2, 4, 6, 8, 12, 18, 28, 30, 32, 34, 36, 44, 46, 50, 52

Chapter 11: Inference for Distributions of Categorical Data

Section 11.1
Chi-Square Goodness-of-Fit Tests

The Practice of Statistics, 4th 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 11.1
Chi-Square Goodness-of-Fit Tests

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 goodness-of-fit test to determine whether sample data are consistent with a specified distribution of a categorical variable
- ✓ EXAMINE individual components of the chi-square statistic as part of a follow-up analysis

1.) State null and alternative hypothesis

Ho: The specified distributions of the [categorical variable] are correct
Ha: The specified distributions of the [categorical variable] are NOT correct

<u>MUST BE IN CONTEXT OF QUESTION!</u>

2.) “If conditions are met we will do a Chi-square Goodness-of-fit test at α = [alpha level]”

3.) Conditions: Random, Independent, Large enough (all expected values at least 5)

4.) Find test statistic $\chi^2 = \Sigma$ (Observed – Expected)^2 / Expected

5.) State degrees of freedom = categories – 1

6.) Use table or χ^2cdf to find p-value

7.) Compare to Alpha α

8.) “We will” Reject (if p-value<α) or Fail to reject (if p-value>α) “null hypothesis”

9.) Write conclusion about alternative hypothesis in context of question.

Introduction

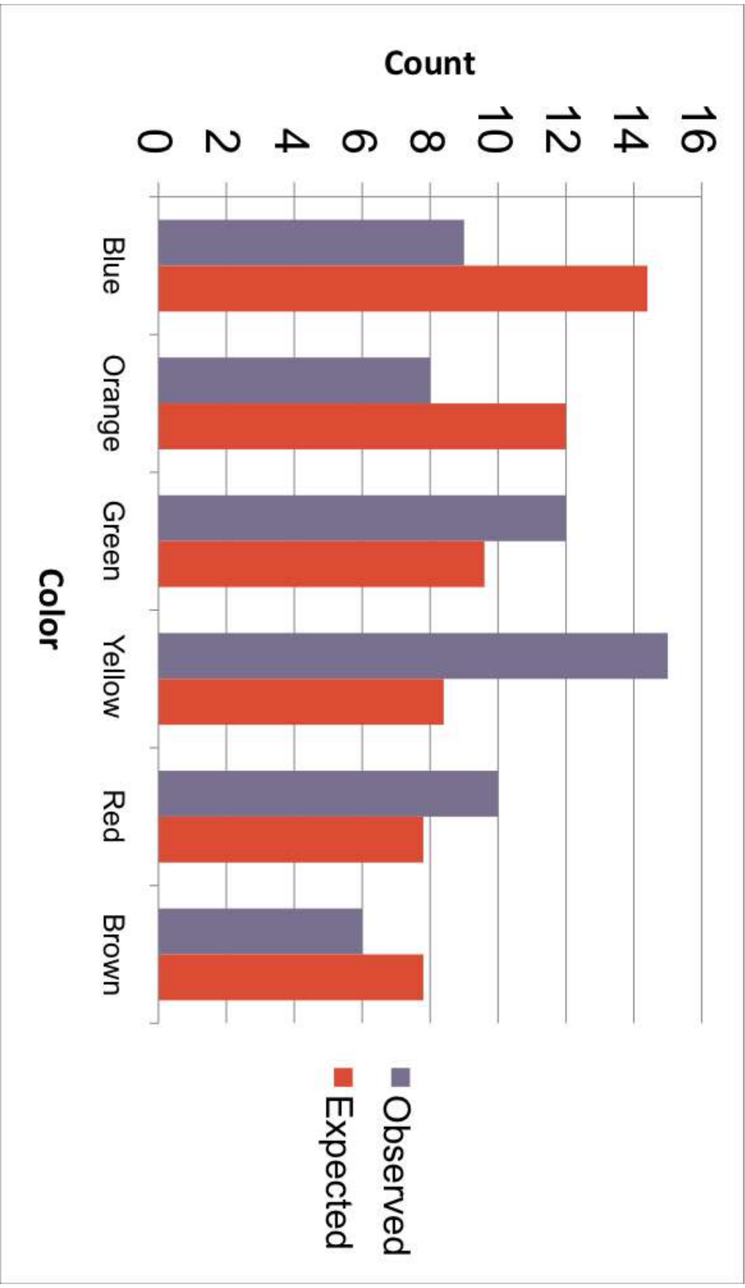
In the previous chapter, we discussed inference procedures for comparing the proportion of successes for two populations or treatments. Sometimes we want to examine the distribution of a single categorical variable in a population. The **chi-square goodness-of-fit test** allows us to determine whether a hypothesized distribution seems valid.

We can decide whether the distribution of a categorical variable differs for two or more populations or treatments using a **chi-square test for homogeneity**. In doing so, we will often organize our data in a two-way table.

It is also possible to use the information in a two-way table to study the relationship between two categorical variables. The **chi-square test for association/independence** allows us to determine if there is convincing evidence of an association between the variables in the population at large.

Activity: The Candy Man Can

- Mars, Incorporated makes milk chocolate candies. Here's what the company's Consumer Affairs Department says about the color distribution of its M&M'S Milk Chocolate Candies: *On average, the new mix of colors of M&M'S Milk Chocolate Candies will contain 13 percent of each of browns and reds, 14 percent yellows, 16 percent greens, 20 percent oranges and 24 percent blues.*



Chi-Square Goodness-of-Fit Tests

The **one-way table** below summarizes the data from a sample bag of M&M'S Milk Chocolate Candies. In general, one-way tables display the distribution of a categorical variable for the individuals in a sample.

Chi-Square Goodness-of-Fit Tests

Color	Blue	Orange	Green	Yellow	Red	Brown	**Total**
Count	9	8	12	15	10	6	**60**

The sample proportion of blue M&M's is $\hat{p} = \frac{9}{60} = 0.15$.

Since the company claims that 24% of all M&M'S Milk Chocolate Candies are blue, we might believe that something fishy is going on. We could use the one-sample *z* test for a proportion from Chapter 9 to test the hypotheses

$$H_0: p = 0.24$$
$$H_a: p \neq 0.24$$

where *p* is the true population proportion of blue M&M'S. We could then perform additional significance tests for each of the remaining colors.

However, performing a one-sample *z* test for each proportion would be pretty inefficient and would lead to the problem of multiple comparisons.

Comparing Observed and Expected Counts

More important, performing one-sample z tests for each color wouldn't tell us how likely it is to get a random sample of 60 candies with a color distribution that differs as much from the one claimed by the company as this bag does (taking *all* the colors into consideration at one time).

For that, we need a new kind of significance test, called a **chi-square goodness-of-fit test.**

The null hypothesis in a chi-square goodness-of-fit test should state a claim about the distribution of a single categorical variable in the population of interest. In our example, the appropriate null hypothesis is

H_0: The company's stated color distribution for M&M'S Milk Chocolate Candies is correct.

The alternative hypothesis in a chi-square goodness-of-fit test is that the categorical variable does *not* have the specified distribution. In our example, the alternative hypothesis is

H_a: The company's stated color distribution for M&M'S Milk Chocolate Candies is not correct.

Comparing Observed and Expected Counts

We can also write the hypotheses in symbols as

H_0: p_{blue} = 0.24, p_{orange} = 0.20, p_{green} = 0.16,
p_{yellow} = 0.14, p_{red} = 0.13, p_{brown} = 0.13,

H_a: At least one of the p_i's is incorrect

where p_{color} = the true population proportion of M&M'S Milk Chocolate Candies of that color.

The idea of the chi-square goodness-of-fit test is this: we compare the **observed counts** from our sample with the counts that would be expected if H_0 is true. The more the observed counts differ from the **expected counts**, the more evidence we have against the null hypothesis.

In general, the expected counts can be obtained by multiplying the proportion of the population distribution in each category by the sample size.

Example: Computing Expected Counts

A sample bag of M&M's milk Chocolate Candies contained 60 candies. Calculate the expected counts for each color.

Assuming that the color distribution stated by Mars, Inc., is true, 24% of all M&M's milk Chocolate Candies produced are blue.

For random samples of 60 candies, the average number of blue M&M's should be (0.24)(60) = 14.40. This is our expected count of blue M&M's.

Using this same method, we can find the expected counts for the other color categories:

Orange:(0.20)(60) = 12.00

Green:(0.16)(60) = 9.60

Yellow:(0.14)(60) = 8.40

Red:(0.13)(60) = 7.80

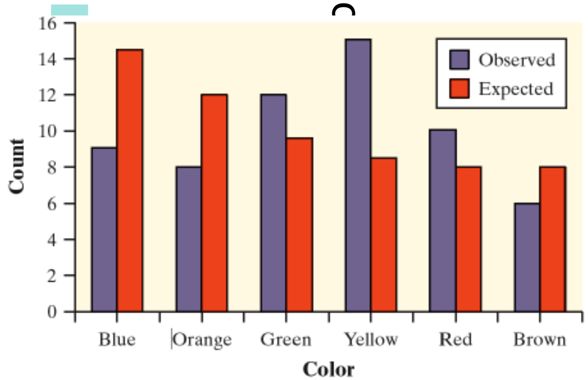
Brown:(0.13)(60) = 7.80

Color	Observed	Expected
Blue	9	14.40
Orange	8	12.00
Green	12	9.60
Yellow	15	8.40
Red	10	7.80
Brown	6	7.80

The Chi-Square Statistic

...o see if the data give convincing evidence against the null hypothesis, we compare the observed counts from our sample with the expected counts assuming H_0 is true. If the observed counts are far from the expected counts, that's the evidence we were seeking.

Chi-Square Goodness-of-Fit Tests



We see some fairly large differences between the observed and expected counts in several color categories. How likely is it that differences this large or larger would occur just by chance in random samples of size 60 from the population distribution claimed by Mars, Inc.?

To answer this question, we calculate a statistic that measures how far apart the observed and expected counts are. The statistic we use to make the comparison is the **chi-square statistic.**

Definition:

The **chi-square statistic** is a measure of how far the observed counts are from the expected counts. The formula for the statistic is

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

where the sum is over all possible values of the categorical variable.

Chi-Square Goodness-of-Fit Tests

Example: Return of the M&M's

The table shows the observed and expected counts for our sample of 60 M&M's Milk Chocolate Candies. Calculate the chi-square statistic.

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

Color	Observed	Expected
Blue	9	14.40
Orange	8	12.00
Green	12	9.60
Yellow	15	8.40
Red	10	7.80
Brown	6	7.80

$$\chi^2 = \frac{(9-14.40)^2}{14.40} + \frac{(8-12.00)^2}{12.00} + \frac{(12-9.60)^2}{9.60} + \frac{(15-8.40)^2}{8.40} + \frac{(10-7.80)^2}{7.80} + \frac{(6-7.80)^2}{7.80}$$

$$\chi^2 = 2.025 + 1.333 + 0.600 + 5.186 + 0.621 + 0.415 = 10.180$$

Think of χ^2 as a measure of the distance of the observed counts from the expected counts. Large values of χ^2 are stronger evidence against H_0 because they say that the observed counts are far from what we would expect if H_0 were true. Small values of χ^2 suggest that the data are consistent with the null hypothesis.

The Chi-Square Distributions and P-Values

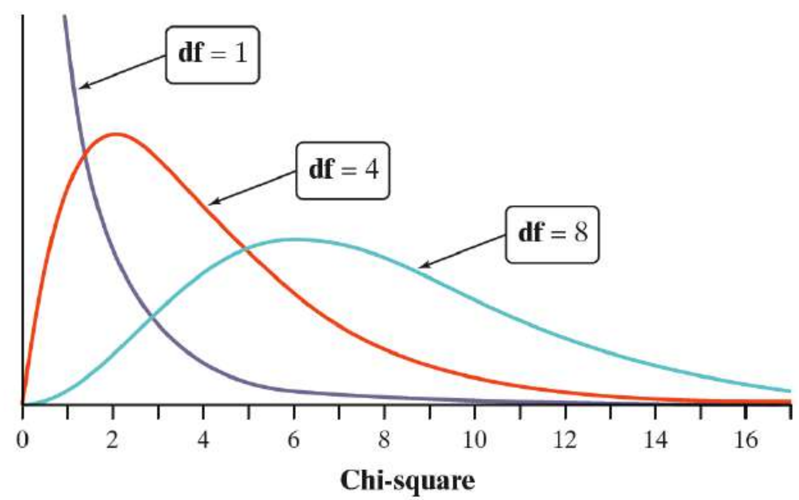
The sampling distribution of the chi-square statistic is not a Normal distribution. It is a right-skewed distribution that allows only positive values because χ^2 can never be negative.

When the expected counts are all at least 5, the sampling distribution of the χ^2 statistic is close to a **chi-square distribution** with degrees of freedom (df) equal to the number of categories minus 1.

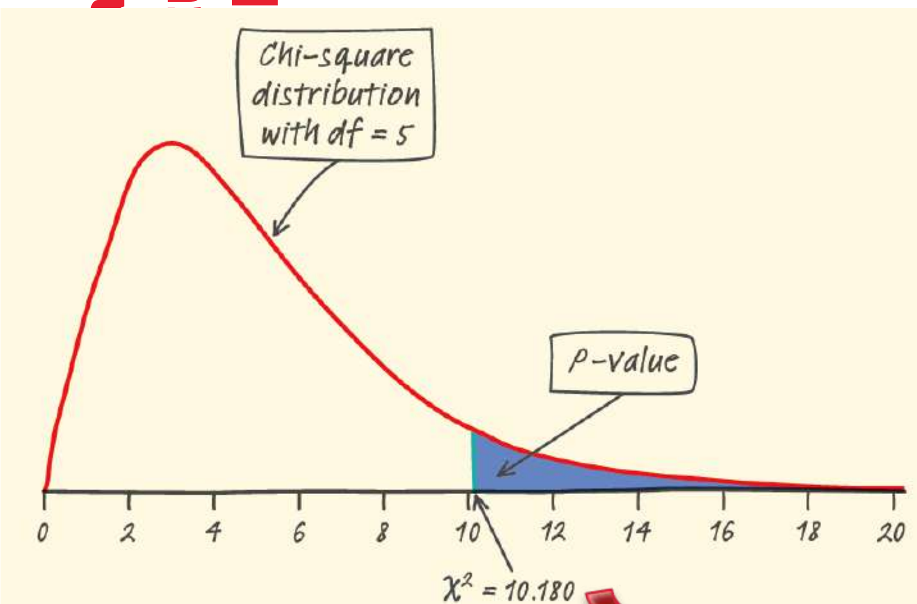
Mean = degrees of freedom
df - 2 = mode

The Chi-Square Distributions

The chi-square distributions are a family of distributions that take only positive values and are skewed to the right. A particular chi-square distribution is specified by giving its degrees of freedom. The chi-square goodness-of-fit test uses the chi-square distribution with degrees of freedom = the number of categories - 1.



We computed the chi-square statistic for our sample of 60 M&M's to be $\chi^2 = 10.180$. Because all of the expected counts are at least 5, the χ^2 statistic will follow a chi-square distribution with df = 6 - 1 = 5 reasonably well when H_0 is true.



To find the *P*-value, use Table C and look in the df = 5 row.

		P	
df	**.15**	**.10**	**.05**
4	6.74	7.78	9.49
5	8.12	9.24	11.07
6	9.45	10.64	12.59

Since our *P*-value is between 0.05 and 0.10, it is greater than $\alpha = 0.05$. Therefore, we fail to reject H_0. We don't have sufficient evidence to conclude that the company's claimed color distribution is incorrect.

Carrying Out a Test

Chi-Square Goodness-of-Fit Test

The Chi-Square Goodness-of-Fit Test

Suppose the Random, Large S mple S ze, and Independent conditions are met. To determine whe r a c gori vari e has specified distribution, expr d a alli nto each possible ca ory.

H_0: The orrect.

H_a: T ect.

We c

true

Before we start using the chi-square goodness-of-fit test, we have two important cautions to offer.

1. The chi-square test statistic compares observed and expected *counts.* Don't try to perform calculations with the observed and expected *proportions* in each category.

2. When checking the Large Sample Size condition, be sure to examine the *expected* counts, not the observed counts.

where the sum is ver t valu the area to the right of χ^2 under the densit urve he c squar distribution with $k-1$ degrees of freedom.

Example: When Were You Born?

Are births even… …ibuted across the
days of the … The one-way table
below show… …istribution of births
across the … the week in a
random san… 140 births from
local record… …arge city. Do these
data give si… …nt evidence that
local births … equally likely on all
days of the

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Births	13	23	24	20	27	18	15

State: We want to perform a test of

H_0: Birth days in this local area are evenly distributed across the days of the week.
H_a: Birth days in this local area are not evenly distributed across the days of the week.

The null hypothesis says that the proportions of births are the same on all days. In that case, all 7 proportions must be 1/7. So we could also write the hypotheses as

H_0: $p_{Sun} = p_{Mon} = p_{Tues} = . . . = p_{Sat} = 1/7$.
H_a: At least one of the proportions is not 1/7.

We will use $\alpha = 0.05$.

Plan: If the conditions are met, we should conduct a chi-square goodness-of-fit test.

• *Random* The data came from a random sample of local births.

• *Large Sample Size* Assuming H_0 is true, we would expect one-seventh of the births to occur on each day of the week. For the sample of 140 births, the expected count for all 7 days would be 1/7(140) = 20 births. Since 20 ≥ 5, this condition is met.

• *Independent* Individual births in the random sample should occur independently (assuming no twins). Because we are sampling without replacement, there need to be at least 10(140) = 1400 births in the local area. This should be the case in a large city.

Example: When Were You Born?

STEP 4

Do: Since the conditions are satisfied, we can perform a chi-square goodness-of-fit test. We begin by calculating the test statistic.

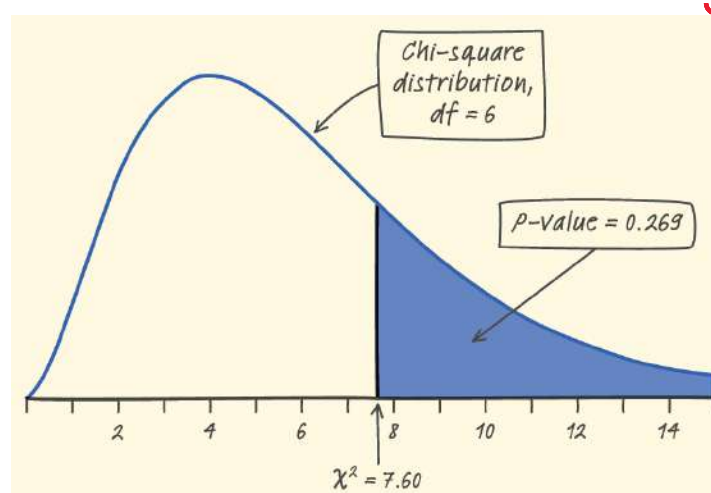
Test statistic:

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

$$= \frac{(13-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(24-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(27-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(15-20)^2}{20}$$

$$= 2.45 + 0.45 + 0.80 + 0.00 + 2.45 + 0.20 + 1.25$$

$$= 7.60$$



***P*-Value**:

Using Table C: χ^2 = 7.60 is less than the smallest entry in the df = 6 row, which corresponds to tail area 0.25. The *P*-value is therefore greater than 0.25.

Using technology: We can find the exact *P*-value with a calculator: $\chi^2\text{cdf}(7.60,1000,6) = 0.269$.

Conclude: Because the *P*-value, 0.269, is greater than α = 0.05, we fail to reject H_0. These 140 births don't provide enough evidence to say that all local births in this area are not evenly distributed across the days of the week.

Inherited ... Ex... Tr...

Biologi... ss pairs of tobacco plar... etic makeup Gg, indi... h plant has one dom...) and one recessive gen... Each offspring plant will ... ene for color from eac...

		Parent 2 passes on:	
		G	**g**
Parent 1 passes on:	**G**	GG	Gg
	g	Gg	gg

The Punnett square suggests that the expected ratio of green (GG) to yellow-green (Gg) to albino (gg) tobacco plants should be 1:2:1.
In other words, the biologists predict that 25% of the offspring will be green, 50% will be yellow-green, and 25% will be albino.

To test their hypothesis about the distribution of offspring, the biologists mate 84 randomly selected pairs of yellow-green parent plants.

Of 84 offspring, 23 plants were green, 50 were yellow-green, and 11 were albino.

Do these data differ significantly from what the biologists have predicted? Carry out an appropriate test at the $\alpha = 0.05$ level to help answer this question.

State: We want to perform a test of

H_0: The biologists' predicted color distribution for tobacco plant offspring is correct. That is, $p_{green} = 0.25$, $p_{yellow\text{-}green} = 0.5$, $p_{albino} = 0.25$

H_a: The biologists' predicted color distribution isn't correct. That is, at least one of the stated proportions is incorrect.

We will use $\alpha = 0.05$.

Plan: If the conditions are met, we should conduct a chi-square goodness-of-fit test.

- *Random* The data came from a random sample of local births.
- *Large Sample Size* We check that all expected counts are at least 5. Assuming H_0 is true, the expected counts for the different colors of offspring are green: (0.25)(84) = 21; yellow-green: (0.50)(84) = 42; albino: (0.25)(84) = 21
The complete table of observed and expected counts is shown below.
- *Independent* Individual offspring inherit their traits independently from one another. Since we are sampling without replacement, there would need to be at least 10(84) = 840 tobacco plants in the population. This seems reasonable to believe.

Offspring color	Observed	Expected
Green	23	21
Yellow-green	50	42
Albino	11	21

Example: Inherited Traits

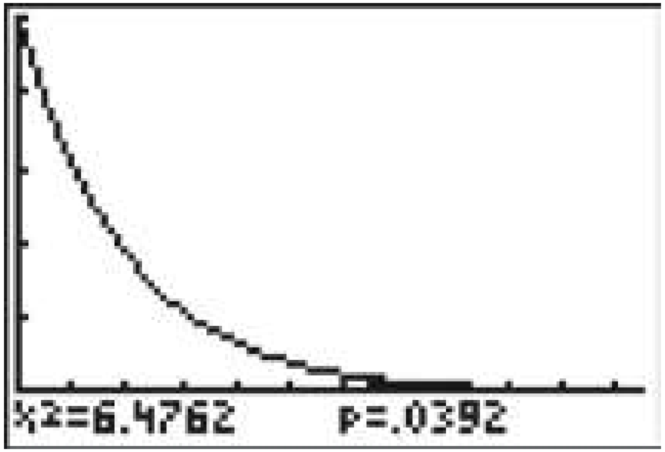
Do: Since the conditions are satisfied, we can perform a chi-square goodness-of-fit test. We begin by calculating the test statistic.

Test statistic:

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

$$= \frac{(23-21)^2}{21} + \frac{(50-42)^2}{50} + \frac{(11-21)^2}{21}$$

$$= 6.476$$



***P*-Value**:
Note that df = number of categories - 1 = 3 - 1 = 2. Using df = 2, the *P*-value from the calculator is 0.0392

Conclude: Because the *P*-value, 0.0392, is less than $\alpha = 0.05$, we will reject H_0. We have convincing evidence that the biologists' hypothesized distribution for the color of tobacco plant offspring is incorrect.

Follow-up Analysis

In the chi-square goodness-of-fit test, we test the null hypothesis that a categorical variable has a specified distribution. If the sample data lead to a statistically significant result, we can conclude that our variable has a distribution different from the specified one.

When this happens, start by examining which categories of the variable show large deviations between the observed and expected counts.

Then look at the individual terms that are added together to produce the test statistic χ^2. These **components** show which terms contribute most to the chi-square statistic.

In the tobacco plant example, we can see that the component for the albino offspring made the largest contribution to the chi - square statitstic.

$$\chi^2 = \frac{(23-21)^2}{21} + \frac{(50-42)^2}{50} + \frac{(11-21)^2}{21}$$

$$= 0.190 + 1.524 + 4.762 = 6.476$$

Offspring color	Observed	Expected
Green	23	21
Yellow-green	50	42
Albino	11	21

Section 11.1
Chi-Square Goodness-of-Fit Tests

Summary

In this section, we learned that…

- ✓ A **one-way table** is often used to display the distribution of a categorical variable for a sample of individuals.
- ✓ The **chi-square goodness-of-fit test** tests the null hypothesis that a categorical variable has a specified distribution.
- ✓ This test compares the **observed count** in each category with the counts that would be expected if H_0 were true. The **expected count** for any category is found by multiplying the specified proportion of the population distribution in that category by the sample size.
- ✓ The **chi-square statistic is**

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

where the sum is over all possible values of the categorical variable.

+

Section 11.1
Chi-Square Goodness-of-Fit Tests

Summary

- ✓ The test compares the value of the statistic χ^2 with critical values from the **chi-square distribution with** degrees of freedom df = number of categories - 1. Large values of χ^2 are evidence against H_0, so the *P*-value is the area under the chi-square density curve to the right of χ^2.

- ✓ The chi-square distribution is an approximation to the sampling distribution of the statistic χ^2. *You can safely use this approximation when all expected* cell counts are at least 5 (**Large Sample Size condition).**

- ✓ Be sure to check that the Random, Large Sample Size, and Independent conditions are met before performing a chi-square goodness-of-fit test.

- ✓ 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 Section…

We'll learn how to perform inference for relationships in distributions of categorical data.

We'll learn about

- ✓ **Comparing Distributions of a Categorical Variable**
- ✓ **The Chi-square Test for Homogeneity**
- ✓ **The Chi-square Test for Association/Independence**
- ✓ **Using Chi-square Tests Wisely**

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

1.) State null and alternative hypothesis

Ho: There is no difference in the categorical variable between the two (or more) samples

Ha: There is a difference in the categorical variable between the two (or more) samples

MUST BE IN CONTEXT OF QUESTION!

2.) “If conditions are met we will do a Chi-square Homogeneity test at α = [alpha level]”

3.) Conditions: Random, Independent, Large enough (all expected values at least 5)

4.) Find test statistic $\chi 2 = \Sigma$ (Observed – Expected)^2 / Expected

5.) State degrees of freedom = (rows - 1)(columns – 1)

6.) Use table or χ2cdf to find p-value

7.) Compare to Alpha α

8.) “We will” Reject (if p-value<α) or Fail to reject (if p-value>α) “null hypothesis”

9.) Write conclusion about alternative hypothesis in context of question.

Chi-Square Goodness-of-Fit Tests

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 ributions

esearchers suspect
ckground music may
he mood and buying
or of customers. One
n a supermarket
red three randomly
ed treatments: no
French accordion
and Italian string
Under each
on, the researchers
ed the numbers of
of French, Italian,
her wine purchased.
s a table that
arizes 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**

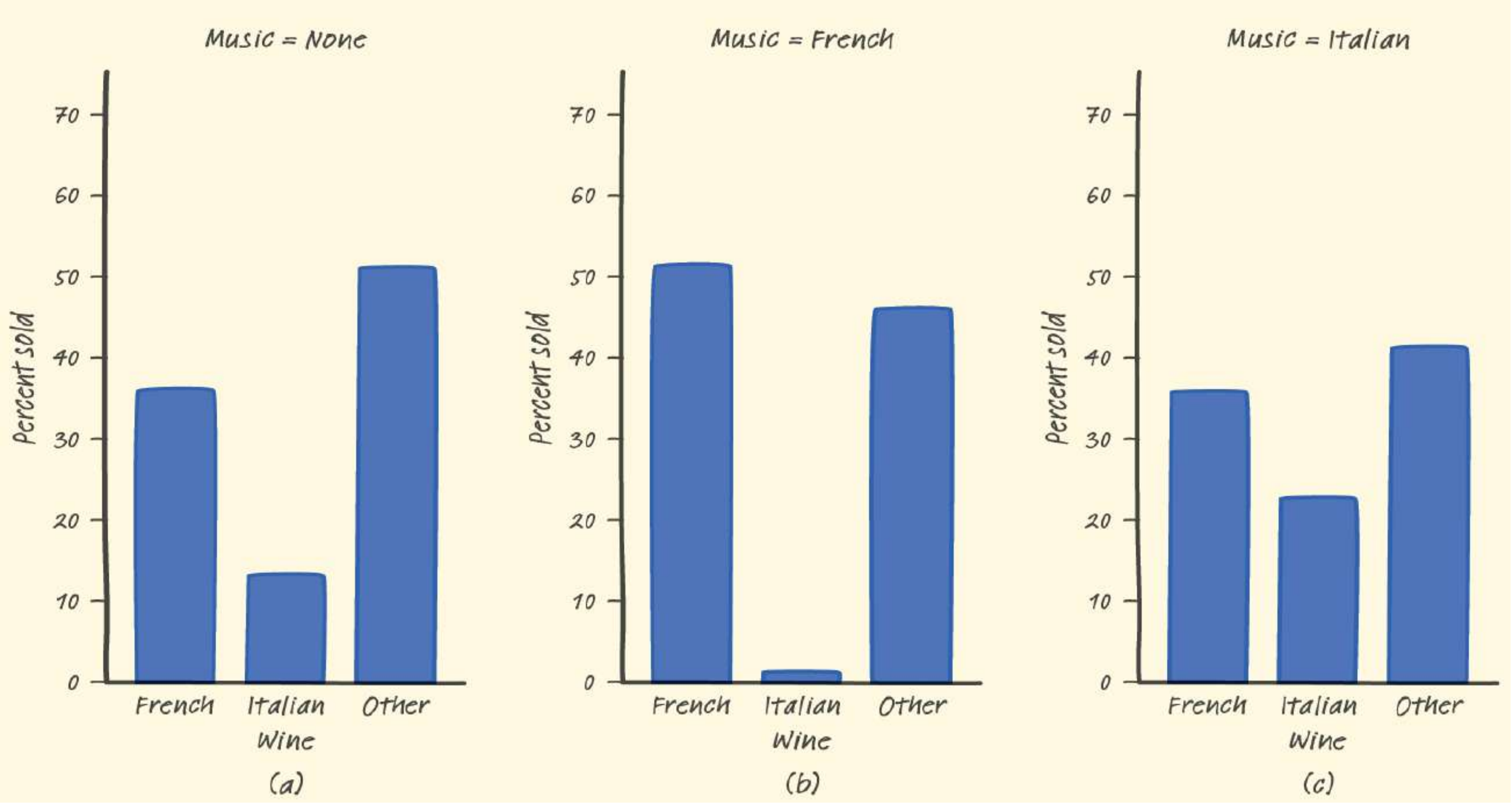
PROBLEM:

(a) Calculate the conditional distribution (in proportions) of the type of wine sold for each treatment.

(b) Make an appropriate graph for comparing the conditional distributions in part (a).

(c) 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.

Infe
renc
e
for
Rel
atio
nshi
ps



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.

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

H₀: There is no difference in the distribution of a categorical variable for several populations or treatments.
Hₐ: 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.

are

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:

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

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

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

Finding Expected Counts

nfe
enc
or
el
tio
nshi
ps

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

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

Observed Counts				
	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 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:

$$\frac{\quad \cdot \quad}{\qquad} = 34.22$$

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.

culatin he Chi- are tistic

es below the ved and cted counts e wine and iment. late the chi- e statistic.

Observed Counts				
	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**

Expected Counts				
	Music			
Wine	**None**	**French**	**Italian**	**Total**
French	34.22	30.56	34.22	**99**
Italian	10.72	9.57	10.72	**31**
Other	39.06	34.88	39.06	**113**
Total	**84**	**75**	**84**	**243**

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

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

The χ^2 statistic is the sum of nine such terms:

$$\chi^2 = \sum \frac{(\text{Observed - Expected})^2}{\text{Expected}} = \frac{(30-34.22)^2}{34.22} + \frac{(39-30.56)^2}{30.56} + \ldots + \frac{(35-39.06)^2}{39.06}$$

$$= 0.52 + 2.33 + \ldots + 0.42 = 18.28$$

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 χ^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 χ^2 under the corresponding chi-square density curve.

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.

	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**

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.

- The distribution of superpower preference for a random sample of 200 children (ages 9-17) from the United Kingdom who filled out a CensusAtSchool survey. Do American children have the same preferences? To find out, a random sample of 215 U.S. children (ages 9-17) was selected from those who filled out a CensusAtSchool survey (amstat.org/censusatschool/). Here are the results from both samples:**UKUSTotal**

Fly544599

Freeze Time524496

Invisibility303767

Super Strength202343

Telepathy4466110

Total200215415

Do these data provide convincing evidence that the distribution of superpower preference differs among U.S. and U.K. children?

We wa… rm a test of the following hypotheses using

Ho: The… erence in the distribution of superpower prefer… S. and U.K. children.

Ha: The… ence in the distribution of superpower prefer… S. and U.K. children.

If conditions are met, we will perform a chi-square test for homogeneity.

• *Random* The data are from separate random samples of U.K. and U.S. children.

• *Large Sample Size* The expected counts (listed below) are all at least 5.

• *Independent* The samples were independently selected. Also, since there are more than 10(200) = 2000 children in the UK and more than 10(215) children in the US, both samples are less than 10% of their populations.

Expected counts	UK	US	Total
Fly	47.7	51.3	99
Freeze Time	46.3	49.7	96
Invisibility	32.3	34.7	67
Super Strength	20.7	22.3	43
Telepathy	53.0	57.0	110
Total	200	215	415

Test Statistic

$$\chi^2 = \frac{(54 - 47.7)^2}{47.7} + \cdots = 6.29$$

P-value Using (5 – 1)(2 – 1) = 4 degrees of freedom,

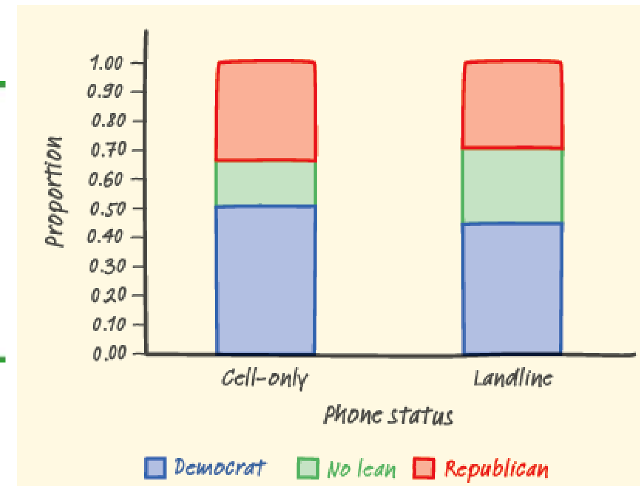
P-value = χ2cdf(6.29,1000,4)= 0.1784.

Because the *P*-value is greater than = 0.05, we fail to reject . We do not have convincing evidence that there is a difference in the distribution of superpower preference for U.S. and U.K. schoolchildren.

ple: Cell-Only hone 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

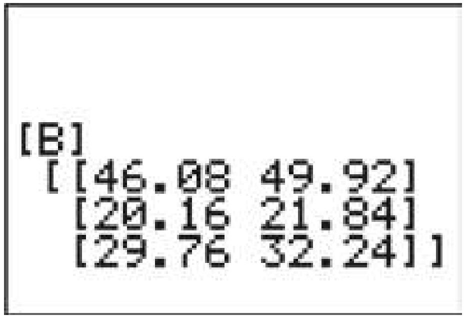
H_0: There is no difference in the distribution of party affiliation in the cell-only and landline populations.
H_a: There is a difference in the distribution of party affiliation in the cell-only and landline populations.

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 ≥ 5.



- *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.

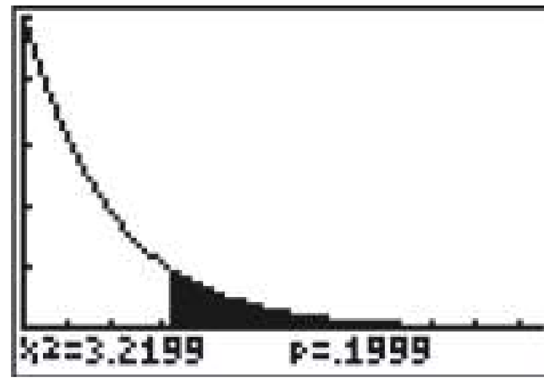
Do: Since the conditions are satisfied, we can a perform chi-test for homogeneity. We begin by calculating the test statistic.

Infe
renc
e
for
Rel
atio
shi
s

Test statistic:

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

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



***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.

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	39	30	99
	34.22	30.56	34.22	
	0.521	2.334	0.521	
2	11	1	19	31
	10.72	9.57	10.72	
	0.008	7.672	6.404	
3	43	35	35	113
	39.06	34.88	39.06	
	0.397	0.000	0.422	
Total	84	75	84	243

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

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 χ^2 = 18.28.

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

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: $p_1 = p_2 = \ldots = p_k$. 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."

The Chi-Square Test for Association/In

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 χ^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 χ^2 under the corresponding chi-square density curve.

1.) State null and alternative hypothesis

Ho: There is no association between the categorical variables in the population of interest

Ha: There is an association between the categorical variables in the population of interest

MUST BE IN CONTEXT OF QUESTION!

2.) "If conditions are met we will do a Chi-square Association test at α = [alpha level]"

3.) Conditions: Random, Independent, Large enough (all expected values at least 5)

4.) Find test statistic $\chi 2 = \Sigma$ (Observed – Expected)^2 / Expected

5.) State degrees of freedom = (rows - 1)(columns – 1)

6.) Use table or χ2cdf to find p-value

7.) Compare to Alpha α

8.) "We will" Reject (if p-value<α) or Fail to reject (if p-value>α) "null hypothesis"

9.) Write conclusion about alternative hypothesis in context of question.

A survey at CensusAtSchool recorded the relationship between gender and having allergies for a random sample of 40 students Here is a two-way table that summarizes the data:

	Female	Male	Total
Allergies	10	8	**18**
No Allergies	13	9	**22**
Total	**23**	**17**	**40**

Do the data provide convincing evidence of an association between gender and having allergies for U.S. high school students who filled out the CensusAtSchool survey?

We want to perform a test of the following hypotheses using = 0.05:

Ho: There is no association between gender and having allergies in the population of U.S. high school students who filled out the CensusAtSchool survey.

Ha: There is an association between gender and having allergies in the population of U.S. high school students who filled out the CensusAtSchool survey.

If conditions are met, we will perform a chi-square test for association.

Random The sample was randomly selected.

Large Sample Size The expected counts (see table above) are all at least 5.

Independent Knowing the responses of one student shouldn't tell us anything about the responses of other students. Also, there are more than 10(40) = 400 high school students in the U.S. who filled out the CensusAtSchool survey.

+

Test Statistic: $\chi^2 = \frac{(10-10.35)^2}{10.35} + \Lambda = 0.051$

P-value Using (2 – 1)(2 – 1) = 1 degrees of freedom,

•*P*-value = χ^2cdf(0.051,1000,1) = 0.821.

Because the *P*-value = 0.821 is much greater than α = 0.05

We fail to reject H_0

We do not have convincing evidence that there is an association between gender and having allergies in the population of U.S. high school students who filled out the CensusAtSchool survey.

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 <u>two</u> or more independent random samples or treatment groups in a randomized experiment, then do a chi-square test for <u>homogeneity</u>.**

✓**If the data come from a <u>single</u> random sample, with the individuals classified according to two categorical variables, use a chi-square test for <u>association/independence</u>.**

Observed Grades

Class\homework	A/B/C	D/F
A	7	0
B	8	0
C	3	2
D	1	4
F	0	1

Expected Grades

Class\homework	A/B/C	D/F
A	5.1	1.9
B	5.8	2.2
C	3.7	1.3
D	3.7	1.3
F	.7	.3

Ho: There is no association between homework grades and class grades in AP Statistics.
Ha: There is an association between homework grades and class grades in AP Statistics.

$X^2 = 15.8$ and the p-value = .003

Since the p-value $.003 < \alpha = 0.05$, we reject the Ho. There is sufficient evidence to support the claim that there is an association between homework grades and class grades in AP Statistics.

Due to our small sample size all our expected grades are not at least 5, we must proceed with caution on our inference

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 χ^2 with critical values from the chi-square distribution with df = (number of rows - 1)(number of columns - 1). Large values of χ^2 are evidence against H_0, so the *P*-value is the area under the chi-square density curve to the right of χ^2.

Section 11.1
Chi-Square Goodness-of-Fit Tests

Summary

- ✓ The chi-square distribution is an approximation to the distribution of the statistic χ^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**