Errata for *Applied Regression Analysis, Linear Models, and Related Methods* (Sage, 1997)

John Fox

19 April 2009

Please note that some (or all) of these errors may be corrected in your printing of the book.

1. Page 25: In the middle of the paragraph, "at the center of the widow," should read "at the center of the window."

2. Pages 61–62: The last sentence beginning on p. 61 should read:

   In particular, descending the "ladder" of powers and roots toward $X^{(-1)}$ compresses the large values of X and spreads out the small ones; ascending the ladder of powers and roots toward $X^{(2)}$ has the opposite effect.

3. Page 67: In the first paragraph, "The inverse of distance traveled in a given period of time is speed," should read "The inverse of time taken to travel a fixed distance is speed."

4. Page 76, Figure 4.12: The point labels for Other and Canada are reversed, as are those for the UK and US. (The point for the US should be furthest to the left; the point for Other should be furthest to the right.)

5. Page 89:

   $$\text{Similarly, } \sum \widehat{Y} E_i.$$

   should read,

   $$\text{Similarly, } \sum \widehat{Y}_i E_i.$$

6. Page 91: The formula

   $$\sum E_i^2 = \sum (Y_i - \widehat{Y})^2$$

   should read

   $$\sum E_i^2 = \sum (Y_i - \widehat{Y}_i)^2$$

7. Page 114, Figure 6.1: The vertical axis in the figure should be labeled $p(Y|x)$, not $P(Y|x)$.

8. Page 120: The text below equation 6.2 should read:

   where R_j^2 is the squared multiple correlation from the regression of X_j on all of the other X's.

9. Page 125: The regression sum of squares (RegSS_0) for the regression of prestige on income alone is 30,665 (not 30,655, as given). The calculation of the incremental sum of squares due to education is correct, however, as is the incremental F.

10. Page 137: In the second paragraph on this page, "income" is erroneously substituted for "education" at several points. The text should read:

    In contrast, the quantitative independent variable *education* and the regressor X are one and the same. Were we to transform education, however, prior to entering it into the regresson equation — say, by taking logs — then there would be a distinction between the independent variable (education) and the regressor (log education).

11. Page 159: The formula at the bottom of the page,

$$\widehat{Y}_{ij} = M + A_i = \overline{Y}_{\cdot} + (\overline{Y}_j - \overline{Y}_{\cdot}) = \overline{Y}_j$$

should read,

$$\widehat{Y}_{ij} = M + A_j = \overline{Y}_{\cdot} + (\overline{Y}_j - \overline{Y}_{\cdot}) = \overline{Y}_j$$

12. Page 160, footnote 4: The formula for the mean of the group means, given as

$$\overline{Y}_{\cdot} = \left(\sum \overline{Y}_j\right)/n$$

should read

$$\overline{Y}_{\cdot} = \left(\sum \overline{Y}_j\right)/m$$

13. Page 179: The second sentence of the last paragraph should read:

For the full model and the main-effects model, we obtain the same sums of squares as before ...

14. Page 194: The analysis-of-covariance model fit to Moore and Krupat's data, given in equation 8.16 as

$$\begin{aligned} \widehat{Y} &= \underset{(2.200)}{13.03} - \underset{(2.200)}{7.767S} - \underset{(0.0485)}{0.02055(X - \overline{X})} \\ &\quad + \underset{(0.0485)}{0.1306[S(X - \overline{X})]} \\ R^2 &= .2942 \end{aligned} \tag{8.16}$$

is in error: The mean $\overline{X}$ was not subtracted from the X values (as indicated). The correct fit is:

$$\begin{aligned} \widehat{Y} &= \underset{(0.681)}{12.14} - \underset{(0.681)}{2.139S} - \underset{(0.0485)}{0.02055(X - \overline{X})} \\ &\quad + \underset{(0.0485)}{0.1306[S(X - \overline{X})]} \\ R^2 &= .2942 \end{aligned} \tag{8.16}$$

15. Page 208: In Equation 9.3, there are ellipses (i.e., $\cdots$) missing between the third and last columns of the row basis of the full-rank model matrix; this equation should read:

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_{m-1} \\ \mu_m \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & -1 & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{m-1} \end{bmatrix}$$

$$\underset{(m\times 1)}{\boldsymbol{\mu}} = \underset{(m\times m)}{\mathbf{X}_B}\,\underset{(m\times 1)}{\boldsymbol{\beta}_F}$$

16. Page 208: In Equation 9.4, the last two entries in the final column of the model matrix $\mathbf{X}_B$ are inverted; the equation should read:

$$\begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \end{bmatrix}$$

$$\underset{(6\times 1)}{\boldsymbol{\mu}} = \underset{(6\times 6)}{\mathbf{X}_B}\,\underset{(6\times 1)}{\boldsymbol{\beta}_F}$$

17. Page 219: The partial derivative of the log-likelihood with respect to σ_ε^2 is in error, missing the factor 2 in the denominator of the second term; the correct derivative is:

$$\frac{\partial \log_e L(\boldsymbol{\beta}, \sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = -\frac{n}{2}\left(\frac{1}{\sigma_\varepsilon^2}\right) + \frac{1}{2\sigma_\varepsilon^4}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

18. Page 226: There are three errors on this page:

    - The formula for the F-statistic for a general linear hypothesis, given as

    $$F_0 = \frac{(\mathbf{Lb} - \mathbf{c})' \mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}' (\mathbf{Lb} - \mathbf{c})}{qS_E^2} \tag{9.15}$$

    should read:

    $$F_0 = \frac{(\mathbf{Lb} - \mathbf{c})' \left[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\right]^{-1} (\mathbf{Lb} - \mathbf{c})}{qS_E^2} \tag{9.15}$$

    The error in Equation 9.15 is repeated in the box on page 236 and in the next-to-last summary point on page 239.

    - The following formula, given as

    $$\mathbf{Lb} \sim N_{k+1}[\mathbf{L}\boldsymbol{\beta},\, \sigma_\varepsilon^2 \mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']$$

    should read

    $$\mathbf{Lb} \sim N_q[\mathbf{L}\boldsymbol{\beta},\, \sigma_\varepsilon^2 \mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']$$

    - And the following formula, given as

    $$(\mathbf{Lb} - \mathbf{c})' \mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}' (\mathbf{Lb} - \mathbf{c}) / \sigma_\varepsilon^2 \sim \chi_q^2$$

    should read

    $$(\mathbf{Lb} - \mathbf{c})' \left[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\right]^{-1} (\mathbf{Lb} - \mathbf{c}) / \sigma_\varepsilon^2 \sim \chi_q^2$$

19. Page 256: The order of the $\mathbf{e}$ vector in the first displayed equation in Section 10.3 should be $(n \times 1)$ not $(n - 1)$; i.e., the equation should read

$$\underset{(n-k-1\times 1)}{\mathbf{z}} \equiv \underset{(n-k-1\times n)(n\times 1)}{\mathbf{G}}\ \underset{}{\mathbf{e}}$$

20. Page 270: The formula

$$\begin{array}{rcccccccc} \widehat{MW} & = & 1.79 & + & 0.969RW & + & 2.07F & - & 0.00953(MW \times F) \\ & & (5.92) & & (0.076) & & (9.30) & & (0.147) \end{array}$$

    should read

$$\begin{array}{rcccccccc} \widehat{MW} & = & 1.79 & + & 0.969RW & + & 2.07F & - & 0.00953(RW \times F) \\ & & (5.92) & & (0.076) & & (9.30) & & (0.147) \end{array}$$

21. Page 277: Near the bottom of the page,

$$\text{DFBETAS}_{3,20} = -24.8$$

    should read

$$\text{DFBETAS}_{3,12} = -24.8$$

22. Page 288, Table 11.1: The value of C for County 13 should be 22.0 (not 2.0).

23. Page 292: The two sets of residuals for the leverage plot are incorrectly identified; the first two bulleted points should read as follows:

    - The residuals around the horizontal line at $V_y = 0$ are the constrained least-squares residuals E_{0i} under the hypothesis H_0.
    - The least-squares line fit to the leverage plot has an intercept of 0 and a slope of 1; the residuals about this line are the unconstrained least-squares residuals, E_i. The incremental sum of squares for H_0 is thus the regression sum of squares for the line.

24. Page 303: The likelihood for the model, given as

$$L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))'\boldsymbol{\Sigma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$

    should be

$$L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$

25. Page 304: The maximum likelihood estimator of the error variance is incorrectly given as

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum(E_i/w_i)^2}{n}$$

    The correct formula is

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum(w_i E_i)^2}{n}$$

26. Page 313: In the second bulleted point, the cubic term for education should be S^3, not E^3: "... and the coefficient for S^3 is not quite statistically significant ...".

27. Page 319: Equation 12.10,

$$Y_i = \alpha' + \gamma_1 D_{i1} + \gamma_3 D_{i3} + \cdots + \gamma_{19} D_{i,20} + \varepsilon_i'$$

    should read:

$$Y_i = \alpha' + \gamma_1 D_{i1} + \gamma_3 D_{i3} + \cdots + \gamma_{20} D_{i,20} + \varepsilon_i'$$

28. Page 326, Figure 12.10: The two Box-Tidwell constructed-variable plots are slightly in error, or at least not quite as described in the text. Each plot was produced from a model that included only one of the two constructed-variable regressors [i.e., $I \log_e I$ for (a), and $S \log_e S$ for (b)], along with the terms in the original model — rather than including both constructed variables in the same model. When both constructed variables are included in the model, the constructed-variable plots are similar, but not identical, to those in Figure 12.10. (*Note*: This figure has not been corrected in subsequent printings of the text because the software that I used to make the graphs is no longer readily available.)

29. Page 344: The variance of the first principal component, given as

$$S_{W_1}^2 = \frac{1}{n-1}\mathbf{w}_1'\mathbf{x}_1 = \frac{1}{n-1}\,\mathbf{a}_1'\mathbf{Z}_X'\mathbf{Z}_X\mathbf{a}_1 = \mathbf{a}_1'\mathbf{R}_{XX}\mathbf{a}_1$$

    should read:

$$S_{W_1}^2 = \frac{1}{n-1}\mathbf{w}_1'\mathbf{w}_1 = \frac{1}{n-1}\,\mathbf{a}_1'\mathbf{Z}_X'\mathbf{Z}_X\mathbf{a}_1 = \mathbf{a}_1'\mathbf{R}_{XX}\mathbf{a}_1$$

    That is, $\frac{1}{n-1}\mathbf{w}_1'\mathbf{x}_1$ should be $\frac{1}{n-1}\mathbf{w}_1'\mathbf{w}_1$.

30. Page 382: The range of the Durbin-Watson statistic D is 0 to 4, not 0 to 2 as stated in the text.

31. Page 408: When the tuning constant for the Huber estimator is made larger, there is *less* resistance to outliers, not *more* as stated in the text.

32. Page 412: There are two equations on this page in which the vector $\mathbf{x}_i'$ should have been written as a column vector, $\mathbf{x}_i$; the correct formulas are:

$$\sum_{i=1}^{n} \psi(Y_i - \mathbf{x}_i'\mathbf{b})\mathbf{x}_i = \mathbf{0}$$

and

$$\sum_{i=1}^{n} w_i(Y_i - \mathbf{x}_i'\mathbf{b})\mathbf{x}_i = \mathbf{0}$$

The first of these equations also appears in the summary box on Page 413.

33. Pages 419 and 435: In point 2.b. on page 419 (*Local WLS fit*), the weighted-least-squares criterion, given incorrectly as $\sum_{j=1}^{m} w_{ij}^2 E_{ij}^2$, should be $\sum_{j=1}^{m} w_{ij} E_{ij}^2$ (i.e., the tricube weight is not squared). The same error is repeated in the first bulleted point under Nonparametric Regression on page 435.

34. Page 442: The bulleted point concerning estimation of the constrained linear probability model and Figure 15.2 are in error. Although the general criticism of the constrained linear probability model is correct, it is not true that the estimated point at which $\pi = 0$ coincides with the leftmost 1 and the point at which $\pi = 1$ with the rightmost 0.

35. Page 467: In the second line of Equation 15.22, the index for the summation should be j, not l:

$$\pi_{im} = 1 - \sum_{j=1}^{m-1} \pi_{ij}$$

36. Page 469: High husband's income and presence of children *decrease the odds* of working full time relative to not working, not *increase the odds*, as stated in the text.

37. Page 482: The degrees of freedom for Models 3 and 4 in Table 15.4 are incorrect; the table should read:

Model	*Terms*	$k+1$	*Deviance:* G^2
1	α, β, γ	6	1356.434
2	α, β	4	1363.552
3	α, γ	4	1368.042
4	β, γ	5	1368.554
5	α	2	1382.658
6	β	3	1371.838

Table 15.5 is correct, however.

38. Page 495: The critical value of t for the 95-percent confidence interval in the example is given as $t_{3,.025} = 4.30$. The correct value is $t_{3,.025} = 3.18$. As a consequence, the confidence interval, which is given as

$$\mu = 2.75 \pm 4.30 \times 2.015 = 2.75 \pm 8.66$$

and as

$$-5.91 < \mu < 11.41$$

should read

$$\mu = 2.75 \pm 3.18 \times 2.015 = 2.75 \pm 6.41$$

and

$$-3.66 < \mu < 9.16$$

39. Page 499, Caption for Table 16.3: "Wives' Iincomes" should read "Wives' Incomes" (obviously!).

40. Page 504: Two equations (below equation 16.2) are given for A_1; the second is actually for A_2.

41. Page 504: The terms in the numerator of Equation 16.2 are reversed; the equation should read:

$$A \equiv \frac{\sum_{i=1}^{n}(\overline{\theta} - \widehat{\theta}_{(-i)})^3}{6[\sum_{i=1}^{n}(\widehat{\theta}_{(-i)} - \overline{\theta})^2]^{3/2}} \tag{16.2}$$

The computation for the example on page 505 is correct, however.

42. Page 509: The formula currently given as

$$T = \mathbf{b}_1' \mathbf{V}_{11} \mathbf{b}_1$$

should read

$$T = \mathbf{b}_1' \mathbf{V}_{11}^{-1} \mathbf{b}_1$$

Likewise, the formula

$$T_b^* = (\mathbf{b}_{b1}^* - \mathbf{b}_1)' \mathbf{V}_{b,11}^* (\mathbf{b}_{b1}^* - \mathbf{b}_1)$$

should read

$$T_b^* = (\mathbf{b}_{b1}^* - \mathbf{b}_1)' \mathbf{V}_{b,11}^{*-1} (\mathbf{b}_{b1}^* - \mathbf{b}_1)$$

43. Page 580: The reference to "Robey, B., Shea, M. A., Rutstein, O., and Morris, L." should be to "Robey, B., Rutstein, S. O., and Morris, L."

44. Page 588: The entry for “Durbin-Watson statistic" in the index should be pp. 382–383, not 482–483.

45. Page 592: The entry for the “premium–protection approach” in the index should be pp. 274–275, not 294–295.