**Stat 145 Homework Solutions: Chapter 20**

Problem 20.2

(a) The proportion of successful students in each group is:

$$\begin{array}{ll} \text{less than 2 hours:} & \frac{11}{11+9} = .550 \\ \text{between 2 and 12 hours:} & \frac{68}{68+23} = .747 \\ \text{more than 12 hours:} & \frac{3}{3+5} = .375 \end{array}$$

(b) A moderate amount of time spent in extracurricular activities (2 to 12 hours) appears to be beneficial.

Problem 20.4

(a) The data is:

| | Extracurricular activities (hours per week) | | | |
|---|---|---|---|---|
| Grade | $< 2$ | 2 to 12 | $> 12$ | Total |
| C or better | 11 | 68 | 3 | 82 |
| D or F | 9 | 23 | 5 | 37 |
| Total | 20 | 91 | 8 | 119 |

The expected cell counts are:

| | Extracurricular activities (hours per week) | | | |
|---|---|---|---|---|
| Grade | $< 2$ | 2 to 12 | $> 12$ | Total |
| C or better | 13.78 | 62.71 | 5.51 | 82 |
| D or F | 6.22 | 28.29 | 2.49 | 37 |
| Total | 20 | 91 | 8 | 119 |

For example, the expected cell count for the first cell is:

$$\frac{82 \times 20}{119} = 13.78$$

(b) The first and third columns have more Ds or Fs than expected, while the second column has more passing grades than expected. It appears that a moderate amount of time spent in extracurricular activities (2 to 12 hours) is beneficial. (This agrees with our answer in 20.2.)

Problem 20.6

(a) The value of the $\chi^2$ statistic is:

$$\begin{aligned}
\chi^2 &= \frac{(11-13.78)^2}{13.78} \\
&+ \frac{(68-62.71)^2}{62.71} \\
&+ \frac{(3-5.51)^2}{5.51} \\
&+ \frac{(9-6.22)^2}{6.22} \\
&+ \frac{(23-28.29)^2}{28.29} \\
&+ \frac{(5-2.49)^2}{2.49} \\
&= 6.926
\end{aligned}$$

(b) From the output in Figure 20.4, the $P$-value is 0.031. Rejecting $H_0$ means that we conclude there is a relationship between time spent in extracurricular activities and performance in the course.

(c) The largest contribution comes from row 2, column 3 (D or F, $>$ 12 hours). This term indicates that too much time spent of extracurricular activities seems to hurt academic performance.

(d) No, because association does not imply causation. This was an observational study, not an experiment. There may be a lurking variable (e.g., a personality trait that "causes" students to do well and also participate in extracurricular activities in moderation) that affects both variables.

Problem 20.8

(a) The degrees of freedom are:

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

(b) The chi-square statistic lies between 5.99 and 7.38. The $P$-value is between .025 and .05.

(c) If $H_0$ were true, we would expect the mean value of $\chi^2$ to be 2 (the degrees of freedom). The observed value $\chi^2 = 6.926$ is significantly larger than 2.

Problem 20.19

The expected cell counts are:

| Rank | Male | Female | Total |
|---|---|---|---|
| Assistant Professors | 70.06 | 268.94 | 339 |
| Associate Professors | 115.73 | 444.27 | 560 |
| Professors | 149.21 | 572.79 | 722 |
| Total | 335 | 1286 | 1621 |

For example, the expected cell count for the first cell is:

$$\frac{339 \times 335}{1621} = 70.06$$

The value of the chi-square statistic is:

$$\begin{aligned}
\chi^2 &= \frac{(126 - 70.06)^2}{70.06} \\
&+ \frac{(213 - 268.94)^2}{268.94} \\
&+ \frac{(149 - 115.73)^2}{115.73} \\
&+ \frac{(411 - 444.27)^2}{444.27} \\
&+ \frac{(60 - 149.21)^2}{149.21} \\
&+ \frac{(662 - 572.79)^2}{572.79} \\
&= 135.592
\end{aligned}$$

The degrees of freedom are:

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

The $P$-value is less than .0005. We reject the null hypothesis and conclude that there is a significant difference between the distribution of ranks for female and male faculty. The biggest contributions to the chi-square statistic come from row 1, column 1 (female assistant professors) and row 3, column 1 (female professors). There are more female assistant professors than expected and fewer female professors than expected.

Problem 20.21

(a) The table below shows the percent of students who smoke in each parent group:

$$\begin{array}{ll} \text{Both parents smoke:} & \frac{400}{400+1380} = 22.5\% \\ \text{One parent smokes:} & \frac{416}{416+1823} = 18.6\% \\ \text{Neither parent smokes:} & \frac{188}{188+1168} = 13.9\% \end{array}$$

A student's likelihood of smoking increases if one parent smokes, and increases even more if both parents smoke.

(b) The null hypothesis says that there is not a relationship between parents' smoking habits and their children's smoking habits.

(c) The expected cell counts are:

| | Student | | |
|---|---|---|---|
| | smokes | does not smoke | Total |
| Both parents smoke | 332.49 | 1447.51 | 1780 |
| One parent smokes | 418.22 | 1820.78 | 2239 |
| Neither parent smokes | 253.29 | 1102.71 | 1356 |
| Total | 1004 | 4371 | 5375 |

For example, the expected cell count for the first cell is:

$$\frac{1780 \times 1004}{5375} = 332.49$$

(d) In row 1, column 1 (both parents smoke, student smokes), the actual count is much larger than the expected count. In row 3, column 1 (neither parent smokes, student smokes), the actual count is much smaller than the expected count. This agrees with our conclusion in part (a): Children of non-smokers are less likely to smoke.

Problem 20.21 (continued)

(e) The value of the chi-square statistic is:

$$\begin{aligned}
\chi^2 &= \frac{(400-332.49)^2}{332.49} \\
&+ \frac{(1380-1447.51)^2}{1447.51} \\
&+ \frac{(416-418.22)^2}{418.22} \\
&+ \frac{(1823-1820.78)^2}{1820.78} \\
&+ \frac{(188-253.29)^2}{253.29} \\
&+ \frac{(1168-1102.71)^2}{1102.71} \\
&= 37.566
\end{aligned}$$

The degrees of freedom are:

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

The $P$-value is less than .0005. We reject the null hypothesis and conclude that there is a relationship between parents' smoking habits and their children's smoking habits.