

CATEGORICAL DATA ANALYSIS

Solutions to Selected Odd-Numbered Problems

Alan Agresti

Version March 15, 2006, ©Alan Agresti 2006

This manual contains solutions and hints to solutions for many of the odd-numbered exercises in *Categorical Data Analysis*, second edition, by Alan Agresti (John Wiley, & Sons, 2002).

Please report errors in these solutions to the author (Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, e-mail AA@STAT.UFL.EDU), so they can be corrected in future revisions of this site. The author regrets that he cannot provide students with more detailed solutions or with solutions of other problems not in this file.

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 $\text{bin}(n, \pi)$. $\text{Var}(Y) = E[\text{Var}(Y | \pi)] + \text{Var}[E(Y | \pi)] > E[\text{Var}(Y | \pi)] = E[n\pi(1 - \pi)]$.

5. $\hat{\pi} = 842/1824 = .462$, so $z = (.462 - .5)/\sqrt{.5(.5)/1824} = -3.28$, for which $P = .001$ for $H_a: \pi \neq .5$. The 95% Wald CI is $.462 \pm 1.96\sqrt{.462(.538)/1824} = .462 \pm .023$, or (.439, .485). The 95% score CI is also (.439, .485).

7. a. $\ell(\pi) = \pi^{20}$, so $\hat{\pi} = 1.0$.

b. Wald statistic $z = (1.0 - .5)/\sqrt{1.0(0)/20} = \infty$. Wald CI is $1.0 \pm 1.96\sqrt{1.0(0)/20} = 1.0 \pm 0.0$, or (1.0, 1.0).

c. $z = (1.0 - .5)/\sqrt{.5(.5)/20} = 4.47$, $P < .0001$. Score CI is (0.839, 1.000).

d. Test statistic $2(20)\log(20/10) = 27.7$, $df = 1$. From problem 1.25a, the CI is $(\exp(-1.96^2/40), 1) = (0.908, 1.0)$.

e. P -value = $2(.5)^{20} = .00000191$. Clopper-Pearson CI is (0.832, 1.000). CI using Blaker method is (0.840, 1.000).

f. $n = 1.96^2(.9)(.1)/(.05)^2 = 138$.

9. The sample mean is 0.61. Fitted probabilities for the truncated distribution are 0.543, 0.332, 0.102, 0.021, 0.003. The estimated expected frequencies are 108.5, 66.4, 20.3, 4.1, and 0.6, and the Pearson $X^2 = 0.7$ with $df = 3$ (0.3 with $df = 2$ if one truncates at 3 and above).

11. $\text{Var}(\hat{\pi}) = \pi(1 - \pi)/n$ decreases as π moves toward 0 or 1 from 0.5.
13. This is the binomial probability of y successes and $k - 1$ failures in $y + k - 1$ trials times the probability of a failure at the next trial.
15. For binomial, $m(t) = E(e^{tY}) = \sum_y \binom{n}{y} (\pi e^t)^y (1 - \pi)^{n-y} = (1 - \pi + \pi e^t)^n$, so $m'(0) = n\pi$.
17. a. $\ell(\mu) = \exp(-n\mu) \mu^{\sum y_i}$, so $L(\mu) = -n\mu + (\sum y_i) \log(\mu)$ and $L'(\mu) = -n + (\sum y_i)/\mu = 0$ yields $\hat{\mu} = (\sum y_i)/n$.
- b. (i) $z_w = (\bar{y} - \mu_0)/\sqrt{\bar{y}/n}$, (ii) $z_s = (\bar{y} - \mu_0)/\sqrt{\mu_0/n}$, (iii) $-2[-n\mu_0 + (\sum y_i) \log(\mu_0) + n\bar{y} - (\sum y_i) \log(\bar{y})]$.
- c. (i) $\bar{y} \pm z_{\alpha/2} \sqrt{\bar{y}/n}$, (ii) all μ_0 such that $|z_s| \leq z_{\alpha/2}$, (iii) all μ_0 such that LR statistic $\leq \chi_1^2(\alpha)$.
19. a. No outcome can give $P \leq .05$, and hence one never rejects H_0 .
- b. When $T = 2$, mid P -value = .04 and one rejects H_0 . Thus, $P(\text{Type I error}) = P(T = 2) = .08$.
- c. P -values of the two tests are .04 and .02; $P(\text{Type I error}) = P(T = 2) = .04$ with both tests.
- d. $P(\text{Type I error}) = E[P(\text{Type I error} | T)] = (5/8)(.08) = .05$.
21. a. With the binomial test the smallest possible P -value, from $y = 0$ or $y = 5$, is $2(1/2)^5 = 1/16$. Since this exceeds .05, it is impossible to reject H_0 , and thus $P(\text{Type I error}) = 0$. With the large-sample score test, $y = 0$ and $y = 5$ are the only outcomes to give $P \leq .05$ (e.g., with $y = 5$, $z = (1.0 - .5)/\sqrt{.5(.5)/5} = 2.24$ and $P = .025$). Thus, for that test, $P(\text{Type I error}) = P(Y = 0) + P(Y = 5) = 1/16$.
- b. For every possible outcome the Clopper-Pearson CI contains .5. e.g., when $y = 5$, the CI is (.478, 1.0), since for $\pi_0 = .478$ the binomial probability of $y = 5$ is $.478^5 = .025$.
23. For π just below $.18/n$, $P(\text{CI contains } \pi) = P(Y = 0) = (1 - \pi)^n = (1 - .18/n)^n \approx \exp(-.18) = 0.84$.
25. a. The likelihood-ratio (LR) CI is the set of π_0 for testing $H_0: \pi = \pi_0$ such that LR statistic = $-2 \log[(1 - \pi_0)^n / (1 - \hat{\pi})^n] \leq z_{\alpha/2}^2$, with $\hat{\pi} = 0.0$. Solving for π_0 , $n \log(1 - \pi_0) \geq -z_{\alpha/2}^2/2$, or $(1 - \pi_0) \geq \exp(-z_{\alpha/2}^2/2n)$, or $\pi_0 \leq 1 - \exp(-z_{\alpha/2}^2/2n)$. Using $\exp(x) = 1 + x + \dots$ for small x , the upper bound is roughly $1 - (1 - z_{.025}^2/2n) = z_{.025}^2/2n = 1.96^2/2n \approx 2^2/2n = 2/n$.
- b. Solve for $(0 - \pi)/\sqrt{\pi(1 - \pi)/n} = -z_{\alpha/2}$.
- c. Upper endpoint is solution to $\pi_0^0(1 - \pi_0)^n = \alpha/2$, or $(1 - \pi_0) = (\alpha/2)^{1/n}$, or $\pi_0 = 1 - (\alpha/2)^{1/n}$. Using the expansion $\exp(x) \approx 1 + x$ for x close to 0, $(\alpha/2)^{1/n} = \exp\{\log[(\alpha/2)^{1/n}]\} \approx 1 + \log[(\alpha/2)^{1/n}]$, so the upper endpoint is $\approx 1 - \{1 + \log[(\alpha/2)^{1/n}]\} = -\log(\alpha/2)^{1/n} = -\log(.025)/n = 3.69/n$.
- d. The mid P -value when $y = 0$ is half the probability of that outcome, so the upper bound for this CI sets $(1/2)\pi_0^0(1 - \pi_0)^n = \alpha/2$, or $\pi_0 = 1 - \alpha^{1/n}$.
29. The right-tail mid P -value equals $P(T > t_o) + (1/2)p(t_o) = 1 - P(T \leq t_o) + (1/2)p(t_o) = 1 - F_{mid}(t_o)$.

31. Since $\partial^2 L / \partial \pi^2 = -(2n_{11} / \pi^2) - n_{12} / \pi^2 - n_{12} / (1 - \pi)^2 - n_{22} / (1 - \pi)^2$, the information is its negative expected value, which is $2n\pi^2 / \pi^2 + n\pi(1 - \pi) / \pi^2 + n\pi(1 - \pi) / (1 - \pi)^2 + n(1 - \pi) / (1 - \pi)^2$, which simplifies to $n(1 + \pi) / \pi(1 - \pi)$. The asymptotic standard error is the square root of the inverse information, or $\sqrt{\pi(1 - \pi) / n(1 + \pi)}$.

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 π_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 $(\hat{\pi} - \pi_0)^2 / [\pi_0(1 - \pi_0) / n] = z_S^2$.

35. If Y_1 is χ^2 with $df = \nu_1$ and if Y_2 is independent χ^2 with $df = \nu_2$, then the *mgf* of $Y_1 + Y_2$ is the product of the *mgfs*, which is $m(t) = (1 - 2t)^{-(\nu_1 + \nu_2)/2}$, which is the *mgf* of a χ^2 with $df = \nu_1 + \nu_2$.

Chapter 2

1. $P(-|C) = 1/4$. It is unclear from the wording, but presumably this means that $P(\bar{C}|+) = 2/3$. Sensitivity = $P(+|C) = 1 - P(-|C) = 3/4$. Specificity = $P(-|\bar{C}) = 1 - P(+|\bar{C})$ can't be determined from information given.

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.

5. Relative risks are 3.3, 5.4, 11.5, 34.7; e.g., 1994 probability of gun-related death in U.S. was 34.7 times that in England and Wales.

7. a. .0012, 10.78; relative risk, since difference of proportions makes it appear there is no association.

b. $(.001304 / .998696) / (.000121 / .999879) = 10.79$; this happens when the proportion in the first category is close to zero.

9. X given Y . Applying Bayes theorem, $P(V = w | M = w) = P(M = w | V = w)P(V = w) / [P(M = w | V = w)P(V = w) + P(M = w | V = b)P(V = b)] = .83 P(V=w) / [.83 P(V=w) + .06 P(V=b)]$. We need to know the relative numbers of victims who were white and black. Odds ratio = $(.94 / .06) / (.17 / .83) = 76.5$.

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.

15. The age distribution is relatively higher in Maine.

17. The odds of carcinoma for the various smoking levels satisfy:

$$\frac{(\text{Odds for high smokers})/(\text{Odds for low smokers})}{(\text{Odds for high smokers})/(\text{Odds for nonsmokers})} = \frac{(\text{Odds for high smokers})/(\text{Odds for nonsmokers})}{(\text{Odds for low smokers})/(\text{Odds for nonsmokers})} = 26.1/11.7 = 2.2.$$

19. $\gamma = .360$ ($C = 1508$, $D = 709$); of the untied pairs, the difference between the proportion of concordant pairs and the proportion of discordant pairs equals $.360$. There is a tendency for wife's rating to be higher when husband's rating is higher.

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

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

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

		Test		
		+	-	Total
Reality	+	.00475	.00025	.005
	-	.04975	.94525	.995

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 $P(D | \bar{E})$).

b. Use Bayes Theorem and result that $RR = P(D | E)/P(D | \bar{E})$.

25. Suppose $\pi_1 > \pi_2$. Then, $1 - \pi_1 < 1 - \pi_2$, and $\theta = [\pi_1/(1 - \pi_1)]/[\pi_2/(1 - \pi_2)] > \pi_1/\pi_2 > 1$. If $\pi_1 < \pi_2$, then $1 - \pi_1 > 1 - \pi_2$, and $\theta = [\pi_1/(1 - \pi_1)]/[\pi_2/(1 - \pi_2)] < \pi_1/\pi_2 < 1$.

27. This simply states that ordinary independence for a two-way table holds in each partial table.

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.

33. This condition is equivalent to the conditional distributions of Y in the first $I - 1$ rows being identical to the one in row I . Equality of the I conditional distributions is equivalent to independence.

37. a. Note that ties on X and Y are counted both in T_X and T_Y , and so T_{XY} must be subtracted. $T_X = \sum_i n_{i+}(n_{i+}-1)/2$, $T_Y = \sum_j n_{+j}(n_{+j}-1)/2$, $T_{XY} = \sum_i \sum_j n_{ij}(n_{ij}-1)/2$.
 c. The denominator is the number of pairs that are untied on X .

39. If in each row the maximum probability falls in the same column, say column 1, then $E[V(Y | X)] = \sum_i \pi_{i+}(1 - \pi_{1|i}) = 1 - \pi_{+1} = 1 - \max\{\pi_{+j}\}$, so $\lambda = 0$. Since the maximum being the same in each row does not imply independence, $\lambda = 0$ can occur even when the variables are not independent.

Chapter 3

3. $X^2 = 0.27$, $G^2 = 0.29$, P-value about 0.6. The free throws are plausibly independent. Sample odds ratio is 0.77, and 95% CI for true odds ratio is (0.29, 2.07), quite wide.

5. The values $X^2 = 7.01$ and $G^2 = 7.00$ ($df = 2$) show considerable evidence against the hypothesis of independence (P -value = .03). The standardized Pearson residuals show that the number of female Democrats and Male Republicans is significantly greater than expected under independence, and the number of female Republicans and Male Democrats is significantly less than expected under independence. e.g., there were 279 female Democrats, the estimated expected frequency under independence is 261.4, and the difference between the observed count and fitted value is 2.23 standard errors.

7. $G^2 = 27.59$, $df = 2$, so $P < .001$. For first two columns, $G^2 = 2.22$ ($df = 1$), for those columns combined and compared to column three, $G^2 = 25.37$ ($df = 1$). The main evidence of association relates to whether one suffered a heart attack.

9. b. Compare rows 1 and 2 ($G^2 = .76$, $df = 1$, no evidence of difference), rows 3 and 4 ($G^2 = .02$, $df = 1$, no evidence of difference), and the 3×2 table consisting of rows 1 and 2 combined, rows 3 and 4 combined, and row 5 ($G^2 = 95.74$, $df = 2$, strong evidences of differences).

11.a. $X^2 = 8.9$, $df = 6$, $P = 0.18$; test treats variables as nominal and ignores the information on the ordering.

b. Residuals suggest tendency for aspirations to be higher when family income is higher.

c. Ordinal test gives $M^2 = 4.75$, $df = 1$, $P = .03$, and much stronger evidence of an association.

13. a. It is plausible that control of cancer is independent of treatment used. (i) P -value is hypergeometric probability $P(n_{11} = 21 \text{ or } 22 \text{ or } 23) = .3808$, (ii) P -value = 0.638 is sum of probabilities that are no greater than the probability (.2755) 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. $.3808 - .5(.2755) = .243$. 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.

15. a. $(0.0, \infty)$, b. $(.618, \infty)$,

17. $P = 0.164$, $P = 0.0035$ takes into account the positive linear trend information in the sample.

21. For proportions π and $1 - \pi$ in the two categories for a given sample, the contribution to the asymptotic variance is $[1/n\pi + 1/n(1 - \pi)]$. The derivative of this with respect to π is $1/n(1 - \pi)^2 - 1/n\pi^2$, which is less than 0 for $\pi < 0.5$ and greater than 0 for $\pi > 0.5$. Thus, the minimum is with proportions $(.5, .5)$ in the two categories.

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

$$L = n_{11} \log(\theta^2) + (n_{12} + n_{21}) \log[\theta(1 - \theta)] + n_{22} \log(1 - \theta)^2$$

$\partial L / \partial \theta = 2n_{11}/\theta + (n_{12} + n_{21})/\theta - (n_{12} + n_{21})/(1 - \theta) - 2n_{22}/(1 - \theta) = 0$ gives $\hat{\theta} = (2n_{11} + n_{12} + n_{21})/2(n_{11} + n_{12} + n_{21} + n_{22}) = (n_{1+} + n_{+1})/2n = (p_{1+} + p_{+1})/2$.

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.

33. By expanding the square and simplifying, one can obtain the alternative formula for X^2 ,

$$X^2 = n \left[\sum_i \sum_j (n_{ij}^2 / n_{i+} n_{+j}) - 1 \right].$$

Since $n_{ij} \leq n_{i+}$, the double sum term cannot exceed $\sum_i \sum_j n_{ij} / n_{+j} = J$, and since $n_{ij} \leq n_{+j}$, the double sum cannot exceed $\sum_i \sum_j n_{ij} / n_{i+} = I$. It follows that X^2 cannot exceed $n[\min(I, J) - 1] = n[\min(I - 1, J - 1)]$.

35. Because G^2 for full table = G^2 for collapsed table + G^2 for table consisting of the two rows that are combined.

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 π , $P(X^2 \geq 6) = \sum_k P(X^2 \geq 6 \text{ and } n_{+1} = k) = \sum_k P(X^2 \geq 6 \mid n_{+1} = k)P(n_{+1} = k)$, and $P(X^2 \geq 6 \mid n_{+1} = k)$ is the P -value for Fisher’s exact test.

45. $P(|\hat{P} - P_o| \leq B) = P(|\hat{P} - P_o| / \sqrt{P_o(1 - P_o)/M} \leq B / \sqrt{P_o(1 - P_o)/M})$. By the approximate normality of \hat{P} , this is approximately $1 - \alpha$ if $B / \sqrt{P_o(1 - P_o)/M} = z_{\alpha/2}$.

Solving for M gives the result.

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.
3. The estimated probability of malformation increases from .0011 at $x = 0$ to $.0025 + .0011(7.0) = .0102$ at $x = 7$. The relative risk is $.0102/.0011 = 9.3$.
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. $\text{logit}(\hat{\pi}) = -3.695 + 1.815(\text{weight})$; at 5.2 kg, predicted logit = 5.74, and $\log(.9968/.0032) = 5.74$.
 d. $\text{probit}(\hat{\pi}) = -2.238 + 1.099(\text{weight})$. $\hat{\pi} = \Phi^{-1}(-2.238 + 1.099(5.2)) = \Phi^{-1}(3.48) = .9997$
7. a. $\exp[-.4288 + .5893(2.44)] = 2.74$.
 b. $.5893 \pm 1.96(.0650) = (.4619, .7167)$.
 c. $(.5893/.0650)^2 = 82.15$.
 d. Need log likelihood value when $\beta = 0$.
 e. Multiply standard errors by $\sqrt{535.896/171} = 1.77$. There is still very strong evidence of a positive weight effect.
9. a. $\log(\hat{\mu}) = -.502 + .546(\text{weight}) + .452c_1 + .247c_2 + .002c_3$, where c_1, c_2, c_3 are dummy variables for the first three color levels.
 b. (i) 3.6 (ii) 2.3. c. Test statistic = 9.1, $df = 3$, $P = .03$.
 d. Using scores 1,2,3,4, $\log(\hat{\mu}) = .089 + .546(\text{weight}) - .173(\text{color})$; predicted values are 3.5, 2.1, and likelihood-ratio statistic for testing color equals 8.1, $df = 1$. Using the ordinality of color yields stronger evidence of an effect, whereby darker crabs tend to have fewer satellites. Compared to the more complex model in (a), likelihood-ratio stat. = 1.0, $df = 2$, so the simpler model does not give a significantly poorer fit.
11. Since $\exp(.192) = 1.21$, a 1 cm increase in width corresponds to an estimated increase of 21% in the expected number of satellites. For estimated mean $\hat{\mu}$, the estimated variance is $\hat{\mu} + 1.11\hat{\mu}^2$, considerably larger than Poisson variance unless $\hat{\mu}$ is very small. The relatively small SE for \hat{k}^{-1} gives strong evidence that this model fits better than the Poisson model and that it is necessary to allow for overdispersion. The much larger SE of $\hat{\beta}$ in this model also reflects the overdispersion.
13. a. $\hat{\alpha} = .456$ ($SE = .029$). O'Neal's estimated probability of making a free throw is .456, and a 95% confidence interval is (.40, .51). However, $X^2 = 35.5$ ($df = 22$) provides evidence of lack of fit ($P = .034$). The exact test using X^2 gives $P = .028$. Thus, there is evidence of lack of fit.
 b. Using quasi likelihood, $\sqrt{X^2/df} = 1.27$, so adjusted $SE = .037$ and adjusted interval

is (.38, .53), reflecting slight overdispersion.

15. Model with main effects and no interaction has fit $\log(\hat{\mu}) = 1.72 + .59x - .23z$. This shows some tendency for a lower rate of imperfections at the high thickness level ($z = 1$), though the std. error of $-.23$ equals $.17$ so it is not significant. Adding an interaction (cross-product) term does not provide a significantly better fit, as the coefficient of the cross product of $.27$ has a std. error of $.36$.

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, $\log[\pi(x)] = \alpha + \beta x$. Since $\log[\pi(x+1)] - \log[\pi(x)] = \beta$, 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 $\alpha + \beta x < 0$.

23. For $j = 1$, $x_{ij} = 0$ for group B, and for observations in group A, $\partial\mu_A/\partial\eta_i$ is constant, so likelihood equation sets $\sum_A (y_i - \mu_A)/\mu_A = 0$, so $\hat{\mu}_A = \bar{y}_A$. For $j = 0$, $x_{ij} = 1$ and the likelihood equation gives

$$\sum_A \frac{(y_i - \mu_A)}{\mu_A} \left(\frac{\partial\mu_A}{\partial\eta_i} \right) + \sum_B \frac{(y_i - \mu_B)}{\mu_B} \left(\frac{\partial\mu_B}{\partial\eta_i} \right) = 0.$$

The first sum is 0 from the first likelihood equation, and for observations in group B, $\partial\mu_B/\partial\eta_i$ is constant, so second sum sets $\sum_B (y_i - \mu_B)/\mu_B = 0$, so $\hat{\mu}_B = \bar{y}_B$.

25. Letting $\phi = \Phi'$, $w_i = [\phi(\sum_j \beta_j x_{ij})]^2 / [\Phi(\sum_j \beta_j x_{ij})(1 - \Phi(\sum_j \beta_j x_{ij})) / n_i]$

27. a. With identity link the GLM likelihood equations simplify to, for each i , $\sum_{j=1}^{n_i} (y_{ij} - \mu_i)/\mu_i = 0$, from which $\hat{\mu}_i = \sum_j y_{ij}/n_i$.

b. Deviance = $2 \sum_i \sum_j [y_{ij} \log(y_{ij}/\bar{y}_i)]$.

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

b. The derivative of Φ 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. $\Phi(\alpha + \beta x) = \Phi\left(\frac{x - (-\alpha/\beta)}{1/\beta}\right)$.

35. For log likelihood $L(\mu) = -n\mu + (\sum_i y_i) \log(\mu)$, the score is $u = (\sum_i y_i - n\mu)/\mu$, $H = -(\sum_i y_i)/\mu^2$, and the information is n/μ . It follows that the adjustment to $\mu^{(t)}$ in Fisher scoring is $[\mu^{(t)}/n][(\sum_i y_i - n\mu^{(t)})/\mu^{(t)}] = \bar{y} - \mu^{(t)}$, and hence $\mu^{(t+1)} = \bar{y}$. For Newton-Raphson, the adjustment to $\mu^{(t)}$ is $\mu^{(t)} - (\mu^{(t)})^2/\bar{y}$, so that $\mu^{(t+1)} = 2\mu^{(t)} - (\mu^{(t)})^2/\bar{y}$. Note

that if $\mu^{(t)} = \bar{y}$, then also $\mu^{(t+1)} = \bar{y}$.

37. $\partial\eta_i/\partial\mu_i = v(\mu_i)^{-1/2}$, so $w_i = (\partial\mu_i/\partial\eta_i)^2/\text{Var}(Y_i) = v(\mu_i)/v(\mu_i = 1)$. For the Poisson, $v(\mu_i) = \mu_i$, so $g'(\mu_i) = \mu_i^{-1/2}$, so $g(\mu_i) = 2\sqrt{\mu}$.

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 $\hat{\beta}\hat{\pi}(1 - \hat{\pi}) = .1449(.068)(.932) = .009$.
 - e. $e^{\hat{\beta}} = e^{.1449} = 1.16$.
 - 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 $= (.1449/.0593)^2 = 5.96$, $df = 1$, P -value $= .0146$ for $H_a:\beta \neq 0$.
 - h. Likelihood-ratio statistic $= 34.37 - 26.07 = 8.30$, $df = 1$, P -value $= .004$.
3. $\text{logit}(\hat{\pi}) = -3.866 + 0.397(\text{snoring})$. Fitted probabilities are .021, .044, .093, .132. Multiplicative effect on odds equals $\exp(0.397) = 1.49$ for one-unit change in snoring, and 2.21 for two-unit change. Goodness-of-fit statistic $G^2 = 2.8$, $df = 2$ shows no evidence of lack of fit.
5. The Cochran–Armitage test uses the ordering of rows and has $df = 1$, and tends to give smaller P -values when there truly is a linear trend.
7. The model does not fit well ($G^2 = 31.7$, $df = 4$), with a particularly large negative residual for the first count. However the fit shows strong evidence of a tendency for the likelihood of lung cancer to increase at higher levels of smoking.
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 $e^{1.3068} = 3.7$ 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 -value $= .025$ for LR statistic.
- d. $G^2 = .38$, $X^2 = .20$, $df = 1$, so model fits well.
11. For main effects logit model with intercourse as response, estimated conditional odds ratios are 3.7 for race and 1.9 for gender; *e.g.*, controlling for gender, the odds of having ever had sexual intercourse are estimated to be $\exp(1.313) = 3.72$ times higher for blacks than for whites. Goodness-of-fit test gives $G^2 = 0.06$, $df = 1$, so model fits well.
13. The main effects model fits very well ($G^2 = .0002$, $df = 1$). Given gender, the odds of a white athlete graduating are estimated to be $\exp(1.015) = 2.8$ times the odds of a black athlete. Given race, the odds of a female graduating are estimated to be $\exp(.352) = 1.4$ times the odds of a male graduating. Both effects are highly significant, with Wald or likelihood-ratio tests (*e.g.*, the race effect of 1.015 has $SE = .087$, and the gender

effect of .352 has $SE = .080$).

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

21. The original variables c and x relate to the standardized variables z_c and z_x by $z_c = (c - 2.44)/.80$ and $z_x = (x - 26.3)/2.11$, so that $c = .80z_c + 2.44$ and $x = 2.11z_x + 26.3$. Thus, the prediction equation is

$$\text{logit}(\hat{\pi}) = -10.071 - .509[.80z_c + 2.44] + .458[2.11z_x + 26.3],$$

The coefficients of the standardized variables are $-.509(.80) = -.41$ and $.458(2.11) = .97$. Controlling for the other variable, a one standard deviation change in x has more than double the effect of a one standard deviation change in c . At $\bar{x} = 26.3$, the estimated logits at $c = 1$ and at $c = 4$ are 1.465 and $-.062$, which correspond to estimated probabilities of .81 and .48.

25. Logit model gives fit, $\text{logit}(\hat{\pi}) = -3.556 + .053(\text{income})$.

29. The odds ratio e^β is approximately equal to the relative risk when the probability is near 0 and the complement is near 1, since

$$e^\beta = [\pi(x+1)/(1-\pi(x+1))]/[\pi(x)/(1-\pi(x))] \approx \pi(x+1)/\pi(x).$$

31. The square of the denominator is the variance of $\text{logit}(\hat{\pi}) = \hat{\alpha} + \hat{\beta}x$. For large n , the ratio of $(\hat{\alpha} + \hat{\beta}x - \text{logit}(\pi_0))$ to its standard deviation is approximately standard normal, and (for fixed π_0) all x for which the absolute ratio is no larger than $z_{\alpha/2}$ are not contradictory.

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

$$\begin{aligned} P(Y = 1|x) &= \rho \exp[-(x-\mu_1)^2/2\sigma^2] / \{\rho \exp[-(x-\mu_1)^2/2\sigma^2] + (1-\rho) \exp[-(x-\mu_0)^2/2\sigma^2]\} \\ &= 1 / \{1 + [(1-\rho)/\rho] \exp\{-[\mu_0^2 - \mu_1^2 + 2x(\mu_1 - \mu_0)]/2\sigma^2\}\} \\ &= 1 / \{1 + \exp[-(\alpha + \beta x)]\} = \exp(\alpha + \beta x) / [1 + \exp(\alpha + \beta x)], \end{aligned}$$

where $\beta = (\mu_1 - \mu_0)/\sigma^2$ and $\alpha = -\log[(1-\rho)/\rho] + [\mu_0^2 - \mu_1^2]/2\sigma^2$.

35. a. Given $\{\pi_i\}$, we can find parameters so model holds exactly. With constraint $\beta_I = 0$, $\log[\pi_I/(1-\pi_I)] = \alpha$ determines α . Since $\log[\pi_i/(1-\pi_i)] = \alpha + \beta_i$, it follows that

$$\beta_i = \log[\pi_i/(1-\pi_i)] - \log[\pi_I/(1-\pi_I)].$$

That is, β_i is the log odds ratio for rows i and I of the table. When all β_i are equal, then the logit is the same for each row, so π_i is the same in each row, so there is independence.

37. d. When y_i is a 0 or 1, the log likelihood is $\sum_i [y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i)]$. 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

$$\begin{aligned} D &= -2 \sum_i [y_i \log \hat{\pi}_i + (1 - y_i) \log(1 - \hat{\pi}_i)] = -2 \sum_i [y_i \log \left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \right) + \log(1 - \hat{\pi}_i)] \\ &= -2 \sum_i [y_i(\hat{\alpha} + \hat{\beta}x_i) + \log(1 - \hat{\pi}_i)]. \end{aligned}$$

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

$$\begin{aligned} D &= -2[\hat{\alpha} \sum_i \hat{\pi}_i + \hat{\beta} \sum_i x_i \hat{\pi}_i + \sum_i \log(1 - \hat{\pi}_i)] \\ &= -2[\sum_i \hat{\pi}_i(\hat{\alpha} + \hat{\beta}x_i) + \sum_i \log(1 - \hat{\pi}_i)] \\ &= -2 \sum_i \hat{\pi}_i \log \left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \right) - 2 \sum_i \log(1 - \hat{\pi}_i). \end{aligned}$$

41. a. Expand $\log[p/(1-p)]$ in a Taylor series for a neighborhood of points around $p = \pi$, and take just the term with the first derivative.

b. Let $p_i = y_i/n_i$. The i th sample logit is

$$\begin{aligned} \log[p_i/(1-p_i)] &\approx \log[\pi_i^{(t)}/(1-\pi_i^{(t)})] + (p_i - \pi_i^{(t)})/\pi_i^{(t)}(1-\pi_i^{(t)}) \\ &= \log[\pi_i^{(t)}/(1-\pi_i^{(t)})] + [y_i - n_i\pi_i^{(t)}]/n_i\pi_i^{(t)}(1-\pi_i^{(t)}) \end{aligned}$$

Chapter 6

1. $\text{logit}(\hat{\pi}) = -9.35 + .834(\text{weight}) + .307(\text{width})$.

a. Like. ratio stat. = 32.9 ($df = 2$), $P < .0001$. There is extremely strong evidence that at least one variable affects the response.

b. Wald statistics are $(.834/.671)^2 = 1.55$ and $(.307/.182)^2 = 2.85$. These each have $df = 1$, and the P -values are .21 and .09. These predictors are highly correlated (Pearson corr. = .887), so this is the problem of multicollinearity.

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).

9. The CMH statistic simplifies to the McNemar statistic of Sec. 10.1, which in chi-squared form equals $(14 - 6)^2/(14 + 6) = 3.2$ ($df = 1$). There is slight evidence of a better response with treatment B ($P = .074$ for the two-sided alternative).

13. $\text{logit}(\hat{\pi}) = -12.351 + .497x$. Prob. at $x = 26.3$ is .674; prob. at $x = 28.4$ (i.e., one std. dev. above mean) is .854. The odds ratio is $[(.854/.146)/(.674/.326)] = 2.83$, so $\lambda = 1.04$, $\delta = 5.1$. Then $n = 75$.

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)²/fitted for the “success” category and the second component is (observed - fitted)²/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.

25. The noncentrality is the same for models $(X + Z)$ and (Z) , so the difference statistic has noncentrality 0. The conditional XY independence model has noncentrality proportional to n , so the power goes to 1 as n increases.

29 a. $P(y = 1) = P(\alpha_1 + \beta_1 x_1 + \epsilon_1 > \alpha_0 + \beta_0 x_0 + \epsilon_0) = P[(\epsilon_0 - \epsilon_1)/\sqrt{2} < ((\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)x)/\sqrt{2}] = \Phi(\alpha^* + \beta^*x)$ where $\alpha^* = (\alpha_1 - \alpha_0)/\sqrt{2}$, $\beta^* = (\beta_1 - \beta_0)/\sqrt{2}$.

31. $(\log \pi(x_2))/(\log \pi(x_1)) = \exp[\beta(x_2 - x_1)]$, so $\pi(x_2) = \pi(x_1)^{\exp[\beta(x_2 - x_1)]}$. For $x_2 - x_1 = 1$, $\pi(x_2)$ equals $\pi(x_1)$ raised to the power $\exp(\beta)$.

Chapter 7

1. a.

$\log(\hat{\pi}_1/\hat{\pi}_2) = (.883 + .758) + (.419 - .105)x_1 + (.342 - .271)x_2 = 1.641 + .314x_1 + 1 + .071x_2$.

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.

3. Both gender and race have significant effects. The logit model with additive effects and no interaction fits well, with $G^2 = 0.2$ based on $df = 2$. The estimated odds of preferring Democrat instead of Republican are higher for females and for blacks, with estimated conditional odds ratios of 1.8 between gender and party ID and 9.8 between race and party ID.

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 $\exp(3.3074)/[1 + \exp(3.3074)] = .965$, $\exp(3.4818)/[1 + \exp(3.4818)] = .970$, $\exp(5.3494)/[1 + \exp(5.3494)] = .995$, $\exp(7.2563)/[1 + \exp(7.2563)] = .9993$, and 1.0. The corresponding response probabilities are .965, .005, .025, .004, and .0007.

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 $G^2 = 4.5$, $df = 6$).

11. a. The model has 12 baseline-category logits (2 for each of the 6 combinations of S and A) and 8 parameters (an intercept, an A effect and two S effects, for each logit), so $df = 4$. The model does not account for ordinality, since the model allows different effects for each logit and treats A as a factor. Also, it does not permit interaction.

b. Since $\exp(.663) = 1.94$, the effect of smoking is 94% higher for the older-aged subjects.

c. For this age group the log odds ratio with smoking is $\beta_1 + \beta_3$ for smoking levels one unit apart and $2(\beta_1 + \beta_3)$ for smoking levels two units apart. Thus, the estimated odds ratios are $\exp(.115 + .663) = 2.18$ and $\exp[2(.115 + .663)] = 4.74$. For the extreme categories of B , which are two units apart, the log odds ratios double and the odds ratios square.

13. For $x_2 = 0$, $\hat{P}(Y > 2) = 1 - \hat{P}(Y \leq 1) = 1 - \Phi(-.161 - .195x_1) = \Phi(.161 + .195x_1) = \Phi([x_1 - .83]/5.13)$, so the shape is that of a normal cdf with $\mu = .83$ and $\sigma = 5.13$. For $x_2 = 1$, $\mu = 2.67$ and $\sigma = 5.13$.

17. a. CMH statistic for correlation alternative, using equally-spaced scores, equals 6.3 ($df = 1$) and has P -value = .012. When there is roughly a linear trend, this tends to be more powerful and give smaller P -values, since it focuses on a single degree of freedom.

b. LR statistic for cumulative logit model with linear effect of operation = 6.7, $df = 1$, $P = .01$; strong evidence that operation has an effect on dumping, gives similar results as in (a).

c. LR statistic comparing this model to model with four separate operation parameters equals 2.8 ($df = 3$), so simpler model is adequate.

$$27. \partial\pi_3(x)/\partial x = \frac{-[\beta_1 \exp(\alpha_1 + \beta_1 x) + \beta_2 \exp(\alpha_2 + \beta_2 x)]}{[1 + \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x)]^2}.$$

a. The denominator is positive, and the numerator is negative when $\beta_1 > 0$ and $\beta_2 > 0$.

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 $j < k$, $\text{logit}[P(Y \leq j | X = x_i)] - \text{logit}[P(Y \leq k | X = x_i)] = (\alpha_j - \alpha_k) + (\beta_j - \beta_k)x$. 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 .

33. The local odds ratios refer to a narrow region of the response scale (categories j and $j - 1$ alone), whereas cumulative odds ratios refer to the entire response scale.

35. From the argument in Sec. 7.2.3, the effect β refers to an underlying continuous variable with a normal distribution with standard deviation 1.

37. a. $df = I(J - 1) - [(J - 1) + (I - 1)] = (I - 1)(J - 2)$.

b. The full model has an extra $I - 1$ parameters.

c. The cumulative probabilities in row a are all smaller or all greater than those in row b depending on whether $\mu_a > \mu_b$ or $\mu_a < \mu_b$.

41. For a given subject, the model has the form

$$\pi_j = \frac{\alpha_j + \beta_j x + \gamma u_j}{\sum_h \alpha_h + \beta_h x + \gamma u_h}.$$

For a given cost, the odds a female selects a over b are $\exp(\beta_a - \beta_b)$ times the odds for males. For a given gender, the log odds of selecting a over b depend on $u_a - u_b$.

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.

3. For either approach, from (8.14), the estimated conditional log odds ratio equals

$$\hat{\lambda}_{11}^{AC} + \hat{\lambda}_{22}^{AC} - \hat{\lambda}_{12}^{AC} - \hat{\lambda}_{21}^{AC}$$

5. a. Let S = safety equipment, E = whether ejected, I = injury. Then, $G^2(SE, SI, EI) = 2.85$, $df = 1$. Any simpler model has $G^2 > 1000$, so it seems there is an association for each pair of variables, and that association can be regarded as the same at each level of the third variable. The estimated conditional odds ratios are .091 for S and E (i.e., wearers of seat belts are much less likely to be ejected), 5.57 for S and I , and .061 for E and I .

b. Loglinear models containing SE are equivalent to logit models with I as response variable and S and E as explanatory variables. The loglinear model (SE, SI, EI) is equivalent to a logit model in which S and E have additive effects on I . The estimated odds of a fatal injury are $\exp(2.798) = 16.4$ times higher for those ejected (controlling for S), and $\exp(1.717) = 5.57$ times higher for those not wearing seat belts (controlling

for E).

7. Injury has estimated conditional odds ratios .58 with gender, 2.13 with location, and .44 with seat-belt use. “No” is category 1 of I , and “female” is category 1 of G , so the odds of no injury for females are estimated to be .58 times the odds of no injury for males (controlling for L and S); that is, females are more likely to be injured. Similarly, the odds of no injury for urban location are estimated to be 2.13 times the odds for rural location, so injury is more likely at a rural location, and the odds of no injury for no seat belt use are estimated to be .44 times the odds for seat belt use, so injury is more likely for no seat belt use, other things being fixed. Since there is no interaction for this model, overall the most likely case for injury is therefore females not wearing seat belts in rural locations.

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

b. Model with A as response and additive factor effects for R and P , $\text{logit}(\pi) = \alpha + \beta_i^R + \beta_j^P$.

c. (i) (GRP, A), $\text{logit}(\pi) = \alpha$, (ii) (GRP, AR), $\text{logit}(\pi) = \alpha + \beta_i^R$,

(iii) (GRP, APR, AG), add term of form β_{ij}^{RP} to logit model in Exercise 5.23.

13. Homogeneous association model (BP, BR, BS, PR, PS, RS) fits well ($G^2 = 7.0$, $df = 9$). Model deleting PR association also fits well ($G^2 = 10.7$, $df = 11$), but we use the full model.

For homogeneous association model, estimated conditional BS odds ratio equals $\exp(1.147) = 3.15$. For those who agree with birth control availability, the estimated odds of viewing premarital sex as wrong only sometimes or not wrong at all are about triple the estimated odds for those who disagree with birth control availability; there is a positive association between support for birth control availability and premarital sex. The 95% CI is $\exp(1.147 \pm 1.645(.153)) = (2.45, 4.05)$.

Model (BPR, BS, PS, RS) has $G^2 = 5.8$, $df = 7$, and also a good fit.

17. b. $\log \theta_{11(k)} = \log \mu_{11k} + \log \mu_{22k} - \log \theta_{12k} - \log \theta_{21k} = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}$; for zero-sum constraints, as in problem 16c this simplifies to $4\lambda_{11}^{XY}$.

e. Use equations such as

$$\lambda = \log(\mu_{111}), \quad \lambda_i^X = \log\left(\frac{\mu_{i11}}{\mu_{111}}\right), \quad \lambda_{ij}^{XY} = \log\left(\frac{\mu_{ij1}\mu_{111}}{\mu_{i11}\mu_{1j1}}\right)$$

$$\lambda_{ijk}^{XYZ} = \log\left(\frac{[\mu_{ijk}\mu_{11k}/\mu_{i1k}\mu_{1jk}]}{[\mu_{ij1}\mu_{111}/\mu_{i11}\mu_{1j1}]}\right)$$

19. a. When Y is jointly independent of X and Z , $\pi_{ijk} = \pi_{+j}\pi_{i+k}$. Dividing π_{ijk} by π_{++k} , we find that $P(X = i, Y = j|Z = k) = P(X = i|Z = k)P(Y = j)$. But when $\pi_{ijk} = \pi_{+j}\pi_{i+k}$, $P(Y = j|Z = k) = \pi_{+jk}/\pi_{++k} = \pi_{+j}\pi_{++k}/\pi_{++k} = \pi_{+j} = P(Y = j)$. 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, $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$. Summing both sides over k , $\pi_{ij+} = \pi_{i++}\pi_{+j+}$, 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.

	Y			Y	
	.125	.125		.125	.125
X	.125	.125		.125	.125
	Z = 1			Z = 2	

This is actually a special case of (X, Y, Z) called the *equiprobability model*.

b.

.15	.10		.15	.10
.10	.15		.10	.15

c.

1/4	1/24		1/12	1/8
1/8	1/12		1/24	1/4

d.

2/16	1/16		4/16	1/16
1/16	4/16		1/16	2/16

e. Any $2 \times 2 \times 2$ table

23. Number of terms = $1 + \binom{T}{1} + \binom{T}{2} + \dots + \binom{T}{T} = \sum_i \binom{T}{i} 1^i 1^{T-i} = (1+1)^T$,
by the Binomial theorem.

25. a. The λ^{XY} term does not appear in the model, so X and Y are conditionally independent. All terms in the saturated model that are not in model (WXZ, WYZ) involve X and Y , so permit an XY conditional association.

b. (WX, WZ, WY, XZ, YZ)

27. $\beta = (\alpha_1, \alpha_2, \beta)'$, \mathbf{X} is the 6×3 matrix with rows $(1, 0, x_1, / 0, 1, x_1, / 1, 0, x_2, / 0, 1, x_2, / 1, 0, x_3, / 0, 1, x_3)$, \mathbf{C} is the 3×12 matrix with rows $(1, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, / 0, 0, 1, -1, 0, 0, 0, 0, 0, 0, 0, 0, / 0, 0, 0, 0, 1, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,$

-1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, -1, 0, 0, / 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, -1), and \mathbf{A} is the 12×9 matrix with rows (1, 0, 0, 0, 0, 0, 0, 0, 0, / 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, / 0, 0, 0, 0, 0, 0, 1, 0, 0, / 0, 0, 0, 0, 0, 0, 0, 1, 1),

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

$$L = n\lambda + \sum_i n_{i++}\lambda_i^X + \sum_j n_{+j+}\lambda_j^Y + \sum_k n_{+++}\lambda_k^Z + \sum_i \sum_j n_{ij+}\lambda_{ij}^{XY} - \sum \sum \sum \mu_{ijk}$$

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

33. For (XY, Z) , $df = IJK - [1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1)]$. For (XY, YZ) , $df = IJK - [1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1) + (J - 1)(K - 1)]$. For (XY, XZ, YZ) , $df = IJK - [1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1) + (J - 1)(K - 1) + (I - 1)(J - 1)]$.

35. a. The formula reported in the table satisfies the likelihood equations $\hat{\mu}_{h+++} = n_{h+++}$, $\hat{\mu}_{+i++} = n_{+i++}$, $\hat{\mu}_{+++j+} = n_{+++j+}$, $\hat{\mu}_{++++k} = n_{++++k}$, and they satisfy the model, which has probabilistic form $\pi_{hijk} = \pi_{h+++}\pi_{+i++}\pi_{+++j+}\pi_{++++k}$, so by Birch's results they are ML estimates.

b. Model (WX, YZ) says that the composite variable (having marginal frequencies $\{n_{hi++}\}$) is independent of the YZ composite variable (having marginal frequencies $\{n_{+++jk}\}$). Thus, $df = [\text{no. categories of } (XY) - 1][\text{no. categories of } (YZ) - 1] = (HI - 1)(JK - 1)$. Model (WXY, Z) says that Z is independent of the WXY composite variable, so the usual results apply to the two-way table having Z in one dimension, HIJ levels of WXY composite variable in the other; e.g., $df = (HIJ - 1)(K - 1)$.

37. $\beta = (\lambda, \lambda_1^X, \lambda_1^Y, \lambda_1^Z)'$, $\mu = (\mu_{111}, \mu_{112}, \mu_{121}, \mu_{122}, \mu_{211}, \mu_{212}, \mu_{221}, \mu_{222})'$, and \mathbf{X} is a 8×4 matrix with rows (1, 1, 1, 1, / 1, 1, 1, 0, / 1, 1, 0, 1, / 1, 1, 0, 0, / 1, 0, 1, 1, / 1, 0, 1, 0, / 1, 0, 0, 1, / 1, 0, 0, 0).

39. Take $\pi_{ij}^{(t+1)} = \pi_{ij}^{(t)}(r_i/\pi_{i+}^{(t)})$, so row totals match $\{r_i\}$, and then $\pi_{ij}^{(t+2)} = \pi_{ij}^{(t+1)}(c_j/\pi_{+j}^{(t+1)})$, so column totals match, for $t = 1, 2, \dots$, where $\{\pi_{ij}^{(0)} = p_{ij}\}$.

Chapter 9

1. a. For any pair of variables, the marginal odds ratio is the same as the conditional odds ratio (and hence 1.0), since the remaining variable is conditionally independent of each of those two.

b. (i) For each pair of variables, at least one of them is conditionally independent of the remaining variable, so the marginal odds ratio equals the conditional odds ratio. (ii)

these are the likelihood equations implied by the λ^{AC} term in the model.

c. (i) Both A and C are conditionally dependent with M , so the association may change when one controls for M . (ii) For the AM odds ratio, since A and C are conditionally independent (given M), the odds ratio is the same when one collapses over C . (iii) These are likelihood equations implied by the λ^{AM} and λ^{CM} terms in the model.

d. (i) no pairs of variables are conditionally independent, so collapsibility conditions are not satisfied for any pair of variables. (ii) These are likelihood equations implied by the three association terms in the model.

5. Model (AC, AM, CM) fits well. It has $df = 1$, and the likelihood equations imply fitted values equal observed in each two-way marginal table, which implies the difference between an observed and fitted count in one cell is the negative of that in an adjacent cell; their SE values are thus identical, as are the standardized Pearson residuals. The other models fit poorly; e.g. for model (AM, CM) , in the cell with each variable equal to yes, the difference between the observed and fitted counts is 3.7 standard errors.

15. With log link, $G^2 = 3.79$, $df = 2$; estimated death rate for older age group is $e^{1.25} = 3.49$ times that for younger group.

17. Do a likelihood-ratio test with and without time as a factor in the model

19. The model appears to fit adequately. The estimated constant collision rate is $\exp() = .0153$ accidents per million miles of travel.

21.a. The ratio of the rate for smokers to nonsmokers decreases markedly as age increases.

b. $G^2 = 12.1$, $df = 4$.

c. For age scores (1,2,3,4,5), $G^2 = 1.5$, $df = 3$. The interaction term = $-.309$, with std. error = $.097$; the estimated ratio of rates is multiplied by $\exp(-.309) = .73$ for each successive increase of one age category.

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.

27. a. Yes – let U be a composite variable consisting of combinations of levels of Y and Z ; then, collapsibility conditions are satisfied as W is conditionally independent of U , given X .

b. No.

33. From the definition, it follows that a joint distribution of two discrete variables is positively likelihood-ratio dependent if all odds ratios of form $\mu_{ij}\mu_{hk}/\mu_{ik}\mu_{hj} \geq 1$, when $i < h$ and $j < k$.

a. For $L \times L$ model, this odds ratio equals $\exp[\beta(u_h - u_i)(v_k - v_j)]$. Monotonicity of scores implies $u_i < u_h$ and $v_j < v_k$, so these odds ratios all are at least equal to 1.0 when $\beta \geq 0$. Thus, when $\beta > 0$, as X increases, the conditional distributions on Y are stochastically increasing; also, as Y increases, the conditional distributions on X are stochastically increasing. When $\beta < 0$, the variables are negatively likelihood-ratio dependent, and the

conditional distributions on Y (X) are stochastically decreasing as X (Y) increases.

b. For row effects model with $j < k$, $\mu_{hj}\mu_{ik}/\mu_{hk}\mu_{ij} = \exp[(\mu_i - \mu_h)(v_k - v_j)]$. When $\mu_i - \mu_h > 0$, all such odds ratios are positive, since scores on Y are monotone increasing. Thus, there is likelihood-ratio dependence for the $2 \times J$ table consisting of rows i and h , and Y is stochastically higher in row i .

35. a. Note the derivative of the log likelihood with respect to β is $\sum_i \sum_j u_i v_j (n_{ij} - \mu_{ij})$, which under indep. estimates is $n \sum_i \sum_j u_i v_j (p_{ij} - p_{i+} p_{+j})$.

b. Use formula (3.9). In this context, $\zeta = \sum \sum u_i v_j (\pi_{ij} - \pi_{i+} \pi_{+j})$ and $\phi_{ij} = u_i v_j - u_i (\sum_b v_b \pi_{+b}) - v_j (\sum_a u_a \pi_{a+})$. Under H_0 , $\pi_{ij} = \pi_{i+} \pi_{+j}$, and $\sum \sum \pi_{ij} \phi_{ij}$ simplifies to $-(\sum u_i \pi_{i+})(\sum v_j \pi_{+j})$. Also under H_0 ,

$$\begin{aligned} \sum_i \sum_j \pi_{ij} \phi_{ij}^2 &= \sum_i \sum_j u_i^2 v_j^2 \pi_{i+} \pi_{+j} + (\sum_j v_j \pi_{+j})^2 (\sum_i u_i^2 \pi_{i+}) + (\sum_i u_i \pi_{i+})^2 (\sum_j v_j^2 \pi_{+j}) \\ &+ 2(\sum_i \sum_j u_i v_j \pi_{i+} \pi_{+j}) (\sum_i u_i \pi_{i+}) (\sum_j v_j \pi_{+j}) - 2(\sum_i u_i^2 \pi_{i+}) (\sum_j v_j \pi_{+j})^2 - 2(\sum_j v_j^2 \pi_{+j}) (\sum_i u_i \pi_{i+})^2. \end{aligned}$$

Then σ^2 in (3.9) simplifies to

$$[\sum_i u_i^2 \pi_{i+} - (\sum_i u_i \pi_{i+})^2] [\sum_j v_j^2 \pi_{+j} - (\sum_j v_j \pi_{+j})^2].$$

The asymptotic standard error is σ/\sqrt{n} , the estimate of which is the same formula with π_{ij} replaced by p_{ij} .

37. a. If parameters do not satisfy these constraints, set $\bar{\lambda}_i^X = \lambda_i^X - \lambda_I^X$, $\bar{\lambda}_j^Y = \lambda_j^Y - \lambda_I^Y + \mu_1(v_j - v_I)$, $\bar{\mu}_i = \mu_i - \mu_I$. Then $\log m_{ij} = \lambda + (\bar{\lambda}_i^X + \lambda_I^X) + [\bar{\lambda}_j^Y + \lambda_I^Y + \mu_I(v_I - v_j)] + (\bar{\mu}_i + \mu_I)v_j = \lambda' + \bar{\lambda}_i^X + \bar{\lambda}_j^Y + \bar{\mu}_i v_j$, where $\lambda' = \lambda + \lambda_I^X + \lambda_I^Y + u_I v_I$. This has row effects form with the indicated constraints.

b. For Poisson sampling, log likelihood is

$$L = n\lambda + \sum_i n_{i+} \lambda_i^X + \sum_j n_{+j} \lambda_j^Y + \sum_i \mu_i [\sum_j n_{ij} v_j] - \sum_i \sum_j \exp(\lambda + \dots)$$

Thus, the minimal sufficient statistics are $\{n_{i+}\}$, $\{n_{+j}\}$, and $\{\sum_j n_{ij} v_j\}$. Differentiating with respect to the parameters and setting results equal to zero gives the likelihood equations. For instance, $\partial L / \partial \mu_i = \sum_j v_j n_{ij} - \sum_j v_j \mu_{ij}$, $i = 1, \dots, I$, from which follows the I equations in the third set of likelihood equations.

39. a. These equations are obtained successively by differentiating with respect to λ^{XZ} , λ^{YZ} , and β . Note these equations imply that the correlation between the scores for X and the scores for Y is the same for the fitted and observed data. This model uses the ordinality of X and Y , and is a parsimonious special case of model (XY, XZ, YZ) .

b. Can calculate this directly, or simply note that the model has one more parameter than the conditional XY independence model (XZ, YZ) , so it has one fewer df than that model.

c. When $I = J = 2$, λ^{XY} has only one nonredundant value. For zero-sum constraints, we can take $u_1 = v_1 = -u_2 = -v_2$ and have $\beta u_i v_j = \lambda_i^{XY}$. (Note the distinction between

ordinal and nominal is irrelevant when there are only two categories; we cannot exploit “trends” until there are at least 3 categories.)

d. The third equation is replaced by the K equations,

$$\sum_i \sum_j u_i v_j \hat{\mu}_{ijk} = \sum_i \sum_j u_i v_j n_{ijk}, \quad k = 1, \dots, K.$$

This model corresponds to fitting $L \times L$ model separately at each level of Z . The G^2 value is the sum of G^2 for separate fits, and df is the sum of $IJ - I - J$ values from separate fits (i.e., $df = K(IJ - I - J)$).

41. a.

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \mu_i v_j,$$

where $\{v_j\}$ are fixed scores and the row effects satisfy a constraint such as $\sum_i \mu_i = 0$. The likelihood equations are $\{\hat{\mu}_{i+k} = n_{i+k}\}$, $\{\hat{\mu}_{+jk} = n_{+jk}\}$, and $\{\sum_j v_j \hat{\mu}_{ij+} = \sum_j v_j n_{ij+}\}$, and residual $df = IJK - [1 + (I-1) + (J-1) + (K-1) + (I-1)(K-1) + (J-1)(K-1) + (I-1)]$. For unit-spaced scores, within each level k of Z , the odds that Y is in category $j + 1$ instead of j are $\exp(\mu_a - \mu_b)$ times higher in level a of X than in level b of X . Note this model treats Y alone as ordinal, and corresponds to an adjacent-categories logit model for Y as a response in which X and Z have additive effects but no interaction. For equally-spaced scores, those effects are the same for each logit.

b. Replace final term in model in (a) by $\mu_{ik} v_j$, where parameters satisfy constraint such as $\sum_i \mu_{ik} = 0$ for each k . Replace final term in df expression by $K(I - 1)$. Within level k of Z , the odds that Y is in category $j + 1$ rather than j are $\exp(\mu_{ak} - \mu_{bk})$ times higher at level a of X than at level b of X .

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., $\hat{\mu}_{11+} = n_{11+}$). 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.’

3. a. Ignoring order, $(A=1, B=0)$ occurred 45 times and $(A=0, B=1)$ occurred 22 times.

The McNemar $z = 2.81$, which has a two-tail P -value of .005 and provides strong evidence that the response rate of successes is higher for drug A.

b. Pearson statistic = 7.8, $df = 1$

5. a. Symmetry model has $X^2 = .59$, based on $df = 3$ ($P = .90$). Independence has $X^2 = 45.4$ ($df = 4$), and quasi independence has $X^2 = .01$ ($df = 1$) and is identical to quasi symmetry. The symmetry and quasi independence models fit well.

b. $G^2(S | QS) = 0.591 - 0.006 = 0.585$, $df = 3 - 1 = 2$. Marginal homogeneity is plausible.

c. Kappa = .389 ($SE = .060$), weighted kappa equals .427 ($SE = .0635$).

13. Under independence, on the main diagonal, fitted = 5 = observed. Thus, kappa = 0, yet there is clearly strong association in the table.

15. a. Good fit, with $G^2 = 0.3$, $df = 1$. The parameter estimates for Coke, Pepsi, and Classic Coke are 0.580 ($SE = 0.240$), 0.296 ($SE = 0.240$), and 0. Coke is preferred to Classic Coke.

b. model estimate = 0.57, sample proportion = $29/49 = 0.59$.

17. a. For testing fit of the model, $G^2 = 4.6$, $X^2 = 3.2$ ($df = 6$). Setting the $\hat{\beta}_5 = 0$ for Sanchez, $\hat{\beta}_1 = 1.53$ for Seles, $\hat{\beta}_2 = 1.93$ for Graf, $\hat{\beta}_3 = .73$ for Sabatini, and $\hat{\beta}_4 = 1.09$ for Navratilova. The ranking is Graf, Seles, Navratilova, Sabatini, Sanchez.

b. $\exp(-.4)/[1 + \exp(-.4)] = .40$. This also equals the sample proportion, $2/5$. The estimate $\hat{\beta}_1 - \hat{\beta}_2 = -.40$ has $SE = .669$. A 95% CI of (-1.71, .91) for $\beta_1 - \beta_2$ translates to (.15, .71) for the probability of a Seles win.

c. Using a 98% CI for each of the 10 pairs, only the difference between Graf and Sanchez is significant.

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.

c. 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} .

25. When $\{\alpha_i\}$ are identical, the individual trials for the conditional model are identical as well as independent, so averaging over them to get the marginal Y_1 and Y_2 gives binomials with the same parameters.

29. Consider the 3×3 table with cell probabilities, by row, (.2, .10, 0, / 0, .3, .10, / .10, 0, .2).

39. a. Since $\pi_{ab} = \pi_{ba}$, it satisfies symmetry, which then implies marginal homogeneity and quasi symmetry as special cases. For $a \neq b$, π_{ab} has form $\alpha_a \beta_b$, identifying β_b with $\alpha_b(1 - \beta)$, so it also satisfies quasi independence.

c. $\beta = \kappa = 0$ is equivalent to independence for this model, and $\beta = \kappa = 1$ is equivalent to perfect agreement.

41. a. The association term is symmetric in a and b . It is a special case of the quasi association model in which the main-diagonal parameters are replaced by a common parameter δ .

c. Likelihood equations are, for all a and b , $\hat{\mu}_{a+} = n_{a+}$, $\hat{\mu}_{+b} = n_{+b}$, $\sum_a \sum_b u_a u_b \hat{\mu}_{ab} = \sum_a \sum_b u_a u_b n_{ab}$, $\sum_a \hat{\mu}_{aa} = \sum_a n_{aa}$.

Chapter 11

1. The sample proportions of yes responses are .86 for alcohol, .66 for cigarettes, and .42 for marijuana. To test marginal homogeneity, the likelihood-ratio statistic equals 1322.3 and the general CMH statistic equals 1354.0 with $df = 2$, extremely strong evidence of differences among the marginal distributions.

3. a. Since $R = G = S_1 = S_2 = 0$, estimated logit is $-.57$ and estimated odds = $\exp(-.57)$.

b. Race does not interact with gender or substance type, so the estimated odds for white subjects are $\exp(0.38) = 1.46$ times the estimated odds for black subjects.

c. For alcohol, estimated odds ratio = $\exp(-.20 + 0.37) = 1.19$; for cigarettes, $\exp(-.20 + .22) = 1.02$; for marijuana, $\exp(-.20) = .82$.

d. Estimated odds ratio = $\exp(1.93 + .37) = 9.97$.

e. Estimated odds ratio = $\exp(1.93) = 6.89$.

7. a. Subjects can select any number of the sources, from 0 to 5, so a given subject could have anywhere from 0 to 5 observations in this table. The multinomial distribution does not apply to these 40 cells.

b. The estimated correlation is weak, so results will not be much different from treating the 5 responses by a subject as if they came from 5 independent subjects. For source A the estimated size effect is 1.08 and highly significant (Wald statistic = 6.46, $df = 1$, $P < .0001$). For sources C, D, and E the size effect estimates are all roughly $-.2$.

c. One can then use such a parsimonious model that sets certain parameters to be equal, and thus results in a smaller SE for the estimate of that effect (.063 compared to values around .11 for the model with separate effects).

9. a. The general CMH statistic equals 14.2 ($df = 3$), showing strong evidence against marginal homogeneity ($P = .003$). Likewise, Bhapkar $W = 12.8$ ($P = .005$)

b. With a linear effect for age using scores 9, 10, 11, 12, the GEE estimate of the age effect is .086 ($SE = .025$), based on the exchangeable working correlation. The P -value

(.0006) is even smaller than in (a), as the test is focused on $df = 1$.

11. GEE estimate of cumulative log odds ratio is 2.52 ($SE = .12$), similar to ML.

13. b. $\hat{\lambda} = 1.08$ ($SE = .29$) gives strong evidence that the active drug group tended to fall asleep more quickly, for those at the two highest levels of initial time to fall asleep.

15. First-order Markov model has $G^2 = 40.0$ ($df = 8$), a poor fit. If we add association terms for the other pairs of ages, we get $G^2 = 0.81$ and $X^2 = 0.84$ ($df = 5$) and a good fit.

21. CMH methods summarize information from the counts in the various strata, treating them as hypergeometric after conditioning on row and column totals in each stratum. When a subject makes the same response for each drug, the stratum for the subject has observations in one column only, and the generalized hypergeometric distribution is degenerate and has variance 0 for each count.

23. Since $\partial\mu_i/\partial\beta = 1$, $u(\beta) = \sum_i \left(\frac{\partial\mu_i}{\partial\beta}\right)' v(\mu_i)^{-1}(y_i - \mu_i) = \sum_i \mu_i^{-1}(y_i - \mu_i) = \sum_i (y_i - \beta)/\beta$. Setting this equal to 0, $\sum_i y_i = n\beta$. Also, $\mathbf{V} = \left[\sum_i \left(\frac{\partial\mu_i}{\partial\beta}\right)' [v(\mu_i)]^{-1} \left(\frac{\partial\mu_i}{\partial\beta}\right)\right]^{-1} = [\sum_i \mu_i^{-1}]^{-1} = [n/\beta]^{-1} = \beta/n$. Also, the actual asymptotic variance that allows for variance misspecification is

$$\mathbf{V} \left[\sum_i \left(\frac{\partial\mu_i}{\partial\beta}\right)' [v(\mu_i)]^{-1} \text{Var}(Y_i) [v(\mu_i)]^{-1} \left(\frac{\partial\mu_i}{\partial\beta}\right) \right] \mathbf{V} = (\beta/n) \left[\sum_i \mu_i^{-1} \mu_i^2 \mu_i^{-1} \right] (\beta/n) = \beta^2/n.$$

Replacing the true variance μ_i^2 in this expression by $(y_i - \bar{y})^2$, the last expression simplifies (using $\mu_i = \beta$) to $\sum_i (y_i - \bar{y})^2/n^2$.

25. Since $v(\mu_i) = \mu_i$ for the Poisson and since $\mu_i = \beta$, the model-based asymptotic variance is

$$\mathbf{V} = \left[\sum_i \left(\frac{\partial\mu_i}{\partial\beta}\right)' [v(\mu_i)]^{-1} \left(\frac{\partial\mu_i}{\partial\beta}\right) \right]^{-1} = \left[\sum_i (1/\mu_i) \right]^{-1} = \beta/n.$$

Thus, the model-based asymptotic variance estimate is \bar{y}/n . The actual asymptotic variance that allows for variance misspecification is

$$\begin{aligned} & \mathbf{V} \left[\sum_i \left(\frac{\partial\mu_i}{\partial\beta}\right)' [v(\mu_i)]^{-1} \text{Var}(Y_i) [v(\mu_i)]^{-1} \left(\frac{\partial\mu_i}{\partial\beta}\right) \right] \mathbf{V} \\ &= (\beta/n) \left[\sum_i (1/\mu_i) \text{Var}(Y_i) (1/\mu_i) \right] (\beta/n) = \left(\sum_i \text{Var}(Y_i) \right) / n^2, \end{aligned}$$

which is estimated by $[\sum_i (Y_i - \bar{y})^2]/n^2$. The model-based estimate tends to be better when the model holds, and the robust estimate tends to be better when there is severe overdispersion so that the model-based estimate tends to underestimate the SE .

27. a. QL assumes only a variance structure but not a particular distribution. They are equivalent if one assumes in addition that the distribution is in the natural exponential

family with that variance function.

b. GEE does not assume a parametric distribution, but only a variance function and a correlation structure.

c. An advantage is being able to extend ordinary GLMs to allow for overdispersion, for instance by permitting the variance to be some constant multiple of the variance for the usual GLM. A disadvantage is not having a likelihood function and related likelihood-ratio tests and confidence intervals.

d. They are consistent if the model for the mean is correct, even if one misspecifies the variance function and correlation structure. They are not consistent if one misspecifies the model for the mean.

31. Yes, because it is still true that given y_t , Y_{t+1} does not depend on y_0, y_1, \dots, y_{t-1} .

33. a. For this model, given the state at a particular time, the probability of transition to a particular state is independent of the time of transition.

Chapter 12

1. a. With a sufficient number of quadrature points, the number of which depends on starting values, $\hat{\beta}$ converges to 0.813 with $SE = 0.127$. For a given subject, the estimated odds of approval for the second question are $\exp(.813) = 2.25$ times the estimated odds for the first question.

b. Same as conditional ML $\hat{\beta} = \log(203/90)$ ($SE = 0.127$).

3. For a given subject, the odds of having used cigarettes are estimated to equal $\exp[1.6209 - (-.7751)] = 11.0$ times the odds of having used marijuana. The large value of $\hat{\sigma} = 3.5$ reflects strong associations among the three responses.

7. $\hat{\beta}_B = 1.99$ ($SE = .35$), $\hat{\beta}_C = 2.51$ ($SE = .37$), with $\hat{\sigma} = 0$. e.g., for a given subject for any sequence, the estimated odds of relief for A are $\exp(-1.99) = .13$ times the estimated odds for B (and odds ratio = .08 comparing A to C and .59 comparing B and C). Taking into account SE values, B and C are better than A .

9. Comparing the simpler model with the model in which treatment effects vary by sequence, double the change in maximized log likelihood is 13.6 on $df = 10$; $P = .19$ for comparing models. The simpler model is adequate. Adding period effects to the simpler model, the likelihood-ratio statistic = .5, $df = 2$, so the evidence of a period effect is weak.

11. a. For a given department, the estimated odds of admission for a female are $\exp(.173) = 1.19$ times the estimated odds of admission for a male.

b. For a given department, the estimated odds of admission for a female are $\exp(.163) = 1.18$ times the estimated odds of admission for a male.

c. The estimated mean log odds ratio between gender and admissions, given department, is .176, corresponding to an odds ratio of 1.19. Because of the extra variance component, permitting heterogeneity among departments, the estimate of β is not as precise.

d. The marginal odds ratio of $\exp(-.07) = .93$ is in a different direction, corresponding

to an odds of being admitted that is lower for females than for males. This is Simpson's paradox, and by results in Chapter 9 on collapsibility is possible when Department is associated both with gender and with admissions.

e. The random effects model assumes the true log odds ratios come from a normal distribution. It smooths the sample values, shrinking them toward a common mean.

15. a. There is more support for increasing government spending on education than on the others.

17. The likelihood-ratio statistic equals $-2(-593 - (-621)) = 56$. The null distribution is an equal mixture of degenerate at 0 and χ_1^2 , and the P -value is half that of a χ_1^2 variate, and is 0 to many decimal places. There is extremely strong evidence that $\sigma > 0$.

19. $\hat{\beta}_{2M} - \hat{\beta}_{1M} = .39$ ($SE = .09$), $\hat{\beta}_{2A} - \hat{\beta}_{1A} = .07$ ($SE = .06$), with $\hat{\sigma}_1 = 4.1$, $\hat{\sigma}_2 = 1.8$, and estimated correlation .33 between random effects.

21. d. When $\hat{\sigma}$ is large, the log likelihood is flat and many N values are consistent with the sample. A narrower interval is not necessarily more reliable. If the model is incorrect, the actual coverage probability may be much less than the nominal probability.

27. When $\hat{\sigma} = 0$, the model is equivalent to the marginal one deleting the random effect. Then, probability = odds/(1 + odds) = $\exp[\text{logit}(q_i) + \alpha]/[1 + \exp[\text{logit}(q_i) + \alpha]]$. Also, $\exp[\text{logit}(q_i)] = \exp[\log(q_i) - \log(1 - q_i)] = q_i/(1 - q_i)$. The estimated probability is monotone increasing in $\hat{\alpha}$. Thus, as the Democratic vote in the previous election increases, so does the estimated Democratic vote in this election.

29. a. $P(Y_{it} = 1 | \mathbf{u}_i) = \Phi(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{u}_i)$, so

$$P(Y_{it} = 1) = \int P(Z \leq \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{u}_i) f(\mathbf{u}; \boldsymbol{\Sigma}) d\mathbf{u}_i,$$

where Z is a standard normal variate that is independent of \mathbf{u}_i . Since $Z - \mathbf{z}'_{it}\mathbf{u}_i$ has a $N(0, 1 + \mathbf{z}'_{it}\boldsymbol{\Sigma}\mathbf{z}_{it})$ distribution, the probability in the integrand is $\Phi(\mathbf{x}'_{it}\boldsymbol{\beta}[1 + \mathbf{z}'_{it}\boldsymbol{\Sigma}\mathbf{z}_{it}]^{-1/2})$, which does not depend on u_i , so the integral is the same.

b. The parameters in the marginal model equal those in the GLMM divided by $[1 + \mathbf{z}'_{it}\boldsymbol{\Sigma}\mathbf{z}_{it}]^{1/2}$, which in the univariate case is $\sqrt{1 + \sigma^2}$.

31. b. Two terms drop out because $\hat{\mu}_{11} = n_{11}$ and $\hat{\mu}_{22} = n_{22}$.

c. Setting $\beta_0 = 0$, the fit is that of the symmetry model, for which $\log(\hat{\mu}_{21}/\hat{\mu}_{12}) = 0$.

35. a. For a given subject, the odds of response ($Y_{it} \leq j$) for the second observation are $\exp(\beta)$ times those for the first observation. The corresponding marginal model has a population-averaged rather than subject-specific interpretation.

b. The estimate given is the conditional ML estimate of β for the collapsed table. It is also the random effects ML estimate if the log odds ratio in that collapsed table is non-negative. The model applies to that collapsed table and those estimates are consistent for it.

c. For the $(I - 1) \times 2$ table of the off-diagonal counts from each collapsed table, the ratio in each row estimates $\exp(\beta)$. Thus, the sum of the counts across the rows gives row totals for which the ratio also estimates $\exp(\beta)$. The log of this ratio is $\tilde{\beta}$.

Chapter 13

1. Since $q = 2$, $I = 2$, and $T = 3$, the model has residual and $df = I^T - qT(I - 1) - q = 0$ and is saturated.

7. With the QL approach with beta-binomial type variance, there is mild overdispersion with $\hat{\rho} = .071$ (logit link). The intercept estimate is -0.186 ($SE = .164$). An estimate of $-.186$ for the common value of the logit corresponds to an estimate of $.454$ for the mean of the beta distribution for π_i . The estimated standard deviation of that distribution is then $\sqrt{.071(.454 \times .546)} = .133$. We estimate that Shaq's probability of success varies from game to game with a mean of $.454$ and standard deviation of $.133$.

9.a. Including litter size as a predictor, its estimate is $-.102$ with $SE = .070$. There is not strong evidence of a litter size effect. b. The $\hat{\rho}$ estimates for the four groups are $.32$, $.02$, $-.03$, and $.02$. Only the placebo group shows evidence of overdispersion.

11. The estimated constant accident rate is $\exp(-4.177) = .015$ per million miles of travel, the same as for a Poisson model. Since the dispersion parameter estimate is relatively small, the SE of $.153$ for the log rate with this model is not much greater than the SE of $.132$ for the log rate for the Poisson model in Problem 9.19.

13. The estimated difference of means is 1.552 for each model, but $SE = .196$ for the Poisson model and $SE = .665$ for the negative binomial model. The Poisson SE is not realistic because of the extreme overdispersion. Using the negative binomial model, a 95% Wald confidence interval for the difference of means is $1.552 \pm 1.96(.665)$, or $(.25, 2.86)$.

15. a. $\log \hat{\mu} = -4.05 + 0.19x$.

b. $\log \hat{\mu} = -5.78 + 0.24x$, with estimated standard deviation 1.0 for the random effect.

17. For the other group, sample mean = $.177$ and sample variance = $.442$, also showing evidence of overdispersion. The only significant difference is between whites and blacks.

25. In the multinomial log likelihood,

$$\sum n_{y_1, \dots, y_T} \log \pi_{y_1, \dots, y_T},$$

one substitutes

$$\pi_{y_1, \dots, y_T} = \sum_{z=1}^q \left[\prod_{t=1}^T P(Y_t = y_t \mid Z = z) \right] P(Z = z).$$

27. The null model falls on the boundary of the parameter space in which the weights given the two components are $(1, 0)$. For ordinary chi-squared distributions to apply, the parameter in the null must fall in the interior of the parameter space.

29. When $\theta = 0$, the beta distribution is degenerate at μ and formula (13.9) simplifies to $\binom{n}{y} \mu^y (1 - \mu)^{n-y}$.

31. $E(Y_{it}) = \pi_i = E(Y_{it}^2)$, so $\text{Var}(Y_{it}) = \pi_i - \pi_i^2$. Also, $\text{Cov}(Y_{it}Y_{is}) = \text{Corr}(Y_{it}Y_{is})\sqrt{[\text{Var}(Y_{it})][\text{Var}(Y_{is})]} = \rho\pi_i(1 - \pi_i)$. Then,

$$\text{Var}\left(\sum_i Y_{it}\right) = \sum_i \text{Var}(Y_{it}) + 2 \sum_{i < j} \text{Cov}(Y_{it}, Y_{is}) = n_i\pi_i(1 - \pi_i) + n_i(n_i - 1)\rho\pi_i(1 - \pi_i) = n_i\pi_i(1 - \pi_i)[1 + (n_i - 1)\rho].$$

33. A binary response must have variance equal to $\mu_i(1 - \mu_i)$ (See, e.g., Problem 13.31), which implies $\phi = 1$ when $n_i = 1$.

37. The likelihood is proportional to

$$\left(\frac{k}{\mu + k}\right)^{nk} \left(\frac{\mu}{\mu + k}\right)^{\sum_i y_i}$$

The log likelihood depends on μ through

$$-nk \log(\mu + k) + \sum_i y_i [\log \mu - \log(\mu + k)]$$

Differentiate with respect to μ , set equal to 0, and solve for μ yields $\hat{\mu} = \bar{y}$.

45. With a normally distributed random effect, with positive probability the Poisson mean (conditional on the random effect) is negative.

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.

13. The vector of partial derivatives, evaluated at the parameter value, is zero. Hence the asymptotic normal distribution is degenerate, having a variance of zero. Using the second-order terms in the Taylor expansion yields an asymptotic chi-squared distribution.

17.a. The vector $\partial\boldsymbol{\pi}/\partial\theta$ equals $(2\theta, 1 - 2\theta, 1 - 2\theta, -2(1 - \theta))'$. Multiplying this by the diagonal matrix with elements $[1/\theta, [\theta(1 - \theta)]^{-1/2}, [\theta(1 - \theta)]^{-1/2}, 1/(1 - \theta)]$ on the main diagonal shows that \mathbf{A} is the 4×1 vector $\mathbf{A} = [2, (1 - 2\theta)/[\theta(1 - \theta)]^{1/2}, (1 - 2\theta)/[\theta(1 - \theta)]^{1/2}, -2]'$.

b. From Problem 3.31, recall that $\hat{\theta} = (p_{1+} + p_{+1})/2$. Since $\mathbf{A}'\mathbf{A} = \mathbf{8} + (\mathbf{1} - 2\boldsymbol{\theta})^2/\theta(\mathbf{1} - \boldsymbol{\theta})$, the asymptotic variance of $\hat{\theta}$ is $(\mathbf{A}'\mathbf{A})^{-1}/\mathbf{n} = 1/n[8 + (1 - 2\theta)^2/\theta(1 - \theta)]$, which simplifies to $\theta(1 - \theta)/2n$. This is maximized when $\theta = .5$, in which case the asymptotic variance is $1/8n$. When $\theta = 0$, then $p_{1+} = p_{+1} = 0$ with probability 1, so $\hat{\theta} = 0$ with probability

1, and the asymptotic variance is 0. When $\theta = 1$, $\hat{\theta} = 1$ with probability 1, and the asymptotic variance is also 0. In summary, the asymptotic normality of $\hat{\theta}$ applies for $0 < \theta < 1$, that is when θ is not on the boundary of the parameter space. This is one of the regularity conditions that is assumed in deriving results about asymptotic distributions.

c. The asymptotic covariance matrix is $(\partial\boldsymbol{\pi}/\partial\theta)(\mathbf{A}'\mathbf{A})^{-1}(\partial\boldsymbol{\pi}/\partial\theta)' = [\theta(1-\theta)/2][2\theta, 1-2\theta, 1-2\theta, -2(1-\theta)][2\theta, 1-2\theta, 1-2\theta, -2(1-\theta)]$.

d. $df = N - t - 1 = 4 - 1 - 1 = 2$.

(Note: This model is equivalent to the Poisson loglinear model $\log m_{ij} = \mu + \lambda_i + \lambda_j$; that is, there is independence plus each marginal distribution has the same parameters, which gives marginal homogeneity. There are four cells and two parameters in the model, so the usual df formula for loglinear models tells us $df = 2$, 1 higher than for the independence model. This is a parsimonious special case of the independence model.)

23. X^2 and G^2 necessarily take very similar values when (1) the model holds, (2) the sample size n is large, and (3) the number of cells N is small compared to the sample size n , so that the expected frequencies in the cells are relatively large.

Chapter 15

17a. From Sec. 15.2.1, the Bayes estimator is $(n_1 + \alpha)/(n + \alpha + \beta)$, in which $\alpha > 0, \beta > 0$. No proper prior leads to the ML estimate, n_1/n .

b. The ML estimator is the limit of Bayes estimators as α and β both converge to 0.

c. This happens with the improper prior, proportional to $[\pi_1(1 - \pi_1)]^{-1}$, which we get from the beta density by taking the improper settings $\alpha = \beta = 0$.