# 11.2 Inference for Two-Way Tables

## 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.

May 4-10:43 AM

## Comparing Distributions of a Categorical Variable

Market researchers suspect that background music may affect the mood and buying behavior of customers.

One study in a Mediterranean restaurant compared three randomly assigned treatments: no music, French accordion music, and Italian string music.

Under each condition, the researchers recorded the numbers of customers who ordered French, Italian, and other entrees.

| | Type of Music | | | |
|---|---|---|---|---|
| **Entree ordered** | **None** | **French** | **Italian** | **Total** |
| French | 30 | 39 | 30 | **99** |
| Italian | 11 | 1 | 19 | **31** |
| Other | 43 | 35 | 35 | **113** |
| **Total** | **84** | **75** | **84** | **243** |

May 4-10:43 AM

## Comparing Distributions of a Categorical Variable

**Problem**: (a) Calculate the conditional distribution (in proportions) of the entree ordered for each treatment.

(a) When no music was playing, the distribution of entree orders was

French: $\frac{30}{84} = 0.357$ Italian: $\frac{11}{84} = 0.131$ Other: $\frac{43}{84} = 0.512$

When French accordion music was playing, the distribution of entree orders was

French: $\frac{39}{75} = 0.520$ Italian: $\frac{1}{75} = 0.013$ Other: $\frac{35}{75} = 0.467$

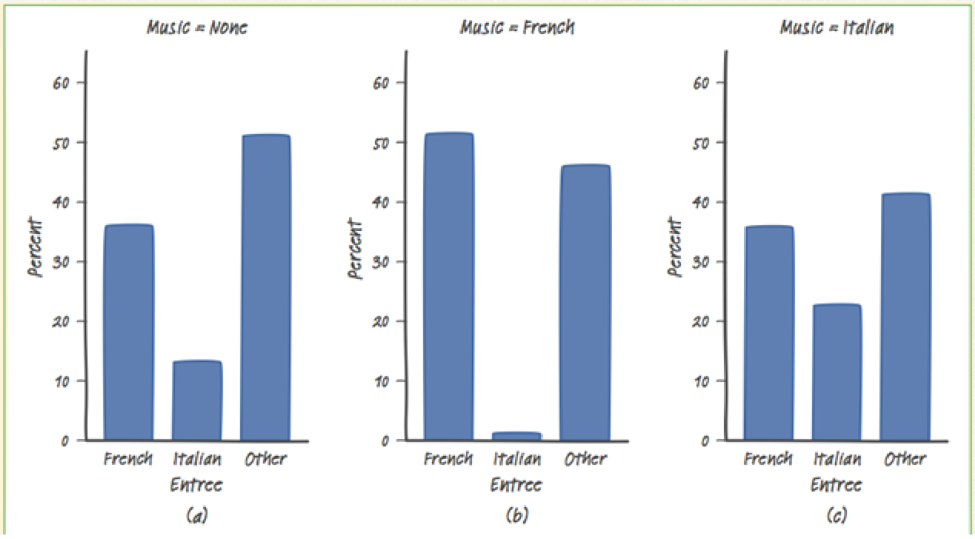
When Italian string music was playing, the distribution of entree orders was

French: $\frac{30}{84} = 0.357$ Italian: $\frac{19}{84} = 0.226$ Other: $\frac{35}{84} = 0.417$

May 4-10:43 AM

## Comparing Distributions of a Categorical Variable

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




May 4-10:44 AM

## Comparing Distributions of a Categorical Variable

**Problem**: (c) Write a few sentences comparing the distributions of entrees ordered under the three music treatments.

The type of entrée that customers buy seems to differ considerably across the three music treatments.

Orders of Italian entrees 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 entrees seem popular in this restaurant, as they are ordered frequently under all music conditions but notably more often when French music is playing.

For all three music treatments, the percent of Other entrees ordered was similar.

May 4-10:44 AM

## Comparing Distributions of a Categorical Variable

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.

May 4-10:44 AM

## Expected Counts and the Chi-Square Statistic

A chi-square test for homogeneity begins with the hypotheses

$H_0$: 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.

May 4-10:44 AM

## Expected Counts and the Chi-Square Statistic

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

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

**Observed Counts**

| | Type of Music | | | |
|---|---|---|---|---|
| Entree ordered | None | French | Italian | Total |
| French | 30 | 39 | 30 | $\boxed{99}$ |
| Italian | 11 | 1 | 19 | **31** |
| Other | 43 | 35 | 35 | **113** |
| **Total** | $\boxed{84}$ | **75** | **84** | $\boxed{243}$ |

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:

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

May 4-10:45 AM

## Expected Counts and the Chi-Square Statistic

**Finding Expected Counts**

When $H_0$ is true, the expected count in any cell of a two-way table is

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

**Conditions for Performing a Chi-Square Test for Homogeneity**

- **Random**: The data come a well-designed random sample or from a randomized experiment.
  - **10%**: When sampling without replacement, check that $n \leq (1/10)N$.
- **Large Counts**: All *expected* counts are greater than 5

May 4-10:46 AM

## Expected Counts and the Chi-Square Statistic

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.

Observed Counts

| | Type of Music | | | |
|---|---|---|---|---|
| Entree ordered | 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

| | Type of Music | | | |
|---|---|---|---|---|
| Entree ordered | 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** |

May 4-10:46 AM

## Expected Counts and the Chi-Square Statistic

| Observed Counts | | | | |
|---|---|---|---|---|
| | Type of Music | | | |
| Entree ordered | 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 | | | | |
|---|---|---|---|---|
| | Type of Music | | | |
| Entree ordered | 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 entrees with no music, the observed count is 30 orders and the expected count is 34.22. The contribution to the $\chi^2$ statistic for this cell is

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

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

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

May 4-10:46 AM

## *P*-value and conclusion

Earlier, we started a significance test of

$H_0$: There is no difference in the true distributions of entrees ordered at this restaurant when no music, French accordion music, or Italian string music is played.

$H_a$: There is a difference in the true distributions of entrees ordered at this restaurant when no music, French accordion music, or Italian string music is played.

| | Type of Music | | | |
|---|---|---|---|---|
| Entree ordered | 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 already checked that the conditions are met. Our calculated test statistic is $\chi^2$ = 18.28.

May 4-10:46 AM

## Example: *P*-value and conclusion

**Problem**: (a) Use Table C to find the *P*-value. Then use your calculator's $\chi^2$cdf command.

(a) Because the two-way table has three rows and three columns that contain the data from the study, we use a chi-square distribution with df = (3 - 1)(3 - 1) = 4 to find the *P*-value.

entrees treatment

df = (rows - 1) times (col - 1)

| | *P* | |
|---|---|---|
| df | .0025 | .001 |
| 4 | 16.42 | 18.47 |

The value $\chi^2$=18.28 falls between the critical values 16.42 and 18.47.

The corresponding *P*-value is between 0.0025 and 0.01.

L U df

*Calculator*: The command $\chi^2$cdf(18.28,1000,4) gives 0.0011.

May 4-10:53 AM

## Example: *P*-value and conclusion

**Problem**: (b) Interpret the *P*-value from the calculator in context.

Assuming that there is no difference in the true distributions of entrees ordered in this restaurant when no music, French accordion music, or Italian string music is played, there is a 0.0011 probability of observing a difference in the distributions of entrees ordered among the three treatment groups as large or larger than the ones in this study.

**Problem**: (c) What conclusion would you draw? Justify your answer

Because the *P*-value, 0.0011, is less than our default α= 0.05 significance level, we reject $H_0$. We have convincing evidence of a difference in the distributions of entrees ordered at this restaurant 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.

May 4-10:55 AM

# Chi-Square Test for Homogeneity

**Chi-Square Test for Homogeneity**

Suppose the 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 count for each category assuming that $H_0$ is true. 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.

May 4-10:55 AM

## 27. TECHNOLOGY CORNER CHI-SQUARE TESTS FOR TWO-WAY TABLES ON THE CALCULATOR

You can use the TI-83/84 or TI-89 to perform calculations for a chi-square test for homogeneity. We'll use the data from the restaurant study to illustrate the process.

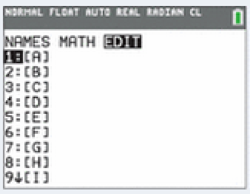
**1.** Enter the observed counts in matrix [A].

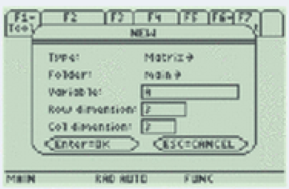
TI-83/84

- Press [2nd] [$x^{-1}$] (MATRIX), arrow to EDIT, and choose A.
- Enter the dimensions of the matrix: 3 × 3.

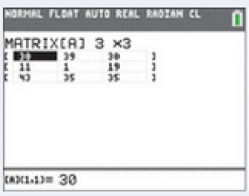
TI-89

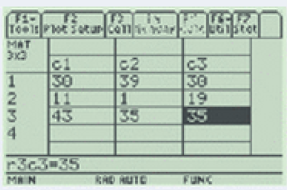
- Press [APPS], select Data/Matrix Editor and then New. . . .
- Adjust your settings to match those shown.



- Enter the observed counts from the two-way table in the same locations in the matrix.



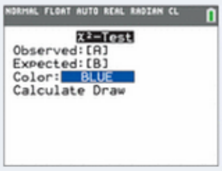



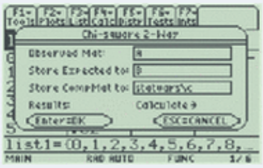

Apr 22-10:39 AM

**2.** Specify the chi-square test, the matrix where the observed counts are found, and the matrix where the expected counts will be stored.

- Press [STAT], arrow to TESTS, and choose $\chi^2$-Test.
- Adjust your settings as shown.

- In the Statistics/List Editor, press [2nd] [F1] ([F6]), and choose Chi2 2-way. . .
- Adjust your settings as shown.



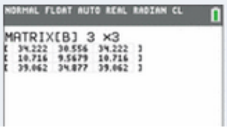



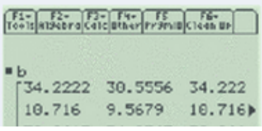
**3.** Choose "Calculate" or "Draw" to carry out the test. If you choose "Calculate," you should get the test statistic, *P*-value, and df shown below. If you specify "Draw," the chi-square distribution with 4 degrees of freedom will be drawn, the area in the tail will be shaded, and the *P*-value will be displayed.

**4.** To see the expected counts, go to the home screen and ask for a display of the matrix [B].

Press [2nd] [$X^{-1}$] (MATRIX), arrow to EDIT, and choose [B].

- Press [2nd] [–] (Var-LINK) and choose B.






Apr 22-10:40 AM

*Birthdays*
Has modern technology changed the distribution of birthdays? With more babies being delivered by planned c-section, Mrs. McDonald's statistics class hypothesized that the day-of-the-week distribution for births would be different for people born after 1993 compared to people born before 1980. After all, why would a doctor plan a c-section for the weekend? To investigate, they selected a random sample of people from each both age categories and recorded the day of the week on which they were born. The results are shown in the table. Is there convincing evidence that the distribution of birth days has changed?

| | Before 1980 | After 1993 | Total |
|---|---|---|---|
| Sunday | 12 | 9 | 21 |
| Monday | 12 | 11 | 23 |
| Tuesday | 14 | 11 | 25 |
| Wednesday | 12 | 10 | 22 |
| Thursday | 7 | 17 | 24 |
| Friday | 9 | 9 | 18 |
| Saturday | 11 | 6 | 17 |
| **Total** | 77 | 73 | 150 |

df $(7-1)(2-1)$
$= 6$

Expected

| | |
|---|---|
| 10.8 | 10.2 |
| 11.8 | 11.2 |
| 12.8 | 12.2 |
| 11.3 | 10.7 |
| 12.3 | 11.7 |
| 9.24 | 8.8 |
| 8.7 | 8.3 |

L.C. all $\geq 5$ ✓

$H_0$: There is no difference in the dist of birthdays before 1980 & after 1993
$H_a$: There is a difference in the dist of birthdays before 1980 & after 1993

random sample ✓
10% ✓ >770 born before 1980
>730 born after 1993

$\chi^2 = 6.55$ pval $= .3646 > \alpha = .05$

Fail to Reject $H_0$
No convincing evidence there is a difference in the dist of birthdays

HW 11.2 P1
27-39 odd
Due W

May 4-11:16 AM

## CHECK YOUR UNDERSTANDING

In the previous Check Your Understanding (page 699), we presented data on the use of Facebook by two randomly selected groups of Penn State students. Here are the data once again.

| Use Facebook | Main campus | Commonwealth |
|---|---|---|
| Several times a month or less | 55 | 76 |
| At least once a week | 215 | 157 |
| At least once a day | 640 | 394 |
| **Total Facebook users** | **910** | **627** |

Do these data provide convincing evidence of a difference in the distributions of Facebook use among students in the two campus settings?

Apr 22-10:40 AM

*The color of candy*

Inspired by the *Does Background Music Influence What Customers Buy?* example, a statistics student decided to investigate other ways to influence a person's behavior. Using 60 volunteers she randomly assigned 20 volunteers to get the "red" survey, 20 volunteers to get the "blue" survey, and 20 volunteers to get a control survey. The first three questions on each survey were the same, but the fourth and fifth questions were different. For example, the fourth question on the "red" survey was "When you think of the color red, what do you think about?" On the blue survey, the question replaced red with blue. On the control survey, the questions were not about color. As a reward, the student let each volunteer choose a chocolate candy in a red wrapper or a chocolate candy in a blue wrapper. Here are the results.

| | **Red survey** | **Blue survey** | **Control survey** | **Total** |
|---|---|---|---|---|
| Red candy | 13 | 5 | 8 | **26** |
| Blue candy | 7 | 15 | 12 | **34** |
| **Total** | **20** | **20** | **20** | **60** |

(a) Compare the distributions of color choice for each of the three treatments.
(b) Do these data provide convincing evidence at the  = 0.05 level that the true distributions of color choice are different for the three types of surveys?

May 4-11:18 AM

## Relationships Between 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 based on two categorical variables.

In that case, our goal is to analyze the relationship between the variables.

indep

Our null hypothesis is that there is no association between the two categorical variables in the population of interest.

not indep

The alternative hypothesis is that there is an association between the variables.

$\chi^2$ GOF

1 var
1 pop
fit dist?

$\chi^2$

Homogeneity
1 var
2 or more grps

Indep
2 var
1 grp

May 4-10:55 AM

## The Chi-Square Test for Independence

The 10% and Large Counts conditions for the chi-square test for independence are the same as for the homogeneity test.

There is a slight difference in the Random condition for the two tests: a test for independence uses data from one sample but a test for homogeneity uses data from two or more samples/groups.

**Conditions for Performing a Chi-Square Test for Independence**

- **Random**: The data come a well-designed random sample or from a randomized experiment.
  - **10%**: When sampling without replacement, check that $n \leq (1/10)N$.
- **Large Counts**: All *expected* counts are greater than 5

May 4-10:56 AM

# Chi-Square Test for Independence

**Chi-Square Test for Independence**

Suppose the conditions are met. You can use the **chi-square test for 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.

Start by finding the expected count for each category assuming that $H_0$ is true. 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.

May 4-10:56 AM

*Finger length*

Is there a relationship between gender and relative finger length? In Chapter 5, we looked at a sample of 452 U.S. high school students who completed a survey. The two-way table shows the gender of each student and which finger was longer on their left hand (index finger or ring finger).

| | Female | Male | **Total** |
|---|---|---|---|
| Index finger | 78 | 45 | **123** |
| Ring finger | 82 | 152 | **234** |
| Same length | 52 | 43 | **95** |
| **Total** | **212** | **240** | **452** |

$df = (3-1)(2-1)$
$= 2$

**Problem:**

(a) Is this an observational study or an experiment? Justify your answer.

(b) Do the data provide convincing evidence of an association between gender and relative finger length for U.S. high school students who filled out the CensusAtSchool survey?

$\chi^2$ Indep

$H_0$: There is no association between gender & relative finger length

$H_a$: There is an association between gender & relative finger length

10% ✓ >4520 US HS students

random sample ✓

Expected all >5 ✓

$$\begin{bmatrix} 57.7 & 65.3 \\ 109.8 & 124.3 \\ 44.6 & 50.4 \end{bmatrix}$$

$\chi^2 = 29.02$

pval ≈ 0

df = 2

pval < $\alpha$ = .05 Reject $H_0$

We have convincing evidence of a relationship between gender & relative finger length

May 4-11:20 AM

## Do Dogs Resemble Their Owners?

Some people look a lot like their pets. Maybe they deliberately choose animals that match their appearance. Or maybe we're perceiving similarities that aren't really there. Researchers at the University of California, San Diego, decided to investigate. They designed an experiment to test whether or not dogs resemble their owners. The researchers believed that resemblance between dog and owner might differ for purebred and mixed-breed dogs.

A random sample of 45 dogs and their owners was photographed separately at three dog parks. Then, researchers "constructed triads of pictures, each consisting of one owner, that owner's dog, and one other dog photographed at the same park." The subjects in the experiment were 28 undergraduate psychology students. Each subject was presented with the individual sets of photographs and asked to identify which dog belonged to the pictured owner. A dog was classified as resembling its owner if more than half of the 28 undergraduate students matched dog to owner.[1]

The table below summarizes the results. There is some support for the researchers' belief that resemblance between dog and owner might differ for purebred and mixed-breed dogs.

| | Breed status | |
|---|---|---|
| **Resemblance?** | **Purebred dogs** | **Mixed-breed dogs** |
| Resemble owner | 16 | 7 |
| Don't resemble | 9 | 13 |

Do these data provide convincing evidence of an association between dogs' breed status and whether or not they resemble their owners? By the end of this chapter, you will have developed the tools you need to answer this question.

Apr 24-3:45 PM

## Example: Choosing the right type of chi-square test

Are men and women equally likely to suffer lingering fear from watching scary movies as children? Researchers asked a random sample of 117 college students to write narrative accounts of their exposure to scary movies before the age of 13. More than one-fourth of the students said that some of the fright symptoms are still present when they are awake. The following table breaks down these results by gender.

| | Gender | | |
|---|---|---|---|
| Fright symptoms? | Male | Female | **Total** |
| Yes | 7 | 29 | **36** |
| No | 31 | 50 | **81** |
| **Total** | **38** | **79** | **117** |

May 4-10:56 AM

## Example: Choosing the right type of chi-square test

Minitab output for a chi-square test using these data is shown below.

```
Chi-Square Test: Male, Female
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts
              Male       Female      Total
1   Ob           7    Ob    29          36
    Ex       11.69    Ex 24.31
 Contrib     1.883    C  0.906
2   Ob          31    Ob    50          81
    Ex       26.31    Ex 54.69
    C        0.837    C  0.403
Total           38          79         117
Chi-Sq = 4.028, DF = 1, P-Value = 0.045
```

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

May 4-10:56 AM

## Example: Choosing the right type of chi-square test

**Problem:** Assume that the conditions for performing inference are met.

(a) Explain why a chi-square test for independence and not a chi-square test for homogeneity should be used in this setting.

The data were produced using a single random sample of college students, who were then classified by gender and whether or not they had lingering fright symptoms.

The chi-square test for homogeneity requires independent random samples from each population.

May 4-10:57 AM

## Example: Choosing the right type of chi-square test

**Problem:** Assume that the conditions for performing inference are met.

(b) State an appropriate pair of hypotheses for researchers to test in this setting.

The null hypothesis is

$H_0$: There is no association between gender and ongoing fright symptoms in the population of college students. The alternative hypothesis is

$H_a$: There is an association between gender and ongoing fright symptoms in the population of college students.

May 4-10:57 AM

## Example: Choosing the right type of chi-square test

**Problem:** Assume that the conditions for performing inference are met.

(d) Interpret the *P*-value in context. What conclusion would you draw at $\alpha = 0.01$?

**Chi-Square Test: Male, Female**

Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

| | Male | Female | Total |
|---|---|---|---|
| 1 | 7 | 29 | 36 |
| | 11.69 | 24.31 | |
| | 1.883 | 0.906 | |
| 2 | 31 | 50 | 81 |
| | 26.31 | 54.69 | |
| | 0.837 | 0.403 | |
| Total | 38 | 79 | 117 |

Chi-Sq = 4.028, DF = 1, P-Value = 0.045

If gender and ongoing fright symptoms really are independent in the population of interest, there is a 0.045 chance of obtaining a random sample of 117 students that gives a chi-square statistic of 4.028 or higher. Because the *P*-value, 0.045, is greater than 0.01, we would fail to reject $H_0$. We do not have convincing evidence that there is an association between gender and fright symptoms in the population of college students.

May 4-10:57 AM

*Online social networking*
An article in the *Arizona Daily Star* (April 9, 2009) included the following table:

| Age (years): | 18–24 | 25–34 | 35–44 | 45–54 | 55–64 | 65+ | Total |
|---|---|---|---|---|---|---|---|
| Use online social networks: | 137 | 126 | 61 | 38 | 15 | 9 | **386** |
| Do not use online social networks: | 46 | 95 | 143 | 160 | 130 | 124 | **698** |
| **Total:** | **183** | **221** | **204** | **198** | **145** | **133** | **1084** |

Suppose that you decide to analyze these data using a chi-square test. However, without any additional information about how the data were collected, it isn't possible to know which chi-square test is appropriate.

**Problem:**
(a) Explain how you know that a test for goodness of fit is *not* appropriate for analyzing these data.
(b) Describe how these data could have been collected so that a test for homogeneity is appropriate.
(c) Describe how these data could have been collected so that a test for independence is appropriate.

a) more than 1 variable (2 way table)

b) Take 2 indep r.s. (for online & not)
ask age OR
Take 6 indep r.s. (for age)
ask online

c) Take 1 r.s.
ask online & age

HW 11.2 P2
41-49 odd
51-55