CHAPTER 11

THE CHI-SQUARE DISTRIBUTION

1. The Chi-Square Distribution

In this chapter we explore two types of hypothesis tests that require the **Chi-Square Distribution**, χ^2_{df}.

The Chi-Square distribution has only one parameter: df = degrees of freedom. The degrees of freedom depends on the application, as we will see later. Here are a few facts about the Chi-Square distribution. If $X^2 \sim \chi^2_{df}$ the following are true of X^2:

- X^2 is a continuous random variable
- $X^2 = Z^2 + Z^2 + \cdots + Z^2$; X^2 is the sum of df independent squared standard normal random variables
- data values can't be negative: $x \in [0, \infty)$
- $\mu = df$ (the mean of the Chi-Square distribution is the degrees of freedom!)
- $\sigma = \sqrt{2 * df}$
- X^2 is skewed right
- the mean (df) is just to the right of the peak of the density curve
- when $df > 90$, X^2 is approximately normal

2. Goodness of Fit Test

We use the goodness of fit test to test if a discrete categorical random variable matches a predetermined "expected" distribution. The hypotheses in a goodness of fit test are

H_0 : the actual distribution fits the expected distribution
H_a : the actual distribution does not fit the expected distribution

REQUIREMENT: In order for a chi-square goodness of fit test to be appropriate, the expected value in each category must be at least 5. It may be possible to combine categories to meet this requirement.

Example 1. Fair die? (Part 1) Suppose we wish to test if a die is weighted. We roll the die 120 times and get the following "observed" results.

Roll	Observed	Expected
1	15	
2	29	
3	16	
4	15	
5	30	
6	15	

(1) What is the expected distribution of the 120 die rolls? Complete the table.
(2) Is the requirement for a chi-square goodness of fit test satisfied? Explain.
(3) Write the null and alternative hypotheses for a goodness of fit test.

(4) I can see that the rolls didn't come out even. What's the point of completing the test?

Chapter 11 Homework #72. (Barbara Illowsky & Susan Dean. Introductory Statistics. OpenStax College, 2013.)

Our goal is to see if the observed values are close enough to the expected values that the differences could be due to random variation or, alternatively, if the differences great enough that we can conclude that the distribution is not as expected. Therefore, our sample statistic (which is also the test statistic in this case) should provide a measure of how far the from "expected" frequencies the "observed" frequencies are, as a group. The **test statistic** for a goodness of fit test is:

$$\sum \frac{(O-E)^2}{E},$$

where O = observed frequency, E = expected frequency, and the sum is taken over all the categories.

Example 2. FAIR DIE? (PART 2) Continuing Example 1, find the test statistic to test if the die is weighted. (Do this by hand using a chart and using the calculator lists.)

The test statistic follows a chi-square distribution, χ^2_{df}, with

$$df = \text{number of categories} - 1.$$

Example 3. FAIR DIE? (PART 3) Continue the same example in which we wish to test if the die is weighted.

(1) Find the distribution of the test statistic.
(2) Find the mean of the distribution.
(3) Find the standard deviation of the distribution.
(4) Sketch the density curve of the distribution.

Studying the test statistic formula, the bigger the differences between the observed and expected frequencies, the larger the test statistic. Since the differences are squared and the expected frequencies are always positive, the test statistic is always 0 or positive. The farther beyond the mean the test statistic is, the more evidence we have that the distribution is not as expected.

The **p-value** is the probability of getting the test statistic or one that is even bigger, which is the area in the right tail of the χ^2_{df} distribution. The calculator function χ^2cdf in the DISTR menu that calculates area under the chi-square distribution:

$$\text{area under the } \chi^2_{df} \text{ density curve between } a \text{ and } b = \chi^2\text{cdf}(a, b, df).$$

$$\begin{aligned} \text{p-value} &= \text{P}(x \geq \text{test statistic} \| the actual distribution fits the expected distribution) \\ &= \chi^2\text{cdf}(test statistic, 10^9, df) \end{aligned}$$

Lower p-values indicate test statistics that are farther from the value assumed in the null hypothesis, therefore providing more evidence that the actual distribution does not fit the expected distribution. How low of a p-value is low enough to conclude there is statistical evidence to support the alternative hypothesis? For this we compare the p-value to the significance level, α. As usual:

If the p is low, the null must go,
if the p is high, the null will fly.

(If $\alpha >$ p-value, reject the null hypothesis,
if $\alpha \leq$ p-value, we can't reject the null hypothesis.)

Example 4. Fair die? (Part 4) Continue the same example in which we wish to test if the die is weighted.

(1) Shade the area representing the p-value on the χ^2_{df} sketch from Example 3.
(2) Find the p-value for this test.
(3) Should we reject or fail to reject the null hypothesis?
(4) Is there sufficient statistical evidence to conclude that the die is weighted?

Example 5. Conduct a hypothesis test to determine if the actual majors of graduating females fit the expected distribution of their majors. The observed data were collected from 5,000 graduating females. Complete a hypothesis test at the $\alpha = 0.05$ significance level to test if the actual distribution of female students to majors matches the expected distribution.

Major	Expected %	Observed Frequency	Expected Frequency
Arts & Humanities	14.0%	670	
Biological Sciences	8.4%	410	
Business	13.1%	685	
Education	13.0%	650	
Engineering	2.6%	145	
Physical Sciences	2.6%	125	
Professional	18.9%	975	
Social Sciences	13.0%	605	
Techical	0.4%	15	
Other	5.8%	300	
Undecided	8.0%	420	

(1) Find the expected frequencies and complete the table.
(2) Are the requirements for a chi-square goodness of fit test satisfied? Explain and adjust the categories if needed.

(3) Write the null and alternative hypotheses.

(4) Find the test statistic.

(5) What is the distribution?

(6) Find the p-value.

(7) Sketch the density curve. Label the mean, the test statistic, and the p-value.

(8) Is there sufficient evidence to conclude that the distribution of majors is not as expected?

Chapter 11 Homework 78. (Barbara Illowsky & Susan Dean. Introductory Statistics. OpenStax College, 2013.)

3. Test of Independence

We apply a Test of Independence to test if two factors are independent. Again the data are categorical with multiple categories. In fact, the data are grouped accoring to two category types, each with multiple categories. The frequency data are organized in contingency tables with rows representing one category type and columns representing the other category type. In a test for independence, we test whether or not the row categories and column categories are independent of each other.

The null and alternative hypotheses are written out with words and always follow this pattern, although the exact wording will change based on the scenario:

H_0 : the row and column categories are independent
H_a : the row and column categories are dependent

Example 6. Independent? (Part 1) In a volunteer group, adults 21 and older volunteer from one to nine hours each week to spend time with a disabled senior citizen. The program recruits among community college students, four-year students, and nonstudents. The following table categorizes 839 volunteers according to volunteer type and number of hours worked.

OBSERVED	1-3 hours	4-6 hours	7-9 hours	Row total
Community college students	111	96	48	255
Four-year college students	96	133	61	290
Nonstudents	91	150	53	294
Column total	298	379	162	839

We are interested in whether or not the number of hours worked depends on the volunteer type. List the null and alternative hypotheses for a test for independence:

H_0 :
H_a :

Example 11.6. (Barbara Illowsky & Susan Dean. Introductory Statistics. OpenStax College, 2013.)

We will use the same **test statistic** as for goodness of fit tests:

$$\sum \frac{(O-E)^2}{E},$$

where O = observed frequency, E = expected frequency, and the sum is taken over all categories in the table.

We have been given the observed frequencies, but we have to calculate the expected frequencies. To do this, we will need to recall that two events A and B are independent if $P(A \text{ AND } B) = P(A)P(B)$. For example, if I want to find the expected number of volunteers who are community college students (A) AND worked 7-9 hours (B) ASSUMING the number of hours worked is independent of volunteer type, then I can calculate,

$$\begin{aligned}\text{expected number of A AND B volunteers} &= P(A \text{ AND } B) * (\text{n}) \\ &= P(A)P(B) * (\text{n}) \\ &= \tfrac{\text{A row total}}{\text{n}} * \tfrac{\text{B row total}}{\text{n}} * (\text{n}) \\ &= \tfrac{(\text{number of cc students})(\text{number of 7-9 hr workers})}{\text{total number of volunteers}},\end{aligned}$$

where n = the total number of volunteers.

The general formula for the expected frequency in row i, column j is

$$E = \frac{(\text{row i total})(\text{column j total})}{n}$$

where $n =$ the sample size.

Example 7. Independent? (Part 2) Complete the table of expected values corresponding to Example 6.
Expected frequencies if number of hours worked is idependent of volunteer type:

EXPECTED (assume indep.)	1-3 hours	4-6 hours	7-9 hours	Row total
Community college students				255
Four-year college students				290
Nonstudents				294
Column total	298	379	162	839

REQUIREMENT for test for independence: Each *expected* frequency must be at least 5.

Example 8. Independent? (Part 3) Referring to the test for independence we began in Example 6:

(1) Is the requirement for test for independence satisfied? Explain.
(2) Calculate the test statistic using calculator lists. It will be convenient to list all the observed frequencies in L1 and all of the corresponding expected frequencies in L2. Then we can compute the test statistic exactly as we did for the goodness of fit test.

Is it easy to see that big values of the test statistic correspond to big differences between observed frequencies and the frequencies we would expect in the case of independence. Therefore, a big test statistic leads us to support the alternative hypothesis (categories are dependent). But how big is big? Again we will use a p-value and compare it to α to decide whether or not to reject the null hypothesis:

$$\begin{aligned} \alpha > \text{p-value} \quad &\Leftrightarrow \quad \text{BIG test statistic} \\ &\Leftrightarrow \quad \text{reject null (reject independence)} \\ &\Leftrightarrow \quad \text{support alternative (support dependence)} \end{aligned}$$

As we know from the last section, this test statistic follows a chi-square distribution: χ^2_{df}. The degrees of freedom formula for a test for independence is different, however:

$$\text{degrees of freedom } (df) = (\text{number of rows } -1)(\text{number of columns } -1).$$

Again the p-value is the right tail probability in the Chi-square distribution:

$$\text{p-value} = \chi^2\text{cdf}(\text{test statistic}, 10^9, df)$$

Example 9. Independent? (Part 4) Complete the same test for independence. Use a significance level of $\alpha = 0.05$.

(1) State the distribution followed by the test statistic.
(2) Find the p-value.
(3) State the conclusion of the test. (Is there statistical evidence that the number of hours worked depends on the volunteer type?)

There is a calculator function that will compute the test statistic and the p-value for a χ^2 test of independence. First enter the matrix of observed values: use MATRIX, EDIT, and select [A]. After entering the table of observed values (not including row and column totals), use STAT, Tests, C: χ^2-Test. As you can see, you have the option of entering a table of expected frequencies. You can do this, but you don't need to. For the case of a test of independence, the expected frequencies will be calculated automatically as the default.

Example 10. INDEPENDENT? (PART 5) Use MATRIX and χ^2-Test on your calculator to find the test statistic and p-value for the test of independence we just completed.

Example 11. TREATING STRESS FRACTURES. With respect to stress fractures in a foot bone, does the success rate of the treatment depend on the treatment method, or do all methods of treatment have basically the same success rate? Use the following data and a significance level of $\alpha = 0.01$ to complete a test of independence.

OBSERVED	Success	Failure	Row total
Surgery	54	12	66
Wt-bearing cast	41	51	92
Non-wt-bearing cast 6 weeks	70	3	73
Non-wt-bearing cast $<$ 6 weeks	17	5	23
Column total	182	71	253

(1) State the null and alternative hypotheses for this test of independence.

(2) Complete the table of expected values assuming the success rate is independent of the treatment method. Use two decimal places of accuracy.

EXPECTED (assuming ind.)	Success	Failure	Row total
Surgery			66
Wt-bearing cast			92
Non-wt-bearing cast 6 weeks			73
Non-wt-bearing cast $<$ 6 weeks			23
Column total	182	71	253

(3) Is the requirement for a test of independence satisfied?

(4) Find the distribution of the test statistic, including the degrees of freedom.

(5) Calculate the test statistic using your preferred method.

(6) Calculate the p-value using your preferred method.

(7) Sketch the density curve, marking and labeling the test statistic and p-value.

(8) What is the outcome of the test for independence? (Can we conclude that the success rate depends on the method of treatment or not?

(Elementary Statistics, Mario Triola.)