

Basic Commands for R – Economics 113

Loading Data:

```
“datasetname”<-read.table(“file name”,header=TRUE)
```

```
Example: x<-read.table("C:/WageData.TXT",header=TRUE)
```

Important!!!! The filename needs to point to where the data are located. "C:/WageData.TXT" is just an example. If your data is not located there, you will get an error!!!

the command `str()` lists the variables in your dataset, and summarizes the type of each variable

```
str(x)
```

the command `summary()` summarizes each variable in the data set, giving the mean, median, min, and max

```
summary(x)
```

to summarize a single variable, write `summary(“datasetname”$“variable name”)`

```
summary(x$wage)
```

The commands, `mean`, `var`, and `cor` calculate means, variance/covariance, and correlation

When using `mean`, `var`, or `cor`, make sure you end the command with `“na.rm=TRUE”`. This drops observations that are labeled as missing. For example:

```
mean(x,na.rm=TRUE), or for wage mean(x$wage,na.rm=TRUE)
```

Regression Commands

The standard regression command in R is

```
lm(“regression equation”, “datasetname”)
```

For example, to regress log wage on education, experience, tenure, mother’s education, and father’s education, we write:

```
Reg<-lm(log(wage)~educ+exper+tenure+meduc+feduc,x)
```

Here, “Reg” is the name of the regression object. Writing “Reg” on the command line will report the regression estimates, but nothing else. However, if you write the following,

```
summary(Reg)
```

you will receive the full results of the regression, including homoskedastic standard errors, t-statistics, R-squared values, and other information.

You can easily run regressions using subsets of the full sample. The following regression estimates the determinants of wages for only urban survey respondents.

```
RegUrban<-lm(log(wage)~educ+exper+tenure+meduc+feduc,subset(x,urban==1))
```

To call residuals, if your regression object is called “Reg”:

```
Reg$residual
```

To call the degrees of freedom:

```
Reg$df
```

If you want use a function of a variable within a regression expression, use the `I()` wrapper:

```
RegSquared<-lm(log(wage)~educ+I(educ^2)+exper+tenure+meduc+feduc,data)
```