# Chapter 16
# Analysis of Categorical Data

## LEARNING OBJECTIVES

The overall objective of this chapter is to give you an understanding of two statistical techniques used to analyze categorical data, thereby enabling you to:

1. Use the chi-square goodness-of-fit test to analyze probabilities of multinomial distribution trials along a single dimension.

2. Use the chi-square test of independence to perform contingency analysis

## CHAPTER OUTLINE

16.1 Chi-Square Goodness-of-Fit Test

16.2 Contingency Analysis: Chi-Square Test of Independence

## KEY TERMS

Categorical Data
Chi-Square Distribution
Chi-Square Goodness-of-Fit Test
Chi-Square Test of Independence
Contingency Analysis
Contingency Table

## STUDY QUESTIONS

1. Statistical techniques based on assumptions about the population from which the sample data are selected are called ________________ statistics.

2. Statistical techniques based on fewer assumptions about the population and the parameters are called ________________ statistics.

3. A chi-square goodness-of-fit test is being used to determine if the observed frequencies from seven categories are significantly different from the expected frequencies from the seven categories. The degrees of freedom for this test are ________________.

4. A value of alpha = .05 is used to conduct the test described in question 3. The critical table chi-square value is ________________.

5. A variable contains five categories. It is expected that data are uniformly distributed across these five categories. To test this, a sample of observed data is gathered on this variable resulting in frequencies of 27, 30, 29, 21, 24. A value of .01 is specified for alpha. The degrees of freedom for this test are ________________.

6. The critical table chi-square value of the problem presented in question 5 is ________________.

7. The observed chi-square value for the problem presented in question five is ________________. Based on this value and the critical chi-square value, a researcher would decide to ________________ the null hypothesis.

8. A researcher believes that a variable is Poisson distributed across six categories. To test this, a random sample of observations is made for the variable resulting in the following data:

   | Number of arrivals | Observed |
   |---|---|
   | 0 | 47 |
   | 1 | 56 |
   | 2 | 38 |
   | 3 | 23 |
   | 4 | 15 |
   | 5 | 12 |

   Suppose alpha is .10, the critical table chi-square value used to conduct this chi-square goodness-of-fit test is ________________.

9. The value of the observed chi-square for the data presented in question 8 is ________________.

   Based on this value and the critical value determined in question 8, the decision of the researcher is to ________________ the null hypothesis.

10. The degrees of freedom used in conducting a chi-square goodness-of-fit test to determine if a distribution is normally distributed are ________________.

11. In using the chi-square goodness-of-fit test, a statistician needs to make certain that none of the expected values are less than ________________.

12. The chi-square _____________________ is used to analyze frequencies of two variables with multiple categories.

13. A two-way frequency table is sometimes referred to as a ________________ table.

14. Suppose a researcher wants to use the data below and the chi-square test of independence to determine if variable one is independent of variable two.

|  |  | Variable One |  |  |
|---|---|---|---|---|
|  |  | A | B | C |
| Variable | D | 25 | 40 | 60 |
| Two | E | 10 | 15 | 20 |

The expected value for the cell of D and B is ________________.

15. The degrees of freedom for the problem presented in question 16 are ________________.

16. If alpha is .05, the critical chi-square value for the problem presented in question 16 is ________________.

17. The observed value of chi-square for the problem presented in question 16 is ________________. Based on this observed value of chi-square and the critical chi-square value determined in question 18, the researcher should decide to ________________ the null hypothesis that the two variables are independent.

18. A researcher wants to statistically determine if variable three is independent of variable four using the observed data given below:

|  |  | Variable Three |  |
|---|---|---|---|
|  |  | A | B |
| Variable | C | 92 | 70 |
| Four | D | 112 | 145 |

If alpha is .01, the critical chi-square table value for this problem is

________________.

19. The observed chi-square value for the problem presented in question 20 is ________________. Based on this value and the critical value determined in question 20, the researcher should decide to ________________ the null hypothesis.

## ANSWERS TO STUDY QUESTIONS

1. Parametric Statistics
2. Nonparametric Statistics
3. 6
4. 12.5916
5. 4
6. 13.2767
7. 2.091, Fail to Reject
8. 7.77944
9. 16.2, Reject
10. $k - 3$
11. 5
12. Test of Independence
13. Contingency
14. 40.44
15. 2
16. 5.9915
17. .19, Fail to Reject
18. 6.6349
19. 6.945, Reject

## SOLUTIONS TO THE ODD-NUMBERED PROBLEMS IN CHAPTER 16

16.1

| $f_0$ | $f_e$ | $\frac{(f_0 - f_e)^2}{f_0}$ |
|---|---|---|
| 53 | 68 | 3.309 |
| 37 | 42 | 0.595 |
| 32 | 33 | 0.030 |
| 28 | 22 | 1.636 |
| 18 | 10 | 6.400 |
| 15 | 8 | 6.125 |

$H_o$: The observed distribution is the same as the expected distribution.

$H_a$: The observed distribution is not the same as the expected distribution.

Observed $\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$ = **18.095**

df = $k$ - 1 = 6 - 1 = 5, $\alpha$ = .05

$\chi^2_{.05,5}$ = 11.0705

Since the observed $\chi^2$ = 18.095 > $\chi^2_{.05,5}$ = 11.0705, the decision is to **reject the null hypothesis**.

The observed frequencies are not distributed the same as the expected frequencies.

16.3

| Number | $f_0$ | (Number)($f_0$) |
|---|---|---|
| 0 | 28 | 0 |
| 1 | 17 | 17 |
| 2 | 11 | 22 |
| 3 | 5 | 15 |
| | 61 | 54 |

$H_o$: The frequency distribution is Poisson.
$H_a$: The frequency distribution is not Poisson.

$\lambda = \frac{54}{61} = 0.9$

| Number | Expected Probability | Expected Frequency |
|---|---|---|
| 0 | .4066 | 24.803 |
| 1 | .3659 | 22.320 |
| 2 | .1647 | 10.047 |
| $\geq 3$ | .0628 | 3.831 |

Since $f_e$ for $\geq 3$ is less than 5, collapse categories 2 and $\geq 3$:

| Number | $f_o$ | $f_e$ | $\frac{(f_0 - f_e)^2}{f_0}$ |
|---|---|---|---|
| 0 | 28 | 24.803 | 0.412 |
| 1 | 17 | 22.320 | 1.268 |
| $\geq 2$ | 16 | 13.878 | 0.324 |
| | 61 | 60.993 | 2.004 |

df = $k$ - 2 = 3 - 2 = 1, $\alpha$ = .05

$\chi^2_{.05,1} = 3.8415$

Observed $\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$ = **2.001**

Since the observed $\chi^2 = 2.001 < \chi^2_{.05,1} = 3.8415$, the decision is to **fail to reject the null hypothesis**.

**There is insufficient evidence to reject the distribution as Poisson distributed**. The conclusion is that the distribution is Poisson distributed.

16.5

| Definition | $f_o$ | Exp.Prop. | $f_e$ | $\frac{(f_0 - f_e)^2}{f_0}$ |
|---|---|---|---|---|
| Happiness | 42 | .39 | 227(.39)= 88.53 | 24.46 |
| Sales/Profit | 95 | .12 | 227(.12)= 27.24 | 168.55 |
| Helping Others | 27 | .18 | 40.86 | 4.70 |
| Achievement/ Challenge | 63 | .31 | 70.37 | 0.77 |
| | 227 | | | 198.48 |

$H_o$: The observed frequencies are distributed the same as the expected frequencies.

$H_a$: The observed frequencies are not distributed the same as the expected frequencies.

Observed $\chi^2$ = **198.48**

df = $k$ – 1 = 4 – 1 = 3, $\alpha$ = .05

$\chi^2_{.05,3} = 7.8147$

Since the observed $\chi^2 = 198.48 > \chi^2_{.05,3} = 7.8147$, the decision is to **reject the null hypothesis**.

**The observed frequencies for men are not distributed the same as the expected frequencies which are based on the responses of women**.

16.7

| Age | $f_o$ | $m$ | $fm$ | $fm^2$ |
|---|---|---|---|---|
| 10-20 | 16 | 15 | 240 | 3,600 |
| 20-30 | 44 | 25 | 1,100 | 27,500 |
| 30-40 | 61 | 35 | 2,135 | 74,725 |
| 40-50 | 56 | 45 | 2,520 | 113,400 |
| 50-60 | 35 | 55 | 1,925 | 105,875 |
| 60-70 | 19 | 65 | 1,235 | 80,275 |
| | 231 | | $\Sigma fm$ = 9,155 | $\Sigma fm^2$ = 405,375 |

$$\bar{x} = \frac{\sum fM}{n} = \frac{9{,}155}{231} = 39.63$$

$$s = \sqrt{\frac{\sum fM^2 - \frac{(\sum fM)^2}{n}}{n-1}} = \sqrt{\frac{405{,}375 - \frac{(9{,}155)^2}{231}}{230}} = 13.6$$

$H_o$: The observed frequencies are normally distributed.
$H_a$: The observed frequencies are not normally distributed.

| For Category 10-20 | Prob |
|---|---|
| $z = \frac{10 - 39.63}{13.6} = -2.18$ | .4854 |
| $z = \frac{20 - 39.63}{13.6} = -1.44$ | -.4251 |
| Expected prob. | .0603 |

| For Category 20-30 | Prob |
|---|---|
| for $x$ = 20, $z$ = -1.44 | .4251 |
| $z = \frac{30 - 39.63}{13.6} = -0.71$ | -.2611 |
| Expected prob. | .1640 |

| For Category 30-40 | Prob |
|---|---|
| for $x$ = 30, $z$ = -0.71 | .2611 |
| $z = \frac{40 - 39.63}{13.6} = 0.03$ | +.0120 |

Expected prob. .2731

| For Category 40-50 | Prob |
|---|---|
| $z = \dfrac{50-39.63}{13.6} = 0.76$ | .2764 |
| for $x$ = 40, $z$ = 0.03 | -.0120 |
| Expected prob. | .2644 |

| For Category 50-60 | Prob |
|---|---|
| $z = \dfrac{60-39.63}{13.6} = 1.50$ | .4332 |
| for $x$ = 50, $z$ = 0.76 | -.2764 |
| Expected prob. | .1568 |

| For Category 60-70 | Prob |
|---|---|
| $z = \dfrac{70-39.63}{13.6} = 2.23$ | .4871 |
| for $x$ = 60, $z$ = 1.50 | -.4332 |
| Expected prob. | .0539 |

For < 10:
Probability between 10 and the mean = .0603 + .1640 + .2611 = .4854
Probability < 10 = .5000 - .4854 = .0146

For > 70:
Probability between 70 and the mean = .0120 + .2644 + .1568 + .0539 = .4871
Probability > 70 = .5000 - .4871 = .0129

| Age | Probability | | $f_e$ |
|---|---|---|---|
| < 10 | .0146 | (.0146)(231) = | 3.37 |
| 10-20 | .0603 | (.0603)(231) = | 13.93 |
| 20-30 | .1640 | | 37.88 |
| 30-40 | .2731 | | 63.09 |
| 40-50 | .2644 | | 61.08 |
| 50-60 | .1568 | | 36.22 |
| 60-70 | .0539 | | 12.45 |
| > 70 | .0129 | | 2.98 |

Categories < 10 and > 70 are less than 5.
Collapse the < 10 into 10-20 and > 70 into 60-70.

| Age | $f_o$ | $f_e$ | $\frac{(f_0 - f_e)^2}{f_0}$ |
|---|---|---|---|
| 10-20 | 16 | 17.30 | 0.10 |
| 20-30 | 44 | 37.88 | 0.99 |
| 30-40 | 61 | 63.09 | 0.07 |
| 40-50 | 56 | 61.08 | 0.42 |
| 50-60 | 35 | 36.22 | 0.04 |
| 60-70 | 19 | 15.43 | 0.83 |
| | | | 2.45 |

df = $k$ - 3 = 6 - 3 = 3, $\alpha$ = .05

$\chi^2_{.05,3} = 7.8147$

Observed $\chi^2$ = **2.45**

Since the observed $\chi^2 < \chi^2_{.05,3} = 7.8147$, the decision is to **fail to reject the null hypothesis**.

**There is no reason to reject that the observed frequencies are normally distributed**.

16.9

| Category | $f_o$ | Prop. from survey | $f_e$ | $\frac{(f_0 - f_e)^2}{f_0}$ |
|---|---|---|---|---|
| Containers/Packaging | 86 | .303 | (.303)(300)=90.9 | 0.279 |
| Nondurable Goods | 74 | .213 | (.213)(300)=63.9 | 1.379 |
| Durable Goods | 70 | .196 | (.196)(300)=58.8 | 1.792 |
| Yard Trimmings/Other | 41 | .149 | (.149)(300)=44.7 | 0.334 |
| Food Scraps | 29 | .139 | (.139)(300)=41.7 | 5.562 |
| | 300 | | | 9.346 |

$H_o$: The distribution of observed frequencies is the same as the distribution of expected frequencies (national EPA survey).

$H_a$: The distribution of observed frequencies is not the same as the distribution of expected frequencies (national EPA survey).

$\alpha = .10$, df = $k$ - 1 = 5 - 1 = 4

$\chi^2_{.10,4} = 7.7794$

The observed $\chi^2$ = **9.346**

Since the observed $\chi^2 = 9.346 > \chi^2_{.10,4} = 7.7794$, the decision is to **reject the null hypothesis**.
There is enough evidence to declare that the distribution of observed frequencies is different from the distribution of expected frequencies.

16.11

| | Variable Two | | | | |
|---|---|---|---|---|---|
| Variable One | 24 | 13 | 47 | 58 | 142 |
| | 93 | 59 | 187 | 244 | 583 |
| | 117 | 72 | 234 | 302 | 725 |

$H_o$: Variable One is independent of Variable Two.
$H_a$: Variable One is not independent of Variable Two.

$e_{11} = \dfrac{(142)(117)}{725} = 22.92$ $\qquad$ $e_{12} = \dfrac{(142)(72)}{725} = 14.10$

$e_{13} = \dfrac{(142)(234)}{725} = 45.83$ $\qquad$ $e_{14} = \dfrac{(142)(302)}{725} = 59.15$

$e_{21} = \dfrac{(583)(117)}{725} = 94.08$ $\qquad$ $e_{22} = \dfrac{(583)(72)}{725} = 57.90$

$e_{23} = \dfrac{(583)(234)}{725} = 188.17$ $\qquad$ $e_{24} = \dfrac{(583)(302)}{725} = 242.85$

| | Variable Two | | | | |
|---|---|---|---|---|---|
| Variable One | (22.92)<br>24 | (14.10)<br>13 | (45.83)<br>47 | (59.15)<br>58 | 142 |
| | (94.08)<br>93 | (57.90)<br>59 | (188.17)<br>187 | (242.85)<br>244 | 583 |
| | 117 | 72 | 234 | 302 | 725 |

$$\chi^2 = \frac{(24-22.92)^2}{22.92} + \frac{(13-14.10)^2}{14.10} + \frac{(47-45.83)^2}{45.83} + \frac{(58-59.15)^2}{59.15} +$$

$$\frac{(93-94.08)^2}{94.08} + \frac{(59-57.90)^2}{57.90} + \frac{(188-188.17)^2}{188.17} + \frac{(244-242.85)^2}{242.85} =$$

.05 + .09 + .03 + .02 + .01 + .02 + .01 + .01 = **0.24**

$\alpha$ = .01, df = ($c$-1)($r$-1) = (4-1)(2-1) = 3, $\chi^2_{.01,3}$ = 11.3449

Since the observed $\chi^2$ = 0.24 < $\chi^2_{.01,3}$ = 11.3449, the decision is to **fail to reject the null hypothesis**.

Variable One is independent of Variable Two.

16.13

| | Type of Music Preferred | | | | | |
|---|---|---|---|---|---|---|
| Region | | Rock | R&B | Coun | Clssic | |
| | NE | 140 | 32 | 5 | 18 | 195 |
| | S | 134 | 41 | 52 | 8 | 235 |
| | W | 154 | 27 | 8 | 13 | 202 |
| | | 428 | 100 | 65 | 39 | 632 |

$H_o$: Type of music preferred is independent of region.
$H_a$: Type of music preferred is not independent of region.

$$e_{11} = \frac{(195)(428)}{632} = 132.6 \qquad e_{23} = \frac{(235)(65)}{632} = 24.17$$

$$e_{12} = \frac{(195)(100)}{632} = 30.85 \qquad e_{24} = \frac{(235)(39)}{632} = 14.50$$

$$e_{13} = \frac{(195)(65)}{632} = 20.06 \qquad e_{31} = \frac{(202)(428)}{632} = 136.80$$

$$e_{14} = \frac{(195)(39)}{632} = 12.03 \qquad e_{32} = \frac{(202)(100)}{632} = 31.96$$

$$e_{21} = \frac{(235)(428)}{632} = 159.15 \qquad e_{33} = \frac{(202)(65)}{632} = 20.78$$

$$e_{22} = \frac{(235)(100)}{632} = 37.18 \qquad e_{34} = \frac{(202)(39)}{632} = 12.47$$

| | Type of Music Preferred | | | | | |
|---|---|---|---|---|---|---|
| Region | | Rock | R&B | Coun | Clssic | |
| | NE | (132.06)<br>140 | (30.85)<br>32 | (20.06)<br>5 | (12.03)<br>18 | 195 |
| | S | (159.15)<br>134 | (37.18)<br>41 | (24.17)<br>52 | (14.50)<br>8 | 235 |
| | W | (136.80)<br>154 | (31.96)<br>27 | (20.78)<br>8 | (12.47)<br>13 | 202 |
| | | 428 | 100 | 65 | 39 | 632 |

$$\chi^2 = \frac{(141-132.06)^2}{132.06} + \frac{(32-30.85)^2}{30.85} + \frac{(5-20.06)^2}{20.06} + \frac{(18-12.03)^2}{12.03} +$$
$$\frac{(134-159.15)^2}{159.15} + \frac{(41-37.18)^2}{37.18} + \frac{(52-24.17)^2}{24.17} + \frac{(8-14.50)^2}{14.50} +$$
$$\frac{(154-136.80)^2}{136.80} + \frac{(27-31.96)^2}{31.96} + \frac{(8-20.78)^2}{20.78} + \frac{(13-12.47)^2}{12.47} =$$

.48 + .04 + 11.31 + 2.96 + 3.97 + .39 + 32.04 + 2.91 + 2.16 + .77 +

7.86 + .02 = **64.91**

$\alpha$ = .01, df = ($c$-1)($r$-1) = (4-1)(3-1) = 6

$\chi^2_{.01,6}$ = 16.8119

Since the observed $\chi^2$ = 64.91 > $\chi^2_{.01,6}$ = 16.8119, the decision is to **reject the null hypothesis**.

Type of music preferred is not independent of region of the country.

16.15

| | Number of Bedrooms | | | | |
|---|---|---|---|---|---|
| | | $\leq 2$ | 3 | $\geq 4$ | |
| Number of Stories | 1 | 116 | 101 | 57 | 274 |
| | 2 | 90 | 325 | 160 | 575 |
| | | 206 | 426 | 217 | 849 |

$H_0$: Number of Stories is independent of number of bedrooms.
$H_a$: Number of Stories is not independent of number of bedrooms.

$$e_{11} = \frac{(274)(206)}{849} = 66.48 \qquad e_{21} = \frac{(575)(206)}{849} = 139.52$$

$$e_{12} = \frac{(274)(426)}{849} = 137.48 \qquad e_{22} = \frac{(575)(426)}{849} = 288.52$$

$$e_{13} = \frac{(274)(217)}{849} = 70.03 \qquad e_{23} = \frac{(575)(217)}{849} = 146.97$$

$$\chi^2 = \frac{(90-139.52)^2}{139.52} + \frac{(101-137.48)^2}{137.48} + \frac{(57-70.03)^2}{70.03} + \frac{(90-139.52)^2}{139.52} +$$

$$\frac{(325-288.52)^2}{288.52} + \frac{(160-146.97)^2}{146.97} =$$

$\chi^2$ = 36.89 + 9.68 + 2.42 + 17.58 + 4.61 + 1.16 = **72.34**

$\alpha = .10$ df = $(c-1)(r-1) = (3-1)(2-1) = 2$

$\chi^2_{.10,2} = 4.6052$

Since the observed $\chi^2 = 72.34 > \chi^2_{.10,2} = 4.6052$, the decision is to **reject the null hypothesis**.

Number of stories is not independent of number of bedrooms.

16.17 $\alpha = .01$, $k = 7$, df = 6

$H_0$: The observed distribution is the same as the expected distribution
$H_a$: The observed distribution is not the same as the expected distribution

Use:

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

critical $\chi^2_{.01,6} = 16.8119$

| $f_0$ | $f_e$ | $(f_0-f_e)^2$ | $\frac{(f_0 - f_e)^2}{f_0}$ |
|---|---|---|---|
| 214 | 206 | 64 | 0.311 |
| 235 | 232 | 9 | 0.039 |
| 279 | 268 | 121 | 0.451 |
| 281 | 284 | 9 | 0.032 |
| 264 | 268 | 16 | 0.060 |
| 254 | 232 | 484 | 2.086 |
| 211 | 206 | 25 | 0.121 |
| | | | 3.100 |

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e} = \mathbf{3.100}$$

Since the observed value of $\chi^2 = 3.1 < \chi^2_{.01,6} = 16.8119$, the decision is to **fail to reject the null hypothesis**. The observed distribution is not different from the expected distribution.

16.19

| | | Location | | | |
|---|---|---|---|---|---|
| | | NE | W | S | |
| Customer | Industrial | 230 | 115 | 68 | 413 |
| | Retail | 185 | 143 | 89 | 417 |
| | | 415 | 258 | 157 | 830 |

$e_{11} = \dfrac{(413)(415)}{830} = 206.5$ $\qquad e_{21} = \dfrac{(417)(415)}{830} = 208.5$

$e_{12} = \dfrac{(413)(258)}{830} = 128.38$ $\qquad e_{22} = \dfrac{(417)(258)}{830} = 129.62$

$e_{13} = \dfrac{(413)(157)}{830} = 78.12$ $\qquad e_{23} = \dfrac{(417)(157)}{830} = 78.88$

| | | Location | | | |
|---|---|---|---|---|---|
| | | NE | W | S | |
| Customer | Industrial | (206.5)<br>230 | (128.38)<br>115 | (78.12)<br>68 | 413 |
| | Retail | (208.5)<br>185 | (129.62)<br>143 | (78.88)<br>89 | 417 |
| | | 415 | 258 | 157 | 830 |

$$\chi^2 = \frac{(230-206.5)^2}{206.5} + \frac{(115-128.38)^2}{128.38} + \frac{(68-78.12)^2}{78.12} +$$

$$\frac{(185-208.5)^2}{208.5} + \frac{(143-129.62)^2}{129.62} + \frac{(89-78.88)^2}{78.88} =$$

$2.67 + 1.39 + 1.31 + 2.65 + 1.38 + 1.30 =$ **10.70**

$\alpha = .10$ and df = $(c - 1)(r - 1) = (3 - 1)(2 - 1) = 2$

$\chi^2_{.10,2} = 4.6052$

Since the observed $\chi^2 = 10.70 > \chi^2_{.10,2} = 4.6052$, the decision is to **reject the null hypothesis**.

Type of customer is not independent of geographic region.

16.21

| | | Gender | | |
|---|---|---|---|---|
| | | M | F | |
| Bought Car | Y | 207 | 65 | 272 |
| | N | 811 | 984 | 1,795 |
| | | 1,018 | 1,049 | 2,067 |

$H_o$: Purchasing a car or not is independent of gender.
$H_a$: Purchasing a car or not is <u>not</u> independent of gender.

$$e_{11} = \frac{(272)(1{,}018)}{2{,}067} = 133.96 \qquad e_{12} = \frac{(27)(1{,}049)}{2{,}067} = 138.04$$

$$e_{21} = \frac{(1{,}795)(1{,}018)}{2{,}067} = 884.04 \qquad e_{22} = \frac{(1{,}795)(1{,}049)}{2{,}067} = 910.96$$

| | | Gender | | |
|---|---|---|---|---|
| | | M | F | |
| Bought Car | Y | (133.96)<br>207 | (138.04)<br>65 | 272 |
| | N | (884.04)<br>811 | (910.96)<br>984 | 1,795 |
| | | 1,018 | 1,049 | 2,067 |

$$\chi^2 = \frac{(207-133.96)^2}{133.96} + \frac{(65-138.04)^2}{138.04} + \frac{(811-884.04)^2}{884.04} +$$

$$\frac{(984-910.96)^2}{910.96} = 39.82 + 38.65 + 6.03 + 5.86 = \mathbf{90.36}$$

$\alpha = .05$ df = ($c$-1)($r$-1) = (2-1)(2-1) = 1

$\chi^2_{.05,1} = 3.8415$

Since the observed $\chi^2 = 90.36 > \chi^2_{.05,1} = 3.8415$, the decision is to **reject the null hypothesis**.

Purchasing a car is not independent of gender.

16.23 $H_o$: The distribution of observed frequencies is the same as the distribution of expected frequencies.
$H_a$: The distribution of observed frequencies is not the same as the distribution of expected frequencies.

| Soft Drink | $f_o$ | proportions | $f_e$ | $\frac{(f_0 - f_e)^2}{f_e}$ |
|---|---|---|---|---|
| Classic Coke | 397 | .26 | (.26)(1726) = 448.76 | 5.97 |
| Pepsi | 310 | .15 | (.15)(1726) = 258.90 | 10.09 |
| Diet Coke | 207 | .15 | 258.90 | 10.40 |
| Mt. Dew | 160 | .10 | 172.60 | 0.92 |
| Diet Pepsi | 130 | .08 | 138.08 | 0.47 |
| Sprite | 126 | .08 | 138.08 | 1.06 |
| Dr. Pepper | 143 | .09 | 155.34 | 0.98 |
| Others | 253 | .09 | 155.34 | 61.40 |
| $\sum f_o =$ | 1,726 | | | 91.29 |

Observed $\chi^2$ = **91.29**

$\alpha = .05$ df = $k$ - 1 = 8 - 1 = 7

$\chi^2_{.05,7} = 14.0671$

Since the observed $\chi^2 = 91.29 > \chi^2_{.05,6} = 14.0671$, the decision is to **reject the null hypothesis**.

The observed frequencies are not distributed the same as the expected frequencies from the national poll.

16.25

| | | Type of College or University | | | |
|---|---|---|---|---|---|
| | | Community College | Large University | Small College | |
| Number of Children | 0 | 25 | 178 | 31 | 234 |
| | 1 | 49 | 141 | 12 | 202 |
| | 2 | 31 | 54 | 8 | 93 |
| | $\geq$3 | 22 | 14 | 6 | 42 |
| | | 127 | 387 | 57 | 571 |

$H_o$: Number of Children is independent of Type of College or University.
$H_a$: Number of Children is not independent of Type of College or University.

$$e_{11} = \frac{(234)(127)}{571} = 52.05 \qquad e_{31} = \frac{(93)(127)}{571} = 20.68$$

$$e_{12} = \frac{(234)(387)}{571} = 158.60 \qquad e_{32} = \frac{(193)(387)}{571} = 63.03$$

$$e_{13} = \frac{(234)(57)}{571} = 23.36 \qquad e_{33} = \frac{(93)(57)}{571} = 9.28$$

$$e_{21} = \frac{(202)(127)}{571} = 44.93 \qquad e_{41} = \frac{(42)(127)}{571} = 9.34$$

$$e_{22} = \frac{(202)(387)}{571} = 136.91 \qquad e_{42} = \frac{(42)(387)}{571} = 28.47$$

$$e_{23} = \frac{(202)(57)}{571} = 20.16 \qquad e_{43} = \frac{(42)(57)}{571} = 4.19$$

| | | Type of College or University | | | |
|---|---|---|---|---|---|
| | | Community College | Large University | Small College | |
| Number of Children | 0 | (52.05)<br>25 | (158.60)<br>178 | (23.36)<br>31 | 234 |
| | 1 | (44.93)<br>49 | (136.91)<br>141 | (20.16)<br>12 | 202 |
| | 2 | (20.68)<br>31 | (63.03)<br>54 | (9.28)<br>8 | 93 |
| | $\geq$3 | (9.34)<br>22 | (28.47)<br>14 | (4.19)<br>6 | 42 |
| | | 127 | 387 | 57 | 571 |

$$\chi^2 = \frac{(25-52.05)^2}{52.05} + \frac{(178-158.6)^2}{158.6} + \frac{(31-23.36)^2}{23.36} + \frac{(49-44.93)^2}{44.93} +$$

$$\frac{(141-136.91)^2}{136.91} + \frac{(12-20.16)^2}{20.16} + \frac{(31-20.68)^2}{20.68} + \frac{(54-63.03)^2}{63.03} +$$

$$\frac{(8-9.28)^2}{9.28} + \frac{(22-9.34)^2}{9.34} + \frac{(14-28.47)^2}{28.47} + \frac{(6-4.19)^2}{4.19} =$$

14.06 + 2.37 + 2.50 + 0.37 + 0.12 + 3.30 + 5.15 + 1.29 + 0.18 +

17.16 + 7.35 + 0.78 = **54.63**

$\alpha$ = .05, df= ($c$ - 1)($r$ - 1) = (3 - 1)(4 - 1) = 6

$\chi^2_{.05,6} = 12.5916$

Since the observed $\chi^2 = 54.63 > \chi^2_{.05,6} = 12.5916$, the decision is to **reject the null hypothesis**.

Number of children is not independent of type of College or University.

16.27 The observed chi-square value for this test of independence is 5.366. The associated $p$-value of .252 indicates failure to reject the null hypothesis. There is not enough evidence here to say that color choice is dependent upon gender. Automobile marketing people do not have to worry about which colors especially appeal to men or to women because car color is independent of gender. In addition, design and production people can determine car color quotas based on other variables.