Statistical packages

(STATS310.3: Simple linear regression)

Gunnar Stefansson

June 10, 2012

The R statistical package

R is freely available on the Internet.
Students can pick this up and install on their home computers.

Example: A typical R example. The following sequence inputs matrix data in columns x, y and z, from the file "test.dat" into R and subsequently prints the data and does a simple linear regression.
The commands also plot a few examples of randomly generated data.

```
dat<-read.table("test.dat",col.names=c("x","y","z"))
print(dat)
summary(lm(y~x,data=dat))
x<-1:100
y<-2+0.5*x+rnorm(100)*5*x
plot(x,y)
plot(dat$x,dat$y)
```

Note that dat becomes a data frame, which is a bit like a matrix, but the columns have names and can be referred to as dat$x etc.

Linear statistical models with R

Mathematical model:

$$y = \alpha + \beta x + \epsilon$$

R definition:

$$y \sim x$$

`lm(y~x)`

Storing the output

`fm<-lm(y~x)` .

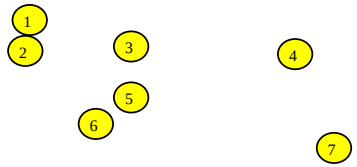

Figure: Example output from a simple linear model fit of the form y=a+bx. Items (1)-(2) are the estimates of a and b respectively. The estimate of the standard error of b is given by (3). The P-value for testing whether the true (underlying) value of b is zero is in (4). Items (5)-(7) give the MSE, R-squared and P-value for the entire model, respectively.

A sequence:

```
fm<-lm(y~x)          Fitting the model
summary(fm)          Traditional summary
drop1(fm)            Effect of dropping each variable
```

The SAS statistical package

SAS is expensive but freely available to students enrolled in courses at licensed universities.

Students are expected to obtain and install SAS and R.