***Note: Problems 12.16–12.17 refer to a professor who claims that his typical grade distribution is 20% As, 25% Bs, 40% Cs, 10% Ds, and 5% Fs.***

**12.17** At the end of the semester, the 85 students in the course are assigned grades as follows: 22 As, 29 Bs, 20 Cs, 10 Ds, and 4 Fs. Test the professor's claim at the $\alpha = 0.05$ significance level.

State the null and alternative hypotheses.

$H_0$ : The faculty member followed his stated grade distribution.
$H_1$ : The faculty member did not follow his stated grade distribution.

Use the following table to calculate $\chi^2$.

| **Grade** | ***O*** | ***E*** | ***O* – *E*** | **(*O* – *E*)²** | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|---|
| A | 22 | 17 | 5 | 25 | 1.47 |
| B | 29 | 21.25 | 7.75 | 60.06 | 2.83 |
| C | 20 | 34 | –14 | 196 | 5.76 |
| D/F | 14 | 12.75 | 1.25 | 1.56 | 0.12 |
| **Total** | | | | | **10.18** |

You have to combine the D and F categories, because F has an expected frequency of 4.25, which is less than 5.

According to Reference Table 3, the critical chi-square value given df = 3 and $\alpha = 0.05$ is 7.815. Because $\chi^2 = 10.18$ is greater than $\chi^2_c = 7.815$, you reject $H_0$ and conclude that the professor did not follow his stated grade distribution.

## Chi-Square Test for Independence

### Are the variables related?

**12.18** Explain how to perform the chi-square test for independence.

The chi-square test for independence is used to determine whether two categorical variables affect each other. Begin by stating a null hypothesis that the variables are independent; the alternative hypothesis states that the variables are not independent—they are related in some way.

A contingency table should be constructed with rows that are the categories of one variable and columns that are the categories of the other. The cells at the intersections of the rows and columns contain the observed frequencies. A contingency table with $r$ rows and $c$ columns contains $r \cdot c$ cells. Calculate the expected frequencies using the following formula.

$$E_{r,c} = \frac{(\text{total of row } r)(\text{total of column } c)}{\text{total number of observations}}$$

Use the chi-square distribution to compare the observed and expected frequencies. The chi-square statistic $\chi^2$ is calculated using the formula first introduced in Problem 12.1.

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

Use Reference Table 3 to determine the critical chi-square value $\chi_c^2$ given $\text{df} = (r-1)(c-1)$ degrees of freedom and the stated significance level $\alpha$. In order to reject the null hypothesis (and therefore claim that the variables are independent), $\chi^2$ must be greater than $\chi_c^2$. Otherwise, you fail to reject the null hypothesis and must conclude that the variables are dependent.

***Note: Problems 12.19–12.20 refer to the data set below, the number of head-to-head tennis matches won by Bob and Deb given warm-up times of 0–10 minutes, 11–20 minutes, and more than 20 minutes.***

| | **0–10 min** | **11–20 min** | **More than 20 min** | **Total** |
|---|---|---|---|---|
| Deb Wins | 4 | 10 | 9 | **23** |
| Bob Wins | 14 | 9 | 4 | **27** |
| **Total** | **18** | **19** | **13** | **50** |

**12.19** Calculate the expected frequency for each cell of the table, assuming that the warm-up time and the match winner are independent variables.

Calculate this for every cell except the bold-faced cells, which represent totals.

Calculate the expected frequency for each cell in the table using the following equation.

$$E_{r,c} = \frac{(\text{total of row } r)(\text{total of column } c)}{\text{total number of observations}}$$

Calculate $E_{1,1}$.

$$E_{1,1} = \frac{(23)(18)}{50} = 8.28$$

You expect Deb to win $E_{1,1} = 8.28$ matches given 0–10 minutes of warm-up time. Calculate the expected frequencies for the other 5 cells using the same technique.

$$E_{1,1} = \frac{(23)(18)}{50} = 8.28 \quad E_{1,2} = \frac{(23)(19)}{50} = 8.74 \quad E_{1,3} = \frac{(23)(13)}{50} = 5.98$$

$$E_{2,1} = \frac{(27)(18)}{50} = 9.72 \quad E_{2,2} = \frac{(27)(19)}{50} = 10.26 \quad E_{2,3} = \frac{(27)(13)}{50} = 7.02$$

***Note: Problems 12.19–12.20 refer to the data set in Problem 12.19, the number of head-to-head tennis matches won by Bob and Deb given warm-up times of 0–10 minutes, 11–20 minutes, and more than 20 minutes.***

**12.20** Determine whether the length of time the players warm up affects the winner of the match at the $\alpha = 0.10$ significance level.

State the null and alternative hypotheses.

$H_0$ : Warm-up time is independent of the eventual winner of the match.
$H_1$ : Warm-up time is not independent of the eventual winner of the match.

Calculate $\chi^2$ using the following table.

This part is no different from Problems 12.2–12.17.

| Row | Column | $O$ | $E$ | $O-E$ | $(O-E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 4 | 8.28 | –4.28 | 18.32 | 2.21 |
| 1 | 2 | 10 | 8.74 | 1.26 | 1.59 | 0.18 |
| 1 | 3 | 9 | 5.98 | 3.02 | 9.12 | 1.53 |
| 2 | 1 | 14 | 9.72 | 4.28 | 18.32 | 1.88 |
| 2 | 2 | 9 | 10.26 | –1.26 | 1.59 | 0.15 |
| 2 | 3 | 4 | 7.02 | –3.02 | 9.12 | 1.30 |
| **Total** | | | | | | **7.25** |

The table in Problem 12.19 (excluding the boldfaced totals) contains $r = 2$ rows (representing the players) and $c = 3$ columns (representing the different warm-up periods). Thus, there are df = (2 – 1)(3 – 1) = 2 degrees of freedom.

According to Reference Table 3, the critical chi-square value given df = 2 and $\alpha = 0.10$ is 4.605. Because $\chi^2 = 7.25$ is more than $\chi^2_c = 4.605$, you reject $H_0$; it appears there is some sort of relationship between the warm-up time and the eventual winner of the match. The variables are not independent at the $\alpha = 0.10$ significance level.

This is the sum of the numbers in the rightmost column of the last table.

***Note: Problems 12.21–12.22 refer to the data set below, the number of men and women who decided to purchase or not to purchase an extended warranty for a digital camera at an electronics store.***

| | Warranty | No Warranty | Total |
|---|---|---|---|
| Men | 7 | 50 | **57** |
| Women | 9 | 19 | **28** |
| **Total** | **16** | **69** | **85** |

**12.21** Calculate the expected frequencies for each cell, assuming that the warranty decision and the gender of customer are independent variables.

*If gender and warranty decision are independent, you would expect men to purchase the warranty 10.73 times.*

Calculate the expected frequencies of each cell by multiplying its row total by its column total and dividing by the overall total.

$$E_{1,1} = \frac{(57)(16)}{85} = 10.73 \qquad E_{1,2} = \frac{(57)(69)}{85} = 46.27$$
$$E_{2,1} = \frac{(28)(16)}{85} = 5.27 \qquad E_{2,2} = \frac{(28)(69)}{85} = 22.73$$

***Note: Problems 12.21–12.22 refer to the data set in Problem 12.21, the number of men and women who decided to purchase or not to purchase an extended warranty for a digital camera at an electronics store.***

**12.22** Determine whether the warranty decision and the gender of the customer are independent variables at the $\alpha = 0.05$ significance level.

State the null and alternative hypotheses.

$H_0$ : Warranty choice is independent of customer gender.
$H_1$ : Warranty choice is not independent of customer gender.

Use the following table to calculate $\chi^2$.

| **Row** | **Column** | ***O*** | ***E*** | ***O* – *E*** | **(*O* – *E*)²** | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 4 | 8.28 | –4.28 | 18.32 | 2.21 |
| 1 | 1 | 7 | 10.73 | –3.73 | 13.91 | 1.30 |
| 1 | 2 | 50 | 46.27 | 3.73 | 13.91 | 0.30 |
| 2 | 1 | 9 | 5.27 | 3.73 | 13.91 | 2.64 |
| 2 | 2 | 19 | 22.73 | –3.73 | 13.91 | 0.61 |
| **Total** | | | | | | **7.06** |

According to Reference Table 3, the critical chi-square value given df = (2 – 1)(2 – 1) = 1 degree of freedom and $\alpha = 0.05$ is 3.841. Because $\chi^2 = 7.06$ is greater than $\chi_c^2 = 3.841$, you reject $H_0$; there is a relationship between warranty decision and gender.

***Note: Problems 12.23–12.24 refer to the data set below, the final exam grade distribution for 215 graduate students and the number of hours the students spent studying for the exam.***

| | A | B | C | Total |
|---|---|---|---|---|
| Less than 3 hr | 18 | 48 | 16 | **82** |
| 3–5 hr | 30 | 28 | 12 | **70** |
| More than 5 hr | 33 | 25 | 5 | **63** |
| **Total** | **81** | **101** | **33** | **215** |

**12.23** Calculate the expected frequencies for each cell, assuming that the final exam grade and the time spent studying are independent variables.

Excluding the total column and the total row, there are $r = 3$ rows and $c = 3$ columns. Calculate the expected frequency for each cell by multiplying its row total by its column total and dividing by the overall total.

$$E_{1,1} = \frac{(82)(81)}{215} = 30.89 \quad E_{1,2} = \frac{(82)(101)}{215} = 38.52 \quad E_{1,3} = \frac{(82)(33)}{215} = 12.59$$

$$E_{2,1} = \frac{(70)(81)}{215} = 26.37 \quad E_{2,2} = \frac{(70)(101)}{215} = 32.88 \quad E_{2,3} = \frac{(70)(33)}{215} = 10.74$$

$$E_{3,1} = \frac{(63)(81)}{215} = 23.73 \quad E_{3,2} = \frac{(63)(101)}{215} = 29.60 \quad E_{3,3} = \frac{(63)(33)}{215} = 9.67$$

If the grade you get on the exam doesn't depend on the length of time you study, then 30.89 students who study less than 3 hours should get an A.

***Note: Problems 12.23–12.24 refer to the data set in Problem 12.23, the final exam grade distribution for 215 graduate students and the number of hours the students spent studying for the exam.***

**12.24** Determine whether the exam grade and the time spent studying for the exam are independent variables at the $\alpha = 0.01$ significance level.

State the null and alternative hypotheses.

$H_0$ : The final exam grade is independent of the hours spent studying for it.
$H_1$ : The final exam grade is not independent of the hours spent studying for it.