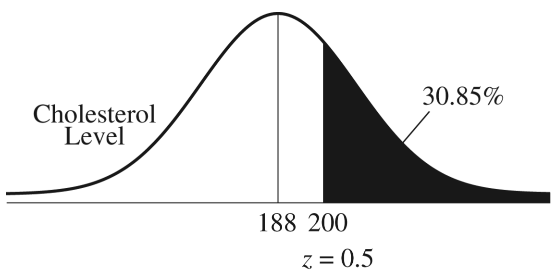
47. Cholesterol.

a) The Normal model for cholesterol levels of adult American women is at the right.



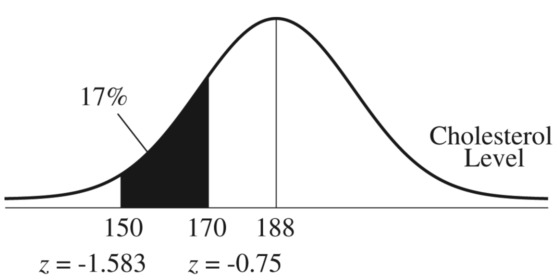
b)

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{200 - 188}{24}$$
$$z = 0.5$$



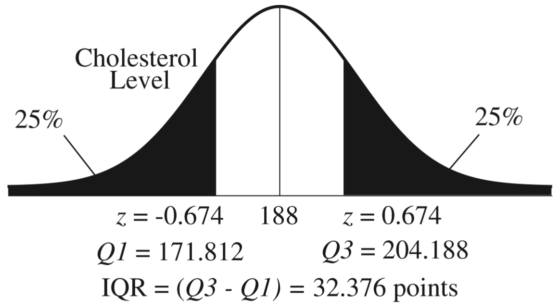
According to the Normal model, 30.85% of American women are expected to have cholesterol levels over 200.

c)



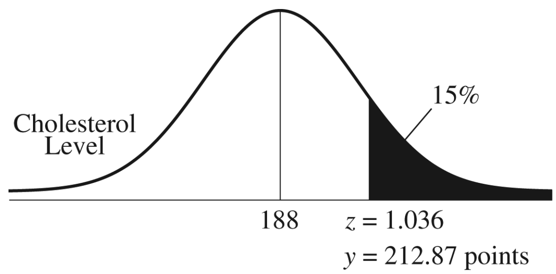
According to the Normal model, 17.00% of American women are expected to have cholesterol levels between 150 and 170.

d)



According to the Normal model, the interquartile range of the distribution of cholesterol levels of American women is approximately 32.38 points.

e)



$$z = \frac{y - \mu}{\sigma}$$

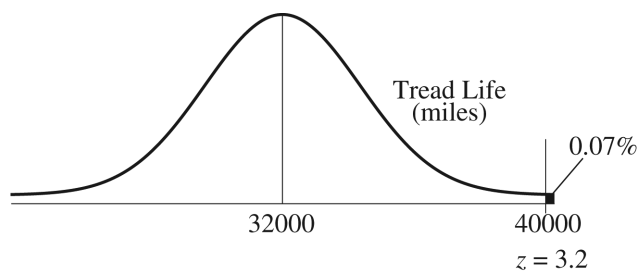
$$1.036 = \frac{y - 188}{24}$$

$$y = 212.87$$

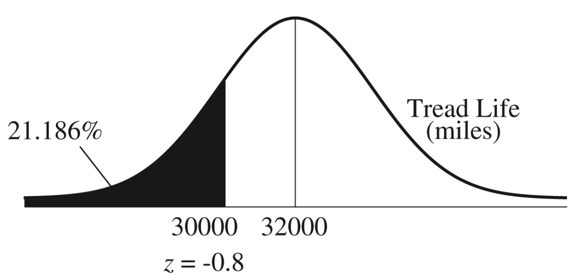
According to the Normal model, the highest 15% of women's cholesterol levels are above approximately 212.87.

48. Tires.

a) A tread life of 40,000 miles is 3.2 standard deviations above the mean tread life of 32,000. According to the Normal model, only approximately 0.07% of tires are expected to have a tread life greater than 40,000 miles. It would not be reasonable to hope that your tires lasted this long.



b)



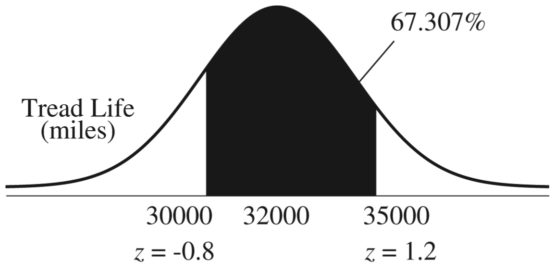
$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{30000 - 32000}{2500}$$

$$z = -0.8$$

According to the Normal model, approximately 21.19% of tires are expected to have a tread life less than 30,000 miles.

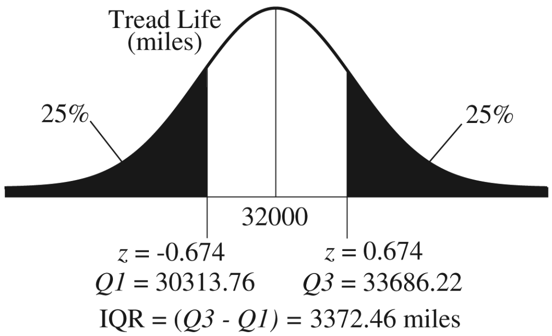
c)



$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{35000 - 32000}{2500}$$
$$z = 1.2$$

According to the Normal model, approximately 67.31% of tires are expected to last between 30,000 and 35,000 miles.

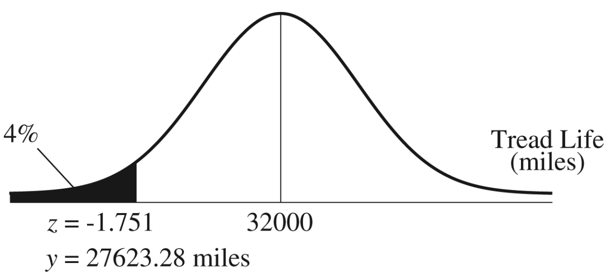
d)



$$z = \frac{y - \mu}{\sigma}$$
$$0.674 = \frac{Q3 - 32000}{2500}$$
$$Q3 = 33686.22$$

According to the Normal model, the interquartile range of the distribution of tire tread life is expected to be 3372.46 miles.

e)



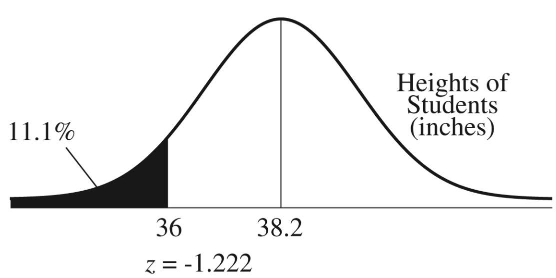
$$z = \frac{y - \mu}{\sigma}$$
$$-1.751 = \frac{y - 32000}{2500}$$
$$y = 27623.28$$

According to the Normal model, 1 of every 25 tires is expected to last less than 27,623.28 miles. If the dealer is looking for a round number for the guarantee, 27,000 miles would be a good tread life to choose.

49. Kindergarten.

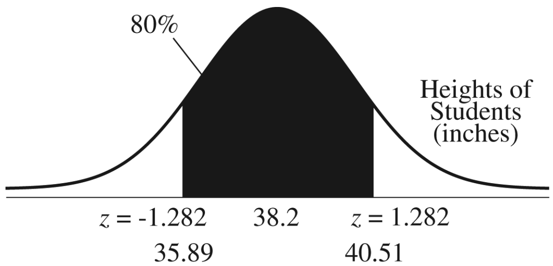
a)

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{36 - 38.2}{1.8}$$
$$z = -1.222$$



According to the Normal model, approximately 11.1% of kindergarten kids are expected to be less than three feet (36 inches) tall.

b)

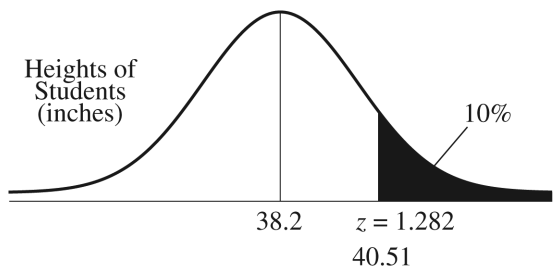


$$z = \frac{y - \mu}{\sigma} \qquad -1.282 = \frac{y_1 - 38.2}{1.8} \qquad y_1 = 35.89$$

$$z = \frac{y - \mu}{\sigma} \qquad 1.282 = \frac{y_2 - 38.2}{1.8} \qquad y_2 = 40.51$$

According to the Normal model, the middle 80% of kindergarten kids are expected to be between 35.89 and 40.51 inches tall. (The appropriate values of $z = \pm 1.282$ are found by using right and left tail percentages of 10% of the Normal model.)

c)

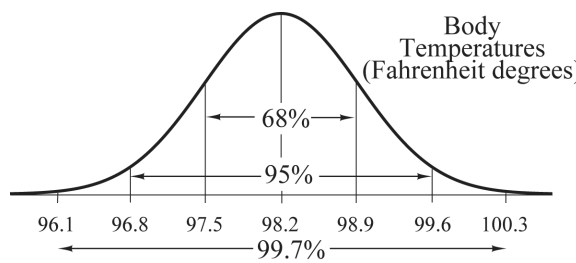


$$z = \frac{y - \mu}{\sigma} \qquad 1.282 = \frac{y - 38.2}{1.8} \qquad y = 40.51$$

According to the Normal model, the tallest 10% of kindergarteners are expected to be at least 40.51 inches tall.

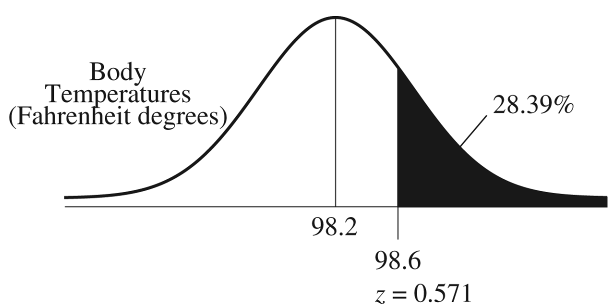
50. Body temperatures.

a) According to the Normal model (and based upon the 68-95-99.7 rule), 95% of people's body temperatures are expected to be between 96.8° and 99.6°. Virtually all people (99.7%) are expected to have body temperatures between 96.1° and 100.3°.



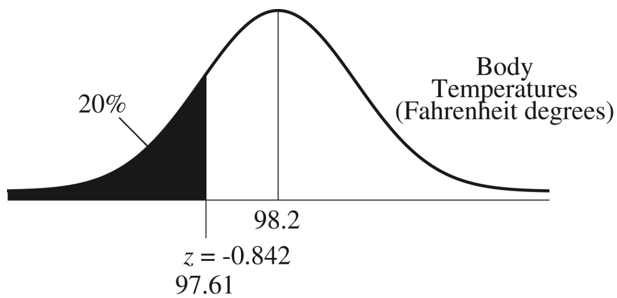
b)

$$z = \frac{y - \mu}{\sigma} \qquad z = \frac{98.6 - 98.2}{0.7} \qquad z = 0.571$$



According to the Normal model, approximately 28.39% of people are expected to have body temperatures above 98.6°.

c)

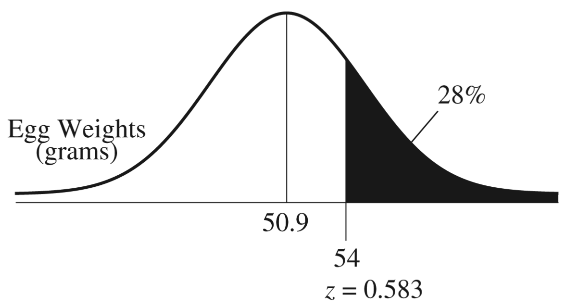


$$z = \frac{y - \mu}{\sigma}$$
$$-0.842 = \frac{y - 98.2}{0.7}$$
$$y = 97.61$$

According to the Normal model, the coolest 20% of all people are expected to have body temperatures below 97.6°.

51. Eggs.

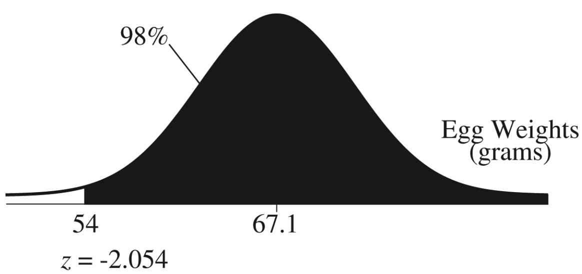
a)



$$z = \frac{y - \mu}{\sigma}$$
$$0.583 = \frac{54 - 50.9}{\sigma}$$
$$0.583\sigma = 3.1$$
$$\sigma = 5.317$$

According to the Normal model, the standard deviation of the egg weights for young hens is expected to be 5.3 grams.

b)



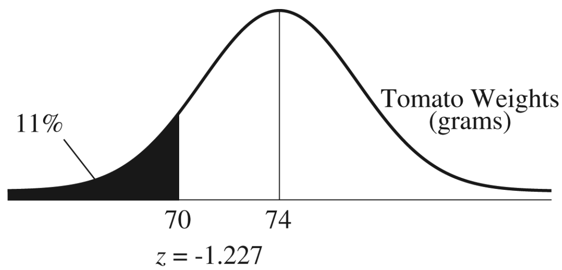
$$z = \frac{y - \mu}{\sigma}$$
$$-2.054 = \frac{54 - 67.1}{\sigma}$$
$$-2.054\sigma = -13.1$$
$$\sigma = 6.377$$

According to the Normal model, the standard deviation of the egg weights for older hens is expected to be 6.4 grams.

c) The younger hens lay eggs that have more consistent weights than the eggs laid by the older hens. The standard deviation of the weights of eggs laid by the younger hens is lower than the standard deviation of the weights of eggs laid by the older hens.

52. Tomatoes.

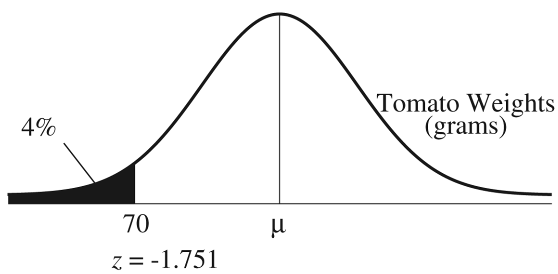
a)



$$z = \frac{y - \mu}{\sigma}$$
$$-1.227 = \frac{70 - 74}{\sigma}$$
$$-1.227\sigma = -4$$
$$\sigma = 3.260$$

According to the Normal model, the standard deviation of the weights of Roma tomatoes now being grown is 3.26 grams.

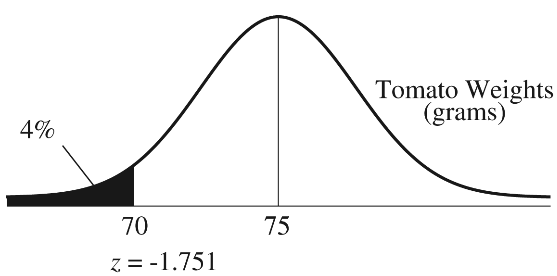
b)



$$z = \frac{y - \mu}{\sigma}$$
$$-1.751 = \frac{70 - \mu}{3.260}$$
$$-5.708 = 70 - \mu$$
$$\mu = 75.71$$

According to the Normal model, the target mean weight for the tomatoes should be 75.71 grams.

c)



$$z = \frac{y - \mu}{\sigma}$$
$$-1.751 = \frac{70 - 75}{\sigma}$$
$$\sigma = 2.856$$

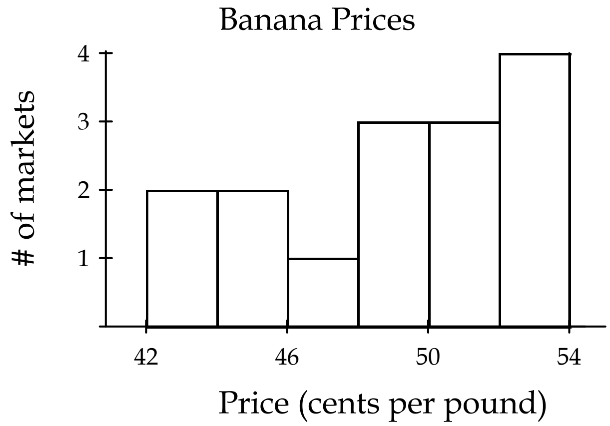
According to the Normal model, the standard deviation of these new Roma tomatoes is expected to be 2.86 grams.

d) The weights of the new tomatoes have a lower standard deviation than the weights of the current variety. The new tomatoes have more consistent weights.

Review of Part I – Exploring and Understanding Data

1. **Bananas.**

   a) A histogram of the prices of bananas from 15 markets, as reported by the USDA, appears at the right.



   b) The distribution of banana prices is skewed to the left, so median and IQR are appropriate measures of center and spread. Median = 49 cents per pound
   IQR = 6 cents per pound

   c) The distribution of the prices of bananas from 15 markets, as reported by the USDA, is unimodal and skewed to the left. The center of the distribution is approximately 50 cents, with the lowest price 42 cents per pound and the highest price 53 cents per pound.

2. **Prenatal care.**

   a) $\frac{5.4+3.9+6.1}{3} = 5.1\overline{3}$, so the overall rate of 5.1 deaths per thousand live births is equal to the average of the rates for Intensive, Adequate, and Inadequate prenatal care, when rounded to the nearest tenth. There is no reason this should be the case unless the number of women receiving each type of prenatal care is approximately the same.

   b) Yes, the results indicate (but do not prove) that adequate prenatal care is important for pregnant women. The mortality rate is quite a bit lower for women with adequate care than for other women.

   c) No, the results do not suggest that a woman pregnant with twins should be wary of seeking too much medical care. Intensive care is given for emergency conditions. The data do not suggest that the level of care is the cause of the higher mortality.

3. **Singers by parts.**

   a) The two statistics could be the same if there were many sopranos of that height.

   b) The distribution of heights of each voice part is roughly symmetric. The basses and tenors are generally taller than the altos and sopranos, with the basses being slightly taller than the tenors. The sopranos and altos have about the same median height. Heights of basses and sopranos are more consistent than altos and tenors.

4. Dialysis.

There are only three patients currently on dialysis. With so few patients, no display is needed. We know that one patient has had his or her toes amputated and that two patients have developed blindness. What we don't know is whether or not the patient that has had his or her toes amputated has also developed blindness. Even if we wanted to, we do not have enough information to make an appropriate display.

5. Beanstalks.

a) The greater standard deviation for the distribution of women's heights means that their heights are more variable than the heights of men.

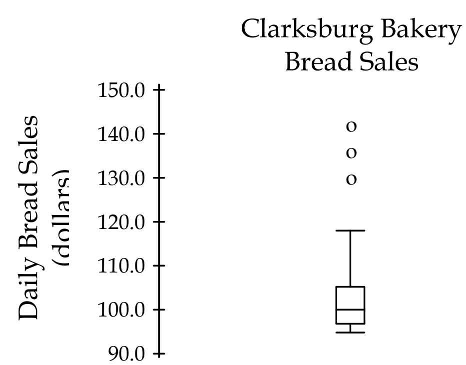
b) The z-score for women to qualify is 2.4 compared with 1.75 for men, so it is harder for women to qualify.

6. Bread.

a) The distribution of the number of loaves sold each day in the last 100 days at the Clarksburg Bakery is unimodal and skewed to the right. The mode is near 100, with the majority of days recording fewer than 120 loaves sold. The number of loaves sold ranges from 95 to 145.

b) The mean number of loaves sold will be higher than the median number of loaves sold, since the distribution of sales is skewed to the right. The mean is sensitive to this skewness, while the median is resistant.

c) Create a boxplot with quartiles at 97 and 105.5, median at 100. The IQR is 8.5 so the upper fence is at 105.5 + 1.5(8.5) = 118.25. There are several high outliers. There are no low outliers because the min at 95 lies well within the lower fence at 97 – 1.5(8.5) = 84.25.
One possible boxplot is at the right.



d) The distribution of daily bread sales is not symmetric, but rather skewed to the right. The Normal model is not appropriate for this distribution. No conclusions can be drawn.

7. State University.

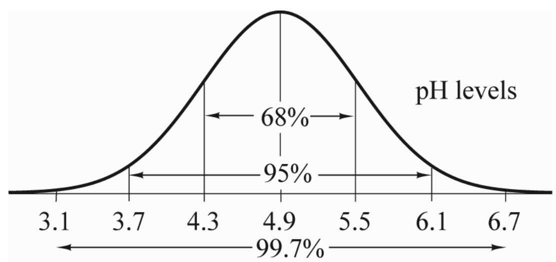
a) *Who* – Local residents near State University. *What* – Age, whether or not the respondent attended college, and whether or not the respondent had a favorable opinion of State University. *When* – Not specified. *Where* – Region around State University. *Why* – The information will be included in a report to the University's directors. *How* – 850 local residents were surveyed by phone.

b) There is one quantitative variable, age, probably measured in years. There are two categorical variables, college attendance (yes or no), and opinion of State University (favorable or unfavorable).

c) There are several problems with the design of the survey. No mention is made of a random selection of residents. Furthermore, there may be a non-response bias present. People with an unfavorable opinion of the university may hang up as soon as the staff member identifies himself or herself. Also, response bias may be introduced by the interviewer. The responses of the residents may be influenced by the fact that employees of the university are asking the questions. There may be greater percentage of favorable responses to the survey than truly exist.

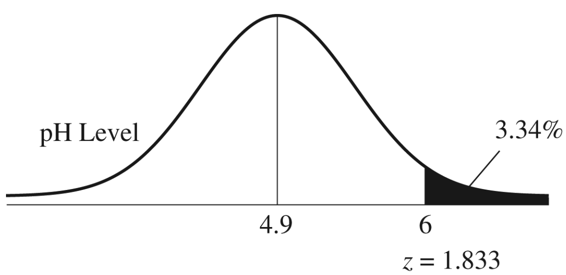
8. Acid Rain.

a) The Normal model for pH level of rainfall in the Shenandoah Mountains is at the right.



b)

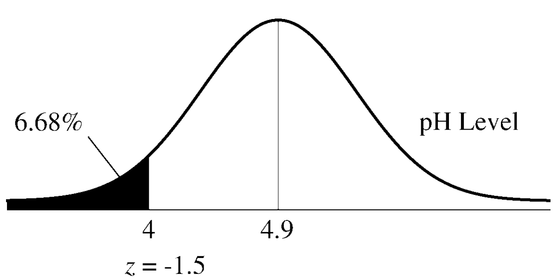
$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{6 - 4.9}{0.6}$$
$$z = 1.833$$



According to the Normal model, 3.34% of the rainstorms are expected to produce rainfall with pH levels above 6.

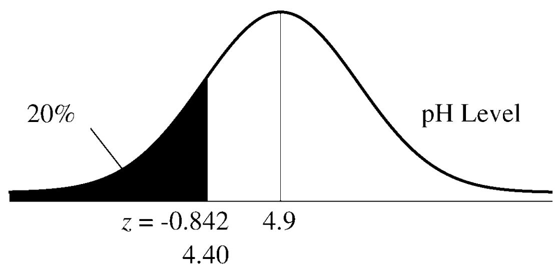
c)

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{4 - 4.9}{0.6}$$
$$z = -1.5$$



According to the Normal model, 6.68% of rainstorms are expected to produce rainfall with pH levels below 4.

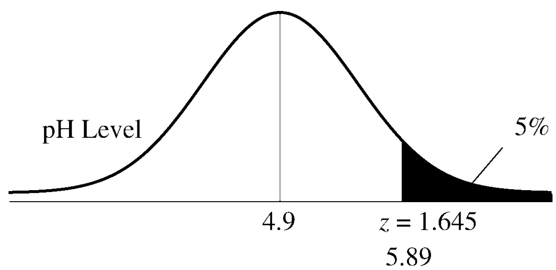
d)



$$z = \frac{y - \mu}{\sigma}$$
$$-0.842 = \frac{y - 4.9}{0.6}$$
$$y = 4.40$$

According to the Normal model, the most acidic 20% of storms have pH below 4.40.

e)

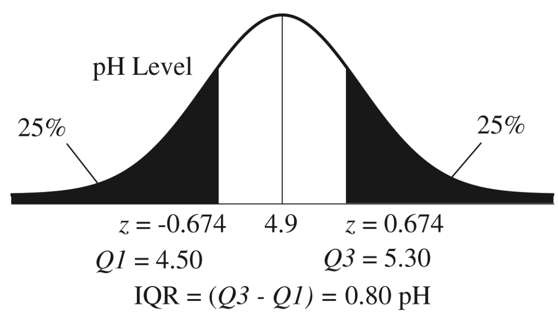


$$z = \frac{y - \mu}{\sigma}$$
$$1.645 = \frac{y - 4.9}{0.6}$$
$$y = 5.89$$

According to the Normal model, the least acidic 5% of storms have pH above 5.89.

f)

$$z = \frac{y - \mu}{\sigma}$$
$$-0.674 = \frac{Q1 - 4.9}{0.6}$$
$$Q1 = 4.50$$



$$z = \frac{y - \mu}{\sigma}$$
$$0.674 = \frac{Q3 - 4.9}{0.6}$$
$$Q3 = 5.30$$

According to the Normal model, the IQR of the pH levels of the rainstorms is 0.80.

9. Fraud detection.

a) Even though they are numbers, the SIC code is a categorical variable. A histogram is a quantitative display, so it is not appropriate.

b) The Normal model will not work at all. The Normal model is for modeling distributions of unimodal and symmetric quantitative variables. SIC code is a categorical variable.

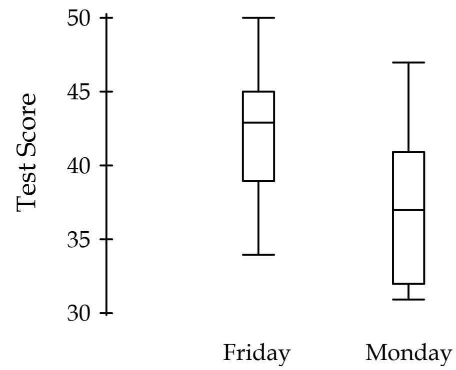
10. Streams.

a) Stream Name – categorical; Substrate – categorical; pH – quantitative; Temperature – quantitative (°C); BCI – quantitative.

b) Substrate is a categorical variable, so a pie chart or a bar chart would be a useful display.

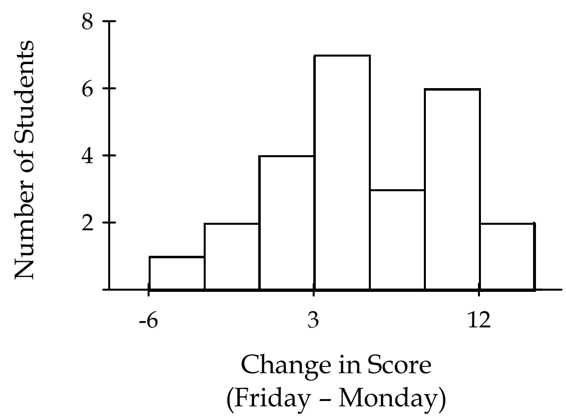
11. Cramming.

a) Comparitive boxplots of the distributions of Friday and Monday scores are at the right.



b) The distribution of scores on Friday was generally higher by about 5 points. Students fared worse on Monday after preparing for the test on Friday. The spreads are about the same, but the scores on Monday are slightly skewed to the right.

c) A histogram of the distribution of change in test score is at the right.



d) The distribution of changes in score is roughly unimodal and symmetric, and is centered near 4 points. Changes ranged from a student who scored 5 points higher on Monday, to two students who scored 13 and 14 points higher on Friday. Only three students did better on Monday.

12. e-Books.

The conclusion is not sound. Many residents who have read a book will have also read an e-book. (In Chapter 12, we will say that these percentages may not be added because they are not disjoint.)

13. Let's play cards.

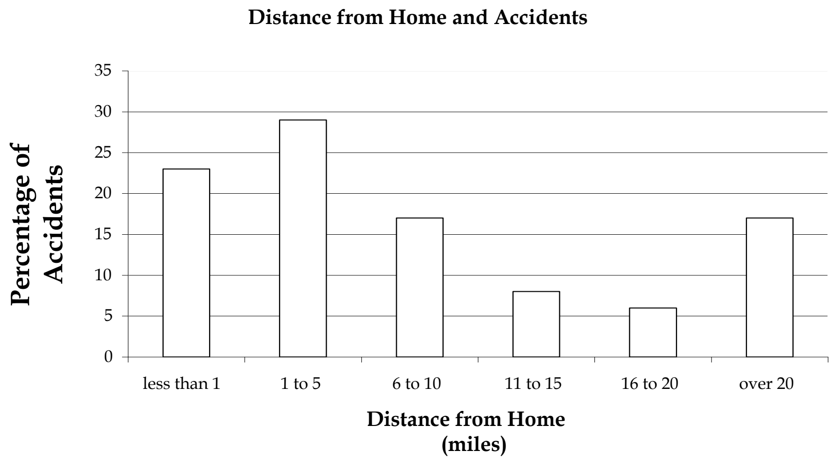
a) Suit is a categorical variable.

b) In the game of Go Fish, the denomination is not ordered. Numbers are merely matched with one another. You may have seen children's Go Fish decks that have symbols or pictures on the cards instead of numbers. These work just fine.

c) In the game of Gin Rummy, the order of the cards is important. During the game, ordered "runs" of cards are assembled (with Jack = 11, Queen = 12, King = 13), and at the end of the hand, points are totaled from the denomination of the card (face cards = 10 points). However, even in Gin Rummy, the denomination of the card sometimes behaves like a categorical variable. When you are collecting 3s, for example, order doesn't matter.

14. Accidents.

a)



The distances from home are organized in categories, so a bar chart is provided at the right. A pie chart would also be useful, since the percentages represent parts of a whole.

b) A greater percentage of accidents happen close to home than further away. But it is likely that people drive more miles close to home as well. These data do not indicate that driving near home is dangerous.

15. Hard water.

a) The variables in this study are both quantitative. Annual mortality rate for males is measured in deaths per 100,000. Calcium concentration is measured in parts per million.

b) The distribution of calcium concentration is skewed right, with many towns having concentrations below 25 ppm. The rest of the towns have calcium concentrations which are distributed in a fairly uniform pattern from 25 ppm to 100 ppm, tapering off to a maximum concentration around 150 ppm.
The distribution of mortality rates is unimodal and symmetric, with center approximately 1500 deaths per 100,000. The distribution has a range of 1000 deaths per 100,000, from 1000 to 2000 deaths per 100,000.

16. Hard water II.

a) The overall mean mortality rate is $\dfrac{34(1631.59)+27(1388.85)}{34+27}=1524.15$ deaths per 100,000.

b) The distribution of mortality rates for the towns north of Derby is generally higher than the distribution of mortality rates for the towns south of Derby. Fully half of the towns south of Derby have mortality rates lower than any of the towns north of Derby. A quarter of the northern towns have rates higher than any of the Southern towns.

17. Seasons.

a) The two histograms have different horizontal and vertical scales. This makes a quick comparison impossible.

b) The center of the distribution of average temperatures in January in is the low 30s, compared to a center of the distribution of July temperatures in the low 70s. The January distribution is also much more spread out than the July distribution. The range is over 50 degrees in January, compared to a range of over 20 degrees in July. The distribution of average temperature in January is skewed slightly to the right, while the distribution of average temperature in July is roughly symmetric.

c) The distribution of difference in average temperature (July – January) for 60 large U.S. cities is slightly skewed to the left, with median at approximately 44 degrees. There are several low outliers, cities with very little difference between their average July and January temperatures. The single high outlier is a city with a large difference in average temperature between July and January. The middle 50% of differences are between approximately 38 and 46 degrees.

18. Old Faithful.

The distribution of time gaps between eruptions of Old Faithful is bimodal. A large cluster of time gaps has a mode at approximately 80 minutes and a slightly smaller cluster of time gaps has a mode at approximately 50 minutes. The distribution around each mode is fairly symmetric.

19. Old Faithful?

a) The distribution of duration of the 222 eruptions is bimodal, with modes at approximately 2 minutes and 4.5 minutes. The distribution is fairly symmetric around each mode.

b) The bimodal shape of the distribution of duration of the 222 eruptions suggests that there may be two distinct groups of eruption durations. Summary statistics would try to summarize these two groups as a single group, which wouldn't make sense.

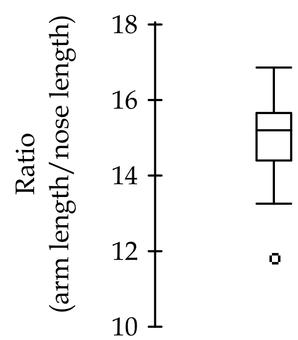
c) The intervals between eruptions are generally longer for long eruptions than the intervals for short eruptions. Over 75% of the short eruptions had intervals of approximately 60 minutes or less, while almost all of the long eruptions had intervals of more than 60 minutes.

20. Teen drivers 2013.

The chance of being a male involved in a fatal accident is different from the chance of being male. Thus those variables are not independent. Also the probability of being drunk in an accident is not the same for males and females, so those variables are not independent.

21. Liberty's nose.

a) The distribution of the ratio of arm length to nose length of 18 girls in a statistics class is unimodal and roughly symmetric, with center around 15. There is one low outlier, a ratio of 11.8. A boxplot is provided at the right. A histogram or stemplot is also an appropriate display.

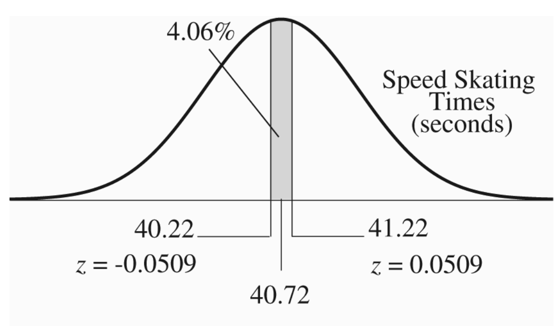


b) In the presence of an outlier, the 5-number summary is the appropriate choice for summary statistics. The 5-number summary is 11.8, 14.4, 15.25, 15.7, 16.9. The IQR is 1.3.

c) The ratio of 9.3 for the Statue of Liberty is very low, well below the lowest ratio in the statistics class, 11.8, which is already a low outlier. Compared to the girls in the statistics class, the Statue of Liberty's nose is very long in relation to her arm.

22. Winter Olympics 2010 speed skating.

a) According to the Normal model, we expect about 4.06% of times to be within 0.5 seconds of the mean of 40.72 seconds.



b) In the actual data set, none of the times are within 0.5 seconds of the mean time.

c) The Normal model is not appropriate, at least not using this mean and standard deviation. There is an extremely high outlier, a skater that took over twice as long to finish as any other skater.

23. Sample.

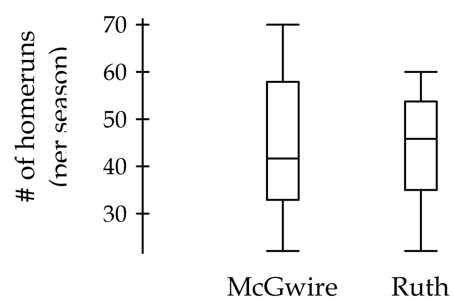
Overall, the follow-up group was insured only 11.1% of the time as compared to 16.6% for the not traced group. At first, it appears that group is associated with presence of health insurance. But for blacks, the follow-up group was quite close (actually slightly higher) in terms of being insured: 8.9% to 8.7%. The same is true for whites. The follow-up group was insured 83.3% of the time, compared to 82.5% of the not traced group. When broken down by race, we see that group is not associated with presence of health insurance for either race. This demonstrates Simpson's paradox, because the overall percentages lead us to believe that there is an association between health insurance and group, but we see the truth when we examine the situation more carefully.

24. Sluggers.

a) The 5-number summary for McGwire's career is 3, 25.5, 36, 50.5, 70. The IQR is 25.

b) By the outlier test, 1.5(IQR) = 37.5. There are no homerun totals less than Q1 – 37.5 or greater than Q3 + 37.5. Technically, there are no outliers. However, the seasons in which McGwire hit fewer than 22 homeruns stand out as a separate group.

c) Parallel boxplots comparing the homerun careers of Mark McGwire and Babe Ruth are at the right.



d) Without the injured seasons, McGwire and Ruth's home run production distributions look similar. (Note: Ruth's seasons as a pitcher were not included.) Ruth's median is a little higher, and he was a little more consistent (less spread), but McGwire had the two highest season totals.

e) A side-by-side stem-and-leaf display of the homerun careers of McGwire and Ruth is at the right.

Ruth		McGwire
	7	0
0	6	5
944	5	28
9766611	4	29
54	3	2399
52	2	2

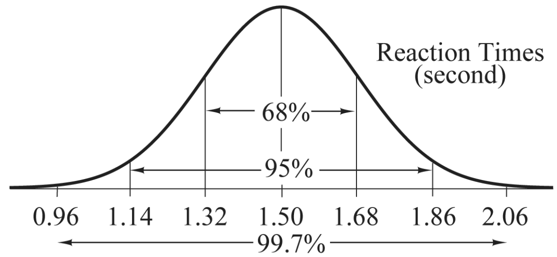
7|0 = 70 homeruns per season

f) From the stem-and-leaf display, we can see that Ruth was much more consistent. During most of his seasons, Ruth had homerun totals in the 40s and 50s. The shape of McGwire's distribution of homeruns is revealed to be skewed to the right.

25. Be quick!

a) The Normal model for the distribution of reaction times is at the right.

b) The distribution of reaction times is unimodal and symmetric, with mean 1.5 seconds, and standard deviation 0.18 seconds. According to the Normal model, 95% of drivers are expected to have reaction times between 1.14 seconds and 1.86 seconds.

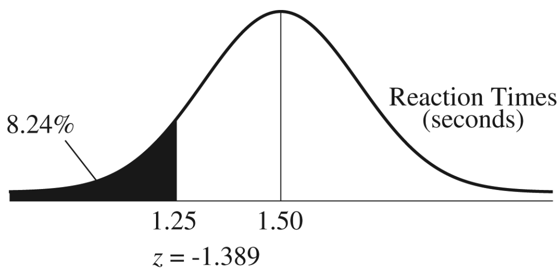


c)

$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{1.25 - 1.50}{0.18}$$

$$z = -1.389$$



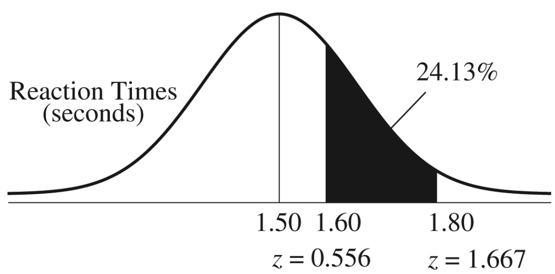
According to the Normal model, 8.24% of drivers are expected to have reaction times below 1.25 seconds.

d)

$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{1.6 - 1.5}{0.18}$$

$$z = 0.556$$



$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{1.8 - 1.5}{0.18}$$

$$z = 1.667$$

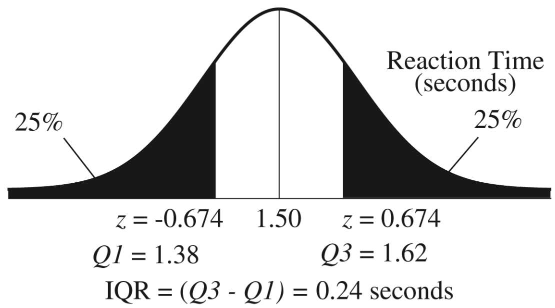
According to the Normal model, 24.13% of drivers are expected to have reaction times between 1.6 seconds and 1.8 seconds.

e)

$$z = \frac{y - \mu}{\sigma}$$

$$-0.674 = \frac{Q1 - 1.5}{0.18}$$

$$Q1 = 1.38$$



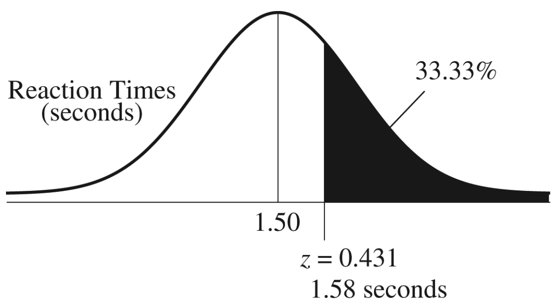
$$z = \frac{y - \mu}{\sigma}$$

$$0.674 = \frac{Q3 - 1.5}{0.18}$$

$$Q3 = 1.62$$

According to the Normal model, the interquartile range of the distribution of reaction times is expected to be 0.24 seconds.

f)



$$z = \frac{y - \mu}{\sigma}$$

$$0.431 = \frac{y - 1.5}{0.18}$$

$$y = 1.58$$

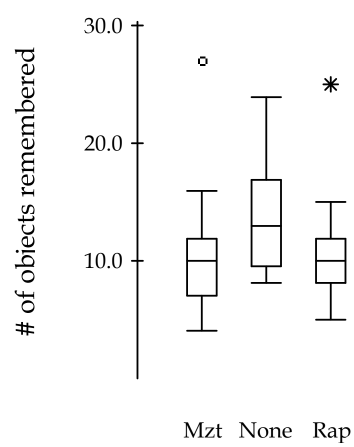
According to the Normal model, the slowest 1/3 of all drivers are expected to have reaction times of 1.58 seconds or more. (Remember that a high reaction time is a SLOW reaction time!)

26. Music and memory.

a) *Who* – 62 people. *What* – Type of music and number of objects remembered correctly. *When* – Not specified. *Where* – Not specified. *Why* – Researchers hoped to determine whether or not music affects memorization ability. *How* – Data were gathered in a completely randomized experiment.

b) Type of music (Rap, Mozart, or None) is a categorical variable. Number of items remembered is a quantitative variable.

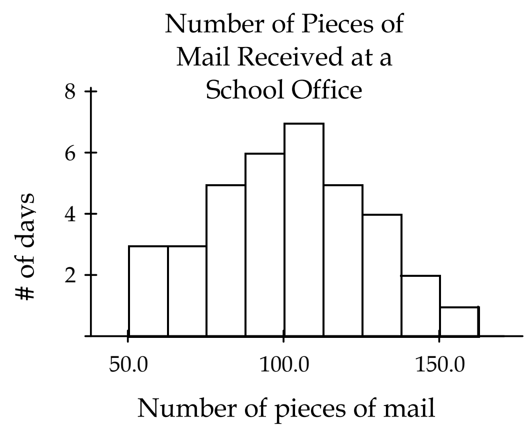
c) Accurate boxplots cannot be constructed, because we do not have all the data. By performing outlier tests, we can determine that there are no low outliers (the minimums are all within the fences), but the Rap group and the Mozart group each have at least one high outlier (the maximum in each group is above the fence). Some possible boxplots are at the right.



d) Mozart and Rap had very similar distributions of the number if objects remembered. The scores for None are, if anything, slightly higher than the other two groups. It is clear that groups listening to music (Rap or Mozart) did **not** score higher than those who listened to None.

27. Mail.

a) A histogram of the number of pieces of mail received at a school office is at the right.



b) Since the distribution of number of pieces of mail is unimodal and symmetric, the mean and standard deviation are appropriate measures of center and spread. The mean number of pieces of mail is 100.25, and the standard deviation is 25.54 pieces.

c) The distribution of the number of pieces of mail received at the school office is unimodal and symmetric, with mean 100.25 and standard deviation 25.54. The lowest number of pieces received in a day was 52 and the highest was 151.

d) 23 of the 36 days (64%) had a number of pieces of mail received within one standard deviation of the mean, or within the interval 74.71 - 125.79. This is fairly close to the 68% predicted by the Normal model. The Normal model may be useful for modeling the number of pieces received by this school office.

28. Birth Order.

a) There were 223 students. Of these, 113, or 50.7%, were oldest or only children.

b) There were 43 Humanities majors. Of these, 15, or 34.9%, were oldest or only children.

c) There were 113 oldest children. Of these, 15, or 13.3%, were Humanities majors.

d) There were 223 students. Of these, 15, or 6.7%, were oldest children majoring in Humanities.

29. Herbal medicine.

a) *Who* – 100 customers. *What* – Researchers asked whether or not the customer had taken the cold remedy and had customers rate the effectiveness of the remedy on a scale from 1 to 10. *When* – Not specified. *Where* – Store where natural health products are sold. *Why* – The researchers were from the Herbal Medicine Council, which sounds suspiciously like a group that might be promoting the use of herbal remedies. *How* – Researchers conducted personal interviews with 100 customers. No mention was made of any type of random selection.

b) "Have you taken the cold remedy?" is a categorical variable. Effectiveness on a scale of 1 to 10 is a categorical variable, as well, with respondents rating the remedy by placing it into one of 10 categories.

c) Very little confidence can be placed in the Council's conclusions. Respondents were people who already shopped in a store that sold natural remedies. They may be pre-disposed to thinking that the remedy was effective. Furthermore, no attempt was made to randomly select respondents in a representative manner. Finally, the Herbal Medicine Council has an interest in the success of the remedy.

30. Birth order revisited.

a) Overall, 25.6% of the students were Math/Science majors, 41.7% were Agriculture majors, 19.3% were Humanities majors, and 13.5% had other majors.

b) Of the oldest children, 30.1% of the students were Math/Science majors, 46.0% were Agriculture majors, 13.3% were Humanities majors, and 10.6% had other majors.

c) Of the second born children, 20.2% of the students were Math/Science majors, 39.1% were Agriculture majors, 24.7% were Humanities majors, and 15.9% had other majors.

d) No, college major does not appear to be independent of birth order. Oldest children are more likely than second born children to major in Math/Science (30.1% to 20.1%), while second born children are more likely than oldest children to major in Humanities (24.7% to 13.3%).

31. Engines.

a) The count of cars is 38.

b) The mean displacement is higher than the median displacement, indicating a distribution of displacements that is skewed to the right. There are likely to be several very large engines in a group that consists of mainly smaller engines.

c) Since the distribution is skewed, the median and IQR are useful measures of center and spread. The median displacement is 148.5 cubic inches and the IQR is 126 cubic inches.

d) Your neighbor's car has an engine that is bigger than the median engine, but 227 cubic inches is smaller than the third quartile of 231, meaning that at least 25% of cars have a bigger engine than your neighbor's car. Don't be impressed!

e) Using the Outlier Rule (more than 1.5 IQRs beyond the quartiles) to find the fences:

Upper Fence: Q3 + 1.5(IQR) = 231 + 1.5(126) = 420 cubic inches.

Lower Fence: Q1 – 1.5(IQR) = 105 – 1.5(126) = - 84 cubic inches.

Since there are certainly no engines with negative displacements, there are no low outliers. Q1 + Range = 105 + 275 = 380 cubic inches. This means that the maximum must be less than 380 cubic inches. Therefore, there are no high outliers (engines over 420 cubic inches).

f) It is not reasonable to expect 68% of the car engines to measure within one standard deviation of the mean. The distribution engine displacements is skewed to the right, so the Normal model is not appropriate.

g) Multiplying each of the engine displacements by 16.4 to convert cubic inches to cubic centimeters would affect measures of position and spread. All of the summary statistics (except the count!) could be converted to cubic centimeters by multiplying each by 16.4.

32. Engines, again.

a) The distribution of horsepower is roughly uniform, with a bit of skew to the right, as the number of cars begins to taper off after about 125 horsepower. The center of the distribution is about 100 horsepower. The lowest horsepower is around 60 and the highest is around 160.

b) The interquartile range is Q3 – Q1 = 125 – 78 = 47 horsepower.

c) Using the Outlier Rule (more than 1.5 IQRs beyond the quartiles) to find the fences:

Upper Fence: Q3 + 1.5(IQR) = 125 + 1.5(47) = 195.5 horsepower

Lower Fence: Q1 – 1.5(IQR) = 78 – 1.5(47) = 7.5 horsepower

From the histogram, we can see that there are no cars with horsepower ratings anywhere near these fences, so there are no outliers.

d) The distribution of horsepower is uniform, not unimodal, and not very symmetric, so the Normal model is probably not a very good model of the distribution of horsepower.

e) Within one standard deviation of the mean is roughly the interval 75 – 125 horsepower. By dividing the bars of the histogram up into boxes representing one car, and taking half of the boxes in the bars representing 70-79 and 120-129, I counted 22 (of the 38) cars in the interval. Approximately 58% of the cars are within one standard deviation of the mean.

f) Adding 10 horsepower to each car would increase the measures of position by 10 horsepower and leave the measures of spread unchanged. Mean, median, 25^{th} percentile and 75^{th} percentile would each increase by 10. The standard deviation, interquartile range, and range would remain the same.

33. Age and party 2011.

a) 3705 of 8414, or approximately 44.0%, of all voters surveyed were Republicans or leaned Republican.

b) This was a representative telephone survey conducted by Pew, a reputable polling firm. It is likely to be a reasonable estimate of the percentage of all voters who are Republicans.

c) 815 + 2416 = 3231 of 8414, or approximately 38.4%, of all voters surveyed were under 30 or over 65 years old.

d) 73 of 8414, or approximately 0.87%, of all voters surveyed were classified as "Neither" and under the age of 30.

e) 73 of the 733 people classified as "Neither", or 9.96%, were under the age of 30.

f) 73 of the 815 respondents under 30, or 8.96%, were classified as "Neither".

34. Pay.

The distribution of hourly wages for Chief Executives has a mean larger than the median, indicating a distribution that is skewed to the right. Likewise, the distribution of hourly wages for General and Operations Managers has a mean higher than the median, indicating a distribution that is skewed to the right.

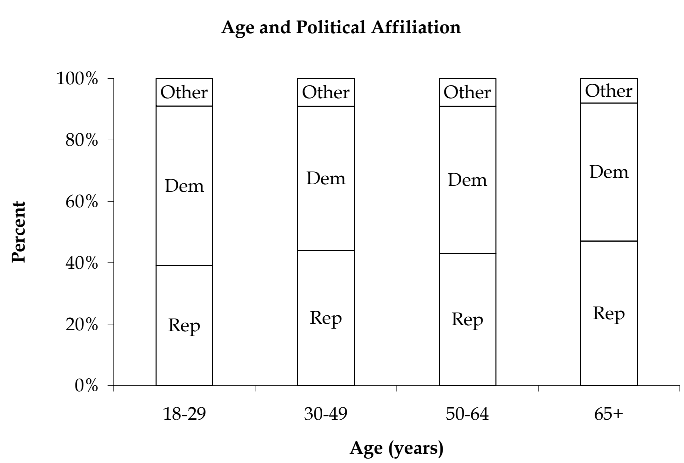
35. Age and party II.

a) The marginal distribution of party affiliation is:
Republican – 44.0% Democrat – 47.3% Neither – 8.7%
(As counts: Republican – 3705 Democrat – 3976 Independent – 733)

b) Graphs are at the right.

c) It appears that older voters are more likely to lean Republican, and younger voters are more likely to lean Deomocrat.

d) No. There is an evidence of an association between party affiliation and age. Younger voters tend to be more Democratic and less Republican.



36. Bike safety 2012.

a) *Who* – Years from 1994 to 2012. *What* – Number of bicycle fatalities reported. *When* – 1994 to 2012. *Where* – United States. *Why* – The information was collected for a report by the Bicycle Helmet Safety Institute. *How* – Although not specifically stated, the information was probably collected from a government agency or hospital records.

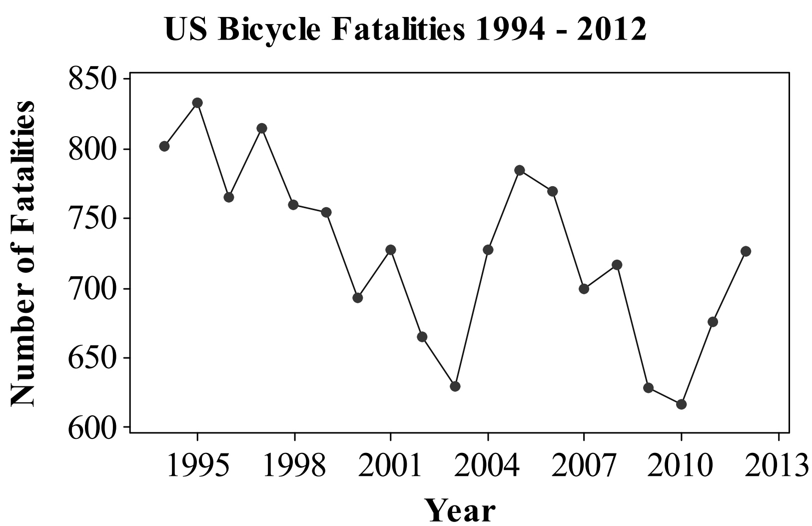
b)

United States Yearly
Bicycle Fatalities (1994 - 2012)

6	233
6	789
7	02333
7	56778
8	013

6|1 = 605 - 614
Bicycle Fatalities

c)



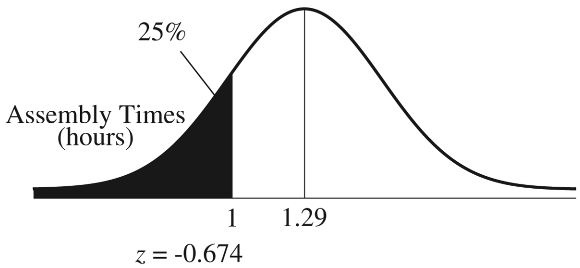
d) The stem-and-leaf display of the number of yearly bicycle fatalities reported in the United States shows that distribution is reasonably symmetric. It also provides some idea about the center and spread of the annual fatalities. This is not visible on the timeplot.

e) The timeplot of the number of yearly bicycle fatalities reported in the United Stats shows that the number of fatalities per year has generally declined over the time, but there are periods where the number of fatalities was increasing.

f) In the 10-year period from 1994 to 2003, reported bicycle fatalities decreased fairly steadily from about 800 per year to around 620 a year, then it increased sharply back to nearly 800 by 2005, decreased to nearly 600 by 2010, and has increased again in 2011 and 2012. Overall, it's not clear whether the decrease is real or just random fluctuation.

37. Some assembly required.

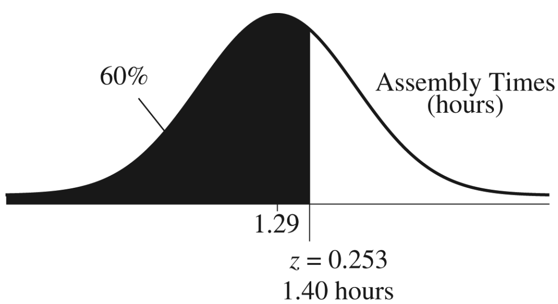
a)



$$z = \frac{y - \mu}{\sigma}$$
$$-0.674 = \frac{1 - 1.29}{\sigma}$$
$$\sigma = 0.43$$

According to the Normal model, the standard deviation is 0.43 hours.

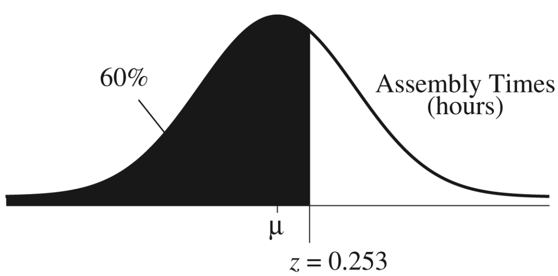
b)



$$z = \frac{y - \mu}{\sigma}$$
$$0.253 = \frac{y - 1.29}{0.43}$$
$$y = 1.40$$

According to the Normal model, the company would need to claim that the desk takes "less than 1.40 hours to assemble", not the catchiest of slogans!

c)



$$z = \frac{y - \mu}{\sigma}$$
$$0.253 = \frac{1 - \mu}{0.43}$$
$$\mu = 0.89$$

According to the Normal model, the company would have to lower the mean assembly time to 0.89 hour (53.4 minutes).

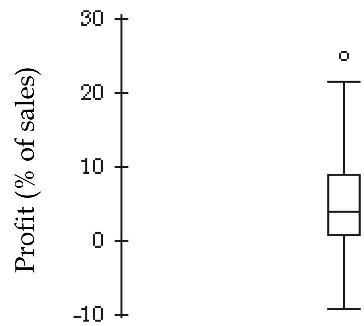
d) The new instructions and part-labeling may have helped lower the mean, but it also may have changed the standard deviation, making the assembly times more consistent as well as lower.

38. Profits.

a) The 5-number summary of the profits as a percent of sales of 29 of the *Forbes* 500 largest US corporations is: -9, 1, 4, 9.5, 25

(If you got –9, 1, 4, 9, 25, don't worry. Some statisticians figure quartiles of small sets differently than others. No one seems to care much which you use, since quartiles are much more useful in large data sets, anyway, where this doesn't matter.)

b) The boxplot of the distribution of the profits as a percent of sales of 29 of the *Forbes* 500 largest US corporations is at the right.



c) The mean profit is 4.72%, and the standard deviation of the distribution of profits is 7.55%.

d) The distribution of profits is unimodal and symmetric, centered around 4% of sales. The middle 50% of companies report profit between 1% and 9.5%. There are two companies with unusually high profits, 22% and 25%, although only 25% is technically an outlier.

Chapter 6 – Scatterplots, Association, and Correlation

Section 6.1

1. Association.

a) Either weight in grams or weight in ounces could be the explanatory or response variable. Greater weights in grams correspond with greater weights in ounces. The association between weight of apples in grams and weight of apples in ounces would be positive, straight, and perfect. Each apple's weight would simply be measured in two different scales. The points would line up perfectly.

b) Circumference is the explanatory variable, and weight is the response variable, since one-dimensional circumference explains three-dimensional volume (and therefore weight). For apples of roughly the same size, the association would be positive, straight, and moderately strong. If the sample of apples contained very small and very large apples, the association's true curved form would become apparent.

c) There would be no association between shoe size and GPA of college freshmen.

d) Number of miles driven is the explanatory variable, and gallons remaining in the tank is the response variable. The greater the number of miles driven, the less gasoline there is in the tank. If a sample of different cars is used, the association is negative, straight, and moderate. If the data is gathered on different trips with the same car, the association would be strong.

2. Association II.

a) Price for each T-shirt is the explanatory variable, and number of T-shirts sold is the response variable. The association would be negative, straight (until the price became too high to sell *any* shirts), and moderate. A very low price would likely lead to very high sales, and a very high price would lead to low sales.

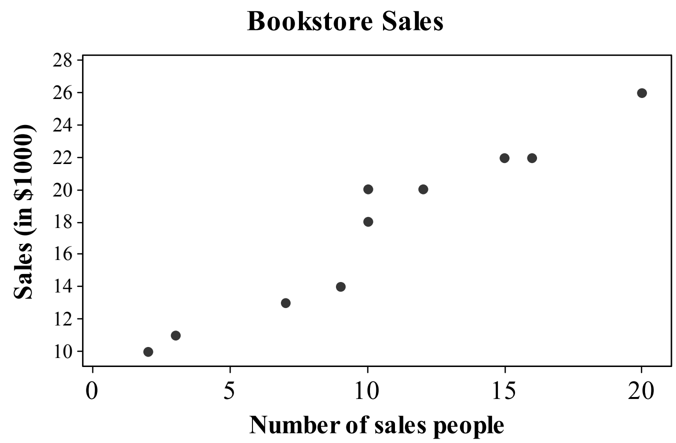
b) Depth of the dive is the explanatory variable, and water pressure is the response variable. The deeper you dive, the greater the water pressure. The association is positive, straight, and strong. For every 33 feet of depth, the pressure increases by one atmosphere (14.7 psi).

c) Depth of the water is the explanatory variable, and visibility is the response variable. The deeper you dive, the lower the visibility. The association is negative, possibly straight, and moderate if a sample of different bodies of water is used. If the same body of water has visibility measured at different depths, the association would be strong.

d) At first, it appears that there should be no association between weight of elementary school students and score on a reading test. However, with weight as the explanatory variable and score as the response variable, the association is positive, straight, and moderate. Students who weigh more are likely to do better on reading tests because of the lurking variable of age. Certainly, older students generally weigh more and generally are better readers. Therefore, students who weigh more are likely to be better readers. This does not mean that weight causes higher reading scores.

3. Bookstore sales.

a) The scatterplot is to the right.



b) There is a positive association between bookstore sales and the number of sales people working.

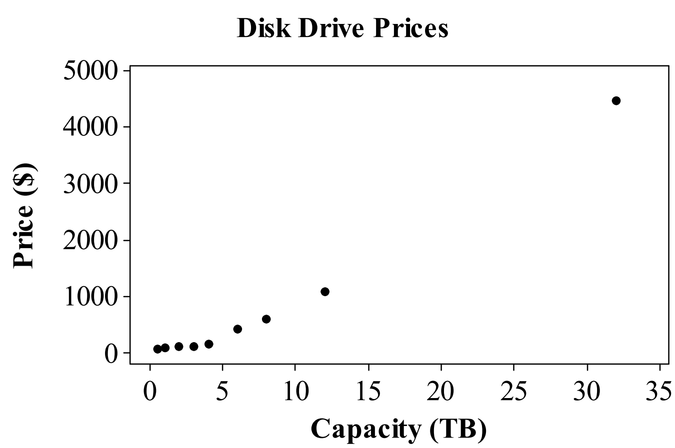
c) There is a linear association between bookstore sales and the number of sales people working.

d) There is a strong association between bookstore sales and the number of sales people working.

e) The relationship between bookstore sales and the number of sales people working has no outliers.

4. Disk drives 2014.

a) The scatterplot is to the right.



b) There is a positive association between price and capacity of disk drives.

c) There is a curved association between price and capacity of disk drives.

d) There is a strong association between price and capacity of disk drives.

e) The relationship between price and capacity of disk drives does not have any outliers.

Section 6.2

5. Correlation facts.

a) True.

b) False. The correlation will remain the same.

c) False. Correlation has no units.

6. Correlation facts II.

a) False. This is a very weak association.

b) False. Standardizing does not change the correlation.

c) True.

Section 6.3

7. Bookstore sales again.

This conclusion is not justified. Correlation does not demonstrate causation. The analyst argues that the number of sales staff working causes sales to be higher. It is possible (perhaps more plausible) that the store hired more people as sales increased. The causation may run in the opposite direction of the analyst's argument.

8. Blizzards.

The director's conclusion is not justified. The lurking variable is the severity of the blizzard. Particularly severe blizzards require more snowplows, and they also prevent people from leaving home, where they are more likely to make online purchases, especially since they have to leave home to go to a store.

Section 6.4

9. Salaries and logs.

Since $\log_{10} 10,000 = 4$, $\log_{10} 100,000 = 5$, and $\log_{10} 1,000,000 = 6$, the plotted points will be (1, 4), (15, 5), and (30, 6). The plot of these three points will lie very close to a straight line.

10. Dexterity scores.

The reciprocal re-expression is straighter. The points to plot for the re-expression are (4, 0.4), (9, 0.5) and (12, 0.67).

Chapter Exercises

11. Association III.

a) Altitude is the explanatory variable, and temperature is the response variable. As you climb higher, the temperature drops. The association is negative, straight, and moderate.

b) At first, it appears that there should be no association between ice cream sales and air conditioner sales. When the lurking variable of temperature is considered, the association becomes more apparent. When the temperature is high, ice cream sales tend to increase. Also, when the temperature is high, air conditioner sales tend to increase. Therefore, there is likely to be an increase in the sales of air conditioners whenever there is an increase in the sales of ice cream. The association is positive, straight, and moderate. Either one of the variables could be used as the explanatory variable.

c) Age is the explanatory variable, and grip strength is the response variable. The association is neither negative nor positive, but is curved, and moderate in strength, due to the variability in grip strength among people in general. The very young would have low grip strength, and grip strength would increase as age increased. After reaching a maximum (when physical prowess peaks), grip strength would decline again, with the elderly having low grip strengths.

d) Blood alcohol content is the explanatory variable, and reaction time is the response variable. As blood alcohol level increases, so does the time it takes to react to a stimulus. The association is positive, probably curved, and strong. The scatterplot would probably be almost linear for low concentrations of alcohol in the blood, and then begin to rise dramatically, with longer and longer reaction times for each incremental increase in blood alcohol content.

12. Association IV.

a) Consultation time is the explanatory variable, and cost is the response variable. The longer the consultation, the more the cost. The association is positive, straight, and moderately strong, since some legal firms charge more than others.

b) Distance from lightning is the explanatory variable, and time delay of the thunder is the response variable. The farther away you are from the strike, the longer it takes the thunder to reach your ears. The association is positive, straight, and fairly strong, since the speed of sound is not a constant. Sound travels at a rate of around 770 miles per hour, depending on the temperature.

c) Distance from the streetlight is the explanatory variable, and brightness is the response variable. The further away from the light you are, the less bright it appears. The association is negative, curved, and strong. Distance and light intensity follow an inverse square relationship. Doubling the distance to the light source reduces the intensity by a factor of four.

d) There is likely very little association between the weight of the car and the age of the owner. However, some might say that older drivers tend to drive larger cars. (Anyone who has seen my grandfather's car can attest to this!) If that is the case, there may be a positive, straight, and very weak association between weight of a car and the age of its owner.

13. Scatterplots.

a) None of the scatterplots show little or no association, although # 4 is very weak.

b) #3 and #4 show negative association. Increases in one variable are generally related to decreases in the other variable.

c) #2, #3, and #4 each show a straight association.

d) #2 shows a moderately strong association.

e) #1 and #3 each show a very strong association. #1 shows a curved association and #3 shows a straight association.

14. Scatterplots II.

a) #1 shows little or no association.

b) #4 shows a negative association.

c) #2 and #4 each show a straight association.

d) #3 shows a moderately strong, curved association.

e) #2 and #4 each show a very strong association, although some might classify the association as merely "strong".

15. Performance IQ scores vs. brain size.

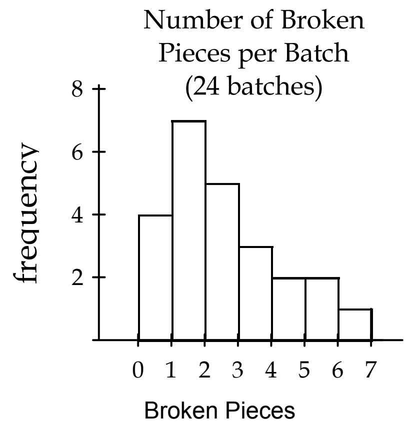
The scatterplot of IQ scores *vs.* Brain Sizes is scattered, with no apparent pattern. There appears to be little or no association between the IQ scores and brain sizes displayed in this scatterplot.

16. Kentucky derby 2014.

Winning speeds in the Kentucky Derby have generally increased over time. The association between year and speed is moderately strong, and seems slightly curved, with a greater rate of increase in winning speed before 1900 and a smaller rate of increase after 1900, suggesting that winning speeds have leveled off over time.

17. Firing pottery.

a) A histogram of the number of broken pieces is at the right.

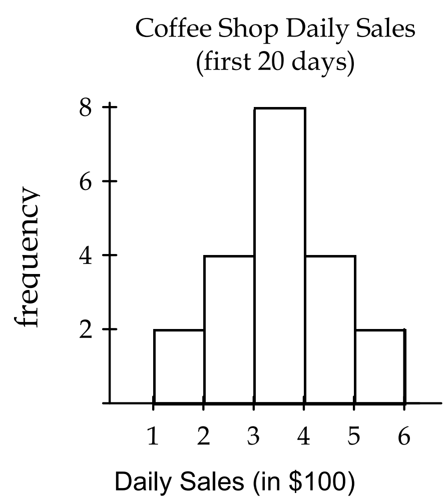


b) The distribution of the number broken pieces per batch of pottery is skewed right, centered around 1 broken piece per batch. Batches had from 0 and 6 broken pieces. The scatterplot does not show the center or skewness of the distribution.

c) The scatterplot shows that the number of broken pieces increases as the batch number increases. If the 8 daily batches are numbered sequentially, this indicates that batches fired later in the day generally have more broken pieces. This information is not visible in the histogram.

18. Coffee sales.

a) A histogram of daily sales is at the right.



b) The scatterplot shows that, in general, the sales have been increasing over time. The histogram does not show this.

c) The histogram shows that the mean of the daily sales for the coffee shop was between \$300 and \$400, and that this happened on 8 days. The scatterplot does not show this.

19. Matching.

a) 0.006 **b)** 0.777 **c)** – 0.923 **d)** – 0.487

20. Matching II.

a) – 0.977 **b)** 0.736 **c)** 0.951 **d)** – 0.021

21. Politics.

The candidate might mean that there is an **association** between television watching and crime. The term correlation is reserved for describing linear associations between quantitative variables. We don't know what type of variables "television watching" and "crime" are, but they seem categorical. Even if the variables are quantitative (hours of TV watched per week, and number of crimes committed, for example), we aren't sure that the relationship is linear. The politician also seems to be implying a cause-and-effect relationship between television watching and crime. Association of any kind does not imply causation.

22. Car thefts.

It might be reasonable to say that there is an **association** between the type of car you own and the risk that it will be stolen. The term correlation is reserved for describing linear associations between quantitative variables. Type of car is a categorical variable.

23. Roller coasters 2014.

a) It is appropriate to calculate correlation. Both height of the drop and speed are quantitative variables, the scatterplot shows an association that is straight enough, and there are not outliers.

b) There is a strong, positive, straight association between drop and speed; the greater the height of the initial drop, the higher the top speed.

24. Antidepressants.

a) It is appropriate to calculate correlation. Both placebo improvement and treated improvement are quantitative variables, the scatterplot shows an association that is straight enough, and there are not outliers.

b) There is a strong, positive, straight association between placebo and treated improvement. Experiments that showed a greater placebo effect also showed a greater mean improvement among patients who took an antidepressant.

25. Streams and hard water.

It is not appropriate to summarize the strength of the association between water hardness and pH with a correlation, since the association is curved, not Straight Enough.

26. Traffic headaches.

It is not appropriate to summarize the strength of the association between highway speed and total delay with a correlation. The scatterplot shows evidence of outliers, and the main cluster of data is not Straight Enough.

27. Cold nights.

The correlation is between the number of days since January 1 and temperature is likely to be near zero. We expect the temperature to be low in January, increase through the spring and summer, then decrease again. The relationship is not Straight Enough, so correlation is not an appropriate measure of strength.

28. Association V.

The researcher should have plotted the data first. A strong, curved relationship may have a very low correlation. In fact, correlation is only a useful measure of the strength of a linear relationship.

29. Prediction units.

The correlation between prediction error and year would not change, since the correlation is based on z-scores. The z-scores are the same whether the prediction errors are measured in nautical miles or miles.

30. More predictions.

The correlation between prediction error and year would not change, since the correlation is based on z-scores. The z-scores of the prediction errors are not changed by adding or subtracting a constant.

31. Correlation errors.

a) If the association between GDP and infant mortality is linear, a correlation of –0.772 shows a moderate, negative association. Generally, as GDP increases, infant mortality rate decreases.

b) Continent is a categorical variable. Correlation measures the strength of linear associations between quantitative variables.

32. More correlation errors.

a) Correlation must be between –1 and 1, inclusive. Correlation can never be 1.22.

b) A correlation, no matter how strong, cannot prove a cause-and-effect relationship.

33. Height and reading.

a) Actually, this *does* mean that taller children in elementary school are better readers. However, this does *not* mean that height causes good reading ability.

b) Older children are generally both taller and are better readers. Age is the lurking variable.

34. Smart phones and life expectancy.

a) No. It simply means that in countries where smart phone use is high, the life expectancy tends to be high as well.

b) General economic conditions of the country could affect both smart phone use and life expectancy. Richer countries generally have more smart phone use and better health care. The economy is a lurking variable.

35. Correlations conclusions I.

a) No. We don't know this from correlation alone. The relationship between age and income may be non-linear, or the relationship may contain outliers.

b) No. We can't tell the form of the relationship between age and income. We need to look at the scatterplot.

c) No. The correlation between age and income doesn't tell us anything about outliers.

d) Yes. Correlation is based on *z*-scores, and is unaffected by changes in units.

36. Correlation conclusions II.

a) No. We don't know this from correlation alone. The relationship between fuel efficiency and price may be non-linear, or the relationship may contain outliers.

b) No. We can't tell the form of the relationship between fuel efficiency and price. We need to look at the scatterplot.

c) No. The correlation between fuel efficiency and price doesn't tell us anything about outliers.

d) No. Correlation is based on *z*-scores, and is unaffected by changes in units.

37. Baldness and heart disease.

Even though the variables baldness and heart disease were assigned numerical values, they are categorical. Correlation is only an appropriate measure of the strength of linear association between quantitative variables. Their conclusion is meaningless.

38. Sample survey.

Even though zipcodes are numbers, they are categorical variables representing different geographic areas. Likewise, even thought the variable *Datasource* has numerical values, it is also categorical, representing the source from which the data were acquired. Correlation is only an appropriate measure of the strength of linear association between quantitative variables.

39. Income and housing.

a) There is a positive, moderately strong, linear relationship between *Housing Cost Index* and *Median Family Income*, with several states whose *Housing Cost Index* seems high for their *Median Family Income*, and one state whose *Housing Cost Index* seems low for their *Median Family Income*.

b) Correlation is based on *z*-scores. The correlation would still be 0.65.

c) Correlation is based on *z*-scores, and is unaffected by changes in units. The correlation would still be 0.65.

d) Washington, D.C. would be a moderately high outlier, with *Housing Cost Index* high for its *Median Family Income*. Since it doesn't fit the pattern, the correlation would decrease slightly if Washington, D.C. were included.

e) No. We can only say that higher *Housing Cost Index* scores are associated with higher *Median Family Income*, but we don't know why. There may be other variables at work.

f) No. We can say that there is a consistent monotone pattern, but correlation—even nonparametric correlation—does not demonstrate causation.

40. Interest rates and mortgages 2013.

a) There is a negative, strong, relationship between *Mortgage Loan Amount* and *Interest Rate*, although the pattern for interest rates between 4 and 6% does not fit the pattern. There are not outliers.

b) Correlation is based on *z*-scores. The correlation would still be –0.86.

c) Correlation is based on *z*-scores, and is unaffected by changes in units. The correlation would still be –0.80.

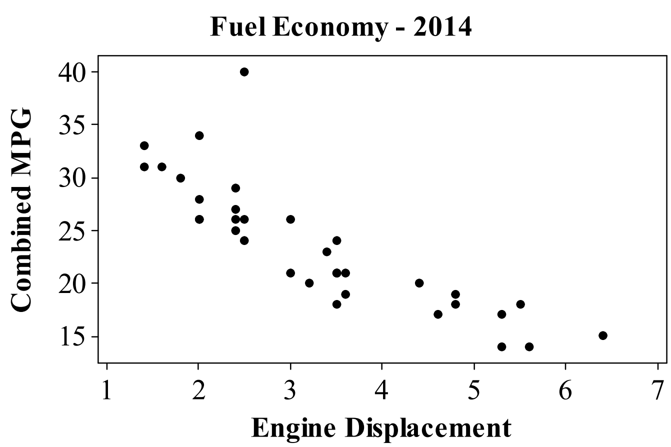
d) The given year has a very high mortgage amount for an interest rate that is that high. It doesn't fit the overall pattern, so the correlation would weaken (get closer to zero).

e) No. We can only say that lower interest rates are associated with larger mortgage amounts, but we don't know why. There may be other economic variables at work.

f) No. We can say that there is a consistent monotone pattern, but we can't say why.

41. Fuel economy 2014.

a) A scatterplot of combined fuel economy vs. engine displacement is at the right.



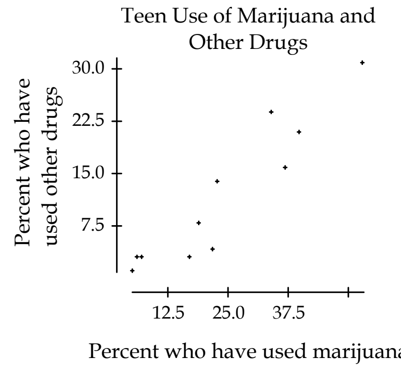
b) There is a strong, negative, straight association between engine displacement and mileage of the selected vehicles. There is one high outlier. Cars with larger engines tend to have lower mileage.

c) Since the relationship is linear, with no outliers, correlation is an appropriate measure of strength. The correlation between engine displacement and mileage of the selected vehicles is $r = -0.848$.

d) There is a strong linear relationship in the negative direction between engine displacement and combined gas mileage. Lower fuel efficiency is associated with larger engine displacement.

42. Drug abuse.

a) A scatterplot of percentage of teens who have used other drugs vs. percentage who have used marijuana in the U.S. and 10 Western European countries is at the right.



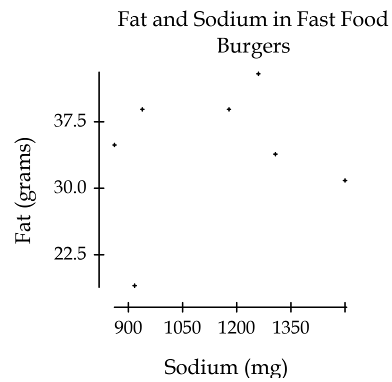
b) The correlation between the percent of teens who have used marijuana and the percent of teens who have used other drugs is $r = 0.934$.

c) The association between the percent of teens who have used marijuana and the percent of teens who have used other drugs is positive, strong, and straight. Countries with higher percentages of teens who have used marijuana tend to have higher percentages of teens that have used other drugs.

d) These results do not confirm that marijuana is a "gateway drug". An association exists between the percent of teens that have used marijuana and the percent of teens that have used other drugs. This does not mean that one caused the other.

43. Burgers.

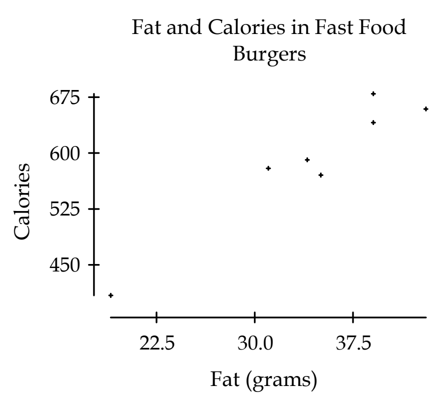
a) There is no apparent association between the number of grams of fat and the number of milligrams of sodium in several brands of fast food burgers. The correlation is only $r = 0.199$, which is close to zero, an indication of no association. One burger had a much lower fat content than the other burgers, at 19 grams of fat, with 920 milligrams of sodium. Without this (comparatively) low fat burger, the correlation would have been $r = -0.325$.



b) Spearman's rho is slightly negative. Using ranks doesn't allow the outlier to have as strong an influence and the remaining points have little or no association.

44. Burgers II.

a) The correlation between the number of calories and the number of grams of fat in several fast food burgers is $r = 0.961$. The association between the number of calories and the number of grams of fat in several fast food burgers is positive, straight, and strong. Typically, burgers with higher fat content have more calories. Even if the outlier at 410



calories and 19 grams of fat is set aside, the correlation is still quite strong at 0.837.

b) Spearman's rho is only 0.83 because it doesn't allow the outlier to have as strong an influence on the association.

45. Attendance 2013.

a) Number of runs scored and attendance are quantitative variables, the relationship between them appears to be straight (though weak), and there are no outliers, so calculating a correlation is appropriate.

b) The association between attendance and runs scored is positive, straight, and weak in strength. Generally, as the number of runs scored increases, so does attendance.

c) There is evidence of an association between attendance and runs scored, but a cause-and-effect relationship between the two is not implied. There may be lurking variables that can account for the increases in each. For example, perhaps winning teams score more runs and also have higher attendance. We don't have any basis to make a claim of causation.

46. Second inning 2013.

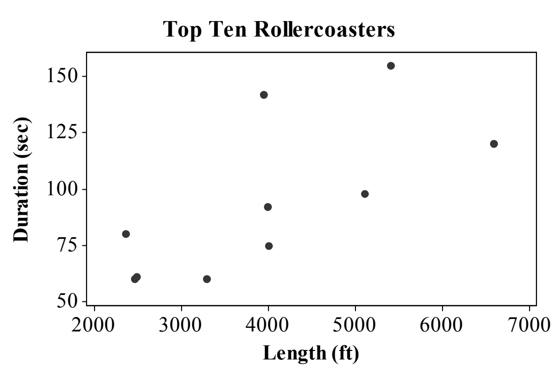
a) Winning teams generally enjoy greater attendance at their home games. The association between home attendance and number of wins is positive, somewhat straight, and weak.

b) The correlations, $r = 0.254$ for wins and attendance, and $r = 0.384$ for runs and attendance, so runs are slightly more strongly correlated with attendance than wins.

c) The correlation between number of runs scored and number of wins is $r = 0.793$, indicating a possible moderate association. However, since there is no scatterplot of wins vs. runs provided, we can't be sure the relationship is straight. Correlation may not be an appropriate measure of the strength of the association.

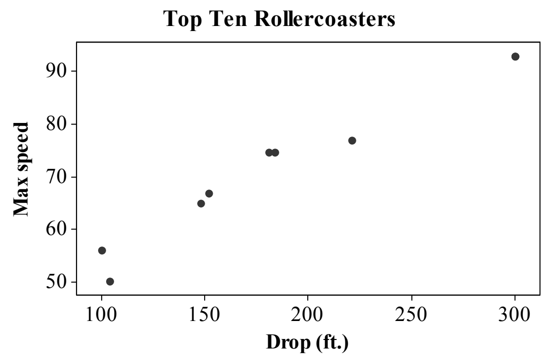
47. Thrills.

The scatterplot at the right shows that the association between duration and length is straight, positive, and moderate, with no outliers. Generally, rides on coasters with a greater length tend to last longer. The correlation between length and duration is 0.698, indicating a moderate association.

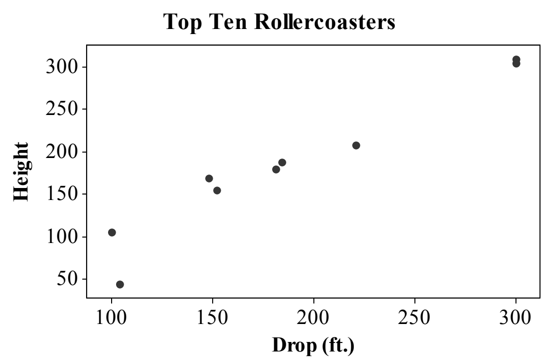


48. Thrills II.

a) With a correlation of 0.980, there is a very strong, positive, and linear relationship between the initial drop of a roller coaster and its maximum speed. It appears that the maximum speed of a roller coaster is directly related to the height of the first drop.



b) Similarly to part (a), the height and initial drop are directly related. The relationship is strong, positive, and reasonably linear and has a correlation of 0.966.

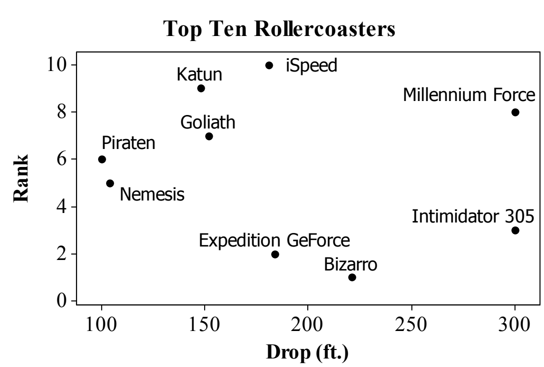


c) The initial drop of a coaster clearly affects the height and speed of the coaster. This is apparent, as most coasters start with a tall ascent that is the largest and the fastest. The initial drop is also moderately correlated with the steepness of the angle ($r = 0.603$) and somewhat strongly correlated with the length of the coaster ($r = 0.903$).

49. Thrills III.

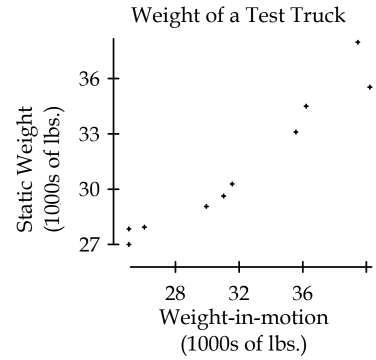
a) We would expect that as one variable (say length of ride) increases, the rank will improve, which means it will decrease.

b) Drop has the strongest correlation ($r = -0.193$), but even that correlation is very weak. The scatterplot shows no apparent association. The number one ranked coaster, Bizarro, has a fairly typical drop. There appear to be other factors that influence the rank of coaster more than any of the ones measured in this data set.



c) There may be other variables that account for the ranking. For example, other quantitative variables such as number of loops and number of corkscrews, categorical variables such as whether the coaster is made of wood or steel and whether or not there are tunnels may all have an affect on the rank.

50. Vehicle weights.



a) A scatterplot of the Static Weight vs. Weight-in-Motion of the test truck is at the right.

b) The association between static weight and weight-in-motion is positive, strong, and roughly straight. There may be a hint of a curve in the scatterplot.

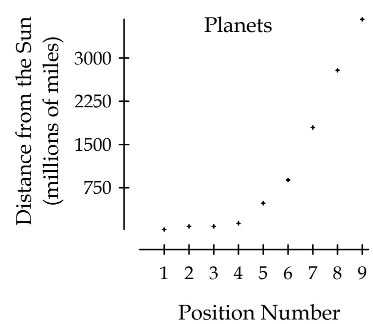
c) As the static weight of the test truck increased, so did the weight-in-motion, but the relationship appears weaker for heavier trucks.

d) The correlation between static weight and weight-in-motion is $r = 0.965$.

e) Weighing the trucks in kilograms instead of pounds would not change the correlation. Correlation, like *z*-score, has no units. It is a numerical measure of the degree of linear association between two variables.

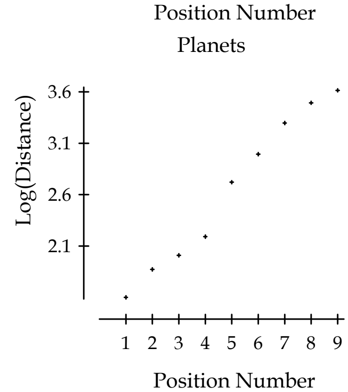
f) When the test truck weighed approximately 35,500 pounds, it weighed higher in motion. The scale may need to be recalibrated. If the scale were calibrated exactly, we would expect the points to line up perfectly, with no curve, and no deviations from the pattern.

51. Planets (more or less).



a) The association between Position Number of each planet and its distance from the sun (in millions of miles) is very strong, positive and curved. The scatterplot is at the right.

b) The relationship between Position Number and distance from the sun is not linear. Correlation is a measure of the degree of *linear* association between two variables.

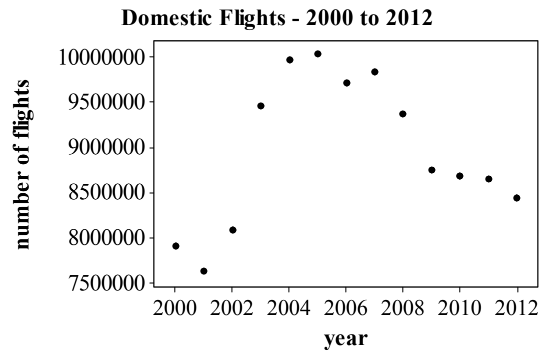


c) The scatterplot of the logarithm of distance versus Position Number (shown at the right) still shows a strong, positive relationship, but it is straighter than the previous scatterplot. It still shows a curve in the scatterplot, but it is straight enough that correlation may now be used as an appropriate measure of the strength of the relationship between logarithm of distance and Position Number, which will in turn give an indication of the strength of the association.

d) Kendall's tau is 1.0 because the points are perfectly monotonically increasing.

52. Flights 2012.

a) The correlation between the year and the number of flights is 0.189.



b) There is a positive, nonlinear association between the year and the number of flights. From 2000 to 2005, the number of flights tends to increase. From 2005 to 2012, the number of flights tends to decrease.

c) Correlation is not appropriate since the relationship is not linear. There are two trends in the data.

d) Kendall's tau is –0.026 and Spearman's rho is 0.154. They are both appropriate and both show that the relationship is very weak.

Chapter 7 – Linear Regression

Section 7.1

1. True or false.

a) False. The line usually touches none of the points. The line minimizes the sum of least squares.

b) True.

c) False. Least squares means the sum of all the squared residuals is minimized.

2. True or false II.

a) True.

b) False. Least squares means the sum of all the squared residuals is minimized.

c) True.

Section 7.2

3. Least square interpretations.

The weight of a newborn boy can be predicted as –5.94 kg plus 0.1875 kg per cm of length. This is a model fit to data. Parents should not be concerned if their newborn's length and weight don't fit this equation. No particular baby should be expected to fit this model exactly.

4. Residual interpretations.

$\widehat{weight} = -5.94 + 0.1875 length = -5.94 + 0.1875(48) = 3.06 \text{ kg}$

Residual $= weight - \widehat{weight} = 3 - 3.06 = -0.06 \text{ kg}$

The newborn was slightly lighter than the weight predicted by his length.

Section 7.3

5. Bookstore sales revisited.

a) The slope of the line of best fit, b_1, is 0.914. (When using technology, the slope is 0.913. When calculated by hand, the standard deviations and correlation must be rounded, resulting in a slight inaccuracy.)

$$b_1 = r\frac{s_y}{s_x}$$

$$b_1 = (0.965)\frac{5.34}{5.64}$$

$$b_1 = 0.914$$

b) The model predicts an increase in sales of 0.914($1000), or $914, for each additional sales person working.

c) The intercept, b_0, is 8.09. (When using technology, the intercept is 8.10. When calculated by hand, the means and slope must be rounded, resulting in a slight inaccuracy.)

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$17.6 = b_0 + (0.914)(10.4)$$
$$b_0 = 8.09$$

d) The model predicts that average sales would be approximately \$8.09(\$1000), or \$809, when there were no sales people working. This doesn't make sense in this context.

e) $\widehat{Sales} = 8.09 + 0.914\,People$ (hand calculation)

$\widehat{Sales} = 8.10 + 0.913\,People$ (technology)

f) $\widehat{Sales} = 8.09 + 0.914\,People = 8.09 + 0.914(18) = 24.542$

According to the model, we would expect sales to be approximately \$24,540 when 18 sales people are working. (\$24,530 using technology)

g) Residual $= sales - \widehat{sales} = \$25{,}000 - \$24{,}540 = \$460$ (Using technology, \$470)

h) Since the residual is positive, we have underestimated the sales.

6. Disk drives 2014 again.

a) The slope of the line of best fit, b_1, is 142.17. (When using technology, the slope is 142.18. When calculated by hand, the standard deviations and correlation must be rounded, resulting in a slight inaccuracy.)

$$b_1 = r\frac{s_y}{s_x}$$
$$b_1 = (0.9876)\frac{1418.67}{9.855}$$
$$b_1 = 142.17$$

b) The model predicts an average increase of \$142 for each additional TB of storage.

c) The intercept, b_0, is –296.25. (When using technology, the intercept is –296.31. When calculated by hand, the means and slope must be rounded, resulting in a slight inaccuracy.)

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$785.819 = b_0 + (142.17)(7.6111)$$
$$b_0 = -296.25$$

d) According to the model, the average cost of a drive with 0 TB capacity is expected to be –\$296.25. This does not make sense in this context.

e) $\widehat{Price} = -296.25 + 142.17\,Capacity$ (hand calculation)

$\widehat{Price} = -296.31 + 142.18\,Capacity$ (technology)

f) $\widehat{Price} = -296.25 + 142.17\,Capacity = -296.25 + 142.17(20) = 2547.15$

According to the model, we would expect the price of a 20 TB drive to average approximately $2547.15. (Answers may vary depending on the values used for the slope and the intercept.)

g) Residual $= Price - \widehat{Price} = 2017.86 - 2547.15 = -\$529.29$. This drive is a good buy. It costs $529.29 less than you expected to pay.

h) Since the residual is negative, the model overestimates the price.

i) No. We saw from the scatterplot that the relationship is curved. The model may not be accurate, or even appropriate.

Section 7.4

7. Sophomore slump?

The winners may be suffering from regression to the mean. Perhaps they weren't really better than other rookie executives, but just happened to have a lucky year the first year. When their second year performance landed them closer to the mean of the others, it looked like their performance had suffered.

8. Sophomore slump again?

Although on average, the performance of funds will cluster around the mean, we can't predict how any particular fund will do.

Section 7.5

9. Bookstore sales once more.

a) The residuals are measured in the same units as the response variable, thousands of dollars.

b) The residual with the largest magnitude, 2.77, contributes most to the sum of the squared residuals.

c) The residual with the smalles magnitude, 0.07, contributes least to the sum of the squared residuals.

10. Disk drive residuals.

a) The drive with a capacity of 12 TB, with a residual of –329.80, contributes the most to the sum of squared residuals, since it has the residual with the largest magnitude.

b) A negative residual means that the drive costs less than we might expect from this model and its capacity. For example, a residual of –329.80 indicates a drive that costs $329.80 less than we might expect.

Section 7.6

11. Bookstore sales last time.

$R^2 = 93.12\%$ Approximately 93% of the variability in bookstore sales can be accounted for by the regression with the number of sales workers.

12. Disk drives encore.

$R^2 = 97.54\%$ Approximately 97.54% of the variability in price of these disk drives can be accounted for by the regression with the capacity.

Section 7.7

13. Residual plots

a) The residual plot has a clear curved pattern. The linearity assumption is violated.

b) One point on the residual plot has a much larger residual than the others. The outlier condition is violated.

c) The residual plot shows a fanned shape. The equal spread condition is violated.

14. Disk drives last time.

a) The residuals do not appear to be straight. The linearity assumption is violated.

b) These data may benefit from re-expression.

Chapter Exercises

15. Cereals.

$\widehat{Potassium} = 38 + 27\,Fiber = 38 + 27(9) = 281$ mg. According to the model, we expect cereal with 9 grams of fiber to have 281 milligrams of potassium.

16. Engine size.

$\widehat{mpg} = 36.25 - 3.867\,EngineSize = 36.25 - 3.867(4) \approx 20.78$ mpg. According to the model, we expect a car with a 4 liter engine to get about 20.78 miles per gallon.

17. More cereal.

A negative residual means that the potassium content is actually lower than the model predicts for a cereal with that much fiber.

18. Engine size, again.

A positive residual means that the car gets better gas mileage than the model predicts for a car with that engine size.

19. Another bowl.

The model predicts that cereals will have approximately 27 more milligrams of potassium for each additional gram of fiber.

20. More engine size.

The model predicts that cars lose an average of 3.867 miles per gallon for each additional liter of engine size.

21. Cereal again.

$R^2 = (0.903)^2 \approx 0.815$ About 81.5% of the variability in potassium content is accounted for by the model.

22. Another car.

$R^2 = (-0.8476)^2 \approx 0.718$ About 71.8% of the variability in fuel economy is accounted for by the model.

23. Last bowl!

True potassium contents of cereals vary from the predicted values with a standard deviation of 30.77 milligrams.

24. Last tank!

True fuel economy varies from the predicted amount with a standard deviation of 2.435 miles per gallon.

25. Regression equations.

$\bar{x}$	s_x	$\bar{y}$	s_y	r	$\hat{y} = b_0 + b_1 x$
a) 10	2	20	3	0.5	$\hat{y} = 12.5 + 0.75x$
b) 2	0.06	7.2	1.2	–0.4	$\hat{y} = 23.2 - 8x$
c) 12	6	152	30	–0.8	$\hat{y} = 200 - 4x$
d) 2.5	1.2	25	100	0.6	$\hat{y} = -100 + 50x$

a)

$$b_1 = r\frac{s_y}{s_x}$$
$$b_1 = (0.5)\frac{3}{2}$$
$$b_1 = 0.75$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$20 = b_0 + 0.75(10)$$
$$b_0 = 12.5$$

b)

$$b_1 = r\frac{s_y}{s_x}$$
$$b_1 = (-0.4)\frac{1.2}{0.06}$$
$$b_1 = -8$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$7.2 = b_0 - 8(2)$$
$$b_0 = 23.2$$

c)

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$\bar{y} = 200 - 4(12)$$
$$\bar{y} = 152$$

$$b_1 = r\frac{s_y}{s_x}$$
$$-4 = (-0.8)\frac{s_y}{6}$$
$$s_y = 30$$

d)

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$\bar{y} = -100 + 50(2.5)$$
$$\bar{y} = 25$$

$$b_1 = r\frac{s_y}{s_x}$$
$$50 = r\frac{100}{1.2}$$
$$r = 0.6$$

26. More regression equations.

$\bar{x}$	s_x	$\bar{y}$	s_y	r	$\hat{y} = b_0 + b_1 x$
a) 30	4	18	6	–0.2	$\hat{y} = 27 - 0.3x$
b) 100	18	60	10	0.9	$\hat{y} = 10 + 0.5x$
c) 4	0.8	50	15	0.8	$\hat{y} = -10 + 15x$
d) 6	1.2	18	4	–0.6	$\hat{y} = 30 - 2x$

a)
$$b_1 = r\frac{s_y}{s_x} \qquad b_1 = (-0.2)\frac{6}{4} \qquad b_1 = -0.3$$
$$\hat{y} = b_0 + b_1 x \qquad \bar{y} = b_0 + b_1\bar{x} \qquad 18 = b_0 - 0.3(30) \qquad b_0 = 27$$

b)
$$b_1 = r\frac{s_y}{s_x} \qquad b_1 = (0.9)\frac{10}{18} \qquad b_1 = 0.5$$
$$\hat{y} = b_0 + b_1 x \qquad \bar{y} = b_0 + b_1\bar{x} \qquad 60 = b_0 + 0.5(100) \qquad b_0 = 10$$

c)
$$\hat{y} = b_0 + b_1 x \qquad \bar{y} = b_0 + b_1\bar{x} \qquad 50 = -10 + 15(\bar{x}) \qquad \bar{x} = 4$$
$$b_1 = r\frac{s_y}{s_x} \qquad 15 = r\frac{15}{0.8} \qquad r = 0.8$$

d)
$$\hat{y} = b_0 + b_1 x \qquad \bar{y} = b_0 + b_1\bar{x} \qquad 18 = 30 - 2(\bar{x}) \qquad \bar{x} = 6$$
$$b_1 = r\frac{s_y}{s_x} \qquad -2 = (-0.6)\frac{4}{s_x} \qquad s_x = 1.2$$

27. Residuals.

a) The scattered residuals plot indicates an appropriate linear model.

b) The curved pattern in the residuals plot indicates that the linear model is not appropriate. The relationship is not linear.

c) The fanned pattern indicates that the linear model is not appropriate. The model's predicting power decreases as the values of the explanatory variable increase.

28. Residuals.

a) The curved pattern in the residuals plot indicates that the linear model is not appropriate. The relationship is not linear.

b) The fanned pattern indicates uneven spread. The models predicting power increases as the value of the explanatory variable increases.

c) The scattered residuals plot indicates an appropriate linear model.

29. Real estate.

a) The explanatory variable (x) is size, measured in square feet, and the response variable (y) is price measured in thousands of dollars.

b) The units of the slope are thousands of dollars per square foot.

c) The slope of the regression line predicting price from size should be positive. Bigger homes are expected to cost more.

30. Roller coasters revisited.

a) The explanatory variable (x) is initial drop, measured in feet, and the response variable (y) is duration, measured in seconds.

b) The units of the slope are seconds per foot.

c) The slope of the regression line predicting duration from initial drop should be positive. Coasters with higher initial drops probably provide longer rides.

31. What slope?

The only slope that makes sense is 300 pounds per foot. 30 pounds per foot is too small. For example, a Honda Civic is about 14 feet long, and a Cadillac DeVille is about 17 feet long. If the slope of the regression line were 30 pounds per foot, the Cadillac would be predicted to outweigh the Civic by only 90 pounds! (The real difference is about 1500 pounds.) Similarly, 3 pounds per foot is too small. A slope of 3000 pounds per foot would predict a weight difference of 9000 pounds (4.5 tons) between Civic and DeVille. The only answer that is even reasonable is 300 pounds per foot, which predicts a difference of 900 pounds. This isn't very close to the actual difference of 1500 pounds, but at least it is in the right ballpark.

32. What slope, again?

The only slope that makes sense is 1 foot in height per inch in circumference. 0.1 feet per inch is too small. A trunk would have to increase in circumference by 10 inches for every foot in height. If that were true, pine trees would be all trunk! 10 feet per inch (and, similarly 100 feet per inch) is too large. If pine trees reach a maximum height of 60 feet, for instance, then the variation in circumference of the trunk would only be 6 inches. Pine tree trunks certainly come in more sizes than that. The only slope that is reasonable is 1 foot in height per inch in circumference.

33. Real estate again.

71.4% of the variability in price can be accounted for by variability in size. (In other words, 71.4% of the variability in price can be accounted for by the linear model.)

34. Roller coasters again.

16.9% of the variability in duration can be accounted for by variability in initial drop. (In other words, 16.9% of the variability in duration can be accounted for by the linear model.)

35. Misinterpretations.

a) R^2 is an indication of the strength of the model, not the appropriateness of the model. A scattered residuals plot is the indicator of an appropriate model.

b) Regression models give predictions, not actual values. The student should have said, "The model predicts that a bird 10 inches tall is expected to have a wingspan of 17 inches."

36. More misinterpretations.

a) R^2 measures the amount of variation accounted for by the model. Literacy rate determines 64% of *the variability* in life expectancy.

b) Regression models give predictions, not actual values. The student should have said, "The slope of the line shows that an increase of 5% in literacy rate *is associated with an expected* 2-year improvement in life expectancy."

37. Real estate redux.

a) The correlation between size and price is $r = \sqrt{R^2} = \sqrt{0.714} = 0.845$. The positive value of the square root is used, since the relationship is believed to be positive.

b) The price of a home that is one standard deviation above the mean size would be predicted to be 0.845 standard deviations (in other words ***r*** standard deviations) above the mean price.

c) The price of a home that is two standard deviations below the mean size would be predicted to be 1.69 (or 2×0.845) standard deviations below the mean price.

38. Another ride.

a) The correlation between drop and duration is $r = \sqrt{R^2} = \sqrt{0.169} = 0.411$. The positive value of the square root is used, since the relationship is believe positive.

b) The duration of a coaster whose initial drop is one standard deviation below the mean drop would be predicted to be about 0.411 standard deviations (in other words, *r* standard deviations) below the mean duration.

c) The duration of a coaster whose initial drop is three standard deviation above the mean drop would be predicted to be about 1.233 (or 3×0.411) standard deviations above the mean duration.

39. ESP.

a) First, since no one has ESP, you must have scored 2 standard deviations above the mean by chance. On your next attempt, you are unlikely to duplicate the extraordinary event of scoring 2 standard deviations above the mean. You will likely "regress" towards the mean on your second try, getting a lower score. If you want to impress your friend, don't take the test again. Let your friend think you can read his mind!

b) Your friend doesn't have ESP, either. No one does. Your friend will likely "regress" towards the mean score on his second attempt, meaning his score will probably go up. If the goal is to get a higher score, your friend should try again.

40. *SI* jinx.

Athletes, especially rookies, usually end up on the cover of Sports Illustrated for extraordinary performances. If these performances represent the upper end of the distribution of performance for this athlete, future performance is likely to regress toward the average performance of that athlete. An athlete's average performance usually isn't notable enough to land the cover of SI. Of course, there are always exceptions, like Michael Jordan, Lebron James, Serena Williams, and others.

41. More real estate.

a) According to the linear model, the price of a home is expected to increase \$61 (0.061 thousand dollars) for each additional square-foot in size.

b)
$$\widehat{Price} = 47.82 + 0.061\,Size$$
$$\widehat{Price} = 47.82 + 0.061(3000)$$
$$\widehat{Price} = 230.82$$

According to the linear model, a 3000 square-foot home is expected to have a price of \$230,820.

c)
$$\widehat{Price} = 47.82 + 0.061\,Size$$
$$\widehat{Price} = 47.82 + 0.061(1200)$$
$$\widehat{Price} = 121.02$$

According to the linear model, a 1200 square-foot home is expected to have a price of \$121,020. The asking price is \$121,020 - \$6000 = \$115,020. \$6000 is the (negative) residual.

42. Last ride.

a) According to the linear model, the duration of a coaster ride is expected to increase by about 0.284 seconds for each additional foot of initial drop.

b)
$$\widehat{Duration} = 99.828 + 0.284\,Drop$$
$$\widehat{Duration} = 99.828 + 0.284(200)$$
$$\widehat{Duration} = 156.628$$

According to the linear model, a coaster with a 200 foot initial drop is expected to last 156.628 seconds.

c)
$$\widehat{Duration} = 99.828 + 0.284\,Drop$$
$$\widehat{Duration} = 99.828 + 0.284(150)$$
$$\widehat{Duration} = 142.428$$

According to the linear model, a coaster with a 150 foot initial drop is expected to last 142.428 seconds. The advertised duration is longer, at 150 seconds.
150 seconds – 142.428 seconds = 7.572 seconds, a positive residual.

43. Cigarettes.

a) A linear model is probably appropriate. The residuals plot shows some initially low points, but there is no clear curvature.

b) 92.4% of the variability in nicotine level is accounted for by variability in tar content. (In other words, 92.4% of the variability in nicotine level is accounted for by the linear model.)

44. Attendance 2013, revisited.

a) The linear model is appropriate. Although the relationship is not strong, it is reasonably straight, and the residuals plot shows no pattern. There may be a bit of thickening on the right, but not enough to keep us from trying the linear model.

b) 6.45% of the variability in attendance is accounted for by variability in the number of wins. (In other words, 6.45% of the variability is accounted for by the model.)

c) The residuals spread out. There is more variation in attendance as the number of wins increases. However, the relationship is very weak.

d) The Yankees attendance was about 11,000 fans more than we might expect given the number of wins. This is a positive residual.

45. Another cigarette.

a) The correlation between tar and nicotine is $r = \sqrt{R^2} = \sqrt{0.924} = 0.961$. The positive value of the square root is used, since the relationship is believed to be positive. Evidence of the positive relationship is the positive coefficient of tar in the regression output.

b) The average nicotine content of cigarettes that are two standard deviations below the mean in tar content would be expected to be about 1.922 (2×0.961) standard deviations below the mean nicotine content.

c) Cigarettes that are one standard deviation above average in nicotine content are expected to be about 0.961 standard deviations (in other words, *r* standard deviations) above the mean tar content.

46. Attendance 2013, revisited.

a) The correlation between attendance and number of wins is $r = \sqrt{R^2} = \sqrt{0.0645} = 0.254$. The positive value of the square root is used, since the relationship is positive.

b) A team that is two standard deviations above the mean in number of wins would be expected to have attendance that is 0.508 (or 2×0.254) standard deviations above the mean attendance.

c) A team that is one standard deviation below the mean in attendance would be expected to have a number of wins that is 0.254 standard deviations (in other words, *r* standard deviations) below the mean number of wins. The correlation between two variables is the same, regardless of the direction in which predictions are made. Be careful, though, since the same is NOT true for predictions made using the slope of the regression equation. Slopes are valid only for predictions in the direction for which they were intended.

47. Last cigarette.

a) $\widehat{Nicotine} = 0.15403 + 0.065052\,Tar$ is the equation of the regression line that predicts nicotine content from tar content of cigarettes.

b)

$$\widehat{Nicotine} = 0.15403 + 0.065052\,Tar$$
$$\widehat{Nicotine} = 0.15403 + 0.065052(4)$$
$$\widehat{Nicotine} = 0.414$$

The model predicts that a cigarette with 4 mg of tar will have about 0.414 mg of nicotine.

c) For each additional mg of tar, the model predicts an increase of 0.065 mg of nicotine.

d) The model predicts that a cigarette with no tar would have 0.154 mg of nicotine.

e)

$$\widehat{Nicotine} = 0.15403 + 0.065052\,Tar$$
$$\widehat{Nicotine} = 0.15403 + 0.065052(7)$$
$$\widehat{Nicotine} = 0.6094$$

The model predicts that a cigarette with 7 mg of tar will have 0.6094 mg of nicotine. If the residual is –0.5, the cigarette actually had 0.1094 mg of nicotine.

48. Last inning 2013.

a) $\widehat{Attendance} = 16484.0 + 147.4\,Wins$ is the equation of the regression line that predicts attendance from the number of games won by American League baseball teams.

b)

$$\widehat{Attendance} = 16484.0 + 147.4\,Wins$$
$$\widehat{Attendance} = 16484.0 + 147.4(50)$$
$$\widehat{Attendance} = 23854$$

The model predicts that a team with 50 wins will have attendance of approximately 23,854 people.

c) For each additional win, the model predicts an increase in attendance of 147.4 people.

d) A negative residual means that the team's actual attendance is lower than the attendance model predicts for a team with as many wins.

e)

$$\widehat{Attendance} = 16484.0 + 147.4(Wins)$$
$$\widehat{Attendance} = 16484.0 + 147.4(97)$$
$$\widehat{Attendance} = 30781.8$$

The predicted attendance for the Cardinals was 30781.8. The actual attendance of 41602 gives a residual of 41,602 – 30781.8 = 10820.2. The Cardinals had over 10,800 more people attending on average than the model predicted.

49. Income and housing revisited.

a) Yes. Both housing cost index and median family income are quantitative. The scatterplot is Straight Enough, although there may be a few outliers. The spread increases a bit for states with large median incomes, but we can still fit a regression line.

b) Using the summary statistics given in the problem, calculate the slope and intercept:

$$b_1 = r\frac{s_{HCI}}{s_{MFI}}$$
$$b_1 = (0.65)\frac{116.55}{7072.47}$$
$$b_1 = 0.0107$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1\bar{x}$$
$$338.2 = b_0 + 0.0107(46234)$$
$$b_0 = -156.50$$

The regression equation that predicts HCI from MFI is $\widehat{HCI} = -156.50 + 0.0107MFI$

c)

$$\widehat{HCI} = -156.50 + 0.0107MFI$$
$$\widehat{HCI} = -156.50 + 0.0107(44993)$$
$$\widehat{HCI} = 324.93$$

The model predicts that a state with median family income of $44993 have an average housing cost index of 324.93

d) The prediction is 223.09 too low. Washington has a positive residual.

e) The correlation is the slope of the regression line that relates *z*-scores, so the regression equation would be $\hat{z}_{HCI} = 0.65z_{MFI}$.

f) The correlation is the slope of the regression line that relates *z*-scores, so the regression equation would be $\hat{z}_{MFI} = 0.65z_{HCI}$.

50. Interest rates and mortgages 2013 again.

a) Yes. Both interest rate and total mortgages are quantitative, and the scatterplot is Straight Enough. The spread is fairly constant, and there are no outliers.

b) Using the summary statistics given in the problem, calculate the slope and intercept:

$$b_1 = r\frac{s_{MortAmt}}{s_{IntRate}} \qquad \hat{y} = b_0 + b_1 x$$

$$b_1 = (-0.80)\frac{1.515}{2.055} \qquad \overline{y} = b_0 + b_1\overline{x}$$

$$3.926 = b_0 - 0.5898(7.194)$$

$$b_1 = -0.5898 \qquad b_0 = 8.1690$$

The regression equation that predicts total mortgage amount from interest rate is $\widehat{MortAmt} = 8.169 - 0.590\,IntRate$

(From the original data, $\widehat{MortAmt} = 8.148 - 0.587\,IntRate$)

c)

$$\widehat{MortAmt} = 8.169 - 0.590\,IntRate$$

$$\widehat{MortAmt} = 8.169 - 0.590(13)$$

$$\widehat{MortAmt} = 0.499$$

If interest rates were 13%, we would expect there to be \$0.499 trillion in total mortgages. (\$0.519 trillion if you worked with the actual data.)

d) We should be very cautious in making a prediction about an interest rate of 13%. It is well outside the range of our original x-variable, and care should always be taken when extrapolating. This prediction may not be appropriate.

e) The correlation is the slope of the regression line that relates z-scores, so the regression equation would be $\hat{z}_{MortAmt} = -0.80 z_{IntRate}$.

f) The correlation is the slope of the regression line that relates z-scores, so the regression equation would be $\hat{z}_{IntRate} = -0.80 z_{MortAmt}$.

51. Online clothes.

a) Using the summary statistics given in the problem, calculate the slope and intercept:

$$b_1 = r\frac{s_{Total}}{s_{Age}} \qquad \hat{y} = b_0 + b_1 x$$

$$b_1 = (0.037)\frac{253.62}{8.51} \qquad \overline{y} = b_0 + b_1\overline{x}$$

$$572.52 = b_0 + 1.1027(29.67)$$

$$b_1 = 1.1027 \qquad b_0 = 539.803$$

The regression equation that predicts total online clothing purchase amount from age is $\widehat{Total} = 539.803 + 1.103\,Age$

b) Yes. Both total purchases and age are quantitative variables, and the scatterplot is Straight Enough, even though it is quite flat. There are no outliers and the plot does not spread throughout the plot.

c)

$$\widehat{Total} = 539.803 + 1.103Age$$
$$\widehat{Total} = 539.803 + 1.103(18)$$
$$\widehat{Total} = 559.66$$

The model predicts that an 18-year-old will have $559.66 in total yearly online clothing purchases.

$$\widehat{Total} = 539.803 + 1.103Age$$
$$\widehat{Total} = 539.803 + 1.103(50)$$
$$\widehat{Total} = 594.95$$

The model predicts that a 50-year-old will have $594.95 in total yearly online clothing purchases.

d) $R^2 = (0.037)^2 \approx 0.0014 = 0.14\%$..

e) This model would not be useful to the company. The scatterplot is nearly flat. The model accounts for almost none of the variability in total yearly purchases.

52. Online clothes II.

a) Using the summary statistics given, calculate the slope and intercept:

$$b_1 = r\frac{s_{Total}}{s_{Income}}$$
$$b_1 = (0.722)\frac{253.62}{16952.50}$$
$$b_1 = 0.01080157$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$572.52 = b_0 + 0.01080157(50343.40)$$
$$b_0 = 28.73$$

The regression equation that predicts total online clothing purchase amount from income is $\widehat{Total} = 28.73 + 0.0108\,Income$

(Since the mean income is a relatively large number, the value of the intercept will vary, based on the rounding of the slope. Notice that it is very close to zero in the context of yearly income.)

b) The assumptions for regression are met. Both variables are quantitative and the plot is straight enough. There are several possible outliers, but none of these points are extreme, and there are 500 data points to establish a pattern. The spread of the plot does not change throughout the range of income.

c)

$$\widehat{Total} = 28.73 + 0.0108\,Income$$
$$\widehat{Total} = 28.73 + 0.0108(20{,}000)$$
$$\widehat{Total} = 244.73$$

The model predicts that a person with $20,000 yearly income will make $244.73 in online purchases. (Predictions may vary, based on rounding of the model.)

$$\widehat{Total} = 28.73 + 0.0108\,Income$$
$$\widehat{Total} = 28.73 + 0.0108(80{,}000)$$
$$\widehat{Total} = \$892.73$$

The model predicts that a person with $80,000 yearly income will make $892.73 in online purchases. (Predictions may vary, based on rounding of the model.)

d) $R^2 = (0.722)^2 \approx 0.521 = 52.1\%$.

e) The model accounts for a 52.1% of the variation in total yearly purchases, so the model would probably be useful to the company. Additionally, the difference between the predicted purchases of a person with $20,000 yearly income and $80,000 yearly income is of practical significance.

53. SAT scores.

a) The association between SAT Math scores and SAT Verbal Scores was linear, moderate in strength, and positive. Students with high SAT Math scores typically had high SAT Verbal scores.

b) One student got a 500 Verbal and 800 Math. That set of scores doesn't seem to fit the pattern.

c) $r = 0.685$ indicates a moderate, positive association between SAT Math and SAT Verbal, but only because the scatterplot shows a linear relationship. Students who scored one standard deviation above the mean in SAT Math were expected to score 0.685 standard deviations above the mean in SAT Verbal. Additionally, $R^2 = (0.685)^2 = 0.469225$, so 46.9% of the variability in math score was accounted for by variability in verbal score.

d) The scatterplot of verbal and math scores shows a relationship that is straight enough, so a linear model is appropriate.

$$b_1 = r\frac{s_{Math}}{s_{Verbal}}$$
$$b_1 = (0.685)\frac{96.1}{99.5}$$
$$b_1 = 0.661593$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1\bar{x}$$
$$612.2 = b_0 + 0.661593(596.3)$$
$$b_0 = 217.692$$

The equation of the least squares regression line for predicting SAT Math score from SAT Verbal score is $\widehat{Math} = 217.692 + 0.662\,Verbal$.

e) For each additional point in verbal score, the model predicts an increase of 0.662 points in math score. A more meaningful interpretation might be scaled up. For each additional 10 points in verbal score, the model predicts an increase of 6.62 points in math score.

f)

$$\widehat{Math} = 217.692 + 0.662\,Verbal$$
$$\widehat{Math} = 217.692 + 0.662(500)$$
$$\widehat{Math} = 548.692$$

According to the model, a student with a verbal score of 500 was expected to have a math score of 548.692.

g)

$$\widehat{Math} = 217.692 + 0.662\,Verbal$$
$$\widehat{Math} = 217.692 + 0.662(800)$$
$$\widehat{Math} = 747.292$$

According to the model, a student with a verbal score of 800 was expected to have a math score of 747.292. She actually scored 800 on math, so her residual was 800 – 747.292 = 52.708 points

54. Success in college

a) A scatterplot showed the relationship between combined SAT score and GPA to be reasonably linear, so a linear model is appropriate.

$$b_1 = r\frac{s_{GPA}}{s_{SAT}}$$
$$b_1 = (0.47)\frac{0.56}{123}$$
$$b_1 \approx 0.0021398$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1\bar{x}$$
$$2.66 = b_0 + 0.0021398(1222)$$
$$b_0 \approx 0.045$$

The regression equation predicting GPA from SAT score is:
$$\widehat{GPA} = 0.045 + 0.00214\,SAT$$

b) The model predicts that a student with an SAT score of 0 would have a GPA of 0.045. The *y*-intercept is not meaningful, since an SAT score of 0 is impossible.

c) The model predicts that students who scored 100 points higher on the SAT tended to have a GPA that was 0.2140 higher.

d)

$$\widehat{GPA} = 0.045 + 0.00214\,SAT$$
$$\widehat{GPA} = 0.045 + 0.00214(1400)$$
$$\widehat{GPA} \approx 3.04$$

According to the model, a student with an SAT score of 1400 is expected to have a GPA of 3.04.

e) According to the model, SAT score is not a very good predictor of college GPA. $R^2 = (0.47)^2 = 0.2209$, which means that only 22.09% of the variability in GPA can be accounted for by the model. The rest of the variability is determined by other factors.

f) A student would prefer to have a positive residual. A positive residual means that the student's actual GPA is higher than the model predicts for someone with the same SAT score.

55. SAT, take 2.

a) $r = 0.685$. The correlation between SAT Math and SAT Verbal is a unitless measure of the degree of linear association between the two variables. It doesn't depend on the order in which you are making predictions.

b) The scatterplot of verbal and math scores shows a relationship that is straight enough, so a linear model is appropriate.

$$b_1 = r\frac{s_{Verbal}}{s_{Math}}$$
$$b_1 = (0.685)\frac{99.5}{96.1}$$
$$b_1 = 0.709235$$

$$\hat{y} = b_0 + b_1 x$$
$$\overline{y} = b_0 + b_1 \overline{x}$$
$$596.3 = b_0 + 0.709235(612.2)$$
$$b_0 = 162.106$$

The equation of the least squares regression line for predicting SAT Verbal score from SAT Math score is: $\widehat{Verbal} = 162.106 + 0.709\,Math$

c) A positive residual means that the student's actual verbal score was higher than the score the model predicted for someone with the same math score.

d)

$$\widehat{Verbal} = 162.106 + 0.709\,Math$$
$$\widehat{Verbal} = 162.106 + 0.709(500)$$
$$\widehat{Verbal} = 516.606$$

According to the model, a person with a math score of 500 was expected to have a verbal score of 516.606 points.

e)

$$\widehat{Math} = 217.692 + 0.662\,Verbal$$
$$\widehat{Math} = 217.692 + 0.662(516.606)$$
$$\widehat{Math} = 559.685$$

According to the model, a person with a verbal score of 516.606 was expected to have a math score of 559.685 points.

f) The prediction in part e) does not cycle back to 500 points because the regression equation used to predict math from verbal is a different equation than the regression equation used to predict verbal from math. One was generated by minimizing squared residuals in the verbal direction, the other was generated by minimizing squared residuals in the math direction. If a math score is one standard deviation above the mean, its predicted verbal score regresses toward the mean. The same is true for a verbal score used to predict a math score.

56. Success, part 2.

$$b_1 = r\frac{s_{SAT}}{s_{GPA}}$$
$$b_1 = (0.47)\frac{123}{0.56}$$
$$b_1 = 103.232$$

$$\hat{y} = b_0 + b_1 x$$
$$\overline{y} = b_0 + b_1\overline{x}$$
$$1222 = b_0 + 103.232(2.66)$$
$$b_0 = 947.403$$

The regression equation to predict SAT score from GPA is:

$$\widehat{SAT} = 947.403 + 103.232(GPA)$$
$$\widehat{SAT} = 947.403 + 103.232(3)$$
$$\widehat{SAT} = 1257.1$$

The model predicts that a student with a GPA of 3.0 is expected to have an SAT score of 1257.1.

57. Wildfires 2012.

a) The scatterplot shows a roughly linear relationship between the year and the number of wildfires, so the linear model is appropriate. The relationship is very weak, however.

b) The model predicts an increase of an average of about 78.5 wildfires per year.

c) It seems reasonable to interpret the intercept. The model predicts about 75,609 wildfires in 1985, which is within the scope of the data, although it isn't very useful since we know the actual number of wildfires in 1985. There isn't much need for a prediction.

d) The standard deviation of the residuals is 11690 fires. That's a large residual, considering that these years show between 60,000 and 90,000 fires per year. The association just isn't very strong.

e) The model only accounts for about 0.032% of the variability in the number of fires each year. The rest of the variability is due to other factors that we don't have data about.

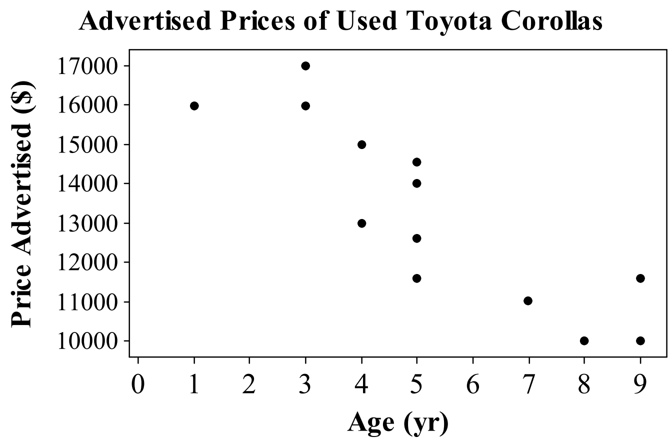
58. Wildfires 2012—sizes.

a) The regression model isn't appropriate for the association between the number of acres per fire and the year. The scatterplot is not straight, and the residuals plot shows a curved pattern, indicating that there is still information contained in the association that is not accounted for by the linear model.

b) R^2 is not of any use when the model is curved, since it is based on r, which is a measure of linear association.

59. Used cars 2014.

a) We are attempting to predict the price in dollars of used Toyota Corollas from their age in years. A scatterplot of the relationship is at the right.



b) There is a strong, negative, linear association between price and age of used Toyota Corollas.

c) The scatterplot provides evidence that the relationship is straight enough. A linear model will likely be an appropriate model.

d) Since R^2 = 0.891, simply take the square root to find r. $\sqrt{0.752} = 0.867$. Since association between age and price is negative, $r = -0.867$.

e) 75.2% of the variability in price of a used Toyota Corolla can be accounted for by variability in the age of the car.

f) The relationship is not perfect. Other factors, such as options, condition, and mileage explain the rest of the variability in price.

60. Drug abuse.

a) The scatterplot shows a positive, strong, linear relationship. It is straight enough to make the linear model the appropriate model.

b) 87.3% of the variability in percentage of other drug usage can be accounted for by percentage of marijuana use.

c) R^2 = 0.873, so $r = \sqrt{0.873} = 0.93434$ (since the relationship is positive).

$$b_1 = r\frac{s_O}{s_M} \qquad \hat{y} = b_0 + b_1 x$$

$$b_1 = (0.93434)\frac{10.2}{15.6} \qquad \overline{y} = b_0 + b_1\overline{x}$$

$$11.6 = b_0 + 0.61091(23.9)$$

$$b_1 = 0.61091 \qquad b_0 = -3.001$$

The regression equation used to predict the percentage of teens that use other drugs from the percentage who have used marijuana is:

$\widehat{Other} = -3.001 + 0.611\,Marijuana$

(Using the data set, and technology: $\widehat{Other} = -3.068 + 0.615\,Marijuana$)

d) According to the model, each additional percent of teens using marijuana is expected to add 0.611 percent to the percentage of teens using other drugs.

e) The results do not *confirm* marijuana as a gateway drug. They do indicate an *association* between marijuana and other drug usage, but association does not imply causation.

61. More used cars 2014.

a) The scatterplot from the previous exercise shows that the relationship is straight, so the linear model is appropriate.
The regression equation to predict the price of a used Toyota Corolla from its age is $\widehat{Price} = 17674 - 844.5\,Years$.

Computer regression output used is at the right.

```
Predictor      Coef  SE Coef       T      P
Constant    17674.0    836.2   21.14  0.000
Age          -844.5    146.1   -5.78  0.000


S = 1224.82   R-Sq = 75.2%   R-Sq(adj) = 73.0%
```

b) According to the model, for each additional year in age, the car is expected to drop \$844.5 in price.

c) The model predicts that a new Toyota Corolla (0 years old) will cost \$17,674.

d)

$\widehat{Price} = 17674 - 844.5\,Years$

$\widehat{Price} = 17674 - 844.5(7)$

$\hat{Price} = 11762.5$

According to the model, an appropriate price for a 7-year old Toyota Corolla is \$11,762.50.

e) Buy the car with the negative residual. Its actual price is lower than predicted.

f) $\widehat{Price} = 17674 - 844.5\,Years$

$\widehat{Price} = 17674 - 844.5(10)$

$\widehat{Price} = 9229$

According to the model, a 10-year-old Corolla is expected to cost \$9229. The car has an actual price of \$8500, so its residual is \$8500 — \$9229 = — \$729

g) The model would not be useful for predicting the price of a 25-year-old Corolla. The oldest car in the list is 9 years old. Predicting a price after 25 years would be an extrapolation.

62. Veggie burgers.

a)
$$\widehat{Fat} = 6.8 + 0.97\,Protein$$
$$\widehat{Fat} = 6.8 + 0.97(14)$$
$$\widehat{Fat} = 20.38$$

According to the model, a burger with 14 grams of protein is expected to have 20.38 grams of fat.

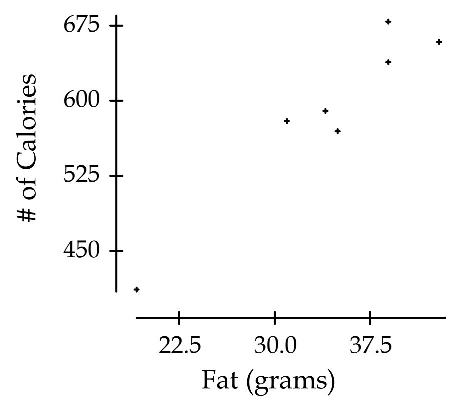
b) From the package, the actual fat content of the veggie burger is 10 grams. The residual is 10 − 20.38 = −10.38 grams of fat. The veggie burgers have about 10.4 fewer grams of fat than is predicted by the model for a regular burger with a similar protein content.

c) The new veggie burger has 14 grams of protein and 10 grams of fat. The veggie burger has about 10.4 fewer grams of fat than a typical regular burger with a similar protein content.

63. Burgers revisited.

a) The scatterplot of calories vs. fat content in fast food hamburgers is at the right. The relationship appears linear, so a linear model is appropriate.

Fat and Calories of Fast Food Burgers



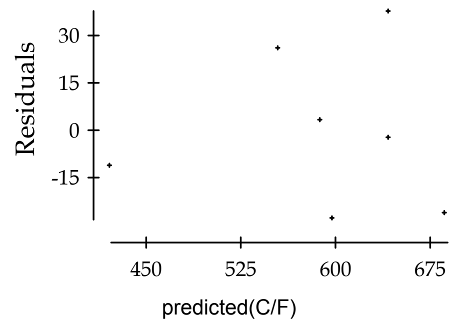
Dependent variable is: **Calories**
No Selector
R squared = 92.3%   R squared (adjusted) = 90.7%
s = 27.33 with 7 - 2 = 5 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	44664.3	1	44664.3	59.8
Residual	3735.73	5	747.146	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	210.954	50.10	4.21	0.0084
Fat	11.0555	1.430	7.73	0.0006

b) From the computer regression output, $R^2 = 92.3\%$. 92.3% of the variability in the number of calories can be explained by the variability in the number of grams of fat in a fast food burger.

c) From the computer regression output, the regression equation that predicts the number of calories in a fast food burger from its fat content is:
$\widehat{Calories} = 210.954 + 11.0555\,Fat$



d) The residuals plot at the right shows no pattern. The linear model appears to be appropriate.

e) The model predicts that a fat free burger would have 210.954 calories. Since there are no data values close to 0, this extrapolation isn't of much use.

f) For each additional gram of fat in a burger, the model predicts an increase of 11.056 calories.

g) $\widehat{Calories} = 210.954 + 11.0555\,Fat = 210.954 + 11.0555(28) = 520.508$
The model predicts a burger with 28 grams of fat will have 520.508 calories. If the residual is +33, the actual number of calories is 520.508 + 33 ≈ 553.5 calories.

64. Chicken.

a) The scatterplot is fairly straight, so the linear model is appropriate.

b) The correlation of 0.947 indicates a strong, linear, positive relationship between fat and calories for chicken sandwiches.

c)
$$b_1 = r\frac{s_{Cal}}{s_{Fat}} \qquad \hat{y} = b_0 + b_1x$$
$$b_1 = (0.947)\frac{144.2}{9.8} \qquad \bar{y} = b_0 + b_1\bar{x}$$
$$b_1 = 13.934429 \qquad 472.7 = b_0 + 13.934429(20.6)$$
$$b_0 = 185.651$$

The linear model for predicting calories from fat in chicken sandwiches is:
$\widehat{Calories} = 185.651 + 13.934\,Fat$

d) For each additional gram of fat, the model predicts an increase in 13.934 calories.

e) According to the model, a fat-free chicken sandwich would have 185.651 calories. This is probably an extrapolation, although without the actual data, we can't be sure.

f) In this context, a negative residual means that a chicken sandwich has fewer calories than the model predicts.

g) For the chicken sandwich:

$$\widehat{Calories} = 185.651 + 13.934\,Fat$$
$$\widehat{Calories} = 185.651 + 13.934(35)$$
$$\widehat{Calories} = 673.341$$

For the burger:

$$\widehat{Calories} = 210.954 + 11.056\,Fat$$
$$\widehat{Calories} = 210.954 + 11.056(35)$$
$$\widehat{Calories} = 597.914$$

A chicken sandwich with 35 grams of fat is predicted to have more calories than a burger with 35 grams of fat.

h) Using the chicken sandwich model:

$$\widehat{Calories} = 185.651 + 13.934\,Fat$$
$$\widehat{Calories} = 185.651 + 13.934(26)$$
$$\widehat{Calories} = 547.935$$

Using the burger model:

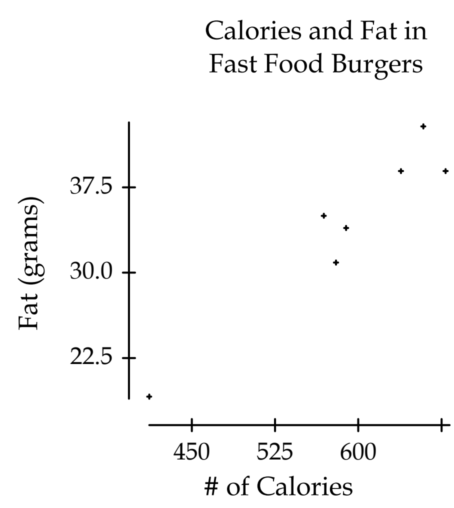
$$\widehat{Calories} = 210.954 + 11.056\,Fat$$
$$\widehat{Calories} = 210.954 + 11.056(26)$$
$$\widehat{Calories} = 498.41$$

A Filet-O-Fish sandwich, at 470 calories, has fewer calories than expected for a burger and many fewer calories than expected for a chicken sandwich. The fish sandwich has a relationship between fat and calories that is similar to the burgers.

65. A second helping of burgers.

a) The model from Exercise 63 was for predicting number of calories from number of grams of fat. In order to predict grams of fat from the number of calories, a new linear model needs to be generated.

b) The scatterplot at the right shows the relationship between number fat grams and number of calories in a set of fast food burgers. The association is strong, positive, and linear. Burgers with higher numbers of calories typically have higher fat contents. The relationship is straight enough to apply a linear model.



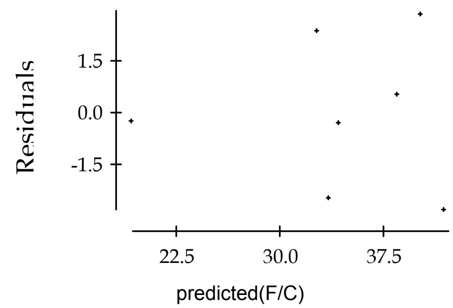
Dependent variable is: Fat
No Selector
R squared = 92.3% R squared (adjusted) = 90.7%
s = 2.375 with 7 - 2 = 5 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	337.223	1	337.223	59.8
Residual	28.2054	5	5.64109	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-14.9622	6.433	-2.33	0.0675
Calories	0.083471	0.0108	7.73	0.0006

The linear model for predicting fat from calories is: $\widehat{Fat} = -14.9622 + 0.083471\,Cal$
The model predicts that for every additional 100 calories, the fat content is expected to increase by about 8.3 grams.

The residuals plot shows no pattern, so the model is appropriate. R^2 = 92.3%, so 92.3% of the variability in fat content can be accounted for by the model.



$$\widehat{Fat} = -14.9622 + 0.083471\,Calories$$
$$\widehat{Fat} = -14.9622 + 0.083471(600)$$
$$\widehat{Fat} \approx 35.1$$

According to the model, a burger with 600 calories is expected to have 35.1 grams of fat.

66. Cost of living 2013.

a) The association between cost of living in 2010 and 2013 is linear, positive and strong. The scatterplot is Straight Enough, so the linear model is appropriate.

b) $R^2 = (0.714)^2 = 0.510$. This means that 51.0% of the variability in cost of living in 2013 can be accounted for by variability in cost of living in 2010.

c) Oslo had a cost of living of 152.85% of New York's in 2010.

$$\widehat{Cost13} = -38.292 + 1.319\,Cost10$$
$$\widehat{Cost13} = -38.292 + 1.319(152.85)$$
$$\widehat{Cost13} = 163.317$$

According to the model, Oslo is predicted to have a cost of living in 2013 that is about 163.317% of New York's. Oslo actually had a cost of living in 2013 that was 170.04% of New York's. Oslo's residual was about +6.72%.

d) Oslo's cost of living in 2013 was about 6.72% more than the cost of living predicted by this model.

67. New York bridges

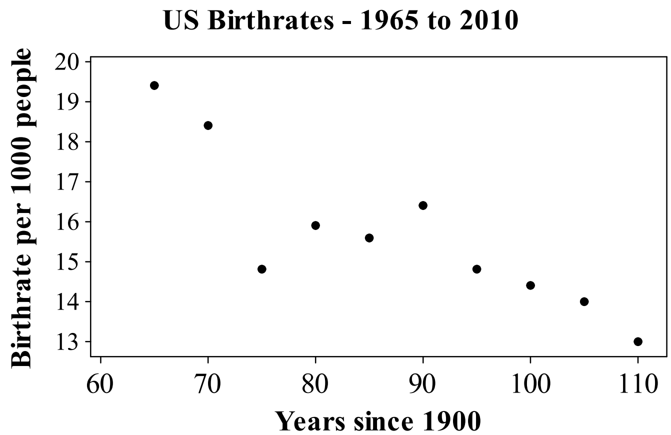
a) Overall, the model predicts the condition score of new bridges to be 4.95, a little below the cutoff of 5. Of course, looking at the scatterplot, we can see that the cluster of new bridges all have scores above 5. Still, the model is not a very encouraging one in regards to the conditions of New York City bridges.

b) According to the model, the condition of the bridges in New York City is decreasing by an average of 0.0048 per year.

c) We shouldn't place too much faith in the model. First of all, the scatterplot shows some curvature, so the linear model may not be appropriate. Even if it is appropriate, R^2 of 2.6% is very low, and the standard deviation of the residuals, 0.6708, is quite high in relation to the scope of the data values themselves. This association is very weak.

68. Birthrates 2010.

a) The scatterplot appears at the right. There is a strong, negative, linear association between year and birthrate. The birthrate has generally decreased over the years.



b) Regression output is at the right.

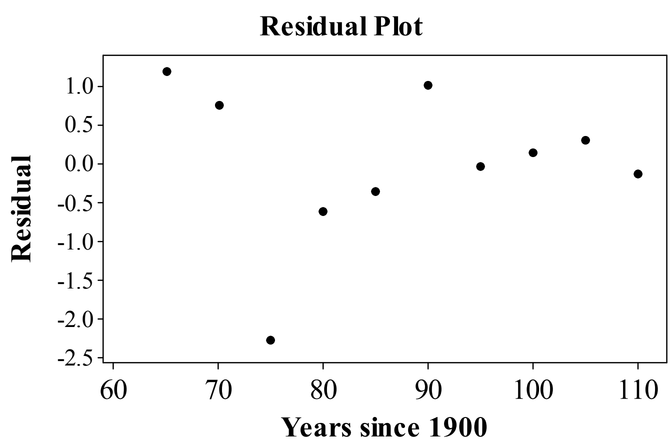
```
Predictor             Coef   SE Coef       T       P
Constant            25.523     2.053   12.43   0.000
Year.since.1900   -0.11261   0.02316   -4.86   0.001


S = 1.05168   R-Sq = 74.7%   R-Sq(adj) = 71.6%
```

The equation of the regression line is: $\widehat{Birthrate} = 25.52 - 0.113(Year - 1900)$

c) There is no apparent pattern in the residuals plot, so the model seems appropriate. There is one low outlier which we may wish to investigate further.

d) The model predicts a decrease in birth rate of about 0.113 births per 1000 people each year.



e) $\widehat{Birthrate} = 25.52 - 0.113(Year - 1900) = 25.52 - 0.113(78) = 16.706$

The model predicts a birth rate of 16.706 births per 1000 population in 1978.

f) The residual is 15.0 – 16.706 = –1.706, which means that the model predicts 1.706 more births per 1000 people than actually occurred.

g) $\widehat{Birthrate} = 25.52 - 0.113(Year - 1900) = 25.52 - 0.113(110) = 13.09$

The model predicts 13.09 (12.91 using technology) births per 1000 women in 2010. This is within the scope of our data, and very close to the actual rate of 13.0.

g) $\widehat{Birthrate} = 25.52 - 0.113(Year - 1900) = 25.52 - 0.113(150) = 8.57$

The model predicts 8.57 (8.63 using technology) births per 1000 women in 2050. We have no faith in this prediction, since it is an extrapolation 40 years beyond the scope of the data.

69. Climate change 2013.

a) The correlation between CO_2level and mean temperature is $r = \sqrt{R^2} = \sqrt{0.736} = 0.858$.

b) 85.8% of the variability in mean temperature can be accounted for by variability in CO_2 level.

c) Since the scatterplot of CO_2 level and mean temperature shows a relationship that is straight enough, use of the linear model is appropriate. The linear regression model that predicts mean temperature from CO_2 level is: $\widehat{MeanTemp} = 10.64 + 0.0103\,CO_2$

d) The model predicts that an increase in CO_2 level of 1 ppm is associated with an increase of 0.0103 °C in mean temperature.

e) According to the model, the mean temperature is predicted to be 10.64 °C when there is no CO_2 in the atmosphere. This is an extrapolation outside of the range of data, and isn't very meaningful in context, since there is always CO_2 in the atmosphere. We want to use this model to study the change in CO_2 level and how it relates to the change in temperature.

f) The residuals plot shows no apparent patterns. The linear model appears to be an appropriate one.

g) $\widehat{MeanTemp} = 10.64 + 0.0103\,CO_2 = 10.64 + 0.0103(400) = 14.76$

According to the model, the temperature is predicted to be 14.76 °C when the CO_2 level is 400 ppm.

h) No. The model describes the association between CO_2 level and temperature. It does not imply that there is a cause and effect relationship between them.

70. Climate change 2013, revisited.

a) The correlation between global average temperature and DJIA is $r = \sqrt{R^2} = \sqrt{0.686} = 0.828$.

b) 68.6% of the variability in mean temperature can be accounted for by variability in DJIA.

c) Since the scatterplot of DJIA and mean temperature shows a relationship that is straight enough, use of the linear model is appropriate. The linear regression model that predicts mean temperature from DJIA is:
$\widehat{MeanTemp} = 14.09 + 0.0000459\,DJIA$.

d) The model predicts an increase of 0.0459 degrees in average temperature for each 1000 point increase in the DJIA.

e) The model predicts that the average temperature would be 14.09 when the DJIA was 0. This doesn't make sense in context.

f) The residuals plot shows no apparent patterns. The linear model appears to be an appropriate one.

g) $\widehat{MeanTemp} = 14.09 + 0.0000459\,DJIA = 14.09 + 0.0000459(20000) = 15.01$

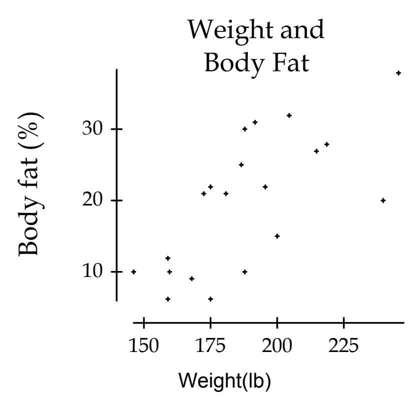
According to the model, the temperature is predicted to be 15.01 °C when the DJIA is 20,000..

h) No. The model describes the association between DJIA and temperature. It does not imply that there is a cause and effect relationship between them.

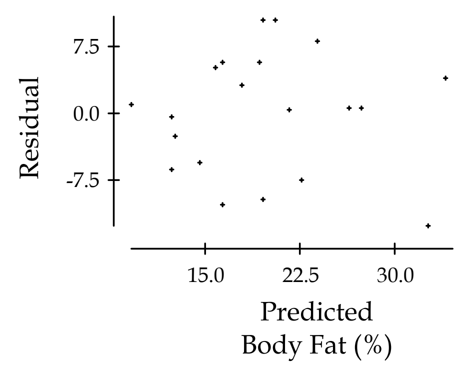
71. Body fat.

a) The scatterplot of % body fat and weight of 20 male subjects, at the right, shows a strong, positive, linear association. Generally, as a subject's weight increases, so does % body fat. The association is straight enough to justify the use of the linear model.

The linear model that predicts % body fat from weight is: $\widehat{\%Fat} = -27.3763 + 0.249874\,Weight$



b) The residuals plot, at the right, shows no apparent pattern. The linear model is appropriate.



c) According to the model, for each additional pound of weight, body fat is expected to increase by about 0.25%.

d) Only 48.5% of the variability in % body fat can be accounted for by the model. The model is not expected to make predictions that are accurate.

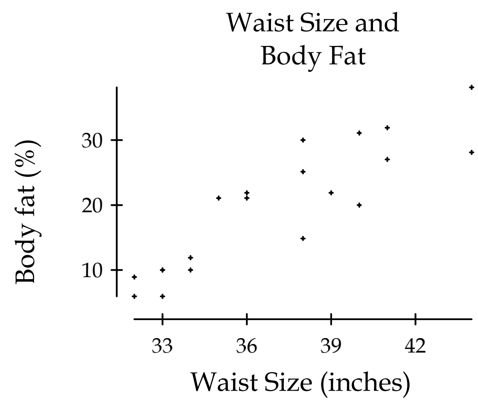
e)

$$\widehat{\%Fat} = -27.3763 + 0.249874\,Weight$$
$$\widehat{\%Fat} = -27.3763 + 0.249874(190)$$
$$\widehat{\%Fat} = 20.09976$$

According to the model, the predicted body fat for a 190-pound man is 20.09976%.
The residual is $21 - 20.09976 \approx 0.9\%$.

72. Body fat again.

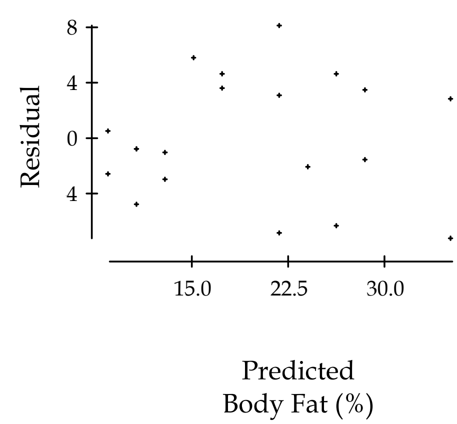
The scatterplot of % body fat and waist size is at the right. The association is strong, linear, and positive. As waist size increases, % body fat has a tendency to increase, as well. The scatterplot is straight enough to justify the use of the linear model.



The linear model for predicting % body fat from waist size is :
$\widehat{\%Fat} = -62.557 + 2.222\,Waist$.

For each additional inch in waist size, the model predicts an increase of 2.222% body fat.

78.7% of the variability in % body fat can be accounted for by waist size. The residuals plot, at right, shows no apparent pattern. The residuals plot and the relatively high value of R^2 indicate an appropriate model with more predicting power than the model based on weight.



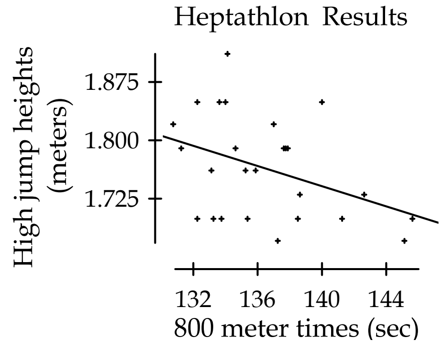
73. Heptathlon revisited.

a) Both high jump height and 800 meter time are quantitative variables, the association is straight enough to use linear regression.

Dependent variable is: **High Jump**
No Selector
R squared = 16.4% R squared (adjusted) = 12.9%
s = 0.0617 with 26 - 2 = 24 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	0.017918	1	0.017918	4.71
Residual	0.091328	24	0.003805	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	2.68094	0.4225	6.35	< 0.0001
800m	-6.71360e-3	0.0031	-2.17	0.0401



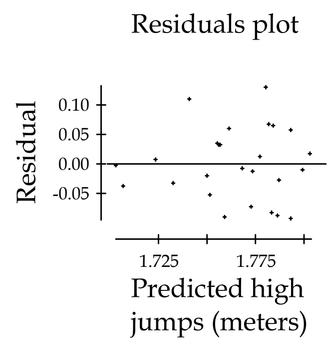
The regression equation to predict high jump from 800m results is: $\widehat{Highjump} = 2.681 - 0.00671\,Time$.

According to the model, the predicted high jump decreases by an average of 0.00671 meters for each additional second in 800 meter time.

b) $R^2 = 16.4\%$. This means that 16.4% of the variability in high jump height is accounted for by the variability in 800 meter time.

c) Yes, good high jumpers tend to be fast runners. The slope of the association is negative. Faster runners tend to jump higher, as well.

d) The residuals plot is fairly patternless. The scatterplot shows a slight tendency for less variation in high jump height among the slower runners than the faster runners. Overall, the linear model is appropriate.



e) The linear model is not particularly useful for predicting high jump performance. First of all, 16.4% of the variability in high jump height is accounted for by the variability in 800 meter time, leaving 83.6% of the variability accounted for by other variables. Secondly, the residual standard deviation is 0.062 meters, which is not much smaller than the standard deviation of all high jumps, 0.066 meters. Predictions are not likely to be accurate.

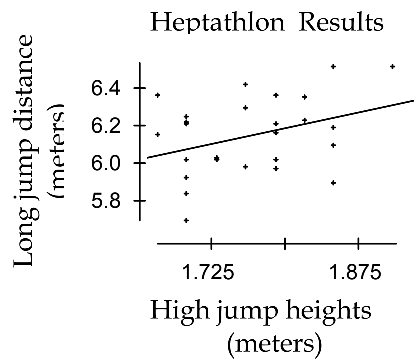
74. Heptathlon revisited again.

a) Both high jump height and long jump distance are quantitative variables, the association is straight enough, and there are no outliers. It is appropriate to use linear regression.

Dependent variable is: **Long Jump**
No Selector
R squared = 12.6% R squared (adjusted) = 9.0%
s = 0.1960 with 26 - 2 = 24 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	0.133491	1	0.133491	3.47
Residual	0.922375	24	0.038432	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	4.20053	1.047	4.01	0.0005
High Jump	1.10541	0.5931	1.86	0.0746



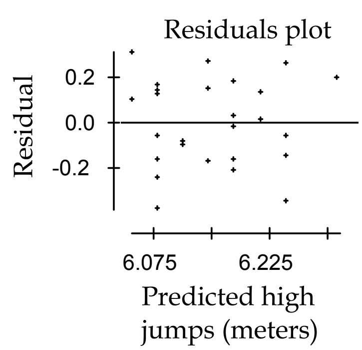
The regression equation to predict long jump from high jump results is:
$\widehat{Longjump} = 4.20053 + 1.10541\,Highjump$.

According to the model, the predicted long jump increases by an average of 1.1054 meters for each additional meter in high jump height.

b) $R^2 = 12.6\%$. This means that only 12.6% of the variability in long jump distance is accounted for by the variability in high jump height.

c) Yes, good high jumpers tend to be good long jumpers. The slope of the association is positive. Better high jumpers tend to be better long jumpers, as well.

d) The residuals plot is fairly patternless. The linear model is appropriate.



e) The linear model is not particularly useful for predicting long jump performance. First of all, only 12.6% of the variability in long jump distance is accounted for by the variability in high jump height, leaving 87.4% of the variability accounted for by other variables. Secondly, the residual standard deviation is 0.196 meters, which is about the same as the standard deviation of all long jumps jumps, 0.206 meters. Predictions are not likely to be accurate.

75. Hard water.

a) There is a fairly strong, negative, linear relationship between calcium concentration (in ppm) in the water and mortality rate (in deaths per 100,000). Towns with higher calcium concentrations tended to have lower mortality rates.

b) The linear regression model that predicts mortality rate from calcium concentration is $\widehat{Mortality} = 1676 - 3.23Calcium$.

c) The model predicts a decrease of 3.23 deaths per 100,000 for each additional ppm of calcium in the water. For towns with no calcium in the water, the model predicts a mortality rate of 1676 deaths per 100,000 people.

d) Exeter had 348.6 fewer deaths per 100,000 people than the model predicts.

e)

$$\widehat{Mortality} = 1676 - 3.23Calcium$$
$$\widehat{Mortality} = 1676 - 3.23(100)$$
$$\widehat{Mortality} = 1353$$

The town of Derby is predicted to have a mortality rate of 1353 deaths per 100,000 people.

f) 43% of the variability in mortality rate can be explained by variability in calcium concentration.

76. Gators.

a) Weight is the proper dependent variable. The researchers can estimate length from the air, and use length to predict weight, as desired.

b) The correlation between an alligator's length and weight is $r = \sqrt{R^2} = \sqrt{0.836} = 0.914$.

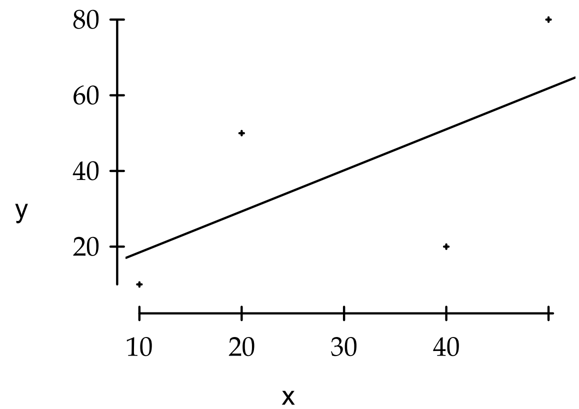
c) The linear regression model that predicts and alligator's weight from its length is $\widehat{Weight} = -393.3 + 5.9Length$.

d) For each additional inch in length, the model predicts an increase of 5.9 pounds in weight.

e) The estimates made using this model should be fairly accurate. The model accounts for 83.6% of the variability in weight. However, care should be taken. With no scatterplot, and no residuals plot, we cannot verify the regression condition of linearity. The association between length and weight may be curved, in which case, the linear model is not appropriate.

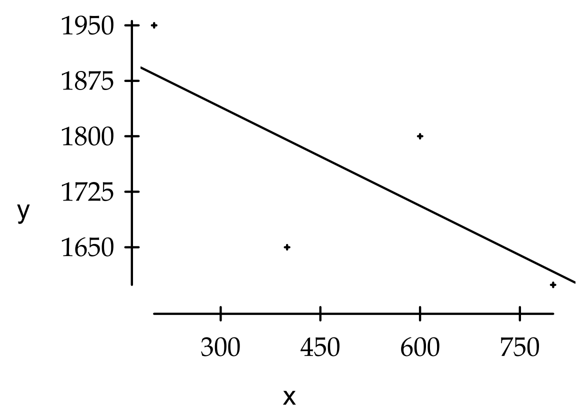
77. Least squares.

If the 4 x-values are plugged into $\hat{y} = 7 + 1.1x$, the 4 predicted values are $\hat{y} = 18$, 29, 51 and 62, respectively. The 4 residuals are –8, 21, –31, and 18. The squared residuals are 64, 441, 961, and 324, respectively. The sum of the squared residuals is 1790. Least squares means that no other line has a sum lower than 1790. In other words, it's the best fit.



78. Least squares.

If the 4 x-values are plugged into $\hat{y} = 1975 - 0.45x$, the 4 predicted values are $\hat{y} = 1885$, 1795, 1705, and 1615, respectively. The 4 residuals are 65, –145, 95, and –15. The squared residuals are 4225, 21025, 9025, and 225, respectively. The sum of the squared residuals is 34,500. Least squares means that no other line has a sum lower than 34,500. In other words, it's the best fit.



Chapter 8 – Regression Wisdom

Section 8.1

1. Credit card spending.

The different segments are not scattered at random throughout the residual plot. Each segment may have a different relationship, which would affect the accuracy of any predictions made with the model that ignores the differences between segments.

2. Revenue and talent cost.

a) There is a positive (revenue increases with talent cost), linear, moderately strong relationship between talent cost and total revenue. There is a possible outlier, an act with both high talent costs and high revenue stands apart from the others.

b) Both venues show an increase of revenue with talent cost.

c) The larger venue has greater variability. Revenue for that venue is more difficult to predict. Additionally, the larger venue typically has both higher total revenue and higher talent cost.

3. Market segments.

Yes, it is clear that the relationship between January and December spending is not the same for all five segments. Using one overall model to predict January spending would be very misleading.

4. Revenue and market sales.

a) There is a positive (revenue increases with ticket sales), linear, strong association between ticket sales and total revenue.

b) Both show a strong, positive association. Each may show some curvature.

c) They differ primarily in magnitude of the values. The larger venue typically has both higher total revenue and higher ticket sales. The larger venue has greater variability. Revenue for that venue is more difficult to predict.

Section 8.2

5. Cell phone costs.

Your friend is extrapolating. It is impossible to know if a trend like this will continue so far into the future.

6. Stopping

Since the model used only data from compact cars, you cannot be certain that this model extends to a large SUV that is much heavier.

7. Revenue and large venues.

a) According to the model, a venue that seats 10,000 would be expected to generate \$354,472 in revenue, if it were to sell out.

$$\widehat{Revenue} = -14,228 + 36.87\,TicketSales$$
$$\widehat{Revenue} = -14,228 + 36.87(10,000)$$
$$\widehat{Revenue} = 354,472$$

b) An extrapolation this far from the data is unreliable. We only have data up to about 3000 seats. 10,000 seats is well above that.

8. Revenue and advanced sales.

The point has high leverage and is an outlier. It will reduce the slope of a fitted regression line and make the line a poor model for the data.

Section 8.3

9. Abalone.

This observation was influential. After it was removed, the correlation and the slope of the regression line both changed by a large amount.

10. Abalone again.

No. Some data points will have higher residuals than others. While large residuals should be looked at carefully, it is not proper to simply remove all those data points. Furthermore, high leverage points often have small residuals, since the can dominate a regression, shifting the regression line toward themselves. Outliers should primarily be indentified by looking at the scatterplot, not the residuals.

Section 8.4

11. Skinned knees.

No. There is a lurking variable, seasonal temperature. In warm weather, more children will go outside and play, and if there are more children playing, there will be more skinned knees.

12. Cell phones.

No. There is a lurking variable, wealth. Wealthier countries typically have more cell phones and better healthcare.

Section 8.5

13. Grading.

Individual student scores will vary greatly. The class averages will have much less variability and may disguise important patterns.

14. Average GPA.

The individual GPAs for each team are going to vary widely. Also, the rest of the team may hide a few individuals with low GPAs. These summaries are a risky method for predicting the students' graduation rates.

Chapter Exercises.

15. Marriage age 2011.

a) The trend in age at first marriage for American women is very strong over the entire time period recorded on the graph, but the direction and form are different for different time periods. The trend appears to be somewhat linear, and consistent at around 22 years, up until about 1940, when the age seemed to drop dramatically, to under 21. From 1940 to about 1970, the trend appears non-linear and slightly positive. From 1975 to the present, the trend again appears linear and positive. The marriage age rose rapidly during this time period.

b) The association between age at first marriage for American women and year is strong over the entire time period recorded on the graph, but some time periods have stronger trends than others.

c) The correlation, or the measure of the degree of linear association is not high for this trend. The graph, as a whole, is non-linear. However, certain time periods, like 1975 to present, have a high correlation.

d) Overall, the linear model is not appropriate. The scatterplot is not Straight Enough to satisfy the condition. You could fit a linear model to the time period from 1975 to 1995, but this seems unnecessary. The ages for each year are reported, and, given the fluctuations in the past, extrapolation seems risky.

16. Smoking 2011.

a) The percent of men 18 – 24 who are smokers decreased dramatically between 1965 and 1990, but the trend has not been consistent since then, though it may be decreasing in a linear fashion since about 1998.

b) The association between percent of men 18 – 24 who smoke and year is very strong from 1965 to 1990, but is erratic after 1990.

c) A linear model is not an appropriate model for the trend in the percent of males 18 – 24 who are smokers. The relationship is not straight.

17. Human Development Index 2012.

a) Fitting a linear model to the association between HDI and GDPPC would be misleading, since the relationship is not straight.

b) If you fit a linear model to these data, the residuals plot will be curved downward.

18. HDI 2012 revisited.

a) Fitting a linear model to the association between the number of cell phones and HDI would be misleading, since the relationship is not straight.

b) The residuals plot will be curved downward.

19. Good model?

a) The student's reasoning is not correct. A scattered residuals plot, not high R^2, is the indicator of an appropriate model. Once the model is deemed appropriate, R^2 is used as a measure of the strength of the model.

b) The model may not allow the student to make accurate predictions. The data may be curved, in which case the linear model would not fit well.

20. Bad model?

a) The student's model may, in fact, be appropriate. Low R^2 simply means that the model is not accurate. The model explains only 13% of the variability in the response variable. If the residuals plot shows no pattern, this model may be appropriate.

b) The predictions are not likely to be very accurate, but they may be the best that the student can get. $R^2 = 13\%$ indicates a great deal of scatter around the regression line, but if the residuals plot is not patterned, there probably isn't a better model. The two variables that are being studied by the student have a weak association.

21. Movie dramas.

a) The units for the slopes of these lines are millions of dollars per minutes of running time.

b) The slopes of the regression lines are about the same. Dramas and movies from other genres have costs for longer movies that increase at the same rate.

c) The regression line for dramas has a lower y-intercept. Regardless of running time, dramas cost about 20 million dollars less than other genres of movies of the same running time.

22. Smoking 2011, women and men.

a) Smoking rates for both men and women in the United States have decreased significantly, but not linearly, over the time period from 1965 to 2011.

b) Smoking rates are generally lower for women than for men.

c) The trend in the smoking rates for women seems a bit straighter than the trend for men. The apparent curvature in the scatterplot for the men could possibly be due to just a few points, and not indicate a serious violation of the linearity condition.

23. Oakland passengers 2013.

a) There are several features to comment on in this plot. There is a strong monthly pattern around the general trend. From 1997 to 2008, passengers increased fairly steadily with a notable exception of Sept. 2001, probably due to the attack on the twin towers. Then sometime in late 2008, departures dropped dramatically, possibly due to the economic crisis. Recently, they have been recovering, but not at the same rate as their previous increase.

b) The trend was fairly linear until late 2008, then passengers dropped suddenly.

c) The trend since 2009 has been linear (overall, ignoring monthly oscillations) If the increase continues to be linear, the predictions should be reasonable for the short term.

24. Tracking hurricanes 2012.

a) According to the linear model, tracking errors averaged about 133 nautical miles in 1970, and have decreased an average of 2 nautical miles per year since then.

b) Residuals based on this model have a standard deviation of 16.13 nautical miles.

c) The linear model for the trend in predicting error is

$\widehat{Error} = 133.024 - 2.0602(Year - 1970)$.

$\widehat{Error} = 133.024 - 2.0602(Year - 1970)$

$\widehat{Error} = 133.024 - 2.0602(45)$

$\widehat{Error} \approx 40.315$

The model predicts an error of 40.3 nautical miles in 2015. This is consistent with the goal of achieving average tracking errors below 45 nautical miles. Of course, this is dependent on a continuation of the pattern.

d) A tracking error of 25 nautical miles is likely to be achieved by about 2023, if the trend fit by the regression model continues, but this is an extrapolation beyond the data. (This is determined by testing different years in the model. This model is NOT meant to be used "backwards", substituting in desired errors.)

e) We should be cautious in assuming that the improvements in prediction will continue at the same rate. They may improve faster, and perhaps the goal will be met. We can't say with any certainty.

25. Unusual points.

a) **1)** The point has high leverage and a small residual.
2) The point is not influential. It has the *potential* to be influential, because its position far from the mean of the explanatory variable gives it high leverage. However, the point is not *exerting* much influence, because it reinforces the association.

3) If the point were removed, the correlation would become weaker. The point heavily reinforces the positive association. Removing it would weaken the association.
4) The slope would remain roughly the same, since the point is not influential.

b) **1)** The point has high leverage and probably has a small residual.
2) The point is influential. The point alone gives the scatterplot the appearance of an overall negative direction, when the points are actually fairly scattered.
3) If the point were removed, the correlation would become weaker. Without the point, there would be very little evidence of linear association.
4) The slope would increase, from a negative slope to a slope near 0. Without the point, the slope of the regression line would be nearly flat.

c) **1)** The point has moderate leverage and a large residual.
2) The point is somewhat influential. It is well away from the mean of the explanatory variable, and has enough leverage to change the slope of the regression line, but only slightly.
3) If the point were removed, the correlation would become stronger. Without the point, the positive association would be reinforced.
4) The slope would increase slightly, becoming steeper after the removal of the point. The regression line would follow the general cloud of points more closely.

d) **1)** The point has little leverage and a large residual.
2) The point is not influential. It is very close to the mean of the explanatory variable, and the regression line is anchored at the point $(\bar{x}, \bar{y})$, and would only pivot if it were possible to minimize the sum of the squared residuals. No amount of pivoting will reduce the residual for the stray point, so the slope would not change.
3) If the point were removed, the correlation would become slightly stronger, decreasing to become more negative. The point detracts from the overall pattern, and its removal would reinforce the association.
4) The slope would remain roughly the same. Since the point is not influential, its removal would not affect the slope.

26. More unusual points.

a) **1)** The point has high leverage and a large residual.
2) The point is influential. It is well away from the mean of the explanatory variable, and has enough leverage to change the slope of the regression line.
3) If the point were removed, the correlation would become stronger. Without the point, the positive association would be reinforced.
4) The slope would increase, becoming steeper after the removal of the point. The regression line would follow the general cloud of points more closely.

b) 1) The point has high leverage and a small residual.
2) The point is influential. The point alone gives the scatterplot the appearance of an overall positive direction, when the points are actually fairly scattered.
3) If the point were removed, the correlation would become weaker. Without the point, there would be very little evidence of linear association.
4) The slope would decrease, from a positive slope to a slope near 0. Without the point, the slope of the regression line would be nearly flat.

c) 1) The point has little leverage and a large residual.
2) The point is not influential. It is very close to the mean of the explanatory variable, and the regression line is anchored at the point $(\bar{x}, \bar{y})$, and would only pivot if it were possible to minimize the sum of the squared residuals. No amount of pivoting will reduce the residual for the stray point, so the slope would not change.
3) If the point were removed, the correlation would become slightly stronger. The point detracts from the overall pattern, and its removal would reinforce the association.
4) The slope would remain roughly the same. Since the point is not influential, its removal would not affect the slope.

d) 1) The point has high leverage and a small residual.
2) The point is not influential. It has the *potential* to be influential, because its position far from the mean of the explanatory variable gives it high leverage. However, the point is not *exerting* much influence, because it reinforces the association.
3) If the point were removed, the correlation would become weaker. The point heavily reinforces the association. Removing it would weaken the association.
4) The slope would remain roughly the same, since the point is not influential.

27. The extra point.

1) Point e is very influential. Its addition will give the appearance of a strong, negative correlation like $r = -0.90$.

2) Point d is influential (but not as influential as point e). Its addition will give the appearance of a weaker, negative correlation like $r = -0.40$.

3) Point c is directly below the middle of the group of points. Its position is directly below the mean of the explanatory variable. It has no influence. Its addition will leave the correlation the same, $r = 0.00$.

4) Point b is almost in the center of the group of points, but not quite. Its addition will give the appearance of a very slight positive correlation like $r = 0.05$.

5) Point a is very influential. Its addition will give the appearance of a strong, positive correlation like $r = 0.75$.

28. The extra point revisited.

1) Point d is influential. Its addition will pull the slope of the regression line toward point d, resulting in the steepest negative slope, a slope of –0.45.

2) Point e is very influential, but since it is far away from the group of points, its addition will only pull the slope down slightly. The slope is –0.30.

3) Point c is directly below the middle of the group of points. Its position is directly below the mean of the explanatory variable. It has no influence. Its addition will leave the slope the same, 0.

4) Point b is almost in the center of the group of points, but not quite. It has very little influence, but what influence it has is positive. The slope will increase very slightly with its addition, to 0.05.

5) Point a is very influential. Its addition will pull the regression line up to its steepest positive slope, 0.85.

29. What's the cause?

1) High blood pressure may cause high body fat.

2) High body fat may cause high blood pressure.

3) Both high blood pressure and high body fat may be caused by a lurking variable, such as a genetic or lifestyle trait.

30. What's the effect?

1) Playing computer games may make kids more violent.

2) Violent kids may like to play computer games.

3) Playing computer games and violence may both be caused by a lurking variable such as the child's home life or a genetic predisposition to aggressiveness.

31. Reading.

a) The principal's description of a strong, positive trend is misleading. First of all, "trend" implies a change over time. These data were gathered during one year, at different grade levels. To observe a trend, one class's reading scores would have to be followed through several years. Second, the strong, positive relationship only indicates the yearly improvement that would be expected, as children get older. For example, the 4^{th} graders are reading at approximately a 4^{th} grade level, on average. This means that the school's students are progressing adequately in their reading, not extraordinarily. Finally, the use of average reading scores instead of individual scores increases the strength of the association.

b) The plot appears very straight. The correlation between grade and reading level is very high, probably between 0.9 and 1.0.

c) If the principal had made a scatterplot of all students' scores, the correlation would have likely been lower. Averaging reduced the scatter, since each grade level has only one point instead of many, which inflates the correlation.

d) If a student is reading at grade level, then that student's reading score should equal his or her grade level. The slope of that relationship is 1. That would be "acceptable", according to the measurement scale of reading level. Any slope greater than 1 would indicate above grade level reading scores, which would certainly be acceptable as well. A slope less than 1 would indicate below grade level average scores, which would be unacceptable.

32. Grades.

Perhaps the best way to start is to discuss the type of graph that would have been useful. The admissions officer should have made a scatterplot with a coordinate for each freshman, matching each individual's SAT score with his or her respective GPA. Then, if the cloud of points was straight enough, the officer could have attempted to fit a linear model, and assessed its appropriateness and strength.

As is, the graph of combined SAT score versus mean Freshman GPA indicates, very generally, that higher SAT achievement is associated with higher mean Freshman GPA, but that's about it.

The first concern is the SAT scores. They have been grouped into categories. We cannot perform any type of regression analysis, because this variable is not quantitative. We don't even know how many students are in each category. There may be one student with an SAT score in the 1500s, and 300 students in the 1200s. On this graph, these possibilities are given equal weight!

Even if the SAT scores were at all useful to us, the GPAs given are averages, which would make the association appear stronger than it actually is.

Finally, a connected line graph isn't a useful model. It doesn't simplify the situation at all, and may, in fact, give the false impression that we could interpolate between the data points.

33. Heating.

a) The model predicts a decrease in $2.13 in heating cost for an increase in temperature of 1° Fahrenheit. Generally, warmer months are associated with lower heating costs.

b) When the temperature is 0° Fahrenheit, the model predicts a monthly heating cost of $133.

c) When the temperature is around 32° Fahrenheit, the predictions are generally too high. The residuals are negative, indicating that the actual values are lower than the predicted values.

d)

$$\widehat{Cost} = 133 - 2.13(Temp)$$
$$\widehat{Cost} = 133 - 2.13(10)$$
$$\widehat{Cost} = \$111.70$$

According to the model, the heating cost in a month with average daily temperature 10° Fahrenheit is expected to be $111.70.

e) The residual for a 10° day is approximately –$6, meaning that the actual cost was $6 less than predicted, or $111.70 – $6 = $105.70.

f) The model is not appropriate. The residuals plot shows a definite curved pattern. The association between monthly heating cost and average daily temperature is not linear.

g) A change of scale from Fahrenheit to Celsius would not affect the relationship. Associations between quantitative variables are the same, no matter what the units.

34. Speed.

a) The model predicts that as speed increases by 1 mile per hour, the fuel economy is expected to decrease by 0.1 miles per gallon.

b) For this model, the *y*-intercept is the predicted mileage at a speed of 0 miles per hour. It's not possible to get 32 miles per gallon if you aren't moving.

c) The residuals are negative for the higher gas mileages. This means that the model is predicting higher than the actual mileage.

d)

$$\widehat{mpg} = 32 - 0.1\,mph$$
$$\widehat{mpg} = 32 - 0.1(50)$$
$$\widehat{mpg} = 27$$

When a car is driven at 50 miles per hour, the model predicts mileage of 27 miles per gallon.

e)

$$\widehat{mpg} = 32 - 0.1\,mph$$
$$\widehat{mpg} = 32 - 0.1(45)$$
$$\widehat{mpg} = 27.5$$

When a car is driven at 45 miles per hour, the model predicts mileage of 27.5 miles per gallon. From the graph, the residual at 27.5 mpg is +1. The actual gas mileage is 27.5 + 1 = 28.5 mpg.

f) The association between fuel economy and speed is probably quite strong, but not linear.

g) The linear model is not the appropriate model for the association between fuel economy and speed. The residuals plot has a clear pattern. If the linear model were appropriate, we would expect scatter in the residuals plot.

35. Interest rates 2014.

a) $r = \sqrt{R^2} = \sqrt{0.776} = 0.88$. The correlation between rate and year is +0.88, since the scatterplot shows a positive association.

b) According to the model, interest rates during this period increased at about 0.25% per year, starting from an interest rate of about 0.61% in 1950.

c) The linear regression equation predicting interest rate from year is

$$\widehat{Rate} = 0.61149 + 0.24788(Year - 1950)$$

$$\widehat{Rate} = 0.61149 + 0.24788(50)$$

$$\widehat{Rate} = 13.00549$$

According to the model, the interest rate is predicted to be about 13% in the year 2000.

d) This prediction is not likely to have been a good one. Extrapolating 20 years beyond the final year in the data would be risky, and unlikely to be accurate.

36. Marriage age, 2011.

a) The correlation between age difference and year is $r = \sqrt{R^2} = \sqrt{0.749} \approx -0.866$. The negative value is used since the scatterplot shows that the association is negative, strong, and linear.

b) The linear regression model that predicts age difference from year is: $\widehat{(Men - Women)} = 32.968 - 0.01549\,Year$. This model predicts that each passing year is associated with a decrease of approximately 0.015 years in the difference between male and female marriage age. A more meaningful comparison might be to say that the model predicts a decrease of approximately 0.15 years in the age difference for every 10 years that pass.

c)

$$\widehat{(Men - Women)} = 32.968 - 0.01549\,Year$$

$$\widehat{(Men - Women)} = 32.968 - 0.01549(2015)$$

$$\widehat{(Men - Women)} = 1.75565$$

According to the model, the age difference between men and women at first marriage is expected to be approximately 1.76 years.
(This figure is very sensitive to the number of decimal places used in the model.)

d) The latest data point is for the year 2011. Extrapolating for 2015 is risky because it depends on the assumption that the trend in age at first marriage will continue in the same manner.

37. Interest rates 2014 revisited.

a) The values of R^2 are approximately the same, so the models fit comparably well, but they have very different slopes.

b) The model that predicts the interest rate on 3-month Treasury bills from the number of years since 1950 is $\widehat{Rate} = 18.87316 - 0.30317(Year - 1950)$. This model predicts the interest rate to be 3.71%, a rate much lower than the prediction from the other model.

c) We can trust the newer prediction, since it is in the middle of the data used to generate the model. Additionally, the model accounts for 77% of the variability in interest rate. We would like to see a residuals plot, though.

d) The first model predicts 17.96% and the second model predicts –2.35%.

e) Since 2020 is 7 years after the last year included in the newer model, it would be extremely risky to use this, or any, model to make a prediction that far into the future. The first model predicts historically high, and the second gives a negative prediction. Neither is a legitimate interest rate prediction.

38. Marriage age 2011 again.

a) The linear model is appropriate, since the scatterplot of the relationship between difference in age at first marriage and the year is reasonably straight, and the residuals plot is scattered.

b) For every 10 years that pass, the model predicts a decrease of approximately 0.23 years in average age difference at first marriage.

c) The y-intercept is the prediction of the model in year 0, over 2000 years ago. An extrapolation that far into the past is not meaningful. The earliest year for which we have data is 1980.

39. Gestation.

a) The association would be stronger if humans were removed. The point on the scatterplot representing human gestation and life expectancy is an outlier from the overall pattern and detracts from the association. Humans also represent an influential point. Removing the humans would cause the slope of the linear regression model to increase, following the pattern of the non-human animals much more closely.

b) The study could be restricted to non-human animals. This appears justifiable, since one could point to a number of environmental factors that could influence human life expectancy and gestation period, making them incomparable to those of animals.

c) The correlation is moderately strong. The model explains 72.2% of the variability in gestation period of non-human animals.

d) For every year increase in life expectancy, the model predicts an increase of approximately 15.5 days in gestation period.

e)

$$\widehat{Gest} = -39.5172 + 15.4980\,LifEx$$
$$\widehat{Gest} = -39.5172 + 15.4980(20)$$
$$\widehat{Gest} \approx 270.4428$$

According to the linear model, monkeys with a life expectancy of 20 years are expected to have gestation periods of about 270.5 days. Care should be taken when assessing the accuracy of this prediction. First of all, the residuals plot has not been examined, so the appropriateness of the model is questionable. Second, it is unknown whether or not monkeys were included in the original 17 non-human species studied. Since monkeys and humans are both primates, the monkeys may depart from the overall pattern as well.

40. Swim the lake 2013.

a) Only 1.2% of the variability in lake swim times is accounted for by the linear model.

b) The slope of the regression, 3.986, means that the model predicts that lake swim times are increasing by about 3.986 minutes per year. This means that lake swimmers are generally getting slower. However, this model has very weak predicting power, and an outlier, so we shouldn't put too much faith in our prediction.

c) Removing this outlier is probably a good idea, since it doesn't belong with the other data points, but its removal probably wouldn't change the regression much. The fact that the point has a large residual indicates that it didn't have much leverage. If it had leverage, it would have dominated the regression, and had a small residual. It would be nice to have a scatterplot to look at, in addition to the residuals plot. There could be other outliers that don't show up in the residuals plot.

41. Elephants and hippos.

a) Hippos are more of a departure from the pattern. Removing that point would make the association appear to be stronger.

b) The slope of the regression line would increase, pivoting away from the hippos point.

c) Anytime data points are removed, there must be a justifiable reason for doing so, and saying, "I removed the point because the correlation was higher without it" is not a justifiable reason.

d) Elephants are an influential point. With the elephants included, the slope of the linear model is 15.4980 days gestation per year of life expectancy. When they are removed, the slope is 11.6 days per year. The decrease is significant.

42. Another swim 2013.

a) The smaller value of s_e means that errors in prediction are smaller for this model than the original model.

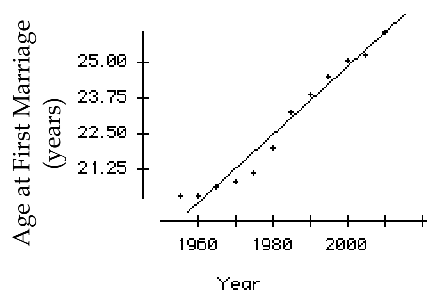
b) The regression accounts for only 4.5% of the variation in lake swim times, but it appears that Lake Ontario swimmers are getting slower, at a rate of about 5.691 seconds per year.

43. Marriage age 2010, revisited.

a) Modeling decisions may vary, but the important idea is using a subset of the data that allows us to make an accurate prediction for the year in which we are interested. We might model a subset to predict the marriage age in 2015, and model another subset to predict the marriage age in 1911.

In order to predict the average marriage age of American women in 2015, use the data points from the most recent trend only. The data points from 1955 – 2010 look straight enough to apply the linear regression model. Even though there is still curvature, it is nowhere near as bad as the curvature when all the data points are used.

Female Marriage Age

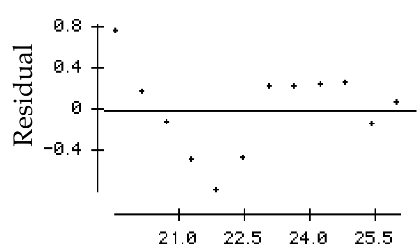


Regression output from a computer program is given below, as well as a residual plot.

Dependent variable is: **Female Marriage Age**
No Selector
R squared = 96.1%   R squared (adjusted) = 95.7%
s = 0.4479 with 12 - 2 = 10 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	49.6368	1	49.6368	247
Residual	2.0057	10	0.20057	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-210.827	14.85	-14.2	≤ 0.0001
Year	0.117832	0.00749	15.7	≤ 0.0001



Predicted Age (years)

The linear model used to predict average female marriage age from year is:

$\widehat{Age} = -210.827 + 0.117832\,Year$. The residuals plot shows a pattern, but the residuals are small, and the value of R^2 is high. 96.1% of the variability in average female age at first marriage is accounted for by variability in the year. The model predicts that each year that passes is associated with an increase of 0.118 years in the average female age at first marriage.

$$\widehat{Age} = -210.827 + 0.117832\,Year$$
$$\widehat{Age} = -210.827 + 0.117832(2020)$$
$$\widehat{Age} = 27.19$$

According to the model, the average age at first marriage for women in 2015 will be 27.19 years old. Care should be taken with this prediction, however. It represents an extrapolation of 10 years beyond the highest year, and the residuals plot shows a pattern.

b) This prediction is for a year that is 10 years higher than the highest year for which we have an average female marriage age. Don't place too much faith in this extrapolation.

c) An extrapolation of more than 50 years into the future would be absurd. There is no reason to believe the trend would continue. In fact, given the situation, it is very unlikely that the pattern would continue in this fashion. The model given in part a) predicts that the average marriage age will be 32.5 years in 2065. Realistically, that seems quite high.

44. Bridges covered.

a) The linear model is $\widehat{Condition} = -37.7775 + 0.02190\,Year$, so a bridge built in 1853 is expected to have a condition of 2.8032. The residual is 4.57 – 2.8032 = 1.77.

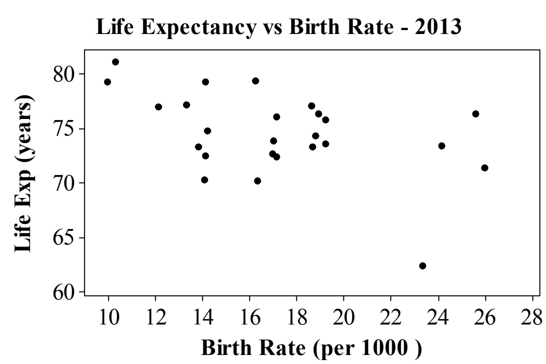
b) A point to the left of the overall group, and higher than expected would have high leverage. This point would pull the regression line toward it, lowering the regression slope.

c) The covered bridge does not fit the pattern we see in the scatterplot, so including it would lower R^2.

d) According to the linear model, a bridge built in 1972 is expected to have condition equal to $-37.7775 + 0.02190(1972) = 5.4093$, which is actually a little higher than the actual condition of 4.523. When you consider the restoration, the bridge isn't remarkable.

45. Life expectancy 2013.

a) The scatterplot of birth rate and life expectancy is at the right. The association is moderate, linear, and negative.



b) There is one outlier, Haiti, with a birthrate of 23.35 births per 1000 people, and a life expectancy of 62.4 years.

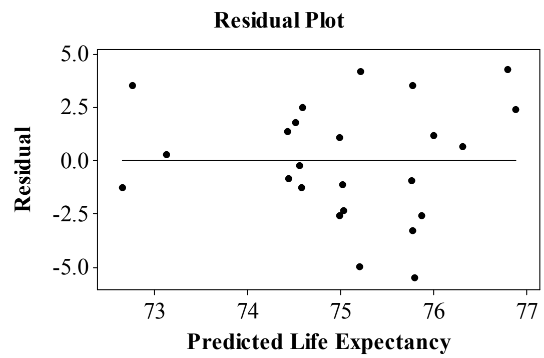
c) Computer regression output is given below.

```
Predictor        Coef  SE Coef       T       P
Constant       79.504    2.418   32.87   0.000
Birth Rate    -0.2640   0.1390   -1.90   0.071


S = 2.79773   R-Sq = 14.1%   R-Sq(adj) = 10.2%
```

c) The linear regression equation that predicts life expectancy from birth rate is:

$\widehat{LifeExp} = 79.504 - 0.2640\,Birthrate$



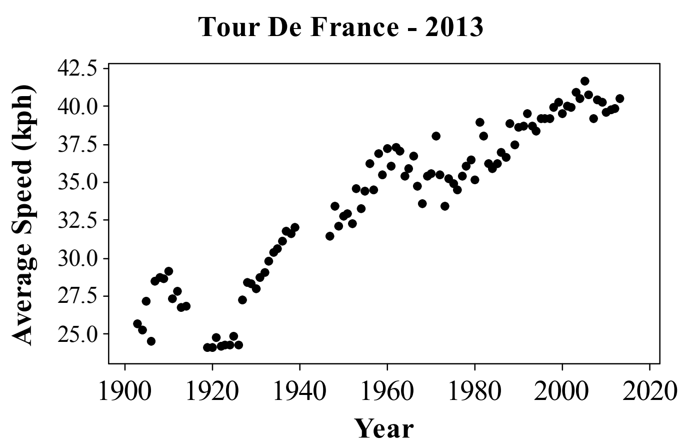
d) The residuals plot, at the right, is reasonably scattered.

e) The linear model is appropriate.

f) $R^2 = 14.1\%$, so 14.1% of the variability in life expectancy is explained by variability in the birthrate.

g) The government leaders should not suggest that women have fewer children in order to raise the life expectancy. Although there is evidence of an association between the birth rate and life expectancy, this does not mean that one causes the other. There may be lurking variables involved, such as economic conditions, social factors, or level of health care.

46. Tour de France 2013.



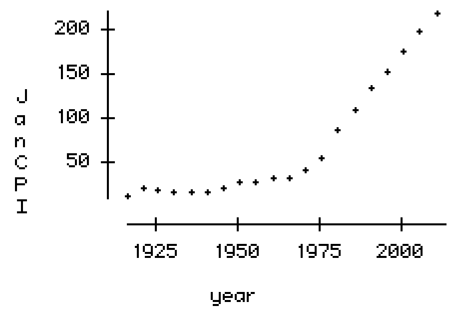
a) The association between average speed and year is positive, moderate, but not quite linear. Generally, average speed of the winner has been increasing over time. There are several periods where the relationship is curved, but since 1950, the relationship has been much more linear. There are no races between 1915 and 1918 or between 1940 and 1946, presumably because of the two World Wars in Europe at the times.

b) $\widehat{Avgspeed} = -258.75 + 0.149246\,Year$

c) The conditions for regression are not met. Although the variables are quantitative, and there are no outliers, the relationship is not straight enough in the early part of the 20th century to fit a regression line.

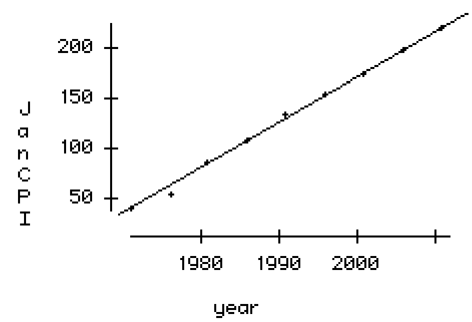
47. Inflation 2011.

a) The trend in Consumer Price Index is strong, non-linear, and positive. Generally, CPI has increased over the years, but the rate of increase has become much greater since approximately 1970. Other characteristics include fluctuations in CPI in the years prior to 1950.

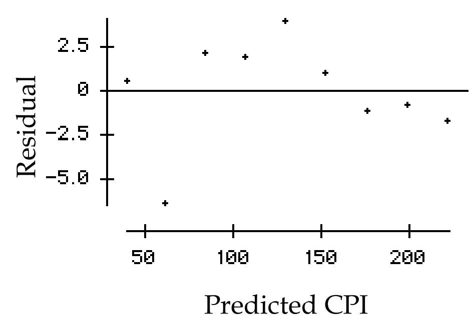


b) In order to effectively predict the CPI in 2016, use only the most recent trend. The trend since 1970 is straight enough to apply the linear model. Prior to 1970, the trend is radically different from that of recent years, and is of no use in predicting CPI for the next decade.

Consumer Price Index



The linear model that predicts CPI from year is $\widehat{CPI} = -8966.98 + 4.56931\,Year$. $R^2 = 99.8\%$, meaning that the model predicts 99.8% of the variability in CPI. The residuals plot shows some pattern, but the residuals are small, so the linear model is appropriate. According to the model, the CPI is expected to increase by \$4.57 each year, for 1970 – 2011.



$$\widehat{CPI} = -8966.98 + 4.56931\,Year$$

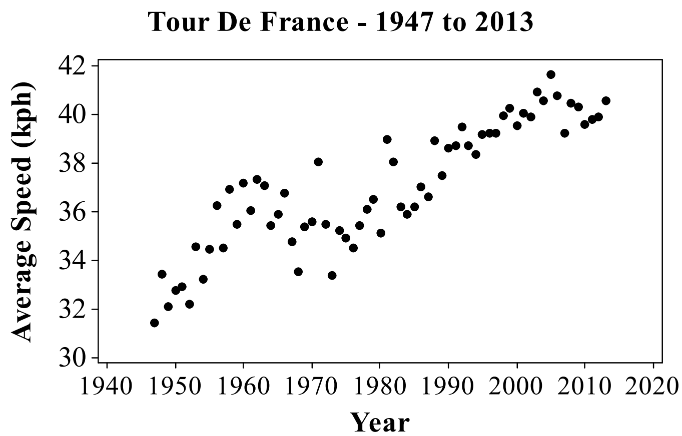
$$\widehat{CPI} = -8966.98 + 4.56931(2020)$$

$$\widehat{CPI} = 263.03$$

As with any model, care should be taken when extrapolating. If the pattern continues, the model predicts that the CPI in 2020 will be approximately \$263.03.

48. Second stage 2013.

a) There is still some curving in the beginning of the period, with speeds increasing a high rate in the post-war years, then they leveled out a bit. However, the relationship is much straighter with this subset of years.



The new linear model is $\widehat{Avgspeed} = -193.12 + 0.116236\,Year$.

b) According to the linear model, the average winning speed increases by about 0.116 kph per year.

c) Hinault's 1981 time has a residual of 2.72 kph. He raced about 2.72 kph faster than the model would predict. Froome's 2013 speed has a residual of –0.31. He raced about than 0.3 kph slower than the model predited. Hinault's performance was more impressive for his era.

Chapter 9 – Re-expressing Data: Get It Straight!

Section 9.1

1. Residuals.

a) The residuals plot shows no pattern. No re-expression is needed.

b) The residuals plot shows a curved pattern. Re-express to straighten the relationship.

c) The residuals plot shows a fan shape. Re-express to equalize spread.

2. More residuals.

a) The residuals plot shows a curved pattern. Re-express to straighten the relationship.

b) The residuals plot shows a fan shape. Re-express to equalize spread.

c) The residuals plot shows no pattern. No re-expression is needed.

3. BK protein.

The goal of this re-expression is to improve homoscedasticity. We desire more equal spread between groups.

4. Life expectancy and TV.

The goal of this re-expression is to straighten the plot.

Section 9.2

5. BK protein again.

The log re-expression is still preferable. The square root doesn't make the spreads as nearly equal. The reciprocal clearly goes too far on the Ladder of Powers.

6. Life expectancy and TV again.

The bend in the plot now goes other way, so we have moved too far on the Ladder of Powers.

Chapter Exercises.

7. Oakland passengers 2013 revisited.

a) The residuals cycle up and down because there is a yearly pattern in the number of passengers departing Oakland, California. There is also a sudden decrease in passenger traffic after 2008.

b) A re-expression should not be tried. A cyclic pattern such as this one cannot be helped by re-expression.

8. Hopkins winds, revisited.

a) The plot shows a wavy pattern, indicating a pattern that continues year to year as part of an annual cycle.

b) A re-expression should not be tried. A cyclic pattern such as this one cannot be helped by re-expression.

9. Models.

a)
$$\begin{aligned} \ln\hat{y} &= 1.2+0.8x \\ \ln\hat{y} &= 1.2+0.8(2) \\ \ln\hat{y} &= 2.8 \\ \hat{y} &= e^{2.8} = 16.44 \end{aligned}$$

b)
$$\begin{aligned} \sqrt{\hat{y}} &= 1.2+0.8x \\ \sqrt{\hat{y}} &= 1.2+0.8(2) \\ \sqrt{\hat{y}} &= 2.8 \\ \hat{y} &= 2.8^2 = 7.84 \end{aligned}$$

c)
$$\begin{aligned} \frac{1}{\hat{y}} &= 1.2+0.8x \\ \frac{1}{\hat{y}} &= 1.2+0.8(2) \\ \frac{1}{\hat{y}} &= 2.8 \\ \hat{y} &= \frac{1}{2.8} = 0.36 \end{aligned}$$

d)
$$\begin{aligned} \hat{y} &= 1.2+0.8\ln x \\ \hat{y} &= 1.2+0.8\ln(2) \\ \hat{y} &= 1.75 \end{aligned}$$

e)
$$\begin{aligned} \log\hat{y} &= 1.2+0.8\log x \\ \log\hat{y} &= 1.2+0.8\log(2) \\ \log\hat{y} &= 1.440823997... \\ \hat{y} &= 10^{1.4408...} \\ \hat{y} &= 27.59 \end{aligned}$$

10. More models.

a)
$$\begin{aligned} \hat{y} &= 1.2+0.8\log x \\ \hat{y} &= 1.2+0.8\log(2) \\ \hat{y} &= 1.44 \end{aligned}$$

b)
$$\begin{aligned} \log\hat{y} &= 1.2+0.8x \\ \log\hat{y} &= 1.2+0.8(2) \\ \log\hat{y} &= 2.8 \\ \hat{y} &= 10^{2.8} = 630.96 \end{aligned}$$

c)
$$\begin{aligned} \ln\hat{y} &= 1.2+0.8\ln x \\ \ln\hat{y} &= 1.2+0.8\ln(2) \\ \ln\hat{y} &= 1.7545... \\ \hat{y} &= e^{1.7545..} = 5.78 \end{aligned}$$

d)
$$\begin{aligned} \hat{y}^2 &= 1.2+0.8x \\ \hat{y}^2 &= 1.2+0.8(2) \\ \hat{y}^2 &= 2.8 \\ \hat{y} &= \sqrt{2.8} = 1.67 \end{aligned}$$

e)
$$\begin{aligned} \frac{1}{\sqrt{\hat{y}}} &= 1.2+0.8x \\ \frac{1}{\sqrt{\hat{y}}} &= 1.2+0.8(2) \\ \frac{1}{\sqrt{\hat{y}}} &= 2.8 \\ \hat{y} &= \frac{1}{2.8^2} = 0.128 \end{aligned}$$

11. Gas mileage.

a) The association between weight and gas mileage of cars is fairly linear, strong, and negative. Heavier cars tend to have lower gas mileage.

b) For each additional thousand pounds of weight, the linear model predicts a decrease of 7.652 miles per gallon in gas mileage.

c) The linear model is not appropriate. There is a curved pattern in the residuals plot. The model tends to underestimate gas mileage for cars with relatively low and high gas mileages, and overestimates the gas mileage of cars with average gas mileage.

12. Crowdedness.

a) The scatterplot shows that the relationship between Crowdedness and GDP is strong, negative, and curved. Re-expression may yield an association that is more linear.

b) Start with logs, since GDP is non-negative. A plot of the log of GDP against Crowdedness score may be straighter.

13. Gas mileage revisited.

a) The residuals plot for the re-expressed relationship is much more scattered. This is an indication of an appropriate model.

b) The linear model that predicts the number of gallons per 100 miles in gas mileage from the weight of a car is: $\widehat{Gal/100} = 0.625 + 1.178(Weight)$.

c) For each additional 1000 pounds of weight, the model predicts that the car will require an additional 1.178 gallons to drive 100 miles.

d)

$$\widehat{Gal/100} = 0.625 + 1.178(Weight)$$
$$\widehat{Gal/100} = 0.625 + 1.178(3.5)$$
$$\widehat{Gal/100} = 4.748$$

According to the model, a car that weighs 3500 pounds (3.5 thousand pounds) is expected to require approximately 4.748 gallons to drive 100 miles, or 0.04748 gallons per mile.

This is $\dfrac{1}{0.04748} \approx 21.06$ miles per gallon.

14. Crowdedness again.

a) This re-expression is not useful. The student has gone too far down the ladder of powers. We now see marked downward curvature and increasing scatter.

b) The next step would be to try a "weaker" re-expression, like reciprocal square root or log of GDP. Having gone too far, the student should move back "up" the ladder of powers.

15. GDP 2013.

a) Although more than 98% of the variation in GDP can be accounted for by the model, the residuals plot should be examined to determine whether or not the model is appropriate.

b) This is not a good model for these data. The residuals plot shows curvature.

16. Interest rates 2014, once more.

Re-expression should not be tried. An erratic trend that is positive then negative cannot be straightened by re-expression.

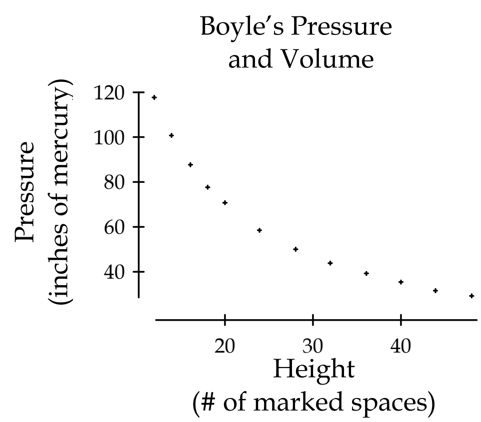
17. Better GDP model?

There is still a pattern in the residuals. This much pattern still indicates an inappropriate model.

18. Pressure.

The scatterplot at the right shows a strong, curved, negative association between the height of the cylinder and the pressure inside. Because of the curved nature of the association, a linear model is not appropriate.

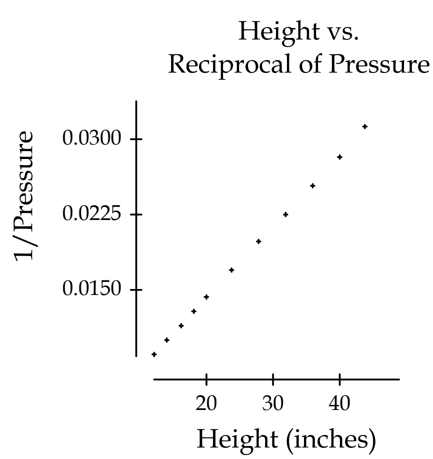
Re-expressing the pressure as the reciprocal of the pressure produces a scatterplot that is much straighter. Computer regression output for the height versus the reciprocal of pressure is below.



Dependent variable is: **recip pressure**
No Selector
R squared = 100.0%   R squared (adjusted) = 100.0%
s = 0.0001 with 12 - 2 = 10 degrees of freedom

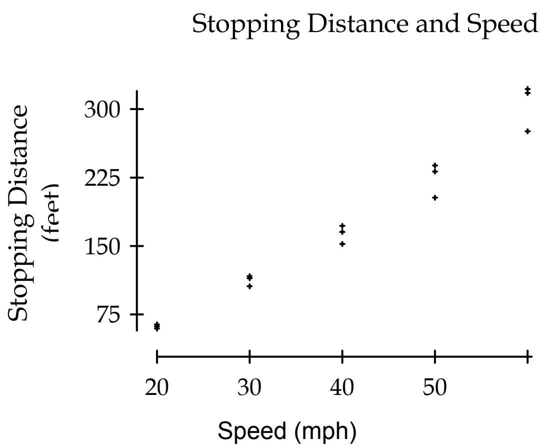
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	0.000841	1	0.000841	75241
Residual	0.000000	10	0.000000	

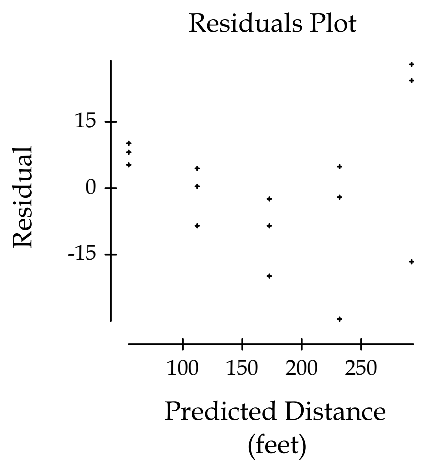
Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-7.66970e-5	0.0001	-0.982	0.3494
Height	7.13072e-4	0.0000	274	< 0.0001



The reciprocal re-expression is very straight (perfectly straight, as far as the statistical software is concerned!). $R^2 = 100\%$, meaning that 100% of the variability in the reciprocal of pressure is explained by the model. The equation of the model is: $\widehat{\dfrac{1}{Pressure}} = -0.000077 + 0.000713(Height)$.

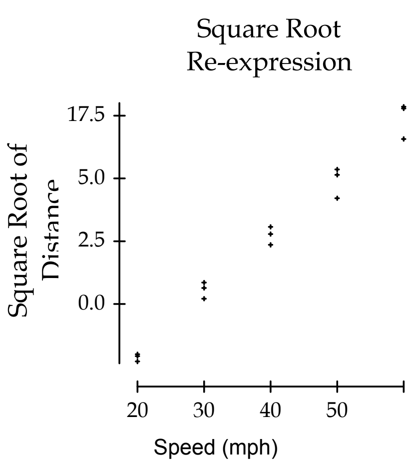
19. Brakes.

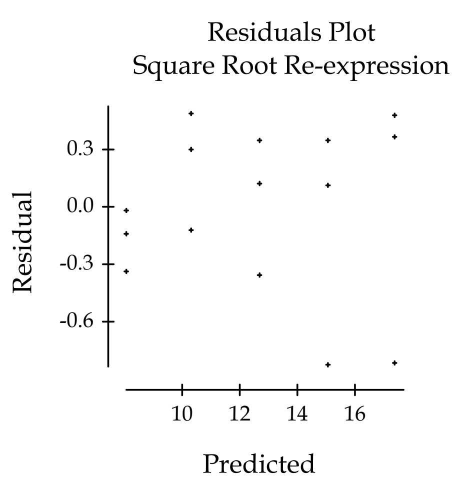




a) The association between speed and stopping distance is strong, positive, and appears straight. Higher speeds are generally associated with greater stopping distances. The linear regression model, with equation $\widehat{Distance} = -65.9 + 5.98(Speed)$, has $R^2 = 96.9\%$, meaning that the model explains 96.9% of the variability in stopping distance. However, the residuals plot has a curved pattern. The linear model is not appropriate. A model using re-expressed variables should be used.

b) Stopping distances appear to be relatively higher for higher speeds. This increase in the rate of change might be able to be straightened by taking the square root of the response variable, stopping distance. The scatterplot of Speed versus $\sqrt{Distance}$ seems like it might be a bit straighter.





c) The model for the re-expressed data is $\widehat{\sqrt{Distance}} = 3.303 + 0.235(Speed)$. The residuals plot shows no pattern, and $R^2 = 98.4\%$, so 98.4% of the variability in the square root of the stopping distance can be explained by the model.

d)

$$\widehat{\sqrt{Distance}} = 3.303 + 0.235(Speed)$$
$$\widehat{\sqrt{Distance}} = 3.303 + 0.235(55)$$
$$\widehat{\sqrt{Distance}} = 16.228$$
$$\widehat{Distance} = 16.228^2 \approx 263.4$$

According to the model, a car traveling 55 mph is expected to require approximately 263.4 feet to come to a stop.

e)

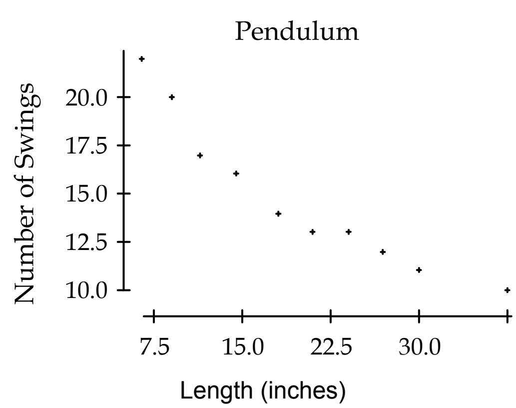
$$\widehat{\sqrt{Distance}} = 3.303 + 0.235(Speed)$$
$$\widehat{\sqrt{Distance}} = 3.303 + 0.235(70)$$
$$\widehat{\sqrt{Distance}} = 19.753$$
$$\widehat{Distance} = 19.753^2 \approx 390.2$$

According to the model, a car traveling 70 mph is expected to require approximately 390.2 feet to come to a stop.

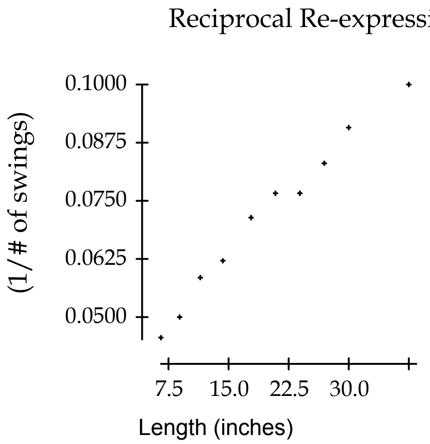
f) The level of confidence in the predictions should be quite high. R^2 is high, and the residuals plot is scattered. The prediction for 70 mph is a bit of an extrapolation, but should still be reasonably close.

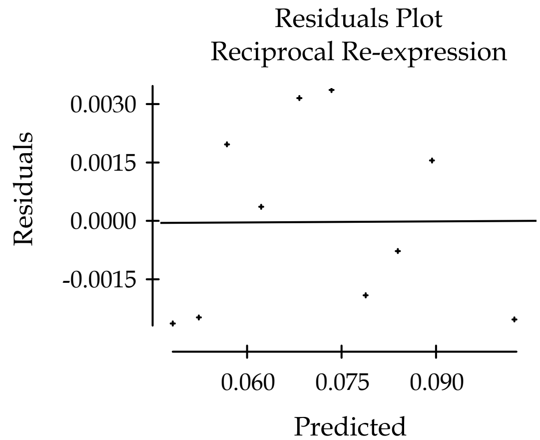
20. Pendulum.

a) The scatterplot shows the association between the length of string and the number of swings a pendulum took every 20 seconds to be strong, negative, and curved. A pendulum with a longer string tended to take fewer swings in 20 seconds. The linear model is not appropriate, because the association is curved.



b) Curvature in a negative relationship sometimes is an indication of a reciprocal relationship. Try re-expressing the response variable with the reciprocal.





c) The reciprocal re-expression yields the model $\widehat{\frac{1}{Swings}} = 0.0367 + 0.00176(Length)$.

The residuals plot is scattered, and R^2 = 98.1%, indicating that the model explains 98.1% of the variability in the reciprocal of the number of swings. The model is both appropriate and accurate.

d) $\widehat{\frac{1}{Swings}} = 0.0367 + 0.00176(Length) = 0.0367 + 0.00176(4) = 0.04374$

$\widehat{Swings} = \frac{1}{0.04374} \approx 22.9$

According to the reciprocal model, a pendulum with a 4″ string is expected to swing approximately 22.9 times in 20 seconds.

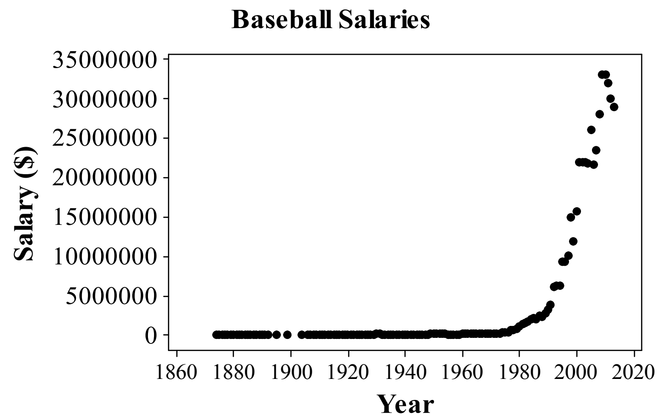
e) $\widehat{\frac{1}{Swings}} = 0.0367 + 0.00176(Length) = 0.0367 + 0.00176(48) = 0.12118$

$\widehat{Swings} = \frac{1}{0.12118} \approx 8.3$. The model predicts 8.3 swings in 20 seconds for a 48″ string.

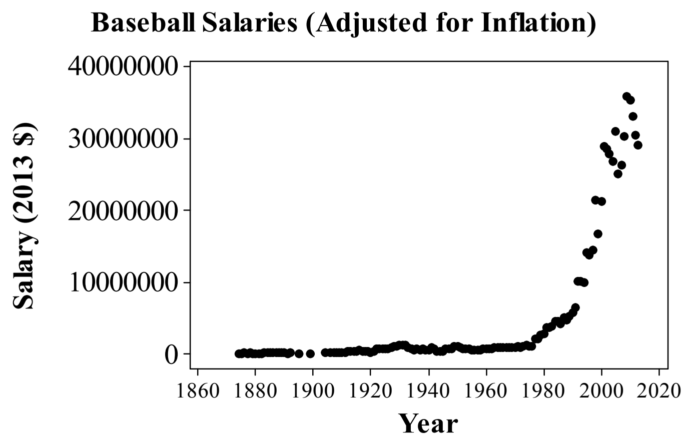
f) Confidence in the predictions is fairly high. The model is appropriate, as indicated by the scattered residuals plot, and accurate, indicated by the high value of R^2. The only concern is the fact that these predictions are slight extrapolations. The lengths of the strings aren't too far outside the range of the data, so the predictions should be reasonably accurate.

21. Baseball salaries 2013.

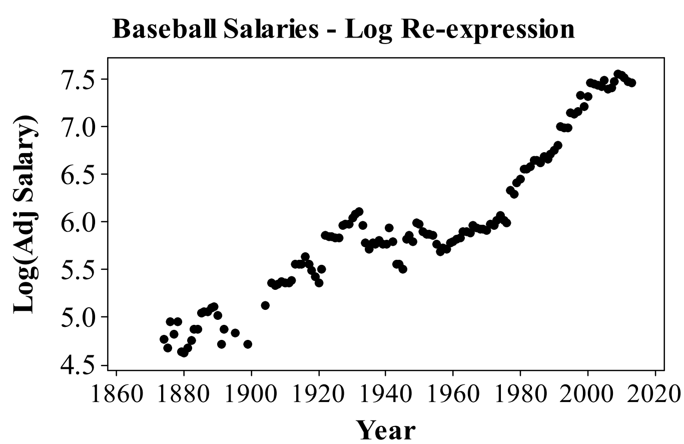
a) The association between year and highest salary is strong and positive. Salaries were flat for many years, and began to increase in the 1970s, then increased more rapidly as in recent years.



b) The association between year and adjusted salary is still very curved.



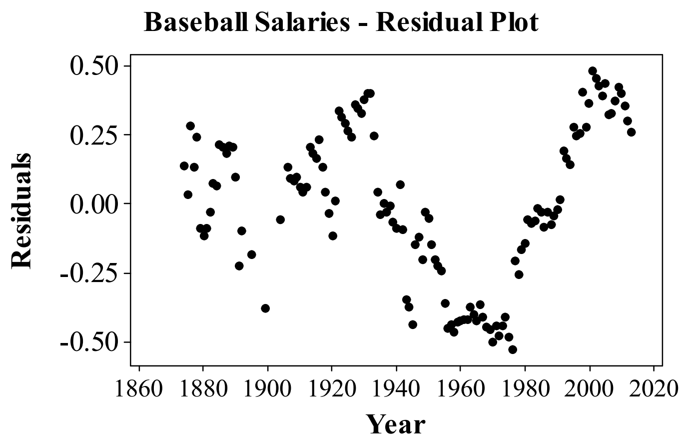
c) Re-expression using the logarithm of the adjusted salaries straightens the plot significantly.



d) $\log(\widehat{AdjSalary}) = -30.022 + 0.0185(Year)$

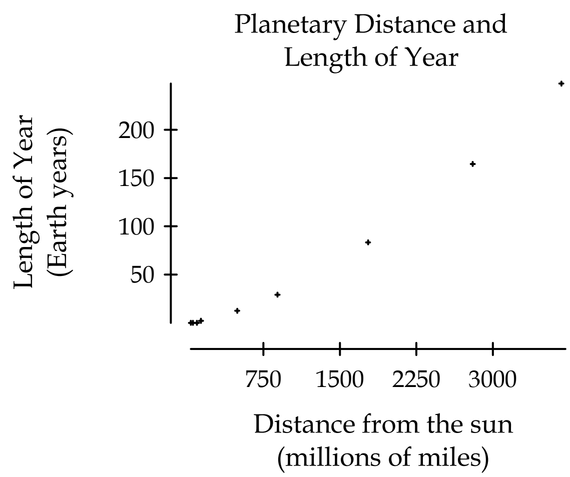
e) $R^2 = 87.7\%$, so the model explains 87.7% of the variability in the logarithm of the inflation-adjusted salary.

f) The plot of year versus the residuals is at the right.



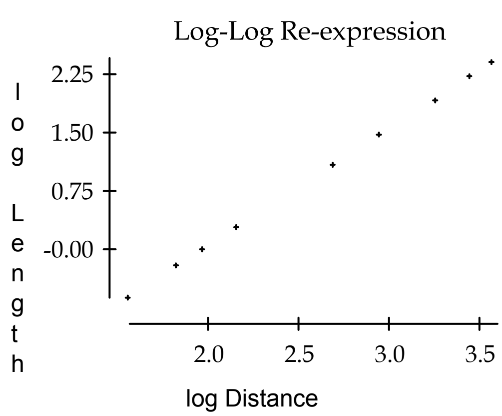
g) The residuals show a cyclic pattern. Salaries were lower than the model predicts at the beginning of the 20th century and again in the 1960s. They may recently hit a high point and started turning lower, but it is difficult to tell. The model based on the logarithmic re-expression may be the best model we can find, but it doesn't explain the pattern in highest baseball salary over time.

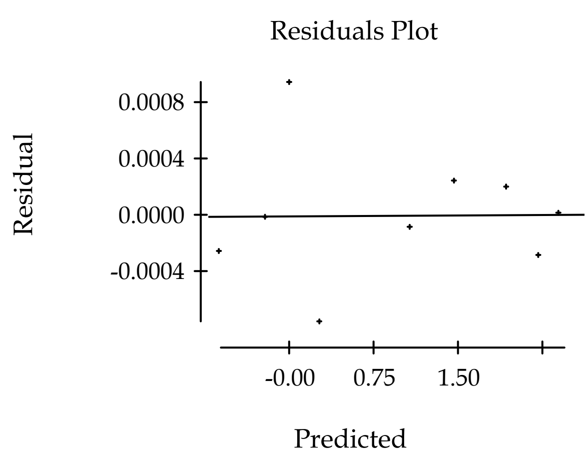
22. Planets, distances and years.



a) The association between distance from the sun and planet year is strong, positive, and curved concave upward. Generally, planets farther from the sun have longer years than closer planets.

b) The rate of change in length of year per unit distance appears to be increasing, but not exponentially. Re-expressing with the logarithm of each variable may straighten a plot such as this. The scatterplot and residuals plot for the linear model relating log(Distance) and log(Length of Year) appear below.



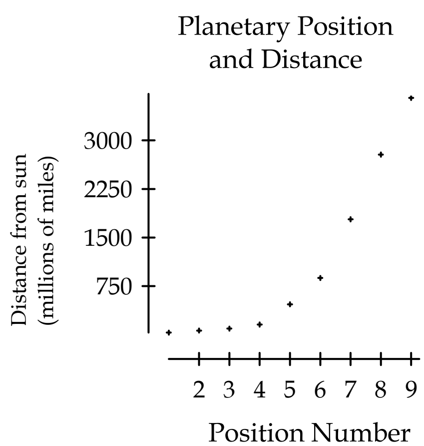


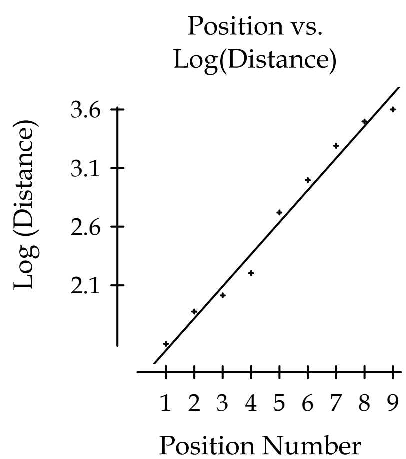
The regression model for the log-log re-expression is :
$\widehat{\log(Length)} = -2.95 + 1.5(\log(Distance))$.

c) R^2 = 100%, so the model explains 100% of the variability in the log of the length of the planetary year, at least according to the accuracy of the statistical software. The residuals plot is scattered, and the residuals are all extremely small. This is a very accurate model.

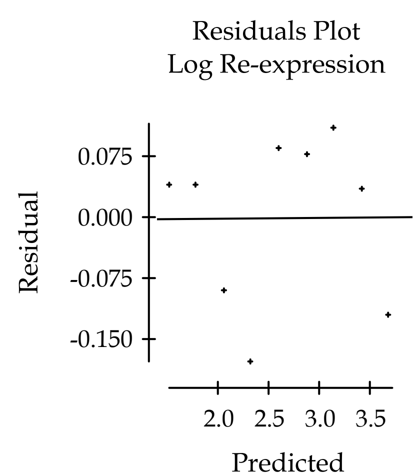
23. Planets, distances and order.

a) The association between planetary position and distance from the sun is strong, positive, and curved (below left). A good re-expression of the data is position versus Log(Distance). The scatterplot with regression line (below center) shows the straightened association. The equation of the model is
$\widehat{\log(Distance)} = 1.245 + 0.271(Position)$. The residuals plot (below right) may have some pattern, but after trying several re-expressions, this is the best that can be done. $R^2 = 98.2\%$, so the model explains 98.2% of the variability in the log of the planet's distance from the sun.





b) At first glance, this model appears to provide little evidence to support the contention of the astronomers. Pluto appears to fit the pattern, although Pluto's distance from the sun is a bit less than expected. A model generated without Pluto does not have a dramatically improved residuals plot, does not have a significantly higher R^2, nor a different slope. Pluto does not appear to be influential.



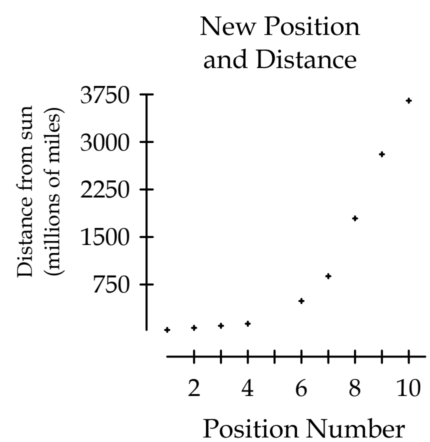
But don't forget that a logarithmic scale is being used for the vertical axis. The higher up the vertical axis you go, the greater the effect of a small change.

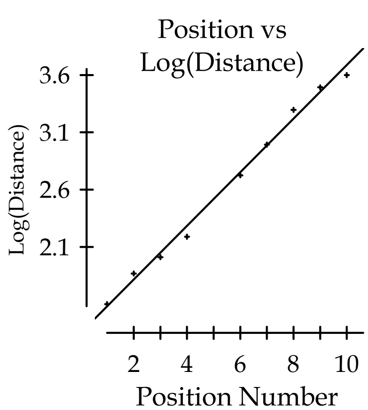
$$\widehat{\log(Distance)} = 1.24418 + 0.271229(Position)$$
$$\widehat{\log(Distance)} = 1.24418 + 0.271229(9)$$
$$\widehat{\log(Distance)} = 3.685241$$
$$\widehat{Distance} = 10^{3.685241} \approx 4844$$

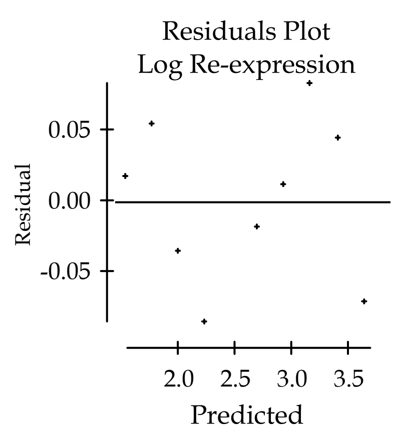
According to the model, the 9th planet in the solar system is predicted to be approximately 4844 million miles away from the sun. Pluto is actually 3707 million miles away.
Pluto doesn't fit the pattern for position and distance in the solar system. In fact, the model made with Pluto included isn't a good one, because Pluto influences those predictions. The model without Pluto,
$\widehat{\log(Distance)} = 1.20267 + 0.283680(Position)$, works much better. It has a high R^2, and scattered residuals plot. This new model predicts that the 9th planet should be a whopping 5699 million miles away from the sun! There is evidence that the IAU is correct. Pluto doesn't behave like planet in its relation to position and distance.

24. Planets, and asteroids.

Using the revised planetary numbering system, and straightening the scatterplot using the same methods as in the previous exercise, the new model, $\widehat{\log(Distance)} = 1.32 + 0.23(Position)$, is a slightly better fit. The residuals plot is more scattered, and R^2 is slightly higher, with the improved model explaining 99.5% of the variability in the log of distance from the sun.







Pluto still doesn't fit very well. The new model predicts that Pluto, as 10th planet, should be about 4169 million miles away. That's about 462 million miles farther away. A better model yet is $\widehat{\log(Distance)} = 1.28514 + 0.238826(Position)$, a model made with the new numbering system and with Pluto omitted.

25. Planets, and Eris.

A planet ninth from the sun was predicted, in a previous exercise, to be about 4844 million miles away from the sun. This distance is much shorter than the actual distance of Eris, about 6300 miles.

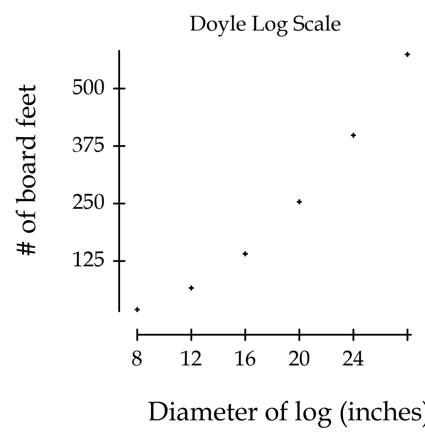
26. Planets, models, and laws.

The re-expressed data relating distance and year length are better described by their model than the re-expressed data relating position and distance. The model relating distance and year length has $R^2 = 100\%$, and a very scattered residuals plot (with miniscule residuals), possibly a natural "law". If planets in another solar system followed the Titius-Bode pattern, this belief would be reinforced. Similarly, if data were acquired from planets in another solar system that did not follow this pattern, we would be unlikely to think that this relationship was a universal law.

27. Logs (not logarithms).

a) The association between the diameter of a log and the number of board feet of lumber is strong, positive, and curved. As the diameter of the log increases, so does the number of board feet of lumber contained in the log.

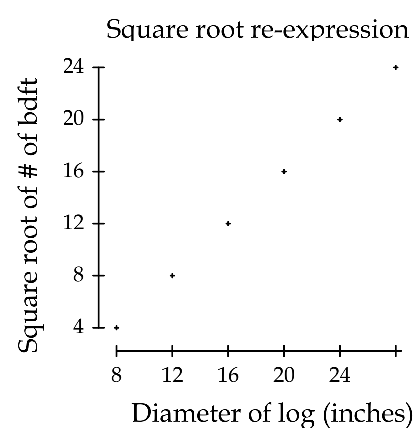
The model used to generate the table used by the log buyers is based upon a square root re-expression. The values in the table correspond exactly to the model $\widehat{\sqrt{BoardFeet}} = -4 + Diameter$.



b)

$$\widehat{\sqrt{BoardFeet}} = -4 + Diameter$$
$$\widehat{\sqrt{BoardFeet}} = -4 + (10)$$
$$\widehat{\sqrt{BoardFeet}} = 6$$
$$\widehat{BoardFeet} = 36$$

According to the model, a log 10" in diameter is expected to contain 36 board feet of lumber.



c)

$$\widehat{\sqrt{BoardFeet}} = -4 + Diameter$$
$$\widehat{\sqrt{BoardFeet}} = -4 + (36)$$
$$\widehat{\sqrt{BoardFeet}} = 32$$
$$\widehat{BoardFeet} = 1024$$

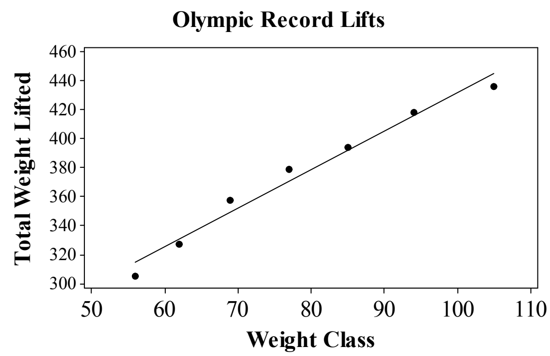
According to the model, a log 36″ in diameter is expected to contain 1024 board feet of lumber.

Normally, we would be cautious of this prediction, because it is an extrapolation beyond the given data, but since this is a prediction made from an exact model based on the volume of the log, the prediction will be accurate.

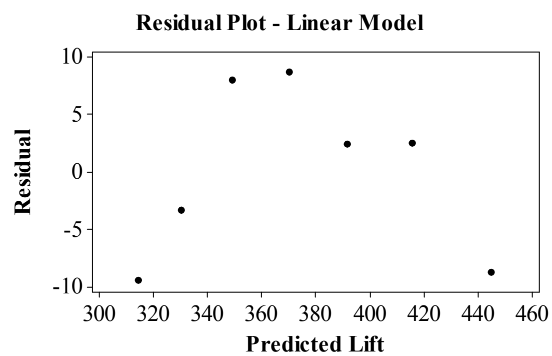
28. Weightlifting 2014.

a) The association between weight class and weight lifted for Olympic record-holders in weightlifting is strong, positive, and curved. The linear model that best fits the data is $\widehat{Lift} = 165.43 + 2.66(WeightClass)$.

Although this model accounts for 97.6% of the variability in weight lifted, it does not fit the data well.

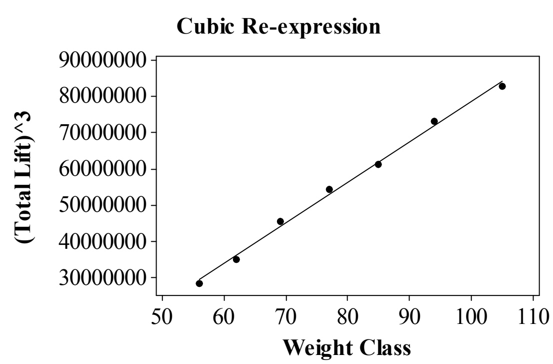


b) The residuals plot for the linear model shows a curved pattern, indicating that the linear model has failed to model the association well. A re-expressed model might fit the association between weight class and weight lifted better than the linear model.

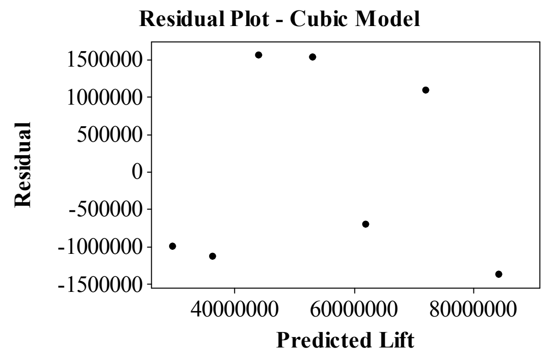


c) Answers may vary. The plot is curved downward, so move up the ladder of powers. Going further than a cube is overly complex. The model is $\widehat{Lift^3} = -33{,}350{,}350 + 1{,}120{,}090(Class)$.

We could also move the variable *Class* down the ladder of powers.



d) The cubic model is a better model, since the residuals plot shows little pattern. Additionally, the model accounts for 99.5% of the variability in weight lifted.



e) $\widehat{Lift} = 165.43 + 2.66(WeightClass)$

$\widehat{Lift} = 165.43 + 2.66(152)$

$\widehat{Lift} = 569.75$

According to the linear model, the Olympic record of a 152 kg lifter is approximately 569.75 kg.

$\widehat{Lift^3} = -33,350,350 + 1,120,090(Class)$

$\widehat{Lift^3} = -33,350,350 + 1,120,090(152)$

$\widehat{Lift^3} = 136,903,330$

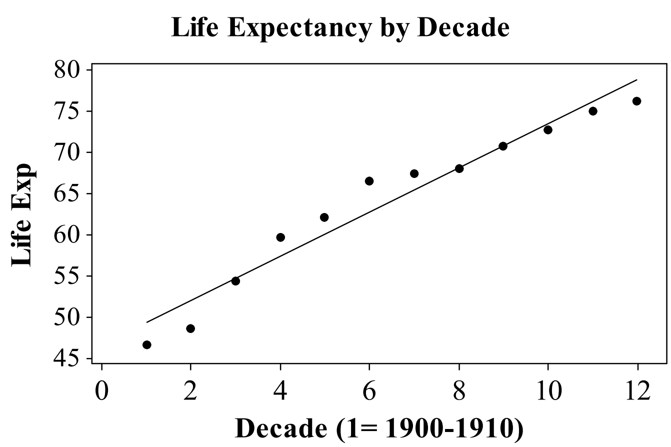
$\widehat{Lift} = \sqrt[3]{136,903,330} \approx 515.39$

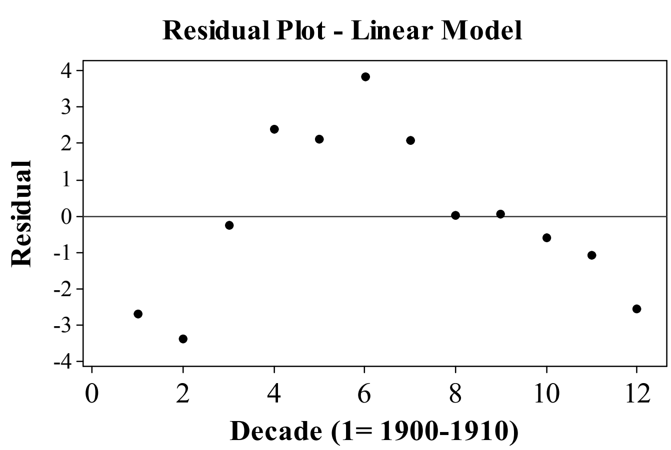
According to the cubic model, the Olympic record of a 152 kg lifter is approximately 515.39 kg.

29. Life expectancy history.

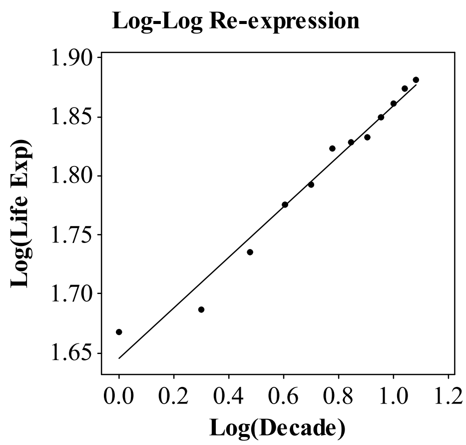
The association between year and life expectancy is strong, curved and positive. As the years passed, life expectancy for white males has increased.

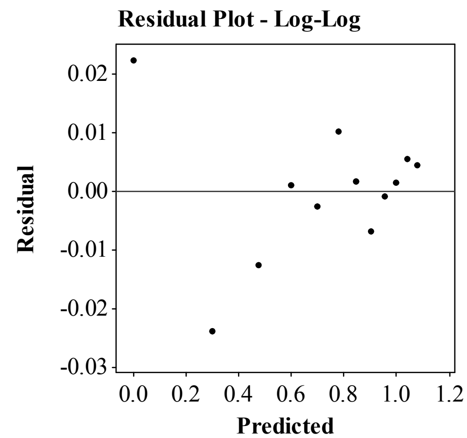
The linear model, $\widehat{LifeExp} = 46.57 + 2.68(Dec)$, explains 95.0% of the variability in the life expectancy of white males, but has a residuals plot that reveals a strong pattern.





The association between Log(*Year*) and Log(*Life Expectancy*) is strong, positive, and reasonably straight. The model is $\widehat{\log_{10}(LifeExp)} = 1.646 + 0.215(\log_{10}(Year))$.



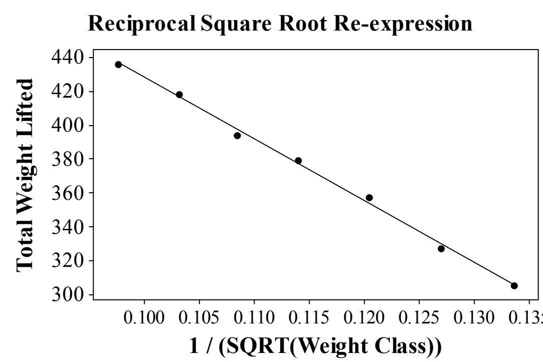


This model explains 97.4% of the variability in the logarithm of the Life Expectancy. The residuals plot shows some pattern, but seems more scattered than the residuals plot for the linear model.

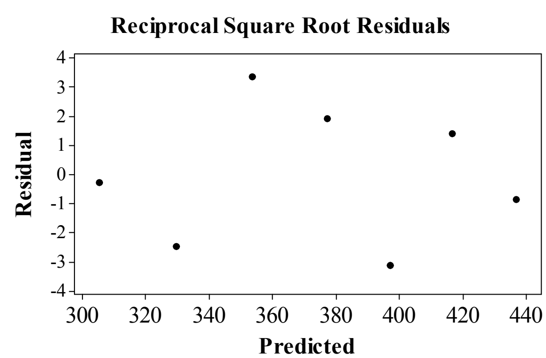
30. Lifting more weight.

a) Answers may vary. The reciprocal square root re-expression seems to straighten the scatterplot significantly.

$$\widehat{Lift} = 793.1 - \left(\frac{3650.9}{\sqrt{WeightClass}}\right)$$



b) The residual plot shows more scatter, and R^2 is higher, at 99.75%. This model appears to be the better one.



c) $$\widehat{Lift} = 793.1 - \left(\frac{3650.9}{\sqrt{WeightClass}}\right)$$

$$\widehat{Lift} = 793.1 - \left(\frac{3650.9}{\sqrt{152}}\right)$$

$$\widehat{Lift} = 496.97$$

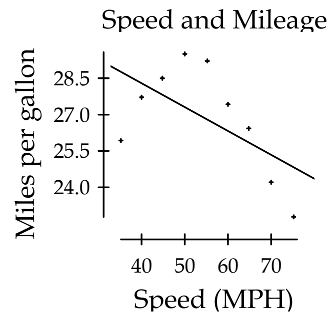
The new model predicts that the 152 kg Olympic record holder will lift approximately 496.97 kg.

d) This prediction is probably better, since the residuals are more scattered and the scatterplot is very straight.

e) All of the models over-predicted, but this one came the closest.

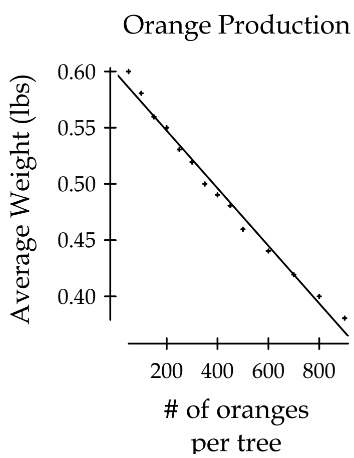
31. Slower is cheaper?

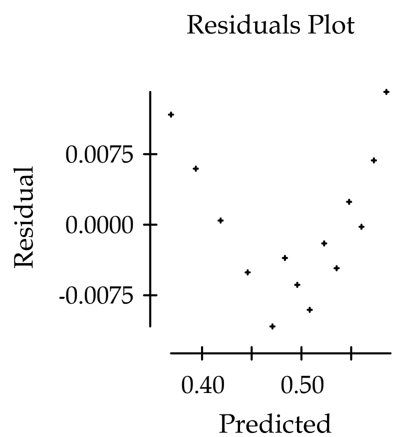
The scatterplot shows the relationship between speed and mileage of the compact car. The association is extremely strong and curved, with mileage generally increasing as speed increases, until around 50 miles per hour, then mileage tends to decrease as speed increases. The linear model is a very poor fit, but the change in direction means that re-expression cannot be used to straighten this association.



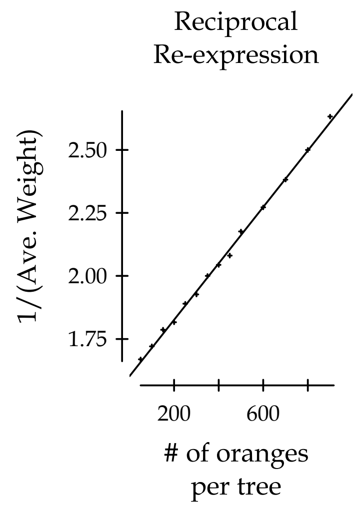
32. Orange production.

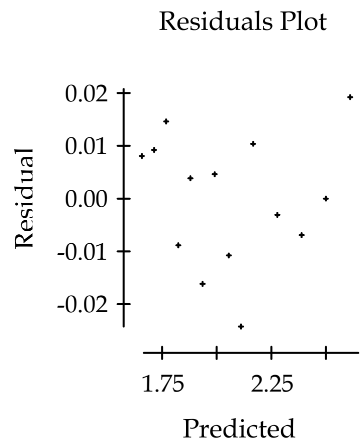
The association between the number of oranges per tree and the average weight is strong, negative, and appears linear at first look. Generally, trees that contain larger numbers of oranges have lower average weight per orange. The residuals plot shows a strong curved pattern. The data should be re-expressed.





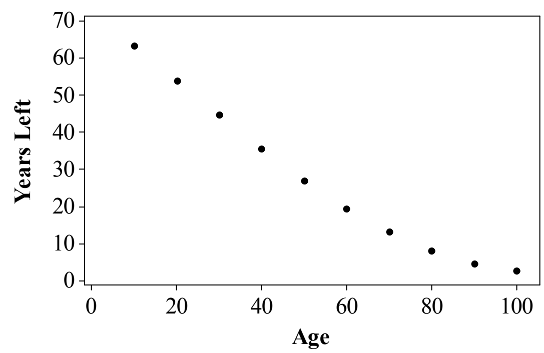
Plotting the number of oranges per tree and the reciprocal of the average weight per orange straightens the relationship considerably. The residuals plot shows little pattern and the value of R^2 indicates that the model explains 99.8% of the variability in the reciprocal of the average weight per orange. The more appropriate model is: $\widehat{\dfrac{1}{Ave.wt}} = 1.603 + 0.00112(\#Oranges/Tree)$.

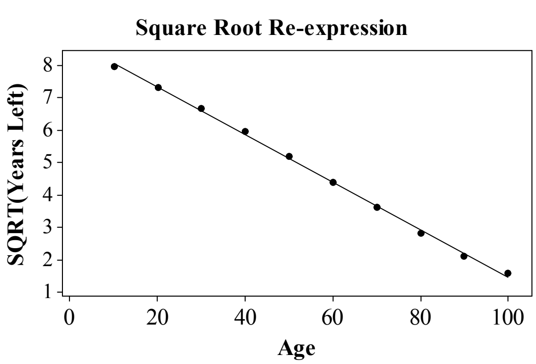


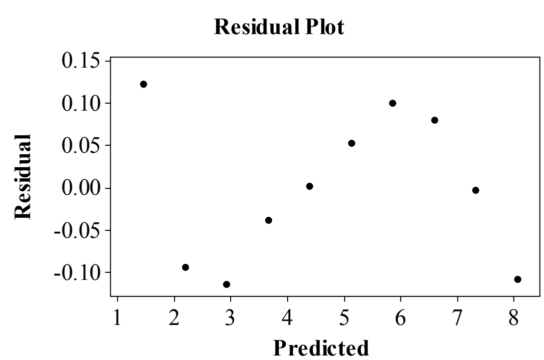


33. Years to live, 2013.

a) The association between the age and estimated additional years of life for black males is strong, curved, and negative. Older men generally have fewer estimated years of life remaining.







The square root re-expression of the data, $\widehat{\sqrt{YearsLeft}} = 8.79 - 0.0733(Age)$, straightens the data considerably, but has an extremely patterned residuals plot. The model is not a mathematically appropriate model, but fits so closely that it should be fine for predictions within the range of data. The model explains 99.8% of the variability in the estimated number of additional years.

b)

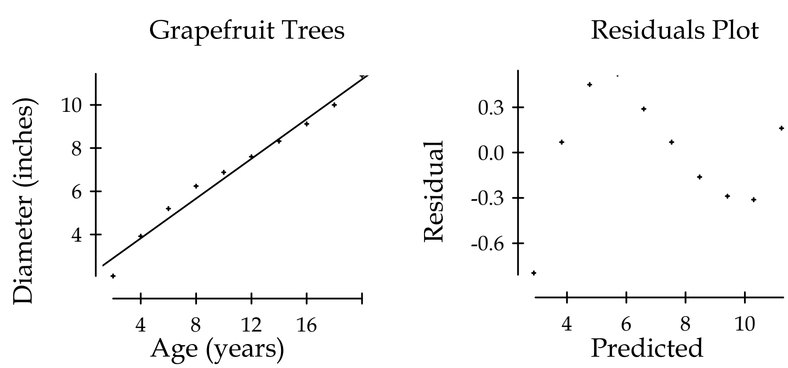
$$\widehat{\sqrt{YearsLeft}} = 8.79 - 0.0733(Age)$$
$$\widehat{\sqrt{YearsLeft}} = 8.79 - 0.0733(18)$$
$$\widehat{\sqrt{YearsLeft}} = 7.4706$$
$$\widehat{YearsLeft} = 7.4706^2 \approx 55.81$$

According to the model, an 18-year-old black male is expected to live and additional 55.81 years, for a total age of 73.81 years.

c) The residuals plot is extremely patterned, so the model is not appropriate. However, the residuals are very small, making for a tight fit. Since 18 years is within the range of the data, the prediction should be at least reasonable.

34. Tree growth.

a) The association between age and average diameter of grapefruit trees is strong, curved, and positive. Generally, older trees have larger average diameters.



The linear model for this association, $\widehat{AverageDiameter} = 1.973 + 0.463(Age)$ is not appropriate. The residuals plot shows a clear pattern.

Because of the change in curvature in the association, these data cannot be straightened by re-expression.

b) If diameters from individual trees were given, instead of averages, the association would have been weaker. Individual observations are more variable than averages.

Review of Part II – Exploring Relationships Between Variables

1. College.

% over 50: $r = 0.69$ The only moderate, positive correlation in the list.
% under 20: $r = -0.71$ Moderate, negative correlation (–0.98 is too strong)
% Full-time Fac.: $r = 0.09$ No correlation.
% Gr. on time: $r = -0.51$ Moderate, negative correlation (not as strong as %under 20)

2. Togetherness.

a) If no meals are eaten together, the model predicts a GPA of 2.73.

b) For an increase of one meal per week eaten together, the model predicts an increase of 0.11 in GPA.

c) The model will predict the mean GPA for the mean number of meals, 3.78.

$\widehat{GPA} = 2.73 + 0.11\,Meals$ The mean GPA is 3.15.

$\widehat{GPA} = 2.73 + 0.11(3.78)$

$\widehat{GPA} = 3.15$

d) A negative residual means that the student's actual GPA was lower than the GPA predicted by the model. The model over-predicted the student's GPA.

e) Although there is evidence of an association between GPA and number of meals eaten together per week, this is not necessarily a cause-and-effect relationship. There may be other variables that are related to GPA and meals, such as parental involvement and family income.

3. Vineyards, more information.

a) There does not appear to be an association between ages of vineyards and the price of products. $r = \sqrt{R^2} = \sqrt{0.027} = 0.164$, indicating a very weak association, at best. The model only explains 2.7% of the variability in case price. Furthermore, the regression equation appears to be influenced by two outliers, products from vineyards over 30 years old, with relatively high case prices.

b) This analysis tells us nothing about vineyards worldwide. There is no reason to believe that the results for the Finger Lakes region are representative of the vineyards of the world.

c) The linear equation used to predict case price from age of the vineyard is:
$\widehat{CasePrice} = 92.765 + 0.567284\,Years$

d) This model is not useful because only 2.7% of the variability in case price is accounted for by the ages of the vineyards. Furthermore, the slope of the regression line seems influenced by the presence of two outliers, products from vineyards over 30 years old, with relatively high case prices.

4. Vineyards again.

a) There is no evidence of an association between vineyard size and case price.

b) One vineyard is approximately 250 acres, with a relatively low case price. This point has high leverage.

c) If the point were removed, the correlation would be expected to increase, from a slightly negative correlation, to a correlation that is slightly positive. The point is an outlier in the *x*-direction and low in the *y*-direction. It is giving the association the artificial appearance of a slightly negative relationship.

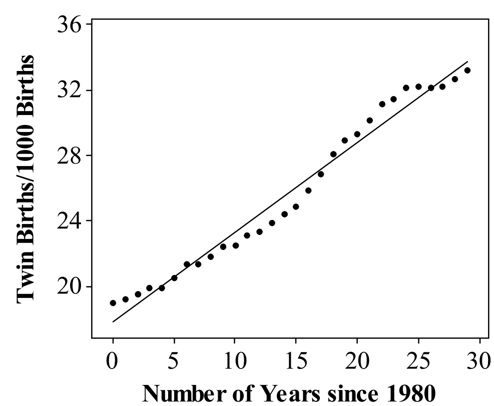
d) If the point were removed, the slope would be expected to increase, from slightly negative to slightly positive. The point is "pulling" the regression line down.

5. More twins 2014?

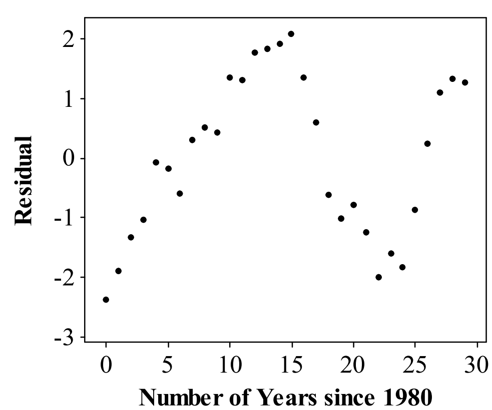
a) The association between year and the twin birth rate is strong, positive, and appears non-linear. Generally, the rate of twin births has increased over the years. The linear model that predicts the rate of twin births from the year is:

$$\widehat{Twins} = 17.77 + 0.551(\textit{Years Since 1980})$$

Twin Birth Rate - 1980 to 2009



Residual Plot



b) For each year that passes, the model predicts that the twin birth rate will increase by an average of approximately 0.55 twin births per 1000 live births.

c)

$$\widehat{Twins} = 17.77 + 0.551(\textit{Years Since 1980})$$

$$\widehat{Twins} = 17.77 + 0.551(34)$$

$$\widehat{Twins} = 36.504$$

According to the model, the twin birth rate is expected to be 36.50 twin births per 1000 live births in the US in 2014. However, the scatterplot appears non-linear, and there is no reason to believe the rate of twin births will keep increasing at the same rate for 4 years beyond the last recorded year. Faith in this prediction is very low.

d) The residuals plot shows a definite curved pattern. The association is not linear, so the linear model is not appropriate.

6. Dow Jones 2014.

a) $r = \sqrt{R^2} = \sqrt{0.883} = 0.940$. Since the slope of the regression equation is positive, we know that the correlation is also positive.

b) The linear model that predicts Dow from year is $\widehat{DJIA} = -2485.66 + 354.24\,Year$.

c) This model predicts that the Dow was expected to be –2485.66 points in the year 1970, which doesn't have contextual meaning. Furthermore, the model predicts that the Dow is expected to increase by approximately 354 points each year, on average.

d) The residuals plot shows a definite pattern. A single linear model is not appropriate. Before attempting to fit a linear model, look at the scatterplot. If it is not straight enough, the linear model cannot be used.

7. Acid rain.

a) $r = \sqrt{R^2} = \sqrt{0.27} = -0.5196$. The association between pH and BCI appears negative in the scatterplot, so use the negative value of the square root.

b) The association between pH and BCI is negative, moderate, and linear. Generally, higher pH is associated with lower BCI. Additionally, BCI appears more variable for higher values of pH.

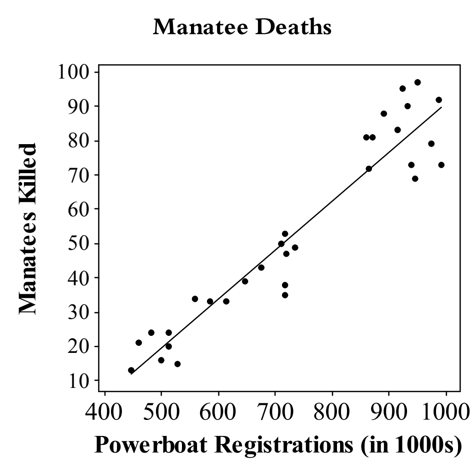
c) In a stream with average pH, the BCI would be expected to be average, as well.

d) In a stream where the pH is 3 standard deviations above average, the BCI is expected to be 1.56 standard deviations below the mean level of BCI. ($r\,(3) = -0.5196(3) = -1.56$)

8. Manatees 2013.

a) The explanatory variable is the number of powerboat registrations. This is the relationship about which the biologists are concerned. They believe that the high number of manatees killed is related to the increase in powerboat registrations.

b) The association between the number of powerboat registrations and the number of manatees killed in Florida is fairly strong, linear, and positive. Higher numbers of powerboat registrations are associated with higher numbers of manatees killed.



c) The correlation between the number of powerboat registrations and the number of manatees killed in Florida is $r = 0.951$.

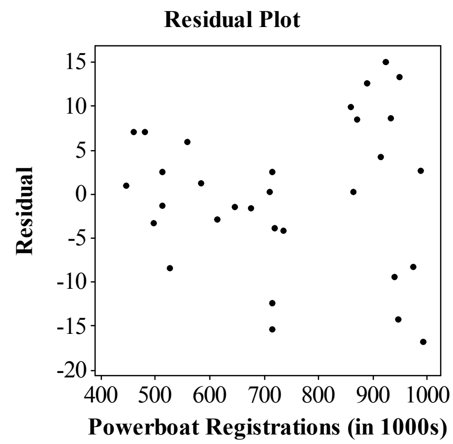
d) $R^2 = 90.5\%$. Variability in the number of powerboat registrations accounts for 90.5% of the variability in the number of manatees killed.

e) There is an association between the number of powerboat registrations and the number of manatees killed, but that is no reason to assume a cause-and-effect relationship. There may be lurking variables that affect one or the other of the variables.

9. A manatee model 2013.

a) The association between the number of powerboat registrations and the number of manatees killed is straight enough to try a linear model.

$\widehat{Kills} = -51.875 + 0.142881\,ThousandBoats$ is the best fitting model. The residuals plot is scattered, so the linear model is appropriate.



b) For every additional 10,000 powerboats registered, the model predicts that an additional 1.429 manatees will be killed on average.

c) The model predicts that if no powerboats were registered, the number of manatee deaths would be approximately −51.875. This is an extrapolation beyond the scope of the data, and doesn't have much contextual meaning.

d)

$$\widehat{Kills} = -51.875 + 0.142881\,ThousandBoats$$

$$\widehat{Kills} = -51.875 + 0.142881(949)$$

$$\widehat{Kills} \approx 83.72$$

The model predicted 83.72 manatee kills in 2009, when the number of powerboat registrations was 949,000. The actual number of kills was 97. The model under-predicted the number of kills by 13.28.

e) Negative residuals are better for the manatees. A negative residual suggests that the actual number of kills was below the number of kills predicted by the model.

f) Over time, the number of powerboat registrations has increased and the number of manatees killed has increased. The trend may continue, resulting in a greater number of manatee deaths in the future. Extrapolation is risky, however. The very trend we are seeing may result in political or societal action to attempt to decrease the number of manatee deaths.

10. Grades.

a)

$$\widehat{Fin} = 10 + 0.9Mid$$
$$\widehat{Fin} = 10 + 0.9(70)$$
$$\widehat{Fin} = 73$$

According to the model, Susan is predicted to earn a score of 73 on the final exam.

b) Susan's residual is 80 – 73 = 7 points. She scored 7 points higher than predicted.

c)

$$b_1 = r\frac{s_y}{s_x}$$
$$0.9 = r\frac{12}{10}$$
$$r = 0.75$$

The correlation between midterm exam score and final exam score is 0.75.

d)

$$\widehat{Fin} = 10 + 0.9Mid$$
$$100 = 10 + 0.9Mid$$
$$Mid = 100$$

In order to have a predicted final exam score of 100, a student would need to have a midterm exam score of 100, as well.

e) This linear model is designed to predict final exam scores based upon midterm exam scores. It does not predict midterm scores from final exam scores. In order to predict in this direction, a linear model would have to be generated with final exam score as the explanatory variable and midterm exam score as the response variable. (Notice that part d is NOT predicting midterm from final, but rather asking what *actual* midterm score is required to result in a *prediction* of 100 for the final exam score.)

f) From part d, a student with a midterm score of 100 is predicted have a final exam score of 100. The students residual is 15 – 100 = –85.

g) The R^2 value of the regression will increase. This student's large negative residual would detract from the overall pattern of the data, allowing the model to explain less of the variability in final exam score. Removing it would increase the strength of the association.

h) The slope of the linear model would increase. This student's large negative residual would "pull" the regression line down, perhaps even making the association appear negative. The removal of this point would allow the line to snap back up to the true positive association.

11. Traffic.

a)

$$b_1 = r\frac{s_y}{s_x}$$
$$-0.352 = r\frac{9.68}{27.07}$$
$$r = -0.984$$

The correlation between traffic density and speed is $r = -0.984$

b) $R^2 = (-0.984)^2 = 0.969$.
The variation in the traffic density accounts for 96.9% of the variation in speed.

c)

$$\widehat{Speed} = 50.55 - 0.352\,Density$$
$$\widehat{Speed} = 50.55 - 0.352(50)$$
$$\widehat{Speed} = 32.95$$

According to the linear model, when traffic density is 50 cars per mile, the average speed of traffic on a moderately large city thoroughfare is expected to be 32.95 miles per hour.

d)

$$\widehat{Speed} = 50.55 - 0.352\,Density$$
$$\widehat{Speed} = 50.55 - 0.352(56)$$
$$\widehat{Speed} = 30.84$$

According to the linear model, when traffic density is 56 cars per mile, the average speed of traffic on a moderately large city thoroughfare is expected to be 30.84 miles per hour. If traffic is actually moving at 32.5 mph, the residual is $32.5 - 30.84 = 1.66$ miles per hour.

e)

$$\widehat{Speed} = 50.55 - 0.352\,Density$$
$$\widehat{Speed} = 50.55 - 0.352(125)$$
$$\widehat{Speed} = 6.55$$

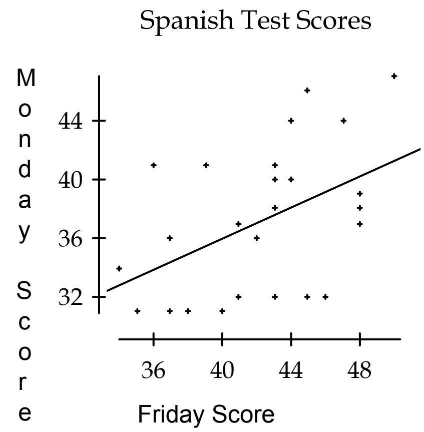
According to the linear model, when traffic density is 125 cars per mile, the average speed of traffic on a moderately large city thoroughfare is expected to be 6.55 miles per hour. The point with traffic density 125 cars per minute and average speed 55 miles per hour is considerably higher than the model would predict. If this point were included in the analysis, the slope would increase.

f) The correlation between traffic density and average speed would become weaker. The influential point (125, 55) is a departure from the pattern established by the other data points.

g) The correlation would not change if kilometers were used instead of miles in the calculations. Correlation is a "unitless" measure of the degree of linear association based on *z*-scores, and is not affected by changes in scale. The correlation would remain the same, $r = -0.984$.

12. Cramming.

a) The correlation between the Friday scores and the Monday scores on the Spanish Test is $r = 0.473$.



b) The scatterplot shows a weak, linear, positive association between Friday score and Monday score. Generally, students who scored high on Friday also tended to score high on Monday.

c) A student with a positive residual scored higher on Monday's test than the model predicted.

d) A student with a Friday score that is one standard deviation below average is expected to have a Monday score that is 0.473 standard deviations below Monday's average score. The distribution of scores for Monday had mean 37.24 points and standard deviation 5.02 points, so the student's score is predicted to be approximately $37.24 - (0.473)(5.02) = 34.87$.

e) The regression equation for the linear model that predicts Monday score from Friday score is: $\widehat{Monday} = 14.59 + 0.536\,Friday$.

f) $\widehat{Monday} = 14.5921 + 0.535666\,Friday$

$\widehat{Monday} = 14.5921 + 0.535666(40)$

$\widehat{Monday} \approx 36.0$

According to the model, a student with a Friday score of 40 is expected to have a Monday score of about 36.0.

13. Car correlations.

a) Weight, with a correlation of –0.903, seems to be most strongly associated with fuel economy, since the correlation has the largest magnitude (distance from zero). However, without looking at a scatterplot, we can't be sure that the relationship is linear. Correlation might not be an appropriate measure of the strength of the association if the association is non-linear.

b) The negative correlation between weight and fuel economy indicates that, generally, cars with higher weights tend to have lower mileages than cars with lower weights. Once again, this is only correct if the association between weight and fuel economy is linear.

c) $R^2 = (-0.903)^2 = 0.815$. The variation in weight accounts for 81.5% of the variation in mileage. Once again, this is only correct if the association between weight and fuel economy is linear.

14. Cars, revisited.

a) Displacement and weight show the strongest association, with a correlation of 0.951. Generally, cars with larger engines are heavier than cars with smaller engines. However, without looking at a scatterplot, we can't be sure that the relationship is linear. Correlation might not be an appropriate measure of the strength of the association if the association is non-linear.

b) The strong correlation between displacement and weight is not necessarily a sign of a cause-and-effect relationship. Price of the car might be confounded with weight and displacement. More expensive luxury cars may have extra features that result in higher weights. One of these features might be a larger engine. Another difficulty with assigning a cause and an effect is that we can't be sure of the direction of the relationship. Certainly, the larger engine adds to the weight of the car, but maybe the larger engine is needed to power heavier cars.

c) The correlation would not change if cubic centimeters or liters were used instead of cubic inches in the calculations. Correlation is a "unitless" measure of the degree of linear association based on z-scores, and is not affected by changes in scale.

d) As long as the association between fuel economy and engine displacement was linear, a car whose engine displacement is one standard deviation above the mean would be predicted to have a fuel economy that is 0.786 standard deviations below the mean. (The correlation between the variables is –0.786, so a change in direction is indicated.) If the relationship were non-linear, the relative fuel economy could not be determined.

15. Cars, one more time!

a) The linear model that predicts the horsepower of an engine from the weight of the car is: $\widehat{Horsepower} = 3.49834 + 34.3144\,Weight$.

b) The weight is measured in thousands of pounds. The slope of the model predicts an increase of about 34.3 horsepower for each additional unit of weight. 34.3 horsepower for each additional thousand pounds makes more sense than 34.3 horsepower for each additional pound.

c) Since the residuals plot shows no pattern, the linear model is appropriate for predicting horsepower from weight.

d)

$$\widehat{Horsepower} = 3.49843 + 34.3144\,Weight$$
$$\widehat{Horsepower} = 3.49843 + 34.3144(2.595)$$
$$\widehat{Horsepower} \approx 92.544$$

According to the model, a car weighing 2595 pounds is expected to have 92.543 horsepower. The actual horsepower of the car is: 92.544 + 22.5 ≈ 115.0 horsepower.

16. Colorblind.

Gender and colorblindness are both categorical variables. Correlation is a measure of the strength of a linear relationship between quantitative variables. The proper terminology is to say gender is associated with colorblindness.

17. Old Faithful again.

a) The association between the duration of eruption and the interval between eruptions of Old Faithful is fairly strong, linear, and positive. Long eruptions are generally associated with long intervals between eruptions. There are also two distinct clusters of data, one with many short eruptions followed by short intervals, the other with many long eruptions followed by long intervals, with only a few medium eruptions and intervals in between.

b) The linear model used to predict the interval between eruptions is:
$\widehat{Interval} = 33.9668 + 10.3582\,Duration$.

c) As the duration of the previous eruption increases by one minute, the model predicts an increase of about 10.4 minutes in the interval between eruptions.

d) $R^2 = 77.0\%$, so the model accounts for 77% of the variability in the interval between eruptions. The predictions should be fairly accurate, but not precise. Also, the association appears linear, but we should look at the residuals plot to be sure that the model is appropriate before placing too much faith in any prediction.

e)

$$\widehat{Interval} = 33.9668 + 10.3582\,Duration$$
$$\widehat{Interval} = 33.9668 + 10.3582(4)$$
$$\widehat{Interval} \approx 75.4$$

According to the model, if an eruption lasts 4 minutes, the next eruption is expected to occur in approximately 75.4 minutes.

f) The actual eruption at 79 minutes is 3.6 minutes later than predicted by the model. The residual is 79 – 75.4 = 3.6 minutes. In other words, the model under-predicted the interval.

18. Crocodile lengths.

a) The associations between the head sizes and body sizes for the crocodiles appear to be strong. 97.2% of the variability in Indian Crocodile length and 98% of the variability in Australian Crocodile length is accounted for by the variability in head size. (This assertion is only valid if the association between head and body length is linear for each crocodile.)

b) The slopes of the two models are similar. Indian Crocodiles are predicted to increase in length 7.4 centimeters for each centimeter increase in head length, and Australian Crocodiles are predicted to increase in length 7.72 centimeters for each centimeter increase in head length. (These predictions are only valid if the association between head and body length is linear for each crocodile.) The values of R^2 are also similar.

c) The two models have different values for the y-intercept. According to the models, the Indian Crocodile is smaller.

d) Indian Crocodile Model

$$\widehat{IBody} = -69.3693 + 7.40004\,IHead$$
$$\widehat{IBody} = -69.3693 + 7.40004(62)$$
$$\widehat{IBody} = 389.43318$$

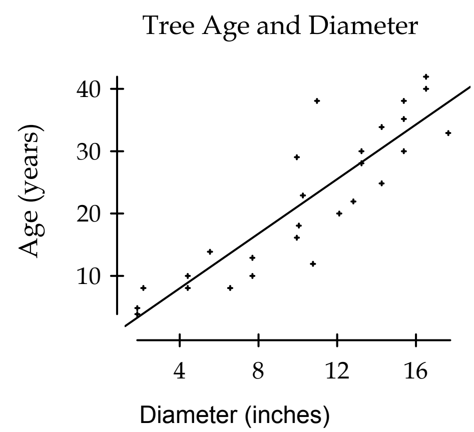
Australian Crocodile Model

$$\widehat{ABody} = -21.3429 + 7.82761\,AHead$$
$$\widehat{ABody} = -21.3429 + 7.82761(62)$$
$$\widehat{ABody} = 463.96892$$

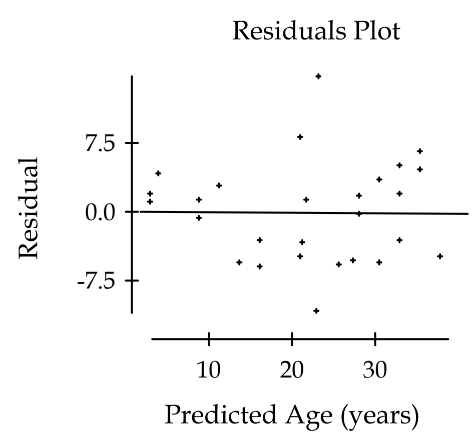
The appropriate models predict body lengths of 389.4 centimeters and 464.0 centimeters for Indian Crocodiles and Australian Crocodiles, respectively. The actual length of 380 centimeters indicates that this is probably an Indian Crocodile. The prediction is closer to the actual length for that model.

19. How old is that tree?

a) The correlation between tree diameter and tree age is $r = 0.888$. Although the correlation is moderately high, this does not suggest that the linear model is appropriate. We must look at a scatterplot in order to verify that the relationship is straight enough to try the linear model. After finding the linear model, the residuals plot must be checked. If the residuals plot shows no pattern, the linear model can be deemed appropriate.



b) The association between diameter and age of these trees is fairly strong, somewhat linear, and positive. Trees with larger diameters are generally older.



c) The linear model that predicts age from diameter of trees is:

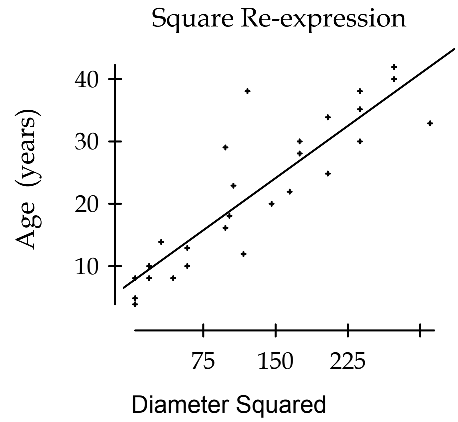
$\widehat{Age} = -0.974424 + 2.20552\,Diameter$. This model explains 78.9% of the variability in age of the trees.

d) The residuals plot shows a curved pattern, so the linear model is not appropriate. Additionally, there are several trees with large residuals.

e) The largest trees are generally above the regression line, indicating a positive residual. The model is likely to underestimate these values.

20. Improving trees.

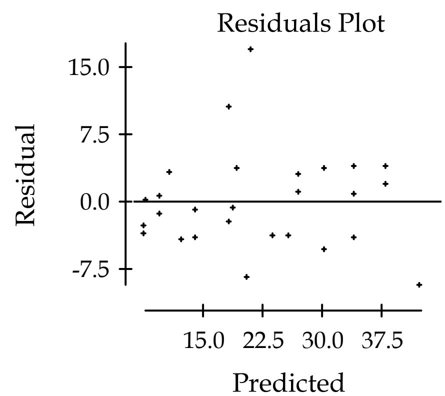
a) The re-expressed data, diameter squared versus age, is straighter than the original. This model appears to fit much better.



b) The linear model that predicts age from diameter squared is:

$\widehat{Age} = 7.23961 + 0.113011\,Diameter^2$. This model explains 78.7% of the variability in the age of the trees.

c) The residuals plot shows random scatter. This model appears to be appropriate, although there are still some trees with large residuals.



d)

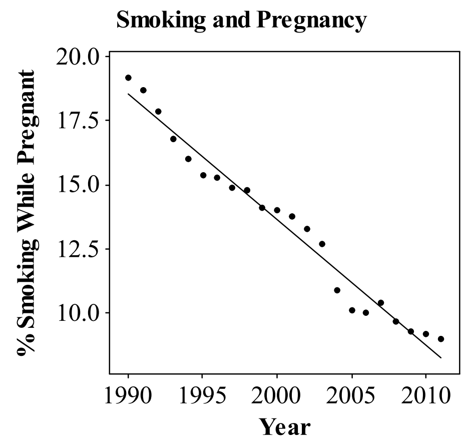
$$\widehat{Age} = 7.23961 + 0.113011\,Diameter^2$$
$$\widehat{Age} = 7.23961 + 0.113011\left(18^2\right)$$
$$\widehat{Age} \approx 43.855$$

According to the model, a tree with a diameter of 18″ is expected to be approximately 43.9 years old.

21. Big screen.

TV screen sizes might vary from 19 to 70 inches. A TV with a screen that was 10 inches larger would be predicted to cost 10(0.03) = +0.3, 10(0.3) = +3, 10(3) = +30, or 10(30) = +300. Notice that the TV costs are measure in hundreds of dollars, so the potential price changes for getting a TV 10 inches larger are \$30, \$300, \$3000, and \$30,000. Only \$300 is reasonable, so the slope must be 0.3.

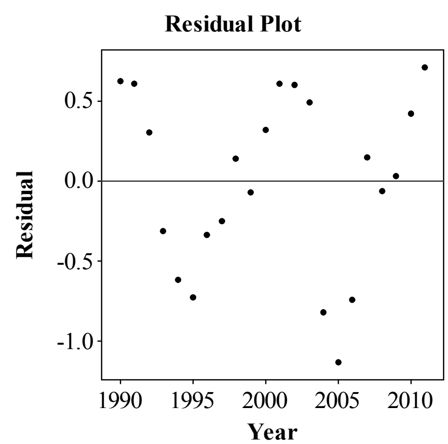
22. Smoking and pregnancy 2011.



a) The association between year and the percent of expectant mothers who smoked cigarettes during their pregnancies is strong, somewhat cyclical, but overall roughly linear, and negative. The percentage has decreased steadily since 1990.

b) The correlation between year and percent of expectant mothers who smoked cigarettes during their pregnancies is $r = -0.985$. This may not be an appropriate measure of strength, since the scatterplot shows a slight bend.

c) The use of averages instead of individual percentages for each of the 50 states results in a correlation that is artificially strong. The correlation of the averaged data is "more negative" than the correlation of the individual percentages would have been.



d) The linear model that predicts the percent of expectant mothers who smoked during their pregnancies from the year is:
$\hat{\%} = 993.35 - 0.48984\,Year$. This model accounts for 97.1% of the variability in the percent. According to this model, for each year that passes, the average percent of women who smoked while pregnant decreases by 0.4898%. This model does not appear to be appropriate, since the residuals plot shows a pattern. However, it is unlikely that we can provide a better model, since the scatterplot shows a bend that cannot be straightened by re-expression.

23. No smoking?

a) The model from Exercise 22 is for predicting the percent of expectant mothers who smoked during their pregnancies from the year, not the year from the percent.

b) The model that predicts the year from the percent of expectant mothers who smoked during pregnancy is: $\widehat{Year} = 2027.13 - 1.98233(\%)$. This model predicts that 0% of mothers will smoke during pregnancy in $2027.13 - 1.98233(0) \approx 2027$.

c) The lowest data point corresponds to 9.0% of expectant mothers smoking during pregnancy in 2011. The prediction for 0% is an extrapolation outside the scope of the data. There is no reason to believe that the model will be accurate at that point.

24. Tips.

a) There is a very weak, linear, positive relationship between tip size and rating of service. Generally, better tips are associated with better ratings of service, but the relationship is so weak that it may simply be due to sampling error.

b) $R^2 = 1.21\%$. Only about 1% of the variability in tip size is explained by variability in the rating of service.

25. U.S. Cities.

There is a strong, roughly linear, negative association between mean January temperature and latitude. U.S. cities with higher latitudes generally have lower mean January temperatures. There are two outliers, cities with higher mean January temperatures than the pattern would suggest.

26. Correlations.

a) Latitude is the better predictor of average January temperature when the relationship between latitude and temperature is linear. The correlation, –0.848, is stronger than the correlation between altitude and temperature, –0.369.

b) The correlation would be the same, –0.848. Correlation is a measure of the degree of linear association between two quantitative variables and is unaffected by changes in units.

c) The correlation would be the same, –0.369. Correlation is a measure of the degree of linear association between two quantitative variables and is unaffected by changes in units.

d) $(-0.369)(2) = -0.738$. If a city has an altitude 2 standard deviations above the mean, its average January temperature is expected to be 0.738 standard deviations below the mean average January temperature.

27. Winter in the city.

a) $R^2 = (-0.848)^2 \approx 0.719$. The variation in latitude explains 71.9% of the variability in average January temperature.

b) The negative correlation indicates that the as latitude increases, the average January temperature generally decreases.

c) $b_1 = r\dfrac{s_y}{s_x}$

$b_1 = (-0.848)\dfrac{13.49}{5.42}$

$b_1 = -2.1106125$

$$\hat{y} = b_0 + b_1 x$$
$$\overline{y} = b_0 + b_1 \overline{x}$$
$$26.44 = b_0 - 2.1106125(39.02)$$
$$b_0 = 108.79610$$

The equation of the linear model for predicting January temperature from latitude is: $\widehat{JanTemp} = 108.796 - 2.111\,Latitude$

d) For each additional degree of latitude, the model predicts a decrease of approximately 2.1°F in average January temperature.

e) The model predicts that the mean January temperature will be approximately 108.8°F when the latitude is 0°. This is an extrapolation, and may not be meaningful.

f)

$$\widehat{JanTemp} = 108.796 - 2.111\,Latitude$$
$$\widehat{JanTemp} = 108.796 - 2.111(40)$$
$$\widehat{JanTemp} \approx 24.4$$

According to the model, the mean January temperature in Denver is expected to be 24.4°F.

g) In this context, a positive residual means that the actual average temperature in the city was higher than the temperature predicted by the model. In other words, the model underestimated the average January temperature.

28. Depression.

First of all, no association between variables can imply a cause-and-effect relationship. There may be lurking variables that explain the increase in both Internet use and depression. Additionally, provided the association is linear, only 4.6% of the variability in depression level can be explained by variability in Internet use. This is a very weak linear association at best.

29. Jumps 2012.

a) The association between Olympic long jump distances and high jump heights is strong, linear, and positive. Years with longer long jumps tended to have higher high jumps. There is one departure from the pattern. The year in which the Olympic gold medal long jump was the longest had a shorter gold medal high jump than we might have predicted.

b) There is an association between long jump and high jump performance, but it is likely that training and technique have improved over time and affected both jump performances.

c) The correlation would be the same, 0.912. Correlation is a measure of the degree of linear association between two quantitative variables and is unaffected by changes in units.

d) In a year when the high jumper jumped one standard deviation better than the average jump, the long jumper would be predicted to jump r = 0.912 standard deviations above the average long jump.

30. Modeling jumps 2012.

a)

$$b_1 = r\frac{s_y}{s_x}$$
$$b_1 = (0.912)\frac{0.1938663}{0.5135913}$$
$$b_1 = 0.3442544015$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$2.148077 = b_0 + 0.3442544015(8.050000)$$
$$b_0 = -0.6232$$

The linear model that predicts high jumps heights from long jump distances is:
$\widehat{High} = -0.6232 + 0.344\,Long$

b) For each additional meter jumped in the long jump, the model predicts an increase of approximately 0.344 meters in the high jump.

c)

$$\widehat{High} = -0.6232 + 0.344\,Long$$
$$\widehat{High} = -0.6232 + 0.344(8.9)$$
$$\widehat{High} \approx 2.4384$$

According to the model, the high jump height is expected to be approximately 2.4384 meters in a year when the long jump distance is 8.9 meters.

d) This equation cannot be used to predict long jump distance from high jump height, because it was specifically designed to predict high jump height from long jump distance.

e)

$$b_1 = r\frac{s_y}{s_x}$$
$$b_1 = (0.912)\frac{0.5135913}{0.1938663}$$
$$b_1 = 2.416073684$$

$$\hat{y} = b_0 + b_1 x$$
$$\bar{y} = b_0 + b_1 \bar{x}$$
$$8.050000 = b_0 + 2.416073684(2.148077)$$
$$b_0 = 2.8601$$

The linear model that predicts long jump distances from high jump distances is:
$\widehat{Long} = 2.8601 + 2.4161\,High$

31. French.

a) Most of the students would have similar weights. Regardless of their individual French vocabularies, the correlation would be near 0.

b) There are two possibilities. If the school offers French at all grade levels, then the correlation would be positive and strong. Older students, who typically weigh more, would have higher scores on the test, since they would have learned more French vocabulary. If French is not offered, the correlation between weight and test score would be near 0. Regardless of weight, most students would have horrible scores.

c) The correlation would be near 0. Most of the students would have similar weights and vocabulary test scores. Weight would not be a predictor of score.

d) The correlation would be positive and strong. Older students, who typically weigh more, would have higher test scores, since they would have learned more French vocabulary.

32. Twins.

a) There is a strong, fairly linear, positive trend in pre-term twin birth rates. As the year of birth increases, the pre-term twin birth rate increases.

b) The highest pre-term twin birth rate is for mothers receiving "adequate" prenatal care, and the lowest pre-term twin birth rate is for mothers receiving "inadequate" prenatal care. The slope is about the same for these relations. Mothers receiving "intensive" prenatal care had a pre-term twin birth rate that was higher than mothers receiving "inadequate" care and lower than mothers receiving "adequate" prenatal care. However, the rate of increase in pre-term twin birth rate is greater for mothers receiving "intensive" prenatal care than the rate of increase for the other groups.

c) Avoiding medical care would not be a good idea. There are likely lurking variables explaining the differences in pre-term twin birth rate. For instance, the level of pre-natal care may actually be determined by complications early in the pregnancy that may result in a pre-term birth.

33. Lunchtime.

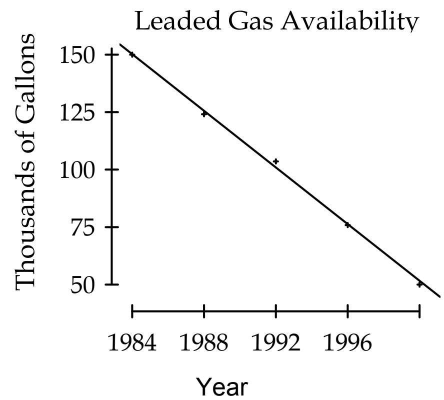
The association between time spent at the table and number of calories consumed by toddlers is moderate, roughly linear, and negative. Generally, toddlers who spent a longer time at the table consumed fewer calories than toddlers who left the table quickly. The scatterplot between time at the table and calories consumed is straight enough to justify the use of the linear model. The linear model that predicts the time number of calories consumed by a toddler from the time spent at the table is $\widehat{Calories} = 560.7 - 3.08\,Time$. For each additional minute spent at the table, the model predicts that the number of calories consumed will be approximately 3.08 fewer. Only 42.1% of the variability in the number of calories consumed can be accounted for by the variability in time spent at the table. The residuals plot shows no pattern, so the linear model is appropriate, if not terribly useful for prediction.

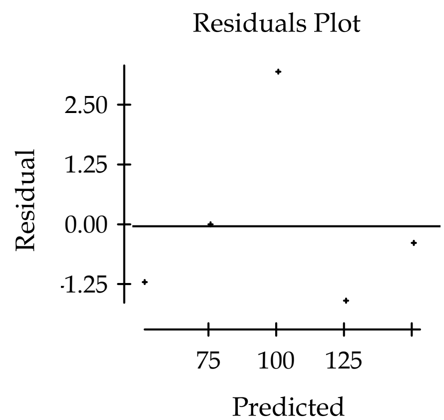
34. Gasoline.

a) The association between the year and the number of gallons of leaded gasoline available is linear, very strong, and negative. As the years have passed, the number of gallons of leaded gasoline has decreased steadily. The linear model that predicts the number of gallons available based upon the year is:

$\widehat{ThousandGallons} = 12{,}451.2 - 6.2\,Year$

The residuals plot shows no pattern, so the linear model is appropriate. $R^2 = 99.8\%$, so the variability in year accounts for 99.8% of the variability in the number of gallons available. According to the model, there will be approximately –41,800 gallons available in 2015. This is an extrapolation, and isn't meaningful.



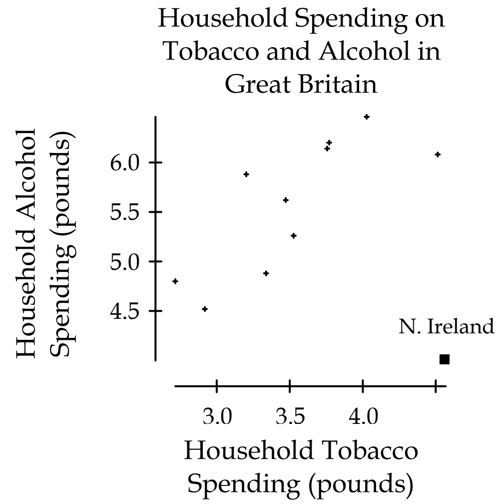


b) The model is designed to predict the number of gallons available based on the year. The question asks the students to predict the year based on the number of gallons available, and models only predict in one direction.

c) The linear regression model that predicts the year from the number of gallons available is: $\widehat{Year} = 2008.22 - 0.161\,ThousandGallons$ The model predicts that 0 gallons of leaded gasoline will be available in about 2008.

d) The association between year and the number of gallons of leaded gasoline available is very strong. In fact, it is so strong that the models actually do a decent job of predicting in the wrong direction! The model designed to minimize the sum of squared residuals in the response direction is actually pretty good at minimizing the sum of the squared residuals in the explanatory direction.

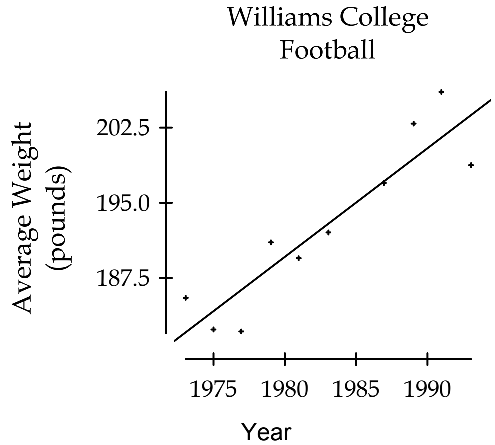
35. Tobacco and alcohol.

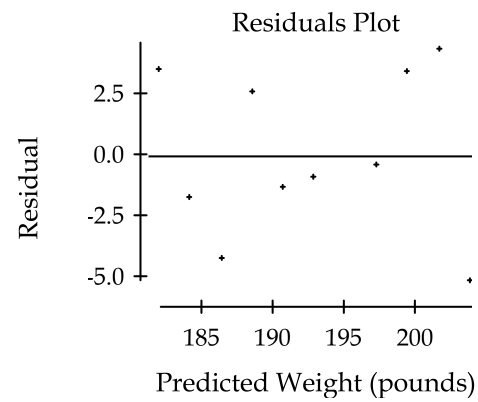


The first concern about these data is that they consist of averages for regions in Great Britain, not individual households. Any conclusions reached can only be about the regions, not the individual households living there. The second concern is the data point for Northern Ireland. This point has high leverage, since it has the highest household tobacco spending and the lowest household alcohol spending. With this point included, there appears to be only a weak positive association between tobacco and alcohol spending. Without the point, the association is much stronger. In Great Britain, with the exception of Northern Ireland, higher levels of household spending on tobacco are associated with higher levels of household spending on tobacco. It is not necessary to make the linear model, since we have the household averages for the regions in Great Britain, and the model wouldn't be useful for predicting in other countries or for individual households in Great Britain.

36. Football weights.

a) The association between year and average weight of the members of the Williams College football team is strong, linear, and positive. Over the years, there has been a general increase in the average weight of the team. The linear model that predicts average weight from year is:
$\widehat{AvgWeight} = -1971.26 + 1.09137\,Year$.
According to the model, average weight has increased by approximately one pound per year since 1973. The model accounts for 83.8% of the variability in average weight.





b) The residuals plot shows no pattern, so the linear model is appropriate. However, there are a couple of things to remember. First, the model is for predicting the average weight of the players on the team. Individual weights would be much more variable. Second, since we are dealing with weights, it is not reasonable to use the model for extrapolations. There is no reason to believe that average weights before 1973 or after 1993 would follow the model.

c) The model predicts that the average weight of the Williams College football team will be approximately 227.85 pounds in 2015. This prediction might be pretty close, but we shouldn't place to much faith in it. 2015 is 22 years later than the last year for which we have data.

$$\widehat{AvgWeight} = -1971.26 + 1.09137\,Year$$
$$\widehat{AvgWeight} = -1971.26 + 1.09137(2015)$$
$$\widehat{AvgWeight} \approx 227.85$$

d) The model predicts that the average weight of the Williams College football team will be approximately 323.88 pounds in 2103. This is not reasonable. The prediction is based upon an extrapolation of 110 years.

e) The model predicts that the average weight of the Williams College football team will be approximately 1306.11 pounds in 3003. This is absurd. The prediction is based upon an extrapolation of 1010 years.

37. Models.

a) $\hat{y} = 2 + 0.8\ln x$

$\hat{y} = 2 + 0.8\ln(10)$

$\hat{y} \approx 3.842$

b) $\log \hat{y} = 5 - 0.23x$

$\log \hat{y} = 5 - 0.23(10)$

$\log \hat{y} = 2.7$

$\hat{y} = 10^{2.7} \approx 501.187$

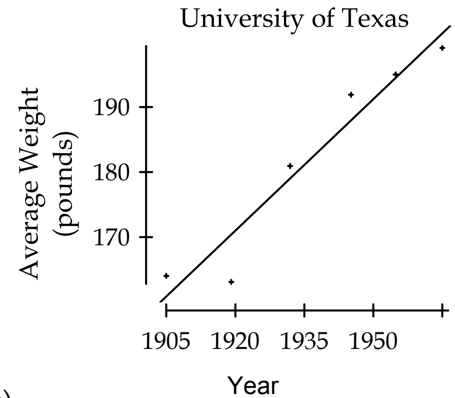
c) $\dfrac{1}{\sqrt{\hat{y}}} = 17.1 - 1.66x$

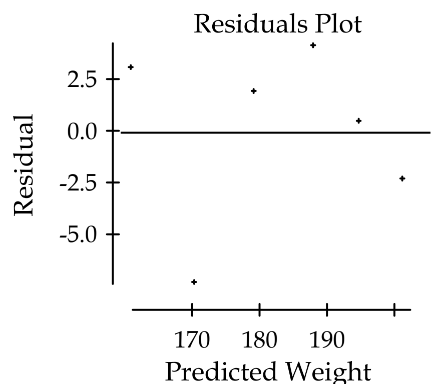
$\dfrac{1}{\sqrt{\hat{y}}} = 17.1 - 1.66(10) = 0.5$

$\hat{y} = \dfrac{1}{0.5^2} = 4$

38. Williams vs. Texas.

a) The association between year and average weight for the University of Texas football team is strong, roughly linear, and positive. The average weight has generally gone up over time. The linear model is: $\widehat{Ave.Weight} = -1121.66 + 0.67326\,Year$. This model explains 92.8% of the variability in average weight, but the residuals plot shows a possible pattern. The linear model may not be appropriate.





b) $-1971.26 + 1.09137\,Year = -1121.66 + 0.67326(Year)$

$$0.41811\,Year = 849.60$$

$$Year \approx 2032$$

According to these models, the predicted weights will be the same some time during the year 2032. The average weight of the Williams College team is predicted to be more than the average weight of the University of Texas team any time after 2032.

c) This information is not likely to be accurate. The year 2032 is an extrapolation for both of the models, each of which has been shown to be of little use for even small extrapolations.

39. Vehicle weights.

a)

$\widehat{Wt} = 10.85 + 0.64\,scale$

$\widehat{Wt} = 10.85 + 0.64(31.2)$

$\widehat{Wt} = 30.818$

According to the model, a truck with a scale weight of 31,200 pounds is expected to weigh 30,818 pounds.

b) If the actual weight of the truck is 32,120 pounds, the residual is 32,120 – 30,818 = 1302 pounds. The model underestimated the weight.

c)

$\widehat{Wt} = 10.85 + 0.64\,scale$

$\widehat{Wt} = 10.85 + 0.64(35.590)$

$\widehat{Wt} = 33.6276$ thousand pounds

The predicted weight of the truck is 33,627.6 pounds. If the residual is –2440 pounds, the actual weight of the truck is 33,627.6 – 2440 = 31,187.6 pounds.

d) $R^2 = 93\%$, so the model explains 93% of the variability in weight, but some of the residuals are 1000 pounds or more. If we need to be more accurate than that, then this model will not work well.

e) Negative residuals will be more of a problem. Police would be issuing tickets to trucks whose weights had been overestimated by the model. The U.S. justice system is based upon the principle of innocence until guilt is proven. These truckers would be unfairly ticketed, and that is worse than allowing overweight trucks to pass.

40. Profit

a) The re-expressed data are more symmetric, with no outliers. That's good for regression because there is less of a chance for influential points. (Additionally, symmetric distributions of the explanatory and response variables will help ensure that the residuals around a line through the data are more unimodal and symmetric. We will learn more about why this is important in a later chapter.)

b) The association between log(sales) and log(profit) is linear, positive, and strong. The residuals plot shows no pattern. This model appears to be appropriate, and is surely better than the model generated with the original data.

c) The linear model is: $\widehat{\log(Profit)} = -0.106259 + 0.647798\,\log(Sales)$

d)

$\widehat{\log(Profit)} = -0.106259 + 0.647798\,\log(Sales)$

$\widehat{\log(Profit)} = -0.106259 + 0.647798\,\log(2500)$

$\widehat{\log(Profit)} = 2.0949197$

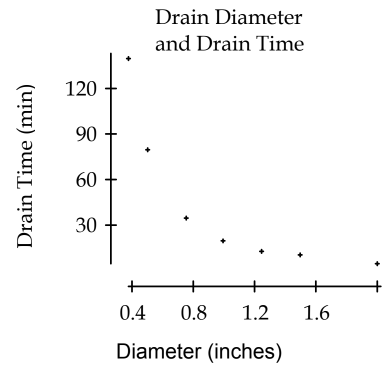
$10^{2.0949197} \approx 124.43$

According to the model, a company with sales of 2.5 billion dollars is expected to have profits of about 124.43 million dollars.

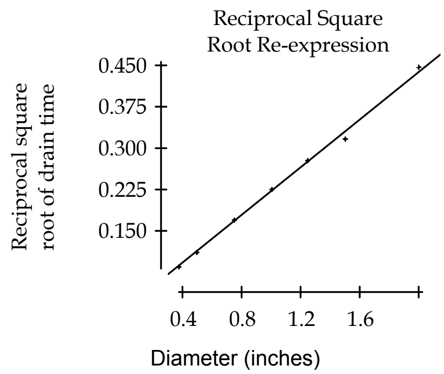
41. Down the drain.

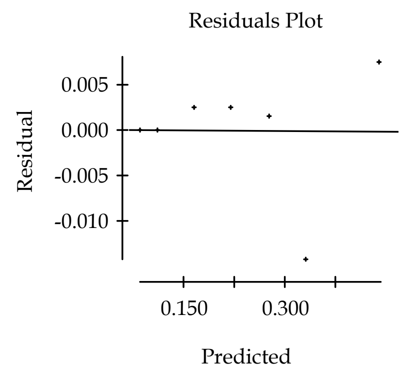
The association between diameter of the drain plug and drain time of this water tank is strong, curved, and negative. Tanks with larger drain plugs have lower drain times. The linear model is not appropriate for the curved association, so several re-expressions of the data were tried. The best one was the reciprocal square root re-expression, resulting in the equation

$$\widehat{\frac{1}{\sqrt{DrainTime}}} = 0.00243 + 0.219\,Diameter.$$



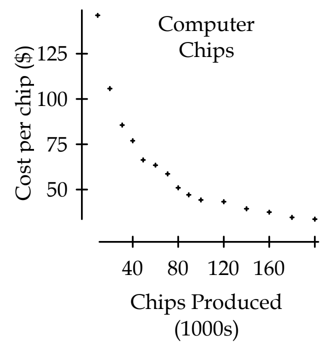
The re-expressed data is nearly linear, and although the residuals plot might still indicate some pattern and has one large residual, this is the best of the models examined. The model explains 99.7% of the variability in drain time.

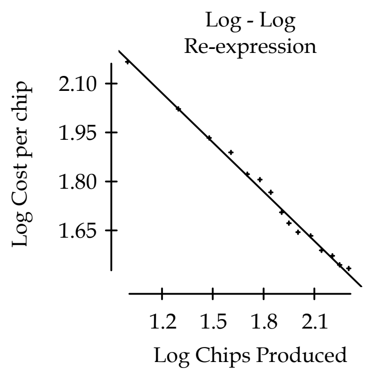


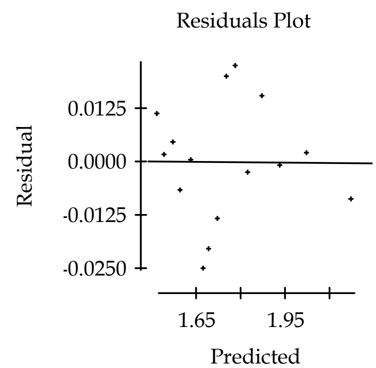


42. Chips.

The association between the number of chips produced and the cost per chip is strong, negative, and curved. As the number of chips produced increases, the cost per chip decreases. The linear model is not appropriate for this curved association, so several re-expressions of the data were tried. Re-expressing each variable using logarithms results in a scatterplot that is nearly linear.







The model, $\widehat{\log(CostperChip)} = 2.67492 - 0.501621 \log(ChipsProduced)$, has a residuals plot that shows no pattern and $R^2 = 99.5\%$. The model accounts for 99.5% of the variability in the cost per chip.

Chapter 10 – Understanding Randomness

Section 10.1

1. Random outcomes.

a) Yes, who takes out the trash cannot be predicted before the flip of a coin.

b) No, it is not random, since you will probably name a favorite team.

c) Yes, your new roommate cannot be predicted before names are drawn.

2. More random outcomes.

a) Yes, the winner cannot be determined before the ticket is drawn.

b) Yes, rolling a die results in random outcomes.

c) Probably not random. Many people tend to always pick one side or the other.

Section 10.2

3. Components.

Rolling the pair of dice is the component.

4. More components.

Obtaining a game piece is the component.

5. Response variable.

To simulate, you could roll dice and note whether or not "doubles" came up. A trial would be completed once "doubles" came up. You would count up the number of rolls until "doubles" for the response variable. Alternatively, you could use the digits 1, 2, 3, 4, 5, 6 on a random digits table and disregard digits 7, 8, 9, and 0. Using the table, note the first legal digit for the first die, and then note the next legal digit for the second die. A double would indicate rolling doubles.

6. Response variable, take two.

You could assign the digits 1, 2, 3 to Burger, 4, 5, 6 to Fries, and 7, 8, 9 to Shake, disregarding the digit 0 in a random digits table. A trial would be obtaining one digit from each game piece. Using the table, note the first legal digit for a game piece and continue until a complete set is collected. The response variable would be the number of game pieces needed to win a free meal.

Chapter Exercises.

7. The lottery.

In state lotteries, a machine pops up numbered balls. If the lottery were truly random, the outcome could not be predicted and the outcomes would be equally likely. It is random only if the balls generate numbers in equal frequencies.

8. Games.

Answers may vary.

Rolling one or two dice: If the dice are fair, then each outcome, 1 through 6 should be equally likely.

Spinning a spinner: Each outcome should be equally likely, but the spinner might be more likely to land on one outcome than another due to friction or design.

Shuffling cards and dealing a hand: If the cards are shuffled adequately (7 times for riffle shuffling), the cards will be approximately equally likely to be in any given hand.

9. Birth defects.

Answers may vary. Generate two-digit random numbers, 00-99. Let 00-02 represent a defect. Let 03-99 represent no defect.

10. Colorblind.

Answers may vary. Generate random digits 0-9. Let 0 represent colorblind. Let 1-9 represent no color perception defect.

11. Geography.

a) Looking at pairs of digits, the first state number is 45, Vermont. The next set is ignored since there is no 92^{nd} state. The next state number is 10, Georgia.

b) Continuing along, the next state number is 17, Kentucky. The next state number, 10, is ignored, since Georgia was already assigned. The final state number is 22, Michigan.

12. Get rich.

Looking at pairs of digits, you would choose 43, ignore 68, since it is not a possible lottery pick, choose 09, ignore 87, choose 50, choose 13, ignore 09, since you already chose that number, choose 27. Your numbers are 43, 9, 50, 13, 27.

13. Play the lottery.

If the lottery is random, it doesn't matter if you play the same favorite "lucky" numbers or if you play different numbers each time. All numbers are equally likely (or, rather, UNLIKELY) to win.

14. Play it again, Sam.

If the lottery is random, it doesn't matter if you play random numbers or not. All numbers are equally likely (or, rather, UNLIKELY) to win.

15. Bad simulations.

a) The outcomes are not equally likely. For example, the probability of getting 5 heads in 9 tosses is not the same as the probability of getting 0 heads, but the simulation assumes they are equally likely.

b) The even-odd assignment assumes that the player is equally likely to score or miss the shot. In reality, the likelihood of making the shot depends on the player's skill.

c) Suppose a hand has four aces. This might be represented by 1,1,1,1, and any other number. The likelihood of the first ace in the hand is not the same as for the second or third or fourth. But with this simulation, the likelihood is the same for each.

16. More bad simulations.

a) The numbers would represent the sums, but the sums are not all equally likely. For example, the probability of rolling a 7 is 6/36, but the probability of getting a 2 is only 1/36. The simulation assumes they are equally likely.

b) The number of boys in a family of 5 children is not equally likely. For example, having a total of 5 boys is less likely than having 3 boys out of 5 children. The simulation assigns the same likelihood to each event.

c) The likelihood for out, single, double, triple, and home run are not the same. The outcome of an at bat depends on the player's skill. The simulation assumes that these outcomes are equally likely.

17. Wrong conclusion.

The conclusion should indicate that the simulation **suggests** that the average length of the line would be 3.2 people. Future results might not match the simulated results exactly.

18. Another wrong conclusion.

The simulation **suggests** that 24% of the people might contract the disease. The simulation does not represent what happened, but what might have happened.

19. Election.

a) Answers will vary. A component is one voter voting. An outcome is a vote for our candidate. Using two random digits, 00-99, let 01-55 represent a vote for your candidate, and let 55-99 and 00 represent a vote for the underdog.

b) A trial is 100 votes. Examine 100 two-digit random numbers and count how many simulated votes are cast for each candidate. Whoever gets the majority of the votes wins the trial.

c) The response variable is whether the underdog wins or not. To calculate the experimental probability, divide the number of trials in which the simulated underdog wins by the total number of trials.

20. Two pair or three of a kind?

a) Answers will vary. A component is picking a single card. An outcome is the suit and denomination of the card. To simulate picking a card, generate two random digits from 00-99, let 01-52 represent the respective cards in the deck. Ignore 53-99 and 00. Alternatively, you could generate a random digit 0-9, and let 1 = spades, 2 = clubs, 3 = hearts, and 4 = diamonds. Ignore 5-9 and 0. Then generate a two digit random number 00-99, representing the denomination (01 = ace, 02 = two,…, 11 = Jack, 12 = Queen, 13 = King), ignoring 14-99 and 00.

b) A trial is a single 5-card hand. Use five sets of random numbers, ignoring repeated cards. If you were actually drawing cards, you couldn't have more than one of each card in your hand.

c) The response variable is whether the simulated hand had Two Pair, Three of a Kind, or neither. To find the experimental probability of any event, divide the number of occurrences of that event by the total number of trials.

21. Cereal.

Answers will vary. A component is the simulation of the picture in one box of cereal. One possible way to model this component is to generate random digits 0-9. Let 0 and 1 represent Hope Solo, 2-4 represent Danica Patrick, and 5-9 represent Blake Griffin. Each trial will consist of 5 random digits, and the response variable will be whether or not a complete set of pictures is simulated. Trials in which at least one of each picture is simulated will be a success. The total number of successes divided by the total number of trials will be the simulated probability of ending up with a complete set of pictures. According to the simulation, the probability of getting a complete set of pictures is expected to be about 51.5%.

22. Cereal, again.

Answers will vary. A component is the simulation of the picture in one box of cereal. One possible way to model this component is to generate random digits 0-9. Let 0 and 1 represent Hope Solo, 2-4 represent Danica Patrick, and 5-9 represent Blake Griffin. Each trial will consist of generating random numbers until a 0 or 1 is generated. The response variable will be the number of digits generated until the first 0 or 1. The total number of digits generated divided by the total number of trials will be the simulated average number of boxes required to get a Hope Solo picture. According to the simulation, in order to be reasonably assured of getting a Hope Solo picture, expect to buy about 5 boxes.

23. Multiple choice

Answers will vary. A component is one multiple-choice question. One possible way to model this component is to generate random digits 0-9. Let digits 0-7 represent a correct answer, and let digits 8 and 9 represent an incorrect answer. Each trial will consist of 6 random digits. The response variable is whether or not all 6 simulated questions are answered correctly (all 6 digits are 0-7). The total number of successes divided by the total number of trials will be the simulated probability of getting all 6 questions right. According to the simulation, the probability of getting all 6 multiple-choice questions correct is expected to be about 26%.

24. Lucky guessing?

Answers will vary. A component is one multiple-choice question. One possible way to model this component is to generate random digits 0-9. Let the digit 0 represent a correct answer, and let digits 1, 2, and 3 represent an incorrect answer. Ignore digits 4-9. Each trial will consist of 6 usable random digits. The response variable is whether or not all 6 simulated questions are answered correctly. The total number of successes divided by the total number of trials will be the simulated probability of getting all 6 questions right. Few simulations will have any trials getting all 6 correct, leading us to conclude that the probability of getting all 6 questions correct is very small. (The true probability is 0.00024). It isn't likely that your friend is telling the truth.

25. Beat the lottery.

a) Answers based on your simulation will vary, but you should win about 10% of the time.

b) You should win at the same rate with any number.

26. Random is as random does.

Answers based on your simulation will vary, but you should win about 10% of the time. Playing randomly selected lottery numbers offers no advantage to picking your own.

27. It evens out in the end.

Answers based on your simulation will vary, but you should win about 10% of the time. Playing lottery numbers that have turned up the least in recent lottery drawers offers no advantage. Each new drawing is independent of recent drawings.

28. Play the winner?

Answers based on your simulation will vary, but you should win about 10% of the time. Playing lottery numbers that have won in recent lottery drawers offers no advantage. Each new drawing is independent of recent drawings.

29. Driving test.

Answers will vary. A component is one drivers test, but this component will be modeled differently, depending on whether or not it is the first test taken. One possible way to model this component is to generate pairs of random digits 00-99. Let 01-34 represent passing the first test and let 35-99 and 00 represent failing the first test. Let 01-72 represent passing a retest, and let 73-99 and 00 represent failing a retest. To simulate one trial, generate pairs of random numbers until a pair is generated that represents passing a test. Begin each trial using the "first test" representation, and switch to the "retest" representation if failure is indicated on the first simulated test. The response variable is the number of simulated tests required to achieve the first passing test. The total number of simulated tests taken divided by the total number of trials is the simulated average number of tests required to pass. According to the simulation, the number of driving tests required to pass is expected to be about 1.9.

30. Still learning?

Answers will vary. A component is one drivers test, but this component will be modeled differently, depending on whether or not it is the first test taken. One possible way to model this component would be to generate pairs of random digits 00-99. Let 01-34 represent passing the first test and let 35-99 and 00 represent failing the first test. Let 01-72 represent passing a retest, and let 73-99 and 00 represent failing a retest. To simulate one trial, generate pairs of random numbers until a pair is generated that represents passing a test. Begin each trial using the "first test" representation, and switch to the "retest" representation if failure is indicated on the first simulated test. The response variable is whether or not the drivers test is passed within two attempts. The total number of simulated ***failed*** tests divided by the total number of trials is the simulated percentage of those tested who do not have a driver's license after two attempts. According to the simulation, the percentage that still do not pass within 2 tests is expected to be about 18%.

31. Basketball strategy.

Answers will vary. A component is one foul shot. One way to model this component would be to generate pairs of random digits 00-99. Let 01-72 represent a made shot, and let 73-99 and 00 represent a missed shot. The response variable is the number of shots made in a "one and one" situation. If the first shot simulated represents a made shot, simulate a second shot. If the first shot simulated represents a miss, the trial is over. The simulated average number of points is the total number of simulated points divided by the number of trials. According to the simulation, the player is expected to score about 1.24 points.

32. Blood donors.

Answers will vary. A component is one donor. One possible way to model this component is to generate pairs of random digits 00-99. Let 01-44 represent a type O donor, and let 45-99 and 00 represent a donor who is not type O. The response variable is the number of pairs of digits generated until 3 type O donors are simulated. Once 3 type O donors are simulated, the trial is over. The simulated average number of donors required is the total number of pairs of digits generated divided by the total number of trials. According to the simulation, about 6.8 donors are required to be reasonably assured of getting 3 type O donors.

33. Free groceries.

Answers will vary. A component is the selection of one card with the prize indicated. One possible way to model the prize is to generate pairs of random digits 00-99. Let 01-10 represent $200, let 11-20 represent $100, let 21-40 represent $50, and let 41-99 and 00 represent $20. Repeated pairs of digits must be ignored. (For this reason, a simulation in which random digits 0-9 are generated with 0 representing $200, 1 representing $100, etc., is NOT acceptable. Each card must be individually represented.) A trial continues until the total simulated prize is greater than $500. The response variable is the number of simulated customers until the payoff is greater than $500. The simulated average number of customers is the total number of simulated customers divided by the number of trials. According to the simulation, about 10.2 winners are expected each week.

34. Find the ace.

Answers will vary. A component is turning over one card. One way to model the cards turned over is to generate random digits 0-9. Let the digit 0 represent the ace, and let digits 1, 2, 3, and 4 each represent one of the other four cards. Ignore digits 5-9. A trial consists of simulating turning over the cards until the ace is drawn. Each card must be represented individually; repeated digits must be ignored. The response variable is the number of simulated cards drawn until the ace is drawn, with $100 being awarded if the ace is drawn first, and $50, $20, $10, or $5 if the ace is drawn second, third, fourth, or fifth, respectively. The simulated average dollar amount of music the store is expected to give away is the total dollar amount of music given away divided by the number of trials. According to the simulation, the dollar amount given away is expected to be about $37.

35. The family.

Answers will vary. Each child is a component. One way to model the component is to generate random digits 0-9. Let 0-4 represent a boy and let 5-9 represent a girl. A trial consists of generating random digits until a child of each gender is simulated. The response variable is the number of children simulated until this happens. The simulated average family size is the number of digits generated in each trial divided by the total number of trials. According to the simulation, the expected number of children in the family is about 3.

36. A bigger family.

Answers will vary. Each child is a component. One way to model the component is to generate random digits 0-9. Let 0-4 represent a boy and let 5-9 represent a girl. A trial consists of generating random digits until two children of each gender are simulated. The response variable is the number of children simulated until this happens. The simulated average family size is the number of digits generated in each trial divided by the total number of trials. According to the simulation, the expected number of children in the family is slightly less than 6.

37. Dice game.

Answers will very. Each roll of the die is a component. One way of modeling this component is to generate random digits 0-9. The digits 1-6 correspond to the numbers on the faces of the die, and digits 7-9 and 0 are ignored. A trial consists of generating random numbers until the sum of the numbers is exactly 10. If the sum exceeds 10, the last roll must be ignored and simulated again, but still counted as a roll. The response variable is the number of rolls until the sum is exactly 10. The simulated average number of rolls until this happens is the total number of rolls simulated divided by the number of trials. According to the simulation, expect to roll the die about 7.5 times.

38. Parcheesi.

Answers will vary. Each roll of two dice is a component. One way of modeling this component is to generate random digits 0-9. The digits 1-6 correspond to the numbers on the faces of the die, and digits 7-9 and 0 are ignored. For this simulation, look at the digits in usable pairs of digits, and consider the sum, as well as the numbers themselves. A trial consists of generating usable pairs of digits until the sum is 3, or until at least one of the dice shows a 3. The response variable is the number of pairs of usable numbers generated until this happens. The simulated average number of rolls is the total number of rolls divided by the number of trials. According to the simulation, expect to roll the dice about 2.6 times.

39. The hot hand.

Answers may vary. Each shot is a component. One way of modeling this component is to generate pairs of random digits 00-99. Let 01-65 represent a made shot, and let 66-99 and 00 represent a missed shot. A trial consists of 20 simulated shots. The response variable is whether or not the 20 simulated shots contained a run of 6 or more made shots. To find the simulated percentage of games in which the player is expected to have a run of 6 or more made shots, divide the total number of successes by the total number of trials. According to the simulation, the player is expected to make 6 or more shots in a row in about 40% of games. This isn't unusual. The announcer was wrong to characterize her performance as extraordinary.

40. The World Series.

Answers may vary. Each game is a component. One way of modeling this component is to generate pairs of random digits 00-99. Let 01-55 represent a win by the favored team, and let 56-99 and 00 represent a win by the underdog. A trial consists of generating pairs until one team has 4 simulated wins. The response variable is whether or not the underdog wins. The simulated percentage of World Series wins is the total number of successes divided by the total number of trials. According to simulation, the underdog is expected to win the World Series about 39% of the time.

41. Teammates.

Answers will vary. Each player chosen is a component. One way to model this component is to generate random numbers 0-9. Let 1 and 2 represent the first couple, 3 and 4 the second couple, 5 and 6 the third couple, and 7 and 8 the fourth couple. Ignore 9 and 0. A trial consists of generating random digits, ignoring repeats, and organizing them into pairs, until pairs representing the first three teams are generated. (The final team is assigned by default.) The response variable is whether or not each of the simulated teams is a pairing other than 1-2, 3-4, 5-6, or 7-8. The simulated percentage of the time this is expected to happen is the total number of successes (times that the pairings are *different* than the couples) divided by the total number of trials. According to the simulation, all players are expected to be paired with someone other than the person with whom he or she came to the party about 37.5% of the time.

42. Second team.

Answers will vary. Each player chosen is a component. One way to model this component is to generate random numbers 0-9. Let digits 1-4 represent the four players who are to be chosen. Ignore digits 5-9 and 0. A trial consists of generating a sequence of random numbers that represents the order in which the cards where chosen. Since each number represents a person, and people cannot be chosen more than once, ignore repeated numbers. The response variable is whether or not any digit in the generated sequence matches the corresponding digit in the sequence 1234 (or any other sequence of the four numbers, as long as it is determined ahead of time). The simulated percentage of the time this is expected to happen is the total number of successes (times that the sequences have no matching corresponding digits) divided by the number of trials. According to the simulation, all players are expected to be paired with someone other than the person with whom he or she came to the party about 37.5% of the time.

43. Job discrimination?

Answers may vary. Each person hired is a component. One way of modeling this component is to generate pairs of random digits 00-99. Let 01-10 represent each of the 10 women, and let 11-22 represent each of the 12 men. Ignore 23-99 and 00. A trial consists of 3 usable pairs of numbers. Ignore repeated pairs of digits, since the same man or woman cannot be hired more than once. The response variable is whether or not all 3 simulated hires are women. The simulated percentage of the time that 3 women are expected to be hired is the number of successes divided by the number of trials. According to the simulation, the 3 people hired will all be women about 7.8% of time. This seems a bit strange, but not quite strange enough to be evidence of job discrimination.

44. Smartphones.

Answers will vary. Each driver is a component. One way to model this component is to generate random digits 00-99. Let 01-10 represent a driver that is using his or her phone, and let 11-99 and 00 represent a driver that is not using his or her phone. A trial consists of 20 pairs of digits. The response variable is whether or not at least 5 of the simulated drivers were using their phones. The simulated percentage of the time that 5 or more drivers of twenty are using their phones if the true rate of usage is 10%is the number of successes divided by the total number of trials. You should expect to find 5 or more drivers using phones among 20 drivers only about 4.3% of the time. Based on what you saw on your drive home, you'd suspect that the legislator's claim of 10% usage is probably too low.

Chapter 11 – Sample Surveys

Section 11.1

1. Texas A&M.

The A&M administrators should take a survey. They should sample a part of the student body, selecting respondents with a randomization method. They should be sure to draw a sufficiently large sample.

2. Satisfied workers.

They gathered data from only a part of the large population of employees, and they selected that part at random. A sample size of several hundred is a reasonable size. It's not really clear, though, whether 437 employees were contacted, or if a larger number were contacted, and only 437 completed the survey. There may be nonresponse bias in the sample.

Section 11.2

3. A&M again.

The proportion in the sample is a statistic. The proportion of all students is the parameter of interest. The statistic estimates that parameter, but is not likely to be exactly the same.

4. Satisfied respondents.

The survey result is a statistic. It estimates the true proportion of satisfied workers, but does not give that value precisely.

Section 11.3

5. Sampling students.

This is not an SRS. Although each student may have an equal chance to be in the survey, groups of friends who choose to sit together will either all be in or out of the sample, so the selection is not independent.

6. Sampling satisfaction.

Yes. Each employee has an equal and independent chance of being sampled.

Section 11.4

7. Sampling A&M students.

a) This is a cluster sample, with each selected dormitory as a cluster.

b) This is a stratified sample, stratified by class year.

c) This systematic sample, with a randomized starting point.

Section 11.5

8. Satisfactory satisfaction samples.

a) This is a stratified sample, stratified by duration of employment.

b) This is a systematic sample, without a random starting point.

c) This is a multi-stage sample.

Section 11.6

9. Survey students.

Several terms are poorly defined. The survey needs to specify the meaning of "family" for this purpose and the meaning of "higher education." The term "seek" may also be poorly defined (for example, would applying to college but not being admitted qualify for seeking more education?)

10. Happy employees.

The survey is likely to be biased because employees won't want to express unhappiness in front of their supervisors or their co-workers. This is a form of response bias.

Section 11.7

11. Student samples.

a) This would suffer from voluntary response bias. Only those students who saw the advertisement could even be part of the sample, and only those who choose to (and are able to) go to the website would actually be in the sample.

b) This would be a convenience sample, as well as suffer from voluntary response bias.

12. Surveying employees.

a) This is a convenience sample. It would also suffer from voluntary response bias. Furthermore, it also may contain response bias. Employees may worry that bosses or other workers are watching to see who takes a survey, which may cause them to answer differently than they would have otherwise.

b) There is likely to be substantial nonresponse bias.

Chapter Exercises.

13. Roper.

a) Roper is not using a simple random sample. The samples are designed to get 500 males and 500 females. This would be very unlikely to happen in a simple random sample.

b) They are using stratified sample, with two strata, males and females.

14. Student center survey.

- **a)** The students are not using a simple random sample. The samples are designed to get 50 students from each grade level. This would be very unlikely to happen in a simple random sample.
- **b)** They are using a stratified sample, with four strata, one for each class year.

15. Drug tests.

- **a)** This is a cluster sample, with teams being the clusters.
- **b)** Cluster sampling is a reasonable solution to the problem of randomly sampling players because an entire team can be sampled easily. It would be much more difficult to randomly sample players from many different teams on the same day.

16. Gallup.

- **a)** The population of interest is all adults in the United States aged 18 and older.
- **b)** The sampling frame is U.S. adults with telephones.
- **c)** Some members of the population (e.g. many college students) don't have land-line telephones, so they could never be chosen in the sample. This may create a bias.

17. Medical treatments.

- **a)** **Population** – Unclear, but possibly all U.S. adults.
- **b)** **Parameter** – Proportion who have used and have benefitted from alternative medical treatments.
- **c)** **Sampling Frame** – All Consumers Union subscribers.
- **d)** **Sample** – Those subscribers who responded.
- **e)** **Method** – Not specified, but probably a questionnaire mailed to all subscribers.
- **f)** **Left Out** – Those who are not Consumers Union subscribers.
- **g)** **Bias** – Voluntary response bias. Those who respond may have strong feelings one way or another.

18. Social life.

- **a)** **Population** – Unclear, but possibly all U.S. young adults.
- **b)** **Parameter** – Proportion who don't or don't really have a social life outside of the Internet.
- **c)** **Sampling Frame** – Visitors to the website. The surveyors did not have a defined sampling frame in mind when designing the survey.

d) Sample – Website visitors that chose to respond to the survey.

e) Method – Nonrandom Internet questionnaire.

f) Left Out – Those who did not visit gamefaqs.com and those who did visit, but did not respond to the survey.

g) Bias – Voluntary response bias. Participants chose whether or not to participate. Additionally, undercoverage limits the scope of any conclusions to the population of young adults, since only those that visited gamefaqs.com could have even chosen to be in the sample.

19. Mayoral race.

a) Population – City voters.

b) Parameter – Not clear. They might be interested in the percentage of voters favoring various issues.

c) Sampling Frame – All city residents

d) Sample – As many residents as they can find in one block from each district. No randomization is specified, but hopefully a block is selected at random within each district.

e) Method – Multistage sampling, stratified by district and clustered by block.

f) Left Out – People not home during the time of the survey.

g) Bias – Convenience sampling. Once the block is randomly chosen as the cluster, every resident living in that block should be surveyed, not just those that were conveniently available. A random sample of each block could be also be taken, but we wouldn't refer to that as "cluster" sampling, but rather multi-stage, with stratification by district, a simple random sample of one block within each district, and another simple random sample of residents within the block.

20. Soil samples.

a) Population – Soil around a former waste dump.

b) Parameter – Proportion with elevated levels of harmful substances, or perhaps a measurement of the actual levels of harmful substances.

c) Sampling Frame – Accessible soil around the dump.

d) Sample – 16 soil samples.

e) Method – Not clear. There is no indication of random sampling.

f) Left Out – Soil in inaccessible areas around the dump.

g) Bias – possible convenience sample. Since there is no indication of randomization, the samples may have been taken from easily accessible areas. Soil in these areas may be more or less polluted than the soil in general.

21. Roadblock.

a) **Population** – All cars.

b) **Parameter** – Proportion of cars with up-to-date (or out-of-date) registrations, insurance, or safety inspections.

c) **Sampling Frame** – Cars on that road.

d) **Sample** – Cars stopped by the roadblock.

e) **Method** – Cluster sample of an area, stopping all cars within the cluster.

f) **Left Out** – Drivers that did not take that road, or traveled that road at a different time.

g) **Bias** – Undercoverage. The cars stopped might not be representative of all cars because of time of day and location. The locations are probably not chosen randomly, so might represent areas in which it is easy to set up a roadblock, resulting in a convenience sample.

22. Snack foods.

a) **Population** – Snack food bags.

b) **Parameter** – Proportion passing inspection, or perhaps weight of bags.

c) **Sampling Frame** – All bags produced each day.

d) **Sample** – 10 bags, one from each of 10 randomly selected cases.

e) **Method** – Multistage sampling. Presumably, they take a simple random sample of 10 cases, followed by a simple random sample of one bag from each case.

f) **Left Out** – Nothing.

g) **Bias** – No indication of bias.

23. Milk samples.

a) **Population** – Dairy farms.

b) **Parameter** – Whether or not the milk contains dirt, antibiotics, or other foreign matter.

c) **Sampling Frame** – All dairy farms

d) **Sample** – Not specified, but probably a random sample of farms.

e) **Method** – not specified

f) **Left Out** – Nothing.

g) **Bias** – Unbiased, as long as the day of inspection is randomly chosen. This might not be the case, however, since the farms might be spread out over a wide geographic area. Inspectors might tend to visit farms that are near one another on the same day, resulting in a convenience sample.

24. Mistaken poll.

The station's faulty prediction is more likely to be the result of bias. Only people watching the news were able to respond, and their opinions were likely to be different from those of other voters. The sampling method may have systematically produced samples that did not represent the population of interest.

25. Another mistaken poll.

The newspaper's faulty prediction was more likely to be due to sampling error. The description of the sampling method suggests that samples should be representative of the voting population. Random chance in selecting the individuals who were polled means that sample statistics will vary from the population parameter, perhaps by quite a bit.

26. Parent opinion, part 1.

a) This is a voluntary response sample. Only those who see the ad, feel strongly about the issue, and have web access will respond.

b) This is cluster sampling, but probably not a good idea. The opinions of parents in one school may not be typical of the opinions of all parents.

c) This is an attempt at a census, and will probably suffer from nonresponse bias.

d) This is stratified sampling. If the follow-up is carried out carefully, the sample should be unbiased.

27. Parent opinion, part 2.

a) This sampling method suffers from voluntary response bias. Only those who see the show and feel strongly will call.

b) Although this method may result in a more representative sample than the method in part **a)**, this is still a voluntary response sample. Only strongly motivated parents attend PTA meetings.

c) This is multistage sampling, stratified by elementary school and then clustered by grade. This is a good design, as long as the parents in the class respond. There should be follow-up to get the opinions of parents who do not respond.

d) This is systematic sampling. As long as a starting point is randomized, this method should produce reliable data.

28. Churches.

a) This is a multistage design, with a cluster sample at the first stage and a simple random sample for each cluster.

b) If any of the three churches you pick at random are not representative of all churches, then your sample will reflect the makeup of that church, not all churches. Also, choosing 100 members at random from each church could introduce bias. The views of the members of smaller churches chosen in the sample will be weighted heavier in your sample than the views of members of larger churches, especially if the views of the members of that small church differ from the views of churchgoers at large. The hope is that random sampling will equalize these sources of variability in the long run.

29. Playground.

The managers will only get responses from people who come to the park to use the playground. Parents who are dissatisfied with the playground may not come.

30. Roller coasters.

a) This is a systematic sample.

b) The sampling frame is patrons willing to wait in line for the roller coaster. The sample should be representative of the people in line, but not of all the people at the park.

c) This sample is likely to be representative of those waiting in line for the roller coaster, especially if those people at the front of the line (after their long wait) respond differently from those at the end of the line.

d) Many people may see the long line, and choose not to go on the ride. These members of the population are excluded.

31. Playground, act two.

The first sentence points our problems that the respondent may not have noticed, and might lead them to feel they should agree. The last phrase mentions higher fees, which could make people reject improvements to the playground.

32. Wording the survey.

a) Responses to these questions will differ. Question 1 will probably get "no" answers, and Question 2 will probably get "yes" answers. This is response bias, based on the wording of the questions.

b) A question with neutral wording might be: "Do you think standardized tests are appropriate for deciding whether a student should be promoted to the next grade?"

33. Banning ephedra.

a) This is a voluntary response survey. The large sample will still be affected by any biases in the group of people that choose to respond.

b) The wording seems fair enough. It states the facts, and gives voice to both sides of the issue.

c) The sampling frame is, at best, those who visit this particular site, and even then depends of their volunteering to respond to the question.

d) This statement is true.

34. Survey questions.

a) The question is biased toward "yes" answers because of the word "pollute". A better question might be: "Should companies be responsible for any costs of environmental clean up?"

b) The question is biased toward "no" because of the preamble "18-year-olds are old enough to serve in the military. A better question might be: "Do you think the drinking age should be lowered from 21?"

35. More survey questions.

a) The question seems unbiased.

b) The question is biased toward "yes" because of the phrase "great tradition". A better question: "Do you favor continued funding for the space program?"

36. Phone surveys.

a) A simple random sample is difficult in this case because there is a problem with undercoverage. People with unlisted phone numbers and those without phones are not in the sampling frame. People who are at work, or otherwise away from home, are included in the sampling frame. These people could never be in the sample itself.

b) One possibility is to generate random phone numbers and call at random times, although obviously not in the middle of the night! This would take care of the undercoverage of people at work during the day, as well as people with unlisted phone numbers, although there is still a problem avoiding undercoverage of people without phones.

c) Under the original plan, those families in which one person stays home are more likely to be included. Under the second plan, many more are included. People without phones are still excluded.

d) Follow-up of this type greatly improves the chance that a selected household is included, increasing the reliability of the survey.

e) Random dialers allow people with unlisted phone numbers to be selected, although they may not be willing participants. There is a reason the number is unlisted. Time of day will still be an issue, as will people without phones.

37. Cell phone survey.

Cell phones are more likely to be used by younger individuals. This will result in an undercoverage bias. As cell phone use grows, this will be less of a problem. Also, many cell phone plans require the users to pay airtime for incoming calls. That seems like a sure way to irritate the respondent, and result in response bias toward negative responses.

38. Arm length.

a) Answers will vary. My arm length is 3 hand widths and 2 finger widths.

b) The parameter estimated by 10 measurements is the true length of your arm. The population is all possible measurements of your arm length.

c) The population is now the arm lengths of your friends. The average now estimates the mean of the arm lengths of your friends.

d) These 10 arm lengths are unlikely to be representative of the community, or the country. Your friends are likely to be of the same age, and not very diverse.

39. Fuel economy.

a) The statistic calculated is the mean mileage for the last six fill-ups.

b) The parameter of interest is the mean mileage for the vehicle.

c) The driving conditions for the last six fill-ups might not be typical of the overall driving conditions. For instance, the last six fill-ups might all be in winter, when mileage might be lower than expected.

d) The EPA is trying to estimate the mean gas mileage for all cars of this make, model, and year.

40. Accounting.

a) Assign numbers 001-120 to each order. Generate 10 random numbers 001-120, and select those orders to recheck.

b) The supervisor should perform a stratified sample, randomly checking a certain percentage of each type of sales, retail and wholesale.

41. Happy workers?

a) A small sample will probably consist mostly laborers, with few supervisors, and maybe no project managers. Also, there is a potential for response bias based on the interviewer if a member of management asks directly about discontent. Workers who want to keep their jobs will likely tell the management that everything is fine!

b) Assign a number from 001 to 439 to each employee. Use a random number table or software to select the sample.

c) The simple random sample might not give a good cross section of the different types of employees. There are relatively few supervisors and project managers, and we want to make sure their opinions are noted, as well as the opinions of the laborers.

d) A better strategy would be to stratify the sample by job type. Sample a certain percentage of each job type.

e) Answers will vary. Assign each person a number from 01-14, and generate 2 usable random numbers from a random number table or software.

42. Quality control.

a) Select three cases at random, then select one jar randomly from each case.

b) Generate three random numbers between 61-80, with no repeats, to select three cases. Then assign each of the jars in the case a number 01-12, and generate one random number for each case to select the three jars, one from each case.

c) This is not a simple random sample, since there are groups of three jars that cannot be the sample. For example, it is impossible for three jars in the same case to be the sample. This would be possible if the sample were a simple random sample.

43. A fish story.

What conclusions they may be able to make will depend on whether fish with discolored scales are equally likely to be caught as those without. It also depends on the level of compliance by fisherman. If fish are not equally likely to be caught, or fishermen more disposed to bring discolored fish, the results will be biased.

44. Sampling methods.

a) This method would probably result in undercoverage of those doctors that are not listed in the Yellow Pages. Using the "line listings" seems fair, as long as all doctors are listed, but using the advertisements would not be a typical list of doctors.

b) This method is not appropriate. This cluster sample will probably contain listings for only one or two types of businesses, not a representative cross-section of businesses.

45. More sampling methods.

a) A petition may pressure people into support. Additionally, some people may not be home on a Saturday, especially those who have taken their kids out to play in a distant park! We are undercovering a group made up of people who probably have a specific opinion.

b) If the food at the largest cafeteria is representative, this should be OK. However, those who really don't like the food won't be eating there. That group is undercovered.

Chapter 12 – Experiments and Observational Studies

Section 12.1

1. Steroids.

This is an observational study because the sports writer is not randomly assigning players to take steroids or not take steroids; the writer is merely observing home run totals between two eras. It would be unwise to conclude steroids caused any increases in home runs because we need to consider other factors besides steroids—factors possibly leading to more home runs include better equipment, players training more in the offseason, smaller ballparks, better scouting techniques, etc.

2. E-commerce.

This is an observational study because the student is not randomly assigning companies to use or not use the Internet for business transactions. If profitability did increase in the 2000s, it could have been due to a number of factors, not specifically the Internet as a means for conducting business.

Section 12.2

3. Tips.

Each of the 40 deliveries is an experimental unit. He has randomized the experiment by flipping a coin to decide whether or not to phone.

4. Tomatoes.

Each tomato plant is an experimental unit. The tastiness and juiciness of the tomatoes will be the response variables.

5. Tips II.

The factor is calling, and the levels are whether or not he calls the customer. The response variable is the tip percentage for each delivery.

6. Tomatoes II.

The factor is the fertilizer, applied at three levels 0, half, and full dose. To measure tastiness and juiciness, we'll need trained tasters.

Section 12.3

7. Tips again.

By calling some customers but not others during the same run, the driver has controlled many variables, such as day of the week, season, and weather. The experiment was randomized because he flipped a coin to determine whether or not to phone and it was replicated because he did this for 40 deliveries.

8. Tomatoes again.

Tomato plants should be grown in the same field, near each other so differences in soil, sun, and rain can be controlled. The experiment is randomized because plants are assigned at random to treatment levels. It is replicated because 6 plants are assigned to each level.

Section 12.4

9. More tips.

Because customers don't know about the experiment, those that are called don't know that others are not, and vice versa. Thus, the customers are blind. That would make this a single-blind study. It can't be double-blind because the delivery driver must know whether or not he phones.

10. More tomatoes.

If the tomato taster is blind, then this is a single-blind study. To make it double-blind, everyone who cares for the tomato plants must be blind to their treatment. This might be done, for example, by treating all plants with solutions that look the same, but applying a "placebo" fertilizer to the plants assigned to receive none.

Section 12.5

11. Block that tip.

Yes. Driver is now a block. The experiment is randomized within each block. This is a good idea because some drivers might generally get higher tips than others, but the goal of the experiment is to study the effect of phone calls. Blocking on driver eliminates the variability in tips inherent to the driver.

12. Blocking tomatoes.

Yes. Garden centers are the blocks. It is important to randomize the assignment of plants to treatments within each block so that any differences between garden centers won't affect the results.

Section 12.6

13. Confounded tips.

Answers may vary. The cost or size of a delivery may confound his results. Larger orders may generally tip a higher or lower percentage of the bill.

14. Tomatoes finis.

Answers may vary. Confounding factors could include variations in soil fertility, sunlight availability, or rainfall. Some plants might become infested with pests.

Chapter Exercises

15. Standardized test scores.

a) No, this is not an experiment. There are no imposed treatments. This is a retrospective observational study.

b) We cannot conclude that the differences in score are caused by differences in parental income. There may be lurking variables that are associated with both SAT score and parental income.

16. Heart attacks and height.

a) No, this is not an experiment. There are no imposed treatments. This is a retrospective observational study.

b) We cannot conclude that shorter men are at higher risk of dying from a heart attack. There may be lurking variables that are associated with both height and risk of heart attack.

17. MS and vitamin D.

a) This is a retrospective observational study.

b) This is an appropriate choice, since MS is a relatively rare disease.

c) The subjects were U.S. military personnel, some of whom had developed MS.

d) The variables were the vitamin D blood levels and whether or not the subject developed MS.

18. Super Bowl commercials.

a) This is a stratified sample. The question was about population values, namely the proportions of men and women who look forward to more commercials. No treatment was applied, so this is not an experiment.

b) Yes, the design was appropriate.

19. Menopause.

a) This was a randomized, comparative, placebo-controlled experiment.

b) Yes, such an experiment is the right way to determine whether black cohosh is an effective treatment for hot flashes.

c) The subjects were 351 women, aged 45 to 55 who reported at least two hot flashes a day.

d) The treatments were black cohosh, a multi-herb supplement, plus advice to consume more soy foods, estrogen, and a placebo. The response was the women's self-reported symptoms, presumably the frequency of hot flashes.

20. Honesty.

a) This is an experiment. The picture is the controlled factor. Randomization may have been used to decide which days each picture appeared.

b) The treatment was the picture behind the coffee station. The response variable was the average contribution.

c) The differences in money contributed were larger than could be reasonably attributed to usual day-to-day variation.

21. a) This is an experiment, since treatments were imposed.
b) The subjects studied were 30 patients with bipolar disorder.
c) The experiment has 1 factor (omega-3 fats from fish oil), at 2 levels (high dose of omega-3 fats from fish oil and no omega-3 fats from fish oil).
d) 1 factor, at 2 levels gives a total of 2 treatments.
e) The response variable is "improvement", but there is no indication of how the response variable was measured.
f) There is no information about the design of the experiment.
g) The experiment is blinded, since the use of a placebo keeps the patients from knowing whether or not they received the omega-3 fats from fish oils. It is not stated whether or not the evaluators of the "improvement" were blind to the treatment, which would make the experiment double-blind.
h) Although it needs to be replicated, the experiment can determine whether or not omega-3 fats from fish oils cause improvements in patients with bipolar disorder, at least over the short term. The experiment design would be stronger is it were double-blind.

22. a) This is an observational study. The researchers are simply studying traits that already exist in the subjects, not imposing new treatments.
b) This is a prospective study. The subjects were identified first, then traits were observed.
c) The subjects were disabled women aged 65 and older, with and without a vitamin B-12 deficiency. The selection process is not stated.
d) The parameter of interest is the percentage of women in each group who suffered severe depression.
e) There is no random assignment, so a cause-and-effect relationship between B-12 deficiency and depression cannot be established. The most that can be determined is an association, if this is supported by the data.

23. a) This is an observational study. The researchers are simply studying traits that already exist in the subjects, not imposing new treatments.
b) This is a prospective study. The subjects were identified first, then traits were observed.

c) The subjects are roughly 200 men and women with moderately high blood pressure and normal blood pressure. There is no information about the selection method.

d) The parameters of interest are difference in memory and reaction time scores between those with normal blood pressure and moderately high blood pressure.

e) An observational study has no random assignment, so there is no way to know that high blood pressure caused subjects to do worse on memory and reaction time tests. A lurking variable, such as age or overall health, might have been the cause. The most we can say is that there was an association between blood pressure and scores on memory and reaction time tests in this group, and recommend a controlled experiment to attempt to determine whether or not there is a cause-and-effect relationship.

24. a) This is an experiment, since treatments were imposed on randomly assigned groups.

b) The subjects were 40 volunteers suffering from insomnia.

c) There are 2 factors in this experiment (dessert and exercise). The dessert factor has 2 levels (no dessert and normal dessert). The exercise factor has 2 levels (no exercise and an exercise program).

d) 2 factors, with 2 levels each, results in 4 treatments.

e) The response variable is improvement in ability to sleep.

f) This experiment is probably completely randomized.

g) This experiment does not use blinding.

h) This experiment indicates that insomniacs who exercise and refrain from desserts will experience improved ability to sleep.

25. a) This is an experiment, since treatments were imposed on randomly assigned groups.

b) 24 post-menopausal women were the subjects in this experiment.

c) There is 1 factor (type of drink), at 2 levels (alcoholic and non-alcoholic). (Supplemental estrogen is not a factor in the experiment, but rather a blocking variable. The subjects were not given estrogen supplements as part of the experiment.)

d) 1 factor, with 2 levels, is 2 treatments.

e) The response variable is an increase in estrogen level.

f) This experiment utilizes a blocked design. The subjects were blocked by whether or not they used supplemental estrogen. This design reduces variability in the response variable of estrogen level that may be associated with the use of supplemental estrogen.

g) This experiment does not use blinding.

h) This experiment indicates that drinking alcohol leads to increased estrogen level among those taking estrogen supplements.

26. a) This is an observational study.
b) The study is retrospective. Results were obtained from pre-existing medical records.
c) The subjects in this study were 981 women who lived near the site of dioxin release.
d) The parameter of interest is the incidence of breast cancer.
e) As there is no random assignment, there is no way to know that the dioxin levels caused the increase in breast cancer. There may have been lurking variables that were not identified.

27. a) This is an observational study.
b) The study is retrospective. Results were obtained from pre-existing church records.
c) The subjects of the study are women in Finland. The data were collected from church records dating 1640 to 1870, but the selection process is unknown.
d) The parameter of interest is difference in average lifespan between mothers of sons and daughters.
e) For this group, having sons was associated with a decrease in lifespan of an average of 34 weeks per son, while having daughters was associated with an unspecified increase in lifespan. As there is no random assignment, there is no way to know that having sons caused a decrease in lifespan.

28. a) This is an experiment, since treatments were imposed on randomly assigned groups.
b) The subjects were volunteers exposed to a cold virus.
c) There is 1 factor (herbal compound), at 2 levels (herbal compound and sugar solution).
d) 1 factor, at 2 levels, results in 2 treatments.
e) The response variable is the severity of cold symptoms.
f) There is no mention of any randomness is the design. Hopefully, subjects were randomly assigned to treatment groups.
g) The experiment uses blinding. The use of a sugar solution as a placebo kept the subjects from knowing whether or not they had received the herbal compound. If the doctors responsible for assessing the severity of the patients' colds were also unaware of the treatment group assignments, then the experiment incorporates double blinding.
h) There is no evidence to suggest that the herbal treatment is effective.

29. a) This is an observational study. (Although some might say that the sad movie was "imposed" on the subjects, this was merely a stimulus used to trigger a reaction, not a treatment designed to attempt to influence some response variable. Researchers merely wanted to observe the behavior of two different groups when each was presented with the stimulus.)

b) The study is prospective. Researchers identified subjects, and then observed them after the sad movie.
c) The subjects in this study were people with and without depression. The selection process is not stated.
d) The parameter of interest is the difference in crying response between depressed and nondepressed people exposed to sad situations.
e) There is no apparent difference in crying response to sad movies for the depressed and nondepressed groups.

30. a) This is an experiment.
b) The subjects were racing greyhounds.
c) There is 1 factor (level of vitamin C in diet). The 3 levels of diet were not specified.
d) One factor, at 3 levels, results in 3 treatments.
e) The response variable is speed.
f) The experiment uses a matched design. Each greyhound was given each of the 3 levels of diet, in random order. The matched design reduces variation due to the racing ability of each greyhound.
g) There is no mention of blinding.
h) Greyhounds that eat diets high in vitamin C run more slowly than greyhounds with diets lower in vitamin C.

31. a) This is an experiment. Subjects were randomly assigned to treatments.
b) The subjects were people experiencing migraines.
c) There are 2 factors (pain reliever and water temperature). The pain reliever factor has 2 levels (pain reliever or placebo), and the water temperature factor has 2 levels (ice water and regular water).
d) 2 factors, at 2 levels each, results in 4 treatments.
e) The response variable is the level of pain relief.
f) The experiment is completely randomized.
g) The subjects are blinded to the pain reliever factor through the use of a placebo. The subjects are not blinded to the water factor. They will know whether they are drinking ice water or regular water.
h) The experiment may indicate whether pain reliever alone or in combination with ice water give pain relief, but patients are not blinded to ice water, so the placebo effect may also be the cause of any relief seen due to ice water.

32. a) This is an experiment. Hopefully, dogs are randomly assigned to different treatment groups.
b) The subjects are inactive dogs.

c) There is 1 factor (type of dog food), at 2 levels (low-calorie and standard). One possible difficulty with this experiment is that some owners might feed their dogs more food than others. We will assume that the dog food company has given the owners specific instructions about the quantity of food required, based on the size of each dog.

d) 1 factor, at 2 levels, results in 2 treatments.

e) The response variable is the weight of the dogs.

f) The experiment uses blocking by size of breed. Blocking by size reduces variation in weight that may be due to overall size of the dog.

g) Assuming that the dog owners do not know which type of dog food their dog is receiving, the experiment is blinded.

h) Assuming the dog owners followed the prescribed feeding levels, there could be a conclusion as to whether or not the dog food helped the dogs maintain a healthy weight.

33. a) This is an experiment. Athletes were randomly assigned to one of two exercise programs.

b) The subjects are athletes suffering hamstring injuries.

c) There is one factor (type of exercise), at 2 levels (static stretching, and agility and trunk stabilization).

d) 1 factor, at 2 levels, results in 2 treatments.

e) The response variable is the time before the athletes were able to return to sports.

f) The experiment is completely randomized.

g) The experiment employs no blinding. The subjects know what kind of exercise they do.

h) Assuming that the athletes actually followed the exercise program, this experiment can help determine which of the two exercise programs is more effective at rehabilitating hamstring injuries.

34. a) This is an observational study. The researchers are simply studying traits that already exist in the subjects, not imposing new treatments.

b) This is a prospective study. Researchers identified two groups, and studied their traits.

c) The subjects are members of the general public, chosen in two random samples.

d) The purpose of this study was to identify variables on which there was a difference , so no response variable(s) could be identified at the start of the study.

e) This study will allow researchers to identify differences between people who can be reached by ordinary 5-day polling methods and those who cannot be reached.

35. Omega-3.

The experimenters need to compare omega-3 results to something. Perhaps bipolarity is seasonal and would have improved during the experiment anyway.

36. Insomnia.

The experimenters need a basis for comparison. Perhaps insomnia is related to the amount of daylight, and that changed during the time when the experiment was conducted.

37. Omega-3 revisited.

a) Subjects' responses might be related to other factors, like diet, exercise, or genetics. Randomization should equalize the two groups with respect to unknown factors.

b) More subjects would minimize the impact of individual variability in the responses, but the experiment would become more costly and time-consuming.

38. Insomnia again.

a) Subjects responses might be related to many other factors, such as diet, medications, or genetics. Randomization should equalize the two groups with respect to unknown factors.

b) More subjects would minimize the impact of individual variability in the responses, but the experiment would become more costly and time-consuming.

39. Omega-3 finis.

The researchers believe that people who engage in regular exercise might respond differently to the omega-3. This additional variability could obscure the effectiveness of the treatment.

40. Insomnia, at last.

The researchers believe that people who are overweight might respond differently to exercise. This additional variability could obscure the effectiveness of the treatment.

41. Injuries.

Answers may vary. Use a random-number generator to randomly select 24 numbers from 01 to 24 without replication. Assign the first 8 numbers to the first group, the second 8 numbers to the second group, and the third 8 numbers to the third group. If an athlete states that he would prefer the other program, he should not be allowed to switch, since a successful experiment requires randomization. In fact, he should probably be eliminated from the study, since there is a risk that he will not follow the directions for his assigned program. He may instead opt to follow the protocols for the other program anyway, since he has an obvious preference. This would invalidate the results of the experiment.

42. Tomatoes II.

Answers may vary. Number the tomatoes plants 1 to 24. Use a random number generator to randomly select 24 numbers from 1 to 24 without replication. Assign the tomato plants matching the first 8 numbers to the first group, the second 8 numbers to the second group, and the third group of 8 numbers to the third group.

43. Shoes.

a) First, the manufacturers are using athletes who have a vested interest in the success of the shoe by virtue of their sponsorship. They should try to find some volunteers that aren't employed by the company! Second, they should randomize the order of the runs, not run all the races with the new shoes second. They should blind the athletes by disguising the shoes, if possible, so they don't know which is which. The experiment could be double blinded, as well, by making sure that the timers don't know which shoes are being tested at any given time. Finally, they should replicate several times since times will vary under both shoe conditions.

b) First of all, the problems identified in part a would have to be remedied before *any* conclusions can be reached. Even if this is the case, the results cannot be generalized to all runners. This experiment compares effects of the shoes on speed for Olympic class runners, not runners in general.

44. Swimsuits.

The "control" in this experiment is not the same for all swimmers. We don't know what "their old swim suit" means. They should compare their new swim suit to the same suit design. The order in which the swims are performed should be randomized. There may be a systematic difference from one swim to the next. For instance, swimmers may be tired after the first swim (or more warmed up). Finally, there is no way to blind this test. The swimmer will know which kind of suit they have on, and this may bias their performance.

45. Hamstrings.

a) Allowing the athletes to choose their own treatments could confound the results. Other issues such as severity of injury, diet, age, etc., could also affect time to heal, and randomization should equalize the two treatment groups with respect to any such variables.

b) A control group could have revealed whether either exercise program was better (or worse) than just letting the injury heal without exercise.

c) Although the athletes cannot be blinded, the doctors who approve their return to sports should not know which treatment the subject had engaged in.

d) It's difficult to say with any certainty, since we aren't sure if the distributions of return times are unimodal and roughly symmetric, and contain no outliers. Otherwise, the use of mean and standard deviation as measures of center and spread is questionable. Assuming mean and standard deviation are appropriate measures, the subjects who exercised with agility and trunk stabilization had a mean return time of 22.2 days compared to the static stretching group, with a mean return time of 37.4 days. The agility and trunk stabilization group also had a much more consistent distribution of return times, with a standard deviation of 8.3 days, compared to the standard deviation of 27.6 days for the static stretching group. This appears to be a statistically significant difference.

46. Diet and blood pressure.

a) Self-selection could result in groups that are very different at the start of the experiment, making it impossible to attribute differences in the results to the diet alone.

b) The meals were prepared by dieticians to ensure that the diets were followed and that all subjects received comparable treatments.

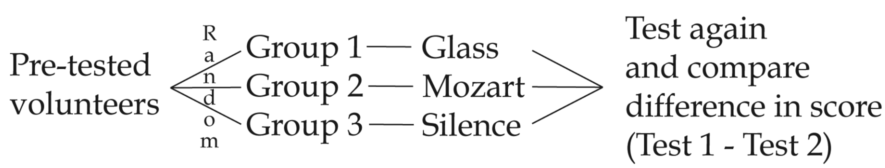
c) The researchers can compare the change in blood pressure observed in the DASH group to the control group. They need to rule out the possibility that external variables (like the season, news events, etc.) affected everyone's blood pressure.

d) We would like to know the standard deviation of the changes, as well. If the standard deviation is very small, then 6.7 points would seem like a significant change. If not, 6.7 points could be due to naturally occurring variability.

47. Mozart.

a) The differences in spatial reasoning scores between the students listening to Mozart and the students sitting quietly were more than would have been expected from ordinary sampling variation.

b)



c) The Mozart group seems to have the smallest median difference in spatial reasoning test score and thus the *least* improvement, but there does not appear to be a significant difference.

d) No, the results do not prove that listening to Mozart is beneficial. If anything, there was generally less improvement. The difference does not seem significant compared with the usual variation one would expect between the three groups. Even if type of music has no effect on test score, we would expect some variation between the groups.

48. Contrast Baths.

a) There is no evidence that the changes in swelling were different among the three treatments. Any differences see between the treatments could be attributed to chance alone. Contrast baths, with or without exercise, did not appear to reduce swelling any more than exercise alone.

b) Patients were assigned to treatments at random. The exercise treatment was the control group.

c) If exercise is the standard treatment, then using it as a control seems appropriate (and would correspond to the Helsinki guidelines.) The use of a placebo could be viewed as unethical, since effective treatment is withheld.

49. Wine.

a) This is a prospective observational study. The researchers followed a group of children born at a Copenhagen hospital between 1959 and 1961.

b) The results of the Danish study report a link between high socioeconomic status, education, and wine drinking. Since people with high levels of education and higher socioeconomic status are also more likely to be healthy, the relation between health and wine consumption might be explained by the confounding variables of socioeconomic status and education.

c) Studies such as these prove none of these. While the variables have a relation, there is no indication of a cause-and-effect relationship. The only way to determine causation is through a controlled, randomized, and replicated experiment.

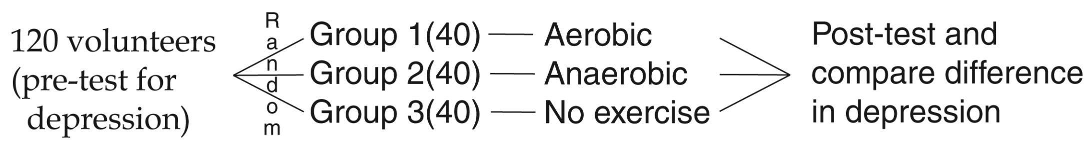
50. Swimming.

a) The swimmers showed a rate of depression that was lower than would be expected from a sample of that size drawn at random from the population. This rate was so low that it was unlikely to be due to natural sampling variation.

b) This is a retrospective observational study. There was no imposition of treatments. The researchers simply identified a group and evaluated them for depression.

c) The news reports made a claim of a cause-and-effect relationship. Causation can only be determined through the use of a controlled, randomized, and replicated experiment, not an observational study. The difference in depression rates might be explained by lurking variables. For example, swimmers might tend to have higher incomes than the general population. Swimmers need to have access to a pool, either by having their own, or paying for a membership to a health club. Perhaps it is their financial situation that makes them happier, not the swimming. Another possible explanation is a reversal of the direction of the relationship implied by the news reports. Perhaps depression makes people not want to swim.

d) Answers may vary. Give the subjects a test to measure depression. Then randomly assign the 120 subjects to one of three groups: the control group (no exercise program), the anaerobic exercise group, and the aerobic exercise group. Monitor subjects' exercise (have them report to a particular gym or pool). At the end of 12 weeks, administer the depression test again. Compare the post-exercise and pre-exercise depression scores.



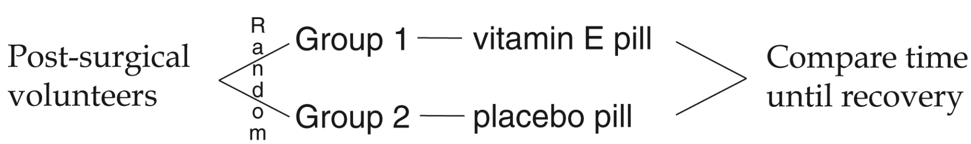
51. Dowsing.

a) Arrange the 20 containers in 20 separate locations. Number the containers 01 – 20, and use a random number generator to identify the 10 containers that should be filled with water.

b) We would expect the dowser to be correct about 50% of the time, just by guessing. A record of 60% (12 out of 20) does not appear to be significantly different than the 10 out of 20 expected.

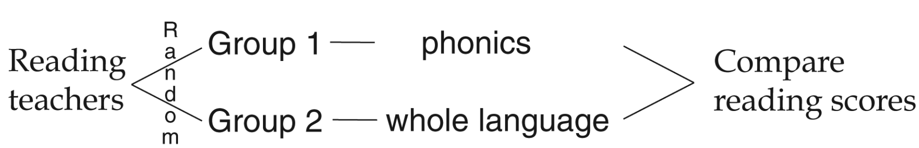
c) Answers may vary. A high level of success would need to be observed. 90% to 100% success (18 to 20 correct identifications) would be convincing.

52. Healing.



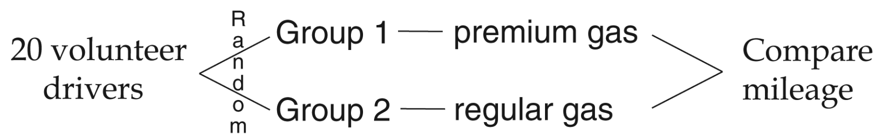
Answers may vary. This double-blind experiment has 1 factor (vitamin E), at 2 levels (vitamin E and no vitamin E), resulting in 2 treatments. The response variable measured is the time it takes the patient to recover from the surgery. Randomly select half of the patients who agree to the study to get large doses of vitamin E after surgery. Give the other patients in the study a similar looking placebo pill. Monitor their progress, recording the time until they have reached an easily agreed upon level of healing. Have the evaluating doctor blinded to whether the patient received the vitamin E or the placebo. Compare the number of days until recovery of the two groups.

53. Reading.



Answers may vary. This experiment has 1 factor (reading program), at 2 levels (phonics and whole language), resulting in 2 treatments. The response variable is reading score on an appropriate reading test after a year in the program. After randomly assigning students to teachers, randomly assign half the reading teachers in the district to use each method. There may be variation in reading score based on school within the district, as well as by grade. Blocking by both school and grade will reduce this variation.

54. Gas mileage.



Answers may vary. This experiment has 1 factor (type of gasoline), at 2 levels (premium and regular), resulting in two treatments. The response variable is gas mileage. An experiment diagram for a simple design appears above. Randomly assign each of the 20 volunteers to the premium or regular groups. Ask them to keep driving logs (the number of miles driven and the gallons of gasoline) for one month. Compare the differences in the fuel economy for the two groups.

Stronger designs would control for several variables that may have an effect on fuel economy, such as size of engine, type of driving (for example, city or highway), and driving style (for example, if the is driver aggressive, or if the driver exceeds the speed limit). With only 20 volunteers, it would be difficult to block for all of these variables, but a matched design would work well. Have each volunteer use regular gasoline for a specified time period and record the mileage, and also use premium for a specified time period. Randomize which type of gasoline is used first.

55. Weekend deaths.

a) The difference between death rate on the weekend and death rate during the week is greater than would be expected due to natural sampling variation.

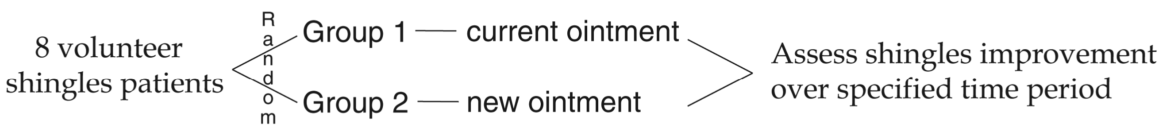
b) This was a prospective observational study. The researchers identified hospitals in Ontario, Canada, and tracked admissions to the emergency rooms. This certainly cannot be an experiment. People can't be assigned to become injured on a specific day of the week!

c) Waiting until Monday, if you were ill on Saturday, would be foolish. There are likely to be confounding variables that account for the higher death rate on the weekends. For example, people might be more likely to engage in risky behavior on the weekend.

d) Alcohol use might have something to do with the higher death rate on the weekends. Perhaps more people drink alcohol on weekends, which may lead to more traffic accidents, and higher rates of violence during these days of the week.

56. Shingles.

a) Answers may vary. This experiment has 1 factor (ointment), at 2 levels (current and new), resulting in 2 treatments. The response variables are the improvements in severity of the case of shingles and the improvements in the pain levels of the patients. Randomly assign the eight patients to either the current ointment or to the new ointment. Before beginning treatment, have doctors assess the severity of the case of shingles for each patient, and ask patients to rate their pain levels. Administer the ointments for a prescribed time, and then have doctors reassess the severity of the case of shingles, and ask patients to once again rate their pain levels. If the neither the patients nor the doctors are told which treatment is given to each patient, the experiment will be double-blind. Compare the improvement levels for each group.



b) Answers may vary. Let numbers 1 through 8 correspond to letter A through H, respectively. Ignore digits 0 and 9, and ignore repeats. The first row contains the random digits, the second row shows the corresponding patient (X indicates an ignored or repeated digit), and the third row shows the resulting group assignment, alternating between Group 1 and Group 2.

```
41098 18329 78458 31685 55259
DAXXH XXCBX GXXEX XXF
11  1   12  2  2    2
```

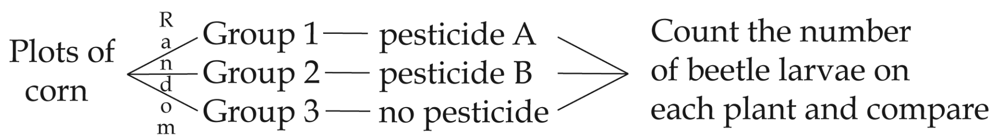
Group 1 (current ointment): D, A, H, C
Group 2 (new ointment): B, G, E, F

c) Assuming that the ointments looked alike, it would be possible to blind the experiment for the patient and the evaluating doctor. If both the subject and the evaluator are blinded, the experiment is double-blind.

d) Before randomly assigning patients to treatments, identify them as male or female. Having blocks for males and females will eliminate variation in improvement due to gender.

57. Beetles.

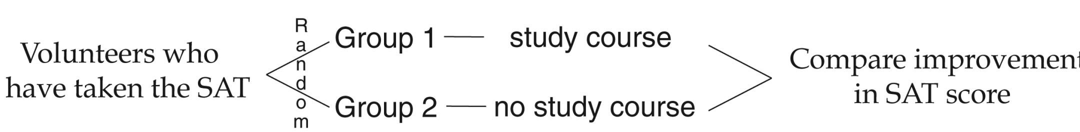
Answers may vary. This experiment has 1 factor (pesticide), at 3 levels (pesticide A, pesticide B, no pesticide), resulting in 3 treatments. The response variable is the number of beetle larvae found on each plant. Randomly select a third of the plots to be sprayed with pesticide A, a third with pesticide B, and a third to be sprayed with no pesticide (since the researcher also wants to know whether the pesticides even work at all). To control the experiment, the plots of land should be as similar as possible, with regard to amount of sunlight, water, proximity to other plants, etc. If not, plots with similar characteristics should be blocked together. If possible, use some inert substance as a placebo pesticide on the control group, and do not tell the counters of the beetle larvae which plants have been treated with pesticides. After a given period of time, count the number of beetle larvae on each plant and compare the results.



58. SAT prep.

a) The students were not randomly assigned to the special study course. Those who signed up for the course may be a special group whose scores would have improved anyway, due to motivation, intelligence, parental involvement, or other reasons.

b) Answers may vary. This experiment has 1 factor (study course), at 2 levels (study course, no study course), resulting in 2 treatments. The response variable is improvement in SAT score on the second test. Find a group of volunteers who are willing to participate. Have all volunteers take the SAT exam. Randomly assign the subjects to the study course or no study course groups. After giving the study course to the appropriate group, have both groups take the SAT again. Check to see if the group given the study course had a significant improvement in scores when compared with the group receiving no study course.



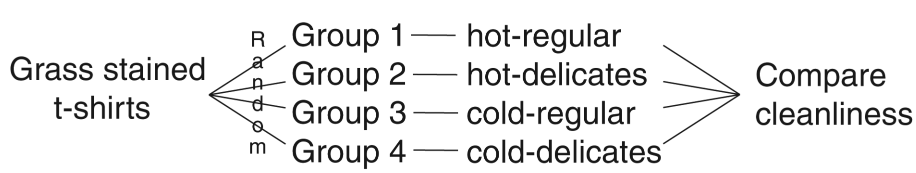
c) After the volunteers have taken the first SAT exam, block the volunteers by Low, Average, and High SAT exam score performance. For each block, replicate the experiment design described in part b.

59. Safety switch.

Answers may vary. This experiment has 1 factor (hand), at 2 levels (right, left), resulting in 2 treatments. The response variable is the difference in deactivation time between left and right hand. Find a group of volunteers. Using a matched design, we will require each volunteer to deactivate the machine with his or her left hand, as well as with his or her right hand. Randomly assign the left or right hand to be used first. Hopefully, this will equalize any variability in time that may result from experience gained after deactivating the machine the first time. Complete the first attempt for the whole group. Now repeat the experiment with the alternate hand. Check the differences in time for the left and right hands. Since the response variable is difference in times for each hand, workers should be blocked into groups based on their dominant hand. Another way to account for this difference would be to use the absolute value of the difference as the response variable. We are interested in whether or not the difference is significantly different from the zero difference we would expect if the machine were just as easy to operate with either hand.

60. Washing clothes.

Answers may vary. This experiment has two factors (water temperature, wash cycle). The factor water temperature has 2 levels (cold, hot), and the factor wash cycle has 2 levels (regular, delicates). 2 factors, at 2 levels each, results in 4 treatments (hot-regular, hot-delicates, cold-regular, cold-delicates). The response variable is the level of cleaning of the grass stains. It would be nice to have 32 shirts with which to experiment, so that we could randomly assign 8 shirts to each treatment group, but equal numbers of shirts in each group are not necessary. After washing, have “laundry experts” rate the cleanliness of each shirt. Compare the level of cleanliness in each group.



61. Skydiving, anyone?

a) There is 1 factor, jumping, with 2 levels, with and without a working parachute.

b) You would need some (dim-witted) volunteers skydivers as the subjects.

c) A parachute that looked real, but didn’t open, would serve as the placebo.

d) 1 factor at 2 levels is 2 treatments, a good parachute and a placebo parachute.

e) The response variable is whether the skydiver survives the jump (or the extent of injuries).

f) All skydivers should jump from the same altitude, in similar weather conditions, and land on similar surfaces.

g) Make sure that you randomly assign the skydivers to the parachutes.

h) The skydivers (and the distributers of the parachutes) shouldn't know who got a working chute. Additionally, the people evaluating the subjects after the jumps should not be told who had a real chute, either.

Review of Part III – Gathering Data

1. The researchers performed a prospective observational study, since the children were identified at birth and examined at ages 8 and 20. There were indications of behavioral differences between the group of "preemies", and the group of full-term babies. The "preemies" were less likely to engage in risky behaviors, like use of drugs and alcohol, teen pregnancy, and conviction of crimes. This may point to a link between premature birth and behavior, but there may be lurking variables involved. Without a controlled, randomized, and replicated experiment, a cause-and-effect relationship cannot be determined.

2. A retrospective observational study was performed. There may be a link between tea drinking and survival after a heart attack. Other variables, like overall health and education might also be involved. Since lurking variables may be involved, a controlled, randomized, and replicated experiment must be performed to determine whether or not a cause-and-effect relationship exists between tea drinking and post heart attack survival.

3. The researchers at the Purina Pet Institute performed an experiment, matched by gender and weight. The experiment had one factor (diet), at two levels (allowing the dogs to eat as much as they want, or restricted diet), resulting in two treatments. One of each pair of similar puppies was randomly assigned to each treatment. The response variable was length of life. The researchers were able to conclude that, on average, dogs with a lower-calorie diet live longer.

4. The officials used a random sample. The population is all homes on the property tax list. The parameter of interest is level of radon contamination. The officials' procedure is not clear, but if they make an effort to get some houses from each area in the city, the sample is stratified by area. If the procedure is followed carefully, the officials can use the results of the sample to make inferences about the radon levels in other houses in the county.

5. This is a completely randomized experiment, with the treatment being receiving folic acid or not (one factor, two levels). Treatments were assigned randomly and the response variable is the number of precancerous growths, or simply the occurrence of additional precancerous growths. Neither blocking nor matching is mentioned, but in a study such as this one, it is likely that researchers and patients are blinded. Since treatments were randomized, it seems reasonable to generalize results to all people with precancerous polyps, though caution is warranted since these results contradict a previous study

6. The research team performed a retrospective observational study. There is evidence that the date of first flowering has generally advanced over the last 47 years, but there may be other variables besides climate change that can account for this. The assertion of the researchers is speculative.

7. The fireworks manufacturers are sampling. No information is given about the sampling procedure, so hopefully the tested fireworks are selected randomly. It would probably be a good idea to test a few of each type of firework, so stratification by type seems likely. The population is all fireworks produced each day, and the parameter of interest is the proportion of duds. With a random sample, the manufacturers can make inferences about the proportion of duds in the entire day's production, and use this information to decide whether or not the day's production is suitable for sale.

8. The researchers used a completely randomized experiment, with three treatments that each subject received, so this could be considered matching. The treatments are receiving the spoiler, incorporating the spoiler into the first paragraph, and no spoiler (control treatment). The response variable is the rating that participants gave each story. Blinding is not mentioned. Results should be generalizable to the public, though since only college students were used, caution should be exercised.

9. This is an observational retrospective study. Researcher can conclude that for anyone's lunch, even when packed with ice, food temperatures are rising to unsafe levels.

10. The medical researchers performed a retrospective observational study. The data were gathered from pre-existing medical records. The study does not *prove* that there is no long-term risk of prostate cancer associated with having a vasectomy, but it does provide evidence to that effect.

11. This is an experiment, with a control group being the genetically engineered mice who received no antidepressant and the treatment group being the mice who received the drug. The response variable is the amount of plaque in their brains after one dose and after four months. There is no mention of blinding or matching. Conclusions can be drawn to the general population of mice and we should assume treatments were randomized. To conclude the same for humans would be risky, but researchers might propose an experiment on humans based on this study.

12. The artisan is performing an experiment. There are 2 factors (glaze type and temperature). The glaze type has 4 levels, and the temperature has 3 levels, resulting in 12 treatments (the different combinations of glazes and temperatures). There is no mention of randomization. The response variable is apparent age of the pottery. Assuming that the evaluator is unbiased, the artisan can make a conclusion about the best combination of glaze and temperature.

13. The researchers performed an experiment. There is one factor (gene therapy), at two levels (gene therapy and no gene therapy), resulting in two treatments. The experiment is completely randomized. The response variable is heart muscle condition. The researchers can conclude that gene therapy is responsible for stabilizing heart muscle in laboratory rats.

14. Observational prospective study. Since the study was based on a large number of adults from eight states over a long period of time, it is reasonable to conclude that smoking and bladder cancer are associated.

15. The orange juice plant depends on sampling to ensure the oranges are suitable for juice. The population is all of the oranges on the truck, and the parameter of interest is the proportion of unsuitable oranges. The procedure used is a random sample, stratified by location in the truck. Using this well-chosen sample, the workers at the plant can estimate the proportion of unsuitable oranges on the truck, and decide whether or not to accept the load.

16. The soft drink manufacture is sampling in order to determine whether or not the machine that caps the bottles is working properly. The population is all of the bottle cap seals. The parameter of interest is the whether or not the bottles are sealing properly. They are using a systematic sample, checking bottles at fixed intervals. If any bottles in the sample are not sealed properly, they can tell that the machine may need adjustment or repair.

17. Observational retrospective study, performed as a telephone-based randomized survey. Based on the excerpt, it seems reasonable to conclude that more education is associated with a higher Emotional Health Index score, but to insist on causality would be faulty reasoning.

18. This statistics professor is performing an experiment, blocked by whether or not the students have taken calculus. However, there is probably no randomization, since students usually select their own courses. Hopefully, the two sections contain similar groups of students. There is one factor (use of software), at two levels (software and no software), resulting in two treatments. The response variable is the final exam score. The experiment incorporates blinding, since the graders do not know which students used software and which did not. The professor can decide if computer software is beneficial, and if so, determine whether or not calculus students perform differently than those who have not had calculus.

19. Point spread.

Answers may vary. Perform a simulation to determine the gambler's expected winnings. A component is one game. To model that component, generate random digits 0 to 9. Since the outcome after the point spread is a tossup, let digits 0-4 represent a loss, and let digits 5-9 represent a win. A run consists of 5 games, so generate 5 random digits at a time. The response variable is the profit the gambler makes, after accounting for the \$10 bet. If the outcome of the run is 0, 1, or 2 simulated wins, the profit is −\$10. If the outcome is 3, 4, or 5 simulated wins, the profit is \$0, \$10, or \$40, respectively. The total profit divided by the number of runs is the average weekly profit. According to the simulation (80 runs were performed), the gambler is expected to break even. His simulated losses equaled his simulated winnings. (In theory, the gambler is expected to lose about \$2.19 per game.)

20. The lottery.

a) Answers may vary. Perform a simulation to determine the number of plays required to win. Pick 3 numbers from 1 to 20. (These don't need to be randomly generated. Players of the lottery aren't required to pick randomly, so there is no reason we should!) Let's use 1, 2, and 3 to keep it simple. A component is the selection of 1 winning number. Simulate the winning number by generating a random pair of digits from 01 to 20. Depending on the type of lottery you simulate, repeated numbers may have to be ignored. Some lotteries choose the numbers from 5 different sets of numbers, while others choose 5 numbers from a single set of numbers. A run consists of 5 winning numbers, so generate 5 such pairs per run. The response variable is whether or not the numbers 1, 2, and 3 appear in the run. Simply count the number of runs it takes to simulate a win. Just for fun, I performed a couple hundred runs of this simulation, and never got a match for my 3 numbers. You are very unlikely to get a match.

b) With more numbers from which to choose, and more matches required to win, the odds of winning go down dramatically. Winning a state lottery is highly improbable.

21. Everyday randomness.

Answers will vary. Most of the time, events described as "random" are anything but truly random.

22. Cell phone risks.

a) This is an experiment, since treatments were imposed on randomly assigned groups. There is one factor (radio waves), at three levels (digital cell phone radio waves, analog cell phone radio waves, and no radio waves), resulting in three treatments. The response variable is the incidence of brain tumors.

b) The differences in the incidence of tumors between the groups of rats were not great enough to convince the researchers that the differences were due to anything other than sampling variability.

c) Since the research was funded by Motorola, there may have been bias present. The researchers may have felt pressure to declare cell phones safe, and introduced bias, intended or not, into the experiment.

23. Tips.

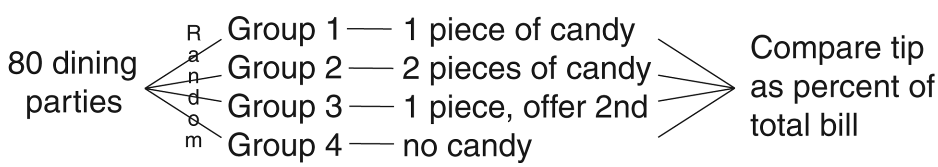
a) The waiters performed an experiment, since treatments were imposed on randomly assigned groups. This experiment has one factor (candy), at two levels (candy or no candy), resulting in two treatments. The response variable is the percentage of the bill given as a tip.

b) If the decision whether to give candy or not was made before the people were served, the server may have subconsciously introduced bias by treating the customers better. If the decision was made just before the check was delivered, then it is reasonable to conclude that the candy was the cause of the increase in the percentage of the bill given as a tip.

c) "Statistically significant" means that the difference in the percentage of tips between the candy and no candy groups was more than expected due to sampling variability.

24. Tips, take 2.

a) A diagram of the tipping experiment appears below.



b) This experiment has 1 factor (candy), at 4 levels (1 piece, 2 pieces, 1 piece with an additional piece offered, and no candy).

c) 1 factor at 4 levels results in 4 treatments.

d) The response variable is the percent of the total bill left as a tip.

e) The diners were not aware that they were part of an experiment, so the experiment was blinded. This experiment did not use double-blinding, but there is probably no way to double blind this experiment, since there is no need to blind the evaluator of the response variable. Biased evaluation of the amount of tip left doesn't seem possible.

f) If the waitress knew which tables were going to receive certain treatments, then she might have treated some tables better than others. The waitress should be unaware of the treatment until after the meal, to avoid the introduction of bias. (Note: This is still only single-blinding. Blinding does not refer to the number of people blinded, but rather the type of blinding employed. If the diners and the waitress are unaware of the assignment of treatments, it is for the same purpose, namely to keep the diners from being systematically influenced.)

25. Timing.

There will be voluntary response bias, and results will mimic those only of the visitors to Sodahead.com and not the general U.S. population. The question is leading responders to answer "yes" though many might understand that the president's timing for his vacation had nothing to do with the events of the week.

26. Laundry.

a) Answers may vary. Water (quality and temperature) and material can vary. These confounding variables may influence results. The treatments in the experiment must be in environments that are identical, with the exception of the factor being studied.

b) These conditions are unrealistic. This will not help the experimenters determine how well *SparkleKleen* will work in the conditions for which it was intended.

c) If the swatches are stained at the same time, the stains on the swatches washed later will have more time to "set in", causing bias towards *SparkleKleen*. Also, unforeseen variables, like changes in water temperature or pressure won't be equalized through randomization.

d) The conditions under which the *SparkleKleen* was tested are unknown. There is no way to keep the conditions comparable. Furthermore, the company that produced *SparkleKleen* may not produce reliable data. They have a vested interest in the success of their product.

27. When to stop?

a) Answers may vary. A component in this simulation is rolling 1 die. To simulate this component, generate a random digit 1 to 6. To simulate a run, simulate 4 rolls, stopping if a 6 is rolled. The response variable is the sum of the 4 rolls, or 0 if a 6 is rolled. The average number of points scored is the sum of all rolls divided by the total number of runs. According to the simulation, the average number of points scored will be about 5.8.

b) Answers may vary. A component in this simulation is rolling 1 die. To simulate this component, generate a random digit 1 to 6. To simulate a run, generate random digits until the sum of the digits is at least 12, or until a 6 is rolled. The response variable is the sum of the digits, or 0 if a 6 is rolled. The average number of points scored is the sum of all rolls divided by the total number of runs. According to the simulation, the average number of points scored will be about 5.8, similar to the outcome of the method described in part a).

c) Answers may vary. Be careful when making your decision about the effectiveness of your strategy. If you develop a strategy with a higher simulated average number of points than the other two methods, this is only an indication that you may win in the long run. If the game is played round by round, with the winner of a particular round being declared as the player with the highest roll made during that round, the game is much more variable. For example, if Player B rolls a 12 in a particular game, Player A will always lose that game, provided he or she sticks to the strategy. A better way to get a feel for your chances of winning this type of game might be to simulate several rounds, recording whether each player won or lost the round. Then estimate the percentage of the time that each player is expected to win, according to the simulation.

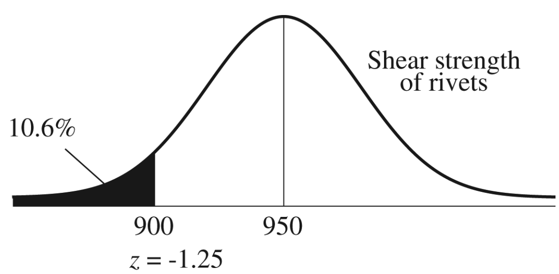
28. Rivets

a)

$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{900 - 950}{40}$$

$$z = -1.25$$



According to the Normal model, approximately 10.6% of rivets are expected to break when tested under a 900-pound load.

b) Answers may vary. The component being simulated is whether or not a rivet will break under a 900-pound load. To model this component, generate triples of random digits, from 000 to 999. Let digits 001 to 106 represent a broken rivet under a 900-pound load. (001 to 106 represents the 10.6% probability of breaking.) Let other triples of digits represent unbroken rivets under a 900-pound load. To simulate a run, generate a triple of random digits. The response variable is whether or not the rivet will break under a 900-pound load. Count the number of simulated rivets until 3 broken rivets are simulated. According to the simulation, you might need to test about 28 rivets before finding 3 that fail at 900-pounds or below.

29. Homecoming.

a) Since telephone numbers were generated randomly, every number that could possibly occur in that community had an equal chance of being selected. This method is "better" than using the phone book, because unlisted numbers are also possible. Those community members who deliberately do not list their phone numbers might not consider this method "better"!

b) Although this method results in a simple random sample of phone numbers, it does not result in a simple random sample of residences. Residences without a phone are excluded, and residences with more than one phone have a greater chance of being included.

c) No, this is not a SRS of local voters. People who respond to the survey may be of the desired age, but not registered to vote. Additionally, some voters who are contacted may choose not to participate.

d) This method does not guarantee an unbiased sample of households. Households in which someone answered the phone may be more likely to have someone at home when the phone call was generated. The attitude about homecoming of these households might not be the same as the attitudes of the community at large.

30. Youthful appearance.

a) The differences is guessed age were greater than differences that could be explained by natural sampling variability.

b) Dr. Weeks is implying that having sex caused the people to have a more youthful appearance. It seems more plausible that younger-looking people are more sexually active than older-looking people, because of their age.

31. Smoking and Alzheimer's.

a) The studies do not prove that smoking offers any protection from Alzheimer's. The studies merely indicate an association. There may be other variables that can account for this association.

b) Alzheimer's usually shows up late in life. Since smoking is known to be harmful, perhaps smokers have died of other causes before Alzheimer's can be seen.

c) The only way to establish a cause-and-effect relationship between smoking and Alzheimer's is to perform a controlled, randomized, and replicated experiment. This is unlikely to ever happen, since the factor being studied, smoking, has already been proven harmful. It would be unethical to impose this treatment on people for the purposes of this experiment. A prospective observational study could be designed in which groups of smokers and nonsmokers are followed for many years and the incidence of Alzheimer's disease is tracked.

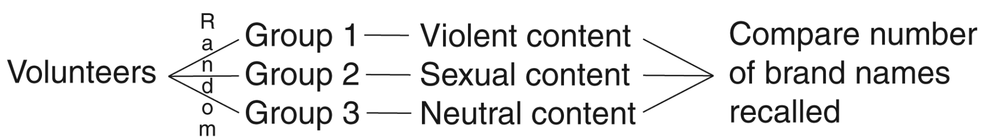
32. Antacids.

a) This is a randomized experiment, blocked by gender.

b) Experiments always use volunteers. This is not a problem, since experiments are testing response to a treatment, not attempting to determine an unknown population parameter. The randomization in an experiment is random assignment to treatment groups, not random selection from a population.

c) Since the experiment is studying the effects of an antacid, the placebo may actually confound the experiment, since the introduction of <u>*any*</u> substance, even a sugar pill, into the digestive system may have an effect on acid reflux. (The use of some sort of placebo is always recommended, but in some cases it may be difficult to find a placebo that truly has no effect, beyond the expected "placebo effect", of course!)

33. Sex and violence.



This experiment has one factor (program content), at three levels (violent, sexual, and neutral), resulting in three treatments. The response variable is the number of brand names recalled after watching the program. Numerous subjects will be randomly assigned to see shows with violent, sexual, or neutral content. They will see the same commercials. After the show, they will be interviewed for their recall of brand names in the commercials.

34. Pubs.

a) *Who* – 900 Englishmen. *What* – The researcher was interested in their reasons for going to the pub. *When* – not stated. *Where* – England. *Why* – The producers of Kaliber alcohol-free beer hoped to show that men went to the pub for reasons other than the alcohol.
How – Researchers surveyed men regarding their reasons for going to the pub.

b) The researcher surveyed believes that the population is all Englishmen.

c) The most important omitted detail is the selection process. How did the researcher acquire the sample of men? Was randomness used? Is the sample representative of the population of Englishmen?

d) Although not stated, it appears that the researcher simply took convenience samples of those in the pubs.

e) The results may be biased. First of all, an alcohol-free beer producer funded the survey. Respondents may have felt subconscious pressure to indicate that alcohol was not the primary reason for going to the pub. Additionally, admitting that you go to the pub merely for the alcohol is a potentially embarrassing admission. The percentage of pub patrons who go for the alcohol may be significantly higher than 10%.

35. Age and party 2008.

a) The number of respondents is roughly the same for each age category. This may indicate a sample stratified by age category, although it may be a simple random sample.

b) 1530 Democrats were surveyed. $\frac{1530}{4002} \approx 38.2\%$ of the people surveyed were Democrats.

c) We don't know. If data were collected from voting precincts that are primarily Democratic or primarily Republican, that would bias the results. Because the survey was commissioned by NBC News, we can assume the data collected are probably reliable.

d) The pollsters were probably attempting to determine whether or not political party is associated with age.

36. Bias?

a) Barone claims that nonresponse bias exists in polls, since conservatives are more likely to refuse to participate than other groups. Nonresponse is less of an issue if it is believed that all groups fail to respond at the same rate.

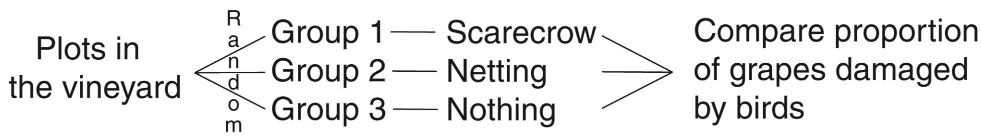
b) The population of interest is all adults in the United States.

c) The column totals do not add up to 100%. There is information missing, and the discrepancies are too large to be attributed to rounding.

d) The differences observed are similar to differences that may have been observed simply due to natural sampling variability. Differences of this size would be probable, even if no bias exists.

37. Save the grapes.

This experiment has one factor (bird control device), at three levels (scarecrow, netting, and no device), resulting in three treatments. Randomly assign different plots in the vineyard to the different treatments, making sure to ensure adequate separation of plots, so that the possible effect of the scarecrow will not be confounded with the other treatments. The response variable to be measured at the end of the season is the proportion of bird-damaged grapes in each plot.



38. Bats.

Answers may vary. This experiment has one factor (type of bat), at two levels (wooden and aluminum), resulting in two treatments. The response variable is the difference is distance the ball is hit with each type of bat. Since players vary in their ability to hit the ball, a matched design should be used, with each batter hitting with both types of bats several times in a randomly chosen order. Find the difference for each batter, and then calculate the average difference in distance between the aluminum and wooden bats.

39. Acupuncture.

a) The "fake" acupuncture was the control group. In an experiment, all subjects must be treated as alike as possible. If there were no "fake" acupuncture, subjects would know that they had not received acupuncture, and might react differently. Of course, all volunteers for the experiment must be aware of the possibility of being randomly assigned to the control group.

b) Experiments always use volunteers. This is not a problem, since experiments are testing response to a treatment, not attempting to determine an unknown population parameter. The randomization in an experiment is random assignment to treatment groups, not random selection from a population. Voluntary response is a problem when sampling, but is not an issue in experimentation. In this case, it is probably reasonable to assume that the volunteers have similar characteristics to others in the population of people with chronic lower back pain.

c) There were differences in the amount of pain relief experienced by the two groups, and these differences were large enough that they could not be explained by natural variation alone. Researchers concluded that both proper and "fake" acupuncture reduced back pain.

40. NBA draft lottery.

Answers will vary. A component in the simulation is drawing one lottery card. To simulate this component, generate random numbers 01 to 66.

Let numbers 01 to 11 represent the team with the worst record, numbers 12 to 21 represent the team with the second worst record, 22 to 30 represent the team with the third worst record (your team), numbers 31 to 38 represent the fourth worst team, numbers 39 to 45 represent the fifth worst team, numbers 46 to 51 representing the sixth worst team, 52 to 56 represent the seventh worst team, 57 to 60 represent the eighth team, 61 to 63 represent the ninth worst team, 64 and 65 represent the tenth worst team, and 66 represent the team with the best record of the teams not making the playoffs.

A run consists of the assignment of one draft pick position to each of the 11 teams. (You can stop after 2 assignments. We are only concerned with whether or not our team gets to pick first or second.) The response variable is whether or not your team receives first or second pick. To simulate a run, generate a random number. If that number is 22 to 30, stop and record a success. If that number is not 22 to 30, note which team received the first pick and generate another random number, ignoring any number that corresponds to the team receiving first pick. If the second number is 22 to 30, record a success. Otherwise record a failure.

The probability that your team gets first or second pick is the total number of successes divided by the number of runs. According to the simulation, your team should get first or second pick approximately 30% of the time.

41. Security.

a) To ensure that passengers from first-class, as well as coach, get searched, select 2 passengers from first-class and 12 from coach. Using this stratified random sample, 10% of the first-class passengers are searched, as are 10% of the coach passengers.

b) Answers will vary. Number the passengers alphabetically, with 2-digit numbers. Bergman = 01, Bowman = 02, and so on, ending with Testut = 20. Read the random digits in pairs, ignoring pairs 21 to 99 and 00, and ignoring repeated pairs.

```
65|43|67|11|              27|04|
XX XX XX Fontana          XX Castillo
```

The passengers selected for search from first-class are Fontana and Castillo.

c) Number the passengers alphabetically, with 3 digit numbers, 001 to 120. Use the random number table to generate 3-digit numbers, ignoring numbers 121 to 999 and 000, and ignoring repeated numbers. Search the passengers corresponding to the first 12 valid numbers generated.

42. Profiling?

Answers will vary. A component in this simulation is the selection of a passenger to be searched. To simulate this component, generate pairs of random digits 01 to 20. Let 01 to 04 represent the businessmen from the Middle East, and let numbers 05 to 20 represent other passengers. A run consists of the selection of 2 passengers for search. To simulate a run, generate 2 pairs of random digits, ignoring repeats. The response variable is whether of not both passengers selected are Middle Eastern businessmen. If both pairs generated are from 01 to 04, record a "success". Otherwise, record a "failure". The total number of successes divided by the total number of runs is the probability that both passengers selected are Middle Eastern businessmen. According to the simulation, this should happen about 3% of the time. Although relatively small, this percentage does not indicate an event that is extremely unlikely. We can't be certain that the Middle Eastern businessmen were not "profiled", but there is little evidence suggesting that they were.

43. Par 4.

Answers may vary. A component in this simulation is a shot. Use pairs of random digits 00 to 99 to represent a shot. The way in which this component is simulated depends on the type of shot.

For the first shot, let pairs of digits 01 to 70 represent hitting the fairway, and let pairs of digits 71 to 99, and 00, represent not hitting the fairway.

If the first simulated shot hits the fairway, let 01 to 80 represent landing on the green on the second shot, and let 81 to 99, and 00, represent not landing on the green on the second shot. If the first simulated shot does not hit the fairway, let 01 to 40 represent landing on the green on the second shot, and let 41 to 99, and 00, represent not landing on the green on the second shot.

If the second simulated shot does not land on the green, let 01 to 90 represent landing on the green, and 91 to 99, and 00, represent not landing on the green. Keep simulating shots until the shot lands on the green.

Once on the green, let 01 to 20 represent sinking the putt on the first putt, and let 21 to 99, and 00, represent not sinking the putt on the first putt. If second putts are required, continue simulating putts until a putt goes in, with 01 to 90 representing making the putt, and 91 to 99, and 00, representing not making it.

A run consists of following the guidelines above until the final putt is made. The response variable is the number of shots required until the final putt is made.

The simulated average score on the hole is the total number of shots required divided by the total number of runs. According to 40 runs of this simulation, a pretty good golfer can be expected to average about 4.2 strokes per hole. Your simulation results may vary.

44. The back nine.

a) Answers may vary. A component in this simulation is a shot. Use pairs of random digits 00 to 99 to represent a shot. The way in which this component is simulated depends on the type of shot.

For the first shot, let pairs of digits 01 to 80 represent hitting the fairway, and let pairs of digits 81 to 99, and 00, represent not hitting the fairway.

If the first simulated shot hits the fairway, let 01 to 80 represent landing on the green on the second shot, and let 81 to 99, and 00, represent not landing on the green on the second shot. If the first simulated shot does not hit the fairway, let 01 to 40 represent landing on the green on the second shot, and let 41 to 99, and 00, represent not landing on the green on the second shot.

If the second simulated shot does not land on the green, let 01 to 90 represent landing on the green, and 91 to 99, and 00, represent not landing on the green. Keep simulating shots until the shot lands on the green.

Once on the green, let 01 to 20 represent sinking the putt on the first putt, and let 21 to 99, and 00, represent not sinking the putt on the first putt. If second putts are required, continues simulating putts until a putt goes in, with 01 to 90 representing making the putt, and 91 to 99, and 00, representing not making the putt.

A run consists of following the guidelines above until the final putt is made. The response variable is the number of shots required until the final putt is made.

The simulated average score on the hole is the total number of shots required divided by the total number of runs. According to 20 runs of this simulation, a pretty good golfer can be expected to average about 3.7 strokes per hole. Your simulation results may vary.

b) Answers may vary. The simulation is set up identically to part a), with the exception of the second shot. Now, let 01 to 10 represent hitting the green, and let 11 to 99, and 00, represent not hitting the green.

According to 20 runs of this simulation, a pretty good golfer can be expected to average about 5.3 strokes per hole. Your simulation results may vary.

c) Answers may vary.

Chapter 13 – From Randomness to Probability

Section 13.1

1. Flipping a coin.

In the long run, a fair coin will generate 50% heads and 50% tails, approximately. But for each flip we cannot predict the outcome.

2. Dice.

In the long run, a fair die will produce roughly equal amounts of the numbers 1 through 6 when rolled. But each roll is unpredictable.

3. Flipping a coin II.

There is no law of averages for the short run. The first five flips do not affect the sixth flip.

4. Dice II.

The dice have no memory and in the short run there is no guarantee about what will happen next.

Section 13.2

5. Wardrobe.

a) There are a total of 10 shirts, and 3 of them are red. The probability of randomly selecting a red shirt is 3/10 = 0.30.

b) There are a total of 10 shirts, and 8 of them are not black. The probability of randomly selecting a shirt that is not black is 8/10 = 0.80.

6. Playlists.

a) There are a total of 20 songs, and 7 of them are rap songs. The probability of randomly selecting a rap song is 7/20 = 0.35.

b) There are a total of 20 songs, and 17 of them are not country songs. The probability of randomly selecting non-country song is 17/20 = 0.85.

Section 13.3

7. Cell phones and surveys.

a) If 25% of homes don't have a landline, then 75% of them do have a landline. The probability that all 5 houses have a landline is $(0.75)^5 \approx 0.237$.

b) $P(\text{at least one without landline}) = 1 - P(\text{all landlines}) = 1 - (0.75)^5 \approx 0.763$.

c) $P(\text{at least one with landline}) = 1 - P(\text{no landlines}) = 1 - (0.25)^5 \approx 0.999$

8. Cell phones and surveys II.

a) The probability that all 4 adults have only a cell phone is $(0.49)^4 \approx 0.0576$.

b) If 49% have only a cell phone and no landline, then 51% don't have this combination of phones. $P(\text{no one with only a cell phone}) = (0.51)^4 \approx 0.0677$.

c)

$$\begin{aligned} P(\text{at least one with only cell phone}) &= 1 - P(\text{cellphone and/or landline}) \\ &= 1-(0.51)^4 \approx 0.9323 \end{aligned}$$

Chapter Exercises.

9. Sample spaces.

a) S = { HH, HT, TH, TT} All of the outcomes are equally likely to occur.

b) S = { 0, 1, 2, 3} All outcomes are not equally likely. A family of 3 is more likely to have, for example, 2 boys than 3 boys. There are three equally likely outcomes that result in 2 boys (BBG, BGB, and GBB), and only one that results in 3 boys (BBB).

c) S = { H, TH, TTH, TTT} All outcomes are not equally likely. For example the probability of getting heads on the first try is $\frac{1}{2}$. The probability of getting three tails is $\left(\frac{1}{2}\right)^3 = \frac{1}{8}$.

d) S = {1, 2, 3, 4, 5, 6} All outcomes are not equally likely. Since you are recording only the larger number (or the number if there is a tie) of two dice, 6 will be the larger when the other die reads 1, 2, 3, 4, or 5. The outcome 2 will only occur when the other die shows 1 or 2.

10. Sample spaces.

a) S = { 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} All outcomes are not equally likely. For example, there are four equally likely outcomes that result in a sum of 5 (1 + 4, 4 + 1, 2 + 3, and 3 + 2), and only one outcome that results in a sum of 2 (1 + 1).

b) S = {BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG} All outcomes are equally likely.

c) S = { 0, 1, 2, 3, 4} All outcomes are not equally likely. For example, there are 4 equally likely outcomes that produce 1 tail (HHHT, HHTH, HTHH, and THHH), but only one outcome that produces 4 tails (TTTT).

d) S = { 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10} All outcomes are not equally likely. A string of 3 heads is much more likely to occur than a string of 10 heads in a row.

11. Roulette.

If a roulette wheel is to be considered truly random, then each outcome is equally likely to occur, and knowing one outcome will not affect the probability of the next. Additionally, there is an implication that the outcome is not determined through the use of an electronic random number generator.

12. Rain.

When a weather forecaster makes a prediction such as a 25% chance of rain, this means that when weather conditions are like they are now, rain happens 25% of the time in the long run.

13. Winter.

Although acknowledging that there is no law of averages, Knox attempts to use the law of averages to predict the severity of the winter. Some winters are harsh and some are mild over the long run, and knowledge of this can help us to develop a long-term probability of having a harsh winter. However, probability does not compensate for odd occurrences in the short term. Suppose that the probability of having a harsh winter is 30%. Even if there are several mild winters in a row, the probability of having a harsh winter is still 30%.

14. Snow.

The radio announcer is referring to the "law of averages", which is not true. Probability does not compensate for deviations from the expected outcome in the recent past. The weather is not more likely to be bad later on in the winter because of a few sunny days in autumn. The weather makes no conscious effort to even things out, which is what the announcer's statement implies.

15. Cold streak.

There is no such thing as being "due to make a shot". This statement is based on the so-called law of averages, which is a mistaken belief that probability will compensate in the short term for odd occurrences in the past. The player's chance of making a shot does not change based on recent successes or failures.

16. Crash.

a) There is no such thing as the "law of averages". The overall probability of an airplane crash does not change due to recent crashes.

b) Again, there is no such thing as the "law of averages". The overall probability of an airplane crash does not change due to a period in which there were no crashes. It makes no sense to say a crash is "due". If you say this, you are expecting probability to compensate for strange events in the past.

17. Auto insurance.

a) It would be foolish to insure your neighbor against automobile accidents for \$1500. Although you might simply collect \$1500, there is a good chance you could end up paying much more than \$1500. That risk is not worth the \$1500.

b) The insurance company insures many people. The overwhelming majority of customers pay the insurance and never have a claim, or have claims that are lower than the cost of their payments. The few customers who do have a claim are offset by the many who simply send their premiums without a claim. The relative risk to the insurance company is low.

18. Jackpot.

a) The Excalibur can afford to give away millions of dollars on a \$3 bet because almost all of the people who bet do not win the jackpot.

b) The press release generates publicity, which entices more people to come and gamble. Of course, the casino wants people to play, because the overall odds are always in favor of the casino. The more people who gamble, the more the casino makes in the long run. Even if that particular slot machine has paid out more than it ever took in, the publicity it gives to the casino more than makes up for it. If the casino is successful, then they will buy more slot machines from the slot machine maker.

19. Spinner.

a) This is a legitimate probability assignment. Each outcome has probability between 0 and 1, inclusive, and the sum of the probabilities is 1.

b) This is a legitimate probability assignment. Each outcome has probability between 0 and 1, inclusive, and the sum of the probabilities is 1.

c) This is not a legitimate probability assignment. Each outcome has probability between 0 and 1, inclusive, but the sum of the probabilities is greater than 1.

d) This is a legitimate probability assignment. Each outcome has probability between 0 and 1, inclusive, and the sum of the probabilities is 1. However, this game is not very exciting!

e) This probability assignment is not legitimate. The sum of the probabilities is 0, and there is one probability, -1.5 , that is not between 0 and 1, inclusive.

20. Scratch off.

a) This is not a legitimate assignment. Although each outcome has probability between 0 and 1, inclusive, the sum of the probabilities is less than 1.

b) This is not a legitimate probability assignment. Although each outcome has probability between 0 and 1, inclusive, the sum of the probabilities is greater than 1.

c) This is a legitimate probability assignment. Each outcome has probability between 0 and 1, inclusive, and the sum of the probabilities is 1.

d) This probability assignment is not legitimate. Although the sum of the probabilities is 1, there is one probability, -0.25 , that is not between 0 and 1, inclusive.

e) This is a legitimate probability assignment. Each outcome has probability between 0 and 1, inclusive, and the sum of the probabilities is 1. This is also known as a 10% off sale!

21. Electronics.

A family may have both a computer and an HDTV. The events are not disjoint, so the Addition Rule does not apply.

22. Homes.

A home may have both a garage and a pool. The events are not disjoint, so the Addition Rule does not apply.

23. Speeders.

When cars are traveling close together, their speeds are not independent. For example, a car following directly behind another can't be going faster than the car ahead. Since the speeds are not independent, the Multiplication Rule does not apply.

24. Lefties.

There may be a genetic factor making handedness of siblings not independent. The Multiplication Rule does not apply.

25. College admissions.

a) Jorge had multiplied the probabilities.

b) Jorge assumes that being accepted to the colleges are independent events.

c) No. Colleges use similar criteria for acceptance, so the decisions are not independent. Students that meet these criteria are more likely to be accepted at all of the colleges. Since the decisions are not independent, the probabilities cannot be multiplied together.

26. College admissions II.

a) Jorge has added the probabilities.

b) Jorge is assuming that getting accepted to the colleges are disjoint events.

c) No. Students can get accepted to more than one of the three colleges. The events are not disjoint, so the probabilities cannot simply be added together.

27. Car repairs.

Since all of the events listed are disjoint, the addition rule can be used.

a) P(no repairs) = 1 – P(some repairs) = 1 – (0.17 + 0.07 + 0.04) = 1 – (0.28) = 0.72

b) P(no more than one repair) = P(no repairs or one repair) = 0.72 + 0.17 = 0.89

c) P(some repairs) = P(one or two or three or more repairs)
= 0.17 + 0.07 + 0.04 = 0.28

28. Stats projects.

Since all of the events listed are disjoint, the addition rule can be used.

a) P(two or more semesters of Calculus) = 1 – (0.55 + 0.32) = 0.13

b) P(some Calculus) = P(one semester or two or more semesters) = 0.32 + 0.13 = 0.45

c) P(no more than one sem.) = P(no Calculus or one semester) = 0.55 + 0.32 = 0.87

29. More repairs.

Assuming that repairs on the two cars are independent from one another, the multiplication rule can be used. Use the probabilities of events from Exercise 27 in the calculations.

a) P(neither will need repair) = (0.72)(0.72) = 0.5184

b) P(both will need repair) = (0.28)(0.28) = 0.0784

c) P(at least one will need repair) = 1 - P (neither will need repair)
= 1 - (0.72)(0.72) = 0.4816

30. Another project.

Since students with Calculus backgrounds are independent from one another, use the multiplication rule. Use the probabilities of events from Exercise 28 in the calculations.

a) P(neither has studied Calculus) = (0.55)(0.55) = 0.3025

b) P(both have studied at least one semester of Calculus) = (0.45)(0.45) = 0.2025

c) P(at least one has had more than one semester of Calculus)
= 1 – P(neither has studied more than one semester of Calculus)
= 1 – (0.87)(0.87) = 0.2431

31. Repairs, again.

a) The repair needs for the two cars must be independent of one another.

b) This may not be reasonable. An owner may treat the two cars similarly, taking good (or poor) care of both. This may decrease (or increase) the likelihood that each needs to be repaired.

32. Final project.

a) The Calculus backgrounds of the students must be independent of one another.

b) Since the professor assigned the groups at random, the Calculus backgrounds are independent.

33. Energy 2013.

a) P(response is "Increase oil, gas, and coal") = 164/529 ≈ 0.310

b) P("Equally important" or "No opinion") = 37/529 + 16/529 = 53/529 ≈ 0.100

34. Failing fathers?

a) P(response is "Harder") = 682/2005 ≈ 0.340

b) P(response is "Same" or "Easier") = 802/2005 + 501/2005 = 1303/2005 ≈ 0.650

35. More energy.

a) P(all three respond "Develop wind and solar") = $\left(\frac{312}{529}\right)\left(\frac{312}{529}\right)\left(\frac{312}{529}\right) \approx 0.205$

b) P(none respond "Equally important") = $\left(\frac{492}{529}\right)\left(\frac{492}{529}\right)\left(\frac{492}{529}\right) \approx 0.805$

c) In order to compute the probabilities, we must assume that responses are independent.

d) It is reasonable to assume that responses are independent, since the three people were chosen at random.

36. Fathers revisited.

a) P(both think being a father is easier) = $\left(\frac{501}{2005}\right)\left(\frac{501}{2005}\right) \approx 0.062$

b) P(neither thinks being a father is easier) = $\left(\frac{1504}{2005}\right)\left(\frac{1504}{2005}\right) \approx 0.563$

c) P(first thinks being a father is easier, the second doesn't) = $\left(\frac{501}{2005}\right)\left(\frac{1504}{2005}\right) \approx 0.187$

d) In order to compute the probabilities, we must assume that responses are independent.

e) It is reasonable to assume that responses are independent, since the two people were chosen at random.

37. Polling.

a)
$$\begin{aligned}&P(\text{household is contacted and household refuses to cooperate})\\&= P(\text{household is contacted})P(\text{household refuses} \mid \text{contacted})\\&= (0.62)(1-0.14) = 0.5332\end{aligned}$$

b) $P(\text{fail to contact household or contacting and not getting interview})$
$= P(\text{fail to contact}) + P(\text{contact household})P(\text{not getting interview} \mid \text{contacted})$
$= (1-0.62)+(0.62)(1-0.14) = 0.9132$

c) The question in part b covers all possible occurrences *except* contacting the house and getting the interview.

$P(\text{failing to contact household or contacting and not getting the interview})$
$= 1 - P(\text{contacting the household and getting the interview})$
$= 1-(0.62)(0.14) = 0.9132$

38. Polling, part II.

a) $P(\text{2012 household is contacted and household cooperates})$
$= P(\text{household is contacted})P(\text{household cooperates} \mid \text{contacted})$
$= (0.62)(0.14) = 0.0868$

b) $P(\text{1997 household is contacted and cooperates})$
$= P(\text{household is contacted})P(\text{household cooperates} \mid \text{contacted})$
$= (0.90)(0.43) = 0.387$

It was more likely for pollsters to obtain an interview at the next household in 1997 than in 2003.

39. M&M's.

a) Since all of the events are disjoint (an M&M can't be two colors at once!), use the addition rule where applicable.

1. P(brown) = 1 – P(not brown) = 1 – P(yellow or red or orange or blue or green) = 1 – (0.20 + 0.20 + 0.10 + 0.10 +0.10) = 0.30
2. P(yellow or orange) = 0.20 + 0.10 = 0.30
3. P(not green) = 1 – P(green) = 1 – 0.10 = 0.90
4. P(striped) = 0

b) Since the events are independent (picking out one M&M doesn't affect the outcome of the next pick), the multiplication rule may be used.

1. P(all three are brown) = (0.30)(0.30)(0.30) = 0.027
2. P(the third one is the first one that is red) = P(not red and not red and red) = (0.80)(0.80)(0.20) = 0.128
3. P(no yellow) = P(not yellow and not yellow and not yellow) = (0.80)(0.80)(0.80) = 0.512

4. P(at least one is green) = 1 – P(none are green) = 1 – (0.90)(0.90)(0.90) = 0.271

40. Blood.

a) Since all of the events are disjoint (a person cannot have more than one blood type!), use the addition rule where applicable.

1. P(Type AB) = 1 – P(not Type AB) = 1 – P(Type O or Type A or Type B)
= 1 – (0.45 + 0.40 + 0.11) = 0.04

2. P(Type A or Type B) = 0.40 + 0.11 = 0.51

3. P(not Type O) = 1 – P(Type O) = 1 – 0.45 = 0.55

b) Since the events are independent (one person's blood type doesn't affect the blood type of the next), the multiplication rule may be used.

1. P(all four are Type O) = (0.45)(0.45)(0.45)(0.45) ≈ 0.041

2. P(no one is Type AB) = P(not AB and not AB and not AB and not AB)
= (0.96)(0.96)(0.96)(0.96) ≈ 0.849

3. P(not all Type A) = 1 – P(all Type A) = 1 – (0.40)(0.40)(0.40)(0.40) = 0.9744

4. P(at least one person is Type B) = 1 – P(no one is Type B)
= 1 – (0.89)(0.89)(0.89)(0.89) ≈ 0.373

41. Disjoint or independent?

a) For one draw, the events of getting a red M&M and getting an orange M&M are disjoint events. Your single draw cannot be both red and orange.

b) For two draws, the events of getting a red M&M on the first draw and a red M&M on the second draw are independent events. Knowing that the first draw is red does not influence the probability of getting a red M&M on the second draw.

c) Disjoint events can never be independent. Once you know that one of a pair of disjoint events has occurred, the other one cannot occur, so its probability has become zero. For example, consider drawing one M&M. If it is red, it cannot possible be orange. Knowing that the M&M is red influences the probability that the M&M is orange. It's zero. The events are not independent.

42. Disjoint or independent?

a) For one person, the events of having Type A blood and having Type B blood are disjoint events. One person cannot be have both Type A and Type B blood.

b) For two people, the events of the first having Type A blood and the second having Type B blood are independent events. Knowing that the first person has Type A blood does not influence the probability of the second person having Type B blood.

c) Disjoint events can never be independent. Once you know that one of a pair of disjoint events has occurred, the other one cannot occur, so its probability has become zero. For example, consider selecting one person, and checking his or her blood type. If the person's blood type is Type A, it cannot possibly be Type B. Knowing that the person's blood type is Type A influences the probability that the person's blood type is Type B. It's zero. The events are not independent.

43. Dice.

a) $P(6) = \frac{1}{6}$, so $P(\text{all 6's}) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) \approx 0.005$

b) $P(\text{odd}) = P(1 \text{ or } 3 \text{ or } 5) = \frac{3}{6}$, so $P(\text{all odd}) = \left(\frac{3}{6}\right)\left(\frac{3}{6}\right)\left(\frac{3}{6}\right) = 0.125$

c) $P(\text{not divisible by 3}) = P(1 \text{ or } 2 \text{ or } 4 \text{ or } 5) = \frac{4}{6}$

$P(\text{none divisible by 3}) = \left(\frac{4}{6}\right)\left(\frac{4}{6}\right)\left(\frac{4}{6}\right) \approx 0.296$

d) $P(\text{at least one 5}) = 1 - P(\text{no 5's}) = 1 - \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right) \approx 0.421$

e) $P(\text{not all 5's}) = 1 - P(\text{all 5's}) = 1 - \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) \approx 0.995$

44. Slot Machine.

Each wheel runs independently of the others, so the multiplication rule may be used.

a) $P(\text{lemon on 1 wheel}) = 0.30$, so $P(\text{3 lemons}) = (0.30)(0.30)(0.30) = 0.027$

b) $P(\text{bar or bell on 1 wheel}) = 0.50$, so $P(\text{no fruit symbols}) = (0.50)(0.50)(0.50) = 0.125$

c) $P(\text{bell on 1 wheel}) = 0.10$, so $P(\text{3 bells}) = (0.10)(0.10)(0.10) = 0.001$

d) $P(\text{no bell on 1 wheel}) = 0.90$, so $P(\text{no bells on 3 wheels}) = (0.90)(0.90)(0.90) = 0.729$

e) $P(\text{no bar on 1 wheel}) = 0.60$.
$P(\text{at least one bar on 3 wheels}) = 1 - P(\text{no bars}) = 1 - (0.60)(0.60)(0.60) = 0.784$

45. Champion bowler.

Assuming each frame is independent of others, the multiplication rule may be used.

a) $P(\text{no strikes in 3 frames}) = (0.30)(0.30)(0.30) = 0.027$

b) $P(\text{makes first strike in the third frame}) = (0.30)(0.30)(0.70) = 0.063$

c) $P(\text{at least one strike in first three frames}) = 1 - P(\text{no strikes}) = 1 - (0.30)^3 = 0.973$

d) $P(\text{perfect game}) = (0.70)^{12} \approx 0.014$

46. The train.

Assuming the arrival time is independent from one day to the next, the multiplication rule may be used.

a) P(gets stopped Monday and gets stopped Tuesday) = (0.15)(0.15) = 0.0225

b) P(gets stopped for the first time on Thursday) = (0.85)(0.85)(0.85)(0.15) ≈ 0.092

c) P(gets stopped every day) = $(0.15)^5 \approx 0.00008$

d) P(gets stopped at least once) = 1 – P(never gets stopped) = $1 - (0.85)^5 \approx 0.556$

47. Voters.

Since you are calling at random, one person's political affiliation is independent of another's. The multiplication rule may be used.

a) P(all Republicans) = (0.29)(0.29)(0.29) ≈ 0.024

b) P(no Democrats) = (1 – 0.37)(1 – 0.37)(1 – 0.37) ≈ 0.25

c) P(at least one Ind.) = 1 – P(no Independents) = 1 – (0.77)(0.77)(0.77) ≈ 0.543

48. Religion.

Since you are calling at random, one person's religion is independent of another's. The multiplication rule may be used.

a) P(all Christian) = (0.62)(0.62)(0.62)(0.62) ≈ 0.148

b) P(no Jews) = (1 – 0.12)(1 – 0.12)(1 – 0.12)(1 – 0.12) ≈ 0.600

c) P(at least one person who is nonreligious) = 1 – P(no nonreligious people)
= 1 – (0.90)(0.90)(0.90)(0.90) = 0.3439

49. Lights.

Assume that the defective light bulbs are distributed randomly to all stores so that the events can be considered independent. The multiplication rule may be used.

P(at least one of five bulbs is defective) = 1 – P(none are defective)
= 1 – (0.94)(0.94)(0.94)(0.94)(0.94) ≈ 0.266

50. Pepsi.

Assume that the winning caps are distributed randomly, so that the events can be considered independent. The multiplication rule may be used.

P(you win something) = 1 – P(you win nothing) = $1 - (0.90)^6 \approx 0.469$

51. 9/11?

a) For any date with a valid three-digit date, the chance is 0.001, or 1 in 1000. For many dates in October through December, the probability is 0. For example, there is no way three digits will make 1015, to match October 15.

b) There are 65 days when the chance to match is 0. (October 10 through October 31, November 10 through November 30, and December 10 through December 31.) That leaves 300 days in a year (that is not a leap year) in which a match might occur.
$P(\text{no matches in 300 days}) = (0.999)^{300} \approx 0.741.$

c) $P(\text{at least one match in a year}) = 1 - P(\text{no matches in a year}) = 1 - 0.741 \approx 0.259$

d) $P(\text{at least one match on 9/11 in one of the 50 states})$
$= 1 - P(\text{no matches in 50 states}) = 1 - (0.999)^{50} \approx 0.049$

52. Red cards.

a) Your thinking is correct. There are 42 cards left in the deck, 26 black and only 16 red.

b) This is not an example of the Law of Large Numbers. There is no "long run". You'll see the entire deck after 52 cards, and you know there will be 26 of each color then.

Chapter 14 – Probability Rules!

Section 14.1

1. Pet ownership.

$$\begin{aligned} P(\text{dog or cat}) &= P(\text{dog}) + P(\text{cat}) - P(\text{dog and cat}) \\ &= 0.25 + 0.29 - 0.12 = 0.42 \end{aligned}$$

2. Cooking and shopping.

$$\begin{aligned} &P(\text{likes to cook or likes to shop}) \\ &= P(\text{likes to cook}) + P(\text{likes to shop}) - P(\text{likes to cook and likes to shop}) \\ &= 0.45 + 0.59 - 0.23 = 0.81 \end{aligned}$$

Section 14.2

3. Sports.

	Football	No Football	Total
Basketball	27	13	40
No Basketball	38	22	60
Total	65	35	100

$$P(\text{football} \mid \text{basketball}) = \frac{P(\text{football and basketball})}{P(\text{basketball})} = \frac{\frac{27}{100}}{\frac{40}{100}} = 0.675$$

(Or, use the table. Of the 40 people who like to watch basketball, 27 people also like to watch football. $27/40 = 0.675$)

4. Sports again.

$$P(\text{football} \mid \text{no basketball}) = \frac{P(\text{football and no basketball})}{P(\text{no basketball})} = \frac{\frac{38}{100}}{\frac{60}{100}} \approx 0.633$$

(Or, use the table. Of the 60 people who don't like to watch basketball, 38 people like to watch football. $38/60 \approx 0.633$)

5. Late to the train.

$$\begin{aligned} P(\text{let out late and missing train}) &= P(\text{let out late}) \times P(\text{missing train} \mid \text{let out late}) \\ &= (0.30)(0.45) = 0.135 \end{aligned}$$

6. Field goals.

$$\begin{aligned} P(\text{make first and make second}) &= P(\text{make first}) \times P(\text{make second} \mid \text{make first}) \\ &= (0.70)(0.90) = 0.63 \end{aligned}$$

Section 14.3

7. Titanic.

The overall survival rate, $P(S)$, was 0.323, yet the survival rate for first class passengers, $P(S|FC)$, was 0.625. Since, $P(S) \neq P(S|FC)$, survival and ticket class are not independent. Rather, survival rate depended on class.

8. Births.

If sex of a child is independent of gender, then $P(\text{girl}) = P(\text{girl}|\text{four boys})$. This means that the probability of a woman giving birth to a girl after having four boys is not greater than it was at her first birth. These probabilities are the same.

Section 14.4

9. Facebook.

	US	Not US	Total
Log on Every Day	0.20	0.30	0.50
Do Not Log on Every Day	0.10	0.40	0.50
Total	0.30	0.70	1.00

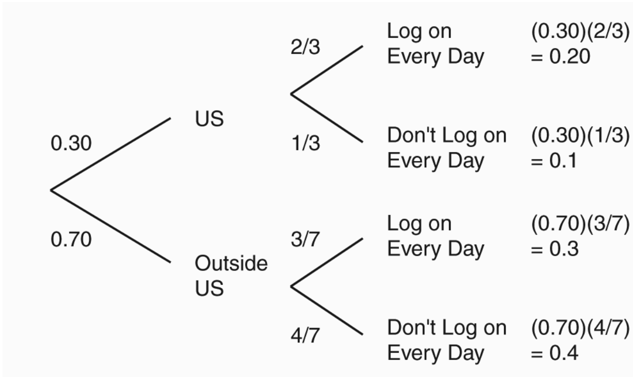
We have joint probabilities and marginal probabilities, not conditional probabilities, so a table is the better choice.

10. Online banking.

	Bank Online	Don't Bank Online	Total
Under 50	0.25	0.15	0.40
50 or older	0.05	0.55	0.60
Total	0.30	0.70	1.00

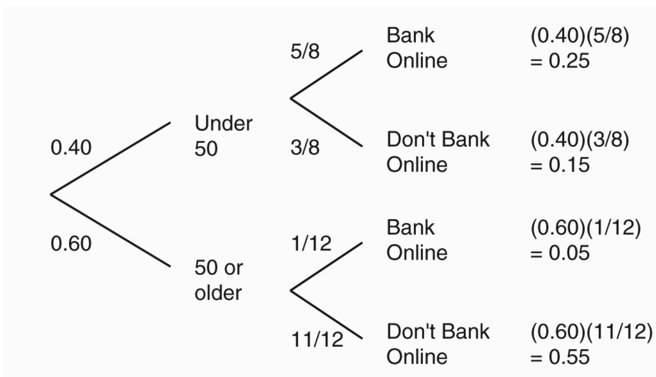
We have joint probabilities and marginal probabilities, not conditional probabilities, so a table is the better choice.

11. Facebook again.



A tree is better because we have conditional and marginal probabilities. The joint probabilities are found at the end of the branches.

12. Online banking again.



A tree is better because we have conditional and marginal probabilities. The joint probabilities are found at the end of the branches.

Section 14.5

13. Facebook final.

$$P(\text{US} \mid \text{Log on every day}) = \frac{P(\text{US and Log on every day})}{P(\text{Log on every day})} = \frac{0.20}{0.20+0.30} = 0.40$$

Knowing that a person logs on every day increases probability that the person is from the United States.

14. Online banking last time.

$$P(\text{Under 50} \mid \text{Bank online}) = \frac{P(\text{Under 50 and Bank online})}{P(\text{Bank online})} = \frac{0.25}{0.25+0.05} \approx 0.833$$

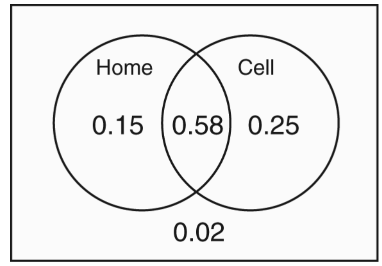
Knowing that someone banks online more than doubles the probability that the person is younger than 50.

Chapter Exercises.

15. Phones.

a) P(home phone or cell phone)
= P(home) + P(cell) – P(home and cell)
= 0.73 + 0.83 – 0.58 = 0.98

Or, from the Venn: 0.15 + 0.58 + 0.25 = 0.98



b) P(neither)= 1 – P(home or cell) = 1 – 0.98 = 0.02

Or, from the Venn: 0.02 (the region outside the circles)

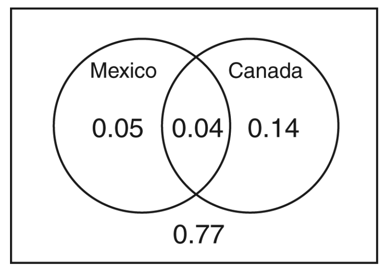
c) P(cell but no home) = P(cell) – P(home and cell) = 0.83 – 0.58 = 0.25

Or, from the Venn: 0.25 (the region inside cell circle, yet outside home circle)

16. Travel.

a) P(Canada and not Mexico)
= P(Canada) – P(Canada and Mexico) = 0.18 – 0.04 = 0.14

Or, from the Venn: 0.14
(region inside the Canada circle, yet outside the Mexico circle)



b) P(either Canada or Mexico)
= P(Canada) + P(Mexico) – P(Canada and Mexico) = 0.18 + 0.09 – 0.04 = 0.23

Or, from the Venn: 0.05 + 0.04 + 0.14 = 0.23 (the regions inside the circles)

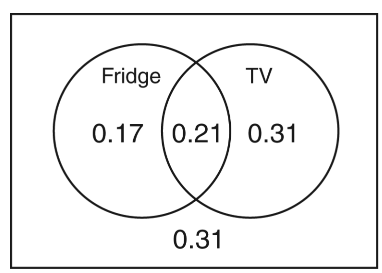
c) P(neither Canada nor Mexico) = 1 – P(either Canada or Mexico) = 1 – 0.23 = 0.77

Or, from the Venn: 0.77 (the region outside the circles)

17. Amenities.

a) P(TV and no refrigerator)
= P(TV) – P(TV and refrigerator)
= 0.52 – 0.21 = 0.31

Or, from the Venn: 0.31
(inside the TV circle, yet outside the Fridge circle)



b) P(refrigerator or TV, but not both) =
= [P(refrigerator) – P(refrigerator and TV)] +[P(TV) – P(refrigerator and TV)]
= [0.38 – 0.21] + [0.52 – 0.21] = 0.48

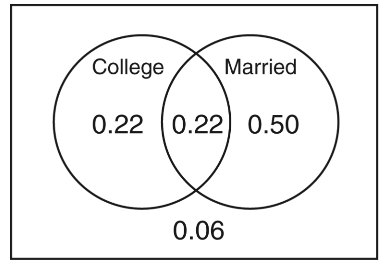
Using the Venn diagram, simply add the probabilities in the two regions for Fridge only and TV only. P(refrigerator or TV, but not both) = 0.17 + 0.31 = 0.48

c) P(neither TV nor refrigerator) $= 1 - P$(either TV or refrigerator)
$= 1 - [P(\text{TV}) + P(\text{ref.}) - P(\text{TV and ref.})]$
$= 1 - [0.52 + 0.38 - 0.21]$
$= 0.31$

Or, from the Venn: 0.31 (the region outside the circles)

18. Workers.

a) P(neither married nor a college graduate)
$= 1 - P$(either married or college graduate)
$= 1 - [P(\text{married}) + P(\text{college graduate}) - P(\text{both})]$
$= 1 - [0.72 + 0.44 - 0.22]$
$= 1 - [0.94]$
$= 0.06$



Or, from the Venn: 0.06 (outside the circles)

b) P(married and not a college grad) $= P$(married) $- P$(married and a college grad)
$= 0.72 - 0.22$
$= 0.50$

Or, from the Venn: 0.50 (inside the Married circle, yet outside the College circle)

c) P(married or a college graduate) $= P$(married) $+ P$(college graduate) $- P$(both)
$= 0.72 + 0.44 - 0.22$
$= 0.94$

Or, from the Venn diagram: 0.22 + 0.22 + 0.50 = 0.94 (inside the circles)

19. Global survey.

a) $P(USA) = \frac{1557}{7690} \approx 0.2025$

b) $P(\text{some high school or primary or less}) = \frac{4195}{7690} + \frac{1161}{7690} \approx 0.6965$

c)

$$P(\text{France or post-graduate}) = P(\text{France}) + P(\text{post-graduate}) - P(\text{both})$$
$$= \frac{1539}{7690} + \frac{379}{7690} - \frac{69}{7690} \approx 0.2404$$

d) $P(\text{France and primary school or less}) = \frac{309}{7690} \approx 0.0402$

20. Birth order.

a) $P(\text{Human Ecology}) = \frac{43}{223} \approx 0.193$ **b)** $P(\text{first-born}) = \frac{113}{223} \approx 0.507$

c) $P(\text{first-born and Human Ecology}) = \dfrac{15}{223} \approx 0.067$

d) $P(\text{first-born or Human Ecology})$

$= P(\text{first-born}) + P(\text{Human Ecology}) - P(\text{first-born and Human Ecology})$

$= \dfrac{113}{223} + \dfrac{43}{223} - \dfrac{15}{223} \approx 0.632$

21. Cards.

a) $P(\text{heart} \mid \text{red}) = \dfrac{P(\text{heart and red})}{P(\text{red})} = \dfrac{13/52}{26/52} = \dfrac{1}{2}$

A more intuitive approach is to think about only the red cards. Half of them are hearts.

b) $P(\text{red} \mid \text{heart}) = \dfrac{P(\text{red and heart})}{P(\text{heart})} = \dfrac{13/52}{13/52} = 1$

Think about only the hearts. They are all red!

c) $P(\text{ace} \mid \text{red}) = \dfrac{P(\text{ace and red})}{P(\text{red})} = \dfrac{2/52}{26/52} = \dfrac{2}{26} \approx 0.077$

Consider only the red cards. Of those 26 cards, 2 of them are aces.

d) $P(\text{queen} \mid \text{face}) = \dfrac{P(\text{queen and face})}{P(\text{face})} = \dfrac{4/52}{12/52} \approx 0.333$

There are 12 faces cards: 4 jacks, 4 queens, and 4 kings. Four of the 12 face cards are queens.

22. Pets.

Organize the counts in a two-way table.

	Cats	Dogs	Total
Male	6	8	14
Female	12	16	28
Total	18	24	42

a) $P(\text{male} \mid \text{cat}) = \dfrac{P(\text{male and cat})}{P(\text{cat})}$

$= \dfrac{6/42}{18/42} = \dfrac{6}{18} \approx 0.333$

Consider only the Cats column. There are 6 male cats, out of a total of 18 cats.

b) $P(\text{cat} \mid \text{female}) = \dfrac{P(\text{cat and female})}{P(\text{female})} = \dfrac{12/42}{28/42} = \dfrac{12}{28} \approx 0.429$

We are interested in the Female row. Of the 28 female animals, 12 are cats.

c) $P(\text{female} \mid \text{dog}) = \dfrac{P(\text{female and dog})}{P(\text{dog})} = \dfrac{16/42}{24/42} = \dfrac{16}{24} \approx 0.667$

Look at only the Dogs column. There are 24 dogs, and 16 of them are female.

23. Health.

Construct a two-way table of the conditional probabilities, including the marginal probabilities.

Blood Pressure

Cholesterol	High	OK	Total
High	0.11	0.21	0.32
OK	0.16	0.52	0.68
Total	0.27	0.73	1.00

a) P(both conditions) = 0.11

b) P(high BP) = 0.11 + 0.16 = 0.27

c) $P(\text{high chol.} \mid \text{high BP}) = \dfrac{P(\text{high chol. and high BP})}{P(\text{high BP})} = \dfrac{0.11}{0.27} \approx 0.407$

Consider only the High Blood Pressure column. Within this column, the probability of having high cholesterol is 0.11 out of a total of 0.27.

d) $P(\text{high BP} \mid \text{high chol.}) = \dfrac{P(\text{high BP and high chol.})}{P(\text{high chol.})} = \dfrac{0.11}{0.32} \approx 0.344$

This time, consider only the high cholesterol row. Within this row, the probability of having high blood pressure is 0.11, out of a total of 0.32.

24. Immigration.

a) Construct a two-way table of the conditional probabilities, including the marginal probabilities.

Stronger Immigration Enforcement

Party	Favor	Oppose	No Opinion	Total
Republican	0.30	0.04	0.03	0.37
Democrat	0.22	0.11	0.02	0.35
Other	0.16	0.07	0.05	0.28
Total	0.68	0.22	0.10	1.00

i) P(favor stronger immigration enforcement)
= 0.30 + 0.22 + 0.16
= 0.68

ii) $P(\text{favor enforcement} \mid \text{Rep.}) = \dfrac{P(\text{favor enforcement and Rep.})}{P(\text{Republican})} = \dfrac{0.30}{0.37} \approx 0.811$

Consider only the Republican row. The probability of favoring stronger immigration enforcement is 0.30 out of a total of 0.37 for that row.

iii) $P(\text{Dem} \mid \text{favor enf.}) = \dfrac{P(\text{Dem and favor enf.})}{P(\text{favor enforcement})} = \dfrac{0.22}{0.68} \approx 0.324$

Consider only the Favor column. The probability of being a Democrat is 0.22 out of a total of 0.68 for that column.

b) $P(\text{Rep. or favor enforcement}) = P(\text{Rep.}) + P(\text{favor enforcement}) - P(\text{both})$
$= 0.37 + 0.68 - 0.30 = 0.75$

The overall probabilities of being a Republican and of favoring enforcement are from the marginal distribution of probability (the totals). The candidate can expect 75% of the votes, provided her estimates are correct.

25. Global survey, take 2.

a) $P(\text{USA and postgraduate work}) = \dfrac{84}{7690} \approx 0.011$

b) $P(\text{USA} \mid \text{post-graduate}) = \dfrac{84}{379} \approx 0.222$

c) $P(\text{post-graduate} \mid \text{USA}) = \dfrac{84}{1557} \approx 0.054$

d) $P(\text{primary} \mid \text{China}) = \dfrac{506}{1502} \approx 0.337$

e) $P(\text{China} \mid \text{primary}) = \dfrac{506}{1161} \approx 0.436$

26. Birth order, take 2.

a) $P(\text{Arts and Science and second child}) = \dfrac{23}{223} \approx 0.103$

b) $P(\text{second child} \mid \text{Arts and Science}) = \dfrac{23}{57} \approx 0.404$

c) $P(\text{Arts and Science} \mid \text{second child}) = \dfrac{23}{110} \approx 0.209$

d) $P(\text{Agriculture} \mid \text{first -born}) = \dfrac{52}{113} \approx 0.460$

e) $P(\text{first-born} \mid \text{Agriculture}) = \dfrac{52}{93} \approx 0.559$

27. Sick kids.

Having a fever and having a sore throat are not independent events, so:

P(fever and sore throat) = P(Fever) P(Sore Throat | Fever) = (0.70)(0.30) = 0.21

The probability that a kid with a fever has a sore throat is 0.21.

28. Sick cars.

Needing repairs and paying more than \$400 for the repairs are not independent events. (What happens to the probability of paying more than \$400, if you don't need repairs?!)

P(needing repairs and paying more than \$400)
= P(needing repairs) P(paying more than \$400 | repairs are needed)
= (0.20)(0.40) = 0.08

29. Cards.

a)

$$\begin{aligned} P(\text{first heart drawn is on the third card}) &= P(\text{no heart})P(\text{no heart})P(\text{heart}) \\ &= \left(\frac{39}{52}\right)\left(\frac{38}{51}\right)\left(\frac{13}{50}\right) \approx 0.145 \end{aligned}$$

b)

$$\begin{aligned} P(\text{all three cards drawn are red}) &= P(\text{red})P(\text{red})P(\text{red}) \\ &= \left(\frac{26}{52}\right)\left(\frac{25}{51}\right)\left(\frac{24}{50}\right) \approx 0.118 \end{aligned}$$

c)

$$\begin{aligned} P(\text{none of the cards are spades}) &= P(\text{no spade})P(\text{no spade})P(\text{no spade}) \\ &= \left(\frac{39}{52}\right)\left(\frac{38}{51}\right)\left(\frac{37}{50}\right) \approx 0.414 \end{aligned}$$

d)

$$\begin{aligned} P(\text{at least one of the cards is an ace}) &= 1 - P(\text{none of the cards are aces}) \\ &= 1 - \left[P(\text{no ace})P(\text{no ace})P(\text{no ace})\right] \\ &= 1 - \left(\frac{48}{52}\right)\left(\frac{47}{51}\right)\left(\frac{46}{50}\right) \approx 0.217 \end{aligned}$$

30. Another hand.

a)

$$\begin{aligned} P(\text{none of the cards are aces}) &= P(\text{no ace})P(\text{no ace})P(\text{no ace}) \\ &= \left(\frac{48}{52}\right)\left(\frac{47}{51}\right)\left(\frac{46}{50}\right) \approx 0.783 \end{aligned}$$

b)

$$
\begin{aligned}
P(\text{all of the cards are hearts}) &= P(\text{heart})P(\text{heart})P(\text{heart}) \\
&= \left(\frac{13}{52}\right)\left(\frac{12}{51}\right)\left(\frac{11}{50}\right) \approx 0.013
\end{aligned}
$$

c)

$$
\begin{aligned}
P(\text{the third card is the first red}) &= P(\text{no red})P(\text{no red})P(\text{red}) \\
&= \left(\frac{26}{52}\right)\left(\frac{25}{51}\right)\left(\frac{26}{50}\right) \approx 0.127
\end{aligned}
$$

d)

$$
\begin{aligned}
P(\text{at least one card is a diamond}) &= 1 - P(\text{no cards are diamonds}) \\
&= 1 - \left[P(\text{no diam.})P(\text{no diam.})P(\text{no diam.})\right] \\
&= 1 - \left(\frac{39}{52}\right)\left(\frac{38}{51}\right)\left(\frac{37}{50}\right) \approx 0.586
\end{aligned}
$$

31. Batteries.

Since batteries are not being replaced, use conditional probabilities throughout.

a)

$$
\begin{aligned}
P(\text{the first two batteries are good}) &= P(\text{good})P(\text{good}) \\
&= \left(\frac{7}{12}\right)\left(\frac{6}{11}\right) \approx 0.318
\end{aligned}
$$

b)

$$
\begin{aligned}
&P(\text{at least one of the first three batteries works}) \\
&= 1 - P(\text{none of the first three batteries work}) \\
&= 1 - \left[P(\text{no good})P(\text{no good})P(\text{no good})\right] \\
&= 1 - \left(\frac{5}{12}\right)\left(\frac{4}{11}\right)\left(\frac{3}{10}\right) \approx 0.955
\end{aligned}
$$

c)

$$
\begin{aligned}
P(\text{the first four batteries are good}) &= P(\text{good})P(\text{good})P(\text{good})P(\text{good}) \\
&= \left(\frac{7}{12}\right)\left(\frac{6}{11}\right)\left(\frac{5}{10}\right)\left(\frac{4}{9}\right) \approx 0.071
\end{aligned}
$$

d)

$$
\begin{aligned}
&P(\text{pick five to find one good}) \\
&= P(\text{not good})P(\text{not good})P(\text{not good})P(\text{not good})P(\text{good}) \\
&= \left(\frac{5}{12}\right)\left(\frac{4}{11}\right)\left(\frac{3}{10}\right)\left(\frac{2}{9}\right)\left(\frac{7}{8}\right) \approx 0.009
\end{aligned}
$$

32. Shirts.

You need two shirts so don't replace them. Use conditional probabilities throughout.

a)

$$P(\text{the first two are not mediums}) = P(\text{not medium})\ P(\text{not medium})$$
$$= \left(\frac{16}{20}\right)\left(\frac{15}{19}\right) \approx 0.632$$

b)

$$P(\text{the first medium is the third shirt})$$
$$= P(\text{not medium})\ P(\text{not medium})\ P(\text{medium})$$
$$= \left(\frac{16}{20}\right)\left(\frac{15}{19}\right)\left(\frac{4}{18}\right) \approx 0.140$$

c)

$$P(\text{the first four shirts are extra} - \text{large}) = P(\text{XL})\ P(\text{XL})\ P(\text{XL})\ P(\text{XL})$$
$$= \left(\frac{6}{20}\right)\left(\frac{5}{19}\right)\left(\frac{4}{18}\right)\left(\frac{3}{17}\right) \approx 0.003$$

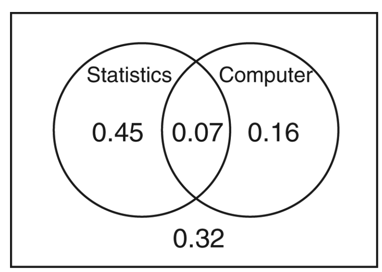
d)

$$P(\text{at least one of four is a medium})$$
$$= 1 - P(\text{none of the first four shirts are mediums})$$
$$= 1 - \left[P(\text{not med.})\ P(\text{not med.})\ P(\text{not med.})\ P(\text{not med.})\right]$$
$$= 1 - \left(\frac{16}{20}\right)\left(\frac{15}{19}\right)\left(\frac{14}{18}\right)\left(\frac{13}{17}\right) \approx 0.624$$

33. Eligibility.

a)

$$P(\text{eligible}) = P(\text{stats}) + P(\text{computer}) - P(\text{both})$$
$$= 0.52 + 0.23 - 0.07$$
$$= 0.68$$



68% of students are eligible for BioResearch, so 100 – 68 = 32% are ineligible.

From the Venn, the region outside the circles represents those students who have taken neither course, and are therefore ineligible for BioResearch.

b)

$$P(\text{computer course} \mid \text{statistics}) = \frac{P(\text{computer and statistics})}{P(\text{statistics})} = \frac{0.07}{0.52} \approx 0.135$$

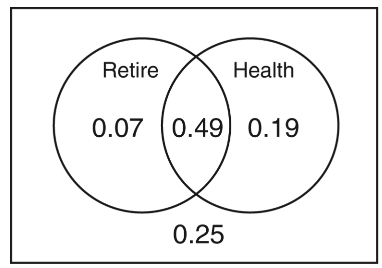
From the Venn, consider only the region inside the Statistics circle. The probability of having taken a computer course is 0.07 out of a total of 0.52 (the entire Statistics circle).

c) Taking the two courses are not disjoint events, since they have outcomes in common. In fact, 7% of juniors have taken both courses.

d) Taking the two courses are not independent events. The overall probability that a junior has taken a computer course is 0.23. The probability that a junior has taken a computer course given that he or she has taken a statistics course is 0.135. If taking the two courses were independent events, these probabilities would be the same.

34. Benefits.

Construct a Venn diagram possible outcomes.



a)
$$\begin{aligned} P(\text{neither benefit}) &= 1 - P(\text{either retirement or health}) \\ &= 1 - [P(\text{ret.}) + P(\text{health}) - P(\text{both})] \\ &= 1 - [\,0.56 + 0.68 - 0.49] \\ &= 0.25 \end{aligned}$$

b)

$$P(\text{health ins.} \mid \text{retirement}) = \frac{P(\text{health insurance and retirement})}{P(\text{retirement})} = \frac{0.49}{0.56} = 0.875$$

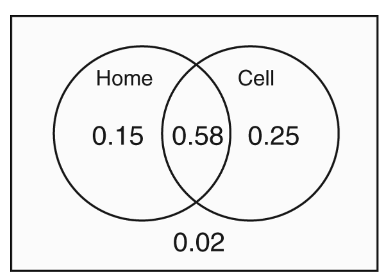
From the Venn, consider only the region inside the Retirement circle. The probability that a worker has health insurance is 0.49 out of a total of 0.56 (the entire Retirement circle).

c) Having health insurance and a retirement plan are not independent events. 68% of all workers have health insurance, while 87.5% of workers with retirement plans also have health insurance. If having health insurance and a retirement plan were independent events, these percentages would be the same.

d) Having these two benefits are not disjoint events, since they have outcomes in common. 49% of workers have both health insurance and a retirement plan.

35. Cell phones in the home.

Construct a Venn diagram of the possible outcomes.



a)

$$P(\text{cell} \mid \text{home}) = \frac{P(\text{cell and home})}{P(\text{home})} = \frac{0.58}{0.73} \approx 0.795$$

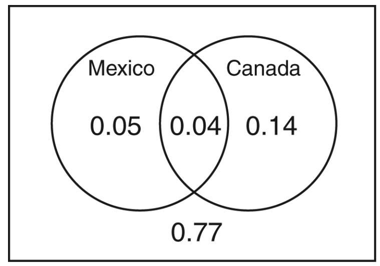
From the Venn, consider only the region inside the Home circle. The probability that the person has a cell phone is 0.58 out of 0.73 (the entire Home circle).

b) Having a home phone and a cell phone are not independent events. 79.5% of people with home phones have cell phones. Overall, 83% of people have cell phones. If having a home phone and cell phone were independent events, these would be the same.

c) No, having a home phone and cell phone are not disjoint events. 58% of people have both.

36. On the road again.

Construct a Venn diagram of the possible outcomes.



a)

$$P(\text{Canada} \mid \text{Mexico}) = \frac{P(\text{Canada and Mexico})}{P(\text{Mexico})} = \frac{0.04}{0.09} \approx 0.444$$

From the Venn, consider only the region inside the Mexico circle. The probability that an American has traveled to Canada is 0.04 out of a total of 0.09 (the entire Mexico circle).

b) No, travel to Mexico and Canada are not disjoint events. 4% of Americans have been to both countries.

c) No, travel to Mexico and Canada are not independent events. 18% of U.S. residents have been to Canada. 44.4% of the U.S. residents who have been to Mexico have also been to Canada. If travel to the two countries were independent, the percentages would be the same.

37. Cards.

Yes, getting an ace is independent of the suit when drawing one card from a well shuffled deck. The overall probability of getting an ace is 4/52, or 1/13, since there are 4 aces in the deck. If you consider just one suit, there is only 1 ace out of 13 cards, so the probability of getting an ace given that the card is a diamond, for instance, is 1/13. Since the probabilities are the same, getting an ace is independent of the suit.

38. Pets, again.

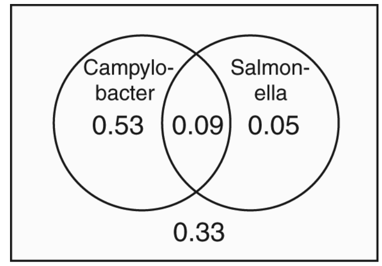
Consider the two-way table from Exercise 22.

	Cats	Dogs	Total
Male	6	8	14
Female	12	16	28
Total	18	24	42

Yes, species and gender are independent events. 8 of 24, or 1/3 of the dogs are male, and 6 of 18, or 1/3 of the cats are male. Since these are the same, species and gender are independent events.

39. Unsafe food.

a) Using the Venn diagram, the probability that a tested chicken was not contaminated with either kind of bacteria is 33%.



b) Contamination with campylobacter and contamination with salmonella are not disjoint events, since 9% of chicken is contaminated with both.

c) Contamination with campylobacter and contamination with salmonella may be independent events. The probability that a tested chicken is contaminated with campylobacter is 0.62. The probability that chicken contaminated with salmonella is also contaminated with campylobacter is $0.09/0.14 \approx 0.64$. If chicken is contaminated with salmonella, it is only slightly more likely to be contaminated with campylobacter than chicken in general. This difference could be attributed to expected variation due to sampling.

40. Birth order, finis.

a) Yes, since the events share no outcomes. Students can enroll in only one college.

b) No, since knowing that one event is true drastically changes the probability of the other. The probability of a student being in the Agriculture college is nearly 42%. The probability of a student being in the Human Ecology college, given that he or she is in the Agriculture college is 0.

c) No, since they share outcomes. 15 students were first-born, Human Ecology students.

d) No, since knowing that one event is true drastically changes the probability of the other. Over 19% of all students enrolled in Human Ecology, but only 13% of first-borns did.

41. Men's health, again.

Consider the two-way table from Exercise 23.

Blood Pressure

Cholesterol	High	OK	Total
High	0.11	0.21	0.32
OK	0.16	0.52	0.68
Total	0.27	0.73	1.00

High blood pressure and high cholesterol are not independent events. 28.8% of men with OK blood pressure have high cholesterol, while 40.7% of men with high blood pressure have high cholesterol. If having high blood pressure and high cholesterol were independent, these percentages would be the same.

42. Politics.

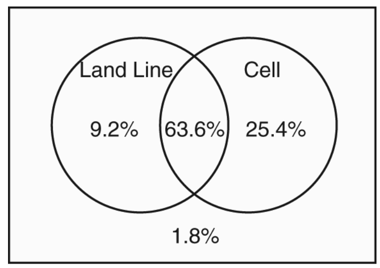
Consider the two-way table from Exercise 24.

Stronger Immigration Enforcement

Party	Favor	Oppose	No Opinion	Total
Republican	0.30	0.04	0.03	0.37
Democrat	0.22	0.11	0.02	0.35
Other	0.16	0.07	0.05	0.28
Total	0.68	0.22	0.10	1.00

Party affiliation and position on the immigration are not independent events. 81.1% of Republicans favor stronger immigration enforcement, but only 62.9% of Democrats favor it. If the events were independent, then these percentages would be the same.

43. Phone service.

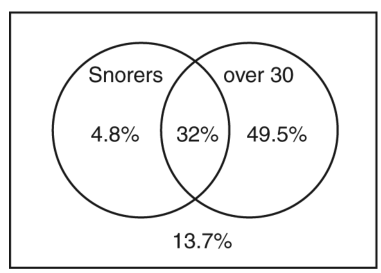


a) Since 25.4% of U.S. adults have only a cell phone, and 1.8% have no phone at all, polling organizations can reach 100 – 25.4 – 1.8 = 72.8% of U.S. adults.

b) Using the Venn diagram, about 72.8% of U.S. adults have a land line. The probability of a U.S. adults having a land line given that they have a cell phone is 63.6/(63.6+25.4) or about 71.5%. It appears that having a cell phone and having a land line are independent, since the probabilities are roughly the same.

44. Snoring.

Organize the percentages in a Venn diagram.



a) 13.7% of the respondents were under 30 and did not snore.

b) According to this survey, snoring is not independent of age. 36.8% of the 995 adults snored, but $32/(32+49.5) \approx 39.3\%$ of those over 30 snored.

45. Gender.

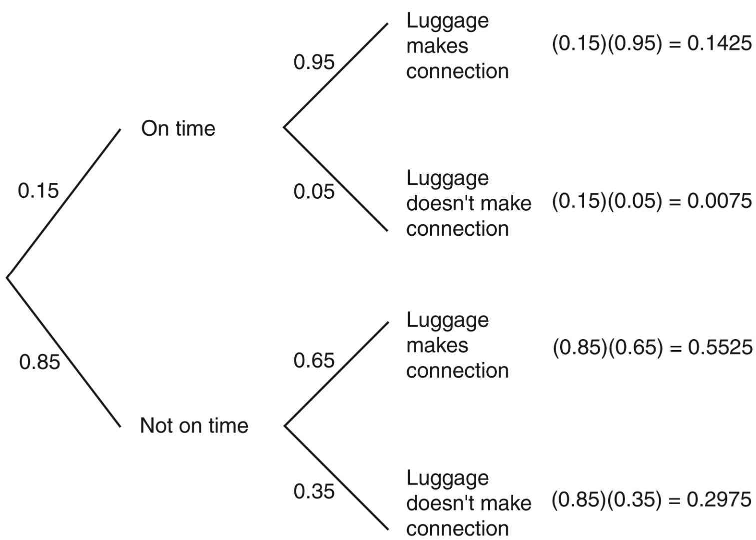
According to the poll, party affiliation is not independent of sex. Overall, $(32+41)/186 \approx 39.25\%$ of the respondents were Democrats. Of the men, only $32/94 \approx 34.04\%$ were Democrats.

46. Cars.

According to the survey, country of origin of the car is not independent of type of driver. $(33+12)/359 \approx 12.5\%$ of the cars were of European origin, but about $33/195 \approx 16.9\%$ of the students drive European cars.

47. Luggage.

Organize using a tree diagram.

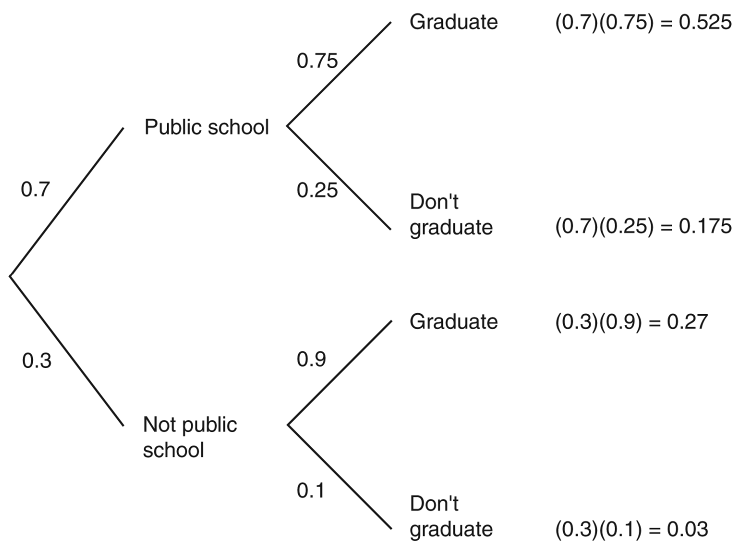


a) No, the flight leaving on time and the luggage making the connection are not independent events. The probability that the luggage makes the connection is dependent on whether or not the flight is on time. The probability is 0.95 if the flight is on time, and only 0.65 if it is not on time.

b)
$$\begin{aligned} P(\text{Luggage}) &= P(\text{On time and Luggage}) + P(\text{Not on time and Luggage}) \\ &= (0.15)(0.95) + (0.85)(0.65) \\ &= 0.695 \end{aligned}$$

48. Graduation.

a) Yes, there is evidence to suggest that a freshman's chances to graduate depend upon what kind of high school the student attended. The graduation rate for public school students is 75%, while the graduation rate for others is 90%. If the high school attended was independent of college graduation, these percentages would be the same.



b)
$$
\begin{aligned}
P(\text{Graduate}) &= P(\text{Public and Graduate}) + P(\text{Not public and Graduate}) \\
&= (0.7)(0.75) + (0.3)(0.9) \\
&= 0.795
\end{aligned}
$$

Overall, 79.5% of freshmen are expected to eventually graduate.

49. Late luggage.

Refer to the tree diagram constructed for Exercise 47.

$$
\begin{aligned}
P(\text{Not on time} \mid \text{No Lug.}) &= \frac{P(\text{Not on time and No Luggage})}{P(\text{No Luggage})} \\
&= \frac{(0.85)(0.35)}{(0.15)(0.05)+(0.85)(0.35)} \approx 0.975
\end{aligned}
$$

If you pick Leah up at the Denver airport and her luggage is not there, the probability that her first flight was delayed is 0.975.

50. Graduation, part II.

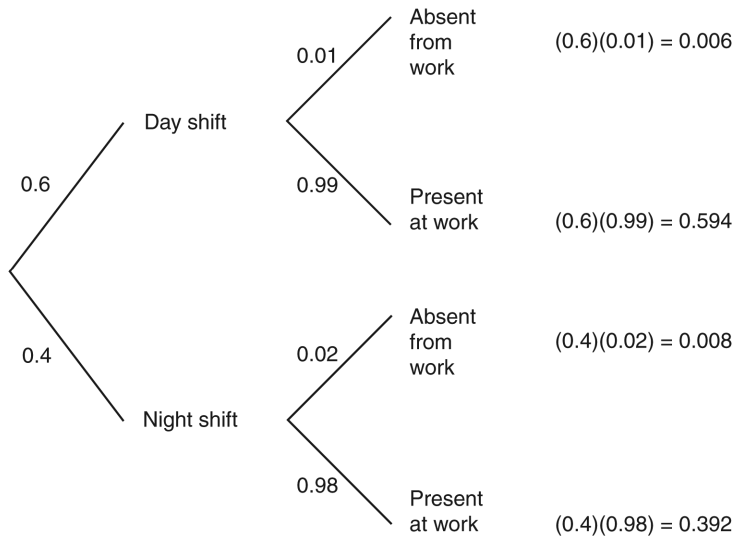
Refer to the tree diagram constructed for Exercise 48.

$$
P(\text{Public} \mid \text{Graduate}) = \frac{P(\text{Public and Graduate})}{P(\text{Graduate})} = \frac{(0.7)(0.75)}{(0.7)(0.75)+(0.3)(0.9)} \approx 0.660
$$

Overall, 66.0% of the graduates of the private college went to public high schools.

51. Absenteeism.

Organize the information in a tree diagram.



a) No, absenteeism is not independent of shift worked. The rate of absenteeism for the night shift is 2%, while the rate for the day shift is only 1%. If the two were independent, the percentages would be the same.

b)

$P(\text{Absent})=P(\text{Day and Absent})+P(\text{Night and Absent})=(0.6)(0.01)+(0.4)(0.02)=0.014$

The overall rate of absenteeism at this company is 1.4%.

52. E-readers.

a) Owning an e-reader and reading at least one book are not independent, since the percentage of people who have read at least one book are different for owners of e-readers and U.S. adults overall. 87.5% of owners of e-readers (0.28/0.32) have read at least one book in the previous year, while only 76% of all U.S. adults have read at least one book in the previous year.

b) If 28% of U.S. adults have read at least one e-book, and 32% have e-readers, then 32% – 28% = 4% of e-reader owners have not read at least one book in the previous year.

53. Absenteeism, part II.

Refer to the tree diagram constructed for Exercise 51.

$$P(\text{Night} \mid \text{Absent})=\frac{P(\text{Night and Absent})}{P(\text{Absent})}=\frac{(0.4)(0.02)}{(0.6)(0.01)+(0.4)(0.02)}\approx 0.571$$

Approximately 57.1% of the company's absenteeism occurs on the night shift.

54. E-readers II.

a) $P(\text{hasn't read at least 1 book} \mid \text{e-reader})$
$$= \frac{P(\text{ hasn't read at least 1 book and e-reader})}{P(\text{e-reader})} = \frac{0.04}{0.32} = 0.125$$

The probability that a randomly selected U.S. adult has read at least one book given that he or she has an e-reader is 12.5%.

b) We know that 24% of U.S. adults have not read any type of book in the previous year, and 4% of U.S. adults were e-reader owners who have not read a book in the previous year. This means that 24% - 4% = 20% of U.S. adults have no e-reader and have read no books.

$P(\text{hasn't read at least 1 book} \mid \text{no e-reader})$
$$= \frac{P(\text{ hasn't read at least 1 book and no e-reader})}{P(\text{no e-reader})} = \frac{0.20}{0.68} \approx 0.294$$

It is more likely that a U.S. adult who does not own an e-reader would have read no books in the previous year.

55. Drunks.

Organize the information into a tree diagram.

a) $P(\text{Detain} \mid \text{Not Drinking}) = 0.2$

b)
$$\begin{aligned} &P(\text{Detain}) \\ &= P(\text{Drinking and Det.}) \\ &\quad + P(\text{Not Drinking and Det.}) \\ &= (0.12)(0.8) + (0.88)(0.2) \\ &= 0.272 \end{aligned}$$

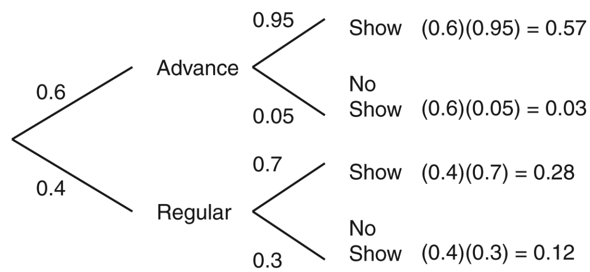
c)
$$\begin{aligned} P(\text{Drunk} \mid \text{Det.}) &= \frac{P(\text{Drunk and Det.})}{P(\text{Detain})} \\ &= \frac{(0.12)(0.8)}{(0.12)(0.8) + (0.88)(0.2)} \\ &\approx 0.353 \end{aligned}$$

d)

$$P(\text{Drunk} \mid \text{Release}) = \frac{P(\text{Drunk and Release})}{P(\text{Release})} = \frac{(0.12)(0.2)}{(0.12)(0.2)+(0.88)(0.8)} \approx 0.033$$

56. No-shows.

Organize the information into a tree diagram.



a)

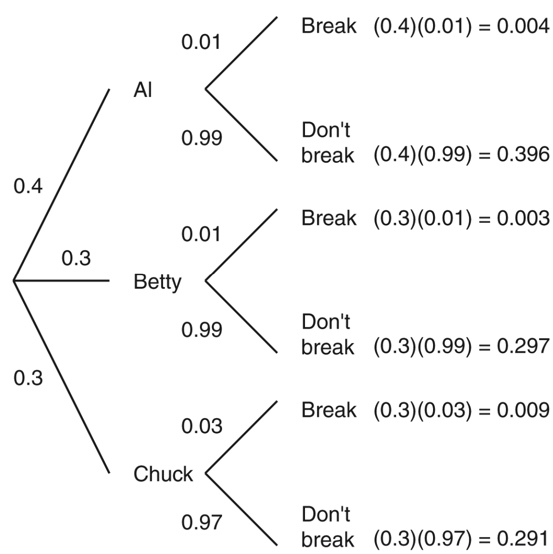
$$\begin{aligned} &P(\text{No Show}) \\ &= P(\text{Advance and No Show}) \\ &\qquad + P(\text{Regular and No Show}) \\ &= (0.60)(0.05) + (0.40)(0.30) \\ &= 0.03 + 0.12 = 0.15 \end{aligned}$$

b) $P(\text{Advance} \mid \text{No Show}) = \dfrac{P(\text{Advance and No Show})}{P(\text{No Show})} = \dfrac{0.03}{0.15} = 0.20$

c) No, being a no show is not independent of the type of ticket a passenger holds. While 30% of regular fare passengers are no shows, only 5% of advanced sale fare passengers are no shows.

57. Dishwashers.

Organize the information in a tree diagram.

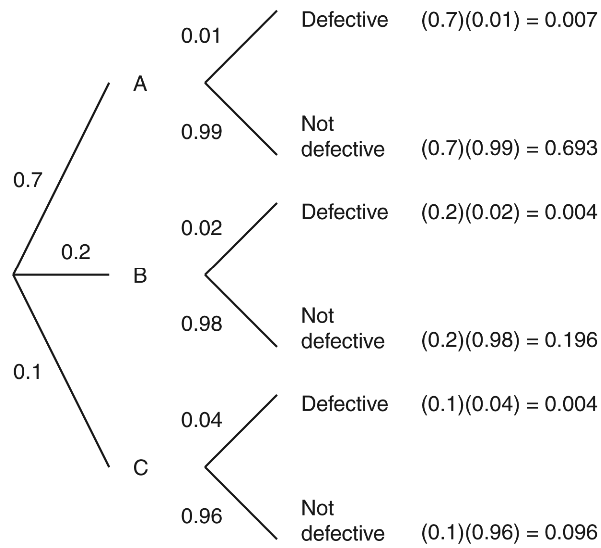


$$\begin{aligned} &P(\text{Chuck} \mid \text{Break}) \\ &= \frac{P(\text{Chuck and Break})}{P(\text{Break})} \\ &= \frac{(0.3)(0.03)}{(0.4)(0.01)+(0.3)(0.01)+(0.3)(0.03)} \\ &\approx 0.563 \end{aligned}$$

If you hear a dish break, the probability that Chuck is on the job is approximately 0.563.

58. Parts.

Organize the information in a tree diagram.

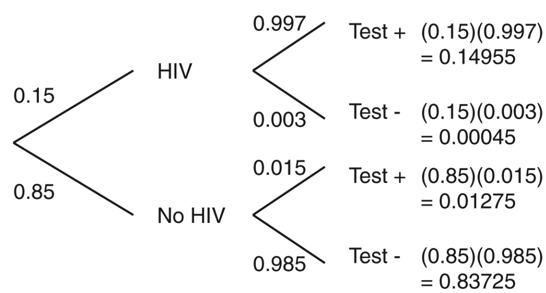


$$\begin{aligned} &P(\text{Supplier A} \mid \text{Defective}) \\ &= \frac{P(\text{Supplier A and Defective})}{P(\text{Defective})} \\ &= \frac{(0.7)(0.01)}{(0.7)(0.01)+(0.2)(0.02)+(0.1)(0.04)} \\ &\approx 0.467 \end{aligned}$$

The probability that a defective component came from supplier A is approximately 0.467.

59. HIV Testing.

Organize the information in a tree diagram.

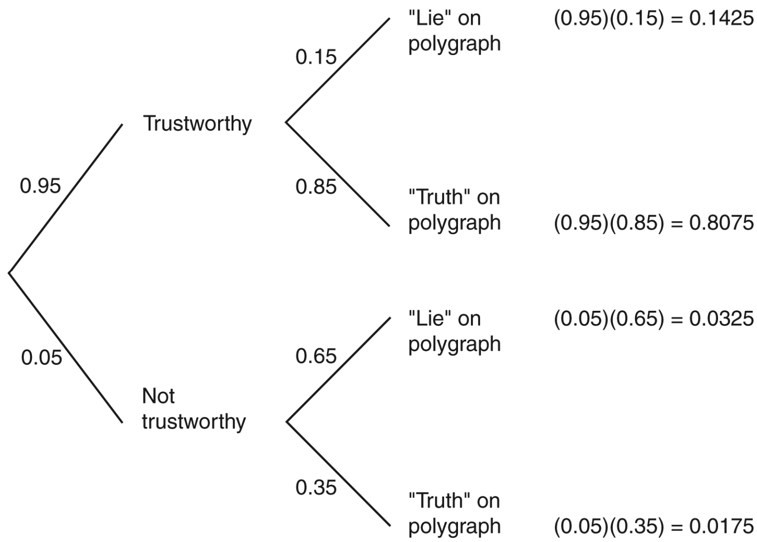


$$\begin{aligned} P(\text{No HIV} \mid \text{Test -}) &= \frac{P(\text{No HIV and Test -})}{P(\text{Test -})} \\ &= \frac{0.83725}{0.00045+0.83725} \\ &\approx 0.999 \end{aligned}$$

The probability that a patient testing negative is truly free of HIV is about 99.9%.

60. Polygraphs.

Organize the information in a tree diagram.



$$P(\text{Trustworthy} \mid \text{"Lie" on polygraph})$$
$$= \frac{P(\text{Trustworthy and "Lie" on polygraph})}{P(\text{"Lie" on polygraph})}$$
$$= \frac{(0.95)(0.15)}{(0.95)(0.15)+(0.05)(0.65)}$$
$$\approx 0.8143$$

The probability that a job applicant rejected under suspicion of dishonesty is actually trustworthy is about 0.8143.

Chapter 15 – Random Variables

Section 15.1

1. Oranges.

$\mu = E(\text{Oranges}) = 25(0.10) + 30(0.40) + 35(0.30) + 40(0.20) = 33$

The citrus farmer can expect an average of 33 oranges per tree.

2. Caffeinated.

$\mu = E(\text{\# of sales}) = 145(0.15) + 150(0.22) + 155(0.37) + 160(0.19) + 170(0.07) = 154.4$

The coffee shop can expect an average of 154.4 daily sales.

Section 15.2

3. Oranges again.

$$\begin{aligned}\sigma^2 &= Var(\text{Oranges})\\ &= (25-33)^2(0.10)+(30-33)^2(0.40)+(35-33)^2(0.30)+(40-33)^2(0.20) = 21\end{aligned}$$

$$\sigma = SD(\text{Oranges}) = \sqrt{Var(\text{Oranges})} \approx \sqrt{21} = 4.58$$

The standard deviation is 4.58 oranges per tree.

4. Caffeinated again.

$$\begin{aligned}\sigma^2 &= Var(\text{\# of sales})\\ &= (145-154.4)^2(0.15)+(150-154.4)^2(0.22)+(155-154.4)^2(0.37)\\ &\quad +(160-154.4)^2(0.19)+(170-154.4)^2(0.07) = 40.64\end{aligned}$$

$$\sigma = SD(\text{\# of sales}) = \sqrt{Var(\text{\# of sales})} = \sqrt{40.64} \approx 6.37$$

The standard deviation is \$6.37 sales per day.

Section 15.3

5. Salary.

$$E(\text{weekday + weekend}) = E(\text{weekday}) + E(\text{weekend}) = 1250 + 450 = \$1750$$

To calculate the standard deviations, we must assume that the weekday and weekend salary expenses are independent.

$$SD(\text{weekday} + \text{weekend}) = \sqrt{Var(\text{weekday}) + Var(\text{weekend})} = \sqrt{129^2 + 57^2} \approx \$141$$

The total weekly salary has a mean of \$1750 and standard deviation of \$141.

6. Golf scores.

$E(18 \text{ holes}) = E(9 \text{ holes}) + E(9 \text{ holes}) = 85 + 85 = 170$

To calculate the standard deviations, we must assume that the scores are independent.

$SD(18 \text{ holes}) = \sqrt{Var(9 \text{ holes}) + Var(9 \text{ holes})} = \sqrt{11^2 + 11^2} \approx 15.6$

The mean and standard deviation for 18 holes are 170 and 15.6.

Section 15.4

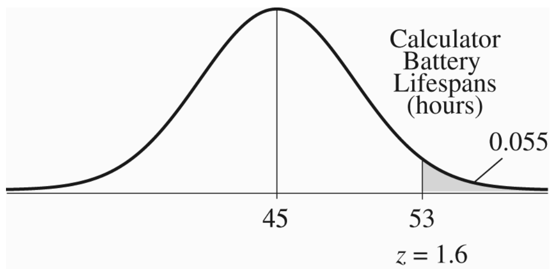
7. Battery.

Let X = the number of hours a calculator battery lasts

$z = \dfrac{x - \mu}{\sigma}$

$z = \dfrac{53 - 45}{5}$

$z = 1.6$



$P(X > 53) = P(z > 1.6) \approx 0.055$

According to the Normal model, the probability that a battery lasts more than 53 hours is approximately 0.055.

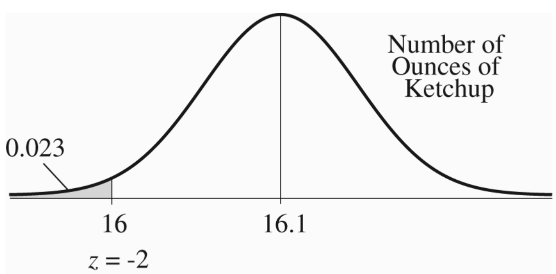
8. Ketchup.

Let X = the number of ounces of ketchup in a bottle

$z = \dfrac{x - \mu}{\sigma}$

$z = \dfrac{16 - 16.1}{0.05}$

$z = -2$



$P(X < 16) = P(z < -2) \approx 0.023$

According to the Normal model, the probability that a bottle of ketchup contains less than 16 ounces is approximately 0.023.

Chapter Exercises

9. Expected value.

a) $\mu = E(Y) = 10(0.3) + 20(0.5) + 30(0.2) = 19$

b) $\mu = E(Y) = 2(0.3) + 4(0.4) + 6(0.2) + 8(0.1) = 4.2$

10. Expected value.

a) $\mu = E(Y) = 0(0.2) + 1(0.4) + 2(0.4) = 1.2$

b) $\mu = E(Y) = 100(0.1) + 200(0.2) + 300(0.5) + 400(0.2) = 280$

11. Pick a card, any card.

a)

Win	\$0	\$5	\$10	\$30
P(amount won)	$\frac{26}{52}$	$\frac{13}{52}$	$\frac{12}{52}$	$\frac{1}{52}$

b) $\mu = E(\text{amount won}) = \$0\left(\frac{26}{52}\right) + \$5\left(\frac{13}{52}\right) + \$10\left(\frac{12}{52}\right) + \$30\left(\frac{1}{52}\right) \approx \$4.13$

c) Answers may vary. In the long run, the expected payoff of this game is \$4.13 per play. Any amount less than \$4.13 would be a reasonable amount to pay in order to play. Your decision should depend on how long you intend to play. If you are only going to play a few times, you should risk less.

12. You bet!

a)

Win	\$100	\$50	\$0
P(amount won)	$\frac{1}{6}$	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = \frac{5}{36}$	$\left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \frac{25}{36}$

b) $\mu = E(\text{amount won}) = \$100\left(\frac{1}{6}\right) + \$50\left(\frac{5}{36}\right) + \$0\left(\frac{25}{26}\right) \approx \$23.61$

c) Answers may vary. In the long run, the expected payoff of this game is \$23.61 per play. Any amount less than \$23.61 would be a reasonable amount to pay in order to play. Your decision should depend on how long you intend to play. If you are only going to play a few times, you should risk less.

13. Kids.

a)

Kids	1	2	3
P(Kids)	0.5	0.25	0.25

b) $\mu = E(\text{Kids}) = 1(0.5) + 2(0.25) + 3(0.25) = 1.75$ kids

c)

Boys	0	1	2	3
P(boys)	0.5	0.25	0.125	0.125

$\mu = E(\text{Boys}) = 0(0.5) + 1(0.25) + 2(0.125) + 3(0.125) = 0.875$ boys

14. Carnival.

a)

Net winnings	\$95	\$90	\$85	\$80	-\$20
# of darts	1 dart	2 darts	3 darts	4 darts (win)	4 darts (lose)
P(Amount won)	$\left(\frac{1}{10}\right) = 0.1$	$\left(\frac{9}{10}\right)\left(\frac{1}{10}\right) = 0.09$	$\left(\frac{9}{10}\right)^2\left(\frac{1}{10}\right) = 0.081$	$\left(\frac{9}{10}\right)^3\left(\frac{1}{10}\right) = 0.0729$	$\left(\frac{9}{10}\right)^4 = 0.6561$

b) $\mu = E(\#\text{ of darts}) = 1(0.1) + 2(0.09) + 3(0.081) + 4(0.0729) + 4(0.6561) \approx 3.44$ darts

c) $\mu = E(\$) = \$95(0.1) + \$90(0.09) + \$85(0.081) + \$80(0.0729) - \$20(0.6561) \approx \$17.20$

15. Software.

Since the company can only get one contract, the probability that they will receive both is 0.

Profit	larger only \$50,000	smaller only \$20,000	both \$70,000	neither \$0
P(profit)	0.30	0.60	0	0.10

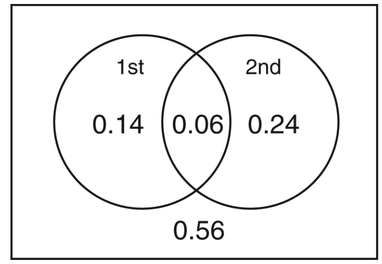
$$\mu = E(\text{profit}) = \$50{,}000(0.30) + \$20{,}000(0.60) + \$70{,}000(0) + \$0(0.10) = \$27{,}000$$

16. Racehorse.

Assuming that the two races are independent events, the probability that the horse wins both races is (0.2)(0.3) = 0.06. Organize the outcomes in a Venn diagram.

Profit	1st only \$30,000	2nd only \$30,000	both \$80,000	neither – \$10,000
P(profit)	0.14	0.24	0.06	0.56

$$\mu = E(\text{profit}) = \$30{,}000(0.14) + \$30{,}000(0.24) + \$80{,}000(0.06) - \$10{,}000(0.56) = \$10{,}600$$



17. Variation 1.

a)

$$\sigma^2 = Var(Y) = (10-19)^2(0.3) + (20-19)^2(0.5) + (30-19)^2(0.2) = 49$$

$$\sigma = SD(Y) = \sqrt{Var(Y)} = \sqrt{49} = 7$$

b)

$$\sigma^2 = Var(Y) = (2-4.2)^2(0.3) + (4-4.2)^2(0.4) + (6-4.2)^2(0.2) + (8-4.2)^2(0.1) = 3.56$$

$$\sigma = SD(Y) = \sqrt{Var(Y)} = \sqrt{3.56} \approx 1.89$$

18. Variation 2.

a)

$$\sigma^2 = Var(Y) = (0-1.2)^2(0.2) + (1-1.2)^2(0.4) + (2-1.2)^2(0.4) = 0.56$$

$$\sigma = SD(Y) = \sqrt{Var(Y)} = \sqrt{0.56} \approx 0.75$$

b)

$$\sigma^2 = Var(Y) = (100-280)^2(0.1) + (200-280)^2(0.2) + (300-280)^2(0.5) + (400-280)^2(0.2) = 7600$$

$$\sigma = SD(Y) = \sqrt{Var(Y)} = \sqrt{7600} \approx 87.18$$

19. Pick another card.

Answers may vary slightly (due to rounding of the mean)

$$\sigma^2 = Var(\text{Won}) = (0-4.13)^2\left(\frac{26}{52}\right) + (5-4.13)^2\left(\frac{13}{52}\right) + (10-4.13)^2\left(\frac{12}{52}\right) + (30-4.13)^2\left(\frac{1}{52}\right) \approx 29.5396$$

$$\sigma = SD(\text{Won}) = \sqrt{Var(\text{Won})} = \sqrt{29.5396} \approx \$5.44$$

20. The die.

Answers may vary slightly (due to rounding of the mean)

$$\sigma^2 = Var(\text{Won}) = (100-23.61)^2\left(\frac{1}{6}\right) + (50-23.61)^2\left(\frac{5}{36}\right) + (0-23.61)^2\left(\frac{25}{36}\right) \approx 1456.4043$$

$$\sigma = SD(\text{Won}) = \sqrt{Var(\text{Won})} \approx \sqrt{1456.4043} \approx \$38.16$$

21. Kids.

$$\sigma^2 = Var(\text{Kids}) = (1-1.75)^2(0.5) + (2-1.75)^2(0.25) + (3-1.75)^2(0.25) = 0.6875$$

$$\sigma = SD(\text{Kids}) = \sqrt{Var(\text{Kids})} = \sqrt{0.6875} \approx 0.83 \text{ kids}$$

22. Darts.

$$\sigma^2 = Var(\text{Winnings}) = (95-17.20)^2(0.1) + (90-17.20)^2(0.09) + (85-17.20)^2(0.081) + (80-17.20)^2(0.0729) + (-20-17.20)^2(0.6561) \approx 2650.057$$

$$\sigma = SD(\text{Winnings}) = \sqrt{Var(\text{Winnings})} \approx \sqrt{2650.057} \approx \$51.48$$

23. Repairs.

a) $\mu = E(\text{Number of Repair Calls}) = 0(0.1) + 1(0.3) + 2(0.4) + 3(0.2) = 1.7 \text{ calls}$

b)

$$\sigma^2 = Var(\text{Calls}) = (0-1.7)^2(0.1) + (1-1.7)^2(0.3) + (2-1.7)^2(0.4) + (3-1.7)^2(0.2) = 0.81$$

$$\sigma = SD(\text{Calls}) = \sqrt{Var(\text{Calls})} = \sqrt{0.81} = 0.9 \text{ calls}$$

24. Red lights.

a) $\mu = E(\text{Red lights}) = 0(0.05) + 1(0.25) + 2(0.35) + 3(0.15) + 4(0.15) + 5(0.05) = 2.25 \text{ red lights}$

b)

$$\sigma^2 = Var(\text{Red lights}) = (0-2.25)^2(0.05) + (1-2.25)^2(0.25) + (2-2.25)^2(0.35) + (3-2.25)^2(0.15) + (4-2.25)^2(0.15) + (5-2.25)^2(0.05) = 1.5875$$

$$\sigma = SD(\text{Red lights}) = \sqrt{Var(\text{Red lights})} = \sqrt{1.5875} \approx 1.26 \text{ red lights}$$

25. Defects.

The percentage of cars with *no* defects is 61%.

$\mu = E(\text{Defects}) = 0(0.61) + 1(0.21) + 2(0.11) + 3(0.07) = 0.64$ defects

$$\sigma^2 = Var(\text{Defects}) = (0-0.64)^2(0.61) + (1-0.64)^2(0.21) + (2-0.64)^2(0.11) + (3-0.64)^2(0.07) \approx 0.8704$$

$\sigma = SD(\text{Defects}) = \sqrt{Var(\text{Defects})} \approx \sqrt{0.8704} \approx 0.93$ defects

26. Insurance.

a)

Profit	\$100	– \$9900	– \$2900
P(Profit)	0.9975	0.0005	0.002

b) $\mu = E(\text{Profit}) = 100(0.9975) - 9900(0.0005) - 2900(0.002) = \$89$

c)

$$\sigma^2 = Var(\text{Profit}) = (100-89)^2(0.9975) + (-9900-89)^2(0.0005) + (-2900-89)^2(0.002) = 67{,}879$$

$\sigma = SD(\text{Profit}) = \sqrt{Var(\text{Profit})} \approx \sqrt{67{,}879} \approx \$260.54$

27. Cancelled flights.

a) $\mu = E(\text{gain}) = (-150)(0.20) + 100(0.80) = \$50$

b)

$\sigma^2 = Var(\text{gain}) = (-150-50)^2(0.20) + (100-50)^2(0.80) = 10{,}000$

$\sigma = SD(\text{gain}) = \sqrt{Var(\text{gain})} \approx \sqrt{10{,}000} = \$100$

28. Day trading.

a) $\mu = E(\text{stock option}) = 1000(0.20) + 0(0.30) + 200(0.50) = \$300$

The trader should buy the stock option. Its expected value is \$300, and she only has to pay \$200 for it.

b) $\mu = E(\text{gain}) = 800(0.20) + (-200)(0.30) + 0(0.50) = \$100$

The trader expects to gain \$100. Notice that this is the same result as subtracting the \$200 price of the stock option from the \$300 expected value.

c) $$\sigma^2 = Var(\text{gain}) = (800-100)^2(0.20) + (-200-100)^2(0.30) + (0-100)^2(0.50) = 130{,}000$$

$\sigma = SD(\text{gain}) = \sqrt{Var(\text{gain})} \approx \sqrt{130{,}000} \approx \$360.56$

Notice that the standard deviation of the trader's gain is the same as the standard deviation in value of the stock option.

29. Contest.

a) The two games are not independent. The probability that you win the second depends on whether or not you win the first.

b)

$$\begin{aligned} P(\text{losing both games}) &= P(\text{losing the first})\, P(\text{losing the second} \mid \text{first was lost}) \\ &= (0.6)(0.7) = 0.42 \end{aligned}$$

c)

$$\begin{aligned} P(\text{winning both games}) &= P(\text{winning the first})\, P(\text{winning the second} \mid \text{first was won}) \\ &= (0.4)(0.2) = 0.08 \end{aligned}$$

d)

X	0	1	2
$P(X = x)$	0.42	0.50	0.08

e)

$$\mu = E(X) = 0(0.42) + 1(0.50) + 2(0.08) = 0.66 \text{ games}$$

$$\sigma^2 = Var(X) = (0 - 0.66)^2(0.42) + (1 - 0.66)^2(0.50) + (2 - 0.66)^2(0.08) = 0.3844$$

$$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{0.3844} \approx 0.62 \text{ games}$$

30. Contracts.

a) The contracts are not independent. The probability that your company wins the second contract depends on whether or not your company wins the first contract.

b)

$$\begin{aligned} P(\text{getting both contracts}) &= P(\text{getting \#1})\, P(\text{getting \#2} \mid \text{got \#1}) \\ &= (0.8)(0.2) \\ &= 0.16 \end{aligned}$$

c)

$$\begin{aligned} P(\text{getting no contract}) &= P(\text{not getting \#1})\, P(\text{not getting \#2} \mid \text{didn't get \#1}) \\ &= (0.2)(0.7) \\ &= 0.14 \end{aligned}$$

d)

X	0	1	2
$P(X = x)$	0.14	0.70	0.16

e) $\mu = E(X) = 0(0.14) + 1(0.70) + 2(0.16) = 1.02$ contracts

$\sigma^2 = Var(X) = (0-1.02)^2(0.14) + (1-1.02)^2(0.70) + (2-1.02)^2(0.16) = 0.2996$

$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{0.2996} \approx 0.55$ contracts

31. Batteries.

a)

# good	0	1	2
$P(\text{\# good})$	$\left(\frac{3}{10}\right)\left(\frac{2}{9}\right) = \frac{6}{90}$	$\left(\frac{3}{10}\right)\left(\frac{7}{9}\right) + \left(\frac{7}{10}\right)\left(\frac{3}{9}\right) = \frac{42}{90}$	$\left(\frac{7}{10}\right)\left(\frac{6}{9}\right) = \frac{42}{90}$

b) $\mu = E(\text{number good}) = 0\left(\frac{6}{90}\right) + 1\left(\frac{42}{90}\right) + 2\left(\frac{42}{90}\right) = 1.4$ batteries

c)

$\sigma^2 = Var(\text{number good}) = (0-1.4)^2\left(\frac{6}{90}\right) + (1-1.4)^2\left(\frac{42}{90}\right) + (2-1.4)^2\left(\frac{42}{90}\right) \approx 0.3733$

$\sigma = SD(\text{number good}) = \sqrt{Var(\text{number good})} \approx \sqrt{0.3733} \approx 0.61$ batteries.

32. Kittens.

a)

Number of males	0	1	2
$P(\text{number of males})$	$\left(\frac{3}{7}\right)\left(\frac{2}{6}\right) = \frac{6}{42}$	$\left(\frac{4}{7}\right)\left(\frac{3}{6}\right) + \left(\frac{3}{7}\right)\left(\frac{4}{6}\right) = \frac{24}{42}$	$\left(\frac{4}{7}\right)\left(\frac{3}{6}\right) = \frac{12}{42}$

b) $\mu = E(\text{number of males}) = 0\left(\frac{6}{42}\right) + 1\left(\frac{24}{42}\right) + 2\left(\frac{12}{42}\right) \approx 1.14$ males

c) Answers may vary slightly (due to rounding of the mean)

$$\sigma^2 = Var(\text{number of males}) = (0-1.14)^2\left(\frac{6}{42}\right) + (1-1.14)^2\left(\frac{24}{42}\right)$$
$$+ (2-1.14)^2\left(\frac{12}{42}\right) \approx 0.4082$$

$\sigma = SD(\text{number of males}) = \sqrt{Var(\text{number of males})} \approx \sqrt{0.4082} \approx 0.64$ males

33. Random variables.

a)

$$\mu = E(3X) = 3(E(X)) = 3(10) = 30$$
$$\sigma = SD(3X) = 3(SD(X)) = 3(2) = 6$$

b)

$$\mu = E(Y+6) = E(Y) + 6 = 20 + 6 = 26$$
$$\sigma = SD(Y+6) = SD(Y) = 5$$

c)

$$\begin{aligned}\mu &= E(X+Y)\\ &= E(X) + E(Y) = 10 + 20 = 30\\ \sigma &= SD(X+Y) = \sqrt{Var(X) + Var(Y)}\\ &= \sqrt{2^2 + 5^2} \approx 5.39\end{aligned}$$

d)

$$\begin{aligned}\mu &= E(X-Y) = E(X) - E(Y)\\ &= 10 - 20 = -10\\ \sigma &= SD(X-Y) = \sqrt{Var(X) + Var(Y)}\\ &= \sqrt{2^2 + 5^2} \approx 5.39\end{aligned}$$

e)

$$\mu = E(X_1 + X_2) = E(X) + E(X) = 10 + 10 = 20$$
$$\begin{aligned}\sigma &= SD(X_1 + X_2) = \sqrt{Var(X) + Var(X)}\\ &= \sqrt{2^2 + 2^2} \approx 2.83\end{aligned}$$

34. Random variables.

a)

$$\begin{aligned}\mu &= E(X-20) = E(X) - 20\\ &= 80 - 20 = 60\\ \sigma &= SD(X-20) = SD(X) = 12\end{aligned}$$

b)

$$\begin{aligned}\mu &= E(0.5Y) = 0.5(E(Y))\\ &= 0.5(12) = 6\\ \sigma &= SD(0.5Y) = 0.5(SD(Y))\\ &= 0.5(3) = 1.5\end{aligned}$$

c)

$$\begin{aligned}\mu &= E(X+Y) = E(X) + E(Y)\\ &= 80 + 12 = 92\\ \sigma &= SD(X+Y) = \sqrt{Var(X) + Var(Y)}\\ &= \sqrt{12^2 + 3^2} \approx 12.37\end{aligned}$$

d)

$$\begin{aligned}\mu &= E(X-Y) = E(X) - E(Y)\\ &= 80 - 12 = 68\\ \sigma &= SD(X-Y) = \sqrt{Var(X) + Var(Y)}\\ &= \sqrt{12^2 + 3^2} \approx 12.37\end{aligned}$$

e)

$$\begin{aligned}\mu &= E(Y_1 + Y_2) = E(Y) + E(Y)\\ &= 12 + 12 = 24\\ \sigma &= SD(Y_1 + Y_2) = \sqrt{Var(Y) + Var(Y)}\\ &= \sqrt{3^2 + 3^2} \approx 4.24\end{aligned}$$

35. Random variables.

a)
$$\mu = E(0.8Y) = 0.8(E(Y)) = 0.8(300) = 240$$
$$\sigma = SD(0.8Y) = 0.8(SD(Y))$$
$$= 0.8(16) = 12.8$$

b)
$$\mu = E(2X - 100) = 2(E(X)) - 100 = 140$$
$$\sigma = SD(2X - 100)$$
$$= 2(SD(X)) = 2(12) = 24$$

c)
$$\mu = E(X + 2Y) = E(X) + 2(E(Y))$$
$$= 120 + 2(300) = 720$$
$$\sigma = SD(X + 2Y) = \sqrt{Var(X) + 2^2 Var(Y)}$$
$$= \sqrt{12^2 + 2^2(16^2)} \approx 34.18$$

d)
$$\mu = E(3X - Y) = 3(E(X)) - E(Y)$$
$$= 3(120) - 300 = 60$$
$$\sigma = SD(3X - Y)$$
$$= \sqrt{3^2 Var(X) + Var(Y)}$$
$$= \sqrt{3^2(12^2) + 16^2} \approx 39.40$$

e)
$$\mu = E(Y_1 + Y_2) = E(Y) + E(Y) = 300 + 300 = 600$$
$$\sigma = SD(Y_1 + Y_2) = \sqrt{Var(Y) + Var(Y)}$$
$$= \sqrt{16^2 + 16^2} \approx 22.63$$

36. Random variables.

a)
$$\mu = E(2Y + 20) = 2(E(Y)) + 20$$
$$= 2(12)+20 = 44$$
$$\sigma = SD(2Y+20) = 2(SD(Y)) = 2(3) = 6$$

b)
$$\mu = E(3X) = 3(E(X)) = 3(80) = 240$$
$$\sigma = SD(3X) = 3(SD(X)) = 3(12) = 36$$

c)
$$\mu = E(0.25X + Y)$$
$$= 0.25(E(X)) + E(Y)$$
$$= 0.25(80)+12 = 32$$
$$\sigma = SD(0.25X + Y)$$
$$= \sqrt{0.25^2 Var(X) + Var(Y)}$$
$$= \sqrt{0.25^2(12^2) + 3^2} \approx 4.24$$

d)
$$\mu = E(X - 5Y)$$
$$= E(X) - 5(E(Y))$$
$$= 80 - 5(12) = 20$$
$$\sigma = SD(X - 5Y)$$
$$= \sqrt{Var(X) + 5^2 Var(Y)}$$
$$= \sqrt{12^2 + 5^2(3^2)} \approx 19.21$$

e)
$$\mu = E(X_1 + X_2 + X_3) = E(X) + E(X) + E(X) = 80 + 80 + 80 = 240$$
$$\sigma = SD(X_1 + X_2 + X_3) = \sqrt{Var(X_1) + Var(X_2) + Var(X_3)}$$
$$= \sqrt{12^2 + 12^2 + 12^2} \approx 20.78$$

37. Eggs.

a) $\mu = E(\text{Broken eggs in 3 doz.}) = 3(E(\text{Broken eggs in 1 doz.})) = 3(0.6) = 1.8$ eggs

b) $\sigma = SD(\text{Broken eggs in 3 dozen}) = \sqrt{0.5^2 + 0.5^2 + 0.5^2} \approx 0.87$ eggs

c) The cartons of eggs must be independent of each other.

38. Garden.

a) $\mu = E(\text{bad seeds in 5 packets}) = 5(E(\text{bad seeds in 1 packet})) = 5(2) = 10$ bad seeds

b) $\sigma = SD(\text{bad seeds in 5 packets}) = \sqrt{1.2^2 + 1.2^2 + 1.2^2 + 1.2^2 + 1.2^2} \approx 2.68$ bad seeds

c) The packets of seeds must be independent of each other. If you buy an assortment of seeds, this assumption is probably OK. If you buy all of one type of seed, the store probably has seed packets from the same batch or lot. If some are bad, the others might tend to be bad as well.

39. Repair calls.

$\mu = E(\text{calls in 8 hours}) = 8(E(\text{calls in 1 hour}) = 8(1.7) = 13.6$ calls

$\sigma = SD(\text{calls in 8 hours}) = \sqrt{8(Var(\text{calls in 1 hour}))} = \sqrt{8(0.9)^2} \approx 2.55$ calls

This is only valid if the hours are independent of one another.

40. Stop!

$$\mu = E(\text{red lights in 5 days})$$
$$= 5(E(\text{red lights each day})) = 5(2.25) = 11.25 \text{ red lights}$$
$$\sigma = SD(\text{red lights in 5 days})$$
$$= \sqrt{5(Var(\text{red lights each day})} = \sqrt{5(1.26)^2} \approx 2.82 \text{ red lights}$$

Standard deviation may vary slightly due to rounding of the standard deviation of the number of red lights each day, and may only be calculated if the days are independent of each other. This seems reasonable.

41. Tickets.

a)

$$\mu = E(\text{tickets for 18 trucks})$$
$$= 18(E(\text{tickets for one truck})) = 18(1.3) = 23.4 \text{ tickets}$$
$$\sigma = SD(\text{tickets for 18 trucks})$$
$$= \sqrt{18(Var(\text{tickets for one truck})} = \sqrt{18(0.7)^2} \approx 2.97 \text{ tickets}$$

b) We are assuming that trucks are ticketed independently.

42. Donations.

a)

$$\mu = E(\text{pledges from 120 people})$$
$$= 120(E(\text{pledge from one person})) = 120(32) = \$3840$$
$$\sigma = SD(\text{pledges from 120 people})$$
$$= \sqrt{120(Var(\text{pledge from one person})} = \sqrt{120(54)^2} \approx \$591.54$$

b) We are assuming that callers make pledges independently from one another.

43. Fire!

a) The standard deviation is large because the profits on insurance are highly variable. Although there will be many small gains, there will occasionally be large losses, when the insurance company has to pay a claim.

b)

$$\mu = E(\text{two policies}) = 2(E(\text{one policy})) = 2(150) = \$300$$
$$\sigma = SD(\text{two policies}) = \sqrt{2(Var(\text{one policy}))} = \sqrt{2(6000^2)} \approx \$8,485.28$$

c)

$$\mu = E(10,000 \text{ policies}) = 10,000(E(\text{one policy})) = 10,000(150) = \$1,500,000$$
$$\sigma = SD(10,000 \text{ policies}) = \sqrt{10,000(Var(\text{one policy}))} = \sqrt{10,000(6000^2)} = \$600,000$$

d) If the company sells 10,000 policies, they are likely to be successful. A profit of \$0, is 2.5 standard deviations below the expected profit. This is unlikely to happen. However, if the company sells fewer policies, then the likelihood of turning a profit decreases. In an extreme case, where only two policies are sold, a profit of \$0 is more likely, being only a small fraction of a standard deviation below the mean.

e) This analysis depends on each of the policies being independent from each other. This assumption of independence may be violated if there are many fire insurance claims as a result of a forest fire, or other natural disaster.

44. Casino.

a) The standard deviation of the slot machine payouts is large because most players will lose their dollar, but a few large payouts are expected. The payouts are highly variable.

b)

$$\mu = E(\text{profit from 5 plays}) = 5(E(\text{profit from one play})) = 5(0.08) = \$0.40$$
$$\sigma = SD(\text{profit from 5 plays}) = \sqrt{5(Var(\text{profit from one play}))} = \sqrt{5(120^2)} \approx \$268.33$$

c)

$$\mu = E(\text{profit from 1000 plays})$$
$$= 1000(E(\text{profit from one play})) = 1000(0.08) = \$80$$
$$\sigma = SD(\text{profit from 1000 plays})$$
$$= \sqrt{1000(Var(\text{profit from one play}))} = \sqrt{1000(120^2)} \approx \$3{,}794.73$$

d) If the machine is played only 1000 times a day, the chance of being profitable isn't as high as the casino might like, since \$80 is only approximately 0.02 standard deviations above 0. But if the casino has many slot machines, the chances of being profitable will go up.

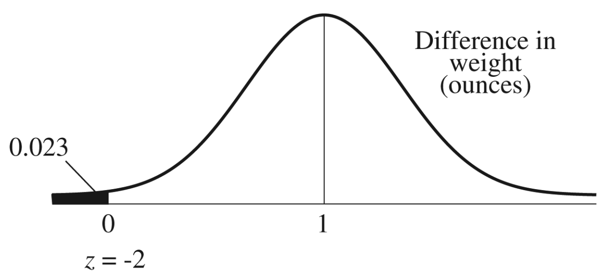
45. Cereal.

a) $E(\text{large bowl} - \text{small bowl}) = E(\text{large bowl}) - E(\text{small bowl}) = 2.5 - 1.5 = 1$ ounce

b) $\sigma = SD(\text{large} - \text{small}) = \sqrt{Var(\text{large}) + Var(\text{small})} = \sqrt{0.4^2 + 0.3^2} = 0.5$ ounces

c)

$$z = \frac{x-\mu}{\sigma}$$
$$z = \frac{0-1}{0.5} = -2$$



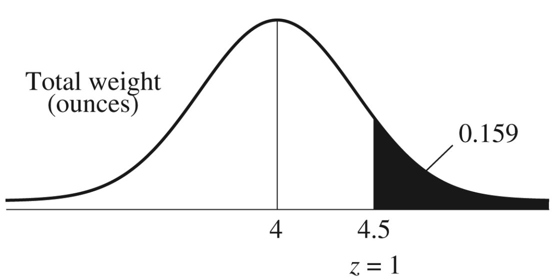
The small bowl will contain more cereal than the large bowl when the difference between the amounts is less than 0. According to the Normal model, the probability of this occurring is approximately 0.023.

d)

$$\mu = E(\text{large} + \text{small}) = E(\text{large}) + E(\text{small}) = 2.5 + 1.5 = 4 \text{ ounce}$$
$$\sigma = SD(\text{large} + \text{small}) = \sqrt{Var(\text{large}) + Var(\text{small})} = \sqrt{0.4^2 + 0.3^2} = 0.5 \text{ ounces}$$

e)

$$z = \frac{x-\mu}{\sigma}$$
$$z = \frac{4.5-4}{0.5} = 1$$



According to the Normal model, the probability that the total weight of cereal in the two bowls is more than 4.5 ounces is approximately 0.159.

f)

$$\mu = E(\text{box} - \text{large} - \text{small}) = E(\text{box}) - E(\text{large}) - E(\text{small})$$
$$= 16.3 - 2.5 - 1.5 = 12.3 \text{ ounces}$$
$$\sigma = SD(\text{box} - \text{large} - \text{small}) = \sqrt{Var(\text{box}) + Var(\text{large}) + Var(\text{small})}$$
$$= \sqrt{0.2^2 + 0.3^2 + 0.4^2} \approx 0.54 \text{ ounces}$$

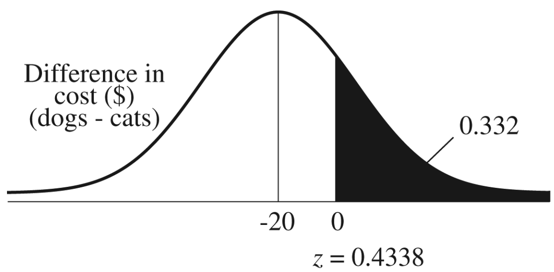
46. Pets.

a) $\mu = E(\text{dogs} - \text{cats}) = E(\text{dogs}) - E(\text{cats}) = 100 - 120 = -\$20$

b) $\sigma = SD(\text{dogs} - \text{cats}) = \sqrt{Var(\text{dogs}) + Var(\text{cats})} = \sqrt{30^2 + 35^2} \approx \$46.10$

c)

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{0 - (-20)}{46.10}$$
$$z = 0.4338$$



The expected cost of the dog is greater than that of the cat when the difference in cost is positive (greater than 0). According to the Normal model, the probability of this occurring is about 0.332.

d) Costs for pets living together may not be independent.

47. More cereal.

a)

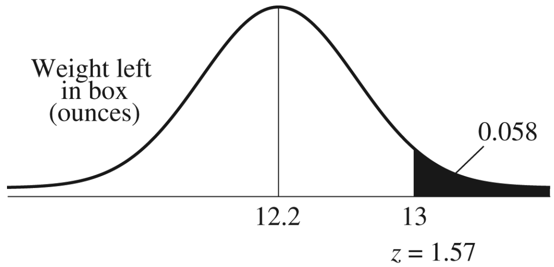
$$\mu = E(\text{box} - \text{large} - \text{small})$$
$$= E(\text{box}) - E(\text{large}) - E(\text{small}) = 16.2 - 2.5 - 1.5 = 12.2 \text{ ounces}$$

b)

$$\sigma = SD(\text{box} - \text{large} - \text{small}) = \sqrt{Var(\text{box}) + Var(\text{large}) + Var(\text{small})}$$
$$= \sqrt{0.1^2 + 0.3^2 + 0.4^2} \approx 0.51 \text{ ounces}$$

c)

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{13 - 12.2}{0.51}$$
$$z = 1.57$$



According to the Normal model, the probability that the box contains more than 13 ounces is about 0.058.

48. More pets.

a) Let $X =$ cost for a dog, and let $Y =$ cost for a cat.

Total cost $= X + X + Y$

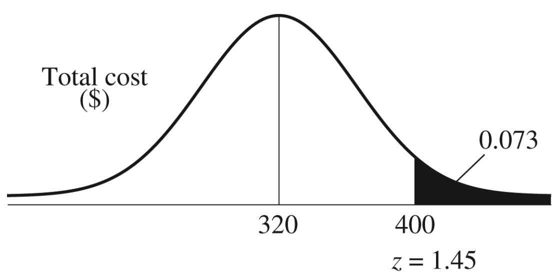
b)

$$\mu = E(X + X + Y) = E(X) + E(X) + E(Y) = 100 + 100 + 120 = \$320$$
$$\sigma = SD(X + X + Y) = \sqrt{Var(X) + Var(X) + Var(Y)}$$
$$= \sqrt{30^2 + 30^2 + 35^2} = \$55$$

Since the models for individual pets are Normal, the model for total costs is Normal with mean \$320 and standard deviation \$55.

c)

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{400 - 320}{55}$$
$$z = 1.45$$



According to the Normal model, the probability that the total cost of two dogs and a cat is more than \$400 is approximately 0.073.

49. Medley.

a)

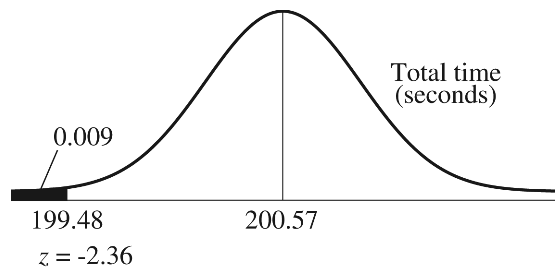
$$\mu = E(\#1 + \#2 + \#3 + \#4) = E(\#1) + E(\#2) + E(\#3) + E(\#4)$$
$$= 50.72 + 55.51 + 49.43 + 44.91 = 200.57 \text{ seconds}$$
$$\sigma = SD(\#1 + \#2 + \#3 + \#4) = \sqrt{Var(\#1) + Var(\#2) + Var(\#3) + Var(\#4)}$$
$$= \sqrt{0.24^2 + 0.22^2 + 0.25^2 + 0.21^2} \approx 0.46 \text{ seconds}$$

b)

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{199.48 - 200.57}{0.461}$$

$$z = -2.36$$



The team is not likely to swim faster than their best time. According to the Normal model, they are only expected to swim that fast or faster about 0.9% of the time.

50. Bikes.

a)

$$\begin{aligned}\mu &= E(\text{unpack} + \text{assembly} + \text{tuning}) \\ &= E(\text{unpack}) + E(\text{assembly}) + E(\text{tuning}) \\ &= 3.5 + 21.8 + 12.3 = 37.6 \text{ minutes}\end{aligned}$$

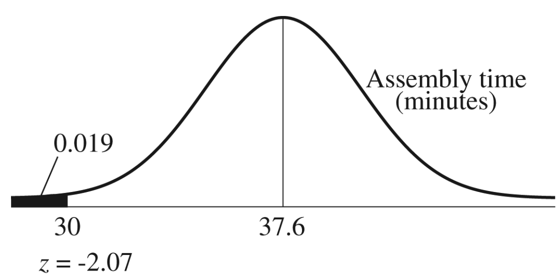
$$\begin{aligned}\sigma &= SD(\text{unpack} + \text{assembly} + \text{tuning}) \\ &= \sqrt{Var(\text{unpack}) + Var(\text{assembly}) + Var(\text{tuning})} \\ &= \sqrt{0.7^2 + 2.4^2 + 2.7^2} \approx 3.7 \text{ minutes}\end{aligned}$$

b)

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{30 - 37.6}{3.68}$$

$$z = -2.07$$



The bike is not likely to be ready on time. According to the Normal model, the probability that an assembly is completed in under 30 minutes is about 0.019.

51. Farmer's market.

a) Let $A =$ price of a pound of apples, and let $P =$ price of a pound of potatoes.

Profit $= 100A + 50P - 2$

b) $\mu = E(100A + 50P - 2) = 100(E(A)) + 50(E(P)) - 2 = 100(0.5) + 50(0.3) - 2 = \$63$

c)

$$\begin{aligned}\sigma &= SD(100A + 50P - 2) \\ &= \sqrt{100^2(Var(A)) + 50^2(Var(P))} = \sqrt{100^2(0.2^2) + 50^2(0.1^2)} \approx \$20.62\end{aligned}$$

d) No assumptions are necessary to compute the mean. To compute the standard deviation, independent market prices must be assumed.

52. Bike sale.

a) Let $B =$ number of basic bikes sold, and let $D =$ number of deluxe bikes sold.

Net Profit $= 120B + 150D - 200$

b)

$$\mu = E(120B + 150D - 200)$$
$$= 120(E(B)) + 150(E(D)) - 200 = 120(5.4) + 150(3.2) - 200 = \$928$$

c)

$$\sigma = SD(120B + 150D - 200) = \sqrt{120^2(Var(B)) + 150^2(Var(D))}$$
$$= \sqrt{120^2(1.2^2) + 150^2(0.8^2)} \approx \$187.45$$

d) No assumptions are necessary to compute the mean. To compute the standard deviation, independent sales must be assumed.

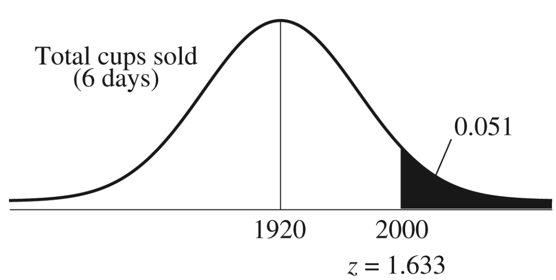
53. Coffee and doughnuts.

a)

$$\mu = E(\text{cups sold in 6 days}) = 6(E(\text{cups sold in 1 day})) = 6(320) = 1920 \text{ cups}$$
$$\sigma = SD(\text{cups sold in 6 days}) = \sqrt{6(Var(\text{cups sold in 1 day})} = \sqrt{6(20)^2} \approx 48.99 \text{ cups}$$

The distribution of total coffee sales for 6 days has distribution $N(1920, 48.99)$.

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{2000 - 1920}{48.99}$$
$$z = 1.633$$



According to the Normal model, the probability that he will sell more than 2000 cups of coffee in a week is approximately 0.051.

b) Let C = the number of cups of coffee sold. Let D = the number of doughnuts sold.

$$\mu = E(0.50C + 0.40D) = 0.50(E(C)) + 0.40(E(D)) = 0.50(320) + 0.40(150) = \$220$$

$$\sigma = SD(0.50C + 0.40D)$$
$$= \sqrt{0.50^2(Var(C)) + 0.40^2(Var(D))} = \sqrt{0.50^2(20^2) + 0.40^2(12^2)} \approx \$11.09$$

The day's profit can be modeled by $N(220, 11.09)$. A day's profit of \$300 is over 7 standard deviations above the mean. This is extremely unlikely. It would not be reasonable for the shop owner to expect the day's profit to exceed \$300.

c) Consider the difference $D - 0.5C$. When this difference is greater than zero, the number of doughnuts sold is greater than half the number of cups of coffee sold.

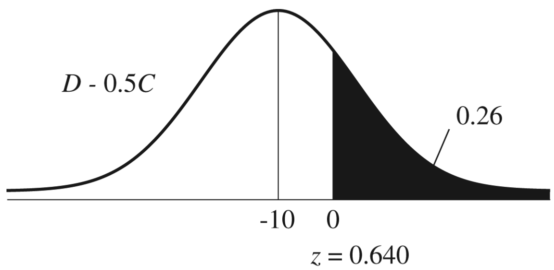
$$\mu = E(D - 0.5C) = (E(D)) - 0.5(E(C)) = 150 - 0.5(320) = -\$10$$

$$\sigma = SD(D - 0.5C) = \sqrt{(Var(D)) + 0.5(Var(C))} = \sqrt{(12^2) + 0.5^2(20^2)} \approx \$15.62$$

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{0 - (-10)}{15.62}$$

$$z = 0.640$$



The difference $D - 0.5C$ can be modeled by $N(-10, 15.62)$.

According to the Normal model, the probability that the shop owner will sell a doughnut to more than half of the coffee customers is approximately 0.26.

54. Weightlifting.

a) Let T = the true weight of a 20-pound weight.
Let F = the true weight of a 5-pound weight.
Let B = the true weight of the bar.

$$\mu = E(\text{Total weight}) = 2E(T) + 4E(F) = 2(20) + 4(5) = 60 \text{ pounds}$$

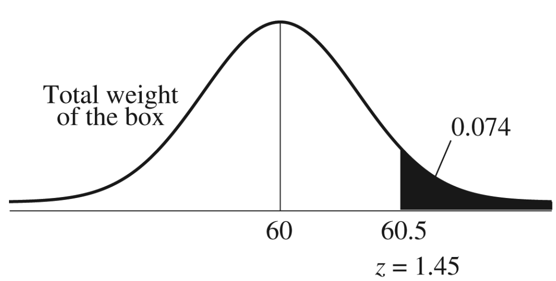
$$\sigma = SD(\text{Total}) = \sqrt{2(Var(T)) + 4(Var(F))} = \sqrt{2(0.2^2) + 4(0.1^2)} = \sqrt{0.12} \approx 0.346 \text{ pounds}$$

Assuming that the true weights of each piece are independent of one another, the total weight of the box can be modeled by $N(60, 0.346)$.

$$z = \frac{x - \mu}{\sigma} \mathfrak{A}$$

$$z = \frac{60.5 - 60}{0.346}$$

$$z = 1.45$$



According to the Normal model, the probability that the total weight in the shipping box exceeds 60.5 pounds is approximately 0.074.

b) $\text{Cost} = 0.40(T+T+F+F+F+F)+0.50(B)+6$

$$\mu = E(Cost) = 0.40(E(T+T+F+F+F+F))+0.50(\text{E}(B))+6$$
$$= 040(60)+0.50(10)+6 = \$35$$

$$\sigma = SD(\text{Cost}) = \sqrt{0.40^2(Var(\text{Total weight of box}))+0.50^2(Var(B))}$$
$$= \sqrt{0.40^2(0.12)+0.50^2(0.25^2)} \approx \$0.187$$

The shipping cost for the starter set has mean \$35 and standard deviation \$0.187.

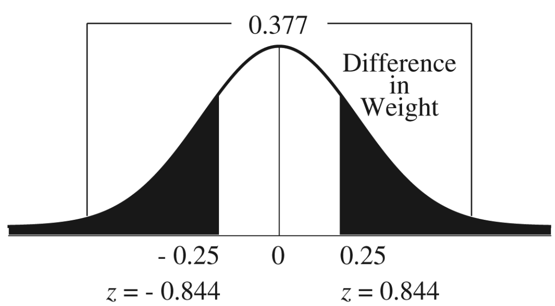
c) Consider the expression $T-(F+F+F+F)$, which represents the difference in weight between a 20-pound weight and four 5-pound weights. We are interested in the distribution of this difference, so that we can find the probability that the difference is greater than 0.25 pounds.

$$\mu = E(T-(F+F+F+F)) = E(T)-4E(F) = 20-4(5) = 0 \text{ pounds}$$

$$\sigma = SD(T-(F+F+F+F))$$
$$= \sqrt{(Var(T))+4(Var(F))} = \sqrt{(0.2^2)+4(0.1^2)} \approx 0.2828 \text{ pounds}$$

The difference in weight can be modeled by $N(0, 0.2828)$.

$$z = \frac{x-\mu}{\sigma}$$
$$z = \frac{0.25-0}{0.2828}$$
$$z = 0.884$$



According to the Normal model, the probability that the difference in weight between one 20-pound weight and four 5-pound weights is greater than 0.25 pounds is 0.377.

Chapter 16 – Probability Models

Section 16.1

1. Bernoulli.

a) These are not Bernoulli trials. The possible outcomes are 1, 2, 3, 4, 5, and 6. There are more than two possible outcomes.

b) These may be considered Bernoulli trials. There are only two possible outcomes, Type A and not Type A. Assuming the 120 donors are representative of the population, the probability of having Type A blood is 43%. The trials are not independent, because the population is finite, but the 120 donors represent less than 10% of all possible donors.

c) These are not Bernoulli trials. The probability of getting a heart changes as cards are dealt without replacement.

d) These are not Bernoulli trials. We are sampling without replacement, so the trials are not independent. Samples without replacement may be considered Bernoulli trials if the sample size is less than 10% of the population, but 500 is more than 10% of 3000.

e) These may be considered Bernoulli trials. There are only two possible outcomes, sealed properly and not sealed properly. The probability that a package is unsealed is constant, at about 10%, as long as the packages checked are a representative sample of all packages. Finally, the trials are not independent, since the total number of packages is finite, but the 24 packages checked probably represent less than 10% of the packages.

2. Bernoulli 2.

a) These may be considered Bernoulli trials. There are only two possible outcomes, getting a 6 and not getting a 6. The probability of getting a 6 is constant at 1/6. The rolls are independent of one another, since the outcome of one die roll doesn't affect the other rolls.

b) These are not Bernoulli trials. There are more than two possible outcomes for eye color.

c) These can be considered Bernoulli trials. There are only two possible outcomes, properly attached buttons and improperly attached buttons. As long as the button problem occurs randomly, the probability of a doll having improperly attached buttons is constant at about 3%. The trails are not independent, since the total number of dolls is finite, but 37 dolls is probably less than 10% of all dolls.

d) These are not Bernoulli trials. The trials are not independent, since the probability of picking a council member with a particular political affiliation changes depending on who has already been picked. The 10% condition is not met, since the sample of size 4 is more than 10% of the population of 19 people.

e) These may be considered Bernoulli trials. There are only two possible outcomes, cheating and not cheating. Assuming that cheating patterns in this school are similar to the patterns in the nation, the probability that a student has cheated is constant, at 74%. The trials are not independent, since the population of all students is finite, but 481 is less than 10% of all students.

Section 16.2

3. Toasters.

Let X = the number of toasters that need repair.

The condition of the toasters can be considered Bernoulli trials. There are only two possible outcomes, needing to be sent back for repairs and not needing to be sent back for repairs. The probability that a toaster needs repair is constant, given as $p = 0.05$. The trials are not independent, since there are a finite number of toasters, but 20 toasters in each carton is less than 10% of all toasters produced.

The distribution of the number of repairs required follows *Binom*(20, 0.05).

$$\begin{aligned} P(\text{exactly 3 toasters need repair}) &= P(X=3) \\ &= {}_{20}C_3(0.05)^3(0.95)^{17} \\ &\approx 0.0596 \end{aligned}$$

According to the Binomial model, the probability that exactly three toasters out of twenty require repair is 0.0596.

4. Soccer.

Let X = the number of goals scored on corner kicks.

The outcome of the corner kicks can be considered Bernoulli trials. There are only two possible outcomes, scoring a goal or not scoring a goal. The probability of scoring a goal is constant at $p = 0.08$. We will assume that the corner kicks are taken independently from each other, and the outcome of one kick will not affect another kick.

The distribution of the number of goals scored follows *Binom*(15, 0.08).

$$\begin{aligned} P(\text{exactly 2 goals scored}) &= P(X=2) \\ &= {}_{15}C_2(0.08)^2(0.92)^{13} \\ &\approx 0.227 \end{aligned}$$

According to the Binomial model, the probability that exactly two corner kicks are made out of fifteen attempts is 0.227.

Section 16.3

5. Toasters again.

A binomial model and a normal model are both appropriate for modeling the number of toasters that need to be sent back for minor repair.

Let X = the number of toasters that need to be sent back.

The condition of the toasters can be considered Bernoulli trials, as verified in the previous exercise. The distribution of the number of repairs required follows *Binom*(10000, 0.05).

$E(X) = np = 10{,}000(0.05) = 500$ toasters.

$SD(X) = \sqrt{npq} = \sqrt{10000(0.05)(0.95)} \approx 21.79$ toasters.

Since np = 500 and nq = 9,500 are both greater than 10, $Binom(10000, 0.05)$ may be approximated by the Normal model, N(500, 21.79).

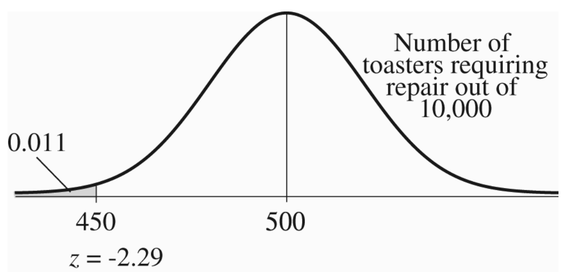
Using *Binom*(10000, 0.05):

$$\begin{aligned} P(\text{fewer than 450 toasters}) &= P(X < 450) \\ &= {}_{10000}C_0\,(0.05)^0\,(0.95)^{10000} + \ldots + {}_{10000}C_{449}\,(0.05)^{449}\,(0.95)^{9551} \\ &\approx 0.009 \end{aligned}$$

According to the Binomial model, the probability that fewer than 450 of 10,000 toasters need repair is approximately 0.009.

Using *N*(500, 21.79):

$$z = \frac{x-\mu}{\sigma}$$
$$z = \frac{450-500}{21.79}$$
$$z = -2.29$$



$P(X < 450) \approx P(z < -2.29) \approx 0.011$

According to the Normal model, the probability that fewer than 450 of 10,000 toasters need repair is approximately 0.011.

6. Soccer again.

A binomial model and a normal model are both appropriate for modeling the number of goals made from corner kicks in a season.

Let X = the number of goals resulting from corner kicks.

Attempting goals from corner kicks can be considered Bernoulli trials, as verified in the previous exercise. The distribution of the number of goals made follows *Binom*(200, 0.08).

$E(X) = np = 200(0.08) = 16$ goals.

$SD(X) = \sqrt{npq} = \sqrt{200(0.08)(0.92)} \approx 3.837$ goals.

Since $np = 16$ and $nq = 184$ are both greater than 10, $Binom(200, 0.08)$ may be approximated by the Normal model, $N(16, 3.837)$.

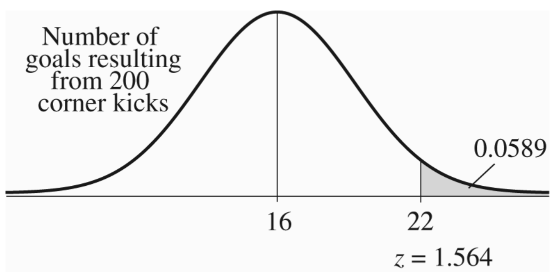
Using *Binom*(200, 0.08):

$$P(\text{more than 22 goals}) = P(X > 22)$$
$$= {}_{200}C_{23}(0.08)^{23}(0.92)^{177} + \ldots + {}_{200}C_{200}(0.08)^{200}(0.92)^{0}$$
$$\approx 0.0507$$

According to the Binomial model, the probability that more than 22 of 200 corner kicks result in goals is approximately 0.0507.

Using *N*(16, 3.837):

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{22 - 16}{3.837}$$
$$z = 1.564$$



$P(X > 22) \approx P(z > 1.564) \approx 0.0589$

According to the Normal model, the probability that more than 22 of 200 corner kicks will result in goals is approximately 0.0589.

Section 16.4

7. Sell!

$E(X) = \lambda = 5$

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$P(X = 3) = \frac{e^{-5}5^{3}}{3!} \approx 0.140$$

According to the Poisson model, the probability of the dealer selling 3 cars is 0.140.

8. Passing on.

$$E(X) = \lambda = 7$$

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$P(X = 10) = \frac{e^{-7}7^{10}}{10!} \approx 0.071$$

According to the Poisson model, the probability that the hospital has 10 fatalities is 0.071.

Section 16.7

9. Telephone numbers.

a) The telephone numbers are distributed uniformly, since all numbers are equally likely to be selected.

b) According to the uniform distribution, the probability of choosing one of the numbers assigned the incubator is 200/10,000 = 0.02.

c) According to the uniform distribution, the probability of choosing a number above 9000 is 1000/10,000 = 0.10

10. Serial numbers.

a) The serial numbers are distributed uniformly, since all numbers are equally likely to be selected.

b) According to the uniform distribution, the probability of choosing one of the last 100 phones to be produced is 100/1000 = 0.10.

c) According to the uniform distribution, the probability of choosing a phone from among the last 200 or first 50 to be produced is (200 + 50)/10000 = 0.25.

11. Component lifetimes.

a) The mean of the exponential model is $\mu = \frac{1}{\lambda}$, so $\lambda = \frac{1}{\mu} = \frac{1}{3}$.

b) Let X = the number of years in the hard drive lifetime.

$$P(X \le 5) = 1 - e^{-\lambda t} = 1 - e^{-(1/3)(5)} \approx 0.811$$

According to the exponential model, the probability that hard drive lasts 5 years or less is approximately 0.811.

12. Website sales.

a) Since 5 sales are expected per hour, we would expect to wait 1/5 of an hour, or 12 minutes.

b) Let X = the number of hours until the next sale. We want to know the probability that the next sale will occur within 6 minutes, or 0.1 hours.

$P(X \le 0.1) = 1 - e^{-\lambda t} = 1 - e^{-(5)(0.1)} \approx 0.393$

According to the exponential model, the probability that the next sale will occur within 6 minutes is 0.393.

13. Simulating the model.

a) Answers will vary. A component is the simulation of the picture in one box of cereal. One possible way to model this component is to generate random digits 0-9. Let 0 and 1 represent Hope Solo and 2-9 a picture of another sports star. Each run will consist of generating random numbers until a 0 or 1 is generated. The response variable will be the number of digits generated until the first 0 or 1.

b) Answers will vary.

c) Answers will vary. To construct your simulated probability model, start by calculating the simulated probability that you get a picture of Hope Solo in the first box. This is the number of trials in which a 0 or 1 was generated first, divided by the total number of trials. Perform similar calculations for the simulated probability that you have to wait until the second box, the third box, etc.

d) Let X = the number of boxes opened until the first Hope Solo picture is found.

X	1	2	3	4	5	6	7	8	≥ 9
P(X)	0.20	$(0.80)(0.20)$ $= 0.16$	$(0.80)^2(0.20)$ $= 0.128$	$(0.80)^3(0.20)$ $= 0.1024$	0.082	0.066	0.052	0.042	0.168

e) Answers will vary.

14. Simulation II.

a) Answers will vary. A component is the simulation of one die roll. One possible way to model this component is to generate random digits 1-6. Let 1 represent getting 1 (the roll you need and let 2-6 represent not getting the roll you need. Each run will consist of generating random numbers until 1 is generated. The response variable will be the number of digits generated until the first 1.

b) Answers will vary.

c) Answers will vary. To construct your simulated probability model, start by calculating the simulated probability that you roll a 1 on the first roll. This is the number of trials in which a 1 was generated first divided by the total number of trials. Perform similar calculations for the simulated probability that you have to wait until the second roll, the third roll, etc.

d) Let X = the number of rolls until the first 1 is rolled.

X	1	2	3	4	5	6	7	8	≥ 9
$P(X)$	$\frac{1}{6}$	$\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)$ ≈ 0.139	$\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)$ ≈ 0.116	$\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)$ ≈ 0.096	$\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)$ ≈ 0.80	0.067	0.056	0.047	0.233

e) Answers will vary.

15. Hope, again.

a) Answers will vary. A component is the simulation of the picture in one box of cereal. One possible way to model this component is to generate random digits 0-9. Let 0 and 1 represent Hope Solo and 2-9 a picture of another sports star. Each run will consist of generating five random numbers. The response variable will be the number of 0s and 1s in the five random numbers.

b) Answers will vary.

c) Answers will vary. To construct your simulated probability model, start by calculating the simulated probability that you get no pictures of Hope Solo in the five boxes. This is the number of trials in which neither 0 nor 1 were generated divided by the total number of trials. Perform similar calculations for the simulated probability that you would get one picture, 2 pictures, etc.

d) Let X = the number of Hope Solo pictures in 5 boxes.

X	0	1	2	3	4	5
$P(X)$	$(0.20)^0(0.80)^5$ ≈ 0.33	${}_5C_1(0.20)^1(0.80)^4$ ≈ 0.41	${}_5C_2(0.20)^2(0.80)^3$ ≈ 0.20	${}_5C_3(0.20)^3(0.80)^2$ ≈ 0.05	${}_5C_4(0.20)^4(0.80)^1$ ≈ 0.01	$(0.20)^5(0.80)^0$ ≈ 0.0

e) Answers will vary.

16. Seatbelts.

a) Answers will vary. A component is the simulation of one driver in a car. One possible way to model this component is to generate pairs of random digits 00-99. Let 01-75 represent a driver wearing his or her seatbelt and let 76-99 and 00 represent a driver not wearing his or her seatbelt. Each run will consist of generating five pairs of random digits. The response variable will be the number of pairs of digits that are 00-75.

b) Answers will vary.

c) Answers will vary. To construct your simulated probability model, start by calculating the simulated probability that none of the five drivers are wearing seatbelts. This is the number of trials in which no pairs of digits were 00-75, divided by the total number of trials. Perform similar calculations for the simulated probability that one driver is wearing his or her seatbelt, two drivers, etc.

d) Let X = the number of drivers wearing seatbelts in 5 cars.

X	0	1	2	3	4	5
P(X)	$(0.75)^0(0.25)^5$ ≈ 0.0	${}_5C_1(0.75)^1(0.25)^4$ ≈ 0.01	${}_5C_2(0.75)^2(0.25)^3$ ≈ 0.09	${}_5C_3(0.75)^3(0.25)^2$ ≈ 0.26	${}_5C_4(0.75)^4(0.25)^1$ ≈ 0.40	$(0.75)^5(0.25)^0$ ≈ 0.24

e) Answers will vary.

17. On time.

These departures cannot be considered Bernoulli trials. Departures from the same airport during a 2-hour period may not be independent. They all might be affected by weather and delays.

18. Lost luggage.

The fate of these bags cannot be considered Bernoulli trials. The fate of 22 pieces of luggage, all checked on the same flight, probably aren't independent.

19. Hoops.

The player's shots may be considered Bernoulli trials. There are only two possible outcomes (make or miss), the probability of making a shot is constant (80%), and the shots are independent of one another (making, or missing, a shot does not affect the probability of making the next).

Let $X =$ the number of shots until the first missed shot.
Let $Y =$ the number of shots until the first made shot.

Since these problems deal with shooting until the first miss (or until the first made shot), a geometric model, either $Geom(0.8)$ or $Geom(0.2)$, is appropriate.

a) Use $Geom(0.2)$. $P(X = 5) = (0.8)^4(0.2) = 0.08192$ (Four shots made, followed by a miss.)

b) Use $Geom(0.8)$. $P(Y = 4) = (0.2)^3(0.8) = 0.0064$ (Three misses, then a made shot.)

c) Use $Geom(0.8)$. $P(Y = 1) + P(Y = 2) + P(Y = 3) = (0.8) + (0.2)(0.8) + (0.2)^2(0.8) = 0.992$

20. Chips.

The selection of chips may be considered Bernoulli trials. There are only two possible outcomes (fail testing and pass testing). Provided that the chips selected are a representative sample of all chips, the probability that a chip fails testing is constant at 2%. The trials are not independent, since the population of chips is finite, but we won't need to sample more than 10% of all chips.

Let $X =$ the number of chips required until the first bad chip.
The appropriate model is $Geom(0.02)$.

a) $P(X = 5) = (0.98)^4(0.02) \approx 0.0184$ (Four good chips, then a bad one.)

b) $P(1 \le X \le 10) = (0.02) + (0.98)(0.02) + (0.98)^2(0.02) + \ldots + (0.98)^9(0.02) \approx 0.183$

(Use the geometric model on a calculator or computer for this one!)

21. More hoops.

As determined in a previous exercise, the shots can be considered Bernoulli trials, and since the player is shooting until the first miss, $Geom(0.2)$ is the appropriate model.

$$E(X) = \frac{1}{p} = \frac{1}{0.2} = 5 \text{ shots}$$

The player is expected to take 5 shots until the first miss.

22. Chips ahoy.

As determined in a previous exercise, the selection of chips can be considered Bernoulli trials, and since the company is selecting until the first bad chip, $Geom(0.02)$ is the appropriate model.

$$E(X) = \frac{1}{p} = \frac{1}{0.02} = 50 \text{ chips}$$

The first bad chip is expected to be the 50th chip selected.

23. Customer center operator.

The calls can be considered Bernoulli trials. There are only two possible outcomes, taking the promotion, and not taking the promotion. The probability of success is constant at 5% (50% of the 10% Platinum cardholders.) The trials are not independent, since there are a finite number of cardholders, but this is a major credit card company, so we can assume we are selecting fewer than 10% of all cardholders. Since we are calling people until the first success, the model $Geom(0.05)$ may be used.

$E(\text{calls}) = \dfrac{1}{p} = \dfrac{1}{0.05} = 20$ calls. We expect it to take 20 calls to find the first cardholder to take the double miles promotion.

24. Cold calls.

The donor contacts can be considered Bernoulli trials. There are only two possible outcomes, giving \$100 or more, and not giving \$100 or more. The probability of success is constant at 1% (5% of the 20% of donors who will make a donation.) The trials are not independent, since there are a finite number of potential donors, but we will assume that she is contacting less than 10% of all possible donors. Since we are contacting people until the first success, the model $Geom(0.01)$ may be used.

$E(\text{contacts}) = \dfrac{1}{p} = \dfrac{1}{0.01} = 100$ contacts. We expect that Justine will have to contact 100 potential donors to find a \$100 donor.

25. Blood.

These may be considered Bernoulli trials. There are only two possible outcomes, Type AB and not Type AB. Provided that the donors are representative of the population, the probability of having Type AB blood is constant at 4%. The trials are not independent, since the population is finite, but we are selecting fewer than 10% of all potential donors. Since we are selecting people until the first success, the model $Geom(0.04)$ may be used.

Let X = the number of donors until the first Type AB donor is found.

a) $E(\mathrm{X}) = \dfrac{1}{p} = \dfrac{1}{0.04} = 25$ people

We expect the 25^{th} person to be the first Type AB donor.

b)

$$
\begin{aligned}
&P(\text{a Type AB donor among the first 5 people checked}) \\
&\quad = P(\mathrm{X}=1)+P(\mathrm{X}=2)+P(\mathrm{X}=3)+P(\mathrm{X}=4)+P(\mathrm{X}=5) \\
&\quad = (0.04)+(0.96)(0.04)+(0.96)^2(0.04)+(0.96)^3(0.04)+(0.96)^4(0.04) \approx 0.185
\end{aligned}
$$

c)

$$
\begin{aligned}
&P(\text{a Type AB donor among the first 6 people checked}) \\
&\quad = P(\mathrm{X}=1)+P(\mathrm{X}=2)+P(\mathrm{X}=3)+P(\mathrm{X}=4)+P(\mathrm{X}=5)+P(\mathrm{X}=6) \\
&\quad = (0.04)+(0.96)(0.04)+(0.96)^2(0.04) \\
&\qquad\qquad +(0.96)^3(0.04)+(0.96)^4(0.04)+(0.96)^5(0.04) \approx 0.217
\end{aligned}
$$

d) $P(\text{no Type AB donor before the 10th person checked}) = P(X > 9) = (0.96)^9 \approx 0.693$

This one is a bit tricky. There is no implication that we actually find a donor on the 10th trial. We only care that nine trials passed with no Type AB donor.

26. Colorblindness.

These may be considered Bernoulli trials. There are only two possible outcomes, colorblind and not colorblind. As long as the men selected are representative of the population of all men, the probability of being colorblind is constant at about 8%. Trials are not independent, since the population is finite, but we won't be sampling more than 10% of the population.

Let $X =$ the number of people checked until the first colorblind man is found.

Since we are selecting people until the first success, the model $Geom(0.08)$, may be used.

a) $E(X) = \frac{1}{p} = \frac{1}{0.08} = 12.5$ people. We expect to examine 12.5 people until finding the first colorblind person.

b) $P(\text{no colorblind men among the first } 4) = P(X > 4) = (0.92)^4 \approx 0.716$

c) $P(\text{first colorblind is the sixth man checked}) = P(X = 6) = (0.92)^5(0.08) \approx 0.0527$

d)

$$\begin{aligned} &P(\text{she finds a colorblind man before the tenth man}) \\ &\quad = P(1 \le X \le 9) \\ &\quad = (0.08) + (0.92)(0.08) + (0.92)^2(0.08) + \ldots + (0.92)^8(0.08) \approx 0.528 \end{aligned}$$

(Use the geometric model on a calculator or computer for this one!)

27. Coins and intuition.

a) Intuitively, we expect 50 heads.

b) $E(\text{heads}) = np = 100(0.5) = 50$ heads.

28. Roulette and intuition.

a) Intuitively, we expect 2 balls to wind up in a green slot.

b) $E(\text{green}) = np = 38\left(\frac{2}{38}\right) = 2$ green.

29. Lefties.

These may be considered Bernoulli trials. There are only two possible outcomes, left-handed and not left-handed. Since people are selected at random, the probability of being left-handed is constant at about 13%. The trials are not independent, since the population is finite, but a sample of 5 people is certainly fewer than 10% of all people.

Let $X =$ the number of people checked until the first lefty is discovered.
Let $Y =$ the number of lefties among $n = 5$.

a) Use $Geom(0.13)$.

$P(\text{first lefty is the fifth person}) = P(X = 5) = (0.87)^4(0.13) \approx 0.0745$

b) Use $Binom(5, 0.13)$.

$$\begin{aligned} P(\text{some lefties among the 5 people}) &= 1 - P(\text{no lefties among the first 5 people}) \\ &= 1 - P(Y = 0) \\ &= 1 - {}_5C_0(0.13)^0(0.87)^5 \\ &\approx 0.502 \end{aligned}$$

c) Use $Geom(0.13)$.

$$\begin{aligned} P(\text{first lefty is second or third person}) &= P(X = 2) + P(X = 3) \\ &= (0.87)(0.13) + (0.87)^2(0.13) \approx 0.211 \end{aligned}$$

d) Use $Binom(5, 0.13)$.

$P(\text{exactly 3 lefties in the group}) = P(Y = 3) = {}_5C_3(0.13)^3(0.87)^2 \approx 0.0166$

e) Use $Binom(5, 0.13)$.

$$\begin{aligned} &P(\text{at least 3 lefties in the group}) \\ &= P(Y = 3) + P(Y = 4) + P(Y = 5) \\ &= {}_5C_3(0.13)^3(0.87)^2 + {}_5C_4(0.13)^4(0.87)^1 + {}_5C_5(0.13)^5(0.87)^0 \\ &\approx 0.0179 \end{aligned}$$

f) Use $Binom(5, 0.13)$.

$$\begin{aligned} P(\text{at most 3 lefties in the group}) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) \\ &= {}_5C_0(0.13)^0(0.87)^5 + {}_5C_1(0.13)^1(0.87)^4 \\ &\quad + {}_5C_2(0.13)^2(0.87)^3 + {}_5C_3(0.13)^3(0.87)^2 \\ &\approx 0.9987 \end{aligned}$$

30. Arrows.

These may be considered Bernoulli trials. There are only two possible outcomes, hitting the bull's-eye and not hitting the bull's-eye. The probability of hitting the bull's-eye is given, $p = 0.80$. The shots are assumed to be independent.

Let $X =$ the number of shots until the first bull's-eye.
Let $Y =$ the number of bull's-eyes in $n = 6$ shots.

a) Use $Geom(0.80)$.

$P\text{ (first bull's-eye is on the third shot) } = P(X = 3) = (0.20)^2(0.80) \approx 0.032$

b) Use $Binom(6, 0.80)$.

$$\begin{aligned} P(\text{at least one miss out of 6 shots}) &= 1 - P(\text{6 out of 6 hits}) \\ &= 1 - P(Y = 6) \\ &= 1 - {}_6C_6(0.80)^6(0.20)^0 \\ &\approx 0.738 \end{aligned}$$

c) Use $Geom(0.80)$.

$$\begin{aligned} P(\text{first hit on fourth or fifth shot}) &= P(X = 4) + P(X = 5) \\ &= (0.20)^3(0.80) + (0.20)^4(0.80) = 0.00768 \end{aligned}$$

d) Use $Binom(6, 0.80)$.

$$\begin{aligned} P(\text{exactly four hits}) &= P(Y = 4) \\ &= {}_6C_4(0.80)^4(0.20)^2 \\ &\approx 0.246 \end{aligned}$$

e) Use $Binom(6, 0.80)$.

$$\begin{aligned} P(\text{at least four hits}) &= P(Y = 4) + P(Y = 5) + P(Y = 6) \\ &= {}_6C_4(0.80)^4(0.20)^2 + {}_6C_5(0.80)^5(0.20)^1 + {}_6C_6(0.80)^6(0.20)^0 \\ &\approx 0.901 \end{aligned}$$

f) Use $Binom(6, 0.80)$.

$$\begin{aligned} P(\text{at most four hits}) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) \\ &= {}_6C_0(0.80)^0(0.20)^6 + {}_6C_1(0.80)^1(0.20)^5 + {}_6C_2(0.80)^2(0.20)^4 \\ &\quad + {}_6C_3(0.80)^3(0.20)^3 + {}_6C_4(0.80)^4(0.20)^2 \\ &\approx 0.345 \end{aligned}$$

31. Lefties redux.

a) In a previous exercise, we determined that the selection of lefties could be considered Bernoulli trials. Since our group consists of 5 people, use $Binom(5, 0.13)$.

Let $Y =$ the number of lefties among $n = 5$.

$E(Y) = np = 5(0.13) = 0.65$ lefties

b) $SD(Y) = \sqrt{npq} = \sqrt{5(0.13)(0.87)} \approx 0.75$ lefties

c) Use $Geom(0.13)$. Let $X =$ the number of people checked until the first lefty is discovered.

$E(X) = \dfrac{1}{p} = \dfrac{1}{0.13} \approx 7.69$ people

32. More arrows.

a) In a previous exercise, we determined that the shots could be considered Bernoulli trials. Since the archer is shooting 6 arrows, use $Binom(6, 0.80)$.

Let $Y =$ the number of bull's-eyes in $n = 6$ shots.

$E(Y) = np = 6(0.80) = 4.8$ bull's-eyes.

b) $SD(Y) = \sqrt{npq} = \sqrt{6(0.80)(0.20)} \approx 0.98$ bull's-eyes.

c) Use $Geom(0.80)$. Let $X =$ the number of arrows shot until the first bull's-eye.

$E(X) = \dfrac{1}{p} = \dfrac{1}{0.80} = 1.25$ shots.

33. Still more lefties.

a) In a previous exercise, we determined that the selection of lefties (and also righties) could be considered Bernoulli trials. Since our group consists of 12 people, and now we are considering the righties, use $Binom(12, 0.87)$.

Let $Y =$ the number of righties among $n = 12$.

$E(Y) = np = 12(0.87) = 10.44$ righties

$SD(Y) = \sqrt{npq} = \sqrt{12(0.87)(0.13)} \approx 1.16$ righties

b)

$$\begin{aligned}P(\text{not all righties}) &= 1 - P(\text{all righties})\\ &= 1 - P(Y = 12)\\ &= 1 - {}_{12}C_{12}(0.87)^{12}(0.13)^{0}\\ &\approx 0.812\end{aligned}$$

c)

$$\begin{aligned}P(\text{no more than 10 righties}) &= P(Y \le 10)\\ &= P(Y = 0) + P(Y = 1) + P(Y = 2) + \ldots + P(Y = 10)\\ &= {}_{12}C_{0}(0.87)^{0}(0.13)^{12} + \ldots + {}_{12}C_{10}(0.87)^{10}(0.13)^{2}\\ &\approx 0.475\end{aligned}$$

d)

$$\begin{aligned}P(\text{exactly six of each}) &= P(Y = 6)\\ &= {}_{12}C_{6}(0.87)^{6}(0.13)^{6}\\ &\approx 0.00193\end{aligned}$$

e)

$$\begin{aligned} P(\text{majority righties}) &= P(Y \geq 7) \\ &= P(Y=7)+P(Y=8)+P(Y=9)+\ldots+P(Y=12) \\ &= {}_{12}C_7(0.87)^7(0.13)^5+\ldots+{}_{12}C_{12}(0.87)^{12}(0.13)^0 \\ &\approx 0.998 \end{aligned}$$

34. Still more arrows.

a) In a previous exercise, we determined that the archer's shots could be considered Bernoulli trials. Since our archer is now shooting 10 arrows, use $Binom(10, 0.80)$.

Let $Y =$ the number of bull's-eyes hit from $n = 10$ shots.
$E(Y) = np = 10(0.80) = 8$ bull's-eyes hit.
$SD(Y) = \sqrt{npq} = \sqrt{10(0.80)(0.20)} \approx 1.26$ bull's-eyes hit.

b)

$$\begin{aligned} P(\text{no misses out of 10 shots}) &= P(\text{all hits out of 10 shots}) \\ &= P(Y=10) \\ &= {}_{10}C_{10}(0.80)^{10}(0.20)^0 \\ &\approx 0.107 \end{aligned}$$

c)

$$\begin{aligned} P(\text{no more than 8 hits}) &= P(Y \leq 8) \\ &= P(Y=0)+P(Y=1)+P(Y=2)+\ldots+P(Y=8) \\ &= {}_{10}C_0(0.80)^0(0.20)^{10}+\ldots+{}_{10}C_8(0.80)^8(0.20)^2 \\ &\approx 0.624 \end{aligned}$$

d)

$$\begin{aligned} P(\text{exactly 8 out of 10 shots}) &= P(Y=8) \\ &= {}_{10}C_8(0.80)^8(0.20)^2 \\ &\approx 0.302 \end{aligned}$$

e)

$$\begin{aligned} P(\text{more hits than misses}) &= P(Y \geq 6) \\ &= P(Y=6)+P(Y=7)+\ldots+P(Y=10) \\ &= {}_{10}C_6(0.80)^6(0.20)^4+\ldots+{}_{10}C_{10}(0.80)^{10}(0.20)^0 \\ &\approx 0.967 \end{aligned}$$

35. Vision.

The vision tests can be considered Bernoulli trials. There are only two possible outcomes, nearsighted or not. The probability of any child being nearsighted is given as $p = 0.12$. Finally, since the population of children is finite, the trials are not independent. However, 169 is certainly less than 10% of all children, and we will assume that the children in this district are representative of all children in relation to nearsightedness. Use $Binom(169, 0.12)$.

$$\mu = E(\text{nearsighted}) = np = 169(0.12) = 20.28 \text{ children.}$$

$$\sigma = SD(\text{nearsighted}) = \sqrt{npq} = \sqrt{169(0.12)(0.88)} \approx 4.22 \text{ children.}$$

36. International students.

The students can be considered Bernoulli trials. There are only two possible outcomes, international or not. The probability of any freshmen being an international student is given as $p = 0.06$. Finally, since the population of freshmen is finite, the trials are not independent. However, 40 is likely to be less than 10% of all students, and we are told that the freshmen in this college are randomly assigned to housing. Use $Binom(40, 0.06)$.

$$\mu = E(\text{international}) = np = 40(0.06) = 2.4 \text{ students.}$$

$$\sigma = SD(\text{international}) = \sqrt{npq} = \sqrt{40(0.06)(0.94)} \approx 1.5 \text{ students.}$$

37. Tennis, anyone?

The first serves can be considered Bernoulli trials. There are only two possible outcomes, successful and unsuccessful. The probability of any first serve being good is given as $p = 0.70$. Finally, we are assuming that each serve is independent of the others. Since she is serving 6 times, use $Binom(6, 0.70)$.

Let $X =$ the number of successful serves in $n = 6$ first serves.

a)

$$\begin{aligned} P(\text{six serves in }) &= P(X = 6) \\ &= {}_6C_6(0.70)^6(0.30)^0 \\ &\approx 0.118 \end{aligned}$$

b)

$$\begin{aligned} &P(\text{exactly four serves in}) \\ &= P(X = 4) \\ &= {}_6C_4(0.70)^4(0.30)^2 \\ &\approx 0.324 \end{aligned}$$

c)

$$\begin{aligned} P(\text{at least four serves in}) &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= {}_6C_4(0.70)^4(0.30)^2 + {}_6C_5(0.70)^5(0.30)^1 + {}_6C_6(0.70)^6(0.30)^0 \\ &\approx 0.744 \end{aligned}$$

d)

$$\begin{aligned}
&P(\text{no more than four serves in})\\
&= P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)\\
&= {}_6C_0(0.70)^0(0.30)^6 + {}_6C_1(0.70)^1(0.30)^5 + {}_6C_2(0.70)^2(0.30)^4\\
&\quad + {}_6C_3(0.70)^3(0.30)^3 + {}_6C_4(0.70)^4(0.30)^2\\
&\approx 0.580
\end{aligned}$$

38. Frogs.

The frog examinations can be considered Bernoulli trials. There are only two possible outcomes, having the trait and not having the trait. If the frequency of the trait has not changed, and the biologist collects a representative sample of frogs, then the probability of a frog having the trait is constant, at p = 0.125. The trials are not independent since the population of frogs is finite, but 12 frogs is fewer than 10% of all frogs. Since the biologist is collecting 12 frogs, use $Binom(12, 0.125)$.

Let $X =$ the number of frogs with the trait, from n = 12 frogs.

a)

$$\begin{aligned}
P(\text{no frogs have the trait}) &= P(X=0)\\
&= {}_{12}C_0(0.125)^0(0.875)^{12}\\
&\approx 0.201
\end{aligned}$$

b)

$$\begin{aligned}
P(\text{at least two frogs}) &= P(X\geq 2)\\
&= P(X=2)+P(X=3)+\ldots+P(X=12)\\
&= {}_{12}C_2(0.125)^2(0.875)^{10}+\ldots+{}_{12}C_{12}(0.125)^{12}(0.875)^0\\
&\approx 0.453
\end{aligned}$$

c)

$$\begin{aligned}
P(\text{three or four frogs have trait}) &= P(X=3)+P(X=4)\\
&= {}_{12}C_3(0.125)^3(0.875)^9 + {}_{12}C_4(0.125)^4(0.875)^8\\
&\approx 0.171
\end{aligned}$$

d)

$$\begin{aligned}
P(\text{no more than four}) &= P(X\leq 4) = P(X=0)+P(X=1)+\ldots+P(X=4)\\
&= {}_{12}C_0(0.125)^0(0.875)^{12}+\ldots+{}_{12}C_4(0.125)^4(0.875)^8\\
&\approx 0.989
\end{aligned}$$

39. And more tennis.

The first serves can be considered Bernoulli trials. There are only two possible outcomes, successful and unsuccessful. The probability of any first serve being good is given as $p = 0.70$. Finally, we are assuming that each serve is independent of the others. Since she is serving 80 times, use $Binom(80, 0.70)$.

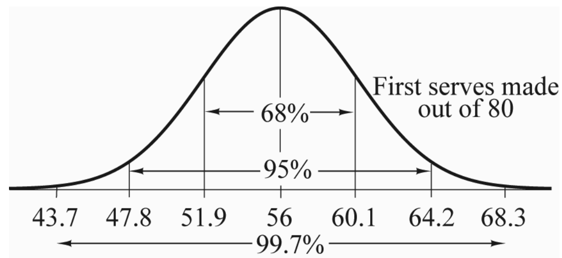
Let $X =$ the number of successful serves in $n = 80$ first serves.

a) $E(X) = np = 80(0.70) = 56$ first serves in.

$SD(X) = \sqrt{npq} = \sqrt{80(0.70)(0.30)} \approx 4.10$ first serves in.

b) Since $np = 56$ and $nq = 24$ are both greater than 10, $Binom(80, 0.70)$ may be approximated by the Normal model, $N(56, 4.10)$.

c) According to the Normal model, in matches with 80 serves, she is expected to make between 51.9 and 60.1 first serves approximately 68% of the time, between 47.8 and 64.2 first serves approximately 95% of the time, and between 43.7 and 68.3 first serves approximately 99.7% of the time.



d) Using *Binom*(80, 0.70):

$$\begin{aligned} P(\text{at least 65 first serves}) &= P(X \geq 65) \\ &= P(X = 65) + P(X = 66) + \ldots + P(X = 80) \\ &= {}_{80}C_{65}(0.70)^{65}(0.30)^{15} + \ldots + {}_{80}C_{80}(0.70)^{80}(0.30)^{0} \\ &\approx 0.0161 \end{aligned}$$

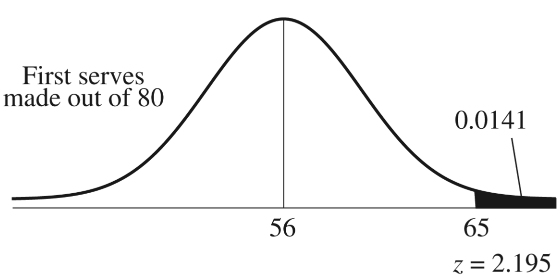
According to the Binomial model, the probability that she makes at least 65 first serves out of 80 is approximately 0.0161.

Using *N*(56, 4.10):

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{65 - 56}{4.10}$$

$$z \approx 2.195$$



According to the Normal model, the probability that she makes at least 65 first serves out of 80 is approximately 0.0141.

$P(X \geq 65) \approx P(z > 2.195) \approx 0.0141$

40. More arrows.

These may be considered Bernoulli trials. There are only two possible outcomes, hitting the bull's-eye and not hitting the bull's-eye. The probability of hitting the bull's-eye is given, $p = 0.80$. The shots are assumed to be independent. Since she will be shooting 200 arrows, use *Binom*(200, 0.80).

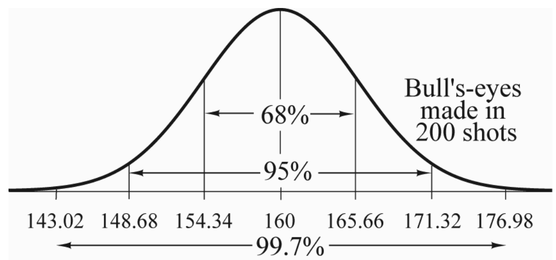
Let $Y =$ the number of bull's-eyes in $n = 200$ shots.

a) $E(Y) = np = 200(0.80) = 160$ bull's-eyes.

$SD(Y) = \sqrt{npq} = \sqrt{200(0.80)(0.20)} \approx 5.66$ bull's-eyes.

b) Since $np = 160$ and $nq = 40$ are both greater than 10, $Binom(200, 0.80)$ may be approximated by the Normal model, $N(160, 5.66)$.

c) According to the Normal model, in matches with 200 arrows, she is expected to get between 154.34 and 165.66 bull's-eyes approximately 68% of the time, between 148.68 and 171.32 bull's-eyes approximately 95% of the time, and between 143.02 and 176.98 bull's-eyes approximately 99.7% of the time.



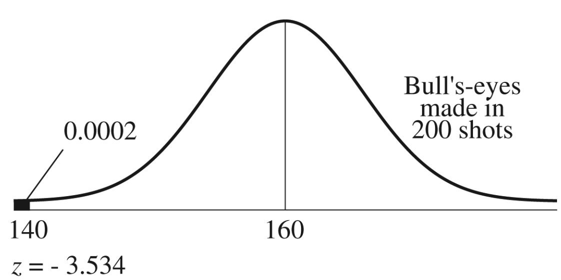
d) Using *Binom*(200, 0.80):

$$\begin{aligned} P(\text{at most 140 hits}) &= P(Y \le 140) \\ &= P(Y = 0) + P(Y = 1) + \ldots + P(Y = 140) \\ &= {}_{200}C_0 (0.80)^0 (0.20)^{200} + \ldots + {}_{200}C_{140} (0.80)^{140} (0.70)^{60} \\ &\approx 0.0005 \end{aligned}$$

According to the Binomial model, the probability that she makes at most 140 bull's-eyes out of 200 is approximately 0.0005.

Using *N*(160, 5.66):

$$z = \frac{y - \mu}{\sigma}$$
$$z = \frac{140 - 160}{5.66}$$
$$z \approx -3.534$$



According to the Normal model, the probability that she hits at most 140 bull's-eyes out of 200 is approximately 0.0002.

$P(Y \le 200) \approx P(z < -3.534) \approx 0.0002$. Using either model, it is apparent that it is very unlikely that the archer would hit only 140 bull's-eyes out of 200.

41. Apples.

a) A binomial model and a normal model are both appropriate for modeling the number of cider apples that may come from the tree.

Let X = the number of cider apples found in the n = 300 apples from the tree.

The quality of the apples may be considered Bernoulli trials. There are only two possible outcomes, cider apple or not a cider apple. The probability that an apple must be used for a cider apple is constant, given as p = 0.06. The trials are not independent, since the population of apples is finite, but the apples on the tree are undoubtedly less than 10% of all the apples that the farmer has ever produced, so model with *Binom*(300, 0.06).

$E(X) = np = 300(0.06) = 18$ cider apples.

$SD(X) = \sqrt{npq} = \sqrt{300(0.06)(0.94)} \approx 4.11$ cider apples.

Since np = 18 and nq = 282 are both greater than 10, $Binom(300, 0.06)$ may be approximated by the Normal model, N(18, 4.11).

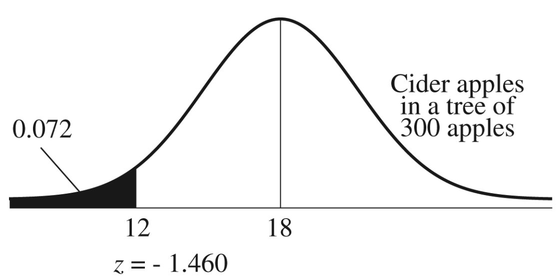
b) Using *Binom*(300, 0.06):

$$\begin{aligned} P(\text{at most 12 cider apples}) &= P(X \le 12) \\ &= P(X=0)+\ldots+P(X=12) \\ &= {}_{300}C_0(0.06)^0(0.94)^{300}+\ldots+{}_{300}C_{12}(0.06)^{12}(0.94)^{282} \\ &\approx 0.085 \end{aligned}$$

According to the Binomial model, the probability that no more than 12 cider apples come from the tree is approximately 0.085.

Using *N*(18, 4.11):

$$z = \frac{x-\mu}{\sigma}$$
$$z = \frac{12-18}{4.11}$$
$$z = -1.460$$



According to the Normal model, the probability that no more than 12 apples out of 300 are cider apples is approximately 0.072.

$P(X \le 12) \approx P(z < -1.460) \approx 0.072$

c) It is extremely unlikely that the tree will bear more than 50 cider apples. Using the Normal model, N(18, 4.11), 50 cider apples is approximately 7.8 standard deviations above the mean.

42. Frogs, part II.

The frog examinations can be considered Bernoulli trials. There are only two possible outcomes, having the trait and not having the trait. If the frequency of the trait has not changed, and the biologist collects a representative sample of frogs, then the probability of a frog having the trait is constant, at $p = 0.125$. The trials are not independent since the population of frogs is finite, but 150 frogs is fewer than 10% of all frogs. Since the biologist is collecting 150 frogs, use $Binom(150, 0.125)$.

Let $X =$ the number of frogs with the trait, from $n = 150$ frogs.

a) $E(X) = np = 150(0.125) = 18.75$ frogs.

$SD(X) = \sqrt{npq} = \sqrt{150(0.125)(0.875)} \approx 4.05$ frogs.

b) Since $np = 18.75$ and $nq = 131.25$ are both greater than 10, $Binom(200, 0.125)$ may be approximated by the Normal model, $N(18.75, 4.05)$.

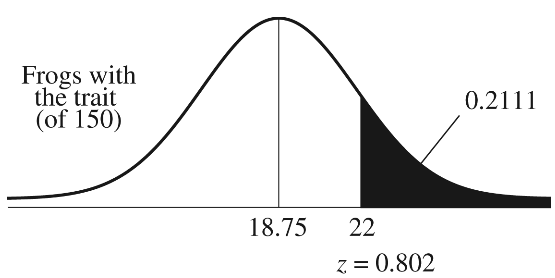
c) **Using *Binom*(150, 0.125):**

$$\begin{aligned} P(\text{at least } 22 \text{ frogs}) &= P(X \geq 22) \\ &= P(X = 22) + \ldots + P(X = 150) \\ &= {}_{150}C_{22}(0.125)^{22}(0.875)^{128} + \ldots + {}_{150}C_{150}(0.125)^{150}(0.875)^{0} \\ &\approx 0.2433 \end{aligned}$$

According to the Binomial model, the probability that at least 22 frogs out of 150 have the trait is approximately 0.2433.

Using *N*(18.75, 4.05):

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{22 - 18.75}{4.05}$$
$$z \approx 0.802$$



According to the Normal model, the probability that at least 22 frogs out of 150 have the trait is approximately 0.2111.

$P(X \geq 22) \approx P(z > 0.802) \approx 0.2111$

Using either model, the probability that the biologist discovers 22 of 150 frogs with the trait simply as a result of natural variability is quite high. This doesn't prove that the trait has become more common.

43. Lefties again.

Let $X =$ the number of righties among a class of n = 188 students.

Using *Binom*(188, 0.87):

These may be considered Bernoulli trials. There are only two possible outcomes, right-handed and not right-handed. The probability of being right-handed is assumed to be constant at about 87%. The trials are not independent, since the population is finite, but a sample of 188 students is certainly fewer than 10% of all people. Therefore, the number of righties in a class of 188 students may be modeled by *Binom*(188, 0.87).

If there are 171 or more righties in the class, some righties have to use a left-handed desk.

$$\begin{aligned} P(\text{at least 171 righties}) &= P(X \geq 171) \\ &= P(X = 171) + \ldots + P(X = 188) \\ &= {}_{188}C_{171}(0.87)^{171}(0.13)^{17} + \ldots + {}_{188}C_{188}(0.87)^{188}(0.13)^{0} \\ &\approx 0.061 \end{aligned}$$

According to the binomial model, the probability that a right-handed student has to use a left-handed desk is approximately 0.061.

Using *N*(163.56, 4.61):

$E(X) = np = 188(0.87) = 163.56$ righties.

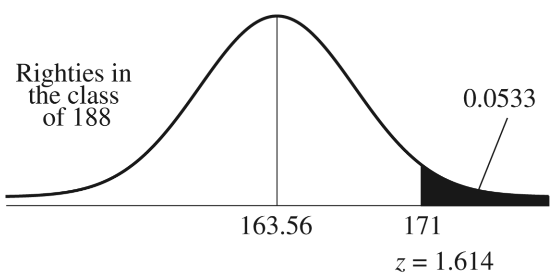
$SD(X) = \sqrt{npq} = \sqrt{188(0.87)(0.13)} \approx 4.61$ righties.

Since np = 163.56 and nq = 24.44 are both greater than 10, $Binom(188, 0.87)$ may be approximated by the Normal model, N(163.56, 4.61).

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{171 - 163.56}{4.61}$$

$$z \approx 1.614$$



According to the Normal model, the probability that there are at least 171 righties in the class of 188 is approximately 0.0533.

$P(X \geq 171) \approx P(z > 1.614) \approx 0.053$

44. No-shows.

Let $X =$ the number of passengers that show up for the flight of n = 275 passengers.

Using *Binom*(275, 0.95):
These may be considered Bernoulli trials. There are only two possible outcomes, showing up and not showing up. The airlines believe the probability of showing up is constant at about 95%. The trials are not independent, since the population is finite, but a sample of 275 passengers is certainly fewer than 10% of all passengers. Therefore, the number of passengers who show up for a flight of 275 may be modeled by *Binom*(275, 0.95).

If 266 or more passengers show up, someone has to get bumped off the flight.

$$\begin{aligned} P(\text{at least } 266 \text{ passengers}) &= P(X \geq 266) \\ &= P(X = 266) + \ldots + P(X = 275) \\ &= {}_{275}C_{266}(0.95)^{266}(0.05)^{9} + \ldots + {}_{275}C_{275}(0.95)^{275}(0.05)^{0} \\ &\approx 0.116 \end{aligned}$$

According to the binomial model, the probability someone on the flight must be bumped is approximately 0.116.

Using *N*(261.25, 3.61):

$E(X) = np = 275(0.95) = 261.25$ passengers.

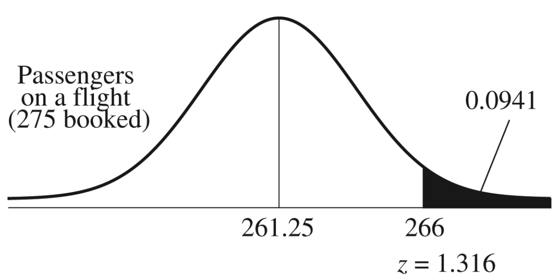
$SD(X) = \sqrt{npq} = \sqrt{275(0.95)(0.05)} \approx 3.61$ passengers.

Since np = 261.25 and nq = 13.75 are both greater than 10, $Binom(275, 0.95)$ may be approximated by the Normal model, *N*(261.25, 3.61).

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{266 - 261.25}{3.61}$$

$$z \approx 1.316$$



According to the Normal model, the probability that at least 266 passengers show up is approximately 0.0941.

$P(X \geq 266) \approx P(z > 1.316) \approx 0.0941$

45. Annoying phone calls.

Let $X =$ the number of sales made after making n = 200 calls.

Using *Binom*(200, 0.12):
These may be considered Bernoulli trials. There are only two possible outcomes, making a sale and not making a sale. The telemarketer was told that the probability of making a sale is constant at about p = 0.12. The trials are not independent, since the population is finite, but 200 calls is fewer than 10% of all calls. Therefore, the number of sales made after making 200 calls may be modeled by *Binom*(200, 0.12).

$$\begin{aligned} P(\text{at most } 10) &= P(X \le 10) \\ &= P(X=0)+\ldots+P(X=10) \\ &= {}_{200}C_0(0.12)^0(0.88)^{200}+\ldots+{}_{200}C_{1900}(0.12)^{10}(0.88)^{190} \\ &\approx 0.0006 \end{aligned}$$

According to the Binomial model, the probability that the telemarketer would make at most 10 sales is approximately 0.0006.

Using *N*(24, 4.60):

$E(X) = np = 200(0.12) = 24$ sales.

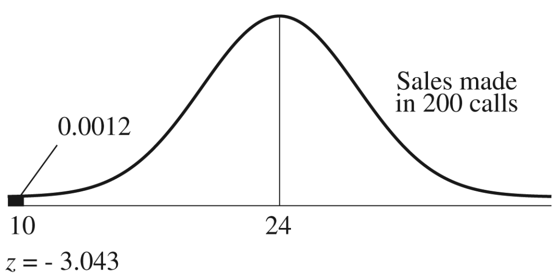
$SD(X) = \sqrt{npq} = \sqrt{200(0.12)(0.88)} \approx 4.60$ sales.

Since $np = 24$ and $nq = 176$ are both greater than 10, $Binom(200, 0.12)$ may be approximated by the Normal model, $N(24, 4.60)$.

$$z = \frac{x-\mu}{\sigma}$$

$$z = \frac{10-24}{4.60}$$

$$z \approx -3.043$$



$P(X \le 10) \approx P(z < -3.043) \approx 0.0012$

According to the Normal model, the probability that the telemarketer would make at most 10 sales is approximately 0.0012.

Since the probability that the telemarketer made 10 sales, given that the 12% of calls result in sales is so low, it is likely that he was misled about the true success rate.

46. The euro.

Let $X =$ the number of heads after spinning a Belgian euro $n = 250$ times.

Using *Binom*(250, 0.5):
These may be considered Bernoulli trials. There are only two possible outcomes, heads and tails. The probability that a fair Belgian euro lands heads is $p = 0.5$. The trials are independent, since the outcome of a spin does not affect other spins. Therefore, *Binom*(250, 0.5) may be used to model the number of heads after spinning a Belgian euro 250 times.

$$\begin{aligned} P(\text{at least } 140) &= P(X \ge 140) \\ &= P(X=140)+\ldots+P(X=250) \\ &= {}_{250}C_{140}(0.5)^{140}(0.5)^{110}+\ldots+{}_{250}C_{250}(0.5)^{250}(0.5)^0 \\ &\approx 0.0332 \end{aligned}$$

According to the Binomial model, the probability that a fair Belgian euro comes up heads at least 140 times is 0.0332.

Using *N*(125, 7.91):

$E(X) = np = 250(0.05) = 125$ heads.

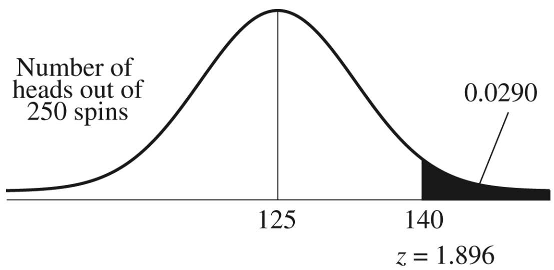
$SD(X) = \sqrt{npq} = \sqrt{250(0.5)(0.5)} \approx 7.91$ heads.

Since np = 125 and nq = 125 are both greater than 10, $Binom(250, 0.5)$ may be approximated by the Normal model, $N(125, 7.91)$.

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{140 - 125}{7.91}$$

$$z \approx 1.896$$



According to the Normal model, the probability that a fair Belgian euro lands heads at least 140 out of 250 spins is approximately 0.0290.

$P(X \geq 140) \approx P(z > 1.896) \approx 0.0290$

Since the probability that a fair Belgian euro lands heads at least 140 out of 250 spins is low, it is unlikely that the euro spins fairly. However, the probability is not extremely low, and we aren't sure of the source of the data, so it might be a good idea to spin it some more.

47. Hurricanes, redux.

a)

$E(X) = \lambda = 2.45$

$$P(\text{no hurricanes next year}) = \frac{e^{-2.45}(2.45)^0}{0!} \approx 0.0863$$

b)

$P(\text{exactly one hurricane in next 2 years})$

$= P(\text{hurr. 1}^{\text{st}}\text{ yr})P(\text{no hurr. 2}^{\text{nd}}\text{ yr}) + P(\text{no hurr. 1}^{\text{st}}\text{ yr})P(\text{hurr. 2}^{\text{nd}}\text{ yr})$

$$= \left(\frac{e^{-2.45}(2.45)^1}{1!}\right)\left(\frac{e^{-2.45}(2.45)^0}{0!}\right) + \left(\frac{e^{-2.45}(2.45)^0}{0!}\right)\left(\frac{e^{-2.45}(2.45)^1}{1!}\right)$$

≈ 0.0365

48. Bank tellers.

a) Because the Poisson model scales according to the sample size, we can calculate the mean for 10 minutes. A Poisson model with mean 2 customers per hour (60 minutes) is equivalent to a Poisson model with mean $\frac{2}{6} = \frac{1}{3}$ customers per $\frac{60}{6} = 10$ minutes.

Let $X =$ the number of customers arriving in 10 minutes.

$$P(X=0)=\frac{e^{-\frac{1}{3}}\left(\frac{1}{3}\right)^0}{0!}\approx 0.7165$$

b) $P(X \geq 2) = 1 - \left[P(X=0) + P(X=1)\right] = 1 - \left[\frac{e^{-\frac{1}{3}}\left(\frac{1}{3}\right)^0}{0!} + \frac{e^{-\frac{1}{3}}\left(\frac{1}{3}\right)^1}{1!}\right] \approx 0.0446$

c) No. The probabilities do not change based on what has just happened. Even though two customers just came in, the probability of no customers in the next 10 minutes will not change. This is neither a better nor a worse time.

49. TB again.

a) $E(X) = \lambda = np = 8000(0.0005) = 4$ cases.

b)

$$\begin{aligned}P(\text{at least one new case}) &= 1 - P(\text{no new cases})\\ &= 1 - \frac{e^{-4}(4)^0}{0!} \approx 0.9817\end{aligned}$$

50. Earthquakes.

a) $E(X) = \lambda = np = 1000\left(\frac{1}{10{,}000}\right) = \frac{1}{10}$

b)

$$\begin{aligned}P(\text{at least one earthquake in next 100 days}) &= 1 - P(\text{no earthquakes})\\ &= 1 - \frac{e^{-\frac{1}{10}}\left(\frac{1}{10}\right)^0}{0!} \approx 0.0952\end{aligned}$$

51. Seatbelts II.

These stops may be considered Bernoulli trials. There are only two possible outcomes, belted or not belted. Police estimate that the probability that a driver is buckled is 80%. (The probability of not being buckled is therefore 20%.) Provided the drivers stopped are representative of all drivers, we can consider the probability constant. The trials are not independent, since the population of drivers is finite, but the police will not stop more than 10% of all drivers.

a) Let X = the number of cars stopped before finding a driver whose seat belt is not buckled. Use $Geom(0.2)$ to model the situation.

$$E(X) = \frac{1}{p} = \frac{1}{0.2} = 5 \text{ cars.}$$

b) $P(\text{First unbelted driver is in the sixth car}) = P(X=6) = (0.8)^5(0.2) \approx 0.066$

c) $P(\text{The first ten drivers are wearing seatbelts}) = (0.8)^{10} \approx .107$

d) Let Y = the number of drivers wearing their seatbelts in 30 cars.
Use $Binom(30, 0.8)$.

$E(Y) = np = 30(0.8) = 24$ drivers.

$SD(Y) = \sqrt{npq} = \sqrt{30(0.8)(0.2)} \approx 2.19$ drivers.

e) Let W = the number of drivers not wearing their seatbelts in 120 cars.

Using *Binom*(120, 0.2):

$$\begin{aligned}P(\text{at least } 20) &= P(W \geq 20)\\ &= P(W = 20) + \ldots + P(W = 120)\\ &= {}_{120}C_{20}(0.2)^{20}(0.8)^{100} + \ldots + {}_{120}C1_{20}(0.2)^{120}(0.8)^{0}\\ &\approx 0.848\end{aligned}$$

According to the Binomial model, the probability that at least 20 out of 120 drivers are not wearing their seatbelts is approximately 0.848.

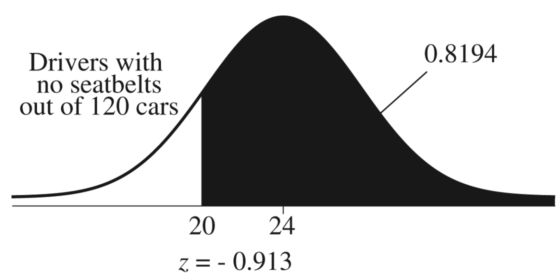
Using *N*(24, 4.38):

$E(W) = np = 120(0.2) = 24$ drivers.

$SD(W) = \sqrt{npq} = \sqrt{120(0.2)(0.8)} \approx 4.38$ drivers.

Since np = 24 and nq = 96 are both greater than 10, $Binom(120, 0.2)$ may be approximated by the Normal model, $N(24, 4.38)$.

$$z = \frac{w - \mu}{\sigma}$$
$$z = \frac{20 - 24}{4.38}$$
$$z \approx -0.913$$



$P(W \geq 120) \approx P(z > -0.913) \approx 0.8194$

According to the Normal model, the probability that at least 20 out of 120 drivers stopped are not wearing their seatbelts is approximately 0.8194.

52. Rickets.

The selection of these children may be considered Bernoulli trials. There are only two possible outcomes, vitamin D deficient or not vitamin D deficient. Recent research indicates that 20% of British children are vitamin D deficient. (The probability of not being vitamin D deficient is therefore 80%.) Provided the students at this school are representative of all British children, we can consider the probability constant. The trials are not independent, since the population of British children is finite, but the children at this school represent fewer than 10% of all British children.

a) Let X = the number of students tested before finding a student who is vitamin D deficient. Use $Geom(0.2)$ to model the situation.

$P(\text{First vit. D def. child is the eighth one tested})=P(X=8)=(0.8)^7(0.2)\approx 0.042$

b) $P(\text{The first ten children tested are okay})=(0.8)^{10}\approx 0.107$

c) $E(X)=\dfrac{1}{p}=\dfrac{1}{0.2}=5$ kids.

d) Let Y = the number of children who are vitamin D deficient out of 50 children. Use $Binom(50, 0.2)$.

$E(Y)=np=50(0.2)=10$ kids. $\quad SD(Y)=\sqrt{npq}=\sqrt{50(0.2)(0.8)}\approx 2.83$ kids

e) Using *Binom*(320, 0.2):

$$
\begin{aligned}
&P(\text{no more than 50 children have the deficiency})\\
&= P(X\le 50)\\
&= P(X=0)+\ldots+P(X=50)\\
&= {}_{320}C_0(0.2)^0(0.8)^{320}+\ldots+{}_{320}C_0(0.2)^{50}(0.8)^{270}\\
&\approx 0.027
\end{aligned}
$$

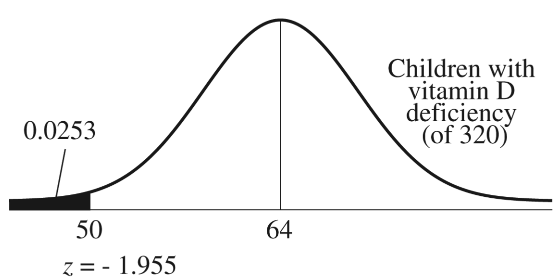
According to the Binomial model, the probability that no more than 50 of the 320 children have the vitamin D deficiency is approximately 0.027**.**

Using *N*(64, 7.16):

$E(Y)=np=320(0.2)=64$ kids. $\quad SD(Y)=\sqrt{npq}=\sqrt{320(0.2)(0.8)}\approx 7.16$ kids.

Since np = 64 and nq = 256 are both greater than 10, $Binom(320, 0.2)$ may be approximated by the Normal model, $N(64, 7.16)$.

$$z=\frac{y-\mu}{\sigma}$$
$$z=\frac{50-64}{7.16}$$
$$z\approx -1.955$$



According to the Normal model, the probability that no more than 50 out of 320 children have the vitamin D deficiency is approximately 0.0253.

$P(Y\le 50)\approx P(z<-1.955)\approx 0.0253$

53. ESP.

Choosing symbols may be considered Bernoulli trials. There are only two possible outcomes, correct or incorrect. Assuming that ESP does not exist, the probability of a correct identification from a randomized deck is constant, at p = 0.20. The trials are independent, as long as the deck is shuffled after each attempt. Since 100 trials will be performed, use *Binom*(100, 0.2).

Let X = the number of symbols identified correctly out of 100 cards.

$E(X) = np = 100(0.2) = 20$ correct identifications.

$SD(X) = \sqrt{npq} = \sqrt{100(0.2)(0.8)} = 4$ correct identifications.

Answers may vary. In order be convincing, the "mind reader" would have to identify at least 32 out of 100 cards correctly, since 32 is three standard deviations above the mean. Identifying fewer cards than 32 could happen too often, simply due to chance.

54. True-False.

Guessing at answers may be considered Bernoulli trials. There are only two possible outcomes, correct or incorrect. If the student was guessing, the probability of a correct response is constant, at p = 0.50. The trials are independent, since the answer to one question should not have any bearing on the answer to the next. Since 50 questions are on the test use *Binom*(500, 0.5).

Let X = the number of questions answered correctly out of 50 questions.

$E(X) = np = 50(0.5) = 25$ correct answers.

$SD(X) = \sqrt{npq} = \sqrt{50(0.5)(0.5)} \approx 3.54$ correct answers.

Answers may vary. In order be convincing, the student would have to answer at least 36 out of 50 questions correctly, since 36 is approximately three standard deviations above the mean. Answering fewer than 36 questions correctly could happen too often, simply due to chance.

55. Hot hand.

A streak like this is not unusual. The probability that he makes 4 in a row with a 55% free throw percentage is $(0.55)(0.55)(0.55)(0.55) \approx 0.09$. We can expect this to happen nearly one in ten times for every set of 4 shots that he makes. One out of ten times is not that unusual.

56. New bow.

A streak like this is not unusual. The probability that she makes 6 consecutive bulls-eyes with an 80% bulls-eye percentage is $(0.8)(0.8)(0.8)(0.8)(0.8)(0.8) \approx 0.26$. If she were to shoot several flights of 6 arrows, she is expected to get 6 bulls-eyes about 26% of the time. An event that happens due to chance about one out of four times is not that unusual.

57. Hotter hand.

The shots may be considered Bernoulli trials. There are only two possible outcomes, make or miss. The probability of success is constant at 55%, and the shots are independent of one another. Therefore, we can model this situation with *Binom*(32, 0.55).

Let X = the number of free throws made out of 40.

$E(X) = np = 40(0.55) = 22$ free throws made.

$SD(X) = \sqrt{npq} = \sqrt{40(0.55)(0.45)} \approx 3.15$ free throws.

Answers may vary. The player's performance seems to have increased. 32 made free throws is $(32-22)/3.15 \approx 3.17$ standard deviations above the mean, an extraordinary feat, unless his free throw percentage has increased. This does NOT mean that the sneakers are responsible for the increase in free throw percentage. Some other variable may account for the increase. The player would need to set up a controlled experiment in order to determine what effect, if any, the sneakers had on his free throw percentage.

58. New bow, again.

The shots may be considered Bernoulli trials. There are only two possible outcomes, hit or miss the bulls-eye. The probability of success is constant at 80%, and the shots are independent of one another. Therefore, we can model this situation with *Binom*(50, 0.8).

Let X = the number of bulls-eyes hit out of 50.

$E(X) = np = 50(0.8) = 40$ bulls-eyes hit.

$SD(X) = \sqrt{npq} = \sqrt{50(0.8)(0.2)} \approx 2.83$ bulls-eyes.

Answers may vary. The archer's performance doesn't seem to have increased. 45 bulls-eyes is $(45-40)/2.83 \approx 1.77$ standard deviations above the mean. This isn't unusual for an archer of her skill level.

59. Web visitors.

a) The Poisson model because it is a good model to use when the data consists of counts of occurrences. The events must be independent and the mean number of occurrences stays constant.

b) Probability that in any one minute at least one purchase is made:

$$P(X=1)+P(X=2)+\ldots=\frac{e^{-3}3^1}{1!}+\frac{e^{-3}3^2}{2!}+\ldots=0.9502$$

c) Probability that no one makes a purchase in the next 2 minutes:

$$P(X=0)=\frac{e^{-6}6^0}{0!}=0.0025$$

60. Quality control.

a) The Poisson model because it is a good model to use when the data consists of counts of occurrences. The events must be independent and the mean number of occurrences stays constant.

b) Probability that no faulty cell phones will be produced tomorrow:

$$P(X=0)=\frac{e^{-2}2^0}{0!}=0.1353$$

c) Probability that 3 or more faulty cell phones were produced today:

$$P(X=3)+P(X=4)+\ldots=\frac{e^{-2}2^3}{3!}+\frac{e^{-2}2^4}{4!}+\frac{e^{-2}3^5}{5!}+\ldots=0.3233$$

61. Web visitors, part 2.

a) The exponential model can be used to model the time between events especially when the number of arrivals of those events can be modeled by a Poisson model.

b) The mean time between purchases: 3 purchases are made per minute so each purchase is made every 1/3 minute.

c) Probability that the time to the next purchase will be between 1 and 2 minutes:

$$e^{-3*1}-e^{-3*2}=0.0473$$

62. Quality control, part 2.

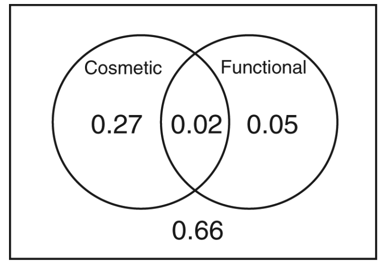
a) The exponential model can be used to model the time between events especially when the number of arrivals of those events can be modeled by a Poisson model.

b) The probability that the time to the next failure is 1 day or less:
$e^{-2*0}-e^{-2*1}=0.8647$

c) The mean time between failures: the mean number of defective cell phones is 2 per day so a defective cell phone occurs every ½ day.

Review of Part IV – Randomness and Probability

1. Quality Control.



a) $P(\text{defect}) = P(\text{cosm.}) + P(\text{func.}) - P(\text{cosm. and func.})$
$= 0.29 + 0.07 - 0.02 = 0.34$

Or, from the Venn: 0.27 + 0.02 + 0.05 = 0.34

b) $P(\text{cosm. and no func.})$

$= P(\text{cosm.}) - P(\text{cosm. and func.})$
$= 0.29 - 0.02 = 0.27$

Or, from the Venn: 0.27 (region inside Cosmetic circle, outside Functional circle)

c) $$P(\text{func.} \mid \text{cosm.}) = \frac{P(\text{func. and cosm.})}{P(\text{cosm.})} = \frac{0.02}{0.29} \approx 0.069$$

From the Venn, consider only the region inside the Cosmetic circle. The probability that the car has a functional defect is 0.02 out of a total of 0.29 (the entire Cosmetic circle).

d) The two kinds of defects are not disjoint events, since 2% of cars have both kinds.

e) Approximately 6.9% of cars with cosmetic defects also have functional defects. Overall, the probability that a car has a cosmetic defect is 7%. The probabilities are estimates, so these are probably close enough to say that they two types of defects are independent.

2. Workers.

Organize the counts in a two-way table.

Job Type	Male	Female	Total
Management	7	6	13
Supervision	8	12	20
Production	45	72	117
Total	60	90	150

a) i) $P(\text{female}) = \dfrac{90}{150} = 0.6$

ii)

$$P(\text{female or production}) = P(\text{female}) + P(\text{pr.}) - P(\text{female and pr.})$$
$$= \frac{90}{150} + \frac{117}{150} - \frac{72}{150} = 0.9$$

iii) Consider only the production row of the table. There are 72 women out of 117 production workers. $72/117 \approx 0.615$. Or, use the formula:

$$P(\text{female} \mid \text{production}) = \frac{P(\text{female and production})}{P(\text{production})} = \frac{72/150}{117/150} \approx 0.615$$

iv) Consider only the female column. There are 72 production workers out of a total of 90 women. 72/90 = 0.8. Or, use the formula:

$$P(\text{production} \mid \text{female}) = \frac{P(\text{production and female})}{P(\text{female})} = \frac{72/150}{90/150} = 0.8$$

b) These data suggest that holding a production position may be associated with whether the worker is male or female.

60% of the plant employees are women, but 61.5% of the production workers are women. However, this is a small difference, and may be due to sampling error.

3. Airfares.

a) Let C = the price of a ticket to China
Let F = the price of a ticket to France.

Total price of airfare = $3C + 5F$

b) $\mu = E(3C + 5F) = 3E(C) + 5E(F) = 3(1000) + 5(500) = \$5500$

$\sigma = SD(3C + 5F) = \sqrt{3^2(Var(C)) + 5^2(Var(F))} = \sqrt{3^2(150^2) + 5^2(100^2)} \approx \$672.68$

c) $\mu = E(C - F) = E(C) - E(F) = 1000 - 500 = \$500$

$\sigma = SD(C - F) = \sqrt{Var(C) + Var(F)}$

$= \sqrt{150^2 + 100^2} \approx \$180.28$

d) No assumptions are necessary when calculating means. When calculating standard deviations, we must assume that ticket prices are independent of each other for different countries but all tickets to the same country are at the same price.

4. Bipolar.

Let X = the number of people with bipolar disorder in a city of n = 10,000 residents.

These may be considered Bernoulli trials. There are only two possible outcomes, having bipolar disorder or not having bipolar disorder. Psychiatrists estimate that the probability that a person has bipolar is about 1 in 100, so p = 0.01. We will assume that the cases of bipolar disorder are distributed randomly throughout the populations. The trials are not independent, since the population is finite, but 10,000 people represent fewer than 10% of all people. Therefore, the number of people with bipolar disorder in a city of 10,000 may be modeled by *Binom*(10000, 0.01).

Since np = 100 and nq = 9900 are both greater than 10, $Binom(10000, 0.01)$ may be approximated by the Normal model, $N(100, 9.95)$.

$E(X) = np = 10,000(0.01) = 100$ residents.

$SD(X) = \sqrt{npq} = \sqrt{10,000(0.01)(0.99)} \approx 9.95$ residents.

We expect 100 city residents to have bipolar disorder. According to the Normal model, 200 cases would be over 10 standard deviations above this mean. The probability of this occurring is essentially zero.

Technology can compute the probability according to the Binomial model. Again, the probability that 200 cases of bipolar disorder exist in the city is essentially zero. We use the Normal model in this case, since it gives us a more intuitive idea of just how unlikely this event is.

5. A game.

a) Let X = net amount won

X	\$0	\$2	– \$2
$P(X)$	0.10	0.40	0.50

$\mu = E(X) = 0(0.10) + 2(0.40) - 2(0.50) = -\$0.20$

$\sigma^2 = Var(X) = (0-(-0.20))^2(0.10) + (2-(-0.20))^2(0.40) + (-2-(-0.20))^2(0.50) = 3.56$

$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{3.56} \approx \$1.89$

b) $X + X$ = the total winnings for two plays.

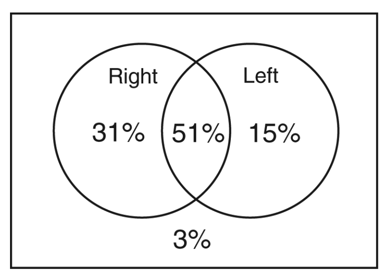
$\mu = E(X+X) = E(X) + E(X) = (-0.20) + (-0.20) = -\$0.40$

$\sigma = SD(X+X) = \sqrt{Var(X) + Var(X)}$

$= \sqrt{3.56 + 3.56} \approx \$2.67$

6. Emergency switch.

Construct a Venn diagram of the disjoint outcomes.



a) From the Venn diagram, 3% of the workers were unable to operate the switch with either hand.

b) $P(\text{left} \mid \text{right}) = \dfrac{P(\text{ left and right})}{P(\text{right})} = \dfrac{0.51}{0.82} \approx 0.622$

About 62% of the workers who could operate the switch with their right hands could also operate it with left hands. Overall, the probability that a worker could operate the switch with his right hand was 66%. Workers who could operate the switch with their right hands were less likely to be able to operate the switch with their left hand, so success is not independent of hand.

c) Success with right and left hands are not disjoint events. 51% of the workers had success with both hands.

7. Twins.

The selection of these women can be considered Bernoulli trials. There are two possible outcomes, twins or no twins. As long as the women selected are representative of the population of all pregnant women, then $p = 1/90$. (If the women selected are representative of the population of women taking Clomid, then $p = 1/10$.) The trials are not independent since the population of all women is finite, but 10 women are fewer than 10% of the population of women.

Let X = the number of twin births from $n = 10$ pregnant women.

Let Y = the number of twin births from $n = 10$ pregnant women taking Clomid.

a) Use $Binom(10, 1/90)$

$$\begin{aligned} P(\text{at least one has twins}) &= 1 - P(\text{none have twins}) \\ &= 1 - P(X = 0) \\ &= 1 - {}_{10}C_0 \left(\frac{1}{90}\right)^0 \left(\frac{89}{90}\right)^{10} \\ &\approx 0.106 \end{aligned}$$

b) Use $Binom(10, 1/10)$

$$\begin{aligned} P(\text{at least one has twins}) &= 1 - P(\text{none have twins}) \\ &= 1 - P(Y = 0) \\ &= 1 - {}_{10}C_0 \left(\frac{1}{10}\right)^0 \left(\frac{9}{10}\right)^{10} \\ &\approx 0.651 \end{aligned}$$

c) Use $Binom(5, 1/90)$ and $Binom(5, 1/10)$.

$$\begin{aligned} &P(\text{at least one has twins}) \\ &= 1 - P(\text{no twins without Clomid})P(\text{no twins with Clomid}) \\ &= 1 - \left({}_5C_0 \left(\frac{1}{90}\right)^0 \left(\frac{89}{90}\right)^5\right)\left({}_5C_0 \left(\frac{1}{10}\right)^0 \left(\frac{9}{10}\right)^5\right) \approx 0.442 \end{aligned}$$

8. Deductible.

$$\mu = E(\text{cost}) = 500(0.005) = \$2.50$$

$$\sigma^2 = Var(\text{cost}) = (2.50 - 500)^2(0.005) + (2.50 - 0)^2(0.995) = 1243.75$$

$$\sigma = SD(\text{cost}) = \sqrt{Var(\text{cost})} = \sqrt{1243.75} \approx \$35.27$$

Expected (extra) cost of the cheaper policy with the deductible is \$2.50, much less than the \$12 surcharge for the policy with no deductible, so on average she will save money by going with the deductible. The standard deviation, at \$35.27, is quite high compared to the \$12 surcharge, indicating a high amount of variability. The value of the car shouldn't influence the decision.

9. More twins.

In the previous exercise, it was determined that these were Bernoulli trials. Use *Binom*(5, 0.10).

Let X = the number of twin births from n = 5 pregnant women taking Clomid.

a)

$$\begin{aligned} P(\text{none have twins}) &= P(X=0) \\ &= {}_5C_0(0.1)^0(0.9)^5 \\ &\approx 0.590 \end{aligned}$$

b)

$$\begin{aligned} P(\text{one has twins}) &= P(X=1) \\ &= {}_5C_1(0.1)^1(0.9)^4 \\ &\approx 0.328 \end{aligned}$$

c)

$$\begin{aligned} P(\text{at least three will have twins}) &= P(X=3)+P(X=4)+P(X=5) \\ &= {}_5C_3(0.1)^3(0.9)^2 + {}_5C_4(0.1)^4(0.9)^1 + {}_5C_5(0.1)^5(0.9)^0 \\ &= 0.00856 \end{aligned}$$

10. At fault.

If we assume that these drivers are representative of all drivers insured by the company, then these insurance policies can be considered Bernoulli trials. There are only two possible outcomes, accident or no accident. The probability of having an accident is constant, p = 0.005. The trials are not independent, since the populations of all drivers is finite, but 1355 drivers represent fewer than 10% of all drivers. Use *Binom*(1355, 0.005).

a) Let X = the number of drivers who have an at-fault accident out of n = 1355.

$E(X) = np = 1{,}355(0.005) = 6.775$ drivers.

$SD(X) = \sqrt{npq} = \sqrt{1{,}355(0.005)(0.995)} \approx 2.60$ drivers.

b) Since np = 6.775 < 10, the Normal model cannot be used to model the number of drivers who are expected to have accidents. The Success/Failure condition is not satisfied.

11. Twins, part III.

In a previous exercise, it was determined that these were Bernoulli trials. Use *Binom*(152, 0.10).

Let X = the number of twin births from n = 152 pregnant women taking Clomid.

a) $E(X) = np = 152(0.10) = 15.2$ births.

$SD(X) = \sqrt{npq} = \sqrt{152(0.10)(0.90)} \approx 3.70$ births.

b) Since np = 15.2 and nq = 136.8 are both greater than 10, the Success/Failure condition is satisfied and *Binom*(152, 0.10) may be approximated by N(15.2, 3.70).

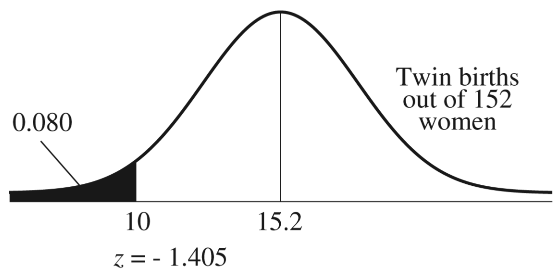
c) Using *Binom*(152, 0.10):

$$\begin{aligned} P(\text{no more than } 10) &= P(X \le 10) \\ &= P(X=0)+\ldots+P(X=10) \\ &= {}_{152}C_0(0.10)^0(0.90)^{152}+\ldots+{}_{152}C_{10}(0.10)^{10}(0.90)^{142} \\ &\approx 0.097 \end{aligned}$$

According to the Binomial model, the probability that no more than 10 women would have twins is approximately 0.097.

Using *N*(15.2, 3.70):

$$z=\frac{x-\mu}{\sigma}$$
$$z=\frac{10-15.2}{3.70}$$
$$z\approx -1.405$$



$P(X \le 10) \approx P(z < -1.405) \approx 0.080$

According to the Normal model, the probability that no more than 10 women would have twins is approximately 0.080.

12. Child's play.

a) Let X = the number indicated on the spinner

X	5	10	20
$P(X)$	0.5	0.25	0.25

b) $\mu = E(X) = 5(0.5)+10(0.25)+20(0.25)=10$

$\sigma^2 = Var(X) = (5-10)^2(0.5)+(10-10)^2(0.25)+(20-10)^2(0.25)=37.5$

$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{37.5} \approx 6.12$

c) Let Y = the number indicated on the die

Y	0	1	2	3	4
$P(Y)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

d) $\mu = E(Y) = 0\left(\frac{1}{3}\right) + 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) = \frac{10}{6} \approx 1.67$

$$Var(Y) = \left(0 - \frac{10}{6}\right)^2\left(\frac{1}{3}\right) + \left(1 - \frac{10}{6}\right)^2\left(\frac{1}{6}\right) + \left(2 - \frac{10}{6}\right)^2\left(\frac{1}{6}\right)$$
$$+ \left(3 - \frac{10}{6}\right)^2\left(\frac{1}{6}\right) + \left(4 - \frac{10}{6}\right)^2\left(\frac{1}{6}\right) \approx 2.22$$

$\sigma = SD(Y) = \sqrt{Var(Y)} = \sqrt{2.22} \approx 1.49$

e) $\mu = E(X+Y) = E(X) + E(Y) \approx 10 + 1.67 \approx 11.67$ spaces

$$\sigma = SD(X+Y) = \sqrt{Var(X) + Var(Y)}$$
$$\approx \sqrt{37.5 + 2.22} \approx 6.30 \text{ spaces}$$

13. Language.

Assuming that the freshman composition class consists of 25 randomly selected people, these may be considered Bernoulli trials. There are only two possible outcomes, having the specified language center or not having the specified language center. The probabilities of the specified language centers are constant at 80%, 10%, or 10%, for right, left, and two-sided language center, respectively. The trials are not independent, since the population of people is finite, but we will select fewer than 10% of all people.

a) Let L = the number of people with left-brain language control from n = 25 people.

Use *Binom*(25, 0.80).

$$\begin{aligned} P(\text{no more than } 15) &= P(L \le 15) \\ &= P(L=0) + \ldots + P(L=15) \\ &= {}_{25}C_0(0.80)^0(0.20)^{25} + \ldots + {}_{25}C_{15}(0.80)^{15}(0.20)^{10} \\ &\approx 0.0173 \end{aligned}$$

According to the Binomial model, the probability that no more than 15 students in a class of 25 will have left-brain language centers is approximately 0.0173.

b) Let T = the number of people with two-sided language control from n = 5 people.

Use *Binom*(5, 0.10).

$$\begin{aligned} P(\text{none have two-sided language control}) &= P(T=0) \\ &= {}_5C_0(0.10)^0(0.90)^5 \\ &\approx 0.590 \end{aligned}$$

c) Use Binomial models:

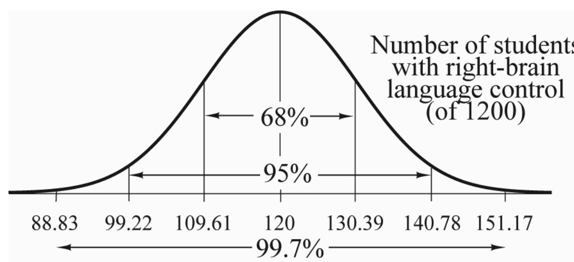
$E(\text{left}) = np_L = 1200(0.80) = 960$ people
$E(\text{right}) = np_R = 1200(0.10) = 120$ people
$E(\text{two}-\text{sided}) = np_T = 1200(0.10) = 120$ people

d) Let R = the number of people with right-brain language control.

$E(R) = np_R = 1200(0.10) = 120$ people

$SD(R) = \sqrt{np_R q_R} = \sqrt{1200(0.10)(0.90)} \approx 10.39$ people.

e) Since $np_R = 120$ and $nq_R = 1080$ are both greater than 10, the Normal model, $N(120, 10.39)$, may be used to approximate $Binom(1200, 0.10)$. According to the Normal model, about 68% of randomly selected groups of 1200 people could be expected to have between 109.61 and 130.39 people with right-brain language control. About 95% of randomly selected groups of 1200 people could be expected to have between 99.22 and 140.78 people with right-brain language control. About 99.7% of randomly selected groups of 1200 people could be expected to have between 88.83 and 151.17 people with right-brain language control.



14. Play again.

$\mu = E(X-Y) = E(X) - E(Y) \approx 10 - 1.67 \approx 8.33$ spaces

$$\sigma = SD(X-Y) = \sqrt{Var(X) + Var(Y)}$$
$$\approx \sqrt{37.5 + 2.22} \approx 6.30 \text{ spaces}$$

15. Beanstalks.

a) The greater standard deviation for men's heights indicates that men's heights are more variable than women's heights.

b) Admission to a Beanstalk Club is based upon extraordinary height for both men and women, but men are slightly more likely to qualify. The qualifying height for women is about 2.4 SDs above the mean height of women, while the qualifying height for men is about 1.75 SDS above the mean height for men.

c) Let M = the height of a randomly selected man from $N(69.1, 2.8)$.

Let W =the height of a randomly selected woman from $N(64.0, 2.5)$.

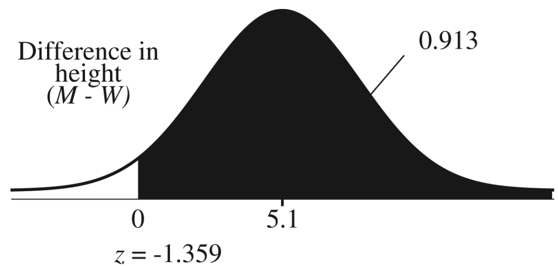
$M - W$ = the difference in height of a randomly selected man and woman.

d) $E(M-W) = E(M) - E(W) = 69.1 - 64.0 = 5.1$ inches

e) $SD(M-W)=\sqrt{Var(M)+Var(W)}=\sqrt{2.8^2+2.5^2}\approx 3.75$ inches

f) Since each distribution is described by a Normal model, the distribution of the difference in height between a randomly selected man and woman is $N(5.1, 3.75)$.

$$z=\frac{y-\mu}{\sigma}$$
$$z=\frac{0-5.1}{\sqrt{2.8^2+2.5^2}}$$
$$z\approx -1.359$$



According to the Normal model, the probability that a randomly selected man is taller than a randomly selected woman is approximately 0.913.

g) If people chose spouses independent of height, we would expect 91.3% of married couples to consist of a taller husband and shorter wife. The 92% that was seen in the survey is close to 91.3%, and the difference may be due to natural sampling variability. Unless this survey is very large, there is not sufficient evidence of association between height and choice of partner.

16. Stocks.

a) $P(\text{market will rise for 3 consecutive years})=(0.73)^3\approx 0.389$

b) Use *Binom*(5, 0.73).

$P(\text{market will rise in 3 out of 5 years})={}_5C_3(0.73)^3(0.27)^2\approx 0.284$

c) $P(\text{fall in at least 1 of next 5 years})=1-P(\text{no fall in 5 years})=1-(0.73)^5\approx 0.793$

d) Let X = the number of years in which the market rises. Use *Binom*(10, 0.73).
$P(\text{rises in the majority of years in a decade})$
$=P(X\geq 6)$
$=P(X=6)+\ldots+P(X=10)$
$={}_{10}C_6(0.73)^6(0.27)^4+\ldots+{}_{10}C_{10}(0.73)^{10}(0.27)^0$
≈ 0.896

17. Multiple choice.

Guessing at questions can be considered Bernoulli trials. There are only two possible outcomes, correct or incorrect. If you are guessing, the probability of success is p = 0.25, and the questions are independent. Use *Binom*(50, 0.25) to model the number of correct guesses on the test.

a) Let $X =$ the number of correct guesses.

$$\begin{aligned} P(\text{at least 30 of 50 correct}) &= P(X \geq 30) \\ &= P(X = 30) + \ldots + P(X = 50) \\ &= {}_{50}C_{30}(0.25)^{30}(0.75)^{20} + \ldots + {}_{50}C_{30}(0.25)^{50}(0.75)^{0} \\ &\approx 0.00000016 \end{aligned}$$

You are **very** unlikely to pass by guessing on every question.

b) Use *Binom*(50, 0.70).

$$\begin{aligned} P(\text{at least 30 of 50 correct}) &= P(X \geq 30) \\ &= P(X = 30) + \ldots + P(X = 50) \\ &= {}_{50}C_{30}(0.70)^{30}(0.30)^{20} + \ldots + {}_{50}C_{50}(0.70)^{50}(0.30)^{0} \\ &\approx 0.952 \end{aligned}$$

According to the Binomial model, your chances of passing are about 95.2%.

c) Use *Geom*(0.70).

$P(\text{first correct on third question}) = (0.30)^2(0.70) = 0.063$

18. Stock strategy.

a) This does not confirm the advice. Stocks have risen 75% of the time after a two-year fall, but there have only been eight occurrences of the two-year fall. The sample size is very small, and therefore highly variable.

b) Stocks have actually risen in 73% of years. This is not much different from the strategy of the advisors, which yielded a rise in 75% of years (from a very small sample of years.)

19. Insurance.

The company is expected to pay \$100,000 only 2.6% of the time, while always gaining \$520 from every policy sold. When they pay, they actually only pay \$99,480.

E(profit) = \$520(0.974) – \$99,480(0.026) = – \$2,080.

The expected profit is actually a **loss** of \$2,080 per policy. The company had better raise its premiums if it hopes to stay in business.

20. Teen smoking.

Randomly selecting high school students can be considered Bernoulli trials. There are only two possible outcomes, smoker or nonsmoker. The probability that a student is a smoker is $p = 0.18$. The trials are not independent, since the population is finite, but we are not sampling more than 10% of all high school students.

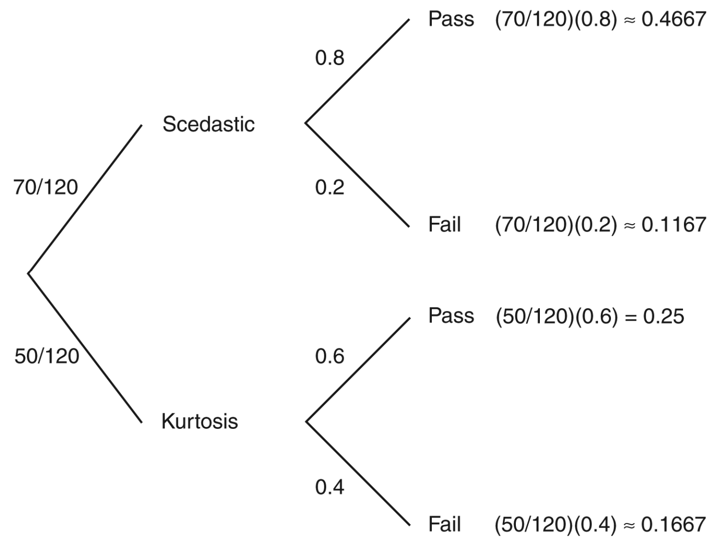
a) $P(\text{none of the first 4 are smokers}) = (0.82)^4 = 0.452$

b) Use *Geom*(0.18).
$P(\text{first smoker is the sixth person}) = (0.82)^5(0.18) \approx 0.067$

c) Use *Binom*(10, 0.18). Let X = the number of smokers among n = 10 students.
$P(\text{no more than 2 smokers of 10})$
$= P(X \le 2)$
$= P(X=0)+P(X=1)+P(X=2)$
$= {}_{10}C_0(0.18)^0(0.82)^{10} + {}_{10}C_1(0.18)^1(0.82)^9 + {}_{10}C_2(0.18)^2(0.82)^8$
≈ 0.737

21. Passing stats.

Organize the information in a tree diagram.



a)
$$\begin{aligned} &P(\text{Passing Statistics}) \\ &\quad = P(\text{Scedastic and Pass}) \\ &\qquad + P(\text{Kurtosis and Pass}) \\ &\quad \approx 0.4667 + 0.25 \\ &\quad \approx 0.717 \end{aligned}$$

b)
$$P(\text{Kurtosis} \mid \text{Fail}) = \frac{P(\text{Kurtosis and Fail})}{P(\text{Fail})} \approx \frac{0.1667}{0.1167+0.1667} \approx 0.588$$

22. Teen smoking II.

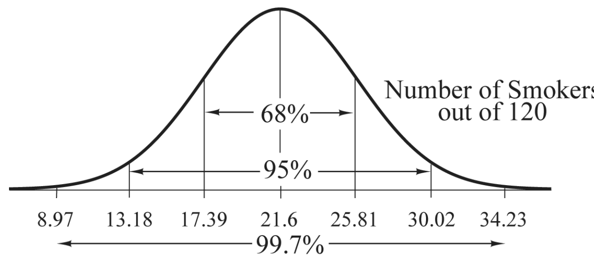
In a previous exercise, it was determined that the selection of students could be considered to be Bernoulli trials.

a) Use *Binom*(120, 0.18) to model the number of smokers out of n = 120 students.
$E(\text{number of smokers}) = np = 120(0.18) = 21.6$ smokers.

b) $SD(\text{number of smokers}) = \sqrt{npq} = \sqrt{120(0.18)(0.82)} \approx 4.21$ smokers.

c) Since np = 21.6 and nq = 98.4 are both greater than 10, the Success/Failure condition is satisfied and *Binom*(120, 0.18) may be approximated by N(21.6, 4.21).

d) According to the Normal model, approximately 68% of samples of size $n = 120$ are expected to have between 17.39 and 25.81 smokers, approximately 95% of the samples are expected to have between 13.18 and 30.02 smokers, and approximately 99.7% of the samples are expected to have between 8.97 and 34.23 smokers.



23. Random variables.

a)
$$\mu = E(X+50) = E(X)+50 = 50+50 = 100$$
$$\sigma = SD(X+50) = SD(X) = 8$$

b)
$$\mu = E(10Y) = 10E(Y) = 10(100) = 1000$$
$$\sigma = SD(10Y) = 10SD(Y) = 60$$

c)
$$\mu = E(X+0.5Y) = E(X)+0.5E(Y) = 50+0.5(100) = 100$$
$$\sigma = SD(X+0.5Y) = \sqrt{Var(X)+0.5^2 Var(Y)} = \sqrt{8^2+0.5^2(6^2)} \approx 8.54$$

d)
$$\mu = E(X-Y) = E(X)-E(Y) = 50-100 = -50$$
$$\sigma = SD(X-Y) = \sqrt{Var(X)+Var(Y)} = \sqrt{8^2+6^2} = 10$$

e)
$$\mu = E(X_1+X_2) = E(X)+E(X) = 50+50 = 100$$
$$\sigma = SD(X_1+X_2) = \sqrt{Var(X)+Var(X)} = \sqrt{8^2+8^2} \approx 11.31$$

24. Merger.

Small companies may run in to trouble in the insurance business. Even if the expected profit from each policy is large, the profit is highly variable. There is a small chance that a company would have to make several huge payouts, resulting in an overall loss, not a profit. By combining two small companies together, the company takes in profit from more policies, making the larger company more resistant to the possibility of a large payout. This is because the total profit is increasing by the expected profit from each additional policy, but the standard deviation is increasing by the square root of the sum of the variances. The larger a company gets, the more the expected profit outpaces the variability associated with that profit.

25. Youth survey.

a) Many boys play computer games and use email, so the probabilities can total more than 100%. There is no evidence that there is a mistake in the report.

b) Playing computer games and using email are not mutually exclusive. If they were, the probabilities would total 100% or less.

c) Emailing friends and being a boy or girl are not independent. 76% of girls emailed friends in the last week, but only 65% of boys emailed. If emailing were independent of being a boy or girl, the probabilities would be the same.

d) Let X = the number of students chosen until the first student is found who does not use the Internet. Use $Geom(0.07)$. $P(X=5)=(0.93)^4(0.07)\approx 0.0524$.

26. Meals.

Let X = the amount the student spends daily.

a) $\mu = E(X+X)=E(X)+E(X)=13.50+13.50=\$27.00$

$$\sigma = SD(X+X)=\sqrt{Var(X)+Var(X)}$$
$$=\sqrt{7^2+7^2}\approx \$9.90$$

b) In order to calculate the standard deviation, we must assume that spending on different days is independent. This is probably not valid, since the student might tend to spend less on a day after he has spent a lot. He might not even have money left to spend!

c) $\mu = E(X+X+X+X+X+X+X)=7E(X)=7(\$13.50)=\$94.50$

$$\sigma = SD(X+X+X+X+X+X+X)$$
$$=\sqrt{Var(X)+Var(X)+Var(X)+Var(X)+Var(X)+Var(X)+Var(X)}$$
$$=\sqrt{7(7^2)}\approx \$18.52$$

d) Assuming once again that spending on different days is independent, it is unlikely that the student will spend less than \$50. This level of spending is about 2.4 standard deviations below the weekly mean. Don't try to approximate the probability! We don't know the shape of this distribution.

27. Travel to Kyrgyzstan.

a) Spending an average of 4237 soms per day, you can stay about $\frac{90{,}000}{4237}\approx 21$ days.

b) Assuming that your daily spending is independent, the standard deviation is the square root of the sum of the variances for 21 days.

$$\sigma = \sqrt{21(360)^2}\approx 1649.73 \text{ soms}$$

c) The standard deviation in your total expenditures is about 1650 soms, so if you don't think you will exceed your expectation by more than 2 standard deviations, bring an extra 3300 soms. This gives you a cushion of about 157 soms for each of the 21 days.

28. Picking melons.

a) $\mu = E(\text{First} - \text{Second}) = E(\text{First}) - E(\text{Second}) = 22 - 18 = 4$ lbs.

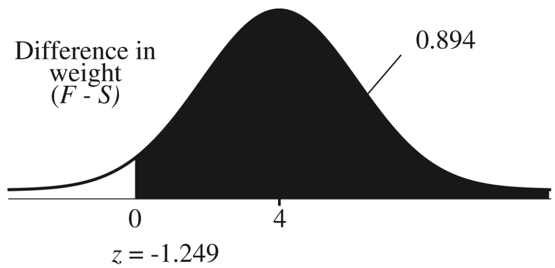
b) $\sigma = SD(\text{First} - \text{Second}) = \sqrt{Var(\text{First}) + Var(\text{Second})} = \sqrt{2.5^2 + 2^2} \approx 3.20$ lbs.

c)

$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{0 - 4}{\sqrt{2.5^2 + 2^2}}$$

$$z \approx -1.249$$



According to the Normal model, the probability that a melon from the first store weighs more than a melon from the second store is approximately 0.894.

29. Home sweet home.

Since the homes are randomly selected, these can be considered Bernoulli trials. There are only two possible outcomes, owning the home or not owning the home. The probability of any randomly selected resident home being owned by the current resident is 0.66. The trials are not independent, since the population is finite, but as long as the city has more than 8200 homes, we are not sampling more than 10% of the population. The Binomial model, *Binom*(820, 0.66), can be used to model the number of homeowners among the 820 homes surveyed. Let H = the number of homeowners found in n = 820 homes.

$E(H) = np = 820(0.66) = 541.2$ homes.

$SD(H) = \sqrt{npq} = \sqrt{820(0.66)(0.34)} \approx 13.56$ homes.

The 523 homeowners found in the candidate's survey represent a number of homeowners that is only about 1.34 standard deviations below the expected number of homeowners. It is not particularly unusual to be 1.34 standard deviations below the mean. There is little support for the candidate's claim of a low level of home ownership.

30. Buying melons.

The mean price of a watermelon at the first store is 22(0.32) = \$7.04.
At the second store the mean price is 18(0.25) = \$4.50.
The difference in the price is expected to be \$7.04 – \$4.50 = \$2.54.

The standard deviation in price at the first store is 2.5(0.32) = \$0.80.
At the second store, the standard deviation in price is 2(0.25) = \$0.50.
The standard deviation of the difference is $\sqrt{0.80^2 + 0.50^2} \approx \$0.94$.

31. Who's the boss?

a) $P(\text{first three owned by women}) = (0.30)^3 \approx 0.027$

b) $P(\text{none of the first four are owned by women}) = (0.70)^4 \approx 0.240$

c) $P(\text{sixth firm called is owned by women} \mid \text{none of the first five were}) = 0.30$
Since the firms are chosen randomly, the fact that the first five firms were owned by men has no bearing on the ownership of the sixth firm.

32. Jerseys.

a) $P(\text{all four kids get the same color}) = 4\left(\dfrac{1}{4}\right)^4 \approx 0.0156$

(There are four different ways for this to happen, one for each color.)

b) $P(\text{all four kids get white}) = \left(\dfrac{1}{4}\right)^4 \approx 0.0039$

c) $P(\text{all four kids get white}) = \left(\dfrac{1}{6}\right)\left(\dfrac{1}{4}\right)^3 \approx 0.0026$

33. When to stop?

a) Since there are only two outcomes, 6 or not 6, the probability of getting a 6 is 1/6, and the trials are independent, these are Bernoulli trials. Use *Geom*(1/6).

$$\mu = \frac{1}{p} = \frac{1}{1/6} = 6 \text{ rolls}$$

b) If 6's are not allowed, the mean of each die roll is $\dfrac{1+2+3+4+5}{5} = 3$. You would expect to get 15 if you rolled 5 times.

c) $P(\text{5 rolls without a 6}) = \left(\dfrac{5}{6}\right)^5 \approx 0.402$

34. Plan B.

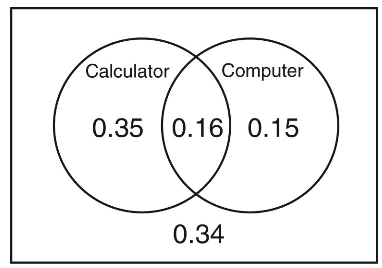
a) If 6's are not allowed, the mean of each die roll is $\dfrac{1+2+3+4+5}{5}=3$.

b) Let X = your current score. You expect to lose it all $\dfrac{1}{6}$ of the time, so your expected loss per roll is $\dfrac{1}{6}X$.

c) Expected gain equals expected loss when $\dfrac{1}{6}X=3$. So, $X=18$.

d) Roll until you get 18 points, then stop.

35. Technology on campus.



a)

$$\begin{aligned} &P(\text{neither tech.}) \\ &=1-P(\text{either tech.}) \\ &=1-\left[P(\text{calculator})+P(\text{computer})-P(\text{both})\right] \\ &=1-[0.51+0.31-0.16]=0.34 \end{aligned}$$

Or, from the Venn: 0.34 (the region outside both circles) This is MUCH easier.

34% of students use neither type of technology.

b) $P(\text{calc. and no comp.})=P(\text{calc.})-P(\text{calc. and comp.})=0.51-0.16=0.35$

Or, from the Venn: 0.35 (region inside the Calculator circle, outside the Computer circle)

35% of students use calculators, but not computers.

c) $P(\text{computer} \mid \text{calculator})=\dfrac{P(\text{comp. and calc.})}{P(\text{calc.})}=\dfrac{0.16}{0.51}\approx 0.314$

About 31.4% of calculator users have computer assignments.

d) The percentage of computer users overall is 31%, while 31.4% of calculator users were computer users. These are very close. There is no indication of an association between computer use and calculator use.

36. Dogs.

Since the outcomes are disjoint, probabilities may be added and subtracted as needed.

a) $P(\text{no dogs})=(0.77)(0.77)=0.5929$

b) $P(\text{some dogs})=1-P(\text{no dogs})=1-(0.77)(0.77)=0.4071$

c) $P(\text{both dogs}) = (0.23)(0.23) = 0.0529$

d) $P(\text{more than one dog in each}) = (0.05)(0.05) \approx 0.0025$

37. O-rings.

a) A Poisson model would be used.

b) If the probability of failure for one O-ring is 0.01, then the mean number of failures for 10 O-rings is $E(X) = \lambda = np = 10(0.01) = 0.1$ O-ring. We are able to calculate this because the Poisson model scales according to sample size.

c) $P(\text{one failed O-ring}) = \dfrac{e^{-0.1}(0.1)^1}{1!} \approx 0.090$

d)

$$\begin{aligned} P(\text{at least one failed O-ring}) &= 1 - P(\text{no failures}) \\ &= 1 - \frac{e^{-0.1}(0.1)^0}{0!} \approx 0.095 \end{aligned}$$

38. Volcanoes.

a) A Poisson model would be used.

b) Since the mean number of volcanic episodes is 2.4 per year, then the mean number of volcanic episodes for 2 years is 4.8.

c) $P(\text{no volcanic epsisodes}) = \dfrac{e^{-4.8}(4.8)^1}{0!} \approx 0.008$

d)

$$\begin{aligned} &P(\text{more than three volcanic episodes}) \\ &= 1 - \left[P(X=0) + P(X=1) + P(X=2) + P(X=3)\right] \\ &= 1 - \left[\frac{e^{-4.8}(4.8)^0}{0!} + \frac{e^{-4.8}(4.8)^1}{1!} + \frac{e^{-4.8}(4.8)^2}{2!} + \frac{e^{-4.8}(4.8)^3}{3!}\right] \\ &\approx 0.706 \end{aligned}$$

(or, use technology. For example: 1 – poissoncdf(4.8,3) on the TI-83.)

39. Socks.

Since we are sampling without replacement, use conditional probabilities throughout.

a) $P(\text{2 blue}) = \left(\dfrac{4}{12}\right)\left(\dfrac{3}{11}\right) = \dfrac{12}{132} = \dfrac{1}{11}$

b) $P(\text{no grey}) = \left(\dfrac{7}{12}\right)\left(\dfrac{6}{11}\right) = \dfrac{42}{132} = \dfrac{7}{22}$

c) $P(\text{at least one black}) = 1 - P(\text{no black}) = 1 - \left(\frac{9}{12}\right)\left(\frac{8}{11}\right) = \frac{60}{132} = \frac{5}{11}$

d) $P(\text{green}) = 0$ (There aren't any green socks in the drawer.)

e) $P(\text{match}) = P(2\text{ blue}) + P(2\text{ grey}) + P(2\text{ black})$

$$= \left(\frac{4}{12}\right)\left(\frac{3}{11}\right) + \left(\frac{5}{12}\right)\left(\frac{4}{11}\right) + \left(\frac{3}{12}\right)\left(\frac{2}{11}\right) = \frac{19}{66}$$

40. Coins.

Coin flips are Bernoulli trials. There are only two possible outcomes, the probability of each outcome is constant, and the trials are independent.

a) Use *Binom*(36, 0.5). Let H = the number of heads in n = 36 flips.

$\mu = E(H) = np = 36(0.5) = 18$ heads.

$\sigma = SD(H) = \sqrt{npq} = \sqrt{36(0.5)(0.5)} = 3$ heads.

b) Two standard deviations above the mean corresponds to 6 "extra" heads observed.

c) The standard deviation of the number of heads when 100 coins are flipped is $\sigma = \sqrt{npq} = \sqrt{100(0.5)(0.5)} = 5$ heads. Getting 6 "extra" heads is not unusual.

d) Following the "two standard deviations" measurement, 10 or more "extra" heads would be unusual.

e) What appears surprising in the short run becomes expected in a larger number of flips. The "Law of Averages" is refuted, because the coin does not compensate in the long run. A coin that is flipped many times is actually less likely to show exactly half heads than a coin flipped only a few times. The Law of Large Numbers is confirmed, because the percentage of heads observed gets closer to the percentage expected due to probability.

41. The Drake equation.

a) $N \cdot f_p$ represents the number of stars in the Milky Way Galaxy expected to have planets.

b) $N \cdot f_p \cdot n_e \cdot f_i$ represents the number of planets in the Milky Way Galaxy expected to have intelligent life.

c) $f_l \cdot f_i$ is the probability that a planet has a suitable environment and has intelligent life.

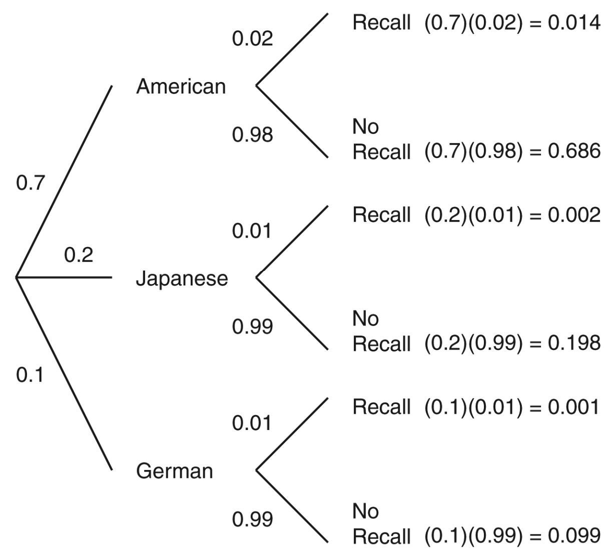
d) $f_l = P(\text{life} \mid \text{suitable environment})$. This is the probability that life develops, if a planet has a suitable environment.

$f_i = P(\text{intelligence} \mid \text{life})$. This is the probability that the life develops intelligence, if a planet already has life.

$f_c = P(\text{communication} \mid \text{intelligence})$. This is the probability that radio communication develops, if a planet already has intelligent life.

42. Recalls.

Organize the information in a tree diagram.



a)

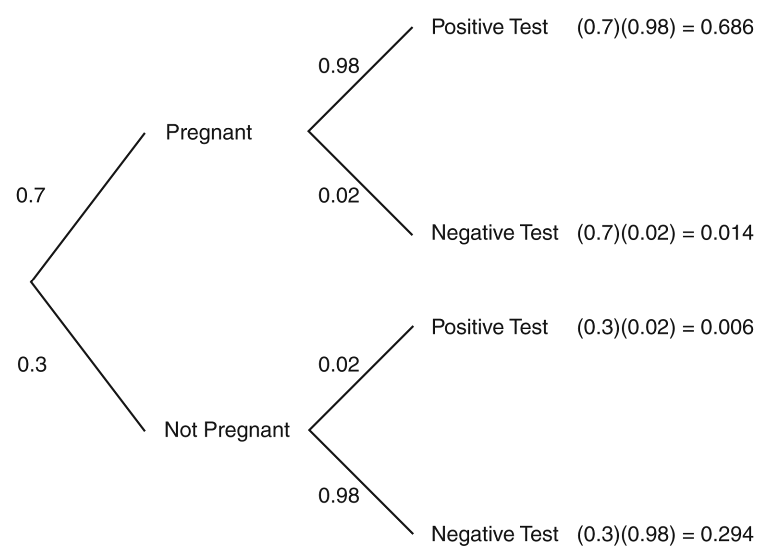
$$\begin{aligned} P(\text{recall}) &= P(\text{American recall}) \\ &\quad + P(\text{Japanese recall}) \\ &\quad + P(\text{German recall}) \\ &= 0.014 + 0.002 + 0.001 \\ &= 0.017 \end{aligned}$$

b)

$$\begin{aligned} &P(\text{American} \mid \text{recall}) \\ &= \frac{P(\text{American and recall})}{P(\text{recall})} \\ &= \frac{0.014}{0.014 + 0.002 + 0.001} \\ &\approx 0.824 \end{aligned}$$

43. Pregnant?

Organize the information in a tree diagram.



$$\begin{aligned} &P(\text{pregnant} \mid + \text{ test}) \\ &= \frac{P(\text{preg. and + test})}{P(+ \text{ test})} \\ &= \frac{0.686}{0.686 + 0.006} \\ &\approx 0.991 \end{aligned}$$

44. Door prize.

a) The probability that the first person in line wins is 1 out of 100, or 0.01.

b) If you are third in line, the two people ahead of you must not win in order for you to win.
The probability is (0.99)(0.99)(0.01) = 0.009801.

c) There must be 100 losers in a row. The probability is $(0.99)^{100} \approx 0.366$.

d) The first person in line has the greatest chance of winning at $p = 0.01$. The probability of winning decreases from there, since winning is dependent upon everyone else in front of you in line losing.

e) Position is irrelevant now. Everyone has the same chance of winning, $p = 0.01$. One way to visualize this is to imagine that one ball is handed out to each person. Only one person out of the 100 people has the red ball. It might be you!

If you insist that the probabilities are still conditional, since you are sampling without replacement, look at it this way:

Consider $P(\text{sixth person wins}) = \left(\frac{99}{100}\right)\left(\frac{98}{99}\right)\left(\frac{97}{98}\right)\left(\frac{96}{97}\right)\left(\frac{95}{96}\right)\left(\frac{1}{95}\right) = \frac{1}{100}$

Chapter 17 – Sampling Distribution Models

Section 17.1

1. Website.

a) Since the sample is drawn at random, and assuming that 200 investors is a small portion of their customers, the sampling distribution for the proportion of 200 investors that use smartphones will be unimodal and symmetric (roughly Normal).

b) The center of the sampling distribution of the sample proportion is 0.36.

c) The standard deviation of the sample proportion is $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.36)(0.64)}{200}} \approx 0.034$.

2. Marketing.

a) The proportion of women in the sample is expected to be 0.51.

b) The standard deviation of the sample proportion is $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.51)(0.49)}{400}} \approx 0.025$.

c) We would expect to find $(0.51)(400) = 204$ women in a sample of 400.

3. Send money.

All of the histograms are centered around $p = 0.05$. As n gets larger, the shape of the histograms get more unimodal and symmetric, approaching a Normal model, while the variability in the sample proportions decreases.

4. Character recognition.

All of the histograms are centered around $p = 0.85$. As n gets larger, the shapes of the histograms get more unimodal and symmetric, approaching a Normal model, while the variability in the sample proportions decreases.

Section 17.2

5. Marriage.

The data come from a random sample, so the randomization condition is met. We don't know the exact value of p, we can estimate it with $\hat{p}$.

$n\hat{p} = (1500)(0.27) = 405$, and $n\hat{q} = (1500)(0.73) = 1095$. So, there are well over 10 successes and 10 failures, meeting the Success/Failure Condition. Since there are more than $(10)(1500) = 15{,}000$ adults in the United States, the 10% Condition is met. A Normal model is appropriate for the sampling distribution of the sample proportion.

6. Campus sample.

Stacy plans to use a random sample, so the randomization condition is met. However, $np = (50)(0.10) = 5$, which is less than 10. The Success/Failure condition is not met. It is not appropriate to use a Normal model for the sampling distribution of the sample proportion.

7. Send more money.

a) The histogram for n = 200 looks quite unimodal and symmetric. We should be able to use the Normal model.

b) The Success/Failure condition requires np and nq to both be at least 10, which is not satisfied until n = 200 for p = 0.05. The theory supports the choice in part a.

8. Character recognition, again.

a) Certainly, the histogram for n = 100 is unimodal and symmetric, but the histogram for n = 75 looks nearly Normal, too. We should be able to use the Normal model for either.

b) The success/failure condition requires np and nq to both be at least 10, which is satisfied for both n = 75 and n = 100 when p = 0.85. The theory supports the choice in part a.

Section 17.3

9. Sample maximum.

a) A Normal model is not appropriate for the sampling distribution of the sample maximum. The histogram is skewed strongly to the left.

b) No. The 95% rule is based on the Normal model, and the Normal model is not appropriate here.

10. Soup.

a) A Normal model is not appropriate for the sampling distribution of the sample variances. The histogram is skewed to the right.

b) No. The 95% rule is based on the Normal model, and the Normal model is not appropriate here.

Section 17.4

11. Market research.

a) The standard deviation of the sample proportion is $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.15)(0.85)}{100}} \approx 0.0357$.

b) To reduce the standard deviation by half, she needs a sample 4 times as large, or 400 people.

12. Market research II.

a) The standard deviation of the sample proportion is $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.76)(0.24)}{500}} \approx 0.0191$.

b) The standard deviation would be $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.76)(0.24)}{125}} \approx 0.0382$. Notice that this is twice the standard deviation in part a, since the sample size is one-fourth as large.

13. Tips.

a) Since the distribution of tips is skewed to the right, we can't use the Normal model to determine the probability that a given party will tip at least $20.

b) No. A sample of 4 parties is probably not a large enough sample for the CLT to allow us to use the Normal model to estimate the distribution of averages.

c) A sample of 10 parties may not be large enough to allow the use of a Normal model to describe the distribution of averages. It would be risky to attempt to estimate the probability that his next 10 parties tip an average of $15. However, since the distribution of tips has $\mu = \$9.60$, with standard deviation $\sigma = \$5.40$, we still know that the mean of the sampling distribution model is $\mu_{\bar{y}} = \$9.60$ with standard deviation $SD(\bar{y}) = \frac{5.40}{\sqrt{10}} \approx \$1.71.$

We don't know the exact shape of the distribution, but we can still assess the likelihood of specific means. A mean tip of $15 is over 3 standard deviations above the expected mean tip for 10 parties. That's not very likely to happen.

14. Groceries.

a) Since the distribution of Sunday purchases is skewed, we can't use the Normal model to determine the probability that a given purchase is at least $40.

b) A sample of 10 customers may not be large enough for the CLT to allow the use of a Normal model for the sampling distribution model. If the distribution of Sunday purchases is only slightly skewed, the sample may be large enough, but if the distribution is heavily skewed, it would be very risky to attempt to determine the probability.

c) **Randomization condition:** Assume that the 50 Sunday purchases can be considered a representative sample of all purchases.

Independence assumption: It is reasonable to think that the Sunday purchases are mutually independent, unless there is a sale or other incentive to purchase more.

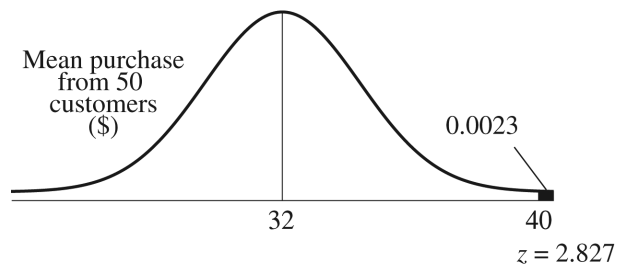
10% condition: The 50 purchases certainly represent less than 10% of all purchases.

Large Enough Sample condition: The sample of 50 purchases is large enough.

The mean Sunday purchase is $\mu = \$32$, with standard deviation $\sigma = \$20$. Since the conditions are met, the CLT allows us to model the sampling distribution of $\bar{y}$ with a Normal model, with $\mu_{\bar{y}} = \$32$ and standard deviation

$SD(\bar{y}) = \dfrac{20}{\sqrt{50}} \approx \$2.83$.

According to the Normal model, the probability the mean Sunday purchase of 50 customers is at least $40 is about 0.0023.



15. More tips.

a) **Randomization condition:** Assume that the tips from 40 parties can be considered a representative sample of all tips.

Independence assumption: It is reasonable to think that the tips are mutually independent, unless the service is particularly good or bad during this weekend.

10% condition: The tips of 40 parties certainly represent less than 10% of all tips.

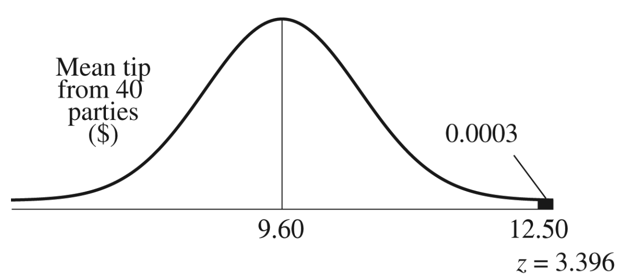
Large Enough Sample condition: The sample of 40 parties is large enough.

The mean tip is $\mu = \$9.60$, with standard deviation $\sigma = \$5.40$. Since the conditions are satisfied, the CLT allows us to model the sampling distribution of $\bar{y}$ with a Normal model, with $\mu_{\bar{y}} = \$9.60$ and standard deviation

$SD(\bar{y}) = \dfrac{5.40}{\sqrt{40}} \approx \$0.8538$.

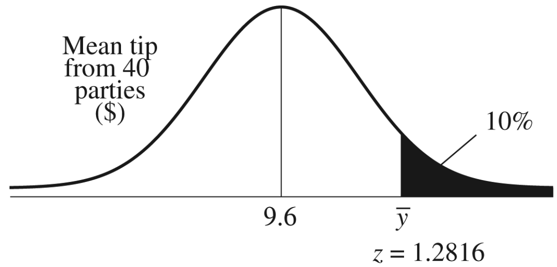
In order to earn at least $500, the waiter would have to average $\dfrac{500}{40} = \$12.50$ per party.

According to the Normal model, the probability that the waiter earns at least $500 in tips in a weekend is approximately 0.0003.



b) According to the Normal model, the waiter can expect to have a mean tip of about \$10.6942, which corresponds to about \$427.77 for 40 parties, in the best 10% of such weekends.

$$z = \frac{\bar{y} - \mu_{\bar{y}}}{SD(\bar{y})}$$
$$1.2816 = \frac{\bar{y} - 9.60}{\frac{5.40}{\sqrt{40}}}$$
$$\bar{y} \approx 10.6942$$



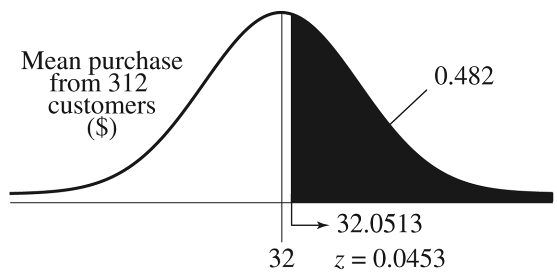
16. More groceries.

a) Assumptions and conditions for the use of the CLT were verified in a previous exercise.

The mean purchase is $\mu = \$32$, with standard deviation $\sigma = \$20$. Since the sample is large, the CLT allows us to model the sampling distribution of $\bar{y}$ with a Normal model, with $\mu_{\bar{y}} = \$32$ and standard deviation $SD(\bar{y}) = \frac{20}{\sqrt{312}} \approx \$1.1323$.

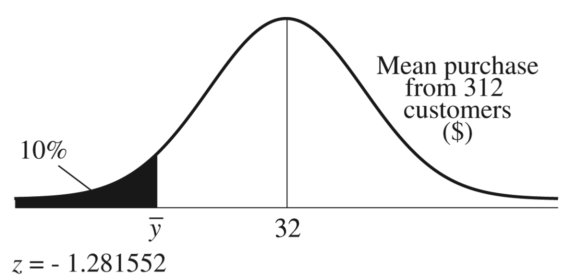
In order to have revenues of at least \$10,000, the mean Sunday purchase must be at least $\frac{10,000}{312} = \$32.0513$.

According to the Normal model, the probability of having a mean Sunday purchase at least that high (and therefore at total revenue of at least \$10,000) is 0.482.



b) According to the Normal model, the mean Sunday purchase on the worst 10% of such days is approximately \$30.548928, so 312 customers are expected to spend about \$9531.27.

$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma(\bar{y})}$$
$$-1.281552 = \frac{\bar{y} - 32}{\frac{20}{\sqrt{312}}}$$
$$\bar{y} \approx 30.548928$$



17. Coin tosses.

a) The histogram of these proportions is expected to be symmetric, but **not** because of the Central Limit Theorem. The sample of 16 coin flips is not large. The distribution of these proportions is expected to be symmetric because the probability that the coin lands heads is the same as the probability that the coin lands tails.

b) The histogram is expected to have its center at 0.5, the probability that the coin lands heads.

c) The standard deviation of data displayed in this histogram should be approximately equal to the standard deviation of the sampling distribution model, $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{16}} = 0.125.$

d) The expected number of heads, $np = 16(0.5) = 8$, which is less than 10. The Success/Failure condition is not met. The Normal model is not appropriate in this case.

18. M&M's.

a) The histogram of the proportions of green candies in the bags would probably be skewed slightly to the right, for the simple reason that the proportion of green M&M's could never fall below 0 on the left, but has the potential to be higher on the right.

b) The Normal model cannot be used to approximate the histogram, since the expected number of green M&M's is $np = 50(0.10) = 5$, which is less than 10. The Success/Failure condition is not met.

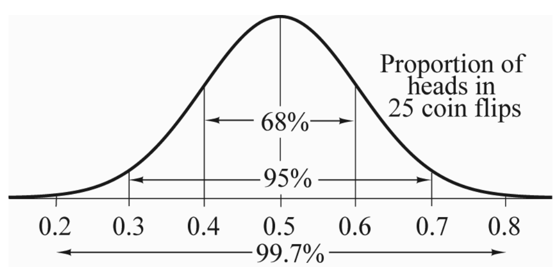
c) The histogram should be centered around the expected proportion of green M&M's, at about 0.10.

d) The proportion should have standard deviation $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.1)(0.9)}{50}} \approx 0.042.$

19. More coins.

a) $\mu_{\hat{p}} = p = 0.5$ and $SD(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{25}} = 0.1$

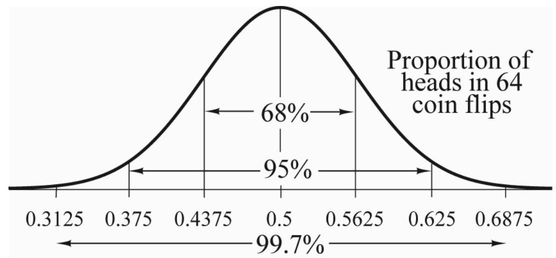
About 68% of the sample proportions are expected to be between 0.4 and 0.6, about 95% are expected to be between 0.3 and 0.7, and about 99.7% are expected to be between 0.2 and 0.8.



b) Coin flips are independent of one another. There is no need to check the 10% Condition. $np = nq = 12.5$, so both are greater than 10. The Success/Failure condition is met, so the sampling distribution model is $N(0.5, 0.1)$.

c) $\mu_{\hat{p}} = p = 0.5$ and $SD(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{64}} = 0.0625$

About 68% of the sample proportions are expected to be between 0.4375 and 0.5625, about 95% are expected to be between 0.375 and 0.625, and about 99.7% are expected to be between 0.3125 and 0.6875.



Coin flips are independent of one another, and $np = nq = 32$, so both are greater than 10. The Success/Failure condition is met, so the sampling distribution model is $N(0.5, 0.0625)$.

d) As the number of tosses increases, the sampling distribution model will still be Normal and centered at 0.5, but the standard deviation will decrease. The sampling distribution model will be less spread out.

20. Bigger bag.

a) **Randomization condition:** The 200 M&M's in the bag can be considered representative of all M&M's, and they are thoroughly mixed.

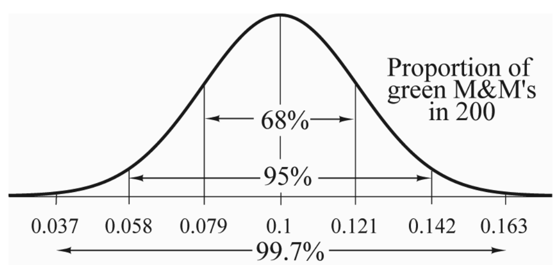
10% condition: 200 is certainly less than 10% of all M&M's.

Success/Failure condition: $np = 20$ and $nq = 180$ are both greater than 10.

b) The sampling distribution model is Normal, with: $\mu_{\hat{p}} = p = 0.1$ and

$SD(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.1)(0.9)}{200}} \approx 0.021$

About 68% of the sample proportions are expected to be between 0.079 and 0.121, about 95% are expected to be between 0.058 and 0.142, and about 99.7% are expected to be between 0.037 and 0.163.



c) If the bags contained more candies, the sampling distribution model would still be Normal and centered at 0.1, but the standard deviation would decrease. The sampling distribution model would be less spread out.

21. Just (un)lucky.

For 200 flips, the sampling distribution model is Normal with $\mu_{\hat{p}} = p = 0.5$ and $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.5)(0.5)}{200}} \approx 0.0354$. Her sample proportion of $\hat{p} = 0.42$ is about 2.26 standard deviations below the expected proportion, which is unusual, but not extraordinary. According to the Normal model, we expect sample proportions this low or lower about 1.2% of the time.

Chapter Exercises

22. Too many green ones?

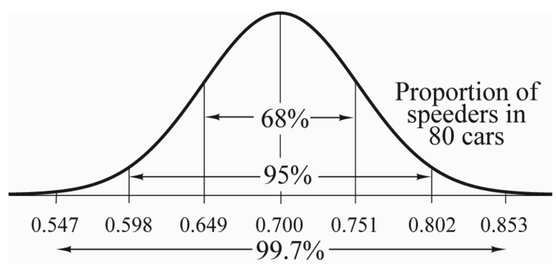
For 500 candies, the sampling distribution model is Normal with $\mu_{\hat{p}} = p = 0.1$ and $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.1)(0.9)}{500}} \approx 0.01342$. The sample proportion of $\hat{p} = 0.12$ is about 1.49 standard deviations above the expected proportion, which is not at all unusual. According to the Normal model, we expect sample proportions this high or higher about 6.8% of the time.

23. Speeding.

a) $\mu_{\hat{p}} = p = 0.70$

$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.7)(0.3)}{80}} \approx 0.051$.

About 68% of the sample proportions are expected to be between 0.649 and 0.751, about 95% are expected to be between 0.598 and 0.802, and about 99.7% are expected to be between 0.547 and 0.853.



b) Randomization condition: The sample may not be representative. If the flow of traffic is very fast, the speed of the other cars around may have some effect on the speed of each driver. Likewise, if traffic is slow, the police may find a smaller proportion of speeders than they expect.

10% condition: 80 cars represent less than 10% of all cars

Success/Failure condition: $np = 56$ and $nq = 24$ are both greater than 10.

The Normal model may not be appropriate. Use caution. (And don't speed!)

24. Smoking 2014.

Randomization condition: 60 people are selected at random

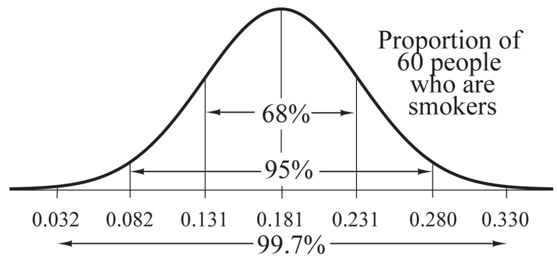
10% condition: 60 is less than 10% of all people.

Success/Failure condition: $np = 10.68$ and $nq = 49.14$ are both greater than 10.

The sampling distribution model is Normal, with: $\mu_{\hat{p}} = p = 0.181$

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.181)(0.819)}{60}} \approx 0.0497$$

There is an approximate chance of 68% that between 13.1% and 23.1% of 60 people are smokers, an approximate chance of 95% that between 8.2% and 28.0% are smokers, and an approximate chance of 99.7% that between 3.2% and 33.0%are smokers.



25. Vision.

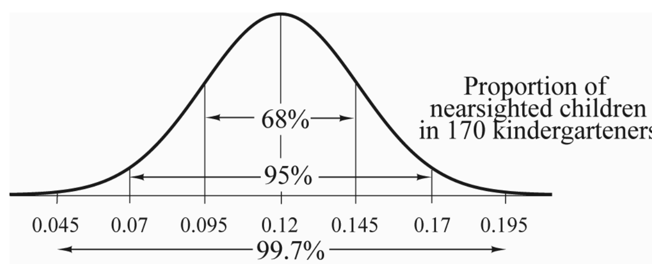
a) **Randomization condition:** Assume that the 170 children are a representative sample of all children.

10% condition: A sample of this size is less than 10% of all children.

Success/Failure condition: np = 20.4 and nq = 149.6 are both greater than 10.

Therefore, the sampling distribution model for the proportion of 170 children who are nearsighted is $N\left(0.12, \sqrt{\frac{(0.12)(0.88)}{170}}\right)$ or $N(0.12, 0.025)$.

b) The Normal model is to the right.



c) They might expect that the proportion of nearsighted students to be within 2 standard deviations of the mean. According to the Normal model, this means they might expect between 7% and 17% of the students to be nearsighted, or between about 12 and 29 students.

26. Mortgages 2013.

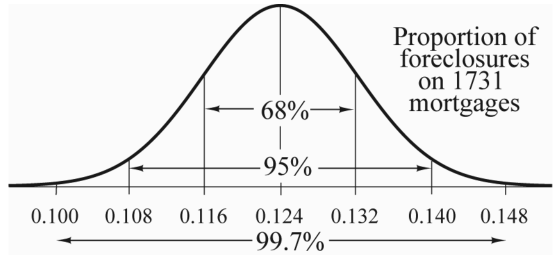
a) **Randomization condition:** Assume that the 1731 mortgages are a representative sample of all mortgages.

10% condition: A sample of this size is less than 10% of all mortgages.

Success/Failure condition: np = 214.6 and nq = 1516.4 are both greater than 10.

Therefore, the sampling distribution model for the proportion of foreclosures on 1731 mortgages is $N\left(0.124, \sqrt{\frac{(0.124)(0.876)}{1731}}\right)$ or $N(0.124, 0.008)$.

b) The Normal model is to the right.



c) They might expect that the proportion of mortgage foreclosures to be within 2 standard deviations of the mean. According to the Normal model, this means they might expect between 10.8% and 14.0% of the mortgages to undergo foreclosure, or between about 187 and 242 foreclosures.

27. Loans.

a) $\mu_{\hat{p}} = p = 7\%$

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.07)(0.93)}{200}} \approx 0.018 \approx 1.8\%$$

b) **Randomization condition:** Assume that the 200 people are a representative sample of all loan recipients.

10% condition: A sample of this size is less than 10% of all loan recipients.

Success/Failure condition: $np = 14$ and $nq = 186$ are both greater than 10.

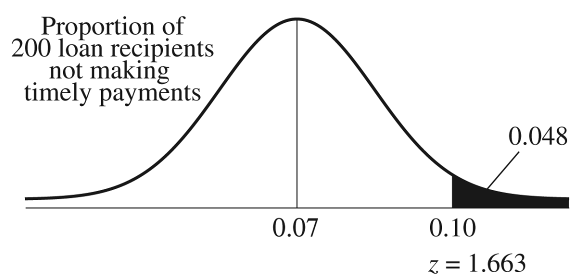
Therefore, the sampling distribution model for the proportion of 200 loan recipients who will not make payments on time is $N(0.07, 0.018)$.

c) According to the Normal model, the probability that over 10% of these clients will not make timely payments is approximately 0.048.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.10 - 0.07}{\sqrt{\frac{(0.07)(0.93)}{200}}}$$

$$z \approx 1.663$$



28. Contacts.

a) **Randomization condition:** 100 students are selected at random.

10% condition: 100 is less than 10% of all of the students at the university, provided the university has more than 1000 students.

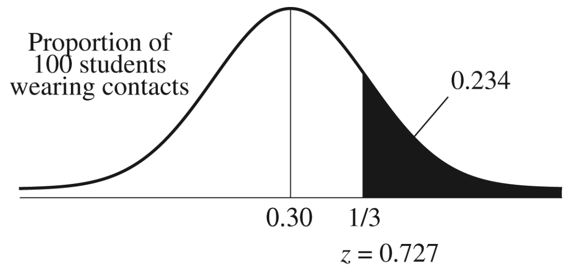
Success/Failure condition: $np = 30$ and $nq = 70$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$\mu_{\hat{p}} = p = 0.30$

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.30)(0.70)}{100}} \approx 0.046$$

b) According to the Normal model, the probability that more than one-third of the students in this sample wear contacts is approximately 0.234.



$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{\frac{1}{3} - 0.30}{\sqrt{\frac{(0.30)(0.70)}{100}}}$$

$$z \approx 0.727$$

29. Back to school? 2013.

Randomization condition: We are considering random samples of 400 students who took the ACT.

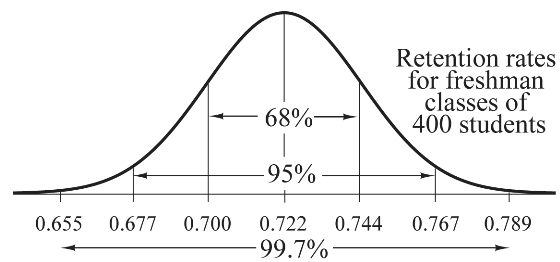
10% Condition: 400 students is less than 10% of all college students.

Success/Failure condition: $np = 288.8$ and $nq = 111.2$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.722$$

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.722)(0.278)}{400}} \approx 0.0224$$



According to the sampling distribution model, about 68% of the colleges are expected to have retention rates between 0.700 and 0.744, about 95% of the colleges are expected to have retention rates between 0.677 and 0.767, and about 99.7% of the colleges are expected to have retention rates between 0.655 and 0.789. However, the conditions for the use of this model may not be met. We should be cautious about making any conclusions based on this model.

30. Binge drinking.

Randomization condition: The students were selected randomly.

10% condition: 200 college students are less than 10% of all college students.

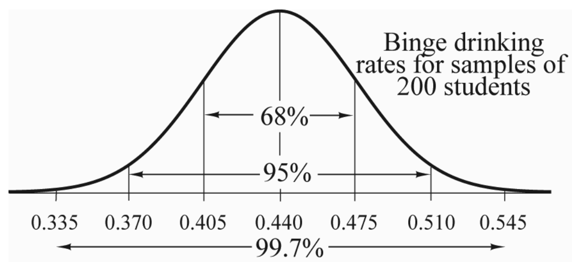
Success/Failure condition: $np = 88$ and $nq = 112$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.44$$

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.44)(0.56)}{200}} \approx 0.035$$

According to the sampling distribution model, about 68% of samples of 200 students are expected to have binge drinking proportions between 0.405 and 0.475, about 95% between 0.370 and 0.510, and about 99.7% between 0.335 and 0.545.



31. Back to school, again.

Provided that the students at this college are typical, the sampling distribution model for the retention rate, $\hat{p}$, is Normal with $\mu_{\hat{p}} = p = 0.722$ and standard deviation $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.722)(0.278)}{603}} \approx 0.0182$

This college has a right to brag about their retention rate. 522/603 = 86.6% is over 7 standard deviations above the expected rate of 72.2%.

32. Binge sample.

Since the sample is random and the Success/Failure condition is met, the sampling distribution model for the binge drinking rate, $\hat{p}$, is Normal with $\mu_{\hat{p}} = p = 0.44$ and standard deviation $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.44)(0.56)}{244}} \approx 0.032$

The binge drinking rate at this college is lower than the national result, but not extremely low. 96/244 = 39.3% is only about 1.5 standard deviations below the national rate of 44%.

33. Polling.

Randomization condition: We must assume that the 400 voters were polled randomly.

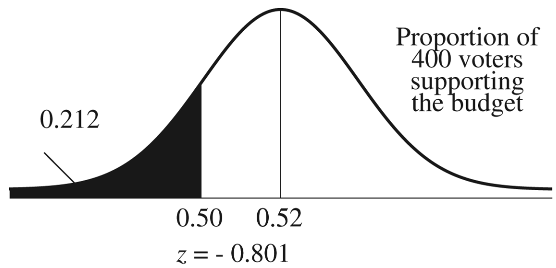
10% condition: 400 voters polled represent less than 10% of potential voters.

Success/Failure condition: np = 208 and nq = 192 are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.52 \qquad SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.52)(0.48)}{400}} \approx 0.025$$

According to the Normal model, the probability that the newspaper's sample will lead them to predict defeat (that is, predict budget support below 50%) is approximately 0.212.



$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.50 - 0.52}{\sqrt{\frac{(0.52)(0.48)}{400}}}$$

$$z \approx -0.801$$

34. Seeds.

Randomization condition: We must assume that the 160 seeds in a pack are a random sample. Since seeds in each pack may *not* be a random sample, proceed with caution.

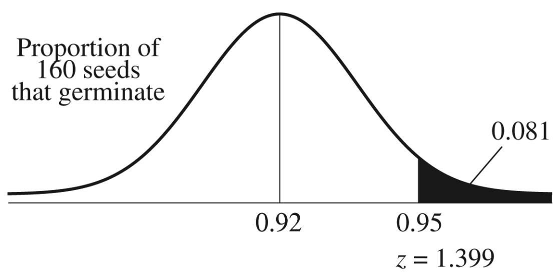
10% condition: The 160 seeds represent less than 10% of all seeds.

Success/Failure condition: np = 147.2 and nq = 12.8 are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.92 \qquad SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.92)(0.08)}{160}} \approx 0.0215$$

According to the Normal model, the probability that more than 95% of the seeds will germinate is approximately 0.081.



$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.95 - 0.92}{\sqrt{\frac{(0.92)(0.08)}{160}}}$$

$$z \approx 1.399$$

35. Gaydar.

Randomization condition: We must assume that the 100 photographs were randomized before they were shown to the women.

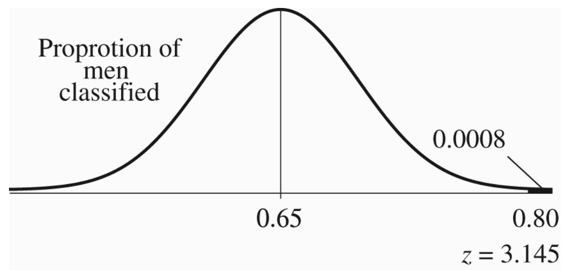
10% condition: This study doesn't involve a sample, just randomization of a set of photos. The 10% condition does not apply.

Success/Failure condition: np = 65 and nq = 35 are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.65 \qquad SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.65)(0.35)}{100}} \approx 0.048$$

According to the Normal model, the probability of correctly classifying 80 of 100 men is approximately 0.0008.



$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.80 - 0.65}{\sqrt{\frac{(0.65)(0.35)}{100}}}$$

$$z \approx 3.145$$

36. Genetic Defect.

Randomization condition: We will assume that the 732 newborns are representative of all newborns.

10% condition: These 732 newborns represent less than 10% of all newborns.

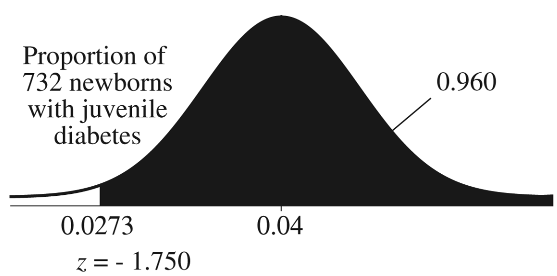
Success/Failure condition: np = 29.28 and nq = 702.72 are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.04 \qquad SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.04)(0.96)}{732}} \approx 0.0072$$

In order to get the 20 newborns for the study, the researchers hope to find at least $\hat{p} = \frac{20}{732} \approx 0.0273$ as the proportion of newborns in the sample with juvenile diabetes.

According to the Normal model, the probability that the researchers find at least 20 newborns with juvenile diabetes is approximately 0.960.



$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{{}^{20}\!/_{732} - 0.04}{\sqrt{\frac{(0.04)(0.96)}{732}}}$$

$$z \approx -1.750$$

37. "No Children" section.

Randomization condition: We will assume that the 120 customers (to fill the restaurant to capacity) are representative of all customers.

10% condition: 120 customers represent less than 10% of all potential customers.

Success/Failure condition: np = 36 and nq = 84 are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$\mu_{\hat{p}} = p = 0.30$ $\qquad$ $SD(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.30)(0.70)}{120}} \approx 0.042$

Answers may vary. We will use 3 standard deviations above the expected proportion of customers with children to be "very sure".

$$\mu_{\hat{p}} + 3\left(\sqrt{\frac{pq}{n}}\right) \approx 0.30 + 3(0.0418) \approx 0.4254$$

Since 120(0.4254) = 51.048, the restaurant needs about 51 seats in the family-friendly section.

38. Meals.

Randomization condition: We will assume that the 180 customers are representative of all customers.

10% condition: 180 customers represent less than 10% of all potential customers.

Success/Failure condition: $np = 36$ and $nq = 144$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$\mu_{\hat{p}} = p = 0.20$

$SD(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.20)(0.80)}{180}} \approx 0.0298$

Answers may vary. We will use 2 standard deviations above the expected proportion of customers who order the steak special to be "pretty sure".

$$\mu_{\hat{p}} + 2\left(\sqrt{\frac{pq}{n}}\right) \approx 0.20 + 2(0.0298) \approx 0.2596$$

Since 180(0.2596) = 46.728, the chef needs at least 47 steaks on hand.

39. Sampling.

a) The sampling distribution model for the sample mean is $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

b) If we choose a larger sample, the mean of the sampling distribution model will remain the same, but the standard deviation will be smaller.

40. Sampling, part II.

a) The sampling distribution model for the sample mean will be skewed to the left as well, centered at μ, with standard deviation $\dfrac{\sigma}{\sqrt{n}}$.

b) When the sample size is increased, the sampling distribution model becomes more Normal in shape, centered at μ, with standard deviation $\frac{\sigma}{\sqrt{n}}$.

c) As we make the sample larger the distribution of data in the sample should more closely resemble the shape, center, and spread of the population.

41. Waist size.

a) The distribution of waist size of 250 men is unimodal and slightly skewed to the right. A typical waist size is approximately 36 inches, and the standard deviation in waist sizes is approximately 4 inches.

b) All of the histograms show distributions of sample means centered near 36 inches. As n gets larger the histograms approach the Normal model in shape, and the variability in the sample means decreases. The histograms are fairly Normal by the time the sample reaches size 5.

42. CEO compensation.

a) The distribution of total compensation for the CEOs for the 800 largest U.S. companies is unimodal, but skewed to the right with several large outliers.

b) All of the histograms are centered near $10,000,000. As n gets larger, the variability in sample means decreases, and histograms approach the Normal shape. However, they are still visibly skewed to the right, with the possible exception of the histogram for n = 200.

c) This rule of thumb doesn't seem to be true for highly skewed distributions.

43. Waist size revisited.

a)

n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical standard deviation
2	36.314	36.33	2.855	$4.019/\sqrt{2} \approx 2.842$
5	36.314	36.33	1.805	$4.019/\sqrt{5} \approx 1.797$
10	36.341	36.33	1.276	$4.019/\sqrt{10} \approx 1.271$
20	36.339	36.33	0.895	$4.019/\sqrt{20} \approx 0.899$

b) The observed values are all very close to the theoretical values.

c) For samples as small as 5, the sampling distribution of sample means is unimodal and symmetric. The Normal model would be appropriate.

d) The distribution of the original data is nearly unimodal and symmetric, so it doesn't take a very large sample size for the distribution of sample means to be approximately Normal.

44. CEOs revisited.

a)

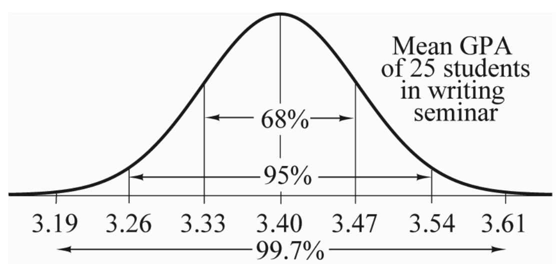
n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical standard deviation
30	10,251.73	10,307.31	3359.64	$17{,}964.62/\sqrt{30} \approx 3279.88$
50	10,343.93	10,307.31	2483.84	$17{,}964.62/\sqrt{50} \approx 2540.58$
100	10,329.94	10,307.31	1779.18	$17{,}964.62/\sqrt{100} \approx 1796.46$
200	10,340.37	10,307.31	1260.79	$17{,}964.62/\sqrt{200} \approx 1270.29$

b) The observed values are all very close to the theoretical values.

c) All the sampling distributions are still quite skewed, with the possible exception of the sampling distribution for n = 200, which is still somewhat skewed. The Normal model would not be appropriate.

d) The distribution of the original data is strongly skewed, so it will take a very large sample size before the distribution of sample means is approximately Normal.

45. GPAs.



Randomization condition: Assume that the students are randomly assigned to seminars.

Independence assumption: It is reasonable to think that GPAs for randomly selected students are mutually independent.

10% condition: The 25 students in the seminar certainly represent less than 10% of the population of students.

Large Enough Sample condition: The distribution of GPAs is roughly unimodal and symmetric, so the sample of 25 students is large enough.

The mean GPA for the freshmen was $\mu = 3.4$, with standard deviation $\sigma = 0.35$. Since the conditions are met, the Central Limit Theorem tells us that we can model the sampling distribution of the mean GPA with a Normal model, with $\mu_{\bar{y}} = 3.4$ and standard deviation $SD(\bar{y}) = \dfrac{0.35}{\sqrt{25}} \approx 0.07$.

The sampling distribution model for the sample mean GPA is approximately $N(3.4, 0.07)$.

46. Home values.

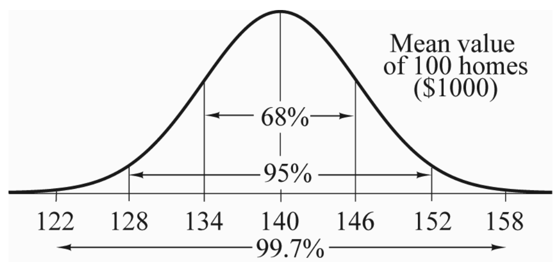
Randomization condition: Homes were selected at random.

Independence assumption: It is reasonable to think that assessments for randomly selected homes are mutually independent.

10% condition: The 100 homes in the sample certainly represent less than 10% of the population of all homes in the city. A small city will likely have more than 1000 homes.

Large Enough Sample condition: A sample of 100 homes is large enough.

The mean home value was $\mu = \$140,000$, with standard deviation $\sigma = \$60,000$. Since the conditions are met, the Central Limit Theorem tells us that we can model the sampling distribution of the mean home value with a Normal model, with $\mu_{\bar{y}} = \$140,000$ and standard deviation $SD(\bar{y}) = \dfrac{60,000}{\sqrt{100}} = \$6000.$



The sampling distribution model for the sample mean home values is approximately $N(140000, 6000)$.

47. Lucky spot?

a) Smaller outlets have more variability than the larger outlets, just as the Central Limit Theorem predicts.

b) If the lottery is truly random, all outlets are equally likely to sell winning tickets.

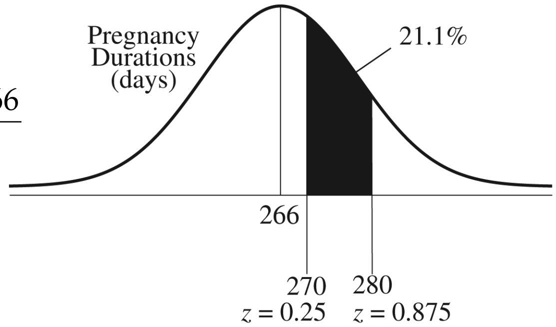
48. Safe cities.

The standard deviation of the sampling model for the mean is $\dfrac{\sigma}{\sqrt{n}}$. So, cities in which the average is based on a smaller number of drivers will have greater variation in their averages and will be more likely to be both safest and least safe.

49. Pregnancy.

a)

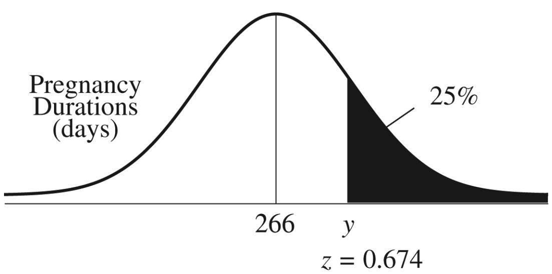
$$z = \frac{y-\mu}{\sigma}$$
$$z = \frac{270-266}{16}$$
$$z = 0.25$$



$$z = \frac{y-\mu}{\sigma}$$
$$z = \frac{280-266}{16}$$
$$z = 0.875$$

According to the Normal model, approximately 21.1% of all pregnancies are expected to last between 270 and 280 days.

b)



$$z = \frac{y-\mu}{\sigma}$$
$$0.674 = \frac{y-266}{16}$$
$$y \approx 276.8 \text{ days}$$

According to the Normal model, the longest 25% of pregnancies are expected to last approximately 276.8 days or more.

c) **Randomization condition:** Assume that the 60 women the doctor is treating can be considered a representative sample of all pregnant women.

Independence assumption: It is reasonable to think that the durations of the patients' pregnancies are mutually independent.

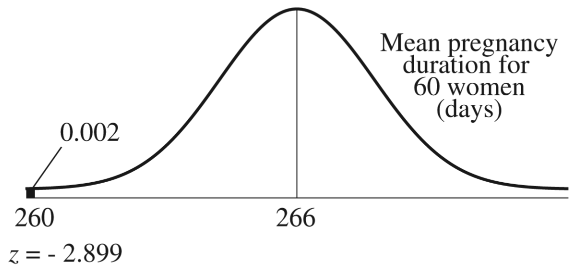
10% condition: The 60 women that the doctor is treating certainly represent less than 10% of the population of all women.

Large Enough Sample condition: The sample of 60 women is large enough. In this case, any sample would be large enough, since the distribution of pregnancies is Normal.

The mean duration of the pregnancies was $\mu = 266$ days, with standard deviation $\sigma = 16$ days. Since the distribution of pregnancy durations is Normal, we can model the sampling distribution of the mean pregnancy duration with a Normal model, with $\mu_{\bar{y}} = 266$ days and standard deviation

$SD(\bar{y}) = \frac{16}{\sqrt{60}} \approx 2.07$ days.

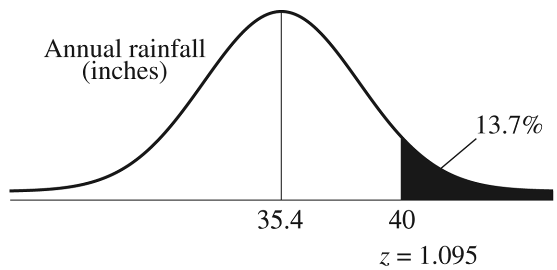
d) According to the Normal model, with mean 266 days and standard deviation 2.07 days, the probability that the mean pregnancy duration is less than 260 days is 0.002.



$$z = \frac{y-\mu}{\sigma}$$
$$z = \frac{260-266}{\frac{16}{\sqrt{20}}}$$
$$z \approx -2.899$$

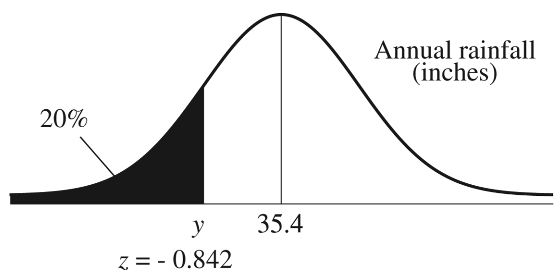
50. Rainfall.

a) According to the Normal model, Ithaca is expected to get more than 40 inches of rain in approximately 13.7% of years.



b) According to the Normal model, Ithaca is expected to get less than 31.9 inches of rain in driest 20% of years.

$$z = \frac{y-\mu}{\sigma}$$
$$-0.842 = \frac{y-35.4}{4.2}$$
$$y \approx 31.9$$



c) **Randomization condition:** Assume that the 4 years in which the student was in Ithaca can be considered a representative sample of all years.

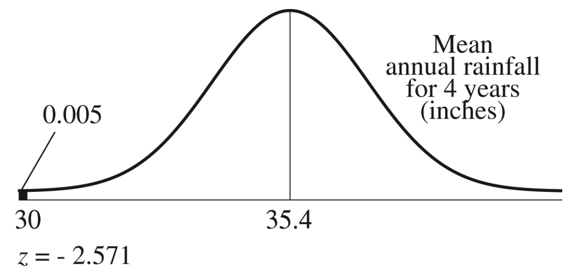
Independence assumption: It is reasonable to think that the rainfall totals for the years are mutually independent.

10% condition: The 4 years in which the student was in Ithaca certainly represent less than 10% of all years.

Large enough sample condition: The sample of 4 years is large enough. In this case, any sample would be large enough, since the distribution of annual rainfall is Normal.

The mean rainfall was $\mu = 35.4$ inches, with standard deviation $\sigma = 4.2$ inches. Since the distribution of yearly rainfall is Normal, we can model the sampling distribution of the mean annual rainfall with a Normal model, with $\mu_{\bar{y}} = 35.4$ inches and standard deviation $SD(\bar{y}) = \frac{4.2}{\sqrt{4}} = 2.1$ inches.

d) According to the Normal model, with mean 35.4 inches and standard deviation 2.4 inches, the probability that those four years averaged less than 30 inches of rain is 0.005.



51. Pregnant again.

a) The distribution of pregnancy durations may be skewed to the left since there are more premature births than very long pregnancies. Modern practice of medicine stops pregnancies at about 2 weeks past normal due date by inducing labor or performing a Caesarean section.

b) We can no longer answer the questions posed in parts a and b. The Normal model is not appropriate for skewed distributions. The answer to part c is still valid. The Central Limit Theorem guarantees that the sampling distribution model is Normal when the sample size is large.

52. At work.

a) The distribution of length of time people work at a job is likely to be skewed to the right, because some people stay at the same job for much longer than the mean plus two or three standard deviations. Additionally, the left tail cannot be long, because a person cannot work at a job for less than 0 years.

b) The Central Limit Theorem guarantees that the distribution of the mean time is Normally distributed for large sample sizes, as long as the assumptions and conditions are satisfied. The CLT doesn't help us with the distribution of individual times.

53. Dice and dollars.

a) Let X = the number of dollars won in one play.

$$\mu = E(X) = 0\left(\frac{3}{6}\right) + 1\left(\frac{2}{6}\right) + 10\left(\frac{1}{6}\right) = \$2$$

$$\sigma^2 = Var(X) = (0-2)^2\left(\frac{3}{6}\right) + (1-2)^2\left(\frac{2}{6}\right) + (10-2)^2\left(\frac{1}{6}\right) = 13$$

$$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{13} \approx \$3.61$$

b) $X + X$ = the total winnings for two plays.

$$\mu = E(X+X) = E(X) + E(X) = 2 + 2 = \$4$$

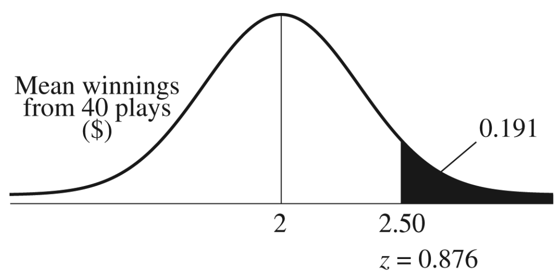
$$\sigma = SD(X+X) = \sqrt{Var(X) + Var(X)}$$
$$= \sqrt{13+13} \approx \$5.10$$

c) In order to win at least \$100 in 40 plays, you must average at least $\frac{100}{40} = \$2.50$ per play.

The expected value of the winnings is $\mu = \$2$, with standard deviation $\sigma = \$3.61$. Rolling a die is random and the outcomes are mutually independent, so the Central Limit Theorem guarantees that the sampling distribution model is Normal with $\mu_{\bar{x}} = \$2$ and standard deviation $SD(\bar{x}) = \frac{\$3.61}{\sqrt{40}} \approx \$0.571$.

According to the Normal model, the probability that you win at least \$100 in 40 plays is approximately 0.191.

(This is equivalent to using $N(80, 22.83)$ to model your total winnings.)



54. New game.

a) Let X = the amount of money won.

X	\$40	\$0	– \$10
$P(X)$	$\frac{1}{6}$	$\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \frac{5}{36}$	$\left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \frac{25}{36}$

b) $\mu = E(X) = 40\left(\frac{1}{6}\right) + 0\left(\frac{5}{36}\right) - 10\left(\frac{25}{36}\right) \approx -\$0.28$

$$\sigma^2 = Var(X) = (40-(-0.28))^2\left(\frac{1}{6}\right) + (0-(-0.28))^2\left(\frac{5}{36}\right) + (-10-(-0.28))^2\left(\frac{25}{36}\right) \approx 336.034$$

$$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{336.034} \approx \$18.33$$

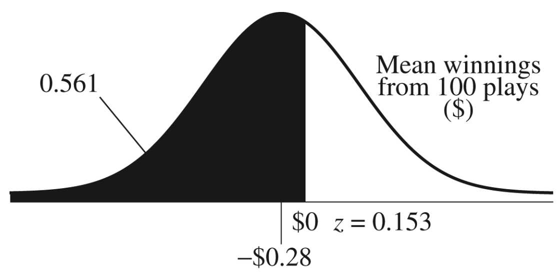
c) $\mu = E(X+X+X+X+X) = 5E(X) = 5(-0.28) = -\$1.40$

$$\sigma = SD(X+X+X+X+X) = \sqrt{5(Var(X))} = \sqrt{5(336.034)} \approx \$40.99$$

d) In order for the person running the game to make a profit, the average winnings of the 100 people must be less than \$0.

The expected value of the winnings is $\mu = -\$0.28$, with standard deviation $\sigma = \$18.33$. Rolling a die is random and the outcomes are mutually independent, so the Central Limit Theorem guarantees that the sampling distribution model is Normal with $\mu_{\bar{x}} = -\$0.28$ and standard deviation $SD(\bar{x}) = \frac{18.33}{\sqrt{100}} \approx \$1.833$.

According to the Normal model, the probability that the person running the game makes a profit is approximately 0.561.



55. AP Stats 2013.

a) $\mu = 5(0.126) + 4(0.202) + 3(0.250) + 2(0.188) + 1(0.234) = 2.798$

$$\sigma = \sqrt{(5-2.798)^2(0.126) + (4-2.798)^2(0.202) + (3-2.798)^2(0.250) + (2-2.798)^2(0.188) + (1-2.798)^2(0.234)} \approx 1.338$$

b) The distribution of scores for 40 randomly selected students would not follow a Normal model. The distribution would resemble the population, mostly uniform for scores 1 – 4, with about half as many 5s.

c) **Randomization condition:** The scores are selected randomly.

Independence assumption: It is reasonable to think that the randomly selected scores are independent of one another.

10% condition: The 40 scores represent less than 10% of all scores.

Large Enough Sample condition: A sample of 40 scores is large enough.

Since the conditions are satisfied, the sampling distribution model for the mean of 40 randomly selected AP Stat scores is Normal, with $\mu_{\bar{y}} = \mu \approx 2.798$ and standard deviation $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{1.338}{\sqrt{40}} \approx 0.2116$.

56. Museum membership.

a) $\mu = 50(0.41) + 100(0.37) + 250(0.14) + 500(0.07) + 1000(0.01) \approx \$137.50$

$$\sigma = \sqrt{\begin{array}{r}(50-137.50)^2(0.41) + (100-137.50)^2(0.37) + (250-137.50)^2(0.14) \\ + (500-137.50)^2(0.07) + (1000-137.50)^2(0.01)\end{array}} \approx \$148.56$$

The calculation for standard deviation is based on a rounded mean. Use technology to calculate the mean and standard deviation to avoid inaccuracy.

b) The distribution of donations for 50 new members would not follow a Normal model. The new members would likely make donations typical of the current member populations. The distribution would resemble the population, skewed to the right.

c) **Randomization condition:** The members are not selected randomly. They are simply the new members that day. However, the donations they make are probably typical of the donations made by current members.

Independence assumption: It is reasonable to think that the donations for the new members would not affect one another.

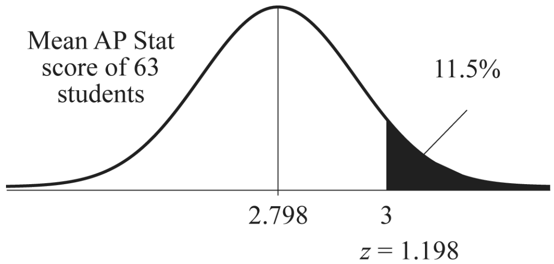
10% condition: The 50 donations represent less than 10% of all donations.

Large Enough Sample condition: The sample of 50 donations is large enough.

Since the conditions are satisfied, the sampling distribution model for the mean of 50 donations is Normal, with $\mu_{\bar{y}} = \mu \approx \$137.50$ and standard deviation $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{148.56}{\sqrt{50}} \approx 21.010$.

57. AP Stats 2013, again.

Since the teacher considers his 63 students "typical", and 63 is less than 10% of all students, the sampling distribution model for the mean AP Stat score for 63 students is Normal, with mean $\mu_{\bar{y}} = \mu \approx 2.798$ and standard deviation $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{1.338}{\sqrt{63}} \approx 0.1686$.

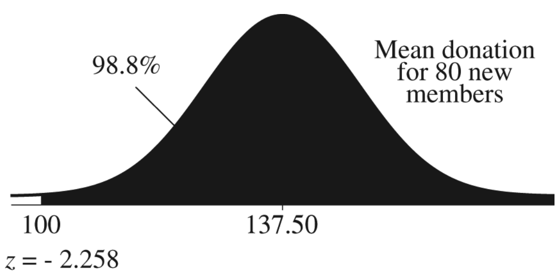


$$z = \frac{\bar{y} - \mu_{\bar{y}}}{SD(\bar{y})}$$
$$z = \frac{3 - 2.798}{\frac{1.338}{\sqrt{63}}}$$
$$z = 1.198$$

According to the sampling distribution model, the probability that the class of 63 students achieves an average of 3 on the AP Stat exam is about 11.5%.

58. Joining the museum.

If the new members can be considered a random sample of all museum members, and the 80 new members are less than 10% of all members, then the sampling distribution model for the mean donation of 80 members is Normal, with $\mu_{\bar{y}} = \mu \approx \$137.50$ and standard deviation $SD(\bar{y}) = \frac{148.56}{\sqrt{80}} = \$16.61$.



$$z = \frac{\bar{y} - \mu_{\bar{y}}}{SD(\bar{y})}$$
$$z = \frac{100 - 137.50}{\frac{148.56}{\sqrt{80}}}$$
$$z = -2.258$$

According to the sampling distribution model, there is a 98.8% probability that the average donation for 80 new members is at least $100.

59. Fuel economy again.

a) **Randomization condition:** Assume that the 150 cars can be considered a representative sample of all cars of this type.

Independence assumption: It is reasonable to think that the mileages for these cars are mutually independent.

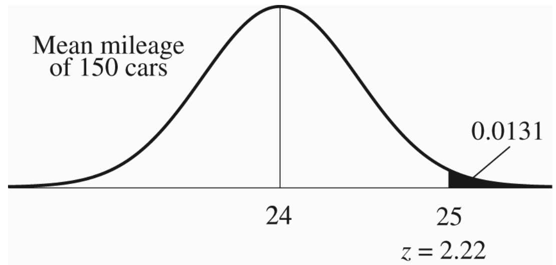
10% condition: The 150 cars in the lot certainly represent less than 10% of all midsize cars.

Large Enough Sample condition: A sample of 150 cars is large enough.

The mean mileage was $\mu = 24$ mpg, with standard deviation $\sigma = 5.5$ mpg. Since the conditions are met, the CLT allows us to model the sampling distribution of the $\overline{y}$ with a Normal model, with $\mu_{\overline{y}} = 24$ mpg and standard deviation

$SD(\overline{y}) = \dfrac{5.5}{\sqrt{150}} \approx 0.45$ mpg.

b) According to the Normal model, the probability that $\overline{y}$ is between 25 and 27 mpg is approximately 0.0131. (27 mpg is about 6.67 standard deviations above the mean, well off the scale of the picture.)



$$z = \frac{\overline{y} - \mu_{\overline{y}}}{SD(\overline{y})}$$

$$z = \frac{\overline{y} - 24}{\frac{5.5}{\sqrt{150}}}$$

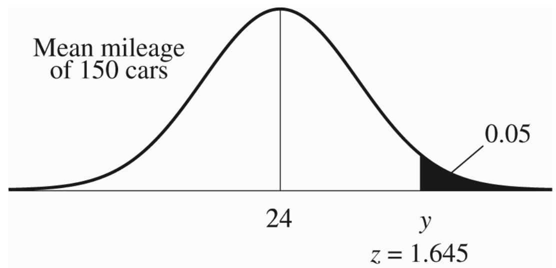
$$z = 2.22$$

c) According to the Normal model, there is only a 5% chance that the mean gas mileage is greater than approximately 24.7 mpg.

$$z = \frac{\overline{y} - \mu_{\overline{y}}}{SD(\overline{y})}$$

$$1.645 = \frac{\overline{y} - 24}{\frac{5.5}{\sqrt{150}}}$$

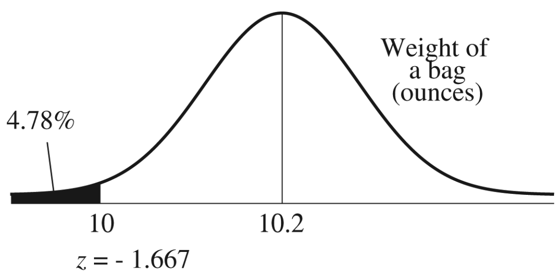
$$\overline{y} \approx 24.7$$



60. Potato chips.

a) According to the Normal model, only about 4.78% of the bags sold are underweight.

b) P(none of the 3 bags are underweight) $= (1 - 0.0478)^3 \approx 0.863$.



c) **Randomization condition:** Assume that the 3 bags can be considered a representative sample of all bags.

Independence assumption: It is reasonable to think that the weights of these bags are mutually independent.

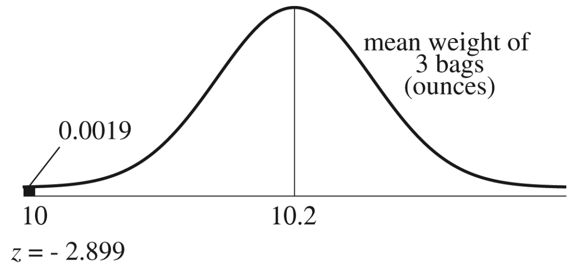
10% condition: The 3 bags certainly represent less than 10% of all bags.

Large Enough Sample condition: Since the distribution of bag weights is believed to be Normal, the sample of 3 bags is large enough.

The mean weight is $\mu = 10.2$ ounces, with standard deviation $\sigma = 0.12$ ounces. Since the conditions are met, we can model the sampling distribution of $\overline{y}$ with a Normal model, with $\mu_{\overline{y}} = 10.2$ ounces and standard deviation

$SD(\overline{y}) = \dfrac{0.12}{\sqrt{3}} \approx 0.069$ ounces.

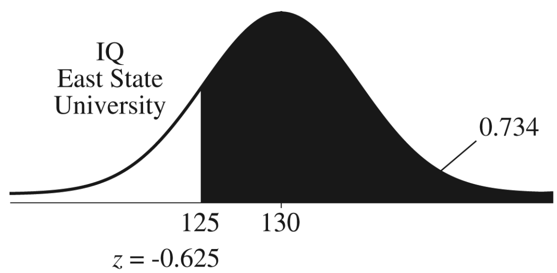
According to the Normal model, the probability that the mean weight of the 3 bags is less than 10 ounces is approximately 0.0019.



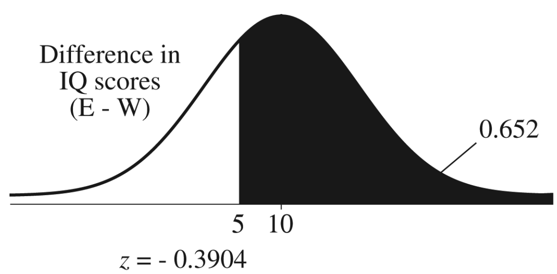
d) For 24 bags, the standard deviation of the sampling distribution model is $SD(\bar{y}) = \frac{0.12}{\sqrt{24}} \approx 0.024$ ounces. Now, an average of 10 ounces is over 8 standard deviations below the mean of the sampling distribution model. This is extremely unlikely.

61. IQs.

a) According to the Normal model, the probability that the IQ of a student from East State is at least 125 is approximately 0.734.



b) First, we will need to generate a model for the difference in IQ between the two schools. Since we are choosing at random, it is reasonable to believe that the students' IQs are independent, which allows us to calculate the standard deviation of the difference.



$$\mu = E(E-W) = E(E) - E(W) = 130 - 120 = 10$$
$$\sigma = SD(E-W) = \sqrt{Var(E) + Var(W)}$$
$$= \sqrt{8^2 + 10^2} \approx 12.806$$

Since both distributions are Normal, the distribution of the difference is $N(10, 12.806)$.

According to the Normal model, the probability that the IQ of a student at ESU is at least 5 points higher than a student at WSU is approximately 0.652.

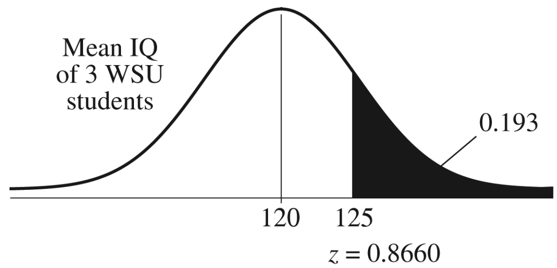
c) Randomization condition: Students are randomly sampled from WSU.

Independence assumption: It is reasonable to think that the IQs are mutually independent.

10% condition: The 3 students certainly represent less than 10% of students.

Large Enough Sample condition: The distribution of IQs is Normal, so the distribution of sample means of samples of any size will be Normal, so a sample of 3 students is large enough.

The mean IQ is $\mu_w = 120$, with standard deviation $\sigma_w = 10$. Since the distribution IQs is Normal, we can model the sampling distribution of $\overline{w}$ with a Normal model, with $\mu_{\overline{w}} = 120$ with standard deviation $SD(\overline{w}) = \frac{10}{\sqrt{3}} \approx 5.7735.$



According to the Normal model, the probability that the mean IQ of the 3 WSU students is above 125 is approximately 0.193.

d) As in part c, the sampling distribution of $\overline{e}$, the mean IQ of 3 ESU students, can be modeled with a Normal model, with $\mu_{\overline{e}} = 130$ with standard deviation

$SD(\overline{e}) = \frac{8}{\sqrt{3}} \approx 4.6188.$

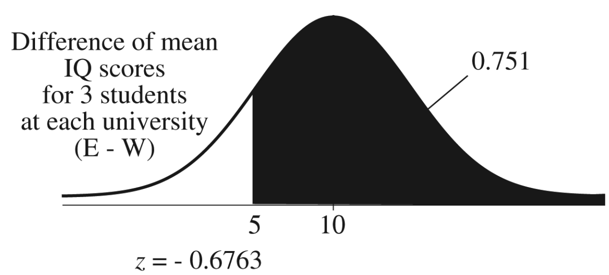
The distribution of the difference in mean IQ is Normal, with the following parameters:

$$\mu_{\overline{e}-\overline{w}} = E(\overline{e} - \overline{w}) = E(\overline{e}) - E(\overline{w}) = 130 - 120 = 10$$

$$\sigma_{\overline{e}-\overline{w}} = SD(\overline{e} - \overline{w}) = \sqrt{Var(\overline{e}) + Var(\overline{w})}$$

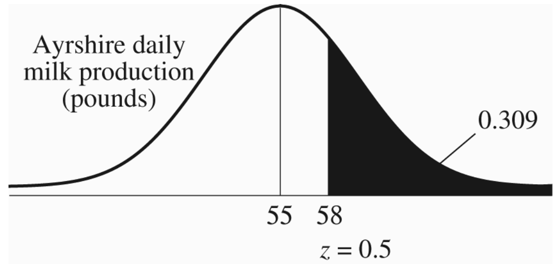
$$= \sqrt{\left(\frac{10}{\sqrt{3}}\right)^2 + \left(\frac{8}{\sqrt{3}}\right)^2} \approx 7.3937$$

According to the Normal model, the probability that the mean IQ of 3 ESU students is at least 5 points higher than the mean IQ of 3 WSU students is approximately 0.751.



62. Milk.

a) According to the Normal model, the probability that an Ayrshire averages more than 58 pounds of milk per day is approximately 0.309.



b) First, we will need to generate a model for the difference in milk production between the two cows. Since we are choosing at random, it is reasonable to believe that the cows' milk productions are independent, which allows us to calculate the standard deviation of the difference.

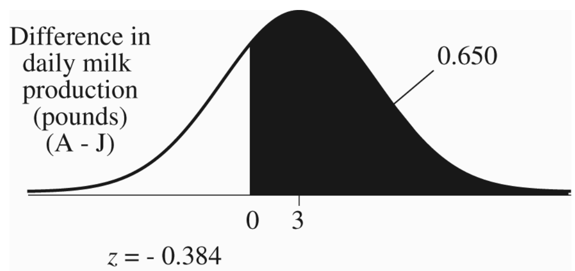
$$\mu = E(A - J) = E(A) - E(J) = 55 - 52 = 3 \text{ pounds}$$

$$\sigma = SD(A - J) = \sqrt{Var(A) + Var(J)}$$

$$= \sqrt{6^2 + 5^2} \approx 7.810 \text{ pounds}$$

Since both distributions are Normal, the distribution of the difference is $N(3, 7.810)$.

According to the Normal model, the probability that the Ayrshire gives more milk than the Jersey is approximately 0.650.



c) **Randomization condition:** Assume that the farmer's 20 Jerseys are a representative sample of all Jerseys.

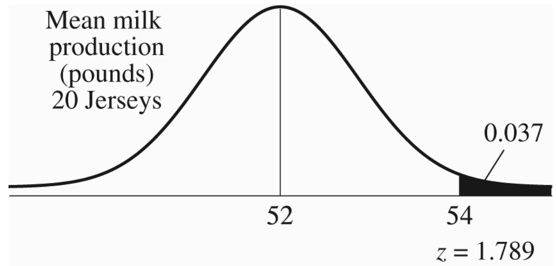
Independence assumption: It is reasonable to think that the cows have mutually independent milk production.

10% condition: The 20 cows certainly represent less than 10% of cows.

Large Enough Sample condition: Since the distribution of daily milk production is Normal, the sample means of samples of any size are Normally distributed, so the sample of 20 cows is certainly large enough.

The mean milk production is $\mu_j = 52$ pounds, with standard deviation $\sigma_j = 5$. Since the distribution of milk production is Normal, we can model the sampling distribution of $\bar{j}$ with a Normal model, with $\mu_{\bar{j}} = 52$ pounds with standard deviation $SD(\bar{j}) = \frac{5}{\sqrt{20}} \approx 1.1180$ pounds.

According to the Normal model, the probability that the mean milk production of the 20 Jerseys is above 45 pounds of milk per day is approximately 0.037.



d) As in part c, the sampling distribution of $\bar{a}$, the mean milk production of 10 Ayrshires, can be modeled with a Normal model, with $\mu_{\bar{a}} = 55$ pounds with standard deviation $SD(\bar{a}) = \frac{6}{\sqrt{10}} \approx 1.8974$ pounds.

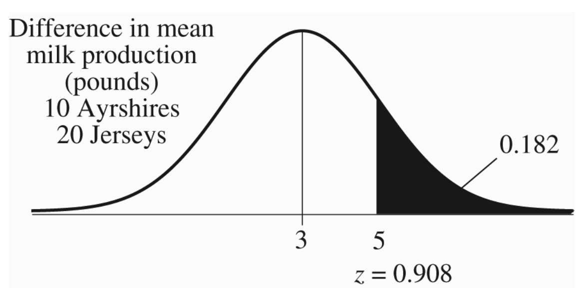
The distribution of the difference in mean milk production is Normal, with the following parameters:

$$\mu_{\bar{a}-\bar{j}} = E(\bar{a}-\bar{j}) = E(\bar{a}) - E(\bar{j}) = 55 - 52 = 3 \text{ pounds}$$

$$\sigma_{\bar{a}-\bar{j}} = SD(\bar{a}-\bar{j}) = \sqrt{Var(\bar{a}) + Var(\bar{j})}$$

$$= \sqrt{\left(\frac{6}{\sqrt{10}}\right)^2 + \left(\frac{5}{\sqrt{20}}\right)^2} \approx 2.2023 \text{ pounds}$$

According to the Normal model, the probability that the mean milk production of 10 Ayrshires is at least 5 pounds higher than the mean milk production of 20 Jerseys is approximately 0.182.



Chapter 18 – Confidence Intervals for Proportions

Section 18.1

1. Lying about age.

a) This means that 49% of the 799 teens in the sample said they have misrepresented their age online. This is our best estimate of p, the proportion of all U.S. teens who would say they have done so.

b) $SE(\hat{p}) = \sqrt{\dfrac{(0.49)(0.51)}{799}} \approx 0.018$

c) Because we don't know p, we use $\hat{p}$ to estimate the standard deviation of the sampling distribution. So the standard error is our estimate of the amount of variation in the sample proportion we expect to see from sample to sample when we ask 799 teens whether they've misrepresented their age online.

2. How's life?

a) This means that 4% of the approximately 1500 people in the sample considered themselves to be suffering. This is our best estimate of p, the proportion of all U.S. people who would consider themselves as suffering.

b) $SE(\hat{p}) = \sqrt{\dfrac{(0.04)(0.96)}{1500}} \approx 0.0051$

c) Because we don't know p, we use $\hat{p}$ to estimate the standard deviation of the sampling distribution. The standard error is our estimate of the amount of variation we expect to see from sample to sample in the sample proportion of the 1500 people rated as suffering.

Section 18.2

3. Lying about age again.

a) We are 95% confident that, if we were to ask all U.S. teens whether they have misrepresented their age online, between 45.6% and 52.5% of them would say they have.

b) If we were to collect many random samples of 799 teens, about 95% of the confidence intervals we construct would contain the proportion of all U.S. teens who admit to misrepresenting their age online.

4. Single moms.

a) We are 95% confident that, if we were to ask all Americans what they think about the trend of more single women having kids, between 2.9% and 5.1% would say it is a "good thing."

b) If we were to collect many random samples of 1229 Americans, about 95% of the confidence intervals we construct would contain the proportion of all Americans who would say the trend of more single women having kids is a "good thing."

Section 18.3

5. Wrong direction.

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645 \times \sqrt{\frac{(0.66)(0.34)}{402}} \approx 0.039$ or 3.9%.

b) The margin of error for 95% confidence would be larger. The critical value of 1.96 is greater than the 1.645 needed for 90% confidence.

6. I asked, "How's life?"

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 2.576 \times \sqrt{\frac{(0.49)(0.51)}{1500}} \approx 0.033$ or 3.3%.

b) The margin of error for 95% confidence would be smaller. The critical value of 1.96 is less than the 2.576 needed for 99% confidence.

Section 18.4

7. Wrong direction again.

a) The sample is a simple random sample. Both $n\hat{p} = (402)(0.66) = 265.32$ and $n\hat{q} = (402)(0.34) = 136.68$ are at least ten. The sample size of 402 is less than 10% of the adult population of Wisconsin. Conditions for the confidence interval are met.

b)

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.02 = 1.645\sqrt{\frac{(0.66)(0.34)}{n}}$$

$$n = \frac{(1.645)^2(0.66)(0.34)}{(0.02)^2}$$

$$n \approx 1519 \text{ people}$$

In order to estimate the proportion to within 2% with 90% confidence, we would need a sample of at least 1519 people. All decimals in the final answer must be rounded up, to the next person.
(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 1692 people.)

8. More single moms.

a) The sample is a simple random sample. Both $n\hat{p} = (1229)(0.04) = 49.16$ and $n\hat{p} = (1229)(0.96) = 1179.84$ are at least ten. The sample size of 1229 is less than 10% of the adult population of the United States. Conditions for the confidence interval are met.

b)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.03 = 1.96\sqrt{\frac{(0.04)(0.96)}{n}}$$

$$n = \frac{(1.96)^2(0.04)(0.96)}{(0.03)^2}$$

$$n \approx 164 \text{ people}$$

In order to estimate the proportion to within 3% with 95% confidence, we would need a sample of at least 164 people. All decimals in the final answer must be rounded up, to the next person.
(A more cautious approach would be to use $\hat{p} = 0.50$, resulting in a sample size of 1068.)

Chapter Exercises

9. Margin of error.

The newscaster believes the true proportion of voters with a certain opinion is within 4% of the estimate, with some degree of confidence, perhaps 95% confidence.

10. Margin of error.

He believes the true percentage of children who are exposed to lead-base paint is within 3% of his estimate, with some degree of confidence, perhaps 95% confidence.

11. Conditions.

a) *Population* – all cars; *sample* – 134 cars actually stopped at the checkpoint; p – proportion of all cars with safety problems; $\hat{p}$ – proportion of cars in the sample that actually have safety problems (10.4%).
Randomization condition: This sample is not random, so hopefully the cars stopped are representative of cars in the area.
10% condition: The 134 cars stopped represent a small fraction of all cars, certainly less than 10%.
Success/Failure condition: $n\hat{p} = 14$ and $n\hat{q} = 120$ are both greater than 10, so the sample is large enough.
A one-proportion z-interval can be created for the proportion of all cars in the area with safety problems.

b) *Population* – the general public; *sample* – 602 viewers that logged on to the Web site;
p –proportion of the general public that support prayer in school; $\hat{p}$ - proportion of viewers that logged on to the Web site and voted that support prayer in schools (81.1%).
Randomization condition: This sample is not random, but biased by voluntary response.
It would be very unwise to attempt to use this sample to infer anything about the opinion of the general public related to school prayer.

c) *Population* – parents at the school; *sample* – 380 parents who returned surveys;
p – proportion of all parents in favor of uniforms; $\hat{p}$ – proportion of those who responded that are in favor of uniforms (60%).
Randomization condition: This sample is not random, but rather biased by nonresponse. There may be lurking variables that affect the opinions of parents who return surveys (and the children who deliver them!).
It would be very unwise to attempt to use this sample to infer anything about the opinion of the parents about uniforms.

d) *Population* – all freshmen enrollees at the college (not just one year); *sample* – 1632 freshmen during the specified year; p – proportion of all students who will graduate on time; $\hat{p}$ – proportion of students from that year who graduate on time (85.05%).
Randomization condition: This sample is not random, but this year's freshmen class is probably representative of freshman classes in other years.
10% condition: The 1632 students in that years freshmen class represent less than 10% of all possible students.
Success/Failure condition: $n\hat{p} = 1388$ and $n\hat{q} = 244$ are both greater than 10, so the sample is large enough.
A one-proportion *z*-interval can be created for the proportion of freshmen that graduate on time from this college.

12. More conditions.

a) *Population* – all customers who recently bought new cars; *sample* – 167 people surveyed about their experience; p – proportion of all new car buyers who are dissatisfied with the salesperson; $\hat{p}$ – proportion of new car buyers surveyed who are dissatisfied with the salesperson (3%).
Success/Failure condition: $n\hat{p} = 167(0.03) = 5$ and $n\hat{q} = 162$. Since only 5 people were dissatisfied, the sample is **not** large enough to use a confidence interval to estimate the proportion of dissatisfied car buyers.

b) *Population* – all college students; *sample* – 2883 who were asked about their cell phones at the football stadium; p – proportion of all college students with cell phones; $\hat{p}$ – proportion of college students at the football stadium with cell phones (84.3%).
Randomization condition: This sample is not random. The best we can hope for is that the students at the football stadium are representative of all college students.
10% condition: The 2883 students at the football stadium represent less than 10% of all college students.
Success/Failure condition: $n\hat{p} = 2430$ and $n\hat{q} = 453$ are both greater than 10, so the sample is large enough.
Extreme caution should be used when using a one-proportion z-interval to estimate the proportion of college students with cell phones. The students at the football stadium may not be representative of all students.

c) *Population* – potato plants in the U.S.; *sample* – 240 potato plants in a field in Maine; p – proportion of all potato plants in the U.S. that show signs of blight; $\hat{p}$ – proportion of potato plants in the sample that show signs of blight (2.9%).
Randomization condition: Although potato plants are randomly selected from the field in Maine, it doesn't seem reasonable to assume that these potato plants are representative of all potato plants in the U.S.
Success/Failure condition: $n\hat{p} = 7$ and $n\hat{q} = 233$. There are only 7 (less than 10!) plants with signs of blight. The sample is not large enough.
Three conditions are not met! Don't use a confidence interval to attempt to estimate the percentage of potato plants in the U.S. that show signs of blight.

d) *Population* – all employees at the company; *sample* – all employees during the specified year; p – proportion of all employees who will have an injury on the job in a year; $\hat{p}$ – proportion of employees who had an injury on the job during the specified year.
Randomization condition: This sample is not random, but this year's employees are probably representative of employees in other years, with regards to injury on the job (3.9%).
10% condition: The 309 employees represent less than 10% of all possible employees over many years.
Success/Failure condition: $n\hat{p} = 12$ and $n\hat{q} = 297$ are both greater than 10, so the sample is large enough.
A one-proportion z-interval can be created for the proportion of employees who are expected to suffer an injury on the job in future years, provided that this year is representative of future years.

13. Conclusions.

- **a)** Not correct. This statement implies certainty. There is no level of confidence in the statement.
- **b)** Not correct. Different samples will give different results. Many fewer than 95% of samples are expected to have *exactly* 88% on-time orders.
- **c)** Not correct. A confidence interval should say something about the unknown population proportion, not the sample proportion in different samples.
- **d)** Not correct. We *know* that 88% of the orders arrived on time. There is no need to make an interval for the sample proportion.
- **e)** Not correct. The interval should be about the proportion of on-time orders, not the days.

14. More conclusions.

- **a)** Not correct. This statement implies certainty. There is no level of confidence in the statement.
- **b)** Not correct. We *know* that 56% of the spins in this experiment landed heads. There is no need to make an interval for the sample proportion.
- **c)** Correct.
- **d)** Not correct. The interval should be about the proportion of heads, not the spins.
- **e)** Not correct. The interval should be about the proportion of heads, not the percentage of euros.

15. Confidence intervals.

- **a)** False. For a given sample size, higher confidence means a *larger* margin of error.
- **b)** True. Larger samples lead to smaller standard errors, which lead to smaller margins of error.
- **c)** True. Larger samples are less variable, which makes us more confident that a given confidence interval succeeds in catching the population proportion.
- **d)** False. The margin of error decreases as the square root of the sample size increases. Halving the margin of error requires a sample four times as large as the original.

16. Confidence intervals, again.

- **a)** True. The smaller the margin of error is, the less confidence we have in the ability of our interval to catch the population proportion.
- **b)** True. Larger samples are less variable, which translates to a smaller margin of error. We can be more precise at the same level of confidence.

c) True. Smaller samples are more variable, leading us to be less confident in the ability of our interval to catch the true population proportion.

d) True. The margin of error decreases as the square root of the sample size increases.

17. Cars.

We are 90% confident that between 29.9% and 47.0% of cars are made in Japan.

18. Parole.

We are 95% confident that between 56.1% and 62.5% of paroles are granted by the Nebraska Board of Parole.

19. Mislabeled seafood.

a) $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{42}{190}\right) \pm 1.960\sqrt{\frac{\left(\frac{42}{190}\right)\left(\frac{148}{190}\right)}{190}} = (0.162, 0.280)$

b) We are 95% confident that between 16.2% and 28.0% of all seafood packages sold in these three states are mislabeled.

c) The size of the population is irrelevant. If *Consumer Reports* had a random sample, 95% of intervals generated by studies like this are expected to capture the true proportion of seafood packages that are mislabeled.

20. Mislabeled seafood, second course.

a) **Randomization condition:** It's not clear how the sample was chosen, but we will assume that the seafood packages, which came from various kinds of establishments in 3 states, is representative of all seafood packages.
10% condition: 22 is far less than 10% of all packages of "red snapper"..
Success/Failure condition: $n\hat{p} = 12$ and $n\hat{q} = 10$ are at least 10, so the sample is large enough.

b) Since the conditions are met, we can use a one-proportion *z*-interval to estimate the percentage of "red snapper" packages that are mislabeled.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{12}{22}\right) \pm 1.960\sqrt{\frac{\left(\frac{12}{22}\right)\left(\frac{10}{22}\right)}{22}} = (0.34, 0.75)$$

c) We are 95% confident that between 34% and 75% of all "red snapper" packages are mislabeled.

21. Baseball fans.

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645 \times \sqrt{\frac{(0.48)(0.52)}{1006}} \approx 0.026$

b) We're 90% confident that this poll's estimate is within 2.6% of the true proportion of people who are baseball fans.

c) The margin of error for 99% confidence would be larger. To be more certain, we must be less precise.

d) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 2.576 \times \sqrt{\frac{(0.48)(0.52)}{1006}} \approx 0.041$

e) Smaller margins of error involve less confidence. The narrower the confidence interval, the less likely we are to believe that we have succeeded in capturing the true proportion of people who are baseball fans.

22. Still lying about age.

a) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.960 \times \sqrt{\frac{(0.49)(0.51)}{799}} \approx 3.5\%$

b) The pollsters are 95% confident that the true proportion of teens who have misrepresented their age is within 3.5% of the estimated 49%.

c) A 90% confidence interval results in a smaller margin of error. If confidence is decreased, a smaller interval is constructed.

d) $ME = z^* \times SE(\hat{p}) = z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.645 \times \sqrt{\frac{(0.49)(0.51)}{799}} \approx 2.9\%$

e) Smaller samples generally produce larger intervals. Smaller samples are more variable, which increases the margin of error.

23. Contributions please.

a) **Randomization condition:** Letters were sent to a random sample of 100,000 potential donors.
10% condition: We assume that the potential donor list has more than 1,000,000 names.
Success/Failure condition: $n\hat{p} = 4781$ and $n\hat{q} = 95{,}219$ are both much greater than 10, so the sample is large enough.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{4781}{100{,}000}\right) \pm 1.960 \sqrt{\frac{\left(\frac{4781}{100{,}000}\right)\left(\frac{95{,}219}{100{,}000}\right)}{100{,}000}} = (0.0465, 0.0491)$$

We are 95% confident that the between 4.65% and 4.91% of potential donors would donate.

b) The confidence interval gives the set of plausible values with 95% confidence. Since 5% is above the interval, it seems to be a bit optimistic.

24. Take the offer.

a) **Randomization condition:** Offers were sent to a random sample of 50,000 cardholders.
10% condition: We assume that there are more than 500,000 cardholders.

Success/Failure condition: $n\hat{p} = 1184$ and $n\hat{q} = 48{,}816$ are both much greater than 10, so the sample is large enough.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{1184}{50{,}000}\right) \pm 1.960 \sqrt{\frac{\left(\frac{1184}{50{,}000}\right)\left(\frac{48{,}816}{50{,}000}\right)}{50{,}000}} = (0.0223, 0.0250)$$

We are 95% confident that the between 2.23% and 2.5% of all cardholders would register for double miles.

b) The confidence interval gives the set of plausible values with 95% confidence. Since 2% is below the interval, there is evidence that the true proportion is above 2%. The campaign should be worth the expense.

25. Teenage drivers.

a) **Randomization condition:** The insurance company randomly selected 582 accidents.
10% condition: 582 accidents represent less than 10% of all accidents.
Success/Failure condition: $n\hat{p} = 91$ and $n\hat{q} = 491$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion *z*-interval to estimate the percentage of accidents involving teenagers.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{91}{582}\right) \pm 1.960 \sqrt{\frac{\left(\frac{91}{582}\right)\left(\frac{491}{582}\right)}{582}} = (12.7\%, 18.6\%)$$

b) We are 95% confident that between 12.7% and 18.6% of all accidents involve teenagers.

c) About 95% of random samples of size 582 will produce intervals that contain the true proportion of accidents involving teenagers.

d) Our confidence interval contradicts the assertion of the politician. The figure quoted by the politician, 1 out of every 5, or 20%, is above the interval.

26. Junk mail.

a) **Independence assumption:** There is no reason to believe that one randomly selected person's response will affect another's.
Randomization condition: The company randomly selected 1000 recipients.
10% condition: 1000 recipients is less than 10% of the population of 200,000 people.
Success/Failure condition: $n\hat{p} = 123$ and $n\hat{q} = 877$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of people who will respond to the new flyer.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{123}{1000}\right) \pm 1.645 \sqrt{\frac{\left(\frac{123}{1000}\right)\left(\frac{877}{1000}\right)}{1000}} = (10.6\%, 14.0\%)$$

b) We are 90% confident that between 10.6% and 14.0% of people will respond to the new flyer.

c) About 90% of random samples of size 1000 will produce intervals that contain the true proportion of people who will respond to the new flyer.

d) Our confidence interval suggests that the company should do the mass mailing. The entire interval is well above the cutoff of 5%.

27. Safe food.

The grocer can conclude nothing about the opinions of all his customers from this survey. Those customers who bothered to fill out the survey represent a voluntary response sample, consisting of people who felt strongly one way or another about irradiated food. The random condition was not met.

28. Local news.

The city council can conclude nothing about general public support for the mayor's initiative. Those who showed up for the meeting are probably a biased group. In addition, a show of hands vote may influence people, affecting the independence of the votes.

29. Death penalty, again.

a) There may be response bias based on the wording of the question.

b) $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.61) \pm 1.960 \sqrt{\frac{(0.61)(0.39)}{1020}} = (58\%, 64\%)$

c) The margin of error based on the pooled sample is smaller, since the sample size is larger.

30. Gambling.

a) The interval based on the survey conducted by the college Statistics class will have the larger margin of error, since the sample size is smaller.

b) **Independence assumption:** There is no reason to believe that one randomly selected voter's response will influence another.
Randomization condition: Both samples were random.
10% condition: Both samples are probably less than 10% of the city's voters, provided the city has more than 12,000 voters.

Success/Failure condition:
For the newspaper, $n_1\hat{p}_1 = (1200)(0.53) = 636$ and $n_1\hat{q}_1 = (1200)(0.47) = 564$
For the Statistics class, $n_2\hat{p}_2 = (450)(0.54) = 243$ and $n_2\hat{q}_2 = (450)(0.46) = 207$
All the expected successes and failures are greater than 10, so the samples are large enough.

Since the conditions are met, we can use one-proportion z-intervals to estimate the proportion of the city's voters that support the gambling initiative.

Newspaper poll: $$\hat{p}_1 \pm z^* \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = (0.53) \pm 1.960\sqrt{\frac{(0.53)(0.47)}{1200}} = (50.2\%, 55.8\%)$$

Statistics class poll: $$\hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}} = (0.54) \pm 1.960\sqrt{\frac{(0.54)(0.46)}{450}} = (49.4\%, 58.6\%)$$

c) The Statistics class should conclude that the outcome is too close to call, because 50% is in their interval.

31. Rickets.

a) **Randomization condition:** The 2700 children were chosen at random.
10% condition: 2700 children are less than 10% of all English children.
Success/Failure condition: $n\hat{p} = (2700)(0.20) = 540$ and $n\hat{q} = (2700)(0.80) = 2160$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the proportion of the English children with vitamin D deficiency.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.20) \pm 2.326\sqrt{\frac{(0.20)(0.80)}{2700}} = (18.2\%, 21.8\%)$$

b) We are 98% confident that between 18.2% and 21.8% of English children are deficient in vitamin D.

c) About 98% of random samples of size 2700 will produce confidence intervals that contain the true proportion of English children that are deficient in vitamin D.

32. Teachers.

a) **Randomization condition:** The poll was conducted from a random sample.
10% condition: 1002 people is less than 10% of all Americans.
Success/Failure condition: $n\hat{p} = 762$ and $n\hat{q} = 240$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the proportion of all Americans who believe that high-achieving high school students should be recruited to become teachers.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{762}{1002}\right) \pm 1.645 \sqrt{\frac{\left(\frac{762}{1002}\right)\left(\frac{240}{1002}\right)}{1002}} = (73.8\%, 78.3\%)$$

b) We are 90% confident that between 73.8% and 78.3% of all Americans believe that high-achieving high school students should be recruited to become teachers.

c) About 90% of random samples of size 1002 will produce confidence intervals that contain the true proportion of all Americans who believe that high-achieving high school students should be recruited to become teachers.

d) These data refute the pundit's claim of that 2/3 of Americans believe this statement, since 66.7% is not in the interval.

33. Privacy or Security?

a) The confidence interval will be wider. The sample size is probably about one-sixth (17%) of the sample size of for all adults, so we would expect the confidence interval to be about two and a half times as wide.

b) The second poll's margin of error should be slightly wider. There are fewer "young" people (13%) in the sample than seniors (17%).

34. Back to campus, again.

a) The confidence interval for the retention rate in private colleges will be narrower than the confidence interval for the retention rate in public colleges, since it is based on a larger sample.

b) Since the overall sample size is larger, the margin of error in retention rate is expected to be smaller.

35. Deer ticks.

a) **Independence assumption:** Deer ticks are parasites. A deer carrying the parasite may spread it to others. Ticks may not be distributed evenly throughout the population.
Randomization condition: The sample is not random and may not represent all deer.
10% condition: 153 deer are less than 10% of all deer.
Success/Failure condition: $n\hat{p} = 32$ and $n\hat{q} = 121$ are both greater than 10, so the sample is large enough.

The conditions are not satisfied, so we should use caution when a one-proportion *z*-interval is used to estimate the proportion of deer carrying ticks.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{32}{153}\right) \pm 1.645 \sqrt{\frac{\left(\frac{32}{153}\right)\left(\frac{121}{153}\right)}{153}} = (15.5\%, 26.3\%)$$

We are 90% confident that between 15.5% and 26.3% of deer have ticks.

b) In order to cut the margin of error in half, they must sample 4 times as many deer.
4(153) = 612 deer.

c) The incidence of deer ticks is not plausibly independent, and the sample may not be representative of all deer, since females and young deer are usually not hunted.

36. Back to campus III.

a) In order to cut the margin of error in half, they must sample 4 times as many college freshmen.
4(1644) = 6576.

b) A sample this large may be more than 10% of the population of all potential students.

37. Graduation.

a)

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.06 = 1.645\sqrt{\frac{(0.25)(0.75)}{n}}$$

$$n = \frac{(1.645)^2(0.25)(0.75)}{(0.06)^2}$$

$$n \approx 141 \text{ people}$$

In order to estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 6% with 90% confidence, we would need a sample of at least 141 people. All decimals in the final answer must be rounded up, to the next person.
(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 188 people.)

b)

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 1.645\sqrt{\frac{(0.25)(0.75)}{n}}$$

$$n = \frac{(1.645)^2(0.25)(0.75)}{(0.04)^2}$$

$$n \approx 318 \text{ people}$$

In order to estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 4% with 90% confidence, we would need a sample of at least 318 people. All decimals in the final answer must be rounded up, to the next person.
(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 423 people.)
Alternatively, the margin of error is now 2/3 of the original, so the sample size must be increased by a factor of 9/4. $141(9/4) \approx 318$ people.

c)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.03 = 1.645\sqrt{\frac{(0.25)(0.75)}{n}}$$
$$n = \frac{(1.645)^2(0.25)(0.75)}{(0.03)^2}$$
$$n \approx 564 \text{ people}$$

In order estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 3% with 90% confidence, we would need a sample of at least 564 people. All decimals in the final answer must be rounded up, to the next person.
(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 752 people.)

Alternatively, the margin of error is now half that of the original, so the sample size must be increased by a factor of 4. $141(4) \approx 564$ people.

38. Hiring.

a)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.05 = 2.326\sqrt{\frac{(0.5)(0.5)}{n}}$$
$$n = \frac{(2.326)^2(0.5)(0.5)}{(0.05)^2}$$
$$n \approx 542 \text{ businesses}$$

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 5% with 98% confidence, we would need a sample of at least 542 businesses. All decimals in the final answer must be rounded up, to the next business.

b)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.03 = 2.326\sqrt{\frac{(0.5)(0.5)}{n}}$$
$$n = \frac{(2.326)^2(0.5)(0.5)}{(0.03)^2}$$
$$n \approx 1503 \text{ businesses}$$

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 3% with 98% confidence, we would need a sample of at least 1503 businesses. All decimals in the final answer must be rounded up, to the next business.

(Alternatively, the margin of error is being decreased to 3/5 of its original size, so the sample size must increase by a factor of 25/9. $542(25/9) \approx 1506$ businesses. A bit off, because 542 was rounded, but close enough!

c)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.01 = 2.326\sqrt{\frac{(0.5)(0.5)}{n}}$$
$$n = \frac{(2.326)^2(0.5)(0.5)}{(0.01)^2}$$
$$n \approx 13,526 \text{ businesses}$$

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 1%with 98% confidence, we would need a sample of at least 13,526 businesses.

(Alternatively, the margin of error has been decreased to 1/5 of its original size, so a sample 25 times as large would be needed. 25(542) = 13,550. Close enough!

It would probably be very expensive and time consuming to sample that many businesses.

39. Graduation, again.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.02 = 1.960\sqrt{\frac{(0.25)(0.75)}{n}}$$
$$n = \frac{(1.960)^2(0.25)(0.75)}{(0.02)^2}$$
$$n \approx 1801 \text{ people}$$

In order to estimate the proportion of non-graduates in the 25-to 30-year-old age group to within 2% with 95% confidence, we would need a sample of at least 1801 people. All decimals in the final answer must be rounded up, to the next person.
(For a more cautious answer, let $\hat{p} = \hat{q} = 0.5$. This method results in a required sample of 2401 people.)

40. Better hiring info.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.04 = 2.576\sqrt{\frac{(0.5)(0.5)}{n}}$$
$$n = \frac{(2.576)^2(0.5)(0.5)}{(0.04)^2}$$
$$n \approx 1037 \text{ businesses}$$

In order to estimate the percentage of businesses planning to hire additional employees within the next 60 days to within 4% with 99% confidence, we would need a sample of at least 1037 businesses. All decimals in the final answer must be rounded up, to the next business.

41. Pilot study.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.03 = 1.645\sqrt{\frac{(0.15)(0.85)}{n}}$$
$$n = \frac{(1.645)^2(0.15)(0.85)}{(0.03)^2}$$
$$n \approx 384 \text{ cars}$$

Use $\hat{p} = \frac{9}{60} = 0.15$ from the pilot study as an estimate.

In order to estimate the percentage of cars with faulty emissions systems to within 3% with 90% confidence, the state's environmental agency will need a sample of at least 384 cars. All decimals in the final answer must be rounded up, to the next car.

42. Another pilot study.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.04 = 2.326\sqrt{\frac{(0.22)(0.78)}{n}}$$
$$n = \frac{(2.326)^2(0.22)(0.78)}{(0.04)^2}$$
$$n \approx 581 \text{ adults}$$

Use $\hat{p} = 0.22$ from the pilot study as an estimate.

In order to estimate the percentage of adults with higher than normal levels of glucose in their blood to within 4% with 98% confidence, the researchers will need a sample of at least 581 adults. All decimals in the final answer must be rounded up, to the next adult.

43. Approval rating.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.025 = z^*\sqrt{\frac{(0.65)(0.35)}{972}}$$
$$z^* = \frac{0.025}{\sqrt{\frac{(0.65)(0.35)}{972}}}$$
$$z^* \approx 1.634$$

Since $z^* \approx 1.634$, which is close to 1.645, the pollsters were probably using 90% confidence. The slight difference in the z^* values is due to rounding of the governor's approval rating.

44. Amendment.

a) This poll is inconclusive because the confidence interval, 52% ± 3% contains 50%. The true proportion of voters in favor of the constitutional amendment is estimated to be between 49% (minority) to 55% (majority). We can't be sure whether or not the majority of voters support the amendment or not.

b)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.03 = z^*\sqrt{\frac{(0.52)(0.48)}{1505}}$$

$$z^* = \frac{0.03}{\sqrt{\frac{(0.52)(0.48)}{1505}}}$$

$$z^* \approx 2.3295$$

Since $z^* \approx 2.3295$, which is close to 2.326, the pollsters were probably using 98% confidence. The slight difference in the z^* values is due to rounding of the amendment's approval rating.

Chapter 19 – Testing Hypotheses about Proportions

Section 19.1

1. Better than aspirin?

a) The new drug is not more effective than aspirin, and reduces the risk of heart attack by 44%. ($p = 0.44$)

b) The new drug is more effective than aspirin, and reduces the risk of heart attack by more than 44%. ($p > 0.44$)

2. Psychic.

a) Your friend's chance of guessing the correct suit is 25%. ($p = 0.25$)

b) Your friend's chance of guessing the correct suit is more than 25%. ($p > 0.25$)

Section 19.2

3. Better than aspirin 2?

a) Since the *P*-value of 0.28 is greater than 0.05, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the new drug is better than aspirin.

b) Since the *P*-value of 0.004 is less than 0.05, reject the null hypothesis. There is evidence that the new drug is more effective than aspirin.

4. Psychic after all?

a) Since the *P*-value of 0.014 is less than 0.05, reject the null hypothesis. There is evidence that your friend's chance of guessing the correct suit is greater than 25%. (You may not want to jump to the conclusion of psychic powers, however.)

b) Since the *P*-value of 0.245 is greater than 0.05, fail to reject the null hypothesis. There is not enough evidence to conclude that your friend's ability to guess the correct suit is greater than 25%.

Section 19.3

5. Hispanic origin.

a) H_0 : The proportion of people in the county that are of Hispanic or Latino origin is 0.16. (p = 0.16)

H_A : The proportion of people in the county that are of Hispanic or Latino origin is different from 0.16. ($p \neq 0.16$)

b) **Randomization condition:** The 437 county residents were a random sample of all county residents.
10% condition: 437 is likely to be less than 10% of all county residents.

Success/Failure condition: $np_0 = (437)(0.16) = 69.92$ and $nq_0 = (437)(0.84) = 367.08$ are both greater than 10, so the sample is large enough.

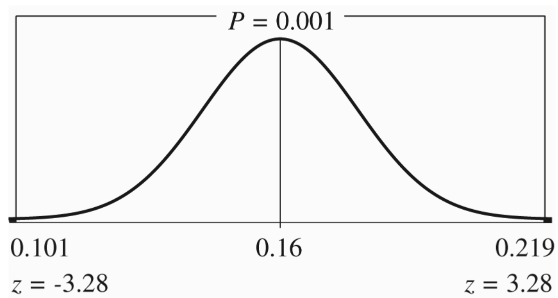
Since the conditions are met, we will model the sampling distribution of $\hat{p}$ with a Normal model and perform a one-proportion z-test.

c) $\hat{p} = \dfrac{44}{437} = 0.101$

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.16)(0.84)}{437}} \approx 0.018$$

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})} = \frac{0.101 - 0.16}{0.018} = -3.28$$

$$P\text{-value} = 2 \bullet P(z < -3.28) = 0.001$$



d) Since the P-value = 0.001 is so low, reject the null hypothesis. There is evidence that the Hispanic/Latino population in this county differs from that as the nation as a whole. These data suggest that the proportion of Hispanic/Latino residents is, in fact, lower than the national proportion.

6. Empty houses.

a) H_0 : The proportion of vacant houses in the county 0.114. ($p = 0.114$)

H_A : The proportion of vacant houses is different from 0.114. ($p \neq 0.114$)

b) **Randomization condition:** The 850 housing units were a random sample of all housing units in the county.
10% condition: 850 is likely less than 10% of all housing units in the county.
Success/Failure condition: $np_0 = (850)(0.114) = 96.9$ and $nq_0 = (850)(0.886) = 753.1$ are both greater than 10, so the sample is large enough.

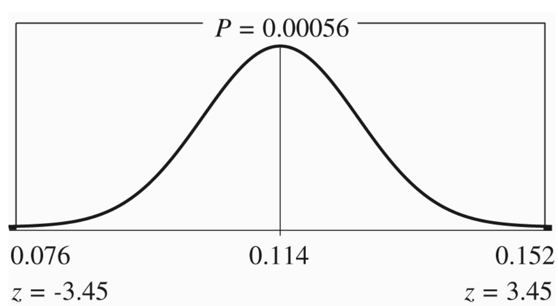
Since the conditions are met, we will model the sampling distribution of $\hat{p}$ with a Normal model and perform a one-proportion z-test.

c) $\hat{p} = \dfrac{129}{850} = 0.152$

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.114)(0.886)}{850}} \approx 0.011$$

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})} = \frac{0.152 - 0.114}{0.011} = 3.45$$

$$P\text{-value} = 2 \bullet P(z < 3.45) = 0.00056$$



d) Since the P-value = 0.00056 is so low, reject the null hypothesis. There is evidence that the proportion of houses that are vacant in this county is different than the national proportion. These data suggest that the proportion of vacant houses is, in fact, higher than the national proportion.

Section 19.4

7. Psychic again (you should have seen this coming).

The alternative hypothesis would be one-sided, because the only evidence that would support the friend's claim is guessing more than 25% of the suits correctly.

8. Hispanic origin II.

The alternative hypothesis would be two-sided, because the county supervisor simply suspected her county differed from the nation as a whole.

Section 19.5

9. Bad medicine.

a) The drug may not be approved for use, and people would miss out on a beneficial product.

b) The drug will go into production and people will suffer the side effect.

10. Expensive medicine.

a) People will pay more money for a drug that is no better than the old drug.

b) The new drug would not be used, so people would miss an opportunity for a more effective drug.

Chapter Exercises

11. Hypotheses.

a) H_0 : The governor's "negatives" are 30%. ($p = 0.30$)
H_A : The governor's "negatives" are less than 30%. ($p < 0.30$)

b) H_0 : The proportion of heads is 50%. ($p = 0.50$)
H_A : The proportion of heads is not 50%. ($p \neq 0.50$)

c) H_0 : The proportion of people who quit smoking is 20%. ($p = 0.20$)
H_A : The proportion of people who quit smoking is greater than 20%. ($p > 0.20$)

12. More hypotheses.

a) H_0 : The proportion of high school graduates is 40%. ($p = 0.40$)
H_A : The proportion of high school graduates is not 40%. ($p \neq 0.40$)

b) H_0 : The proportion of cars needing transmission repair is 20%. ($p = 0.20$)
H_A : The proportion of cars is less than 20%. ($p < 0.20$)

c) H_0 : The proportion of people who like the flavor is 60%. ($p = 0.60$)
H_A : The proportion of people who like the flavor is greater than 60%. ($p > 0.60$)

13. Negatives.

Statement d is the correct interpretation of a *P*-value.

14. Dice.

Statement d is the correct interpretation of a *P*-value.

15. Relief.

It is *not* reasonable to conclude that the new formula and the old one are equally effective. Furthermore, our inability to make that conclusion has nothing to do with the *P*-value. We can not prove the null hypothesis (that the new formula and the old formula are equally effective), but can only fail to find evidence that would cause us to reject it. All we can say about this *P*-value is that there is a 27% chance of seeing the observed effectiveness from natural sampling variation if the new formula and the old one are equally effective.

16. Cars.

It is reasonable to conclude that a greater proportion of high schoolers have cars. If the proportion were no higher than it was a decade ago, there is only a 1.7% chance of seeing such a high sample proportion just from natural sampling variability.

17. He cheats?

a) Two losses in a row aren't convincing. There is a 25% chance of losing twice in a row, and that is not unusual.

b) If the process is fair, three losses in a row can be expected to happen about 12.5% of the time. $(0.5)(0.5)(0.5) = 0.125$.

c) Three losses in a row is still not a convincing occurrence. We'd expect that to happen about once every eight times we tossed a coin three times.

d) Answers may vary. Maybe 5 times would be convincing. The chances of 5 losses in a row are only 1 in 32, which seems unusual.

18. Candy.

a) $P(\text{first three vanilla}) = \left(\frac{6}{12}\right)\left(\frac{5}{11}\right)\left(\frac{4}{10}\right) \approx 0.091$

b) It seems reasonable to think there really may have been six of each. We would expect to get three vanillas in a row about 9% of the time. That's unusual, but not *that* unusual.

c) If the fourth candy was also vanilla, we'd probably start to think that the mix of candies was not 6 vanilla and 6 peanut butter. The probability of 4 vanilla candies in a row is:

$$P(\text{first four vanilla}) = \left(\frac{6}{12}\right)\left(\frac{5}{11}\right)\left(\frac{4}{10}\right)\left(\frac{3}{9}\right) \approx 0.03$$

We would only expect to get four vanillas in a row about 3% of the time. That's unusual.

19. Smartphones.

1) Null and alternative hypotheses should involve p, not $\hat{p}$.

2) The question is about *failing* to meet the goal. H_A should be $p < 0.96$.

3) The student failed to check $nq_0 = (200)(0.04) = 8$. Since $nq_0 < 10$, the Success/Failure condition is violated. Similarly, the 10% Condition is not verified.

4) $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.96)(0.04)}{200}} \approx 0.014$. The student used $\hat{p}$ and $\hat{q}$.

5) Value of z is incorrect. The correct value is $z = \frac{0.94 - 0.96}{0.014} \approx -1.43$.

6) P-value is incorrect. $P = P(z < -1.43) = 0.076$

7) For the P-value given, an incorrect conclusion is drawn. A P-value of 0.12 provides no evidence that the new system has failed to meet the goal. The correct conclusion for the corrected P-value is: Since the P-value of 0.076 is fairly low, there is weak evidence that the new system has failed to meet the goal.

20. Obesity 2008.

1) Null and alternative hypotheses should involve p, not $\hat{p}$.

2) The question asks if there is evidence that the 34% figure is *not accurate*, so a two-sided alternative hypothesis should be used. H_A should be $p \neq 0.34$.

3) The conditions are SRS, 750 < 10% of county population, $np_0 = (750)(0.34) = 255 \geq 10$, $nq_0 = (750)(0.66) = 495 \geq 10$

4) $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.34)(0.66)}{750}} \approx 0.017$. Rounded values of $\hat{p}$ and $\hat{q}$ were used.

5) Value of z is incorrect. The correct value is $z = \frac{0.304 - 0.34}{0.017} \approx -2.12$.

6) The correct, two-tailed P-value is $P = 2P(z < -2.12) = 0.034$.

7) The P-value is misinterpreted. Since the P-value is so low, there is moderately strong evidence that the proportion of adults in the county who are obese is different than the claimed 34%. In fact, our sample suggests that the proportion may be lower. There is only a 3.4% chance of observing a $\hat{p}$ as far from 0.34 as this simply from natural sampling variation.

21. Dowsing.

a) H_0 : The percentage of successful wells drilled by the dowser is 30%. ($p = 0.30$)
H_A : The percentage of successful wells drilled is greater than 30%. ($p > 0.30$)

b) **Independence assumption:** There is no reason to think that finding water in one well will affect the probability that water is found in another, unless the wells are close enough to be fed by the same underground water source.
Randomization condition: This sample is not random, so hopefully the customers you check with are representative of all of the dowser's customers.
10% condition: The 80 customers sampled may be considered less than 10% of all possible customers.
Success/Failure condition: $np_0 = (80)(0.30) = 24$ and $nq_0 = (80)(0.70) = 56$ are both greater than 10, so the sample is large enough.

c) The sample of customers may not be representative of all customers, so we will proceed cautiously. A Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.30$ and

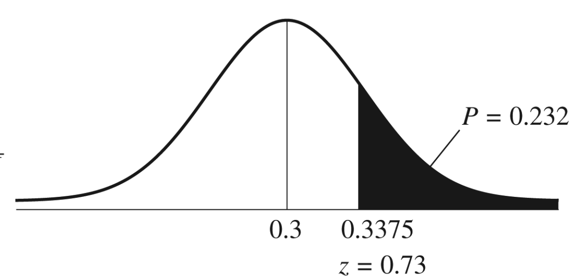
$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.30)(0.70)}{80}} \approx 0.0512.$$

We can perform a one-proportion z-test. The observed proportion of successful wells is $\hat{p} = \frac{27}{80} = 0.3375$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.3375 - 0.30}{\sqrt{\frac{(0.30)(0.70)}{80}}}$$

$$z \approx 0.73$$



d) If his dowsing has the same success rate as standard drilling methods, there is more than a 23% chance of seeing results as good as those of the dowser, or better, by natural sampling variation.

e) With a P-value of 0.232, we fail to reject the null hypothesis. There is no evidence to suggest that the dowser has a success rate any higher than 30%.

22. Abnormalities.

a) H_0 : The percentage of children with genetic abnormalities is 5%.(p = 0.05)
H_A : The percentage of with genetic abnormalities is greater than 5%. (p > 0.05)

b) **Randomization condition:** This sample may not be random, but genetic abnormalities are plausibly independent. The sample is probably representative of all children, with regards to genetic abnormalities.
10% condition: The sample of 384 children is less than 10% of all children.
Success/Failure condition: $np_0 = (384)(0.05) = 19.2$ and $nq_0 = (384)(0.95) = 364.8$ are both greater than 10, so the sample is large enough.

c) The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 =$ 0.05 and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.05)(0.95)}{384}} \approx 0.0111.$$

We can perform a one-proportion z-test. The observed proportion of children with genetic abnormalities is $\hat{p} = \frac{46}{384} \approx 0.1198$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

$$z = \frac{0.1198 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{384}}}$$

$$z \approx 6.28$$

The value of z is approximately 6.28, meaning that the observed proportion of children with genetic abnormalities is over 6 standard deviations above the hypothesized proportion. The P-value associated with this z score is 2×10^{-10}, essentially 0.

d) If 5% of children have genetic abnormalities, the chance of observing 46 children with genetic abnormalities in a random sample of 384 children is essentially 0.

e) With a P-value of this low, we reject the null hypothesis. There is strong evidence that more than 5% of children have genetic abnormalities.

f) We don't know that environmental chemicals cause genetic abnormalities. We merely have evidence that suggests that a greater percentage of children are diagnosed with genetic abnormalities now, compared to the 1980s.

23. Absentees.

a) H_0 : The percentage of students in 2000 with perfect attendance the previous month is 34% (p = 0.34)
H_A : The percentage of students in 2000 with perfect attendance the previous month is different from 34% ($p \neq 0.34$)

b) **Randomization condition:** Although not specifically stated, we can assume that the National Center for Educational Statistics used random sampling.
10% condition: The 8302 students are less than 10% of all students.
Success/Failure condition: $np_0 = (8302)(0.34) = 2822.68$ and $nq_0 = (8302)(0.66) = 5479.32$ are both greater than 10, so the sample is large enough.

c) Since the conditions for inference are met, a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.34$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.34)(0.66)}{8302}} \approx 0.0052$$

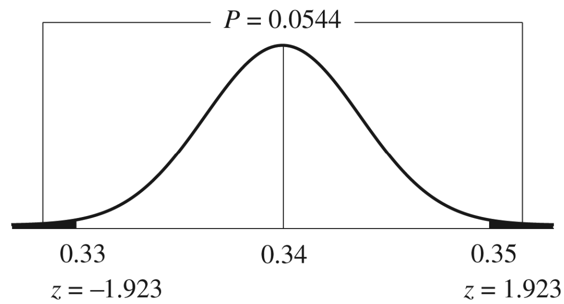
We can perform a two-tailed one-proportion z-test. The observed proportion of perfect attendees is $\hat{p} = 0.33$.

d) With a P-value of 0.0544, we reject the null hypothesis. There is some evidence to suggest that the percentage of students with perfect attendance in the previous month has changed in 2000.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.33 - 0.34}{\sqrt{\frac{(0.34)(0.66)}{8302}}}$$

$$z \approx -1.923$$



e) This result is not meaningful. A difference this small, although statistically significant, is of little practical significance.

24. Educated mothers.

a) H_0 : The percentage of students in 2000 whose mothers had graduated college is 31% ($p = 0.31$)
H_A : The percentage of students is different than 31% ($p \neq 0.31$)

b) **Randomization condition:** Although not specifically stated, we can assume that the National Center for Educational Statistics used random sampling.
10% condition: The 8368 students are less than 10% of all students.
Success/Failure condition: $np_0 = (8368)(0.31) = 2594.08$ and $nq_0 = (8368)(0.69) = 5773.92$ are both greater than 10, so the sample is large enough.

c) Since the conditions for inference are met, a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.31$ and

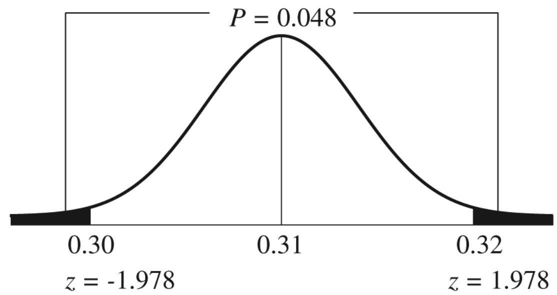
$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.31)(0.69)}{8368}} \approx 0.0051$$

We can perform a one-proportion two-tailed z-test. The observed proportion of students whose mothers are college graduates is $\hat{p} = 0.32$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.32 - 0.31}{\sqrt{\frac{(0.31)(0.69)}{8368}}}$$

$$z \approx 1.978$$



d) With a P-value of 0.048, we reject the null hypothesis. There is evidence to suggest that the percentage of students whose mothers are college graduates has changed since 1996. In fact, the evidence suggests that the percentage has increased.

e) This result is not meaningful. A difference this small, although statistically significant, is of little practical significance.

25. Contributions, please, part II.

a) H_0 : The contribution rate is 5% ($p = 0.05$)
H_A : The contribution rate is less than 5% ($p < 0.05$)

b) **Randomization condition:** Potential donors were randomly selected.
10% condition: We will assume the entire mailing list has over 1,000,000 names.
Success/Failure condition: $np_0 = 5000$ and $nq_0 = 95{,}000$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.05$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.05)(0.95)}{100{,}000}} \approx 0.0007.$$

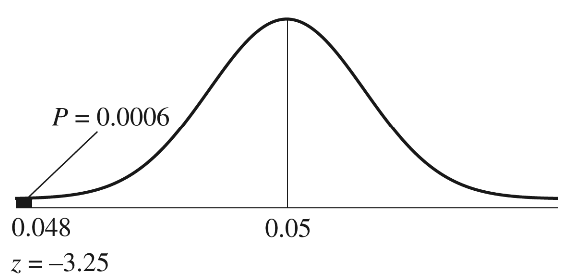
We can perform a one-proportion z-test. The observed contribution rate is $\hat{p} = \frac{4781}{100{,}000} = 0.04781$.

c) Since the P-value = 0.0006 is low, we reject the null hypothesis. There is strong evidence that contribution rate for all potential donors is lower than 5%.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.048 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{100{,}000}}}$$

$$z \approx -3.25$$



26. Take the offer, part II.

a) H_0 : The success rate is 2% ($p = 0.02$)
H_A : The success rate is something other than 2% ($p \neq 0.02$)

b) **Randomization condition:** The sample was 50,000 randomly selected cardholders.
10% condition: We will assume that the number of cardholders is more than 500,000.
Success/Failure condition: $np_0 = 1000$ and $nq_0 = 49{,}000$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.02$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.02)(0.98)}{50{,}000}} \approx 0.0006.$$

We can perform a one-proportion z-test. The observed success rate is
$$\hat{p} = \frac{1184}{50{,}000} = 0.02368.$$

c) Since the P-value is less than 0.0001, we reject the null hypothesis. There is strong evidence that success rate for all cardholders is not 2%. In fact, this sample suggests that the success rate is higher than 2%.

27. Law school 2011.

a) H_0 : The law school acceptance rate for *LSATisfaction* is 71% ($p = 0.71$)
H_A : The law school acceptance rate is greater than 71% ($p > 0.71$)

b) **Randomization condition:** These 240 students may be considered representative of the population of law school applicants.
10% condition: There are certainly more than 2400 law school applicants.
Success/Failure condition: $np_0 = 170.4$ and $nq_0 = 69.6$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.71$ and

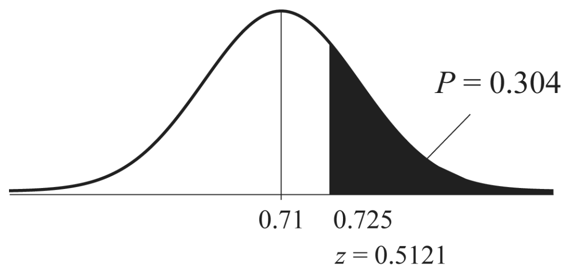
$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.71)(0.29)}{240}} \approx 0.0293.$$

We can perform a one-proportion z-test. The observed success rate is
$$\hat{p} = \frac{174}{240} = 0.725.$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.725 - 0.71}{\sqrt{\frac{(0.71)(0.29)}{240}}}$$

$$z \approx 0.5121$$



c) Since the P-value = 0.304 is high, we fail to reject the null hypothesis. There is no evidence that the law school acceptance rate is higher for *LSATisfaction* applicants than for applicants nationwide.

28. Med school 2013.

a) H_0 : The med school acceptance rate for Striving College is 42% ($p = 0.42$)
H_A : The med school acceptance rate is less than 42% ($p < 0.42$)

b) **Randomization condition:** Assume that these 81 students are representative of all applicants from this college.
10% condition: 81 students represent less than 10% of all applicants.
Success/Failure condition: $np_0 = 34.02$ and $nq_0 = 46.98$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.42$ and

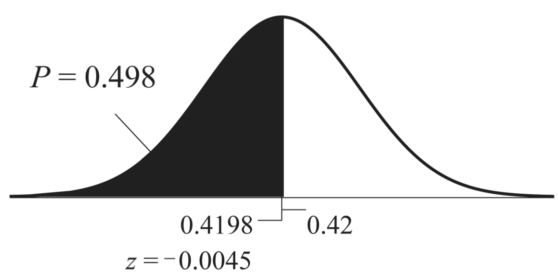
$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.42)(0.58)}{81}} \approx 0.0548.$$

We can perform a one-proportion z-test. The observed success rate is
$\hat{p} = \frac{34}{81} = 0.4198$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.4198 - 0.42}{\sqrt{\frac{(0.42)(0.58)}{81}}}$$

$$z \approx -0.0045$$



c) Since the P-value = 0.498 is high, we fail to reject the null hypothesis. There is no evidence that the med school acceptance rate at Striving College is significantly lower than 42%. This could simply be year-to-year variation, as the president suggests.

29. Pollution.

H_0 : The percentage of cars with faulty emissions is 20%. ($p = 0.20$)
H_A : The percentage of cars with faulty emissions is greater than 20%. ($p > 0.20$)

Two conditions are not satisfied. 22 is greater than 10% of the population of 150 cars, and $np_0 = (22)(0.20) = 4.4$, which is not greater than 10. It's not advisable to proceed with a test.

30. Scratch and dent.

H_0 : The percentage of damaged machines is 2%, and the warehouse is meeting the company goal. ($p = 0.02$)
H_A : The percentage of damaged machines is greater than 2%, and the warehouse is failing to meet the company goal. ($p > 0.02$)

An important condition is not satisfied. $np_0 = (60)(0.02) = 1.2$, which is not greater than 10. The Normal model is not appropriate for modeling the sampling distribution.

31. Twins.

H_0 : The percentage of twin births to teenage girls is 3%. ($p = 0.03$)
H_A : The percentage of twin births to teenage girls differs from 3%. ($p \neq 0.03$)

Independence assumption: One mother having twins will not affect another. Observations are plausibly independent.
Randomization condition: This sample may not be random, but it is reasonable to think that this hospital has a representative sample of teen mothers, with regards to twin births.
10% condition: The sample of 469 teenage mothers is less than 10% of all such mothers.
Success/Failure condition: $np_0 = (469)(0.03) = 14.07$ and $nq_0 = (469)(0.97) = 454.93$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.03$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.03)(0.97)}{469}} \approx 0.0079.$$

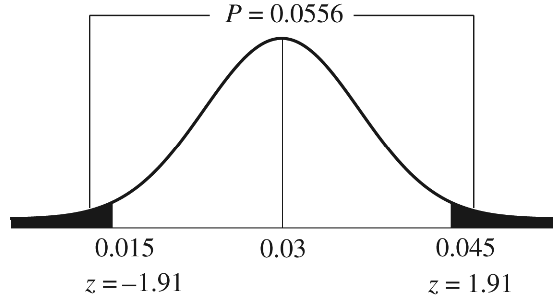
We can perform a one-proportion z-test. The observed proportion of twin births to teenage mothers is $\hat{p} = \frac{7}{469} \approx 0.015$.

Since the P-value = 0.0556 is fairly low, we reject the null hypothesis. There is some evidence that the proportion of twin births for teenage mothers at this large city hospital is lower than the proportion of twin births for all mothers.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.015 - 0.03}{\sqrt{\frac{(0.03)(0.97)}{469}}}$$

$$z \approx -1.91$$



32. Football 2013.

H_0 : The percentage of home team wins is 50%. (p = 0.50)
H_A : The percentage of home team wins is greater than 50%. (p > 0.50)

Independence assumption: Results of one game should not affect others.
Randomization condition: This season should be representative of other seasons, with regards to home team wins.
10% condition: 245 games represent less than 10% of all games, in all seasons.
Success/Failure condition: $np_0 = (245)(0.50) = 122.5$ and $nq_0 = (245)(0.50) = 122.5$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.50$ and

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.5)(0.5)}{245}} \approx 0.0319.$$

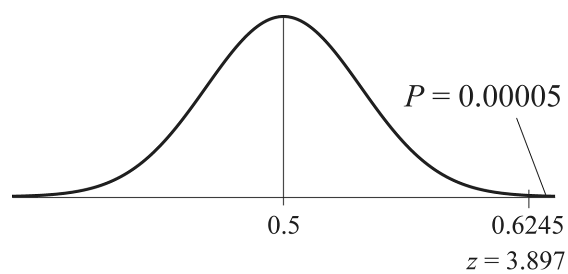
We can perform a one-proportion z-test. The observed proportion of home team wins is $\hat{p} = \frac{153}{245} = 0.6245$.

Since the P-value = 0.00005 is low, we reject the null hypothesis. There is strong evidence that the proportion of home teams wins is greater than 50%. This provides evidence of a home team advantage.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.6245 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{245}}}$$

$$z \approx 3.897$$



33. WebZine.

H_0 : The percentage of readers interested in an online edition is 25%. ($p = 0.25$)
H_A : The percentage of readers interested is greater than 25%. ($p > 0.25$)

Randomization condition: The magazine conducted an SRS of 500 current readers.
10% condition: 500 readers are less than 10% of all potential subscribers.
Success/Failure condition: $np_0 = (500)(0.25) = 125$ and $nq_0 = (500)(0.75) = 375$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.25$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.25)(0.75)}{500}} \approx 0.0194.$$

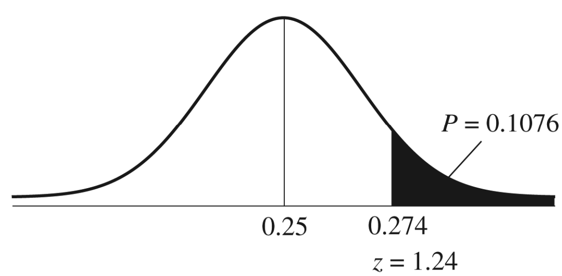
We can perform a one-proportion z-test. The observed proportion of interested readers is $\hat{p} = \frac{137}{500} = 0.274$.

Since the P-value = 0.1076 is high, we fail to reject the null hypothesis. There is little evidence to suggest that the proportion of interested readers is greater than 25%. The magazine should not publish the online edition.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.274 - 0.25}{\sqrt{\frac{(0.25)(0.75)}{500}}}$$

$$z \approx 1.24$$



34. Seeds.

H_0 : The germination rate of the green bean seeds is 92%. ($p = 0.92$)
H_A : The germination rate of the green bean seeds is less than 92%. ($p < 0.92$)

Independence assumption: Seeds in a single packet may not germinate independently. They have been treated identically with regards to moisture exposure, temperature, etc. They may have higher or lower germination rates than seeds in general.
Randomization condition: The cluster sample of one bag of seeds was not random.
10% condition: 200 seeds is less than 10% of all seeds.
Success/Failure condition: $np_0 = (200)(0.92) = 184$ and $nq_0 = (200)(0.08) = 16$ are both greater than 10, so the sample is large enough.

The conditions have *not* been satisfied. We will assume that the seeds in the bag are representative of all seeds, and cautiously use a Normal model to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.92$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.92)(0.08)}{200}} \approx 0.0192.$$

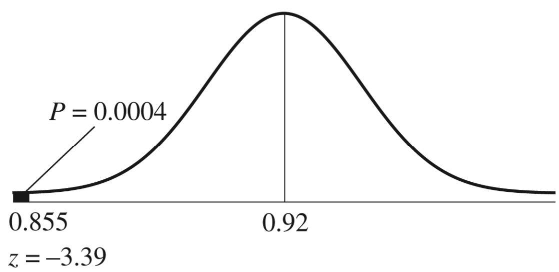
We can perform a one-proportion z-test. The observed proportion of germinated seeds is $\hat{p} = \frac{171}{200} = 0.855$.

Since the P-value = 0.0004 is very low, we reject the null hypothesis. There is strong evidence that the germination rate of the seeds in less than 92%. We should use extreme caution in generalizing these results to all seeds, but the manager should be safe, and avoid selling faulty seeds. The seeds should be thrown out.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.855 - 0.92}{\sqrt{\frac{(0.92)(0.08)}{200}}}$$

$$z \approx -3.39$$



35. Women executives.

H_0 : The proportion of female executives is similar to the overall proportion of female employees at the company. ($p = 0.40$)
H_A : The proportion of female executives is lower than the overall proportion of female employees at the company. ($p < 0.40$)

Independence assumption: It is reasonable to think that executives at this company were chosen independently.
Randomization condition: The executives were not chosen randomly, but it is reasonable to think of these executives as representative of all potential executives over many years.
10% condition: 43 executives are less than 10% of all executives at the company.
Success/Failure condition: $np_0 = (43)(0.40) = 17.2$ and $nq_0 = (43)(0.60) = 25.8$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.40$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.40)(0.60)}{43}} \approx 0.0747.$$

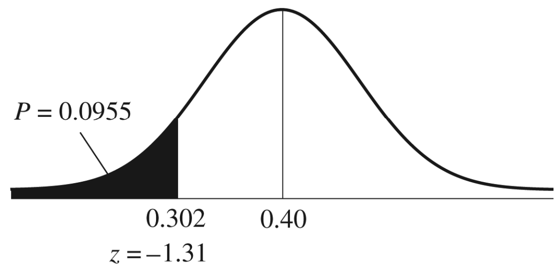
Perform a one-proportion z-test. The observed proportion is $\hat{p} = \frac{13}{43} \approx 0.302$.

Since the P-value = 0.0955 is high, we fail to reject the null hypothesis. There is little evidence to suggest proportion of female executives is any different from the overall proportion of 40% female employees at the company.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.302 - 0.40}{\sqrt{\frac{(0.40)(0.60)}{43}}}$$

$$z \approx -1.31$$



36. Jury.

H_0 : The proportion of Hispanics called for jury duty is similar to the proportion of Hispanics in the county, 19%. ($p = 0.19$)
H_A : The proportion of Hispanics called for jury duty is less than the proportion of Hispanics in the county, 19%. ($p < 0.19$)

Independence assumption /Randomization condition: Assume that potential jurors were called randomly from all of the residents in the county. This is really what we are testing. If we reject the null hypothesis, we will have evidence that jurors are not called randomly.
10% condition: 72 people are less than 10% of all potential jurors in the county.
Success/Failure condition: $np_0 = (72)(0.19) = 13.68$ and $nq_0 = (72)(0.81) = 58.32$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.19$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.19)(0.81)}{72}} \approx 0.0462.$$

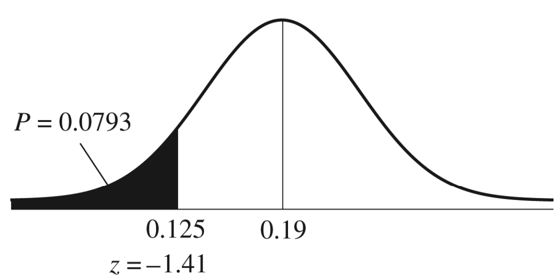
We can perform a one-proportion z-test. The observed proportion of Hispanics called for jury duty is $\hat{p} = \frac{9}{72} = 0.125$.

Since the P-value = 0.0793 is somewhat high, we fail to reject the null hypothesis. We are not convinced that Hispanics are underrepresented in the jury selection system. However, this P-value isn't extremely high. There is some evidence that the selection process may be biased. We should examine some other groups called for jury duty and take a closer look.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.125 - 0.19}{\sqrt{\frac{(0.19)(0.81)}{72}}}$$

$$z \approx -1.41$$



37. Dropouts.

H_0 : The proportion of dropouts at this high school is similar to 10.3%, the proportion of dropouts nationally. (p = 0.103)
H_A : The proportion of dropouts at this high school is greater than 10.3%, the proportion of dropouts nationally. (p > 0.103)

Independence assumption /Randomization condition: Assume that the students at this high school are representative of all students nationally. This is really what we are testing. The dropout rate at this high school has traditionally been close to the national rate. If we reject the null hypothesis, we will have evidence that the dropout rate at this high school is no longer close to the national rate.
10% condition: 1782 students are less than 10% of all students nationally.
Success/Failure condition: $np_0 = (1782)(0.103) = 183.546$ and $nq_0 = (1782)(0.897) = 1598.454$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, $\mu_{\hat{p}} = p_0 = 0.103$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.103)(0.897)}{1782}} \approx 0.0072.$$

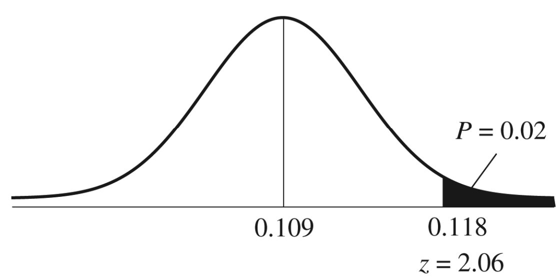
We can perform a one-proportion z-test. The observed proportion of dropouts is $\hat{p} = \frac{210}{1782} \approx 0.117845$.

Since the P-value = 0.02 is low, we reject the null hypothesis. There is evidence that the dropout rate at this high school is higher than 10.3%.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.117845 - 0.109}{\sqrt{\frac{(0.103)(0.897)}{1782}}}$$

$$z \approx 2.06$$



38. Acid rain.

H_0 : The proportion of trees with acid rain damage in Hopkins Forest is 15%, the proportion of trees with acid rain damage in the Northeast. (p = 0.15)
H_A : The proportion of trees with acid rain damage in Hopkins Forest is greater than 15%, the proportion of trees with acid rain damage in the Northeast. (p > 0.15)

Independence assumption /Randomization condition: Assume that the trees in Hopkins Forest are representative of all trees in the Northeast. This is really what we are testing. If we reject the null hypothesis, we will have evidence that the proportion of trees with acid rain damage is greater in Hopkins Forest than the proportion in the Northeast.
10% condition: 100 trees are less than 10% of all trees.
Success/Failure condition: $np_0 = (100)(0.15) = 15$ and $nq_0 = (100)(0.85) = 85$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.15$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.15)(0.85)}{100}} \approx 0.0357 .$$

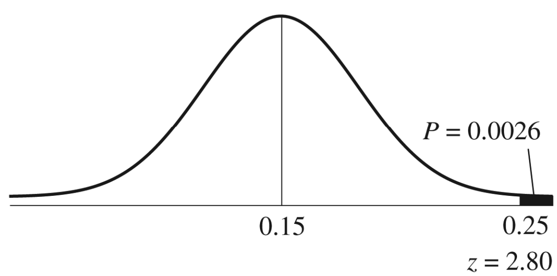
We can perform a one-proportion z-test. The observed proportion of damaged trees is $\hat{p} = \frac{25}{100} = 0.25$.

Since the P-value = 0.0026 is low, we reject the null hypothesis. There is strong evidence that the trees in Hopkins forest have a greater proportion of acid rain damage than the 15% reported for the Northeast.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.25 - 0.150}{\sqrt{\frac{(0.150)(0.850)}{100}}}$$

$$z \approx 2.80$$



$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{25}{100}\right) \pm 1.96 \sqrt{\frac{\left(\frac{25}{100}\right)\left(\frac{75}{100}\right)}{100}} = (0.165, 0.335)$$

We are 95% confident that the interval 0.165 to 0.335 captures the true proportion of trees in the Hopkins forest that are damaged by acid rain, which is higher than the 15% reported for the Northeast.

39. Lost luggage.

H_0 : The proportion of lost luggage returned the next day is 90%. ($p = 0.90$)
H_A : The proportion of lost luggage returned is lower than 90%. ($p < 0.90$)

Independence assumption: It is reasonable to think that the people surveyed were independent with regards to their luggage woes.
Randomization condition: Although not stated, we will hope that the survey was conducted randomly, or at least that these air travelers are representative of all air travelers for that airline.

10% condition: 122 air travelers are less than 10% of all air travelers on the airline.
Success/Failure condition: $np_0 = (122)(0.90) = 109.8$ and $nq_0 = (122)(0.10) = 12.2$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.90$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.90)(0.10)}{122}} \approx 0.0272.$$

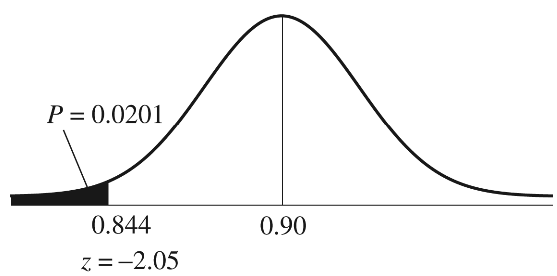
We can perform a one-proportion z-test. The observed proportion of lost luggage is $\hat{p} = \frac{103}{122} \approx 0.844$.

Since the P-value = 0.0201 is low, we reject the null hypothesis. There is evidence that the proportion of lost luggage returned the next day is lower than the 90% claimed by the airline.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.844 - 0.90}{\sqrt{\frac{(0.90)(0.10)}{122}}}$$

$$z \approx -2.05$$



40. TV ads.

H_0 : The proportion of respondents who recognize the name is 40%.($p = 0.40$)
H_A : The proportion is more than 40%. ($p > 0.40$)

Randomization condition: The pollster contacted the 420 adults randomly.
10% condition: A sample of 420 adults is less than 10% of all adults.
Success/Failure condition: $np_0 = (420)(0.40) = 168$ and $nq_0 = (420)(0.60) = 252$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.40$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.40)(0.60)}{420}} \approx 0.0239.$$

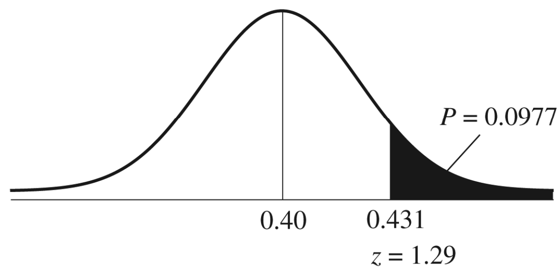
We can perform a one-proportion z-test. The observed proportion is $\hat{p} = \frac{181}{420} \approx 0.431$.

Since the P-value = 0.0977 is fairly high, we fail to reject the null hypothesis. There is little evidence that more than 40% of the public recognizes the product. Don't run commercials during the Super Bowl!

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.431 - 0.40}{\sqrt{\frac{(0.40)(0.60)}{420}}}$$

$$z \approx 1.29$$



41. John Wayne.

a) H_0 : The death rate from cancer for people working on the film was similar to that predicted by cancer experts, 30 out of 220.
H_A : The death rate from cancer for people working on the film was higher than the rate predicted by cancer experts.

The conditions for inference are not met, since this is not a random sample. We will assume that the cancer rates for people working on the film are similar to those predicted by the cancer experts, and a Normal model can be used to model the sampling distribution of the rate, with $\mu_{\hat{p}} = p_0 = 30/220$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{\left(\frac{30}{220}\right)\left(\frac{190}{220}\right)}{220}} \approx 0.0231.$$

We can perform a one-proportion z-test. The observed cancer rate is $\hat{p} = \frac{46}{220} \approx 0.209$.

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})}$$

$$z = \frac{\frac{46}{220} - \frac{30}{220}}{\sqrt{\frac{\left(\frac{30}{220}\right)\left(\frac{190}{220}\right)}{220}}}$$

$$z = 3.14$$

Since the P-value = 0.0008 is very low, we reject the null hypothesis. There is strong evidence that the cancer rate is higher than expected among the workers on the film.

b) This does not prove that exposure to radiation may increase the risk of cancer. This group of people may be atypical for reasons that have nothing to do with the radiation.

42. AP Stats.

H_0 : These students achieve scores of 3 or higher at a similar rate to the nation. ($p = 0.579$)
H_A : These students achieve these scores at a higher rate than the nation. ($p > 0.579$)

Independence assumption: There is no reason to believe that students' scores would influence others.
Randomization condition: The teacher considers this class typical of other classes.
10% condition: A sample of 54 students is less than 10% of all students.
Success/Failure condition: $np_0 = (54)(0.579) = 31.266$ and $nq_0 = (54)(0.421) = 22.734$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p_0 = 0.579$ and

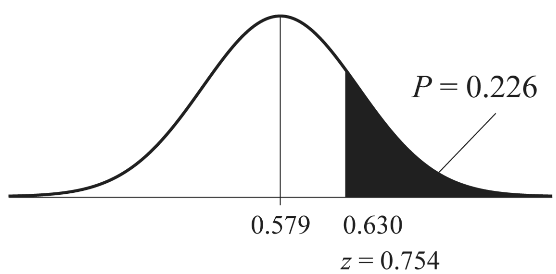
$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.579)(0.421)}{54}} \approx 0.0672.$$

We can perform a one-proportion z-test. The observed pass rate is $\hat{p} = 0.630$.

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})}$$

$$z = \frac{\frac{34}{54} - 0.579}{\sqrt{\frac{(0.579)(0.421)}{54}}}$$

$$z = 0.754$$



Since the P-value = 0.226 is high, we fail to reject the null hypothesis. There is little evidence that the rate at which these students score 3 or higher on the AP Stats exam is any higher than the national rate.

The teacher has no cause to brag. Her students did have a higher rate of scores of 3 or higher, but not so high that the results could not be attributed to sampling variability.

Chapter 20 – Inferences About Means

Section 20.1

1. Salmon.

a) The shipment of 4 salmon has $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{4}} = 1$ pound.

The shipment of 16 salmon has $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{16}} = 0.5$ pounds.

The shipment of 100 salmon has $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2$ pounds.

b) The Normal model would better characterize the shipping weight of the pallets than the shipping weight of the boxes, since the pallets contain a large number of salmon. The Central Limit Theorem tells us that the distribution of means (and therefore totals) approaches the Normal model, regardless of the underlying distribution. As samples get larger, the approximation gets better.

2. LSAT.

a) Since the distribution of LSAT scores for all test takers is unimodal and symmetric, the distribution of scores for the test takers at these test preparation organizations are probably at least free of outliers and skewness. Therefore, the distribution of the mean score of classes of size 9 and 25 should each be roughly Normal, and each would have a mean of 151 points. The standard deviation of the distribution of the means would be $\frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{9}} = 3$ points for the class of size 9 and $\frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{25}} = 1.8$ points for the class of size 25.

b) The organization with the smaller class has a larger standard deviation of the mean. A class mean score of 160 is 3 standard deviations above the population mean, which is rare, but could happen. For the larger organization, a class mean of 160 is 5 standard deviations above the population mean, which would be highly unlikely

c) The smaller organization is at a greater risk of having to pay for LSAT retakes. They are more likely to have a low class mean for the same reason they are more likely to have a high class mean. The variability in the class mean score is greater when the class size is small.

Section 20.2

3. *t*-models, part I.

a) 1.74 **b)** 2.37

4. *t*-models, part II.

a) 2.36 **b)** 2.62

5. *t*-models, part III.

As the number of degrees of freedom increases, the shape and center of *t*-models do not change. The spread of *t*-models decreases as the number of degrees of freedom increases, and the shape of the distribution becomes closer to Normal.

6. *t*-models, part IV.

As the number of degrees of freedom increases, the critical value of t for a 95% confidence interval gets smaller, approaching approximately 1.960, the critical value of z for a 95% confidence interval.

Section 20.3

7. Home sales.

a) The estimates of home value losses must be independent. This is verified using the Randomization condition, since the houses were randomly sampled. The distribution of home value losses must be Normal. A histogram of home value losses in the sample would be checked to verify this, using the Nearly Normal condition. Even if the histogram is not unimodal and symmetric, the sample size of 36 should allow for some small departures from Normality.

b) $\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 9560 \pm t^*_{35}\left(\frac{1500}{\sqrt{36}}\right) \approx (9052.50,\ 10067.50)$

c) We are 95% confident that average home value loss is between \$9052.50 and \$10,067.50.

8. Home sales, again.

a) A larger standard deviation in home value losses would increase the width of the confidence interval.

b) Your classmate is correct. A lower confidence level results in a narrower interval.

c) A larger sample would reduce the standard error, since larger samples result in lower variability in the distribution of means than smaller samples, which makes the interval narrower. This is more statistically appropriate, since we could narrow the interval without sacrificing confidence. However, it may be difficult or expensive to increase the sample size, so it may not be practical.

Section 20.4

9. ***t*-models, again.**

a) 0.0524 **b)** 0.0889

10. ***t*-models, last time.**

c) 0.9829 **d)** 0.0381

11. Home prices.

H_0: The mean loss in home value has not decreased from the mean loss in home value in 2010 of \$10,200. $(\mu = 10,200)$

H_A: The mean loss in home value has decreased significantly from the mean loss in home value in 2010 of \$10,200. $(\mu < 10,200)$

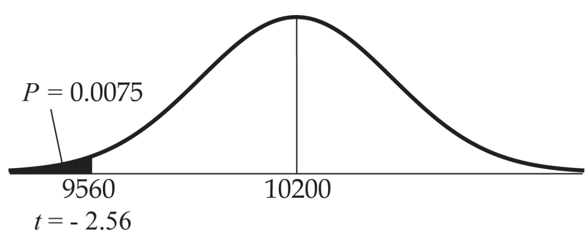
Randomization condition: The home value losses were collected from 36 randomly selected homes.
Nearly Normal condition: We don't have the data, so we can't make a scatterplot, but with a sample size of 36 homes, the Central Limit Theorem should allow us to use a Normal model to describe the sampling distribution of mean home value losses.

In 2010, home value losses had a mean of \$10,200 and we will use the sample standard deviation of \$1500. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean home value loss with a Student's t model, with 36 – 1 = 35 degrees of freedom, $t_{35}\left(10200, \frac{1500}{\sqrt{36}}\right)$. We will perform a one-sample t-test.

$$t = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{9560 - 10200}{\frac{1500}{\sqrt{36}}}$$

$$t = -2.56$$



Since the P-value = 0.0075 is low, we reject the null hypothesis. There is evidence that the loss of home values in this community has decreased.

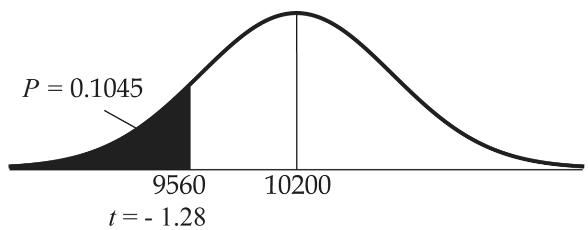
12. Home prices II.

In 2010, home value losses had a mean of \$10,200 and we will use the sample standard deviation of \$3000.Since the conditions for inference are satisfied, we can model the sampling distribution of the mean home value loss with a Student's t model, with 36 – 1 = 35 degrees of freedom, $t_{35}\left(10200, \frac{3000}{\sqrt{36}}\right)$. We will perform a one-sample t-test.

$$t = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{9560 - 10200}{\frac{3000}{\sqrt{36}}}$$

$$t = -1.28$$



Since the P-value = 0.1045 is high, we fail to reject the null hypothesis. There is no evidence that the loss of home values in this community has decreased.

Section 20.5

13. Jelly.

A preliminary calculation with $z^* = 2.576$ gives an approximate sample size of 27. Refining the approximation with $df = n - 1 = 27 - 1 = 26$ and $t_{26}^* = 2.779$ gives an approximate sample size of 31. The consumer advocate should collect a sample of at least 31 jelly jars to achieve a margin of error of no more than 2 grams.

$$ME = z^*\left(\frac{s}{\sqrt{n}}\right) \qquad ME = t_{n-1}^*\left(\frac{s}{\sqrt{n}}\right)$$

$$2 = 2.576\left(\frac{4}{\sqrt{n}}\right) \qquad 2 = 2.779\left(\frac{4}{\sqrt{n}}\right)$$

$$n = \frac{(2.576)^2(4)^2}{(2)^2} \qquad n = \frac{(2.779)^2(4)^2}{(2)^2}$$

$$n \approx 27 \qquad n \approx 31$$

14. A good book.

A preliminary calculation with $z^* = 1.96$ gives an approximate sample size of 43. Refining the approximation with $df = n - 1 = 43 - 1 = 42$ and $t_{42}^* = 2.018$ gives an approximate sample size of 46. The professor should survey at least 46 students to achieve a margin of error of at most 3 books.

$$ME = z^*\left(\frac{s}{\sqrt{n}}\right) \qquad ME = t_{n-1}^*\left(\frac{s}{\sqrt{n}}\right)$$

$$3 = 1.96\left(\frac{10}{\sqrt{n}}\right) \qquad 3 = 2.018\left(\frac{10}{\sqrt{n}}\right)$$

$$n = \frac{(1.96)^2(10)^2}{(3)^2} \qquad n = \frac{(2.018)^2(10)^2}{(2)^2}$$

$$n \approx 43 \qquad n \approx 46$$

Chapter Exercises

15. Cattle.

- **a)** Not correct. A confidence interval is for the mean weight gain of the population of all cows. It says nothing about individual cows. This interpretation also appears to imply that there is something special about the interval that was generated, when this interval is actually one of many that could have been generated, depending on the cows that were chosen for the sample.
- **b)** Not correct. A confidence interval is for the mean weight gain of the population of all cows, not individual cows.
- **c)** Not correct. We don't need a confidence interval about the average weight gain for cows in this study. We are certain that the mean weight gain of the cows in this study is 56 pounds. Confidence intervals are for the mean weight gain of the population of all cows.
- **d)** Not correct. This statement implies that the average weight gain varies. It doesn't. We just don't know what it is, and we are trying to find it. The average weight gain is either between 45 and 67 pounds, or it isn't.
- **e)** Not correct. This statement implies that there is something special about our interval, when this interval is actually one of many that could have been generated, depending on the cows that were chosen for the sample. The correct interpretation is that 95% of samples of this size will produce an interval that will contain the mean weight gain of the population of all cows.

16. Teachers.

- **a)** Not correct. Actually, 9 out of 10 samples will produce intervals that will contain the mean salary for Nevada teachers. Different samples are expected to produce different intervals.
- **b)** Correct! This is the one!
- **c)** Not correct. A confidence interval is about the mean salary of the population of Nevada teachers, not the salaries of individual teachers.
- **d)** Not correct. A confidence interval is about the mean salary of the population of Nevada teachers and doesn't tell us about the sample, nor does it tell us about individual salaries.
- **e)** Not correct. The population is teachers' salaries in Nevada, not the entire United States.

17. Meal plan.

- **a)** Not correct. The confidence interval is not about the individual students in the population.

b) Not correct. The confidence interval is not about individual students in the sample. In fact, we know exactly what these students spent, so there is no need to estimate.

c) Not correct. We know that the mean cost for students in this sample was $1467.

d) Not correct. A confidence interval is not about other sample means.

e) This is the correct interpretation of a confidence interval. It estimates a population parameter.

18. Snow.

a) Not correct. The confidence interval is not about the winters in the sample.

b) Not correct. The confidence interval does not predict what will happen in any one winter.

c) Not correct. The confidence interval is not based on a sample of days.

d) This is the correct interpretation of a confidence interval. It estimates a population parameter.

e) Not correct. We know exactly what the mean was in the sample. The mean snowfall was 23″ per winter over the last century.

19. Pulse rates.

a) We are 95% confident the interval 70.9 to 74.5 beats per minute contains the true mean heart rate.

b) The width of the interval is about 74.5 – 70.9 = 3.6 beats per minute. The margin of error is half of that, about 1.8 beats per minute.

c) The margin of error would have been larger. More confidence requires a larger critical value of t, which increases the margin of error.

20. Crawling.

a) We are 95% confident that the interval 29.2 to 31.8 weeks contains the true mean age at which babies begin to crawl.

b) The width of the interval is about 31.8 – 29.2 = 2.6 weeks. The margin of error is half of that, about 1.3 weeks.

c) The margin of error would have been smaller. Less confidence requires a smaller critical value of t, which decreases the margin of error.

21. CEO compensation.

We should be hesitant to trust this confidence interval, since the conditions for inference are not met. The distribution is highly skewed and there is an outlier.

22. Credit card charges.

The analysts did not find the confidence interval useful because the conditions for inference were not met. There is one cardholder who spent over $3,000,000 on his card. This made the standard deviation, and therefore the standard error, huge. The t-interval is too wide to be of any use.

23. Normal temperature.

a) **Randomization condition:** The adults were randomly selected.
Nearly Normal condition: The sample of 52 adults is large, and the histogram shows no serious skewness, outliers, or multiple modes.

The people in the sample had a mean temperature of 98.2846° and a standard deviation in temperature of 0.682379°. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's t model, with 52 – 1 = 51 degrees of freedom. We will use a one-sample t-interval with 98% confidence for the mean body temperature.
(By hand, use $t_{50}^* \approx 2.403$ from the table.)

b) $\bar{y} \pm t_{n-1}^* \left(\frac{s}{\sqrt{n}} \right) = 98.2846 \pm t_{51}^* \left(\frac{0.682379}{\sqrt{52}} \right) \approx (98.06, 98.51)$

c) We are 98% confident that the interval 98.06°F to 98.51°F contains the true mean body temperature for adults. (If you calculated the interval by hand, using $t_{50}^* \approx 2.403$ from the table, your interval may be slightly different than intervals calculated using technology. With the rounding used here, they are identical. Even if they aren't, it's not a big deal.)

d) 98% of all random samples of size 52 will produce intervals that contain the true mean body temperature of adults.

e) Since the interval is completely below the body temperature of 98.6°F, there is strong evidence that the true mean body temperature of adults is lower than 98.6°F.

24. Parking.

a) **Randomization condition:** The weekdays were not randomly selected. We will assume that the weekdays in our sample are representative of all weekdays.
Nearly Normal condition: We don't have the actual data, but since the sample of 44 weekdays is fairly large it is okay to proceed.

The weekdays in the sample had a mean revenue of $126 and a standard deviation in revenue of $15. The sampling distribution of the mean can be modeled by a Student's t model, with 44 – 1 = 43 degrees of freedom. We will use a one-sample t -interval with 90% confidence for the mean daily income of the parking garage. (By hand, use $t_{40}^* \approx 1.684$)

b) $\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 126 \pm t_{43}^{*}\left(\frac{15}{\sqrt{44}}\right) \approx (122.2,\ 129.8)$

c) We are 90% confident that the interval \$122.20 to \$129.80 contains the true mean daily income of the parking garage. (If you calculated the interval by hand, using $t_{40}^{*} \approx 1.684$ from the table, your interval will be (122.19, 129.81), ever so slightly wider from the interval calculated using technology. This is not a big deal.)

d) 90% of all random samples of size 44 will produce intervals that contain the true mean daily income of the parking garage.

e) Since the interval is completely below the \$130 predicted by the consultant, there is evidence that the average daily parking revenue is lower than \$130.

25. Normal temperatures, part II.

a) The 90% confidence interval would be narrower than the 98% confidence interval. We can be more precise with our interval when we are less confident.

b) The 98% confidence interval has a greater chance of containing the true mean body temperature of adults than the 90% confidence interval, but the 98% confidence interval is less precise (wider) than the 90% confidence interval.

c) The 98% confidence interval would be narrower if the sample size were increased from 52 people to 500 people. The smaller standard error would result in a smaller margin of error.

d) Our sample of 52 people gave us a 98% confidence interval with a margin of error of (98.51 – 98.05)/2 = 0.225°F. In order to get a margin of error of 0.1, less than half of that, we need a sample over 4 times as large. It should be safe to use $t_{100}^{*} \approx 2.364$ from the table, since the sample will need to be larger than 101. Or we could use $z^{*} \approx 2.326$, since we expect the sample to be large. We need a sample of about 252 people in order to estimate the mean body temperature of adults to within 0.1°F.

$$ME = t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right)$$

$$0.1 = 2.326\left(\frac{0.682379}{\sqrt{n}}\right)$$

$$n = \frac{(2.326)^2(0.682379)^2}{(0.1)^2}$$

$$n \approx 252$$

26. Parking II.

a) The 95% confidence interval would be wider than the 90% confidence interval. We can be more confident that our interval contains the mean parking revenue when we are less precise. This would be better for the city because the 95% confidence interval is more likely to contain the true mean parking revenue.

b) The 95% confidence interval is wider than the 90% confidence interval, and therefore less precise. It would be difficult for budget planners to use this wider interval, since they need precise figures for the budget.

c) By collecting a larger sample of parking revenue on weekdays, they could create a more precise interval without sacrificing confidence.

d) The confidence interval that was calculated in Exercise 24 won't help us to estimate the sample size. That interval was for 90% confidence. Now we want 95% confidence. A quick estimate with a critical value of $z^* = 2$ (from the 68-95-99.7 rule) gives us a sample size of 100, which will probably work fine. Let's be a bit more precise, just for fun! Conservatively, let's choose t^* with fewer degrees of freedom, which will give us a wider interval. From the table, the next available number of degrees of freedom is $t^*_{80} \approx 1.990$, not much different than the estimate of 2 that was used before. If we substitute 1.990 for t^*, we can estimate a sample size of about 99. Why not play it a bit safe? Use $n = 100$.

$$ME = t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right)$$
$$3 = 2\left(\frac{15}{\sqrt{n}}\right)$$
$$n = \frac{(2)^2(15)^2}{(3)^2}$$
$$n = 100$$

27. Speed of Light.

a) $\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 756.22 \pm t^*_{22}\left(\frac{107.12}{\sqrt{23}}\right) \approx (709.9,\ 802.5)$

b) We are 95% confident that the interval 299,709.9 to 299,802.5 km/sec contains the speed of light.

c) We have assumed that the measurements are independent of each other and that the distribution of the population of all possible measurements is Normal. The assumption of independence seems reasonable, but it might be a good idea to look at a display of the measurements made by Michelson to verify that the Nearly Normal Condition is satisfied.

28. Better light.

a) $SE(\bar{y}) = \left(\frac{s}{\sqrt{n}}\right) = \left(\frac{79.0}{\sqrt{100}}\right) = 7.9\ \text{km/sec.}$

b) The interval should be narrower. There are three reasons for this: the larger sample size results in a smaller standard error (reducing the margin of error), the larger sample size results in a greater number of degrees of freedom (decreasing the value of t^*, reducing the margin of error), and the smaller standard deviation in measurements results in a smaller standard error (reducing the margin of error). Additionally, the interval will have a different center, since the sample mean is different.

c) We must assume that the measurements are independent of one another. Since the sample size is large, the Nearly Normal Condition is overridden, but it would still be nice to look at a graphical display of the measurements. A one-sample t-interval for the speed of light can be constructed, with 100 – 1 = 99 degrees of freedom, at 95% confidence.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 852.4 \pm t^*_{99}\left(\frac{79.0}{\sqrt{100}}\right) \approx (836.72,\ 868.08)$$

We are 95% confident that the interval 299,836.72 to 299,868.08 km/sec contains the speed of light.

Since the interval for the new method does not contain the true speed of light as reported by Stigler, 299,710.5 km/sec., there is no evidence to support the accuracy of Michelson's new methods.

The interval for Michelson's old method (from Exercise 27) does contain the true speed of light as reported by Stigler. There is some evidence that Michelson's previous measurement technique was a good one, if not very precise.

29. Flight on time 2013.

a) **Randomization condition:** Since there is no time trend, the monthly on-time departure rates should be independent. This is not a random sample, but should be representative.
Nearly Normal condition: The histogram looks unimodal, and slightly skewed to the left. Since the sample size is 238, this should not be of concern.

b) The on-time departure rates in the sample had a mean of 78.02%, and a standard deviation in of 5.045%. Since the conditions have been satisfied, construct a one-sample t-interval, with 238 – 1 = 237 degrees of freedom, at 90% confidence.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 78.02 \pm t^*_{237}\left(\frac{5.045}{\sqrt{238}}\right) \approx (77.48, 78.56)$$

c) We are 90% confident that the interval from 77.48% to 78.56% contains the true mean monthly percentage of on-time flight departures.

30. Flight on time 2013 revisited.

a) **Randomization condition:** Since there is no time trend, the monthly delay rates should be independent. This is not a random sample, but should be representative.
Nearly Normal condition: The histogram looks unimodal and symmetric.

b) The delay rates in the sample had a mean of 19.753%, and a standard deviation in of 4.273%. Since the conditions have been satisfied, construct a one-sample t-interval, with 238 – 1 = 237 degrees of freedom, at 99% confidence.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 19.753 \pm t^*_{237}\left(\frac{4.273}{\sqrt{238}}\right) \approx (19.034, 20.472)$$

c) We are 99% confident that the interval from 19.03% to 20.47% contains the true mean monthly percentage of delayed flights.

31. Farmed salmon, second look.

The 95% confidence interval lies entirely above the 0.08 ppm limit. This is evidence that mirex contamination is too high and consistent with rejecting the null hypothesis. We used an upper-tail test, so the *P*-value should be smaller than $\frac{1}{2}(1-0.95)=0.025$, and it was.

32. Hot dogs.

The 90% confidence interval contains the 325 mg limit. They can't assert that the mean sodium content is less than 325 mg, consistent with not rejecting the null hypothesis. They used an upper-tail test, so the P-value should be more than $\frac{1}{2}(1-0.90)=0.05$, and it was.

33. Pizza.

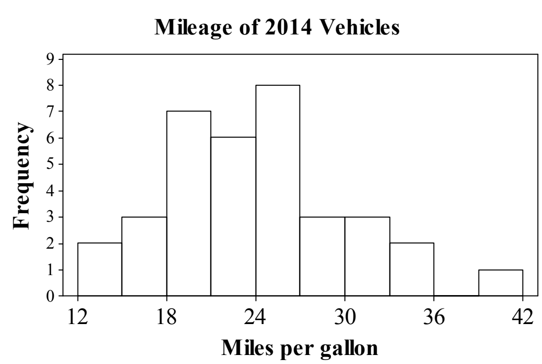
If in fact the mean cholesterol of pizza eaters does not indicate a health risk, then only 7 out of every 100 samples would be expected to have mean cholesterol as high or higher than the mean cholesterol observed in the sample.

34. Golf balls.

If in fact this ball meets the velocity standard, then 34% of samples tested would be expected to have mean speeds at least as high as the mean speed recorded in the sample.

35. Fuel economy 2014 revisited.

a) **Randomization condition:** The 35 cars were not selected randomly. We will have to assume that they are representative of all 2014 automobiles.
Nearly Normal condition: The distribution doesn't appear to be unimodal and symmetric, but the sample size is reasonably large.



The mileages in the sample had a mean of 23.5143 mpg, and a standard deviation in of 6.0603 mpg. Since the conditions have been satisfied, construct a one-sample t-interval, with 35 – 1 = 34 degrees of freedom, at 95% confidence.

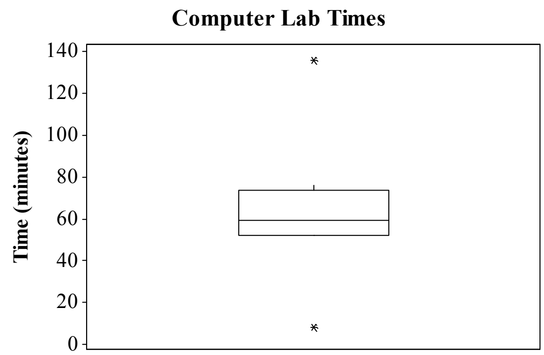
$$\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 23.5143 \pm t_{34}^{*}\left(\frac{6.0603}{\sqrt{35}}\right) \approx (21.43, 25.60)$$

We are 95% confident that the interval from 21.43 to 25.60 contains the true mean mileage of 2014 automobiles.

b) The data is a mix of small, mid-size, and large vehicles. Without knowing how the data were selected, we are cautious about generalizing to all 2014 cars.

36. Computer lab fees.

a) The 8 minute and 136 minute times were extreme outliers. If the outliers are included, the conditions for inference, specifically the Nearly Normal condition, are not met.



b) With outliers: $\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 63.25 \pm t_{11}^{*}\left(\frac{28.927}{\sqrt{12}}\right) \approx (44.9, 81.6)$

Without outliers: $\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 61.5 \pm t_{9}^{*}\left(\frac{9.595}{\sqrt{10}}\right) \approx (54.6, 68.4)$

In either case, we would be reluctant to conclude that the mean is above 55 minutes. The small sample size and the presence of two large outliers causes us to be cautious about conclusions from this sample.

37. Marriage.

a) H_0: The mean age at which American men first marry is 23.3 years. $(\mu = 23.3)$

H_A: The mean age is greater than 23.3 years. $(\mu > 23.3)$

b) **Randomization condition:** The 40 men were selected randomly.
Nearly Normal condition: The population of ages of men at first marriage is likely to be skewed to the right. It is much more likely that there are men who marry for the first time at an older age than at an age that is very young. We should examine the distribution of the sample to check for serious skewness and outliers, but with a large sample of 40 men, it should be safe to proceed.

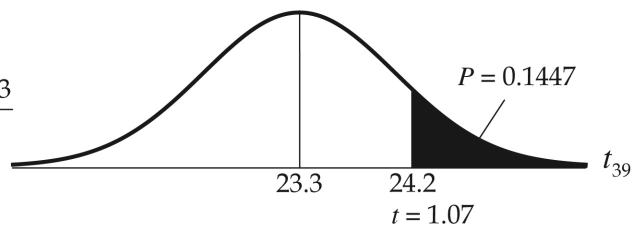
c) Since the conditions for inference are satisfied, we can model the sampling distribution of the mean age of men at first marriage with $N\left(23.3, \frac{\sigma}{\sqrt{n}}\right)$. Since we do not know σ, the standard deviation of the population, $\sigma(\bar{y})$ will be estimated by $SE(\bar{y}) = \frac{s}{\sqrt{n}}$, and we will use a Student's *t* model, with 40 – 1 = 39 degrees of freedom, $t_{39}\left(23.3, \frac{5.3}{\sqrt{40}}\right)$.

d) The mean age at first marriage in the sample was 24.2 years, with a standard deviation in age of 5.3 years. Use a one-sample *t*-test, modeling the sampling distribution of $\bar{y}$ with $t_{39}\left(23.3, \frac{5.3}{\sqrt{40}}\right)$. The *P*-value is 0.1447.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{24.2 - 23.3}{\frac{5.3}{\sqrt{40}}}$$

$$t \approx 1.07$$



e) If the mean age at first marriage is still 23.3 years, there is a 14.5% chance of getting a sample mean of 24.2 years or older simply from natural sampling variation.

f) Since the *P*-value = 0.1447 is high, we fail to reject the null hypothesis. There is no evidence to suggest that the mean age of men at first marriage has changed from 23.3 years, the mean in 1960.

38. Saving gas.

a) H_0: The mean mileage of the cars in the fleet is 30.2 mpg. $(\mu = 30.2)$

H_A: The mean mileage of the cars in the fleet is greater than 30.2 mpg. $(\mu > 30.2)$

b) **Randomization condition:** The 50 trips were selected randomly.
Nearly Normal condition: We don't have the actual data, so we cannot look at the distribution of the data, but the sample is large, so we can proceed.

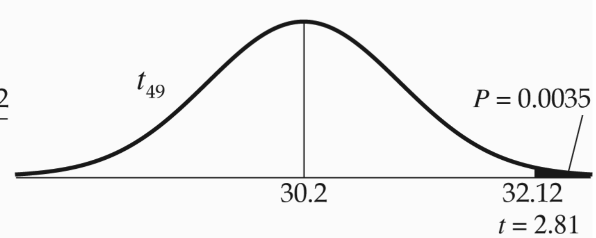
c) Since the conditions for inference are satisfied, we can model the sampling distribution of the mean mileage of cars in the fleet with $N\left(30.2, \frac{\sigma}{\sqrt{n}}\right)$. Since we do not know σ, the standard deviation of the population, $\sigma(\bar{y})$ will be estimated by $SE(\bar{y}) = \frac{s}{\sqrt{n}}$, and we will use a Student's t model, with 50 – 1 = 49 degrees of freedom, $t_{49}\left(30.2, \frac{4.83}{\sqrt{50}}\right)$.

d) The trips in the sample had a mean mileage of 32.12 mpg, with a standard deviation of 4.83 mpg. Use a one-sample t-test, modeling the sampling distribution of $\bar{y}$ with $t_{49}\left(30.2, \frac{4.83}{\sqrt{50}}\right)$. The P-value is 0.0035.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{32.12 - 30.2}{\frac{4.83}{\sqrt{50}}}$$

$$t \approx 2.81$$

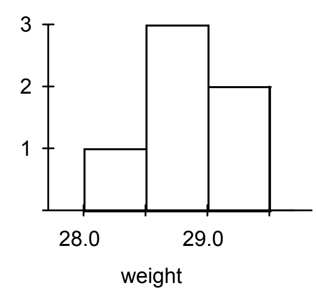


e) If the mean mileage of cars in the fleet is 30.2 mpg, the chance that a sample mean of a sample of size 50 is 32.12 mpg or greater simply due to sampling error is 0.35%.

f) Since the P-value = 0.0035 is is low, we reject the null hypothesis. There is evidence to suggest that the mean mileage of cars in the fleet is more than 30.2 mpg. The company appears to be meeting their goal.

39. Ruffles.

a) **Randomization condition:** The 6 bags were not selected at random, but it is reasonable to think that these bags are representative of all bags of chips.
10% condition: 6 bags are less than 10% of all bags of chips.
Nearly Normal condition: The histogram of the weights of chips in the sample is nearly normal.



b) $\bar{y} \approx 28.78$ grams, $s \approx 0.40$ grams

c) Since the conditions for inference have been satisfied, use a one-sample t-interval, with 6 – 1 = 5 degrees of freedom, at 95% confidence.

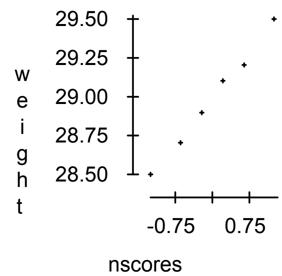
$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 28.78 \pm t^*_5\left(\frac{0.40}{\sqrt{6}}\right) \approx (28.36,\ 29.21)$$

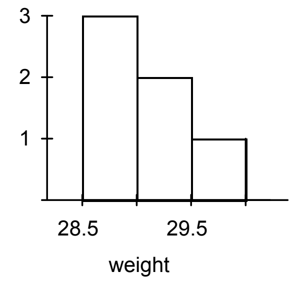
d) We are 95% confident that the mean weight of the contents of Ruffles bags is between 28.36 and 29.21 grams.

e) Since the interval is above the stated weight of 28.3 grams, there is evidence that the company is filling the bags to more than the stated weight, on average.

40. Doritos.

a) **Randomization condition:** The 6 bags were not selected at random, but it is reasonable to think that these bags are representative of all bags.
Nearly Normal condition: The Normal probability plot is reasonably straight. Although the histogram of the weights of chips in the sample is not symmetric, any apparent "skewness" is the result of a single bag of chips. It is safe to proceed.





b) $\bar{y} \approx 28.98$ grams, $s \approx 0.36$ grams

c) Since the conditions for inference have been satisfied, use a one-sample t-interval, with $6 - 1 = 5$ degrees of freedom, at 95% confidence.

$$\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 28.98 \pm t_{5}^{*}\left(\frac{0.36}{\sqrt{6}}\right) \approx (28.61,\ 29.36)$$

d) We are 95% confident that the interval 28.61 to 29.36 grams contains the true mean weight of the contents of Doritos bags.

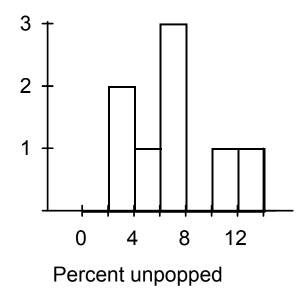
e) Since the interval is above the stated weight of 28.3 grams, there is evidence that the company is filling the bags to more than the stated weight, on average.

41. Popcorn.

H_0: The mean proportion of unpopped kernels is 10%. $(\mu = 10)$

H_A: The mean proportion of unpopped kernels is lower than 10%. $(\mu < 10)$

Randomization condition: The 8 bags were randomly selected.
Nearly Normal condition: The histogram of the percentage of unpopped kernels is unimodal and roughly symmetric.



The bags in the sample had a mean percentage of unpopped kernels of 6.775 percent and a standard deviation in percentage of unpopped kernels of 3.637 percent. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean percentage of unpopped kernels with a Student's t model, with $8 - 1 = 7$ degrees of freedom, $t_7\left(6.775, \frac{3.637}{\sqrt{8}}\right)$.

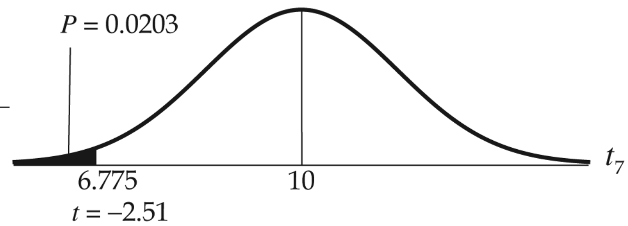
We will perform a one-sample t-test.

Since the P-value = 0.0203 is low, we reject the null hypothesis. There is evidence to suggest the mean percentage of unpopped kernels is less than 10% at this setting.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{6.775 - 10}{\frac{3.637}{\sqrt{8}}}$$

$$t \approx -2.51$$

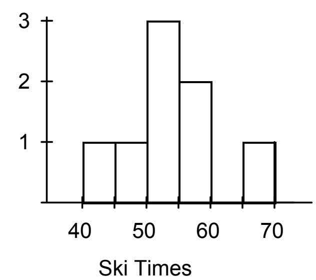


42. Ski wax.

H0: The mean time was 55 seconds. $(\mu = 55)$

HA: The mean time was less than 55 seconds. $(\mu < 55)$

Independence assumption: Since the times are not randomly selected, we will assume that the times are independent, and representative of all times.
Nearly Normal condition: The histogram of the times is unimodal and roughly symmetric.



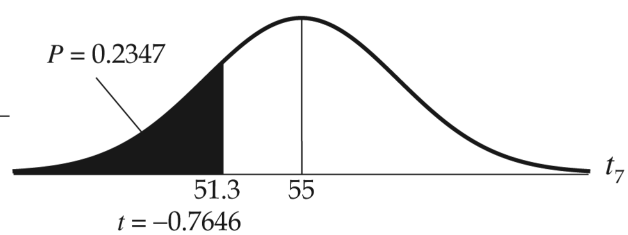
The times in the sample had a mean of 53.1 seconds and a standard deviation of 7.029 seconds. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean time with a Student's t model, with 8 – 1 = 7 degrees of freedom, $t_7\left(53.1, \frac{7.029}{\sqrt{8}}\right)$. We will perform a one-sample t-test.

Since the P-value = 0.2347 is high, we fail to reject the null hypothesis. There is no evidence to suggest the mean time is less than 55 seconds. He should not buy the new ski wax.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{53.1 - 55}{\frac{7.029}{\sqrt{8}}}$$

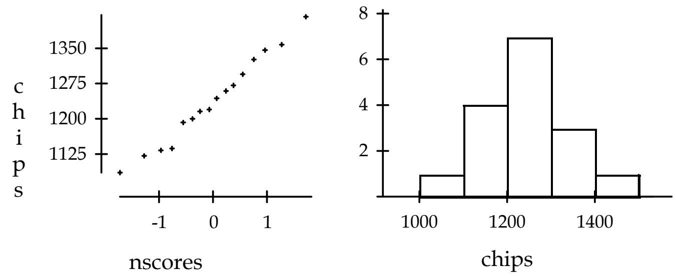
$$t \approx -0.7646$$



43. Chips ahoy.

a) Randomization condition: The bags of cookies were randomly selected.

Nearly Normal condition: The Normal probability plot is reasonably straight, and the histogram of the number of chips per bag is unimodal and symmetric.



b) The bags in the sample had a mean number of chips of 1238.19, and a standard deviation of 94.282 chips. Since the conditions for inference have been satisfied, use a one-sample *t*-interval, with 16 – 1 = 15 degrees of freedom, at 95% confidence.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 1238.19 \pm t^*_{15}\left(\frac{94.282}{\sqrt{16}}\right) \approx (1187.9,\ 1288.4)$$

We are 95% confident that the mean number of chips in an 18-ounce bag of Chips Ahoy cookies is between 1187.9 and 1288.4.

c) H_0: The mean number of chips per bag is 1000. $(\mu = 1000)$

H_A: The mean number of chips per bag is greater than 1000. $(\mu > 1000)$

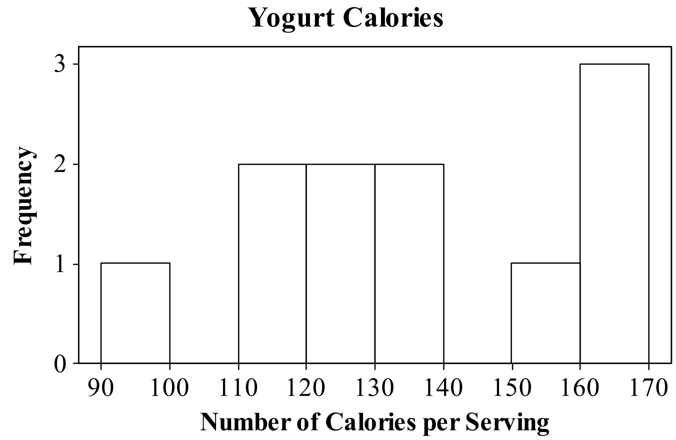
Since the confidence interval is well above 1000, there is strong evidence that the mean number of chips per bag is well above 1000.

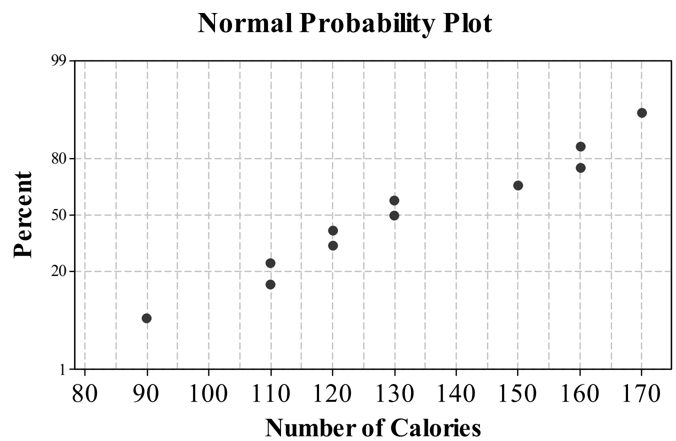
However, since the "1000 Chip Challenge" is about individual bags, not means, the claim made by Nabisco may not be true. If the mean was around 1188 chips, the low end of our confidence interval, and the standard deviation of the population was about 94 chips, our best estimate obtained from our sample, a bag containing 1000 chips would be about 2 standard deviations below the mean. This is not likely to happen, but not an outrageous occurrence. These data do not provide evidence that the "1000 Chip Challenge" is true.

44. Yogurt.

a) **Randomization condition:** The brands of vanilla yogurt may not be a random sample, but they are probably representative of all brands of yogurt. **Independence assumption:** The Randomization Condition is designed to check the reasonableness of the assumption of independence. We had some trouble verifying this condition. But is the calorie content per serving of one brand of yogurt likely to be associated with that of another brand? Probably not. We're okay.

Nearly Normal condition: The Normal probability plot is reasonably straight, and the histogram of the number of calories per serving is plausibly unimodal and symmetric.





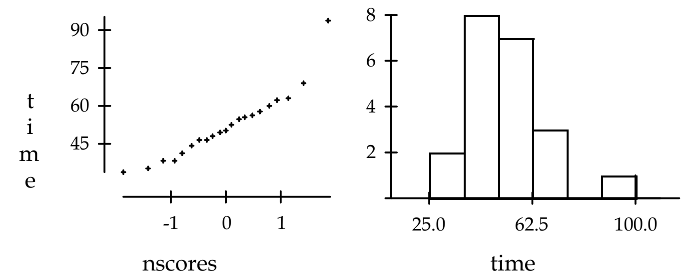
b) The brands in the sample had a mean calorie content of 131.82 calories, and a standard deviation of 25.23 calories. Since the conditions for inference have been satisfied, use a one-sample *t*-interval, with 11 – 1 = 10 degrees of freedom, at 95% confidence.

$$\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 131.82 \pm t_{10}^{*}\left(\frac{25.23}{\sqrt{11}}\right) \approx (114.87, 148.77)$$

c) We are 95% confident that the mean calorie content in a serving of vanilla yogurt is between 114.87 and 148.77 calories. The reported average of 120 calories is plausible. The 95% confidence interval contains 120 calories.

45. Maze.

a) **Independence assumption:** It is reasonable to think that the rats' times will be independent, as long as the times are for different rats.
Nearly Normal condition: There is an outlier in both the Normal probability plot and the histogram that should probably be eliminated before continuing the test. One rat took a long time to complete the maze.



b) H_0: The mean time for rats to complete this maze is 60 seconds. $(\mu = 60)$

H_A: The mean time for rats to complete this maze is not 60 seconds. $(\mu \neq 60)$

The rats in the sample finished the maze with a mean time of 52.21 seconds and a standard deviation in times of 13.5646 seconds. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean time in which rats complete the maze with a Student's t model, with

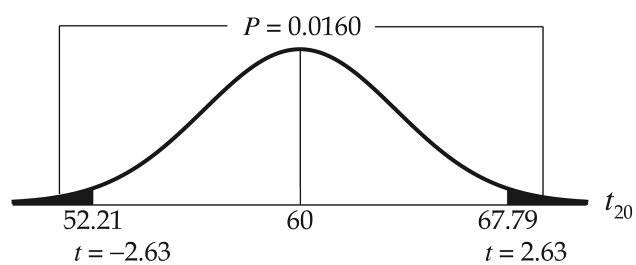
21 – 1 = 20 degrees of freedom, $t_{20}\left(60, \frac{13.5646}{\sqrt{21}}\right)$. We will perform a one-sample t-test.

Since the P-value = 0.0160 is low, we reject the null hypothesis. There is evidence that the mean time required for rats to finish the maze is not 60 seconds. Our evidence suggests the mean time is less than 60 seconds.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{52.21 - 60}{\frac{13.56}{\sqrt{21}}}$$

$$t \approx -2.63$$



c) Without the outlier, the rats in the sample finished the maze with a mean time of 50.13 seconds and standard deviation in times of 9.90 seconds. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean time in which rats complete the maze with a Student's t model, with

20 – 1 = 19 degrees of freedom, $t_{19}\left(60, \frac{9.90407}{\sqrt{20}}\right)$. We will use a one-sample t-test.

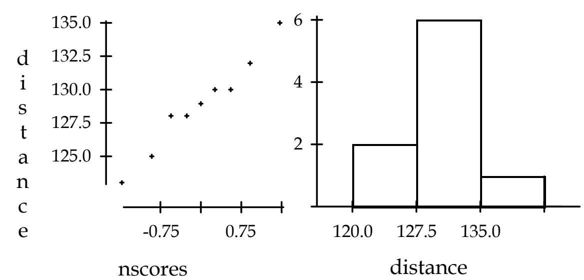
This test results in a value of $t = -4.46$, and a two-sided P-value = 0.0003. Since the P-value is low, we reject the null hypothesis. There is evidence that the mean time required for rats to finish the maze is not 60 seconds. Our evidence suggests that the mean time is actually less than 60 seconds.

d) According to both tests, there is evidence that the mean time required for rats to complete the maze is different than 60 seconds. The maze does not meet the "one-minute average" requirement. It should be noted that the test without the outlier is the appropriate test. The one slow rat made the mean time required seem much higher than it probably was.

46. Braking.

H_0: The mean braking distance is 125 feet. The tread pattern works adequately. $(\mu = 125)$

H_A: The mean braking distance is greater than 125 feet, and the new tread pattern should not be used. $(\mu > 125)$



Independence assumption: It is reasonable to think that the braking distances on the test track are independent of each other.
Nearly Normal condition: The braking distance of 102 feet is an outlier. After it is removed, the Normal probability plot is reasonably straight, and the histogram of braking distances unimodal and symmetric.

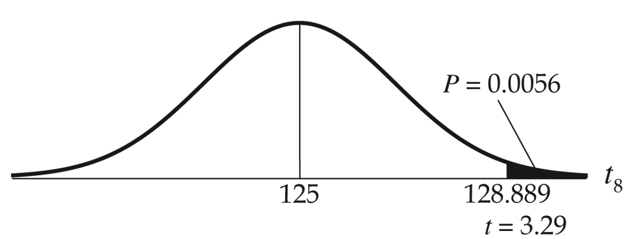
The braking distances in the sample had a mean of 128.889 feet, and a standard deviation of 3.55121 feet. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean braking distance with a Student's t model, with 9 – 1 = 8 degrees of freedom, $t_9\left(125, \frac{3.55121}{\sqrt{9}}\right)$. We will perform a one-sample t-test.

Since the P-value = 0.0056 is low, we reject the null hypothesis. There is strong evidence that the mean braking distance of cars with these tires is greater than 125 feet. The new tread pattern should not be adopted.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{128.889 - 125}{\frac{3.55121}{\sqrt{9}}}$$

$$t \approx 3.29$$



47 Golf Drives III 2013.

a) $\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 289.24 \pm t^*_{154}\left(\frac{11.18}{\sqrt{155}}\right) \approx (287.47, 291.01)$

b) These data are not a random sample of golfers. The top professionals are not representative of all golfers and were not selected at random. We might consider the 2013 data to represent the population of all professional golfers, past, present, and future.

c) The data are means for each golfer, so they are less variable than if we looked at separate drives, and inference is invalid.

48. Wind power.

a) H_0: The mean wind speed is 8 mph. It's not windy enough for a turbine. $(\mu = 8)$

H_A: The mean wind speed is greater than 8 mph. It's windy enough. $(\mu > 8)$

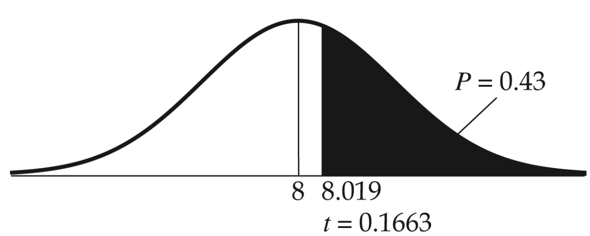
Independence assumption: The timeplot shows no pattern, so it seems reasonable that the measurements are independent.
Randomization condition: This is not a random sample, but an entire year is measured. These wind speeds should be representative of all wind speeds at this location.
10% condition: These wind speeds certainly represent fewer than 10% of all wind speeds.
Nearly Normal condition: The Normal probability plot is reasonably straight, and the histogram of the wind speeds is unimodal and reasonably symmetric.

The wind speeds in the sample had a mean of 8.091 mph, and a standard deviation of 3.813 mph. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean wind speed with a Student's *t* model, with 1114 – 1 = 1113 degrees of freedom, $t_{1113}\left(8, \frac{3.813}{\sqrt{1114}}\right)$. We will perform a one-sample *t*-test.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{8.019 - 8}{\frac{3.813}{\sqrt{1114}}}$$

$$t \approx 0.1663$$



Since the *P*-value = 0.43 is high, we fail to reject the null hypothesis. There is no evidence that the mean wind speed at this site is higher than 8 mph. Even thought the mean wind speed for these 1114 measurements is 8.019 mph, I wouldn't recommend building a wind turbine at this site.

Chapter 21 – More About Tests and Intervals

Section 21.1

1. Parameters and hypotheses.

a) Let $p =$ probability of winning on a slot machine.
$H_0: p = 0.01$ vs. $H_A: p \neq 0.01$

b) Let $\mu =$ mean spending per customer this year.
$H_0: \mu = \$35.32$ vs. $H_A: \mu \neq \$35.32$

c) Let $p =$ proportion of patients cured by the new drug.
$H_0: p = 0.3$ vs. $H_A: p \neq 0.3$

d) Let $p =$ proportion of clients now using the website.
$H_0: p = 0.4$ vs. $H_A: p \neq 0.4$

2. Hypotheses and parameters.

a) Let $p =$ proportion using seat belts in MA.
$H_0: p = 0.65$ vs. $H_A: p \neq 0.65$

b) Let $p =$ proportion of employees willing to pay for onsite day care.
$H_0: p = 0.45$ vs. $H_A: p \neq 0.45$

c) Let $p =$ probability of default for Gold card customers.
$H_0: p = 0.067$ vs. $H_A: p \neq 0.067$

d) Let $\mu =$ mean time (in months) that regular customers have been customers of the bank. $H_0: \mu = 17.3$ vs. $H_A: \mu \neq 17.3$

Section 21.2

3. P-values.

a) False. A low *P*-value provides evidence for rejecting the null hypothesis.

b) False. It results in rejecting the null hypothesis, but does not prove that it is false.

c) False. A high *P*-value shows that the data are consistent with the null hypothesis but does not prove that the null hypothesis is true.

d) False. Whether a *P*-value provides enough evidence to reject the null hypothesis depends on the risk of a type I error that one is willing to assume (the α level).

4. More P-values.

a) True.

b) False. A high *P*-value shows that the data are consistent with the null hypothesis but does not prove that the null hypothesis is true.

c) False. No *P*-value ever shows that the null hypothesis is true (or false).

d) False. If the null hypothesis is true, you will get a *P*-value below 0.01 about once in a hundred hypothesis tests.

5. Hypotheses.

a) Two-sided. Let $p =$ proportion of accounting reports that are on time.
H_0 : 40% of accounting reports were on time this year. ($p = 0.40$)
H_A : The percentage of reports submitted on time is not 40%. ($p \neq 0.40$)

b) One-sided. Let $\mu =$ mean click-through rate.
H_0 : The mean click-through rate for the new website is 5.4 minutes. ($\mu = 5.4$)
H_A : The mean click-through rate is less than 5.4 minutes. ($\mu < 5.4$)

c) One-sided. Let $p =$ percentage of employees enrolled in at least one wellness class.
H_0 : 42% of employees are enrolled in at least one wellness class. ($p = 0.42$)
H_A : The percentage of employees enrolled in at least one wellness class is greater than 42%. ($p > 0.42$)

d) One-sided. Let $p =$ percentage of voters who will vote for the candidate.
H_0 : 50% of voters will vote for the candidate. ($p = 0.50$)
H_A : More than 50% of voters will vote for the candidate. ($p > 0.50$)

6. More hypotheses.

a) One-sided. Let $p =$ percentage of warranty problems.
H_0 : 20% of computers have warranty problems. ($p = 0.20$)
H_A : The percentage of computers with warranty problems is less than 20%. ($p < 0.20$)

b) One-sided. Let $p =$ percentage of responses to a donation request.
H_0 : 4.75% of people will respond to a donation request. ($p = 0.0475$)
H_A : The percentage of people who will respond to a donation request is greater than 4.75%. ($p > 0.0475$)

c) One-sided. Let $\mu =$ average age of a website user.
H_0 : The average age of a website user is 35.2 years. ($\mu = 35.2$)
H_A : The average age of a website user is less than 35.2 years. ($\mu < 35.2$)

d) Two-sided. Let $p =$ proportion of students who prefer Coke to Pepsi.
H_0: The proportion of students who prefer Coke to Pepsi is 50%. ($p = 0.50$)
H_A: The proportion of students who prefer Coke to Pepsi is not 50%. ($p \neq 0.50$)

Section 21.3

7. Alpha true and false.

a) True.

b) False. The alpha level is set independently and does not depend on the sample size.

c) False. The *P*-value would have to be less than 0.01 to reject the null hypothesis.

d) False. It simply means we do not have enough evidence at that alpha level to reject the null hypothesis.

8. Alpha false and true.

a) False. A lower alpha level lowers the probability of a Type I error, but increases the probability of a Type II error.

b) True.

c) True.

d) True.

Section 21.4

9. Critical values.

a) $z = \pm 1.96$

b) $z = 1.645$

c) $t = \pm 2.03$

d) $z = 2.33$ (n is not relevant for critical values of z)

e) $z = -2.33$

10. More critical values.

a) $t = \pm 2.00$

b) $z = 1.645$

c) $z = \pm 2.58$

d) $z = -2.33$ (n is not relevant for critical values of z)

e) $z = -2.33$

Section 21.5

11. Errors.

a) Type I error. The actual value is not greater than 0.3 but they rejected the null hypothesis.

b) No error. The actual value is 0.50, which was not rejected.

c) Type II error. The actual value was 55.3 points, which is greater than 52.5.

d) Type II error. The null hypothesis was not rejected, but it was false. The true relief rate was greater than 0.25.

12. More errors.

a) Type I error. The actual mean was not greater than 25.

b) No error. The actual proportion is greater than 0.80 so they were correct in not rejecting the null hypothesis.

c) No error. The actual proportion is not equal to 0.5.

d) Type II error. They should have rejected the null hypothesis since 0.60 is less than 0.70.

Chapter Exercises

13. One sided or two?

a) Two sided. Let p be the percentage of students who prefer Diet Coke.

H_0 : 50% of students prefer Diet Coke. ($p = 0.50$)
H_A : The percentage of students who prefer Diet Coke is not 50%. ($p \neq 0.50$)

b) One sided. Let p be the percentage of teenagers who prefer the new formulation.

H_0 : 50% of students prefer the new formulation. ($p = 0.50$)
H_A : More than 50%of students prefer the new formulation. ($p > 0.50$)

c) One sided. Let p be the percentage of people who plan to vote for the override.

H_0 : 2/3 of the residents intend to vote for the override. ($p = 2/3$)
H_A : More than 2/3 of the residents intend to vote for the override. ($p > 2/3$)

d) Two sided. Let p be the percentage of days the market goes up.

H_0 : The market goes up on 50% of days. ($p = 0.50$)
H_A : The percentage of days the market goes up is not 50%. ($p \neq 0.50$)

14. Which alternative?

a) Two sided. Let p be the percentage of students who prefer plastic.

H_0 : 50% of students prefer plastic. ($p = 0.50$)
H_A : The percentage of students who prefer plastic is not 50%. ($p \neq 0.50$)

b) Two sided. Let p be the percentage of juniors planning to study abroad.

H_0 : 10% of juniors plan to study abroad. ($p = 0.10$)
H_A : The percentage of juniors plan to study abroad is not 10%. ($p \neq 0.10$)

c) One sided. Let p be the percentage of people who experience relief.

H_0 : 22% of people experience headache relief with the drug. ($p = 0.22$)
H_A : More than 22% of people experience headache relief with the drug. ($p > 0.22$)

d) One sided. Let p be the percentage of hard drives that pass all performance tests.

H_0 : 60% of hard drives pass all performance tests. ($p = 0.60$)
H_A : The percentage of drives is greater than 60%. ($p > 0.60$)

15. P-value.

If the effectiveness of the new poison ivy treatment is the same as the effectiveness of the old treatment, the chance of observing an effectiveness this large or larger in a sample of the same size is 4.7% by natural sampling variation alone.

16. Another P-value.

If the rate of seat belt usage after the campaign is the same as the rate of seat belt usage before the campaign, there is a 17% chance of observing a rate of seat belt usage after the campaign this large or larger in a sample of the same size by natural sampling variation alone.

17. Alpha.

Since the null hypothesis was rejected at $\alpha = 0.05$, the P-value for the researcher's test must have been less than 0.05. He would have made the same decision at $\alpha = 0.10$, since the P-value must also be less than 0.10. We can't be certain whether or not he would have made the same decision at $\alpha = 0.01$, since we only know that the P-value was less than 0.05. It may have been less than 0.01, but we can't be sure.

18. Alpha again.

Since the environmentalists failed to reject the null hypothesis at $\alpha = 0.05$, the P-value for the environmentalists' test must have been greater than 0.05. We can't be certain whether or not they would have made the same decision at $\alpha = 0.10$, since we only know that the P-value was greater than 0.05. It may have been greater than 0.10 as well, but we can't be sure. They would have made them same decision at $\alpha = 0.01$, since the P-value must also be greater than 0.01.

19. Significant?

a) If 98% of children have really been vaccinated, there is practically no chance of observing 97.4% of children (in a sample of 13,000) vaccinated by natural sampling variation alone.

b) We conclude that the proportion of children who have been vaccinated is below 98%, but a 95% confidence interval would show that the true proportion is between 97.1% and 97.7%. Most likely a decrease from 98% to 97.7% would not be considered important. The 98% figure was probably an approximate figure anyway. However, if the 98% figure was not as estimate, and with 1,000,000 kids per year vaccinated, even 0.1% represents 1,000 kids, so this may be important.

20. Significant again?

a) If 15.9% is the true percentage of children who did not attain the grade level standard, there is only a 2.3% chance of observing 15.1% of children (in a sample of 8500) not attaining grade level by natural sampling variation alone.

b) Under old methods, 1352 students would not be expected to read at grade level. With the new program, 1284 would not be expected to read at grade level. This is only a decrease of 68 students. The costs of switching to the new program might outweigh the potential benefit. It is also important to realize that this is only a *potential* benefit.

21. Groceries.

a) **Randomization condition:** We will assume that the Yahoo survey was conducted randomly.
10% condition: 2400 is less than 10% of all men.
Success/Failure condition: $n\hat{p} = 1224$ and $n\hat{q} = 1176$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of men who identify themselves as the primary grocery shopper in their household.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{1224}{2400}\right) \pm 2.326\sqrt{\frac{\left(\frac{1224}{2400}\right)\left(\frac{1176}{2400}\right)}{2400}} = (48.6\%, 53.4\%)$$

We are 98% confident that between 48.6% and 53.4% of all men identify themselves as the primary grocery shopper in their household.

b) Since 45% is not in the interval, there is strong evidence that more than 45% of all men identify themselves as the primary grocery shopper in their household.

c) The significance level of this test is $\alpha = 0.01$. It's an upper tail test based on a 98% confidence interval.

22. Is the Euro fair?

a) Independence assumption: The Euro spins are independent. One spin is not going to affect the others. (With true independence, it doesn't make sense to try to check the randomization condition and the 10% condtion. These verify our assumption of independence, and we don't need to do that!)

Success/Failure condition: $n\hat{p} = 140$ and $n\hat{q} = 110$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the proportion of heads in Euro spins.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{140}{250}\right) \pm 1.960 \sqrt{\frac{\left(\frac{140}{250}\right)\left(\frac{110}{250}\right)}{250}} = (0.498, 0.622)$$

We are 95% confident that the true proportion of heads when a Euro is spun is between 0.498 and 0.622.

b) Since 0.50 is within the interval, there is no evidence that the coin in unfair. 50% is a plausible value for the true proportion of heads. (That having been said, I'd want to spin this coin a few hundred more times. It's close!)

c) The significance level is $\alpha = 0.05$. It's a two-tail test based on a 95% confidence interval.

23. Approval 2014.

a) Randomization condition: The adults were randomly selected.
10% condition: 1500 adults represent less than 10% of all adults.
Success/Failure condition: $n\hat{p}$ = (1500)(0.40) = 600 and $n\hat{q}$ = (1500)(0.60) = 900 are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate Barack Obama's approval rating.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.40) \pm 1.960 \sqrt{\frac{(0.40)(0.60)}{1500}} = (0.375, 0.425)$$

We are 95% confident that Barack Obama's approval rating is between 37.5% and 42.5%.

b) Since 45% is not within the interval, this is not a plausible value for Barack Obama's approval rating. There is evidence against the null hypothesis.

24. Hard times.

a) Randomization condition: The men were contacted through a random poll.
10% condition: 800 men represent less than 10% of all men.

Success/Failure condition: $n\hat{p} = (800)(0.09) = 72$ and $n\hat{q} = (800)(0.91) = 728$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of men who have taken a second job to help pay the bills.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.09) \pm 1.96\sqrt{\frac{(0.09)(0.91)}{800}} = (7.0\%, 11.0\%)$$

We are 95% confident that between 7.0% and 11.0% of all men have taken a second job to help pay the bills.

b) Since 6% is not in the interval, the pundit's claim is not plausible.

25. Dogs.

a) We cannot construct a confidence interval for the rate of occurrence of early hip dysplasia among 6-month old puppies because only 5 of 42 puppies were found with early hip dysplasia. The Success/Failure condition is not satisfied.

b) **Independence assumption:** If hip dysplasia is hereditary, and puppies brought to the vaccination clinic are from the same litter, independence might be an issue.
Randomization condition: The veterinarian considers the 42 puppies to be a random sample of all puppies.
10% condition: 42 puppies represent less than 10% of all puppies.
Success/Failure condition: As previously mentioned, this condition is not met, since there aren't at least 10 puppies with hip dysplasia in the sample.

Since the other conditions are met, we can construct a one-proportion plus-four z-interval to estimate the percentage of puppies with early hip dysplasia.

$$\tilde{p} = \frac{y+2}{n+4} = \frac{7}{46} = 0.152$$

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}} = (0.152) \pm 1.96\sqrt{\frac{(0.152)(0.848)}{46}} = (4.8\%, 25.6\%)$$

We are 95% confident that between 4.8% and 25.6% of puppies have early hip dysplasia.

26. Fans.

a) We cannot construct a confidence interval for the percentage of home team fans entering the stadium, since only 9 people were not home fans. The Success/Failure condition is not satisfied.

b) **Randomization condition:** The people standing in line were randomly selected.
10% condition: 81 people are likely to be less than 10% of the fans in the stadium.
Success/Failure condition: As previously mentioned, this condition is not met, since there aren't at least 10 non-home fans in the sample.

Since the other conditions are met, we can construct a one-proportion plus-four z-interval to estimate the percentage of home fans at the game.

$$\tilde{p} = \frac{y+2}{n+4} = \frac{75}{85} = 0.882$$

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}} = (0.882) \pm 1.96\sqrt{\frac{(0.882)(0.118)}{85}} = (81.4\%, 95.1\%)$$

We are 95% confident that between 81.4% and 95.1% of all people are home fans.

27. Loans.

a) The bank has made a Type II error. The person was not a good credit risk, and the bank failed to notice this.

b) The bank has made a Type I error. The person was a good credit risk, and the bank was convinced that he/she was not.

c) By making it easier to get a loan, the bank has reduced the alpha level. It takes less evidence to grant the person the loan.

d) The risk of Type I error is decreased and the risk of Type II error has increased.

28. Spam.

a) Type II. The filter decided that the message was safe, when in fact it was spam.

b) Type I. The filter decided that the message was spam, when in fact it was not.

c) This is analogous to lowering alpha. It takes more evidence to classify a message as spam.

d) The risk of Type I error is decreased and the risk of Type II error has increased.

29. Second loan.

a) Power is the probability that the bank denies a loan that could not have been repaid.

b) To increase power, the bank could raise the cutoff score.

c) If the bank raised the cutoff score, a larger number of trustworthy people would be denied credit, and the bank would lose the opportunity to collect the interest on these loans.

30. More spam.

a) The power of the test is the ability of the filter to detect spam.

b) To increase the filter's power, lower the cutoff score.

c) If the cutoff score is lowered, a larger number of real messages would end up in the junk mailbox.

31. Homeowners 2013.

a) The null hypothesis is that the level of home ownership does not rise. The alternative hypothesis is that it rises.

b) In this context, a Type I error is when the city concludes that home ownership is on the rise, but in fact, the tax breaks don't help.

c) In this context, a Type II error is when the city abandons the tax breaks, thinking they don't help, when in fact they were helping.

d) A Type I error causes the city to forego tax revenue, while a Type II error withdraws help from those who might have otherwise been able to buy a house.

e) The power of the test is the city's ability to detect an actual increase in home ownership.

32. Alzheimer's.

a) The null hypothesis is that a person is healthy. The alternative is that they have Alzheimer's disease. There is no parameter of interest here.

b) A Type I error is a false positive. It has been decided that the person has Alzheimer's disease when they don't.

c) A Type II error is a false negative. It has been decided that the person is healthy, when they actually have Alzheimer's disease.

d) A Type I error would require more testing, resulting in time and money lost. A Type II error would mean that the person did not receive the treatment they needed. A Type II error is much worse.

e) The power of this test is the ability of the test to detect patients with Alzheimer's disease. In this case, the power can be computed as $1 - P(\text{Type II error}) = 1 - 0.08 = 0.92$.

33. Testing cars.

H_0 : The shop is meeting the emissions standards.
H_A : The shop is not meeting the emissions standards.

a) Type I error is when the regulators decide that the shop is not meeting standards when they actually are meeting the standards.

b) Type II error is when the regulators certify the shop when they are not meeting the standards.

c) Type I would be more serious to the shop owners. They would lose their certification, even though they are meeting the standards.

d) Type II would be more serious to environmentalists. Shops are allowed to operate, even though they are allowing polluting cars to operate.

34. Quality control.

H_0 : The assembly process is working fine.
H_A : The assembly process is producing defective items.

a) Type I error is when the production managers decide that there has been an increase in the number of defective items and stop the assembly line, when the assembly process is working fine.

b) Type II error is when the production managers decide that the assembly process is working fine, but defective items are being produced.

c) The factory owner would probably consider Type II error to be more serious, depending of the costs of shutting the line down. Generally, because of warranty costs and lost customer loyalty, defects that are caught in the factory are much cheaper to fix than defects found after items are sold.

d) Customers would consider Type II error to be more serious, since customers don't want to buy defective items.

35. Cars again.

a) The power of the test is the probability of detecting that the shop is not meeting standards when they are not.

b) The power of the test will be greater when 40 cars are tested. A larger sample size increases the power of the test.

c) The power of the test will be greater when the level of significance is 10%. There is a greater chance that the null hypothesis will be rejected.

d) The power of the test will be greater when the shop is out of compliance "a lot". Larger problems are easier to detect.

36. Production.

a) The power of the test is the probability that the assembly process is stopped when defective items are being produced.

b) An advantage of testing more items is an increase in the power of the test to detect a problem. The disadvantages of testing more items are the additional cost and time spent testing.

c) An advantage of lowering the alpha level is that the probability of stopping the assembly process when everything is working fine (committing a Type I error) is decreased. A disadvantage is that the power of the test to detect defective items is also decreased.

d) The power of the test will increase as a day passes. Bigger problems are easier to detect.

37. Equal opportunity?

H_0 : The company is not discriminating against minorities.
H_A : The company is discriminating against minorities.

a) This is a one-tailed test. They wouldn't sue if "too many" minorities were hired.

b) Type I error would be deciding that the company is discriminating against minorities when they are not discriminating.

c) Type II error would be deciding that the company is not discriminating against minorities when they actually are discriminating.

d) The power of the test is the probability that discrimination is detected when it is actually occurring.

e) The power of the test will increase when the level of significance is increased from 0.01 to 0.05.

f) The power of the test is lower when the lawsuit is based on 37 employees instead of 87. Lower sample size leads to less power.

38. Stop signs.

H_0 : The new signs provide the same visibility than the old signs.
H_A : The new signs provide greater visibility than the old signs.

a) The test is one-tailed, because we are only interested in whether or not the signs are more visible. If the new design is less visible, we don't care how much less visible it is.

b) Type I error happens when the engineers decide that the new signs are more visible when they are not more visible.

c) Type II error happens when the engineers decide that the new signs are not more visible when they actually are more visible.

d) The power of the test is the probability that the engineers detect a sign that is truly more visible.

e) When the level of significance is dropped from 5% to 1%, power decreases. The null hypothesis is harder to reject, since more evidence is required.

f) If a sample of size 20 is used instead of 50, power will decrease. A smaller sample size has more variability, lowering the ability of the test to detect falsehoods.

39. Software for learning.

a) The test is one-tailed. We are testing to see if an increase in average score is associated with the software.

b) H_0 : The average score does not change following the use of software. ($\mu = 105$)
H_A : The average score increases following the use of the software. ($\mu > 105$)

c) The professor makes a Type I error if he buys the software when the average score has not actually increased.

d) The professor makes a Type II error if he doesn't buy the software when the average has actually increased.

e) The power of the test is the probability of buying the software when the average score has actually increased.

40. Ads.

a) H_0 : The percentage of residents that have heard the ad and recognize the product is 20%. (p = 0.20)
H_A : The percentage of residents that have heard the ad and recognize the product is greater than 20%. (p > 0.20)

b) The company wants more evidence that the ad is effective before deciding it really is. By lowering the level of significance from 10% to 5%, the probability of Type I error is decreased. The company is less likely to think that the ad is effective when it actually is not effective.

c) The power of the test is the probability of correctly deciding more than 20% have heard the ad and recognize the product when it's true.

d) The power of the test will be higher for a level of significance of 10%. There is a greater chance of rejecting the null hypothesis.

e) Increasing the sample size to 600 will lower the risk of Type II error. A larger sample size decreases variability, which helps us notice what is really going on. The company will be more likely to notice when the ad really works.

41. Software, part II.

a) H_0 : The average score does not change following the use of software. (μ = 105)
H_A : The average score increases following the use of the software. (μ > 105)

Randomization condition: This year's class of 203 students is probably representative of all stats students.
Nearly Normal condition: We don't have the scores from the 203 individuals, so we can't check a plot of the data. However, with a sample this large, the Central Limit Theorem allows us to model the sampling distribution of the means with a t-distribution.

The mean score was 108 points, with a standard deviation of 8.7 points. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean score with a Student's t model, with 203 – 1 = 202 degrees of freedom, $t_{202}\left(108, \frac{8.7}{\sqrt{203}}\right)$. We will perform a one-sample t-test.

The value of t is approximately 4.91, which results in a P-value of less than 0.0001, so we reject the null hypothesis. There is strong evidence that the mean score has increased since use of the software program was implemented. As long as the professor feels confident that this class of stats students is representative of all potential students, then he should buy the program.

$$t = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{108 - 105}{\frac{8.7}{\sqrt{203}}}$$

$$t = 4.91$$

If you used a 95% confidence interval to assess the effectiveness of the program: $\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 108 \pm t^*_{202}\left(\frac{8.7}{\sqrt{203}}\right) \approx (106.8,\ 109.2)$

We are 95% confident that the mean score is between 106.8 and 109.2. Since 105 is above the interval, this provides evidence that the mean score has increased following the implementation of the software program.

b) The mean score on the exam only increased by 1 to 4 points. This small difference might not be enough to be worth the cost of the program.

42. Testing the ads.

a) H_0 : The percentage of residents that remember the ad is 20%. (p = 0.20)
H_A : The percentage of residents that remember is greater than 20%. (p > 0.20)

Independence assumption: It is reasonable to think that randomly selected residents would remember the ad independently of one another.
Randomization condition: The sample consisted of 600 randomly selected residents.
10% condition: The sample of 600 is less than 10% of the population of the city.
Success/Failure condition: $np = (600)(0.20) = 120$ and $nq = (600)(0.80) = 480$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p =$ 0.20 and

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.20)(0.80)}{600}} \approx 0.0163.$$

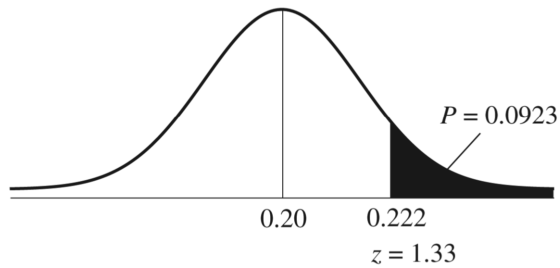
We can perform a one-proportion z-test. The observed proportion of residents who remembered the ad is $\hat{p} = \frac{133}{600} \approx 0.222$.

Since the P-value = 0.0923 is somewhat high, we fail to reject the null hypothesis. There is little evidence that more than 20% of people remember the ad.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.222 - 0.20}{\sqrt{\frac{(0.20)(0.80)}{600}}}$$

$$z \approx 1.33$$



The company should not renew the contract.

b) There is a 9.23% chance of having 133 or fewer of 600 people in a random sample remember the ad, if in fact only 20% of people in the population do.

43. TV safety.

a) This is an upper-tail test. We hope to show that the TV stand will hold 500 or more pounds easily.

b) The inspectors will commit a Type I error if they decide the TV stands are safe when they are not.

c) The inspectors will commit a Type II error if they decide the TVs are unsafe when they are actually safe.

44. Catheters.

a) This is a two-sided test. If catheters are too big, they won't fit through the vein. If the catheters are too small, they might not have enough structural integrity to work well.

b) The quality control people will commit a Type I error if they decide that the catheters are not the correct size when they actually have the correct diameter. The manufacturing process is stopped needlessly.

c) The quality control people will commit a Type II error if they decide that the catheters have the correct diameter when they actually are too small or too large. Catheters that do not meet specifications are produced and sold, possibly injuring patients, or simply not working properly.

45. TV safety, revisited.

a) To decrease the likelihood of producing an unsafe TV stand, they should decrease α. This lower the chance of making a Type I error.

b) The power of the test is the ability of the inspectors to determine that the TV stand is safe when it is actually capable of holding more than 500 pounds.

c) The company could increase the power of the test by lowering the standard deviation by testing more stands. This could prove costly and would require more time to test. They could also increase α, but then they will commit Type I errors with greater frequency, approving stands that cannot hold 500 pounds of more. Finally, they could require that TV stands have a higher weight capacity than 500 pounds as the standard. Again, that might prove costly, since they would be rejecting many more stands that were safe.

46. Catheters, again.

a) If the significance level were changed to $\alpha = 0.01$, this would increase the probability of Type II error. Requiring a higher standard of proof would mean that more catheters would be rejected, even when they met the diameter specification.

b) The power of the test is the probability of detecting a catheter than does not meet the 2.00 mm specification.

c) As the diameters got farther and farther away from 2.00 mm, the power would increase. It would become easier to detect that the catheters did not meet the diameter specification when the were much too big or small.

d) To increase the power of the test, they could sample more catheters or increase the significance level.

47. Two coins.

a) The alternative hypothesis is that your coin produces 30% heads.

b) Reject the null hypothesis if the coin comes up tails. Otherwise, fail to reject.

c) There is a 10% chance that the coin comes up tails if the null hypothesis is true, so alpha is 10%.

d) Power is our ability to detect the 30% coin. That coin will come up tails 70% of the time. That's the power of our test.

e) To increase the power and lower the probability of Type I error at the same time, simply flip the coin more times.

48. Faulty or not?

a) The null hypothesis is that the drive is good. The alternative hypothesis is that the drive is bad.

b) Reject the null hypothesis if the computer fails the test. Otherwise, fail to reject.

c) There is a 4% chance that the computer fails the test, even if the drive is good, so alpha is 4%.

d) Power is the ability to detect faulty drives. Faulty drives fail the test 65% of the time. That's the power of our test.

49. Hoops.

H_0 : The player's foul-shot percentage is only 60%. (p = 0.60)
H_A : The player's foul-shot percentage is better than 60%. (p > 0.60)

a) The player's shots can be considered Bernoulli trials. There are only two possible outcomes, make the shot and miss the shot. The probability of making any shot is constant at p = 0.60. Assume that the shots are independent of each other. Use *Binom*(10, 0.60).

Let X = the number of shots made out of n = 10.

$$\begin{aligned} P(\text{makes at least 9 out of 10}) &= P(X \geq 9) \\ &= P(X = 9) + P(X = 10) \\ &= {}_{10}C_9(0.60)^9(0.40)^1 + {}_{10}C_{10}(0.60)^{10}(0.40)^0 \\ &\approx 0.0464 \end{aligned}$$

b) The coach made a Type I error.

c) The power of the test can be calculated for specific values of the new probability of success. Each true value of p has a power calculation associated with it. In this case, we are finding the power of the test to detect an 80% foul-shooter. Use *Binom*(10, 0.80).

Let X = the number of shots made out of n = 10.

$$\begin{aligned} P(\text{makes at least 9 out of 10}) &= P(X \geq 9) \\ &= P(X = 9) + P(X = 10) \\ &= {}_{10}C_9(0.80)^9(0.20)^1 + {}_{10}C_{10}(0.80)^{10}(0.20)^0 \approx 0.376 \end{aligned}$$

The power of the test to detect an increase in foul-shot percentage from 60% to 80% is about 37.6%.

d) The power of the test to detect improvement in foul-shooting can be increased by increasing the number of shots, or by keeping the number of shots at 10 but increasing the level of significance by declaring that 8, 9, or 10 shots made will convince the coach that the player has improved. In other words, the coach can increase the power of the test by lowering the standard of proof.

50. Pottery.

H_0 : The new clay is no better than the old, and breaks 40% of the time. (p = 0.40)
H_A : The new clay breaks less than 40% of the time. (p < 0.40)

a) The fired pieces can be considered Bernoulli trials. There are only two possible outcomes, broken and unbroken. The probability of breaking is constant at p = 0.40. It is reasonable to think that the pieces break independently of each other. Use *Binom*(10, 0.40).

Let X = the number of broken pieces out of n = 10.

$$\begin{aligned} P(\text{at most one breaks}) &= P(X \leq 1) \\ &= P(X=0) + P(X=1) \\ &= {}_{10}C_0(0.40)^0(0.60)^{10} + {}_{10}C_1(0.40)^1(0.60)^9 \approx 0.0464 \end{aligned}$$

b) The artist made a Type I error.

c) The probability Type II error can be calculated for specific values of the new probability of success. Each true value of p has a Type II error calculation associated with it. In this case, we are finding the probability of Type II error if the pieces break only 20% of the time instead of 40% of the time. She won't notice that the clay is better if 2 or more pieces break. Use *Binom*(10, 0.80).

Let X = the number of broken pieces out of n = 10.

$$\begin{aligned} P(\text{at least 2 break}) &= P(X \geq 2) \\ &= P(X=2) + \cdots + P(X=10) \\ &= {}_{10}C_2(0.20)^2(0.80)^8 + \cdots + {}_{10}C_{10}(0.20)^{10}(0.80)^0 \\ &\approx 0.6242 \end{aligned}$$

The probability that she makes a Type II error (not noticing that the clay is better) is approximately 0.6242.

d) The power of the test to detect improvement in the clay can be increased by increasing the number of pieces fired, or by keeping the number of pieces at 10 but increasing the level of significance by declaring that 0, 1, or 2 broken pieces will convince the artist that the player has improved. In other words, the artist can improve the power by lowering her standard of proof.

Review of Part V – From the Data at Hand to the World at Large

1. Crohn's disease.

H_0: The relapse rate for those using Omega-3 fatty acids is the same as the relapse rate for those who are not. $\left(p_{Omega-3} = p_{Not} \text{ or } p_{Omega-3} - p_{Not} = 0\right)$

H_A: The relapse rate for those using Omega-3 fatty acids is lower than the relapse rate for those who are not. $\left(p_{Omega-3} < p_{Not} \text{ or } p_{Omega-3} - p_{Not} < 0\right)$

2. Color-blind.

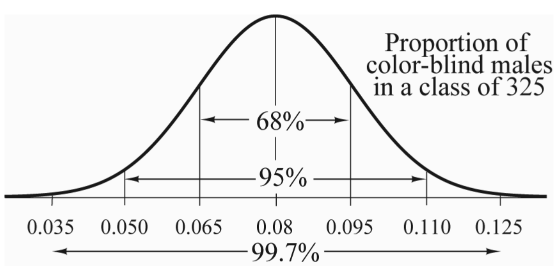
a) **Randomization condition:** The 325 male students are probably representative of all males.
10% condition: 325 male students are less than 10% of the population of males.
Success/Failure condition: np = (325)(0.08) = 26 and nq = (325)(0.92) = 299 are both greater than 10, so the sample is large enough.

Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion of colorblind men among 325 students.

b) $\mu_{\hat{p}} = p = 0.08$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.08)(0.92)}{325}} \approx 0.015$$

c)



d) According to the Normal model, we expect about 68% of classes with 325 males to have between 6.5% and 9.5% colorblind males. We expect about 95% of such classes to have between 5% and 11% colorblind males. About 99.7% of such classes are expected to have between 3.5% and 12.5% colorblind males.

3. Hamsters.

a) **Randomization condition:** Assume these litters are representative of all litters.
Nearly Normal condition: We don't have the actual data, so we can't look at a graphical display. However, since the sample size is large, the Central Limit Theorem guarantees that the distribution of averages will be approximately Normal, as long as there are no outliers.

The litters in the sample had a mean size of 7.72 baby hamsters and a standard deviation of 2.5 baby hamsters. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's t- model, with 47 – 1 = 46 degrees of freedom. We will use a one-sample t-interval with 90% confidence for the mean number of baby hamsters per litter.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 7.72 \pm t^*_{46}\left(\frac{2.5}{\sqrt{47}}\right) \approx (7.11,\ 8.33)$$

We are 90% confident that the mean number of baby hamsters per litter is between 7.11 and 8.33.

b) A 98% confidence interval would have a larger margin of error. Higher levels of confidence come at the price of less precision in the estimate.

c) A quick estimate using z gives us a sample size of about 25 litters. Using this estimate, $t^*_{24} = 2.064$ at 95% confidence. We need a sample of about 27 litters in order to estimate the number of baby hamsters per litter to within 1 baby hamster.

$$ME = t^*_{24}\left(\frac{s}{\sqrt{n}}\right)$$

$$1 = 2.064\left(\frac{2.5}{\sqrt{n}}\right)$$

$$n = \frac{(2.064)^2(2.5)^2}{(1)^2}$$

$$n \approx 27$$

4. Polling 2004.

a) No, the number of votes would not always be the same. We expect a certain amount of variability when sampling.

b) This is NOT a problem about confidence intervals. We already know the true proportion of voters who voted for Bush. This problem deals with the sampling distribution of that proportion.

We would expect 95% of our sample proportions of Bush voters to be within 1.960 standard deviations of the true proportion of Bush voters, 50.7%.

$$\sigma(\hat{p}_B) = \sqrt{\frac{p_B q_B}{n}} = \sqrt{\frac{(0.507)(0.493)}{1000}} \approx 1.58\%$$

So, we expect 95% of our sample proportions to be within 1.960(1.58%) = 3.1% of 40.7%, or between 47.6% and 53.8%.

c) Since we only expect 0.004(1000) = 4 votes for Ralph Nader, we cannot represent the sampling model with a Normal model. The Success/Failure condition is not met.

d) The sample proportion of Nader voters is expected to vary less than the sample proportion of Bush voters. Proportions farther away from 50% have smaller standard errors. (Look at the standard deviations calculated for parts b and c.)

5. Leaky gas tanks.

a) H_0: The proportion of leaky gas tanks is 40%. ($p = 0.40$)
H_A: The proportion of leaky gas tanks is less than 40%. ($p < 0.40$)

b) **Randomization condition:** A random sample of 27 service stations in California was taken.
10% condition: 27 stations are less than 10% of all service stations in California.
Success/Failure condition: $np = (27)(0.40) = 10.8$ and $nq = (27)(0.60) = 16.2$ are both greater than 10, so the sample is large enough.

c) Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.40$ and

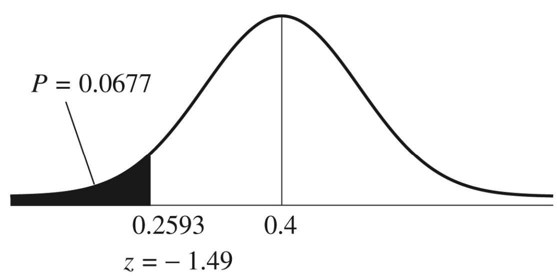
$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.40)(0.60)}{27}} \approx 0.09428$. We can perform a one-proportion z-test.

The observed proportion of leaky gas tanks is $\hat{p} = \dfrac{7}{27} \approx 0.2593$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{pq}{n}}}$$

$$z = \frac{0.2593 - 0.40}{\sqrt{\dfrac{(0.40)(0.60)}{27}}}$$

$$z \approx -1.49$$



d) Since the P-value = 0.0677 is relatively high, we fail to reject the null hypothesis. There is little evidence that the proportion of leaky gas tanks is less than 40%. The new program doesn't appear to be effective in decreasing the proportion of leaky gas tanks.

e) If the program actually works, we haven't done anything *wrong*. Our methods are correct. Statistically speaking, we have committed a Type II error.

f) In order to decrease the probability of making this type of error, we could lower our standards of proof, by raising the level of significance. This will increase the power of the test to detect a decrease in the proportion of leaky gas tanks. Another way to decrease the probability that we make a Type II error is to sample more service stations. This will decrease the variation in the sample proportion, making our results more reliable.

g) Increasing the level of significance is advantageous, since it decreases the probability of making a Type II error, and increases the power of the test. However, it also increases the probability that a Type I error is made, in this case, thinking that the program is effective when it really is not effective. Increasing the sample size decreases the probability of making a Type II error and increases power, but can be costly and time-consuming.

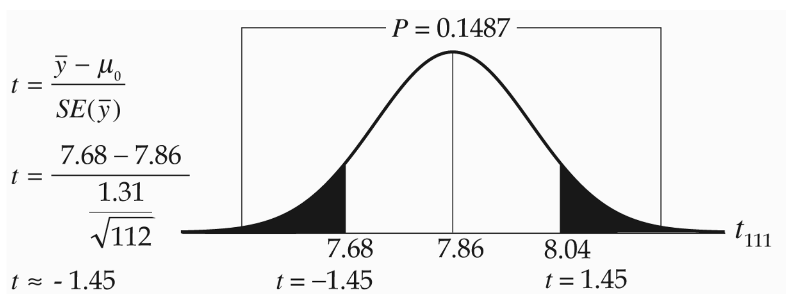
6. Babies.

H_0: The mean weight of newborns in the U.S. is 7.86 pounds, the same as the mean weight of Australian babies. $(\mu = 7.86)$

H_A: The mean weight of newborns in the U.S. is not the same as the mean weight of Australian babies. $(\mu \neq 7.86)$

Randomization condition: Assume that the babies at this Missouri hospital are representative of all U.S. newborns. (Given)
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The babies in the sample had a mean weight of 7.68 pounds and a standard deviation in weight of 1.31 pounds. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean weight of U.S. newborns with a Student's t model, with 112 – 1 = 111 degrees of freedom, $t_{111}\left(7.86, \frac{1.31}{\sqrt{112}}\right)$. We will perform a one-sample t-test.



Since the P-value = 0.1487 is high, we fail to reject the null hypothesis. If we believe that the babies at this Missouri hospital are representative of all U.S. babies, there is little evidence to suggest that the mean weight of U.S. babies is different than the mean weight of Australian babies.

7. Scrabble.

a) The researcher believes that the true proportion of As is within 10% of the estimated 54%, namely, between 44% and 64%.

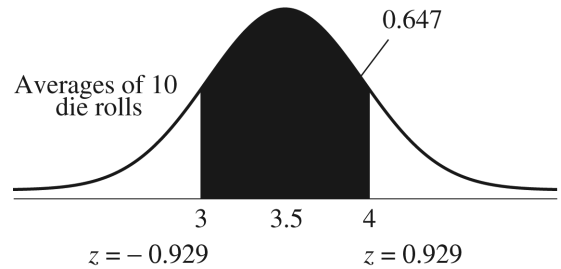
b) A large margin of error is usually associated with a small sample, but the sample consisted of "many" hands. The margin of error is large because the standard error of the sample is large. This occurs because the true proportion of As in a hand is close to 50%, the most difficult proportion to predict.

c) This provides no evidence that the simulation is faulty. The true proportion of As is contained in the confidence interval. The researcher's results are consistent with 63% As.

8. Dice.

Die rolls are truly independent, and the distribution of the outcomes of die rolls is not skewed (it's uniform). According to the CLT, the sampling distribution of $\overline{y}$, the average for 10 die rolls, can be approximated by a Normal model, with $\mu_{\overline{y}} = 3.5$ and standard deviation $\sigma(\overline{y}) = \dfrac{1.7}{\sqrt{10}} \approx 0.538$, even though 10 rolls is a fairly small sample.

According to the Normal model, the probability that the average of 10 die rolls is between 3 and 4 (and therefore the probability of the sum of 10 die rolls is between 30 and 40) is approximately 0.647.



9. Net-Newsers.

a) The Pew Research Foundation believes that the true proportion of people who obtain news from the Internet is between 11% and 15%.

b) The smaller sample size in the cell sample would result in a larger standard error. This would make the margin of error larger, as well.

c) $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}} = (0.82) \pm 1.960\sqrt{\dfrac{(0.82)(0.18)}{470}} = (78.5\%, 85.5\%)$

We are 95% confident that between 78.5% and 85.5% of Net-Newsers get news during the course of the day.

d) The sample of 470 Net-Newsers is smaller than either of the earlier samples. This results in a larger margin of error.

10. Gay marriage.

a) **Randomization condition:** Pew Research randomly selected 1821 U.S. adults.
10% condition: 1821 results is less than 10% of all U.S. adults.
Success/Failure condition: $n\hat{p}$ = (1821)(0.54) = 983 and $n\hat{q}$ = (1821)(0.46) = 838 are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion *z*-interval to estimate the percentage of U.S. adults who support marriage equality.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.54) \pm 1.960\sqrt{\frac{(0.54)(0.46)}{1821}} = (51.7\%, 56.3\%)$$

We are 95% confident that between 51.7% and 56.3% of U.S. adults support marriage equality.

b) Since the interval is entirely above 50%, there is evidence that a majority of U.S. adults support marriage equality.

c)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.02 = 2.326\sqrt{\frac{(0.50)(0.50)}{n}}$$
$$n = \frac{(2.326)^2(0.50)(0.50)}{(0.02)^2}$$
$$n \approx 3382 \text{ people}$$

We do not know the true proportion of U.S. adults who support marriage equality, so use $\hat{p} = \hat{q} = 0.50$, for the most cautious estimate. In order to determine the proportion of U.S. adults who support marriage equality to within 2% with 98% confidence, we would have to sample at least 3382 people.

11. Bimodal.

a) The *sample's* distribution (NOT the *sampling* distribution), is expected to look more and more like the distribution of the population, in this case, bimodal.

b) The expected value of the sample's mean is expected to be μ, the population mean, regardless of sample size.

c) The variability of the sample mean, $\sigma(\bar{y})$, is $\frac{\sigma}{\sqrt{n}}$, the population standard deviation divided by the square root of the sample size, regardless of the sample size.

d) As the sample size increases, the sampling distribution model becomes closer and closer to a Normal model.

12. Vitamin D 2012.

a) Certainly, the 11,218 Australian adults are less than 10% of all Australian adults, and $n\hat{p}$ = (11,218)(0.31) = 3478 and $n\hat{q}$ = (11,218)(0.69) = 7740 are both greater than 10, so the sample is large enough. We would like to know that the sample is random. This would help assure us that these people were chosen independently.

b) $\hat{p} \pm z^*\sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.31) \pm 1.960\sqrt{\frac{(0.31)(0.69)}{11218}} = (30.1\%, 31.9\%)$.

c) We are 95% confident that between 30.1% and 31.9% of Australian adults have a vitamin D deficiency.

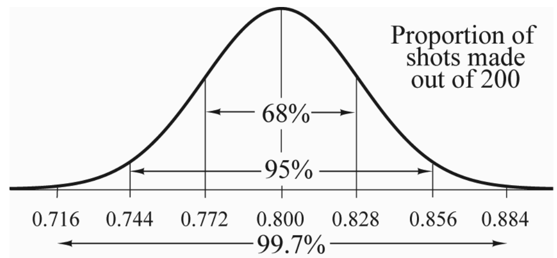
d) 95% of all random samples of this size will produce intervals that contain the true proportion of Australia adults who have a vitamin D deficiency.

13. Archery.

a) $\mu_{\hat{p}} = p = 0.80$ $\qquad \sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.80)(0.20)}{200}} \approx 0.028$

b) $np = (200)(0.80) = 160$ and $nq = (200)(0.20) = 40$ are both greater than 10, so the Normal model is appropriate.

c) The Normal model of the sampling distribution of the proportion of bull's-eyes she makes out of 200 is at the right.

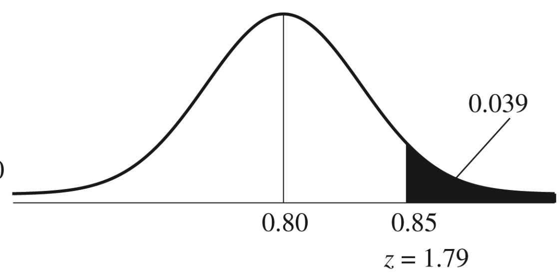


Approximately 68% of the time, we expect her to hit the bull's-eye on between 77.2% and 82.8% of her shots. Approximately 95% of the time, we expect her to hit the bull's-eye on between 74.4% and 85.6% of her shots. Approximately 99.7% of the time, we expect her to hit the bull's-eye on between 71.6% and 88.4% of her shots.

d) According to the Normal model, the probability that she hits the bull's-eye in at least 85% of her 200 shots is approximately 0.039.

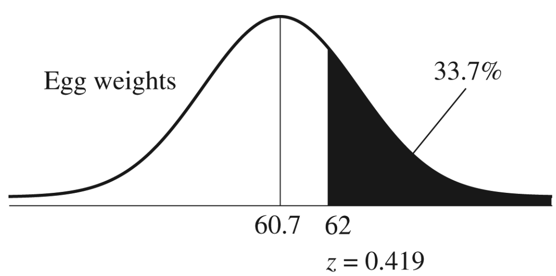
$$z = \frac{0.85 - 0.80}{0.028}$$

$$z \approx 1.79$$



14. Eggs.

a) According to the Normal model, approximately 33.7% of these eggs weigh more than 62 grams.



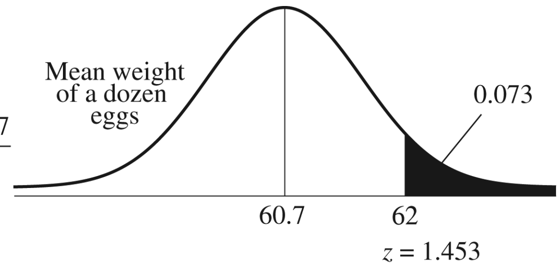
b) **Randomization condition:** The dozen eggs are selected randomly.
10% condition: The dozen eggs are less than 10% of all eggs.

The mean egg weight is $\mu = 60.7$ grams, with standard deviation $\sigma = 3.1$ grams. Since the distribution of egg weights is Normal, we can model the sampling distribution of the mean egg weight of a dozen eggs with a Normal model, with $\mu_{\bar{y}} = 60.7$ grams and standard deviation $\sigma(\bar{y}) = \frac{3.1}{\sqrt{12}} \approx 0.895$ grams.

According to the Normal model, the probability that a randomly selected dozen eggs have a mean greater than 62 grams is approximately 0.073.

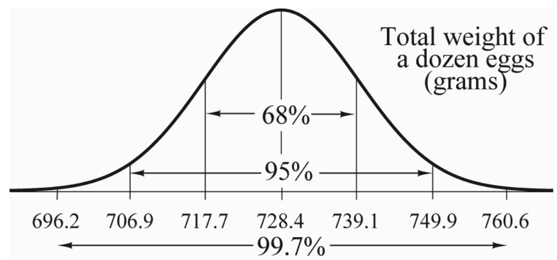
$$z = \frac{62 - 60.7}{\frac{3.1}{\sqrt{12}}}$$

$z \approx 1.453$



c) The average weight of a dozen eggs can be modeled by $N(60.7, 0.895)$, so the total weight of a dozen eggs can be modeled by $N(728.4, 10.74)$.

Approximately 68% of the cartons of a dozen eggs would weigh between 717.1 and 739.1 grams. Approximately 95% of the cartons would weigh between 706.9 and 749.9 grams. Approximately 99.7% of the cartons would weigh between 696.2 and 760.6 grams.



15. Polling disclaimer.

a) It is not clear what specific question the pollster asked. Otherwise, they did a great job of identifying the W's.

b) A sample that was stratified by age, sex, region, and education was used.

c) The margin of error was 4%.

d) Since "no more than 1 time in 20 should chance variations in the sample cause the results to vary by more than 4 percentage points", the confidence level is 19/20 = 95%.

e) The subgroups had smaller sample sizes than the larger group. The standard errors in these subgroups were larger as a result, and this caused the margins of error to be larger.

f) They cautioned readers about response bias due to wording and order of the questions.

16. Enough eggs?

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.02 = 1.960\sqrt{\frac{(0.75)(0.25)}{n}}$$

$$n = \frac{(1.960)^2(0.75)(0.25)}{(0.02)^2}$$

$$n \approx 1801 \text{ eggs}$$

ISA Babcock needs to collect data on about 1800 hens in order to advertise the production rate for the B300 Layer with 95% confidence with a margin of error of $\pm 2\%$.

17. Teen deaths 2005.

a) H_0 : The percentage of fatal accidents involving teenage girls is 12.6%, the same as the overall percentage of fatal accidents involving teens . (p = 0.126)
H_A : The percentage of fatal accidents involving teenage girls is lower than 12.6%, the overall percentage of fatal accidents involving teens . (p < 0.126)

Randomization condition: Assume that the 388 accidents observed are representative of all accidents.
10% condition: The sample of 388 accidents is less than 10% of all accidents.
Success/Failure condition: np = (388)(0.126) = 48.888 and nq = (388)(0.874) = 339.112 are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p =$ 0.126 and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.126)(0.874)}{388}} \approx 0.01685.$$

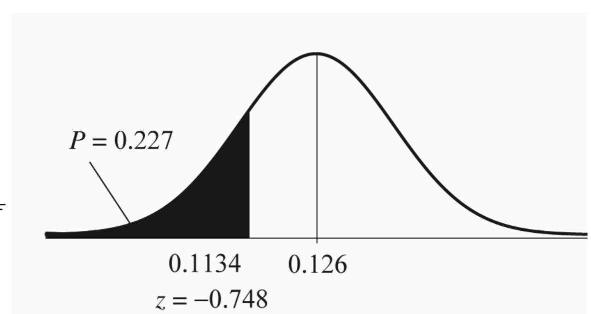
We can perform a one-proportion z-test. The observed proportion of fatal accidents involving teen girls is $\hat{p} = \frac{44}{388} \approx 0.1134$.

Since the P-value = 0.227 is high, we fail to reject the null hypothesis. There is little evidence that the proportion of fatal accidents involving teen girls is less than the overall proportion of fatal accidents involving teens.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.13 - 0.126}{\sqrt{\frac{(0.126)(0.874)}{388}}}$$

$$z \approx -0.748$$



b) If the proportion of fatal accidents involving teenage girls is really 12.6%, we expect to see the observed proportion, 11.34%, in about 22.7% of samples of size 388 simply due to sampling variation.

18. Largemouth bass.

a) One would expect many small fish, with a few large fish.

b) We cannot determine the probability that a largemouth bass caught from the lake weighs over 3 pounds because we don't know the exact shape of the distribution. We know that it is NOT Normal.

c) It would be quite risky to attempt to determine whether or not the mean weight of 5 fish was over 3 pounds. With a skewed distribution, a sample of size 5 is not large enough for the Central Limit Theorem to guarantee that a Normal model is appropriate to describe the distribution of the mean.

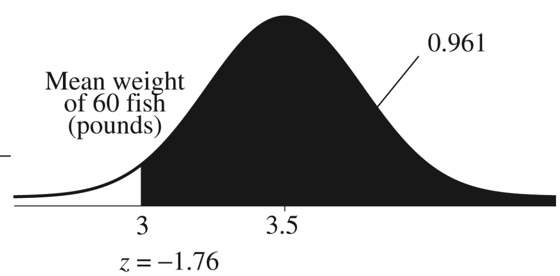
d) A sample of 60 randomly selected fish is large enough for the Central Limit Theorem to guarantee that a Normal model is appropriate to describe the sampling distribution of the mean, as long as 60 fish is less than 10% of the population of all the fish in the lake.

The mean weight is $\mu = 3.5$ pounds, with standard deviation $\sigma = 2.2$ pounds. Since the sample size is sufficiently large, we can model the sampling distribution of the mean weight of 60 fish with a Normal model, with $\mu_{\bar{y}} = 3.5$ pounds and standard deviation $\sigma(\bar{y}) = \dfrac{2.2}{\sqrt{60}} \approx 0.284$ pounds.

According to the Normal model, the probability that 60 randomly selected fish average more than 3 pounds is approximately 0.961.

$$z = \frac{3 - 3.5}{\frac{2.2}{\sqrt{60}}}$$

$$z \approx -1.76$$



19. Cheating.

a) **Randomization condition:** The 4500 students were selected randomly.
10% condition: 4500 students is less than 10% of all students.
Success/Failure condition: $n\hat{p} = (4500)(0.74) = 3330$ and $n\hat{q} = (4500)(0.26) = 1170$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of students who have cheated at least once.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.74) \pm 1.645\sqrt{\frac{(0.74)(0.26)}{4500}} = (72.9\%, 75.1\%)$$

b) We are 90% confident that between 72.9% and 75.1% of high school students have cheated at least once.

c) About 90% of random samples of size 4500 will produce intervals that contain the true proportion of high school students who have cheated at least once.

d) A 95% confidence interval would be wider. Greater confidence requires a larger margin of error.

20. Language.

a) **Randomization condition:** 60 people were selected at random.
10% condition: The 60 people represent less than 10% of all people.
Success/Failure condition: $np = (60)(0.80) = 48$ and $nq = (60)(0.20) = 12$ are both greater than 10.

Therefore, the sampling distribution model for the proportion of 60 randomly selected people who have left-brain language control is Normal, with

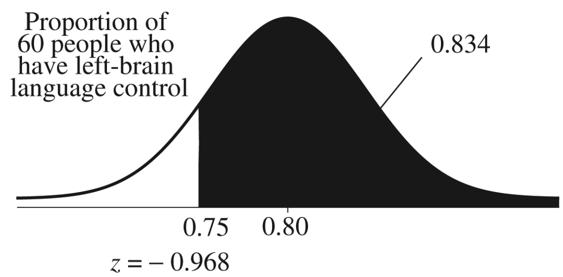
$\mu_{\hat{p}} = p = 0.80$ and standard deviation $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.80)(0.20)}{60}} \approx 0.0516$.

b) According to the Normal model, the probability that over 75% of these 60 people have left-brain language control is approximately 0.834.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.75 - 0.80}{\sqrt{\frac{(0.80)(0.20)}{60}}}$$

$$z \approx -0.968$$



c) If the sample had consisted of 100 people, the probability would have been higher. A larger sample results in a smaller standard deviation for the sample proportion.

d) Answers may vary. Let's consider three standard deviations below the expected proportion to be "almost certain". It would take a sample of (exactly!) 576 people to make sure that 75% would be 3 standard deviations below the expected percentage of people with left-brain language control.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$-3 = \frac{0.75 - 0.80}{\sqrt{\frac{(0.80)(0.20)}{n}}}$$

$$n = \frac{(-3)^2(0.80)(0.20)}{(0.75 - 0.80)^2} = 576$$

Using round numbers for n instead of z, about 500 people in the sample would make the probability of choosing a sample with at least 75% of the people having left-brain language control is a whopping 0.997. It all depends on what "almost certain" means to you.

21. Cigarettes 2009.

a) H_0: 17.7% of high school students smoke. ($p = 0.177$)
H_A: More than 17.7% of high school students smoke. ($p > 0.177$)

b) **Randomization condition:** The CDC randomly sampled 5080 high school students.
10% condition: The sample of 5080 students is less than 10% of all high school students.
Success/Failure condition: $np = (5080)(0.177) = 899.16$ and $nq = (5080)(0.823) = 4180.84$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.177$ and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.177)(0.823)}{5080}} \approx 0.0054.$$

c) We can perform a one-proportion z-test. The observed proportion of high school students who smoke is $\hat{p} = 0.195$ This proportion is about 3.36 standard deviations above the hypothesized proportion of smokers.

The P-value of this test is 0.0004.

d) If the proportion of students who smoke is actually 17.7%, the probability that a sample of this size would have a sample proportion of 19.5% or higher is 0.0004.

e) Since the P-value = 0.0004 is low, we reject the null hypothesis. There is strong evidence that greater than 17.7% of all high school students smoked in 2009. The goal is not on track.

f) If the conclusion is incorrect, a Type I error has been made.

22. Church going.

a) **Randomization condition:** We will assume that Pew Research used a random sample of American adults.
10% condition: 2303 is less than 10% of all American adults.
Success/Failure condition: The number of church-goers and non-church-goers in the sample, 921 and 1382 respectively, are both more than 10, so the sample is large enough.

Since the conditions have been satisfied, we will find a one-proportion z-interval.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{921}{2303}\right) \pm 1.960 \sqrt{\frac{\left(\frac{921}{2303}\right)\left(\frac{1382}{2303}\right)}{2303}} = (38.0\%,\ 42.0\%)$$

b) We are 95% confident that the proportion of all American adults who are active in church, religious, or spiritual organizations is between 38% and 42%.

c) 95% of all random samples of size 2303 will produce confidence intervals that contain the true proportion of American adults who are active in church, religious, or spiritual organizations.

23. Teen smoking 2009.

Randomization condition: Assume that the freshman class is representative of all teenagers. This may not be a reasonable assumption. There are many interlocking relationships between smoking, socioeconomic status, and college attendance. This class may not be representative of all teens with regards to smoking simply because they are in college. Be cautious with your conclusions!
10% condition: The freshman class is less than 10% of all teenagers.
Success/Failure condition: $np = (522)(0.195) = 101.79$ and $nq = (522)(0.805) = 420.21$ are both greater than 10.

Therefore, the sampling distribution model for the proportion of 522 students who smoke is Normal, with $\mu_{\hat{p}} = p = 0.195$, and standard deviation

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.195)(0.805)}{522}} \approx 0.0173.$$

25% is about 3.2 standard deviations above the expected proportion of smokers. If the true proportion of smokers is 19.5%, the Normal model predicts that the probability that more than 25% of these students smoke is approximately 0.0008. It is very unlikely that more than 25% the freshman class smokes.

24. Alcohol abuse.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 1.645\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$n = \frac{(1.645)^2 (0.5)(0.5)}{(0.04)^2}$$

$$n \approx 423$$

The university will have to sample at least 423 students in order to estimate the proportion of students who have been drunk with in the past week to within ± 4%, with 90% confidence.

25. Errors.

a) Since a treatment (the additive) is imposed, this is an experiment.

b) The company is only interested in a decrease in the percentage of cars needing repairs, so they will perform a one-sided test.

c) The independent laboratory will make a Type I error if they decide that the additive reduces the number of repairs, when it actually makes no difference in the number of repairs.

d) The independent laboratory will make a Type II error if they decide that the additive makes no difference in the number of repairs, when it actually reduces the number of repairs.

e) The additive manufacturer would consider a Type II error more serious. The lab claims that the manufacturer's product doesn't work, and it actually does.

f) Since this was a controlled experiment, the company can conclude that the additive is the reason that the cabs are running better. They should be cautious recommending it for all cars. There is evidence that the additive works well for cabs, which get heavy use. It might not be effective in cars with a different pattern of use than cabs.

26. Safety.

a) $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.14) \pm 1.960 \sqrt{\frac{(0.14)(0.86)}{814}} = (11.6\%, 16.4\%)$

We are 95% confident that between 11.6% and 16.4% of Texas children wear helmets when biking, roller skating, or skateboarding.

b) These data might not be a random sample.

c)

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 2.326 \sqrt{\frac{(0.14)(0.86)}{n}}$$

$$n = \frac{(2.326)^2 (0.14)(0.86)}{(0.04)^2}$$

$$n \approx 408$$

If we use the 14% estimate obtained from the first study, the researchers will need to observe at least 408 kids in order to estimate the proportion of kids who wear helmets to within 4%, with 98% confidence.
(If you use a more cautious approach, estimating that 50% of kids wear helmets, you need a whopping 846 observations. Are you beginning to see why pilot studies are conducted?)

27. Fried PCs.

a) H_0: The computer is undamaged.
H_A: The computer is damaged.

b) The biggest advantage is that all of the damaged computers will be detected, since, historically, damaged computers never pass all the tests. The disadvantage is that only 80% of undamaged computers pass all the tests. The engineers will be classifying 20% of the undamaged computers as damaged.

c) In this example, a Type I error is rejecting an undamaged computer. To allow this to happen only 5% of the time, the engineers would reject any computer that failed 3 or more tests, since 95% of the undamaged computers fail two or fewer tests.

d) The power of the test in part c is 20%, since only 20% of the damaged machines fail 3 or more tests.

e) By declaring computers "damaged" if they fail 2 or more tests, the engineers will be rejecting only 7% of undamaged computers. From 5% to 7% is an increase of 2% in α. Since 90% of the damaged computers fail 2 or more tests, the power of the test is now 90%, a substantial increase.

28. Power.

a) Power will increase, since the variability in the sampling distribution will decrease. We are more certain of all our decisions when there is less variability.

b) Power will decrease, since we are rejecting the null hypothesis less often.

29. Approval 2008.

H_0 : George W. Bush's April 2008 disapproval rating was 66%. ($p = 0.66$)
H_A : George W. Bush's disapproval rating was higher than 66%. ($p > 0.66$)

Randomization condition: The adults were chosen randomly.
10% condition: 1016 adults are less than 10% of all adults.
Success/Failure condition: $np = (1016)(0.66) = 670.56$ and $nq = (1016)(0.34) = 345.44$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.66$ and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.66)(0.34)}{1016}} \approx 0.015.$$

We can perform a one-proportion z-test. The observed approval rating is $\hat{p} = 0.69$.

The value of z is 2.02. Since the P-value = 0.022 is low, we reject the null hypothesis. There is strong evidence that President George W. Bush's April 2008 disapproval rating was higher than the 66% disapproval rating of President Richard Nixon.

30. Grade inflation.

H_0 : In 2000, 20% of students at the university had a GPA of at least 3.5. ($p = 0.20$)
H_A : In 2000, more than 20% of students had a GPA of at least 3.5. ($p > 0.20$)

Randomization condition: The GPAs were chosen randomly.
10% condition: 1100 GPAs are less than 10% of all GPAs.
Success/Failure condition: $np = (1100)(0.20) = 220$ and $nq = (1100)(0.80) = 880$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.20$ and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.20)(0.80)}{1100}} \approx 0.0121.$$

We can perform a one-proportion z-test. The observed approval rating is $\hat{p} = 0.25$.

Since the P-value is less than 0.0001, which is low, we reject the null hypothesis. There is strong evidence the percentage of students whose GPAs are at least 3.5 is higher in 2000 than in 1996.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.25 - 0.20}{\sqrt{\frac{(0.20)(0.80)}{1100}}}$$

$$z \approx 4.15$$

31. Name recognition.

a) The company wants evidence that the athlete's name is recognized more often than 25%.

b) Type I error means that fewer than 25% of people will recognize the athlete's name, yet the company offers the athlete an endorsement contract anyway. In this case, the company is employing an athlete that doesn't fulfill their advertising needs.

Type II error means that more than 25% of people will recognize the athlete's name, but the company doesn't offer the contract to the athlete. In this case, the company is letting go of an athlete that meets their advertising needs.

c) If the company uses a 10% level of significance, the company will hire more athletes that don't have high enough name recognition for their needs. The risk of committing a Type I error is higher.

At the same level of significance, the company is less likely to lose out on athletes with high name recognition. They will commit fewer Type II errors.

32. Name recognition, part II.

a) The 2% difference between the 27% name recognition in the sample, and the desired 25% name recognition may have been due to sampling error. It's possible that the actual percentage of all people who recognize the name is lower than 25%, even though the percentage in the sample of 500 people was 27%. The company just wasn't willing to take that chance. They'll give the endorsement contract to an athlete that they are convinced has better name recognition.

b) The company committed a Type II error. The null hypothesis (that only 25% of the population would recognize the athlete's name) was false, and they didn't notice.

c) The power of the test would have been higher if the athlete were more famous. It would have been difficult not to notice that an athlete had, for example, 60% name recognition if they were only looking for 25% name recognition.

33. Dropouts.

Randomization condition: Assume that these subjects are representative of all anorexia nervosa patients.
10% condition: 198 is less than 10% of all patients.
Success/Failure condition: The number of dropouts, 105, and the number of subjects that remained, 93, are both greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a one-proportion z-interval.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{105}{198}\right) \pm 1.960 \sqrt{\frac{\left(\frac{105}{198}\right)\left(\frac{93}{198}\right)}{198}} = (46\%,\ 60\%)$$

We are 95% confident that between 46% and 60% of anorexia nervosa patients will drop out of treatment programs. However, this wasn't a random sample of all patients. They were assigned to treatment programs rather than choosing their own. They may have had different experiences if they were not part of an experiment.

34. Women.

H_0 : The percentage of businesses in the area owned by women is 26%. ($p = 0.26$)
H_A : The percentage of businesses owned by women is not 26%. ($p \neq 0.26$)

Random condition: This is a random sample of 410 businesses.
10% condition: The sample of 410 businesses is less than 10% of all businesses.
Success/Failure condition: $np = (410)(0.26) = 106.6$ and $nq = (410)(0.74) = 303.4$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.26$ and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.26)(0.74)}{410}} \approx 0.02166.$$

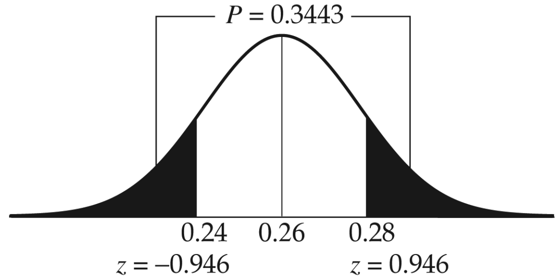
We can perform a one-proportion z-test. The observed proportion of businesses owned by women is $\hat{p} = \dfrac{115}{410} \approx 0.2805$.

Since the P-value = 0.3443 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of businesses in the Denver area owned by women is any different than the national figure of 26%.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.2805 - 0.26}{\sqrt{\frac{(0.26)(0.74)}{410}}}$$

$$z \approx 0.946$$



35. Speeding.

a) H_0 : The percentage of speeding tickets issued to black drivers is 16%, the same as the percentage of registered drivers who are black. ($p = 0.16$)

H_A : The percentage of speeding tickets issued to black drivers is greater than 16%, the percentage of registered drivers who are black. ($p > 0.16$)

Random condition: Assume that this month is representative of all months with respect to the percentage of tickets issued to black drivers.
10% condition: 324 speeding tickets are less than 10% of all tickets.
Success/Failure condition: $np = (324)(0.16) = 52$ and $nq = (324)(0.84) = 272$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.16$ and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.16)(0.84)}{324}} \approx 0.02037.$$

We can perform a one-proportion z-test.
The observed proportion of tickets issued is $\hat{p} = 0.25$.

Since the P-value = 4.96×10^{-6} is very low, we reject the null hypothesis. There is strong evidence that the percentage of speeding tickets issued to black drivers is greater than 16%.

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})}$$

$$z \approx \frac{0.25 - 0.16}{\sqrt{\frac{(0.16)(0.84)}{324}}}$$

$$z \approx 4.42$$

b) There is strong evidence of an association between the receipt of a speeding ticket and race. Black drivers appear to be issued tickets at a higher rate than expected. However, this does not prove that racial profiling exists. There may be other factors present.

c) Answers may vary. The primary statistic of interest is the percentage of black motorists on this section of the New Jersey Turnpike. For example, if 80% of drivers on this section are black, then 25% of the speeding tickets being issued to black motorists is not an usually high percentage. In fact, it is probably unusually low. On the other hand, if only 3% of the motorists on this section of the turnpike are black, then there is even more evidence that racial profiling may be occurring.

36. Petitions.

a) $\frac{1772}{2000} = 0.886 = 88.6\%$ of the sample signatures were valid.

b) $\frac{250,000}{304,266} \approx 0.822 \approx 82.2\%$ of the petition signatures must be valid in order to have the initiative certified by the Elections Committee.

c) If the Elections Committee commits a Type I error, a petition would be certified when there are not enough valid signatures.

d) If the Elections Committee commits a Type II error, a valid petition is not certified.

e) H_0: The percentage of valid signatures is 82.2% ($p = 0.822$)
H_A: The percentage of valid signatures is greater than 82.2% ($p > 0.822$)

Random Condition: This is a simple random sample of 2000 signatures.
10% condition: The sample of 2000 signatures is less than 10% of all signatures.
Success/Failure condition: $np = (2000)(0.822) = 1644$ and $nq = (2000)(0.178) = 356$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.822$ and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.822)(0.178)}{2000}} \approx 0.00855.$$

We can perform a one-proportion z-test.
The observed proportion of valid signatures is

$$\hat{p} = \frac{1772}{2000} \approx 0.886.$$

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})}$$

$$z \approx \frac{0.886 - 0.822}{0.008553}$$

$$z \approx 7.48$$

Since the P-value $= 3.64 \times 10^{-14}$ is low, we reject the null hypothesis. There is strong evidence that the percentage of valid signatures is greater than 82.2%. The petition should be certified.

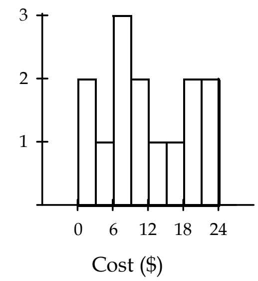
f) In order to increase the power of their test to detect valid petitions, the Elections Committee could sample more signatures.

37. Meals.

H_0: The college student's mean daily food expense is \$10. $(\mu = 10)$

H_A: The college student's mean daily food expense is greater than \$10. $(\mu > 10)$

Randomization condition: Assume that these days are representative of all days.
Nearly Normal condition: The histogram of daily expenses is fairly unimodal and symmetric. It is reasonable to think that this sample came from a Normal population.



The expenses in the sample had a mean of 11.4243 dollars and a standard deviation of 8.05794 dollars. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean daily expense with a Student's t model, with 14 – 1 = 13 degrees of freedom, $t_{13}\left(10, \dfrac{8.05794}{\sqrt{14}}\right)$.

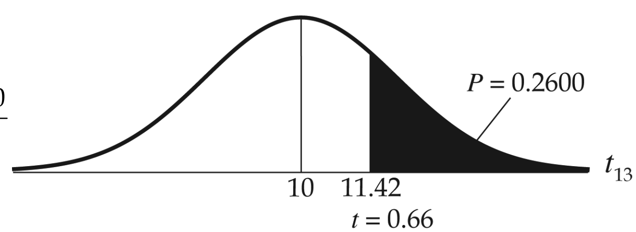
We will perform a one-sample t-test.

Since the P-value = 0.2600 is high, we fail to reject the null hypothesis. There is no evidence that the student's average spending is more than \$10 per day.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{11.4243 - 10}{\frac{8.05794}{\sqrt{14}}}$$

$$t \approx 0.66$$



38. Occupy Wall Street.

Randomization condition: The 901 American adults were sampled randomly.
10% condition: 901 is less than 10% of all American adults.
Success/Failure condition: $n\hat{p}$ = (901)(0.599) = 540 and $n\hat{q}$ = (901)(0.401) = 361 are both greater than 10, so the sample is large enough.

Since the conditions have been satisfied, we will find a one-proportion z-interval.

$$\hat{p} \pm z^*\sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{540}{901}\right) \pm 1.960\sqrt{\frac{\left(\frac{540}{901}\right)\left(\frac{361}{901}\right)}{901}} = (56.7\%, 63.1\%)$$

We are 95% confident that between 56.7% and 63.1% of all American adults agree with the statement "The Occupy Wall Street protesters offered new insights on social issues."

39. Streams.

Random condition: The researchers randomly selected 172 streams.
10% condition: 172 is less than 10% of all streams.
Success/Failure condition: $n\hat{p} = 69$ and $n\hat{q} = 103$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of Adirondack streams with a shale substrate.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{69}{172}\right) \pm 1.960\sqrt{\frac{\left(\frac{69}{172}\right)\left(\frac{103}{172}\right)}{172}} = (32.8\%, 47.4\%)$$

We are 95% confident that between 32.8% and 47.4% of Adirondack streams have a shale substrate.

40. Skin cancer.

a) **Independence assumption:** We must assume that the 152 patients are representative of others with skin cancer.
10% condition: 152 is less than 10% of all skin cancer patients.
Success/Failure condition: $n\hat{p} = (152)(0.53) = 81$ and $n\hat{q} = (152)(0.47) = 71$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of all skin cancer patients that would have a partial or complete response to vemurafenib.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{81}{152}\right) \pm 1.960\sqrt{\frac{\left(\frac{81}{152}\right)\left(\frac{71}{152}\right)}{152}} = (45.36\%, 61.22\%)$$

We are 95% confident that between 45.36% and 61.22% of all patients with metastatic melanoma would have a partial or complete response to vemurafenib.

b)

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.06 = 1.960\sqrt{\frac{(0.53)(0.47)}{n}}$$

$$n = \frac{(1.960)^2 (0.53)(0.47)}{(0.06)^2}$$

$$n \approx 266$$

If we use the 53% estimate obtained from the first study, the researchers will need to study at least 266 patients in order to estimate the proportion of skin cancer patients who would have a partial or complete response to vemurafenib to within 6% at 95% confidence. (If they use a more cautious approach with an estimate of 50%, they would need 267 patients.)

41. Bread.

a) Since the histogram shows that the distribution of the number of loaves sold per day is skewed strongly to the right, we can't use the Normal model to estimate the number of loaves sold on the busiest 10% of days.

b) **Randomization condition:** Assume that these days are representative of all days.
Nearly Normal condition: The histogram is skewed strongly to the right. However, since the sample size is large, the Central Limit Theorem guarantees that the distribution of averages will be approximately Normal.

The days in the sample had a mean of 103 loaves sold and a standard deviation of 9 loaves sold. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's *t*- model, with 100 – 1 = 99 degrees of freedom. We will use a one-sample *t*-interval with 95% confidence for the mean number of loaves sold. (By hand, use $t_{50}^* \approx 2.403$ from the table.)

c) $\bar{y} \pm t_{n-1}^*\left(\frac{s}{\sqrt{n}}\right) = 103 \pm t_{99}^*\left(\frac{9}{\sqrt{100}}\right) \approx (101.2,\ 104.8)$

We are 95% confident that the mean number of loaves sold per day at the Clarksburg Bakery is between 101.2 and 104.8.

d) We know that in order to cut the margin of error in half, we need to a sample four times as large. If we allow a margin of error that is twice as wide, that would require a sample only one-fourth the size. In this case, our original sample is 100 loaves; so 25 loaves would be a sufficient number to estimate the mean with a margin of error twice as wide.

e) Since the interval is completely above 100 loaves, there is strong evidence that the estimate was incorrect. The evidence suggests that the mean number of loaves sold per day is greater than 100. This difference is statistically significant, but may not be practically significant. It seems like the owners made a pretty good estimate!

42. Fritos.

a) H_0: The mean weight of bags of Fritos is 35.4 grams. $(\mu = 35.4)$

H_A: The mean weight of bags of Fritos is less than 35.4 grams. $(\mu < 35.4)$

b) **Randomization condition:** It is reasonable to think that the 6 bags are representative of all bags of Fritos.
Nearly Normal condition: The histogram of bags weights shows one unusually heavy bag. Although not technically an outlier, it probably should be excluded for the purposes of the test. (We will leave it in for the preliminary test, then remove it and test again.)

c) The bags in the sample had a mean weight of 35.5333 grams and a standard deviation in weight of 0.450185 grams. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean weight of bags of Fritos with a Student's t model, with 6 – 1 = 5 degrees of freedom,

$t_5\left(35.4, \dfrac{0.450185}{\sqrt{6}}\right)$.

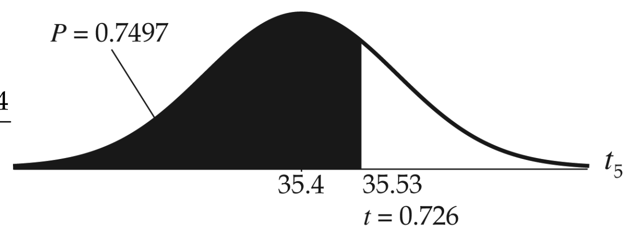
We will perform a one-sample t-test.

Since the P-value = 0.7497 is high, we fail to reject the null hypothesis. There is no evidence to suggest that the mean weight of bags of Fritos is less than 35.4 grams.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{35.5333 - 35.4}{\frac{0.450185}{\sqrt{6}}}$$

$$t = 0.726$$



d) With the one unusually high value removed, the mean weight of the 5 remaining bags is 35.36 grams, with a standard deviation in weight of 0.167332 grams. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean weight of bags of Fritos with a Student's t model, with 5 – 1 = 4 degrees of freedom, $t_4\left(35.4, \dfrac{0.167332}{\sqrt{5}}\right)$.

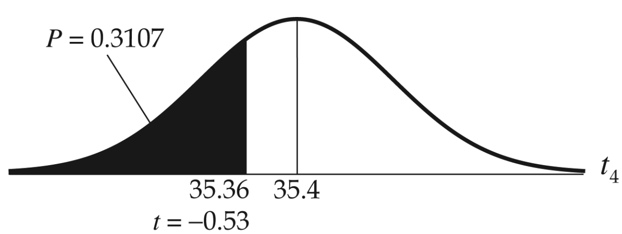
We will perform a one-sample t-test.

Since the P-value = 0.3107 is high, we fail to reject the null hypothesis. There is no evidence to suggest that the mean weight of bags of Fritos is less than 35.4 grams.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{35.36 - 35.4}{\frac{0.167332}{\sqrt{5}}}$$

$$t = -0.53$$



e) Neither test provides evidence that the mean weight of bags of Fritos is less than 35.4 grams. It is reasonable to believe that the mean weight of the bags is the same as the stated weight. However, the sample sizes are very small, and the tests have very little power to detect lower mean weights. It would be a good idea to weigh more bags.

43. And it means?

a) The margin of error is $\dfrac{(2391-1644)}{2} = \$373.50$.

b) The insurance agent is 95% confident that the mean loss claimed by clients after home burglaries is between \$1644 and \$2391.

c) 95% of all random samples of this size will produce intervals that contain the true mean loss claimed.

44. Batteries.

a) Different samples have different means. Since this is a fairly small sample, the difference may be due to natural sampling variation. Also, we have no idea how to quantify "a lot less" without considering the variation as measured by the standard deviation.

b) H_0: The mean life of a battery is 100 hours. $(\mu = 100)$

H_A: The mean life of a battery is less than 100 hours. $(\mu < 100)$

c) **Randomization condition:** It is reasonable to think that these 16 batteries are representative of all batteries of this type
Normal population assumption: Since we don't have the actual data, we can't check a graphical display, and the sample is not large. Assume that the population of battery lifetimes is Normal.

d) The batteries in the sample had a mean life of 97 hours and a standard deviation of 12 hours. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean battery life with a Student's t model, with 16 – 1 = 15 degrees of freedom, $t_{15}\left(100, \frac{12}{\sqrt{16}}\right)$, or $t_{15}(100, 3)$

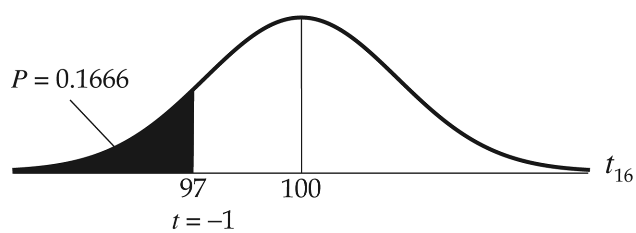
We will perform a one-sample t-test.

Since the P-value = 0.1666 is greater than $\alpha = 0.05$, we fail to reject the null hypothesis. There is no evidence to suggest that the mean battery life is less than 100 hours.

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

$$t = \frac{97 - 100}{\frac{12}{\sqrt{16}}}$$

$$t = -1$$



e) If the mean life of the company's batteries is only 98 hours, then the mean life is less than 100, and the null hypothesis is false. We failed to reject a false null hypothesis, making a Type II error.

Chapter 22 – Comparing Groups

Section 22.1

1. **Canada.**

$$SE\left(\hat{p}_{Can}-\hat{p}_{Am}\right)=\sqrt{\frac{\hat{p}_{Can}\hat{q}_{Can}}{n_{Can}}+\frac{\hat{p}_{Am}\hat{q}_{Am}}{n_{Am}}}=\sqrt{\frac{\left(\frac{192}{960}\right)\left(\frac{768}{960}\right)}{960}+\frac{\left(\frac{170}{1250}\right)\left(\frac{1080}{1250}\right)}{1250}}=0.0161$$

2. **Non-profits.**

$$SE\left(\hat{p}_{Non}-\hat{p}_{For}\right)=\sqrt{\frac{\hat{p}_{Non}\hat{q}_{Non}}{n_{Non}}+\frac{\hat{p}_{For}\hat{q}_{For}}{n_{For}}}=\sqrt{\frac{\left(\frac{377}{422}\right)\left(\frac{45}{422}\right)}{422}+\frac{\left(\frac{431}{518}\right)\left(\frac{87}{518}\right)}{518}}=0.0223$$

Section 22.2

3. **Canada, deux.**

   We must assume the data were collected randomly and that the Americans selected are independent of the Canadians selected. Both assumptions should be met. Also, for both groups, we have at least 10 national-born and foreign-born citizens and the sample sizes are less than 10% of the population sizes. All conditions for inference are met.

4. **Non-profits, part 2.**

   We have random samples and we must assume the samples were collected independently of each other. This should be met. For both groups, we have at least 10 people who are highly satisfied and 10 people who are not. Finally, the sample sizes are less than 10% of the population sizes—all conditions for inference are met.

Section 22.3

5. **Canada, trois.**

   We are 95% confident that, based on these data, the proportion of foreign-born Canadians is between 3.24% and 9.56% more than the proportion of foreign-born Americans.

6. **Non-profits, part 3.**

   We are 95% confident that, based on these data, the proportion of people who are "highly satisfied" working at non-profits is between 1.77% and 10.50% higher than the proportion of people who are "highly satisfied" working at for-profit companies.

7. Canada, encore.

If we were to take repeated samples of Canadians and Americans, we would expect 95% of the intervals to contain the true difference in the proportion of foreign-born citizens.

8. Non-profits, again.

If we were to take repeated samples of people who work at non-profits and for-profits, we would expect 95% of the intervals to contain the true difference in the proportion of those who are highly satisfied.

Section 22.4

9. Canada, la fin.

a) $\hat{p}_{Can} - \hat{p}_{Am} = \dfrac{192}{960} - \dfrac{170}{1250} = 0.064$

b) $\hat{p}_{pooled} = \dfrac{192+170}{960+1250} = \dfrac{362}{2210} \approx 0.1638$

c) $SE_{pooled}\left(\hat{p}_{Can} - \hat{p}_{Am}\right) = \sqrt{\dfrac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_{Can}} + \dfrac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_{Am}}} = \sqrt{\dfrac{\left(\frac{362}{2210}\right)\left(\frac{1848}{2210}\right)}{960} + \dfrac{\left(\frac{362}{2210}\right)\left(\frac{1848}{2210}\right)}{1250}} \approx 0.0159$

d) $z = \dfrac{\hat{p}_{Can} - \hat{p}_{Am}}{SE_{pooled}\left(\hat{p}_{Can} - \hat{p}_{Am}\right)} = \dfrac{0.064}{0.0159} \approx 4.03$

e) Since the P-value is < 0.001, which is very low, reject the null hypothesis. There is very strong evidence that the proportion of foreign born citizens is different in Canada than it is in the United States. According to this data, the proportion of foreign born Canadians is likely to be the higher of the two.

10. Non-profits, last one.

a) $\hat{p}_{Non} - \hat{p}_{For} = \dfrac{377}{422} - \dfrac{431}{518} \approx 0.061$

b) $\hat{p}_{pooled} = \dfrac{377+431}{422+518} = \dfrac{808}{940} \approx 0.860$

c) $SE_{pooled}\left(\hat{p}_{Non} - \hat{p}_{For}\right) = \sqrt{\dfrac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_{Non}} + \dfrac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_{For}}} = \sqrt{\dfrac{\left(\frac{808}{940}\right)\left(\frac{132}{940}\right)}{422} + \dfrac{\left(\frac{808}{940}\right)\left(\frac{132}{940}\right)}{518}} \approx 0.0228$

d) $z = \dfrac{\hat{p}_{Non} - \hat{p}_{For}}{SE_{pooled}\left(\hat{p}_{Non} - \hat{p}_{For}\right)} = \dfrac{0.061}{0.0228} \approx 2.68$

e) Since the P-value = 0.007 is low, reject the null hypothesis. There is strong evidence that the proportion of highly satisfied workers at non-profits is higher than the proportion of highly satisfied workers at for-profits. These data suggest that the proportion higher at non-profits than at for-profits.

Section 22.5

11. Cost of shopping.

We must assume the samples were random or otherwise independent of each other. We also assume that the distributions are roughly normal, so it would be a good idea to check a histogram to make sure there isn't strong skewness or outliers.

12. Athlete ages.

We must assume the samples are unbiased and independent of each other. These conditions seem met. We also assume that the distributions are roughly normal, so it would be a good idea to check a histogram to make sure there isn't strong skewness or outliers.

13. Cost of shopping, again.

We are 95% confident that the mean purchase amount at Walmart is between \$1.85 and \$14.15 less than the mean purchase amount at Target.

14. Athlete ages, again.

We are 95% confident that the mean age of MLB players is between 0.41 years younger and 3.09 years older than the mean age of NFL players.

Section 22.6

15. Cost of shopping, three.

The difference is –\$8 with an SE of 3.115, so the t-stat is –2.569. With 162.75 (or 163) df, the P-value is 0.011 which is less than 0.05. Reject the null hypothesis that the means are equal. There is evidence that the mean purchase amounts at the two stores are not the same. These data suggest that the mean purchase amount at Target is lower than the mean purchase price at Walmart.

16. Athlete ages, three.

The difference is 1.34 years with an SE of 0.8711 so the t-stat is 1.538. With 51.83 (or 52) df, the P-value is 0.13 which is greater than 0.05. Fail to reject the null hypothesis that the means are equal. There is no evidence of a difference in mean age of MLB and NFL players.

Section 20.7

17. Cost of shopping, once more.

The *t*-statistic is –2.561 using the pooled estimate of the standard deviation, 20.06. There are 163 df so the P-value is still 0.011. We reach the same conclusion as before. Because the sample standard deviations are nearly the same and the sample sizes are large, the pooled test is essentially the same as the two-sample *t*-test.

18. Athlete ages, one more time.

The *t*-statistic is now 1.556 using the pooled estimate of the standard deviation. There are 60 df so the P-value is now 0.125. We reach the same conclusion as before. Because the sample standard deviations are nearly the same, the pooled test is nearly the same as the two-sample *t*-test.

19. Cost of shopping, last look.

No. The two-sample test is almost always the safer choice. In this case, the variances are likely to be quite different. The purchase prices of Italian sports cars are much higher and may be more variable than the domestic prices. They should use the two-sample t-test.

20. Athletes, final inning.

No. The two-sample test is almost always the safer choice. In this case, the variances are likely to be quite different, since third grade Little League players will all be nearly the same age. They should use the two-sample *t*-test.

21. Online social networking.

It is very unlikely that samples would show an observed difference this large if, in fact, there was no real difference between the proportion of American adults who visited Facebook on a daily basis in 2013 and the proportion of American adults who visited Facebook on a daily basis 2010.

22. Science news.

If, in fact, there is no difference between the percentage 18-29-year-olds who read newspapers and the percentage of 30-49-year-olds who read newspapers , then it's not unusual to observe a difference of 4 percentage points by sampling.

Chapter Exercises.

23. Revealing information.

This test is not appropriate for these data, since the responses are not from independent groups, but are from the same individuals. The independent samples condition has been violated.

24. Regulating access.

The 790 parents are a subset of the 935 parents, so the two groups are not independent. This violates the independent samples condition.

25. Gender gap.

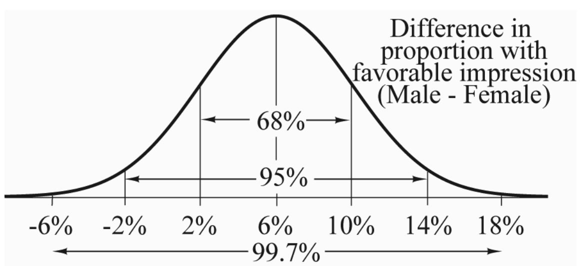
a) This is a stratified random sample, stratified by gender.

b) We would expect the difference in proportions in the sample to be the same as the difference in proportions in the population, with the percentage of respondents with a favorable impression of the candidate 6 percentage points higher among males.

c) The standard deviation of the difference in proportions is:

$$SD(\hat{p}_M - \hat{p}_F) = \sqrt{\frac{\hat{p}_M\hat{q}_M}{n_M} + \frac{\hat{p}_F\hat{q}_F}{n_F}} = \sqrt{\frac{(0.59)(0.41)}{300} + \frac{(0.53)(0.47)}{300}} \approx 4\%$$

d)



e) The campaign could certainly be misled by the poll. According to the model, a poll showing little difference could occur relatively frequently. That result is only 1.5 standard deviations below the expected difference in proportions.

26. Buy it again?

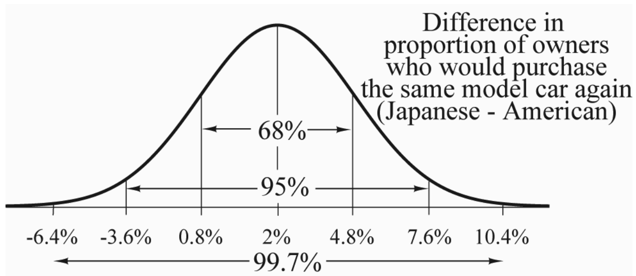
a) This is a stratified random sample, stratified by country of origin of the car.

b) We would expect the difference in proportions in the sample to be the same as the difference in proportions in the population, with the percentage of respondents who would purchase the same model again 2 percentage points higher among owners of Japanese cars than among owners of American cars.

c) The standard deviation of the difference in proportions is:

$$SD(\hat{p}_J - \hat{p}_A) = \sqrt{\frac{\hat{p}_J\hat{q}_J}{n_J} + \frac{\hat{p}_A\hat{q}_A}{n_A}} = \sqrt{\frac{(0.78)(0.22)}{450} + \frac{(0.76)(0.24)}{450}} \approx 2.8\%$$

d)



e) The magazine could certainly be misled by the poll. According to the model, a poll showing greater satisfaction among owners of American cars could occur relatively frequently. That result is a little more than one standard deviation below the expected difference in proportions.

27. Arthritis.

a) **Randomization condition:** Americans age 65 and older were selected randomly.
10% condition: 1012 men and 1062 women are less than 10% of all men and women.
Independent groups assumption: The sample of men and the sample of women were drawn independently of each other.
Success/Failure condition: $n\hat{p}$ (men) = 411, $n\hat{q}$ (men) = 601, $n\hat{p}$ (women) = 535, and $n\hat{q}$ (women) = 527 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b)

$$(\hat{p}_F - \hat{p}_M) \pm z^* \sqrt{\frac{\hat{p}_F\hat{q}_F}{n_F} + \frac{\hat{p}_M\hat{q}_M}{n_M}} = \left(\tfrac{535}{1062} - \tfrac{411}{1012}\right) \pm 1.960\sqrt{\frac{\left(\frac{535}{1062}\right)\left(\frac{527}{1062}\right)}{1062} + \frac{\left(\frac{411}{1012}\right)\left(\frac{601}{1012}\right)}{1012}}$$

$$= (0.055,\ 0.140)$$

c) We are 95% confident that the proportion of American women age 65 and older who suffer from arthritis is between 5.5% and 14.0% higher than the proportion of American men the same age who suffer from arthritis.

d) Since the interval for the difference in proportions of arthritis sufferers does not contain 0, there is strong evidence that arthritis is more likely to afflict women than men.

28. Graduation.

a) **Randomization condition:** Assume that the samples are representative of all recent graduates.
10% condition: Although large, the samples are less than 10% of all graduates.

Independent groups assumption: The sample of men and the sample of women were drawn independently of each other.
Success/Failure condition: The samples are very large, certainly large enough for the methods of inference to be used.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b) $\left(\hat{p}_F - \hat{p}_M\right) \pm z^* \sqrt{\dfrac{\hat{p}_F\hat{q}_F}{n_F} + \dfrac{\hat{p}_M\hat{q}_M}{n_M}}$

$$= (0.881 - 0.849) \pm 1.960\sqrt{\frac{(0.881)(0.119)}{12{,}678} + \frac{(0.849)(0.151)}{12{,}460}} = (0.024,\ 0.040)$$

c) We are 95% confident that the proportion of 24-year-old American women who have graduated from high school is between 2.4 and 4.0 percentage points higher than the proportion of American men the same age who have graduated from high school.

d) Since the interval for the difference in proportions of high school graduates does not contain 0, there is strong evidence that women are more likely than men to complete high school.

29. Pets.

a) $SE\left(\hat{p}_{Herb} - \hat{p}_{None}\right) = \sqrt{\dfrac{\hat{p}_{Herb}\hat{q}_{Herb}}{n_{Herb}} + \dfrac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}} = \sqrt{\dfrac{\left(\frac{473}{827}\right)\left(\frac{354}{827}\right)}{827} + \dfrac{\left(\frac{19}{130}\right)\left(\frac{111}{130}\right)}{130}} = 0.035$

b) **Randomization condition:** Assume that the dogs studied were representative of all dogs.
10% condition: 827 dogs from homes with herbicide used regularly and 130 dogs from homes with no herbicide used are less than 10% of all dogs.
Independent groups assumption: The samples were drawn independently of each other.
Success/Failure condition: $n\hat{p}$ (herb) = 473, $n\hat{q}$ (herb) = 354, $n\hat{p}$ (none) = 19, and $n\hat{q}$ (none) = 111 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Herb} - \hat{p}_{None}\right) \pm z^* \sqrt{\frac{\hat{p}_{Herb}\hat{q}_{Herb}}{n_{Herb}} + \frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}}$$

$$= \left(\tfrac{473}{827} - \tfrac{19}{130}\right) \pm 1.960\sqrt{\frac{\left(\frac{473}{827}\right)\left(\frac{354}{827}\right)}{827} + \frac{\left(\frac{19}{130}\right)\left(\frac{111}{130}\right)}{130}} = (0.356,\ 0.495)$$

c) We are 95% confident that the proportion of pets with a malignant lymphoma in homes where herbicides are used is between 35.6 and 49.5 percentage points higher than the proportion with lymphoma in homes where no pesticides are used.

30. Carpal Tunnel.

a) $SE\left(\hat{p}_{Surg}-\hat{p}_{Splint}\right)=\sqrt{\frac{\hat{p}_{Surg}\hat{q}_{Surg}}{n_{Surg}}+\frac{\hat{p}_{Splint}\hat{q}_{Splint}}{n_{Splint}}}=\sqrt{\frac{(0.80)(0.20)}{88}+\frac{(0.48)(0.52)}{88}}=0.068$

b) **Randomization condition:** It's not clear whether or not this study was an experiment. If so, assume that the subjects were randomly allocated to treatment groups. If not, assume that the subjects are representative of all carpal tunnel sufferers.
10% condition: 88 subjects in each group are less than 10% of all carpal tunnel sufferers.
Independent groups assumption: The improvement rates of the two groups are not related.
Success/Failure condition: $n\hat{p}$ (surg) = (88)(0.80) = 70, $n\hat{q}$ (surg) = (88)(0.20) = 18, $n\hat{p}$ (splint) = (88)(0.48) = 42, and $n\hat{q}$ (splint) = (88)(0.52) = 46 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Surg}-\hat{p}_{Splint}\right)\pm z^{*}\sqrt{\frac{\hat{p}_{Surg}\hat{q}_{Surg}}{n_{Surg}}+\frac{\hat{p}_{Splint}\hat{q}_{Splint}}{n_{Splint}}}$$
$$=(0.80-0.48)\pm 1.960\sqrt{\frac{(0.80)(0.20)}{88}+\frac{(0.48)(0.52)}{88}}=(0.184,\ 0.452)$$

c) We are 95% confident that the proportion of patients who show improvement in carpal tunnel syndrome with surgery is between 18.4 and 45.2 percentage points higher than the proportion who show improvement with wrist splints.

31. Prostate cancer.

a) This is an experiment. Men were randomly assigned to have surgery or not.

b) **Randomization condition:** Men were randomly assigned to treatment groups.
Independent groups assumption: The survival rates of the two groups are not related.
Success/Failure condition: $n\hat{p}$ (no surgery) = 31 , $n\hat{q}$ (no surgery) = 317, $n\hat{p}$ (surgery) = 16, and $n\hat{q}$ (surgery) = 331 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{None}-\hat{p}_{Surg}\right)\pm z^{*}\sqrt{\frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}+\frac{\hat{p}_{Surg}\hat{q}_{Surg}}{n_{Surg}}}$$
$$=\left(\tfrac{31}{348}-\tfrac{16}{347}\right)\pm 1.960\sqrt{\frac{\left(\frac{31}{348}\right)\left(\frac{317}{348}\right)}{348}+\frac{\left(\frac{16}{347}\right)\left(\frac{331}{347}\right)}{347}}=(0.006,\ 0.080)$$

c) We are 95% confident that the survival rate of patients who have surgery for prostate cancer is between 0.6 and 8.0 percentage points higher than the survival rate of patients who do not have surgery. Since the interval is completely above zero, there is evidence that surgery may be effective in preventing death from prostate cancer.

32. Race and smoking 2013.

a) **Randomization condition:** We will assume that these samples were random.
10% condition: 11.112 and 1916 are both less than 10% of their respective populations.
Independent groups assumption: The samples were taken randomly, so the groups are independent.
Success/Failure condition: $n\hat{p}$ (white) = (11,112)(0.147) = 1633 , $n\hat{q}$ (white) = (11,112)(0.853) = 9479, $n\hat{p}$ (black) = (1916)(0.188) = 360, and $n\hat{q}$ (black) = (1916)(0.812) = 1556 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Black} - \hat{p}_{White}\right) \pm z^* \sqrt{\frac{\hat{p}_{Black}\hat{q}_{Black}}{n_{Black}} + \frac{\hat{p}_{White}\hat{q}_{White}}{n_{White}}}$$
$$= (0.188 - 0.147) \pm 1.645\sqrt{\frac{(0.188)(0.812)}{1916} + \frac{(0.147)(0.853)}{11,112}} = (0.025,\ 0.057)$$

b) We are 90% confident that the smoking rate for blacks is between 2.5 and 5.7 percentage points higher than the smoking rate for whites in New Jersey. We can use this interval to test the hypothesis that there is no difference in smoking rates between blacks and whites in New Jersey. Since the interval does not contains zero, there is evidence of a difference in smoking rates based on race.

c) The interval had 90% confidence, so $\alpha = (1 - 0.90) = 0.10$ for the two-sided test.

33. Ear infections.

a) **Randomization condition:** The babies were randomly assigned to the two treatment groups.
Independent groups assumption: The groups were assigned randomly, so the groups are not related.
Success/Failure condition: $n\hat{p}$ (vaccine) = 333, $n\hat{q}$ (vaccine) = 2122, $n\hat{p}$ (none) = 499, and $n\hat{q}$ (none) = 1953 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b) $\left(\hat{p}_{None} - \hat{p}_{Vacc}\right) \pm z^* \sqrt{\dfrac{\hat{p}_{None}\hat{q}_{None}}{n_{None}} + \dfrac{\hat{p}_{Vacc}\hat{q}_{Vacc}}{n_{Vacc}}}$

$$= \left(\tfrac{499}{2452} - \tfrac{333}{2455}\right) \pm 1.960\sqrt{\frac{\left(\frac{499}{2452}\right)\left(\frac{1953}{2452}\right)}{2452} + \frac{\left(\frac{333}{2455}\right)\left(\frac{2122}{2455}\right)}{2455}} = (0.047,\ 0.089)$$

c) We are 95% confident that the proportion of unvaccinated babies who develop ear infections is between 4.7 and 8.9 percentage points higher than the proportion of vaccinated babies who develop ear infections. The vaccinations appear to be effective, especially considering the 20% infection rate among the unvaccinated. A reduction of 5% to 9% is meaningful.

34. Anorexia.

a) **Randomization condition:** The women were randomly assigned to the groups.
Independent groups assumption: The groups were assigned randomly, so the groups are not related.
Success/Failure condition: $n\hat{p}$ (Prozac) = 35, $n\hat{q}$ (Prozac) = 14, $n\hat{p}$ (placebo) = 32, and $n\hat{q}$ (placebo) = 12 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

b) $\left(\hat{p}_{Proz} - \hat{p}_{Plac}\right) \pm z^* \sqrt{\dfrac{\hat{p}_{Proz}\hat{q}_{Proz}}{n_{Proz}} + \dfrac{\hat{p}_{Plac}\hat{q}_{Plac}}{n_{Plac}}}$

$$= \left(\tfrac{35}{49} - \tfrac{32}{44}\right) \pm 1.960\sqrt{\frac{\left(\frac{35}{49}\right)\left(\frac{14}{49}\right)}{49} + \frac{\left(\frac{32}{44}\right)\left(\frac{12}{44}\right)}{44}} = (-0.20,\ 0.17)$$

c) We are 95% confident that the proportion of women taking Prozac deemed healthy is between 20 percentage points lower and 17 percentage points higher than the proportion women taking a placebo. Prozac does not appear to be effective, since 0 is in the confidence interval. There is no evidence of a difference in the effectiveness of Prozac and a placebo.

35. Another ear infection.

a) H_0: The proportion of vaccinated babies who get ear infections is the same as the proportion of unvaccinated babies who get ear infections.
$\left(p_{Vacc} = p_{None} \text{ or } p_{Vacc} - p_{None} = 0\right)$

H_A: The proportion of vaccinated babies who get ear infections is the lower than the proportion of unvaccinated babies who get ear infections.
$\left(p_{Vacc} < p_{None} \text{ or } p_{Vacc} - p_{None} < 0\right)$

b) Since 0 is not in the confidence interval, reject the null hypothesis. There is evidence that the vaccine reduces the rate of ear infections.

c) If we think that the vaccine really reduces the rate of ear infections and it really does not reduce the rate of ear infections, then we have committed a Type I error.

d) Babies would be given ineffective vaccines.

36. Anorexia again.

a) H_0: The proportion of women taking Prozac who are deemed healthy is the same as the proportion of women taking the placebo who are deemed healthy.
$\left(p_{Prozac} = p_{Placebo} \text{ or } p_{Prozac} - p_{Placebo} = 0\right)$

H_A: The proportion of women taking Prozac who are deemed healthy is greater than the proportion of women taking the placebo who are deemed healthy.
$\left(p_{Prozac} > p_{Placebo} \text{ or } p_{Prozac} - p_{Placebo} > 0\right)$

b) Since 0 is in the confidence interval, fail to reject the null hypothesis. There is no evidence that Prozac is an effective treatment for anorexia.

c) If we think that Prozac is not effective and it is, we have committed a Type II error.

d) We might overlook a potentially helpful treatment.

37. Teen smoking.

a) This is a prospective observational study.

b) H_0: The proportion of teen smokers among the group whose parents disapprove of smoking is the same as the proportion of teen smokers among the group whose parents are lenient about smoking.
$\left(p_{Dis} = p_{Len} \text{ or } p_{Dis} - p_{Len} = 0\right)$

H_A: The proportion of teen smokers among the group whose parents disapprove of smoking is lower than the proportion of teen smokers among the group whose parents are lenient about smoking.
$\left(p_{Dis} < p_{Len} \text{ or } p_{Dis} - p_{Len} < 0\right)$

c) **Randomization condition:** Assume that the teens surveyed are representative of all teens.
10% condition: 284 and 41 are both less than 10% of all teens.
Independent groups assumption: The groups were surveyed independently.
Success/Failure condition: $n\hat{p}$ (disapprove) = 54, $n\hat{q}$ (disapprove) = 230, $n\hat{p}$ (lenient) = 11, and $n\hat{q}$ (lenient) = 30 are all greater than 10, so the samples are both large enough.

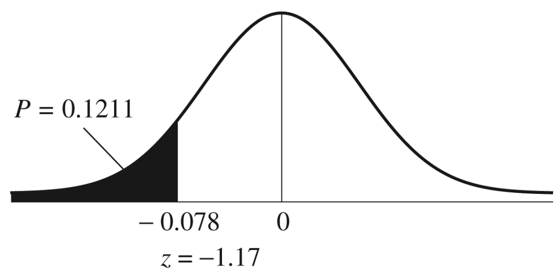
Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{Dis}-\hat{p}_{Len}\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Dis}}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Len}}}=\sqrt{\frac{\left(\frac{65}{325}\right)\left(\frac{260}{325}\right)}{284}+\frac{\left(\frac{65}{325}\right)\left(\frac{260}{325}\right)}{41}}=0.0668\,.$$

d) The observed difference between the proportions is 0.190 – 0.268 = – 0.078.

Since the P-value = 0.1211 is high, we fail to reject the null hypothesis. There is little evidence to suggest that parental attitudes influence teens' decisions to smoke.

$$z=\frac{-0.078-0}{0.0668}$$
$$z\approx-1.17$$



e) If there is no difference in the proportions, there is about a 12% chance of seeing the observed difference or larger by natural sampling variation.

f) If teens' decisions about smoking *are* influenced, we have committed a Type II error.

g) Since the conditions have already been satisfied in a previous exercise, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Dis}-\hat{p}_{Len}\right)\pm z^{*}\sqrt{\frac{\hat{p}_{Dis}\hat{q}_{Dis}}{n_{Dis}}+\frac{\hat{p}_{Len}\hat{q}_{Len}}{n_{Len}}}$$
$$=\left(\tfrac{54}{284}-\tfrac{11}{41}\right)\pm 1.960\sqrt{\frac{\left(\frac{54}{284}\right)\left(\frac{230}{284}\right)}{284}+\frac{\left(\frac{11}{41}\right)\left(\frac{30}{41}\right)}{41}}=\left(-0.221,\,0.065\right)$$

h) We are 95% confident that the proportion of teens whose parents disapprove of smoking who will eventually smoke is between 22.1 percentage points lower and 6.5 percentage points higher than for teens with parents who are lenient about smoking.

i) We expect 95% of random samples of this size to produce intervals that contain the true difference between the proportions.

38. Depression.

a) This is a prospective observational study.

b) H_0: The proportion of cardiac patients without depression who died within the 4 years is the same as the proportion of cardiac patients with depression who died during the same time period.

$$\left(p_{None}=p_{Dep}\ \text{ or }\ p_{None}-p_{Dep}=0\right)$$

H_A : The proportion of cardiac patients without depression who died within the 4 years is the less than the proportion of cardiac patients with depression who died during the same time period.

$$\left(p_{None} < p_{Dep} \text{ or } p_{None} - p_{Dep} < 0\right)$$

c) **Randomization condition:** Assume that the cardiac patients followed by the study are representative of all cardiac patients.
10% condition: 361 and 89 are both less than 10% of all teens.
Independent groups condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (no depression) = 67, $n\hat{q}$ (no depression) = 294, $n\hat{p}$ (depression) = 26, and $n\hat{q}$ (depression) = 63 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

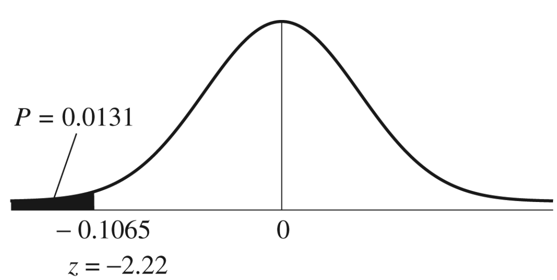
$$SE_{\text{pooled}}\left(\hat{p}_{None} - \hat{p}_{Dep}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{None}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Dep}}} = \sqrt{\frac{\left(\frac{93}{450}\right)\left(\frac{357}{450}\right)}{361} + \frac{\left(\frac{93}{450}\right)\left(\frac{357}{450}\right)}{89}} \approx 0.0479.$$

d) The observed difference between the proportions is: 0.1856 – 0.2921 = – 0.1065.

Since the P-value = 0.0131 is low, we reject the null hypothesis. There is strong evidence to suggest that the proportion of non-depressed cardiac patients who die within 4 years is less than the proportion of depressed cardiac patients who die within 4 years.

$$z = \frac{-0.1065 - 0}{0.0479}$$

$$z \approx -2.22$$



e) If there is no difference in the proportions, we will see an observed difference this large or larger only about 1.3% of the time by natural sampling variation.

f) If cardiac patients without depression don't actually have a lower proportion of deaths than those with depression, then we have committed a Type I error.

g) Since the conditions have already been satisfied in a previous exercise, we will find a two-proportion z-interval.

$$\left(\hat{p}_{None} - \hat{p}_{Dep}\right) \pm z^* \sqrt{\frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}} + \frac{\hat{p}_{Dep}\hat{q}_{Dep}}{n_{Dep}}}$$

$$= \left(\tfrac{67}{361} - \tfrac{26}{89}\right) \pm 1.960\sqrt{\frac{\left(\frac{67}{361}\right)\left(\frac{294}{361}\right)}{361} + \frac{\left(\frac{26}{89}\right)\left(\frac{63}{89}\right)}{89}} = \left(-0.209, -0.004\right)$$

h) We are 95% confident that the proportion of cardiac patients who die within 4 years is between 0.4 and 20.9 percentage points higher for depressed patients than for non-depressed patients.

i) We expect 95% of random samples of this size to produce intervals that contain the true difference between the proportions.

39. Birthweight.

a) H_0: The proportion of low birthweight is the same. $\left(p_{Exp} = p_{Not} \text{ or } p_{Exp} - p_{Not} = 0\right)$

H_A: The proportion of low birthweight is higher for women exposed to soot and ash. $\left(p_{Exp} > p_{Not} \text{ or } p_{Exp} - p_{Not} > 0\right)$

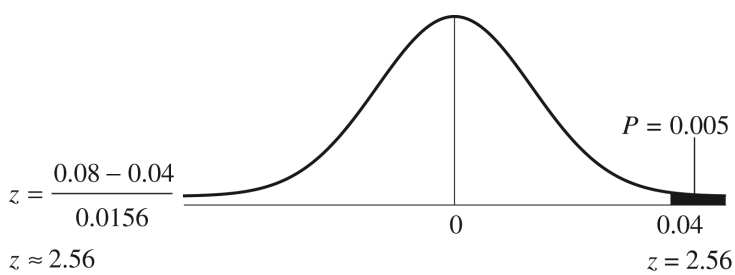
Randomization condition: Assume the women are representative of all women.
10% condition: 182 and 2300 are both less than 10% of all women.
Independent groups assumption: The groups don't appear to be associated, with respect to soot and ash exposure, but all of the women were in New York. There may be a confounding variable explaining any relationship between exposure and birthweight.
Success/Failure condition: $n\hat{p}$ (Exposed) = 15, $n\hat{q}$ (Exposed) = 167, $n\hat{p}$ (Not) = 92, and $n\hat{q}$ (Not) = 2208 are all greater than 10. All of the samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{Exp} - \hat{p}_{Not}\right) = \sqrt{\frac{\hat{p}_p\hat{q}_p}{n_{Exp}} + \frac{\hat{p}_p\hat{q}_p}{n_{Not}}} = \sqrt{\frac{\left(\frac{107}{2482}\right)\left(\frac{2375}{2482}\right)}{182} + \frac{\left(\frac{107}{2482}\right)\left(\frac{2375}{2482}\right)}{2300}} \approx 0.0156.$$

The observed difference between the proportions is: 0.08 – 0.04 = 0.04.

Since the P-value = 0.005 is low, we reject the null hypothesis. There is strong evidence that the proportion of low birthweight babies is higher in the women exposed to soot and ash after the World Trade Center attacks.



b) $\left(\hat{p}_{Exp} - \hat{p}_{Not}\right) \pm z^* \sqrt{\dfrac{\hat{p}_{Exp}\hat{q}_{Exp}}{n_{Exp}} + \dfrac{\hat{p}_{Not}\hat{q}_{Not}}{n_{Not}}}$

$$= \left(\tfrac{15}{182} - \tfrac{92}{2300}\right) \pm 1.960\sqrt{\frac{\left(\tfrac{15}{182}\right)\left(\tfrac{167}{182}\right)}{182} + \frac{\left(\tfrac{92}{2300}\right)\left(\tfrac{2208}{2300}\right)}{2300}} = (0.002,\ 0.083)$$

We are 95% confident that the proportion of low birthweight babies is between 0.2 and 8.3 percentage points higher for mothers exposed to soot and ash after the World Trade Center attacks, than the proportion of low birthweight babies for mothers not exposed.

40. Politics and sex.

a) H_0: The proportion of voters in support of the candidate is the same before and after news of his extramarital affair got out. $(p_B = p_A \text{ or } p_B - p_A = 0)$

H_A: The proportion of voters in support of the candidate has decreased after news of his extramarital affair got out. $(p_B > p_A \text{ or } p_B - p_A > 0)$

Randomization condition: Voters were randomly selected.
10% condition: 630 and 1010 are both less than 10% of all voters.
Independent groups assumption: Since the samples were random, the groups are independent.
Success/Failure condition: $n\hat{p}$ (before) = (630)(0.54) = 340, $n\hat{q}$ (before) = (630)(0.46) = 290, $n\hat{p}$ (after) = (1010)(0.51) = 515, and $n\hat{q}$ (after) = (1010)(0.49) = 505 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_B - \hat{p}_A\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_B} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_A}}$$

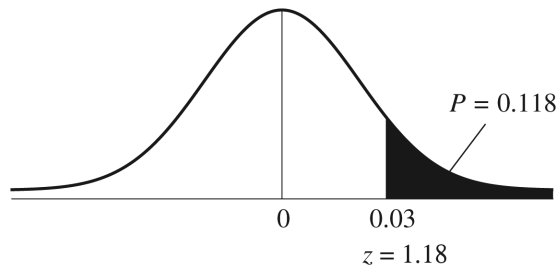
$$= \sqrt{\frac{(0.5215)(0.4785)}{630} + \frac{(0.5215)(0.4785)}{1010}} \approx 0.02536.$$

The observed difference between the proportions is: 0.54 – 0.51 = 0.03.

Since the P-value = 0.118 is fairly high, we fail to reject the null hypothesis. There is little evidence of a decrease in the proportion of voters in support of the candidate after the news of his extramarital affair got out.

$$z = \frac{0.03 - 0}{0.02536}$$

$$z \approx 1.18$$



b) No evidence of a decrease in the proportion of voters in support of the candidate was found. If there is actually a decrease, and we failed to notice, that's a Type II error.

41. Mammograms.

a) H_0: The proportion of deaths from breast cancer is the same for women who never had a mammogram as for women who had mammograms.
$\left(p_N = p_M \text{ or } p_N - p_M = 0\right)$

H_A: The proportion of deaths from breast cancer is greater for women who never had a mammogram than for women who had mammograms.
$\left(p_N > p_M \text{ or } p_N - p_M > 0\right)$

Randomization condition: Assume the women are representative of all women.
10% condition: 30,565 and 30,131 are both less than 10% of all women.
Independent groups assumption: We must assume that these groups are independent.
Success/Failure condition: $n\hat{p}$ (never) = 196, $n\hat{q}$ (never) = 30,369, $n\hat{p}$ (mammogram) = 153, and $n\hat{q}$ (mammogram) = 29,978 are all greater than 10, so both samples are large enough.

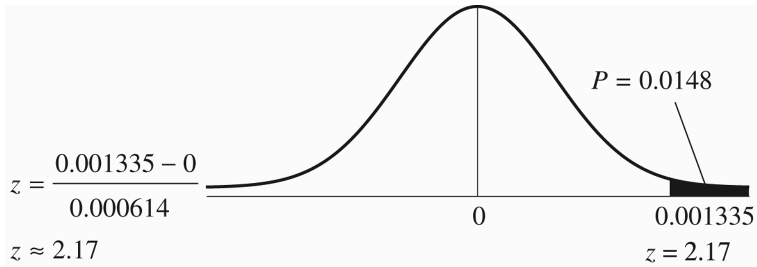
Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_N - \hat{p}_M\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_M}}$$

$$= \sqrt{\frac{\left(\frac{349}{60,696}\right)\left(\frac{60,347}{60,696}\right)}{30,565} + \frac{\left(\frac{349}{60,696}\right)\left(\frac{60,347}{60,696}\right)}{30,131}} \approx 0.000614$$

The observed difference between the proportions is:
0.006413 – 0.005078 = 0.001335.

Since the P-value = 0.0148 is low, we reject the null hypothesis. There is strong evidence that the breast cancer mortality rate is higher among women that have never had a mammogram. However, the large sample sizes involved may have yielded a result that is statistically significant, but not practically significant. These data suggest a difference in mortality rate of only about 0.1 percentage points.



b) If there is actually no difference in the mortality rates, we have committed a Type I error.

42. Mammograms redux.

a) H_0: The proportion of deaths from breast cancer is the same for women who never had a mammogram as for women who had mammograms.
$(p_N = p_M \text{ or } p_N - p_M = 0)$

H_A: The proportion of deaths from breast cancer is greater for women who never had a mammogram than for women who had mammograms.
$(p_N > p_M \text{ or } p_N - p_M > 0)$

Randomization condition: Assume the women are representative of all women.
10% condition: 21,195 and 21,088 are both less than 10% of all women.
Independent groups assumption: We must assume that these groups are independent.
Success/Failure condition: $n\hat{p}$ (never) = 66, $n\hat{q}$ (never) = 21,129, $n\hat{p}$ (mammogram) = 63, and $n\hat{q}$ (mammogram) = 21,025 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}(\hat{p}_N - \hat{p}_M) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_M}}$$

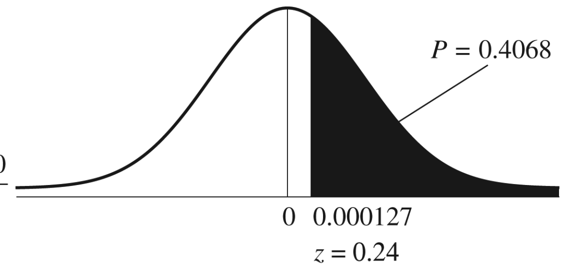
$$= \sqrt{\frac{\left(\frac{129}{42,283}\right)\left(\frac{42,154}{42,283}\right)}{21,195} + \frac{\left(\frac{129}{42,283}\right)\left(\frac{42,154}{42,283}\right)}{21,088}} \approx 0.000536$$

The observed difference between the proportions is:
0.003114 – 0.002987 = 0.000127.

Since the P-value = 0.4068 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of breast cancer deaths for women who have never had a mammogram is greater than the proportion of deaths from breast cancer for women who underwent screening by mammogram.

$$z = \frac{0.000127 - 0}{0.000536}$$

$$z \approx 0.24$$



b) If the proportion of deaths from breast cancer for women who have not had mammograms is actually greater than the proportion of deaths from breast cancer for women who have had mammograms, we have committed a Type II error.

43. Pain.

a) **Randomization condition:** The patients were randomly selected AND randomly assigned to treatment groups. If that's not random enough for you, I don't know what is!
10% condition: 112 and 108 are both less than 10% of all people with joint pain.
Success/Failure condition: $n\hat{p}$ (A) = 84, $n\hat{q}$ (A) = 28, $n\hat{p}$ (B) = 66, and $n\hat{q}$ (B) = 42 are all greater than 10, so both samples are large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of patients who may get relief from medication A.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{84}{112}\right) \pm 1.960\sqrt{\frac{\left(\frac{84}{112}\right)\left(\frac{28}{112}\right)}{112}} = (67.0\%, 83.0\%)$$

We are 95% confident that between 67.0% and 83.0% of patients with joint pain will find medication A to be effective.

b) Since the conditions were met in part a, we can use a one-proportion z-interval to estimate the percentage of patients who may get relief from medication B.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{66}{108}\right) \pm 1.960\sqrt{\frac{\left(\frac{66}{108}\right)\left(\frac{42}{108}\right)}{108}} = (51.9\%, 70.3\%)$$

We are 95% confident that between 51.9% and 70.3% of patients with joint pain will find medication B to be effective.

c) The 95% confidence intervals overlap, which might lead one to believe that there is no evidence of a difference in the proportions of people who find each medication effective. However, if one was lead to believe that, one should proceed to part…

d) Most of the conditions were checked in part a. We only have one more to check:
Independent groups assumption: The groups were assigned randomly, so there is no reason to believe there is a relationship between them.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_A - \hat{p}_B\right) \pm z^* \sqrt{\frac{\hat{p}_A\hat{q}_A}{n_A} + \frac{\hat{p}_B\hat{q}_B}{n_B}}$$

$$= \left(\tfrac{84}{112} - \tfrac{66}{108}\right) \pm 1.960\sqrt{\frac{\left(\frac{84}{112}\right)\left(\frac{28}{112}\right)}{112} + \frac{\left(\frac{66}{108}\right)\left(\frac{42}{108}\right)}{112}} = \left(0.017,\ 0.261\right)$$

We are 95% confident that the proportion of patients with joint pain who will find medication A effective is between 1.70 and 26.1 percentage points higher than the proportion of patients who will find medication B effective.

e) The interval does not contain zero. There is evidence that medication A is more effective than medication B.

f) The two-proportion method is the proper method. By attempting to use two, separate, confidence intervals, you are adding standard deviations when looking for a difference in proportions. We know from our previous studies that *variances* add when finding the standard deviation of a difference. The two-proportion method does this.

44. Gender gap.

a) **Randomization condition:** The poll was probably random, although not specifically stated.
10% condition: 473 and 522 are both less than 10% of all voters.
Success/Failure condition: $n\hat{p}$ (men) = 246, $n\hat{q}$ (men) = 227, $n\hat{p}$ (women) = 235, and $n\hat{q}$ (women) = 287 are all greater than 10, so both samples are large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of men who may vote for the candidate.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(0.52\right) \pm 1.960\sqrt{\frac{\left(0.52\right)\left(0.48\right)}{473}} = (47.5\%,\ 56.5\%)$$

We are 95% confident that between 47.5% and 56.5% of men may vote for the candidate.

b) Since the conditions were met in part a, we can use a one-proportion z-interval to estimate the percentage of women who may vote for the candidate.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(0.45\right) \pm 1.960\sqrt{\frac{\left(0.45\right)\left(0.55\right)}{522}} = (40.7\%,\ 49.3\%)$$

We are 95% confident that between 40.7% and 49.3% of women may vote for the candidate.

c) The 95% confidence intervals overlap, which might make you think that there is no evidence of a difference in the proportions of men and women who may vote for the candidate. However, if you think that, don't delay! Move on to part…

d) Most of the conditions were checked in part a. We only have one more to check:
Independent groups assumption: There is no reason to believe that the samples of men and women influence each other in any way.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_M - \hat{p}_W\right) \pm z^* \sqrt{\frac{\hat{p}_M\hat{q}_M}{n_M} + \frac{\hat{p}_W\hat{q}_W}{n_W}}$$
$$= \left(0.52 - 0.45\right) \pm 1.960\sqrt{\frac{(0.52)(0.48)}{473} + \frac{(0.45)(0.55)}{522}} = \left(0.008,\ 0.132\right)$$

We are 95% confident that the proportion of men who may vote for the candidate is between 0.8 and 13.2 percentage points higher than the proportion of women who may vote for the candidate.

e) The interval does not contain zero. There is evidence that the proportion of men may vote for the candidate is greater than the proportion of women who may vote for the candidate.

f) The two-proportion method is the proper method. By attempting to use two, separate, confidence intervals, you are adding standard deviations when looking for a difference in proportions. We know from our previous studies that *variances* add when finding the standard deviation of a difference. The two-proportion method does this.

45. Convention bounce.

Randomization condition: The polls were conducted randomly.
10% condition: 1500 and 1500 are both less than 10% of all voters.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (after) = 705, $n\hat{q}$ (after) = 795, $n\hat{p}$ (before) = 735, and $n\hat{q}$ (women) = 765 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{After} - \hat{p}_{Before}\right) \pm z^* \sqrt{\frac{\hat{p}_{After}\hat{q}_{After}}{n_{After}} + \frac{\hat{p}_{Before}\hat{q}_{Before}}{n_{Before}}}$$
$$= \left(0.49 - 0.47\right) \pm 1.960\sqrt{\frac{(0.49)(0.51)}{1500} + \frac{(0.47)(0.53)}{1500}} = \left(-0.016,\ 0.056\right)$$

We can be 95% confident that the proportion of likely voters who favored John Kerry was between 1.6 percentage points lower and 5.6 percentage points higher after the convention, compared to before the convention. Since zero is contained in the interval, it is plausible that there was no difference in Kerry support. The poll showed no evidence of a convention bounce.

46. Stay-at-home dads.

a) **Randomization condition:** We will assume that the polls were conducted randomly.
10% condition: 161 and 358 are both less than 10% of all black and Latino men, respectively.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (black) = 11, $n\hat{q}$ (black) = 150, $n\hat{p}$ (Latino) = 20, and $n\hat{q}$ (Latino) = 338 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{Black} - \hat{p}_{Latino}\right) \pm z^* \sqrt{\frac{\hat{p}_{Black}\hat{q}_{Black}}{n_{Black}} + \frac{\hat{p}_{Latino}\hat{q}_{Latino}}{n_{Latino}}}$$

$$= \left(\tfrac{11}{161} - \tfrac{20}{358}\right) \pm 1.960 \sqrt{\frac{\left(\frac{11}{161}\right)\left(\frac{150}{161}\right)}{161} + \frac{\left(\frac{20}{358}\right)\left(\frac{338}{358}\right)}{358}} = \left(-0.033,\ 0.058\right)$$

We are 95% confident that the percentage of black men who are stay-at-home dads is between 3.3 percentage points lower and 5.8 percentage points higher than the percentage of Latino men who are stay-at-home dads.

b) The margin of error is higher for the interval we constructed because the sample size is smaller than the entire *Time* magazine survey. Our interval only involved black and Latino men.

47. Sensitive men.

H_0: The proportion of 18-24-year-old men who are comfortable talking about their problems is the same as the proportion of 25-34-year old men.
$\left(p_{Young} = p_{Old} \text{ or } p_{Young} - p_{Old} = 0\right)$

H_A: The proportion of 18-24-year-old men who are comfortable talking about their problems is higher than the proportion of 25-34-year old men.
$\left(p_{Young} > p_{Old} \text{ or } p_{Young} - p_{Old} > 0\right)$

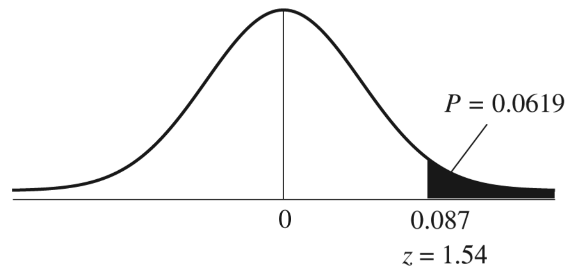
Randomization condition: We assume the respondents were chosen randomly.
10% condition: 129 and 184 are both less than 10% of all people.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (young) = 80, $n\hat{q}$ (young) = 49, $n\hat{p}$ (old) = 98, and $n\hat{q}$ (old) = 86 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated
by:.

$$SE_{\text{pooled}}\left(\hat{p}_{Young}-\hat{p}_{Old}\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_Y}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_O}}=\sqrt{\frac{\left(\frac{178}{313}\right)\left(\frac{135}{313}\right)}{129}+\frac{\left(\frac{178}{313}\right)\left(\frac{135}{313}\right)}{184}}\approx 0.05687$$

The observed difference between the proportions is:
0.620 – 0.533 = 0.087.

$$z=\frac{0.087-0}{0.05687}$$
$$z\approx 1.54$$



Since the P-value = 0.0619 is high, we fail to reject the null hypothesis. There is little evidence that the proportion of 18-24-year-old men who are comfortable talking about their problems is higher than the proportion of 25-34-year-old men who are comfortable. *Time* magazine's interpretation is questionable.

48. Carbs.

H_0: The proportion of U.S. adults who actively avoid carbohydrates in their diet is the same now as it was in 2002. $\left(p_{Now}=p_{2002}\ \text{ or }\ p_{Now}-p_{2002}=0\right)$

H_A: The proportion of U.S. adults who actively avoid carbohydrates in their diet is greater now than it was in 2002. $\left(p_{Now}>p_{2002}\ \text{ or }\ p_{Now}-p_{2002}>0\right)$

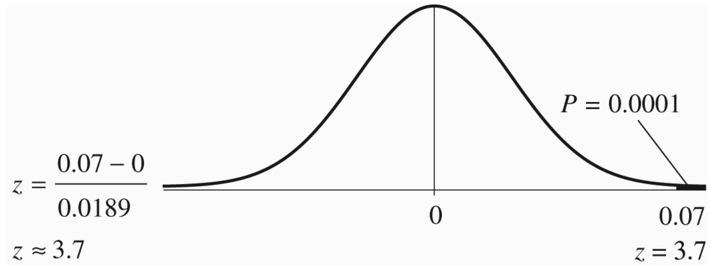
Randomization condition: We assume the respondents were chosen randomly.
10% condition: 1005 and 1005 are both less than 10% of all people.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (now) = 271, $n\hat{q}$ (now) = 734, $n\hat{p}$ (2002) = 201, and $n\hat{q}$ (2002) = 804 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_{Now}-\hat{p}_{2002}\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Now}}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{2002}}}=\sqrt{\frac{\left(\frac{472}{2010}\right)\left(\frac{1538}{2010}\right)}{1005}+\frac{\left(\frac{472}{2010}\right)\left(\frac{1538}{2010}\right)}{1005}}\approx 0.0189$$

The observed difference between the proportions is: 0.27 – 0.20 = 0.07.

Since the P-value = 0.0001 is low, we reject the null hypothesis. There is strong evidence that the percentage of U.S. adults who actively try to avoid carbs has increased since 2002.



49. Food preference.

H_0: The proportion of people who agree with the statement is the same in rural and urban areas. $\left(p_{Urban} = p_{Rural} \text{ or } p_{Urban} - p_{Rural} = 0\right)$

H_A: The proportion of people who agree with the statement differs in rural and urban areas. $\left(p_{Urban} \neq p_{Rural} \text{ or } p_{Urban} - p_{Rural} \neq 0\right)$

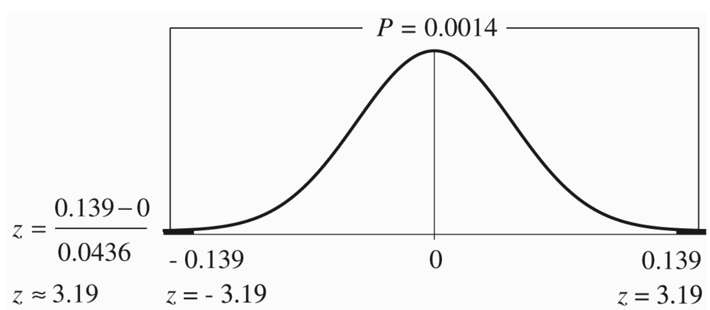
Randomization condition: The respondents were chosen randomly.
10% condition: 646 and 154 are both less than 10% of all urban and rural people, respectively.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (urban) = 417, $n\hat{q}$ (urban) = 229, $n\hat{p}$ (rural) = 78, and $n\hat{q}$ (rural) = 76 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_{Urban} - \hat{p}_{Rural}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Urban}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Rural}}} = \sqrt{\frac{\left(\frac{495}{800}\right)\left(\frac{305}{800}\right)}{646} + \frac{\left(\frac{495}{800}\right)\left(\frac{305}{800}\right)}{154}} \approx 0.0436$$

The observed difference between the proportions is: $\frac{417}{646} - \frac{78}{154} \approx 0.139$

Since the P-value = 0.0014 is low, reject the null hypothesis. There is evidence that the proportion of people who agree with the statement is not the same in urban and rural areas. These data suggest that the proportion is higher in urban areas than in rural areas.



50. Fast food.

H_0: The proportion of people who agree with the statement is the same for the two age groups. $(p_{>35} = p_{\le 35} \text{ or } p_{>35} - p_{\le 35} = 0)$

H_A: The proportion of people who agree with the statement is different for the two age groups. $(p_{>35} \ne p_{\le 35} \text{ or } p_{>35} - p_{\le 35} \ne 0)$

Randomization condition: The respondents were chosen randomly.
10% condition: 389 and 411 are both less than 10% of all people in these age groups, respectively.
Independent groups assumption: The groups were chosen independently.
Success/Failure condition: $n\hat{p}$ (>35) = 246, $n\hat{q}$ (>35) = 143, $n\hat{p}$ (≤ 35) = 197, and $n\hat{q}$ (≤ 35) = 214 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}(\hat{p}_{>35} - \hat{p}_{\le 35}) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{>35}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{\le 35}}} = \sqrt{\frac{\left(\frac{443}{800}\right)\left(\frac{357}{800}\right)}{389} + \frac{\left(\frac{443}{800}\right)\left(\frac{357}{800}\right)}{411}} \approx 0.0352$$

The observed difference between the proportions is: $\frac{246}{389} - \frac{197}{411} \approx 0.153$

$$z = \frac{0.153 - 0}{0.0352}$$
$$z \approx 4.35$$

Since the P-value = 0.00001 is low, reject the null hypothesis. There is evidence that the proportion of people who agree with the statement is not the same for the two age groups. These data suggest that a greater percentage of those 35 or older will say they avoid fast foods.

51. Hot dogs.

Yes, the 95% confidence interval would contain 0. The high P-value means that we lack evidence of a difference, so 0 is a possible value for $\mu_{Meat} - \mu_{Beef}$.

52. Washers.

Yes, the 95% confidence interval would contain 0. The high P-value means that we lack evidence of a difference, so 0 is a possible value for $\mu_{Top} - \mu_{Front}$.

53. Hot dogs, second helping.

a) Plausible values for $\mu_{Meat} - \mu_{Beef}$ are all negative, so the mean fat content is probably higher for beef hot dogs.

b) The fact that the confidence interval does not contain 0 indicates that the difference is significant.

c) The corresponding alpha level is 10%.

54. Second load of wash.

a) Plausible values for $\mu_{Top} - \mu_{Front}$ are all negative, so the mean cycle time is probably higher for front loading machines.

b) The fact that the confidence interval does not contain 0 indicates that the difference is significant.

c) The corresponding alpha level is 2%.

55. Hot dogs, last one.

a) False. The confidence interval is about means, not about individual hot dogs.

b) False. The confidence interval is about means, not about individual hot dogs.

c) True.

d) False. Confidence intervals based on other samples will also try to estimate the true difference in population means. There's not reason to expect other samples to conform to this result.

e) True.

56. Third load of wash.

a) False. The confidence interval is about means, not about individual loads.

b) False. The confidence interval is about means, not about individual loads.

c) False. Confidence intervals based on other samples will also try to estimate the true difference in population means. There's not reason to expect other samples to conform to this result.

d) True.

e) True.

57. Learning math.

a) The margin of error of this confidence interval is (11.427 – 5.573)/2 = 2.927 points.

b) The margin of error for a 98% confidence interval would have been larger. The critical value of t^* is larger for higher confidence levels. We need a wider interval to increase the likelihood that we catch the true mean difference in test scores within our interval. In other words, greater confidence comes at the expense of precision.

c) We are 95% confident that the mean score for the CPMP math students will be between 5.573 and 11.427 points higher on this assessment than the mean score of the traditional students.

d) Since the entire interval is above 0, there is strong evidence that students who learn with CPMP will have higher mean scores is algebra than those in traditional programs.

58. Stereograms test.

a) We are 90% confident that the mean time required to "fuse" the image for people who receive no information or verbal information only will be between 0.55 and 5.47 seconds longer than the mean time required to "fuse" the image for people who receive both verbal and visual information.

b) Since the entire interval is above 0, there is evidence that viewing the picture of the image helps people "see" the 3D image.

c) The margin of error for this interval is (5.47 – 0.55)/2 = 2.46 seconds.

d) 90% of all random samples of this size will produce intervals that will contain the true value of the mean difference between the times of the two groups.

e) A 99% confidence interval would be wider. The critical value of t^* is larger for higher confidence levels. We need a wider interval to increase the likelihood that we catch the true mean difference in test scores within our interval. In other words, greater confidence comes at the expense of precision.

f) The conclusion reached may very well change. A wider interval may contain the mean difference of 0, failing to provide evidence of a difference in mean times.

59. CPMP, again.

a) H_0: The mean score of CPMP students is the same as the mean score of traditional students. $\left(\mu_C = \mu_T \text{ or } \mu_C - \mu_T = 0\right)$

H_A: The mean score of CPMP students is different from the mean score of traditional students. $\left(\mu_C \neq \mu_T \text{ or } \mu_C - \mu_T \neq 0\right)$

b) **Independent groups assumption:** Scores of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to either CPMP or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, we can use a two-sample *t*-test with 583 degrees of freedom (from the computer).

c) If the mean scores for the CPMP and traditional students are really equal, there is less than a 1 in 10,000 chance of seeing a difference as large or larger than the observed difference just from natural sampling variation.

d) Since the P-value < 0.0001, reject the null hypothesis. There is strong evidence that the CPMP students have a different mean score than the traditional students. The evidence suggests that the CPMP students have a higher mean score.

60. CPMP and word problems.

H_0: The mean score of CPMP students is the same as the mean score of traditional students. $(\mu_C = \mu_T \text{ or } \mu_C - \mu_T = 0)$

H_A: The mean score of CPMP students is different from the mean score of traditional students. $(\mu_C \neq \mu_T \text{ or } \mu_C - \mu_T \neq 0)$

Independent groups assumption: Scores of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to either CPMP or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 590.05 degrees of freedom (from the approximation formula).

We will perform a two-sample t-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_C - \bar{y}_T) = \sqrt{\frac{32.1^2}{320} + \frac{28.5^2}{273}} \approx 2.489$.

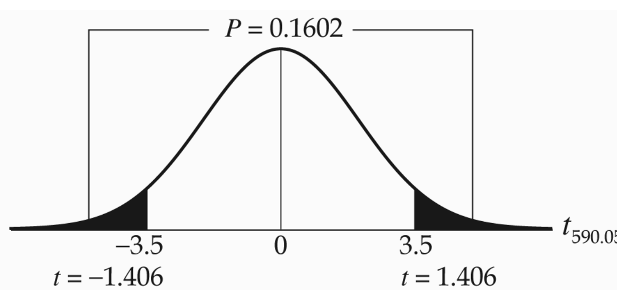
The observed difference between the mean scores is 57.4 – 53.9 = 3.5.

Since the P-value = 0.1602, we fail to reject the null hypothesis. There is no evidence that the CPMP students have a different mean score on the word problems test than the traditional students.

$$t = \frac{(\bar{y}_C - \bar{y}_T) - (0)}{SE(\bar{y}_C - \bar{y}_T)}$$

$$t \approx \frac{3.5}{2.489}$$

$$t \approx 1.406$$



61. Commuting.

a) **Independent groups assumption:** Since the choice of route was determined at random, the commuting times for Route A are independent of the commuting times for Route B.
Randomization condition: The man randomly determined which route he would travel on each day.
Nearly Normal condition: The histograms of travel times for the routes are roughly unimodal and symmetric. (Given)

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 33.1 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\overline{y}_B - \overline{y}_A) \pm t^*_{df}\sqrt{\frac{s_B^2}{n_B} + \frac{s_A^2}{n_A}} = (43 - 40) \pm t^*_{33.1}\sqrt{\frac{2^2}{20} + \frac{3^2}{20}} \approx (1.36,\ 4.64)$$

We are 95% confident that Route B has a mean commuting time between 1.36 and 4.64 minutes longer than the mean commuting time of Route A.

b) Since 5 minutes is beyond the high end of the interval, there is no evidence that the Route B is an average of 5 minutes longer than Route A. It appears that the old-timer may be exaggerating the average difference in commuting time.

62. Pulse rates.

a) The boxplots suggest that the mean pulse rates for men and women are roughly equal, but that females' pulse rates are more variable.

b) **Independent groups assumption:** There is no reason to believe that the pulse rates for men and women are related.
Randomization condition: There is no mention of randomness, but we can assume that the researcher chose a representative sample of men and women with regards to pulse rate.

Nearly Normal condition: The boxplots are reasonably symmetric. Let's hope the distributions of the samples are unimodal, too.

The conditions for inference are satisfied, so we can analyze these data using the methods discussed in this chapter.

c) Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 40.2 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 90% confidence.

$$(\bar{y}_M - \bar{y}_F) \pm t^*_{df}\sqrt{\frac{s_M^2}{n_M} + \frac{s_F^2}{n_F}}$$

$$= (72.75 - 72.625) \pm t^*_{40.2}\sqrt{\frac{5.37225^2}{28} + \frac{7.69987^2}{24}} \approx (-3.025,\ 3.275)$$

We are 90% confident that the mean pulse rate for men is between 3.025 points lower and 3.275 points higher than the mean pulse rate for women.

d) Since 0 is in the interval, there is no evidence of a difference in mean pulse rate for men and women. This confirms our answer to part a.

63. View of the water.

Independent groups assumption: Since the 170 properties were randomly selected, the groups should be independent.
Randomization condition: The 170 properties were selected randomly.
Nearly Normal condition: The boxplots of sale prices are roughly symmetric. The plots show several outliers, but the sample sizes are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 105.48 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_W - \bar{y}_N) \pm t^*_{df}\sqrt{\frac{s_W^2}{n_W} + \frac{s_N^2}{n_N}}$$

$$= (319{,}906.40 - 219{,}896.60) \pm t^*_{105.48}\sqrt{\frac{153{,}303.80^2}{70} + \frac{94{,}627.15^2}{100}}$$

$$\approx (\$59121,\ \$140898)$$

We are 95% confident that waterfront property has a mean selling price that is between $59,121 and $140,898 higher, on average, than non-waterfront property.

64. New construction.

Independent groups assumption: Since the 200 properties were randomly selected, the groups should be independent.
Randomization condition: The 200 properties were selected randomly.
Nearly Normal condition: The boxplots of sale prices are roughly symmetric. The plots show several outliers, but the sample sizes are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 197.8 degrees of freedom (from the approximation formula). We will construct a two-sample t-interval, with 95% confidence.

$$(\bar{y}_{New} - \bar{y}_{Old}) \pm t^*_{df} \sqrt{\frac{s^2_{New}}{n_{New}} + \frac{s^2_{Old}}{n_{Old}}}$$

$$= (267{,}878.10 - 201{,}707.50) \pm t^*_{197.8} \sqrt{\frac{93{,}302.18^2}{100} + \frac{96{,}116.88^2}{100}}$$

$$\approx (\$39754,\ \$92587)$$

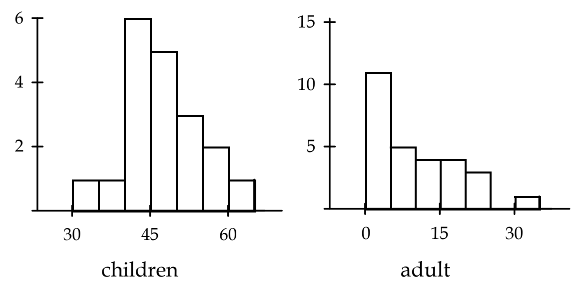
We are 95% confident that newly-constructed property has a mean selling price that is between $39,754 and $92,587 higher, on average, than property that not newly-constructed.

65. Cereal.

Independent groups assumption: The percentage of sugar in the children's cereals is unrelated to the percentage of sugar in adult's cereals.

Randomization condition: It is reasonable to assume that the cereals are representative of all children's cereals and adult cereals, in regard to sugar content.

Nearly Normal condition: The histogram of adult cereal sugar content is skewed to the right, but the sample sizes are of reasonable size. The Central Limit Theorem allows us to proceed.



Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 42 degrees of freedom (from the approximation formula). We will construct a two-sample t-interval, with 95% confidence.

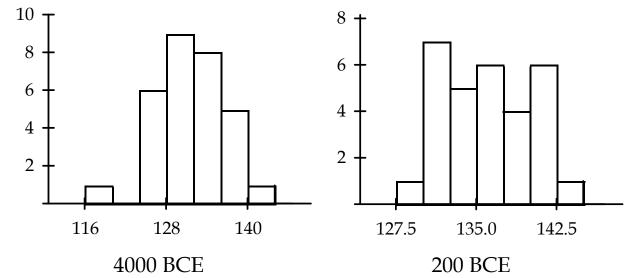
$$(\bar{y}_C - \bar{y}_A) \pm t^*_{df} \sqrt{\frac{s^2_C}{n_C} + \frac{s^2_A}{n_A}} = (46.8 - 10.1536) \pm t^*_{42} \sqrt{\frac{6.41838^2}{19} + \frac{7.61239^2}{28}} \approx (32.49,\ 40.80)$$

We are 95% confident that children's cereals have a mean sugar content that is between 32.49% and 40.80% higher than the mean sugar content of adult cereals.

66. Egyptians.

a) Independent groups assumption: The skull breadth of Egyptians in 4000 B.C.E is independent of the skull breadth of Egyptians almost 4 millennia later!

Randomization condition: It is reasonable to assume that the skulls measured have skull breadths that are representative of all Egyptians of the time.



Nearly Normal condition: The histograms of skull breadths are both unimodal and symmetric.

b) Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 54 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_{200} - \bar{y}_{4K}) \pm t^*_{df}\sqrt{\frac{s^2_{200}}{n_{200}} + \frac{s^2_{4K}}{n_{4K}}}$$

$$= (135.633 - 131.367) \pm t^*_{54}\sqrt{\frac{4.03846^2}{30} + \frac{5.12925^2}{30}} \approx (1.88,\ 6.66)$$

We are 95% confident that Egyptian males in 200 B.C.E. had a mean skull breadth between 1.88 and 6.66 mm larger than the mean skull breadth of Egyptian males in 4000 B.C.E.

c) Since the interval is completely above 0, there is evidence that the mean breadth of males' skulls has changed over this time period. The evidence suggests that the mean skull breadth has increased.

d) The two samples of skulls are independent, the 200 B.C.E sample contains the largest measurement, and the 4000 B.C.E sample contains the smallest measurement, so we can use Tukey's test. There are 10 values that exceed the maximum and minimum values in the other group, and two ties, for a total of 11. With a *P*-value between 0.01 and 0.001, these data provide evidence that the mean skull breadth of Egyptian males in 200 B.C.E exceeds that of Egyptian males in 4000 B.C.E.

e) The Rank-sum test has a *P*-value of 0.0014, which is low, so these data provide evidence that the mean skull breadth of Egyptian males in 200 B.C.E exceeds that of Egyptian males in 4000 B.C.E.

67. Reading.

H$_0$: The mean reading comprehension score of students who learn by the new method is the same as the mean score of students who learn by traditional methods. $(\mu_N = \mu_T \;\text{ or }\; \mu_N - \mu_T = 0)$

H$_A$: The mean reading comprehension score of students who learn by the new method is greater than the mean score of students who learn by traditional methods. $(\mu_N > \mu_T \;\text{ or }\; \mu_N - \mu_T > 0)$

Independent groups assumption: Student scores in one group should not have an impact on the scores of students in the other group.
Randomization condition: Students were randomly assigned to classes.
Nearly Normal condition: The stem-and-leaf plots show distributions of scores that are unimodal and symmetric.

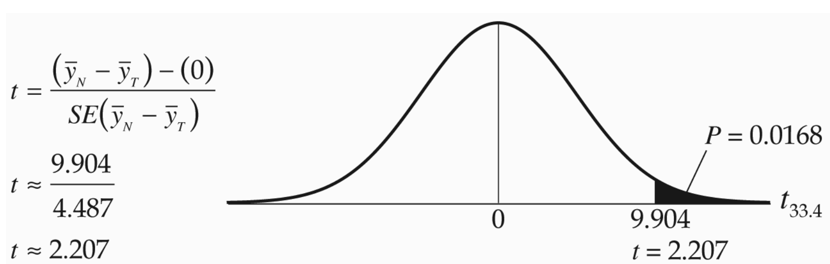
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 33.4 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test. We know:

$$\bar{y}_N = 51.7222 \qquad \bar{y}_T = 41.8182$$
$$s_N = 11.7062 \qquad s_T = 16.5979$$
$$n_N = 18 \qquad n_T = 22$$

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_N - \bar{y}_T) = \sqrt{\frac{11.7062^2}{18} + \frac{16.5979^2}{22}} \approx 4.487.$$

The observed difference between the mean scores is $51.7222 - 41.8182 \approx 9.904$.



Since the *P*-value = 0.0168 is low, we reject the null hypothesis. There is evidence that the students taught using the new activities have a higher mean score on the reading comprehension test than the students taught using traditional methods.

The control group contains both the highest and lowest scores, so Tukey's test is not appropriate. The Rank-sum test gives a *P*-value of 0.0045, so we reject the null hypothesis. There is evidence that the students taught using the new activities have a higher mean score on the reading comprehension test than the students taught using traditional methods.

68. Streams.

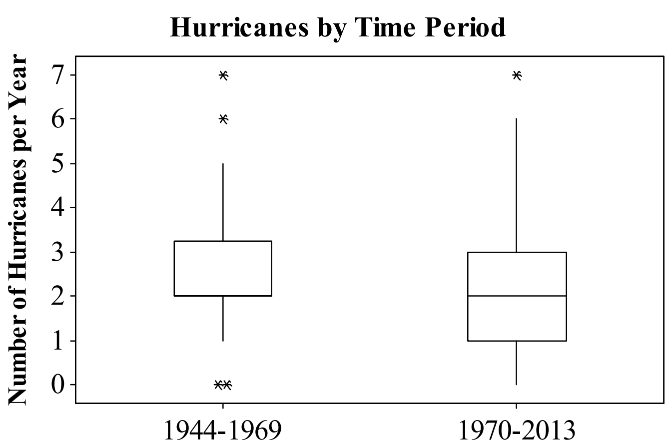
a) H_0: Streams with limestone substrates and streams with shale substrates have the same mean pH level. $\left(\mu_L = \mu_S \text{ or } \mu_L - \mu_S = 0\right)$

H_A: Streams with limestone substrates and streams with shale substrates have different mean pH levels. $\left(\mu_L \neq \mu_S \text{ or } \mu_L - \mu_S \neq 0\right)$

b) **Independent groups assumption:** pH levels from the two types of streams are independent.
Independence assumption: Since we don't know if the streams were chosen randomly, assume that the pH level of one stream does not affect the pH of another stream. This seems reasonable.
Nearly Normal condition: The boxplots provided show that the pH levels of the streams may be skewed (since the median is either the upper or lower quartile for the shale streams and the lower whisker of the limestone streams is stretched out), and there are outliers. However, since there are 133 degrees of freedom, we know that the sample sizes are large. It should be safe to proceed.

c) Since the P-value $\leq$ 0.0001 is low, we reject the null hypothesis. There is strong evidence that the streams with limestone substrates have mean pH levels different than those of streams with shale substrates. The limestone streams are less acidic on average.

69. Hurricanes 2013.



There are several concerns here. First, we don't have a random sample. We have to assume that the actual number of hurricanes in a given year is a random sample of the hurricanes that might occur under similar weather conditions. Also, the data for 1944–1969 are not symmetric and have four outliers. The outliers will tend to make the average for the period 1944–1969 higher. These data are not appropriate for inference. The boxplots provide little evidence of a change in the mean number of hurricanes in the two periods.

70. Memory.

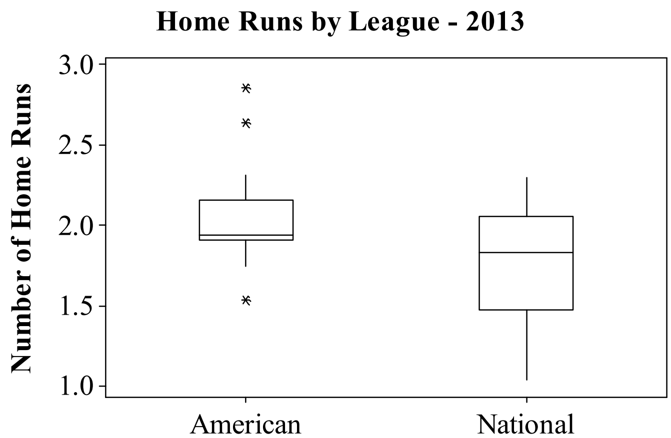
a) If there is no difference between ginkgo and the placebo, there is a 93.74% chance of seeing a difference as large as or larger than that observed, just from natural sampling variation.

b) There is no evidence based on this study that ginkgo biloba improves memory, as the difference in mean memory score was not significant and, in fact, negative.

c) If we fail to notice the effectiveness of ginkgo biloba, we have committed a Type II error.

71. Home runs 2013.

a) The boxplots of the average number of home runs hit at the ballparks in the two leagues are at the right. Both distributions appear at least roughly symmetric, with roughly the same center, around 2 home runs. The distribution of average number of home runs hit is more spread out for the National League. There are outliers in the American League's distribution. (There may be only a high outlier, depending on how quartiles are calculated.)



b) $\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 2.0693 \pm t_{14}^{*}\left(\frac{0.3324}{\sqrt{15}}\right) \approx (1.89,\ 2.25)$

We are 95% confident that the mean number of home runs hit per game in American League stadiums is between 1.89 and 2.25.

c) The average of 1.96 home runs hit per game in Coors Field is not unusual. It isn't even the highest average in the National League.

d) If you attempt to use two confidence intervals to assess a difference in means, you are actually adding standard deviations. But it's the variances that add, not the standard deviations. The two-sample difference of means procedure takes this into account.

e)

$$(\bar{y}_A - \bar{y}_N) \pm t^*_{df}\sqrt{\frac{s_A^2}{n_A} + \frac{s_N^2}{n_N}}$$

$$= (2.0693 - 1.7653) \pm t^*_{26.77}\sqrt{\frac{0.3324^2}{15} + \frac{0.3833^2}{15}} \approx (0.035,\ 0.573)$$

f) We are 95% confident that the mean number of home runs in American League stadiums is between 0.035 and 0.573 home runs higher than the mean number of home runs in National League stadiums.

g) Since the interval does not contain 0, there is evidence of a difference in the mean number of home runs hit per game in the stadiums of the two leagues. American League stadiums have a significantly higher average number of home runs than National League stadiums.

72. Hard water Derby.

a) H_0: The mean mortality rate is the same for towns North and South of Derby. $(\mu_N = \mu_S \text{ or } \mu_N - \mu_S = 0)$

H_A: The mean mortality rate is different for towns North and South of Derby. $(\mu_N \neq \mu_S \text{ or } \mu_N - \mu_S \neq 0)$

Independent groups assumption: The towns were sampled independently.
Independence assumption: Assume that the mortality rates are in each town are independent of the mortality rates in the others.
Nearly Normal condition: We don't have the actual data, so we can't look at histograms of the distributions, but the samples are fairly large. It should be okay to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 53.49 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_N - \bar{y}_S) = \sqrt{\frac{138.470^2}{34} + \frac{151.114^2}{27}} \approx 37.546\,.$$

The observed difference between the mean scores is 1631.59 – 1388.85 = 242.74.

Since the *P*-value = 3.2×10^{-8} is low, we reject the null hypothesis. There is strong evidence that the mean mortality rate different for towns north and south of Derby. There is evidence that the mortality rate north of Derby is higher.

$$t = \frac{(\bar{y}_N - \bar{y}_S) - (0)}{SE(\bar{y}_N - \bar{y}_S)}$$

$$t \approx \frac{242.74}{37.546}$$

$$t \approx 6.47$$

b) Since there is an outlier in the data north of Derby, the conditions for inference are not satisfied, and it is risky to use the two-sample t-test. The outlier should be removed, and the test should be performed again. Without the actual data, we are not able to do this. The test without the outlier would *probably* help us reach the same conclusion, but there is no way to be sure.

73. Job satisfaction.

A two-sample t-procedure is not appropriate for these data, because the two groups are not independent. They are before and after satisfaction scores for the same workers. Workers that have high levels of job satisfaction before the exercise program is implemented may tend to have higher levels of job satisfaction than other workers after the program as well.

74. Summer school.

A two-sample t-procedure is not appropriate for these data, because the two groups are not independent. They are before and after scores for the same students. Students with high scores before summer school may tend to have higher scores after summer school as well.

75. Sex and violence.

a) Since the P-value = 0.136 is high, we fail to reject the null hypothesis. There is no evidence of a difference in the mean number of brands recalled by viewers of sexual content and viewers of violent content.

b) H_0: The mean number of brands recalled is the same for viewers of sexual content and viewers of neutral content. $(\mu_S = \mu_N \text{ or } \mu_S - \mu_N = 0)$

H_A: The mean number of brands recalled is different for viewers of sexual content and viewers of neutral content. $(\mu_S \neq \mu_N \text{ or } \mu_S - \mu_N \neq 0)$

Independent groups assumption: Recall of one group should not affect recall of another.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: The samples are large.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 214 degrees of freedom (from the approximation formula). We will perform a two-sample t-test.

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_S - \bar{y}_N) = \sqrt{\frac{1.76^2}{108} + \frac{1.77^2}{108}} \approx 0.24.$$

The observed difference between the mean scores is 1.71 – 3.17 = – 1.46.

Since the P-value = 5.5×10^{-9} is low, we reject the null hypothesis. There is strong evidence that the mean number of brand names recalled is different for viewers of sexual content and viewers of neutral content. The evidence suggests that viewers of neutral ads remember more brand names on average than viewers of sexual content.

$$t = \frac{(\bar{y}_S - \bar{y}_N) - (0)}{SE(\bar{y}_S - \bar{y}_N)}$$

$$t \approx \frac{-1.46}{0.24}$$

$$t \approx -6.08$$

76. Ad campaign.

a) We are 95% confident that the mean number of ads remembered by viewers of shows with violent content will be between 1.6 and 0.6 lower than the mean number of brand names remembered by viewers of shows with neutral content.

b) If they want viewers to remember their brand names, they should consider advertising on shows with neutral content, as opposed to shows with violent content.

77. Hungry?

H_0: The mean number of ounces of ice cream people scoop is the same for large and small bowls. $(\mu_{big} = \mu_{small}$ or $\mu_{big} - \mu_{small} = 0)$

H_A: The mean number of ounces of ice cream people scoop is the different for large and small bowls. $(\mu_{big} \neq \mu_{small}$ or $\mu_{big} - \mu_{small} \neq 0)$

Independent groups assumption: The amount of ice cream scooped by individuals should be independent.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: Assume that this condition is met.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 34 degrees of freedom (from the approximation formula). Perform a two-sample t-test.

The sampling distribution model has mean 0, with standard error:

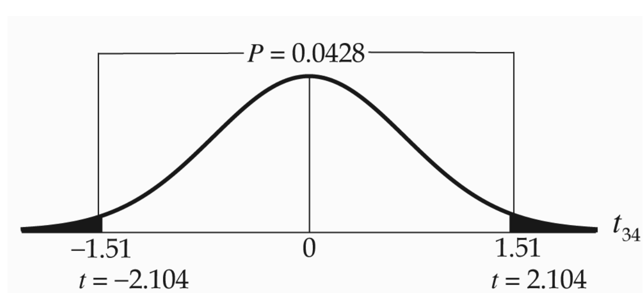
$$SE(\bar{y}_{big} - \bar{y}_{small}) = \sqrt{\frac{2.91^2}{22} + \frac{1.84^2}{26}} \approx 0.7177 \text{ oz.}$$

The observed difference between the mean amounts is 6.58 – 5.07 = 1.51 oz.

$$t = \frac{(\bar{y}_{big} - \bar{y}_{small}) - (0)}{SE(\bar{y}_{big} - \bar{y}_{small})}$$

$$t \approx \frac{1.51}{0.7177}$$

$$t \approx 2.104$$



Since the P-value of 0.0428 is low, we reject the null hypothesis. There is strong evidence that the that the mean amount of ice cream people put into a bowl is related to the size of the bowl. People tend to put more ice cream into the large bowl, on average, than the small bowl.

78. Thirsty?

H_0: The mean number of milliliters of liquid people pour when asked to pour a "shot" is the same for highballs and tumblers.
$\left(\mu_{tumbler} = \mu_{highball} \text{ or } \mu_{tumbler} - \mu_{highball} = 0\right)$

H_A: The mean number of milliliters of liquid people pour when asked to pour a "shot" is different for highballs and tumblers.
$\left(\mu_{tumbler} \neq \mu_{highball} \text{ or } \mu_{tumbler} - \mu_{highball} \neq 0\right)$

Independent groups assumption: The amount of liquid poured by individuals should be independent.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: Assume that this condition is met.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 194 degrees of freedom (from the approximation formula). We will perform a two-sample t-test.

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_{tumbler} - \bar{y}_{highball}) = \sqrt{\frac{17.9^2}{99} + \frac{16.2^2}{99}} \approx 2.4264 \text{ ml}.$$

The observed difference between the mean amounts is 60.9 – 42.2 = 18.7 ml.

$$t = \frac{\left(\bar{y}_{tumbler} - \bar{y}_{highball}\right) - (0)}{SE\left(\bar{y}_{tumbler} - \bar{y}_{highball}\right)}$$

$$t \approx \frac{18.7}{2.4264}$$

$$t \approx 7.71$$

Since the P-value (less than 0.0001) is low, we reject the null hypothesis. There is strong evidence that the mean amount liquid people pour into a glass is related to the shape of the glass. People tend to pour more, on average, into a small, wide tumbler than into a tall, narrow highball glass.

79. Swim the Lake 2013 revisited.

a)

$$(\bar{y}_M - \bar{y}_W) \pm t^*_{df}\sqrt{\frac{s_M^2}{n_M} + \frac{s_W^2}{n_W}}$$

$$= (1226.04 - 1257.09) \pm t^*_{35.7}\sqrt{\frac{382.85^2}{23} + \frac{261.10^2}{34}} \approx (-216.74, 154.64)$$

If the assumptions and conditions are met, we can be 95% confident that the interval –216.74 to 154.64 minutes contains the true difference in mean crossing times between men and women. Because the interval includes zero, we cannot be confident that there is any difference at all.

b) **Independent groups assumption:** The times from the two groups are likely to be independent of one another, provided that these were all individual swims.
Randomization condition: The times are not a random sample from any identifiable population, but it is likely that the times are representative of times from swimmers who might attempt a challenge such as this. Hopefully, these times were recorded from different swimmers.
Nearly Normal condition: The boxplots show two high outliers for the men and some skewness for both. Removing the outliers may make the difference in times larger, but there is no justification for doing so. The histograms are unimodal; but somewhat skewed to the right.

We are reluctant to draw any conclusions about the difference in mean times it takes men or women to swim the lake. The sample is not random, we have no way of knowing if it is representative, and the data are skewed with some outliers.

80. Still swimming.

a)

$$(\bar{y}_W - \bar{y}_M) \pm t^*_{df}\sqrt{\frac{s_W^2}{n_W} + \frac{s_M^2}{n_M}}$$

$$= (1195.34 - 1226.04) \pm t^*_{32}\sqrt{\frac{216.326^2}{29} + \frac{382.85^2}{23}} \approx (-212.54, 151.14)$$

We are 95% confident that the interval –187.2 to 144.2 minutes contains the true difference in mean crossing times between men and women. Because the interval includes zero, we cannot be confident that there is any difference at all.

b)

$$(\bar{y}_W - \bar{y}_M) \pm t^*_{df}\sqrt{\frac{s_W^2}{n_W} + \frac{s_M^2}{n_M}}$$

$$= (1195.34 - 1132.14) \pm t^*_{42}\sqrt{\frac{216.326^2}{29} + \frac{220.36^2}{21}} \approx (-63.18, 189.58)$$

We are 95% confident that the interval –63.18 to 189.58 minutes contains the true difference in mean crossing times between men and women. Because the interval includes zero, we cannot be confident that there is any difference at all. Even giving men the benefit of the doubt, there is no evidence that the means are any difference. However, since this is not a random sample, we should be cautious in making any conclusions.

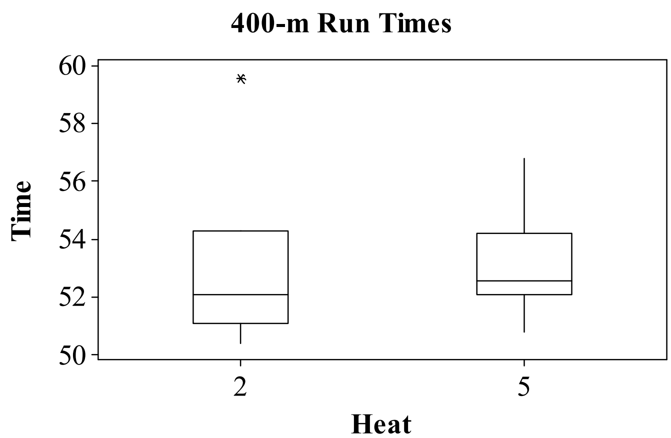
c) It is reasonable to assume that the same swimmer crossing the lake on two different occasions might perform similarly, so these are not all independent events. That could reduce the variability within each of the groups, reducing the standard error. We don't place much faith in our earlier analyses.

81. Running heats London.

H0: The mean time to finish is the same for heats 2 and 5. $(\mu_2 = \mu_5 \text{ or } \mu_2 - \mu_5 = 0)$

HA: The mean time is not the same for heats 2 and 5. $(\mu_2 \neq \mu_5 \text{ or } \mu_2 - \mu_5 \neq 0)$

Independent groups assumption: The two heats were independent.
Randomization condition: Runners were randomly assigned.
Nearly Normal condition: The boxplots show an outlier in the distribution of times in heat 2. We will perform the test twice, with and without the outlier.



Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 8.39 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{3.3618249^2}{6} + \frac{2.0996325^2}{6}} \approx 1.6181.$$

The observed difference between mean times is 53.2267 – 53.1483 = 0.0784.

$$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)}$$
$$t \approx \frac{0.0784}{1.6181}$$
$$t \approx 0.048$$

Since the P-value = 0.9625 is high, we fail to reject the null hypothesis. There is no evidence that the mean time to finish differs between the two heats.

Without the outlier, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 8.79 degrees of freedom (from the approximation formula). We will perform a two-sample t-test.

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\frac{1.4601267^2}{5} + \frac{2.0996325^2}{6}} \approx 1.0776 \,.$$

The observed difference between mean times is 51.962 – 53.1483 = –1.1863.

$$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)}$$
$$t \approx \frac{-1.1863}{1.0776}$$
$$t \approx -1.10$$

Since the P-value = 0.3001 is high, we fail to reject the null hypothesis. There is no evidence that the mean time to finish differs between the two heats.

There are only two values that exceed the maximum and minimum in the other set, so Tukey's test does no find a significant difference in mean 400 meter times between the two heats.

The Rank-sum test has a P-value of 0.70, so there is no evidence of a difference in mean 400 meter times here, either.

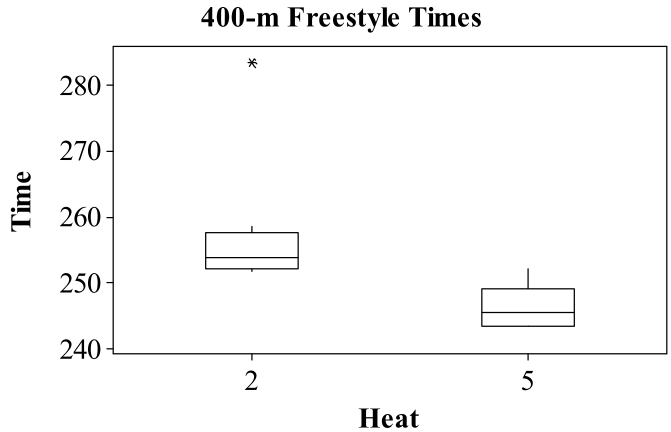
82. Swimming heats London.

H_0: The mean time to finish is the same for heats 2 and 5. $(\mu_2 = \mu_5 \text{ or } \mu_2 - \mu_5 = 0)$

H_A: The mean time is not the same for heats 2 and 5. $(\mu_2 \neq \mu_5 \text{ or } \mu_2 - \mu_5 \neq 0)$

Independent groups assumption: The two heats were independent.
Randomization condition: Swimmers were not randomly assigned, but if we consider these heats to be representative of seeded heats, we may be able to generalize the results.
Nearly Normal condition: The boxplots of the times in each heat show distributions that are reasonably symmetric, but there is one high outlier in heat 2. We will perform the test twice, with and without the outlier.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 8.41 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.



The sampling distribution model has mean 0, with

standard error: $SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\dfrac{10.6634^2}{8} + \dfrac{3.4062^2}{8}} \approx 3.9578$.

$$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)}$$

$$t \approx \frac{11.375}{3.9578}$$

$$t \approx 2.874$$

The observed difference between the mean times is 257.8 – 246.425 = 11.375.

Since the P-value = 0.0196, we reject the null hypothesis. There is strong evidence that the mean time to finish differs between the two heats. In fact, the mean time in heat two was higher than the mean time in heat five.

Without the outlier, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 12.82 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

$SE(\bar{y}_2 - \bar{y}_5) = \sqrt{\dfrac{2.6177^2}{7} + \dfrac{3.4062^2}{8}} \approx 1.5586$.

$$t = \frac{(\bar{y}_2 - \bar{y}_5) - (0)}{SE(\bar{y}_2 - \bar{y}_5)}$$

$$t \approx \frac{7.704}{1.5586}$$

$$t \approx 4.943$$

The observed difference between the mean times is 254.129 – 246.425 = 7.704.

Since the P-value = 0.0003, we reject the null hypothesis. There is strong evidence that the mean time to finish differs between the two heats. In fact, the mean time in heat two was higher than the mean time in heat five.

There are 12 values that exceed the maximum and minimum in the other set, so Tukey's test does finds a significant difference in mean 400 meter times between the two heats, at the 0.01 level.

The Rank-sum test has a P-value of 0.0016, so there is evidence of a difference in mean 400 meter times here as well.

83. Tees.

H_0: The mean ball velocity is the same for regular and Stinger tees. $(\mu_S = \mu_R \text{ or } \mu_S - \mu_R = 0)$

H_A: The mean ball velocity is higher for the Stinger tees. $(\mu_S > \mu_R \text{ or } \mu_S - \mu_R > 0)$

Assuming the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 7.03 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_S - \bar{y}_R) = \sqrt{\frac{0.41^2}{6} + \frac{0.89^2}{6}} \approx 0.4000.$$

The observed difference between the mean velocities is 128.83 – 127 = 1.83.

Since the *P*-value = 0.0013, we reject the null hypothesis. There is strong evidence that the mean ball velocity for stinger tees is higher than the mean velocity for regular tees.

$$t = \frac{(\bar{y}_S - \bar{y}_R) - (0)}{SE(\bar{y}_S - \bar{y}_R)}$$

$$t \approx \frac{1.83}{0.4000}$$

$$t \approx 4.57$$

84. Golf again.

H_0: The mean distance is the same for regular and Stinger tees. $(\mu_S = \mu_R \text{ or } \mu_S - \mu_R = 0)$

H_A: The mean distance is greater for the Stinger tees. $(\mu_S > \mu_R \text{ or } \mu_S - \mu_R > 0)$

Assuming the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 9.42 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_S - \bar{y}_R) = \sqrt{\frac{2.76^2}{6} + \frac{2.14^2}{6}} \approx 1.426.$$

The observed difference between mean distances is 241 – 227.17 = 13.83.

Since the *P*-value <0.0001, we reject the null hypothesis. There is strong evidence that the mean distance for Stinger tees is higher than the mean distance for regular tees.

$$t = \frac{(\bar{y}_S - \bar{y}_R) - (0)}{SE(\bar{y}_S - \bar{y}_R)}$$

$$t \approx \frac{13.83}{1.426}$$

$$t \approx 9.70$$

85. Music and memory.

a) H_0: The mean memory test score is the same for those who listen to Mozart as it is for those who listen to rap music. $(\mu_M = \mu_R \text{ or } \mu_M - \mu_R = 0)$

H_A: The mean memory test score is greater for those who listen to Mozart than it is for those who listen to rap music. $(\mu_M > \mu_R \text{ or } \mu_M - \mu_R > 0)$

Independent groups assumption: The groups are not related in regards to memory score.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: We don't have the actual data. We will assume that the distributions of the populations of memory test scores are Normal.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 45.88 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

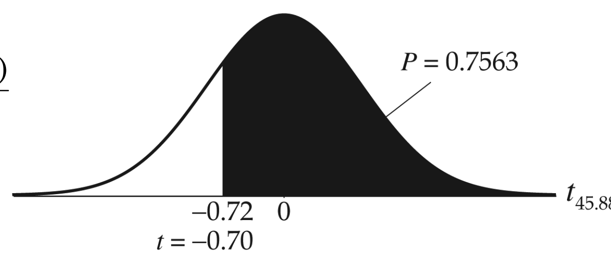
$$SE(\bar{y}_M - \bar{y}_R) = \sqrt{\frac{3.19^2}{20} + \frac{3.99^2}{29}} \approx 1.0285.$$

The observed difference between the mean number of objects remembered is 10.0 – 10.72 = – 0.72.

$$t = \frac{(\bar{y}_M - \bar{y}_R) - (0)}{SE(\bar{y}_M - \bar{y}_R)}$$

$$t \approx \frac{-0.72}{1.0285}$$

$$t \approx -0.70$$



Since the *P*-value = 0.7563 is high, we fail to reject the null hypothesis. There is no evidence that the mean number of objects remembered by those who listen to Mozart is higher than the mean number of objects remembered by those who listen to rap music.

b) $$(\bar{y}_M - \bar{y}_N) \pm t_{df}^* \sqrt{\frac{s_M^2}{n_M} + \frac{s_N^2}{n_N}} = (10.0 - 12.77) \pm t_{19.09}^* \sqrt{\frac{3.19^2}{20} + \frac{4.73^2}{13}} \approx (-5.351, -0.189)$$

We are 90% confident that the mean number of objects remembered by those who listen to Mozart is between 0.189 and 5.352 objects lower than the mean of those who listened to no music.

86. Rap.

a) H_0: The mean memory test score is the same for those who listen to rap as it is for those who listen to no music. $(\mu_R = \mu_N \text{ or } \mu_R - \mu_N = 0)$

H_A: The mean memory test score is lower for those who listen to rap than it is for those who listen to no music. $(\mu_R < \mu_N \text{ or } \mu_R - \mu_N < 0)$

Independent groups assumption: The groups are not related in regards to memory score.
Randomization condition: Subjects were randomly assigned to groups.
Nearly Normal condition: We don't have the actual data. We will assume that the distributions of the populations of memory test scores are Normal.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 20.00 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test.

The sampling distribution model has mean 0, with standard error:

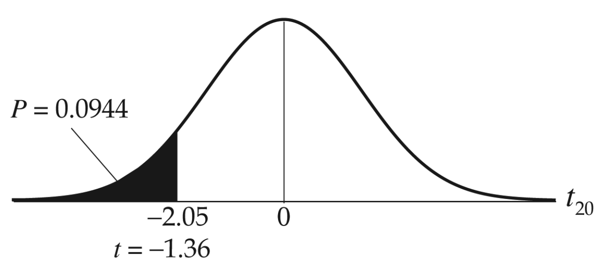
$$SE(\bar{y}_R - \bar{y}_N) = \sqrt{\frac{3.99^2}{29} + \frac{4.73^2}{13}} \approx 1.5066.$$

The observed difference between the mean number of objects remembered is 10.72 – 12.77 = – 2.05.

$$t = \frac{(\bar{y}_R - \bar{y}_N) - (0)}{SE(\bar{y}_R - \bar{y}_N)}$$

$$t \approx \frac{-2.05}{1.5066}$$

$$t \approx -1.36$$



Since the *P*-value = 0.0944 is high, we fail to reject the null hypothesis. There is little evidence that the mean number of objects remembered by those who listen to rap is lower than the mean number of objects remembered by those who listen to no music.

b) We did not conclude that there was a difference in the number of items remembered.

Chapter 23 – Paired Samples and Blocks

Section 23.1

1. Which method?

a) Paired. Each individual has two scores, so they are certainly associated.

b) Not paired. The scores of males and females are independent.

c) Paired. Each student was surveyed twice, so there responses are associated.

2. Which method II?

a) Paired. Each respondent is paired with his or her spouse.

b) Not paired. The respondents in the different treatment groups are not associated.

c) Not paired. Opinions of freshman and sophomores are not associated.

Section 23.2

3. Cars and trucks.

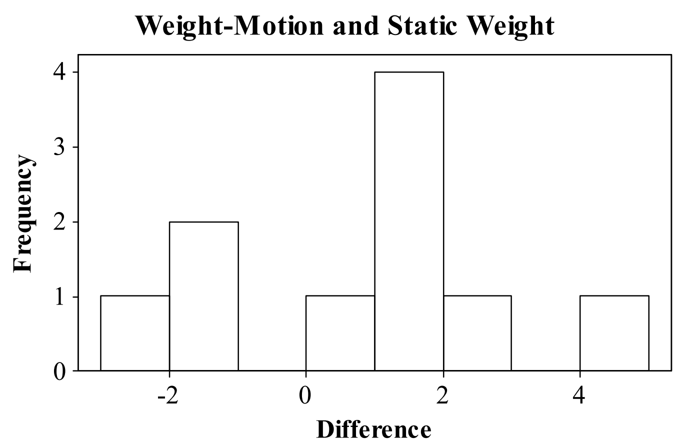
a) No. The vehicles have no natural pairing.

b) Possibly. The data are quantitative and paired by vehicle.

c) The sample size is large, but there is at least one extreme outlier that should be investigated before applying these methods.

4. Weighing trucks.

Yes. The data are quantitative and paired. A graph of the differences is roughly symmetric.



Section 23.3

5. Cars and trucks again.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 7.37 \pm t^*_{631}\left(\frac{2.52}{\sqrt{632}}\right) \approx (7.17, 7.57)$$

We are 95% confident that the interval 7.17 mpg to 7.57 mpg captures the true improvements in highway gas mileage compared to city gas mileage.

6. Weighing trucks II.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 0.740 \pm t^*_9\left(\frac{2.280}{\sqrt{10}}\right) \approx (-1.29,\ 2.77)$$

We are 98% confident that the interval from –1290 pounds to 2770 pounds captures the true mean difference in measured weights.

Section 23.4

7. Blocking cars and trucks.

The difference between fuel efficiency of cars and that of trucks can be large, but isn't relevant to the question asked about highway vs. city driving. Pairing places each vehicle in its own block to remove that variation from consideration.

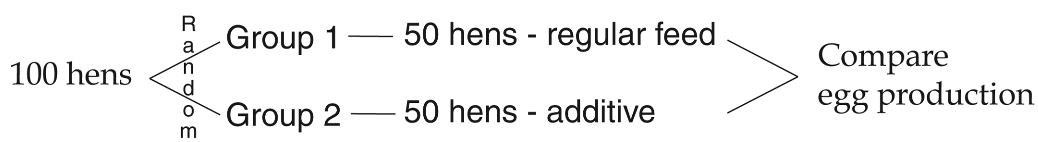
8. Weighing trucks III.

No. We are not concerned with the measurements of each scale. We want to know the difference between the methods for each truck. So it is the paired differences that we are concerned with.

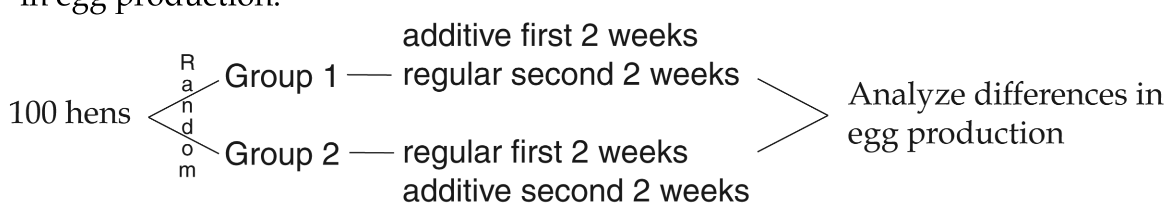
Chapter Exercises

9. More eggs?

a) Randomly assign 50 hens to each of the two kinds of feed. Compare the mean egg production of the two groups at the end of one month.



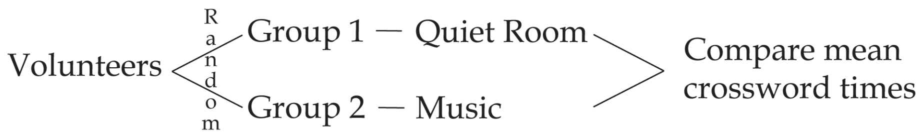
b) Randomly divide the 100 hens into two groups of 50 hens each. Feed the hens in the first group the regular feed for two weeks, then switch to the additive for 2 weeks. Feed the hens in the second group the additive for two weeks, and then switch to the regular feed for two weeks. Subtract each hen's "regular" egg production from her "additive" egg production, and analyze the mean difference in egg production.



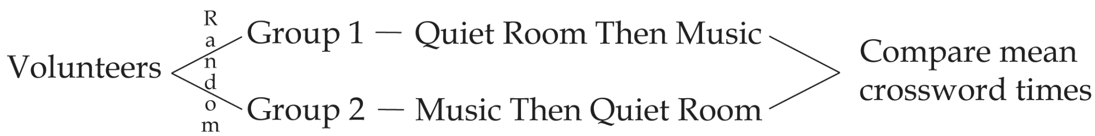
c) The matched pairs design in part b is the stronger design. Hens vary in their egg production regardless of feed. This design controls for that variability by matching the hens with themselves.

10. Music.

a) Randomly assign half of the volunteers to do the puzzles in a quiet room, and assign the other half to do the puzzles with music in headphones. Compare the mean time of the group in the quiet room to the mean time of the group listening to music.



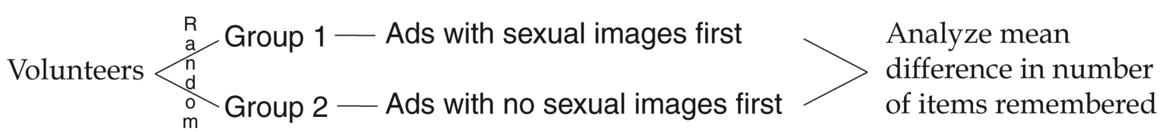
b) Randomly assign half of the volunteers to do a puzzle in a quiet room, and assign the other half to do the puzzles with music. Then have each do a puzzle under the other condition. Subtract each volunteer's "quiet" time from his or her "music" time, and analyze the mean difference in times.



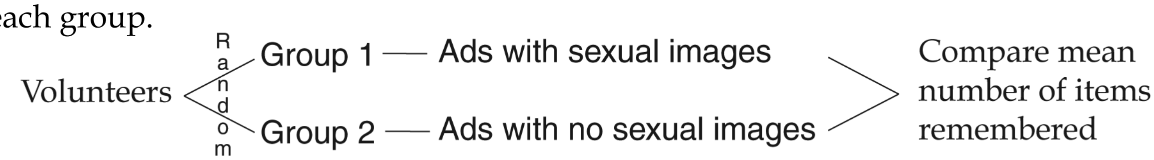
c) The matched pairs design in part b is the stronger design. People vary in their ability to do crossword puzzles. This design controls for that variability by matching the volunteers with themselves.

11. Sex sells?

a) Randomly assign half of the volunteers to watch ads with sexual images, and assign the other half to watch ads without the sexual images. Record the number of items remembered. Then have each group watch the other type of ad. Record the number of items recalled. Examine the difference in the number of items remembered for each person.



b) Randomly assign half of the volunteers to watch ads with sexual images, and assign the other half to watch ads without the sexual images. Record the number of items remembered. Compare the mean number of products remembered by each group.



12. Freshman 15?

a) Select a random sample of freshmen. Weigh them when college starts in the fall, and again when they leave for home in the spring. Examine the difference in weight for each student.

b) Select a random sample of freshman as they enter college in the fall to determine their average weight. Select a new random sample of students at the end of the spring semester to determine their average weight. Compare the mean weights of the two groups.

13. Women.

a) The paired t-test is appropriate. The labor force participation rate for two different years was paired by city.

b) Since the P-value = 0.0244, there is evidence of a difference in the average labor force participation rate for women between 1968 and 1972. The evidence suggests an increase in the participation rate for women.

14. Cloud seeding.

a) The two-sample t-test is appropriate for these data. The seeded and unseeded clouds are not paired in any way. They are independent.

b) Since the P-value = 0.0538, there is some evidence that the mean rainfall from seeded clouds is greater than the mean rainfall from unseeded clouds.

15. Friday the 13th, traffic.

a) The paired t-test is appropriate, since we have pairs of Fridays in 5 different months. Data from adjacent Fridays within a month may be more similar than randomly chosen Fridays.

b) Since the P-value = 0.0212, there is evidence that the mean number of cars on the M25 motorway on Friday the 13th is less than the mean number of cars on the previous Friday.

c) We don't know if these Friday pairs were selected at random. Obviously, if these are the Fridays with the largest differences, this will affect our conclusion. The Nearly Normal condition appears to be met by the differences, but the sample size of five pairs is small.

16. Friday the 13th, accidents.

a) The paired t-test is appropriate, since we have pairs of Fridays in 6 different months. Data from adjacent Fridays within a month may be more similar than randomly chosen Fridays.

b) Since the P-value = 0.0211, there is evidence that the mean number of admissions to hospitals found on Friday the 13th is more than on the previous Friday.

c) We don't know if these Friday pairs were selected at random. Obviously, if these are the Fridays with the largest differences, this will affect our conclusion. The Nearly Normal condition appears to be met by the differences, but the sample size of six pairs is small.

17. Online insurance I.

Adding variances requires that the variables be independent. These price quotes are for the same cars, so they are paired. Drivers quoted high insurance premiums by the local company will be likely to get a high rate from the online company, too.

18. Wind speed, part I.

Adding variances requires that the variables be independent. The wind speeds were recorded at nearby sites, to they are likely to be both high or both low at the same time.

19. Online insurance II.

a) The histogram would help you decide whether the online company offers cheaper insurance. We are concerned with the difference in price, not the distribution of each set of prices.

b) Insurance cost is based on risk, so drivers are likely to see similar quotes from each company, making the differences relatively smaller.

c) The price quotes are paired. They were for a random sample of the agent's customers and the histogram of differences looks approximately Normal.

20. Wind speed, part II.

a) The outliers are particularly windy days, but they were windy at both sites, making the difference in wind speeds less unusual.

b) The histogram and summaries of the differences are more appropriate because they are paired observations and all we care about is which site is windier.

c) The wind measurements at the same times at two nearby sites are paired. We should be concerned that there might be a lack of independence from one time to the next, but the times were 6 hours apart and the differences in speeds are likely to be independent. Although not random, we can regard a sample this large as generally representative of wind speed s at these sites. The histogram of differences is unimodal, symmetric and bell-shaped.

21. Online insurance III.

H_0: The mean difference between online and local insurance rates is zero. $(\mu_{Local-Online} = 0)$

H_A: The mean difference is greater than zero. $(\mu_{Local-Online} > 0)$

Since the conditions are satisfied (in a previous exercise), the sampling distribution of the difference can be modeled with a Student's t-model with 10 – 1 = 9 degrees of freedom, $t_9\left(0, \frac{175.663}{\sqrt{10}}\right)$.

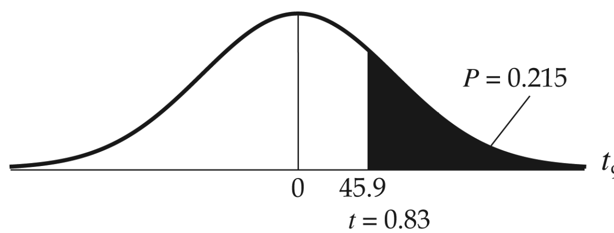
We will use a paired t-test, with $\bar{d} = 45.9$.

Since the P-value = 0.215 is high, we fail to reject the null hypothesis. There is no evidence that online insurance premiums are lower on average.

$$t = \frac{\bar{d} - 0}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{45.9 - 0}{\frac{175.663}{\sqrt{10}}}$$

$$t \approx 0.83$$



22. Wind speed, part III.

H_0: The mean difference between wind speeds at the two sites is zero. $(\mu_{2-4} = 0)$

H_A: The mean difference between at the two sites is different than zero. $(\mu_{2-4} \neq 0)$

Since the conditions are satisfied (in a previous exercise), the sampling distribution of the difference can be modeled with a Student's t-model with 1114 – 1 = 1113 degrees of freedom, $t_{1113}\left(0, \frac{2.551}{\sqrt{1114}}\right)$.

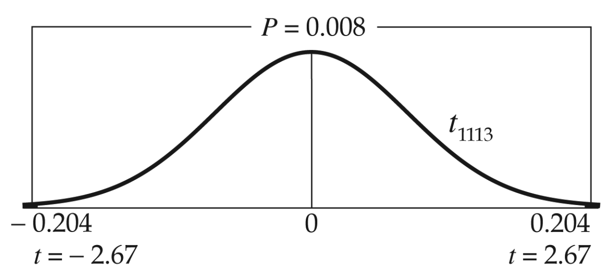
We will use a paired t-test, with $\bar{d} = 0.204$.

Since the P-value = 0.008 is low, we reject the null hypothesis. There is strong evidence that the average wind speed is higher at site 2.

$$t = \frac{\bar{d} - 0}{\frac{s}{\sqrt{n}}}$$

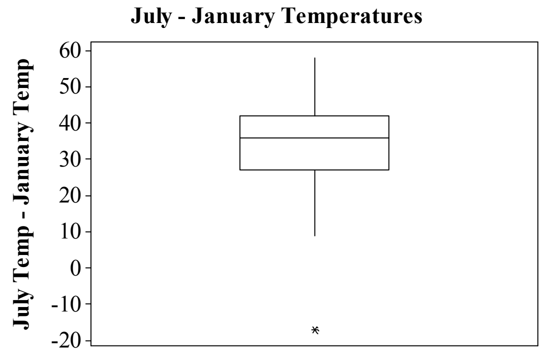
$$t = \frac{0.204 - 0}{\frac{2.551}{\sqrt{1114}}}$$

$$t \approx 2.67$$



23. City Temperatures.

a) **Paired data assumption:** The data are paired by city.
Randomization condition: These cities might not be representative of all cities, so be cautious in generalizing the results.
Normal population assumption: The histogram of differences between January and July mean temperature is roughly unimodal and symmetric, but shows a low outlier. This is Auckland, New Zealand, a city in the southern hemisphere. Seasons here would be the opposite of the rest of the cities, which are in the northern hemisphere. It should be set aside.



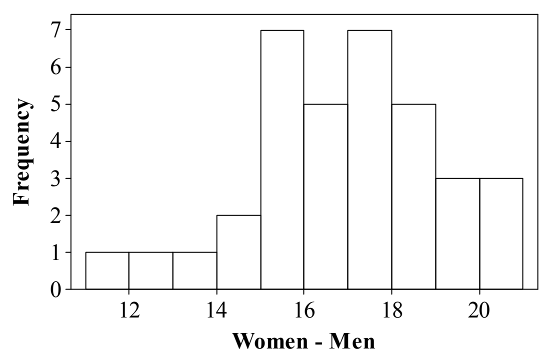
b) With Auckland set aside, the conditions are satisfied. The sampling distribution of the difference can be modeled with a Student's *t*-model with 11 – 1 = 10 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 33.7429 \pm t^*_{34}\left(\frac{14.9968}{\sqrt{34}}\right) \approx (30.94,\ 39.53)$$

We are 95% confident that the average high temperature in northern hemisphere cities in July is an average of between 30.94° to 39.53° higher than in January.

24. NY Marathon 2013.

Paired data assumption: The data are paired by year.
Randomization condition: Assume these years, at this marathon, are representative of all differences.
Normal population assumption: The histogram of differences between women's and men's times is roughly unimodal and symmetric.

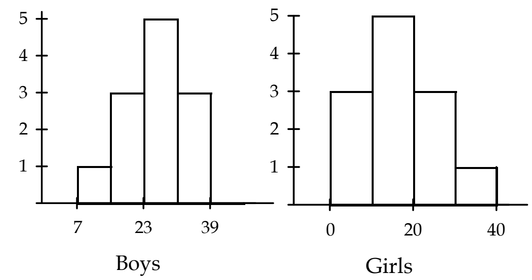


Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 35 – 1 = 34 df. We will find a paired *t*-interval, with 90% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 16.786 \pm t^*_{34}\left(\frac{2.148}{\sqrt{35}}\right) \approx (16.048,\ 17.524)$$

We are 90% confident women's winning marathon times are an average of between 16.05 and 17.52 minutes higher than men's winning times.

25. Push-ups.



Independent groups assumption: The group of boys is independent of the group of girls.
Randomization condition: Assume that students are assigned to gym classes at random.
Nearly Normal condition: The histograms of the number of push-ups from each group are roughly unimodal and symmetric.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 21 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 90% confidence.

$$(\bar{y}_B - \bar{y}_G) \pm t^*_{df}\sqrt{\frac{s_B^2}{n_B} + \frac{s_G^2}{n_G}} = (23.8333 - 16.5000) \pm t^*_{21}\sqrt{\frac{7.20900^2}{12} + \frac{8.93919^2}{12}} \approx (1.6,\ 13.0)$$

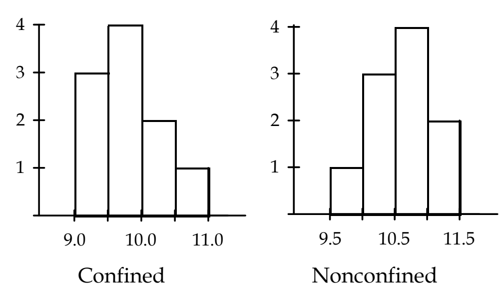
We are 90% confident that, at Gossett High, the mean number of push-ups that boys can do is between 1.6 and 13.0 more than the mean for the girls.

26. Brain waves.

a) H_0: The mean alpha-wave frequency for nonconfined inmates is the same as the mean alpha wave frequency for confined inmates. $(\mu_{NC} = \mu_C \text{ or } \mu_{NC} - \mu_C = 0)$

H_A: The mean alpha-wave frequency for nonconfined inmates is different from the mean frequency for confined inmates. $(\mu_{NC} \neq \mu_C \text{ or } \mu_{NC} - \mu_C \neq 0)$

b) **Independent Groups Assumption:** The two groups of inmates were placed under different conditions, solitary confinement and not confined.
Randomization Condition: Inmates were randomly assigned to groups.
Nearly Normal Condition: The histograms of the alpha-wave frequencies are unimodal and symmetric.



c) Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 16.9 degrees of freedom (from the approximation formula). We will perform a two-sample *t*-test. We know:

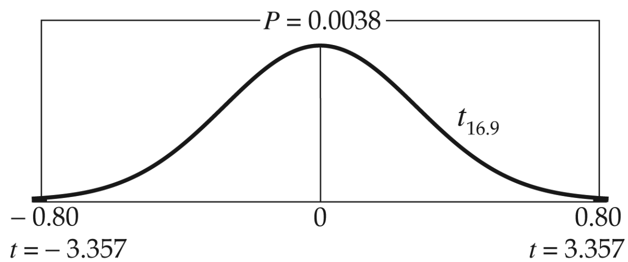
$$\bar{y}_{NC} = 10.58 \qquad \bar{y}_C = 9.78$$
$$s_{NC} = 0.458984 \qquad s_C = 0.597774$$
$$n_{NC} = 10 \qquad n_C = 10$$

The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_{NC} - \bar{y}_C) = \sqrt{\frac{0.458984^2}{10} + \frac{0.597774^2}{10}} \approx 0.2383.$$

The observed difference between the mean scores is 10.58 – 9.78 ≈ 0.80.

$$t = \frac{(\bar{y}_{NC} - \bar{y}_C) - (0)}{SE(\bar{y}_{NC} - \bar{y}_C)}$$
$$t \approx \frac{0.80}{0.2382}$$
$$t \approx 3.357$$



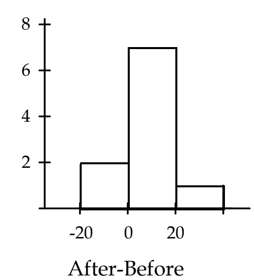
Since the P-value = 0.0038 is low, we reject the null hypothesis. There is evidence the mean alpha-wave frequency is different for nonconfined inmates and confined inmates.

d) The evidence suggests that the mean alpha-wave frequency for inmates subjected to confinement is different than the mean alpha-wave frequency for inmates that are not confined. This experiment suggests that mean alpha-wave frequency is lower for confined inmates.

27. Job satisfaction.

a) Use a paired t-test.

Paired data assumption: The data are before and after job satisfaction rating for the same workers.
Randomization condition: The workers were randomly selected to participate.
Nearly Normal conditon: The histogram of differences between before and after job satisfaction ratings is roughly unimodal and symmetric.



b) H_0: The mean difference in before and after job satisfaction scores is zero, and the exercise program is not effective at improving job satisfaction. $(\mu_d = 0)$

H_A: The mean difference is greater than zero, and the exercise program is effective at improving job satisfaction. $(\mu_d > 0)$

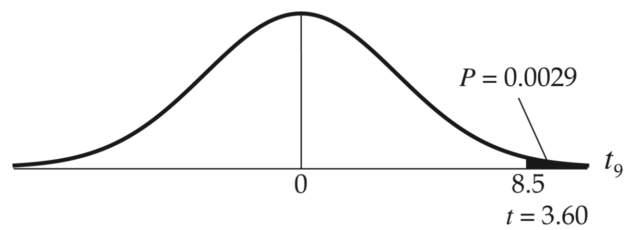
Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 10 – 1 = 9 degrees of freedom, $t_9\left(0, \frac{7.47217}{\sqrt{10}}\right)$. We will use a paired t-test, with $\bar{d} = 8.5$.

Since the P-value = 0.0029 is low, we reject the null hypothesis. There is evidence that the mean job satisfaction rating has increased since the implementation of the exercise program.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{8.5 - 0}{\frac{7.47217}{\sqrt{10}}}$$

$$t \approx 3.60$$



c) We concluded that there was an increase job satisfaction rating. If we are wrong, and there actually was no increase, we have committed a Type I error.

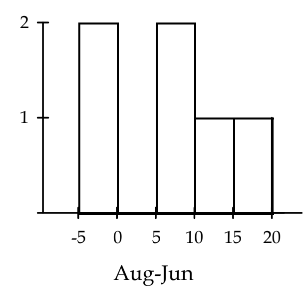
d) The P-value of the sign test is 0.1094, which would lead us to fail to reject the null hypothesis that the median difference was 0. This is a different conclusion than the paired t-test. However, since the conditions for the paired t-test are met, we should use those results. The t-test has more power.

28. Summer school.

a) H_0: The mean difference between August and June scores is zero, and the summer school program is not worthwhile. $(\mu_d = 0)$

H_A: The mean difference between August and June scores is greater than zero, and the summer school program is worthwhile. $(\mu_d > 0)$

Paired data assumption: The scores are paired by student.
Randomization condition: Assume that these students are representative of students who attend this school in other years.
Normal population assumption: The histogram of differences between August and June scores shows a distribution that could have come from a Normal population.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 6 – 1 = 5 degrees of freedom, $t_5\left(0, \frac{7.44759}{\sqrt{6}}\right)$.

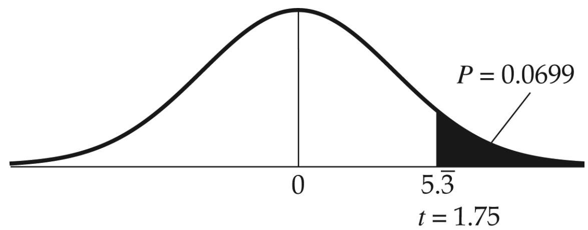
We will use a paired t-test, with $\bar{d} = 5.\bar{3}$.

Since the P-value = 0.0699 is fairly high, we fail to reject the null hypothesis. There is not strong evidence that scores increased on average. The summer school program does not appear worthwhile, but the P-value is low enough that we should look at a larger sample to be more confident in our conclusion.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{5.\overline{3} - 0}{\frac{7.44759}{\sqrt{6}}}$$

$$t \approx 1.75$$



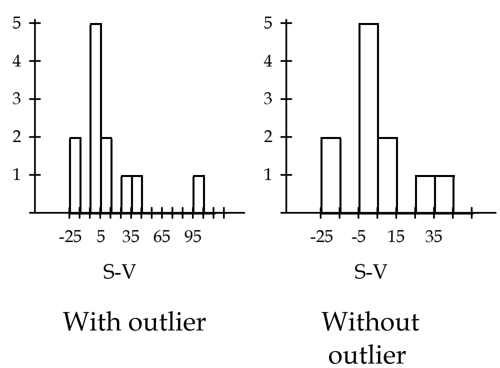
b) We concluded that there was no evidence of an increase. If there actually was an increase, we have committed a Type II error.

29. Yogurt.

H_0: The mean difference in calories between servings of strawberry and vanilla yogurt is zero. $(\mu_d = 0)$

H_A: The mean difference in calories between servings of strawberry and vanilla yogurt is different from zero. $(\mu_d \neq 0)$

Paired data assumption: The yogurt is paired by brand.
Randomization condition: Assume that these brands are representative of all brands.
Normal population assumption: The histogram of differences in calorie content between strawberry and vanilla shows an outlier, Great Value. When the outlier is eliminated, the histogram of differences is roughly unimodal and symmetric.



When Great Value yogurt is removed, the conditions are satisfied. The sampling distribution of the difference can be modeled with a Student's t-model with 11 – 1 = 10 degrees of freedom, $t_{10}\left(0, \frac{18.0907}{\sqrt{11}}\right)$.

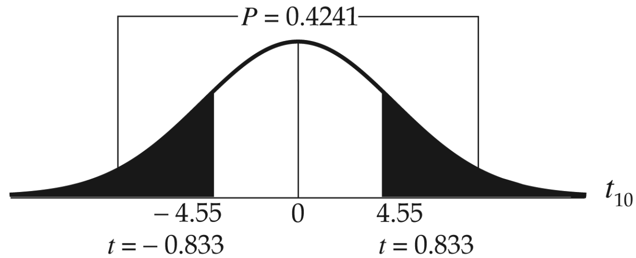
We will use a paired t-test, with $\bar{d} \approx 4.54545$.

Since the P-value = 0.4241 is high, we fail to reject the null hypothesis. There is no evidence of a mean difference in calorie content between strawberry yogurt and vanilla yogurt.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{4.54545 - 0}{\frac{18.0907}{\sqrt{11}}}$$

$$t \approx 0.833$$

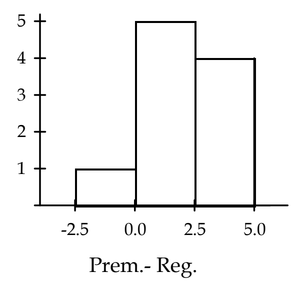


30. Gasoline.

a) H_0: The mean difference in mileage between premium and regular is zero. $(\mu_d = 0)$

H_A: The mean difference in mileage between premium and regular is greater than zero. $(\mu_d > 0)$

Paired data assumption: The mileage is paired by car.
Randomization condition: We randomized the order in which the different types of gasoline were used in each car.
Normal population assumption: The histogram of differences between premium and regular is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 10 – 1 = 9 degrees of freedom, $t_9\left(0, \frac{1.41421}{\sqrt{10}}\right)$.

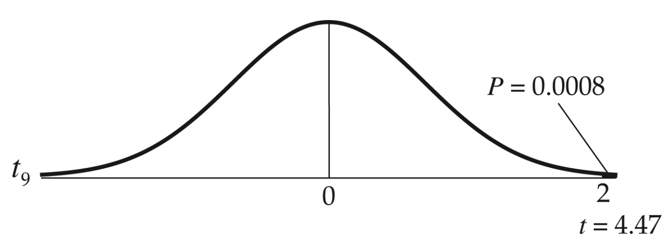
We will use a paired t-test, with $\bar{d} = 2$.

Since the P-value = 0.0008 is very low, we reject the null hypothesis. There is strong evidence of a mean increase in gas mileage between regular and premium.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{2 - 0}{\frac{1.41421}{\sqrt{10}}}$$

$$t \approx 4.47$$



b) $\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 2 \pm t^*_9\left(\frac{1.41421}{\sqrt{10}}\right) \approx (1.18,\ 2.82)$

We are 90% confident that the mean increase in gas mileage when using premium rather than regular gasoline is between 1.18 and 2.82 miles per gallon.

c) Premium costs more than regular. This difference might outweigh the increase in mileage.

d) With $t = 1.25$ and a *P*-value = 0.1144, we would have failed to reject the null hypothesis, and conclude that there was no evidence of a mean difference in mileage. The variation in performance of individual cars is greater than the variation related to the type of gasoline. This masked the true difference in mileage due to the gasoline. (Not to mention the fact that the two-sample test is not appropriate because we don't have independent samples!)

e) The sign test has a *P*-value of 0.0215, which would lead us to reject the hypothesis that the median difference was 0. This conclusion is consistent with the paired *t*-test.

31. Braking test.

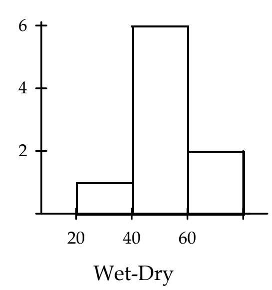
a) **Randomization Condition:** These cars are not a random sample, but are probably representative of all cars in terms of stopping distance.
Nearly Normal Condition: A histogram of the stopping distances is skewed to the right, but this may just be sampling variation from a Normal population. The "skew" is only a couple of stopping distances. We will proceed cautiously.

The cars in the sample had a mean stopping distance of 138.7 feet and a standard deviation of 9.66149 feet. Since the conditions have been satisfied, construct a one-sample t-interval, with 10 – 1 = 9 degrees of freedom, at 95% confidence.

$$\bar{y} \pm t_{n-1}^{*}\left(\frac{s}{\sqrt{n}}\right) = 138.7 \pm t_{9}^{*}\left(\frac{9.66149}{\sqrt{10}}\right) \approx (131.8,\ 145.6)$$

We are 95% confident that the mean dry pavement stopping distance for cars with this type of tires is between 131.8 and 145.6 feet. This estimate is based on an assumption that these cars are representative of all cars and that the population of stopping distances is Normal.

b) **Paired data assumption:** The data are paired by car.
Randomization condition: Assume that the cars are representative of all cars.
Normal population assumption: The difference in stopping distance for car #4 is an outlier, at only 12 feet. After excluding this difference, the histogram of differences is unimodal and symmetric.



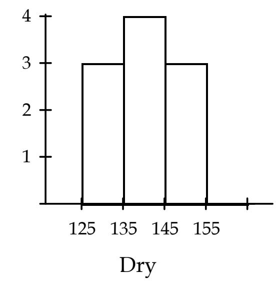
Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 9 – 1 = 8 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

$$\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = 55 \pm t_{8}^{*}\left(\frac{10.2103}{\sqrt{9}}\right) \approx (47.2,\ 62.8)$$

With car #4 removed, we are 95% confident that the mean increase in stopping distance on wet pavement is between 47.2 and 62.8 feet. (If you leave the outlier in, the interval is 38.8 to 62.6 feet, but you should remove it! This procedure is sensitive to outliers!)

32. Braking test 2.

a) **Randomization Condition:** These stops are probably representative of all such stops for this type of car, but not for all cars.
Nearly Normal Condition: A histogram of the stopping distances is roughly unimodal and symmetric.

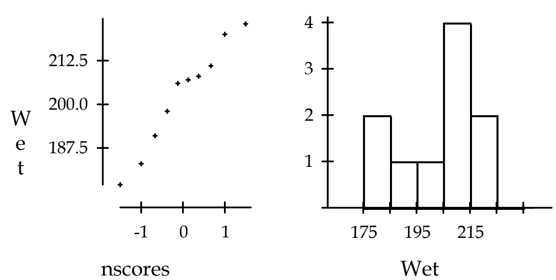


The stops in the sample had a mean stopping distance of 139.4 feet, and a standard deviation of 8.09938 feet. Since the conditions have been satisfied, construct a one-sample t-interval, with 10 – 1 = 9 degrees of freedom, at 95% confidence.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 139.4 \pm t^*_9\left(\frac{8.09938}{\sqrt{10}}\right) \approx (133.6,\ 145.2)$$

We are 95% confident that the mean dry pavement stopping distance for this type of car is between 133.6 and 145.2 feet.

b) **Independent Groups Assumption:** The wet pavement stops and dry pavement stops were made under different conditions and not paired in any way.
Randomization Condition: These stops are probably representative of all such stops for this type of car, but not for all cars.
Nearly Normal Condition: The histogram of dry pavement stopping distances is roughly unimodal and symmetric (from part a), but the histogram of wet pavement stopping distances is a bit skewed. Since the Normal probability plot looks fairly straight, we will proceed.



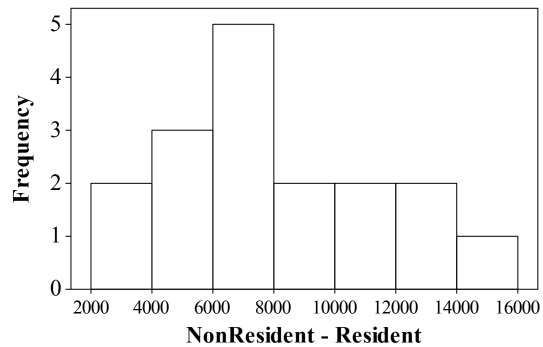
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 13.8 degrees of freedom (from the approximation formula). We will construct a two-sample t-interval, with 95% confidence.

$$(\bar{y}_W - \bar{y}_D) \pm t^*_{df}\sqrt{\frac{s_W^2}{n_W} + \frac{s_D^2}{n_D}} = (202.4 - 139.4) \pm t^*_{13.8}\sqrt{\frac{15.07168^2}{10} + \frac{8.09938^2}{10}} \approx (51.4,\ 74.7)$$

We are 95% confident that the mean stopping distance on wet pavement is between 51.4 and 74.6 feet longer than the mean stopping distance on dry pavement.

33. Tuition 2013.

a) **Paired data assumption:** The data are paired by college.
Randomization condition: The colleges were selected randomly.
Normal population assumption: The tuition difference for UC Irvine, at \$22,878, is an outlier, as is the tuition difference for New College of Florida at \$23,029. Once these have been set aside, the histogram of the differences is roughly unimodal and slightly skewed to the right. This should be fine for inference in a sample of 17 colleges.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 17 – 1 = 16 degrees of freedom. We will find a paired t-interval, with 90% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 7963.65 \pm t^*_{16}\left(\frac{3560.53}{\sqrt{17}}\right) \approx (6456,\ 9471)$$

b) With outliers removed, we are 90% confident that the mean increase in tuition for nonresidents versus residents is between about \$6456 and \$9471. (If you left the outliers in your data, the interval is about \$7235 to \$11,848, but you should set them aside! This procedure is sensitive to the presence of outliers!)

c) There is no evidence to suggest that the magazine made a false claim. An increase of \$7000 for nonresidents is contained within our 90% confidence interval.

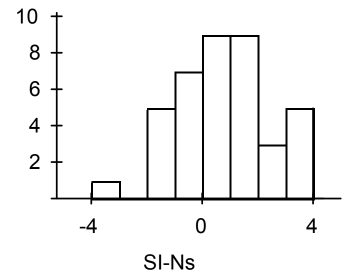
34. Sex sells, part II.

H_0: The mean difference in number of items remembered for ads with sexual images and ads without sexual images is zero. $(\mu_d = 0)$

H_A: The mean difference in number of items remembered for ads with sexual images and ads without sexual images is not zero. $(\mu_d \neq 0)$

Paired data assumption: The data are paired by subject.
Randomization condition: The ads were in random order.

Normal population assumption: The histogram of differences is roughly unimodal and symmetric.



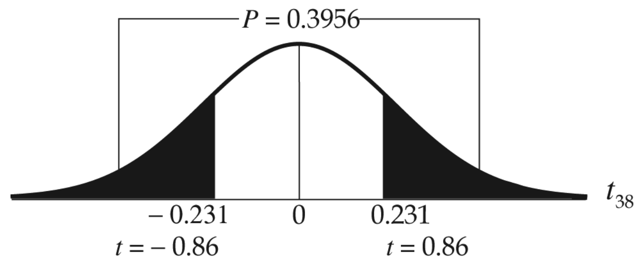
Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 39 – 1 = 38 degrees of freedom, $t_{38}\left(0, \frac{1.677}{\sqrt{39}}\right)$. We will use a paired t-test, with $\bar{d} = 0.231$.

Since the P-value = 0.3956 is high, we fail to reject the null hypothesis. There is no evidence of a mean difference in the number of objects remembered with ads with sexual images and without.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{0.231 - 0}{\frac{1.67747}{\sqrt{39}}}$$

$$t \approx 0.86$$



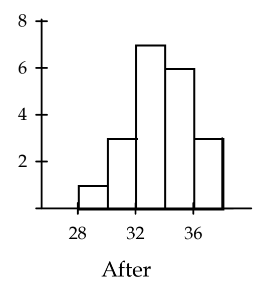
35. Strikes.

a) Since 60% of 50 pitches is 30 pitches, the Little Leaguers would have to throw an average of more than 30 strikes in order to give support to the claim made by the advertisements.

H_0: The mean number of strikes thrown by Little Leaguers who have completed the training is 30. $(\mu_A = 30)$

H_A: The mean number of strikes thrown by Little Leaguers who have completed the training is greater than 30. $(\mu_A > 30)$

Randomization Condition: Assume that these players are representative of all Little League pitchers.
Nearly Normal Condition: The histogram of the number of strikes thrown after the training is roughly unimodal and symmetric.



The pitchers in the sample threw a mean of 33.15 strikes, with a standard deviation of 2.32322 strikes. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean number of strikes thrown with a Student's t model, with 20 – 1 = 19 degrees of freedom, $t_{19}\left(30, \frac{2.32322}{\sqrt{20}}\right)$.

We will perform a one-sample t-test.

Since the P-value = 3.92×10^{-6} is very low, we reject the null hypothesis. There is strong evidence that the mean number of strikes that Little Leaguers can throw after the training is more than 30. (This test says nothing about the effectiveness of the training; just that Little Leaguers can throw more than 60% strikes on average after completing the training. This might not be an improvement.)

$$t = \frac{\bar{y}_A - \mu_0}{\frac{s_A}{\sqrt{n_A}}}$$

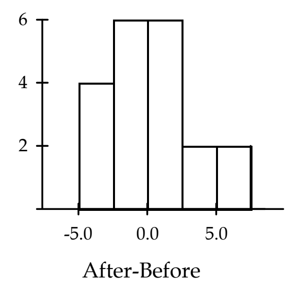
$$t = \frac{33.15 - 30}{\frac{2.32322}{\sqrt{20}}}$$

$$t = 6.06$$

b) H_0: The mean difference in number of strikes thrown before and after the training is zero. $(\mu_d = 0)$

H_A: The mean difference in number of strikes thrown before and after the training is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by pitcher.
Randomization condition: Assume that these players are representative of all Little League pitchers.
Normal population assumption: The histogram of differences is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 20 – 1 = 19 degrees of freedom, $t_{19}\left(0, \frac{3.32297}{\sqrt{19}}\right)$.

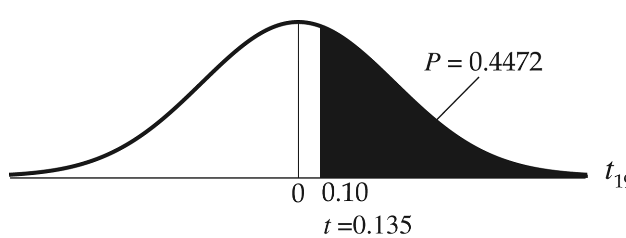
We will use a paired t-test, with $\bar{d} = 0.1$.

Since the P-value = 0.4472 is high, we fail to reject the null hypothesis. There is no evidence of a mean difference in number of strikes thrown before and after the training. The training does not appear to be effective.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

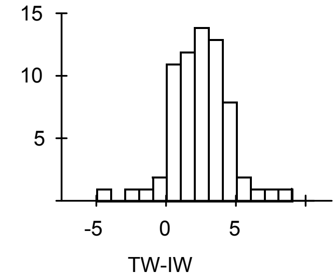
$$t = \frac{0.1 - 0}{\frac{3.32297}{\sqrt{20}}}$$

$$t \approx 0.135$$



c) The sign test has a P-value of 0.7597, which would lead us to fail to reject the hypothesis that the median difference was 0. This conclusion is consistent with the paired t-test.

36. Freshman 15, revisited.



Paired data assumption: The data are paired by student.
Randomization condition: The students matched the rest of the class in terms of demographic variables.
Normal population assumption: The histogram of differences is roughly unimodal and symmetric.

Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 68 – 1 = 67 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 1.91176 \pm t^*_{67}\left(\frac{2.12824}{\sqrt{68}}\right) \approx (1.40,\ 2.43)$$

We are 95% confident that freshmen at Cornell have a mean weight gain of between 1.40 and 2.43 pounds during the first 12 weeks of college. This interval does not contain zero, so there is evidence of a weight gain among freshmen, although it is quite small. These data certainly do not support the idea of the "freshman 15".

37. Wheelchair marathon 2013.

a) The data are certainly paired. Even if the individual times show a trend of improving speed over time, the differences may well be independent of each other. They are subject to random year-to-year fluctuations, and we may believe that these data are representative of similar races. We don't have any information with which to check the Nearly Normal condition.

b) $$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = -4.88 \pm t^*_{36}\left(\frac{35.227}{\sqrt{37}}\right) \approx (-16.63,\ 6.87)$$

We are 95% confident that the interval –16.63 to 6.87 minutes contains the true mean time difference for women's wheelchair times and men's running times.

c) The interval contains zero, so we would not reject a null hypothesis of no mean difference at a significance level of 0.05. We are unable to discern a difference between the female wheelchair times and the male running times.

38. Marathon start-up years 2010.

a) The data are certainly paired. Even if the individual times show a trend of improving speed over time, the differences may well be independent of each other. They are subject to random year-to-year fluctuations, and we may believe that these data are representative of similar races. After removing the three initial years, the remaining part of the histogram is a bit skewed and possibly bimodal, but we can probably use paired-*t* methods with caution. The sample size of 34 is still fairly large.

b) $\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = -13.40 \pm t^*_{33}\left(\frac{20.57}{\sqrt{34}}\right) \approx (-20.58, -6.22)$

We are 95% confident that female wheelchair marathoners average between 6.22 and 20.58 minutes faster than male runners in races such as this.

c) Since the interval does not contain zero, we would reject the null hypothesis of no difference at a significance level of 0.05. There is strong evidence that female wheelchair marathoners finish faster than male runners, on average.

39. BST.

a) **Paired data assumption:** We are testing the same cows, before and after injections of BST.
Randomization condition: These cows are likely to be representative of all Ayrshires.
Normal population assumption: We don't have the list of individual differences, so we can't look at a histogram. The sample is large, so we may proceed.

Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 60 – 1 = 59 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

b) $\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 14 \pm t^*_{59}\left(\frac{5.2}{\sqrt{60}}\right) \approx (12.66, 15.34)$

c) We are 95% confident that the mean increase in daily milk production for Ayshire cows after BST injection is between 12.66 and 15.34 pounds.

d) 25% of 47 pounds is 11.75 pounds. According to the interval generated in part b, the average increase in milk production is more than this, so the farmer can justify the extra expense for BST.

40. BST II.

Although the data from each herd of cows are paired, we are asked to compare the paired differences from each herd. The herds are independent, so we will use a two-sample *t*-test.

H_0: The mean increase in milk production due to BST is the same for both breeds. $\left(\mu_{dA} = \mu_{dJ} \text{ or } \mu_{dA} - \mu_{dJ} = 0\right)$

H_A: The mean increase in milk production due to BST is different for the two breeds. $\left(\mu_{dA} \neq \mu_{dJ} \text{ or } \mu_{dA} - \mu_{dJ} \neq 0\right)$

Independent groups assumption: The cows are from different herds.
Randomization condition: Assume that the cows are representative of their breeds.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the two sets of differences. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means (actually, the difference in mean differences!) with a Student's *t*-model, with 109.55 degrees of freedom (from the approximation formula).

We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{d}_A - \bar{d}_J) = \sqrt{\frac{5.2^2}{60} + \frac{4.8^2}{52}} \approx 0.945$.

The observed difference between the mean differences is 14 – 9 = 5.

$$t = \frac{(\bar{d}_A - \bar{d}_J) - (0)}{SE(\bar{d}_A - \bar{d}_J)}$$

$$t \approx \frac{5}{0.945}$$

$$t \approx 5.29$$

Since the *P*-value = 6.4×10^{-7} is very small, we reject the null hypothesis. There is strong evidence that the mean increase for each breed is different. The average increase for Ayshires is significantly greater than the average increase for Jerseys.

Chapter 24 – Comparing Counts

Section 24.1

1. Human births.

a) If there were no "seasonal effect" we would expects 25% of births to occur in each season. The expected number of births is 0.25(120) = 30 births per season.

b) $\chi^2 = \frac{(25-30)^2}{30} + \frac{(35-30)^2}{30} + \frac{(32-30)^2}{30} + \frac{(28-30)^2}{30} \approx 1.933$

c) There are 4 seasons, so there are 4 – 1 = 3 degrees of freedom.

2. Bank cards.

a) If the historical percentages hold, we would expect 0.60(200) = 120, 0.30(200) = 60, and 0.10(200) = 20 customers to apply for Silver, Gold, and Platinum cards, respectively.

b) $\chi^2 = \frac{(110-120)^2}{120} + \frac{(55-60)^2}{60} + \frac{(35-20)^2}{20} = 12.5$

c) There are 3 types of cards, so there are 3 – 1 = 2 degrees of freedom.

3. Human births, again.

a) The mean of the χ^2 distribution is the number of degrees of freedom, so we would expect the χ^2 statistic to be 3 if there were no seasonal effect.

b) Since 1.933 is less than the mean of 3, it does not seem large in comparison.

c) We should fail to reject the null hypothesis. These data do not provide evidence of a seasonal effect on human births.

d) The critical value of χ^2 with 3 degrees of freedom and $\alpha = 0.05$ is 7.815.

e) Since $\chi^2 = 1.933$ is less than the critical value, 7.815, we fail to reject the null hypothesis. There is no evidence of a seasonal effect on human births.

4. Bank cards, again.

a) The mean of the χ^2 distribution is the number of degrees of freedom, so we would expect the χ^2 statistic to be 2 if customers applied for bank cards according to the historical proportions.

b) The χ^2 statistic calculated in the previous exercise, 12.5, in much greater than the mean of 2.

c) We should reject the null hypothesis. We are unlikely to see a value of χ^2 this high if customers applied for bank cards according to the historical proportions.

d) The critical value of χ^2 with 2 degrees of freedom and $\alpha = 0.05$ is 5.991.

e) Since $\chi^2 = 12.5$ is greater than the critical value, 5.991, we reject the null hypothesis. There is strong evidence that the customers are not applying for bank cards according to the historical proportions.

5. Customer ages.

a) The null hypothesis is that the age distributions of the customers are the same at the two branches.

b) There are three age groups and one variable, location of the branch. This is a χ^2 test of homogeneity.

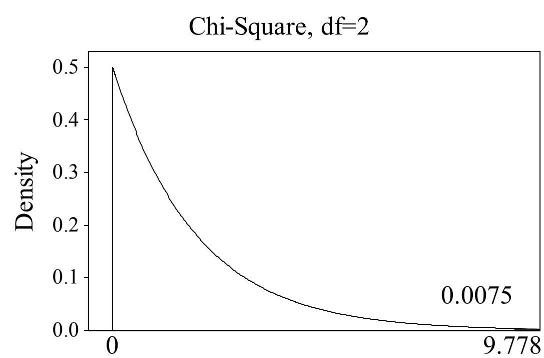
c)

Expected Counts	Age			
	Less than 30	30-55	56 or older	Total
In-Town Branch	25	45	30	100
Mall Branch	25	45	30	100
Total	50	90	60	200

d) $\chi^2 = \frac{(20-25)^2}{25} + \frac{(40-45)^2}{45} + \frac{(40-30)^2}{30} + \frac{(30-25)^2}{25} + \frac{(50-45)^2}{45} + \frac{(20-30)^2}{30} \approx 9.778$

e) There are 2 rows and 3 columns, so there are (2 – 1)(3 – 1) = 2 degrees of freedom.

f) The probability of having a χ^2 value over 9.778 with df = 2 is 0.0075.



g) Since the P-value is so low, reject the null hypothesis. There is strong evidence that the distribution of ages is not the same at the two branches. The mall branch had more customers under 30, and fewer customers 56 or older, than expected. The in-town branch had more customers 56 and older, and fewer customers under 30, than expected.

6. Bank cards, once more.

a) The null hypothesis is that the proportion of customers applying for the three types of cards is the same for each of the three mailings.

b) There are three mailings and one variable, the type of card. This is a χ^2 test of homogeneity.

c) Expected Counts

	Type of Card			
	Silver	Gold	Platinum	Total
Mailing 1	133.33	51.67	35	200
Mailing 2	133.33	51.67	35	200
Mailing 3	133.33	51.67	35	200
Total	340	155	105	600

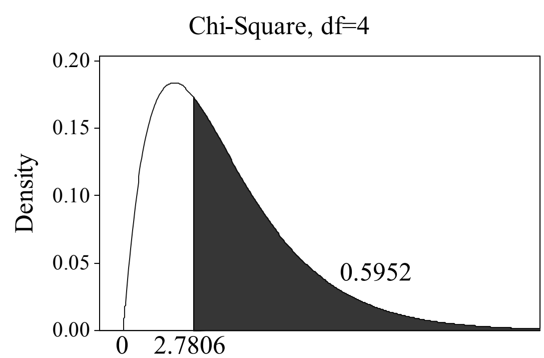
d)

$$\chi^2 = \frac{(120-113.33)^2}{113.33} + \frac{(50-51.67)^2}{51.67} + \frac{(30-35)^2}{35} + \frac{(115-113.33)^2}{113.33} + \frac{(50-51.67)^2}{51.67} + \frac{(35-35)^2}{35} + \frac{(105-113.33)^2}{113.33} + \frac{(55-51.67)^2}{51.67} + \frac{(40-35)^2}{35} \approx 2.7806$$

e) There are 3 rows and 3 columns, so there are (3 – 1)(3 – 1) = 4 degrees of freedom.

f) The probability of having a χ^2 value over 2.7806 with df = 4 is 0.5952.

g) Since the P-value is high, fail to reject the null hypothesis. There is no evidence that the proportions of card type differ for the three mailings.



Section 24.2

7. Human births, last time.

a) The standardized residuals are:

$$c_1 = \frac{(25-30)}{\sqrt{30}} \approx -0.913 \qquad c_2 = \frac{(35-30)}{\sqrt{30}} \approx 0.913$$

$$c_3 = \frac{(32-30)}{\sqrt{30}} \approx 0.365 \qquad c_4 = \frac{(28-30)}{\sqrt{30}} \approx -0.365$$

b) None of the standardized residuals are large. Since they are *z*-scores, they are actually quite small.

c) We did not reject the null hypothesis, so we should expect the standardized residuals to be relatively small.

8. Bank cards, last time.

a) The standardized residuals are:

$$c_1 = \frac{(110-120)}{\sqrt{120}} \approx -0.913 \qquad c_2 = \frac{(55-60)}{\sqrt{60}} \approx -0.645 \qquad c_3 = \frac{(35-20)}{\sqrt{20}} \approx 3.354$$

b) The standardized residual for the Platinum card is quite large. A z-score of 3.354 is large.

c) This large standardized residual suggests that the main difference in the proportion for this group is the increase in the number applying for the Platinum card.

Section 24.3

9. Internet use poll.

a) The null hypothesis is that the response to the question is independent of race.

Expected Counts	Did you use the Internet yesterday?		
	Yes	No	Total
White	2506.38	895.62	3402
Black	338.90	121.10	460
Hispanic/Other	445.73	159.27	605
Total	3291	1176	4467

b)

$$\chi^2 = \frac{(2546-2506.38)^2}{2506.38} + \frac{(856-895.62)^2}{895.62} + \frac{(314-338.90)^2}{338.90} + \frac{(146-121.10)^2}{121.10} + \frac{(431-445.73)^2}{445.73} + \frac{(174-159.27)^2}{159.27} \approx 11.176$$

c) There are 3 rows and 2 columns, so there are (3 – 1)(2 – 1) = 2 degrees of freedom.

d) The probability of having a χ^2 value over 11.176 with df = 2 is 0.0037.

e) Since the P-value is low, reject the null hypothesis. There is evidence of an association between the response to the question and race. Black respondents indicated that they had not used the Internet yesterday more frequently than we expected.

10. Internet use poll, II.

a) The null hypothesis is that the response to the question is independent of education level.

Expected Counts	Did you use the Internet yesterday? Yes	No	Total
Less than HS	250.13	89.87	340
High school	1090.28	391.72	1482
Some college	959.33	344.67	1304
College grad	1246.25	447.75	1694
Total	3546	1274	4820

b)

$$\chi^2 = \frac{(209-250.13)^2}{250.13} + \frac{(131-89.87)^2}{89.87} + \frac{(932-1090.28)^2}{1090.28} + \frac{(550-391.72)^2}{391.72} + \frac{(958-959.33)^2}{959.33} + \frac{(346-344.67)^2}{344.67} + \frac{(1447-1246.25)^2}{1246.25} + \frac{(247-447.75)^2}{447.75} \approx 234.881$$

c) There are 4 rows and 2 columns, so there are (4 – 1)(2 – 1) = 3 degrees of freedom.

d) The probability of having a χ^2 value over 234.881 with df = 3 is less than 0.0001.

e) Since the P-value is low, reject the null hypothesis. There is evidence of an association between the response to the question and education level. College grads were much more likely to answer "yes" than expected, while those with high school educations or less were much more likely to answer "no" than expected.

Chapter Exercises.

11. Which test?

a) Chi-square test of Independence. We have one sample and two variables. We want to see if the variable *account type* is independent of the variable *trade type.*

b) Some other statistics test. The variable *account size* is quantitative, not counts.

c) Chi-square test of Homogeneity. We have two samples (residential and non-residential students), and one variable, *courses.* We want to see if the distribution of *courses* is the same for the two groups.

12. Which test again?

a) Chi-square goodness-of-fit test. We want to see if the distribution of defects is uniform over the variable *day*.

b) Some other statistical test. *Cholesterol level* is a quantitative variable, not counts.

c) Chi-square test of Independence. We have data on two variables, *political leaning* and *major*, for one group of students.

13. Dice.

a) If the die were fair, you'd expect each face to show 10 times.

b) Use a chi-square test for goodness-of-fit. We are comparing the distribution of a single variable (outcome of a die roll) to an expected distribution.

c) H_0: The die is fair. (All faces have the same probability of coming up.)

H_A: The die is not fair. (Some faces are more or less likely to come up than others.)

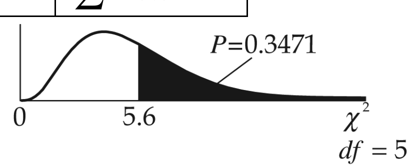
d) **Counted data condition:** We are counting the number of times each face comes up.
Randomization condition: Die rolls are random and independent of each other.
Expected cell frequency condition: We expect each face to come up 10 times, and 10 is greater than 5.

e) Under these conditions, the sampling distribution of the test statistic is χ^2 on 6 – 1 = 5 degrees of freedom. We will use a chi-square goodness-of-fit test.

f)

Face	Observed	Expected	Residual = $(Obs - Exp)$	$(Obs - Exp)^2$	Component = $\frac{(Obs - Exp)^2}{Exp}$
1	11	10	1	1	0.1
2	7	10	– 3	9	0.9
3	9	10	– 1	1	0.1
4	15	10	5	25	2.5
5	12	10	2	4	0.4
6	6	10	– 4	16	1.6
					$\sum = 5.6$

g) Since the *P*-value = 0.3471 is high, we fail to reject the null hypothesis. There is no evidence that the die is unfair.



14. M&M's

a) There are 29 + 23 + 12 + 14 + 8 + 20 = 106 M&M's in the bag. The expected number of M&M's of each color is: 106(0.20) = 21.2 red, 106(0.20) = 21.2 yellow, 106(0.10) = 10.6 orange, 106(0.10) = 10.6 blue, 106(0.10) = 10.6 green, and 106(0.30) = 31.8 brown.

b) Use a chi-square test for goodness-of-fit. We are comparing the distribution of a single variable (color) to an expected distribution.

c) H_0: The distribution of colors of M&M's is as specified by the company.

H_A: The distribution of colors of M&M's is different than specified by the company.

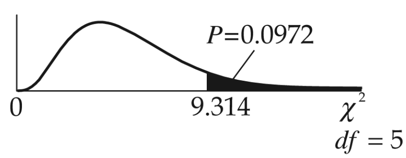
d) **Counted data condition:** The author counted the M&M's in the bag. **Randomization condition:** These M&M's are mixed thoroughly at the factory. **Expected cell frequency condition:** The expected counts (calculated in part b) are all greater than 5.

e) Since there are 6 different colors, there are 6 – 1 = 5 degrees of freedom.

f) Under these conditions, the sampling distribution of the test statistic is χ^2 on 6 – 1 = 5 degrees of freedom. We will use a chi-square goodness-of-fit test.

Color	Observed	Expected	Residual = $(Obs - Exp)$	$(Obs - Exp)^2$	Component = $\frac{(Obs - Exp)^2}{Exp}$
yellow	29	21.2	7.8	60.84	2.8698
red	23	21.2	1.8	3.24	0.1528
orange	12	10.6	1.4	1.96	0.1849
blue	14	10.6	3.4	11.56	1.0906
green	8	10.6	– 2.6	6.76	0.6377
brown	20	31.8	– 11.8	139.24	4.3786
					$\sum \approx$ 9.314

$\chi^2 = 9.314$. Since the P-value = 0.0972 is high, we fail to reject the null hypothesis.



g) There is no evidence that the distribution of colors is anything other than the distribution specified by the company.

15. Nuts.

a) The weights of the nuts are quantitative. Chi-square goodness-of-fit requires counts.

b) In order to use a chi-square test, you could count the number of each type of nut. However, it's not clear whether the company's claim was a percentage by number or a percentage by weight.

16. Mileage.

The average number of miles traveled is quantitative date, not categorical. Chi-square is for comparing counts.

17. NYPD and race.

H_0: The distribution of ethnicities in the police department represents the distribution of ethnicities of the youth of New York City.

H_A: The distribution of ethnicities in the police department does not represent the distribution of ethnicities of the youth of New York City.

Counted data condition: The percentages reported must be converted to counts.
Randomization condition: Assume that the current NYPD is representative of recent departments with respect to ethnicity.
Expected cell frequency condition: The expected counts are all much greater than 5.

(Note: The observed counts should be whole numbers. They are actual policemen. The expected counts may be decimals, since they behave like averages.)

Ethnicity	Observed	Expected
White	16965	7644.852
Black	3796	7383.042
Latino	5001	8247.015
Asian	367	2382.471
Other	52	523.620

Under these conditions, the sampling distribution of the test statistic is χ^2 on 5 – 1 = 4 degrees of freedom. We will use a chi-square goodness-of-fit test.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} = \frac{(16965 - 7644.852)^2}{7644.852} + \frac{(3796 - 7383.042)^2}{7383.042} + \frac{(5001 - 8247.015)^2}{8247.015}$$
$$+ \frac{(367 - 2382.471)^2}{2382.471} + \frac{(52 - 523.620)^2}{523.620} \approx 16,500$$

With χ^2 of over 16,500, on 4 degrees of freedom, the *P*-value is essentially 0, so we reject the null hypothesis. There is strong evidence that the distribution of ethnicities of NYPD officers does not represent the distribution of ethnicities of the youth of New York City. Specifically, the proportion of white officers is much higher than the proportion of white youth in the community. As one might expect, there are also lower proportions of officers who are black, Latino, Asian, and other ethnicities than we see in the youth in the community.

18. Violence against women 2009.

H_0: The weapon use rates in murders of women have the same distribution as the weapon use rates of all murders.

H_A: The weapon use rates in murders of women have a different distribution than the weapon use rates of all murders.

Counted data condition: The percentages reported must be converted to counts.
Randomization condition: Assume that the weapon use rates from 2009 are representative of the weapon use rates for all recent years.
Expected cell frequency condition: The expected counts are all much greater than 5.

weapon	Observed	Expected
guns	861	1048.636
knives	364	216.674
other	214	277.872
personal	215	110.818

(Note: The observed counts should be whole numbers. They are actual murders. The expected counts may be decimals, since they behave like averages.)

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 – 1 = 3 degrees of freedom. We will use a chi-square goodness-of-fit test.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} = \frac{(861-1048.636)^2}{1048.636} + \frac{(364-216.674)^2}{216.674} + \frac{(214-277.872)^2}{277.872} + \frac{(215-110.818)^2}{110.818} \approx 246.4$$

With $\chi^2 \approx 246.4$, on 3 degrees of freedom, the *P*-value is essentially 0, so we reject the null hypothesis. There is strong evidence that the distribution of weapon use rates is different for murders of women than for all murders. Women are much more likely to be killed by personal attacks and knives, and less likely to be killed with guns or other weapons.

19. Fruit flies.

a) H_0: The ratio of traits in this type of fruit fly is 9:3:3:1, as genetic theory predicts.

H_A: The ratio of traits in this type of fruit fly is not 9:3:3:1.

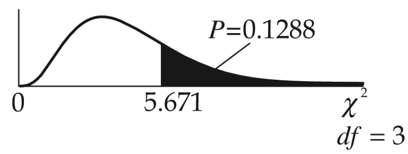
Counted data condition: The data are counts.
Randomization condition: Assume that these flies are representative of all fruit flies of this type.
Expected cell frequency condition: The expected counts are all greater than 5.

trait	Observed	Expected
YN	59	56.25
YS	20	18.75
EN	11	18.75
ES	10	6.25

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 – 1 = 3 degrees of freedom. We will use a chi-square goodness-of-fit test.

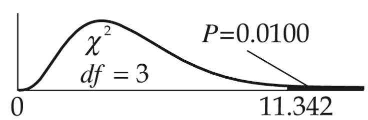
$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} = \frac{(59-56.25)^2}{56.25} + \frac{(20-18.75)^2}{18.75} + \frac{(11-18.75)^2}{18.75} + \frac{(10-6.25)^2}{6.25} \approx 5.671$$

With $\chi^2 \approx 5.671$, on 3 degrees of freedom, the P-value = 0.1288 is high, so we fail to reject the null hypothesis. There is no evidence that the ratio of traits is different than the theoretical ratio predicted by the genetic model. The observed results are consistent with the genetic model.



b) With $\chi^2 \approx 11.342$, on 3 degrees of freedom, the P-value = 0.0100 is low, so we reject the null hypothesis. There is strong evidence that the ratio of traits is different than the theoretical ratio predicted by the genetic model. Specifically, there is evidence that the normal wing length occurs less frequently than expected and the short wing length occurs more frequently than expected.

trait	Observed	Expected
YN	118	112.5
YS	40	37.5
EN	22	37.5
ES	20	12.5



c) At first, this seems like a contradiction. We have two samples with exactly the same ratio of traits. The smaller of the two provides no evidence of a difference, yet the larger one provides strong evidence of a difference. This is explained by the sample size. In general, large samples decrease the proportion of variation from the true ratio. Because of the relatively small sample in the first test, we are unwilling to say that there is a difference. There just isn't enough evidence. But the larger sample allows us to be more certain about the difference.

20. Pi.

H_0: Digits of π are uniformly distributed (all occur with frequency 1/10).

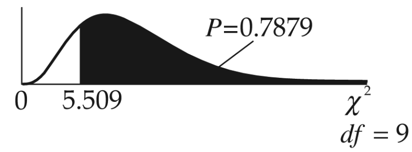
H_A: Digits of π are not uniformly distributed.

Counted data condition: The data are counts.
Randomization condition: Assume that the first million digits of π are representative of all digits.
Expected cell frequency condition: The expected counts are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 10 – 1 = 9 degrees of freedom. We will use a chi-square goodness-of-fit test.

Digit	Observed	Expected
0	99959	100000
1	99758	100000
2	100026	100000
3	100229	100000
4	100230	100000
5	100359	100000
6	99548	100000
7	99800	100000
8	99985	100000
9	100106	100000

With $\chi^2 \approx 5.509$, on 9 degrees of freedom, the P-value = 0.7879 is high, so we fail to reject the null hypothesis. There is no evidence that the digits of π are not uniformly distributed. These data are consistent with the null hypothesis.



21. Hurricane frequencies.

a) We would expect 96/16 = 6 hurricanes per time period.

b) We are comparing the distribution of the number of hurricanes, a single variable, to a theoretical distribution. A Chi-square test for goodness-of-fit is appropriate.

c) H_0: The number of large hurricanes remains constant over decades.

H_A: The number of large hurricanes has changed.

d) There are 16 time periods, so there are 16 – 1 = 15 degrees of freedom.

e) $P(\chi^2_{df=15} > 12.67) \approx 0.63$

f) The very high P-value means that these data offer no evidence that the number of hurricanes large hurricanes has changed.

22. Lottery numbers.

a) We are comparing the distribution of the number of times each lottery number has occurred, either as a regular number, or as the Powerball. We will use a Chi-square test for goodness-of-fit.

b) We expect the Powerballs to be distributed uniformly over the 49 numbers. We expect each number to be the bonus ball 655/49 = 13.367 times.

c) H_0: All numbers are equally likely to be the bonus ball.

H_A: Some numbers are more likely than others to be the bonus ball.

d) There are 49 numbers, so there are 49 – 1 = 48 degrees of freedom.

e) $P(\chi^2_{df=48} > 34.5) \approx 0.93$

f) The very high P-value means that these data offer no evidence that some lottery numbers are more likely than others to be the bonus ball.

23. Childbirth, part 1.

a) There are two variables, breastfeeding and having an epidural, from a single group of births. We will perform a Chi-square test for Independence.

b) H_0: Breastfeeding success is independent of having an epidural.

H_A: There is an association between breastfeeding success and having an epidural.

24. Does your doctor know?

a) There is one variable, whether or not statistics is used, over three time periods in which the articles were published. We will perform a Chi Square test of Homogeneity.

b) H_0: The same proportion of articles used statistics in the three time periods.

H_A: The proportion of articles that used statistics was different in the three time periods.

25. Childbirth, part 2.

a) The table has 2 rows and 2 columns, so there are $(2-1)\times(2-1)=1$ degree of freedom.

b) We expect $\frac{474}{1178} \approx 40.2\%$ of all babies to not be breastfeeding after 6 months, so we expect that 40.2% of the 396 epidural babies, or 159.34, to not be breastfeeding after 6 months.

c) Breastfeeding behavior should be independent for these babies. They are fewer than 10% of all babies, and we assume they are representative of all babies. We have counts, and all the expected cells are at least 5.

26. Does your doctor know? (part 2).

a) The table has 2 rows and 3 columns, so there are $(2-1)\times(3-1)=2$ degrees of freedom.

b) We expect $\frac{144}{758} \approx 19\%$ of all articles to contain no statistics, so we expect 19% of the 115 articles from 1989, or 21.85, to contain no statistics.

c) These are counted data. One article shouldn't affect another article (except perhaps for a follow-up article to another article included in the study). We can regard the selected years as representative of other years, and the authors seem to want to regard these articles as representative of those appearing in similar-quality medical journals, so they are fewer than 10% of all articles. All expected counts are greater than 5.

27. Childbirth, part 3.

a) $\frac{(Obs-Exp)^2}{Exp} = \frac{(190-159.34)^2}{159.34} = 5.90$ **b)** $P(\chi^2_{df=1} > 14.87) < 0.005$

c) The *P*-value is very low, so reject the null hypothesis. There's strong evidence of an association between having an epidural and subsequent success in breastfeeding.

28. Does your doctor know? (part 3).

a) $\frac{(Obs-Exp)^2}{Exp} = \frac{(14-21.85)^2}{21.85} = 2.82$ **b)** $P(\chi^2_{df=2} > 25.28) < 0.001$

c) The P-value is very low, so reject the null hypothesis. There's strong evidence that the proportion of medical journal articles that contain statistics is different for the three time periods.

29. Childbirth, part 4.

a) $c = \frac{Obs-Exp}{\sqrt{Exp}} = \frac{190-159.34}{\sqrt{159.34}} = 2.43$

b) It appears that babies whose mothers had epidurals during childbirth are much more likely to be breastfeeding 6 months later.

30. Does your doctor know? (part 4).

a) $c = \frac{Obs-Exp}{\sqrt{Exp}} = \frac{14-21.85}{\sqrt{21.85}} = -1.68$

b) The residuals for No stats are decreasing and those for Stats are increasing over time, indicating that, over time, a smaller proportion of articles are appearing without statistics.

31. Childbirth, part 5.

These factors would not have been mutually exclusive. There would be yes or no responses for every baby for each.

32. Does your doctor know? (part 5).

These methods would not have been mutually exclusive. Articles might use more than one statistical method.

33. Titanic.

a) $P(\text{crew}) = \frac{885}{2201} \approx 0.402$ **b)** $P(\text{third and alive}) = \frac{178}{2201} \approx 0.081$

c) $P(\text{alive} \mid \text{first}) = \frac{P(\text{alive and first})}{\text{P(first)}} = \frac{^{202}\!/_{2201}}{^{325}\!/_{2201}} = \frac{202}{325} \approx 0.622$

d) The overall chance of survival is $\frac{710}{2201} \approx 0.323$, so we would expect about 32.3% of the crew, or about 285.48 members of the crew, to survive.

e) H_0: Survival was independent of status on the ship.

H_A: Survival depended on status on the ship.

f) The table has 2 rows and 4 columns, so there are $(2-1)\times(4-1)=3$ degrees of freedom.

g) With $\chi^2 \approx 187.8$, on 3 degrees of freedom, the *P*-value is essentially 0, so we reject the null hypothesis. There is strong evidence survival depended on status. First-class passengers were more likely to survive than any other class or crew.

34. NYPD and sex discrimination.

a) $P(\text{female}) = \dfrac{5613}{37,379} \approx 0.150$

b) $P(\text{detective}) = \dfrac{4864}{37,379} \approx 0.130$

c) The overall percentage of females is 15%, so we would expect about 15% of the detectives, or about 729.6 detectives, to be female.

d) We have one group, categorized according to two variables, rank and gender, so we will perform a chi-square test for independence.

e) H_0: Rank is independent of gender in the NYPD.

H_A: Rank is associated with gender in the NYPD.

Expected counts

Rank	Male	Female
Officer	22249.5	3931.5
Detective	4133.6	730.4
Sergeant	3665.3	647.7
Lieutenant	1208.5	213.5
Captain	315.3	55.7
Higher ranks	193.8	34.2

f) **Counted data condition:** The data are counts.
Randomization condition: These data are not a random sample, but all NYPD officers. Assume that these officers are representative with respect to the recent distribution of sex and rank in the NYPD.
Expected cell frequency condition: The expected counts are all greater than 5.

g) The table has 6 rows and 2 columns, so there are $(6-1)\times(2-1)=5$ degrees of freedom.

h) With $\chi^2 = 343.9$,the P-value is very low. We reject the null hypothesis. There is strong evidence of an association between the sex and rank of NYPD officers.

35. Titanic again.

First class passengers were most likely to survive, while third class passengers and crew were under-represented among the survivors.

36. NYPD again.

Women are over-represented at the lower ranks and under-represented at every rank from sergeant up.

37. Cranberry juice.

a) This is an experiment. Volunteers were assigned to drink a different beverage.

b) We are concerned with the proportion of urinary tract infections among three different groups. We will use a chi-square test for homogeneity.

c) H_0: The proportion of urinary tract infection is the same for each group.

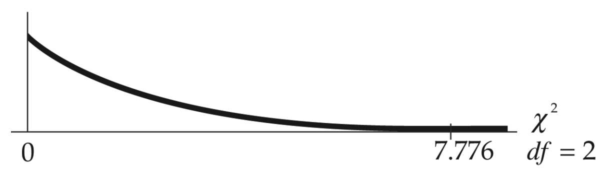
H_A: The proportion of urinary tract infection is different among the groups.

d) **Counted data condition:** The data are counts.
Randomization condition: Although not specifically stated, we will assume that the women were randomly assigned to treatments.
Expected cell frequency condition: The expected counts are all greater than 5.

	Cranberry (*Obs / Exp*)	Lactobacillus (*Obs / Exp*)	Control (*Obs / Exp*)
Infection	8 / 15.333	20 / 15.333	18 / 15.333
No infection	42 / 34.667	30 / 34.667	32 / 34.667

e) The table has 2 rows and 3 columns, so there are $(2-1)\times(3-1)=2$ degrees of freedom.

f) $\chi^2 = \sum_{all\,cells} \frac{(Obs-Exp)^2}{Exp} \approx 7.776$



P-value ≈ 0.020.

g) Since the P-value is low, we reject the null hypothesis. There is strong evidence of difference in the proportion of urinary tract infections for cranberry juice drinkers, lactobacillus drinkers, and women that drink neither of the two beverages.

h) A table of the standardized residuals is below, calculated by using $c = \frac{Obs-Exp}{\sqrt{Exp}}$.

	Cranberry	Lactobacillus	Control
Infection	–1.87276	1.191759	0.681005
No infection	1.245505	–0.79259	–0.45291

There is evidence that women who drink cranberry juice are less likely to develop urinary tract infections, and women who drank lactobacillus are more likely to develop urinary tract infections.

38. Car origins.

a) We have two groups, staff and students (selected from different lots), and we are concerned with the distribution of one variable, origin of car. We will perform a chi-square test for homogeneity.

b) H_0: The distribution of car origin is the same for students and staff.

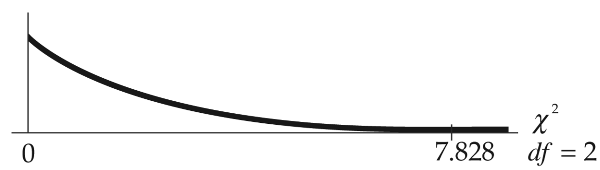
H_A: The distribution of car origin is different for students and staff.

c) **Counted data condition:** The data are counts.
Randomization condition: Cars were surveyed randomly.
Expected cell frequency condition: The expected counts are all greater than 5.

	Driver	
Origin	**Student** *(Obs / Exp)*	**Staff** *(Obs / Exp)*
American	107 / 115.15	105 / 96.847
European	33 / 24.443	12 / 20.557
Asian	55 / 55.404	47 / 46.596

d) Under these conditions, the sampling distribution of the test statistic is χ^2 on 2 degrees of freedom. We will use a chi-square test for homogeneity.

With $\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 7.828$, on 2 degrees of freedom, the P-value ≈ 0.020.



e) Since P-value = 0.020 is low, we reject the null hypothesis. There is strong evidence that the distribution of car origins at this university differs between students and staff. Students are more likely to drive European cars than staff and less likely than staff to drive American cars.

39. Montana.

a) We have one group, categorized according to two variables, political party and being male or female, so we will perform a chi-square test for independence.

b) H_0: Political party is independent of being male or female in Montana.

H_A: There is an association between political party and being male or female in Montana.

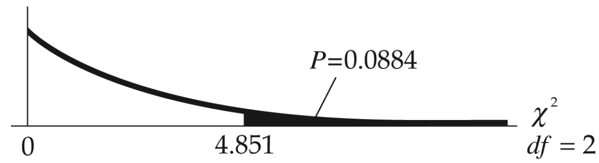
c) **Counted data condition:** The data are counts.
Randomization condition: Although not specifically stated, we will assume that the poll was conducted randomly.
Expected cell frequency condition: The expected counts are all greater than 5.

	Democrat *(Obs / Exp)*	**Republican** *(Obs / Exp)*	**Independent** *(Obs / Exp)*
Male	36 / 43.663	45 / 40.545	24 / 20.792
Female	48 / 40.337	33 / 37.455	16 / 19.208

d) Under these conditions, the sampling distribution of the test statistic is χ^2 on 2 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 4.851$$

The P-value ≈ 0.0884



e) Since the P-value ≈ 0.0884 is fairly high, we fail to reject the null hypothesis. There is little evidence of an association between being male or female and political party in Montana.

40. Fish diet.

a) This is an observational prospective study. Swedish men were selected, and then followed for 30 years.

b) We have one group, categorized according to two variables, fish consumption and incidence of prostate cancer, so we will perform a chi-square test for independence.

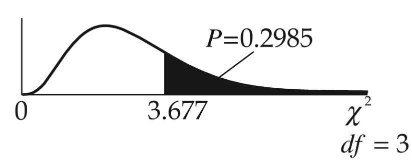
c) H_0: Prostate cancer and fish consumption are independent.

H_A: There is an association between prostate cancer and fish consumption.

Counted data condition: The data are counts.
Randomization condition: Assume that these men are representative of all men.
Expected cell frequency condition: The expected counts are all greater than 5.

Fish Consumption	**Prostate Cancer *(Obs / Exp)***	**No Prostate Cancer *(Obs / Exp)***
Never/Seldom	14 / 9.21	110 / 114.79
Small part	201 /194.74	2420 / 2426.3
Moderate part	209 / 221.26	2769 / 2756.7
Large part	42 / 40.79	507 / 508.21

Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom. We will use a chi-square test for independence.



$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 3.677,$$

and the P-value ≈ 0.2985.

Since the P-value ≈ 0.2985 is high, we fail to reject the null hypothesis. There is no evidence of an association between prostate cancer and fish consumption.

d) This does not prove that eating fish does not prevent prostate cancer. There is merely a lack of evidence of a relationship. Furthermore, association (or lack thereof) does not prove a cause-and-effect relationship. We would need to conduct a controlled experiment before anything could be proven.

41. Montana revisited.

H_0: Political party is independent of region in Montana.

H_A: There is an association between political party and region in Montana.

Counted data condition: The data are counts.
Randomization condition: Although not specifically stated, we will assume that the poll was conducted randomly.

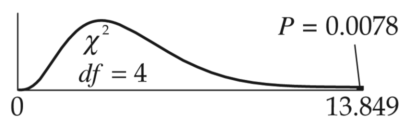
Expected cell frequency condition: All expected counts are greater than 5.

	Democrat *(Obs/Exp)*	**Republican** *(Obs/Exp)*	**Independent** *(Obs/Exp)*
West	39 / 28.277	17 / 26.257	12 / 13.465
Northeast	15 / 23.703	30 / 22.01	12 / 11.287
Southeast	30 /32.02	31 / 29.733	16 / 15.248

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 13.849,$$

and the P-value ≈ 0.0078



Since the P-value ≈ 0.0078 is low, reject the null hypothesis. There is strong evidence of an association between region and political party in Montana. Residents in the West are more likely to be Democrats than Republicans, and residents in the Northeast are more likely to be Republicans than Democrats.

42. Working parents.

a) This is a survey of adults.

b) We have two groups, males and females, and we are concerned with the distribution of one variable, attitude about the child care options. We will perform a chi-square test for homogeneity.

c) H_0: The distribution of attitudes about child care is the same for men and women.

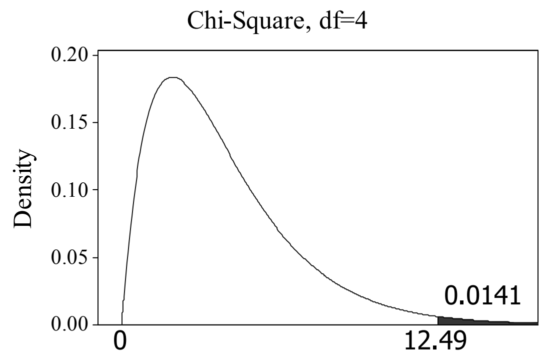
H_A: The distribution of attitudes about child care is not the same for men and women.

Counted data condition: The data are counts.
Randomization condition: Adults were surveyed randomly.
Expected cell frequency condition: The expected counts are all greater than 5.

	Male (Obs / Exp)	Female (Obs / Exp)
Both work full time	161 / 150.5	140/ 150.5
One works full time, other part time	259/ 283.5	308 / 283.5
One works, other stays at home	189 / 175	161/ 175
Both parents work part time	49 / 56	63 / 56
No opinion	42 / 35	28 / 35

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 degrees of freedom. We will use a chi-square test for homogeneity.

With $\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 12.490$, the P-value ≈ 0.0141.



Since P-value = 0.0141 is low, we reject the null hypothesis. There is strong evidence of a difference in the distribution of attitudes about child care options between men and women. Men are more likely than woman to prefer a situation where both parents work full time, and women are more likely to prefer a situation where one parent works full time and the other works part time.

43. Grades.

a) We have two groups, students of Professor Alpha and students of Professor Beta, and we are concerned with the distribution of one variable, grade. We will perform a chi-square test for homogeneity.

b) H_0: The distribution of grades is the same for the two professors.

H_A: The distribution of grades is different for the two professors.

c) The expected counts are organized in the table below:

	Prof. Alpha	Prof. Beta
A	6.667	5.333
B	12.778	10.222
C	12.222	9.778
D	6.111	4.889
F	2.222	1.778

Since three cells have expected counts less than 5, the chi-square procedures are not appropriate. Cells would have to be combined in order to proceed. (We will do this in another exercise.)

44. Full moon.

a) We have two groups, weeks of six full moons and six other weeks, and we are concerned with the distribution of one variable, type of offense. We will perform a chi-square test for homogeneity.

b) H_0: The distribution of type of offense is the same for full moon weeks as it is for weeks in which there is not a full moon.

H_A: The distribution of type of offense is different for full moon weeks than it is for weeks in which there is not a full moon.

c) The expected counts are organized in the table below:

Offense	Full Moon	Not Full
Violent	2.558	2.442
Property	19.442	18.558
Drugs / Alcohol	23.535	22.465
Domestic Abuse	12.791	12.209
Other offenses	7.674	7.326

Since two cells have expected counts less than 5, the chi-square procedures are not appropriate. Cells would have to be combined in order to proceed. (We will do this in another exercise.)

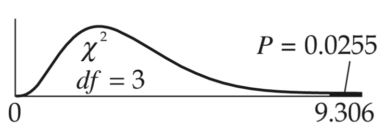
45. Grades again.

a) **Counted data condition:** The data are counts.
Randomization condition: Assume that these students are representative of all students that have ever taken courses from the professors.
Expected cell frequency condition: The expected counts are all greater than 5.

	Prof. Alpha *(Obs / Exp)*	Prof. Beta *(Obs / Exp)*
A	3 / 6.667	9 / 5.333
B	11 / 12.778	12 / 10.222
C	14 / 12.222	8 / 9.778
Below C	12 / 8.333	3 / 6.667

b) Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom, instead of 4 degrees of freedom before the change in the table. We will use a chi-square test for homogeneity.

c) With $\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 9.306$, the P-value ≈ 0.0255.



Since the P-value = 0.0255 is low, we reject the null hypothesis. There is evidence that the grade distributions for the two professors are different. Professor Alpha gives fewer As and more grades below C than Professor Beta.

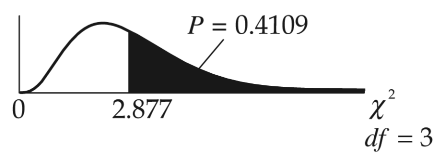
46. Full moon, next phase.

a) **Counted data condition:** The data are counts.
Randomization condition: Assume that these weeks are representative of all weeks.
Expected cell frequency condition: It seems reasonable to combine the violent offenses and domestic abuse, since both involve some sort of violence. Combining violent crimes with the "other offenses" is okay, but that may put very minor offenses in with violent offenses, which doesn't seem right. Once the cells are combined, all expected counts are greater than 5.

Offense	Full Moon (*Obs/Exp*)	Not Full (*Obs/Exp*)
Violent/ Domestic Abuse	13 / 15.349	17 / 14.651
Property	17 / 19.442	21 / 18.558
Drugs/Alcohol	27 / 23.535	19 / 22.465
Other offenses	9 / 7.674	6 / 7.326

b) Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom, instead of 4 degrees of freedom before the change in the table. We will use a chi-square test for homogeneity.

With $\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 2.877$, the P-value $\approx$ 0.4109.



Since P-value = 0.4109 is high, we fail to reject the null hypothesis. There is no evidence that the distribution of offenses is different during the full moon than during other phases.

47. Racial steering.

H_0: There is no association between race and section of the complex in which people live.

H_A: There is an association between race and section of the complex in which people live.

Counted data condition: The data are counts.
Randomization condition: Assume that the recently rented apartments are representative of all apartments in the complex.
Expected cell frequency condition: The expected counts are all greater than 5.

	White (*Obs/Exp*)	Black (*Obs/Exp*)
Section A	87 / 76.179	8 / 18.821
Section B	83 / 93.821	34 / 23.179

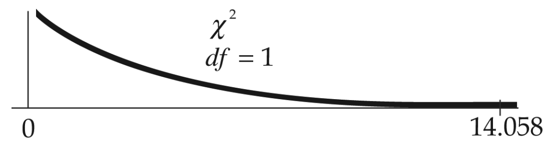
Under these conditions, the sampling distribution of the test statistic is χ^2 on 1 degree of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx \frac{(87-76.179)^2}{76.179} + \frac{(8-18.821)^2}{18.821} + \frac{(83-93.821)^2}{93.821} + \frac{(34-23.179)^2}{23.179}$$

$$\approx 1.5371 + 6.2215 + 1.2481 + 5.0517$$

$$\approx 14.058$$

With $\chi^2 \approx 14.058$, on 1 degree of freedom, the P-value ≈ 0.0002.



Since the P-value ≈ 0.0002 is low, we reject the null hypothesis. There is strong evidence of an association between race and the section of the apartment complex in which people live. An examination of the components shows us that whites are much more likely to rent in Section A (component = 6.2215), and blacks are much more likely to rent in Section B (component = 5.0517).

48. Survival on the *Titanic*.

H_0: Survival was independent of gender on the *Titanic*.

H_A: There is an association between survival and gender on the *Titanic*.

Counted data condition: The data are counts.
Randomization condition: We have the entire population of the *Titanic*.
Expected cell frequency condition: The expected counts are all greater than 5.

	Female ***(Obs / Exp)***	**Male** ***(Obs / Exp)***
Alive	343 / 151.613	367 / 558.387
Dead	127 / 318.387	1364 / 1172.613

Under these conditions, the sampling distribution of the test statistic is χ^2 on 1 degree of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx \frac{(343-151.613)^2}{151.613} + \frac{(367-558.387)^2}{558.387}$$

$$+ \frac{(127-318.387)^2}{318.387} + \frac{(1364-1172.613)^2}{1172.613}$$

$$\approx 241.5953 + 65.5978 + 115.0455 + 31.2371 \approx 453.476$$

With $\chi^2 \approx 453.476$, on 1 degree of freedom, the P-value is essentially 0.

Since the P-value is so low, we reject the null hypothesis. There is strong evidence of an association between survival and gender on the *Titanic*. Females were much more likely to survive than males.

49. Steering revisited.

a) H_0: The proportion of whites who live in Section A is the same as the proportion of blacks who live in Section A. $\left(p_W = p_B \text{ or } p_W - p_B = 0\right)$

H_A: The proportion of whites who live in Section A is different than the proportion of blacks who live in Section A. $\left(p_W \neq p_B \text{ or } p_W - p_B \neq 0\right)$

Independence assumption: Assume that people rent independent of the section.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (white) = 87, $n\hat{q}$ (white) = 83, $n\hat{p}$ (black) = 8, and $n\hat{q}$ (black) = 34. These are not all greater than 10, since the number of black renters in Section A is only 8, but it is close to 10, and the others are large. It should be safe to proceed.

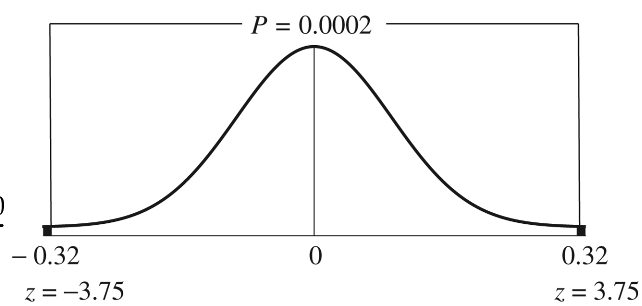
Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_W - \hat{p}_B\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_W} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_B}} = \sqrt{\frac{\left(\frac{95}{212}\right)\left(\frac{117}{212}\right)}{170} + \frac{\left(\frac{95}{212}\right)\left(\frac{117}{212}\right)}{42}} \approx 0.0856915.$$

The observed difference between the proportions is:
0.5117647 – 0.1904762 = 0.3212885.

$$z = \frac{0.3212885 - 0}{0.0856915}$$

$z \approx 3.74936$



(You have to use a *ridiculous* number of decimal places to get this to come out “right”. This is to done to illustrate the point of the question. DO NOT DO THIS! Use technology.)

Since the *P*-value = 0.0002 is low, we reject the null hypothesis. There is strong evidence of a difference in the proportion of whites and blacks living in Section A. The evidence suggests that the proportion of all whites living in Section A is much higher than the proportion of all black residents living in Section A.

The value of z for this test was approximately 3.74936. $z^2 \approx (3.74936)^2 \approx 14.058$, the same as the value for χ^2 in the previous exercise.

b) The resulting P-values were both approximately 0.0002. The two tests are equivalent.

50. Survival on the *Titanic*, one more time.

a) H_0: The proportion of females who survived is the same as the proportion of males who survived. $\left(p_F = p_M \text{ or } p_F - p_M = 0\right)$

H_A: The proportion of females who survived is different than the proportion of males who survived. $\left(p_F \neq p_M \text{ or } p_F - p_M \neq 0\right)$

Independence assumption: Assume that survival and sex are independent.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (female) = 343, $n\hat{q}$ (female) = 127, $n\hat{p}$ (male) = 367, and $n\hat{q}$ (male) = 1364. All are greater than 10.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_F - \hat{p}_M\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_F} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_M}}$$

$$= \sqrt{\frac{\left(\frac{710}{2201}\right)\left(\frac{1491}{2201}\right)}{470} + \frac{\left(\frac{710}{2201}\right)\left(\frac{1491}{2201}\right)}{1731}} \approx 0.0243142295$$

The observed difference between the proportions is:
0.729787234 – 0.212016176 = 0.517771058.

$$z = \frac{0.517771058 - 0}{0.0243142295}$$
$$z \approx 21.29498$$

(You have to use a *ridiculous* number of decimal places to get this to come out "right". This is to done to illustrate the point of the question. DO NOT DO THIS! Use technology.)

Since the P-value is essentially 0, we reject the null hypothesis. There is strong evidence of a difference between the proportions of women and men who survived on the Titanic. Women survived at a much higher rate.

b) The value of z for this test was approximately 21.29498.
$z^2 \approx (21.29498)^2 \approx 453.476$, the same as the value for χ^2 in the previous exercise.

c) The resulting P-values were both essentially 0. The two tests are equivalent.

51. Pregnancies.

H_0: Pregnancy outcome is independent of age.

H_A: There is an association between pregnancy outcome and age.

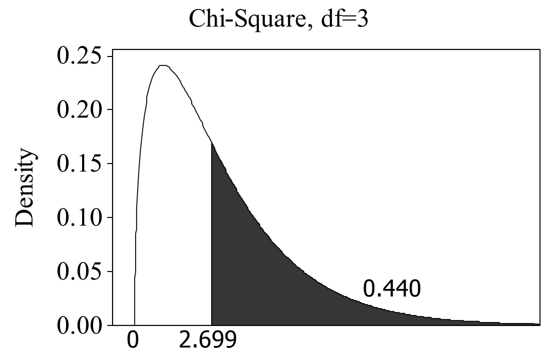
Counted data condition: The data are counts.
Randomization condition: Assume that these women are representative of all pregnant women.
Expected cell frequency condition: The expected counts are all greater than 5.

	Early Preterm (Obs/Exp)	Late Preterm (Obs/Exp)
Under 20	129 / 116.55	270 / 282.45
20 – 29	243 / 249.75	612 / 605.25
30 – 39	165 / 172.05	424 / 416.95
40 and over	18 / 16.65	39 / 40.35

Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom. We will use a chi-square test for independence.

With $\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 2.699$, on 3 degrees of freedom, the P-value ≈ 0.440.

Since the P-value is high, we fail to reject the null hypothesis. There is no evidence of an association between pregnancy outcome and age.



52. Education by age.

H_0: The distribution of education level attained is the same for different age groups.

H_A: The distribution of education level attained is different for different age groups.

Counted data condition: The data are counts.
Randomization condition: Assume that the sample was taken randomly.
Expected cell frequency condition: The expected counts are all greater than 5.

	25 – 34 (Obs/Exp)	35 – 44 (Obs/Exp)	45 – 54 (Obs/Exp)	55 – 64 (Obs/Exp)	65 and older (Obs/Exp)
Not HS Grad	27 / 60.2	50 / 60.2	52 / 60.2	71 / 60.2	101 / 60.2
HS	82 / 66.2	19 / 66.2	88 / 66.2	83 / 66.2	59 / 66.2
1 – 3 years college	43 / 33	56 / 33	26 / 33	20 / 33	20 / 33
4+ years college	48 / 40.6	75 / 40.6	34 / 40.6	26 / 40.6	20 / 40.6

Under these conditions, the sampling distribution of the test statistic is χ^2 on 12 degrees of freedom. We will use a chi-square test for homogeneity. (There are 200 people in each age group, an indication that we are examining 5 age groups, with respect to one variable, education level attained.)

With $\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 178.453$, on 12 degrees of freedom, the P-value is essentially 0.

Since the P-value is so low, we reject the null hypothesis. There is strong evidence that the distribution of education level attained is different between the groups. Generally, younger people are more likely to have higher levels of education than older people, who are themselves over represented at the lower education levels. Specifically, people in the 35 – 44 age group were less likely to have only a high school diploma, and more likely to have at least four years of college.

Chapter 25 – Inferences for Regression

Section 25.2

1. Graduation rates 2013.

The scatterplot shows a linear relationship with equal spread about a regression line throughout the range of *Acceptance Rates*. The residual plot has no structure, and there are not any striking outliers. The histogram of residuals is symmetric and bell-shaped. All conditions are satisfied to proceed with the regression analysis.

2. Shoot to score.

The scatterplot shows a linear relationship between *Shots* and *Goals*. The residual plot looks scattered, but there may be some increase in variation of Goals as Shots increases. We will proceed with caution. The normal probability plot looks straight. All conditions seem satisfied to go ahead with the regression analysis.

Section 25.3

3. Graduation rates 2013, part II.

The error standard deviation is s = 00254 or 2.54 percentage points. Since the range of *Graduation Rates* is only about 17 percentage points, this number is fairly small and helps us to understand the amount of spread about the regression model. The standard deviation of the residuals is only about 2.54 percentage points.

4. Shoot to score another one.

The error standard deviation is s = 5.13. This number indicates the amount of variation in the data points about the linear regression model. The standard deviation of the residuals is about 5.13 goals.

5. Graduation rates 2013, part III.

The standard error for the slope tells us how much the slope of the regression equation would vary from sample to sample. If we took many samples, and found a regression equation for each, we would expect most of these slopes to be within a couple of standard deviations of the true slope of the association between graduation rate and acceptance rate.

6. Shoot to score, hat trick.

The standard error for the slope tells us how much the slope of the regression equation would vary from sample to sample. If we took many samples, and found a regression equation for each, we would expect most of these slopes to be within a couple of standard deviations of the true slope of the association between number of shots and number of goals. This number is very small, which means we should be able to make precise predictions.

Section 25.4

7. Graduation 2013, part IV.

Since the *P*-value is so low, we reject the null hypothesis. The administrators can conclude that there is evidence of a relationship between *Graduation Rates* and *Admission Rates* for these prestigious schools. It seems the lower the *Admission Rate*, the higher the *Graduation Rate*.

8. Shoot to score, number four.

Since the *P*-value is low, we reject the null hypothesis. The coach can conclude there is evidence of a relationship between shooting and scoring. In other words, as a player takes more *Shots*, he should expect to score more *Goals*, on average.

9. Graduation 2013, part V.

The administrators can conclude, with 95% confidence, that schools with *Admission Rates* that are lower by 1 percentage point will have, on average, *Graduation Rates* that are higher by between 0.238 and 0.332 percentage points.

10. Shoot to score, overtime.

With 95% confidence, the coach can conclude that, on average, players who take an additional *Shot* can expect to score between 0.11 to 0.13 *Goals* more.

Section 25.5

11. Graduation 2013, part VI.

The standard error for the mean will be smaller. We can be much more precise about the prediction for the mean graduation rate of all colleges compared to the prediction for just one.

12. Shoot to score, double overtime.

The standard error for the mean will be smaller. We can be much more precise about the prediction for the mean of all players who took 104 shots than for the prediction of an individual player.

Section 25.6

13. Graduation 2013, part VII.

The administrators can expect, with 95% confidence, that schools that admit 33% of applicants will have average graduation rates between 87.35% and 88.88%.

14. Shoot to score, triple overtime.

We can conclude with 95% confidence that players who take 104 shots should average between 9.8 and 11.5 goals.

15. Graduation 2013, part VIII.

The administrators can be 95% confident that if their *Admission Rate* is 33%, their *Graduation Rate* would be between 82.96% and 93.28%.

16. Shoot to score, again.

The coach can be 95% confident that a player who takes 150 shots will score between 6.03 and 24.25 goals during the season.

Chapter Exercises.

17. Graduation 2013 party.

The sentence in the exercise is implying a cause-and-effect relationship. It seems to suggest that if they reduce the admission rate, then the graduation rate must increase. These data don't support that strong a conclusion. We have no evidence of what effects a change in admission rate might have.

18. Shoot to score, shootout.

To imply that shooting more will equate to scoring more is faulty logic. Though the two variables are related, we cannot infer a cause-and-effect relationship. Other factors may determine how many goals a hockey team will score over the course of an entire season.

19. Tracking hurricanes 2012.

a) The equation of the line of best fit for these data points is

$\widehat{24Error} = 133.0 - 2.06(Year)$, where *Year* is measured in years since 1970.

According to the linear model, the error made in predicting a hurricane's path was about 133 nautical miles, on average, in 1970. It has been declining at rate of about 2.06 nautical miles per year.

b) H_0: There has been no change in prediction accuracy. $(\beta_1 = 0)$

H_A: There has been a change in prediction accuracy. $(\beta_1 \neq 0)$

c) Assuming the conditions have been met, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 43 – 2 = 41 degrees of freedom. We will use a regression slope *t*-test.

The value of $t = -10.4$. The *P*-value $\leq$ 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence that the prediction accuracies have in fact been changing during the time period.

d) 72.5% of the variation in the prediction accuracy is accounted for by the linear model based on year.

20. Drug use 2013.

a) The equation of the line of best fit for these data points is $\widehat{\%Cocaine} = 0.297 + 0.124(\%Cannabis)$. According to the linear model, the average percentage of people in these countries who use cocaine increases by about 0.124 percentage points for each additional 1% of people who use cannabis.

b) H_0: There is no linear relationship between marijuana use and use of other drugs. $(\beta_1 = 0)$

H_A: There is a linear relationship between marijuana use and use of other drugs. $(\beta_1 \neq 0)$

c) Assuming the conditions have been met, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (32 – 2) = 30 degrees of freedom. We will use a regression slope *t*-test.

The value of $t = 3.23$. The *P*-value of 0.003 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence that the percentage of people who use cocaine is related to the percentage of people who use cannabis. Countries with a high percentage of people using cannabis tend to have a high percentage of people using cocaine.

d) 25.8% of the variation in the percentage of people using cocaine can be accounted for by the percentage of people using cannabis.

e) The use of cocaine is associated with cannabis use, but there is no proof of a cause-and-effect relationship between the two variables. There may be lurking variables present.

21. Movie budgets.

a) $\widehat{Budget} = -63.9981 + 1.02648(RunTime)$. The model suggests that each additional minute of run time for a movie costs about $1,026,000.

b) A negative intercept makes no sense, but the *P*-value of 0.07 indicates that we can't discern a difference between our estimated value and zero. It makes sense that a movie of zero length should cost $0.

c) Amounts by which movie costs differ from predictions made by this model vary, with a standard deviation of about $33 million.

d) The standard error of the slope is 0.1541 million dollars per minute.

e) If we constructed other models based on different samples of movies, we'd expect the slopes of the regression lines to vary, with a standard deviation of about $154,000 per minute.

22. Saratoga house prices.

a) $\widehat{Price} = -3.11686 + 94.4539(Size)$. The model suggests that the prices of Saratoga homes increase by about \$94.5 for each additional square foot.

b) A negative intercept makes no sense, but the *P*-value of 0.50 indicates that we can't discern a difference between our estimated value and zero. It makes sense that a house of zero square feet should cost \$0.

c) Amounts by which house prices differ from predictions made by this model vary, with a standard deviation of about \$54,000 per thousand square feet.

d) The standard error of the slope is 2.393 dollars per square foot.

e) If we constructed other models based on different samples of homes, we'd expect the slopes of the regression lines to vary, with a standard deviation of about 2.39 dollars per square foot.

23. Movie budgets, the sequel.

a) **Straight enough condition:** The scatterplot is straight enough, and the residuals plot looks unpatterned.
Independence assumption: The residuals plot shows no evidence of dependence.
Does the plot thicken? condition: The residuals plot shows no obvious trends in the spread.
Nearly Normal condition, Outlier condition: The Normal probability plot is reasonably straight.

b) Since conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (120 – 2) = 118 degrees of freedom.

$$b_1 \pm t_{n-2}^* \times SE(b_1) = 1.02648 \pm (t_{118}^*) \times 0.1541 \approx (0.72,\ 1.33)$$

We are 95% confident that the cost of making longer movies increases at a rate of between 0.72 and 1.33 million dollars per minute.

24. Second home.

a) **Straight enough condition:** The scatterplot is straight enough, and the residuals plot looks unpatterned.
Randomization condition: The houses were selected at random.
Does the plot thicken? condition: The residuals plot shows no obvious trends in the spread.
Nearly Normal condition, Outlier condition: The histogram of residuals is unimodal and symmetric, and shows no outliers.

b) Since conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (1064 – 2) = 1062 degrees of freedom.

$b_1 \pm t^*_{n-2} \times SE(b_1) = 94.4539 \pm (t^*_{1062}) \times 2.393 \approx (89.8,\ 99.2)$

We are 95% confident that Saratoga housing costs increase at a rate of between \$89.8 and \$99.2 per square foot.

25. Hot dogs.

a) H_0: There's no association between calories and sodium content of all-beef hot dogs. $(\beta_1 = 0)$

H_A: There is an association between calories and sodium content. $(\beta_1 \neq 0)$

b) Assuming the conditions have been met, the sampling distribution of the regression slope can be modeled by a Student's t-model with (13 – 2) = 11 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{Sodium} = 90.9783 + 2.29959(Calories)$

The value of t = 4.10. The P-value of 0.0018 means that the association we see in the data is very unlikely to occur by chance alone. We reject the null hypothesis, and conclude that there is evidence of a linear association between the number of calories in all-beef hotdogs and their sodium content. Because of the positive slope, there is evidence that hot dogs with more calories generally have higher sodium contents.

26. Cholesterol.

a) H_0: There is no linear relationship between age and cholesterol. $(\beta_1 = 0)$

H_A: Cholesterol levels change with age. $(\beta_1 \neq 0)$

b) Assuming the conditions have been met, the sampling distribution of the regression slope can be modeled by a Student's t-model with (1406 – 2) = 1404 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{Cholesterol} = 194.232 + 0.772(Age)$

The value of t = 3. The P-value of 0.0056 means that the association we see in the data is very unlikely to occur by chance alone. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between age and cholesterol. Because of the positive slope, there is evidence that cholesterol levels tend to increase with age.

27. Second frank.

a) Among all-beef hot dogs with the same number of calories, the sodium content varies, with a standard deviation of about 60 mg.

b) The standard error of the slope of the regression line is 0.5607 milligrams of sodium per calorie.

c) If we tested many other samples of all-beef hot dogs, the slopes of the resulting regression lines would be expected to vary, with a standard deviation of about 0.56 mg of sodium per calorie.

28. More cholesterol.

a) Among adults of the same age, cholesterol levels vary, with a standard deviation of about 46 points.

b) The standard error of the slope of the regression line is 0.2574 cholesterol points per year of age.

c) If we tested many other samples of adults, the slopes of the resulting regression lines would be expected to vary with a standard deviation of 0.26 cholesterol points per year of age.

29. Last dog.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = 2.29959 \pm (2.201) \times 0.5607 \approx (1.06, 3.54)$$

We are 95% confident that for every additional calorie, all-beef hot dogs have, on average, between 1.06 and 3.54 mg more sodium.

30. Cholesterol, finis.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = 0.771639 \pm (t^*_{1404}) \times 0.2574 \approx (0.27, 1.28)$$

We are 95% confident that, on average, adult cholesterol levels increase by between 0.27 and 1.28 points per year of age.

31. Marriage age 2011.

a) H_0: The difference in age between men and women at first marriage has not been decreasing since 1975. $(\beta_1 = 0)$

H_A: The difference in age between men and women at first marriage has been decreasing since 1975. $(\beta_1 < 0)$

b) **Straight enough condition:** The scatterplot is straight enough.
Independence assumption: We are examining a relationship over time, so there is reason to be cautious, but the residuals plot shows no evidence of dependence.
Does the plot thicken? condition: The residuals plot shows no obvious trends in the spread.
Nearly Normal condition, Outlier condition: The histogram is reasoanbly unimodal and symmetric, and shows no obvious skewness or outliers.

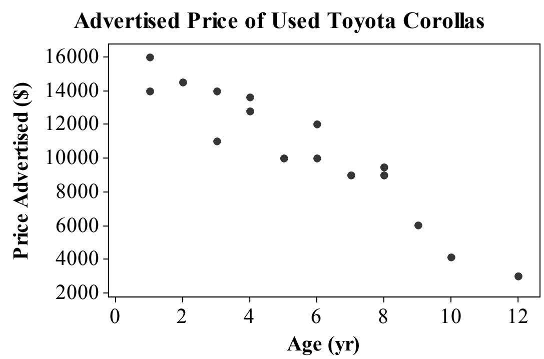
c) Since conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with 37 – 2 = 35 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{(Men - Women)} = 42.566 - 0.0203(Year)$

The value of t = – 6.14. The P-value of less than 0.0001 (even though this is the value for a two-tailed test, it is still very small) means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a negative linear relationship between difference in age at first marriage and year. The difference in marriage age between men and women appears to be decreasing over time.

32. Used cars 2010.

a) A scatterplot of the used cars data is at the right.

b) A linear model is probably appropriate. The plot appears to be linear.



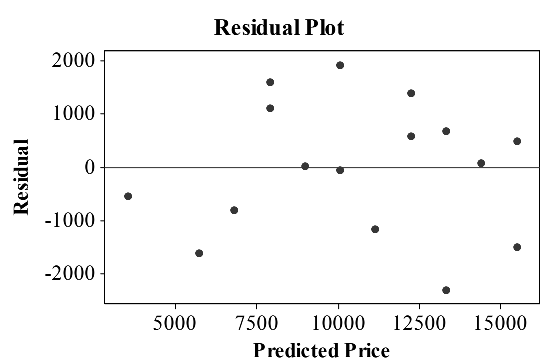
c)
```
Predictor      Coef  SE Coef       T      P
Constant   16566.72    649.2   25.52  0.000
Age (yr)   -1085.95    101.5  -10.70  0.000

S = 1283.27   R-Sq = 89.1%   R-Sq(adj) = 88.3%
```

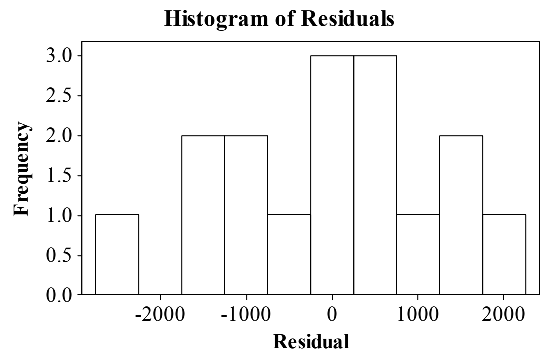
The equation of the regression line is: $\widehat{Price} = 16566.72 - 1089.95(Age)$.

According to the model, the average asking price for a used Toyota Corolla decreases by about $1086 dollars for each additional year in age. Let's take a closer look.

d) **Straight enough condition:** The scatterplot is straight enough to try a linear model.
Independence assumption: Prices of Toyota Corollas of different ages might be related, but the residuals plot looks fairly scattered. (The fact that there are several prices for some years draws our eyes to some patterns that may not exist.)



Does the plot thicken? condition: The residuals plot shows no obvious patterns in the spread.
Nearly Normal condition, Outlier condition: The histogram is reasonably unimodal, but might show some slight skew to the left.



Since conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (16 – 2) = 14 degrees of freedom.

33. Marriage age 2011, again.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = -0.020 \pm (2.030) \times 0.0033 \approx (-0.027, -0.013)$$

We are 95% confident that the mean difference in age between men and women at first marriage decreases by between 0.013 and 0.027 years in age for each year that passes.

34. Used cars 2010, again.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = -1086 \pm (2.145) \times 101.5 \approx (-1304, -868)$$

We are 95% confident that the advertised price of a used Toyota Corolla is decreasing by an average of between \$868 and \$1304 for each additional year in age.

35. Fuel economy.

a) H_0: There is no linear relationship between the weight of a car and its mileage. $(\beta_1 = 0)$

H_A: There is a linear relationship between the weight of a car and its mileage. $(\beta_1 \neq 0)$

b) **Straight enough condition:** The scatterplot is straight enough to try a linear model.
Independence assumption: The residuals plot is scattered.
Does the plot thicken? condition: The residuals plot indicates some possible "thickening" as the predicted values increases, but it's probably not enough to worry about.
Nearly Normal condition, Outlier condition: The histogram of residuals is unimodal and symmetric, with one possible outlier. With the large sample size, it is okay to proceed.

Since conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 50 – 2 = 48 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{MPG} = 48.7393 - 8.2136(Weight)$, where *Weight* is measured in thousands of pounds.

The value of $t = -12.2$. The *P*-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between weight of a car and its mileage. Cars that weigh more tend to have lower gas mileage.

36. SAT scores.

a) H_0: There is no linear relationship between SAT Verbal and Math scores. $(\beta_1 = 0)$

H_A: There is a linear relationship between SAT Verbal and Math scores. $(\beta_1 \neq 0)$

b) **Straight enough condition:** The scatterplot is straight enough to try a linear model.
Independence assumption: The residuals plot is scattered.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition, Outlier condition: The histogram of residuals is unimodal and symmetric, with one possible outlier. With the large sample size, it is okay to proceed.

Since conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (162 – 2) = 160 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{Math} = 209.554 + 0.675(Verbal)$.

The value of $t = 11.9$. The *P*-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between SAT Verbal and Math scores. Students with higher SAT-Verbal scores tend to have higher SAT-Math scores.

37. Fuel economy, part II.

a) Since conditions have been satisfied in Exercise 35, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (50 – 2) = 48 degrees of freedom. (Use $t_{45}^{*} = 2.014$ from the table.) We will use a regression slope *t*-interval, with 95% confidence.

$b_1 \pm t_{n-2}^{*} \times SE(b_1) = -8.2136 \pm (2.014) \times 0.674 \approx (-9.57, -6.86)$

b) We are 95% confident that the mean mileage of cars decreases by between 6.86 and 9.57 miles per gallon for each additional 1000 pounds of weight.

38. SATs, part II.

a) Since conditions have been satisfied in Exercise 36, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (162 – 2) = 160 degrees of freedom. (Use $t^*_{140} = 1.656$ from the table.) We will use a regression slope *t*-interval, with 90% confidence.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = 0.675 \pm (1.656) \times 0.0568 \approx (0.583,\ 0.767)$$

b) We are 90% confident that the mean Math SAT scores increase by between 0.583 and 0.767 point for each additional point scored on the Verbal test.

39. Fuel economy, part III.

a) The regression equation predicts that cars that weigh 2500 pounds will have a mean fuel efficiency of $48.7393 - 8.2136(2.5) = 28.2053$ miles per gallon.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n}}$$

$$= 28.2053 \pm (2.014)\sqrt{0.674^2 \cdot (2.5 - 2.8878)^2 + \frac{2.413^2}{50}} \approx (27.35,\ 28.20)$$

We are 95% confident that cars weighing 2500 pounds will have mean fuel efficiency between 27.35 and 28.20 miles per gallon.

b) The regression equation predicts that cars that weigh 3450 pounds will have a mean fuel efficiency of $48.7393 - 8.2136(3.45) = 20.40238$ miles per gallon.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n} + s_e^2}$$

$$= 20.40238 \pm (2.014)\sqrt{0.6738^2 \cdot (3.45 - 2.8878)^2 + \frac{2.413^2}{50} + 2.413^2} \approx (15.44,\ 25.37)$$

We are 95%confident that a car weighing 3450 pounds will have fuel efficiency between 15.44 and 25.37 miles per gallon.

40. SATs, again.

a) The regression equation predicts that students with an SAT-Verbal score of 500 will have a mean SAT-Math score of $209.554 + 0.675(500) = 547.054$.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n}}$$

$$= 547.054 \pm (1.656)\sqrt{0.0568^2 \cdot (500 - 596.296)^2 + \frac{71.75^2}{162}} \approx (534.0,\ 560.18)$$

We are 90% confident that students with scores of 500 on the SAT-Verbal will have a mean SAT-Math score between 534.0 and 560.18.

b) The regression equation predicts that students with an SAT-Verbal score of 710 will have a mean SAT-Math score of $209.554 + 0.675(710) = 688.804$.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}$$

$$= 688.804 \pm (1.656)\sqrt{0.0568^2 \cdot (710 - 596.296)^2 + \frac{71.75^2}{162} + 71.75^2} \approx (569,\ 808)$$

We are 90%confident that a student scoring 710 on the SAT-Verbal would have an SAT-Math score of between 569.14 and 808.47. Since we are talking about individual scores, and not means, it is reasonable to restrict ourselves to possible scores, so we are 90% confident that the class president scored between 570 and 800 on the SAT-Math test.

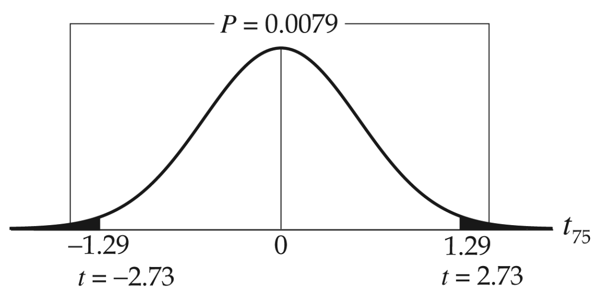
41. Cereals.

a) H_0: There is no linear relationship between the number of calories and the sodium content of cereals. $(\beta_1 = 0)$

H_A: There is a linear relationship between the number of calories and the sodium content of cereals. $(\beta_1 \neq 0)$

Since these data were judged acceptable for inference, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 77 – 2 = 75 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{Sodium} = 21.4143 + 1.29357(Calories)$.

The value of *t* = 2.73. The *P*-value of 0.0079 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between the number of calories and sodium content of cereals. Cereals with higher numbers of calories tend to have higher sodium contents.



b) Only 9% of the variability in sodium content can be explained by the number of calories. The residual standard deviation is 80.49 mg, which is pretty large when you consider that the range of sodium content is only 320 mg. Although there is strong evidence of a linear association, it is too weak to be of much use. Predictions would tend to be very imprecise.

42. Brain size.

a) H_0: There is no linear relationship between brain size and IQ. $(\beta_1 = 0)$

H_A: There is a linear relationship between brain size and IQ. $(\beta_1 \neq 0)$

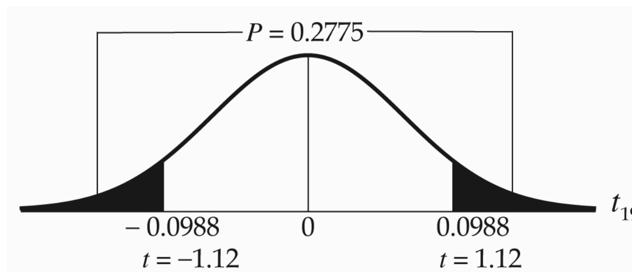
Since these data were judged acceptable for inference, the sampling distribution of the regression slope can be modeled by a Student's t-model with (21 – 2) = 19 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{IQ\_Verbal} = 24.1835 + 0.0988(Size)$.

The value of $t \approx 1.12$. The P-value of 0.2775 means that the association we see in the data is likely to occur by chance. We fail to reject the null hypothesis, and conclude that there is no evidence of a linear relationship between brain size and verbal IQ score.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$

$$t = \frac{0.0988 - 0}{0.0884}$$

$$t \approx 1.12$$



b) Since $R^2 = 6.5\%$, only 6.5% of the variability in verbal IQ can be accounted for by brain size. This association is very weak. There are three students with large brains who scored high on the IQ test. Without them, there appears to be no association at all.

43. Cereals, part 2.

Straight enough condition: The scatterplot is not straight.
Independence assumption: The residuals plot shows a curved pattern.
Does the plot thicken? condition: The spread of the residuals is not consistent. The residuals plot "thickens" as the predicted values increase.
Nearly Normal condition, Outlier condition: The histogram of residuals is skewed to the right, with an outlier.

These data are not appropriate for inference.

44. Winter.

Straight enough condition: The scatterplot is not straight.
Independence assumption: The residuals plot shows a curved pattern.
Does the plot thicken? condition: The spread of the residuals is not consistent. The residuals plot shows decreasing variability as the predicted values increase.
Nearly Normal condition, Outlier condition: The histogram of residuals is skewed to the right, with an outlier.

These data are not appropriate for inference.

45. Streams.

a) H_0: There is no linear relationship between BCI and pH. $(\beta_1 = 0)$
H_A: There is a linear relationship between BCI and pH. $(\beta_1 \neq 0)$

b) Assuming the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 163 – 2 = 161 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{BCI} = 2733.37 - 197.694(pH)$.

c) The value of $t \approx -7.73$. The *P*-value (two-sided!) of essentially 0 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between BCI and pH. Streams with higher pH tend to have lower BCI.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$
$$t = \frac{-197.694 - 0}{25.57}$$
$$t \approx -7.73$$

46. Climate change 2013.

a) The regression equation is $\widehat{Temp} = 10.6442 + 0.0103(CO_2)$, with CO_2 concentration measured in parts per million from the top of Mauna Loa in Hawaii, and temperature in degrees Celsius.

b) H_0: There is no linear relationship between temperature and CO_2 concentration. $(\beta_1 = 0)$
H_A: There is a linear relationship between temperature and CO_2 concentration. $(\beta_1 \neq 0)$

Since the scatterplots showed that the data were appropriate for inference, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (44 – 2) = 42 degrees of freedom. We will use a regression slope *t*-test. We must assume independence, since we have no plot of the residuals.

The value of *t* = 10.8. The *P*-value (two-sided!) of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between CO_2 concentration and temperature. Years with higher CO_2 concentration tend to be warmer, on average.

c) The standard deviation of the residuals is 0.1316 degrees Celsius, so we don't expect the model to predict with accuracy greater than about ± 0.26 degrees Celsius. The range of the temperatures is under a degree, so that's not very precise.

d) The regression does not prove that increasing CO_2 levels are causing global warming. There is a strong association, but it could be coincidental. However, the known physical relationship between greenhouse gases and temperature provides some reason to suspect causation.

47. Ozone and population.

a) H_0: There is no linear relationship between population and ozone level. $(\beta_1 = 0)$

H_A: There is a positive linear relationship between population and ozone level. $(\beta_1 > 0)$

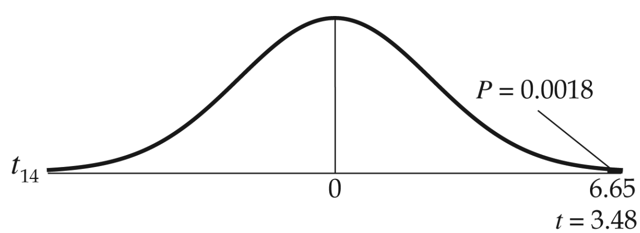
Assuming the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with 16 – 2 = 14 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{Ozone} = 18.892 + 6.650(Population)$, where ozone level is measured in parts per million and population is measured in millions.

The value of $t \approx 3.48$. The P-value of 0.0018 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a positive linear relationship between ozone level and population. Cities with larger populations tend to have higher ozone levels.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$

$$t = \frac{6.650 - 0}{1.910}$$

$$t \approx 3.48$$



b) City population is a good predictor of ozone level. Population explains 84% of the variability in ozone level and s is just over 5 parts per million.

48. Sales and profits.

a) H_0: There is no linear relationship between sales and profit. $(\beta_1 = 0)$

H_A: There is a linear relationship between sales and profit. $(\beta_1 \neq 0)$

Assuming the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with 79 – 2 = 77 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{Profits} = -176.644 + 0.092498(Sales)$, with both profits and sales measured in millions of dollars.

The value of $t \approx 12.33$. The P-value of essentially 0 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between sales and profits. Companies with higher sales tend to have higher profits.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$
$$t = \frac{0.092498 - 0}{0.0075}$$
$$t \approx 12.33$$

b) A company's sales may be of some help in predicting profits. $R^2 = 66.2\%$, so 66.2% of the variability in profits can be accounted for by sales. Although s is nearly half a billion dollars, the mean profit for these companies is over 4 billion dollars.

49. Ozone, again

a) $b_1 \pm t^*_{n-2} \times SE(b_1) = 6.65 \pm (1.761) \times 1.910 \approx (3.29, 10.01)$

We are 90% confident that each additional million people will increase mean ozone levels by between 3.29 and 10.01 parts per million.

b) The regression equation predicts that cities with a population of 600,000 people will have ozone levels of 18.892 + 6.650(0.6) = 22.882 parts per million.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n}}$$
$$= 22.882 \pm (1.761)\sqrt{1.91^2 \cdot (0.6 - 1.7)^2 + \frac{5.454^2}{16}} \approx (18.47, 27.29)$$

We are 90% confident that the mean ozone level for cities with populations of 600,000 will be between 18.47 and 27.29 parts per million.

50. More sales and profits.

a) There are 77 degrees of freedom, so use $t^*_{75} = 1.992$ as a conservative estimate from the table.

$b_1 \pm t^*_{n-2} \times SE(b_1) = 0.092498 \pm (1.992) \times 0.0075 \approx (0.078, 0.107)$

We are 95% confident that each additional million dollars in sales will increase mean profits by between \$78,000 and \$107,000.

b) The regression equation predicts that corporations with sales of \$23,000 million dollars will have profits of $-176.644 + 0.092498(23,000) = 1950.603$ million dollars.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n} + s_e^2}$$
$$= 1950.603 \pm (1.992)\sqrt{0.0075^2 \cdot (23,000 - 4178.29)^2 + \frac{466.2^2}{79} + 466.2^2} \approx (974.98, 2926.63)$$

We are 95% confident that Eli Lilly's profits will be between 974,980,000 and \$2,926,630,000. This interval is too wide to be of any use.

51. Tablet computers 2014.

a) Since there are 34 – 2 = 32 degrees of freedom, there were 34 tablet computers tested.

b) **Straight enough condition:** The scatterplot is roughly straight, but scattered.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals is reasonably straight.

c) H_0: There is no linear relationship between maximum brightness and battery life. $(\beta_1 = 0)$

H_A: There is a positive linear relationship between maximum brightness and battery life. $(\beta_1 > 0)$

Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 34 – 2 = 32 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{Hours} = 5.38719 + 0.00904(ScreenBrightness)$, battery life measure in hours and screen brightness measured in cd/m^2.

The value of $t \approx 2.02$. The *P*-value of 0.0522 means that the association we see in the data is not unlikely to occur by chance. We fail to reject the null hypothesis, and conclude that there is little evidence of a positive linear relationship between battery life and screen brightness.

d) Since $R^2 = 11.3\%$, only 11.3% of the variability in battery life can be accounted for by screen brightness. The residual standard deviation is 2.128 hours. That's pretty large, considering the range of battery life is only about 9 hours. Even if we concluded that there was some evidence of a linear association, it is too weak to be of much use. Predictions would tend to be very imprecise.

e) The equation of the line of best fit for these data points is:
$\widehat{Hours} = 5.38719 + 0.00904(ScreenBrightness)$, battery life measure in hours and screen brightness measured in cd/m^2.

f) There are 32 degrees of freedom, so use $t_{32}^{*} = 1.694$.

$$b_1 \pm t_{n-2}^{*} \times SE(b_1) = 0.00904 \pm (1.694) \times 0.0045 \approx (0.00142,\ 0.0167)$$

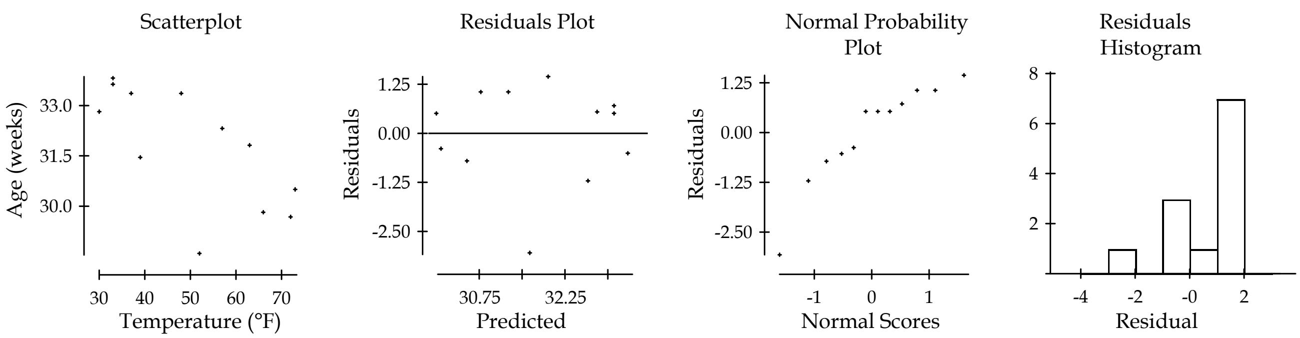
g) We are 90% confident that the mean battery life increases by between 0.00142 and 0.0167 hours for each additional cd/m^2 of screen brightness.

52. Crawling.

a) If the data had been plotted for individual babies, the association would appear to be weaker, since individuals are more variable than averages.

b) H_0: There is no linear relationship between 6-month temperature and crawling age. $(\beta_1 = 0)$

H_A: There is a linear relationship. $(\beta_1 \neq 0)$



Straight enough condition: The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern, but there may be an outlier. If the month of May were just one data point, it would be removed. However, since it represents the average crawling age of several babies, there is no justification for its removal.
Does the plot thicken? condition: The spread of the residuals is consistent
Nearly Normal condition: The Normal probability plot of residuals isn't very straight, largely because of the data point for May. The histogram of residuals also shows this outlier.

Since we had difficulty with the conditions for inference, we will proceed cautiously. These data may not be appropriate for inference. The sampling distribution of the regression slope can be modeled by a Student's *t*-model with 12 – 2 = 10 degrees of freedom. We will use a regression slope *t*-test.

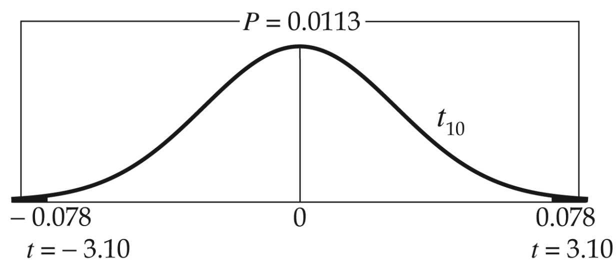
Dependent variable is: **Age**
No Selector
R squared = 49.0% R squared (adjusted) = 43.9%
s = 1.319 with 12 - 2 = 10 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	16.6933	1	16.6933	9.59
Residual	17.4028	10	1.74028	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	35.6781	1.318	27.1	≤ 0.0001
Temp	-0.077739	0.0251	-3.10	0.0113

The equation of the line of best fit for these data points is:
$\widehat{Age} = 35.6781 - 0.077739(Temp)$, with average crawling age measured in weeks and average temperature in °F.

The value of $t \approx -3.10$. The P-value of 0.0113 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between average temperature and average crawling age. Babies who reach six months of age in warmer temperatures tend to crawl at earlier ages than babies who reach six months of age in colder temperatures.



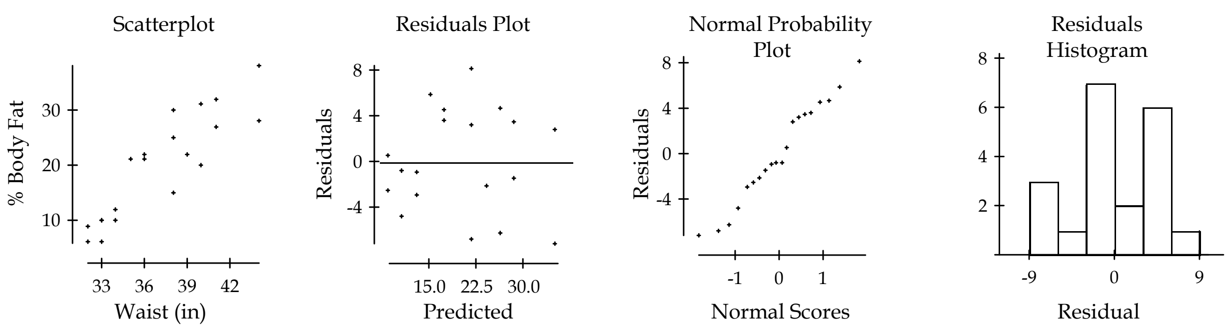
c) $b_1 \pm t^*_{n-2} \times SE(b_1) = -0.077739 \pm (2.228) \times 0.0251 \approx (-0.134, -0.022)$

We are 95% that the average crawling age decreases by between 0.022 weeks and 1.34 weeks when the average temperature increases by 10°F.

53. Body fat.

a) H_0: There is no linear relationship between waist size and percent body fat. $(\beta_1 = 0)$

H_A: There is a linear relationship between waist size and percent body fat. $(\beta_1 \neq 0)$



Straight enough condition: The scatterplot is straight enough.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition, Outlier condition: The Normal probability plot of residuals is straight, and the histogram of the residuals is unimodal and symmetric with no outliers.

Since the conditions for inference are inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (20 – 2) = 18 degrees of freedom. We will use a regression slope t-test.

Dependent variable is: Body Fat %
No Selector
R squared = 78.7% R squared (adjusted) = 77.5%
s = 4.540 with 20 - 2 = 18 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	1366.79	1	1366.79	66.3
Residual	370.960	18	20.6089	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-62.5573	10.16	-6.16	$\le$ 0.0001
Waist (in)	2.22152	0.2728	8.14	$\le$ 0.0001

The equation of the line of best fit for these data points is:
$\widehat{\%BodyFat} = -62.5573 + 2.22152(Waist)$.

The value of $t \approx 8.14$. The P-value of essentially 0 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between waist size and percent body fat. People with larger waists tend to have a higher percentage of body fat.

b) The regression equation predicts that people with 40-inch waists will have $-62.5573 + 2.22152(40) = 26.3035\%$ body fat. The average waist size of the people sampled was approximately 37.05 inches.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n}}$$

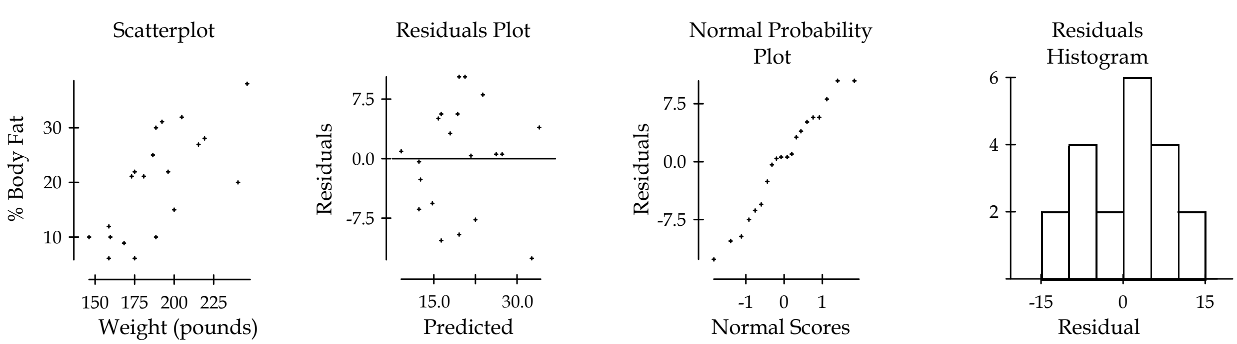
$$= 26.3035 \pm (2.101)\sqrt{0.2728^2 \cdot (40 - 37.05)^2 + \frac{4.54^2}{20}}$$

$$\approx (23.58,\ 29.03)$$

We are 95% confident that the mean percent body fat for people with 40-inch waists is between 23.58% and 29.03%.

54. Body fat, again.

a)



Straight enough condition: The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.

Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals is straight, and the histogram of the residuals is unimodal and symmetric with no outliers.

Since the conditions for inference are inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 20 – 2 = 18 degrees of freedom. We will use a regression slope *t*-interval.

Dependent variable is: **Body Fat %**
No Selector
R squared = 48.5% R squared (adjusted) = 45.7%
s = 7.049 with 20 - 2 = 18 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	843.325	1	843.325	17.0
Residual	894.425	18	49.6903	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-27.3763	11.55	-2.37	0.0291
Weight (lb)	0.249874	0.0607	4.12	0.0006

The equation of the line of best fit for these data points is:
$\widehat{\%BodyFat} = -27.3763 + 0.249874(Weight)$.

$$b_1 \pm t^*_{n-2} \times SE\left(b_1\right) = 0.249874 \pm (1.734) \times 0.0607 \approx (0.145,\ 0.355)$$

b) We are 90% confident that the mean percent body fat increases between 1.45% and 3.55% for an additional 10 pounds in weight.

c) The regression equation predicts that a person weighing 165 pounds would have $-27.3763 + 0.249874(165) = 13.85291\%$ body fat. The average weight of the people sampled was 188.6 pounds.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}$$

$$= 13.85291 \pm (2.101)\sqrt{0.0607^2 \cdot (165 - 188.6)^2 + \frac{7.049^2}{20} + 7.049^2} \approx (-1.61,\ 29.32)$$

We are 95% confident that a person weighing 165 pounds would have between 0% (–1.61%) and 29.32% body fat.

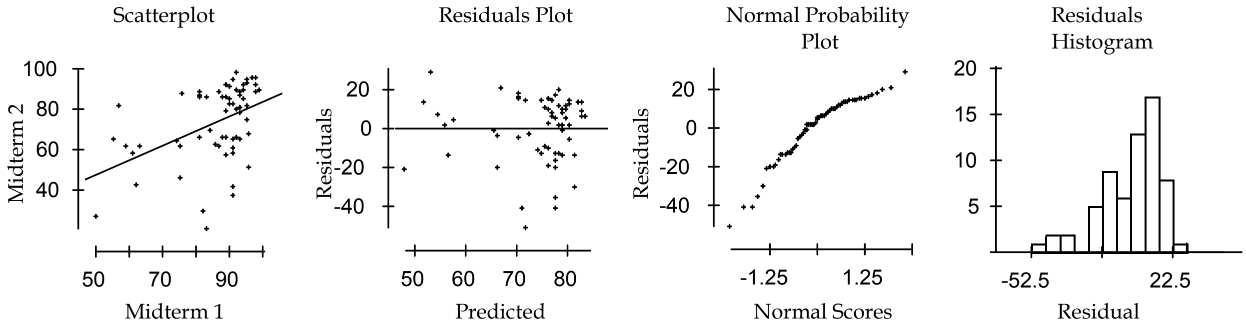
55. Grades.

a) The regression output is to the right. The model is:
$\widehat{Midterm2} = 12.005 + 0.721(Midterm1)$

Dependent variable is: **Midterm 2**
No Selector
R squared = 19.9% R squared (adjusted) = 18.6%
s = 16.78 with 64 - 2 = 62 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	4337.14	1	4337.14	15.4
Residual	17459.5	62	281.604	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	12.0054	15.96	0.752	0.4546
Midterm 1	0.720990	0.1837	3.92	0.0002



b) **Straight enough condition:** The scatterplot shows a weak, positive relationship between Midterm 2 score and Midterm 1 score. There are several outliers, but removing them only makes the relationship slightly stronger. The relationship is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern..
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition, Outlier condition: The histogram of the residuals is unimodal, slightly skewed with several possible outliers. The Normal probability plot shows some slight curvature.

Since we had some difficulty with the conditions for inference, we should be cautious in making conclusions from these data. The small *P*-value of 0.0002 for the slope would indicate that the slope is statistically distinguishable from zero, but the R^2 value of 0.199 suggests that the relationship is weak. Midterm 1 isn't a useful predictor of Midterm 2.

c) The student's reasoning is not valid. The R^2 value is only 0.199 and the value of *s* is 16.8 points. Although correlation between Midterm 1 and Midterm 2 may be statistically significant, it isn't of much practical use in predicting Midterm 2 scores. It's too weak.

56. Grades?

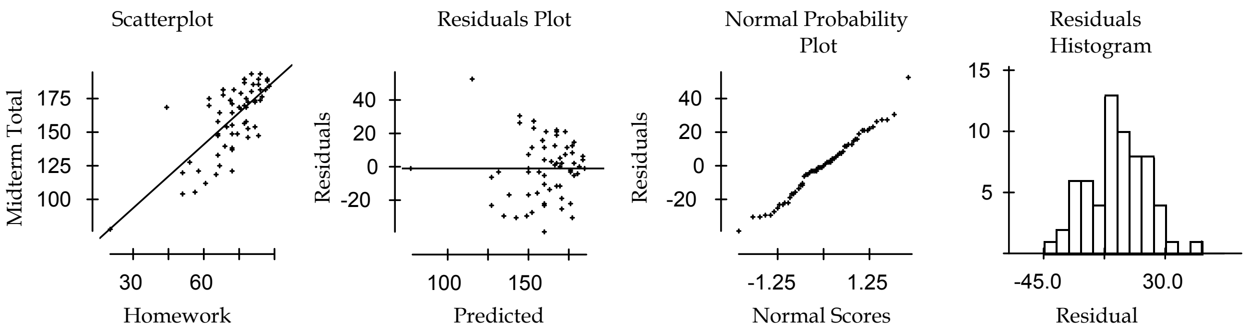
a) The regression output is to the right. The model is:

$$\widehat{MTtotal} = 46.062 + 1.580(Homework)$$

Dependent variable is: **M1+M2**
No Selector
R squared = 50.7%   R squared (adjusted) = 49.9%
s = 18.30 with 64 - 2 = 62 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	21398.1	1	21398.1	63.9
Residual	20773.0	62	335.048	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	46.0619	14.46	3.19	0.0023
Homework	1.58006	0.1977	7.99	$\le$ 0.0001



b) **Straight enough condition:** The scatterplot shows a moderate, positive relationship between Midterm total and homework. There are two outliers, but removing them does not significantly change the model. The relationship is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern..
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The histogram of the residuals is unimodal and symmetric, and the Normal probability plot is reasonably straight..

Since the conditions are met, linear regression is appropriate. The small *P*-value for the slope would indicate that the slope is statistically distinguishable from zero.

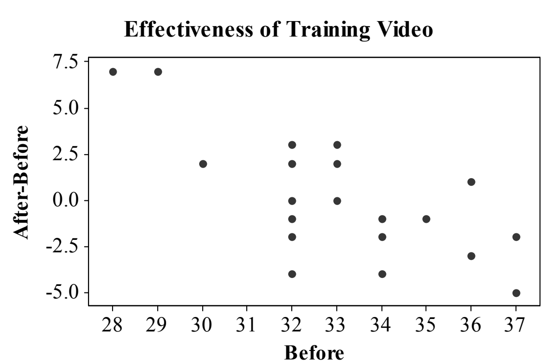
c) The R^2 value of 0.507 suggests that the overall relationship is fairly strong. However, this does not mean that midterm total is accurately predicted from homework scores. The error standard deviation of 18.30 indicates that a prediction in midterm total could easily be off by 20 to 30 points. If this is significant number of points for deciding grades, then homework score alone will not suffice.

57. Strike two.

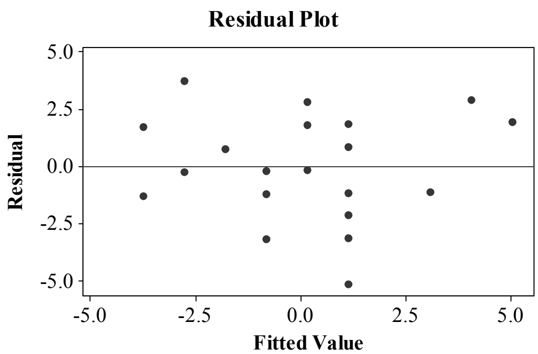
H_0: Effectiveness is independent of the player's initial ability. $(\beta_1 = 0)$

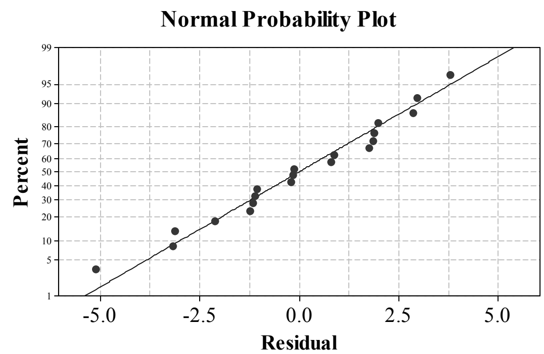
H_A: Effectiveness of the video depends on the player's initial ability. $(\beta_1 \neq 0)$

Straight enough condition: The scatterplot is straight enough.



Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot is straight.





Since the conditions for inference are inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 20 – 2 = 18 degrees of freedom. We will use a regression slope *t*-test.

The equation of the line of best fit for these data points is:

$\widehat{(After - Before)} = 32.316 - 0.9748(Before)$, where we are counting the number of strikes thrown before and after the training program.

The value of $t \approx -4.34$. Since the *P*-value is 0.004, reject the null hypothesis, and conclude that there is evidence of a linear relationship between the player's intial ability and the effectiveness of the program. The negative slope indicates that the method is more effective for those whose initial performance was poorest and less effective for those whose initial performance was better. This appears to be a case of regression to the mean. Those who were above average initially tended to do worse after training. Those who were below average initially tended to improve.

58. All the efficiency money can buy 2011.

a) We'd like to know if there is a linear association between price and fuel efficiency in cars. We have data on 2011 model year cars, with information on highway MPG and retail price.

H_0: There is no linear relationship between MPG and retail price. $(\beta_1 = 0)$

H_A: Highway MPG and retail price are linearly associated. $(\beta_1 \neq 0)$

b) The scatterplot fails the Straight enough condition. There is no evidence of any association whatsoever between highway MPG and retail price. We can't really perform a test.

c) Since the conditions are not satisfied, we cannot continue this analysis.

59. Education and mortality.

a) **Straight enough condition:** The scatterplot is straight enough.
Independence assumption: The residuals plot shows no pattern. If these cities are representative of other cities, we can generalize our results.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition, Outlier condition: The histogram of the residuals is unimodal and symmetric with no outliers.

b) H_0: There is no linear relationship between education and mortality. $(\beta_1 = 0)$

H_A: There is a linear relationship between education and mortality. $(\beta_1 \neq 0)$

Since the conditions for inference are inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with 58 – 2 = 56 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is:
$\widehat{Mortality} = 1493.26 - 49.9202(Education)$.

The value of $t \approx -6.24$. The P-value of essentially 0 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between the level of education in a city and its mortality rate. Cities with lower education levels tend to have higher mortality rates.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$
$$t = \frac{-49.9202 - 0}{8.000}$$
$$t \approx -6.24$$

c) We cannot conclude that getting more education is likely to prolong your life. Association does not imply causation. There may be lurking variables involved.

d) For 95% confidence, $t^*_{56} \approx 2.00327$.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = -49.9202 \pm (2.003) \times 8.000 \approx (-65.95, -33.89)$$

e) We are 95% confident that the mean number of deaths per 100,000 people decreases by between 33.89 and 65.95 deaths for an increase of one year in average education level.

f) The regression equation predicts that cities with an adult population with an average of 12 years of school will have a mortality rate of $1493.26 - 49.9202(12) = 894.2176$ deaths per 100,000. The average education level was 11.0328 years.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n}}$$
$$= 894.2176 \pm (2.003)\sqrt{8.00^2 \cdot (12 - 11.0328)^2 + \frac{47.92^2}{58}} \approx (874.239, 914.196)$$

We are 95% confident that the mean mortality rate for cities with an average of 12 years of schooling is between 874.239 and 914.196 deaths per 100,000 residents.

60. Property assessments.

a) **Straight enough condition:** The scatterplot is straight enough.
Independence assumption: The residuals plot shows no pattern. If these cities are representative of other cities, we can generalize our results.
Does the plot thicken? condition: The spread of the residuals is consistent
Nearly Normal condition: The Normal probability plot is fairly straight.

b) H_0: There is no linear relationship between size and assessed valuation. $(\beta_1 = 0)$

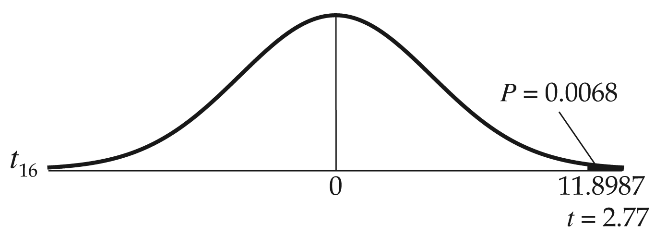
H_A: Larger houses have higher assessed values. $(\beta_1 > 0)$

Since the conditions for inference are inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (18 – 2) = 16 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is:
$\widehat{Assess} = 37{,}108.8 + 11.8987(SqFt)$.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$

$$t = \frac{11.8987 - 0}{4.290}$$

$$t \approx 2.77$$



The value of $t \approx 2.77$. The P-value of 0.0068 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between the size of a home and its assessed value. Larger homes tend to have higher assessed values.

c) $R^2 = 32.5\%$. This model accounts for 32.5% of the variability in assessments.

d) For 90% confidence, $t_{16}^* \approx 1.746$.

$$b_1 \pm t_{n-2}^* \times SE(b_1) = 11.8987 \pm (1.746) \times 4.290 \approx (4.41,\ 19.39)$$

e) We are 90% confident that the mean assessed value increases by between \$441 and \$1939 for each additional 100 square feet in size.

f) The regression equation predicts that houses measuring 2100 square feet will have an assessed value of $37108.8 + 11.8987(2100) = \$62,096.07$. The average size of the houses sampled is 2003.39 square feet.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n} + s_e^2}$$

$$= 62096.07 \pm (2.120)\sqrt{4.290^2 \cdot (2100 - 2003.39)^2 + \frac{4682^2}{18} + 4682^2} \approx (51860,\ 72332)$$

We are 95% confident that the assessed value of a home measuring 2100 square feet will have an assessed value between \$51,860 and \$72,332. There is no evidence that this home has an assessment that is too high. The assessed value of \$70,200 falls within the prediction interval.

The homeowner might counter with an argument based on the mean assessed value of all homes such as this one.

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n}}$$

$$= 62096.07 \pm (2.120)\sqrt{4.290^2 \cdot (2100 - 2003.39)^2 + \frac{4682^2}{18}} \approx (\$59,597,\ \$64,595)$$

The homeowner might ask the city assessor to explain why his home is assessed at \$70,200, if a typical 2100-square-foot home is assessed at between \$59,597 and \$64,595.

61. Right-to-work laws.

a) $\widehat{\text{Logit}\left(Right-to-work\right)} = 6.19951 - 0.106155\,publ - 0.222957\,pvt$

b) Yes, logistic regression seems appropriate here. The response variable is dichotomous and there are no outliers in the predictor variables.

62. Cost of higher education.

a) $\widehat{\text{Logit}\left(Type\right)} = -13.1461 + 0.08455\,Top10\% + 0.000259\$ / Student$

b) Yes, the percent of students in the top 10% is statistically significant, since the P-value of 0.033 is less than $\alpha = 0.05$.

c) Yes, the amount of money spent per student is statistically significant, since the P-value of 0.003 is less than $\alpha = 0.05$.

Review of Part VI – Accessing Associations Between Variables

1. **Herbal cancer.**

   H_0: The cancer rate for those taking the herb is the same as the cancer rate for those not taking the herb. $(p_{Herb} = p_{Not} \text{ or } p_{Herb} - p_{Not} = 0)$

   H_A: The cancer rate for those taking the herb is higher than the cancer rate for those not taking the herb. $(p_{Herb} > p_{Not} \text{ or } p_{Herb} - p_{Not} > 0)$

2. **Birth days.**

   a) If births are distributed uniformly across all days, we expect the number of births on each day to be $np = (72)(1/7) \approx 10.29$.

   b) **Randomization condition:** The 72 births are likely to be representative of all births at the hospital with regards to day of birth.
   10% condition: 72 births are less than 10% of the births.
   Success/Failure condition: The expected number of births on a particular day of the week is $np = (72)(1/7) \approx 10.29$ and the expected number of births not on that particular day is $nq = (72)(6/7) \approx 61.71$. These are both greater than 10, so the sample is large enough.

   Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion of 72 births that occur on a given day of the week.

   $\mu_{\hat{p}} = p = 1/7 \approx 0.1429$

   $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(1/7)(6/7)}{72}} \approx 0.04124$

   There were 7 births on Mondays, so $\hat{p} = \frac{7}{72} \approx 0.09722$. This is only about a 1.11 standard deviations below the expected proportion, so there's no evidence that this is unusual.

   c) The 17 births on Tuesdays represent an unusual occurrence. For Tuesdays, $\hat{p} = \frac{17}{72} \approx 0.2361$, which is about 2.26 standard deviations above the expected proportion of births. There is evidence to suggest that the proportion of births on Tuesdays is higher than expected, if births are distributed uniformly across days.

   d) Some births are scheduled for the convenience of the doctor and/or the mother.

3. Surgery and germs.

a) Lister imposed a treatment, the use of carbolic acid as a disinfectant. This is an experiment.

b) H_0: The survival rate when carbolic acid is used is the same as the survival rate when carbolic acid is not used. $(p_C = p_N \text{ or } p_C - p_N = 0)$

H_A: The survival rate when carbolic acid is used is greater than the survival rate when carbolic acid is not used. $(p_C > p_N \text{ or } p_C - p_N > 0)$

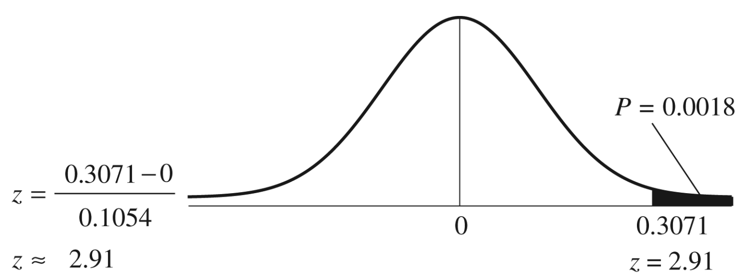
Randomization condition: There is no mention of random assignment. Assume that the two groups of patients were similar, and amputations took place under similar conditions, with the use of carbolic acid being the only variable.
10% condition: 40 and 35 are both less than 10% of all possible amputations.
Independent samples condition: It is reasonable to think that the groups were not related in any way.
Success/Failure condition: $n\hat{p}$ (carbolic acid) = 34, $n\hat{q}$ (carbolic acid) = 6, $n\hat{p}$ (none) = 19, and $n\hat{q}$ (none) = 16. The number of patients who died in the carbolic acid group is only 6, but the expected number of deaths using the pooled proportion, $n\hat{q}_{\text{pooled}} = (40)(\frac{22}{75}) = 11.7$, so the samples are both large enough.

Since the conditions have been satisfied, we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}(\hat{p}_C - \hat{p}_N) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_C} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N}} = \sqrt{\frac{\left(\frac{53}{75}\right)\left(\frac{22}{75}\right)}{40} + \frac{\left(\frac{53}{75}\right)\left(\frac{22}{75}\right)}{35}} \approx 0.1054.$$

The observed difference between the proportions is:
0.85 – 0.5429 = 0.3071.

Since the P-value = 0.0018 is low, we reject the null hypothesis. There is strong evidence that the survival rate is higher when carbolic acid is used to disinfect the operating room than when carbolic acid is not used.



c) We don't know whether or not patients were randomly assigned to treatments, and we don't know whether or not blinding was used.

4. Free throws 2011.

a) **Randomization condition:** Assume that these free throws are representative of the free throw ability of these players.
10% condition: 227 and 419 are less than 10% of all possible free throws.
Independent samples condition: The free throw abilities of these two players should be independent.
Success/Failure condition: $n\hat{p}$ (Curry) = 212, $n\hat{q}$ (Curry) = 15, $n\hat{p}$ (Billups) = 384, and $n\hat{q}$ (Billups) = 35 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$(\hat{p}_C - \hat{p}_B) \pm z^* \sqrt{\frac{\hat{p}_C\hat{q}_C}{n_C} + \frac{\hat{p}_B\hat{q}_B}{n_B}}$$

$$= \left(\tfrac{212}{227} - \tfrac{384}{419}\right) \pm 1.960\sqrt{\frac{\left(\frac{212}{227}\right)\left(\frac{15}{227}\right)}{227} + \frac{\left(\frac{384}{419}\right)\left(\frac{35}{419}\right)}{419}} = (-0.0243,\ 0.0592)$$

We are 95% confident that Stephen Curry's true free throw percentage is between 2.4% worse and 5.9% better than Chauncey Billups's.

b) Since the interval for the difference in percentage of free throws made includes 0, it is uncertain who the better free throw shooter is.

5. Twins.

H_0: The proportion of preterm twin births in 2000 is the same as the proportion of preterm twin births in 2010. $(p_{2000} = p_{2010} \text{ or } p_{2000} - p_{2010} = 0)$

H_A: The proportion of preterm twin births in 1990 is the less than the proportion of preterm twin births in 2000. $(p_{2000} < p_{2010} \text{ or } p_{2000} - p_{2010} < 0)$

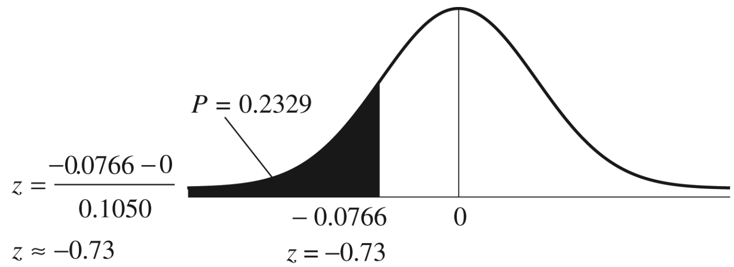
Randomization condition: Assume that these births are representative of all twin births.
10% condition: 43 and 48 are both less than 10% of all twin births.
Independent samples condition: The samples are from different years, so they are unlikely to be related.
Success/Failure condition: $n\hat{p}$ (2000) = 20, $n\hat{q}$ (2000) = 23, $n\hat{p}$ (2010) = 26, and $n\hat{q}$ (2010) = 22 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}(\hat{p}_{2000} - \hat{p}_{2010}) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{2000}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{2010}}} = \sqrt{\frac{\left(\frac{46}{91}\right)\left(\frac{45}{91}\right)}{43} + \frac{\left(\frac{46}{91}\right)\left(\frac{45}{91}\right)}{48}} \approx 0.1050$$

The observed difference between the proportions is: 0.4651 – 0.5417 = – 0.0766

Since the P-value = 0.2329 is high, we fail to reject the null hypothesis. There is no evidence of an increase in the proportion of preterm twin births from 2000 to 2010, at least not at this large city hospital.



6. **Eclampsia.**

a) **Randomization condition:** Although not specifically stated, these results are from a large-scale experiment, which was undoubtedly properly randomized.
10% condition: 4999 and 4993 are less than 10% of all pregnant women.
Independent samples condition: Subjects were randomly assigned to the treatments.
Success/Failure condition: $n\hat{p}$ (mag. sulf.) = 1201, $n\hat{q}$ (mag. sulf.) = 3798, $n\hat{p}$ (placebo) = 228, and $n\hat{q}$ (placebo) = 4765 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{MS} - \hat{p}_N\right) \pm z^* \sqrt{\frac{\hat{p}_{MS}\hat{q}_{MS}}{n_{MS}} + \frac{\hat{p}_N\hat{q}_N}{n_N}}$$

$$= \left(\tfrac{1201}{4999} - \tfrac{228}{4993}\right) \pm 1.960 \sqrt{\frac{\left(\frac{1201}{4999}\right)\left(\frac{3798}{4999}\right)}{4999} + \frac{\left(\frac{228}{4993}\right)\left(\frac{4765}{4993}\right)}{4993}} = \left(0.181,\ 0.208\right)$$

We are 95% confident that the proportion of pregnant women who will experience side effects while taking magnesium sulfide will be between 18.1% and 20.8% higher than the proportion of women that will experience side effects while not taking magnesium sulfide.

b) H_0: The proportion of pregnant women who will develop eclampsia is the same for women taking magnesium sulfide as it is for women not taking magnesium sulfide. $\left(p_{MS} = p_N \text{ or } p_{MS} - p_N = 0\right)$

H_A: The proportion of pregnant women who will develop eclampsia is lower for women taking magnesium sulfide than for women not taking magnesium sulfide. $\left(p_{MS} < p_N \text{ or } p_{MS} - p_N < 0\right)$

Success/Failure condition: $n\hat{p}$ (mag. sulf.) = 40, $n\hat{q}$ (mag. sulf.) = 4959, $n\hat{p}$ (placebo) = 96, and $n\hat{q}$ (placebo) = 4897 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied (some in part a), we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_{MS}-\hat{p}_{N}\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{MS}}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{N}}}=\sqrt{\frac{\left(\frac{136}{9992}\right)\left(\frac{9856}{9992}\right)}{4999}+\frac{\left(\frac{136}{9992}\right)\left(\frac{9856}{9992}\right)}{4993}}\approx 0.002318$$

The observed difference between the proportions is 0.00800 – 0.01923 = – 0.01123, which is approximately 4.84 standard errors below the expected difference in proportion of 0.

Since the P-value = 6.4×10^{-7} is very low, we reject the null hypothesis. There is strong evidence that the proportion of pregnant women who develop eclampsia will be lower for women taking magnesium sulfide than for those not taking magnesium sulfide.

7. Eclampsia deaths.

a) H_0: The proportion of pregnant women who die after developing eclampsia is the same for women taking magnesium sulfide as it is for women not taking magnesium sulfide. $\left(p_{MS}=p_{N}\ \text{ or }\ p_{MS}-p_{N}=0\right)$

H_A: The proportion of pregnant women who die after developing eclampsia is lower for women taking magnesium sulfide than for women not taking magnesium sulfide. $\left(p_{MS}<p_{N}\ \text{ or }\ p_{MS}-p_{N}<0\right)$

b) **Randomization condition:** Although not specifically stated, these results are from a large-scale experiment, which was undoubtedly properly randomized.
10% condition: 40 and 96 are less than 10% of all pregnant women.
Independent samples condition: Subjects were randomly assigned to the treatments.
Success/Failure condition: $n\hat{p}$ (mag. sulf.) = 11, $n\hat{q}$ (mag. sulf.) = 29, $n\hat{p}$ (placebo) = 20, and $n\hat{q}$ (placebo) = 76 are all greater than 10, so both samples are large enough.

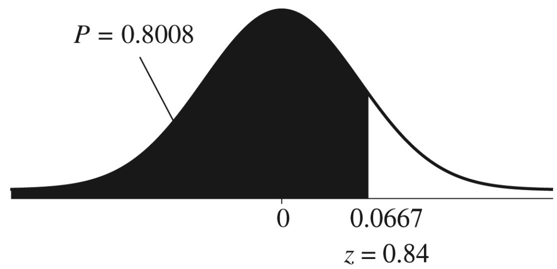
Since the conditions have been satisfied, we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_{MS}-\hat{p}_{N}\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{MS}}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{N}}}=\sqrt{\frac{\left(\frac{31}{136}\right)\left(\frac{105}{136}\right)}{40}+\frac{\left(\frac{31}{136}\right)\left(\frac{105}{136}\right)}{96}}\approx 0.07895.$$

c) The observed difference between the proportions is: 0.275 – 0.2083 = 0.0667

Since the P-value = 0.8008 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of women who may die after developing eclampsia is lower for women taking magnesium sulfide than for women who are not taking the drug.

$$z = \frac{0.0667 - 0}{0.07895}$$
$$z \approx 0.84$$



d) There is not sufficient evidence to conclude that magnesium sulfide is effective in preventing death when eclampsia develops.

e) If magnesium sulfide is effective in preventing death when eclampsia develops, then we have made a Type II error.

f) To increase the power of the test to detect a decrease in death rate due to magnesium sulfide, we could increase the sample size or increase the level of significance.

g) Increasing the sample size lowers variation in the sampling distribution, but may be costly. The sample size is already quite large. Increasing the level of significance increases power by increasing the likelihood of rejecting the null hypothesis, but increases the chance of making a Type I error, namely thinking that magnesium sulfide is effective when it is not.

8. Perfect pitch.

a) H_0: The proportion of Asian students with perfect pitch is the same as the proportion of non-Asians with perfect pitch. $\left(p_A = p_N \text{ or } p_A - p_N = 0\right)$

H_A: The proportion of Asian students with perfect pitch is the different than the proportion of non-Asians with perfect pitch. $\left(p_A \neq p_N \text{ or } p_A - p_N \neq 0\right)$

b) Since $P < 0.0001$, which is very low, we reject the null hypothesis. There is strong evidence of a difference in the proportion of Asians with perfect pitch and the proportion of non-Asians with perfect pitch. There is evidence that Asians are more likely to have perfect pitch.

c) If there is no difference in the proportion of students with perfect pitch, we would expect the observed difference of 25% to be seen simply due to sampling variation in less than 1 out of every 10,000 samples of 2700 students.

d) The data do not prove anything about genetic differences causing differences in perfect pitch. Asians are merely more likely to have perfect pitch. There may be lurking variables other than genetics that cause the higher rate of perfect pitch.

9. More errors.

a) Since a treatment (the additive) is imposed, this is an experiment.

b) The company is only interested in an increase in fuel economy, so they will perform a one-sided test.

c) The company will make a Type I error if they decide that the additive increases the fuel economy, when it actually makes no difference in the fuel economy.

d) The company will make a Type II error if they decide that the additive does not increase the fuel economy, when it actually increases fuel economy.

e) The additive manufacturer would consider a Type II error more serious. If the test caused the company to conclude that the manufacturer's product didn't work, and it actually did, the manufacturer would lose sales, and the reputation of their product would suffer.

f) Since this was a controlled experiment, the company can conclude that the additive is the reason that the fuel economy has increased. They should be cautious recommending it for all cars. There is evidence that the additive works well for fleet vehicles, which get heavy use. It might not be effective in cars with a different pattern of use than fleet vehicles.

10. Preemies.

a) **Randomization condition:** Assume that these kids are representative of all kids.
10% condition: 242 and 233 are less than 10% of all kids.
Independent samples condition: The groups are independent.
Success/Failure condition: $n\hat{p}$ (preemies) = (242)(0.74) = 179, $n\hat{q}$ (preemies) = (242)(0.26) = 63, $n\hat{p}$ (normal weight) = (233)(0.83) = 193, and $n\hat{q}$ (normal weight) = 40 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_N - \hat{p}_P\right) \pm z^* \sqrt{\frac{\hat{p}_N\hat{q}_N}{n_N} + \frac{\hat{p}_P\hat{q}_P}{n_P}}$$

$$= (0.83 - 0.74) \pm 1.960\sqrt{\frac{(0.83)(0.17)}{233} + \frac{(0.74)(0.26)}{242}} = (0.017,\ 0.163)$$

We are 95% confident that between 1.7% and 16.3% more normal birth-weight children graduated from high school than children who were born premature.

b) Since the interval for the difference in percentage of high school graduates is above 0, there is evidence normal birth-weight children graduate from high school at a greater rate than premature children.

c) If preemies do not have a lower high school graduation rate than normal birth-weight children, then we made a Type I error. We rejected the null hypothesis of "no difference" when we shouldn't have.

11. Crawling.

a) H_0: The mean age at which babies begin to crawl is the same whether the babies were born in January or July. $\left(\mu_{Jan} = \mu_{July} \text{ or } \mu_{Jan} - \mu_{July} = 0\right)$

H_A: There is a difference in the mean age at which babies begin to crawl, depending on whether the babies were born in January or July. $\left(\mu_{Jan} \neq \mu_{July} \text{ or } \mu_{Jan} - \mu_{July} \neq 0\right)$

Independent groups assumption: The groups of January and July babies are independent.
Randomization condition: Although not specifically stated, we will assume that the babies are representative of all babies.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are fairly large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 43.68 degrees of freedom (from the approximation formula).

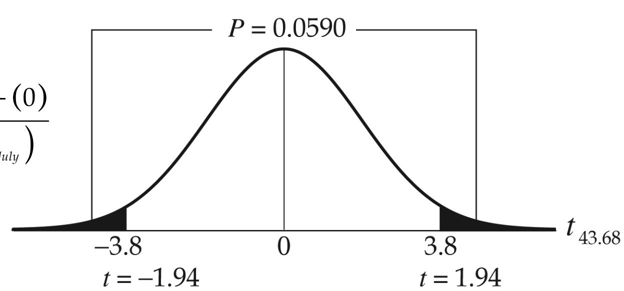
We will perform a two-sample t-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_{Jan} - \bar{y}_{July}) = \sqrt{\frac{7.08^2}{32} + \frac{6.91^2}{21}} \approx 1.9596$.

The observed difference between the mean ages is 29.84 – 33.64 = – 3.8 weeks.

$$t = \frac{\left(\bar{y}_{Jan} - \bar{y}_{July}\right) - (0)}{SE\left(\bar{y}_{Jan} - \bar{y}_{July}\right)}$$

$$t \approx \frac{-3.8 - 0}{1.9596}$$

$$t \approx -1.94$$



Since the P-value = 0.0590 is fairly low, we reject the null hypothesis. There is some evidence that mean age at which babies crawl is different for January and July babies. July babies appear to crawl a bit earlier than July babies, on average. Since the evidence is not strong, we might want to do some more research into this claim.

b) H_0: The mean age at which babies begin to crawl is the same whether the babies were born in April or October. $\left(\mu_{Apr} = \mu_{Oct} \text{ or } \mu_{Apr} - \mu_{Oct} = 0\right)$

H_A: There is a difference in the mean age at which babies begin to crawl, depending on whether the babies were born in April or October. $\left(\mu_{Apr} \neq \mu_{Oct} \text{ or } \mu_{Apr} - \mu_{Oct} \neq 0\right)$

The conditions (with minor variations) were checked in part a.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 59.40 degrees of freedom (from the approximation formula).

We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_{Apr} - \bar{y}_{Oct}) = \sqrt{\frac{6.21^2}{26} + \frac{7.29^2}{44}} \approx 1.6404$.

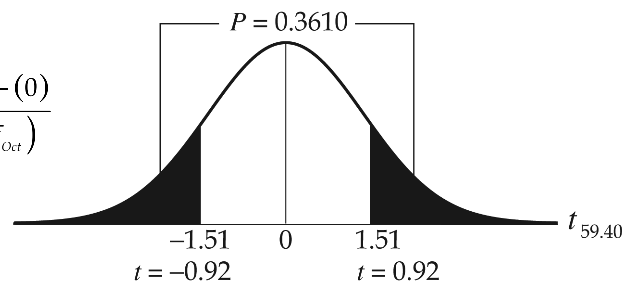
The observed difference between the mean ages is 31.84 – 33.35 = – 1.51 weeks.

Since the *P*-value = 0.3610 is high, we fail to reject the null hypothesis. There is no evidence that mean age at which babies crawl is different for April and October babies.

$$t = \frac{\left(\bar{y}_{Apr} - \bar{y}_{Oct}\right) - (0)}{SE\left(\bar{y}_{Apr} - \bar{y}_{Oct}\right)}$$

$$t \approx \frac{-1.51 - 0}{1.6404}$$

$$t \approx -0.92$$



c) These results are not consistent with the researcher's claim. We have slight evidence in one test and no evidence in the other. The researcher will have to do better than this to convince us!

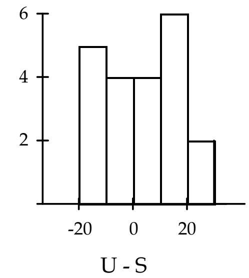
12. Mazes and smells.

H_0: The mean difference in maze times with and without the presence of a floral aroma is zero. $\left(\mu_d = 0\right)$

H_A: The mean difference in maze times with and without the presence of a floral aroma (unscented – scented) is greater than zero. $\left(\mu_d > 0\right)$

Paired data assumption: Each subject is paired with himself or herself.
Randomization condition: Subjects were randomized with respect to whether they did the scented trial first or second.

Nearly Normal condition: The histogram of differences between unscented and scented scores shows a distribution that could have come from a Normal population, and the sample size is fairly large.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with $21 - 1 = 20$ degrees of freedom, $t_{20}\left(0, \frac{13.0087}{\sqrt{21}}\right)$.

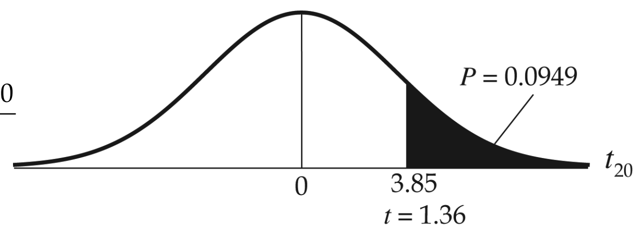
We will use a paired t-test, (unscented – scented) with $\overline{d} = 3.85238$ seconds.

Since the P-value = 0.0949 is fairly high, we fail to reject the null hypothesis. There is little evidence that the mean difference in time required to complete the maze is greater than zero. The floral scent didn't appear to cause lower times.

$$t = \frac{\overline{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

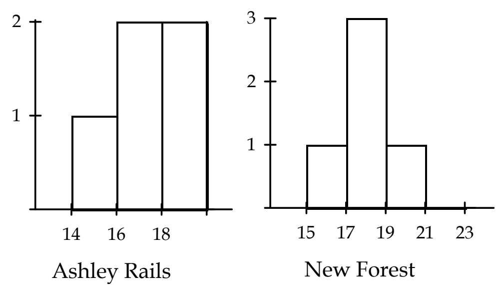
$$t = \frac{3.85238 - 0}{\frac{13.0087}{\sqrt{21}}}$$

$$t \approx 1.36$$



13. Pottery.

Independent groups assumption: The pottery samples are from two different sites.
Randomization condition: It is reasonable to think that the pottery samples are representative of all pottery at that site with respect to aluminum oxide content.
Nearly Normal condition: The histograms of aluminum oxide content are roughly unimodal and symmetric.



Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 7 degrees of freedom (from the approximation formula). We will construct a two-sample t-interval, with 95% confidence.

$$(\overline{y}_{AR} - \overline{y}_{NF}) \pm t^*_{df}\sqrt{\frac{s^2_{AR}}{n_{AR}} + \frac{s^2_{NF}}{n_{NF}}}$$

$$= (17.32 - 18.18) \pm t^*_7\sqrt{\frac{1.65892^2}{5} + \frac{1.77539^2}{5}} \approx (-3.37, 1.65)$$

We are 95% confident that the difference in the mean percentage of aluminum oxide content of the pottery at the two sites is between –3.37% and 1.65%. Since 0 is in the interval, there is no evidence that the aluminum oxide content at the two sites is different. It would be reasonable for the archaeologists to think that the same ancient people inhabited the sites.

14. Grant writing.

H_0: The proportion of NIH grants accepted is the same for white applicants and black applicants. $\left(p_{White} = p_{Black} \text{ or } p_{White} - p_{Black} = 0\right)$

H_A: The proportion of NIH grants accepted is the greater for white applicants than for black applicants. $\left(p_{White} > p_{Black} \text{ or } p_{White} - p_{Black} > 0\right)$

Independence assumption: This is not a random sample of applicants, but is more likely to be all applications. We will assume that the applications are independent of each other.
Independent samples condition: We will assume the groups are independent.
Success/Failure cond.: $n\hat{p}$ (white) = 15,700, $n\hat{q}$ (white) = 42,448, $n\hat{p}$ (black) = 198, and $n\hat{q}$ (black) = 966 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_p\left(\hat{p}_{White} - \hat{p}_{White}\right) = \sqrt{\frac{\hat{p}_p\hat{q}_p}{n_{White}} + \frac{\hat{p}_p\hat{q}_p}{n_{Black}}} = \sqrt{\frac{\left(\frac{15,898}{59,312}\right)\left(\frac{43,414}{59,312}\right)}{58,148} + \frac{\left(\frac{15,898}{59,312}\right)\left(\frac{43,414}{59,312}\right)}{1164}} \approx 0.01311$$

The observed difference between the proportions is 0.2700 – 0.1701 = 0.0999.

$$z = \frac{\left(\hat{p}_W - \hat{p}_B\right) - 0}{SE\left(\hat{p}_W - \hat{p}_B\right)}$$

$$z = \frac{0.0999}{0.01311} \approx 7.62$$

Since the P-value < 0.0001 is very low, we reject the null hypothesis. There is strong evidence that the proportion of white applicants who are receiving NIH grants is higher than the proportion of black applicants receiving grants.

b) This was a retrospective observational study. In order to make the inference, we must assume that the researchers used in the study are representative of all researchers would could possibly apply for funding from NIH.

15. Feeding fish.

a) If there is no difference in the average fish sizes, the chance of observing a difference this large, or larger, just by natural sampling variation is 0.1%.

b) There is evidence that largemouth bass that are fed a natural diet are larger. The researchers would advise people who raise largemouth bass to feed them a natural diet.

c) If the advice is incorrect, the researchers have committed a Type I error.

16. Driving fatalities 2011.

We have two independent samples, but the sample sizes are very small and it would be hard to generalize the results for New England and the Southwestern states based on summary numbers from each state. We also don't know which states the figures are coming from. These values are not appropriate for inference.

17. Age.

a) **Independent groups assumption:** The group of patients with and without cardiac disease are not related in any way.
Randomization condition: Assume that these patients are representative of all people.
Normal population assumption: We don't have the actual data, so we will assume that the population of ages of patients is Normal.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 670 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_{Card} - \bar{y}_{None}) \pm t^*_{df}\sqrt{\frac{s^2_{Card}}{n_{Card}} + \frac{s^2_{None}}{n_{None}}} = (74.0 - 69.8) \pm t^*_{670}\sqrt{\frac{7.9^2}{450} + \frac{8.7^2}{2397}} \approx (3.39,\ 5.01)$$

We are 95% confident that the mean age of patients with cardiac disease is between 3.39 and 5.01 years higher than the mean age of patients without cardiac disease.

b) Older patients are at greater risk for a variety of health problems. If an older patient does not survive a heart attack, the researchers will not know to what extent depression was involved, because there will be a variety of other possible variables influencing the death rate. Additionally, older patients may be more (or less) likely to be depressed than younger ones.

18. Smoking.

Randomization condition: Assume the patients are representative of all people.
10% condition: 2397 patients without cardiac disease and 450 patients with cardiac disease are both less than 10% of all people.
Independent groups assumption: The group of patients with and without cardiac disease are not related in any way.

Success/Failure condition: $n\hat{p}$(cardiac) = (450)(0.32) = 144, $n\hat{q}$(cardiac) = (450)(0.68) = 306, $n\hat{p}$(none) = (2397)(0.237) = 568, and $n\hat{q}$(none) = (2397)(0.763) = 1829 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion *z*-interval.

$$(\hat{p}_{Card} - \hat{p}_{None}) \pm z^* \sqrt{\frac{\hat{p}_{Card}\hat{q}_{Card}}{n_{Card}} + \frac{\hat{p}_{None}\hat{q}_{None}}{n_{None}}}$$

$$= (0.32 - 0.237) \pm 1.960\sqrt{\frac{(0.32)(0.68)}{450} + \frac{(0.237)(0.763)}{2397}} = (0.0367,\ 0.1293)$$

We are 95% confident that the proportion of smokers is between 3.67 percentage points and 12.93 percentage points higher for patients with cardiac disease than for patients without cardiac disease.

b) Since the confidence interval does not contain 0, there is evidence that cardiac patients have a higher rate of smokers than the patients without cardiac disease. The two groups are different.

c) Smoking could be a confounding variable. Smokers have a higher risk of other health problems that may be associated with their ability to survive a heart attack.

19. Eating disorders.

a) **Randomization condition:** Hopefully, the students were selected randomly.
10% condition: 150 and 200 are less than 10% of all students.
Independent samples condition: The groups are independent.
Success/Failure condition: $n\hat{p}$(Muslim) = 46, $n\hat{q}$(Muslim) = 104, $n\hat{p}$(Christian) = 34, and $n\hat{q}$(Christian) = 166 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion *z*-interval.

$$(\hat{p}_M - \hat{p}_C) \pm z^* \sqrt{\frac{\hat{p}_M\hat{q}_M}{n_M} + \frac{\hat{p}_C\hat{q}_C}{n_C}}$$

$$= \left(\tfrac{46}{150} - \tfrac{34}{200}\right) \pm 1.960\sqrt{\frac{\left(\tfrac{46}{150}\right)\left(\tfrac{104}{150}\right)}{150} + \frac{\left(\tfrac{34}{200}\right)\left(\tfrac{166}{200}\right)}{200}} = (0.046,\ 0.227)$$

We are 95% confident that the percentage of Muslim students who have an eating disorder is between 4.6 and 22.7 percentage points higher than the percentage of Christian students who have an eating disorder.

b) Although caution in generalizing must be used since the study was restricted to the Spanish city of Ceuta, it appears there is a true difference in the prevalence of eating disorders. We can conclude this because the entire interval is above 0.

20. Cesareans.

H_0: The proportion of births involving cesarean deliveries is the same in Vermont and New Hampshire. $(p_{VT} = p_{NH}$ or $p_{VT} - p_{NH} = 0)$

H_A: The proportion of births involving cesarean deliveries is different in Vermont and New Hampshire. $(p_{VT} \neq p_{NH}$ or $p_{VT} - p_{NH} \neq 0)$

Random condition: Hospitals were randomly selected.
10% condition: 223 and 186 are both less than 10% of all births in these states.
Independent samples condition: Vermont and New Hampshire are different states!
Success/Failure cond.: $n\hat{p}$ (VT) = (223)(0.166) = 37, $n\hat{q}$ (VT) = (223)(0.834) = 186, $n\hat{p}$ (NH) = (186)(0.188) = 35, and $n\hat{q}$ (NH) = (186)(0.812) = 151 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}(\hat{p}_{VT} - \hat{p}_{NH}) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{VT}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{NH}}}.$$

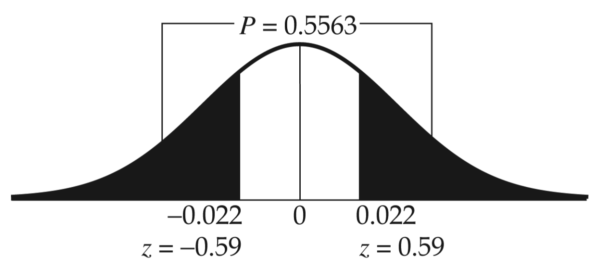
$$= \sqrt{\frac{(0.176)(0.824)}{223} + \frac{(0.176)(0.824)}{186}} \approx 0.03782$$

The observed difference between the proportions is 0.166 – 0.188 = – 0.022.

Since the P-value = 0.5563 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of cesarean births in Vermont is different from the proportion of cesarean births in New Hampshire.

$$z = \frac{-0.22 - 0}{0.03782}$$

$$z \approx -0.59$$



21. Teach for America.

H_0: The mean score of students with certified teachers is the same as the mean score of students with uncertified teachers. $(\mu_C = \mu_U$ or $\mu_C - \mu_U = 0)$

H_A: The mean score of students with certified teachers is greater than as the mean score of students with uncertified teachers. $(\mu_C > \mu_U$ or $\mu_C - \mu_U > 0)$

Independent groups assumption: The certified and uncertified teachers are independent groups.

Randomization condition: Assume the students studied were representative of all students.
Nearly Normal condition: We don't have the actual data, so we can't look at the graphical displays, but the sample sizes are large, so we can proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 86 degrees of freedom (from the approximation formula).

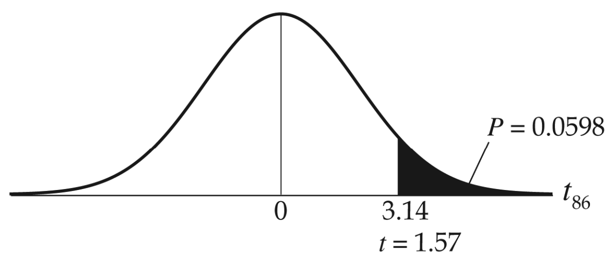
We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_C - \bar{y}_U) = \sqrt{\frac{9.31^2}{44} + \frac{9.43^2}{44}} \approx 1.9977$.

The observed difference between the mean scores is 35.62 – 32.48 = 3.14.

$$t = \frac{(\bar{y}_C - \bar{y}_U) - (0)}{SE(\bar{y}_C - \bar{y}_U)}$$

$$t \approx \frac{3.14}{1.9977}$$

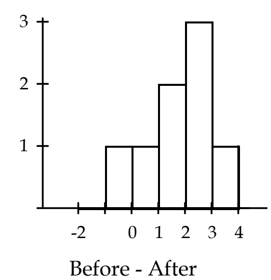
$$t \approx 1.57$$



Since the *P*-value = 0.0598 is fairly low, we reject the null hypothesis. There is some evidence that students with certified teachers had mean scores higher than students with uncertified teachers. However, since the *P*-value is not extremely low, further investigation is recommended.

22. Legionnaires' disease.

Paired data assumption: The data are paired by room.
Randomization condition: We will assume that these rooms are representative of all rooms at the hotel.
Nearly Normal condition: The histogram of differences between before and after measurements is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's *t*-model with 8 – 1 = 7 degrees of freedom. We will find a paired *t*-interval, with 95% confidence.

$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 1.6125 \pm t^*_7\left(\frac{1.23801}{\sqrt{8}}\right) \approx (0.58,\ 2.65)$$

We are 95% confident that the mean difference in the bacteria counts is between 0.58 and 2.65 colonies per cubic foot of air. Since the entire interval is above 0, there is evidence that the new air-conditioning system was effective in reducing average bacteria counts.

23. Teach for America, Part II.

H_0: The mean score of students with certified teachers is the same as the mean score of students with uncertified teachers. $(\mu_C = \mu_U \text{ or } \mu_C - \mu_U = 0)$

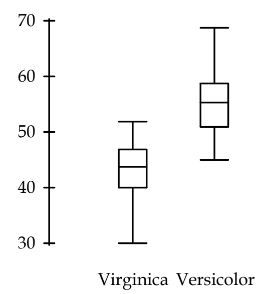
H_A: The mean score of students with certified teachers is different than the mean score of students with uncertified teachers. $(\mu_C \neq \mu_U \text{ or } \mu_C - \mu_U \neq 0)$

Mathematics: Since the *P*-value = 0.002 is low, we reject the null hypothesis. There is strong evidence that students with certified teachers have different mean math scores than students with uncertified teachers. Students with certified teachers do better.

Language: Since the *P*-value = 0.045 is fairly low, we reject the null hypothesis. There is evidence that students with certified teachers have different mean language scores than students with uncertified teachers. Students with certified teachers do better. However, since the *P*-value is not extremely low, further investigation is recommended.

24. Irises.

a) Parallel boxplots of the distributions of petal lengths for the two species of flower are at the right. No units are specified, but millimeters seems like a reasonable guess.



b) The petals of *versicolor* are generally longer than the petals of *virginica.* Both distributions have about the same range, and both distributions are fairly symmetric.

c) **Independent groups assumption:** The two species of flowers are independent.
Randomization condition: It is reasonable to assume that these flowers are representative of their species.
Nearly Normal condition: The boxplots show distributions of petal lengths that are reasonably symmetric with no outliers. Additionally, the samples are large.
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 97.92 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_{Ver} - \bar{y}_{Vir}) \pm t_{df}^{*}\sqrt{\frac{s_{Ver}^{2}}{n_{Ver}} + \frac{s_{Vir}^{2}}{n_{Vir}}} = (55.52 - 43.22) \pm t_{97.92}^{*}\sqrt{\frac{5.519^{2}}{50} + \frac{5.362^{2}}{50}} \approx (10.14,\ 14.46)$$

d) We are 95% confident the mean petal length of *versicolor* irises is between 10.14 and 14.46 millimeters longer than the mean petal length of *virginica* irises.

e) Since the interval is completely above 0, there is strong evidence that the mean petal length of *versicolor* irises is greater than the mean petal length of *virginica* irises.

25. Insulin and diet.

a) H_0: People with high dairy consumption have IRS at the same rate as those with low dairy consumption. $\left(p_{High} = p_{Low} \text{ or } p_{High} - p_{Low} = 0\right)$

H_A: People with high dairy consumption have IRS at a different rate than those with low dairy consumption. $\left(p_{High} \neq p_{Low} \text{ or } p_{High} - p_{Low} \neq 0\right)$

Random condition: Assume the people studied are representative of all people.
10% condition: 102 and 190 are both less than 10% of all people.
Independent samples condition: The two groups are not related.
Success/Failure condition: $n\hat{p}$ (high) = 24, $n\hat{q}$ (high) = 78, $n\hat{p}$ (low) = 85, and $n\hat{q}$ (low) = 105 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_{High} - \hat{p}_{Low}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{High}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{Low}}}.$$

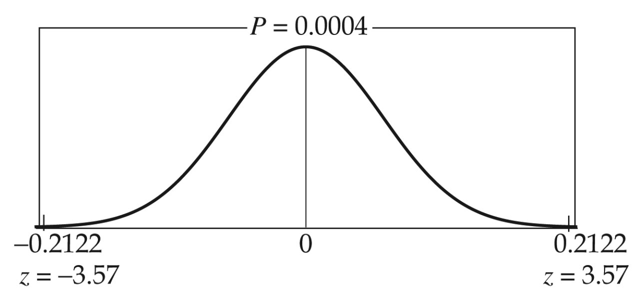
$$= \sqrt{\frac{(0.373)(0.627)}{102} + \frac{(0.373)(0.627)}{190}} \approx 0.05936$$

The observed difference between the proportions is 0.2352 – 0.4474 = – 0.2122.

Since the *P*-value = 0.0004 is very low, we reject the null hypothesis. There is strong evidence that the proportion of people with IRS is different for those who with high dairy consumption compared to those with low dairy consumption. People who consume dairy products more than 35 times per week appear less likely to have IRS than those who consume dairy products fewer than 10 times per week.

$$z = \frac{-0.2122 - 0}{0.05936}$$

$$z \approx -3.57$$



b) There is evidence of an association between the low consumption of dairy products and IRS, but that does not prove that dairy consumption influences the development of IRS. This is an observational study, and a controlled experiment is required to prove cause and effect.

26. Rainmakers?

Independent groups assumption: The two groups of clouds are independent.
Randomization condition: Researchers randomly assigned clouds to be seeded with silver iodide or not seeded.
Nearly Normal condition: We don't have the actual data, so we can't look at the distributions, but the means of group are significantly higher than the medians. This is an indication that the distributions are skewed to the right, with possible outliers. The samples sizes of 26 each are fairly large, so it should be safe to proceed, but we should be careful making conclusions, since there may be outliers.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 33.86 degrees of freedom (from the approximation formula). We will construct a two-sample *t*-interval, with 95% confidence.

$$(\bar{y}_S - \bar{y}_U) \pm t^*_{df}\sqrt{\frac{s_S^2}{n_S} + \frac{s_U^2}{n_U}}$$

$$= (441.985 - 164.588) \pm t^*_{33.86}\sqrt{\frac{650.787^2}{26} + \frac{278.426^2}{26}} \approx (-4.76,\ 559.56)$$

We are 95% confident the mean amount of rainfall produced by seeded clouds is between 4.76 acre-feet less than and 559.56 acre-feet more than the mean amount of rainfall produced by unseeded clouds.

Since the interval contains 0, there is little evidence that the mean rainfall produced by seeded clouds is any different from the mean rainfall produced by unseeded clouds. However, we shouldn't place too much faith in this conclusion. It is based on a procedure that is sensitive to outliers, and there may have been outliers present.

27. Genetics.

H_0: The proportions of traits are as specified by the ratio 1:3:3:9.
H_A: The proportions of traits are not as specified.

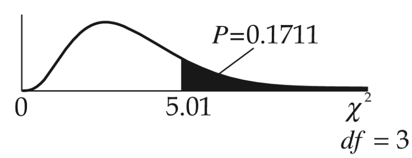
Counted data condition: The data are counts.
Randomization condition: Assume that these students are representative of all people.

Expected cell frequency condition: The expected counts (shown in the table) are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 – 1 = 3 degrees of freedom. We will use a chi-square goodness-of-fit test.

Trait	Observed	Expected	Residual = $(Obs - Exp)$	$(Obs - Exp)^2$	Component = $\frac{(Obs - Exp)^2}{Exp}$
Attached, noncurling	10	7.625	2.375	5.6406	0.73975
Attached, curling	22	22.875	– 0.875	0.7656	0.03347
Free, noncurling	31	22.875	8.125	66.0156	2.8859
Free, curling	59	68.625	– 9.625	92.6406	1.35
					$\sum \approx$ 5.01

$\chi^2 = 5.01$. Since the P-value = 0.1711 is high, we fail to reject the null hypothesis.



There is no evidence that the proportions of traits are anything other than 1:3:3:9.

28. Tableware.

a) Since there are 57 degrees of freedom, there were 59 different products included.

b) 84.5% of the variation in retail price is explained by the polishing time.

c) Assuming the conditions have been met, the sampling distribution of the regression slope can be modeled by a Student's t-model with (59 – 2) = 57 degrees of freedom. We will use a regression slope t-interval. For 95% confidence, use $t_{57}^* \approx 2.0025$, or estimate from the table $t_{50}^* \approx 2.009$.

$$b_1 \pm t_{n-2}^* \times SE(b_1) = 2.49244 \pm (2.0025) \times 0.1416 \approx (2.21,\ 2.78)$$

d) We are 95% confident that the average price increases between \$2.21 and \$2.78 for each additional minute of polishing time.

29. Hard water.

a) H_0: There is no linear relationship between calcium concentration in water and mortality rates for males. $(\beta_1 = 0)$

H_A: There is a linear relationship between calcium concentration in water and mortality rates for males. $(\beta_1 \neq 0)$

b) Assuming the conditions have been satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (61 – 2) = 59 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{Mortality} = 1676 - 3.23(Calcium)$, where mortality is measured in deaths per 100,000, and calcium concentration is measured in parts per million.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$
$$t = \frac{-3.23 - 0}{0.48}$$
$$t \approx -6.73$$

The value of $t = -6.73$. The *P*-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between calcium concentration and mortality. Towns with higher calcium concentrations tend to have lower mortality rates.

c) For 95% confidence, use $t^*_{59} \approx 2.001$, or estimate from the table $t^*_{50} \approx 2.009$.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = -3.23 \pm (2.001) \times 0.48 \approx (-4.19,\ -2.27)$$

d) We are 95% confident that the average mortality rate decreases by between 2.27 and 4.19 deaths per 100,000 for each additional part per million of calcium in drinking water.

30. Wealth distribution.

H_0: Income level and feelings about wealth distribution are independent.

H_A: There is an association between income level and feelings about wealth distribution.

Counted data condition: The data are counts.
Randomization condition: Although not specifically stated, the Gallup Poll was likely to be random.

	Should Redistribute *(Obs/Exp)*	Should Not *(Obs/Exp)*	No Opinion *(Obs/Exp)*
High Income	170 / 267.19	371 / 266.87	9 / 15.942
Middel Income	306 / 291.48	282 / 291.13	12 / 17.391
Low Income	362 / 279.33	184 / 279	29 / 16.667

Expected cell frequency condition: The expected counts are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 147.63,\text{ and the } P\text{-value} < 0.0001.$$

Since the P-value is low, we reject the null hypothesis. There is strong evidence of an association between income level and opinion on wealth distribution. Examination of the components shows that the low-income respondents are more likely to approve of redistribution when compared to the high-income respondents.

31. Wild horses.

a) Since there are 36 degrees of freedom, 38 herds of wild horses were studied.

b) **Straight enough condition:** The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The histogram of residuals is unimodal and symmetric.

c) Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (38 – 2) = 36 degrees of freedom. We will use a regression slope t-interval, with 95% confidence. Use $t_{35}^{*} \approx 2.030$ as an estimate.

$$b_1 \pm t_{n-2}^{*} \times SE(b_1) = 0.153969 \pm (2.030) \times 0.0114 \approx (0.131,\ 0.177)$$

d) We are 95% confident that the mean number of foals in a herd increases by between 0.131 and 0.177 foals for each additional adult horse.

e) The regression equation predicts that herds with 80 adults will have $-1.57835 + 0.153969(80) = 10.73917$ foals. The average size of the herds sampled is 110.237 adult horses. Use $t_{36}^{*} \approx 1.6883$, or use an estimate of $t_{35}^{*} \approx 1.690$, from the table.

$$\hat{y}_\nu \pm t_{n-2}^{*}\sqrt{SE^2(b_1)\cdot(x_\nu - \overline{x})^2 + \frac{s_e^2}{n} + s_e^2}$$
$$= 10.73917 \pm (1.6883)\sqrt{0.0114^2 \cdot (80 - 110.237)^2 + \frac{4.941^2}{38} + 4.941^2}$$
$$\approx (2.26,\ 19.21)$$

We are 90% confident that number of foals in a herd of 80 adult horses will be between 2.26 and 19.21. This prediction interval is too wide to be of much use.

32. Lefties and music.

H_0: The proportion of right-handed people who can match the tone is the same as the proportion of left-handed people who can match the tone.
$(p_L = p_R \text{ or } p_L - p_R = 0)$

H_A : The proportion of right-handed people who can match the tone is different from the proportion of left-handed people who can match the tone. $(p_L \neq p_R \text{ or } p_L - p_R \neq 0)$

Random condition: Assume that the people tested are representative of all people.
10% condition: 76 and 53 are both less than 10% of all people.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (right) = 38, $n\hat{q}$ (right) = 38, $n\hat{p}$ (left) = 33, and $n\hat{q}$ (left) = 20 are all greater than 10, so the samples are both large enough.

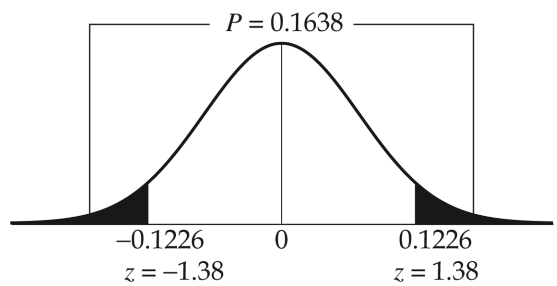
Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_L - \hat{p}_R\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_L} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_R}} = \sqrt{\frac{\left(\frac{71}{129}\right)\left(\frac{58}{129}\right)}{53} + \frac{\left(\frac{71}{129}\right)\left(\frac{58}{129}\right)}{76}} \approx 0.089.$$

The observed difference between the proportions is:
0.6226 – 0.5 = 0.1226.

Since the P-value = 0.1683 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of people able to match the tone differs between right-handed and left-handed people.

$$z = \frac{0.1226 - 0}{0.089}$$
$$z \approx 1.38$$



33. AP Statistics scores 2010.

a) H_0: The distribution of AP Statistics scores at Ithaca High School is the same as it is nationally.
H_A: The distribution of AP Statistics scores at Ithaca High School is different than it is nationally.

Counted data condition: The data are counts.
Randomization condition: Assume that this group of students is representative of all years at Ithaca High School.
Expected cell frequency condition: The expected counts (shown in the table) are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 5 – 1 = 4 degrees of freedom. We will use a chi-square goodness-of-fit test.

Score	Observed	Expected	Residual = $(Obs - Exp)$	Standardized Residual = $\frac{(Obs - Exp)}{\sqrt{Exp}}$	Component = $\frac{(Obs - Exp)^2}{Exp}$
5	26	12.416	4.445	3.8551	14.862
4	36	21.728	2.792	3.0618	9.3745
3	19	22.795	– 3.795	– 0.7949	0.63181
2	10	17.654	– 7.654	– 1.822	3.3184
1	6	22.407	– 16.41	– 3.466	12.014
					$\sum \approx 40.20$

$\chi^2 \approx 40.20$. Since the *P*-value is essentially 0, we reject the null hypothesis.

There is strong evidence that the distribution of scores at Ithaca High School is different than the national distribution. Students at IHS get fewer scores of 2 and 1 than expected, and more scores of 4 and 5 than expected.

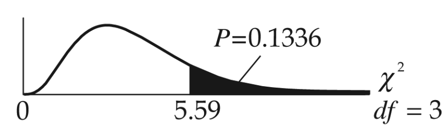
b) H_0: Gender and AP Statistics score are independent at Ithaca High School.

H_A: There is an association between gender and AP Statistics score at Ithaca High School.

Counted data condition: The data are counts.
Randomization condition: Assume this year's students are representative of all years.
Expected cell frequency condition: After combining the cells for scores of 2 and 1, the expected counts are all greater than 5.

	Boys **(*Obs/Exp*)**	**Girls** **(*Obs/Exp*)**
5	13 / 13.67	13 / 12.33
4	21 / 18.928	15 / 17.072
3	6 / 9.9897	13 / 9.0103
2 or 1	11 / 8.4124	5 / 7.5876

Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom. We will use a chi-square test for independence. (This is a test for independence, since we have one group that has been classified according to two variables, gender and score. However, if you said it was a test for homogeneity, since you were comparing two groups, no one would get terribly upset!)



With $\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 5.59$,

the *P*-value ≈ 0.1336.

Since P-value ≈ 0.1336 is high, we fail to reject the null hypothesis. There is no evidence of an association between gender and score at Ithaca High School. The boys seem to do just as well as the girls.

34. Twins.

H_0: There is no association between duration of pregnancy and level of prenatal care.

H_A: There is an association between duration of pregnancy and level of prenatal care.

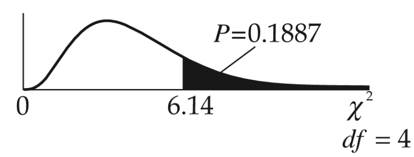
Counted data condition: The data are counts.
Randomization condition: Assume that these pregnancies are representative of all twin births.
Expected cell frequency condition: The expected counts are all greater than 5.

	Preterm (induced or Cesarean) *(Obs / Exp)*	**Preterm (without procedures)** *(Obs / Exp)*	**Term or postterm** *(Obs / Exp)*
Intensive	18 /16.676	15 / 15.579	28 / 28.745
Adequate	46 / 42.101	43 / 39.331	65 / 72.568
Inadequate	12 / 17.223	13 / 16.090	38 / 29.687

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 6.14,$$

and the P-value ≈ 0.1887.



Since the P-value ≈ 0.1887 is high, we fail to reject the null hypothesis. There is no evidence of an association between duration of pregnancy and level of prenatal care in twin births.

35. Twins, again.

H_0: The distributions of pregnancy durations are the same for the three years.

H_A: The distributions of pregnancy durations are different for the three years.

Counted data condition: The data are counts.
Independence assumption: Assume that the durations of the pregnancies are mutually independent.

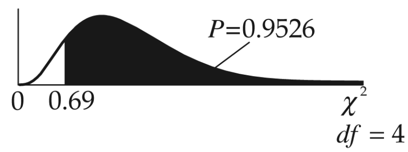
Expected cell frequency condition: The expected counts are all greater than 5.

	1990 *(Obs / Exp)*	**1995** *(Obs / Exp)*	**2000** *(Obs / Exp)*
Preterm (induced or Cesarean)	11 /12.676	13 / 13.173	19 / 17.150
Preterm (without procedures)	13 / 13.266	14 / 13.786	18 / 17.948
Term or postterm	27 / 25.058	26 / 26.04	32 / 33.902

Under these conditions, the sampling distribution of the test statistic is χ^2 on 4 degrees of freedom. We will use a chi-square test for homogeneity.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 0.69,$$

and the P-value $\approx$ 0.9526.



Since the P-value $\approx$ 0.9526 is high, we fail to reject the null hypothesis. There is no evidence that the distributions of the durations of pregnancies are different for the three years. It does not appear that the way the hospital deals with twin pregnancies has changed.

36. Retirement planning.

H_0: The proportion of men who are "a lot behind schedule" in retirement planning is the same as the proportion of women. $(p_M = p_W \text{ or } p_M - p_W = 0)$

H_A : The proportion of men who are "a lot behind schedule" in retirement planning is lower than the proportion of women. $(p_M < p_W \text{ or } p_M - p_W < 0)$

Random condition: Assume the survey was conducted randomly.
10% condition: 722 and 701 are both less than 10% of all man and women.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (Men) = 267, $n\hat{q}$ (Men) = 455, $n\hat{p}$ (Women) = 301, and $n\hat{q}$ (Women) = 400 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_M - \hat{p}_W\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_M} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_W}} = \sqrt{\frac{\left(\frac{568}{1423}\right)\left(\frac{855}{1423}\right)}{722} + \frac{\left(\frac{568}{1423}\right)\left(\frac{855}{1423}\right)}{701}} \approx 0.0260.$$

The observed difference between the proportions is: 0.3698 – 0.4294 = 0.0596.

Since the P-value = 0.0109 is low, we reject the null hypothesis. There is evidence that the proportion of women who will say they are "a lot behind schedule" for retirement planning is higher than the percentage of men that would say the same thing.

$$z = \frac{(\hat{p}_M - \hat{p}_W) - 0}{SE(\hat{p}_M - \hat{p}_W)}$$

$$z = \frac{-0.0596}{0.0260} \approx -2.29$$

37. Age and party.

a) There is one sample, classified according to two different variables, so we will perform a chi-square test for independence.

b) H_0: There is no association between age and political party for white voters.

H_A: There is an association between age and political party for white voters.

Counted data condition: The data are counts.
Randomization condition: These data are from a representative phone survey.
Expected cell frequency condition: The expected counts are all greater than 5.

	Leaning Republican *(Obs / Exp)*	Leaning Democrat *(Obs / Exp)*	Other *(Obs / Exp)*
18 – 29	274 / 271.97	216 /205.84	36 / 48.197
30 – 49	888 / 835.03	581 / 631.99	146 / 147.98
50 – 64	1173 /1213	962 / 918.05	211 / 214.96
65 +	1062 / 1077	812 / 815.13	209 / 190.86

Under these conditions, the sampling distribution of the test statistic is χ^2 on 6 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 16.54, \text{ and the } P\text{-value} \approx 0.0111.$$

c) Since the P-value ≈ 0.0111 is low, we reject the null hypothesis. There is evidence of an association between age and political party for white voters.

The table of standardized residuals is useful for the analysis of the differences. Looking for the largest standardized residuals, we can see there are more Republicans and fewer Democrats among those 30 – 49 than we expect. There are also fewer voters classified as "Other" than we would expect.

	Standardized Residuals		
	Leaning Republican	**Leaning Democrat**	**Other**
18 – 29	0.12	0.71	-1.76
30 – 49	1.83	-2.03	-0.16
50 – 64	-1.15	1.45	-0.27
65 +	-0.46	-0.11	1.31

38. Eye and hair color.

a) This is an attempt at linear regression. Regression inference is meaningless here, since eye and hair color are categorical variables.

b) This is an analysis based upon a chi-square test for independence.

H_0: Eye color and hair color are independent.

H_A: There is an association between eye color and hair color.

Since we have two categorical variables, this analysis seems appropriate. However, if you check the expected counts, you will find that 4 of them are less than 5. We would have to combine several cells in order to perform the analysis. (Always check the conditions!)

Since the value of chi-square is so high, it is likely that we would find an association between eye and hair color, even after the cells were combined. There are many cells of interest, but some of the most striking differences that would not be affected by cell combination involve people with fair hair. Blonds are likely to have blue eyes, and not likely to have brown eyes. Those with red hair are not likely to have brown eyes. Additionally, those with black hair are much more likely to have brown eyes than blue.

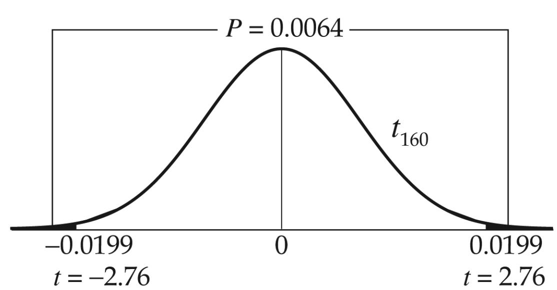
39. Depression and the Internet.

a) H_0: There is no linear relationship between depression and Internet usage. $(\beta_1 = 0)$

H_A: There is a linear relationship. $(\beta_1 \neq 0)$

Since the conditions for inference are satisfied (given), the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (162 – 2) = 160 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is:

$\widehat{DepressionAfter} = 0.565485 + 0.019948(InternetUsage)$.



The value of $t \approx 2.76$. The *P*-value of 0.0064 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between depression and Internet usage. Those with high levels of Internet usage tend to have high levels of depression. It should be noted, however, that although the evidence is strong, the association is quite weak, with $R^2 = 4.6\%$. The regression analysis only explains 4.6% of the variation in depression level.

b) The study says nothing about causality, merely association. Furthermore, there are almost certainly other factors involved. In fact, if 4.6% of the variation in depression level is related to Internet usage, the other 95.4% of the variation must be related to something else!

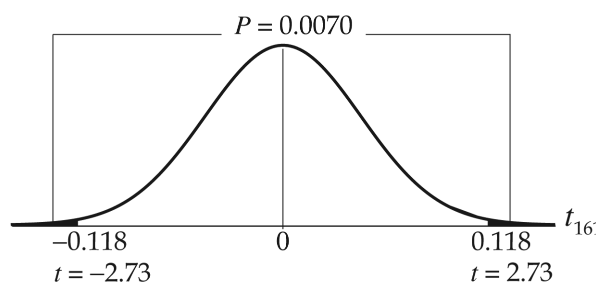
c) H_0: The mean difference in depression before and after the experiment is zero. $(\mu_d = 0)$

H_A: The mean difference in depression before and after the experiment is different than zero. $(\mu_d \neq 0)$

Since the conditions are satisfied (given), the sampling distribution of the difference can be modeled with a Student's *t*-model with 162 – 1 = 161 degrees of freedom, $t_{161}\left(0, \frac{0.552417}{\sqrt{162}}\right)$.

We will use a paired *t*-test, with $\bar{d} = -0.118457$.

Since the *P*-value = 0.0070 is very low, we reject the null hypothesis. There is strong evidence that the mean depression level changed over the course of the experiment. These data suggest that depression levels actually decreased.



40. Pregnancy.

a) H_0: The proportion of live births is the same for women under the age of 38 as it is for women over the age of 38. $(p_{<38} = p_{\geq 38}$ or $p_{<38} - p_{\geq 38} = 0)$

H_A : The proportion of live births is different for women under the age of 38 than for women over the age of 38. $(p_{<38} \neq p_{\geq 38}$ or $p_{<38} - p_{\geq 38} \neq 0)$

Random condition: Assume that the women studied are representative of all women.
10% condition: 157 and 89 are both less than 10% of all women.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n\hat{p}$ (under 38) = 42, $n\hat{q}$ (under 38) = 115, $n\hat{p}$ (38 and over) = 7, and $n\hat{q}$ (38 and over) = 82 are not all greater than 10, since the observed number of live births is only 7. However, if we check the pooled value, $n\hat{p}_{\text{pooled}}$ (38 and over) = (89)(0.191) = 17. All of the samples are large enough.

Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

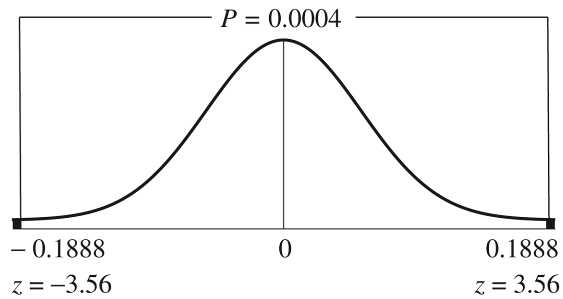
$$SE_{\text{pooled}}\left(\hat{p}_{<38}-\hat{p}_{\geq 38}\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{<38}}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{\geq 38}}}=\sqrt{\frac{\left(\frac{49}{246}\right)\left(\frac{197}{246}\right)}{157}+\frac{\left(\frac{49}{246}\right)\left(\frac{197}{246}\right)}{89}}\approx 0.0530.$$

The observed difference between the proportions is:
0.2675 – 0.0787 = 0.1888.

Since the P-value = 0.0004 is low, we reject the null hypothesis. There is strong evidence to suggest a difference in the proportion of live births for women under 38 and women 38 and over at this clinic. In fact, the evidence suggests that women under 38 have a higher proportion of live births.

$$z=\frac{0.1888-0}{0.0530}$$

$$z\approx 3.56$$



b) H_0: Age and birth rate are independent.

H_A: There is an association between age and birth rate

Counted data condition: The data are counts.
Randomization condition: Assume that these women are representative of all women.

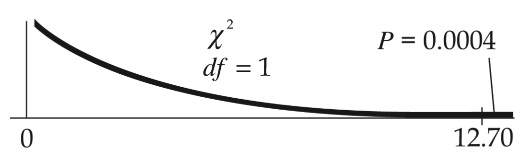
	Live birth (*Obs / Exp*)	No live birth (*Obs / Exp*)
Under 38	42 / 31.272	115 / 125.73
38 and over	7 / 17.728	82 / 71.27

Expected cell frequency condition: The expected counts are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 1 degree of freedom. We will use a chi-square test for independence.

$$\chi^2=\sum_{\text{all cells}}\frac{(Obs-Exp)^2}{Exp}\approx 12.70,$$

and the P-value ≈ 0.0004.



Since the P-value ≈ 0.0004 is low, we reject the null hypothesis. There is strong evidence of an association between age and birth rate. Younger mothers tend to have higher birth rates.

c) A two-proportion z-test and a chi-square test for independence with 1 degree of freedom are equivalent. $z^2 = (3.563944)^2 = 12.70 = \chi^2$. The P-values are both the same.

41. Family planning.

H_0: Unplanned pregnancies and education level are independent.

H_A: There is an association between unplanned pregnancies and education level.

Counted data condition: The percentages must be converted to counts.
Randomization condition: Assume that these women are representative of all women.
Expected cell frequency condition: The expected counts are all greater than 5.

Under these conditions, the sampling distribution of the test statistic is χ^2 on 1 degree of freedom. We will use a chi-square test for independence.

$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 40.71$, and the P-value is essentially 0.

Since the P-value is essentially 0, we reject the null hypothesis. There is strong evidence of an association between unplanned pregnancies and education level. More educated women tend to have fewer unplanned pregnancies.

42. Old Faithful.

a) There is a moderate, linear, positive association between duration of the previous eruption and interval between eruptions for Old Faithful. Relatively long eruptions appear to be associated with relatively long intervals until the next eruption.

b) H_0: There is no linear relationship between duration of the eruption and interval until the next eruption. $(\beta_1 = 0)$

H_A: There is a linear relationship between duration of the eruption and interval until the next eruption. $(\beta_1 \neq 0)$

c) **Straight enough condition:** The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The histogram of residuals is unimodal and symmetric.

Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with 222 – 2 = 220 degrees of freedom. We will use a regression slope *t*-test. The equation of the line of best fit for these data points is: $\widehat{Interval} = 33.9668 + 10.3582(Duration)$.

d) The value of $t \approx 27.1$. The *P*-value of essentially 0 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between duration and interval. Relatively long eruptions tend to be associated with relatively long intervals until the next eruption.

e) The regression equation predicts that an eruption with duration of 2 minutes will have an interval until the next eruption of $33.9668 + 10.3582(2) = 54.6832$ minutes. ($t^*_{220} \approx 1.9708$)

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n}}$$
$$= 54.6832 \pm (1.9708)\sqrt{0.3822^2 \cdot (2 - 3.57613)^2 + \frac{6.159^2}{222}}$$
$$\approx (53.24,\ 56.12)$$

We are 95% confident that, after a 2-minute eruption, the mean length of time until the next eruption will be between 53.24 and 56.12 minutes.

f) The regression equation predicts that an eruption with duration of 4 minutes will have an interval until the next eruption of $33.9668 + 10.3582(4) = 75.3996$ minutes. ($t^*_{220} \approx 1.9708$)

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}$$
$$= 75.3996 \pm (1.9708)\sqrt{0.3822^2 \cdot (4 - 3.57613)^2 + \frac{6.159^2}{222} + 6.159^2}$$
$$\approx (63.23,\ 87.57)$$

We are 95%confident that the length of time until the next eruption will be between 63.23 and 87.57 minutes, following a 4-minute eruption.

43. Togetherness.

a) H_0: There is no linear relationship number of meals eaten as a family and grades. $(\beta_1 = 0)$

H_A: There is a linear relationship. $(\beta_1 \neq 0)$

Since the conditions for inference are satisfied (given), the sampling distribution of the regression slope can be modeled by a Student's t-model with (142 – 2) = 140 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{GPA} = 2.7288 + 0.1093(Meals\,/\,Week)$.

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$
$$t = \frac{0.1093 - 0}{0.0263}$$
$$t \approx 4.16$$

The value of $t \approx 4.16$. The P-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between grades and the number of meals eaten as a family. Students whose families eat together relatively frequently tend to have higher grades than those whose families don't eat together as frequently.

b) This relationship would not be particularly useful for predicting a student's grade point average. $R^2 = 11.0\%$, which means that only 11% of the variation in GPA can be explained by the number of meals eaten together per week.

c) These conclusions are not contradictory. There is strong evidence that the slope is not zero, and that means strong evidence of a linear relationship. This does not mean that the relationship itself is strong, or useful for predictions.

44. Learning math.

a) H_0: The mean score of Accelerated Math students is the same as the mean score of traditional students. $(\mu_A = \mu_T \text{ or } \mu_A - \mu_T = 0)$

H_A: The mean score of Accelerated Math students is different from the mean score of traditional students. $(\mu_A \neq \mu_T \text{ or } \mu_A - \mu_T \neq 0)$

Independent groups assumption: Scores of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 459.24 degrees of freedom (from the approximation formula).

We will perform a two-sample t-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_A - \bar{y}_T) = \sqrt{\frac{84.29^2}{231} + \frac{74.68^2}{245}} \approx 7.3158$.

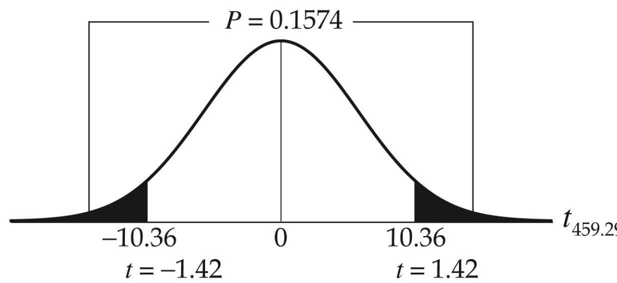
The observed difference between the mean scores is 560.01 – 549.65 = 10.36

Since the P-value = 0.1574, we fail to reject the null hypothesis. There is no evidence that the Accelerated Math students have a different mean score on the pretest than the traditional students.

$$t = \frac{(\bar{y}_A - \bar{y}_T) - (0)}{SE(\bar{y}_A - \bar{y}_T)}$$

$$t \approx \frac{10.36}{7.3158}$$

$$t \approx 1.42$$



b) H_0: Accelerated Math students do not show significant improvement in test scores. The mean individual gain for Accelerated Math is zero. $(\mu_d = 0)$

H_A: Accelerated Math students show significant improvement in test scores. The mean individual gain for Accelerated Math is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The Accelerated Math students had a mean individual gain of $\bar{d}$ = 77.53 points and a standard deviation of 78.01 points. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean individual gain with a Student's t model, with 231 – 1 = 230 degrees of freedom, $t_{230}\left(0, \frac{78.01}{\sqrt{231}}\right)$.

We will perform a paired t-test.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{77.53 - 0}{\frac{78.01}{\sqrt{231}}}$$

$$t \approx 15.11$$

Since the P-value is essentially 0, we reject the null hypothesis. There is strong evidence that the mean individual gain is greater than zero. The Accelerated Math students showed significant improvement.

c) H_0: Students taught using traditional methods do not show significant improvement in test scores. The mean individual gain for traditional methods is zero. $(\mu_d = 0)$

H_A: Students taught using traditional methods show significant improvement in test scores. The mean individual gain for traditional methods is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The students taught using traditional methods had a mean individual gain of $\bar{d}$ = 39.11 points and a standard deviation of 66.25 points. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean individual gain with a Student's t model, with 245 – 1 = 244 degrees of freedom, $t_{244}\left(0, \frac{66.25}{\sqrt{245}}\right)$. We will perform a paired t-test.

Since the P-value is essentially 0, we reject the null hypothesis. There is strong evidence that the mean individual gain is greater than zero. The students taught using traditional methods showed significant improvement.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{39.11 - 0}{\frac{66.25}{\sqrt{245}}}$$

$$t \approx 9.24$$

d) H_0: The mean individual gain of Accelerated Math students is the same as the mean individual gain of traditional students. $(\mu_{dA} = \mu_{dT} \text{ or } \mu_{dA} - \mu_{dT} = 0)$

H_A: The mean individual gain of Accelerated Math students is greater than the mean individual gain of traditional students. $(\mu_{dA} > \mu_{dT} \text{ or } \mu_{dA} - \mu_{dT} > 0)$

Independent groups assumption: Individual gains of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.

Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 452.10 degrees of freedom (from the approximation formula).

We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{d}_A - \bar{d}_T) = \sqrt{\frac{78.01^2}{231} + \frac{66.25^2}{245}} \approx 6.6527$.

The observed difference between the mean scores is 77.53 – 39.11 = 38.42

Since the *P*-value is less than 0.0001, we reject the null hypothesis. There is strong evidence that the Accelerated Math students have an individual gain that is significantly higher than the individual gain of the students taught using traditional methods.

$$t = \frac{(\bar{d}_A - \bar{d}_T) - 0}{SE(\bar{d}_A - \bar{d}_T)}$$
$$t = \frac{38.42 - 0}{6.6527}$$
$$t \approx 5.78$$

45. Juvenile offenders.

a) **Randomization condition:** We will assume that the youths studied are representative of other youths that might receive this therapy.
10% condition: 125 and 125 are less than 10% of all such youths.
Independent samples condition: The groups are independent.
Success/Failure condition: $n\hat{p}$ (Ind) = 19, $n\hat{q}$ (Ind) = 106, $n\hat{p}$ (MST) = 5, and $n\hat{q}$ (MST) = 120 are all not greater than 10, but with only one equal to 5, the test two-proportion interval should be reliable.

Since the conditions have been satisfied, we will find a two-proportion *z*-interval.

$$(\hat{p}_I - \hat{p}_M) \pm z^* \sqrt{\frac{\hat{p}_I \hat{q}_I}{n_I} + \frac{\hat{p}_M \hat{q}_M}{n_M}}$$
$$= \left(\tfrac{19}{125} - \tfrac{5}{125}\right) \pm 2.576 \sqrt{\frac{\left(\frac{19}{125}\right)\left(\frac{106}{125}\right)}{125} + \frac{\left(\frac{5}{125}\right)\left(\frac{120}{125}\right)}{125}} = (0.0178,\ 0.206)$$

We are 99% confident that the percentage of violent felony arrest among juveniles who receive individual therapy is between 1.78 and 20.6 percentage points higher than the percentage of violent felony arrest among juveniles who receive MST.

b) Since the entire interval is above 0, we can conclude that MST is successful in reducing the proportion of juvenile offenders who commit violent felonies. The population of interest is adolescents with mental health problems.

46. Dairy sales.

a) Since the CEO is interested in the association between cottage cheese sales and ice cream sales, the regression analysis is appropriate.

b) There is a moderate, linear, positive association between cottage cheese and ice cream sales. For each additional million pounds of cottage cheese sold, an average of 1.19 million pounds of ice cream are sold.

c) The regression will not help here. A paired *t*-test will tell us whether there is an average difference in sales.

d) There is evidence that the company sells more cottage cheese than ice cream, on average.

e) In part a, we are assuming that the relationship is linear, that errors are independent with constant variation, and that the distribution of errors is Normal.

In part c, we are assuming that the observations are independent and that the distribution of the differences is Normal. This may not be a valid assumption, since the histogram of differences looks bimodal.

f) The equation of the regression line is $\widehat{IceCream} = -26.5306 + 1.19334(CottageCheese)$. In a month in which 82 million pounds of ice cream are sold we expect to sell:

$\widehat{IceCream} = -26.5306 + 1.19334(82) = 71.32$ million pounds of ice cream.

g) Assuming the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (12 – 2) = 10 degrees of freedom. We will use a regression slope *t*-interval, with 95% confidence.

$$b_1 \pm t^*_{n-2} \times SE(b_1) = 1.19334 \pm (2.228) \times 0.4936 \approx (0.09,\ 2.29)$$

h) We are 95% confident that the mean number of pounds of ice cream sold increases by between 0.09 and 2.29 pounds for each additional pound of cottage cheese sold.

47. Infliximab.

H_0: The remission rates are the same for the three groups.

H_A: The remission rates are different for the three groups.

Counted data condition: The data are counts.
Randomization condition: Assume that these patients are representative of all patients.

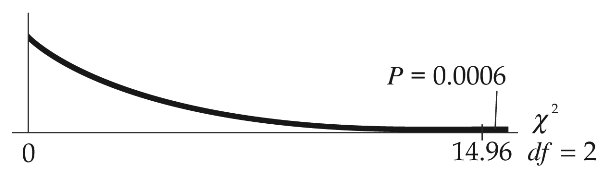
Expected cell frequency condition: The expected counts are all greater than 5.

	Placebo (Obs / Exp)	**5 mg (Obs / Exp)**	**10 mg (Obs / Exp)**
Remission	23 / 38.418	44 / 39.466	50 / 39.116
No Remission	87 / 71.582	69 / 73.534	62 / 72.884

Under these conditions, the sampling distribution of the test statistic is χ^2 on 2 degrees of freedom. We will use a chi-square test for homogeneity.

$$\chi^2 = \sum_{all\,cells} \frac{(Obs - Exp)^2}{Exp} \approx 14.96,$$

and the P-value ≈ 0.0006.



Since the P-value ≈ 0.0006 is low, we reject the null hypothesis. There is strong evidence that the remission rates are different in the three groups. Patients receiving 10 mg of Infliximab have higher remission rates than the other groups. These data indicate that continued treatment with Infliximab is of value to Crohn's disease patients who exhibit a positive initial response to the drug.

48. Education vs. income 2009.

a) **Straight enough condition:** The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot is reasonably straight.

Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (50 – 2) = 48 degrees of freedom. We will use a regression slope t-test. The equation of the line of best fit for these data points is: $\widehat{Income} = 22294.24 + 853(Education)$.

b) The value of $t \approx 5.91$. The P-value of less than 0.0001 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between education level and income. Cities in which the percentage of residents with Bachelor's degrees is relatively high have relatively high median incomes.

c) If the data were plotted for individuals, the association would appear to be weaker. Individuals vary more than averages.

d) $b_1 \pm t^*_{n-2} \times SE(b_1) = 853 \pm (2.0106) \times 144.38 \approx (563,\ 1143)$

We are 95% confident that each additional percentage point increase in the percent of residents with a Bachelor's degree in a city is associated with an increase of between \$563 and \$1143 in median income.

e) The regression equation predicts that a city with a median education level of 11 years of school will have a median income of $22,294.24 + 853(25) = \$43,619.24$
($t^*_{48} \approx 1.677$)

$$\hat{y}_\nu \pm t^*_{n-2}\sqrt{SE^2(b_1)\cdot(x_\nu - \bar{x})^2 + \frac{s_e^2}{n}}$$
$$= 43619.24 \pm (1.677)\sqrt{144.38^2 \cdot (25 - 28.402)^2 + \frac{11881.87^2}{50}}$$
$$\approx (40683,\ 46556)$$

We are 90% confident that cities with 25% of residents with Bachelor's degrees will have an average median household income of between \$40,683 and \$46,556.

49. Diet.

H_0: Cracker type and bloating are independent.

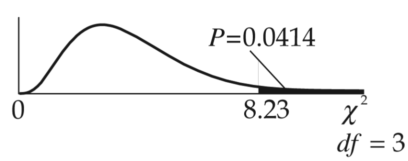
H_A: There is an association between cracker type and bloating.

Counted data condition: The data are counts.
Randomization condition: Assume that these women are representative of all women.
Expected cell frequency condition: The expected counts are all (almost!) greater than 5.

	Bloat	
	Little/None (Obs / Exp)	**Moderate/Severe (Obs / Exp)**
Bran	11 / 7.6471	2 / 5.3529
Gum Fiber	4 / 7.6471	9 / 5.3529
Combination	7 / 7.6471	6 / 5.3529
Control	8 / 7.0588	4 / 4.9412

Under these conditions, the sampling distribution of the test statistic is χ^2 on 3 degrees of freedom. We will use a chi-square test for independence.

$$\chi^2 = \sum_{all\,cells}\frac{(Obs - Exp)^2}{Exp} \approx 8.23, \text{ and the } P\text{-value} \approx 0.0414.$$

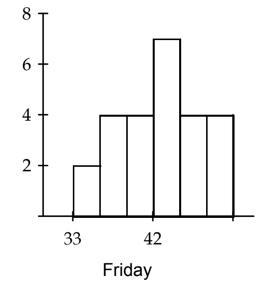


Since the *P*-value is low, we reject the null hypothesis. There is evidence of an association between cracker type and bloating. The gum fiber crackers had a higher rate of moderate/severe bloating than expected. The company should head back to research and development and address the problem before attempting to market the crackers.

50. Cramming.

a) H_0: The mean score of week-long study group students is the same as the mean score of overnight cramming students. $(\mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0)$

H_A: The mean score of week-long study group students is greater than the mean score of overnight cramming students.
$(\mu_1 > \mu_2 \text{ or } \mu_1 - \mu_2 > 0)$



Independent Groups Assumption: Scores of students from different classes should be independent.
Randomization Condition: Assume that the students are assigned to each class in a representative fashion.
Nearly Normal Condition: The histogram of the crammers is unimodal and symmetric. We don't have the actual data for the study group, but the sample size is large enough that it should be safe to proceed.

$$\bar{y}_1 = 43.2 \qquad \bar{y}_2 = 42.28$$
$$s_1 = 3.4 \qquad s_2 = 4.43020$$
$$n_1 = 45 \qquad n_2 = 25$$

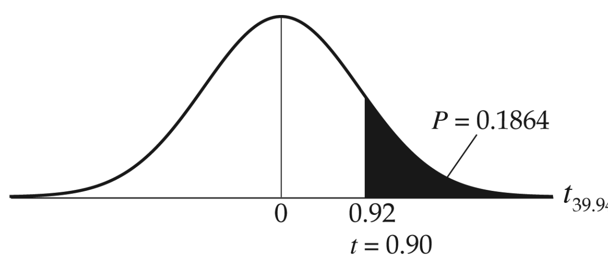
Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 39.94 degrees of freedom (from the approximation formula). We will perform a two-sample t-test. The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{3.4^2}{45} + \frac{4.43020^2}{25}} \approx 1.02076 .$$

The observed difference between the mean scores is 43.2 – 42.28 = 0.92.

Since the P-value = 0.1864 is high, we fail to reject the null hypothesis. There is no evidence that students with a week to study have a higher mean score than students who cram the night before.

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (0)}{SE(\bar{y}_1 - \bar{y}_2)}$$
$$t \approx \frac{0.92}{1.02076}$$
$$t \approx 0.90$$



b) H_0: The proportion of study group students who will pass is the same as the proportion of crammers who will pass. $(p_1 = p_2 \text{ or } p_1 - p_2 = 0)$

H_A : The proportion of study group students who will pass is different from the proportion of crammers who will pass. $(p_1 \neq p_2 \text{ or } p_1 - p_2 \neq 0)$

Random condition: Assume students are assigned to classes in a representative fashion.
10% condition: 45 and 25 are both less than 10% of all students.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n_1\hat{p}_1 = 15$, $n_1\hat{q}_1 = 30$, $n_2\hat{p}_2 = 18$, and $n_2\hat{q}_2 = 7$ are not all greater than 10, since only 7 crammers didn't pass. However, if we check the pooled value, $n_2\hat{p}_{\text{pooled}} = (25)(0.471) = 11.775$. All of the samples are large enough.

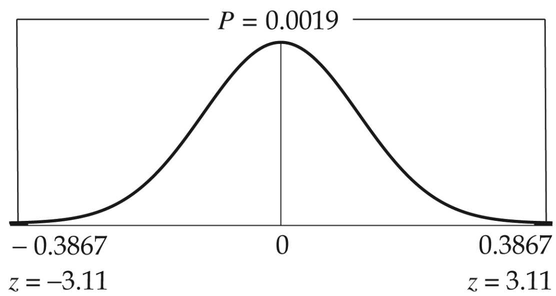
Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}\left(\hat{p}_1-\hat{p}_2\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_1}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_2}}=\sqrt{\frac{\left(\frac{33}{70}\right)\left(\frac{37}{70}\right)}{45}+\frac{\left(\frac{33}{70}\right)\left(\frac{37}{70}\right)}{25}}\approx 0.1245.$$

The observed difference between the proportions is:
$0.3333 - 0.72 = -0.3867$.

Since the P-value = 0.0019 is low, we reject the null hypothesis. There is strong evidence to suggest a difference in the proportion of passing grades for study group participants and overnight crammers. The crammers generally did better.

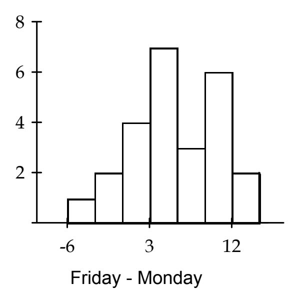
$$z=\frac{-0.3867-0}{0.1245}$$
$$z\approx -3.11$$



c) H_0: There is no mean difference in the scores of students who cram, after 3 days. $(\mu_d = 0)$

H_A: The scores of students who cram decreases, on average, after 3 days. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Assume that students are assigned to classes in a representative fashion.
Nearly Normal condition: The histogram of differences is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 25 – 1 = 24 degrees of freedom, $t_{24}\left(0,\frac{4.8775}{\sqrt{25}}\right)$. We will use a paired t-test, with $\bar{d}=5.04$.

Since the P-value is less than 0.0001, we reject the null hypothesis. There is strong evidence that the mean difference is greater than zero. Students who cram seem to forget a significant amount after 3 days.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{5.04 - 0}{\frac{4.8775}{\sqrt{25}}}$$

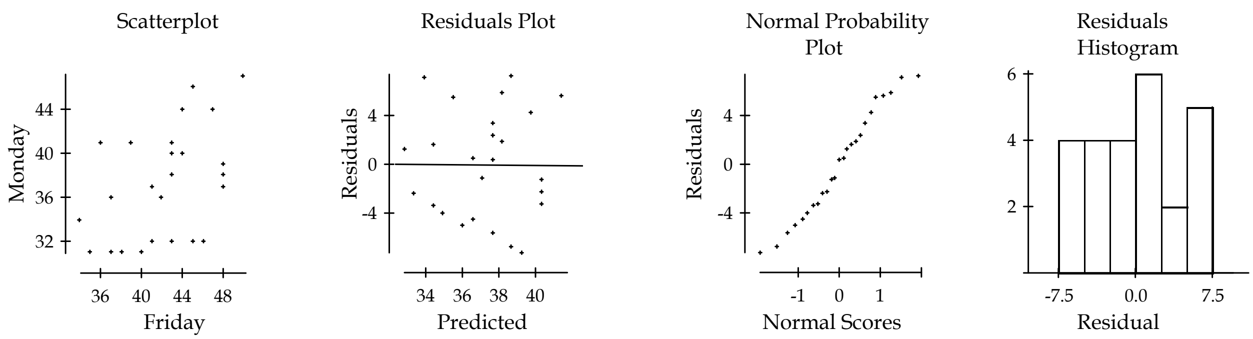
$$t \approx 5.17$$

d) $\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 5.04 \pm t^*_{24}\left(\frac{4.8775}{\sqrt{25}}\right) \approx (3.03, 7.05)$

We are 95% confident that students who cram will forget an average of 3.03 to 7.05 words in 3 days.

e) H_0: There is no linear relationship between Friday score and Monday score. $(\beta_1 = 0)$

H_A: There is a linear relationship between Friday score and Monday score. $(\beta_1 \neq 0)$



Straight enough condition: The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals is reasonably straight, and the histogram of the residuals is roughly unimodal and symmetric.

Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (25 – 2) = 23 degrees of freedom. We will use a regression slope t-test.

Dependent variable is: **Monday**
No Selector
R squared = 22.4% R squared (adjusted) = 19.0%
s = 4.518 with 25 - 2 = 23 degrees of freedom

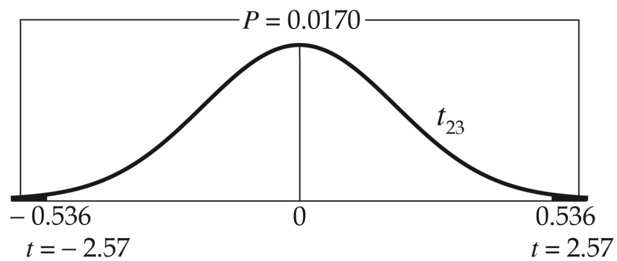
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	135.159	1	135.159	6.62
Residual	469.401	23	20.4087	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	14.5921	8.847	1.65	0.1127
Friday	0.535666	0.2082	2.57	0.0170

The equation of the line of best fit for these data points is:

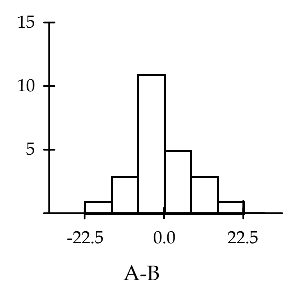
$$\widehat{Monday} = 14.5921 + 0.535666(Friday).$$

The value of $t \approx 2.57$. The P-value of 0.0170 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between Friday score and Monday score. Students who do better in the first place tend to do better after 3 days. However, since R^2 is only 22.4%, Friday score is not a very good predictor of Monday score.



51. Hearing.

Paired data assumption: The data are paired by subject.
Randomization condition: The order of the tapes was randomized.
Normal population assumption: The histogram of differences between List A and List B is roughly unimodal and symmetric.



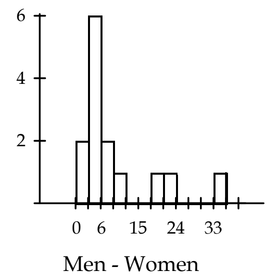
Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 24 – 1 = 23 degrees of freedom. We will find a paired t-interval, with 95% confidence.

$$\bar{d} \pm t_{n-1}^{*}\left(\frac{s_d}{\sqrt{n}}\right) = -0.\bar{3} \pm t_{23}^{*}\left(\frac{8.12225}{\sqrt{24}}\right) \approx (-3.76,\ 3.10)$$

We are 95% confident that the mean difference in the number of words a person might misunderstand using these two lists is between –3.76 and 3.10 words. Since 0 is contained in the interval, there is no evidence to suggest that that the two lists are different for the purposes of the hearing test when there is background noise. It is reasonable to think that the two lists are still equivalent.

52. Newspapers.

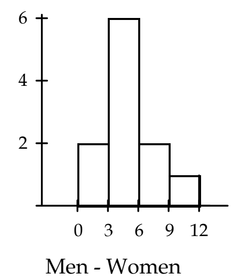
a) An examination of a graphical display reveals Spain, Portugal, and Italy to be outliers. They are all Mediterranean countries, and all have a significantly higher percentage of men than women reading a newspaper daily.



b) H_0: The mean difference in the percentage of men and women who read a daily newspaper in these countries is zero. $(\mu_d = 0)$

H_A: The mean difference in the percentage of men and women who read a daily newspaper in these countries is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by country.
Randomization condition: Samples in each country were random.
Nearly Normal condition: With three outliers removed, the distribution of differences is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 11 – 1 = 10 degrees of freedom, $t_{10}\left(0, \frac{2.83668}{\sqrt{11}}\right)$.

We will use a paired t-test, (Men – Women) with $\bar{d} = 4.75455$.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{11}}}$$

$$t \approx \frac{4.75455 - 0}{\frac{2.83668}{\sqrt{11}}}$$

$$t \approx 5.56$$

Since the P-value = 0.0001 is very low, we reject the null hypothesis. There is strong evidence that the mean difference is greater than zero. The percentage of men in these countries who read the paper daily appears to be greater than the percentage of women who do so.

Chapter 26 – Analysis of Variance

Section 26.1

1. Popcorn

a) H_0: The mean number of unpopped kernels is the same for all four brands of popcorn. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean number of unpopped kernels is not the same for all four brands of popcorn.

b) MS_T has $k - 1 = 4 - 1 = 3$ degrees of freedom.
MS_E has $N - k = 16 - 4 = 12$ degrees of freedom.

c) The *F*-statistic is 13.56 with 3 and 12 degrees of freedom, resulting in a *P*-value of 0.00037. We reject the null hypothesis and conclude that there is strong evidence that the mean number of unpopped kernels is not the same for all four brands of popcorn.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

2. Skating.

a) H_0: The mean score is the same for each focus. $(\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5)$

H_A: The mean scores for each focus are not all the same.

b) MS_T has $k - 1 = 5 - 1 = 4$ degrees of freedom.
MS_E has $N - k = 30 - 5 = 25$ degrees of freedom.

c) The *F*-statistic is 7.43 with 4 and 25 degrees of freedom, resulting in a *P*-value of 0.00044. We reject the null hypothesis and conclude that there is strong evidence that the mean scores for each focus are not all the same.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

3. Gas mileage.

a) H_0: The mean gas mileage is the same for each muffler. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean gas mileages for each muffler are not all the same.

b) MS_T has $k-1=3-1=2$ degrees of freedom.
MS_E has $N-k=24-3=21$ degrees of freedom.

c) The *F*-statistic is 2.35 with 2 and 21 degrees of freedom, resulting in a *P*-value of 0.1199. We fail to reject the null hypothesis and conclude that there is no evidence to suggest that gas mileage associated with any single muffler is different than the others.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

e) By failing to notice that one of the mufflers resulted in significantly different gas mileage, you have committed a Type II error.

4. Darts.

a) H_0: The mean distance from dart to bull's-eye is the same for all four stances. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean distances from dart to bull's-eye are not all the same.

b) MS_T has $k-1=4-1=3$ degrees of freedom.
MS_E has $N-k=40-4=36$ degrees of freedom.

c) The *F*-statistic is 1.41 with 3 and 36 degrees of freedom, resulting in a *P*-value of 0.2557. We fail to reject the null hypothesis and conclude that there is no evidence that the mean distances from dart to target are not all the same.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

e) By failing to notice that one of the stances had a mean distance from dart to bull's-eye significantly different than the others, you have made a Type II error.

Section 26.2

5. Activating baking yeast.

a) H_0: The mean activation time is the same for all four recipes. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean activation times for each recipe are not all the same.

b) The *F*-statistic is 44.7392 with 3 and 12 degrees of freedom, resulting in a *P*-value less than 0.0001. We reject the null hypothesis and conclude that there is strong evidence that the mean activation times for each recipe are not all the same.

c) Yes, it would be appropriate to follow up with multiple comparisons, because we have rejected the null hypothesis.

6. Frisbee throws.

a) H_0: The mean distance is the same for each grip. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean distances for each grip are not all the same.

b) The *F*-statistic is 2.0453 with 2 and 21 degrees of freedom, resulting in a *P*-value equal to 0.1543. We fail to reject the null hypothesis and conclude that there is not enough evidence that the mean distances for each type of grip are not all the same.

c) No, it would not be appropriate to follow up with multiple comparisons, because we have failed to reject the null hypothesis.

Section 26.3

7. Fuel economy by cylinders, revisited.

a) H_0: The mean mileage is the same for each engine type (number of cylinders). $(\mu_4 = \mu_5 = \mu_6 = \mu_8)$

H_A: The mean mileages for each engine type are not all the same.

b) The Similar Variance condition is not met, because the boxplots show distributions with radically different spreads. A re-expression of the response variable may equalize the spreads, allowing us to proceed with an analysis of variance.

8. Finger Lakes wines 2014.

a) H_0: The mean bottle price is the same at all three locations. $(\mu_C = \mu_K = \mu_S)$

H_A: The mean bottle prices for each location are not all the same.

b) The Similar Variance condition is not met, because the boxplots show distributions with radically different spreads, with several outliers in the Seneca wines.

Section 26.4

9. Tellers.

a) H_0: The mean time to serve a customer is the same for each teller. $(\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6)$

H_A: The mean times to serve a customer for each teller are not all the same.

b) The F-statistic is 1.508 with 5 and 134 degrees of freedom, resulting in a P-value equal to 0.1914. We fail to reject the null hypothesis and conclude that there is not enough evidence that the mean times to serve a customer for each teller are not all the same.

c) No, it would not be appropriate to follow up with multiple comparisons, because we have failed to reject the null hypothesis.

10. Hearing.

a) H_0: The mean hearing score is the same for all four lists. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean hearing scores for each list are not all the same.

b) The F-statistic is 4.9192 with 3 and 92 degrees of freedom, resulting in a P-value equal to 0.0033. We reject the null hypothesis and conclude that there is strong evidence that the mean hearing scores for each list are not all the same.

c) Yes, it would be appropriate to follow up with multiple comparisons, because we have rejected the null hypothesis.

Section 26.5.

11. Eye and hair color.

An analysis of variance is not appropriate, because eye color is a categorical variable. The students could consider a chi-square test of independence.

12. Zip codes, revisited.

An analysis of variance is not appropriate, because zip code is a categorical variable. The students could consider categorizing zip codes by the first digit, corresponding to the region of the country (0 on the east coast to 9 on the west coast), and performing a chi-square test of independence.

Chapter Exercises.

13. Yogurt.

a) MS_T has $k-1=3-1=2$ degrees of freedom.

$$MS_T = \frac{SS_T}{df_T} = \frac{17.300}{2} = 8.65$$

MS_E has $N-k=9-3=6$ degrees of freedom.

$$MS_E = \frac{SS_E}{df_E} = \frac{0.4600}{6} \approx 0.0767$$

b) $F\text{-statistic} = \dfrac{MS_T}{MS_E} = \dfrac{8.65}{0.0767} \approx 112.78$

c) With a P-value equal to 0.000017, there is very strong evidence that the mean taste test scores for each method of preparation are not all the same.

d) We have assumed that the experimental runs were performed in random order, that the variances of the treatment groups are equal, and that the errors are Normal.

e) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

f) $s_p = \sqrt{MS_E} = \sqrt{0.0767} \approx 0.277$ points

14. Smokestack scrubbers.

a) MS_T has $k-1 = 4-1 = 3$ degrees of freedom.

$$MS_T = \frac{SS_T}{df_T} = \frac{81.2}{3} \approx 27.067$$

MS_E has $N - \text{k} = 20 - 4 = 16$ degrees of freedom.

$$MS_E = \frac{SS_E}{df_E} = \frac{30.8}{16} = 1.925$$

b) $F\text{-statistic} = \dfrac{MS_T}{MS_E} = \dfrac{27.067}{1.925} \approx 14.0606$

c) With a P-value equal to 0.00000949, there is very strong evidence that the mean particulate emissions for each smokestack scrubber are not all the same.

d) We have assumed that the experimental runs were performed in random order, that the variances of the treatment groups are equal, and that the errors are Normal.

e) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

f) $s_p = \sqrt{MS_E} = \sqrt{1.925} \approx 1.387$ ppb

15. Eggs.

a) H_0: The mean taste test scores are the same for both real and substitute eggs. $(\mu_R = \mu_S)$

H_A: The mean taste test scores are different. $(\mu_R \neq \mu_S)$

b) The F-statistic is 31.0712 with 1 and 6 degrees of freedom, resulting in a P-value equal to 0.0014. We reject the null hypothesis and conclude that there is strong evidence that the mean taste test scores for real and substitute eggs are different. The real eggs have a higher mean taste test score.

c) The Similar Variance assumption is a bit of a problem. The spread of the distribution of taste test scores for real eggs looks greater than the spread of the distribution of taste test scores for substitute eggs. Additionally, we cannot check the Nearly Normal condition because we don't have a residuals plot. Caution should be used in making any conclusions.

d) H_0: The mean taste test score of brownies made with real eggs is the same as the mean taste test score of brownies made with substitute eggs.
$(\mu_R = \mu_S \text{ or } \mu_R - \mu_S = 0)$

H_A: The mean taste test score of brownies made with real eggs is different than the mean taste test score of brownies made with substitute eggs.
$(\mu_R \neq \mu_S \text{ or } \mu_R - \mu_S \neq 0)$

Independent groups assumption: Taste test scores for each type of brownie should be independent.
Randomization condition: Brownies were tasted in random order.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. We will need to assume that the distribution of taste test scores is unimodal. The boxplots show distributions that are at least symmetric.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 6 degrees of freedom ($n_R + n_S - 2 = 6$ for a pooled t-test). We will perform a two-sample t-test.

The pooled sample variance is $s_p^2 = \dfrac{(4-1)(0.651)^2 + (4-1)(0.395)^2}{(4-1)+(4-1)} \approx 0.2899$

The sampling distribution model has mean 0, with standard error:

$$SE_{pooled}(\bar{y}_R - \bar{y}_S) = \sqrt{\frac{0.2899}{4} + \frac{0.2899}{4}} \approx 0.3807 .$$

The observed difference between the mean scores is 6.78 – 4.66 = 2.12.

$$t = \frac{(\bar{y}_R - \bar{y}_T) - (0)}{SE_{pooled}(\bar{y}_R - \bar{y}_T)} \approx \frac{2.12 - 0}{0.3807} \approx 5.5687$$

With a t = 5.5687 and 6 degrees of freedom, the 2-sided P-value is 0.0014. Since the P-value is low, we reject the null hypothesis. There is strong evidence that the mean taste test score for brownies made with real eggs is different from the mean taste test score for brownies made with substitute eggs. In fact, there is evidence that the brownies made with real eggs taste better.

The P-value for the 2-sample pooled t-test, 0.0014, which is the same as the p-value for the analysis of variance test. The F-statistic, 31.0712, was approximately the same as $t^2 = (5.5727)^2 \approx 31.05499$. Also, the pooled estimate of the variance, 0.2899, is approximately equal to MS_E, the mean squared error. This is because an analysis of variance for two samples is equivalent to a 2-sample pooled t-test.

16. Auto noise filters.

a) H_0: The mean noise level is the same for both types of filters. $(\mu_1 = \mu_2)$

H_A: The mean noise levels are different. $(\mu_1 \neq \mu_2)$

b) The F-statistic is 0.7673 with 1 and 33 degrees of freedom, resulting in a P-value equal to 0.3874. We fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that the mean noise levels are different for each type of filter.

c) The Similar Variance assumption seems reasonable. The spreads of the two distributions look similar. We cannot check the Nearly Normal condition because we don't have a residuals plot.

d) H_0: The mean noise level is the same for both types of filters.
$(\mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0)$

H_A: The mean noise level is different for each type of filters.
$(\mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0)$

Independent groups assumption: Noise levels for the different types of filters should be independent.
Randomization condition: Assume that this study was conducted using random allocation of the noise filters.
Nearly Normal condition: This might cause some difficulty. The distribution of the noise levels for the new device is skewed. However, the sample sizes are somewhat large, so the CLT will help to minimize any difficulty.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 33 degrees of freedom ($n_1 + n_2 - 2 = 33$ for a pooled t-test). We will perform a two-sample t-test.

The pooled sample variance is $s_p^2 = \dfrac{(18-1)(3.2166)^2 + (17-1)(2.43708)^2}{(18-1)+(17-1)} \approx 8.2097$

The sampling distribution model has mean 0, with standard error:

$$SE_{pooled}(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{8.2097}{18} + \frac{8.2097}{17}} \approx 0.9690.$$

The observed difference between the mean scores is 81.5556 – 80.7059 = 0.8497.

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (0)}{SE_{pooled}(\bar{y}_1 - \bar{y}_2)} \approx \frac{0.8497 - 0}{0.9690} \approx 0.8769$$

With a t = 0.877 and 33 degrees of freedom, the 2-sided P-value is 0.3874. Since the P-value is high, we fail to reject the null hypothesis. There is no evidence that the mean noise levels for the two filters are different.

The P-value for the 2-sample pooled t-test, 0.3874, which is the same as the P-value for the analysis of variance test. The F-statistic, 0.7673, was approximately the same as $t^2 = (0.8770)^2 \approx 0.7673$. Also, the pooled estimate of the variance, 8.2097, is approximately equal to MS_E, the mean squared error. This is because an analysis of variance for two samples is equivalent to a 2-sample pooled t-test.

17. School system.

a) H_0: The mean math test score is the same at each of the 15 schools.
$\left(\mu_A = \mu_B = \ldots = \mu_O\right)$

H_A: The mean math test scores are not all the same at each of the 15 schools.

b) The F-statistic is 1.0735 with 14 and 105 degrees of freedom, resulting in a P-value equal to 0.3899. We fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that the mean math scores are not all the same for the 15 schools.

c) This does not match our findings in part b. Because the intern performed so many t-tests, he may have committed several Type I errors. (Type I error is the probability of rejecting the null hypothesis when there is actually no difference between the schools.) Some tests would be expected to result in Type I error due to chance alone. The overall Type I error rate is higher for multiple t-tests than for an analysis of variance, or a multiple comparisons method. Because we failed to reject the null hypothesis in the analysis of variance, we should not do multiple comparison tests.

18. Fertilizers.

a) H_0: The mean height of beans is the same for each of the 10 fertilizers. $(\mu_A = \mu_B = \ldots = \mu_J)$

H_A: The mean heights of beans are not all the same for each of the 10 fertilizers.

b) The *F*-statistic is 1.1882 with 9 and 110 degrees of freedom, resulting in a *P*-value equal to 0.3097. We fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that the mean bean heights are not all the same for the 10 fertilizers.

c) This does not match our findings in part b. Because the lab partner performed so many *t*-tests, he may have committed several Type I errors. (Type I error is the probability of rejecting the null hypothesis when there is actually no difference between the fertilizers.) Some tests would be expected to result in Type I error due to chance alone. The overall Type I error rate is higher for multiple *t*-tests than for an analysis of variance, or a multiple comparisons method. Because we failed to reject the null hypothesis in the analysis of variance, we should not do multiple comparison tests.

19. Cereals once more.

a) H_0: The mean sugar content is the same for each of the 3 shelves. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean sugar contents are not all the same for each of the 3 shelves.

b) The *F*-statistic is 7.2721 with 2 and 74 degrees of freedom, resulting in a *P*-value equal to 0.0013. We reject the null hypothesis and conclude that there is strong evidence to suggest that the mean sugar content of cereal on at least one shelf differs from that on other shelves.

c) We cannot conclude that cereals on shelf two have a higher mean sugar content than cereals on shelf three, or that cereals on shelf two have a higher mean sugar content than cereals on shelf one. We can conclude only that the mean sugar contents are not all equal.

d) We can conclude that the mean sugar content on shelf two is significantly different from the mean sugar contents on shelves one and three. In fact, there is evidence that the mean sugar content on shelf two is greater than that of shelf one and shelf three.

20. Cereals redux.

a) H_0: The mean protein content is the same for each of the 3 shelves. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean protein contents are not all the same for each of the 3 shelves.

b) The F-statistic is 5.8445 with 2 and 74 degrees of freedom, resulting in a P-value equal to 0.0044. We reject the null hypothesis and conclude that there is strong evidence to suggest that the mean protein content of cereal on at least one shelf differs from that on other shelves.

c) We cannot conclude that cereals on shelf two have a higher mean protein content than cereals on shelf three, or that cereals on shelf two have a higher mean protein content than cereals on shelf one. We can conclude only that the mean protein contents are not all equal.

d) We can conclude that the mean protein content on shelf two is significantly different from the mean protein contents on shelf three. In fact, there is evidence that the mean protein content on shelf three is greater than that of shelf two. The other pairwise comparisons are not significant at $\alpha = 0.05$.

21. Downloading.

a) H_0: The mean download time is the same for each of the three times of day. $\left(\mu_{Early} = \mu_{Evening} = \mu_{Late}\right)$

H_A: The mean download times are not all the same for each of the three times of day.

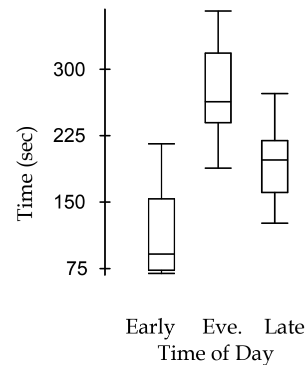
b)

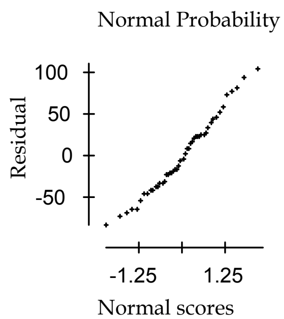
Analysis of Variance

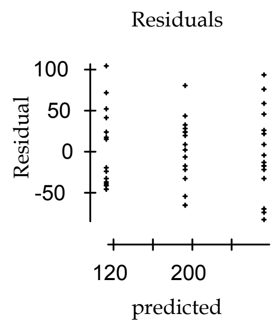
Source	**DF**	**Sum of Squares**	**Mean Square**	**F-ratio**	**P-Value**
Time of Day	2	204641	102320	46.035	<0.0001
Error	45	100020	2222.67		
Total	47	304661			

The F-statistic is 46.035 with 2 and 45 degrees of freedom, resulting in a P-value less than 0.0001. We reject the null hypothesis and conclude that there is strong evidence to suggest that the mean download time is different in at least one of the three times of day.

c) **Randomization condition:** The runs were not randomized, but it is likely that they are representative of all download times at these times of day.
Similar Variance condition: The boxplots show similar spreads for the distributions of download times for the different times of day.
Nearly Normal condition: The normal probability plot of residuals is reasonably straight.







d) A Bonferroni test shows that all three pairs are different from each other at $\alpha = 0.05$.

22. Analgesics.

a) H_0: The mean pain level reported is the same for each of the three drugs.
$(\mu_A = \mu_B = \mu_C)$

H_A: The mean pain levels reported are not all the same for the three drugs.

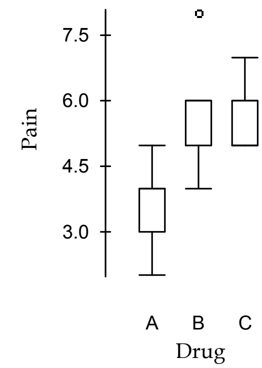
b)

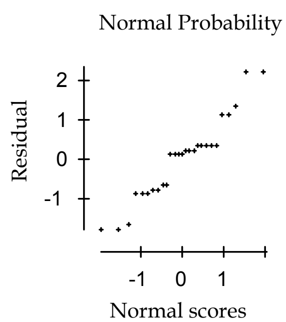
Analysis of Variance

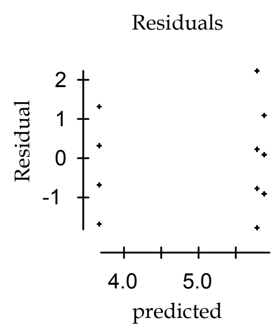
Source	**DF**	**Sum of Squares**	**Mean Square**	**F-ratio**	**P-Value**
Drug	2	28.2222	14.1111	11.9062	0.0003
Error	24	28.4444	1.1852		
Total	26	56.6666			

The F-statistic is 11.9062 with 2 and 24 degrees of freedom, resulting in a P-value equal to 0.0003. We reject the null hypothesis and conclude that there is strong evidence that the mean pain level reported is different in at least one of the three drugs.

c) **Randomization condition:** The volunteers were randomly allocated to treatment groups.
Similar Variance condition: The boxplots show similar spreads for the distributions of pain levels for the different drugs, although the outlier for drug B may cause some concern.
Nearly Normal condition: The normal probability plot of residuals is reasonably straight.







d) A Bonferroni test shows that drug A's mean pain level as reported is significantly below the other two, but that drug B's and C's means are indistinguishable at the $\alpha = 0.05$ level.

Chapter 27 – Multifactor Analysis of Variance

Section 27.1

1. Student TV time.

$$TV_{ijk} = \mu + Sex_j + Athlete_k + \varepsilon_{ijk}$$

2. Cookies.

$$Score_{ijk} = \mu + Chips_j + Sugar_k + \varepsilon_{ijk}$$

Section 27.1

3. Student TV assumptions.

The boxplots show different distributions for Women and Men, so the additivity assumption is probably not satisfied. The plots show that, at least for men, the equal variance assumption seems unlikely. And both plots show outliers, making the normality assumption questionable. If the survey was randomized, the responses may be mutually independent, but we don't have enough information to tell.

4. Cookie assumptions.

The boxplots show some differences, which may indicate a failure of additivity, but the differences are not very great. The ratings of the cookies should be independent. There is some difference in the variabilities, but it is not great. There are a few outliers, which may challenge the normality assumption.

Section 27.3

5. Student TV interactions.

The interaction term reflects the fact that there is very little difference in TV watching among women between athletes and non-athletes, but there is a difference for men with athletes watching more TV.

6. Have ANOVA cookie.

Milk chocolate chips combined with the middle amount of sugar made a low-scoring cookie. The other recipe combinations were additive.

Chapter Exercises.

7. Popcorn revisited.

a) H_0: The effect due to power level is the same for each level. $(\gamma_{Low} = \gamma_{Med} = \gamma_{High})$

H_A: Not all of the power levels have the same effect on popcorn popping.

H_0: The effect due to popping time is the same for each time. $(\tau_3 = \tau_4 = \tau_5)$

H_A: Not all of the popping times have the same effect on popcorn popping.

b) The *power* sum of squares has 3 – 1 = 2 degrees of freedom.
The *popping time* sum of squares has 3 – 1 = 2 degrees of freedom.
The error sum of squares has (9 – 1) – 2 – 2 = 4 degrees of freedom.

c) Because the experiment did not include replication, there are no degrees of freedom left for the interaction term. The interaction term would require 2(2) = 4 degrees of freedom, leaving none for the error term, making any tests impossible.

8. Gas mileage revisited.

a) H_0: The effect on gas mileage due to tire pressure is the same for each level. $(\gamma_{Low} = \gamma_{Med} = \gamma_{Full})$

H_A: Not all of the tire pressure levels have the same effect on gas mileage.

H_0: The effect on gas mileage due to acceleration is the same for each level. $(\tau_S = \tau_P)$

H_A: The two acceleration levels have different effects.

b) The *tire pressure* sum of squares has 3 – 1 = 2 degrees of freedom.
The *acceleration* sum of squares has 2 – 1 = 1 degrees of freedom.
The error sum of squares has (24 – 1) – 2 – 1 = 20 degrees of freedom.

c) Yes, he should consider fitting an interaction term. Because this is a replicated experiment, there are enough degrees of freedom left for an interaction term. The effect of tire pressure may change, depending on the level of acceleration.

d) Interaction degrees of freedom equals (3 – 1) (2 – 1) = 2.

9. Popcorn again.

a) The *F*-statistic for *power* is 13.56 with 2 and 4 degrees of freedom, resulting in a *P*-value equal to 0.0165. The *F*-statistic for *time* is 9.36 with 2 and 4 degrees of freedom, resulting in a *P*-value equal to 0.0310.

b) With a *P*-value equal to 0.0165 we reject the null hypothesis that *power* has no effect and conclude that the mean number of uncooked kernels is not equal across all 3 *power* levels. With a *P*-value equal to 0.0310 we reject the null hypothesis that *time* has no effect and conclude that the mean number of uncooked kernels is not equal across all 3 *time* levels.

c) **Randomization condition:** The bags should be randomly assigned to treatments.
Similar Variance condition: The side-by-side boxplots should have similar spreads. The residuals plot should show no pattern, and no change in spread.
Nearly Normal condition: The Normal probability plot of the residuals should be straight, the histogram of the residuals should be unimodal and symmetric.

10. Gas mileage again.

a) With a *P*-value equal to 0.241 there is no interaction effect of *tire pressure* with *acceleration*. The *F*-statistic for *tire pressure* is 4.29 with 2 and 20 degrees of freedom, resulting in a *P*-value equal to 0.030, we reject the null hypothesis and conclude that *tire pressure* has a significant effect on gas mileage. The *F*-statistic for *acceleration* is 2.35 with 1 and 20 degrees of freedom, resulting in a *P*-value equal to 0.143, we fail to reject the null hypothesis, there is not enough evidence to conclude that acceleration has any effect on gas mileage.

b) **Randomization condition:** The trials are randomized.
Similar Variance condition: The side-by-side boxplots should have similar spreads. The residuals plot should show no pattern, and no systematic change in spread.
Nearly Normal condition: The Normal probability plot of the residuals should be straight, the histogram of the residuals should be unimodal and symmetric.

c) You have made a Type II error.

11. Crash analysis.

a) H_0: The effect on head injury severity is the same for both seats. $(\gamma_D = \gamma_P)$

H_A: The effects on head injury severity are different for the two seats.

H_0: The size of the vehicle has no effect on head injury severity.
$(\tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = \tau_6)$

H_A: The size of the vehicle does have an effect on head injury severity.

b) **Randomization condition:** Assume that the cars are representative of all cars.
Additive Enough condition: The interaction plot is reasonably parallel.
Similar Variance condition: The side-by-side boxplots have similar spreads. The residuals plot shows no pattern, and no systematic change in spread.
Nearly Normal condition: The Normal probability plot of the residuals should be straight, the histogram of the residuals should be unimodal and symmetric.

c) With a *P*-value equal to 0.838 the interaction term between *seat* and *vehicle size* is not significant. With a *P*-value less than 0.0001 for both *seat* and *car size* we reject both null hypotheses and conclude that both *seat* and *car size* affect the severity of head injury. By looking at one of the partial boxplots we see that the mean head injury severity is higher for the driver's side. The effect of driver's *seat* seems to be roughly the same for all six cars.

12. Sprouts.

a) H_0: The temperature level has no effect on biomass. $(\gamma_{32} = \gamma_{34} = \gamma_{36})$

H_A: The effects of the temperature levels are not all the same.

H_0: The effect on biomass is the same for all salinity levels. $(\tau_0 = \tau_4 = \tau_8 = \tau_{12})$

H_A: The effects of salinity levels are not all the same.

b) **Randomization condition:** Assume that the mung beans are representative of all mung beans.
Additive Enough condition: The interaction plot is reasonably parallel.
Similar Variance condition: The side-by-side boxplots have similar spreads. The residuals plot shows no pattern, and no systematic change in spread.
Nearly Normal condition: The Normal probability plot of the residuals should be straight, the histogram of the residuals should be unimodal and symmetric.

c) With a *P*-value of 0.3244 there is no evidence of an interaction between *temperature* and *salinity* on biomass. The *P*-values for *temperature* and *salinity* are both less than 0.0001. We reject the null hypotheses and conclude that the mean biomass is different for different *temperature* levels and different *salinity* levels. We conclude that both *salinity* and *temperature* affect the amount of mung bean sprout biomass produced.

13. Baldness and Heart Disease.

a) A two-factor ANOVA must have a quantitative response variable. Here the response is whether they exhibited baldness or not, which is a categorical variable. A two-factor ANOVA is not appropriate.

b) We could use a chi-square analysis to test whether baldness and heart disease are independent.

14. Fish and prostate.

a) A two-factor ANOVA must have a quantitative response variable. Here the response is whether the subjects suffered prostate cancer or not, a categorical variable. A two-factor ANOVA is not appropriate.

b) We could use a chi-square analysis to test whether amount of fish in the diet and incidence of prostate cancer are independent.

15. Baldness and heart disease again.

a) A chi-square test of independence gives a chi-square statistic of 14.510 with a *P*-value equal to 0.0023. We reject the hypothesis that baldness and heart disease are independent.

b) No, the fact that these are not independent does not mean that one causes the other. There could be a lurking variable (such as age) that influences both.

16. Fish and prostate.

A chi-square test of independence finds a chi-square statistic of 3.677 with 3 degrees of freedom and a *P*-value equal to 0.2985. We fail to reject the null hypothesis that *fish* consumption and incidence of *prostate* cancer are independent. The data provide no evidence that these are associated.

17. Basketball shots.

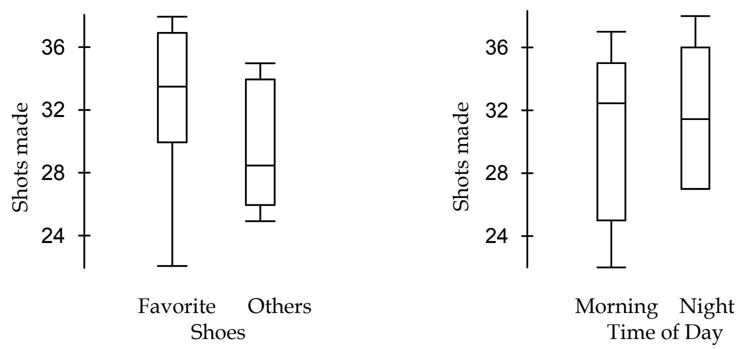
a) H_0: The time of day has no effect on the number of shots made. $(\gamma_M = \gamma_N)$

H_A: The time of day does have an effect on the number of shots made.

H_0: The type of shoe has no effect on the number of shots made. $(\tau_F = \tau_O)$

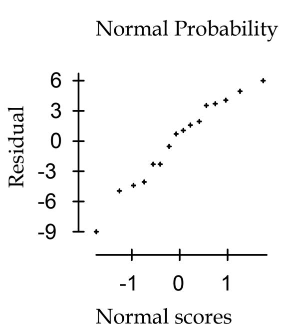
H_A: The type of shoe does have an effect on the number of shots made.

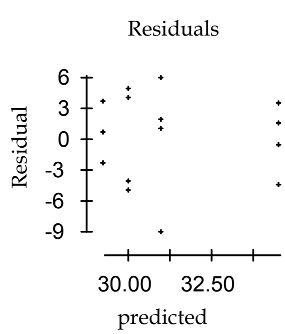
b)

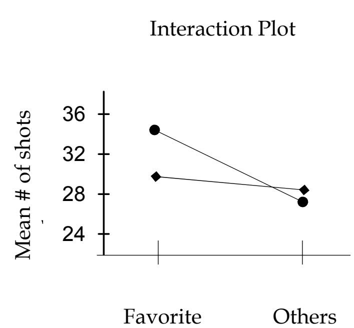


The partial boxplots show little effect of either *time of day* or *shoes* on the number of shots made.

Randomization condition: We assume that the number of shots made were independent from one treatment condition to the next.
Similar Variance condition: The side-by-side boxplots have similar spreads. The residuals plot shows no pattern, and no systematic change in spread.
Nearly Normal condition: The Normal probability plot of the residuals is straight.



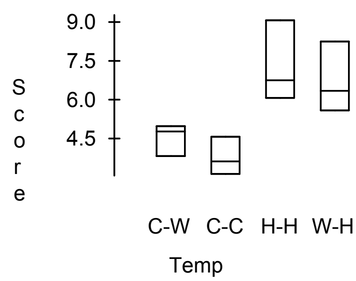


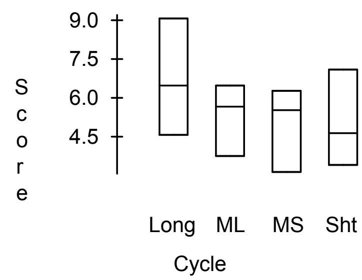


The interaction plot shows a possible interaction effect. It looks as though the favorite shoes may make more of a difference at night. However, we fail to reject the null hypothesis that there is no interaction effect. In fact, none of the effects appears to be significant. It looks as though she cannot conclude that either *shoes* or *time of day* affect her mean free throw percentage.

18. Washing.

We will test the effects of *temperature* settings and *cycle* lengths on washing quality. The partial boxplots show that both factors may affect washing quality.





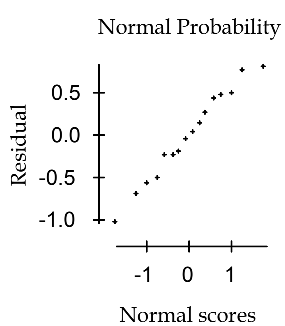
H$_0$: The temperature has no effect on the washing quality. $(\gamma_{CC} = \gamma_{CW} = \gamma_{WH} = \gamma_{HH})$

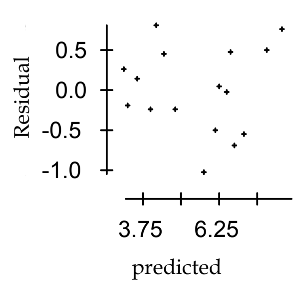
H$_A$: The temperature does have an effect on the washing quality.

H$_0$: The cycle has no effect on the washing quality. $(\tau_S = \tau_{MS} = \tau_{ML} = \tau_L)$

H$_A$: The cycle does have an effect on the washing quality.

Randomization condition: We assume that he randomly assigned handkerchiefs to treatments.
Similar Variance condition: The side-by-side boxplots have relatively similar spreads. The residuals plot does not show enough of a pattern to cause concern.
Nearly Normal condition: The Normal probability plot of the residuals is reasonably straight.





We assume that the interaction effects are negligible. The experiment was unreplicated, so no degrees of freedom are available to estimate them.

The ANOVA confirms what we saw in the boxplots: both *temperature* and *cycle* have significant effects on washing quality. With such small *P*-values we reject both null hypotheses and conclude that both the *temperature* settings and *cycle* lengths affect the cleanliness score.

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F-ratio**	**P-Value**
Temp	3	33.2519	11.084	23.468	0.0001
Cycle	3	7.19688	2.39896	5.0794	0.0250
Error	9	4.25062	0.472292		
Total	15	44.6994			

19. Sprouts again.

a) H_0: The temperature level has no effect on the number of sprouts. $(\gamma_{32} = \gamma_{34} = \gamma_{36})$

H_A: The temperature level does have an effect on the number of sprouts.

H_0: The salinity level has no effect on the number of sprouts. $(\tau_0 = \tau_4 = \tau_8 = \tau_{12})$

H_A: The salinity level does have an effect on the number of sprouts.

b) **Randomization condition:** We assume that the sprouts were representative of all sprouts.
Similar Variance condition: There appears to be more spread in the number of sprouts for the lower salinity levels. This is cause for some concern, but most likely does not affect the conclusions.
Nearly Normal condition: The Normal probability plot of the residuals should be straight.

The partial boxplots show that *salinity* level appears to have an effect on the number of bean sprouts while *temperature* does not. The ANOVA supports this. With a p-value less than 0.0001 for *salinity*, there is strong evidence that the salinity level affects the number of bean sprouts. However, it appears that neither the interaction term nor *temperature* have a significant effect on the number of bean sprouts. The interaction term has a *P*-value equal to 0.7549 and *temperature* has a *P*-value equal to 0.3779.

20. Containers revisited.

a) H_0: The type of liquid has no effect on the change in temperature. $(\gamma_W = \gamma_C)$

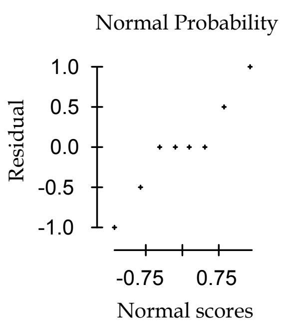
H_A: The type of liquid does have an effect on the change in temperature.

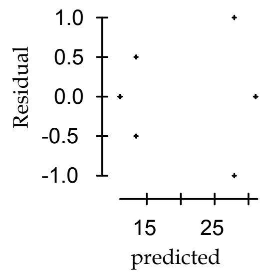
H_0: The environment has no effect on the change in temperature. $(\tau_O = \tau_R)$

H_A: The environment does have an effect on the change in temperature.

b) Both *liquid* and *environment* are significant with *P*-values of 0.0079 and less than 0.0001, respectively. The interaction of *liquid* and *environment* is not significant.

c) The Normal probability plot is reasonably straight and the residuals plot shows no definite pattern. There appear to be no violations of the Nearly Normal condition nor the Similar Variance condition.





d) Both the type of *liquid* and the *environment* seem to affect how well the containers maintain heat. On average the changes is about 2.75 degrees less for coffee than water. On average, the change is about 17.25 degrees less at room temperature than outside.

21. Gas additives.

H_0: The car type has no effect on gas mileage. $(\gamma_H = \gamma_M = \gamma_S)$

H_A: The car type does have an effect on gas mileage.

H_0: The mean gas mileage is the same for both additives. $(\tau_G = \tau_R)$

H_A: The mean gas mileage is different for the additives.

The ANOVA table shows that both *car type* and *additive* affect gas mileage, with *P*-values less than 0.0001. There is a significant interaction effect as well that makes interpretation of the main effects problematic. However, the residual plot shows a strong increase in variance, which makes the whole analysis suspect. The Similar Variance condition appears to be violated.

22. Chromatography

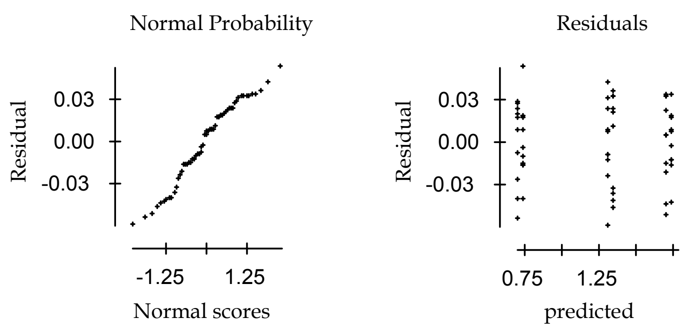
H_0: The mean counts is the same for both flow rates. $(\gamma_F = \gamma_S)$
H_A: The mean counts are not the same for the flow rates.

H_0: The mean counts is the same for each concentration. $(\tau_L = \tau_M = \tau_H)$
H_A: The mean counts are not all the same for the concentrations.

The ANOVA table shows that both *concentration* and *flow rate* affect counts, with *P*-values less than 0.0001. There is a significant interaction effect as well that makes interpretation of the main effects problematic. However, the residual plot shows a strong increase in variance, which makes the whole analysis suspect. The Similar Variance condition appears to be violated.

23. Gas additives again.

After the re-expression of the response, gas mileage, the Normal probability plot of residuals looks straight and the residuals plot shows constant spread over the predicted values.



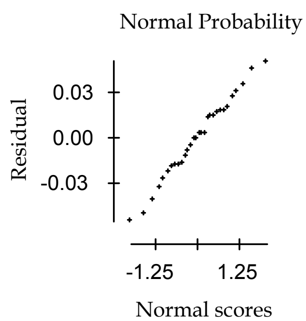
Analysis of Variance

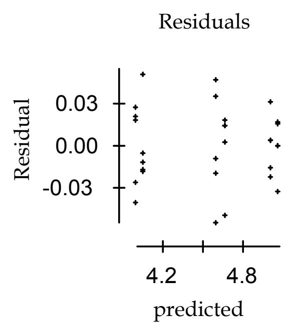
Source	DF	Sum of Squares	Mean Square	F-ratio	P-Value
Type	2	10.1254	5.06268	5923.1	<0.0001
Additive	1	0.026092	0.0260915	30.526	<0.0001
Interaction	2	7.57E-05	3.78E-05	0.044265	0.9567
Error	54	0.046156	8.55E-04		
Total	59	10.1977			

After the re-expression, the ANOVA table only shows the main effects to be significant, while the interaction term is not. We can conclude that both the *car type* and *additive* have an effect on mileage and that the effects are constant (in log(*mpg*)) over the values of the various levels of the other factor.

24. Chromatography again.

After the re-expression of the response, total counts, the Normal probability plot of residuals looks straight and the residuals plot shows constant spread over the





predicted values.

After the re-expression, the ANOVA table only shows the main effects to be significant, while the interaction term is not. We can conclude that both the *concentration* and *flow rate* have an effect on counts and that the effects are constant (in log(*total counts*)) over the values of the various levels of the other factors.

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F-ratio**	**P-Value**
Concentration	2	5.14171	2.57085	2992	<0.0001
Flow rate	1	0.022372	0.0223722	26.038	<0.0001
Interaction	2	2.88E-04	1.44E-04	0.16736	0.8469
Error	24	0.020622	8.59E-04		
Total	29	5.18499			

25. Batteries again.

a) H_0: The mean times are the same under both environments. $(\gamma_C = \gamma_{RT})$

H_A: The mean times are different under the two environments.

H_0: The brand of battery has no effect on time. $(\tau_A = \tau_B = \tau_C = \tau_D)$

H_A: The brand of batter does have an effect on time.

b) From the partial boxplots it appears that *environment* does have an effect on time, but it is unclear whether *brand* has an effect on time.

c) Yes, the ANOVA does match our intuition based on the boxplots. The *brand* effect has a *P*-value equal to 0.099. While this is not significant at the α = 0.05 level, it is significant at the α = 0.10 level. As it appeared on the boxplot, the *environment* is clearly significant with a *P*-value less than 0.0001.

d) There is also an interaction, however, which makes the statement about *brands* problematic. Not all *brands* are affected by the environment in the same way. Brand C, which works best in the cold, performs worst at room temperature.

e) I would be uncomfortable recommending brand C because it performs the worst of the four at room temperature.

26. Peas.

Varying levels of *Quickgrow* affect the growth of sweet peas as measured by weight (*P*-value = 0.0013). The levels of *water* used did not change the weight enough to be statistically significant (*P*-value = 0.095). Using *little* amounts of *Quickgrow* increased weight by an average 369 mg over using *moderate* or *full* amounts. Using *full* amounts of water increased weight an average of almost 200 mg over *little* water, but this increase was not statistically significant.

27. Batteries once more.

In this one-way ANOVA, we can see that the means vary across treatments. (However, boxplots with only 2 observations are not appropriate.) By looking closely, it seems obvious that the four flashlights at room temperature lasted much longer than the ones in the cold. It is much harder to see whether the means of the four brands are different, or whether they differ by the same amounts across both environmental conditions. The two-way ANOVA with interaction makes these distinctions clear.

28. Containers one more time.

With only two observations in each group, a boxplot is not appropriate. There are clearly differences (as shown in the ANOVA table below), but it is difficult to assess how much change is due to changes in *environment* and how much is due to the *liquid*. A two-factor ANOVA makes this much clearer by making it possible to separate and better interpret the effects of brand and environment.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F-ratio	P-Value
Treatment	3	610.375	203.4583	325.533	<0.0001
Error	4	2.5	0.625		
Total	7	612.575			

Chapter 28 – Multiple Regression

Section 28.1

1. House prices.

a)

$$\begin{aligned}\widehat{Price} &= 20{,}986.09 - 7483.10\,Bedrooms + 93.84\,LivingArea \\ &= 20{,}986.09 - 7483.10(2) + 93.84(1000) \\ &= 99{,}859.89\end{aligned}$$

According to the multiple linear regression model, we would expect an Upstate New York home with 2 bedrooms with 1000 square feet of living space to have a price of approximately $99,859.89.

b) The residual is $135,000 – $99,859.89 = $35,140.11.

c) The house sold for about $35,100 more than our estimate.

2. Candy sales.

a)

$$\begin{aligned}\widehat{Calories} &= 28.4 + 11.37\,Fat(g) + 2.91\,Sugar(g) \\ &= 28.4 + 11.37(15) + 2.91(20) \\ &= 257.15\end{aligned}$$

According to the multiple linear regression model, we would expect a chocolate bar with 15g of fat and 20g of sugar to have approximately 257.15 calories.

b) The residual is 227 – 257.15 = –30.15 calories.

c) Her candy has about 30 fewer calories than expected from the model.

Section 28.2

3. Movie profits.

a) $\widehat{USGross} = -22.9898 + 1.12442\,Budget + 24.9724\,Stars - 0.403296\,RunTime$

b) After allowing for the effects of *RunTime* and *Stars*, each million dollars spent making a film yields about 1.13 million dollars in gross revenue.

4. Movie profits again.

The manager is incorrectly interpreting the coefficient causally. The model says that longer films had smaller gross incomes (after allowing for *Budget* and *Stars*), but it doesn't say that making a movie shorter will increase its gross. In fact, cutting arbitrarily would, for example, probably reduce the *Star* rating.

Section 28.3

5. More movies.

a) Linearity: The plot is reasonably linear with no bends.

b) Equal spread: The plot fails the Equal Spread condition. It is much more spread out to the right than on the left.

c) Normality: A scatterplot of two of the variables doesn't tell us anything about the distribution of the residuals.

6. More residuals.

a) Linearity: A histogram doesn't show whether a relationship is linear or not.

b) Nearly Normal: The histogram is unimodal and slightly skewed to the right. But it does seem to have an outlier on the right, which would violate the Nearly Normal condition

c) Equal Spread: A histogram doesn't show whether there is equal spread in different parts of the data; we need a scatterplot for that.

Section 28.4

7. Movie tests.

a) $H_0 : \beta_{Stars} = 0$

b) $t = 4.24$

c) $P < 0.0001$

d) Since the *P*-value is low, reject the null hypothesis. There is strong evidence that the coefficient of *Stars* is not zero.

8. More movie tests.

a) $H_0 : \beta_{RunTime} = 0$

b) $t = -1.60$

c) The coefficient is negative, so the *t*-statistic is negative as well.

d) $P = 0.1113$

e) Since the *P*-value is high, fail to reject the null hypothesis at $\alpha = 0.05$. There is no evidence that the coefficient of *RunTime* is significantly different from zero.

Section 28.5

9. Movie R-squared.

a) $R^2 = 0.474$ or 47.4%. About 47.4% of the variation *USGross* is accounted for by the least squares regression on *Budget*, *RunTime*, and *Stars.*

b) Adjusted R^2 accounts for the number of predictors, so it differs from R^2, which does not.

10. Movie output.

a) $R^2 = \frac{SSR}{SST} = \frac{224995}{(224995 + 249799)}$

b) $F = 34.8$

c) $H_0 : \beta_{Budget} = \beta_{RunTime} = \beta_{Stars} = 0$

d) Yes. Since the *P*-value is very small, we should reject the null hypothesis. There is strong evidence that at least one of the coefficients is significantly different from zero.

Chapter Exercises.

11. Interpretations.

a) There are two problems with this interpretation. First, the other predictors are not mentioned. Secondly, the prediction should be stated in terms of a mean, not a precise value.

b) This is a correct interpretation.

c) This interpretation attempts to predict in the wrong direction. This model cannot predict *lotsize* from *price.*

d) R^2 concerns the fraction of variability accounted for by the regression model, not the fraction of data values.

12. More interpretations.

a) Regression predicts mean values, not a precise value.

b) *Sales* increases should be stated in terms of a mean, not a precise value. In addition, we cannot assume causality. The word "makes" could be replaced by the phrase "is associated with".

c) Changes in one predictor variable are not necessarily associated with changes in another predictor variable.

d) This is a correct interpretation.

13. Predicting final exams.

a) $\widehat{Final} = -6.7210 + 0.2560(Test1) + 0.3912(Test2) + 0.9015(Test3)$

b) R^2 = 77.7%, which means that 77.7% of the variation in *Final* grade is accounted for by the multiple regression model.

c) According to the multiple regression model, each additional point on *Test3* is associated with an average increase of 0.9015 points on the final, for students with given *Test1* and *Test2* scores.

d) Test scores are probably collinear. If we are only concerned about predicting the final exam score, *Test1* may not add much to the regression. However, we would expect it to be associated with the final exam score.

14. Scottish hill races 2008.

a) $\widehat{Time} = -10.3723 + 0.034227(Climb) + 4.04204(Distance)$. Men's record *Time* is associated with both *Distance* and *Climb.* An increase of one kilometer in *Distance* is associated with an average increase of 4.04 minutes in the men's record *Time* for races with a given *Climb*. An increase in one meter of *Climb* is associated with an average increase of 0.0342 minutes in the men's record *Time* for races with a given *Distance*.

b) R^2 = 98%, which means that 98% of the variation in men's record *Time* is accounted for by the multiple regression model.

c) According to the multiple regression model, an increase in one meter of *Climb* is associated with an average increase of 0.0342 minutes in the men's record *Time* for races with a given distance.

15. Home prices.

a) $\widehat{price} = -152037 + 9530(baths) + 139.87(sqft)$

b) R^2 = 71.1%, which means that 71.1% of the variation in asking *price* is accounted for by the multiple regression model.

c) According to the multiple regression model, the asking *price* increases, on average, by about \$139.87 for each additional square foot, for homes with the same number of bathrooms.

d) The number of bathrooms is probably correlated with the size of the house, even after considering the square footage of the bathroom itself. This correlation may account for the coefficient of *baths* not being discernibly different from 0. Moreover, the regression model does not predict what will happen when a house is modified, for example, by converting existing space into a bathroom.

16. More hill races 2008.

a) The two models are similar. In both models, additional *Distance* and additional *Climb* are associated with increases in average race men's record *Time*.

b) The residuals appear to fan out as the predicted value increases. This violates the "does the plot thicken?" condition.

17. Predicting finals II.

Straight enough condition: The plot of residuals versus predicted values looks curved, rising in the middle, and falling on both ends. This is a potential difficulty.
Randomization condition: It is reasonable to think of this class as a representative sample of all classes.
Nearly Normal condition: The Normal probability plot and the histogram of residuals suggest that the highest five residuals are extraordinarily high.
Does the plot thicken? condition: The spread is not consistent over the range of predicted values.

These data may benefit from a re-expression.

18. Home prices II.

Straight enough condition: The plot of residuals versus predicted values looks curved. Multiple regression is not appropriate.
Randomization condition: A random sample of homes for sale in the area was chosen from the Internet.
Nearly Normal condition: The Normal probability plot is not straight, and the histogram of residuals is skewed to the right. Multiple regression is not appropriate.
Does the plot thicken? condition: There may be a slight increase in variability as the fitted values increase.

19. Admin performance.

a) $$\widehat{salary} = 9.788 + 0.11(service) + 0.053(education) + 0.071(score) + 0.004(speed) + 0.065(dictation)$$

b) $\widehat{salary} = 9.788 + 0.11(120) + 0.053(9) + 0.071(50) + 0.004(60) + 0.065(30) \approx \$29,200$

c) H_0: When including the other potential predictors, typing speed does not increase our ability to predict salary. $(\beta_4 = 0)$

H_A: When including the other potential predictors, typing speed increases our ability to predict salary. $(\beta_4 \neq 0)$

With $t = 0.013$, and $30 - 6 = 24$ degrees of freedom, the *P*-value equals 0.9897. Since the *P*-value is so large, we fail to reject the null hypothesis. There is no evidence to suggest that the coefficient for typing *speed* is anything other than 0.

d) Omitting typing *speed* would simplify the model, and probably result in a model that was almost as good. Other predictors might also be omitted, but we cannot make that decision from the information given.

e) *Age* may be collinear with other predictors in the model. In particular, it is likely to be highly associated with months of *service*.

20. GPA and SATs.

a) $\widehat{GPA} = 0.574968 + 0.001394(SATV) + 0.001978(SATM)$

b) $\widehat{GPA} = 0.574968 + 0.001394(500) + 0.001978(550) \approx 2.36$

c) **Straight enough condition:** The scatterplots of the response versus each predicted value should be reasonably straight.
Randomization condition: Hopefully, this section of statistics is representative of all sections, and can be considered a random sample. Otherwise, the regression residuals plots should show no pattern.
Nearly Normal condition: The Normal probability plot should be straight, and the histogram of residuals should be unimodal and symmetric.
Does the plot thicken? condition: The spread of the residuals plots should be constant.

21. Body fat revisited.

a) H_0: There is no linear relationship between weight and percent body fat. $(\beta_W = 0)$

H_A: There is a linear relationship between weight and percent body fat. $(\beta_W \neq 0)$

With t = 12.4, and 250 – 2 = 248 degrees of freedom, the *P*-value is less than 0.0001. Since the *P*-value is so small, we reject the null hypothesis. There is strong evidence of a linear relationship between *weight* and *%body fat*.

b) According to the linear model, each pound of *weight* is associated with a 0.189% increase in *%body fat*.

c) After removing the linear effects of *waist* and *height*, each pound of *weight* is associated, on average, with a decrease of 0.10% in *%body fat*. The change in coefficient and sign is a result of including the other predictors. We expect *weight* to be correlated with both *waist* and *height*. It may be collinear with them.

d) The *P*-value of 0.1567 says that if the coefficient of *height* in this model is truly 0, we could expect to observe a sample regression coefficient at least as far from 0 as the one we have here about 15.7% of the time.

22. Breakfast cereals.

a)

$$\widehat{calories} = -0.879994 + 3.60495(protein) + 8.56877(fat)$$
$$0.309180(fiber) + 4.13996(carbo) + 4.00677(sugars)$$

b) R^2 = 93.6%, which means that 93.6% of the variation in *calories* is accounted for by the multiple regression model. Since we are told that the assumptions are met, the model should do a good job of predicting calories.

c) **Straight enough condition:** The scatterplots of the response versus each predicted value should be reasonably straight.
Randomization condition: Hopefully, these cereals are representative of all cereals, and can be considered a random sample. Otherwise, the regression residuals plots should show no pattern.
Nearly Normal condition: The Normal probability plot should be straight, and the histogram of residuals should be unimodal and symmetric.
Does the plot thicken? condition: The spread of the residuals plots should be constant.

d) After allowing for the linear effects of *protein, fiber, carbo,* and *sugars*, the multiple regression model predicts that each gram of *fat* is, on average, associated with an increase of 8.56877 *calories.*

23. Body fat one more time.

a) H_0: There is no linear relationship between chest size and percent body fat.
$(\beta_C = 0)$

H_A: There is a linear relationship between chest size and percent body fat.
$(\beta_C \neq 0)$

With t = 15.5, and 250 – 2 = 248 degrees of freedom, the p-value is less than 0.0001. Since the p-value is so small, we reject the null hypothesis. There is strong evidence of a linear relationship between *chest* and *%body fat*.

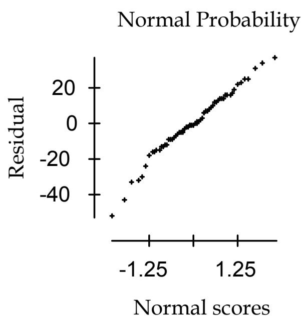
b) According to the linear model, each additional inch in *chest* is associated with an average increase of 0.71272% in *%body fat*.

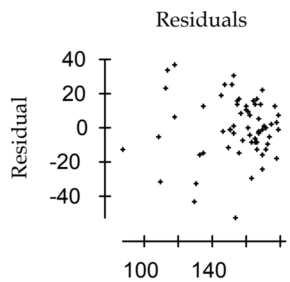
c) After allowing for the linear effects of *waist* and *height*, the multiple regression model predicts an average decrease of 0.233531% in *%body fat* for each additional inch *chest* size.

d) Each of the variables appears to contribute to the model. There does not appear to be an advantage to removing any of them.

24. Grades.

$$\widehat{final} = 8.15541 + 1.38222(Midterm1) + 0.373241(Midterm2)$$





The regression of *final* on *Midterm 1* and *Midterm 2* has an $R^2 = 56.7\%$ and significant *t*-statistics on both coefficients. The Normal probability plot is reasonably straight and the residuals plot shows no pattern. This model is stronger than all other models using only 2 predictors. *Homework* is another promising predictor in some regression models, but it is skewed and has a low outlier.

25. Fifty states 2009.

a) The only model that seems to do poorly is the one that omits *murder.* The other three are hard to choose among.

$$\widehat{Lifeexp} = 75.437 - 0.41458(Murder) - 0.00062(HSgrad) + 0.00010898(Income)$$
$$R^2 = 68.3\%$$
$$\widehat{Lifeexp} = 77.867 - 0.47180(Murder) + 0.00976(HSgrad) + 4.477(Illiteracy)$$
$$R^2 = 56.3\%$$
$$\widehat{Lifeexp} = 75.3472 - 0.42699(Murder) + 0.00010727(Income) + 1.335(Illiteracy)$$
$$R^2 = 68.5\%$$

b) Each of the models has at least one coefficient with a large *P*-value. This predictor variable could be omitted to simplify the model without degrading it too much.

c) No. Regression models cannot be interpreted that way. Association is not the same thing as causation.

d) Plots of the residuals highlight some states as possible outliers. You may want to consider setting them aside to see if the model changes.

26. Breakfast cereals again.

a) $\widehat{calories} = 83.0469 + 0.05721(sodium) - 0.01933(potassium) + 2.38757(sugars)$

b) $R^2 = 38.4\%$. This is not very large. The model would make only moderately accurate predictions.

c) The *P*-value for *potassium* is large, suggesting that this predictor could be removed.

d) **Straight enough condition:** The scatterplots of the response versus each predicted value should be reasonably straight.
Randomization condition: Hopefully, these cereals are representative of all cereals, and can be considered a random sample. Otherwise, the regression residuals plots should show no pattern.
Nearly Normal condition: The Normal probability plot should be straight, and the histogram of residuals should be unimodal and symmetric.
Does the plot thicken? condition: The spread of the residuals plots should be constant.

27. Burger King.

a) With an $R^2 = 99.88\%$, the model should make excellent predictions.

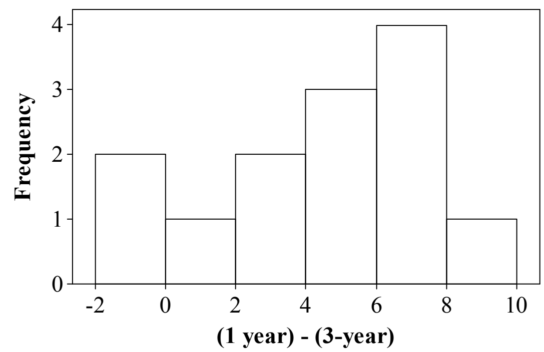
b) The value of *s*, 8.178 calories, is very small compared to the initial standard deviation of *calories*. This means that the model fits the data quite well, leaving very little variation unaccounted for.

c) No, the residuals are not all 0. Indeed, we know that their standard deviation is $s = 8.178$ calories. They are very small compared with the original values.

Review of Part VII – Inference When Variables Are Related

1. Mutual funds 2013.

a) **Paired data assumption:** These data are paired by mutual fund.
Randomization condition: Assume that these funds are representative of all mutual funds.
Nearly Normal condition: The histogram of differences shows plausible normality, even if not symmetric.

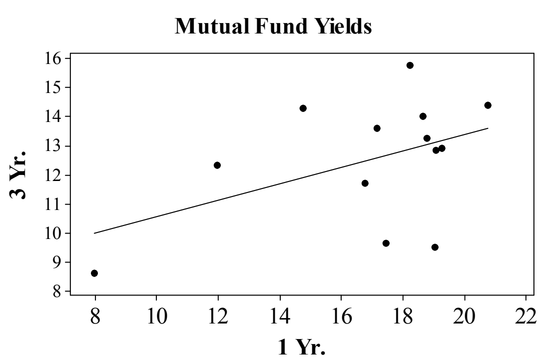


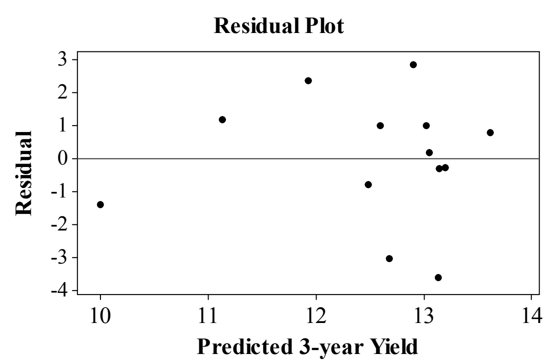
Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 13 – 1 = 12 degrees of freedom. We will find a paired t-interval, with 95% confidence.

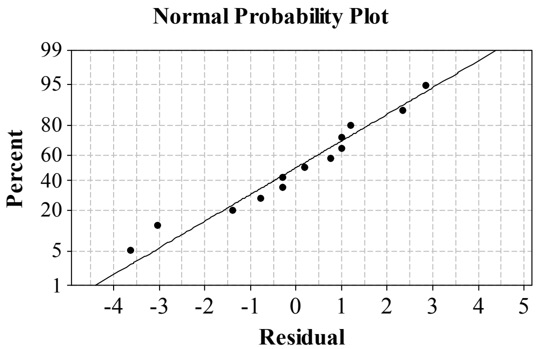
$$\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 4.3823 \pm t^*_{12}\left(\frac{3.14529}{\sqrt{13}}\right) \approx (2.48,\ 6.28)$$

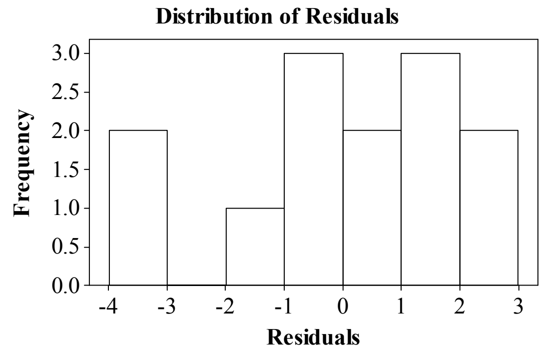
Provided that these mutual funds are representative of all mutual funds, we are 95% confident that, on average, 1-year yields are between 2.48% and 6.28% higher than 3-year yields.

b) H_0: There is no linear relationship between 1-year and 3-year rates of return. $(\beta_1 = 0)$
H_A: There is a linear relationship between 1-year and 3-year rates of return. $(\beta_1 \neq 0)$







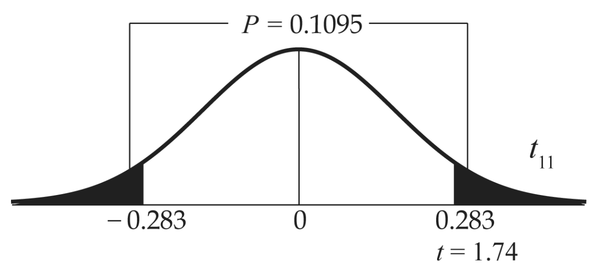


Straight enough condition: The scatterplot is straight enough to try linear regression, though one point seems to be influencing the association.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals isn't very straight and the histogram of residuals isn't symmetric. With a sample size of 13, it's difficult to determine the true nature of the distribution, so it is probably okay to proceed, but we won't have too much faith in our results.

Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (13 – 2) = 11 degrees of freedom. We will use a regression slope t-test.

The equation of the line of best fit for these data points is:
$\widehat{(3year)} = 7.749 + 0.283(1year)$.

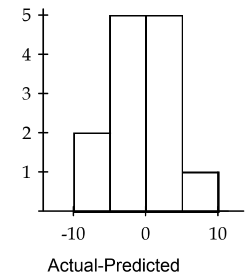
The value of $t \approx 1.74$. The P-value of 0.1095 means that the association we see in the data is likely to occur by chance. We fail to reject the null hypothesis, and conclude that there is no evidence of a linear relationship between the rates of return for 1-year and 3-year periods. Furthermore, we don't know if this sample of mutual funds is a random sample (or even representative) of all mutual funds.



2. Polling.

a) H_0: The mean difference in the number of predicted Democrats and the number of actual Democrats is zero. $(\mu_d = 0)$

H_A: The mean difference in the number of predicted Democrats and the number of actual Democrats is different than zero. $(\mu_d \neq 0)$



Paired data assumption: The data are paired by year.
Randomization condition: Assume these predictions are representative of other predictions.
Nearly Normal condition: The histogram of differences between the predicted number of Democrats and the actual number of Democrats is roughly unimodal and symmetric. The year 1958 is an outlier, and was removed.

Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 13 – 1 = 12 degrees of freedom, $t_{12}\left(0, \frac{3.57878}{\sqrt{13}}\right)$.

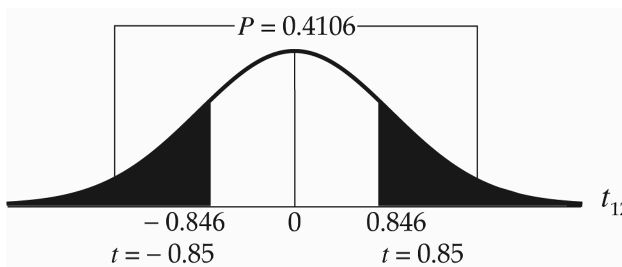
We will use a paired t-test, with $\bar{d} = -0.846$.

Since the P-value = 0.4106 is high, we fail to reject the null hypothesis. There is no evidence that the mean difference between the actual and predicted number of Democrats was different than 0.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

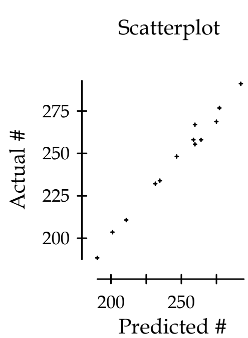
$$t = \frac{-0.846 - 0}{\frac{3.57878}{\sqrt{13}}}$$

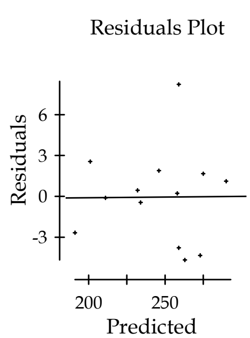
$$t \approx -0.85$$

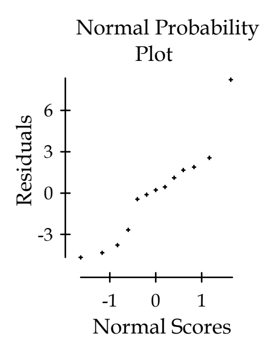


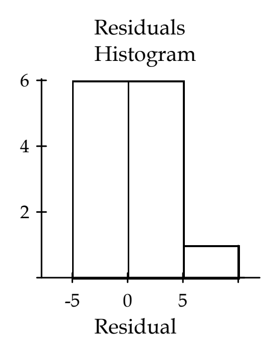
b) H_0: There is no linear relationship between Gallup's predictions and the actual number of Democrats. $(\beta_1 = 0)$

H_A: There is a linear relationship between Gallup's predictions and the actual number of Democrats. $(\beta_1 \neq 0)$









Straight enough condition: The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.

Nearly Normal condition: After an outlier in 1958 is removed, the Normal probability plot of residuals still isn't very straight. However, the histogram of residuals is roughly unimodal and symmetric. With a sample size of 13, it is probably okay to proceed.

Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's *t*-model with (13 – 2) = 11 degrees of freedom. We will use a regression slope *t*-test.

Dependent variable is: **Actual**
No Selector
R squared = 98.7% R squared (adjusted) = 98.6%
s = 3.628 with 13 - 2 = 11 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	10874.4	1	10874.4	826
Residual	144.805	11	13.1641	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	6.00180	8.395	0.715	0.4895
Predicted	0.972206	0.0338	28.7	≤ 0.0001

The equation of the line of best fit for these data points is:
$\widehat{Actual} = 6.00180 + 0.972206(Predicted)$.

The value of $t \approx 28.7$. The *P*-value of essentially 0 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between the number of Democrats predicted by Gallup and the number of Democrats actually in the House of Representatives. Years in which the predicted number was high tend to have high actual numbers also. The high value of $R^2 = 98.7\%$ indicates a very strong model. Gallup polls seem very accurate.

3. Football.

a) He should have performed a one-way ANOVA and an *F*-test.

b) H_0: The mean distance thrown is the same for each grip. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean distances thrown are not all the same.

c) With a *P*-value equal to 0.0032, we reject the null hypothesis and conclude that the distance thrown is not the same for all grips, on average.

d) **Randomization condition:** The grips used were randomized.
Similar Variance condition: The boxplots should show similar spread.
Nearly Normal condition: The histogram of the residuals should be unimodal and symmetric.

e) He might want to perform a multiple comparison test to see which grip is best.

4. Golf.

a) She should have performed a one-way ANOVA and an *F*-test.

b) H_0: The mean distance from the target is the same for all four clubs. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean distances from the target are not all the same.

c) With a P-value equal to 0.0245, we reject the null hypothesis and conclude that the mean distances from the target are not the same for all four clubs.

d) **Randomization condition:** The clubs used were randomized.
Similar Variance condition: The boxplots should show similar spread.
Nearly Normal condition: The histogram of the residuals should be unimodal and symmetric.

e) She should perform a multiple comparison test to determine which club is best.

5. Horses again, but less wild.

a) *Sterilized* is an indicator variable.

b) According to the model, there were, on average, 6.4 fewer foals in herds in which some stallions were sterilized, after allowing for the number of adults in the herd.

c) The p-value for *sterilized* is equal to 0.096. That is not significant at the $\alpha = 0.05$ level, but it does seem to indicate some effect.

6. North and hard.

a) *Derby* is an indicator variable.

b) According to the model. The mortality rate north of Derby is, on average, 158.9 deaths per 100,000 higher than south of Derby, after allowing for the linear effects of the water hardness.

c) The regression with the indicator variable appears to be a better regression. The indicator has a highly significant (P-value equal to 0.0001) coefficient, and the R^2 for the overall model has increased from 43% to 56.2%.

7. Lost baggage.

The **Counted data condition** is not met. We cannot use a chi-square test.

8. Bowling.

a) H_0: The mean number of pins knocked down is the same for all three weights. $(\gamma_{Low} = \gamma_{Med} = \gamma_{High})$

H_A: The mean number of pins knocked down are not all the same.

H_0: The mean number of pins knocked down is the same whether walking or standing. $(\tau_S = \tau_W)$

H_A: The mean number of pins knocked down are not the same for the two approaches.

b) The *weight* sum of squares has 3 – 1 = 2 degrees of freedom.
The *approach* sum of squares has 2 – 1 = 1 degrees of freedom.
The error sum of squares has (24 – 1) – 2 – 1 = 20 degrees of freedom.

c) If the interaction plot shows any evidence of not being parallel, she should fit an interaction term, using 2 (1) = 2 degrees of freedom.

9. Video racing.

a) H_0: The mean time is the same for all three types of mouse. $(\gamma_{Ergo} = \gamma_{\mathrm{Re}g} = \gamma_{Cord})$

H_A: The mean times are not all the same for the three types of mouse.

H_0: The mean time is the same whether the lights are on or off. $(\tau_{On} = \tau_{Off})$

H_A: The mean times are not the same for lights on and lights off.

b) The *mouse* sum of squares has 3 – 1 = 2 degrees of freedom.
The *light* sum of squares has 2 – 1 = 1 degrees of freedom.
The error sum of squares has (6 – 1) – 2 – 1 = 2 degrees of freedom.

c) No, he should not fit an interaction term. The interaction term would use 2 degrees of freedom. This would not leave any degrees of freedom for the error sum of squares and thus would not allow any interpretation of the factors.

10. Resume fraud.

In order to estimate the true percentage of people have misrepresented their backgrounds to within ± 5%, the company would have to perform about 406 random checks.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.05 = 2.326\sqrt{\frac{(0.25)(0.75)}{n}}$$

$$n = \frac{(2.326)^2(0.25)(0.75)}{(0.05)^2}$$

$n \approx$ 406 random checks

11. Paper airplanes.

a) It is reasonable to think that the flight distances are independent of one another. The histogram of flight distances (given) is unimodal and symmetric. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's t model, with 11 – 1 = 10 degrees of freedom. We will use a one-sample t -interval with 95% confidence for the mean flight distance.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 48.3636 \pm t^*_{10}\left(\frac{18.0846}{\sqrt{11}}\right) \approx (36.21,\ 60.51)$$

We are 95% confident that the mean distance the airplane may fly is between 36.21 and 60.51 feet.

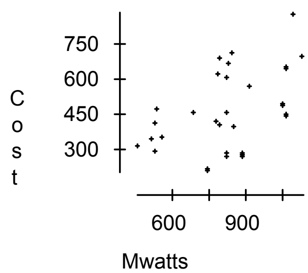
b) Since 40 feet is contained within our 95% confidence interval, it is plausible that the mean distance is 40 feet.

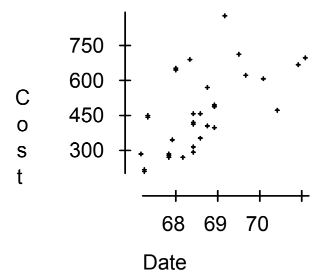
c) A 99% confidence interval would be wider. Intervals with greater confidence are less precise.

d) In order to cut the margin of error in half, she would need a sample size four times as large, or 44 flights.

12. Nuclear plants.

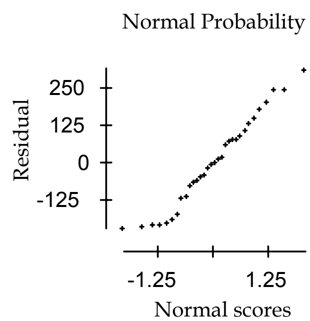
a) Both scatterplots are straight enough. Larger plants are more expensive. Plants got more expensive over time.

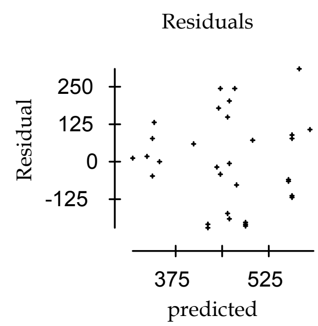




b) $\widehat{cost} = 111.741 + 0.423831(mwatts)$. According to the linear model, the cost of nuclear plants increased by $42,383, on average, for each Megawatt of power.

c) **Straight enough condition:** The plot of residuals versus predicted values looks reasonably straight.
Independence assumption: It is reasonable to think of these nuclear power plants as representative of all nuclear power plants.
Nearly Normal condition: The Normal probability plot looks straight enough.
Does the plot thicken? condition: The residuals plot shows reasonably constant spread.





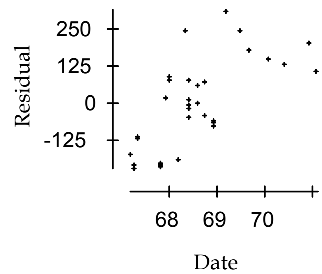
d) H_0: The power plant capacity (*mwatts)* is not linearly associated with the cost of the plant. $(\beta_{Mwatts} = 0)$

H_A: The power plant capacity (*mwatts*) is linearly associated with the cost of the plant. $(\beta_{Mwatts} \neq 0)$

With $t = 2.93$, and 30 degrees of freedom, the P-value equals 0.0064. Since the P-value is so small, we reject the null hypothesis, *mwatts* is linearly associated with the cost of the plant.

e) From the regression model we can find the cost of a 1000 *mwatt* plant: $111.741 + 0.42383(1000) \approx 535.57$ which is \$535,570,000.

f) The R^2 means that 22.3% of the variation in *cost* of nuclear plants can be accounted for by the linear relationship between *cost* and the size of the plant, measured in megawatts (*mwatts*). A scatterplot of residuals against *date* shows a strong linear patter, so it looks like *date* could account for more of the variation.



g) The coefficients of *mwatts* changed very little from the simple linear regression to the multiple regression.

h) Because the coefficient changed little when we added *date* to the model, we can expect that *date* and *mwatts* are relatively uncorrelated. In fact, their correlation is 0.02.

13. Barbershop music.

a) With an R^2 of 90.9%, your friend is right about being able to predict singing scores.

b) H_0: When including the other predictor, performance does not change our ability to predict singing scores. $(\beta_{Prs} = 0)$

H_A: When including the other predictor, performance changes our ability to predict singing scores. $(\beta_{Prs} \neq 0)$

With $t = 8.13$, and 31 degrees of freedom, the *P*-value is less than 0.0001, so we reject the null hypothesis and conclude that *performance* is useful in predicting singing scores.

c) Based on the multiple regression, both *performance* and *music* (even with a *P*-value equal to 0.0766) are useful in predicting singing scores. According to the residuals plot the spread is constant across the predicted values. The histogram of residuals is unimodal and symmetric. Based on the information provided we have met the Similar Variance and Nearly Normal conditions.

14. Sleep.

a) H_0: The mean hours slept is the same for both sexes. $(\gamma_F = \gamma_M)$

H_A: The mean hours slept are not the same for both sexes.

H_0: The mean hours slept is the same for all classes. $(\tau_{Fr} = \tau_{So} = \tau_{Jr} = \tau_{Sr})$

H_A: The mean hours slept are not the same for all classes.

b) Based on the interaction plot, the lines are not parallel, so the Additive Enough condition is not met. The interaction term should be fit.

c) With all three *P*-values so large, none of the effects appear to be significant.

d) There are a few outliers. Based on the residuals plot, the Similar Variance condition appears to be met, but we do not have a Normal probability plot of residuals. The main concern is that we may not have enough power to detect differences between groups. In particular it may be that upper class women (juniors and seniors) sleep more than other groups, but we would need more data to tell.

15. Study habits.

a) H_0: The mean hours studied is the same for both sexes. $(\gamma_F = \gamma_M)$

H_A: The mean hours studied not the same for both sexes.

H_0: The mean hours studied is the same for all classes. $(\tau_{Fr} = \tau_{So} = \tau_{Jr} = \tau_{Sr})$

H_A: The mean hours studied are not the same for all classes.

b) Based on the interaction plot, the lines are not parallel, so the Additive Enough condition is not met. The interaction term should be fit.

c) With all three *P*-values so large, none of the effects appear to be significant.

d) There are a few outliers. Based on the residuals plot, the Similar Variance condition appears to be met, but we do not have a Normal probability plot of residuals. The main concern is that we may not have enough power to detect differences between groups. We would need more data to increase the power.

16. Is Old Faithful getting older?

a) $\widehat{interval} = 35.2463 + 10.4348(duration) - 0.126316(day)$

b) H_0: After allowing for the effects of duration, *interval* does not change over time. $\left(\beta_{Day} = 0\right)$

H_A: After allowing for the effects of duration, *interval* does change over time. $\left(\beta_{Day} \neq 0\right)$

With a *P*-value equals 0.0166, we reject the null hypothesis. It appears there is a change over time, after we account for the *duration* of the previous eruption.

c) The coefficient for *day* is not about the two variable association, but about the relationship between *interval* and *day* after allowing for the linear effects of *duration*.

d) The amount of change is only about –0.126316 (60) = –7.58 seconds per day. This doesn’t seem particularly meaningful, although we expect a change of about 46 minutes per year.

17. Teen traffic deaths 2007.

a) **Straight enough condition:** The scatterplots of the response versus each predicted value should be reasonably straight.
Independence: These data are measured over time. We should plot the data and the residuals against time to look for failures of independence.
Nearly Normal condition: The Normal probability plot should be straight, and the histogram of residuals should be unimodal and symmetric.
Similar Variance condition: The spread of the residuals plot is constant.

b) $R^2 = 57.4\%$, so 57.4% of the variation in female teen traffic deaths is accounted for by the linear model.

c) $\widehat{deaths} = 45074.0 - 21.6006(year)$

d) The model predicts that female teen traffic deaths have been declining at a rate of approximately 21.6 per year.

18. Births.

a) $\widehat{births} = 4422.98 - 15.1898(age) - 1.89830(year)$

b) According to the model, births seem to be declining, on average, at a rate of 1.89 births per 1000 women each year, after allowing for differences across the age of women. According to the model, from 1990 to 1999, birth rate decreased by about 17 births per 1000 women.

c) The scatterplot shows both clumping and a curved relationship. We might want to re-express *births* or add a quadratic term to the model. The clumping is dues to having data for each year of the decade of the 90s for the age bracket of women. It probably does not indicate a failure of the linearity assumption.

19. Typing.

a) H_0: Typing speed is the same whether the gloves are on or off. $(\gamma_{On} = \gamma_{Off})$

H_A: Typing speeds are not the same when the gloves are on and off.

H_0: Typing speed is the same at hot or cold temperatures. $(\tau_H = \tau_C)$

H_A: Typing speeds are not the same at hot and cold temperatures.

b) Based on the boxplots, it appears that both the effects of *temperature* and *gloves* affect typing speed.

c) The interaction plot is not parallel so the additive enough condition is not met. Therefore, an interaction term should be fit.

d) Yes, I think it will be significant because a difference in typing speed due to temperature seems to be significant only when he is not wearing the gloves.

e) The P-values for all three effects: *gloves, temperature* and the interaction term, are all very small, so all three effects are significant.

f) Gloves decreased typing speed at both temperatures. A hot temperature increases typing speed with the gloves off, but has little effect with the gloves on.

g) $s = \sqrt{\frac{58.75}{28}} = 1.45$ words per minute. Yes, this seems consistent with the size of the variation shown on the partial boxplots.

h) Tell him to type in a warm room without wearing his gloves.

i) The Normal probability plot of residuals should be straight. The residuals plot shows constant spread across the predicted values. The Similar Variance condition appears to be met. For the levels of the factors that he used, it seems clear that a warm room and not wearing gloves are beneficial for his typing.

20. Typing again.

a) H_0: Typing speed is the same when the television is on or when it is off. $(\gamma_{On} = \gamma_{Off})$

H_A: Typing speeds are not the same when the television is on and off.

H_0: Typing speed is the same when the music is on or when it is off. $(\tau_{On} = \tau_{Off})$

H_A: Typing speeds are not the same when the music is on and off.

b) Based on the partial boxplots, both effects appear to be small and probably not statistically significant.

c) The interaction plot is not parallel so the additive enough condition is not met. Therefore, an interaction term should be fit.

d) The interaction term may be real. It appears that the effect of television is stronger with the music off than on.

e) All three effects have large P-values. *Music* has a P-value equal to 0.4312, *television* has a P-value equal to 0.2960, and the interaction effect has a P-value equal to 0.5986. None of these effects appear to be significant.

f) None of the effects seem strong. He seems to type just as well with the music and/or the television on.

g) $s = \sqrt{\frac{197.5}{28}} = 2.66$ words per minute. This seems consistent with the size of the variation shown in the partial boxplots.

h) Turning the television and/or the music on will not increase his typing speed, nor will it decrease it.

i) The Normal probability plot of residuals should be straight. The residuals plot shows constant spread across the predicted values. The Similar Variance condition appears to be met. For the levels of the factors that he used, it seems that to the level of experimental error present, neither the music nor the television affects his typing speed. If he wanted to see smaller effects, he would have to increase his sample size.

21. New York Marathon.

a) The number of finishers per minute appears to increase by about 0.519037 finishers per minute.

b) There is a definite pattern in the residuals. This is a clear violation of the linearity assumption. While there does seem to be a strong association between time and number of finishers, it is not linear.

22. Learning math.

a) H_0: The mean score of Accelerated Math students is the same as the mean score of traditional students. $(\mu_A = \mu_T \text{ or } \mu_A - \mu_T = 0)$

H_A: The mean score of Accelerated Math students is different from the mean score of traditional students. $(\mu_A \neq \mu_T \text{ or } \mu_A - \mu_T \neq 0)$

Independent groups assumption: Scores of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 459.24 degrees of freedom (from the approximation formula).

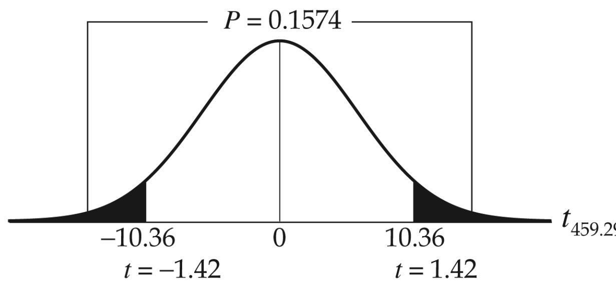
We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{y}_A - \bar{y}_T) = \sqrt{\frac{84.29^2}{231} + \frac{74.68^2}{245}} \approx 7.3158$.

The observed difference between the mean scores is 560.01 – 549.65 = 10.36

$$t = \frac{(\bar{y}_A - \bar{y}_T) - (0)}{SE(\bar{y}_A - \bar{y}_T)}$$

$$t \approx \frac{10.36}{7.3158}$$

$$t \approx 1.42$$



Since the P-value = 0.1574, we fail to reject the null hypothesis. There is no evidence that the Accelerated Math students have a different mean score on the pretest than the traditional students.

b) H_0: Accelerated Math students do not show significant improvement in test scores. The mean individual gain for Accelerated Math is zero. $(\mu_d = 0)$

H_A: Accelerated Math students show significant improvement in test scores. The mean individual gain for Accelerated Math is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The Accelerated Math students had a mean individual gain of $\bar{d}$ = 77.53 points and a standard deviation of 78.01 points. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean individual gain with a Student's t model, with 231 – 1 = 230 degrees of freedom, $t_{230}\left(0, \frac{78.01}{\sqrt{231}}\right)$.

We will perform a paired t-test.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{77.53 - 0}{\frac{78.01}{\sqrt{231}}}$$

$$t \approx 15.11$$

Since the P-value is essentially 0, we reject the null hypothesis. There is strong evidence that the mean individual gain is greater than zero. The Accelerated Math students showed significant improvement.

c) H_0: Students taught using traditional methods do not show significant improvement in test scores. The mean individual gain for traditional methods is zero. $(\mu_d = 0)$

H_A: Students taught using traditional methods show significant improvement in test scores. The mean individual gain for traditional methods is greater than zero. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The students taught using traditional methods had a mean individual gain of $\bar{d}$ = 39.11 points and a standard deviation of 66.25 points. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean individual gain with a Student's *t* model, with 245 – 1 = 244 degrees of freedom, $t_{244}\left(0, \frac{66.25}{\sqrt{245}}\right)$. We will perform a paired *t*-test.

Since the *P*-value is essentially 0, we reject the null hypothesis. There is strong evidence that the mean individual gain is greater than zero. The students taught using traditional methods showed significant improvement.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{39.11 - 0}{\frac{66.25}{\sqrt{245}}}$$

$$t \approx 9.24$$

d) H_0: The mean individual gain of Accelerated Math students is the same as the mean individual gain of traditional students. $(\mu_{dA} = \mu_{dT} \text{ or } \mu_{dA} - \mu_{dT} = 0)$

H_A: The mean individual gain of Accelerated Math students is greater than the mean individual gain of traditional students. $(\mu_{dA} > \mu_{dT} \text{ or } \mu_{dA} - \mu_{dT} > 0)$

Independent groups assumption: Individual gains of students from different classes should be independent.
Randomization condition: Although not specifically stated, classes in this experiment were probably randomly assigned to learn either Accelerated Math or traditional curricula.

Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. However, the samples are large. The Central Limit Theorem allows us to proceed.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 452.10 degrees of freedom (from the approximation formula).

We will perform a two-sample *t*-test. The sampling distribution model has mean 0, with standard error: $SE(\bar{d}_A - \bar{d}_T) = \sqrt{\frac{78.01^2}{231} + \frac{66.25^2}{245}} \approx 6.6527$.

The observed difference between the mean scores is 77.53 – 39.11 = 38.42

Since the *P*-value is less than 0.0001, we reject the null hypothesis. There is strong evidence that the Accelerated Math students have an individual gain that is significantly higher than the individual gain of the students taught using traditional methods.

$$t = \frac{(\bar{d}_A - \bar{d}_T) - 0}{SE(\bar{d}_A - \bar{d}_T)}$$

$$t = \frac{38.42 - 0}{6.6527}$$

$$t \approx 5.78$$

23. Pesticides.

H_0 : The percentage of males born to workers at the plant is 51.2%.($p = 0.512$)
H_A : The percentage of males is less than 51.2%. ($p < 0.512$)

Independence assumption: It is reasonable to think that the births are independent.
Success/Failure Condition: $np = (227)(0.512) = 116$ and $nq = (227)(0.488) = 111$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.512$ and

$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.512)(0.488)}{227}} \approx 0.0332$. We can perform a one-proportion *z*-test.

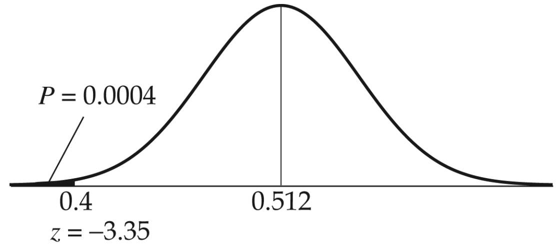
The observed proportion of males is $\hat{p} = 0.40$.

The value of $z \approx -3.35$, meaning that the observed proportion of males is over 3 standard deviations below the expected proportion. The *P*-value associated with this *z* score is approximately 0.0004.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.4 - 0.512}{\sqrt{\frac{(0.512)(0.488)}{227}}}$$

$$z \approx -3.35$$



With a P-value this low, we reject the null hypothesis. There is strong evidence that the percentage of males born to workers is less than 51.2%. This provides evidence that human exposure to dioxin may result in the birth of more girls.

24. Video pinball.

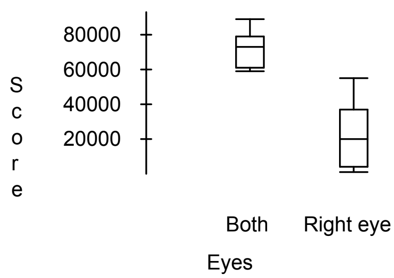
a) H_0: Pinball score is the same whether the tilt is on or off. $(\gamma_{On} = \gamma_{Off})$

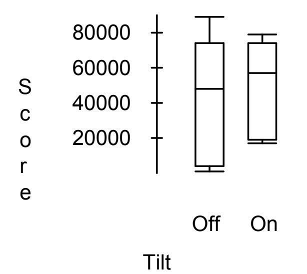
H_A: Pinball scores are different when the tilt is on and when it is off.

H_0: Pinball score is the same whether both eyes are open or the right is closed. $(\tau_B = \tau_R)$

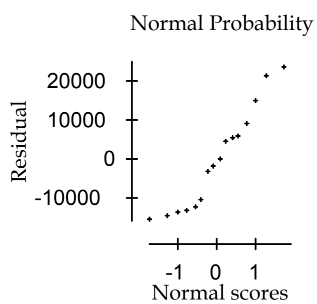
H_A: Pinball scores are different when both eyes are open and when the right is closed.

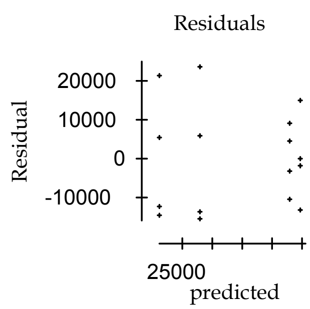
b) The partial boxplots show that the *eye* effect is strong, but there appears to be little or no effect due to *tilt*.

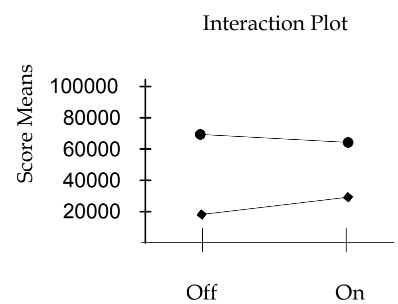




Randomization condition: The experiment was performed in random order.
Similar Variance condition: The side-by-side boxplots have similar spreads. The residuals plot shows no pattern, and no systematic change in spread.
Nearly Normal condition: The Normal probability plot of the residuals is reasonably straight.







Analysis of Variance For **Score**
No Selector

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Tlt	1	156806745	156806745	0.75080	0.4032
Eye	1	9385201568	9385201568	44.937	≤ 0.0001
Tlt*Eye	1	450532463	450532463	2.1572	0.1676
Error	12	2506236979	208853082		
Total	15	12498777755			

According to the ANOVA, only eye effect is significant with a p-value less than 0.0001. Both the *tilt* and interaction effects are not significant. We conclude that keeping both eyes open improves score but that using the tilt does not.

25. Javelin.

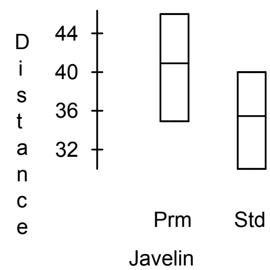
a) H_0: Type of javelin has no effect on the distance of her throw. $(\gamma_P = \gamma_S)$

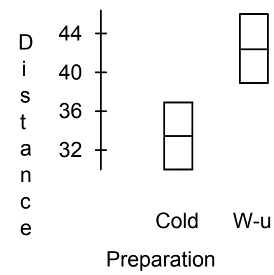
H_A: Type of javelin does have an effect on the distance of her throw.

H_0: Preparation has no effect on the distance of her throw. $(\tau_W = \tau_C)$

H_A: Preparation does have an effect on the distance of her throw.

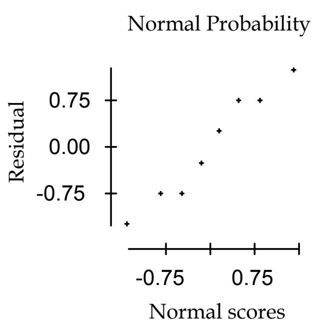
b) The partial boxplots show that both factors seem to have an effect on the distance of her throw.

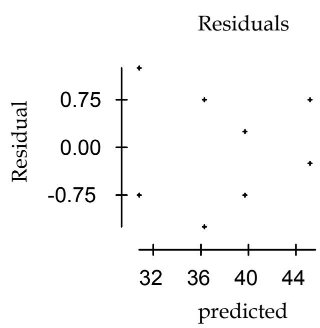


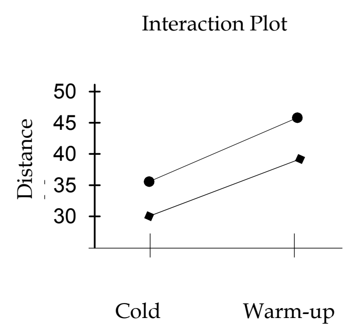


The interaction plot is parallel. The additive enough condition is met. No interaction term is needed.

Randomization condition: The experiment was performed in random order. **Similar Variance condition:** The side-by-side boxplots have similar spreads. The residuals plot shows no pattern, and no systematic change in spread. **Nearly Normal condition**: The Normal probability plot of the residuals should be straight.







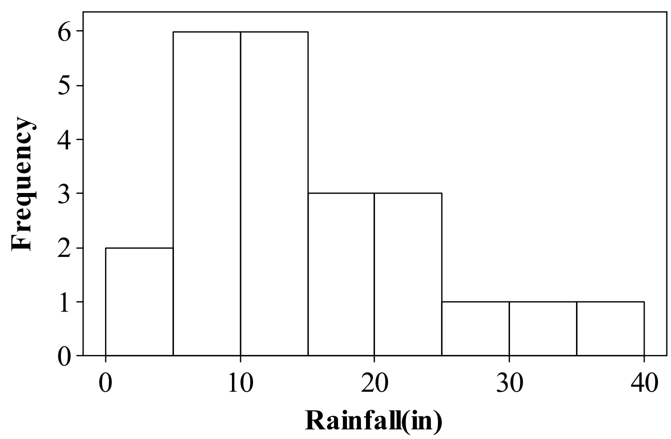
Analysis of Variance For **Distance (m)**
No Selector

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	11552	11552	10502	≤ 0.0001
Prn	1	162	162	147.27	≤ 0.0001
Jvn	1	60.5000	60.5000	55.000	0.0007
Error	5	5.50000	1.10000		
Total	7	228			

With very small *P*-values, both effects are significant. It appears that depending on the cost, the premium javelin may be worth it – it increases the distance about 5.5 meters, on average. Warming up increases distance about 9 meters, on average. She should always warm up and consider using the premium javelin.

26. LA rainfall 2012.

a) Independence assumption: Annual rainfall is independent from year to year.
Nearly Normal condition: The histogram of the rainfall totals is skewed to the right, but the sample is fairly large, so it is safe to proceed.



The mean annual rainfall is 15.0835 inches, with a standard deviation 8.62983 inches. Since the conditions have been satisfied, construct a one-sample t-interval, with 23 – 1 = 22 degrees of freedom, at 90% confidence.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 15.0835 \pm t^*_{22}\left(\frac{8.62983}{\sqrt{23}}\right) \approx (11.99,\ 18.17)$$

We are 90% confident that the mean annual rainfall in LA is between 11.99 and 18.17 inches.

b) Start by making an estimate, either using $z^* = 1.645$ or $t^*_{22} = 1.717$ from above. Either way, your estimate is around 55 years. Make a better estimate using $t^*_{54} = 1.674$. You would need about 53 years' data to estimate the annual rainfall in LA to within 2 inches.

$$ME = t^*_{40}\left(\frac{s}{\sqrt{n}}\right)$$

$$2 = 1.674\left(\frac{8.62983}{\sqrt{n}}\right)$$

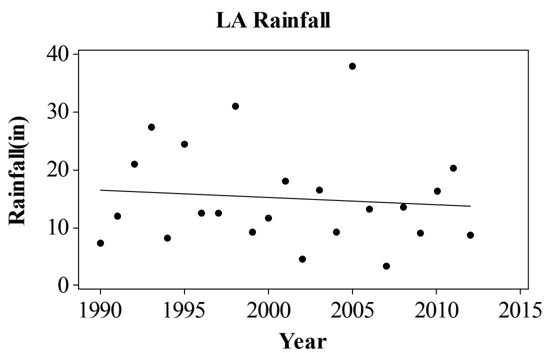
$$n = \frac{(1.674)^2(8.62983)^2}{(2)^2}$$

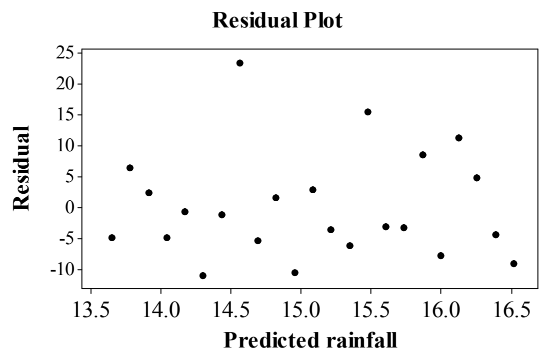
$$n \approx 53 \text{ years}$$

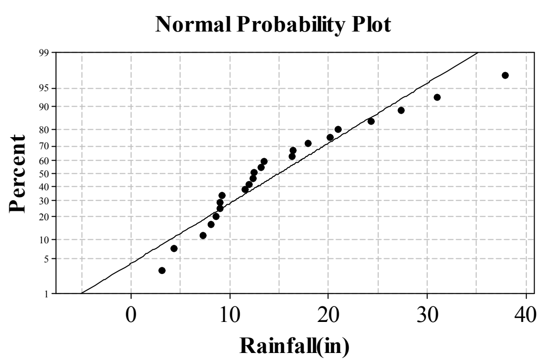
c) H_0: There is no linear relationship between year and annual LA rainfall. $(\beta_1 = 0)$

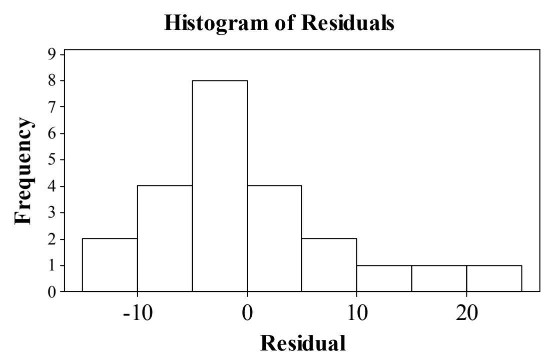
H_A: There is a linear relationship between year and annual LA rainfall. $(\beta_1 \neq 0)$

Straight enough condition: The scatterplot is straight enough to try linear regression, although there is no apparent pattern.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals is not straight, but a sample of 23 years is large enough to proceed.









Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (23 – 2) = 21 degrees of freedom. We will use a regression slope t-test.

The equation of the line of best fit for these data points is:
$\widehat{Rain} = 275.8 - 0.1303(Year)$.

The value of $t \approx -0.472$. The P-value of 0.64 means that the association we see in the data is quite likely to occur by chance. We fail to reject the null hypothesis, and conclude that there is no evidence of a linear relationship between the annual rainfall in LA and the year.

27. TV and athletics.

a) H_0: The mean hours of television watched is the same for all 3 groups. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean hours of television watched are not the same for all 3 groups.

b) The variance for the *none* group appears to be slightly smaller, and there are outliers in all three groups. We do not have a Normal probability plot of the residuals, but we suspect that the data may not be Normal enough.

c) The ANOVA *F*-test indicates that athletic participation is significant with a p-value equal to 0.0167, we conclude that the number of television hours watched is not the same for all three types of athletic participation.

d) It seems that the differences are evident even when the outliers are removed. The conclusions seems valid.

28. Weight and athletics.

a) H_0: Mean weight is the same for all 3 groups. $(\mu_1 = \mu_2 = \mu_3)$

H_A: Mean weights are not the same for all 3 groups.

b) According to the boxplots the spread appears constant for all three groups. There is one outlier. We do not have a Normal probability plot, but we suspect that the data may be Normal enough.

c) With a *P*-value equal to 0.0042, we reject the null hypothesis, there is evidence that the mean weights are not the same for all three groups. It may be that more men are involved with athletics, which might explain the weight differences. On the other hand, it may simply be that those who weigh more are more likely to be involved with sports.

d) It seems that differences are evident even when the outlier is removed. It seems that conclusion is valid.

29. Weight loss.

Randomization Condition: The respondents were randomly selected from among the clients of the weight loss clinic.
Nearly Normal Condition: The histogram of the number of pounds lost for each respondent is unimodal and symmetric, with no outliers.

The clients in the sample had a mean weight loss of 9.15 pounds, with a standard deviation of 1.94733 pounds. Since the conditions have been satisfied, construct a one-sample t-interval, with 20 – 1 = 19 degrees of freedom, at 95% confidence.

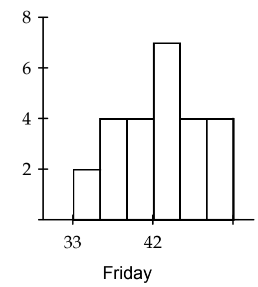
$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 9.15 \pm t^*_{19}\left(\frac{1.94733}{\sqrt{20}}\right) \approx (8.24,\ 10.06)$$

We are 95% confident that the mean weight loss experienced by clients of this clinic is between 8.24 and 10.06 pounds. Since 10 pounds is contained within the interval, the claim that the program will allow clients to lose 10 pounds in a month is plausible. Answers may vary, depending on the chosen level of confidence.

30. Cramming.

a) H_0: The mean score of week-long study group students is the same as the mean score of overnight cramming students. $(\mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0)$

H_A: The mean score of week-long study group students is greater than the mean score of overnight cramming students.
$(\mu_1 > \mu_2 \text{ or } \mu_1 - \mu_2 > 0)$



Independent Groups Assumption: Scores of students from different classes should be independent.
Randomization Condition: Assume that the students are assigned to each class in a representative fashion.
Nearly Normal Condition: The histogram of the crammers is unimodal and symmetric. We don't have the actual data for the study group, but the sample size is large enough that it should be safe to proceed.

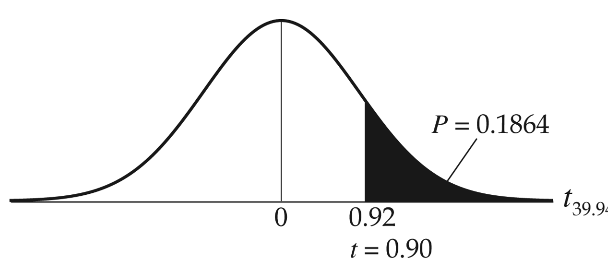
$$\bar{y}_1 = 43.2 \qquad \bar{y}_2 = 42.28$$
$$s_1 = 3.4 \qquad s_2 = 4.43020$$
$$n_1 = 45 \qquad n_2 = 25$$

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 39.94 degrees of freedom (from the approximation formula). We will perform a two-sample t-test. The sampling distribution model has mean 0, with standard error:

$$SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{3.4^2}{45} + \frac{4.43020^2}{25}} \approx 1.02076 .$$

The observed difference between the mean scores is 43.2 – 42.28 = 0.92.

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (0)}{SE(\bar{y}_1 - \bar{y}_2)}$$
$$t \approx \frac{0.92}{1.02076}$$
$$t \approx 0.90$$



Since the P-value = 0.1864 is high, we fail to reject the null hypothesis. There is no evidence that students with a week to study have a higher mean score than students who cram the night before.

b) H_0: The proportion of study group students who will pass is the same as the proportion of crammers who will pass. $(p_1 = p_2 \text{ or } p_1 - p_2 = 0)$

H_A : The proportion of study group students who will pass is different from the proportion of crammers who will pass. $(p_1 \neq p_2 \text{ or } p_1 - p_2 \neq 0)$

Random condition: Assume students are assigned to classes in a representative fashion.
10% condition: 45 and 25 are both less than 10% of all students.
Independent samples condition: The groups are not associated.
Success/Failure condition: $n_1\hat{p}_1$ = 15, $n_1\hat{q}_1$ = 30, $n_2\hat{p}_2$ = 18, and $n_2\hat{q}_2$ = 7 are not all greater than 10, since only 7 crammers didn't pass. However, if we check the pooled value, $n_2\hat{p}_{\text{pooled}}$ = (25)(0.471) = 11.775. All of the samples are large enough.

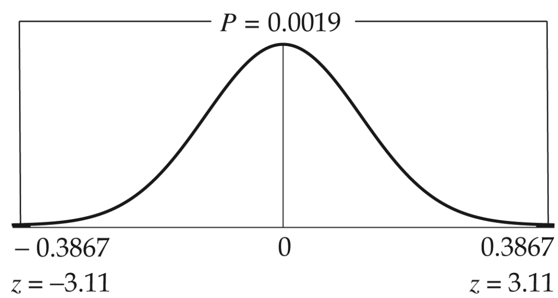
Since the conditions have been satisfied, we will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by

$$SE_{\text{pooled}}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_1} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_2}} = \sqrt{\frac{\left(\frac{33}{70}\right)\left(\frac{37}{70}\right)}{45} + \frac{\left(\frac{33}{70}\right)\left(\frac{37}{70}\right)}{25}} \approx 0.1245.$$

The observed difference between the proportions is:
0.3333 – 0.72 = –0.3867.

Since the P-value = 0.0019 is low, we reject the null hypothesis. There is strong evidence to suggest a difference in the proportion of passing grades for study group participants and overnight crammers. The crammers generally did better.

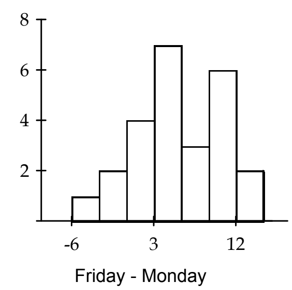
$$z = \frac{-0.3867 - 0}{0.1245}$$
$$z \approx -3.11$$



c) H_0: There is no mean difference in the scores of students who cram, after 3 days. $(\mu_d = 0)$

H_A: The scores of students who cram decreases, on average, after 3 days. $(\mu_d > 0)$

Paired data assumption: The data are paired by student.
Randomization condition: Assume that students are assigned to classes in a representative fashion.
Nearly Normal condition: The histogram of differences is roughly unimodal and symmetric.



Since the conditions are satisfied, the sampling distribution of the difference can be modeled with a Student's t-model with 25 – 1 = 24 degrees of freedom, $t_{24}\left(0, \frac{4.8775}{\sqrt{25}}\right)$. We will use a paired t-test, with $\bar{d} = 5.04$.

Since the P-value is less than 0.0001, we reject the null hypothesis. There is strong evidence that the mean difference is greater than zero. Students who cram seem to forget a significant amount after 3 days.

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{5.04 - 0}{\frac{4.8775}{\sqrt{25}}}$$

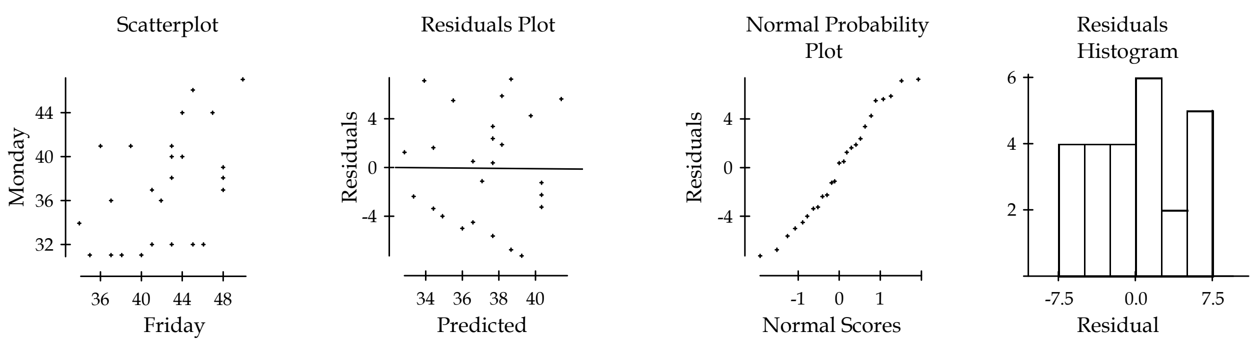
$$t \approx 5.17$$

d) $\bar{d} \pm t^*_{n-1}\left(\frac{s_d}{\sqrt{n}}\right) = 5.04 \pm t^*_{24}\left(\frac{4.8775}{\sqrt{25}}\right) \approx (3.03, 7.05)$

We are 95% confident that students who cram will forget an average of 3.03 to 7.05 words in 3 days.

e) H_0: There is no linear relationship between Friday score and Monday score. $(\beta_1 = 0)$

H_A: There is a linear relationship between Friday score and Monday score. $(\beta_1 \neq 0)$



Straight enough condition: The scatterplot is straight enough to try linear regression.
Independence assumption: The residuals plot shows no pattern.
Does the plot thicken? condition: The spread of the residuals is consistent.
Nearly Normal condition: The Normal probability plot of residuals is reasonably straight, and the histogram of the residuals is roughly unimodal and symmetric.

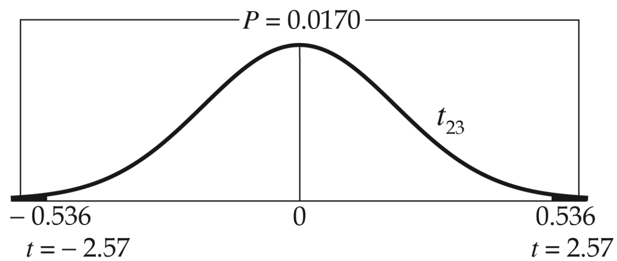
Since the conditions for inference are satisfied, the sampling distribution of the regression slope can be modeled by a Student's t-model with (25 – 2) = 23 degrees of freedom. We will use a regression slope t-test.

Dependent variable is: **Monday**
No Selector
R squared = 22.4%   R squared (adjusted) = 19.0%
s = 4.518 with 25 - 2 = 23 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	135.159	1	135.159	6.62
Residual	469.401	23	20.4087	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	14.5921	8.847	1.65	0.1127
Friday	0.535666	0.2082	2.57	0.0170

The equation of the line of best fit for these data points is:
$\widehat{Monday} = 14.5921 + 0.535666(Friday)$.



The value of $t \approx 2.57$. The P-value of 0.0170 means that the association we see in the data is unlikely to occur by chance. We reject the null hypothesis, and conclude that there is strong evidence of a linear relationship between Friday score and Monday score. Students who do better in the first place tend to do better after 3 days. However, since R^2 is only 22.4%, Friday score is not a very good predictor of Monday score.

Chapter 29 – Multiple Regression Wisdom

Section 29.1

1. Indicators.

a) Use and indicator. Code Male = 0, Female = 1 (or the reverse).

b) Treat it as a quantitative predictor.

c) Use and indicator. Code older than 65 = 1, Younger than 65 = 0.

2. More indicators.

a) Treat it as quantitative predictor.

b) Use an indicator. Code 1 for buildings with elevators, 0 otherwise.

c) Use two indicators. One could be called *Child* and would be 1 for children and 0 otherwise. The other could be called *Senior* and would be 1 for those over 65 and 0 for others. You can't use three indicators, but must leave out one of the set. Other choices are possible.

Section 29.2

3. Residual, leverage, influence.

a) Likely high leverage.

b) Likely outlier.

4. Residual, leverage, influence, 2.

a) Likely influential. It has high leverage for the unusual combination of SAT scores and an unusually high GPA

b) Likely large (negative) residual. We aren't told that the country is otherwise unusual.

Section 29.3

5. Significant coefficient?

No. The predictors are almost certainly collinear, which would affect the coefficient of *Age*.

6. Better model?

The smaller model is likely better. The boss could check the Adjusted R^2 statistic, which is likely to be larger for the smaller regression.

Chapter Exercises.

7. Climate change 2013 again.

a) The distribution of Studentized residuals is unimodal and fairly symmetric. The association should be fine for regression.

b) That would not be a correct interpretation. Since CO_2 levels have risen steadily over recent years, CO_2 and *Year* are likely to be collinear.

8. Pizza.

a) According to the multiple regression model, the pizza score is higher by 15.6 points for cheese pizza than for pepperoni, after allowing for the effects of *calories* and *fat*.

b) We should plot the residuals against predicted values or against each predictor, looking for pattern or outliers. We should check for evidence that the residuals are nearly Normal with a Normal probability plot or histogram.

9. Healthy breakfast, sick data.

a) The slope of a partial regression plot is the coefficient of the corresponding predictor, in this case, –1.020.

b) Quaker oatmeal makes the slope more strongly negative. It appears to have substantial influence on this slope.

c) Not surprisingly, omitting Quaker oatmeal changes the coefficient of *fiber*. It is now positive (although not significantly different from 0). This second regression model has a higher R^2, suggesting that it fits the data better. Without the influential point, the second regression is probably the better model.

d) The coefficient of *fiber* is not discernibly different from 0. We have no evidence that it contributes significantly to calories.

10. Fifty states.

a) Yes, they should be influential. Points that have both large leverage and large Studentized residuals are bound to be influential.

b) The *t*-ratios for indicator variables are *t*-tests of whether those cases fit the regression model established by the other cases. Both of the indicators have *t*-ratios that are large enough to be significant at the 0.05 level.

c) The *t*-ratios for *illiteracy* and *income* are not very large. The coefficient for *income* is near zero. Either predictor might be considered for removal from the regression model.

11. Cereals, part 2.

a) After allowing for the effects of *sodium* and *sugars*, the model predicts a decrease of 0.019 calories for each additional gram of *potassium*.

b) Those points pull the slope of the relationship down. Omitting them should increase the value of the coefficient of *potassium*. It would likely become positive, since the remaining points show a positive slope in the partial regression plot.

c) These appear to be influential points. They have both high leverage and large residuals, and the partial regression plot shows their influence.

d) If our goal is to understand the relationships among these variables, then it might be best to omit these cereals because they seem to behave as if they are from a separate subgroup.

12. Scottish hill races 2008.

a) The scatterplot of the residuals against the predicted values shows a fan shape with the spread increasing for longer races, and two races have large residuals. The Lairig Ghru race has high leverage.

b) Both races may be outliers. Their larger residuals have inflated the residual standard deviation and may have reduced the value of R^2, but we can't tell if they have affected the coefficients without examining partial regression plots.

c) The partial regression plots show that these races have had little effect on the coefficients other than on the intercept, which may have been increased. A scatterplot of Distance vs. Climb shows that the Lairig Ghru race is unusually long and has very little climb (for these races), accounting for its large leverage.

13. Traffic delays 2011.

a) This set of indicators uses *medium* size as its base. The coefficients of *small, large,* and *very large* estimate the average change in amount of *delay/person* relative to the amount of *delay/person* for *medium* size cities found for each of the other three sizes. If there were an indicator variable for *medium* as well, the four indicators would be collinear, so the coefficients could not be estimated.

b) The costs of congestion per commuter in *large* cities are, on average, about $106 greater than in *medium* size cities, after allowing for the effects of *congested%*.

14. Gourmet pizza.

a) Reggio's and Michelina's received lower scores than we would otherwise have expected from the model.

b) The *t*-ratio for the indicator variable for *Michelina's* is – 4.03, which is large. We can reject the null hypothesis that Michelina's fits the regression model, and we can conclude that it is an outlier.

15. More traffic 2011.

a) An assumption required for indicator variables to be useful is that the regression models fit for the different groups identified by the indicators are parallel, i.e. they have the same slope. These lines are not parallel.

b) The coefficient of $Sml \times C\%V$ adjusts the slope of the regression model fit for the small cities. We would say that the slope of *congestion cost per person* on *congested% of motor vehicle miles* for *small* cities (after allowing for the linear effects of the other variables in the model) is \$27.815 – 4.42 = \$23.395.

c) The regression model seems to do a good job. The R^2 shows that 65% of the variability in *congestion cost per person* is accounted for by the model. Most of the P-values for the coefficients are small.

16. Another slice of pizza.

a) Cheese and pepperoni pizzas do not appear to be described by the same model.

b) The slope of taste *score* on *calories*, after allowing for the linear effects of fat and removing the influence of the two outlying pizzas, is estimated to be 1.92 points per gram for pepperoni pizzas and 1.92 – 0.4615 = 1.45 points per gram for cheese pizzas.

c) This should be a better regression model. We have identified a consistent difference between pepperoni and cheese pizzas and incorporated it into the model. All coefficients are significantly different from zero, and both the R^2 and adjusted R^2 are higher than the model in Exercise 14.

17. Influential traffic?

Laredo has high leverage, but it does not have a particularly high Studentized residual. It appears that the point has influence but does not exert it. Removing this case from the regression may not result in a large change in the model, so the case is probably not influential.

18. The final slice.

a) The coefficient for the indicator for Weight Watchers is not significantly discernible from zero at the 0.05 level, but with a P-value of 0.09, it may still improve the model. This model has a slightly higher R^2 and adjusted R^2, but it is not enough improved to be grounds for choosing between the models. But the t-ratios are larger and P-values smaller for a number of the coefficients. That is a sign of improvement.

b) Looking at the other coefficients in the model (and especially at the coefficient for *calories* –not too surprising, considering the identity of the newly isolated pizza), the addition of an indicator for the Weight Watcher's pizza has made several of them more significantly different from zero. This seems to be a cleaner model and one that might lead to a better understanding.

c) The tasters score cheese pizzas substantially higher than pepperoni pizzas. Even after allowing for that, additional fat lowers scores, but higher calories pizzas score a bit better.