CATEGORICAL DATA ANALYSIS, 3rd edition

Solutions to Selected Exercises

Alan Agresti

This file contains solutions and hints to solutions for some of the exercises in *Categorical Data Analysis*, third edition, by Alan Agresti (John Wiley, & Sons, 2012). The solutions given are partly those that are also available at the website `www.stat.ufl.edu/~aa/cda2/cda.html` for many of the odd-numbered exercises in the second edition of the book (some of which are now even-numbered). I intend to expand the document with additional solutions, when I have time.

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 exercises not in this file.

Chapter 1

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

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

7. a. $\ell(\pi) = \pi^{20}$, so it is not close to quadratic.
b. $\hat{\pi} = 1.0$. 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). These are not sensible.
c. $z = (1.0 - .5)/\sqrt{.5(.5)/20} = 4.47$, $P < 0.0001$. Score CI is (0.839, 1.000).
d. Test statistic $2(20)\log(20/10) = 27.7, df = 1$. The CI is $(\exp(-1.96^2/40), 1) = (0.908, 1.0)$.
e. P-value $= 2(.5)^{20} = .00000191$.

9. The chi-squared goodness-of-fit test of the null hypothesis that the binomial proportions equal (0.75, 0.25) has expected frequencies (827.25, 275.75), and $X^2 = 3.46$ based on $df = 1$. The P-value is 0.063, giving moderate evidence against the null.

10. 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 we truncate at 3 and above). The fit seems adequate.

11. 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 0.05, it is impossible to reject H_0, and thus P(Type I error) = 0. With the large-sample score test, $y = 0$ and $y = 5$ are the only outcomes to give $P \leq 0.05$ (e.g., with $y = 5$, $z = (1.0 - 0.5)/\sqrt{0.5(0.5)/5} = 2.24$ and $P = 0.025$). Thus, for that test, P(Type I error) $= P(Y = 0) + P(Y = 5) = 1/16$.
12. a. No outcome can give $P \leq .05$, and hence one never rejects H_0.
b. When $T = 2$, mid P-value = 0.04 and one rejects H_0. Thus, P(Type I error) = P(T = 2) = 0.08.
c. P-values of the two tests are 0.04 and 0.02; P(Type I error) = P(T = 2) = 0.04 with both tests.
d. P(Type I error) = E[P(Type I error | T)] = (5/8)(0.08) = 0.05. Randomized tests are not sensible for practical application.

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

17. a. $\text{Var}(Y) = n\pi(1-\pi)$, binomial.

b. $\text{Var}(Y) = \sum \text{Var}(Y_i) + 2\sum_{i<j} \text{Cov}(Y_i, Y_j) = n\pi(1-\pi) + 2\rho\pi(1-\pi)\binom{n}{2} > n\pi(1-\pi)$.

c. $\text{Var}(Y) = E[\text{Var}(Y|\pi)] + \text{Var}[E(Y|\pi)] = E[n\pi(1-\pi)] + \text{Var}(n\pi) = n\rho - nE(\pi^2) + [n^2E(\pi^2) - n^2\rho^2] = n\rho + (n^2 - n)[E(\pi^2) - \rho^2] - n\rho^2 = n\rho(1-\rho) + (n^2-n)Var(\pi) > n\rho(1-\rho)$.

18. 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.

19. Using results shown in Sec. 16.1, $\text{Cov}(n_j, n_k)/\sqrt{\text{Var}(n_j)\text{Var}(n_k)} = -n\pi_j\pi_k/\sqrt{n\pi_j(1-\pi_j)n\pi_k(1-\pi_k)}$. When $c = 2$, $\pi_1 = 1 - \pi_2$ and correlation simplifies to -1.

20. a. 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 $m'(0) = n\pi$.

21. $t_o = -2\log[(\text{prob. under } H_0)/(\text{prob. under } H_a)]$, so (prob. under H_0)/(prob. under H_a) $= \exp(-t_o/2)$.

22. 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 the LR statistic $\leq \chi_1^2(\alpha)$.

23. Conditional on $n = y_1 + y_2$, y_1 has a *bin*(n, π) distribution with $\pi = \mu_1/(\mu_1 + \mu_2)$, which is 0.5 under H_0. The large sample score test uses $z = (y_1/n - 0.5)/\sqrt{0.5(0.5)/n}$. If (ℓ, u) denotes a CI for π (e.g., the score CI), then the CI for $\pi/(1-\pi) = \mu_1/\mu_2$ is $[\ell/(1-\ell), u/(1-u)]$.

25. $g(\pi) = \pi(1-\pi)/n^*$ is a concave function of π, so if π is random, $g(E\pi) \geq Eg(\pi)$ by Jensen's inequality. Now $\tilde{\pi}$ is the expected value of π for a distribution putting probability $n/(n + z_{\alpha/2}^2)$ at $\hat{\pi}$ and probability $z_{\alpha/2}^2/(n + z_{\alpha/2}^2)$ at 1/2.

26. a. The likelihood-ratio (LR) CI is the set of π_0 for testing H_0: $\pi = \pi_0$ such that LR statistic $= -2\log[(1-\pi_0)^n/(1-\hat{\pi})^n] \leq z^2_{\alpha/2}$, with $\hat{\pi} = 0.0$. Solving for π_0, $n\log(1-\pi_0) \geq -z^2_{\alpha/2}/2$, or $(1-\pi_0) \geq \exp(-z^2_{\alpha/2}/2n)$, or $\pi_0 \leq 1 - \exp(-z^2_{\alpha/2}/2n)$. Using $\exp(x) = 1 + x + ...$ for small x, the upper bound is roughly $1 - (1 - z^2_{0.025}/2n) = z^2_{0.025}/2n = 1.96^2/2n \approx 2^2/2n = 2/n$.
b. Solve for $(0-\pi)/\sqrt{\pi(1-\pi)/n} = -z_{\alpha/2}$.

27. If we form the P-value using the right tail, then mid P-value $= \pi_j/2 + \pi_{j+1} + \cdots$. Thus, $E(\text{mid } P\text{-value}) = \sum_j \pi_j(\pi_j/2 + \pi_{j+1} + \cdots) = (\sum_j \pi_j)^2/2 = 1/2$.

28. The right-tail mid P-value equals $P(T > t_o) + (1/2)p(t_o) = 1 - P(T \leq t_o) + (1/2)p(t_o) = 1 - F_{mid}(t_o)$.

29. 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 θ.
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.

30. 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 $\sqrt{\pi(1-\pi)/n(1+\pi)}$.

32. 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, $X^2 = (n_1 - n\pi_0)^2/n\pi_0 + (n_2 - n(1-\pi_2))^2/n(1-\pi_0)$
$= n[(\hat{\pi}-\pi_0)^2(1-\pi_0) + ((1-\hat{\pi}) - (1-\pi_0))^2\pi_0]/\pi_0(1-\pi_0)$,
which equals $(\hat{\pi}-\pi_0)^2/[\pi_0(1-\pi_0)/n] = z_S^2$.

33. Let X be a random variable that equals $\pi_{j0}/\hat{\pi}_j$ 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 $G^2 = 2nE(-\log X) \geq 0$.

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 *mgf*s, which is $m(t) = (1-2t)^{-(\nu_1+\nu_2)/2}$, which is the *mgf* of a χ^2 with *df* $= \nu_1 + \nu_2$.

40a. 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. The ML estimator is the limit of Bayes estimators as α and β both converge to 0.
b. 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 $\alpha = \beta = 0$.

Chapter 2

3. $P(-|C) = 1/4$. 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.

5. a. Relative risk.
b. (i) $\pi_1 = 0.55\pi_2$, so $\pi_1/\pi_2 = 0.55$.
(ii) $1/0.55 = 1.82$.

11. a. $(0.847/0.153)/(0.906/0.094) = 0.574$.
b. This is interpretation for relative risk, not the odds ratio. The actual relative risk = $0.847/0.906 = 0.935$; i.e., 60% should have been 93.5%.

12. 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, 0.00130; Heart disease, 0.00256. (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 $0.00130N$ fewer deaths per year from lung cancer if they had never smoked, and $0.00256N$ fewer deaths per year from heart disease. Thus elimination of cigarette smoking would have biggest impact on deaths due to heart disease.

15. Marginal odds ratio = 1.84, but most conditional odds ratios are close to 1.0 except in Department A where odds ratio = 0.35. 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 aply 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.

17. a. 0.18 for males and 0.32 for females; e.g., for male children, the odds that a white was a murder victim were 0.18 times the odds that a nonwhite was a murder victim.
b. 0.21.

18. The age distribution is relatively higher in Maine. Death rates are higher at older ages, and Maine tends to have an older population than South Carolina.

19. 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 0.0, 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. 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.

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

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

b. $0.95(0.005)/[0.95(0.005) + 0.05(0.995)] = 0.087$.

		Test		
		+	−	Total
Reality	+	0.00475	0.00025	0.005
	−	0.04975	0.94525	0.995

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+.

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

29. a. For instance, if first row becomes first column and second row becomes second column, the table entries become

$$\begin{matrix} n_{11} & n_{21} \\ n_{12} & n_{22} \end{matrix}$$

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.

30. 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 $\theta = [\pi_1/(1 - \pi_1)]/[\pi_2/(1 - \pi_2)] < \pi_1/\pi_2 < 1$.

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

36. 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. Use an argument similar to that in Sec. 1.2.5. Since Y_{i+} is sum of independent Poissons, it is Poisson. In the denominator for the calculation of the conditional probability, the distribution of $\{Y_{i+}\}$ is a product of Poissons with means $\{\mu_{i+}\}$. The multinomial distributions are obtained by identifying $\pi_{j|i}$ with μ_{ij}/μ_{i+}.

40. If in each row the maximum probability falls in the same column, say column 1, then

$E[V(Y \mid X)] = \sum_i \pi_{i+}(1-\pi_{1|i}) = 1-\pi_{+1} = 1-\max\{\pi_{+j}\}$, 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

14. b. Compare rows 1 and 2 ($G^2 = 0.76$, $df = 1$, no evidence of difference), rows 3 and 4 ($G^2 = 0.02$, $df = 1$, 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).

16.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 $M^2 = 4.75$, $df = 1$, $P = 0.03$, and much stronger evidence of an association.

18. a. It is plausible that control of cancer is independent of treatment used. (i) P-value is hypergeometric probability $P(n_{11} = 21 \text{ or } 22 \text{ or } 23) = 0.3808$, (ii) P-value $= 0.638$ is sum of probabilities that are no greater than the probability (0.2755) of the observed table.
b. 0.3808 - 0.5(0.2755) = 0.243. 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 0.50; however, it does not guarantee that the actual error probability is no greater than the nominal value.

25. 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 $\pi < 0.50$ and greater than 0 for $\pi > 0.50$. Thus, the minimum is with proportions (0.5, 0.5) in the two categories.

29. Use formula (3.9), noting that the partial derivative of the measure with respect to π_i is just η_i/δ^2.

30. 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,d. 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.

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

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

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.

37. $\sum_j p_{+j}\hat{r}_j = \sum_j p_{+j}[\sum_{k<j} p_{+k} + p_{+j}/2] = \sum_S p_{+j}p_{+k} + \sum_j p_{+j}^2/2$, where $S = \{(j,k) : k < j\}$. This equals

$$(1/2)[\sum p_{+j}^2 + 2\sum_S p_{+j}p_{+k}] = (1/2)(\sum p_{+j})^2 = (1/2)(1)^2 = 0.50.$$

Also,

$$\sum p_{i+}\hat{R}_i = \sum p_{i+}(\sum_j \hat{r}_j p_{j|i}) = \sum_i \sum_j \hat{r}_j p_{ij} = \sum_j \hat{r}_j p_{+j} = 0.50.$$

43. The observed table has $X^2 = 6$. The probability of this table is highest at $\pi = 0.50$. 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.

Chapter 4

1. a. Roughly 3%.
b. Estimated proportion $\hat{\pi} = -0.0003 + 0.0304(0.0774) = 0.0021$. The actual value is 3.8 times the predicted value, which together with Fig. 4.7 suggests it is an outlier.
2. a. The estimated probability of malformation increases from 0.0011 at $x = 0$ to 0.0025 + 0.0011(7.0) = 0.0102 at $x = 7$. The relative risk is 0.0102/0.0025 = 4.1.

3. The fit of the linear probability model is (a) 0.018 + 0.018(snoring), (b) 0.018 + 0.036(snoring), (c) -0.019 + 0.036(snoring). Slope depends on distance between scores; doubling the distance halves the slope estimate. The fitted values and goodness-of-fit statistics are identical for any linear transformation.

4. a. $\hat{\pi}$ = -0.145 + 0.323(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(weight); at 5.2 kg, predicted logit = 5.74, and log(0.9968/0.0032) = 5.74.
6. a. a. $0.5893 \pm 1.96(0.0650) = (0.4619, 0.7167)$.

b. $(0.5893/0.0650)^2 = 82.15$.
Need log likelihood value when $\beta = 0$.
c. Multiply standard errors by $\sqrt{535.896/171} = 1.77$. There is still very strong evidence of a positive weight effect.

7. a. $\log(\hat{\mu}) = 1.6094 + 0.5878x$. Since $\beta = \log(\mu_B/\mu_A)$, $\exp(\hat{\beta}) = \hat{\mu}_B/\hat{\mu}_A = 1.80$; i.e., the mean is predicted to be 80% higher for treatment B. (In fact, this estimate is simply the ratio of sample means.)
b. Wald test gives $z = 0.588/0.176 = 3.33$, $z^2 = 11.1$ ($df = 1$), $P < 0.001$. Likelihood-ratio statistic equals 27.86 - 16.27 = 11.6 with $df = 1$, $P < 0.001$. There is strong evidence against H_0 and of a higher defect rate for treatment B.
c. Exponentiate 95% CI for β of $0.588 \pm 1.96(0.176)$ to get Wald CI of $\exp(0.242, 0.934) = (1.27, 2.54)$.

d. Normal approximation to binomial yields $z = (50 - 70)/\sqrt{140(0.5)(0.5)} = -3.4$ and very strong evidence against H_0.

9. Since $\exp(0.192) = 1.21$, 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.

11.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 = -0.309, with std. error = 0.097; the estimated ratio of rates is multiplied by $\exp(-0.309) = 0.73$ for each successive increase of one age category.

15. 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 estimated 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.

16. 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 $\alpha + \beta x < 0$.

17. 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. 121 has mean μ and standard deviation $\tau\pi/\sqrt{3}$. Writing $\alpha + \beta x$ in the

form $(x-\alpha/\beta)/(1/\beta)$, we identify μ with α/β and τ with $1/\beta$, so the standard deviation is $\pi/\beta\sqrt{3}$ when $\beta > 0$.

19. 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 $\hat{\mu}_B = \bar{y}_B$.

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

22. a. Since ϕ is symmetric, $\Phi(0) = 0.5$. 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 0.25 at $x = 0$; the standard normal pdf equals $1/\sqrt{2\pi}$ at $x = 0$.
c. $\Phi(\alpha + \beta x) = \Phi(\frac{x-(-\alpha/\beta)}{1/\beta})$.

23. a. Cauchy. You can see this by taking the derivative and noting it has the form of a Cauchy density. The GLM with Bernoulli random component, systematic component $\alpha + \beta x$, link function tan[pi($\pi(x) - 1/2$)] (where pi = 3.14...), would work well when the rate of convergence of π to 0 and 1 is slower than with the logit or probit link (Recall that Cauchy density has thick tails compared to logistic and normal densities).

26. 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 $= 2\sum_i \sum_j [y_{ij} \log(y_{ij}/\bar{y}_i)$.

31. 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 $\mu^{(t+1)} = \bar{y}$.

Chapter 5

2. a. $\hat{\pi} = e^{-3.7771+0.1449(8)}/[1 + e^{-3.7771+0.1449(8)}]$.
b. $\hat{\pi} = 0.5$ at $-\hat{\alpha}/\hat{\beta} = 3.7771/0.1449 = 26$.
c. At LI = 8, $\hat{\pi} = 0.068$, so rate of change is $\hat{\beta}\hat{\pi}(1 - \hat{\pi}) = 0.1449(0.068)(0.932) = 0.009$.
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 $= (0.1449/0.0593)^2 = 5.96$, $df = 1$, P-value $= 0.0146$ for H_a:$\beta \neq 0$.
h. Likelihood-ratio statistic = 34.37 - 26.07 = 8.30, $df = 1$, P-value = 0.004.

5. a. At 26.3, estimated odds $= \exp[-12.351 + 0.497(26.3)] = 2.06$, and at 27.3 the estimated odds $= \exp[-12.351 + 0.497(27.3)] = 3.38$, and $3.38 = 1.64(2.06)$. For each 1-unit increase in x, the odds multiply by 1.64 (i.e., increase by 64%).
b. The approximate rate of change when $\pi = 0.5$ is $\beta\pi(1-\pi) = \beta/4$. The 95% Wald CI for β of (0.298, 0.697) translates to one for $\beta/4$ of (0.07, 0.17).

7. $\text{logit}(\hat{\pi})$ = -3.866 + 0.397(snoring). Fitted probabilities are 0.021, 0.044, 0.093, 0.132. 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.

9. 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.

11. Estimated odds of contraceptive use for those with at least 1 year of college were $e^{0.501} = 1.65$ times the estimated odds for those with less than 1 year of college. The 95% Wald CI for the true odds ratio is $\exp[0.501 \pm 1.96(0.077)] = (e^{0.350}, e^{0.652}) = (1.42, 1.92)$.

14. The original variables c and x relate to the standardized variables z_c and z_x by $z_c = (c - 2.44)/0.80$ and $z_x = (x - 26.3)/2.11$, so that $c = 0.80z_c + 2.44$ and $x = 2.11z_x + 26.3$. Thus, the prediction equation is
$\text{logit}(\hat{\pi}) = -10.071 - 0.509[0.80z_c + 2.44] + 0.458[2.11z_x + 26.3]$,
The coefficients of the standardized variables are -0.509(0.80) = -0.41 and 0.458(2.11) = 0.97. Adjusting 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 -0.062, which correspond to estimated probabilities of 0.81 and 0.48.

15. a. Black defendants with white victims had estimated probability $e^{-3.5961+2.4044}/[1 + e^{-3.5961+2.4044}] = 0.23$.
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 $(-0.8678/0.3671)^2 = 5.6$, LR statistic = 5.0, each with $df = 1$. P-value = 0.025 for LR statistic.
d. $G^2 = 0.38$, $X^2 = 0.20$, $df = 1$, so model fits well.

19. $R = 1$: $\text{logit}(\hat{\pi}) = -6.7 + 0.1A + 1.4S$. $R = 0$: $\text{logit}(\hat{\pi}) = -7.0 + 0.1A + 1.2S$.
The YS conditional odds ratio is exp(1.4) = 4.1 for blacks and exp(1.2) = 3.3 for whites. Note that 0.2, 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 $P < 0.01$ for smoking represents the result of the test that the log odds ratio between Y and S for whites is 0.

22. Logit model gives fit, $\text{logit}(\hat{\pi})$ = -3.556 + 0.053(income).

25. The derivative equals $\beta \exp(\alpha + \beta x)/[1 + \exp(\alpha + \beta x)]^2 = \beta\pi(x)(1 - \pi(x))$.

26. 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
$e^\beta = [\pi(x+1)/(1-\pi(x+1))]/[\pi(x)/(1-\pi(x))] \approx \pi(x+1)/\pi(x)$.

27. $\partial\pi(x)/\partial x = \beta\pi(x)[1-\pi(x)]$, and $\pi(1-\pi) \leq 0.25$ with equality at $\pi = 0.5$. For multiple explanatory variable case, the rate of change as x_i changes with other variables held constant is greatest when $\pi = 0.5$.

28. 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.

29. 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^α equals the odds for the first draft pick.
b. As a function of d, the odds decreases more quickly for pro basketball.

30. a. Let ρ = 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 $\alpha = -\log[(1-\rho)/\rho] + [\mu_0^2-\mu_1^2]/2\sigma^2$.

32. 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.

35. 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
$D = -2\sum_i[y_i\log\hat{\pi}_i + (1-y_i)\log(1-\hat{\pi}_i)] = -2\sum_i[y_i\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) + \log(1-\hat{\pi}_i)]$
$= -2\sum_i[y_i(\hat{\alpha}+\hat{\beta}x_i) + \log(1-\hat{\pi}_i)]$.
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
$D = -2[\hat{\alpha}\sum_i\hat{\pi}_i + \hat{\beta}\sum_i x_i\hat{\pi}_i + \sum_i\log(1-\hat{\pi}_i)]$
$= -2[\sum_i\hat{\pi}_i(\hat{\alpha}+\hat{\beta}x_i) + \sum_i\log(1-\hat{\pi}_i)]$
$= -2\sum_i\hat{\pi}_i\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) - 2\sum_i\log(1-\hat{\pi}_i)$.

40. a. Expand $\log[p/(1-p)]$ in a Taylor series for a neighborhood of points around p = π, and take just the term with the first derivative.
b. Let $p_i = y_i/n_i$. The ith sample logit is

$$\log[p_i/(1-p_i)] \approx \log[\pi_i^{(t)}/(1-\pi_i^{(t)})] + (p_i-\pi_i^{(t)})/\pi_i^{(t)}(1-\pi_i^{(t)})$$

$$= \log[\pi_i^{(t)}/(1-\pi_i^{(t)})] + [y_i - n_i\pi_i^{(t)}]/n_i\pi_i^{(t)}(1-\pi_i^{(t)})$$

Chapter 6

1. logit($\hat{\pi}$) = -9.35 + 0.834(weight) + 0.307(width).
a. Like. ratio stat. = 32.9 (df = 2), $P < 0.0001$. There is extremely strong evidence that at least one variable affects the response.
b. Wald statistics are $(0.834/0.671)^2 = 1.55$ and $(0.307/0.182)^2 = 2.85$. These each have df = 1, and the P-values are 0.21 and 0.09. These predictors are highly correlated (Pearson corr. = 0.887), so this is the problem of multicollinearity.

12. The estimated odds of admission were 1.84 times higher for men than women. However, $\hat{\theta}_{AG(D)} = 0.90$, so given department, the estimated odds of admission were 0.90 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. 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).

17. The CMH statistic simplifies to the McNemar statistic of Sec. 11.1, 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 = 0.074 for the two-sided alternative).

27. logit($\hat{\pi}$) $= -12.351 + 0.497x$. Prob. at $x = 26.3$ is 0.674; prob. at $x = 28.4$ (i.e., one std. dev. above mean) is 0.854. The odds ratio is $[(0.854/0.146)/(0.674/0.326)] = 2.83$, so $\lambda = 1.04$, $\delta = 5.1$. Then $n = 75$.

31. 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.

35. 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.

41. a. $E(Y) = \alpha + \beta_1 X + \beta_2 Z$. The slope β_1 of the line for the partial relationship between $E(Y)$ and X is the same at all fixed levels of Z.

b. With dummy (indicator) variables for Z, one has parallelism of lines. That is, the slope of the line relating $E(Y)$ and X is the same for each category of Z.
c. Use dummy variables for X and Z, but no interaction terms. The difference between $E(Y)$ at two categories of X is the same at each fixed category of Z.
d. For logistic models, the odds ratio relating Y and X is the same at each category of Z.

Chapter 7

21. Log likelihood for the probit model is

$$\log\left\{\prod_{i=1}^{N}\left[\Phi\Big(\sum_j \beta_j x_{ij}\Big)\right]^{y_i}\left[1-\Phi\Big(\sum_j \beta_j x_{ij}\Big)\right]^{1-y_i}\right\}$$

$$=\sum_i y_i \log\left[\Phi\Big(\sum_j \beta_j x_{ij}\Big)\right]+\sum_i (1-y_i)\log\left[1-\Phi\Big(\sum_j \beta_j x_{ij}\Big)\right]$$

$$=\sum_i y_i \log\left[\frac{\Phi\big(\sum_j \beta_j x_{ij}\big)}{1-\Phi\big(\sum_j \beta_j x_{ij}\big)}\right]+\sum_i \log\left[1-\Phi\Big(\sum_j \beta_j x_{ij}\Big)\right]$$

For the probit model,

$$\frac{\partial L}{\partial \beta_j}=\sum_i y_i\left[\frac{1-\Phi\big(\sum_j \beta_j x_{ij}\big)}{\Phi\big(\sum_j \beta_j x_{ij}\big)}\right]\frac{\left[\phi\big(\sum_j \beta_j x_{ij}\big)x_{ij}\big(1-\Phi\big(\sum_j \beta_j x_{ij}\big)\big)+x_{ij}\phi\big(\sum_j \beta_j x_{ij}\big)\Phi\big(\sum_j \beta_j x_{ij}\big)\right]}{\left[1-\Phi\big(\sum_j \beta_j x_{ij}\big)\right]^2}$$

$$-\sum_i \frac{x_{ij}\phi\big(\sum_j \beta_j x_{ij}\big)}{1-\Phi\big(\sum_j \beta_j x_{ij}\big)} \quad = \quad 0$$

$$\Longrightarrow \quad \sum_i \frac{y_i x_{ij}\phi\big(\sum_j \hat\beta_j x_{ij}\big)}{\left[1-\Phi\big(\sum_j \hat\beta_j x_{ij}\big)\right]\Phi\big(\sum_j \hat\beta_j x_{ij}\big)}-\sum_i \frac{x_{ij}\phi\big(\sum_j \hat\beta_j x_{ij}\big)}{1-\Phi\big(\sum_j \hat\beta_j x_{ij}\big)} \quad = \quad 0$$

$$\Longrightarrow \quad \sum_i \frac{y_i x_{ij}\phi\big(\sum \hat\beta_j x_{ij}\big)}{\hat\pi_i(1-\hat\pi_i)}-\sum_i \frac{x_{ij}\phi\big(\sum_j \hat\beta_j x_{ij}\big)\hat\pi_i}{\hat\pi_i(1-\hat\pi_i)} \quad = \quad 0$$

$$\Longrightarrow \quad \sum_i \big(y_i-\hat\pi_i\big)x_{ij}z_i \quad = \quad 0, \quad \text{where} \quad z_i=\phi(\Sigma_j\hat\beta_j x_{ij})/\hat\pi_i(1-\hat\pi_i).$$

For logistic regression, from (4.28) with $\{n_i=1\}$, $\sum_i \big(y_i-\hat\pi_i\big)x_{ij}=0$.

25. $(\log\pi(x_2))/(\log\pi(x_1))=\exp[\beta(x_2-x_1)]$, 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 8

3. Both gender and race have significant effects. The logistic 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.

7. For any collapsing of the response, for Democrats the estimated odds of response in the liberal direction are $\exp(0.975) = 2.65$ times the estimated odds for Republicans. The estimated probability of a very liberal response equals $\exp(-2.469)/[1+\exp(-2.469)] = 0.078$ for Republicans and $\exp(-2.469+0.975)/[1+\exp(-2.469+0.975)] = 0.183$ for Democrats.

8. 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 $\exp(3.3074)/[1+\exp(3.3074)] = 0.965$, $\exp(3.4818)/[1+\exp(3.4818)] = 0.970$, $\exp(5.3494)/[1+\exp(5.3494)] = 0.995$, $\exp(7.2563)/[1+\exp(7.2563)] = 0.9993$, and 1.0. The corresponding response probabilities are 0.965, 0.005, 0.025, 0.004, and 0.0007.
b. Wald CI is $\exp[-0.5463 \pm 1.96(0.0272)] = (\exp(-0.600), \exp(-0.493)) = (0.549, 0.611)$. Give seat belt use and location, the estimated odds of injury below any fixed level for a female are between 0.549 and 0.611 times the estimated odds for a male.
c. Estimated odds ratio equals $\exp(-0.7602 - 0.1244) = 0.41$ in rural locations and $\exp(-0.7602) = 0.47$ in urban locations. The interaction effect -0.1244 is the difference between the two log odds ratios.

10. a. Setting up indicator variables (1,0) for (male, female) and (1,0) for (sequential, alternating), we get treatment effect = -0.581 ($SE = 0.212$) and gender effect = -0.541 ($SE = 0.295$). The estimated odds ratios are 0.56 and 0.58. 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 0.56 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 $G^2 = 4.5$, $df = 6$).

15. The estimated odds a Democrat is classified in the more liberal instead of the more conservative of two adjacent categories are $\exp(0.435) = 1.54$ times the estimated odds for a Republican. For the two extreme categories, the estimated odds ratio equals $\exp[4(0.435)] = 5.7$.

17.a. Using scores 3.2, 3.75, 4.5, 5.2, the proportional odds model has a treatment effect of 0.805 with $SE = 0.206$; for the treatment group, the estimated odds that ending cholesterol is below any fixed level are $\exp(0.805) = 2.24$ times the odds for the control group. The psyllium treatment seems to have had a strong, beneficial effect.

18. CMH statistic for correlation alternative, using equally-spaced scores, equals 6.3 ($df = 1$) and has P-value = 0.012. 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. LR statistic for cumulative logit model with linear effect of operation = 6.7, $df = 1$, $P =$

0.01; strong evidence that operation has an effect on dumping, gives similar results as in (a). LR statistic comparing this model to model with four separate operation parameters equals 2.8 (df = 3), so simpler model is adequate.

29. The multinomial mass function factors as the multinomial coefficient times $\pi_J^n \exp[\sum_{i=1}^{J-1} n_i \log(\pi_i/\pi_J)]$, which has the form a function of the data times a function of the parameters (namely $(1 - \pi_1 - ... - \pi_{J-1})^n$) times an exponential function of a sum of the observations times the canonical parameters, which are the baseline-category logits.

32. $\partial\pi_3(x)/\partial x = \frac{-[\beta_1 \exp(\alpha_1+\beta_1 x)+\beta_2 \exp(\alpha_2+\beta_2 x)]}{[1+\exp(\alpha_1+\beta_1 x)+\exp(\alpha_2+\beta_2 x)]^2}$.
a. The denominator is positive, and the numerator is negative when $\beta_1 > 0$ and $\beta_2 > 0$.

36. The baseline-category logit model refers to individual categrories rather than cumulative probabilities. There is not linear structure for baseline-category logits that implies identical effects for each cumulative logit.

37. a. 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.

39. a. $df = I(J-1) - [(J-1) + (I-1)] = (I-1)(J-2)$.
c. The full model has an extra $I - 1$ parameters.
d. 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 $\mu_a < \mu_b$.

43. 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 9

1. G^2 values are 2.38 (df = 2) for (GI, HI), and 0.30 (df = 1) for (GI, HI, GH).
b. Estimated log odds ratios is -0.252 (SE = 0.175) for GH association, so CI for odds ratio is $\exp[-0.252 \pm 1.96(0.175)]$. Similarly, estimated log odds ratio is 0.464 (SE = 0.241) for GI association, leading to CI of $\exp[0.464 \pm 1.96(0.241)]$. 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.

4. 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. $G^2 = 31.7$, 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. $\log(\mu_{11cl}\mu_{33cl}/\mu_{13cl}\mu_{31cl}) = \log(\mu_{11cl}) + \log(\mu_{33cl}) - \log(\mu_{13cl}) - \log(\mu_{31cl})$.
Substitute model formula, and simplify. The estimated odds ratio equals $\exp(2.142) = 8.5$. 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 H to be too much than for those who judge spending on H to be too low. The 95% CI is $\exp[2.142 \pm 1.96(0.523)]$, or (3.1, 24.4.). Though it is very wide, it is clear that the true association is strong.

7. 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 0.091 for S and E (i.e., wearers of seat belts are much less likely to be ejected), 5.57 for S and I, and 0.061 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).

8. Injury has estimated conditional odds ratios 0.58 with gender, 2.13 with location, and 0.44 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 0.58 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 0.44 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. (DVF, YD, YV, YF).
b. Model with Y as response and additive factor effects for D and V, $\text{logit}(\pi) = \alpha + \beta_i^D + \beta_j^V$.
c. (i) (DVF, Y), $\text{logit}(\pi) = \alpha$, (ii) (DVF, YF), $\text{logit}(\pi) = \alpha + \beta_i^F$,
(iii) (DVF, YDV, YF), add term of form β_{ij}^{DV} to logit model.

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 $\exp(1.147 \pm 1.645(0.153)) = (2.45, 4.05)$.
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

$$\begin{aligned}
\lambda &= \log(\mu_{111}), \quad \lambda_i^X = \log\left(\frac{\mu_{i11}}{\mu_{111}}\right), \quad \lambda_{ij}^{XY} = \log\left(\frac{\mu_{ij1}\mu_{111}}{\mu_{i11}\mu_{1j1}}\right) \\
\lambda_{ijk}^{XYZ} &= \log\left(\frac{[\mu_{ijk}\mu_{11k}/\mu_{i1k}\mu_{1jk}]}{[\mu_{ij1}\mu_{111}/\mu_{i11}\mu_{1j1}]}\right)
\end{aligned}$$

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. 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).
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	
	0.125	0.125	0.125	0.125
X	0.125	0.125	0.125	0.125
	Z = 1		Z = 2	

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

0.15	0.10	0.15	0.10
0.10	0.15	0.10	0.15

c.

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

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

23. Number of terms $= 1 + \binom{T}{1} + \binom{T}{2} + \ldots + \binom{T}{T} = \sum_i \binom{T}{i} 1^i 1^{T-i} = (1+1)^T$, 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. For independent Poisson sampling,

$$L = \sum_i \sum_j n_{ij} \log \mu_{ij} - \sum_i \sum_j \mu_{ij} = n\lambda + \sum_i n_{i+}\lambda_i^X + \sum_j n_{+j}\lambda_j^Y - \sum_i \sum_j \exp(\log \mu_{ij})$$

It follows that $\{n_{i+}\}$, $\{n_{+j}\}$ are minimal sufficient statistics, and the likelihood equations are $\hat{\mu}_{i+} = n_{i+}$, $\hat{\mu}_{+j} = n_{+j}$ for all i and j. Since the model is $\mu_{ij} = \mu_{i+}\mu_{+j}/n$, the fitted values are $\hat{\mu}_{ij} = \hat{\mu}_{i+}\hat{\mu}_{+j}/n = n_{i+}n_{+j}/n$. The residual degrees of freedom are $IJ - [1 + (I-1) + (J-1)] = (I-1)(J-1)$.

28. 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. 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* = [no. categories of (XY)-1][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* $= (HIJ-1)(K-1)$.

Chapter 10

2. 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.

7. 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.

19. 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.

20. 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.

21. b. Using the Haberman result, it follows that

$$\begin{aligned}\sum \hat{\mu}_{1i} \log(\hat{\mu}_{0i}) &= \sum \hat{\mu}_{0i} \log(\hat{\mu}_{0i}) \\ \sum n_i \log(\hat{\mu}_{ai}) &= \sum \hat{\mu}_{ai} \log(\hat{\mu}_{ai}), \qquad a = 0, 1.\end{aligned}$$

The first equation is obtained by letting $\{\hat{\mu}_i\}$ be the fitted values for M_0. The second pair of equations is obtained by letting M_1 be the saturated model. Using these, one can obtain the result.

25. 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.

27. 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,

$$\sum_i \sum_j \pi_{ij}\phi_{ij}^2 = \sum_i \sum_j u_i^2 v_j^2 \pi_{i+}\pi_{+j} + (\sum_j v_j \pi_{+j})^2 (\sum_i u_i^2 \pi_{i+}) + (\sum_i u_i \pi_{i+})^2 (\sum_j v_j^2 \pi_{+j})$$

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

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 $\sigma/\sqrt{n}$, the estimate of which is the same formula with π_{ij} replaced by p_{ij}.

28. For Poisson sampling, log likelihood is

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

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, $\partial L / \partial \mu_i = \sum_j v_j n_{ij} - \sum_j v_j \mu_{ij}$, $i = 1, ..., I$, from which follows the I equations in the third set of likelihood equations.

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

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

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* $= K(IJ - I - J)$).

31. Deleting the XY superscript to simplify notation,

$$\log \theta_{ij(k)} = (\lambda_{ij} + \lambda_{i+1,j+1} - \lambda_{i,j+1} - \lambda_{i+1,j}) + \beta(u_{i+1} - u_i)(v_{j+1} - v_j) w_k.$$

This has form $\alpha_{ij} + \beta_{ij} w_k$, a linear function of the scores for the levels of Z. Thus, the conditional association between X and Y changes linearly across the levels of Z.

36. 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.

37. That value for the sufficient statistic becomes more likely as the model parameter moves toward infinity.

Chapter 11

6. 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 0.005 and provides strong evidence that the response rate of successes is higher for drug A.
b. Pearson statistic = 7.8, df = 1

7. b. $z^2 = (3-1)^2/(3+1) = 1.0 =$ CMH statistic.
e. The P-value equals the binomial probability of 3 or more successes out of 4 trials when the success probability equals 0.5, which equals 5/16.

9. a. Symmetry has $G^2 = 22.5$, $X^2 = 20.4$, with $df = 10$. The lack of fit results primarily from the discrepancy between n_{13} and n_{31}, for which the adjusted residual is $(44-17)/\sqrt{44+17} = 3.5$.
b. Compared to quasi symmetry, $G^2(S \mid QS) = 22.5 - 10.0 = 12.5$, $df = 4$, for a P-value of .014. The McNemar statistic for the 2×2 table with row and column categories (High Point, Others) is $z = (78-42)/\sqrt{78+42} = 3.3$. The 95% CI comparing the proportion choosing High Point at the two times is 0.067 ± 0.039.
c. Quasi independence fits much better than independence, which has $G^2 = 346.4$ ($df = 16$). Given a change in brands, the new choice of coffee brand is plausibly independent of the original choice.

12. a. Symmetry model has $X^2 = 0.59$, based on $df = 3$ ($P = 0.90$). Independence has $X^2 = 45.4$ ($df = 4$), and quasi independence has $X^2 = 0.01$ ($df = 1$) and is identical to quasi symmetry. The symmetry and quasi independence models fit well.
b. $G^2(S \mid QS) = 0.591 - 0.006 = 0.585$, $df = 3 - 1 = 2$. Marginal homogeneity is plausible.
c. Kappa = 0.389 ($SE = 0.060$), weighted kappa equals 0.427 ($SE = 0.0635$).

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

16. 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. $G^2 = 4.29$, $X^2 = 4.65$, $df = 3$; With JRSS-B parameter = 0, other estimates are

-0.269 for Biometrika, -0.748 for JASA, -3.218 for Communications, so prestige ranking is: 1. *JRSS-B*, 2. *Biometrika*, 3. *JASA*, 4. *Commun. Stat.*

26. 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.

28. 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.
d. This is simply the mean of the expected values of the individual binary observations.
e. 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}.

30. 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.

31. Since $\beta_M = \log[\pi_{+1}(1-\pi_{1+})/(1-\pi_{+1})\pi_{1+}]$, $\partial\beta_M/\partial\pi_{+1} = 1/\pi_{+1} + 1/(1-\pi_{+1}) = 1/\pi_{+1}(1-\pi_{+1})$ and $\partial\beta_M/\partial\pi_{1+} = -1/(1-\pi_{1+}) - 1/\pi_{1+} = -1/\pi_{1+}(1-\pi_{1+})$. The covariance matrix of $\sqrt{n}(p_{+1}, p_{1+})$ has variances $\pi_{+1}(1-\pi_{+1})$ and $\pi_{1+}(1-\pi_{1+})$ and covariance $(\pi_{11}\pi_{22} - \pi_{12}\pi_{21})$. By the delta method, the asymptotic variance of $\sqrt{n}[\log(p_{+1}p_{2+}/p_{+2}p_{1+}) - \log(\pi_{+1}\pi_{2+}/\pi_{+2}\pi_{1+})]$ is

$$[1/\pi_{+1}(1-\pi_{+1}), -1/\pi_{1+}(1-\pi_{1+})]\mathbf{Cov}[\sqrt{n}(p_{+1}, p_{1+})][1/\pi_{+1}(1-\pi_{+1}), -1/\pi_{1+}(1-\pi_{1+})]'$$

which simplifies to the expression given in the problem. Under independence, the last term in the variance expression drops out (since an odds ratio of 1.0 implies $\pi_{11}\pi_{22} = \pi_{12}\pi_{21}$) and the variance simplifies to $(\pi_{1+}\pi_{2+})^{-1} + (\pi_{+1}\pi_{+2})^{-1}$. Similarly, with the delta method the asymptotic variance of $\sqrt{n}(\hat{\beta}_C)$ is $\pi_{12}^{-1} + \pi_{21}^{-1}$ (which leads to the SE in (11.10) for $\hat{\beta}_C$)); under independence, this is $(\pi_{1+}\pi_{+2})^{-1} + (\pi_{+1}\pi_{2+})^{-1}$. For each variance, combining the two parts to get a common denominator, then expressing marginal probabilities in each numerator in terms of cell probabilities and comparing the two numerators gives the result.

34. Consider the 3×3 table with cell probabilities, by row, (0.20, 0.10, 0, / 0, 0.30, 0.10, / 0.10, 0, 0.20).

41. 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.

43. a. $\log(\Pi_{ac}/\Pi_{ca}) = \beta_a - \beta_c = (\beta_a - \beta_b) + (\beta_b - \beta_c) = \log(\Pi_{ab}/\Pi_{ba}) + \log(\Pi_{bc}/\Pi_{cb})$.
b. No, this is not possible, since if a is preferred to b then $\beta_a > \beta_b$, and if b is preferred to c then $\beta_b > \beta_c$; then, it follows that $\beta_a > \beta_c$, so a is preferred to c.

45. The kernel of the log likelihood simplifies to $\sum_{a<b} n_{ab}(\beta_a - \beta_b)$, which further simplifies to $\sum_a \beta_a n_{a+} - \sum \beta_a(\sum N_{ab})$, so the minimal sufficient statistics are $\{n_{a+}\}$.

Chapter 12

2. The sample proportions of yes responses are 0.86 for alcohol, 0.66 for cigarettes, and 0.42 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 -0.57 and estimated odds = $\exp(-0.57)$.
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 = $\exp(-0.20+0.37) = 1.19$; for cigarettes, $\exp(-0.20+0.22) = 1.02$; for marijuana, $\exp(-0.20) = 0.82$.
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, $df = 1$, $P < 0.0001$). For sources C, D, and E the size effect estimates are all roughly -0.2.
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 (0.063 compared to values around 0.11 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 $W = 12.8$ ($P = 0.005$)
b. With a linear effect for age using scores 9, 10, 11, 12, the GEE estimate of the age effect is 0.086 ($SE = 0.025$), based on the exchangeable working correlation. The P-value (0.0006) 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 ($SE = 0.12$), similar to ML.

13. b. $\hat{\lambda} = 1.08$ ($SE = 0.29$) 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.

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

variance is

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

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

$$\mathbf{V}\Big[\sum_i \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)' [v(\mu_i)]^{-1} \text{Var}(Y_i)[v(\mu_i)]^{-1} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)\Big]\mathbf{V}$$

$$= (\beta/n)[\sum_i (1/\mu_i)\text{Var}(Y_i)(1/\mu_i)](\beta/n) = (\sum_i \text{Var}(Y_i))/n^2,$$

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.

23. Since $\partial \mu_i/\partial \beta = 1$, $u(\beta) = \sum_i \left(\frac{\partial \mu_i}{\partial \beta}\right)' v(\mu_i)^{-1}(y_i - \mu_i) = (1/\sigma^2)\sum_i (y_i - \mu_i) = (1/\sigma^2)\sum_i (y_i - \beta)$. Setting this equal to 0, $\hat{\beta} = (\sum_i y_i)/n = \bar{y}$. Also,

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

Also, the actual asymptotic variance that allows for variance misspecification is

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

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

26. a. GEE does not assume a parametric distribution, but only a variance function and a correlation structure.
b. 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.
c. 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. b. For $0 < i < I$, $\pi_{i+1|i}(t) = \pi$, $\pi_{i-1|i}(t) = 1 - \pi$, and $\pi_{j|i} = 0$ otherwise. For the absorbing states, $\pi_{0|0} = 1$ and $\pi_{I|I} = 1$.

Chapter 13

3. For a given subject, the odds of having used cigarettes are estimated to equal $\exp[1.6209 - (-0.7751) = 11.0$ times the odds of having used marijuana. The large value of $\hat{\sigma} = 3.5$ reflects strong associations among the three responses.

6. a. $\hat{\beta}_B = 1.99$ ($SE = 0.35$), $\hat{\beta}_C = 2.51$ ($SE = 0.37$), with $\hat{\sigma} = 0$. e.g., for a given subject for any sequence, the estimated odds of relief for A are $\exp(-1.99) = 0.13$ times the estimated odds for B (and odds ratio = 0.08 comparing A to C and 0.59 comparing B and C). Taking into account SE values, B and C are better than A.

b. 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 $df = 10$; $P = 0.19$ for comparing models. The simpler model is adequate. Adding period effects to the simpler model, the likelihood-ratio statistic = 0.5, $df = 2$, so the evidence of a period effect is weak.

7. a. For a given department, the estimated odds of admission for a female are $\exp(0.173) = 1.19$ times the estimated odds of admission for a male. For the random effects model, for a given department, the estimated odds of admission for a female are $\exp(0.163) = 1.18$ times the estimated odds of admission for a male.
b. The estimated mean log odds ratio between gender and admissions, given department, is 0.176, 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. (Note that the marginal odds ratio of $\exp(-0.07) = 0.93$ 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.)
c. 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.

11. When $\hat{\sigma}$ is large, subject-specific estimates of random effects models tend to be much larger than population-averaged estimates from marginal models. See Sec. 13.2.3.

13. $\hat{\beta}_{2M} - \hat{\beta}_{1M} = 0.39$ ($SE = 0.09$), $\hat{\beta}_{2A} - \hat{\beta}_{1A} = 0.07$ ($SE = 0.06$), with $\hat{\sigma}_1 = 4.1, \hat{\sigma}_2 = 1.8$, and estimated correlation 0.33 between random effects.

15. b. 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.

23. When $\hat{\sigma} = 0$, the model is equivalent to the marginal one deleting the random effect. Then, probability = odds/(1 + odds) = $\exp[logit(q_i) + \alpha]/[1 + \exp[logit(q_i) + \alpha]]$. Also, $\exp[logit(q_i)] = \exp[\log(q_i) - \log(1 - q_i)] = q_i/(1 - q_i)$. 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.

25. 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 $N(0, 1+\mathbf{z}_{it}^{'}\mathbf{\Sigma}\mathbf{z}_{it})$ distribution, t he probability in the integrand is $\Phi(\mathbf{x}_{it}^{'}\boldsymbol{\beta}[1+\mathbf{z}_{it}^{'}\mathbf{\Sigma}\mathbf{z}_{it}]^{-1/2})$, 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 $[1 + \mathbf{z}_{it}^{'}\mathbf{\Sigma}\mathbf{z}_{it}]^{1/2}$, which in the univariate case is $\sqrt{1+\sigma^2}$.

27. 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 $\log(\hat{\mu}_{21}/\hat{\mu}_{12}) = 0$.

Chapter 14

3. With exchangeable association, the same parameter β applies for each association term. One can express each association term as that parameter times a product of dummy variables for the capture outcome for each time. The predicted missing count is 22.5, and the predicted population size is 90.5.

6.a. Including litter size as a predictor, its estimate is -0.102 with $SE = 0.070$. There is not strong evidence of a litter size effect. b. The $\hat{\rho}$ estimates for the four groups are 0.32, 0.02, -0.03, and 0.02. Only the placebo group shows evidence of overdispersion.

7.a. Estimate should be minus infinity.
b. Using the QL approach with beta-binomial form of variance, the $\hat{\rho} = 0.12$.

10. a. No, because the Poisson distribution has variance equal to the mean and has mode equal to the integer part of the mean.
b. A difference of 0.308 in fitted log means corresponds to a fitted ratio of $\exp(0.308) = 1.36$. The Wald CI is $\exp[0.308 \pm 1.96(0.038)] = \exp(0.2335, 0.3825) = (1.26, 1.47)$.
c. $\exp[0.308 \pm 1.96(0.127)] = (1.06, 1.75)$. The CI is much wider for the negative binomial model, since the Poisson model exhibits considerable overdispersion and fits much worse.

13. 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. In the multinomial log likelihood,

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

one substitutes

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

18. 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.

20. $E(Y_{it}) = \pi_i = E(Y_{it}^2)$, so $\mathrm{Var}(Y_{it}) = \pi_i - \pi_i^2$. Also, $\mathrm{Cov}(Y_{it}Y_{is}) = \mathrm{Corr}(Y_{it}Y_{is})\sqrt{[\mathrm{Var}(Y_{it})][\mathrm{Var}(Y_{is})]} =$

$\rho\pi_i(1-\pi_i)$. Then,

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

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

b . A binary response must have variance equal to $\mu_i(1-\mu_i)$, which implies $\phi = 1$ when $n_i = 1$.

26. 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 $\hat{\mu} = \bar{y}$.

27. For the beta binomial, $\text{Var}(Y) = E[n\pi(1-\pi)] + \text{Var}(n\pi) =$
$nE(\pi) - nE(\pi^2) + n[(E\pi)^2 - (E\pi)^2] + n^2\text{Var}(\pi)$
$= nE(\pi)[1-E(\pi)] + n(n-1)\text{Var}(\pi) = m\mu(1-\mu) + n(n-1)\mu(1-\mu)\theta/(1+\theta)$
$= n\mu(1-\mu)[1+(n-1)\theta/(1+\theta)]$.
For the negative binomial,

$$\text{Var}(Y) = E(\lambda) + \text{Var}(\lambda) = \mu + \mu^2/k.$$

29. Using the variance decomposition of Exercise 14.27,

$$\text{Var}(Y) = E(\lambda) + \text{Var}(\lambda) = \mu + \mu/k.$$

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

Chapter 15

2.a. Predict disenrollment for those of age > 70 and ≤ 83 if they have dementia or Parkinson's.
b. Predict disenrollment for no one.

4.a. Predict at least one satellite for the 88 crabs of width ≥ 25.85 cm and color in the three lowest categories, and for the 33 crabs of width < 25.85 cm and color in the two

lowest categories. Predict no satellites for the other cases.
b. Those with the greater width (between 25.55 and 25.85) are predicted not to have satellites, whereas the logistic model predicts that width has a positive effect on the likelihood of a satellite (given the color).
10.b. For the Jaccard dissimilarity, the first step combines states with Jaccard dissimilarity 0, which is only California and Florida, and the other states stay as separate clusters. At the second step, New York and Texas are combined, having Jaccard dissimilarity 0.333. (At subsequent steps, Illinois is joined with Massachusetts, then Michigan is joined with the California/Florida cluster, thenthe New York/Texas cluster is joined with the Illinois/Massachusetts cluster, then Washington is joined with the California/Florida/Michigan cluster, and then finally the two clusters left are joined to give a single cluster.
15. For such an application, the ratio d/p is very close to 1 (since for any given product, there is a high probability that neither person buys it). So, the dissimilarity $(b+c)/p$ would typically be very close to 0, even if there is no agreement between the people in what they buy.

Chapter 16

1. a. P-value is 0.22 for Pearson test and 0.208 for the one-sided Fisher's exact test P-value, and 0.245 for the two-sided Fisher's exact test P-value based on summing all probabilities no greater than observed.
b. Large-sample CI for odds ratio is (0.51, 15.37), and exact based on Cornfield approach is (0.39, 31.04).

3. a. The margins all equal 1.0. Every table with these margins has each estimated expected frequency equal to $1/I$, has I cell entries equal to 1 and (I^2 - I) entries equal to 0, and has the same probability (namely, $1/I!$, from formula 3.26). Thus the sum of probabilities of tables no more likely to occur than the observed table equals 1.0. This is the P-value for the test that orders the tables by their probabilities under H_0. (Also, each table has the same X^2 value. Thus the probability of observing X^2 at least as large as the given one is $P = 1.0$.)
b. There are $I!$ different tables with the given margins, each equally likely. The largest possible value of $C - D$ occurs for the observed table, so the null $P[C - D \geq \text{observed}]$ $= (1/I!)$. Thus using the information on ordering can result in a much different P-value than ignoring it.

9. 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.

11.a. 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.

16.a. 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.

17. 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.

19.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 $\mathbf{A} = [\mathbf{2}, (\mathbf{1}-\mathbf{2}\theta)/[\theta(\mathbf{1}-\theta)]^{\mathbf{1/2}}, (\mathbf{1}-\mathbf{2}\theta)/[\theta(\mathbf{1}-\theta)]^{\mathbf{1/2}}, -\mathbf{2}]'$. Recall that $\hat{\theta} = (p_{1+} + p_{+1})/2$. Since $\mathbf{A'A} = \mathbf{8} + (\mathbf{1} - \mathbf{2}\theta)^{\mathbf{2}}/\theta(\mathbf{1} - \theta)$, the asymptotic variance of $\hat{\theta}$ is $(\mathbf{A'A})^{-\mathbf{1}}/\mathbf{n} = 1/n[8 + (1-2\theta)^2/\theta(1-\theta)]$, 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.
b. The asymptotic covariance matrix is $(\partial\boldsymbol{\pi}/\partial\theta)(\mathbf{A'A})^{-\mathbf{1}}(\partial\boldsymbol{\pi}/\partial\theta)' = [\theta(1-\theta)/2][2\theta, 1-2\theta, 1-2\theta, -2(1-\theta)]'[2\theta, 1-2\theta, 1-2\theta, -2(1-\theta)]$.

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.

29. $P(|\hat{P} - P_o| \leq B) = P(|\hat{P} - P_o|/\sqrt{P_o(1-P_o)/M} \leq B/\sqrt{P_o(1-P_o)/M}$. 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.

31. For every possible outcome the Clopper-Pearson CI contains 0.5. e.g., when $y = 5$, the CI is (0.478, 1.0), since for $\pi_0 = 0.478$ the binomial probability of $y = 5$ is $0.478^5 = 0.025$.

33.

$$L(\theta) = \log\left[\frac{\binom{n_{1+}}{n_{11}}\binom{n-n_{1+}}{n_{+1}-n_{11}}\theta^{n_{11}}}{\sum_{u=m_-}^{m_+}\binom{n_{1+}}{u}\binom{n-n_{1+}}{n_{+1}-u}\theta^u}\right], \qquad m_- \leq n_{11} \leq m_+$$

$$= n_{11}\log(\theta) - \log\left[\sum\binom{n_{1+}}{u}\binom{n-n_{1+}}{n_{+1}-u}\theta^u\right] + \cdots.$$

$$\frac{\partial L}{\partial \theta} = \frac{n_{11}}{\theta} - \frac{\sum u \binom{n_{1+}}{u} \binom{n - n_{1+}}{n_{+1} - u} \theta^{u-1}}{\sum \binom{n_{1+}}{u} \binom{n - n_{1+}}{n_{+1} - u} \theta^{u}},$$

so $\hat{\theta}$ satisfies

$$n_{11} = n_{1+} \frac{\sum \binom{n_{1+} - 1}{u - 1} \binom{n - n_{1+}}{n_{+1} - u} \hat{\theta}^{u}}{\sum \binom{n_{1+}}{u} \binom{n - n_{1+}}{n_{+1} - u} \hat{\theta}^{u}}.$$

Now,

$$E(n_{11}) = \frac{\sum u \binom{n_{1+}}{u} \binom{n - n_{1+}}{n_{+1} - u} \theta^{u}}{\sum \binom{n_{1+}}{u} \binom{n - n_{1+}}{n_{+1} - u} \theta^{u}} = n_{1+} \frac{\sum \binom{n_{1+} - 1}{u - 1} \binom{n - n_{1+}}{n_{+1} - u} \theta^{u}}{\sum \binom{n_{1+}}{u} \binom{n - n_{1+}}{n_{+1} - u} \theta^{u}}$$

Thus ML estimate satisfies $E(n_{11}) = n_{11}$.