1 Errata

x. In the paragraph beginning “In S-Plus and R, . . .”, add a new second sentence:

Most of our S (S-Plus and R) graphics examples use trellis graphics and quietly assume that you have explicitly opened a trellis device (see page 637, page 639 item 5, and page 640 Table B.2).

The page 637 reference is in this errata.
Add a new fourth sentence:

You may wish to open a trellis device in your .First function. See file splus.library/First-w-6-rmh-hp.s for an example in S-Plus.

xii. In the Acknowledgements, change “Each of gratefully” to “Each of us gratefully”

{Reminder:

18. The first line has several words in a narrow font. }

23. Change “occurences” to “occurrences”.


28. Change Equation (3.10) to:

\[ \sum_{x \leq \eta} p(x) \geq .5 \quad \text{and} \quad \sum_{x < \eta} p(x) \leq .5 \quad \text{for discrete distributions.} \]

35. The equation defining \( \rho_{ij} \) should be numbered (3.12). This is the equation referenced on line 3 of page 195.
36. add three words:
...bivariate normal runs in S-PLUS, but not R, with the function ...

37. Add this material after line 4:
It follows from the preceding formula that if $X_1, X_2, X_3, X_4$ are univariate random variables then
\[ \text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2) + 2 \text{cov}(X_1, X_2) \]
and
\[ \text{cov}(X_1 + X_3, X_2 + X_4) = \text{cov}(X_1, X_2) + \text{cov}(X_1, X_4) + \text{cov}(X_3, X_2) + \text{cov}(X_3, X_4) \]

38. Section 3.4.2, paragraph 2: change mean $mu$ and standard deviation $\sigma$. to mean $\mu$ and standard deviation $\sigma$.

40. Section 3.5, paragraph 2, last line should be statements are made with reference to the Student’s $t$ distribution.

44. Equation (3.12) should be renumbered to (3.13). The reference on page 91 to Equation (3.12) correctly points to page 44 and should be changed to point to the renumbered (3.13).

50. On line 2, change $\mu \geq 10$ to $\mu > 10$.

61. Revise Exercise 3.23d to read: Approximate the power of this test for the alternative $\mu_1 = 53$ by using the normal distribution as an approximation for the test statistic in part c, assuming $\alpha = .05$.

Revise Exercise 3.24d to read: Approximate the probability of committing a Type II error for the alternative $\mu_1 = 15$. Use the normal distribution to approximate the test statistic in part c, assuming $\alpha = .05$.

61. Revise the third sentence of Exercise 3.24 to read: Be aware that the cereal Quaker Oatmeal shows a missing value for carbohydrates. Be sure that you inform your data analysis package of the missingness and that the package does something sensible with it. Elimination of the observation is one possible response to missingness.

67. In two places, change $P_5 = 30$ to $P_5 = 30$.

75. Add at the end of the caption of Figure 4.9:
Variable pp.per.tv has two missing values. We notice this immediately in panel ppl.per.tv ~ ppl.per.phys, where the two points at ppl.per.phys = 25000 in the bottom three rows of the ppl.per.phys
column do not appear. Similarly these points are missing in Figures 4.10, 4.13, and 4.14.

76. See the caption for Figure 4.9 on the missing values in variable \texttt{pp.per.tv}.

79. See the caption for Figure 4.9 on the missing values in variable \texttt{pp.per.tv}.

80. See the caption for Figure 4.9 on the missing values in variable \texttt{pp.per.tv}.

91. The reference on page 91 to Equation (3.12) correctly points to page 44. It should be changed to point to (3.13) when we renumber the equation on page 44 to (3.13).

98.a Just below Equation (5.9), add the year: (Agresti and Caffo, 2000).

98.b Number the Agresti and Caffo confidence interval at the end of Section 5.4.1 as Equation (5.9a).

98.c The test version of Equation (5.9) is missing. Add a new paragraph at the end of Section 5.4.1:

To test the null hypothesis \( H_0: \hat{p}_1 - \hat{p}_2 \) the appropriate statistic is

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}
\]

(5.9b)

where \( \hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} \).

Notice the distinction between the standard error portions of Equations (5.9) and (5.9b). The standard error in the test statistic (5.9b) is calculated under the assumption that the null hypothesis is true. The larger standard error in (5.9) cannot utilize this assumption.

121. Exercise 5.13. Change “illusration” to “illustration”.

Exercise 5.13b. Change “5” to “47”.

152. Exercise 6.13. Change \texttt{contrasts(turkey$dist)} to \texttt{contrasts(turkey$diet)}.

159. The \texttt{data} step displayed in Table 7.1 is incorrect for the format of the \texttt{weightloss.dat} dataset. The online file (\texttt{mcomp/code/weightloss.sas}) is correct. The correct file display is
SAS (mcomp/code/weightloss.sas):

```sas
data weightloss;
  infile "&hh/datasets/weightloss.dat" firstobs=2;
  input loss @@; group="A"; output;
  input loss @@; group="B"; output;
  input loss @@; group="C"; output;
  input loss @@; group="D"; output;
  input loss @@; group="E"; output;
run;

proc anova;
  class group;
  model loss = group;
  means group / dunnett('D');
run;
```

164. In Table 7.4 we have improved the labeling of the interaction contrast by changing the label from \(A \times b\) to \(A.vs.B.by.amount\). The change was made by changing the files turkey.contrasts2.sas, turkey.contrasts2.lst, and turkey.contrasts2a.lst.

165. a. Revise the first complete paragraph to:

Figures 7.5 and 7.7 contain overprinting of the confidence lines and labels for several of their comparisons of level means. The overprinting in Figure 7.5 is due to almost identical mean values for levels B1 and A2. The overprinting in Figure 7.7 is a consequence of the almost identical mean values, now reflected as identical heights for the contrasts because the interaction of the \(A.vs.B\) and the \(amount\) comparisons is not significant. In situations with such overprinting, we augment the mean–mean display with a traditional S-Plus display of these same confidence intervals. This “tiebreaker” plot lists the contrasts in the same vertical order as in the mean–mean plot. The conclusions here, based on the fact that 9 of the 10 intervals lie entirely above zero, are

165. b. Add a new paragraph at the end of page 165:

Table 7.3 and Figure 7.4 show three of the user-defined contrasts to have negative estimates. Figures 7.7 and 7.8 show those contrasts to be reversed to have positive contrasts. We believe that multiple comparisons are most easily interpreted when the means are sequenced in numerical order (not lexicographic order), and consequently that all displayed contrasts should compare the larger value to the smaller value. That is, all displayed contrast values should be positive. Such reversal of the direction of a contrast creates no problem when assessing how contrasts
relate to zero so long as the reversal is noted. We note the reversal by appending a “−” to the names of the reversed contrasts.

171 The captions of Figure 7.9 and 7.10 should read: Sample sizes were 5 from populations A and D and 100 from populations B and C.

182. Missing line in first paragraph:
The means for the six groups are in file (datasets/pulmonary.dat).

195. \( \text{df}_{\text{Total}} \) should be \( n - 1 \).

\{(\text{Reminder: (conc/conc.tex), lines 717 and 720, \texttt{\textbackslash eqnarray} should not be *'ed. This equation on page 35 should be numbered because it is referenced on line 3 of page 195. rmh added errata items both for pages 35 and 195 for the new (3.12) and for pages 44 and 91 where the current (3.12) needs to be changed to (3.13).} \}

195. The line 3 reference to Equation (3.12) is intended to point to the currently unnumbered equation on page 35 that defines \( \rho_{ij} \). The reference should have parentheses around the number.

196. In the Table 8.3 formula for the Standard Error of \( \hat{\beta}_0 \), change \( \sigma \) to \( \hat{\sigma} \).
In Table 8.4, and again in Equation (9.27) on page 249, the formula for \( R^2_{\text{adj}} \) is incorrect. It should read:

\[
1 - \left( \frac{n - 1}{n - p - 1} \right) (1 - R^2)
\]
or, equivalently

\[
R^2 - \left( \frac{p}{n - p - 1} \right) (1 - R^2)
\]

We use \( p = 1 \) to denote the number of degrees of freedom for regression in the ANOVA table in Table 8.1. More generally, in Chapter 9 and beyond, \( p \) is the number of predictor \( x \)-variables in the model. The notation is presented in Section 9.2.

198. Add this paragraph immediately before the present last paragraph on this page:
It follows from this that a \( 100(1 - \alpha)\% \) confidence interval on \( \beta_1 \) is

\[
\hat{\beta}_1 \pm t_{df, \alpha/2} \hat{\sigma}_{\hat{\beta}_1}
\]
where \( df = df_{\text{Res}} \) degrees of freedom.

199. Add this sentence prior to the last sentence in Section 8.3.4:
Nevertheless, on occasion we use the term “error” as a synonym for “residual” to match the continued use by SAS of Error Mean Square rather than the preferred Residual Mean Square.

{Reminder: Burt added this new item and removed all December 2004 changes in this file from Error to Residual. The list of these Dec 2004 changes is retained in errata.notes.122204.}

202. In Table 8.6 caption, change “surrounding” to “surrounding”.

204. In Figure 8.4 title, change 0.95 to 95%.

Chapters 9 and 11. There are six occurrences of “threshhold” when it should be “threshold”.

{Reminder: When first defining a hat matrix we say on page 219: places a hat \( \hat{\beta} \). This rationale for the terminology hat matrix is unnecessarily repeated in two places: on page 220 and page 315. I think once is enough.

This redundancy doesn’t bother rmh.}

212. New Exercise:

8.11 The estimator of the predicted value when \( x = x_0 \) is given by \( \hat{y} + \hat{\beta}_1(x_0 - \bar{x}) \). The estimated variance of the estimator is

\[
\sigma^2 \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) = h_0 \sigma^2 \text{ from page 201 item 1.d and Equation (8.18).}
\]

i. Referring to \( \hat{y} = \frac{\sum y_i}{n} \) and \( \hat{\beta}_1 \) as defined in Equation (8.7), use results in Section 3.3.5 (specifically the additional equations listed in the errata for page 37) to establish that these estimators are uncorrelated. That is, show that

\[
cov(\hat{y}, \hat{\beta}_1) = 0
\]

ii. Use part 0.i to show

\[
\text{var}(\hat{y}|x_0) = \text{var}(\hat{y}) + \text{var} \left( \hat{\beta}_1(x_0 - \bar{x}) \right) + 2 \text{cov} \left( \hat{y}, \hat{\beta}_1(x_0 - \bar{x}) \right) \\
= \frac{\sigma^2}{n} + \frac{\sigma^2}{\sum (x_i - \bar{x})^2} (x_0 - \bar{x})^2 + 2(0)
\]

249. \( R_{\text{adj}}^2 \) is also misdefined on page 249. See the correction at page 196.

280. Table 10.6

a. Typo:

\[
tmp.min <- \text{apply(abs(tmp), 2, min)}
\]
should be

tmp.min <- apply(abs(tmp.c), 2, min)
The file regbb/code/fabricwear.s is correct.

b. The algorithm for simplifying the display of the orthogonal contrasts works only for an even number of levels. A corrected algorithm that detects the zero and almost zero values and doesn’t use them is posted in a new file regbb/code/fabricwear2.s.

286. Figure 10.6. Caption has wrong formula. Correct formula is:

    ancova(Sodium ~ Type, x=Calories)

The intent of the notation is twofold: The arithmetic of the analysis is based on the one-way ANOVA of Sodium ~ Type. The graph is more complex. The points in the graph show $y=$Sodium plotted against $x=$Calories separately for each level of Type. The horizontal line in each panel is the mean of the levels of Sodium at each level of Type.

287. Figure 10.7. Caption has wrong formula. Correct formula is:

    Sodium ~ Calories, groups=Type

The intent of the notation is twofold: The arithmetic of the analysis is based on the simple linear regression of Sodium ~ Calories. The graph is more complex. The points in the graph show $y=$Sodium plotted against $x=$Calories separately for each level of Type. The common regression line in all panels ignores Type.

304. Table 11.4 uses the wrong formula anova(rent.lm12m$aov) The correct formula is anova(rent.lm12m)

306. The dffits panel didn’t label all the points that crossed the threshold. The corrected Figure 11.6 is here.

316. The dffits panel didn’t label all the points that crossed the threshold. The corrected Figure 11.16 is here.

322. Delete the sentence immediately preceding Section 11.3.6. No cases exceeds Cook’s flag.

344. In the second paragraph, change to blocking factor {\tt id}.

{Reminder: Notice how the type is scrunched at the end of the first line of page 373. (It overflowed into the margins on the dvi file we sent. They changed the font to make it fit into the margins.) In the next printing, I propose to change the sentence as: The file (errata/transcript/feed2a.lst) includes the level means of temp and supp. }

353. In paragraph 3, change Table 12.1 to Table 12.3.
FIGURE 11.6. Diagnostics from ANCOVA (rnt.alf/rnt.till) ~ cow.dens | lime. See Table 11.4 and Figure 11.4.

(regc/code/rent4.s), (regc/figure/rent.diag.lm12m.eps.gz)

{Reminder: Files supporting the analysis of the (datasets/skateslc.dat) exercise on page 377 and the (datasets/ironpot.dat) exercise on page 378 are in (tway/code/*) and (tway/figure/*). They should be moved to (tway/answer/*) in future postings of the book’s files.}

366. {Reminder: File (tway/code/rhiz-clov-mmc.s) does not currently use the correct critical value for the simple effects comparisons. Figures 12.13, 12.14, 12.15 need narrower confidence intervals. They need \[ q_{tukey(1-.05/2, 6, 48)}/sqrt(2) \]. Currently the \[ sqrt(2) \] is missing.
FIGURE 11.16. DFFITS for Model (11.1) for rent data.
(regc/code/rent4b.s), (regc/figure/dffits.eps.gz)

See (h2/mmc/code/rhiz-clov-mmc.s) for the revision. The discussion
needs to be modified to match. See dsc-2005 for the revision. }

372. In line 1 of the last paragraph, change Table 12.18 to Table 12.19.

382. In paragraph 3, line 5, change “respoinse” to “response”.

384. Replace the file (dsgn/transcript/cc176-1.st) in Table 13.1, which
uses a model containing the covariate and the three-way interaction
of the factors. Use file (dsgn/transcript/cc176-full.st) which also
contains the interaction of the covariate and the factors. The four-way
interaction is used as the error term. The caption is now:

Muscle data. ANCOVA and adjusted means. The covariate wt.n, the
linear effect of n.treats, and the current are the significant treatment
effects. The covariate wt.n shows strong interaction with the current
factor. The table shows possible interaction with two other factors, but
further investigation allowed us to suppress them from the final model.

388. Figure 13.5 corresponds to the final model

cc176.aov <- aov(wt.d ~ rep + wt.n + n.treats + wt.n*current, data=cc176)

This model suppresses the non-significant effects from the ANOVA table
on page 384. The table is shown in (dsgn/transcript/cc176-model.st).
389. Last paragraph, next to the last line should begin with \((r - 1)^2 - 3(r - 1) = (r - 1)(r - 2)\) df.

404. In Equation (13.2), change \(mn\sigma^2_T\) to \(nm\sigma^2_T\).

\{Reminder: This change is aesthetic only, to conform with what is in Table 13.11. \}

417. Typographical error in section header 13.6.

13.6 Sequential and Conditional Tests

The error also appears in the page head on pages 417 and 419, and the listing in the Table of Contents on page xix.

442–444. File (dsgstwo/code/2.8-2-full-generate.s)

- reads (datasets/2.8-2.dat),
- generates the factors from the text representation of the treatment combinations,
- randomly generates a response variable,

and

- writes (datasets/2.8-2-full.dat).

File (dsgstwo/code/2.8-2-full.s) reads the data in (datasets/2.8-2-full.dat) and writes (dsgstwo/transcript/2.8-2-full.st).

The dummy variables are orthogonal, their correlation matrix is the identity, and the correlation matrix of the estimated coefficients is also the identity.

File (dsgstwo/code/2.8-2-full.sas) reads the data in (datasets/2.8-2-full.dat) and writes (dsgstwo/transcript/2.8-2-full.lst).

443. The last item “adcdfg” in block 2 of Table 14.9 is incorrect. The correct value is “abcdfg”.

File (datasets/2.8-2.dat) shows the corrected Table 14.9.

443. The second line in Table 14.10 should be

Blocks 3 blocks are aliased with ACF, BDG, and CDH.
444. The SAS sample code is unnecessarily complex. It would be better to use the SAS syntax (the @2) for generating the complete set of 2-factor interactions.

SAS (dsgntwo/code/2.8-2b.sas):
model y = blocks a|b|c|d|e|f|g|h@2 ;

467. In last paragraph, change $\hat{s}^2$ to $s^2$.

472. Figure 14.10. Revise the captions for a and b to (using R notation, with the response variable to the left of the ~):
a. final.adj ~ order | standing + sex
b. final.adj ~ standing | order + sex

474. Add a new paragraph:
There are several difficulties we see in Figure 14.11:
a. There are two factors, driver age and passenger presence. They are not treated symmetrically.
b. The marginal main effects of the factors are not displayed.
c. The response variable is a rate. Barplots were designed for counts and don’t make much sense for rates.
d. The majority of the plotting surface is used to display the region where the data doesn’t appear.
e. The symmetry in uncertainty is hidden by the asymmetry between the heavy bar below each observed value and the empty region above.
f. Superimposed error bars (not shown here) on a barplot add to the asymmetry.
Figure 14.14a addresses all these difficulties.

509. Table 15.9: the formatting code “[-1.5ex]” does not belong there.

510. The caption in Table 15.10 places the file (twtb/transcript/intubate.s) in the (twtb/transcript) subdirectory instead of the correct (twtb/code) subdirectory. The file is now in both places—the first to agree with the text and the second because that is where it really belongs.

515. Line −8, change “is less than” to “is greater than”.

525. The transcript portion of the Table should be headed SAS, not S-PLUS.

530. Line 3, change $23 \times 6$ to $23 \times 6$. 
538. There is no Equation (17.8).

538–539. Replace last sentence on p.538 (beginning with “We can rewrite ....”) and Equation (17.11) on p.539 with:

Substitute Equation (17.7) into Equation (17.10) and differentiate with respect to $\beta_0$ and $\beta_1$ to get:

$$\frac{\partial}{\partial \beta_0} \ell(\beta|y; X) = \sum_{i=1}^{n} (y_i - p_i) = 0$$

$$\frac{\partial}{\partial \beta_1} \ell(\beta|y; X) = \sum_{i=1}^{n} x_i (y_i - p_i) = 0$$

(17.11)

540. The code snippet should be replaced by

S-PLUS: logi/code/spap.glmd.s:

```
coef(summary(spacshu.bin.glm))
```

544. The next to the last sentence should read: However, we choose to retain grade because a rerun of the model without grade actually increases the $p$-value of acid.ph.

559. In Exercise 17.3, part a, change the second sentence to: Estimate the change in the odds of response = 1 resulting from one additional day of radiotherapy.

577–581. The product example is defective and will be replaced in a future listing of errata.

Lines 123–148 of product.st indicate that the optimizer didn’t converge. Its values are similar to those for product.lst which apparently converged. But product.lst reveals that the parameter estimates are closely correlated, all approximately ±1. While the statement on page 580 that ARIMA(2,1,1) is the best model may be true, the high correlation among the parameter estimates makes the model unreliable.

579. In the caption for Figure 18.3, change

```
ARIMA(p,1,q)
```
to

```
ARIMA(p, 1, q)
```

589. In instruction b., replace “Describe” with “Describe”.

618. paragraph 3 line 10. “AS” doesn’t belong there.

629. On line 4, change

“each with a subdirectories”
to
“each with subdirectories”

635 and 637. In Step 3, paragraph 2 on both pages, replace the sentences

The HH library splus.library/HH should be the eighth item on
the search list. We need to be the eighth item because we have
replacement function definitions for several of the R-supplied
functions in the lattice and grid packages.

with

The HH library hh/splus.library/HH/.RData should appear
prior to the lattice and grid packages on the search() list
because we have replacement function definitions for several of
the R-supplied functions in the lattice and grid packages.

637. Add a new second paragraph to Section B.2.

Many of our graphics functions modify the trellis parameters. They therefore require that the user has explicitly opened
a trellis.device. In S-PLUS you may use one of the function
calls listed in Table B.2. We do this for ourselves in our .First
function. See for example, file

hh/splus.library/First-w-6-rmh-hp.s

In R you may use the color scheme shown in file

hh/splus.library/trellis.device.hh.r

If you do not explicitly open a trellis device, you may occasionally
see the uninterpretable S-PLUS error message

Problem in trellis.par.set:
No data to interpret as logical value:
if(p[i] != q[i])
stop("inconsistent component names")

which means

Current graphics device is not a trellis device.
Please open a trellis device, for example with:
trellis.device()

638. Replace the sentence

You can source them, for example source(hh("grap/code/grap.read.le.s")), or open
them in your editor and execute one line at a time (for example, by using C-c C-n in Emacs, or by copy and paste from an
editor that doesn’t work smoothly with S).
with

Files defining functions in the HH library, primarily the files (`splus.library/*.s`) and (`splus.library/*.r`), can be entered by sourcing them. See Section B.9 for details. Most users will not need to `source()` these files as we distribute the (`splus.library/HH/.Data`) directory and the (`splus.library/HH/.RData`) file.

Files illustrating use of the functions and files that construct the Figures and Tables in the book, that is, most of the files in the individual chapters with names of the form (`*/code/*.s`) and (`*/code/*.r`), are designed to be opened in your editor and executed one line at a time (for example, by using `C-c C-n` in Emacs, or by copy and paste from an editor that doesn’t work smoothly with S). Many of these files produce trellis graphs. The graphs will not be displayed should the file be sourced. The graphs will be displayed only when individual lines are pasted into the Commands window.

We assume in the construction of our (`*/code/*.s`) and (`*/code/*.r`) files that the graph window retains all previous graphs in the current session. We describe how to do this in S-Plus for Windows on page 638 in Section B.3.1.e. We describe how to do this in R in Section B.10.

{Reminder: At this writing Section B.10 appears only in the errata.}

648. New section:

B.10 R—Recommended Options

1. We recommend that you turn on graphical history in your interactive graphical device. We therefore recommend that you include the line

   `options(graphics.record = T)`

in your `.First()` function. We have the line in a comment in our sample files (`splus.library/First-u-r.s`) and (`splus.library/First-w-r.s`). The line is uncommented in (`splus.library/First-w-r-rmh-hp.s`), the file which has been customized for my personal machine.

676. Excel 2003 has made some improvements to its variance algorithm. It now gets the correct answer through $10^{14}$. It also gets the correct answer for $10^{15}$, but it is not obvious how it could have done so since it doesn’t
store the data correctly for $10^{15}$. It still shows disastrous overflow for $10^{16}$.

\[
(1 : 3) + 10^k
\]

<table>
<thead>
<tr>
<th>$k$</th>
<th>$10^k$</th>
<th>$1 + 10^k$</th>
<th>$2 + 10^k$</th>
<th>$3 + 10^k$</th>
<th>var $((1 : 3) + 10^k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$10^4$</td>
<td>10001</td>
<td>10002</td>
<td>10003</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>$10^{14}$</td>
<td>10000000000001</td>
<td>10000000000002</td>
<td>10000000000000003</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>$10^{15}$</td>
<td>1000000000000000</td>
<td>10000000000000000</td>
<td>10000000000000000000</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>$10^{16}$</td>
<td>10000000000000000</td>
<td>100000000000000000</td>
<td>10000000000000000000</td>
<td>3.610^{16}</td>
</tr>
<tr>
<td>17</td>
<td>$10^{17}$</td>
<td>100000000000000000</td>
<td>10000000000000000000</td>
<td>10000000000000000000000</td>
<td>0</td>
</tr>
</tbody>
</table>

Microsoft® Excel 2003 (11.6113.5703)

---

692. Apply the equation numbering (F.6) to the equation for $E(x'Ax)$.

695. Add to the end of Section F.4.10: All projection matrices are idempotent matrices. That is, they are unchanged when multiplied by themselves: $P_X P_X = P_X$.


713. (Harrison et al., 2004) is Spine, 29:2485–2492.

715. Uppercase “Tukey” in (Hsu and Peruggia, 1994).

717. The ESS site was moved.  
(Rossini et al., 2004b): The www.analytics address is obsolete. It has been replaced with http://ESS.R-project.org

---

2 Software Status—hh071505-CD.tar.gz

2.1 S-PLUS

Most of the graphs in the book were prepared with S-PLUS. Therefore all *.s files in the online files work well.
2.2 R

R is similar, but not identical, to S-PLUS. The most important consequence of the difference, and the major reason for the different state of our S-PLUS and R code, is that we have designed many new graphical techniques that we implemented as panel functions in the trellis graphics system of the S language. The internal workings of the panel functions for S-PLUS trellis and R lattice are very different and often require serious rethinking of the programming details.

We began systematic checking of all our *.s files with R after the manuscript went to press. Some of our *.s files worked in R with no change. Some required minor changes—which we were able to make with an if.R() statement in the *.s file in many cases. Other *.s files required more extensive changes—which we made by writing parallel *.r files. The new files are listed in the table below.

About 96% of the R examples work as of July 15, 2005. The rest will work within a few months. There are two classes of examples that don’t work yet in R.

1. Not yet done.

   a. The ARIMA-trellis functions depend on the S-PLUS rts class of time series, a class that is not exactly matched in R. The functions are heavily dependent on new panel functions. We have not yet duplicated the new panel functions in R.

   b. Some of the xysplom examples, primarily those depending on panel.cartesian, have not yet been translated to the R lattice coding conventions.

   c. The MMC plots (Mean–Mean Multiple Comparisons) work for many cases in R. Specifically, they work for one-way designs and factorial designs. They seem to work for designs with covariates as long as there is no interaction between a covariate and a factor. We believe the difficulty is due to an incompleteness in the simint function in the R multcomp package. We will confirm whether that is indeed the source of the problem, and if so, resolve it with the R developers.

2. Many examples will work after we make simple adjustments to the files. We have not yet made all those adjustments. Instead, we have focussed our attention on making the new techniques work at all.

   a. We prefer boxplots with the response on the vertical axis. The mechanics of doing that are different in S-PLUS (we use the t.trellis function that we provide) and in R (we use the horizontal=FALSE
argument that they provide). I think I made that simple adjustment to all of the example files, but it is possible that I missed some.

b. Many of the R examples work, and provide technically correct results, but the graphs are not yet pretty. We haven’t yet made the adjustments for different formatting decisions between the S-PLUS and R developers.

c. Some files we just haven’t gotten to yet. These are primarily in Chapters 16 and 17. We don’t expect any difficulty, we just haven’t verified that they work or made the (anticipated) minor adjustments.

We will be distributing the R library as an .RData file to be attached. The HH library is not currently a standard R package.

2.3 Documentation

All Figures and Tables in the HH text have a reference to the S-PLUS, R, or SAS code files that generated them. These commented examples of use of the HH library are the primary documentation.

There are not currently any formal help files for the HH functions. The implicit documentation is in the HH book and in the comments in the *.s files defining the functions and illustrating their use.

3 Additional Files

These files are not listed in the book, but are included in the online files hh071505-CD.tar.gz. The notation “incomplete” in the table means we know the files are not yet complete.
con clad conc code tv-graphs.r
con clad conc code conc.oc.r
datasets/2.8-2-full.dat
datasets/2.8-2.dat
dsgn code turkey.f2.r
dsgn code vulcan.sas
dsgn transcript vulcan.lst
dsgntwo code 2.8-2-full-generate.s
dsgntwo code 2.8-2-full.s
dsgntwo code 2.8-2-full.sas
dsgntwo code 2.8-2b.sas
dsgntwo code apple.sas
dsgntwo code testscore.sas
dsgntwo transcript/2.8-2-full.lst
dsgntwo transcript/2.8-2-full.st
dsgntwo transcript/apple.lst
dsgntwo transcript/testsascore.lst
grap code draft70mn-read.r
grap code grap.f10.r
grap code grap.f3.r
grap code grap.f5.r
grap code grap.f6.r
grap code grap.read.1e.r
grap code njgolf-read.r
iinf code har3.sas
iinf code teachers2.r
iinf transcript/har3.lst
logi code budworm.sas incomplete
logi code lymph.sas incomplete
logi code spaceshu.sas
logi transcript/budworm.lst incomplete
logi transcript/lymph.lst incomplete
logi transcript/spaceshu.lst incomplete
mcomp code catalyst-mmc3.r
mcomp code mmc.explain.r
mcomp code turkey-mmc.r
mcomp code weightloss.r
mcomp figure mmc1-b0.eps.gz
tway/code/plasma.sas
tway/transcript/plasma.lst
tway/code/rhiz-bwplot.t.r
tway/code/rhiz-bwplot.ti.r
twtb/code/intubate.s
nparcode/vocab2.r
mcomp/code/pulmonary2.s
dsgntwo/code/jsm2004.crash.s
regbb/code/jsm2004.hotdog.s
dsgn/code/jsm.cc176.s
regc/code/regression-diagnostics.s