Statistics 243: *class notes*

William J. De Meo

October 29, 1997

1 S

A data frame allows you to think of columns as named variables and the rows as observations. It requires that all columns be the same length. A list is a completely general way of storing information.

1.1 Getting Data Into S

Assignment Operator ←
x ← 5
x ← c(2,3,4)
If you want to add a 20 to the end of x,
x ← c(x,20)
x ← c('hey', 'you')

The scan() function
x ← scan()
1 10 12:15
17 18

Typing a blank line and then return will tell S that you are done.
Suppose the file filename contains
1 2 3 4
5 6 7 8
You could read this into a matrix with
x ← scan("filename")
x ← matrix(x,nrow=2, ncol=4)
Which would give you
1 3 5 7
2 4 6 8
So we should instead do
x ← matrix(x, nrow=2, ncol=4, byrow=T)
Instead we could compress the two steps into one with
x ← matrix(scan("filename"),nr=2,nc=4,byr=T)

The read.table function
x ← read.table("filename",header=T)

When invoking Splus with the -e option, you should have the environmental variable set to your editor preference:
setenv SEDITOR emacs

all data ←list(x,y,z)
x, y and z can be any kind of argument.

1.2 Subscripts

vectors
empty subscript x[]
EX:
x ← matrix(0, 5, 5) is a five by five matrix with all entries 0.
To make a matrix of identical columns you would write:
y ← matrix(c(1,5,7,9,11), 5, 5)
Suppose now we want to change the entries of x to all ones:
x ← 1 would just set x to the scalar 1.
x[] ← 1 puts all the elements of the 5 by 5 matrix to 1.
0 subscript is ignored.
positive numeric subscripts give you what you would expect.
negative subscripts give you the matrix with that subscript value removed.
EX:
x ← c(1,3,5,7)
x[-2] would be 1,5,7
logical subscripts T or F
EX:
x ← 1:10 gives x the values 1,2,3,...,10.
x ¡ 5 will give you
T T T T F F F F F F so you can use
x[x ¿ 5] to extract from x the first four entries.