Chapter 14 Solutions

14.1. If $p = 0.25$, then odds $= \dfrac{p}{1-p} = \dfrac{0.25}{0.75} = \dfrac{1}{3}$, or "1 to 3."

14.2. If odds $= 2$, then $p = \dfrac{\text{odds}}{\text{odds}+1} = \dfrac{2}{2+1} = \dfrac{2}{3}$.

14.3. We have $\hat{p}_{\text{men}} = \frac{80}{140} = \frac{4}{7}$, and $\hat{p}_{\text{women}} = \frac{45}{100} = \frac{9}{20}$. Therefore:

$$\text{odds}_{\text{men}} = \frac{4/7}{3/7} = \frac{4}{3} \doteq 1.3333, \text{ or "4 to 3," and}$$

$$\text{odds}_{\text{women}} = \frac{9/20}{11/20} = \frac{9}{11} \doteq 0.8182, \text{ or "9 to 11"}$$

Note: *The odds can also be computed without first finding $\hat{p}$; for example, 80 men preferred Commercial A and 60 preferred Commercial B, so $\text{odds}_{men} = \frac{80}{60} = \frac{4}{3}$.*

14.4. The odds for selecting Commercial B would be the reciprocal of the odds for Commercial A: $\text{odds}^*_{\text{men}} = \dfrac{3}{4} = 0.75$ and $\text{odds}^*_{\text{women}} = \dfrac{11}{9} \doteq 1.2222$.

14.5. With $\text{odds}_{\text{men}} = \frac{4}{3}$ and $\text{odds}_{\text{women}} = \frac{9}{11}$, we have $\log(\text{odds}_{\text{men}}) = \log(4/3) \doteq 0.2877$ and $\log(\text{odds}_{\text{women}}) = \log(9/11) \doteq -0.2007$.

Note: *You may wish to remind students to use the natural logarithm, called "ln" by* Excel *and most calculators. A student who mistakenly uses the common (base-10) logarithm instead of the natural logarithm will get 0.1249 and −0.0872 as answers.*

14.6. With $\text{odds}^*_{\text{men}} = \frac{3}{4}$ and $\text{odds}^*_{\text{women}} = \frac{11}{9}$, we have $\log(\text{odds}^*_{\text{men}}) = \log(3/4) \doteq -0.2877$ and $\log(\text{odds}^*_{\text{women}}) = \log(11/9) \doteq 0.2007$.

Note: *Because these odds were the reciprocals of those from Exercise 14.3, the log odds are the opposites (negations) of those found in Exercise 14.5. A student who mistakenly uses the common (base-10) logarithm instead of the natural logarithm will get −0.1249 and 0.0872 as answers.*

14.7. The model is $y = \log(\text{odds}) = \beta_0 + \beta_1 x$. If $x = 1$ for men and 0 for women, we need:

$$\log\left(\frac{p_{\text{men}}}{1-p_{\text{men}}}\right) = \beta_0 + \beta_1 \quad \text{and} \quad \log\left(\frac{p_{\text{women}}}{1-p_{\text{women}}}\right) = \beta_0$$

We estimate $b_0 = \log(\text{odds}_{\text{women}}) \doteq -0.2007$ and $b_1 = \log(\text{odds}_{\text{men}}) - b_0 \doteq 0.4884$, so the regression equation is $\hat{y} = -0.2007 + 0.4884x$.

If $x = 0$ for men and 1 for women, we estimate $b_0 = \log(\text{odds}_{\text{men}}) \doteq 0.2877$ and $b_1 = \log(\text{odds}_{\text{women}}) - b_0 \doteq -0.4884$, so the regression equation is $\hat{y} = 0.2877 - 0.4884x$.

The estimated odds ratio is either:

$$e^{0.4884} \doteq \frac{\text{odds}_{\text{men}}}{\text{odds}_{\text{women}}} \doteq 1.6296 \quad \text{if } x = 1 \text{ for men, or}$$

$$e^{-0.4884} \doteq \frac{\text{odds}_{\text{women}}}{\text{odds}_{\text{men}}} \doteq 0.6136 \quad \text{if } x = 1 \text{ for women}$$

14.8. Because of the relationships between the (log) odds for selecting Commercial A and the (log) odds for selecting Commercial B, noted in the solutions to Exercises 14.4 and 14.6, these coefficients are the opposites (negations) of, and the odds ratios are reciprocals of, those found in the solution to the previous exercise.

The model is $y = \log(\text{odds}) = \beta_0 + \beta_1 x$. If $x = 1$ for men and 0 for women, we need:

$$\log\left(\frac{p_{\text{men}}}{1 - p_{\text{men}}}\right) = \beta_0 + \beta_1 \quad \text{and} \quad \log\left(\frac{p_{\text{women}}}{1 - p_{\text{women}}}\right) = \beta_0$$

We estimate $b_0 = \log(\text{odds}^*_{\text{women}}) \doteq 0.2007$ and $b_1 = \log(\text{odds}^*_{\text{men}}) - b_0 \doteq -0.4884$, so the regression equation is $\hat{y} = 0.2007 - 0.4884x$.

If $x = 0$ for men and 1 for women, we estimate $b_0 = \log(\text{odds}^*_{\text{men}}) \doteq -0.2877$ and $b_1 = \log(\text{odds}^*_{\text{women}}) - b_0 \doteq 0.4884$, so the regression equation is $\hat{y} = -0.2877 + 0.4884x$.

The estimated odds ratio is either:

$$e^{-0.4884} \doteq \frac{\text{odds}_{\text{men}}}{\text{odds}_{\text{women}}} \doteq 0.6136 \quad \text{if } x = 1 \text{ for men, or}$$

$$e^{0.4884} \doteq \frac{\text{odds}_{\text{women}}}{\text{odds}_{\text{men}}} \doteq 1.6296 \quad \text{if } x = 1 \text{ for women}$$

14.9. **(a)** The appropriate test would be a chi-square test with df = 6. **(b)** H_0 should refer to β_1 (the population slope) rather than b_1 (the estimated slope). **(c)** The logistic regression model has no error term.

14.10. **(a)** $\hat{p}_1 = \frac{108}{142} \doteq 0.7606$ for exclusive-territory firms. **(b)** $\hat{p}_2 = \frac{15}{28} \doteq 0.5357$ for other firms. **(c)** $\text{odds}_1 = \frac{\hat{p}_1}{1-\hat{p}_1} \doteq 3.1765$ and $\text{odds}_2 = \frac{\hat{p}_2}{1-\hat{p}_2} \doteq 1.1538$. **(d)** $\log(\text{odds}_1) \doteq 1.1558$ and $\log(\text{odds}_2) \doteq 0.1431$. (Be sure to use the *natural* logarithm for this computation.)

14.11. **(a)** $\hat{p}_{\text{w}} = \frac{63}{296} \doteq 0.2128$ for women and $\hat{p}_{\text{m}} = \frac{27}{251} \doteq 0.1076$ for men.
(b) $\text{odds}_{\text{w}} = \frac{\hat{p}_{\text{w}}}{1-\hat{p}_{\text{w}}} \doteq 0.2704$ and $\text{odds}_{\text{m}} = \frac{\hat{p}_{\text{m}}}{1-\hat{p}_{\text{m}}} \doteq 0.1205$.
(c) $\log(\text{odds}_{\text{w}}) \doteq -1.3079$ and $\log(\text{odds}_{\text{m}}) \doteq -2.1158$.

14.12. **(a)** $b_0 = \log(\text{odds}_2) \doteq 0.1431$ and $b_1 = \log(\text{odds}_1) - \log(\text{odds}_2) \doteq 1.0127$. **(b)** The fitted model is $\log(\text{odds}) = 0.1431 + 1.0127x$. **(c)** The odds ratio is $\text{odds}_1/\text{odds}_2 = e^{b_1} \doteq 2.7529$.

Minitab output

```
Predictor      Coef   SE Coef     Z      P  Ratio  Lower  Upper
Constant   0.143101  0.378932  0.38  0.706
Exclusive
 Yes        1.01267  0.426920  2.37  0.018   2.75   1.19   6.36
```

14.13. **(a)** $b_0 = \log(\text{odds}_{\text{m}}) \doteq -2.1158$ and $b_1 = \log(\text{odds}_{\text{w}}) - \log(\text{odds}_{\text{m}}) \doteq 0.8079$.
(b) The fitted model is $\log(\text{odds}) = -2.1158 + 0.8079x$. **(c)** The odds ratio is $\text{odds}_{\text{w}}/\text{odds}_{\text{m}} = e^{b_1} \doteq 2.2432$.

Minitab output

```
                                          Odds     95% CI
Predictor      Coef   SE Coef       Z      P  Ratio  Lower  Upper
Constant   -2.11581  0.203719  -10.39  0.000
gender
 w         0.807905  0.248327    3.25  0.001   2.24   1.38   3.65
```

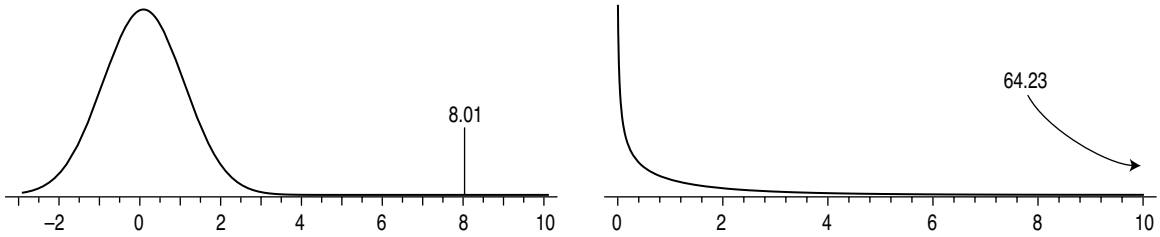
14.14. Recall that, by properties of exponents, $\frac{e^a}{e^b} = e^{a-b}$. Therefore:

$$\frac{\text{odds}_{x+1}}{\text{odds}_x} = \frac{e^{-13.71} \times e^{2.25(x+1)}}{e^{-13.71} \times e^{2.25x}} = e^{2.25(x+1)-2.25x} = e^{2.25}$$

14.15. With $b_1 \doteq 3.1088$ and $\text{SE}_{b_1} \doteq 0.3879$, the 99% confidence interval is $b_1 \pm 2.576\text{SE}_{b_1} \doteq b_1 \pm 0.9992$, or 2.1096 to 4.1080.

14.16. To find the confidence interval for the odds ratio, we first make a confidence interval for the slope b_1 and then transform (exponentiate) it: $b_1 \pm z^*\text{SE}_{b_1} = 3.1088 \pm (1.96)(0.3879) \doteq 2.3485$ to 3.8691, so the odds ratio interval is $e^{2.3485} \doteq 10.470$ to $e^{3.8691} \doteq 47.898$. Up to rounding error, this agrees with the software output.

14.17. **(a)** $z = \frac{3.1088}{0.3879} \doteq 8.01$. **(b)** $z^2 \doteq 64.23$, which agrees with the value of X^2 given by SPSS and SAS. **(c)** The sketches are below. For both the Normal and chi-square distribution, the test statistics are quite extreme, consistent with the reported P-value.



14.18. **(a)** For each explanatory variable, we test H_0: $\beta_i = 0$ vs. H_a: $\beta_i \neq 0$. **(b)** Under the null hypotheses, the z statistic has a standard Normal distribution. Therefore, we reject H_0 at the 5% level if $|z| > 1.96$. We do not reject H_0 for "Men's magazines" but have very strong evidence that all other coefficients (as well as the constant) are not zero. **(c)** The probability that the model's clothing is sexual is higher for magazines targeted at young adults (as the problem states) when the model is female and for magazines aimed at men or at men and women. **(d)** The fitted model is $\log(\text{odds}) = -2.32 + 0.50x_1 + 1.31x_2 - 0.05x_3 + 0.45x_4$.

14.19. **(a)** For men's magazines, the odds ratio confidence interval includes 1. This indicates that this explanatory variable has no effect on the probability that a model's clothing is not sexual, which is consistent with our failure to reject H_0 for men's magazines in the previous exercise. For all other explanatory variables, the odds ratio interval does not include 1, equivalent to the significant evidence against H_0 for those variables. **(b)** The odds that the model's clothing is not sexual are 1.27 to 2.16 times higher for magazines targeted at mature adults, 2.74 to 5.01 times higher when the model is male, and 1.11 to 2.23 times higher for magazines aimed at women. (These statements can also be made in terms of the odds that the model's clothing *is* sexual; for example, those odds are 1.27 to 2.16 times higher for magazines targeted at *young* adults, and so forth.) **(c)** The summary might note that it is easier to interpret the odds ratio rather than the regression coefficients because of the difficulty of thinking in terms of a log-odds scale.

14.20. (a) $\hat{p}_1 = \frac{487}{811} \doteq 0.6005$. **(b)** $\text{odds}_1 = \frac{\hat{p}_1}{1-\hat{p}_1} \doteq 1.5031$. **(c)** $\hat{p}_2 = \frac{324}{811} \doteq 0.3995$. **(d)** $\text{odds}_2 = \frac{\hat{p}_2}{1-\hat{p}_2} \doteq 0.6653$. **(e)** The probabilities in (a) and (c) are *complements* (their sum is 1), and the odds in (b) and (d) are *reciprocals* (their product is 1).

14.21. (a) $\hat{p}_{\text{hi}} = \frac{73}{91} \doteq 0.8022$ and $\text{odds}_{\text{hi}} = \frac{\hat{p}_{\text{hi}}}{1-\hat{p}_{\text{hi}}} = 4.0\overline{5}$. **(b)** $\hat{p}_{\text{non}} = \frac{75}{109} \doteq 0.6881$ and $\text{odds}_{\text{non}} = \frac{\hat{p}_{\text{non}}}{1-\hat{p}_{\text{non}}} \doteq 2.2059$. **(c)** The odds ratio is $\text{odds}_{\text{hi}}/\text{odds}_{\text{non}} \doteq 1.8385$. The odds of a high-tech company offering stock options are about 1.84 times those for a non-high-tech firm.

14.22. (a) $\log(\text{odds}_{\text{hi}}) \doteq 1.4001$ and $\log(\text{odds}_{\text{non}}) \doteq 0.7911$. **(b)** In the model, $\log(\text{odds}_{\text{non}}) = \beta_0$ and $\log(\text{odds}_{\text{hi}}) = \beta_0 + \beta_1$, so we find the estimates of β_0 and β_1 from the observed log-odds: $b_0 = \log(\text{odds}_{\text{non}}) \doteq 0.7911$ and $b_1 = \log(\text{odds}_{\text{hi}}) - \log(\text{odds}_{\text{non}}) \doteq 0.6090$. **(c)** $e^{b_1} \doteq e^{0.6090} \doteq 1.8385$, as we found in 16.13(c).

14.23. (a) With $b_1 \doteq 0.6090$ and $\text{SE}_{b_1} \doteq 0.3347$, the 95% confidence interval is $b_1 \pm 1.96\text{SE}_{b_1} \doteq b_1 \pm 0.6560$, or -0.0470 to 1.2650. **(b)** Exponentiating the confidence limits gives the interval 0.9540 to 3.5430. **(c)** Because the confidence interval for β_1 contains 0, or equivalently because 1 is in the interval for the odds ratio, we could not reject H_0: $\beta_1 = 0$ at the 5% level. There does not appear to be a significant difference between the odds of stock options for high-tech and other firms.

Note: *Software reports $z = 1.820$ and a P-value of 0.0688, which are nearly identical to the results for a two-proportion z test with the same counts ($z = -1.832$ and $P = 0.0669$)—see the solution to Exercise 8.51. For large samples, these two tests should give similar results.*

Minitab output

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	0.791128	0.206749	3.83	0.000			
HT							
Yes	0.608960	0.334663	1.82	0.069	1.84	0.95	3.54

14.24. All proportions, odds, odds ratios, and parameter estimates (b_0 and b_1) are unchanged. Because the standard error is smaller, the 95% confidence interval is narrower: $b_1 \pm 1.96\text{SE}_{b_1} \doteq b_1 \pm 0.4637$, or 0.1452 to 1.0727. The odds-ratio interval is therefore 1.1563 to 2.9233. Because 0 is not in the confidence interval for β_1 and 1 is not in the odds-ratio interval, we have significant evidence of a difference in the odds between the two types of companies.

Note: *For testing H_0: $\beta_1 = 0$, software reports $z = 2.573$ and $P = 0.0101$. For comparison, the test of $p_1 = p_2$ yields $z = 2.591$ and $P = 0.0096$.*

Minitab output

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	0.791128	0.146194	5.41	0.000			
HT							
Yes	0.608960	0.236642	2.57	0.010	1.84	1.16	2.92

14.25. **(a)** For the high blood pressure group, $\hat{p}_{\text{hi}} = \frac{55}{3338} \doteq 0.01648$, giving $\text{odds}_{\text{hi}} = \frac{\hat{p}_{\text{hi}}}{1-\hat{p}_{\text{hi}}} \doteq 0.01675$, or about 1 to 60. (If students give odds in the form "a to b," their choices of a and b might be different.) **(b)** For the low blood pressure group, $\hat{p}_{\text{lo}} = \frac{21}{2676} \doteq 0.00785$, giving $\text{odds}_{\text{lo}} = \frac{\hat{p}_{\text{lo}}}{1-\hat{p}_{\text{lo}}} \doteq 0.00791$, or about 1 to 126 (or 125). **(c)** The odds ratio is $\text{odds}_{\text{hi}}/\text{odds}_{\text{lo}} \doteq 2.1181$. Odds of death from cardiovascular disease are about 2.1 times greater in the high blood pressure group.

14.26. **(a)** For female references, $\hat{p}_{\text{w}} = \frac{48}{60} = 0.8$, giving $\text{odds}_{\text{w}} = \frac{\hat{p}_{\text{w}}}{1-\hat{p}_{\text{w}}} = 4$ ("4 to 1"). **(b)** For male references, $\hat{p}_{\text{m}} = \frac{52}{132} = 0.\overline{39}$, giving $\text{odds}_{\text{m}} = \frac{\hat{p}_{\text{m}}}{1-\hat{p}_{\text{m}}} = 0.65$ ("13 to 20"). **(c)** The odds ratio is $\text{odds}_{\text{w}}/\text{odds}_{\text{m}} \doteq 6.1538$. (The odds of a juvenile reference are more than six times greater for females.)

14.27. **(a)** The interval is $b_1 \pm 1.96\text{SE}_{b_1}$, or 0.2452 to 1.2558. **(b)** $X^2 = \left(\frac{0.7505}{0.2578}\right)^2 \doteq 8.47$. This gives a P-value between 0.0025 and 0.005. **(c)** We have strong evidence that there is a real (significant) difference in risk between the two groups.

14.28. **(a)** The interval is $b_1 \pm 1.96\text{SE}_{b_1}$, or 1.0946 to 2.5396. **(b)** $X^2 = \left(\frac{1.8171}{0.3686}\right)^2 \doteq 24.3$. This gives $P < 0.0005$. **(c)** We have strong evidence that there is a real (significant) difference in juvenile references between male and female references.

14.29. **(a)** The estimated odds ratio is $e^{b_1} \doteq 2.1181$ (as we found in Exercise 14.25). Exponentiating the interval for β_1 in Exercise 14.27(a) gives the odds-ratio interval from about 1.28 to 3.51. **(b)** We are 95% confident that the odds of death from cardiovascular disease are about 1.3 to 3.5 times greater in the high blood pressure group.

Minitab output

```
                                                Odds     95% CI
Predictor      Coef   SE Coef       Z      P  Ratio  Lower  Upper
Constant   -4.83968  0.219078  -22.09  0.000
BP
 hi        0.750498  0.257840    2.91  0.004   2.12   1.28   3.51
```

14.30. **(a)** The estimated odds ratio is $e^{b_1} \doteq 6.1538$ (as we found in Exercise 14.26). Exponentiating the interval for β_1 in Exercise 14.28(a) gives the odds-ratio interval from about 2.99 to 12.67. **(b)** We are 95% confident that the odds of a juvenile reference are about 3 to 13 times greater among females.

Minitab output

```
                                                Odds     95% CI
Predictor       Coef   SE Coef       Z      P  Ratio  Lower  Upper
Constant   -0.430783  0.178131   -2.42  0.016
gender
 Female      1.81708  0.368641    4.93  0.000   6.15   2.99  12.67
```

14.31. **(a)** The model is $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$, where $x_i = 1$ if the ith person is over 40, and 0 if he/she is under 40. **(b)** p_i is the probability that the ith person is terminated; this model assumes that the probability of termination depends on age (over/under 40). In this case, that seems to have been the case, but we might expect that other factors were taken

into consideration. **(c)** The estimated odds ratio is $e^{b_1} \doteq 3.859$. (Of course, we can also get this from $\frac{41/765}{7/504}$.) We can also find, for example, a 95% confidence interval for b_1: $b_1 \pm 1.96\text{SE}_{b_1} = 0.5409$ to 2.1599. Exponentiating this translates to a 95% confidence interval for the odds: 1.7176 to 8.6701. The odds of being terminated are 1.7 to 8.7 times greater for those over 40. **(d)** Use a multiple logistic regression model, for example, $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$.

14.32. Before the change, the odds of a repair order being returned within five days were $\text{odds}_{\text{before}} = \frac{32}{168} \doteq 0.1905$. After the change, the odds rose to $\text{odds}_{\text{after}} = \frac{180}{2} = 9$. Using a logistic regression model $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$, with $x = 0$ for before and $x = 1$ for after, we estimate $b_0 = \log(\text{odds}_{\text{before}}) \doteq -1.6582$ and $b_1 = \log(\text{odds}_{\text{after}}) - \log(\text{odds}_{\text{before}}) \doteq 3.8555$. The odds ratio is $e^{b_1} \doteq 47.25$; the change increased the odds by a factor of about 47. (Of course, this ratio can be found without fitting the logistic regression model.)

14.33. There were a total of $1132 + 852 = 1984$ people in the sample, of which $643 + 349 = 992$ had completed college and the rest (also 992) had not. Among college graduates, 643 use the Internet for travel arrangements, so $\hat{p}_{\text{c}} = \frac{643}{992} \doteq 0.6482$ and $\text{odds}_{\text{c}} = \frac{643}{349} \doteq 1.8424$. Among those who have not completed college, $1132 - 643 = 489$ use the Internet, so $\hat{p}_{\text{n}} = \frac{489}{992} \doteq 0.4929$ and $\text{odds}_{\text{n}} = \frac{489}{503} \doteq 0.9722$. With the model $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$, where $x = 1$ for college graduates, we estimate $b_0 = \log(\text{odds}_{\text{n}}) \doteq -0.0282$, and $b_1 = \log(\text{odds}_{\text{c}}) - \log(\text{odds}_{\text{n}}) \doteq 0.6393$. (We can also find b_1 as the logarithm of the odds ratio $\text{odds}_{\text{c}}/\text{odds}_{\text{n}} \doteq 1.8952$.) The odds of a college graduate using the Internet for travel arrangements are about 1.90 times higher than those for a noncollege graduate.

Note: *Exercise 8.65 uses the same data but asks about proportions of college graduates among each group (users and nonusers), rather than proportions of users among graduates and nongraduates.*

14.34. A total of $493 + 378 + 477 + 200 = 1548$ people responded to the income question, of which $493 + 477 = 970$ had incomes below \$50,000 and the rest (578) had incomes over \$50,000. In the high-income group, 378 use the Internet for travel arrangements, so $\hat{p}_{\text{hi}} = \frac{378}{578} \doteq 0.6540$ and $\text{odds}_{\text{hi}} = \frac{378}{200} = 1.89$. In the low-income group, 493 use the Internet, so $\hat{p}_{\text{lo}} = \frac{493}{970} \doteq 0.5082$ and $\text{odds}_{\text{lo}} = \frac{493}{477} \doteq 1.0335$. With the model $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$, where $x = 1$ for income over \$50,000, we estimate $b_0 = \log(\text{odds}_{\text{lo}}) \doteq 0.0330$ and $b_1 = \log(\text{odds}_{\text{hi}}) - \log(\text{odds}_{\text{lo}}) \doteq 0.6036$. (We can also find b_1 as the logarithm of the odds ratio $\text{odds}_{\text{hi}}/\text{odds}_{\text{lo}} \doteq 1.8287$.) For someone with income over \$50,000, the odds of using the Internet for travel arrangements are about 1.83 times higher than those for someone in the lower-income group.

Note: *Exercise 8.66 uses the same data, but asks about income proportions in each group (users and nonusers), rather than proportions of users in each income group.*

14.35. For women, the proportion, odds, and log-odds for testing positive are $\hat{p}_f = \frac{27}{191} \doteq 0.1414$, $\text{odds}_f = \frac{\hat{p}_f}{1-\hat{p}_f} \doteq 0.1646$, and $\log(\text{odds}_f) \doteq -1.8040$. For men, these numbers are $\hat{p}_m = \frac{515}{1520} \doteq 0.3388$, $\text{odds}_m = \frac{\hat{p}_m}{1-\hat{p}_m} \doteq 0.5124$, and $\log(\text{odds}_m) \doteq -0.6686$. The fitted logistic regression model depends on whether we use $x = 1$ to indicate male or female (the exercise does not specify which to use). If $x = 1$ for males, we have:

$\log(\text{odds}) = -1.8040 + 1.1355x$ and odds ratio 3.1126

If $x = 1$ for females, we have:

$\log(\text{odds}) = -0.6686 - 1.1355x$ and odds ratio 0.3213

The odds of testing positive for alcohol are about 3 times higher for men than for women. (Or, those odds are about one-third as high for women as for men.)

Minitab output

```
- - - - - - - - - - - - - - - - - - -  Male = 1  - - - - - - - - - - - - - - - - - -
                                                          Odds      95% CI
Predictor      Coef    SE Coef      Z      P     Ratio  Lower  Upper
Constant   -1.80403   0.207682   -8.69  0.000
 Male       1.13545   0.214636    5.29  0.000   3.11    2.04   4.74
 - - - - - - - - - - - - - - - - - - -  Female = 1  - - - - - - - - - - - - - - - - - -
                                                          Odds      95% CI
Predictor      Coef    SE Coef      Z      P     Ratio  Lower  Upper
Constant   -0.668576  0.0541920  -12.34  0.000
 Female    -1.13545   0.214636    -5.29  0.000   0.32    0.21   0.49
```

14.36. For the model $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$, we obtain the fitted model $\log(\text{odds}) = b_0 + b_1 x = -7.2789 + 0.9399x$. (Here p is the probability that the cheese is acceptable, and x is the value of H2S.) We have $b_1 = 0.9399$ and $\text{SE}_{b_1} = 0.3443$, so we estimate that the odds ratio increases by a factor of $e^{b_1} \doteq 2.56$ for every unit increase in H2S. For testing $\beta_1 = 0$, we find $X^2 = 7.45$ ($P = 0.0063$), so we conclude that $\beta_1 \neq 0$. We are 95% confident that β_1 is in the interval $b_1 \pm 1.96\text{SE}_{b_1} = 0.2651$ to 1.6147; exponentiating this tells us that the odds ratio increases by a factor between 1.3035 and 5.0265 (with 95% confidence) for each unit increase in H2S.

Minitab output

```
                                                    Odds     95% CI
Predictor      Coef   SE Coef      Z      P     Ratio  Lower  Upper
Constant   -7.27894   2.52165  -2.89  0.004
h2s        0.939916  0.344331   2.73  0.006   2.56   1.30   5.03
```

GLMStat output

```
             estimate   se(est)  z ratio  Prob>|z|
1 Constant   -7.279     2.519    -2.889   0.0039
2 h2s        0.9399     0.3443   2.730    0.0063

             odds ratio  lower 95% ci   upper 95% ci
1 Constant   6.899e-4    4.945e-6       9.625e-2
2 h2s        2.560       1.304          5.024
```

14.37. For the model $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$, we obtain the fitted model $\log(\text{odds}) = b_0 + b_1 x = -10.7799 + 6.3319x$. (Here p is the probability that the cheese is acceptable and x is the value of Lactic.) We have $b_1 = 6.3319$ and $\text{SE}_{b_1} = 2.4532$, so we estimate that the odds ratio increases by a factor of $e^{b_1} \doteq 562.22$ for every unit increase in Lactic. For testing $\beta_1 = 0$, we find $X^2 = 6.66$ ($P = 0.0098$), so we conclude that $\beta_1 \neq 0$. We are 95%

confident that β_1 is in the interval $b_1 \pm 1.96\text{SE}_{b_1} = 1.5236$ to 11.1402; exponentiating this tells us that the odds ratio increases by a factor between 4.5889 and about 68,884 (with 95% confidence) for each unit increase in Lactic.

Minitab output

```
                                                          95% CI
Predictor      Coef  SE Coef      Z      P  Odds Ratio  Lower     Upper
Constant   -10.7799  3.97542  -2.71  0.007
lactic      6.33193  2.45317   2.58  0.010      562.24   4.59  68880.95
```

GLMStat output

```
             estimate   se(est)  z ratio  Prob>|z|
1 Constant   -10.78     3.973    -2.713   0.0067
2 lactic     6.332      2.452    2.583    0.0098

             odds ratio  lower 95% ci   upper 95% ci
1 Constant   2.081e-5    8.644e-9       5.012e-2
2 lactic     562.2       4.603          6.867e+4
```

14.38. The seven models are summarized below. The P-value in the right column is for the null hypothesis that all slopes equal 0 (i.e., the significance of the regression); all are significant.

For the models with two predictors, all have only one coefficient significantly different from 0 (in the last case, arguably neither coefficient is nonzero). The standard errors are given in parentheses below each coefficient; the six respective P-values are 0.4276, 0.0238; 0.3094, 0.0355; 0.0567, 0.1449.

In summary, we might conclude that H2S has the greatest effect: It had the smallest P-value among the three single-predictor models, and in the three multiple logistic regression models in which it was used, it had the minimum P-value. (It was the closest to being significant in the last two models in the table below.)

Note: *In Exercises 11.51 through 11.59, in considering multiple regression models to predict taste from Acetic, H2S, and Lactic, the H2S/Lactic model was chosen as the best. In every model that included H2S, its t statistic was the largest.*

Fitted Model				P
$\log(\text{odds}) = -13.71$	$+ 2.249$ Acetic			0.0285
$\log(\text{odds}) = -7.279$		$+ 0.9399$ H2S		0.0063
$\log(\text{odds}) = -10.78$			$+ 6.332$ Lactic	0.0098
$\log(\text{odds}) = -12.85$	$+ 1.096$ Acetic (1.382)	$+ 0.8303$ H2S (0.3673)		0.0008
$\log(\text{odds}) = -16.56$	$+ 1.309$ Acetic (1.288)		$+ 5.257$ Lactic (2.500)	0.0016
$\log(\text{odds}) = -11.72$		$+ 0.7346$ H2S (0.3866)	$+ 3.777$ Lactic (2.596)	0.0003
$\log(\text{odds}) = -14.26$	$+ 0.584$ Acetic	$+ 0.6849$ H2S	$+ 3.468$ Lactic	0.0010

14.39. Portions of SAS and GLMStat output are given below. **(a)** The X^2 statistic for testing this hypothesis is 33.65 (df $=$ 3), which has $P = 0.0001$. We conclude that at least one coefficient is not 0. **(b)** The fitted model is log(odds) $=$ $-6.053 + 0.3710$ HSM $+ 0.2489$ HSS $+ 0.03605$ HSE. The standard errors of the three coefficients are 0.1302, 0.1275, and 0.1253, giving respective 95% confidence intervals 0.1158 to 0.6262, -0.0010 to 0.4988, and -0.2095 to 0.2816. **(c)** Only the coefficient of HSM is significantly different from 0, though HSS may also be useful.

Note: *In the multiple regression case study of Chapter 11, HSM was also the only significant explanatory variable among high school grades, and HSS was not even close to significant. See Figure 11.4 on page 694 of the text.*

SAS output

```
                        Intercept
             Intercept      and
Criterion      Only      Covariates   Chi-Square for Covariates
-2 LOG L       295.340      261.691      33.648 with 3 DF (p=0.0001)

               Analysis of Maximum Likelihood Estimates

                Parameter  Standard     Wald        Pr >     Standardized
Variable  DF     Estimate    Error   Chi-Square  Chi-Square    Estimate

INTERCPT  1      -6.0528    1.1562     27.4050      0.0001            .
HSM       1       0.3710    0.1302      8.1155      0.0044     0.335169
HSS       1       0.2489    0.1275      3.8100      0.0509     0.233265
HSE       1       0.0361    0.1253      0.0828      0.7736     0.029971
```

GLMStat output

```
            estimate   se(est)   z ratio   Prob>|z|
1 Constant -6.053      1.156     -5.236     <0.0001
2 HSM       0.3710     0.1302     2.849      0.0044
3 HSS       0.2489     0.1275     1.952      0.0509
4 HSE       3.605e-2   0.1253     0.2877     0.7736
```

14.40. Portions of SAS and GLMStat output are given on the next page. **(a)** The X^2 statistic for testing this hypothesis is 14.2 (df $=$ 2), which has $P = 0.0008$. We conclude that at least one coefficient is not 0. **(b)** The model is log(odds) $=$ $-4.543 + 0.003690$ SATM $+ 0.003527$ SATV. The standard errors of the two coefficients are 0.001913 and 0.001751, giving respective 95% confidence intervals -0.000059 to 0.007439, and 0.000095 to 0.006959. (The first coefficient has a P-value of 0.0537 and the second has $P = 0.0440$.) **(c)** We (barely) cannot reject $\beta_{\text{SATM}} = 0$—though because 0 is just in the confidence interval, we are reluctant to discard SATM. Meanwhile, we conclude that $\beta_{\text{SATV}} \neq 0$.

Note: *By contrast, with multiple regression of GPA on SAT scores, we found SATM useful but not SATV. See Figure 11.7 on page 698 of the text.*

SAS output

```
                          Intercept
               Intercept      and
Criterion        Only      Covariates   Chi-Square for Covariates
-2 LOG L         295.340      281.119      14.220 with 2 DF (p=0.0008)

                Analysis of Maximum Likelihood Estimates

                 Parameter  Standard     Wald        Pr >     Standardized
Variable  DF     Estimate     Error    Chi-Square  Chi-Square    Estimate

INTERCPT  1       -4.5429    1.1618     15.2909      0.0001            .
SATM      1       0.00369   0.00191      3.7183      0.0538      0.175778
SATV      1       0.00353   0.00175      4.0535      0.0441      0.180087
```

GLMStat output

```
            estimate   se(est)   z ratio   Prob>|z|
1 Constant -4.543      1.161     -3.915     <0.0001
2 SATM      3.690e-3   1.913e-3   1.929      0.0537
3 SATV      3.527e-3   1.751e-3   2.014      0.0440
```

14.41. The coefficients and standard errors for the fitted model are below. Note that the tests requested in (a) and (b) are not available with all software packages. **(a)** The X^2 statistic for testing this hypothesis is given by SAS (below) as 19.2256 (df = 3); because $P = 0.0002$, we reject H_0 and conclude that high school grades add a significant amount to the model with SAT scores. **(b)** The X^2 statistic for testing this hypothesis is 3.4635 (df = 2); because $P = 0.1770$, we cannot reject H_0; SAT scores do not add significantly to the model with high school grades. **(c)** For modeling the odds of HIGPA, high school grades (specifically HSM, and to a lesser extent HSS) are useful, while SAT scores are not.

SAS output

```
                Analysis of Maximum Likelihood Estimates

                 Parameter  Standard     Wald        Pr >     Standardized
Variable  DF     Estimate     Error    Chi-Square  Chi-Square    Estimate

INTERCPT  1       -7.3732    1.4768     24.9257      0.0001            .
HSM       1        0.3427    0.1419      5.8344      0.0157      0.309668
HSS       1        0.2249    0.1286      3.0548      0.0805      0.210704
HSE       1        0.0190    0.1289      0.0217      0.8829      0.015784
SATM      1      0.000717   0.00220      0.1059      0.7448      0.034134
SATV      1       0.00289   0.00191      2.2796      0.1311      0.147566

                      Linear Hypotheses Testing

                         Wald                          Pr >
        Label         Chi-Square              DF    Chi-Square
        HS               19.2256               3        0.0002
        SAT               3.4635               2        0.1770
```

GLMStat output

```
            estimate   se(est)   z ratio   Prob>|z|
1 Constant -7.373      1.477     -4.994     <0.0001
2 SATM      7.166e-4   2.201e-3   0.3255     0.7448
3 SATV      2.890e-3   1.914e-3   1.510      0.1311
4 HSM       0.3427     0.1419     2.416      0.0157
5 HSS       0.2249     0.1286     1.748      0.0805
6 HSE       1.899e-2   0.1289     0.1473     0.8829
```

14.42. **(a)** The fitted model is log(odds) $= -0.6124 + 0.0609$ Gender; the coefficient of gender is not significantly different from 0 ($z = 0.21$, $P = 0.8331$). **(b)** Now, log(odds) $= -5.214 + 0.3028$ Gender $+ 0.004191$ SATM $+ 0.003447$ SATV. In this model, gender is still not significant ($P = 0.3296$). **(c)** Gender is not useful for modeling the odds of HIGPA.

GLMStat output

Gender only

	estimate	se(est)	z ratio	Prob>|z|
1 Constant	-0.6124	0.4156	-1.474	0.1406
2 Gender	6.087e-2	0.2889	0.2107	0.8331

Gender and SAT scores

	estimate	se(est)	z ratio	Prob>|z|
1 Constant	-5.214	1.362	-3.828	0.0001
2 Gender	0.3028	0.3105	0.9750	0.3296
3 SATM	4.191e-3	1.987e-3	2.109	0.0349
4 SATV	3.447e-3	1.760e-3	1.958	0.0502

14.43. The models reported below are for the odds of *death*, as requested in the instructions. If a student models odds of survival, or codes the indicator variables for hospital and condition differently, his or her answers will be slightly different from these (but the conclusions should be the same). **(a)** The fitted model is log(odds) $= -3.892 + 0.4157$ Hospital, using 1 for Hospital A and 0 for Hospital B. With $b_1 \doteq -0.4157$ and $\text{SE}_{b_1} \doteq 0.2831$, we find that $z = -1.47$ or $X^2 = 2.16$ ($P = 0.1420$), so we do not have evidence to suggest that β_1 is not 0. A 95% confidence interval for β_1 is -0.1392 to 0.9706 (this interval includes 0). We estimate the odds ratio to be $e^{b_1} \doteq 1.515$, with confidence interval 0.87 to 2.64 (this includes 1, since β_1 might be 0). **(b)** The fitted model is log(odds) $= -3.109 - 0.1320$ Hospital $- 1.266$ Condition; as before, use 1 for Hospital A and 0 for Hospital B, 1 for good condition and 0 for poor. The estimated odds ratio is $e^{b_1} \doteq 0.8764$, with confidence interval 0.48 to 1.60. **(c)** In neither case is the effect of Hospital significant. However, we can see the effect of Simpson's paradox in the coefficient of Hospital, or equivalently in the odds ratio. In the model with Hospital alone, this coefficient was positive and the odds ratio was greater than 1, meaning Hospital A patients have higher odds of death. When condition is added to the model, this coefficient is negative and the odds ratio is less than 1, meaning Hospital A patients have lower odds of death.

GLMStat output

Hospital only

	estimate	se(est)	z ratio	Prob>|z|
1 Constant	-3.892	0.2525	-15.41	<0.0001
2 Hosp	0.4157	0.2831	-1.469	0.1420

	odds ratio	lower 95% ci	upper 95% ci
1 Constant	2.041e-2	1.244e-2	3.348e-2
2 Hosp	1.515	0.8701	2.639

Hospital and condition

	estimate	se(est)	z ratio	Prob>|z|
1 Constant	-3.109	0.2959	-10.51	<0.0001
2 Hosp	-0.1320	0.3078	-0.4288	0.6681
3 Cond	-1.266	0.3218	-3.935	<0.0001

	odds ratio	lower 95% ci	upper 95% ci
1 Constant	4.463e-2	2.499e-2	7.971e-2
2 Hosp	0.8764	0.4794	1.602
3 Cond	0.2820	0.1501	0.5298