Corrections to second edition

An Introduction to Categorical Data Analysis

p. 71, Figure 3.2: The label for the bottom figure should be $\beta < 0$.

p. 93, Exercise 3.10: In line 2, "finding the remission value" should be "finding the labeling index value".

p. 95, Exercise 3.17: The article cited has actual volume number 136 and upper limit of 883 for the page numbers.

p. 109: $\text{Corr}(\hat{\alpha}, \hat{\beta})$ is very near 1.0, and the estimated covariance matrix reported is not shown with enough digits to replicate the calculation shown for $\text{Var}(\hat{\alpha} + \hat{\beta}x)$. Computationally, a better approach to get this interval at $x = 26.5$ is to use new predictor $x^* = (x - 26.5)$, so that the relevant variance for the logit (which equals 0.826) is then $\text{Var}(\hat{\alpha})$, which is 0.0356 (rather than 0.038 as reported in the text). The confidence interval for the logit is then (0.456, 1.196), which translates to (0.61, 0.77) for the probability, as stated in the text.

p. 142: In Table 5.3, the prediction counts using $\pi_0 = 0.50$ when $y = 0$ should be 34 and 28, not 37 and 25. These tables were obtained using PROC LOGISTIC in SAS; it gets the predicted probability for an observation by using the cross-validation approach of fitting the model without that observation (more precisely, using a one-step approximation for the ML estimates for the model fitted without that observation).

p. 182: Table 6.7 should have been presented as a 2×5 table relating political party to political ideology (i.e., collapsed over gender), because the models discussed in Sections 6.2.2 and 6.3.2 do not include gender as an explanatory variable.

p. 194: In line 4 (and also 4 lines from the bottom), the deviance reported of 13.9 is actually 13.952 and rounds to 14.0 and has df = 19, whereas the model without the income effect has df = 20 (not 19).

p. 207: In line 6 of text from the bottom, "gender" should be "race"

p. 252: For the way Table 8.3 is shown (with the cases as columns rather than as rows), the conditional ML estimate of $37/16 = 2.3$ should be expressed on line 5 as n_{21}/n_{12} rather than as n_{12}/n_{21}.

In the Appendix A pages for SAS, in Tables A.1, A.2, A.3, A.5, A.6, A.8, A.9, A.10,

A.12, A.14, A.15, the @@ signs mysteriously changed to copyright signs when the book was printed!

p. 358: Corrected answers for Exercise 2.1:
(a) The probability that is 2/3 is $P(+|\bar{C})$.
(c) $P(C, +) = 0.0075$, $P(C, -) = 0.0025$, $P(\bar{C}, +) = 0.66$, $P(\bar{C}, -) = 0.33$.
(d) $P(+) = 0.6675$, $P(-) = 0.3325$.
(e) $0.0075/0.6675 = 0.01236$.

p. 362: For Exercise 13 in Chapter 4, answer to part (b) for expected count should be 1.76. For Exercise 19 in Chapter 4, answer to part (b) should be (i) 0.08, (ii) 0.71, and answer to part (c) should be $\hat{\beta}_2^G = -0.16$ (so that it still is true that $\hat{\beta}_1^G - \hat{\beta}_2^G = 0.16$ and the odds ratio is $e^{0.16} = 1.17$).

p. 370: The answer given for Exercise 9.11 is incorrect, as it refers to a different data set. The answer should read: For a GEE approach with independence working correlation structure for a cumulative logit model with constraint $\beta_E = 0$ for the environment, the estimates are $\hat{\beta}_C = -2.338 (SE = 0.121), \hat{\beta}_L = -0.465 (SE = 0.119), \hat{\beta}_H = -0.076 (SE = 0.116)$. It appears that there is less support for spending on cities than for the other types of spending.

The first two printings also have the error that the printout in Table 4.11 (p. 126) is not for the death penalty data in Exercise 4.12 but rather the death penalty data in Table 2.10 on p. 50. So, in the statement of the exercise, please replace 19 by 53, 151 by 467, 9 by 16, 63 by 48, 6 by 4 and 103 by 143.