Chapter 14

Chi-Square Procedures

Objectives:

- Explain what is meant by a chi-square goodness of fit test.
- Conduct a chi-square goodness of fit test.
- Given a two-way table, compute conditional distributions.
- Conduct a chi-square test for homogeneity of populations,
- Conduct a chi-square test for association/independence.
- Use technology to conduct a chi-square significance test.

Case Closed: Does acupuncture promote pregnancy?

A study reported in the medical journal *Fertility and Sterility* in 2002 sought to determine if the ancient Chinese art of acupuncture could help infertile women become pregnant.[1] One hundred sixty healthy women undergoing treatment with in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) were recruited for the study. The purpose of the study was to determine if acupuncture improves the clinical pregnancy rate after IVF or ICSI treatment.

Only patients with good embryo quality were accepted into the study. The ages of the subjects ranged from 21 to 43. The cause of the infertility was the same for both groups. Half of the subjects (80) were randomly assigned to an experimental (acupuncture) treatment group, and the remaining 80 were assigned to a control group.

The subjects in the treatment/acupuncture group received acupuncture treatment 25 minutes before embryo transfer and again 25 minutes after the transfer. Subjects in the control group were instructed to lie still for 25 minutes after the embryo transfer. Results are tabulated in the table below.

	Acupuncture group	Control group
Pregnant	34	21
Not pregnant	46	59
Total	**80**	**80**

When you finish this chapter, you will be able to apply an appropriate inference procedure to answer the question the researchers were investigating.

In the previous chapter, we discussed inference procedures for comparing two population proportions. Sometimes we want to examine the distribution of proportions in a single population. The *chi-square test for goodness of fit* allows us to determine whether a specified population distribution seems valid. We can compare two or more population proportions using a *chi-square test for homogeneity of populations.* In doing so, we will organize our data in a two-way table. It is also possible to use the information provided in a two-way table to determine whether the distribution of one variable has been influenced by another variable. The *chi-square test of association/independence* helps us decide this issue.

The methods of this chapter help us answer questions such as these:

- Are you more likely to have a motor vehicle collision when using a cell phone?
- Does background music influence wine purchases?
- How does the presence of an exclusive-territory clause in a franchisee's contract relate to the success of the business

Test For Goodness of Fit

Example 1: Hypotheses for goodness of fit test

Are you more likely to have a motor vehicle collision when using a cell phone? A study of 699 drivers who were using a cell phone when they were involved in a collision examined this question. These drivers made 26,798 cell phone calls during a 14-month study period. Each of the 699 collisions was classified in various ways. Here are the counts for each day of the week:

Day:	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
Number:	20	133	126	159	136	113	12	699

We have a total of 699 accidents involving drivers who were using a cell phone at the time of their accident. Let's explore the relationship between these accidents and the day of the week. Are the accidents equally likely to occur on any day of the week?

We can think of this table of counts as a ***one-way table*** with seven ***cells,*** each with a count of the number of accidents that occurred on the particular day of the week.

What are the appropriate hypotheses for the distribution?

H_0: Accidents are equally likely to occur on the day of the week

H_a: The probabilities of accidents vary from day to day

The idea of the goodness of fit test is this: we compare the observed counts for our sample with the counts that would be expected if the number of motor vehicle accidents involving cell phones occurred uniformly throughout the week. The uniform distribution of accidents is the population. The more the observed counts differ from the expected counts, the more evidence we have to reject H_0 and conclude that the probabilities of an accident involving cell phone use are not the same for each day of the week.

Before actually carrying out the χ^2 test for goodness of fit, make a graph of the data. Categorical data are very often displayed by a bar graph. In this case, a double bar graph comparing the observed vs. expected counts would be appropriate.

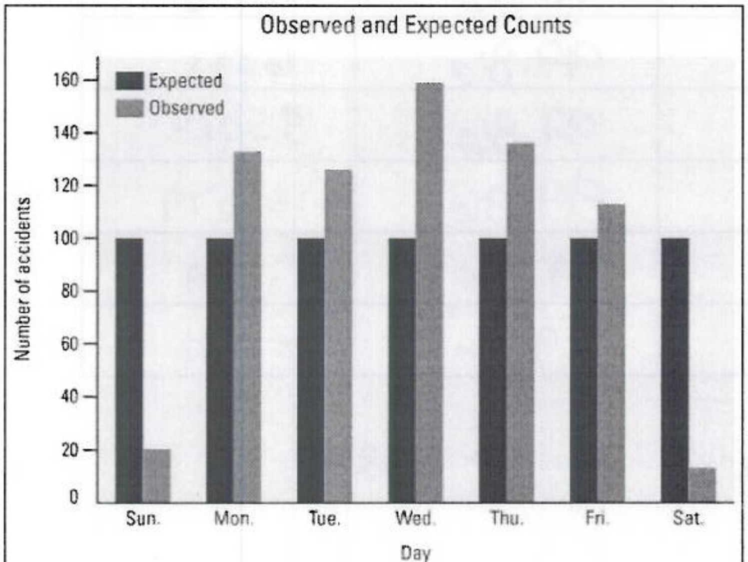


Figure 14.1 ***Bar graph for the data in*** *Example 14.2.*

Expected Count:

The expected count of any categorical variable is found by multiplying the proportion of the distribution for each category by the sample size.

Chi-Squared Statistic:

$$\sum \frac{(\text{observed} - \text{Expected})^2}{\text{Expected}}$$

Example 2: Find the chi-squared statistic for the cell phone example above.

Day	Observed count, O	Expected Count, E	$\frac{(O-E)^2}{E}$
Sunday	20	99.86	63.8766
Monday	133	99.86	10.998
Tuesday	126	99.86	6.843
Wednesday	159	99.86	35.024
Thursday	136	99.86	13.079
Friday	113	99.86	1.729
Saturday	12	99.86	77.362

Chi-squared (X^2) is the sum of the last column. χ^2 208.84

The larger the differences between the expected and observed outcomes, the larger X^2 will be and the more evidence against the null we will have.

Like the t-distribution, there is an entire family of curves called the chi—squared family. We use the degrees of freedom to determine which curve to use in order to assess the the evidence against the null hypothesis.

Degrees of freedom (df): Number of categories -1

The chi-square test applied to the hypothesis that a categorical variable has a specified distribution is called the ***test for goodness of fit***. The idea is that the test assesses whether the observed counts "fit" the hypothesized distribution. Here are the details.

The Chi-Square Test for Goodness of Fit

A **goodness of fit test** is used to help determine whether a population has a certain hypothesized distribution, expressed as proportions of individuals in the population falling into various outcome categories. Suppose that the hypothesized distribution has *k* outcome categories. To test the hypothesis

H_0: the actual population proportions are *equal to* the hypothesized proportions first calculate the **chi-square test statistic**

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} = \sum \frac{(O-E)^2}{E}$$

Then, X^2 has approximately a χ^2 distribution with $(n - 1)$ degrees of freedom.

For a test of H_0 against the alternative hypothesis:

H_a: at least two of the actual population proportions *differ from* their hypothesized proportions the *P*-value is $P(\chi^2 \geq X^2)$.

Conditions: You may use this test with critical values from the chi-square distribution when all individual **expected counts are at least 1 and no more than 20% of the expected counts are less than 5.**

Notice the conditions for using the chi-square goodness of fit test. Look again at the difference terms

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

We don't want to divide by zero, and since we're working with *counts,* we require that *all expected counts* be greater than zero. **In checking conditions, remember that it's the expected counts that are critical, not the observed counts.** In our example note that all of the expected cell counts are greater than 5.

<u>Properties of the Chi-squared Distribution</u>

<u>Chi-Square Distributions:</u>

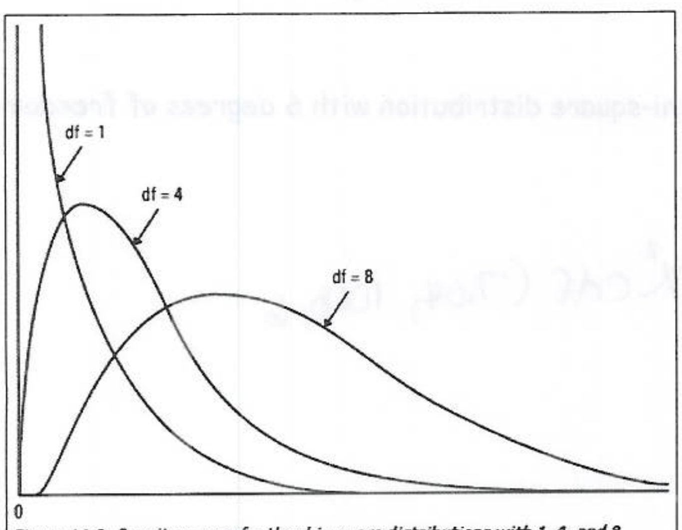


Figure 14.2 *Density curves for the chi-square distributions with 1, 4, and 8 degrees of freedom. Chi-square distributions take only positive values.*

The **chi-square distributions** are a family of distributions that take only positive values and are skewed to the right. A specific chi-square distribution is specified by one parameter, called the *degrees of freedom.*

Chi Square density Curves:

1. The total area under a chi-square curve is equal to 1.

2. Each chi-square curve (except when degrees of freedom = 1) begins at 0 on the horizontal axis, increases to a peak, and then approaches the horizontal axis asymptotically from above.

3. Each chi-square curve is skewed to the right. As the number of degrees of freedom increase, the curve becomes more and more symmetrical and looks more like a Normal curve.

To find P-Value: P-value is the area to the right of the χ^2 Statistic

Finding P-Values

You can use the TI calculator to find your P-value for a X^2statistic just as we do for a t or z statistic. The command is in the distribution menu and is X^2cdf(lower bound, upper bound, degrees of freedom)

Example 3: (a) Find the *P*-value corresponding to X^2 = 1.41 for a chi-square distribution with 1 degree of freedom with your graphing calculator.

χ^2cdf(1.41, 1000, 1)

P-value: .2351

(b) Find the area to the right of X^2 = 19.62 under the chi-square curve with 9 degrees of freedom with your graphing calculator.

Area: .0204 χ^2cdf (19.62, 1000, 9)

(c) Find the *P*-value corresponding to X^2 = 7.04 for a chi-square distribution with 6 degrees of freedom with your graphing calculator.

P-Value: .3172 χ^2cdf (7.04, 1000, 6

Example 4: Goodness of fit test

Biologists wish to mate two fruit flies having genetic makeup RrCc, indicating that each has one dominant gene (R) and one recessive gene (r) for eye color, along with one dominant (C) and one recessive (c) gene for wing type. Each offspring will receive one gene for each of the two traits from each parent. The following table, often called a Punnett square, shows the possible combinations of genes received by the offspring:

		Parent 2 passes on			
		RC	Rc	rC	rc
Parent 1 passes on	RC	RRCC (x)	RRCc (x)	RrCC (x)	RrCc (x)
	Rc	RRCc (x)	RRcc (y)	RrCc (x)	Rrcc (y)
	rC	RrCC (x)	RrCc (x)	rrCC (z)	rrCc (z)
	rc	RrCc (x)	Rrcc (y)	rrCc (z)	rrcc (w)

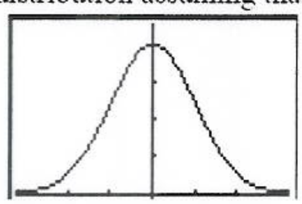
Any offspring receiving an R gene will have red eyes, and any offspring receiving a C gene will have straight wings. So based on this Punnett square, the biologists predict a ratio of 9 red-eyed, straight-winged (x) : 3 red-eyed, curly-winged (y) : 3 white-eyed, straight-winged (z) : 1 white-eyed, curly-winged (w) offspring. To test their hypothesis about the distribution of offspring, the biologists mate the fruit flies. Of 200 offspring, 99 had red eyes and straight wings, 42 had red eyes and curly wings, 49 had white eyes and straight wings, and 10 had white eyes and curly wings. Do these data differ significantly from what the biologists have predicted?

We return to the familiar structure of the Inference Toolbox to carry out the significance test.

type of offspring

	x	y	z	w
proportion	$9/16$	$3/16$	$3/16$	$1/16$
Observed:	$99/200$	$42/200$	$49/200$	$10/200$
expected:	$9/16 \cdot 200$	$3/16(200)$	$3/16(200)$	$1/16(200)$
	112.5	37.5	37.5	12.5

Population: Fruit Flies

1.	Parameter of Interest:	proportion of Flies in each Category	
2.	Choice of test:	χ^2 Goodness of Fit	
3.	Check of conditions:	Statement of necessary conditions: All expected counts are at least 5.	Verification of satisfaction:
4.	Null Hypothesis:	H_0: (English) H_0: (symbols) $p_x = .5625$ $p_y = 3/16$ $p_z = 3/16$ $p_w = 1/16$	
5.	Alternative Hypothesis:	H_a: (English) H_a: (symbols) At least one p is incorrect	
6.	Test Statistic:	Formula: $\chi^2 = 6.187$	Value:
7.	Test: Level of Significance α = ________	Sketch of sampling distribution assuming that H_0 is true.  Identify the location of the test statistic in the sketch and shade the appropriate region for the p-value	

8.	p-value: df = 3	Exact p-value: .1023	Reconciliation with Critical Value of Rejection:
9.	Recommended Decision:	Regarding significance: Not significant at even α = .10	Regarding H_0: Fail to reject
10.	Interpretation: (English)	the chance of our sample or one more extreme is .1023 if the hypothesized values are correct this is Not good evidence to reject the biologists predicted distribution.	

In the chi-square test for goodness of fit, we test the null hypothesis that a categorical variable has a specified distribution. If we find significance, we can conclude that our variable has a distribution different from the specified one. In this case, it's always a good idea to determine which categories of the variable provide the greatest differences between observed and expected counts. To do this, look a the individual terms $(O - E)^2/E$ that are added together to produce the test statistic X^2.

In the above example, the offspring contributing the largest amount to the X^2 statistic are the white-eyed, straight-winged flies (3.5267). This is known as the largest ***component*** of the chi-square statistic. In the cell phone example, the largest components of X^2 are the ones for Saturday (77.30) and Sunday (63.86).

Example 5: Most students in a large college statistics course are taught by teaching assistants (TAs). One section is taught by the course supervisor, a full-time professor. The distribution of grades for the hundreds of students taught by TAs this semester was

Grade:	A	B	C	D/F
Probability:	0.32	0.41	0.20	0.07

The grades assigned by the professor to the 91 students in his section were

Grade:	A	B	C	D/F
Count:	22	38	20	11

|91

(These data are real. We won't say when and where, but the professor was not any of the authors of this book.)

(a) What percents of students in the professor's section earned A, B, C, and D/F? In what ways does this distribution of grades differ from the TA distribution?

A	B	C	D/F
.242	.42	.22	.12

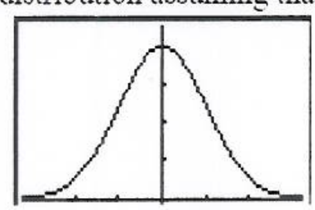
Less A's B's are about the same. C's are slightly Higher & D/F are slightly lower

(b) Because the TA distribution is based on hundreds of students, we are willing to regard it as a fixed probability distribution. If the professor's grading follows this distribution, what are the expected counts of each grade in his section?

A	B	C	D/F
29.12	37.31	18.2	6.37

(c) Does the chi-square test for goodness of fit give good evidence that the professor's grade distribution differs from the TA distributions? Use the Inference Toolbox.

Population Statistics Students

1.	Parameter of Interest:	p proportions of students grades	
2.	Choice of test:	Chi-Squared Goodness of fit	
3.	Check of conditions:	Statement of necessary conditions: all expected counts No more than 20%	Verification of satisfaction: are above 1 and are below 5
4.	Null Hypothesis:	H_0: (English) The distribution of grades are H_0: (symbols) $p_A = .32$ $p_B = .41$ $p_C = .2$ $p_D = .07$	
5.	Alternative Hypothesis:	H_a: (English) At least one of these proportions are different H_a: (symbols)	
6.	Test Statistic: $\chi^2 = 5.3$	Formula: $\chi^2 = \frac{(22-29.12)^2}{29.12} + \frac{(38-37.31)^2}{37.31} + \frac{(20-18.2)^2}{18.2} + \frac{(11-6.37)^2}{6.37}$	Value:
7.	Test: Level of Significance $\alpha =$ ________	Sketch of sampling distribution assuming that H_0 is true.  Identify the location of the test statistic in the sketch and shade the appropriate region for the p-value	

8.	p-value: df=3	Exact p-value: .1511	Reconciliation with Critical Value of Rejection:
9.	Recommended Decision:	Regarding significance: Not Sig	Regarding H_0: Fail to Reject
10.	Interpretation: (English)	if the proportions are like the TA's, then the professors distribution happens 15.11% of the time. This is Not	

Significant. It appears the professors distribution is similar to the TA's.

Inference for two-way tables

The two-sample z procedures of Chapter 13 allow us to compare the proportions of successes in two groups (either two populations or two treatment groups in an experiment). What if we want to compare more than two groups? We need a new statistical test. The new test starts by presenting the data as a two-way table. Two-way tables have more general uses than comparing the proportions of successes in several groups. As we saw in Chapter 4, they can be used to describe relationships between any two ategorical variables. The same test that compares several proportions also tests whether the row and column variables are related in any two-way table. We will start with the problem of comparing several proportions.

Example 6: Does background music influence wine purchases?

Conditional distributions

Market researchers know that background music can influence the mood and purchasing behavior of customers. One study in a supermarket in Northern Ireland compared three treatments: no music, French accordion music, and Italian string music. Under each condition, the researchers recorded the numbers of bottles of French, Italian, and other wine purchased.[5] Here is a table that summarizes the data:

	Music			
Wine	None	French	Italian	Total
French	30	39	30	99
Italian	11	1	19	31
Other	43	35	35	113
Total	84	75	84	243

The conditional distributions of types of wine sold for each kind of music being played and the marginal distribution of the types of wine sold are shown in the following table: