An R Companion to Applied Regression, Second Edition
Errata and Changes

John Fox and Sanford Weisberg

2015-07-10

Note: Some or all of these errors and changes may be corrected in your printing of the book.

- p. 33: At the bottom of the page,

  ```
  > rownames(Duncan)[c(6, 16)]
  ```

  ```
  [1] "minister"  "contractor"
  ```

  should read

  ```
  > rownames(Duncan)[c(6, 16)]
  ```

  ```
  [1] "minister"  "conductor"
  ```

  Explanation: Because the book was prepared using Sweave, this error should have been impossible, but the mistake was introduced during copy-editing of the LaTeX source.

- p. 110: The text makes reference to the `stemleaf` function in the **aplpack** package; the correct name of the function is `stem.leaf`. We thank David Arnold for bringing this error to our attention.

- p. 134: The formula for the empirical logit, given as $\log[(x + 1/2)/(n + 1)]$, is in error; the correct formula is $\log[(x + 1/2)/(n - x + 1/2)]$. (The formula given in error is the proportion on which the empirical logit is based.) We are grateful to Bill Venables for pointing out this error.

- p. 143: Near the bottom of the page, the rounded values used by `powerTransform` are given as $(1, 0, 0.5, .33, -0.5, -0.33, 2, -2)$ but are actually $(1, 0, -1, 0.5, .33, -0.5, -0.33, 2, -2)$ (i.e., -1 is missing). Thank you to Liviu Andronic for reporting this omission.

- p. 159: In the last paragraph, "The first column is a columns of 1s" should read "the first column is a column of 1s" (i.e., "column," singular, rather than "columns," plural). We are grateful to Liviu Andronic for noticing this typo.

- p. 160: In the paragraph above the command `some(Baumann)`, the text "two new methods called `DTRA` and `Strat`," should read, "two new methods called `DRTA` and `Strat`," (i.e., `DRTA`, not `DTRA`); we thank David Arnold for noticing this error.

- p. 178: At the bottom of the page, the sentence, "Natural splines with 6 *df* have four knots, dividing the data into five regions." should read, "Natural splines with 6 *df* have five knots, dividing the data into six regions." Liviu Andronic kindly pointed out this error.

- p. 179: Near the bottom of the page, "the B-spline regressors do not have a simple description," should read, "the natural-spline regressors do not have a simple description." Similarly, the caption to Figure 4.10 on p. 180 should make reference to natural splines rather than to B-splines. We're grateful to Liviu Andronic for bringing the latter point to our attention.

- p. 185: The HC3 sandwich estimator is based on $\mathrm{diag}[e_i^2/(1-h_i)^2]$, not on $\mathrm{diag}[e_i^2/(1-h_i{}^2)]$ as given in the text. The `hccm` function in the **car** package computes the HC3 estimator correctly, however. We're grateful to Paul Dudgeon for bringing this error to our attention.

- p. 188: In the first paragraph, "simply the SD of the bootstap estimates" should read "simply the SD of the bootstrap estimates." Later in the same paragraph, `bootBias` is given as "the difference between the usual estimate and the average of the bootstrap estimates ...," but actually is computed as "the difference between the average of the bootstrap estimates and the usual estimate ... ." We thank Liviu Andronic for reporting these errors.

- p. 189: In the first block of code on the page, the comments for the two confidence intervals are reversed: the first should read, `"# CI for income"` and the second `"# CI for education"`. We are grateful to Liviu Andronic for noticing this error.

- p. 192: In the first sentence of Section 4.4.3, "As we gave seen" should read "As we have seen"; we thank Liviu Andronic for noticing this typo.

- p. 202: The arguments for the function `deltaMethod` were changed in August, 2011. Four lines above Section 4.4.7, change `deltaMethod(trans.mod, "b1/b2")` to `deltaMethod(trans.mod, "b1/b2", parameterNames=c("b0", "b1", "b2"))`.

- p. 221: In the first paragraph, "Summaries such as *df* ..." would be clearer as "Summaries such as residual *df* ... ." We thank Liviu Andronic for asking for a clarification.

- p. 260: The variable x_k is missing from the first displayed equation, which should read:

$$\eta\left(\mathbf{x}\right) = \beta_{0j} + \beta_{1j}x_1 + \cdots + \beta_{kj}x_k$$

- p. 317: In the first paragraph of Section 6.6.1, the differences defining the ordinary residuals and errors in linear models are reversed: The residuals are given as $\widehat{y} - y$, and should be $y - \widehat{y}$; the errors are given as $\varepsilon = E(y|\eta) - y$ and should be $\varepsilon = y - E(y|\eta)$. We thank Liviu Andronic for reporting these errors.

- p. 357: In the second paragraph, the bases of the **iplots** and **playwith** packages are reversed; the text should read, "the former via the **rJava** package, which links R to the Java computing platform, and the latter via the **RGtk2** package, which links R to the GTK+ GUI toolkit." We are grateful to Liviu Andronic for noticing this error.

- p. 403: At the bottom of the page, the command `coef(mod.2)`, and the output that follows, was intended to be:

```
> coef(mod.glm)

(Intercept)          X1          X2          X3          X4          X5
  1.1194903   1.0927021   1.0320458   1.0694751   0.9978092   0.9832376
         X6          X7          X8          X9         X10
  0.9575331   1.1162508   1.0137408   1.1149053   1.1113571
```

We thank Mark Leeds for noticing this error.

- p. 412: The R input beneath "Let us verify that the new method works properly:" (near the middle of the page) repeats the definition of the `summary` method near the bottom of the page. What was intended here was the following test of the `show` method:

```
> mod.mroz.4

   Constant          k5        k618         age          wc          hc
 3.18214046 -1.46291304 -0.06457068 -0.06287055  0.80727378  0.11173357
        lwg         inc
 0.60469312 -0.03444643
```

Again, we are grateful to Mark Leeds for discovering this error.

- pp. 421–422: For Tukey's test to be possible, the squared fitted values $\widehat{y}^2$ must not be obtainable as a linear combination of the regressors. This will happen, for example, for one-way analysis of variance models. The revised version of the function shown on this page adds a check for this condition and returns `NA` when it occurs:

```
tukeyNonaddTest <- function(model){
    tol <- model$qr$tol
    qr <- model$qr
    fitsq <- predict(model, type="response")^2
    fitsq <- qr.resid(qr, fitsq/sqrt(sum(fitsq^2)))
    if(sum(fitsq^2) < tol){
        return(c(Test=NA, Pvalue=NA))
    } else {
        r <- residuals(model, type="pearson")
        m1 <- lm(r ~ fitsq, weights=weights(model))
        df.correction <- sqrt((df.residual(model) - 1)/df.residual(m1))
        tukey <- summary(m1)$coef[2, 3] * df.correction
        c(Test=tukey, Pvalue=2*pnorm(-abs(tukey)))
    }
}
```