



# Chapter 11: Inference for Distributions of Categorical Data

Section 11.2

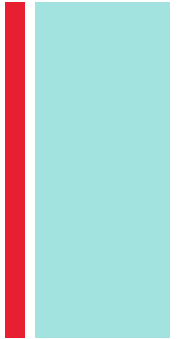
Inference for Relationships

The Practice of Statistics, 4<sup>th</sup> edition – For AP\*  
STARNES, YATES, MOORE



# Chapter 11

## Inference for Distributions of Categorical Data

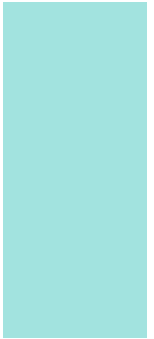


- **11.1** Chi-Square Goodness-of-Fit Tests
- **11.2** Inference for Relationships



# Section 11.2

## Inference for Relationships



### Learning Objectives

After this section, you should be able to...

- ✓ COMPUTE expected counts, conditional distributions, and contributions to the chi-square statistic
- ✓ CHECK the Random, Large sample size, and Independent conditions before performing a chi-square test
- ✓ PERFORM a chi-square test for homogeneity to determine whether the distribution of a categorical variable differs for several populations or treatments
- ✓ PERFORM a chi-square test for association/independence to determine whether there is convincing evidence of an association between two categorical variables
- ✓ EXAMINE individual components of the chi-square statistic as part of a follow-up analysis
- ✓ INTERPRET computer output for a chi-square test based on a two-way table

## ■ Introduction

The two-sample  $z$  procedures of Chapter 10 allow us to compare the proportions of successes in two populations or for two treatments. What if we want to compare more than two samples or groups? More generally, what if we want to compare the distributions of a single categorical variable across several populations or treatments? We need a new statistical test. The new test starts by presenting the data in a two-way table.

Two-way tables have more general uses than comparing distributions of a single categorical variable. They can be used to describe relationships between any two categorical variables.

- ✓ In this section, we will start by developing a test to determine whether the distribution of a categorical variable is the same for each of several populations or treatments.
- ✓ Then we'll examine a related test to see whether there is an *association* between the row and column variables in a two-way table.

## ■ Example: Comparing Conditional Distributions

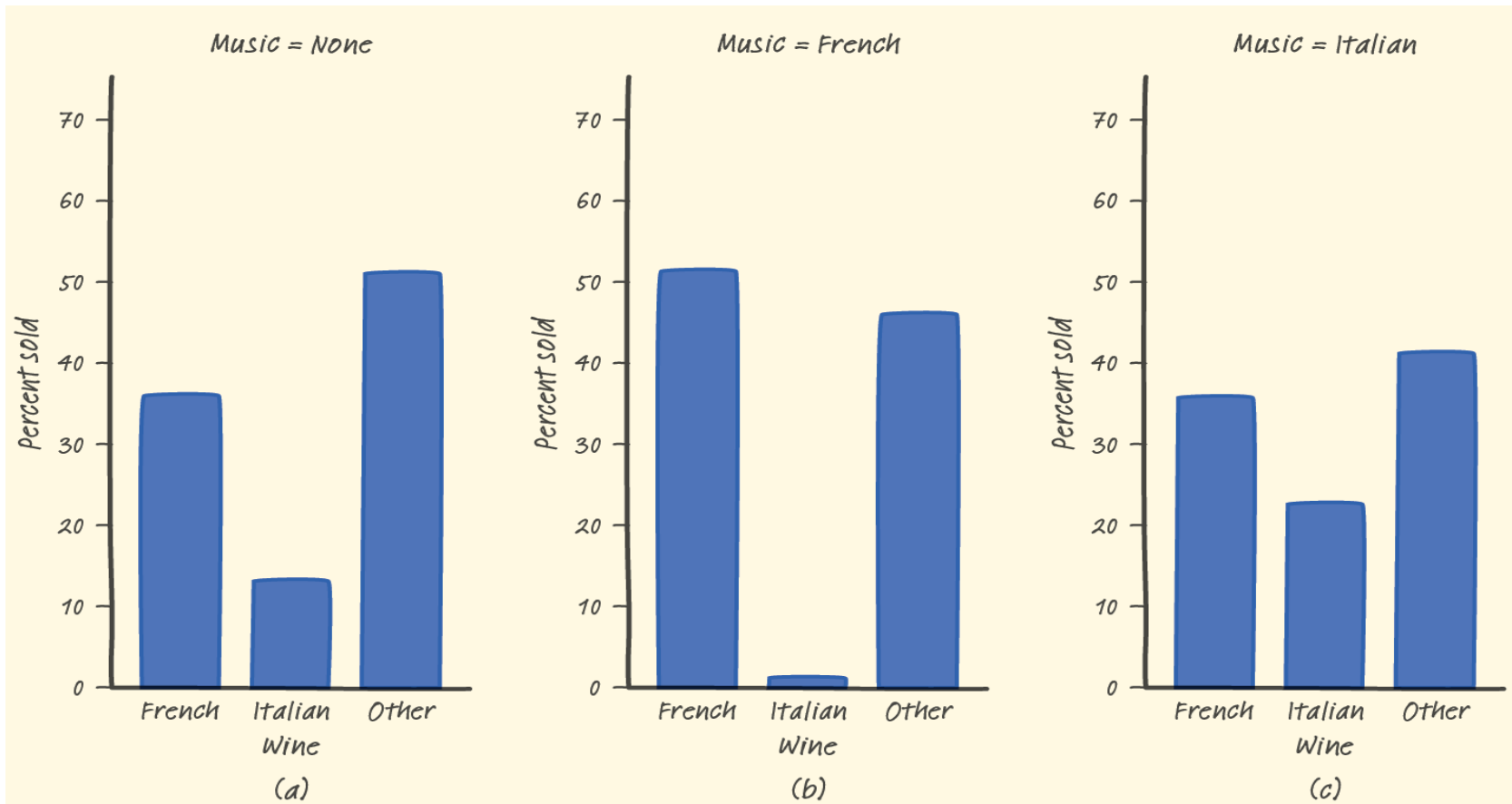
Market researchers suspect that background music may affect the mood and buying behavior of customers. One study in a supermarket compared three randomly assigned treatments: no music, French accordion music, and Italian string music. Under each condition, the researchers recorded the numbers of bottles of French, Italian, and other wine purchased. Here is a table that summarizes the data:

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

### PROBLEM:

- Calculate the conditional distribution (in proportions) of the type of wine sold for each treatment.
- Make an appropriate graph for comparing the conditional distributions in part (a).
- Are the distributions of wine purchases under the three music treatments similar or different? Give appropriate evidence from parts (a) and (b) to support your answer.

## ■ Example: Comparing Conditional Distributions



The type of wine that customers buy seems to differ considerably across the three music treatments. Sales of Italian wine are very low (1.3%) when French music is playing but are higher when Italian music (22.6%) or no music (13.1%) is playing. French wine appears popular in this market, selling well under all music conditions but notably better when French music is playing. For all three music treatments, the percent of Other wine purchases was similar.

## ■ Alternate Example: Saint-John's-wort and depression

An article in the *Journal of the American Medical Association* (vol. 287, no. 14, April 10, 2002) reports the results of a study designed to see if the herb Saint-John's-wort is effective in treating moderately severe cases of depression. The study involved 338 subjects who were being treated for major depression. The subjects were randomly assigned to receive one of three treatments – Saint-John's-wort, Zoloft (a prescription drug), or a placebo – for an eight-week period. The table summarizes the results of the experiment.

	Saint-John's-wort	Zoloft	Placebo	Total
Full Response	27	27	37	91
Partial Response	16	26	13	55
No Response	70	56	66	192
Total	113	109	116	338

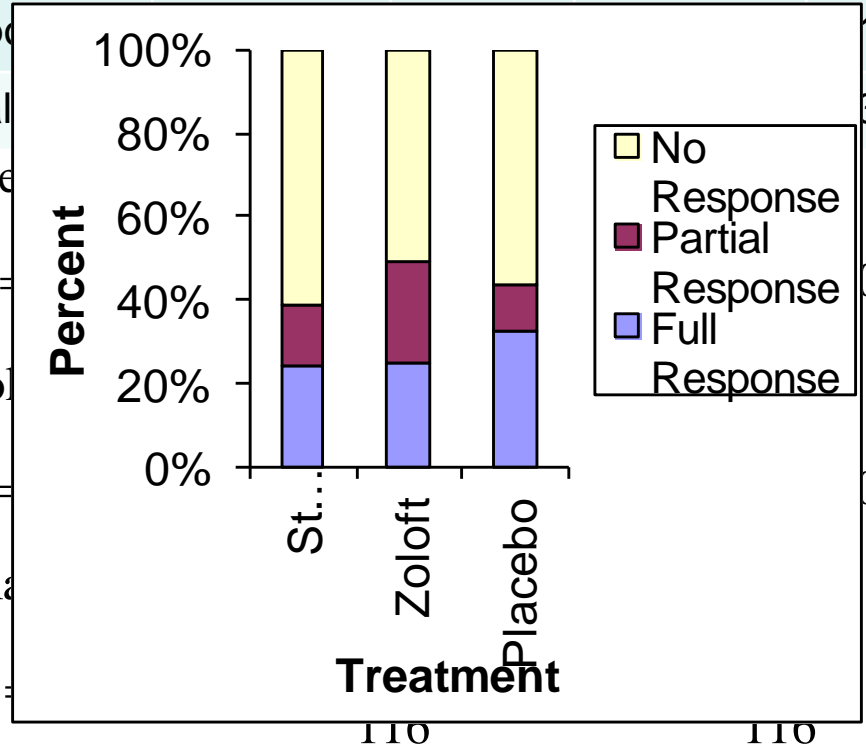
### PROBLEM:

- Calculate the conditional distribution (in proportions) of the type of response for each treatment.
- Make an appropriate graph for comparing the conditional distributions in part (a).
- Compare the distributions of response for each treatment.

# ■ Alternate Example: Saint-John's-wort and depression

	Saint-John's-wort	Zoloft	Placebo	Total
Full Response	27	27	37	91
Partial Response	16	26	13	55
No Response	50	57	62	169
Total	93	110	110	338

(a) For the Saint-John's-wort group:  
 Full:  $\frac{27}{93} = 0.291$   
 For the Zoloft group:  
 Full:  $\frac{27}{110} = 0.245$   
 For the placebo group:  
 Full:  $\frac{37}{110} = 0.336$



(b) Overall proportions:  
 Full:  $\frac{91}{338} = 0.269$   
 Partial:  $\frac{55}{338} = 0.163$   
 No Response:  $\frac{192}{338} = 0.568$

(c) Surprisingly, a higher proportion of subjects receiving the placebo had a full response than subjects receiving Saing-John's-wort or Zoloft. Overall, a higher proportion of Zoloft users had at least some response, followed by placebo users, and then Saint-John's-wort users.

+ Inference for Relationships



## ■ Expected Counts and the Chi-Square Statistic

The problem of how to do many comparisons at once with an overall measure of confidence in all our conclusions is common in statistics. This is the problem of **multiple comparisons**. Statistical methods for dealing with multiple comparisons usually have two parts:

- 1. An *overall test* to see if there is good evidence of any differences among the parameters that we want to compare.
- 2. A detailed *follow-up analysis* to decide which of the parameters differ and to estimate how large the differences are.

The overall test uses the familiar chi-square statistic and distributions.

To perform a test of

$H_0$ : There is no difference in the distribution of a categorical variable for several populations or treatments.

$H_a$ : There is a difference in the distribution of a categorical variable for several populations or treatments.

we compare the observed counts in a two-way table with the counts we would expect if  $H_0$  were true.

## ■ Expected Counts and the Chi-Square Statistic

The overall proportion of French wine bought during the study was  $99/243 = 0.407$ . So the expected counts of French wine bought under each treatment are:

$$\text{No music: } \frac{99}{243} \cdot 84 = 34.22 \quad \text{French music: } \frac{99}{243} \cdot 75 = 30.56 \quad \text{Italian music: } \frac{99}{243} \cdot 84 = 34.22$$

The overall proportion of Italian wine bought during the study was  $31/243 = 0.128$ . So the expected counts of Italian wine bought under each treatment are:

$$\text{No music: } \frac{31}{243} \cdot 84 = 10.72 \quad \text{French music: } \frac{31}{243} \cdot 75 = 9.57 \quad \text{Italian music: } \frac{31}{243} \cdot 84 = 10.72$$

The overall proportion of Other wine bought during the study was  $113/243 = 0.465$ . So the expected counts of Other wine bought under each treatment are:

$$\text{No music: } \frac{113}{243} \cdot 84 = 39.06 \quad \text{French music: } \frac{113}{243} \cdot 75 = 34.88 \quad \text{Italian music: } \frac{113}{243} \cdot 84 = 39.06$$

## ■ Alternate Example: Saint-John's-wort and depression

Here is a summary of the results of the experiment comparing the effects of Saint-John's-wort, Zoloft, and a placebo.

**Problem:** Calculate the expected counts for the three treatments, assuming that all three treatments are equally effective.

**Solution:** Since  $91/338 = 26.9\%$  of all patients had a full response, we expect 26.9% of patients in each treatment group to have a full response.

$$\text{Saint - John's - wort} : \frac{91}{338} \cdot 113 = 30.4$$

$$\text{Zoloft} : \frac{91}{338} \cdot 109 = 29.3$$

$$\text{Placebo} : \frac{91}{338} \cdot 116 = 31.2$$

Similarly, we expect  $55/338 = 16.3\%$  of patients in each treatment group to have a partial response. This gives expected counts of Saint-John's-wort, 18.4; Zoloft, 17.7; and placebo, 18.9.

	Saint-John's-wort	Zoloft	Placebo	Total
Full Response	27	27	37	91
Partial Response	16	26	13	55
No Response	70	56	66	192
Total	113	109	116	338

Finally, we expect  $192/338 = 56.8\%$  of patients in each treatment group to have no response. This gives expected counts of Saint-John's-wort, 64.2; Zoloft, 61.9; and placebo, 65.9.

## Finding Expected Counts

Consider the expected count of French wine bought when no music was playing:

$$\frac{99}{243} \cdot 84 = 34.22$$

Observed Counts				
Wine	Music			Total
	None	French	Italian	
French	30	39	30	99
Italian	11	1	19	31
Other	43	35	35	113
<b>Total</b>	<b>84</b>	<b>75</b>	<b>84</b>	<b>243</b>

The values in the calculation are the row total for French wine, the column total for no music, and the table total. We can rewrite the original calculation as:

$$\frac{99 \cdot 84}{243} = 34.22$$

This suggests a general formula for the expected count in any cell of a two-way table:

### Finding Expected Counts

The expected count in any cell of a two-way table when  $H_0$  is true is

$$\text{expected count} = \frac{\text{row total} \cdot \text{column total}}{\text{table total}}$$

## ■ Calculating the Chi-Square Statistic

In order to calculate a chi-square statistic for the wine example, we must check to make sure the conditions are met:

- ✓ All the expected counts in the music and wine study are at least 5. This satisfies the Large Sample Size condition.
- ✓ The Random condition is met because the treatments were assigned at random.
- ✓ We're comparing three independent groups in a randomized experiment. But are individual observations (each wine bottle sold) independent? If a customer buys several bottles of wine at the same time, knowing that one bottle is French wine might give us additional information about the other bottles bought by this customer. In that case, the Independent condition would be violated. But if each customer buys only one bottle of wine, this condition is probably met. We can't be sure, so we'll proceed to inference with caution.

Just as we did with the chi-square goodness-of-fit test, we compare the observed counts with the expected counts using the statistic

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

This time, the sum is over all cells (not including the totals!) in the two-way table.

## ■ Calculating The Chi-Square Statistic

The tables below show the observed and expected counts for the wine and music experiment. Calculate the chi-square statistic.

Observed Counts					Expected Counts				
Music					Music				
Wine	None	French	Italian	Total	Wine	None	French	Italian	Total
French	30	39	30	<b>99</b>	French	34.22	30.56	34.22	<b>99</b>
Italian	11	1	19	<b>31</b>	Italian	10.72	9.57	10.72	<b>31</b>
Other	43	35	35	<b>113</b>	Other	39.06	34.88	39.06	<b>113</b>
<b>Total</b>	<b>84</b>	<b>75</b>	<b>84</b>	<b>243</b>	<b>Total</b>	<b>84</b>	<b>75</b>	<b>84</b>	<b>243</b>

For the French wine with no music, the observed count is 30 bottles and the expected count is 34.22. The contribution to the  $\chi^2$  statistic for this cell is

$$\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \frac{(30 - 34.22)^2}{34.22} = 0.52$$

The  $\chi^2$  statistic is the sum of nine such terms

$$\begin{aligned} \chi^2 &= \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \frac{(30 - 34.22)^2}{34.22} + \frac{(39 - 30.56)^2}{30.56} + \dots + \frac{(35 - 39.06)^2}{39.06} \\ &= 0.52 + 2.33 + \dots + 0.42 = 18.28 \end{aligned}$$

## ■ Alternate Example: Saint-John's-wort and depression

Here is a summary of the results of the experiment comparing the effects of Saint-John's-wort, Zoloft, and a placebo. Expected counts are listed in parentheses below the observed counts.

	Saint-John's-wort	Zoloft	Placebo	Total
Full Response	27 (30.4)	27 (29.3)	37 (31.2)	91
Partial Response	16 (18.4)	26 (17.7)	13 (18.9)	55
No Response	70 (64.2)	56 (61.9)	66 (65.9)	192
Total	113	109	116	338

Inference for Relationships

**Problem:** Calculate the chi-square statistic. Show your work.

**Solution:**

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \frac{(27 - 30.4)^2}{30.4} + \frac{(27 - 29.3)^2}{29.3} + \dots = 8.72$$

# ■ The Chi-Square Test for Homogeneity

## Chi-Square Test for Homogeneity

Suppose the Random, Large Sample Size, and Independent conditions are met. You can use the **chi-square test for homogeneity** to test

$H_0$ : There is no difference in the distribution of a categorical variable for several populations or treatments.

$H_a$ : There is a difference in the distribution of a categorical variable for several populations or treatments.

Start by finding the expected counts. Then calculate the chi-square statistic

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

where the sum is over all cells (not including totals) in the two-way table. If  $H_0$  is true, the  $\chi^2$  statistic has approximately a chi-square distribution with degrees of freedom = (number of rows – 1) (number of columns - 1). The  $P$ -value is the area to the right of  $\chi^2$  under the corresponding chi-square density curve.



## ■ Example: Does Music Influence Purchases?

Earlier, we started a significance test of

$H_0$ : There is no difference in the distributions of wine purchases at this store when no music, French accordion music, or Italian string music is played.

$H_a$ : There is a difference in the distributions of wine purchases at this store when no music, French accordion music, or Italian string music is played.

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

We decided to proceed with caution because, although the Random and Large Sample Size conditions are met, we aren't sure that individual observations (type of wine bought) are independent.

Our calculated test statistic is  $\chi^2 = 18.28$ .

To find the  $P$ -value, use Table C and look in the  $df = (3 - 1)(3 - 1) = 4$  row.

$P$		
df	.0025	.001
4	16.42	18.47

The small  $P$ -value gives us convincing evidence to reject  $H_0$  and conclude that there is a difference in the distributions of wine purchases at this store when no music, French accordion music, or Italian string music is played. Furthermore, the random assignment allows us to say that the difference is caused by the music that's played.

## ■ Alternate Example: Saint-John's-wort and depression

Earlier, we started a significance test of

$H_0$ : There is no difference in the distributions of responses for patients with moderately severe cases of depression when taking Saint-John's-wort, Zoloft, or a placebo.

$H_a$ : There is a difference in the distributions of responses for patients with moderately severe cases of depression when taking Saint-John's-wort, Zoloft, or a placebo.

Our calculated test statistic is  $\chi^2 = 18.28$ . **Solution:**

### Problem:

- (a) Verify that the conditions for this test are satisfied.
  - (b) Calculate the P-value for this test.
  - (c) Interpret the P-value in context.
  - (d) What is your conclusion?
- (b)  $df = (3-1)(3-1) = 4$ ,

$$P\text{-value} = \chi^2 cdf(8.72, 99999, 4) = 0.0685$$

- (c) Assuming that the treatments are equally effective, the probability of observing a difference in the distributions of responses among the three treatment groups as large or larger than the one in the study is about 0.07.
- (d) Since the P-value is greater than  $\alpha = 0.05$ , we fail to reject the null hypothesis. We do not have convincing evidence that there is a difference in the distributions of responses for patients with moderately severe cases of depression when taking Saint-John's-wort, Zoloft, or a placebo.

(a) *Random*: The treatments were randomly assigned.

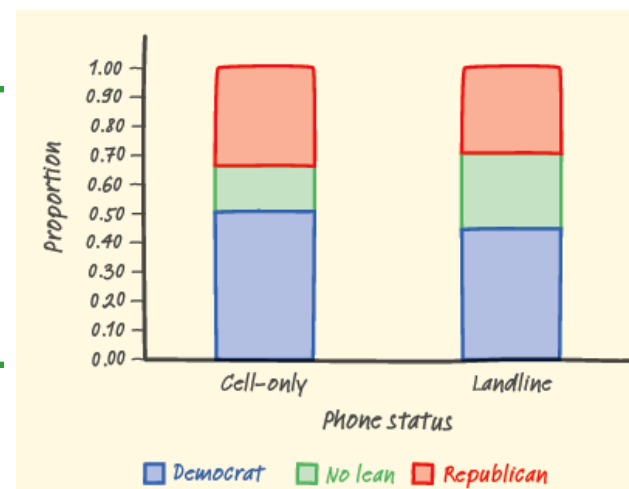
*Large Sample Size*: The expected counts are all at least 5.

*Independent*: Knowing the response of one patient should not provide any additional information about the response of any other patient.

## Example: Cell-Only Telephone Users

Random digit dialing telephone surveys used to exclude cell phone numbers. If the opinions of people who have only cell phones differ from those of people who have landline service, the poll results may not represent the entire adult population. The Pew Research Center interviewed separate random samples of cell-only and landline telephone users who were less than 30 years old. Here's what the Pew survey found about how these people describe their political party affiliation.

	Cell-only sample	Landline sample
Democrat or lean Democratic	49	47
Refuse to lean either way	15	27
Republican or lean Republican	32	30
Total	96	104



**State:** We want to perform a test of

$H_0$ : There is no difference in the distribution of party affiliation in the cell-only and landline populations.

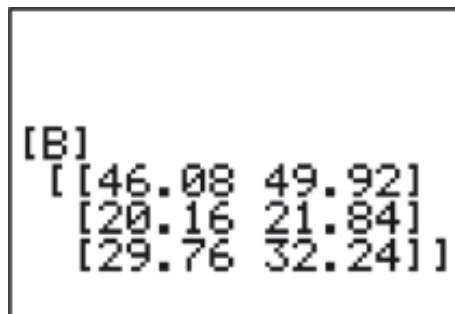
$H_a$ : There is a difference in the distribution of party affiliation in the cell-only and landline populations.

We will use  $\alpha = 0.05$ .

## ■ Example: Cell-Only Telephone Users

**Plan:** If the conditions are met, we should conduct a chi-square test for homogeneity.

- *Random* The data came from separate random samples of 96 cell-only and 104 landline users.
- *Large Sample Size* We followed the steps in the Technology Corner (page 705) to get the expected counts. The calculator screenshot confirms all expected counts  $\geq 5$ .



```
[B]
[[46.08 49.92]
 [20.16 21.84]
 [29.76 32.24]]
```

- *Independent* Researchers took independent samples of cell-only and landline phone users. Sampling without replacement was used, so there need to be at least  $10(96) = 960$  cell-only users under age 30 and at least  $10(104) = 1040$  landline users under age 30. This is safe to assume.

## ■ Example: Cell-Only Telephone Users

**Do:** Since the conditions are satisfied, we can perform a chi-test for homogeneity. We begin by calculating the test statistic.

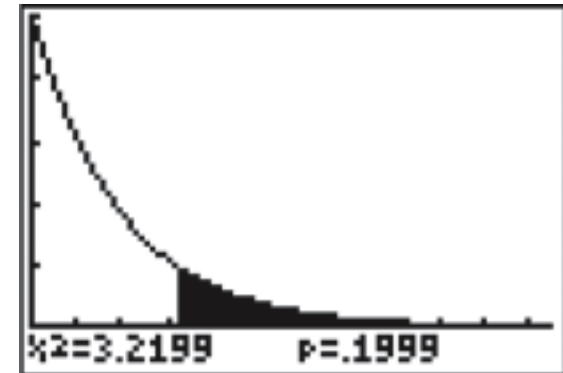
**Test statistic:**

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

$$= \frac{(49 - 46.08)^2}{46.08} + \frac{(47 - 49.92)^2}{49.92} + \dots + \frac{(30 - 32.24)^2}{32.24} = 3.22$$

**P-Value:**

Using  $df = (3 - 1)(2 - 1) = 2$ , the  $P$ -value is 0.20.



**Conclude:** Because the  $P$ -value, 0.20, is greater than  $\alpha = 0.05$ , we fail to reject  $H_0$ . There is not enough evidence to conclude that the distribution of party affiliation differs in the cell-only and landline user populations.

## ■ Follow-up Analysis

The chi-square test for homogeneity allows us to compare the distribution of a categorical variable for any number of populations or treatments. If the test allows us to reject the null hypothesis of no difference, we then want to do a follow-up analysis that examines the differences in detail.

Start by examining which cells in the two-way table show large deviations between the observed and expected counts. Then look at the individual components to see which terms contribute most to the chi-square statistic.

Minitab output for the wine and music study displays the individual components that contribute to the chi-square statistic.

Looking at the output, we see that just two of the nine components that make up the chi-square statistic contribute about 14 (almost 77%) of the total  $\chi^2 = 18.28$ .

We are led to a specific conclusion: *sales of Italian wine are strongly affected by Italian and French music.*

### Chi-Square Test: None, Franch, Italian

Expected counts are printed below observed counts  
Chi-Square contributions are printed below expected counts

	None	French	Italian	Total
1	30 34.22 0.521	39 30.56 2.334	30 34.22 0.521	99
2	11 10.72 0.008	1 9.57 7.672	19 10.72 6.404	31
3	43 39.06 0.397	35 34.88 0.000	35 39.06 0.422	113
Total	84	75	84	243

Chi-Sq = 18.279, DF = 4, P-Value = 0.001

## ■ Comparing Several Proportions

Many studies involve comparing the proportion of successes for each of several populations or treatments.

- The two-sample z test from Chapter 10 allows us to test the null hypothesis  $H_0: p_1 = p_2$ , where  $p_1$  and  $p_2$  are the actual proportions of successes for the two populations or treatments.
- The chi-square test for homogeneity allows us to test  $H_0: p_1 = p_2 = \dots = p_k$ . This null hypothesis says that there is no difference in the proportions of successes for the  $k$  populations or treatments. The alternative hypothesis is  $H_a$ : at least two of the  $p_i$ 's are different.

### **Caution:**

Many students *incorrectly state*  $H_a$  as “all the proportions are different.”

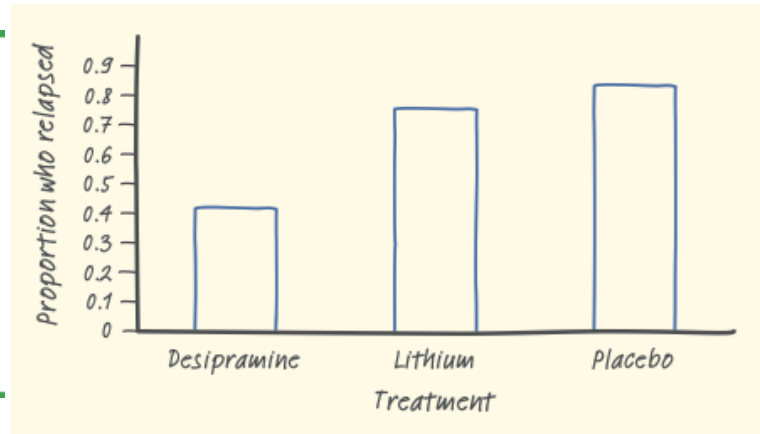
Think about it this way: the opposite of “all the proportions are equal” is “some of the proportions are not equal.”

# Example: Cocaine Addiction is Hard to Break

Cocaine addicts need cocaine to feel any pleasure, so perhaps giving them an antidepressant drug will help. A three-year study with 72 chronic cocaine users compared an antidepressant drug called desipramine with lithium (a standard drug to treat cocaine addiction) and a placebo. One-third of the subjects were randomly assigned to receive each treatment. Here are the results:

Inference for Relationships

Group	Treatment	Cocaine Relapse?	
		Yes	No
1	Desipramine	10	14
2	Lithium	18	6
3	Placebo	20	4



**State:** We want to perform a test of

$$H_0: p_1 = p_2 = p_3$$

there is no difference in the relapse rate for the three treatments.

$H_a$ : at least two of the  $p_i$ 's are different

there is a difference in the relapse rate for the three treatments.

where  $p_i$  = the actual proportion of chronic cocaine users like the ones in this experiment who would relapse under treatment  $i$ . We will use  $\alpha = 0.01$ .



## ■ Example: Cocaine Addiction is Hard to Break

**Plan:** If the conditions are met, we should conduct a chi-square test for homogeneity.

- *Random* The subjects were randomly assigned to the treatment groups.
- *Large Sample Size* We can calculate the expected counts from the two-way table assuming  $H_0$  is true.

Expected count who relapse under each treatment

$$\frac{24 \cdot 48}{72} = 16$$

Expected count who don't relapse under each treatment

$$\frac{24 \cdot 24}{72} = 8$$

Group	Treatment	Cocaine Relapse?	
		Yes	No
1	Desipramine	10	14
2	Lithium	18	6
3	Placebo	20	4

All the expected counts are  $\geq 5$  so the condition is met.

- *Independent* The random assignment helps create three independent groups. If the experiment is conducted properly, then knowing one subject's relapse status should give us no information about another subject's outcome. So individual observations are independent.

## ■ Example: Cocaine Addiction is Hard to Break

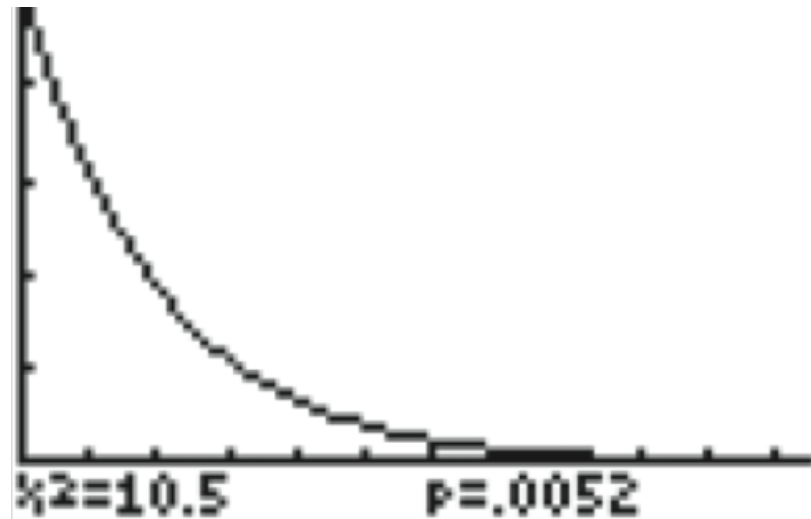
**Do:** Since the conditions are satisfied, we can perform a chi-test for homogeneity. We begin by calculating the test statistic.

**Test statistic:**

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = 10.5$$

**P-Value:**

Using  $df = (3 - 1)(2 - 1) = 2$ , the calculator gives a  $P$ -value of 0.0052.



**Conclude:** Because the  $P$ -value, 0.0052, is less than  $\alpha = 0.01$ , we reject  $H_0$ . We have sufficient evidence to conclude that the true relapse rates for the three treatments are not all the same.

## ■ Relationships Between Two Categorical Variables

Another common situation that leads to a two-way table is when a *single* random sample of individuals is chosen from a *single* population and then classified according to two categorical variables. In that case, our goal is to analyze the relationship between the variables.

A study followed a random sample of 8474 people with normal blood pressure for about four years. All the individuals were free of heart disease at the beginning of the study. Each person took the Spielberger Trait Anger Scale test, which measures how prone a person is to sudden anger. Researchers also recorded whether each individual developed coronary heart disease (CHD). This includes people who had heart attacks and those who needed medical treatment for heart disease. Here is a two-way table that summarizes the data:

---

	<b>Low anger</b>	<b>Moderate anger</b>	<b>High anger</b>	<b>Total</b>
CHD	53	110	27	190
No CHD	3057	4621	606	8284
Total	3110	4731	633	8474

---

## ■ Example: Angry People and Heart Disease

We're interested in whether angrier people tend to get heart disease more often. We can compare the percents of people who did and did not get heart disease in each of the three anger categories:

	Low anger	Moderate anger	High anger	Total
CHD	53	110	27	190
No CHD	3057	4621	606	8284
Total	3110	4731	633	8474

Low anger:	CHD $\frac{53}{3110} = 0.0170 = 1.70\%$	no CHD $\frac{3057}{3110} = 0.9830 = 98.30\%$
Moderate anger:	CHD $\frac{110}{4731} = 0.0233 = 2.33\%$	no CHD $\frac{4621}{4731} = 0.9767 = 97.67\%$
High anger:	CHD $\frac{27}{633} = 0.0427 = 4.27\%$	no CHD $\frac{606}{633} = 0.9573 = 95.73\%$



There is a clear trend: as the anger score increases, so does the percent who suffer heart disease. A much higher percent of people in the high anger category developed CHD (4.27%) than in the moderate (2.33%) and low (1.70%) anger categories.

## ■ The Chi-Square Test for Association/Independence

We often gather data from a random sample and arrange them in a two-way table to see if two categorical variables are associated. The sample data are easy to investigate: turn them into percents and look for a relationship between the variables.

Our null hypothesis is that there is no association between the two categorical variables. The alternative hypothesis is that there is an association between the variables. For the observational study of anger level and coronary heart disease, we want to test the hypotheses

$H_0$ : There is no association between anger level and heart disease in the population of people with normal blood pressure.

$H_a$ : There is an association between anger level and heart disease in the population of people with normal blood pressure.

No association between two variables means that the values of one variable do not tend to occur in common with values of the other. That is, the variables are independent. An equivalent way to state the hypotheses is therefore

$H_0$ : Anger and heart disease are independent in the population of people with normal blood pressure.

$H_a$ : Anger and heart disease are not independent in the population of people with normal blood pressure.

# ■ The Chi-Square Test for Association/Independence

## Chi-Square Test for Association/Independence

Suppose the Random, Large Sample Size, and Independent conditions are met. You can use the **chi-square test for association/independence** to test

$H_0$ : There is no association between two categorical variables in the population of interest.

$H_a$ : There is an association between two categorical variables in the population of interest.

Or, alternatively

$H_0$ : Two categorical variables are independent in the population of interest.

$H_a$ : Two categorical variables are not independent in the population of interest.

Start by finding the expected counts. Then calculate the chi-square statistic

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

where the sum is over all cells (not including totals) in the two-way table. If  $H_0$  is true, the  $\chi^2$  statistic has approximately a chi-square distribution with degrees of freedom = (number of rows - 1) (number of columns - 1). The  $P$ -value is the area to the right of  $\chi^2$  under the corresponding chi-square density curve.

## ■ Example: Angry People and Heart Disease

Here is the complete table of observed and expected counts for the CHD and anger study side by side. Do the data provide convincing evidence of an association between anger level and heart disease in the population of interest?

	Observed			Expected		
	Low	Moderate	High	Low	Moderate	High
CHD	53	110	27	69.73	106.08	14.19
No CHD	3057	4621	606	3040.27	4624.92	618.81

**State:** We want to perform a test of

$H_0$ : There is no association between anger level and heart disease in the population of people with normal blood pressure.

$H_a$ : There is an association between anger level and heart disease in the population of people with normal blood pressure.

We will use  $\alpha = 0.05$ .



## ■ Example: Angry People and Heart Disease

**Plan:** If the conditions are met, we should conduct a chi-square test for association/independence.

- *Random* The data came from a random sample of 8474 people with normal blood pressure.
- *Large Sample Size* All the expected counts are at least 5, so this condition is met.
- *Independent* Knowing the values of both variables for one person in the study gives us no meaningful information about the values of the variables for another person. So individual observations are independent. Because we are sampling without replacement, we need to check that the total number of people in the population with normal blood pressure is at least  $10(8474) = 84,740$ . This seems reasonable to assume.



## ■ Example: Cocaine Addiction is Hard to Break

**Do:** Since the conditions are satisfied, we can perform a chi-test for association/independence. We begin by calculating the test statistic.

**Test statistic:**

$$\begin{aligned}\chi^2 &= \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \\ &= \frac{(53 - 69.73)^2}{69.73} + \frac{(110 - 106.08)^2}{106.08} + \dots + \frac{(606 - 618.81)^2}{618.81} \\ &= 4.014 + 0.145 + \dots + 0.265 = 16.077\end{aligned}$$

**P-Value:**

The two-way table of anger level versus heart disease has 2 rows and 3 columns. We will use the chi-square distribution with  $df = (2 - 1)(3 - 1) = 2$  to find the  $P$ -value.

*Table:* Look at the  $df = 2$  line in Table C. The observed statistic  $\chi^2 = 16.077$  is larger than the critical value 15.20 for  $\alpha = 0.0005$ . So the  $P$ -value is less than 0.0005.

*Technology:* The command  $\chi^2\text{cdf}(16.077, 1000, 2)$  gives 0.00032.

**Conclude:** Because the  $P$ -value is clearly less than  $\alpha = 0.05$ , we reject  $H_0$  and conclude that anger level and heart disease are associated in the population of people with normal blood pressure.

## ■ Using Chi-Square Tests Wisely

Both the chi-square test for homogeneity and the chi-square test for association/independence start with a two-way table of observed counts. They even calculate the test statistic, degrees of freedom, and  $P$ -value in the same way. *The questions that these two tests answer are different, however.*

- A chi-square test for homogeneity tests whether the distribution of a categorical variable is the same for each of several populations or treatments.
- The chi-square test for association/independence tests whether two categorical variables are associated in some population of interest.

Instead of focusing on the question asked, it's much easier to look at how the data were produced.

- ✓ If the data come from two or more independent random samples or treatment groups in a randomized experiment, then do a chi-square test for homogeneity.
- ✓ If the data come from a single random sample, with the individuals classified according to two categorical variables, use a chi-square test for association/independence.



## Section 11.2

# Inference for Relationships

### Summary

In this section, we learned that...

- ✓ We can use a two-way table to summarize data on the relationship between two categorical variables. To analyze the data, we first compute percents or proportions that describe the relationship of interest.
- ✓ If data are produced using independent random samples from each of several populations of interest or the treatment groups in a randomized comparative experiment, then each observation is classified according to a categorical variable of interest. The null hypothesis is that the distribution of this categorical variable is the same for all the populations or treatments. We use the **chi-square test for homogeneity** to test this hypothesis.
- ✓ If data are produced using a single random sample from a population of interest, then each observation is classified according to two categorical variables. The **chi-square test of association/independence** tests the null hypothesis that there is no association between the two categorical variables in the population of interest. Another way to state the null hypothesis is  $H_0$ : The two categorical variables are independent in the population of interest.



## Section 11.1

# Chi-Square Goodness-of-Fit Tests

### Summary

- ✓ The expected count in any cell of a two-way table when  $H_0$  is true is

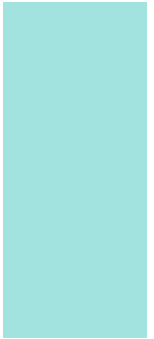
$$\text{expected count} = \frac{\text{row total} \cdot \text{column total}}{\text{table total}}$$

- ✓ The chi-square statistic is

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

where the sum is over all cells in the two-way table.

- ✓ The chi-square test compares the value of the statistic  $\chi^2$  with critical values from the chi-square distribution with  $df = (\text{number of rows} - 1)(\text{number of columns} - 1)$ . Large values of  $\chi^2$  are evidence against  $H_0$ , so the  $P$ -value is the area under the chi-square density curve to the right of  $\chi^2$ .



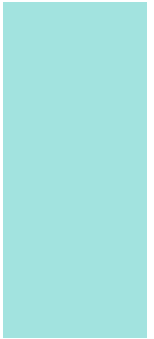


## Section 11.1

# Chi-Square Goodness-of-Fit Tests

### Summary

- ✓ The chi-square distribution is an approximation to the distribution of the statistic  $\chi^2$ . You can safely use this approximation when all expected cell counts are at least 5 (the Large Sample Size condition).
- ✓ Be sure to check that the Random, Large Sample Size, and Independent conditions are met before performing a chi-square test for a two-way table.
- ✓ If the test finds a statistically significant result, do a follow-up analysis that compares the observed and expected counts and that looks for the largest components of the chi-square statistic.





# Looking Ahead...



## In the next Chapter...

We'll learn more about regression.

We'll learn about

- ✓ **Inference for Linear Regression**
- ✓ **Transforming to Achieve Linearity**