CATEGORICAL DATA ANALYSIS

Outlines of Solutions to Selected Homework Problems

Alan Agresti

January 5, 2004, ©Alan Agresti 2004

This handout contains solutions and hints to solutions for many of the STA 6505 homework exercises from Categorical Data Analysis, second edition, by Alan Agresti (John Wiley, & Sons, 2002). It should not be distributed elsewhere without permission of the author. Additional solutions for odd-numbered exercises are available at the website for the text, http://www.stat.ufl.edu/~aa/cda/cda.html. Please report any errors in these solutions so they can be corrected.

Chapter 1

1. a. nominal, b. ordinal, c. interval, d. nominal, e. ordinal, f. nominal, g. ordinal.

3. π varies from batch to batch, so the counts come from a mixture of binomials rather than a single $\text{bin}(n, \pi)$. $\text{Var}(Y) = E[\text{Var}(Y \mid \pi)] + \text{Var}[E(Y \mid \pi)] > E[\text{Var}(Y \mid \pi)] = E[n\pi(1-\pi)].$

7. a. $\ell(\pi) = \pi^2$, so $\hat{\pi} = 1.0$.
b. Wald statistic $z = (1.0 - .5)/\sqrt{1.0(0)/20} = 1.0 \pm 0.0$, Wald CI is $1.0 \pm 1.96\sqrt{1.0(0)/20} = 0.0 \pm 0.0$, or $(1.0, 1.0)$.
c. $z = (1.0 - .5)/\sqrt{5.0(0)/20} = 0.96, P < .0001$. Score CI is $(0.839, 0.100)$.
d. Test statistic $2(20)\log(20/10) = 27.7$, $df = 1$. From problem 1.25a, the CI is $(\exp(1.96^2/40), 1) = (0.908, 1.0)$.

11. $\text{Var}(\hat{\pi}) = \pi(1-\pi)/n$ decreases as π moves toward 0 or 1 from 0.5.

12. a. $\text{Var}(Y) = n\pi(1-\pi)$, binomial.
b. $\text{Var}(Y) = \sum \text{Var}(Y_i) + 2 \sum_{i<j} \text{Cov}(Y_i, Y_j) = n\pi(1-\pi) + 2\rho\pi(1-\pi) \left( \frac{n}{2} \right) > n\pi(1-\pi)$.
c. $\text{Var}(Y) = E[\text{Var}(Y \mid \pi)] + \text{Var}[E(Y \mid \pi)] = E[n\pi(1-\pi)] + \text{Var}(n\pi) = n\rho - nE(\pi^2) + [n^2E(\pi^2) - n^2\rho^2] = n\rho + (n^2 - n)[E(\pi^2) - \rho^2] = n\rho(1-\rho) + (n^2 - n)\text{Var}(\pi) > n\rho(1-\rho)$.
d. Conditionally, $Y$ is the sum of non-identical Bernoulli trials, so is not binomial. Conditionally, the probability of a particular sequence is $\prod \pi_i^{y_i}(1-\pi_i)^{1-y_i}$. Since the responses are independent, the unconditional probability of that sequence is $\prod(E\pi_i)^{y_i}(1-E(\pi_i))^{1-y_i}$, which corresponds to a sequence of identical, independent trials.

18. Conditional on $n = y_1 + y_2$, $y_1$ has a $\text{bin}(n, \pi)$ distribution with $\pi = \mu_1/(\mu_1 + \mu_2)$, which is .5 under $H_0$. The large sample score test uses $z = (y_1/n - .5)/\sqrt{.5(0)/n}$. If $(\ell, u)$ denotes a CI for $\pi$ (e.g., the score CI), then the CI for $\pi/(1-\pi) = \mu_1/\mu_2$ is $[\ell/(1-\ell), u/(1-u)]$.

30. a. The kernel of the log likelihood is $L(\theta) = n_1\log \theta^2 + n_2\log[2\theta(1-\theta)] + n_3\log(1-\theta)^2$. Take $\partial L/\partial \theta = 2n_1/\theta + n_2/\theta - n_2/(1-\theta) - 2n_3/(1-\theta) = 0$ and solve for $\theta$.
b. Find the expectation using $E(n_1) = n\theta^2$, etc. Then, the asymptotic variance is the inverse information $= \theta(1-\theta)/2n$, and thus the estimated $SE = \sqrt{\hat{\theta}(1-\hat{\theta})/2n}$.
c. The estimated expected counts are $[n\hat{\theta}^2, 2n\hat{\theta}(1-\hat{\theta}), n(1-\hat{\theta})^2]$. Compare these to the observed counts $(n_1, n_2, n_3)$ using $X^2$ or $G^2$, with $df = (3-1) - 1 = 1$, since 1 parameter is estimated.

33. c. Let $\hat{\pi} = n_1/n$, and $(1 - \hat{\pi}) = n_2/n$, and denote the null probabilities in the two categories by $\pi_0$ and $(1-\pi_0)$. Then, $X^2 = (n_1 - n\pi_0)^2/n\pi_0 + (n_2 - n(1-\pi_0))^2/n(1-\pi_0)$.
\( n[(\hat{\pi} - \pi_0)^2(1 - \pi_0) + ((1 - \hat{\pi}) - (1 - \pi_0))^2 \pi_0]/\pi_0(1 - \pi_0), \)

which equals \( (\pi - \pi_0)^2/[\pi_0(1 - \pi_0)/n] = \frac{z^2}{S^2} \).

34. Let \( X \) be a random variable that equals \( \pi_{j_0}/\hat{\pi}_j \) with probability \( \hat{\pi}_j \). By Jensen’s inequality, since the negative log function is convex, \( E(-\log X) \geq -\log(EX) \). Hence, \( E(-\log X) = \sum \hat{\pi}_j \log(\hat{\pi}_j/p_{j0}) \geq -\log[\sum \hat{\pi}_j(\pi_{j0}/\hat{\pi}_j)] = -\log(\sum \pi_{j0}) = -\log(1) = 0 \). Thus \( G^2 = 2nE(-\log X) \geq 0 \).

Chapter 2

3. The odds ratio is \( \hat{\theta} = 7.965 \); the relative risk of fatality for ‘none’ is 7.897 times that for ‘seat belt’; difference of proportions = .0085. The proportion of fatal injuries is close to zero for each row, so the odds ratio is similar to the relative risk.

4. a. Relative risk.
   b. (i) \( \pi_1 = .55\pi_2 \), so \( \pi_1/\pi_2 = .55 \).
   (ii) \( 1/.55 = 1.82 \).

8. a. The quoted interpretation is that of the relative risk. Should substitute odds for probability.
   b. For females, probability = \( 2.9/(1 + 2.9) = .744 \). Odds for males = \( 2.9/11.4 = .25 \), so probability = \( .25/(1 + .25) = .20 \).

10. a. \( (.847/.153)/(906/.094) = .574 \).
   b. This is interpretation for relative risk, not the odds ratio. The actual relative risk = \( .847/.906 = .935 \); i.e., 60% should have been 93.5%.

11. a. Relative risk: Lung cancer, 14.00; Heart disease, 1.62. (Cigarette smoking seems more highly associated with lung cancer)
   Difference of proportions: Lung cancer, .00130; Heart disease, .00256. (Cigarette smoking seems more highly associated with heart disease)
   Odds ratio: Lung cancer, 14.02; Heart disease, 1.62. e.g., the odds of dying from lung cancer for smokers are estimated to be 14.02 times those for nonsmokers. (Note similarity to relative risks.)
   b. Difference of proportions describes excess deaths due to smoking. That is, if \( N = \) no. smokers in population, we predict there would be \( .00130N \) fewer deaths per year from lung cancer if they had never smoked, and \( .00256N \) fewer deaths per year from heart disease. Thus elimination of cigarette smoking would have biggest impact on deaths due to heart disease.

12. Marginal odds ratio = 1.84, but most conditional odds ratios are close to 1.0 except in Department A where odds ratio = .35. Note that males tend to apply in greater numbers to Departments A and B, in which admissions rates are relatively high, and females tend to apply in greater numbers to Departments C, D, E, F, in which admissions rates are relatively low. This results in the marginal association whereby the odds of admission for males are 84% higher than those for females.

16. Kentucky: Counts are (31, 360 / 7, 50) when victim was white and (0, 18 / 2, 106) when victim was black. Conditional odds ratios are 0.62 and 0, whereas marginal odds ratio is 1.42. Simpson’s paradox occurs. Whites tend to kill whites and blacks tend to kill blacks, and killing a white is more likely to result in the death penalty.

21. a. Let “pos” denote positive diagnosis, “dis” denote subject has disease.

\[
P(\text{dis|pos}) = \frac{P(\text{pos|dis})P(\text{dis})}{P(\text{pos|dis})P(\text{dis}) + P(\text{pos|no dis})P(\text{no dis})}
\]

b. \( .95(.005)/[.95(.005) + .05(.995)] = .087 \).

<table>
<thead>
<tr>
<th>Test</th>
<th>Reality</th>
<th>+</th>
<th>-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reality</td>
<td>+</td>
<td>.00475</td>
<td>.00025</td>
<td>.005</td>
</tr>
<tr>
<td>-</td>
<td>.04975</td>
<td>.94525</td>
<td>.995</td>
<td></td>
</tr>
</tbody>
</table>
Nearly all (99.5%) subjects are not HIV+. The 5% errors for them swamp (in frequency) the 95% correct cases for subjects who truly are HIV+. The odds ratio = 361; i.e., the odds of a positive test result are 361 times higher for those who are HIV+ than for those not HIV+.

23. a. The numerator is the extra proportion that got the disease above and beyond what the proportion would be if no one had been exposed (which is $P(D \mid \bar{E})$).
24. a. For instance, if first row becomes first column and second row becomes second column, the table entries become

$$
n_{11} \quad n_{21} \\
n_{12} \quad n_{22}
$$

The odds ratio is the same as before. The difference of proportions and relative risk are only invariant to multiplication of cell counts within rows by a constant.

29. Yes, this would be an occurrence of Simpson’s paradox. One could display the data as a $2 \times 2 \times K$ table, where rows = (Smith, Jones), columns = (hit, out) response for each time at bat, layers = (year 1, . . . , year K). This could happen if Jones tends to have relatively more observations (i.e., “at bats”) for years in which his average is high.

### Chapter 3

3. $X^2 = 0.27$, $G^2 = 0.29$, P-value about 0.6. The free throws are plausibly independent. However, sample odds ratio is 0.77, and 95% CI for true odds ratio is (0.29, 2.07), quite wide.
4. a. $G^2 = 90.3$, df = 2; very strong evidence of association ($P < .0001$).
   c. $G^2 = 7.2$ for comparing races on (Democrat, Independent) choice, and $G^2 = 83.2$ for comparing races on (Dem. + Indep., Republican) choice; extremely strong evidence that whites are more likely than blacks to be Republicans. (To get independent components, combine the two groups compared in the first analysis and compare them to the other group in the second analysis.)

13. a. It is plausible that control of cancer is independent of treatment used. (i) P-value = .3808 is hypergeometric probability $P(n_{11} = 21$ or 22 or 23), (ii) P-value = 0.638 is sum of probabilities that are no greater than the probability (2.755) of the observed table.

b. The asymptotic CI (.31, 14.15) uses the delta method formula (3.1) for the $SE$. The ‘exact’ CI (.21, 27.55) is the Cornfield tail-method interval that guarantees a coverage probability of at least .95.

c. $.3808 - .5(2.755) = .243$. With this type of $P$-value, the actual error probability tends to be closer to the nominal value, the sum of the two one-sided $P$-values is 1, and the null expected value is 0.5; however, it does not guarantee that the actual error probability is no greater than the nominal value.

22. Note the delta method was used to derive the standard error of the sample logit in Sec. 3.1.6. If the endpoints of the CI for the logit are $(\ell, u)$, then the corresponding endpoints of the CI for $\pi$ are its inverse, $[e^\ell/(1 + e^\ell), e^u/(1 + e^u)]$.

24. One can apply the delta method directly, or else use what we already know about the log odds ratio, $L = \log(\theta)$. Since $\theta = \exp(L)$, $\partial \theta / \partial L = \exp(L) = \theta$, so

$$
\text{Var}(\hat{\theta}) \approx (\partial \theta / \partial L)^2 \text{Var}(L) \approx (\theta^2 / n)(\sum 1 / \pi_{ij}).
$$

Now $Q = (\theta - 1)/(\theta + 1)$, $\partial Q / \partial \theta = 2/(1 + \theta)^2$, so

$$
\text{Var}(Q) \approx (\partial Q / \partial \theta)^2 \text{Var}(\theta) = (2/(1 + \theta)^2)^2(\theta^2 / n)(\sum 1 / \pi_{ij}).
$$

Now $\theta = (1 + Q)/(1 - Q)$, $\theta/(1 + \theta)^2 = (1 - Q^2)/4$, so

$$
\text{Var}(\hat{Q}) \approx [(1 - Q^2)^2/4n](\sum 1 / \pi_{ij}).
$$

26. Use formula (3.9), noting that the partial derivative of the measure with respect to $\pi_i$ is just $\eta_i / \delta^2$.

29. For any “reasonable” significance test, whenever $H_0$ is false, the test statistic tends to be larger and the $P$-value tends to be smaller as the sample size increases. Even if $H_0$ is just slightly false, the $P$-value will be small if the sample size is large enough. Most statisticians feel we learn more by estimating parameters using confidence intervals than by conducting significance tests.

31. a. Note $\theta = \pi_{1+} = \pi_{+1}$.
b. The log likelihood has kernel
\[ L = n_{11} \log(\theta^2) + (n_{12} + n_{21}) \log[\theta(1 - \theta)] + n_{22} \log(1 - \theta)^2 \]
\[ \frac{\partial L}{\partial \theta} = 2n_{11}/\theta + (n_{12} + n_{21})/\theta - (n_{12} + n_{21})/(1 - \theta) - 2n_{22}/(1 - \theta) = 0 \]
gives \[ \hat{\theta} = \frac{2n_{11} + n_{12} + n_{21}}{2(n_{11} + n_{12} + n_{21} + n_{22})} = \frac{n_{1+} + n_{+1}}{2n} = (p_1 + p_+1)/2. \]
c. Calculate estimated expected frequencies (e.g., \[ \hat{\mu}_{11} = n\hat{\theta}^2 \]), and obtain Pearson \( X^2 \), which is 2.8. We estimated one parameter, so \( df = (4-1)-1 = 2 \) (one higher than in testing independence without assuming identical marginal distributions). The free throws are plausibly independent and identically distributed.

34. a. \[ \sum n_i[(n_i - \mu_i)/\mu_i] = \sum(n_i - \mu_i + \mu_i)((n_i - \mu_i)/\mu_i) = X^2 + \sum(n_i - \mu_i) = X^2. \]
b. Use fact that \[ \log(t) = \lim_{h \to 0}(t^h - 1)/h. \] In (b) identify \( t \) with \( n_i/\hat{\mu}_i \) and \( h \) with \( \lambda \). In (c) identify \( h \) with \( \lambda + 1 \), and express the term following the summation as \( \hat{\mu}_i[(n_i/\hat{\mu}_i)^{\lambda+1} - 1] + (\hat{\mu}_i - n_i). \) The latter term sums to zero, and the first term divided by \( (\lambda + 1) \) tends to \( \hat{\mu}_i \log(n_i/\hat{\mu}_i) = -\hat{\mu}_i \log(\hat{\mu}_i/n_i); \) multiplying this by \( 2/\lambda \) for \( \lambda = -1 \) and summing gives the result.

e. \[ -8\sum \sqrt{\mu_i} - \sqrt{n_i}/2 = 8\sum(\sqrt{\mu_i} - \sqrt{n_i})^2 - 8\sum\sqrt{\mu_i}(\sqrt{\mu_i} - \sqrt{n_i}) = 4\sum(\sqrt{\mu_i} - \sqrt{n_i})^2, \] since \( \sum\sqrt{\mu_i} = \sum n_i = n. \)

35. Because \( G^2 \) for full table is \( G^2 \) for collapsed table + \( G^2 \) for table consisting of the two rows that are combined.

40. For most \( \alpha \) values, \( P(P\text{-value} \leq \alpha | n_{1+}, n_{+1}) < \alpha \), since the attainable \( P \) values do not include \( \alpha \). The average of this over all possible marginal configurations is \( < \alpha \). Because of this, Fisher’s exact test and other discrete tests are “conservative”, the actual probability of rejecting a true null hypothesis being smaller than a fixed significance level (such as .05) that one chooses to conduct the test.

42. For the unconditional test, for common value \( \pi \) under \( H_0 \), the probability of the observed table is \( \pi^3(1-\pi)^3 \). This is the most extreme result, and the \( P \)-value is the supremum of this over all \( \pi \), which is \( .5^6 = 1/64 \). For Fisher’s exact test, the \( P \)-value is the hypergeometric probability of this table, or \( 1/20 \).

43. The observed table has \( X^2 = 6 \). As noted in problem 42, the probability of this table is highest at \( \pi = .5 \). For given \( \pi \), \( P(X^2 \geq 6) = \sum_k P(X^2 \geq 6 \text{ and } n_{+1} = k) = \sum_k P(X^2 \geq 6 | n_{+1} = k)P(n_{+1} = k), \) and \( P(X^2 \geq 6 | n_{+1} = k) \) is the \( P \)-value for Fisher’s exact test.

Chapter 4

1. a. Roughly 3%.
   b. Estimated proportion \( \hat{\pi} = -.0003 + .0304(.0774) = .0021 \). The actual value is 3.8 times the predicted value, which together with Fig. 4.8 suggests it is an outlier.
   c. \( \hat{\pi} = e^{-6.2182}/[1 + e^{-6.2182}] = .0020 \). Palm Beach County is an outlier.

2. a. With each decade, the estimated probability of a complete game decreases by .069.
   b. Predicted probabilities are .664, -.006, and -.075; not plausible to have negative estimates.
   c. Predicted probabilities are .119, .090, and .067 (e.g., at \( x = 12 \), \( \log(\hat{\pi}) = 1.148 - 0.315(12) = -2.636 \), and \( \hat{\pi} = e^{(-2.636)/(1 + e^{(-2.636)})} = 0.067 \), which are more plausible.

5. a. \( \hat{\pi} = -.145 + .323(\text{weight}); \) at weight = 5.2, predicted probability = 1.53, much higher than the upper bound of 1.0 for a probability.
   c. \( \logit(\hat{\pi}) = -3.695 + 1.815(\text{weight}); \) at 5.2 kg, predicted logit = 5.74, and \( \log(.9968/\text{.0032}) \) = 5.74.
   d. \( \text{probit}(\hat{\pi}) = -2.238 + 1.099(\text{weight}). \) \( \hat{\pi} = \Phi^{-1}(-2.238 + 1.099(5.2)) = \Phi^{-1}(3.48) = .9997 \)

17. The link function determines the function of the mean that is predicted by the linear predictor in a GLM. The identity link models the binomial probability directly as a linear function of the predictors. It is not often used, because probabilities must fall between 0 and 1, whereas straight lines provide predictions that can be any real number. When the probability is near 0 or 1 for some predictor values or when there are several predictors, it is not unusual to get predicted probabilities below 0 or above 1.
With the logit link, any real number predicted value for the linear model corresponds to a probability between 0 and 1. Similarly, Poisson means must be nonnegative. If we use an identity link, we could get negative predicted values. With the log link, a predicted negative log mean still corresponds to a positive mean.

19. With single predictor, \( \log[p(x)] = \alpha + \beta x \). Since \( \log[p(x + 1)] - \log[p(x)] = \beta \), the relative risk is \( \pi(x + 1)/\pi(x) = \exp(\beta) \). A restriction of the model is that to ensure \( 0 < \pi(x) < 1 \), it is necessary that \( \alpha + \beta x < 0 \).

20. a. \( \partial \pi(x)/\partial x = \beta e^{\alpha + \beta x}/[1 + e^{\alpha + \beta x}]^2 \), which is positive if \( \beta > 0 \). Note that the general logistic cdf on p. 125 has mean \( \mu \) and standard deviation \( \sqrt{\pi/\beta} \). Writing \( \alpha + \beta x \) in the form \( (x - \alpha/\beta)/(1/\beta) \), we identify \( \mu \) with \( \alpha/\beta \) and \( \tau \) with \( 1/\beta \), so the standard deviation is \( \pi/\beta \sqrt{\beta} \) when \( \beta > 0 \).

22. From the product of the \( N \) binomial mass functions, the model is then equivalent to having a single binomial variate \( \hat{\pi} = \sum_n y_i \) from a \( \text{bin}(\sum_n n_i, \pi) \) distribution, for which the ML estimate is \( \hat{\pi} = (\sum_i y_i)/\sum n_i \). When all \( n_i = 1 \), each row has fitted values \( \hat{\pi} \) and \( (1 - \hat{\pi}) \). If \( y_i = 1 \), then the contribution to \( X^2 \) in that row is

\[
\frac{(1 - \hat{\pi})^2}{\pi} + \frac{[0 - (1 - \hat{\pi})]^2}{1 - \hat{\pi}} = \frac{1 - \hat{\pi}}{\pi}.
\]

If \( y_i = 0 \), then the contribution to \( X^2 \) is

\[
\frac{(0 - \hat{\pi})^2}{\pi} + \frac{[1 - (1 - \hat{\pi})]^2}{1 - \hat{\pi}} = \frac{\hat{\pi}}{1 - \hat{\pi}}.
\]

There are \( \sum_i y_i = N\hat{\pi} \) rows of the first type and \( N(1 - \hat{\pi}) \) of the second type so the value of \( X^2 \) is

\[
N\hat{\pi}\left(\frac{1 - \hat{\pi}}{\pi} + N(1 - \hat{\pi})\frac{\hat{\pi}}{1 - \hat{\pi}} = N.
\]

29. a. Since \( \phi \) is symmetric, \( \Phi(0) = .5 \). Setting \( \alpha + \beta x = 0 \) gives \( x = -\alpha/\beta \).

b. The derivative of \( \Phi \) at \( x = -\alpha/\beta \) is \( \beta \phi(\alpha + \beta(-\alpha/\beta)) = \beta \phi(0) \). The logistic pdf has \( \phi(x) = e^x/(1+e^x)^2 \) which equals .25 at \( x = 0 \); the standard normal pdf equals \( 1/\sqrt{2\pi} \) at \( x = 0 \).

c. \( \Phi(\alpha + \beta x) = \Phi(x - (\alpha/\beta)). \)

30. Setting \( \sigma = 1 \) to illustrate, \( f(y; \mu) = [(2\pi)^{-1/2} \exp(-\mu^2/2)][\exp(-y^2/2)] \exp(\mu y) \), so the natural parameter is the mean \( \mu \). In the usual regression model, the random component is normal, and the systematic component is the linear predictor (e.g., \( \alpha + \beta x \)). The link is the identity link, since the mean is directly modeled.

34. a. Note ML estimate \( \hat{\beta} \) satisfies \( L'(\hat{\beta}) = 0 \), so dropping higher-order terms in the expansion yields

\[
0 \approx L'(\beta^{(0)}) + (\hat{\beta} - \beta^{(0)})L''(\beta^{(0)}).
\]

Solving for \( \hat{\beta} \) yields the approximation

\[
\hat{\beta} = \beta^{(0)} - L'(\beta^{(0)})/L''(\beta^{(0)}),
\]

which we use as \( \beta^{(1)} \), the next approximation for \( \hat{\beta} \). Note this is just the one-dimensional version of the Newton-Raphson method.

b. Use same argument as in (a), replacing \( \beta^{(0)} \) by \( \beta^{(t)} \).

Chapter 5

1. a. \( \hat{\pi} = e^{-3.7771+.1449(8)}/[1 + e^{-3.7771+.1449(8)}] \).

b. \( \hat{\pi} = .5 \) at \( -\hat{\alpha}/\hat{\beta} = 3.7771/.1449 = 26. \)
c. At LI = 8, \( \hat{\pi} = .068 \), so rate of change is \( \hat{\beta}(1 - \hat{\pi}) = .1449(.068)(.932) = .009 \).

e. \( e^{\hat{\beta}} = e^{1.449} = 1.16 \).

f. The odds of remission at LI = \( x + 1 \) are estimated to fall between 1.029 and 1.298 times the odds of remission at LI = \( x \).

g. Wald statistic = \( (.1449/0.0593)^2 = 5.96, df = 1, P\text{-value} = .0146 \) for \( H_0: \beta \neq 0 \).

h. Likelihood-ratio statistic = 34.37 - 26.07 = 8.30, \( df = 1, P\text{-value} = .004 \).

4. a. Since \( \log(\hat{\pi}) = -.573 + .0043(\text{age}) \), Likelihood ratio statistic = .55, Wald statistic = \( (.0043/0.058)^2 = .54 \), both based on \( df = 1 \), showing no evidence of an age effect.

c. The likelihood ratio statistic for the age squared term equals 6.3 (\( df = 1 \)), showing strong evidence of an effect. The plot shows that the estimated probability of kyphosis decreases at relatively low and relatively high age levels.

8. Estimated odds of contraceptive use for those with at least 1 year of college were \( e^{.501} = 1.65 \) times the estimated odds for those with less than 1 year of college. The 95% Wald CI for the true odds ratio is \( \exp[.501 \pm 1.96(.077)] = (e^{.350}, e^{.652}) = (1.42, 1.92) \).

9. a. Black defendants with white victims had estimated probability \( e^{-3.5961+2.4044}/[1+e^{-3.5961+2.4044}] = .23 \).

b. For a given defendant’s race, the odds of the death penalty when the victim was white are estimated to be between \( e^{3.068} = 3.7 \) and \( e^{3.7175} = 41.2 \) times the odds when the victim was black.

c. Wald statistic (-.8678/3671)^2 = 5.6, LR statistic = 5.0, each with \( df = 1 \). \( P\text{-value} = .025 \) for LR statistic.

d. \( G^2 = .38, X^2 = .20, df = 1 \), so model fits well.

12. Logit model with additive factor effects for age and gender has \( G^2 = 0.1 \) and \( X^2 = 0.1 \) with \( df = 2 \). Estimated odds of females still being missing are \( \exp(0.38) = 1.46 \) times those for males, given age. Estimated odds are considerably higher for those aged at least 19 than for other two age groups, given gender.

15. \( R = 1: \logit(\hat{\pi}) = -6.7 + .1A + 1.4S \). \( R = 0: \logit(\hat{\pi}) = -7.0 + .1A + 1.2S \).

The YS conditional odds ratio is \( \exp(1.4) = 4.1 \) for blacks and \( \exp(1.2) = 3.3 \) for whites. Note that .2, the coeff. of the cross-product term, is the difference between the log odds ratios 1.4 and 1.2. The coeff. of \( S \) of 1.2 is the log odds ratio between \( Y \) and \( S \) when \( R = 0 \) (whites), in which case the \( RS \) interaction does not enter the equation. The \( P\text{-value} of P < .01 \) for smoking represents the result of the test that the log odds ratio between \( Y \) and \( S \) for whites is 0.

28. The derivative equals \( \beta \exp(\alpha + \beta x)/[1 + \exp(\alpha + \beta x)] = \beta \pi(x)(1 - \pi(x)) \).

29. The odds ratio \( e^{\beta} \) is approximately equal to the relative risk when the probability is near 0 and the complement is near 1, since \( e^{\beta} = [\pi(x + 1)/(1 - \pi(x + 1))]/[\pi(x)/(1 - \pi(x))] \approx \pi(x + 1)/\pi(x) \).

30. \( \partial \pi(x)/\partial x = \beta \pi(x)(1 - \pi(x)) \), and \( \pi(1 - \pi) \leq .25 \) with equality at \( \pi = .5 \). For multiple explanatory variable case, the rate of change as \( x \) changes with other variables held constant is greatest when \( \pi = .5 \).

32. a. Since \( \log[\pi/(1 - \pi)] = \alpha + \log(d^{\beta}) \), exponentiating yields \( \pi/(1 - \pi) = e^\alpha e^{\log(d^{\beta})} = e^\alpha d^{\beta} \). Letting \( d = 1, e^\alpha \) equals the odds for the first draft pick.

b. As a function of \( d \), the odds decreases more quickly for pro basketball.

33. a. Let \( \rho = P(Y=1) \). By Bayes Theorem,

\[
P(Y = 1|x) = \rho \exp[-(x - \mu_1)^2/2\sigma^2]/\{\rho \exp[-(x - \mu_1)^2/2\sigma^2 + (1 - \rho) \exp[-(x - \mu_0)^2/2\sigma^2]\}
\]

\[
= 1/[1 + [(1 - \rho)/\rho] \exp[-(\mu_0^2 - \mu_1^2 + 2x(\mu_1 - \mu_0))/2\sigma^2]}
\]

\[
= 1/[1 + \exp[-(\alpha + \beta x)] = \exp(\alpha + \beta x)/[1 + \exp(\alpha + \beta x)],
\]

where \( \beta = (\mu_1 - \mu_0)/\sigma^2 \) and \( \alpha = -\log[(1 - \rho)/\rho] + [\mu_0^2 - \mu_1^2]/2\sigma^2 \).
36. When the explanatory variable takes only two values \((x = 0 \text{ or } 1)\), the model for \(P(Y = 1)\) simplifies to \(\exp(\alpha + \beta)/[1 + \exp(\alpha + \beta)]\) at \(x = 1\) and to \(\exp(\alpha)/[1 + \exp(\alpha)]\) at \(x = 0\). Note that the odds ratio is

\[
\frac{P(Y = 1|X = 1)/P(Y = 0|X = 1)}{P(Y = 1|X = 0)/P(Y = 0|X = 0)} = \exp(\beta),
\]

so \(\beta\) represents the log odds ratio. Let \(y_i\) denote the number of successes for the \(n_1\) trials when \(x = 1\), and let \(y_2\) denote the number of successes for the \(n_2\) trials when \(x = 0\). By directly taking the log of the likelihood (or by using (4.25) with \(I = 2\)), we get

\[
L = (y_1 + y_2) \alpha + (y_1) \beta - \{n_1 \log(1 + \exp(\alpha + \beta)) + n_2 \log(1 + \exp(\alpha))\}. 
\]

The likelihood equations are

\[
\frac{\partial L}{\partial \beta} = y_1 - n_1 \exp(\alpha + \beta)/[1 + \exp(\alpha + \beta)] = 0
\]

and

\[
\frac{\partial L}{\partial \alpha} = (y_1 + y_2) - n_1 \exp(\alpha + \beta)/[1 + \exp(\alpha + \beta)] - n_2 \exp(\alpha)/[1 + \exp(\alpha)] = 0.
\]

Subtracting gives \(y_2 = n_2 \exp(\alpha)/[1 + \exp(\alpha)]\), or \(\hat{\alpha} = \log[y_2/(n_2 - y_2)]\). The first equation gives \(\hat{\alpha} + \hat{\beta} = \log[y_1/(n_1 - y_1)]\), so that \(\hat{\beta} = \log([y_1/(n_1 - y_1)]/[y_2/(n_2 - y_2)])\), which is the sample log odds ratio.

Or, the likelihood equations are

\[
y_0 = n_0 e^\alpha/(1 + e^\alpha), \quad y_1 = n_1 e^{\alpha + \beta}/(1 + e^{\alpha + \beta}),
\]

equating sufficient statistics to their expected values. Solving the first gives \(\hat{\alpha} = \log(y_0/n_0)\). Solving the second gives \(\hat{\alpha} + \hat{\beta} = \log(y_1/n_1)\), from which \(\hat{\beta}\) is the log odds ratio.

37. d. When \(y_i\) is a 0 or 1, the log likelihood is \(\sum_i[y_i \log \hat{\pi}_i + (1 - y_i) \log(1 - \hat{\pi}_i)]\). For the saturated model, \(\hat{\pi}_i = y_i\), and the log likelihood equals 0. So, in terms of the ML fit and the ML estimates \(\hat{\pi}_i\) for this linear trend model, the deviance equals

\[
D = -2 \sum_i[y_i \log \hat{\pi}_i + (1 - y_i) \log(1 - \hat{\pi}_i)] = -2 \sum_i[y_i \log \left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) + \log(1 - \hat{\pi}_i)]
\]

\[
= -2 \sum_i[y_i (\hat{\alpha} + \hat{\beta} x_i) + \log(1 - \hat{\pi}_i)].
\]

For this model, the likelihood equations are \(\sum_i y_i = \sum_i \hat{\pi}_i\) and \(\sum_i x_i y_i = \sum_i x_i \hat{\pi}_i\). So, the deviance simplifies to

\[
D = -2(\hat{\alpha} \sum_i \hat{\pi}_i + \hat{\beta} \sum_i x_i \hat{\pi}_i + \sum_i \log(1 - \hat{\pi}_i)]
\]

\[
= -2\sum_i \hat{\pi}_i (\hat{\alpha} + \hat{\beta} x_i) + \sum_i \log(1 - \hat{\pi}_i])
\]

\[
= -2\sum_i \hat{\pi}_i \log \left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) - 2 \sum_i \log(1 - \hat{\pi}_i).
\]

38. a. The ratio of the test statistics equals the ratio of squared standard errors, or \(n\hat{\pi}(1 - \hat{\pi})/n(\hat{\pi})\), which is \(4\hat{\pi}(1 - \hat{\pi})/n\). If \(|\alpha|\) is large, then \(\hat{\pi}\) or \((1 - \hat{\pi})\) tends to be near 0 and the sample logit tends to be large, but the ratio of statistics tends to approach 0. That is, the Wald test tends to lose power compared to the test using null standard error.

b. Now the ratio of test statistics equals \(1/4\hat{\pi}(1 - \hat{\pi})\), and the reverse behavior occurs.

c. If \(y = 0\) or \(n\), then \(\hat{\pi} = 0\) or 1, and the estimated asymptotic variance is \(\infty\) in (a) and 0 in (b).

Chapter 6

5. a. The estimated odds of admission were 1.84 times higher for men than women. However, \(\hat{\theta}_{AG(D)} = .90\), so given department, the estimated odds of admission were .90 times as high for men as for women. Simpson’s paradox strikes again! Men applied relatively more often to Departments A and B, whereas women applied relatively more often to Departments C, D, E, F. At the same time, admissions rates were relatively high for Departments A and B and relatively low for C, D, E, F. These two effects combine to give a relative advantage to men for admissions when we study the marginal association.

c. The values of \(G^2\) are 2.68 for the model with no \(G\) effect and 2.56 for the model with \(G\) and \(D\) main
effects. For the latter model, CI for conditional AG odds ratio is (0.87, 1.22).

14. a. \( \pi(x) = \Phi(-2.3178 + .0878x) \) is the cdf of a normal distribution with mean 2.3178/.0878 = 26.4 and standard deviation 1/.0878 = 11.4
b. Estimated rate of change = .40 \( \bar{\beta} = .035 \).
c. At \( x = 14 \), \( \hat{\pi} = \Phi(-1.09) = .138 \); at \( x = 28 \), \( \hat{\pi} = \Phi(14) = .556 \). Thus, \( \hat{\pi} \) increases by .42 over the middle half of LI values, a considerable increase.

22. a. \( G^2(M_j \mid M_k) = G^2(M_j) - G^2(M_k) = [G^2(M_j) - G^2(M_k)] + [G^2(M_k) - G^2(M_k)] = G^2(M_j \mid M_k) + G^2(M_k \mid M_k) \), so \( G^2(M_j \mid M_k) \geq G^2(M_j \mid M_k) \).
b. \( P(G^2(M_j \mid M_k) > \chi^2(\alpha)) \leq P(G^2(M_j \mid M_k) > \chi^2(\alpha)) \), which converges to \( \alpha \) as \( n \to \infty \), since this is a fit statistic for testing a model that truly holds.

23. We consider the contribution to the \( X^2 \) statistic of its two components (corresponding to the two levels of the response) at level \( i \) of the explanatory variable. For simplicity, we use the notation of (4.21) but suppress the subscripts. Then, that contribution is \( (y - n\pi)^2/n\pi + [(n - y) - n(1 - \pi)]^2/n(1 - \pi) \), where the first component is (observed - fitted)/fitted for the “success” category and the second component is (observed - fitted)/fitted for the “failure” category. Combining terms gives \( (y - n\pi)^2/n\pi(1 - \pi) \), which is the square of the residual. Adding these chi-squared components therefore gives the sum of the squared residuals.

26. The latter test describes dependence by a single parameter (\( \beta \)), rather than \( I - 1 \) parameters (e.g., \( \mu_1 - \mu_1, \ldots, \mu_{I-1} - \mu_I \)). It is not more powerful, though, if there is actually a \( U \)-shaped pattern to the means, the linear model fitting very poorly.

28. The model for the underlying latent normal variable can be expressed in GLM form as \( E(y^* \mid x_i) = \alpha + \beta x_i \). That is, if \( x_i \uparrow \), then \( E(y^* \mid x_i) \uparrow \) when \( y^* \) is scaled to have standard deviation 1. So, \( \beta \) refers to the change in \( E(y^*) \) for a one-unit change in \( x_i \). The model for the actual observed response \( y_i \) is \( P(y_i = 1) = P(y_i^* > 0) = P(-\epsilon_i < \alpha + \beta x_i) = \Phi(\alpha + \beta x_i) \) (since \( -\epsilon_i \) has a standard normal distribution). This is the probit model. If \( y^* \) were linearly transformed to have arbitrary standard deviation \( \sigma \) (e.g., my multiplying all the possible \( y^* \) values by \( \sigma \)), the change would then translate to \( \beta \sigma \) in \( E(y^*) \). So, the probit parameter can be interpreted as the number of standard deviation change in an underlying continuous response for each one-unit change in \( x_i \). For the beetle data (discussed on p. 247 of the text), the probit fit slope has 19.7. If there were an underlying continuous normal rv that determined whether a beetle died, then the expected value of that rv increases by .197 standard deviations for a .01 increase in the log dose.

29 a. \( P(y = 1) = P(\alpha + \beta x_i + \epsilon_i > \alpha_0 + \beta_0 x_0 + \epsilon_0) = P((\epsilon_0 - \epsilon_1)/\sqrt{2} < ((\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)x)/\sqrt{2}) = \Phi(\alpha^* + \beta^* x) \) where \( \alpha^* = (\alpha_1 - \alpha_0)/\sqrt{2}, \beta^* = (\beta_1 - \beta_0)/\sqrt{2}. \)

30. a. \( \log[-\log(0.5)] = -.3665 = \alpha + \beta x \), from which \( x = -(.3665 + \alpha)/\beta. \)
b. Using (6.12), \( \partial \pi/\partial x \) can be expressed as \( -\beta(1 - \pi) \log(1 - \pi) \). Using standard calculus, the function \( -y \log(y) \) is maximized at \( y = e^{-1} \), so that \( -(1 - \pi) \log(1 - \pi) \) is maximized at \( \pi = 1 - e^{-1} \) (compared to \( \pi = .5 \) for symmetric links such as the logit and probit). From (6.12), \( \pi = 1 - e^{-1} \) when \( \pi = 1 - e^{-1} \), or \( \alpha + \beta x = 0, \) or \( x = -\alpha/\beta. \) For the log-log link, the answer for \( x \) is the same, but the value of \( \pi \) at that \( x \)-value is \( e^{-1}. \)

32. a. Log likelihood for probit model is

\[
\log \left\{ \prod_{i=1}^{N} \left[ \Phi\left( \sum_{j} \beta_{j}x_{ij} \right) \right]^{y_{i}} \left[ 1 - \Phi\left( \sum_{j} \beta_{j}x_{ij} \right) \right]^{1-y_{i}} \right\ } = \sum_{i=1}^{N} \left[ y_{i} \log \left[ \Phi\left( \sum_{j} \beta_{j}x_{ij} \right) \right] + (1-y_{i}) \log \left[ 1 - \Phi\left( \sum_{j} \beta_{j}x_{ij} \right) \right] \right]
\[\sum_i y_i \log \left[ \Phi\left( \sum_j \beta_j x_{ij} \right) \right] + \sum_i \log \left[ 1 - \Phi\left( \sum_j \beta_j x_{ij} \right) \right]\]

b. For probit model,
\[
\frac{\partial L}{\partial \beta_j} = \sum_i y_i \left[ \frac{1 - \Phi\left( \sum_j \beta_j x_{ij} \right)}{\Phi\left( \sum_j \beta_j x_{ij} \right)} \right] \phi\left( \sum_j \beta_j x_{ij} \right) x_{ij} \left( 1 - \Phi\left( \sum_j \beta_j x_{ij} \right) \right) + x_{ij} \phi\left( \sum_j \beta_j x_{ij} \right) \Phi\left( \sum_j \beta_j x_{ij} \right) \frac{\Phi\left( \sum_j \beta_j x_{ij} \right)}{\left( 1 - \Phi\left( \sum_j \beta_j x_{ij} \right) \right)^2} - \sum_i x_{ij} \phi\left( \sum_j \beta_j x_{ij} \right) = 0
\]
\[
\Rightarrow \sum_i \frac{y_i x_{ij} \phi\left( \sum_j \beta_j x_{ij} \right)}{1 - \Phi\left( \sum_j \beta_j x_{ij} \right)} \Phi\left( \sum_j \beta_j x_{ij} \right) - \sum_i \frac{x_{ij} \phi\left( \sum_j \beta_j x_{ij} \right)}{1 - \Phi\left( \sum_j \beta_j x_{ij} \right)} = 0
\]
\[
\Rightarrow \sum_i \frac{y_i x_{ij} \phi\left( \sum_j \beta_j x_{ij} \right)}{\hat{\pi}_i \left( 1 - \hat{\pi}_i \right)} - \sum_i \frac{x_{ij} \phi\left( \sum_j \beta_j x_{ij} \right)}{\hat{\pi}_i \left( 1 - \hat{\pi}_i \right)} = 0
\]
\[
\Rightarrow \sum_i \left( y_i - \hat{\pi}_i \right) x_{ij} z_i = 0, \text{ where } z_i = \phi(\sum_j \beta_j x_{ij}) / \hat{\pi}_i \left( 1 - \hat{\pi}_i \right).
\]

For logistic regression, from (4.26) with \( \{n_i = 1\} \), \( \sum_i \left( y_i - \hat{\pi}_i \right) x_{ij} = 0 \).

Chapter 7

1. a.
\[\log(\hat{\pi}_1 / \hat{\pi}_2) = (.883 + .758) + (.419 - .105)x_1 + (.342 - .271)x_2 = 1.641 + .314x + 1 + .071x_2.\]
b. The estimated odds for females are \(\exp(0.419) = 1.5\) times those for males, controlling for race; for whites, they are \(\exp(0.342) = 1.4\) times those for blacks, controlling for gender.
c. \(\hat{\pi}_1 = \exp(.883 + .419 + .342) / [1 + \exp(.883 + .419 + .342) + \exp(-.758 + .105 + .271)] = .76.\)
f. For each of 2 logits, there are 4 gender-race combinations and 3 parameters, so \(df = 2(4) - 2(3) = 2.\) The likelihood-ratio statistic of 7.2, based on \(df = 2\), has a \(P\)-value of .03 and shows evidence of a gender effect.

5. For any collapsing of the response, for Democrats the estimated odds of response in the liberal direction are \(\exp(975) = 2.65\) times the estimated odds for Republicans. The estimated probability of a very liberal response equals \(\exp(-2.469) / [1 + \exp(-2.469)] = .078\) for Republicans and \(\exp(-2.469 + .975) / [1 + \exp(-2.469 + .975)] = .183\) for Democrats.

7. a. Four intercepts are needed for five response categories. For males in urban areas wearing seat belts, all dummy variables equal 0 and the estimated cumulative probabilities are \(\exp(3.3074) / [1 + \exp(3.3074)] = .965, \exp(3.4818) / [1 + \exp(3.4818)] = .970, \exp(5.3494) / [1 + \exp(5.3494)] = .995, \exp(7.2563) / [1 + \exp(7.2563)] = .9993, \text{ and } 1.0.\) The corresponding response probabilities are \(.965, \.905, \.904, \.904, \text{ and } .9007.\)
b. Wald CI is \(\exp[-.5463 \pm 1.96(.0272)] = (\exp(-.600), \exp(-.493)) = (.549, .611).\) Give seat belt use and location, the estimated odds of injury below any fixed level for a female are between .549 and .611 times the estimated odds for a male.
c. Estimated odds ratio equals \(\exp(-.7602 - .1244) = .41\) in rural locations and \(\exp(-.7602) = .47\) in urban locations. The interaction effect -.1244 is the difference between the two log odds ratios.

9. a. Setting up dummy variables (1,0) for (male, female) and (1,0) for (sequential, alternating), we get treatment effect = -.581 (SE = 0.212) and gender effect = -.541 (SE = .295). The estimated odds
ratios are .56 and .58. The sequential therapy leads to a better response than the alternating therapy; the estimated odds of response with sequential therapy below any fixed level are .56 times the estimated odds with alternating therapy.

b. The main effects model fits well ($G^2 = 5.6$, $df = 7$), and adding an interaction term does not give an improved fit (The interaction model has $G^2 = 4.5$, $df = 6$).

29. No, because the baseline-category logit model refers to individual categories rather than cumulative probabilities. There is not linear structure for baseline-category logits that implies identical effects for each cumulative logit.

31. For $j < k$, logit[$P(Y \leq j \mid X = x_i)$] - logit[$P(Y \leq k \mid X = x_i)$] = $(\alpha_j - \alpha_k) + (\beta_j - \beta_k)x$. This difference of cumulative probabilities cannot be positive since $P(Y \leq j) \leq P(Y \leq k)$; however, if $\beta_j > \beta_k$ then the difference is positive for large $x$, and if $\beta_j > \beta_k$ then the difference is positive for small $x$.

Chapter 8

1. a. $G^2$ values are 2.38 ($df = 2$) for (GI, HI), and .30 ($df = 1$) for (GI, HI, GH).

b. Estimated log odds ratios is -.252 ($SE = .175$) for GH association, so CI for odds ratio is $\exp[-.252 \pm 1.96(.175)]$. Similarly, estimated log odds ratio is .464 ($SE = .241$) for GI association, leading to CI of $[\exp(.464 \pm 1.96(.241)]$. Since the intervals contain values rather far from 1.0, it is safest to use model (GH, GI, HI), even though simpler models fit adequately.

6. a. $G^2 = 31.7$, $df = 48$. The data are sparse, but the model seems to fit well. It is plausible that the association between any two items is the same at each combination of levels of the other two items.

b. Model with $A$ and $H$ much than for those who judge spending on

Substitute model formula, and simplify. The estimated odds ratio equals $\exp(2.38) = 8.5$. There is a strong positive association. Given responses on $C$ and $L$, the estimated odds of judging spending on $E$ to be too much instead of too little are 8.5 times as high for those who judge spending on $E$ to be too much than for those who judge spending on $H$ to be too low. The 95% CI is $[\exp(2.142 \pm 1.96(.523)]$, or (3.1, 24.4.). Though it is very wide, it is clear that the true association is strong.

c. 2.4 for $C$ and $L$, 6.5 for $H$ and $L$, .8 for $C$ and $H$, .9 for $E$ and $L$, 3.3 for $C$ and $E$. The associations seem to be strongest between $E$ and $H$ and between $H$ and $L$. In fact, the simpler model deleting the CH and EL terms fits well ($G^2 = 39.4$, $df = 56$).

9. a. (GRP, AG, AR, AP). Set $\beta^R_0 = 0$ in model in previous logit model.

b. Model with $A$ as response and additive factor effects for $R$ and $P$, logit($\pi$) = $\alpha + \beta^R_i + \beta^P_j$.

c. (i) (GRP, A), logit($\pi$) = $\alpha$, (ii) (GRP, AR), logit($\pi$) = $\alpha + \beta^R_i$, (iii) (GRP, APR, AG), add term of form $\beta^R_i\beta^P_j$ to logit model in Exercise 5.23.

14. a. For parameters defined using zero-sum constraints, $\lambda^Y_a - \lambda^Y_b = [\log \pi_{i+a} - (\sum \log \pi_{i+h})/J] - \log \pi_{i+b} - (\sum \log \pi_{i+h}/J) = \log(\pi_{i+a}/\pi_{i+b})$, so $\lambda^Y_a > \lambda^Y_b$ if $\pi_{i+a} > \pi_{i+b}$.

b. From (a) if all $\{\lambda^Y_i\}$ are equal, then so are all $\{\pi_{i+j}\}$, and hence each $\pi_{i+j} = 1/J$.

18. You can show, for instance, that $P(X = i, Y = j) = P(X = i)P(Y = j)$ for all $i$ and $j$, and similarly for $X$ and $Z$, and for $Y$ and $Z$. However, $P(X = yes, Y = yes, Z = no) = 0 \neq P(X = yes)P(Y = yes)P(Z = no) = 1/8$.

19. a. When $Y$ is jointly independent of $X$ and $Z$, $\pi_{ijk} = \pi_{ij+}\pi_{i+k}$. Dividing $\pi_{ijk}$ by $\pi_{i+k}$, we find that $P(X = i, Y = j | Z = k) = P(X = i | Z = k)P(Y = j)$. But when $\pi_{ijk} = \pi_{ij+}\pi_{i+k}$, $P(Y = j | Z = k) = \pi_{i+k}/\pi_{i+k} = \pi_{ij+k}/\pi_{i+k} = \pi_{i+j+k}/\pi_{i+k} = \pi_{i+j} = P(Y = j)$. Hence, $P(X = i, Y = j | Z = k) = P(X = i | Z = k)P(Y = j) = P(X = i | Z = k)P(Y = j | Z = k)$ and there is $XY$ conditional independence.

b. For mutual independence, $\pi_{ijk} = \pi_{i+k+}\pi_{i+k+}$. Summing both sides over $k$, $\pi_{ij+} = \pi_{i+k+}\pi_{i+j+}$, which is marginal independence in the $XY$ marginal table.

c. No. For instance, model ($Y, XZ$) satisfies this, but $X$ and $Z$ are dependent (the conditional association being the same as the marginal association in each case, for this model).
d. When \( X \) and \( Y \) are conditionally independent, then an odds ratio relating them using two levels of each variable equals 1.0 at each level of \( Z \). Since the odds ratios are identical, there is no three-factor interaction.

21. Use the definitions of the models, in terms of cell probabilities as functions of marginal probabilities. When one specifies sufficient marginal probabilities that have the required one-way marginal probabilities of 1/2 each, these specified marginal distributions then determine the joint distribution. Model \((XY, XZ, YZ)\) is not defined in the same way; for it, one needs to determine cell probabilities for which each set of partial odds ratios do not equal 1.0 but are the same at each level of the third variable.

   a. 
   \[
   \begin{array}{cc|cc}
   Y & \empty & Y & \empty \\
   \empty .125 & .125 & \empty .125 & .125 \\
   X & .125 & .125 & .125 & .125 \\
   Z = 1 & \empty & \empty & \empty & \empty \\
   Z = 2 & \empty & \empty & \empty & \empty \\
   \end{array}
   \]
   This is actually a special case of \((X,Y,Z)\) called the **equiprobability model**.

   b. 
   \[
   \begin{array}{cc|cc}
   \empty & .15 & .10 & \empty \\
   \empty .10 & .15 & \empty .10 & .15 \\
   \end{array}
   \]

c. 
\[
\begin{array}{cc|cc}
1/4 & 1/24 & 1/12 & 1/8 \\
1/8 & 1/12 & 1/24 & 1/4 \\
\end{array}
\]

d. 
\[
\begin{array}{cc|cc}
2/16 & 1/16 & 4/16 & 1/16 \\
1/16 & 4/16 & 1/16 & 2/16 \\
\end{array}
\]

e. Any \(2 \times 2 \times 2\) table

29. For this model, in a given row the \( J \) cell probabilities are equal. The likelihood equations are \( \hat{\mu}_{i+} = n_{i+} \) for all \( i \). The fitted values that satisfy the model and the likelihood equations are \( \hat{\mu}_{ij} = n_{i+}/J \).

31. For model \((XY, Z)\), log likelihood is
\[
L = n\lambda + \sum_i n_{i+} \lambda_i^X + \sum_j n_{+j} \lambda_j^Y + \sum_k n_{++k} \lambda_k^Z + \sum_i \sum_j n_{ij+} \lambda_{ij}^{XY} - \sum_i \sum_j \sum_k \mu_{ijk}
\]

The minimal sufficient statistics are \(\{n_{i+}\}, \{n_{++k}\}\). Differentiating with respect to \(\lambda_{ij}^{XY}\) and \(\lambda_k^Z\) gives the likelihood equations \(\hat{\mu}_{i+} = n_{i+} \) and \(\hat{\mu}_{++k} = n_{++k} \) for all \( i, j \), and \( k \). For this model, since \(\pi_{ijk} = \pi_{ij} \pi_{+jk} \pi_{++k} \), \(\hat{\mu}_{ijk} = \hat{\mu}_{ij} \hat{\mu}_{+jk} / n = n_{ij} n_{++k} / n \). Residual \(df = IJK - [1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1)] = (IJ - 1)(K - 1)\).

32. Note that \(\{\hat{\mu}_{ijk} = n_{i+j+k+n_{i+jk}} / n_{++k}, i = 1, ..., I, j = 1, ..., J\} \) are fitted values for the usual test of independence between \(X\) and \(Y\), when the counts at a fixed level \(k\) of \(Z\) are viewed as a single two-way table. Thus the contribution that partial table makes toward \(X^2\) for testing model \((XZ,YZ)\) is identical to \(X^2\) that would be computed if independence were tested for that table alone.

38. For \((X,Y,Z)\), start with initial values of 1.0. Then if we first fit the one-way \(X\) margin, at the end of the first cycle the estimates are \(\{n_{i+}/JK\}\), which match the observed data in the \(X\) margin. In the next cycle we fit the \(YZ\) margin, and obtain \(n_{i+++j+k}/n\). All successive cycles give the same value, which is the ML estimate.

Chapter 9
18. a. CI for log rate is $2.549 \pm 1.96(0.0495)$, so CI for rate is $(11.7, 14.0)$.
b. Using log link, the log of the accident rate equals $\alpha$ for females, and it equals $\alpha + \beta$ for males. The rates are identical if $\beta = 0$. The estimate of $\alpha$ in the model is simply the sample log(rate) for females, namely $\log(10.12) = 2.31$; the estimate of $\alpha + \beta$ is the sample log(rate) for males, namely $\log(14.95) = 2.70$. The estimated difference is $\hat{\beta} = 0.39$. The estimated accident rate for men was $\exp(\hat{\beta}) = \exp(0.39) = 1.48$ times the rate for women. That is, $14.95/10.12 = 1.48$, the sample rate being 48% higher for men. For $H_0: \beta = 0$, GLM software reports a $SE$ for $\hat{\beta} = 0.39$ of 0.09, so there is strong evidence that the accident rate was higher for males (i.e., that $\beta > 0$).

24. No, whenever the conditional odds ratios are identical in each stratum, there is no three-factor interaction. No, $Z$ is not conditionally independent of $X$ or of $Y$, since if it were, then collapsibility conditions would be satisfied and the $XY$ marginal odds ratio would equal the $XY$ conditional odds ratios.

25. $W$ and $Z$ are separated using $X$ alone or $Y$ alone or $X$ and $Y$ together. $W$ and $Y$ are conditionally independent given $X$ and $Z$ (as the model symbol implies) or conditional on $X$ alone since $X$ separates $W$ and $Y$. $X$ and $Z$ are conditionally independent given $W$ and $Y$ or given only $Y$ alone.

26. a.

$$
\mu_{ij+} = \sum_k \exp(\lambda_i X + \lambda_j Y + \lambda_{ik} Z + \lambda_{ij} X Y + \lambda_{ik} X Z)
$$

$$
= \exp(\lambda_i X + \lambda_j Y + \lambda_{ij} X Y) \sum_k \exp(\lambda_{ik} Z + \lambda_{ik} X Z)
$$

which implies that $\log \mu_{ij+} = \lambda_i X + \lambda_j Y + \lambda_{ij} X Y + \xi_i$, where $\xi_i = \log(\sum_k \exp(\lambda_{ik} Z + \lambda_{ik} X Z))$ and can be combined with the $\lambda_i X$ term to give the main effect for $X$. Hence, $\{\lambda_{ij} X Y\}$ are the same in the two-way and three-way tables. Since the $XY$ conditional odds ratios are functions of $\{\lambda_{ij} X Y\}$, they are the same in both marginal and partial tables.

b. The term summed over $k$ involves both $i$ and $j$, so it gets absorbed with the term $\lambda_{ij} X Y$ to yield a new term involving both $i$ and $j$ for the $XY$ marginal association.

47. Suppose ML estimates did exist, and let $c = \hat{\mu}_{111}$. Then $c > 0$, since we must be able to evaluate the logarithm for all fitted values. But then $\hat{\mu}_{112} = n_{112} - c$, since likelihood equations for the model imply that $\hat{\mu}_{111} + \hat{\mu}_{112} = n_{111} + n_{112}$ (i.e., $\hat{\mu}_{111} = n_{111} - c$). Using similar arguments for other two-way margins implies that $\hat{\mu}_{122} = n_{122} + c$, $\hat{\mu}_{212} = n_{212} + c$, and $\hat{\mu}_{222} = n_{222} - c$. But since $n_{222} = 0$, $\hat{\mu}_{222} = -c < 0$, which is impossible. Thus we have a contradiction, and it follows that ML estimates cannot exist for this model.

Chapter 10

1. a. Sample marginal proportions are $1300/1825 = 0.712$ and $1187/1825 = 0.650$. The difference of .062 has an estimated variance of $[(90 + 203)/1825 - (90 - 203)^2/1825^2]/1825 = .000086$, for $SE = .0093$. The $95\%$ Wald CI is .062 $\pm$1.96(.0093), or .062 $\pm$ .018, or (.044, .080).
b. McNemar chi-squared = $(203 - 90)^2/(203 + 90) = 43.6$, $df = 1$, $P < .0001$; there is strong evidence of a higher proportion of ‘yes’ responses for ‘let patient die.’
c. $\hat{\beta} = \log(203/90) = \log(2.26) = 0.81$. For a given respondent, the odds of a ‘yes’ response for ‘let patient die’ are estimated to equal 2.26 times the odds of a ‘yes’ response for ‘suicide.’

21. The matched-pairs $t$ test compares means for dependent samples, and McNemar’s test compares proportions for dependent samples. The $t$ test is valid for interval-scale data (with normally-distributed differences, for small samples) whereas McNemar’s test is valid for binary data.

23. a. This is a conditional odds ratio, conditional on the subject, but the other model is a marginal model so its odds ratio is not conditional on the subject.
b. This is simply the mean of the expected values of the individual binary observations.
d. In the three-way representation, note that each partial table has one observation in each row. If each response in a partial table is identical, then each cross-product that contributes to the M-H estimator equals 0, so that table makes no contribution to the statistic. Otherwise, there is a contribution of 1 to the numerator or the denominator, depending on whether the first observation is a success and the second a failure, or the reverse. The overall estimator then is the ratio of the numbers of such pairs, or in terms of the original 2×2 table, this is $n_{12}/n_{21}$.

Chapter 14

5. Using the delta method, the asymptotic variance is $(1 - 2\pi)^2/4n$. This vanishes when $\pi = 1/2$, and the convergence of the estimated standard deviation to the true value is then faster than the usual rate.

7a. By the delta method with the square root function, $\sqrt{n}[\sqrt{T_n/n} - \sqrt{\mu}]$ is asymptotically normal with mean 0 and variance $(1/2\sqrt{\mu})^2(\mu)$, or in other words $\sqrt{T_n} - \sqrt{n\mu}$ is asymptotically N(0, 1/4).

b. If $g(p) = \arcsin(\sqrt{p})$, then $g'(p) = (1/\sqrt{1-p})(1/2\sqrt{p}) = 1/2\sqrt{p(1-p)}$, and the result follows using the delta method. Ordinary least squares assumes constant variance.