Corrections and Elucidations (May 8, 2010) – 2^{nd} edn (1^{st} and 2^{nd}printing) of Data Analysis and Graphics Using R – An Example-Based Approach

John Maindonald (email: john.maindonald@anu.edu.au) and John Braun

Webpage: `http://www.maths.anu.edu.au/~johnm/r-book/r-bookEd2.html`.

Chapter 1

p.16, line -6 (ss 1.4.1)

Replace `subset(Cars93.summary, subset=1:2)`
by

```
subset(Cars93.summary, subset=c(TRUE, TRUE, FALSE, FALSE, FALSE, FALSE))
```

The `subset` argument requires a logical expression indicating which rows to keep. Rows may not be identified by number.

p.33, line -10 (Section 1.8)

Remove the unwanted `)` from `save.image())`

p.39, exercise 6(b) (Section 1.11)

Replace
"The `log="y"` setting is automatic, after its initial use with `plot()`, for the subsequent use of `text()`."
by
"The `log="y"` setting carries across to the subsequent `text()` commands. See Subsection 2.1.5 for an example."

p.40, exercise 14 (Section 1.11)

Replace `census1889` by `census1886`, in each of the three places where it appears.

p.42, exercise 20 (Section 1.11)

Replace this exercise by:
The help page for `iris` (type `help(iris)`) gives code that converts the data in `iris3` (*datasets* package) to case-by-variable format, with column names "Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width", and "Species". Look up the help pages for the functions that are used, and make sure that you understand them. Then add annotation to the code that explains each step in the computation.
This should be an asterisked question.

Chapter 2

p.48, code segment in final 8 lines (ss 2.1.2)

line -2: Omit two redundant spaces; make `p` in `plot` is lower case:

```
plot(window(measles, start=1840, end=1882), ylim=c(0, 4600), yaxt="n")
```

p.49, footnote 4 (ss 2.1.2)

In order not to interfere with the plotting of subsequent graphs, the code should be followed by

```
par(fig=c(0,1,0,1))
```

p.55, lines -15 to -14 (ss 2.1.5)

Replace
" Labeling is done on the untransformed (`Number`) scale, with values on the logarithmic scale given in brackets:
"
by
" Labeling is done on the transformed (`log(Number)`) scale. The scale to the right of the graph relates `log(Number)` to `Number`. "

line -6
Replace c(2,3) by c(3,2).

lines -3 to -1
Omit see Subsection 14.3.4. Alternatively,

Add as the final sentence in the paragraph:
Subsection 14.3.4 describes how to label the x-axis with dates, in the format `Jan95`, `Apr95`, ...

p.65, fnote 11 (Section 2.2.2)

Omit the final two lines of the footnote, i.e., omit

```
# Note that parameter settings were given both in the function call and
# in the list supplied to key.  [With auto.key, this is unnecessary.]
```

p.70, l.14 (Subsection 2.3.1)

Replace 6.5 by 13.2.

p.72, line 14 (ss 2.3.4)

Replace "use as by "use are".

p.72, line 16 (Subsection 2.3.4)

Replace , discussed briefly in Subsection 2.2.1,
by (see Subsection 2.2.1 and Section 13.2)"

p.76, Exercise 7 (Section 2.6)

Replace Download and load by Install and attach.
[The package must be installed, which usually involves an initial downloading step. The action of `library()` is better described as attaching, rather than loading, the relevant package.]

Chapter 4

p.105, line 16 (ss 4.1.5)

Replace "with 3 d.f." by "with 8 and 3 d.f.".

p.106, line -3 (ss 4.1.6)

Replace 90% by 95%.

p.108, line -11 (ss 4.2.1)

Replace -6.10 /2.03 by -6.33/2.03

p.110, Table 4.2, column "Test statistic" (ss 4.2.1)

In line 3, omit the vertical bar that appears as the final character in the numerator.

p.111, line -9 (ss 4.2.1)

Replace equal by unequal.

p.115, lines -10 to -9 (Section 4.3)

Replace

```
> chisq.test(table(nsw74psid3$trt, nsw74psid3$nodeg))
>   # Specify correct=FALSE
. . . .
X-squared = 12, df = 1, p-value = 0.0004975
```

by

```
> # Specify correct=FALSE for easy comparison with hand calculation
> with(nsw74psid3, chisq.test(trt, nodeg, correct=FALSE))
. . . .
X-squared = 12.9666, df = 1, p-value = 0.0003171
```

p.121, line 14 (Section 4.4)

Replace using the `anova()` function, thus:
by
using the `aov()` function to fit the analysis of variance model, then calling the `anova()` function with the resulting `aov` object as argument, thus:

Chapter 5

p.172 (Section 5.11)

Exercise 12 should be asterisked.

Chapter 6

p.189 (ss 6.3.3), line 3

This should read
" Figure 6.9B does seem marginally more consistent than Figure 6.9A with the assumption of a linear relationship between predictors. "
(i.e., interchange A with B)

p.192 (Section 6.3.3)

lines 3-4: Replace "The model without the interaction has a slightly smaller AIC. For this reason, and because it is a simpler model, it is the preferrred model "
by
"The model without the interaction has a slightly larger AIC. Because however the difference is slight, the simpler model (no interaction) is the preferred model.

p.196, lines -9 to -4 (2nd para under Section 6.5)

This should be abbreviated to:
" The way that data are sampled can likewise affect the coefficients. This section will examine data, sampled in a deliberately biased way, on the effect of book dimensions (thickness, height and width) on book weight. "

p.207 (ss 6.8.1), line 3

In line 3 of Subsection 6.8.1, replace Such error, ...
with
Discussion will be limited to a relatively simple 'classical "errors in x" model. For this model the error in x, ...

p.209 (ss 6.8.1), final paragraph

Replace Variables that are measured inaccurately ... interpreting results.
with
Again, attention will be limited to the classical "errors in x" model. Where one only of several variables is measured inaccurately, its coefficient may on that account not appear statistically significant, or be severely attenuated. For remaining variables (measured without error) possible scenarios include: the coefficient suggests a relationship when there is none, or the coefficient is reversed in sign. Where several variables are measured with error, there is even more room for misleading and counter-intuitive coefficient values.

p.211 (ss 6.8.3), Table 6.3, final line

Replace " 14 = 28.5 - 16.5"
by
with
"12 = 28.5 - 16.5"

p.218 (Section 6.11)

In "Additional Notes", there is a further exercise (exercise 12) that illustrates how measurement error in one variable may lead to a spurious effect for a variable that is measured without error.

Chapter 7

p.222 (ss 7.1.1)

line -6:
Replace " 4 results/trt" by 3 results/trt.

lines -4 to -3:
There are 8 degrees of freedom (not 66) for the residual standard error. These lines should read.

```
> qtukey(p=.95, nmeans=4, df=8) * sem
[1] 20.26
```

p.229 (sec 7.3), line -7)

Replace `summary(leaf.lm2)` by `summary(leaf.lm3)`.

p.236 (ss 7.5.1)

line 2: "... the first with two knots, and the second with three knots.'

line 10: Replace "Figure 7.7 A " by "'Figure 7.7B".

fnote 4, lines 2 & 3 Replace 4 by 3 (3 occurrences). In line 2, replace
"`# panel B: nspline, df=4` " by "`# panel B: ns(juice,4)`" .
The code then reads:

```
fruit.lmb3 <- lm(ohms ~ ns(juice,3))   # panel B: ns(juice,4)
plot(fruit.lmb3)
```

p.237 (ss 7.5.1)

Figure 7.7, caption: Add: The degrees of freedom ('df' or 'degree') shown are those supplied to `ns()` or `poly()`. These must in each case be increased by one to allow for the intercept.

Figure 7.8, caption: Replace "Figure 7.7 A " by "'Figure 7.7B".

p.237 (ss 7.5.1)

Figure 7.7, caption: Add: The degrees of freedom ('df' or 'degree') shown are those supplied to `ns()` or `poly()`. These must in each case be increased by one to allow for the intercept.

Figure 7.8, caption: Replace
"Figure 7.7 A " by "'Figure 7.7B".

p.238, Figure 7.9, caption (ss 7.5.1)

Replace " B -spline (one knot) fitted in Figure 7.7 A " by "N-spline (three knots) fitted in Figure 7.7B"

Chapter 8

p.266 (ss 8.4.2), line3 -3 and -2

Replace `exp(-1.282)` by `exp(-1.283)`

Replace Bank by Lowerside. Omit the final sentence; it adds nothing to what has gone before.

p.268 (ss 8.4.2), lines 11-12

Omit these lines (they repeat lines 9-10), i.e., omit

```
> ANW.glm <- glm(A ~ habitatNW + log(meters), family = quasipoisson,
+                  data=moths)
```

p.282 (ss 8.7.5), line 5

Omit (or from $\beta = 0$) . (The parenthetical comment is superfluous).

Chapter 10

p.302, line 2

Replace *nline* by *nlme*.

p.308 (ss 10.1.3)

In line 13 (""lower" of the ..."), replace 0 by 1. In line 14, replace 1 by 2.

p.309 (ss 10.1.3), line 5:

Following "In", insert hierarchical
i.e., "In hierarchical multi-level ..."

pp. 310, 314 & 343 – output from `mcmcsamp()`:

Changes to the structure of objects created by `mcmcsamp()` affect code on pages 310 (lower half-page), 314 (lines -16 to -10) and 343 (line 2). The code on page 310 starts with the two lines:

```
ant111b.lmer <- lmer(harvwt   1 + (1 | site), data=ant111b)
ant111b.samp <- mcmcsamp(ant111b.lmer, n=1000)
```

The code in subsequent lines is no longer valid. Instead, specify:

```
HPDinterval(VarCorr(ant111b.samp, type="varcov"))
```

This does not, currently, give results that are believable for this example.

p.311 (ss 10.1.3)

line 5: Replace 0 by 3, 1 by 2, and 2 by 1. Thus the sentence becomes:
" We now have three levels of variation: level 3 is house, level 2 is suburb, and level 1 is city."

line 8: Replace reflected by identified.

p.314 (ss 10.2.1), lines -17 to -10

Replace by

```
science1.lmer <- lmer(like ~ sex + PrivPub + (1 | school:class),
                      data = science, na.action=na.exclude)
science1.samp <- mcmcsamp(science1.lmer, n=1000)
HPDinterval(VarCorr(ant111b.samp, type="varcov"))
```

Here, the results do make sense.

p.325 (final 2 lines) & p.326 (first 2 lines), ss 10.3.5

(First paragraph under *Plots of residuals*)
Replace these first 4 lines, i.e.

Recall that by default, fitted values adjust for all except random variation between individual vines, i.e., they account for treatment, block and plot effects. For this, set `level=2` when calculating fitted values, or the equivalent residuals. Other choices are to calculate fitted values as treatment plus block (`level=1`) or as treatment effects only ...

by

In this hierarchical model there are three levels of variation: level 1 is between blocks, level 2 is between plots, and level 3 is between vines. The function `fitted()` adjusts for all levels of random variation except between individual vines, i.e. fitted values are at level 2.

Because the older `lme()` function (*nlme*) was designed for use with hierarchical models, the `fitted` method for `lme` model objects does accept the parameter `level`, here with a choice of 0, 1 or 2. ...

Thus, the paragraph reads
"In this hierarchical model there are three levels of variation: level 1 is between blocks, level 2 is between plots, and level 3 is between vines. The function `fitted()` adjusts for all levels of random variation except between individual vines, i.e. fitted values are at level 2.

Because the older `lme()` function (*nlme*) was designed for use with hierarchical models, the `fitted` method for `lme` model objects does accept the parameter `level`, here with a choice of 0, 1 or 2. The block *effects* are differences between fitted values at level 1 and fitted values at level 0, while the plot effects are differences between fitted values at level 2 and fitted values at level 1."

p.343 (Section 10.5.2), lines 1–3

Code to handle the use of `mcmcsamp()`, for models of this type, has not yet, at the time of writing, been adapted for use with the current version of `lmer()`

p.348 (Section 10.10), exercise 1

Replace the final four lines of code, starting `vcov <- show(VarCorr(kiwishade.lmer))`, with

```
vcov <- VarCorr(kiwishade.lmer)
vars <- c("(block:plot)^2"=as.vector(vcov[["block:plot"]]),
          "sigma^2"=as.vector(attributes(vcov, "sigmaREML")$sc^2))
print(vars)
```

p.348 (Section 10.10), exercise 5

For assessing the accuracy of the components of variance, consider using `mcmcsamp()` as demonstrated on p.314.

Chapter 11

p.365, line 9 (ss 11.5.1)

Replace "(=0.832+0.0 54)" by "(=0.832+0.045)"

p.371, line -12 (Section 11.7)

Replace regression by classification.

Chapter 12

Figure 12.1 (p.377) & Figure 12.3 (p.380), figure legends

In these figures, females are in gray, and males in black.

Chapter 14

p.421, line -9 (l. -2 of Subsection 14.1.1)

Replace library() by attach().

p.426, line -11 (Section 14.2)

Following "The command", insert `datafile("bostonc")`

p.433, lines 7 & 11 (ss 14.3.4)

Replace "1Jan1990" by "1Jan1995".

p.462, Figure 14.2 (ss 14.11.3)

In the legend in the top left of the plotting area, replace $y = -1.112x^{2.274}$ by $y = 0.384\ x^{2.274}$.

p.463, line -8 (Section 14.12)

Replace `trellis.settings` by `par.settings`.

p.464, line 9 (ss 14.3.4)

Replace `levels (groups)` provides the legends by the levels of the `groups` argument supply the text strings.

Index of R symbols and functions, pp.485-490

On page 487, remove "189" from the entries against "identify".

On page 488, add the index entry
par.settings, 57, 58, 463

On page 489, under trellis.par.set, replace 462, 464 by 463.

Index of Terms, pp.491-500

p.494

Replace the heading at the top of column 1, i.e.
distribution (*cont.*)
by

examples, analyses & analysis issues (*cont.*)

Index of Authors, pp.501-502

References to pages 43 to 75 (Chapter 2) are mostly, in each case, too large by 1. Thus Aldrich, 64 should be Aldrich, 63.

Add Box & Jenkins, 299; Chatfield, 97; Hyndman et al, 299; McCullagh & Nelder, 283; Nightingale, 76; Ord et al, 299; Snijders & Bosker, 347; Tukey, 43.

Redundant spaces

Omit redundant spaces as follows:

p.48: `log10(c(1,5000*1000)` (line -7)
[NB also in line -2, omit two redundant spaces & `plot` (not `Plot`)]

p.196 final line: `oddbooks` (not `odd books`)

p.310 midpage: `names(ant111b.samp)`

p.463 : `trellis.par.set()` (line -13); `xyplot` (line -7), following `list` (lines -6, -5 and -4)

p.464 : `help(xyplot)` (line 19), `simpleKey()` (lines 11-12 and 17)
[NB also: "help" should be in typewriter font.]

[The inclusion of a space between a function name and the `(` that precedes the argument is contrary to established practice. It does no harm, however.]

Acknowledgement

We are indebted to Raymond Balise (Stanford University) for a number of these corrections.