Chapter 1
Introduction to using R with Mind on Statistics

Manual overview

The purpose of this manual is to help you learn how to use the free R software to perform the graphs, simulations, and data analyses presented by Mind on Statistics. This chapter will describe how to obtain R and give a brief introduction to the software. The following chapters will provide specific commands for the book’s examples.

Line-by-line examples of R code will be provided for many of the examples demonstrated in the text. Two methods will be used to demonstrate the R code. The first method uses screen copies. The screen copy shows exactly what you will see when running R. The red text shown in the screen copy is what you type. You will not need to type the initial “>” prompt at the start of each line. The blue text in the screen copy is output produced by R. A simple example is the first bullet under “Basic Commands” of this chapter. The second method is simply a display of the text you would type. You would type the command after the “>” prompt. R would respond after you type the [Enter] or [Return] key. An example of this is the third bullet under “Basic Commands” of this chapter.

What is R?

R is a computer language and environment that was developed with statistical graphics and analysis in mind. Consequently it is commonly thought of as a statistical software package, like the proprietary Minitab and SPSS packages. In the growing atmosphere of free software, scientists are constantly making available new packages that enable R to perform very advanced modern statistics. This manual, however, will focus on the more elementary aspects of R needed to learn statistical concepts and successfully perform the statistics that you are likely to encounter in your future careers.

There are several consequences of R being free software developed by scientists for scientists. First of all, it is very powerful. If you decide to continue in a career that depends heavily on statistics such as economics, biology, medicine, marketing, etc., R will allow you to develop your own statistical functions specific to your own immediate needs. Secondly, it was created as a tool for scientists rather than for mass marketing to make money. Thus it is line command driven and lacks features as pull down menus and point-and-click commands. This results in a software has a high “nerd factor” as you will notice when looking at the help commands and manuals. S-Plus, a proprietary software package, is almost identical to R with respect to line commands, but includes pull down menus and some point-and-click commands.

Obtaining R

R is freely available via the website www.r-project.org as are its online manuals. To install R on your computer, follow the links which originate from the “CRAN” link found on the left-hand side of the R project homepage. This link leads you to a web page where you select the location of a mirror site closest to your location; e.g., University of California at Berkeley, California, USA. If you are using Microsoft Windows, for example, you would then click on the “Download R for Windows” link, followed by the “base” link, and then the “Download R 2.13.1 for Windows” link to install R.
Basic commands

The most essential features and commands to keep in mind to use R are:

- R is a line command driven software with the commands typed at the > prompt followed by the [Enter] or [Return] key. Values are assigned using the R the two key arrow <-. It is created by typing < followed by a dash. To see the value of a variable, simply type it and hit [Enter]. The following screen copy shows the basic command of assigning the value 10 to the variable n.

(Note: The bracketed [1] is used to count the values in the output as you will notice when working with larger datasets.)

- The # sign is used to add comments to a command line. R ignores everything on the command line typed after #. For example, at the prompt, type the following command to assign 4, 6, and 3 to the variable “heights”.

> heights <- c(4,6,3) #c() assigns a string of numbers

Although comments are provided throughout this manual to give extra instruction, it is not necessary for you to type the comments to run the commands.
R commands are followed by parentheses with variables and options put within the parentheses. Example:

```r
> sort( heights )
[1] 3 4 6
```

Typing the command without the parentheses will result in the software code flashing on the screen, but causes no harm.

- **R is case sensitive** which means that R will distinguish between the variables `Heights` and `heights`.
- To list what objects you have available, use the `ls()` command. Example:
  ```r
  > ls()
  [1] "heights" "Heights" "n"
  ```
- To delete an object use the `remove()` command. Example:
  ```r
  > remove( heights )
  > ls()
  [1] "Heights" "n"
  ```
- If you want to change a previous command you can hit the up arrow key and edit your old commands.
- If you type `[Enter] before a command is completed, R will go to the next line and respond with a “+” to denote the command is not finished and you can continue typing. To terminate the command, type the [Escape] or [Esc] key.
- Help can be obtained using either the pull-down help menu, using the function `help( function )` or the `help.search("key word")` commands. For example, `help(boxplot)` will provide information about the function `boxplot`. To get a list of commands that regard regression, try `help.search("regression")`.
- To quit R, type `q()`.

**Importing data from the Mind on Statistics cd into R**

R is best designed to import and export text data. The *Mind on Statistics* cd provides R data. First you need to install the datasets from the *Mind on Statistics* cd to your computer. First, go to “Install Datasets” and at the next level also click on “Install Datasets”. Select “R” to load the data on your computer. For this manual, the data were saved to the C drive in the location C:\Mind on Statistics\R.

The `load()` command is used to import data. R uses forwards rather than backward slashes to denote a folder location. Try importing the Temperature dataset.

```r
> load("c:/Mind on Statistics/R/temperature.RData")
```

**Storing data: Objects, vectors, matrices, and data frames.**

Anything stored by R is called an object. Thus the function `sort( )` is an object, as were the variables `height` and `n` in the previous examples. We will focus primarily on objects that store data, such as did `height` and `n`. Data will most commonly be stored in either vectors or data frames.

A vector is simply a string of numbers. `n` is a vector of length 1 and `heights` a vector of length 3. If you are typing a variable with only a few data values you will often simply type the values into a vector. R allows mathematical operations to be carried out on an entire vector. Example:
Elements of a vector can be specified by use of the hard brackets []. Either a single element or more than one can be specified. Examples:

```r
> x <- c( 5,2,4 )
> x + 6
[1] 11 8 10
```

A matrix is a string of vectors of the same mode (numeric, character, or factor) and of the same length bound together. The dimension of a matrix is typically described by its number of rows and columns. Analogous to a vector, the elements within a matrix are described by their row and column position in the matrix; e.g., X[ row, column ]. Leaving either the row or column unspecified is the same as specifying them all. Examples:

```r
> X <- c(2,4,6,8,10,12,14,16,18,20,22,24) #vector
> X <- matrix( X, nrow=3, byrow=F ) # turn into matrix
> X
[1,]   2   8  14  20
[2,]   4  10  16  22
[3,]   6  12  18  24
> X[2,3] # row 2, column 3
[1] 16
> X[1:2, 3:4] # rows 1 through 2 and columns 3 and 4.
     [,1] [,2]
[1,]  14  20
[2,]  16  22
> X[1,] # row 1, all columns
[1] 2 8 14 20
> X[,3] # all columns, row 3
[1] 14 16 18
> X+3 # add 3 to all values of X
[1,]   5  11  17  23
[2,]   7  13  19  25
[3,]   9  15  21  27
```

A data frame is like a matrix, but one column of the data frame may consist of numbers and another column words. The load() function automatically puts the data into a data frame.
The first column (1, 2, …, 20) are simply the row labels while the data are the columns listed under “c”, “l”, etc. Columns within a data frame can be accessed directly using the $ symbol.

The variables within a data frame can be accessed more directly by attaching the data frame. For example, suppose you typed j. R would say that it could not find the variable. That is because it looks in the directory (displayed by ls() ) and sees the data frame temperature, but does not look inside temperature for the variable j. Using the attach(temperature) command tells R to also look inside data frame temperature if it cannot find j in the directory. Once finished using the dataset, detach the data frame. Examples:

```r
> j
Error: Object ‘j’ not found
> attach(temperature)
> j
> detach("temperature")
> j
Error: Object ‘j’ not found
```
Editing data

Objects in R can be edited using the `edit()` function where the output is assigned either to a new object or the original object. A spreadsheet editor will appear for data frames and a simple text editor for vectors. To end the editing session, simply click on the X icon in the upper right hand corner of the window. Examples:

```
> newTemp <- edit( temperature )
> x <- edit( x )
```

Exporting data

The function `save()` is used to export a data frame to an R file on your computer. There are many advanced optional features to the command, but you must provide an R object to export and a file destination for the object. Example:

```
save(temperature, file="c:/StatData/temperature.RData")
```

Learning more about R

In this manual, you will continue to learn more about R as you progress through the examples corresponding to each chapter in *Mind on Statistics*. Although this manual will focus on more introductory statistics, the potential for using R for statistical analysis is almost endless. There always seems to be more about this software and statistics that a person can learn no matter how introductory or advanced the user. Besides the R manuals available through the Help icon at the top of R, there are a number of books written at introductory and advanced levels which describe how to use R and the similar S-Plus package.