Section 2 Answer Key:

0) Find the median and quartiles of each of the following sets of numbers. These represent the "four cases" that you should be able to compute using the rules in this course.

(a) 23, 35, 28, 33, 5, 12, 40, 25, 20, 18, 1, 16

1, 5, <u>12, 16</u>, 18, 20 23, 25, <u>28, 33</u>, 35, 40

n = 12 M = (20 + 23)/ 2 = 21.5

Q1 = (12 + 16) / 2 = 14 Q3 = (28 + 33) /2 = 30.5

(b) 22, 33, 25, 28, 5, 12, 35, 23, 20, 18, 1, 40, 16

1, 5, <u>12, 16</u>, 18, 20 | 22 | 23, 25, <u>28, 33</u>, 35, 40

n = 13 M = 22 Q1 = 14 Q3 = 30.5

(c) 20, 28, 23, 25, 3, 5, 33, 22, 18, 16, 40, 1, 35, 12

1, 3, 5, <u>12,</u> 16, 18, 20, 22, 23, 25<u>, 28</u>, 33, 35, 40

n = 14 M = (20 + 22)/2 = 21 Q1 = 12 Q3 = 28

(d) 20, 28, 23, 25, 3, 5, 30, 22, 18, 40, 16, 35, 1, 33, 12

1, 3, 5, <u>12</u>, 16, 18, 20 | 22 | 23, 25, 28, <u>30</u>, 33, 35, 40

n = 15 M = 22 Q1 = 12 Q3 = 30

1) A certain test is used to measure the reading ability of children. Here are the scores of 44 third grade students.

40	26	39	14	42	18	25	43	56	27	19
47	19	26	35	34	15	44	40	38	31	46
52	25	35	35	33	28	34	41	49	28	52
47	35	48	22	33	41	51	27	14	54	45

(a) Find the mean, median, five number summary, IQR, and standard deviation of the children's reading ability.

$$\sum x = x_1 + x_2 + \cdots + x_{44} = 40 + \cdots + 45 = 1553$$

$$\sum x^2 = {x_1}^2 + {x_2}^2 + \cdots + {x_{44}}^2 = 40^2 + \cdots + 45^2 = 60527$$

$$\bar{x} = \frac{1553}{44} = 35.30$$

$$s = \sqrt{\frac{(\Sigma x^2) - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{60527 - \frac{(1553)^2}{44}}{43}} = 11.53$$

Ordering the numbers, the average of the middle two blanks is 35. 22 numbers on each side means that the quartiles (26.5 and 34) are also averages of two numbers. Five number summary is {14,26.5,35,44.5,56}. IQR = 44.5 – 26.5 = 18.

(b) The test scores are redesigned so that the previous scores are doubled, and everyone gets an additional 10 points added in. How will the quantities in (a) be affected, without recomputing?

The unit change is $y = 2x + 10$. Mean, median, min, max, quartiles will be run through this formula. Standard deviation and IQR will only be multiplied by 2.

2) Refer to problem 1.4:

Statistics Exam Scores:

60	47	82	95	88	72	67	66	68	98	90
77	86	58	64	95	74	72	88	74	77	39
90	63	68	97	70	64	70	70	58	78	89
44	55	85	82	83	72	77	72	86	50	94
92	80	91	75	76	78					

(a) Find the mean, median, five number summary, IQR, and standard deviation of the statistics exam scores. According to the rule of thumb we use, are there any outliers?

$$\sum x = 60 + \cdots + 78 = 3746$$

$$\sum x^2 = 60^2 + \cdots + 78^2 = 290524$$

$$\bar{x} = \frac{4746}{50} = 74.92$$

$$s = \sqrt{\frac{290524 - \frac{(3746)^2}{50}}{49}} = 14.2$$

Remember that we have a stemplot, so we've already ordered them. Five number summary is {39,67,75.5,86,98}, so IQR = 86 – 67 = 19.

Outliers are (1) more than 86 + 1.5*IQR = 114.5 or (2) less than 67 – 1.5*IQR = 38.5, so no outliers.

(b) Suppose the teacher adds five points to everyone's scores. What effect will this have (without recomputing) on the quantities in (a)?

The unit change is $y = x + 5$. Mean, median, min, max, quartiles will be run through this formula. Standard deviation and IQR will be unaffected.

3) Refer to problem 1.6:

(a) Find the mean, median, five number summary, IQR, and standard deviation of the physical-fitness scores.

Third graders at Roth Elementary School were given a physical-fitness test. Their scores were:

12	22	6	9	2	9	5	9	3	5	16	1	22	18
6	12	21	23	9	10	24	21	17	11	18	19	17	5
14	16	19	19	18	3	4	21	16	20	15	14	17	4
5	22	12	15	18	20	8	10	13	20	6	9	2	17
15	9	4	15	14	19	3	24						

Remember that we have a stemplot, so we've already ordered them. Five number summary is {1,7,14,18.5,24}, so IQR=18.5 – 7 = 11.5.

$$\sum x = 12 + \cdots + 24 = 832$$

$$\sum x^2 = 12^2 + \cdots + 24^2 = 13572$$

$$\bar{x} = \frac{4746}{64} = 13$$

$$s = \sqrt{\frac{13572 - \frac{(832)^2}{64}}{63}} = 6.6$$

(b) Are there any outliers according to our "rule of thumb"?

Outliers are (1) more than 18.5 + 1.5*11.5 = 35.75 or (2) less than 14 – 1.5*11.5R = -3.25, so no outliers.

(c) Replacing 24 with a larger number will not affect either the third quartile or the IQR, so what is the smallest whole number with which we could replace 24 and have it be considered an outlier?

We have seen that any number *larger* than 35.75 would be considered an outlier. 36.8, for instance, would be considered an outlier. But the smallest *whole* number which would be considered an outlier is 36.

4. The following are the golf scores of 12 members of a women's golf team.

89	90	87	95	86	81	102	105	83	88	91	79

(a) Compute the mean, median, five number summary, IQR, and standard deviation of the scores. Are there any outliers, according to our rule of thumb?

Five number summary is {79,84.5,88.5,93,105}, so IQR = 93– 84.5 = 8.5.

Outliers are (1) more than 93 + 1.5*IQR = 105.75 or (2) less than 84.5 – 1.5*IQR = 71.75, so no outliers.

$$\sum x = 89 + \cdots + 79 = 1076$$

$$\sum x^2 = 89^2 + \cdots + 79^2 = 97156$$

$$\bar{x} = \frac{4746}{12} = 89.7$$

$$s = \sqrt{\frac{97156 - \frac{(1076)^2}{12}}{11}} = 7.83$$

(b) If the 105 is replaced by 200, how will the mean, median, standard deviation, and IQR be affected?

The mean and standard deviation will be affected by the replacement (mean will be greater because the 200 will "pull" it, and standard deviation will be greater because a more extreme replacement has been made; the "average distance from the center" will increase. Replacing the largest of 12 numbers with a different one which is still the largest of 12 numbers will not affect median or IQR.

5. Last year a small accounting firm paid each of its five clerks $25,000, two junior accountants $60,000 each, and the firm's owner $255,000. Find the mean and median salaries at this firm. How many employees earn less than the mean? Find the standard deviation and IQR of the salaries. If next year everyone's salaries are tripled, and they're given a $1000 bonus, how will this affect the mean, median, IQR, and standard deviations of the salaries?

Data is

25000	25000	25000	25000	25000	60000	60000	255000

Five number summary is {25000, 25000, 25000, 60000, 255000}, so IQR = 60000– 25000 = 35000.

$$\sum x = 25000 + \cdots + 255000 = 500{,}000$$

$$\sum x^2 = 25000^2 + \cdots + 255000^2 = 75{,}350{,}000{,}000$$

$$\bar{x} = \frac{500{,}000}{8} = 62{,}500$$

$$s = \sqrt{\frac{75{,}350{,}000{,}000 - \frac{(500{,}000)^2}{8}}{7}} = 79{,}372.54$$

Seven out of eight employees earn less than the mean (the obvious outlier pulls the mean up considerably).

The unit change is $y = 3x + 1000$. Mean, median, min, max, quartiles will be run through this formula. Standard deviation and IQR will only be multiplied by 3.

Notice that these calculations can be greatly simplified by doing everything in “thousands of dollars” rather than “dollars”; i.e., a unit change. In this case, of course, the final formula would be $y = 3x + 1$, since the quantities on both sides are “thousands of dollars”.

6. In each case, find a formula for the linear transformation from X to Y and answer the questions.

(a) You're writing a report on the power of car engines. Your sources use horsepower X. Your professor requires you to re-express power in watts Y. One horsepower is 746 watts. What is the power in watts of a 140 horsepower engine?

$$140\ \text{\sout{hp}} * \frac{746\ \text{watts}}{1\ \text{\sout{hp}}} = 104{,}440\ \text{watts}$$

$$Y = 746 * X$$

(b) Convert speed X in miles/hour to into Y, kilometers/hour. (1 km = 0.62 mile) What is 65 miles per hour in km/hr's? What is 82 km/hour in miles/hr?

In one hour, we travel:

$$65\ \text{\sout{miles}} * \frac{1\ \text{km}}{0.62\ \text{\sout{mile}}} = 104.65\ \text{km}$$

$$Y = \frac{X}{0.62} = 1.61 * X, \text{which looks nicer.}$$

$$82\ \text{\sout{km}} * \frac{0.62\ \text{mile}}{1\ \text{\sout{km}}} = 50.93\ \text{miles}$$

(c) The recommended daily allowance (RDA) of vitamin C is 30 milligrams. Convert milligrams of vitamin C (X) into Y, percent of RDA.

0 mg would be 0 RDA.

30 mg would be the total proportion of RDA (1).

15 mg would be half the RDA (0.5).

60 mg would be twice the RDA (2), etc.

There is a change of scale (*, ÷) but no shift (+, –). To convert mg into RDA we're simply dividing by the total (30):

$$prop\ RDA = X/30$$

To convert to percent, we're multiplying by 100:

$$percent\ RDA = 100 * prop\ RD = 100 * \frac{X}{30} = \frac{10}{3}X$$

Note that as usual, percents are easier to talk about informally, but more confusing in formulas, so we tend to use percents only in casual conversation, not calculations.

(d) High school students can take SAT's and ACT's, two different standardized tests. The two exams use two very different grading scales, and yet we can compare performance by a rule of thumb: multiply an ACT score by 40 and add 150 points to estimate the equivalent SAT score. If a large number of ACT scores has five number summary 19, 24,28,30,36, mean 27, standard deviation 3, find the corresponding quantities on the SAT exam. Find a formula for converting SAT scores to ACT scores.

We are told that:

$$X = ACT, \qquad Y = SAT$$

$$Y = 40X + 150$$

Mean, median, min, max, quartiles will be run through this formula. Standard deviation and IQR will only be multiplied by 40. “”

For instance, if you want to verify for IQR:

$$IQR_Y = (40 * 30 + \cancel{150}) - (40 * 24 + \cancel{150}) = 40(30 + 24) = 40 * IQR_X$$

Solving for X, we get:

$$Y - 150 = 40X \quad \text{so} \quad \frac{Y - 150}{40} = X$$

In $“a + b * variable”$ form, if you like, this is:

$$X = \frac{Y}{40} - \frac{150}{40} = \frac{1}{40}Y - 3.75$$

7. Give an example of a set of twelve numbers (distribution) with IQR = 0, min = 20, max = 50.

We require 12 ordered blanks with:

20											50

The median will be the average of:

20					***	***					50

The quartiles will be the average of:

20		***	***					***	***		50

For the $IQR = Q_3 - Q_1 = 0,$ we need $Q_3 = Q_1$.

This must mean that all four of these blanks are the same.

Since they're in order, all the intervening blanks must be the same. This can be any particular number between 20 and 50, inclusive. For instance,

20		32	32	32	32	32	32	32	32		50

The remaining blanks can be filled with anything as long as the numbers remain in order. For instance,

20	25	32	32	32	32	32	32	32	32	47	50

In passing, notice that median and IQR give plenty of room for alternative examples (compare #8, #11).

8. Give an example of a set of twelve numbers with standard deviation = 0, min = 28

We require 12 ordered blanks with:

28											

Standard deviation ("average distance from the center") being exactly 0 implies that every distance from the center is 0; i.e., that all the numbers are the same. (The moment two numbers are different, there is a strictly positive "average distance from the center"). So the list must be:

28	28	28	28	28	28	28	28	28	28	28	28

Notice that specifying conditions on mean and/or standard deviation gives you a lot less room to maneuver (compare #7).

9. Consider the coin distribution in 1.3 above. Which would be more useful to summarize the results from many, many coins: the mean and standard deviation, or the median and interquartile range? Which will be higher, the mean or median coin year?

The mean/standard deviation "balance" analysis has the first claim. If there is something about the list that will render it suspect (outliers, long tails/skewed, generally asymmetric) then that's the reason we would prefer the median, IQR, Five Number Summary "order" analysis. Since this example was very left-skewed, we'll go with the latter. The long tail toward smaller numbers will pull down the mean compared to the median. A coin from ancient Rome will considerably pull down the average year; one more coin in the small side of the median has very little or no effect.

10. The ages in years of everyone who visits a certain shopping mall on Saturday afternoon is strongly right-skewed. (a) How will this characteristic be affected if we measure in months? (b) Which will be higher, the mean or median age, or will they be the same?

Unit changes don't affect the generally shape of a distribution, only how pronounced it physically looks. Since months result in bigger numbers than years do, the long tail to the right will look longer. But bear in mind that this is very cosmetic; how far apart the "tick marks" are placed when you're drawing it is a very arbitrary decision. The general shape (whether it possesses the various characteristics from section 1, unimodal, bimodal, symmetric, right-skewed, left-skewed, uniform) is not affected.

Since it is right-skewed, the mean will be pulled up by the tail, the median relatively unaffected by data that is very far away. So the mean is higher.

11. Give an example of a distribution of numbers with mean 15 and standard deviation 2

This is going to be very hard to do by trial and error. But we can start with any set of numbers we like and think what unit change to run it through (shift, rescale) to get what we want. The math is easiest if we start with the simplest mean (0) and the simplest standard deviation (1). This is achieved by the numbers

				mean	st. dev.
X	−1	0	1	0	1

This mean and standard deviation are easily checked.

Remember that shifts (+, – numbers) do not affect standard deviation. So once the standard deviation is correct, we can re-center without changing it. So get the standard deviation correct first. Multiplication has the obvious affect on standard deviation, so we'll double every number to get the correct standard deviation:

				mean	st. dev.
X	−1	0	1	0	1
2X	−2	0	2	2*0 = 0	2*1 = 2

Now we can re-center. Adding a number will have the obvious effect on mean but no effect on standard deviation. So we'll add 15 to change the mean to 15:

				mean	st. dev.
X	−1	0	1	0	1
2X	−2	0	2	2*0 = 0	2*1 = 2
2X+15	−13	15	17	2*0+15 = 15	2*1 = 2