

Workshop 2019 on R

Reading data into R

From the textbook, *Econometrics*, Stephen J. Schmidt, 2005, McGraw-Hill, New York, NY

Example from Chapter 6, Ordinary Least Squares, p. 115

In his textbook, Stephen J. Schmidt (SJS), states:

Definition 1.1 *Econometrics* is the study of the application of statistical methods to economic problems.

One of the most standard statistical methods is the application of OLS to a simple linear regression model.

Example

From page 115 of the textbook is Exercise 7.

7. Load the data file TAXRECPT, which contains data from 1960 to 1995 on two variables, RECEIPTS is the U.S. federal government's total tax receipts; GDP is the U.S. gross domestic product. Both are measured in billions of dollars.

Estimate the following equation: $RECEIPTS_i = \beta_0 + \beta_1 GDP_i + \varepsilon_i$

- a. If GDP rises by \$1 billion, how much do tax receipts rise?

The data for the above comes in a CSV file: TAXRECPT.csv

The file contents are:

```
OBS, GDP, RECEIPTS
1960, 527.375, 92.84999847
1961, 545.6250153, 94.42499924
1962, 586.5250092, 102.3249989
1963, 618.6750031, 110.1750011
1964, 664.375, 110.1500015
1965, 720.1000061, 119.2749996
1966, 789.2999878, 136.3250008
1967, 834.0750122, 144.8999977
1968, 911.4500122, 168.4749985
1969, 985.3500061, 190.0750008
1970, 1039.675003, 184.2749977
1971, 1128.600006, 189.8249969
1972, 1240.424988, 217.4750023
1973, 1385.550018, 248.4750023
1974, 1501, 277.3249969
1975, 1635.174988, 276.0749969
1976, 1823.925018, 318.8499985
1977, 2031.399963, 359.875
1978, 2295.875, 417.2749939
1979, 2566.375, 478.2750015
1980, 2795.549988, 522.7999954
1981, 3131.350037, 605.625
1982, 3259.225037, 599.5
1983, 3534.950012, 623.8750153
1984, 3932.749939, 688.0749969
```

```
1985,4213,747.4500122
1986,4452.849976,786.375
1987,4742.474976,870.4750061
1988,5108.324951,928.8749847
1989,5489.050049,1010.325012
1990,5803.249878,1055.725006
1991,5986.224976,1072.249969
1992,6318.949951,1121.300018
1993,6642.325073,1197.325012
1994,7054.300049,1293.674988
1995,7400.549927,1383.699951
```

Suppose the data file's location was local-directory C:\Users\root\rcoding this means the file on the computer is C:\Users\root\rcoding\TAXRECPT.csv. Check the working directory with `getwd()` and set working directory with `setwd()`

The commands in R would look like:

```
R Console (64-bit)
File Edit Misc Packages Windows Help

Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> #Step 1, check the working directory
> getwd()
[1] "C:/Users/root/Documents"
> #Step 2, set working directory
> setwd("C:/Users/root/rcoding")
> getwd()
[1] "C:/Users/root/rcoding"
> gdpdataset = read.csv("TAXRECPT.csv", header=T,colClasses = c("numeric", "numeric", "numeric"))
> exampleReg = lm(RECEIPTS~GDP, data=gdpdataset)
> summary(exampleReg)

Call:
lm(formula = RECEIPTS ~ GDP, data = gdpdataset)

Residuals:
    Min       1Q   Median       3Q      Max
-28.038  -9.127   0.655   8.990  39.083

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.969811   4.182008  -1.427   0.163
GDP           0.182832   0.001162 157.297 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.04 on 34 degrees of freedom
Multiple R-squared:  0.9986,    Adjusted R-squared:  0.9986
F-statistic: 2.474e+04 on 1 and 34 DF,  p-value: < 2.2e-16

>
```