Appendix A

Recommended Books

A.1 Books on R

There are currently no books written specifically for R, although several guides can be downloaded from the R web site.

R is very similar to S-plus so most material on S-plus applies immediately to R. I highly recommend Venables and Ripley (1999). Alternative introductory books are Spector (1994) and Krause and Olson (2000). You may also find Becker, Chambers, and Wilks (1998) and Chambers and Hastie (1991), useful references to the S language. Ripley and Venables (2000) is a more advanced text on programming in S or R.

A.2 Books on Regression and Anova

There are many books on regression analysis. Weisberg (1985) is a very readable book while Sen and Srivastava (1990) contains more theoretical content. Draper and Smith (1998) is another well-known book. One popular textbook is Kutner, Nachtschiem, Wasserman, and Neter (1996). This book has everything spelled out in great detail and will certainly strengthen your biceps (1400 pages) if not your knowledge of regression.
Appendix B

R functions and data

R may be obtained from the R project web site at www.r-project.org.

This book uses some functions and data that are not part of base R. You may wish to download these functions from the R web site. The additional packages used are

MASS leaps ggobi ellipse nlme

MASS and nlme are part of the “recommended” R installation so depending on what installation option you choose, you may already have these without additional effort. Use the command

> library()

to see what packages you have. The MASS functions are part of the VR package that is associated with the book Venables and Ripley (1999). The ggobi data visualization application may also need to be installed. This may be obtained from www.ggobi.org This is not essential so don’t worry if you can’t install it. In addition, you will need the splines, mva and lqs packages but these come with basic R installation so no extra work is necessary.

I have packaged the data and functions that I have used in this book as an R package that you may obtain from my web site — www.stat.lsa.umich.edu/~faraway. The functions available are

halfnorm Half normal plot
Cpplot Cp plot
qqnorml Case-labeled Q-Q plot
maxadjr Models with maximum adjusted R^2
vif Variance Inflation factors
prplot Partial residual plot

In addition the following datasets are used:

breaking Breaking strengths of material by day, supplier, operator
cathedral Cathedral nave heights and lengths in England
chicago Chicago insurance redlining
chiczip Chicago zip codes north/south
chmiss Chicago data with some missing values
coagulation Blood coagulation times by diet
corrosion Corrosion loss in Cu-Ni alloys
eco Ecological regression example
gala Species diversity on the Galapagos Islands
Where add-on packages are needed in the text, you will find the appropriate `library()` command. However, I have assumed that the `faraway` library is always loaded. You can add a line reading `library(faraway)` to your `.Rprofile` file if you expect to use this package in every session. Otherwise you will need to remember to type it each time.

I set the following options to achieve the output seen in this book

```r
> options(digits=5,show.signif.stars=FALSE)
```

The `digits=5` reduces the number of digits shown when printing numbers from the default of seven. Note that this does not reduce the precision with which these numbers are internally stored. One might take this further — anything more than 2 or 3 significant digits in a displayed table is usually unnecessary and more important, distracting.
Appendix C

Quick introduction to R

C.1 Reading the data in

The first step is to read the data in. You can use the `read.table()` or `scan()` function to read data in from outside R. You can also use the `data()` function to access data already available within R.

```r
> data(stackloss)
> stackloss
   Air.Flow Water.Temp Acid.Conc. stack.loss
 1    80       27       89          42
 2    80       27       88          37
   ... stuff deleted ...
 21   70       20       91          15
```

Type

```r
> help(stackloss)
```

We can check the dimension of the data:

```r
> dim(stackloss)
[1] 21 4
```

C.2 Numerical Summaries

One easy way to get the basic numerical summaries is:

```r
> summary(stackloss)
   Air.Flow Water.Temp Acid.Conc. stack.loss
Min.   :50.0  Min.   :17.0  Min.   :72.0  Min.   : 7.0
1st Qu.:56.0  1st Qu.:18.0  1st Qu.:82.0  1st Qu.:11.0
Median :58.0  Median :20.0  Median :87.0  Median :15.0
Mean   :60.4  Mean   :21.1  Mean   :86.3  Mean   :17.5
3rd Qu.:62.0  3rd Qu.:24.0  3rd Qu.:89.0  3rd Qu.:19.0
Max.   :80.0  Max.   :27.0  Max.   :93.0  Max.   :42.0
```
We can compute these numbers separately also:

```r
> stackloss$Air.Flow
[1] 80 80 75 62 62 62 62 58 58 58 58 58 58 50 50 50 50 50 56 70
> mean(stackloss$Air.Flow)
[1] 60.429
> median(stackloss$Air.Flow)
[1] 58
> range(stackloss$Air.Flow)
[1] 50 80
> quantile(stackloss$Air.Flow)
   0%  25%  50%  75% 100%
50  56  58  62  80
```

We can get the variance and sd:

```r
> var(stackloss$Air.Flow)
[1] 84.057
> sqrt(var(stackloss$Air.Flow))
[1] 9.1683
```

We can write a function to compute sd’s:

```r
> sd <- function(x) sqrt(var(x))
> sd(stackloss$Air.Flow)
[1] 9.1683
```

We might also want the correlations:

```r
> cor(stackloss)
     Air.Flow Water.Temp Acid.Conc. stack.loss
Air.Flow  1.00000   0.78185   0.50014   0.91966
Water.Temp 0.78185   1.00000   0.39094   0.87550
Acid.Conc. 0.50014   0.39094   1.00000   0.39983
stack.loss 0.91966   0.87550   0.39983   1.00000
```

Another numerical summary with a graphical element is the stem plot:

```r
> stem(stackloss$Air.Flow)

The decimal point is 1 digit(s) to the right of the |

5 | 000006888888
6 | 22222
7 | 05
8 | 00
C.3  Graphical Summaries

We can make histograms and boxplot and specify the labels if we like:

```r
> hist(stackloss$Ai)
> hist(stackloss$Ai,main="Histogram of Air Flow",
   xlab="Flow of cooling air")
> boxplot(stackloss$Ai)
```

Scatterplots are also easily constructed:

```r
> plot(stackloss$Ai,stackloss$W)
> plot(Water.Temp ˜ Air.Flow,stackloss,xlab="Air Flow",
   ylab="Water Temperature")
```

We can make a scatterplot matrix:

```r
> plot(stackloss)
```

We can put several plots in one display

```r
> par(mfrow=c(2,2))
> boxplot(stackloss$Ai)
> boxplot(stackloss$Wa)
> boxplot(stackloss$Ac)
> boxplot(stackloss$s)
> par(mfrow=c(1,1))
```

C.4  Selecting subsets of the data

Second row:

```r
> stackloss[2,]
   Air.Flow Water.Temp Acid.Conc. stack.loss
2   80     27    88    37
```

Third column:

```r
> stackloss[,3]
[1]  89  88  90  87  87  93  93  87  80  89  88  82  93  89  86  72  79  80  82  91
```

The 2,3 element:

```r
> stackloss[2,3]
[1] 88
```

c() is a function for making vectors, e.g.

```r
> c(1,2,4)
[1] 1 2 4
```
Select the first, second and fourth rows:

```r
> stackloss[c(1,2,4),]
Air.Flow Water.Temp Acid.Conc. stack.loss
1  80  27  89  42
2  80  27  88  37
4  62  24  87  28
```

The `:` operator is good for making sequences e.g.

```r
> 3:11
[1] 3 4 5 6 7 8 9 10 11
```

We can select the third through sixth rows:

```r
> stackloss[3:6,]
Air.Flow Water.Temp Acid.Conc. stack.loss
3  75  25  90  37
4  62  24  87  28
5  62  22  87  18
6  62  23  87  18
```

We can use `"-"` to indicate "everything but", e.g all the data except the first two columns is:

```r
> stackloss[,-c(1,2)]
Acid.Conc. stack.loss
1  89  42
2  88  37
... stuff deleted ...
21  91  15
```

We may also want to select the subsets on the basis of some criterion e.g. which cases have an air flow greater than 72.

```r
> stackloss[stackloss$Ai > 72,]
Air.Flow Water.Temp Acid.Conc. stack.loss
1  80  27  89  42
2  80  27  88  37
3  75  25  90  37
```

C.5 Learning more about R

While running R you can get help about a particular command - eg - if you want help about the `stem()` command just type `help(stem)`.

If you don’t know what the name of the command is that you want to use then type:

```r
help.start()
```

and then browse. You may be able to learn the language simply by example in the text and referring to the help pages.

You can also buy the books mentioned in the recommendations or download various guides on the web — anything written for S-plus will also be useful.
Bibliography


