Microeconometrics Using Stata

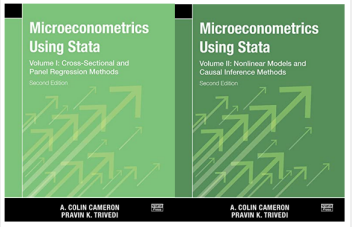
Data management and graphics: Exercises

Jhon R. Ordoñez [1] [2] [3]

[1] National University of San Cristóbal de Huamanga
[2] Faculty of Economic, Administrative and Accounting Sciences
[3] Professional School of Economics

Research Assistant

September 20, 2024



Outline

Exercise 1

Exercise 1

Exercise 1

Type the command **display %10.5f 123.321**. Compare the results with those you obtain when you change the format **%10.5f** to, respectively, **%10.5e**, **%10.5g**, **%-10.5f**, and **%10,5f** and when you do not specify a format.

Exercise 2

Exercise 2

Consider the example of **section 2.3** except with the variables reordered. Specifically, the variables are in the order **age**, **name**, **income**, and **female**. The three observations are **29 "Barry" 40.990 0**; **30 "Carrie" 37.000 1**; and **31 "Gary" 48.000 0**. Use **input** to read these data, along with names, into Stata, and list the results. Use a text editor to create a comma-separated values file that includes variable names in the first line, read this file into Stata by using **import delimited**, and list the results. Then, drop the first line in the text file, read in the data by using **import delimited** with variable names assigned, and list the results. Finally, replace the commas in the text file with blanks, read the data in by using **infix**, and list the results.

Exercise 3

Exercise 3

Consider the dataset in **section 2.4.** The **er32049** variable is the last known marital status. Rename this variable as **marstatus**, give the variable the label "marital status", and tabulate **marstatus**. From the codebook, marital status is married (1), never married (2), widowed (3), divorced or annulment (4), separated (5), not answered or do not know (8), and no marital history collected (9). Set **marstatus** to missing where appropriate. Use **label define** and **label values** to provide descriptions for the remaining categories, and tabulate **marstatus**. Create a binary indicator variable equal to 1 if the last known marital status is married and equal to 0 otherwise, with appropriate handling of any missing data. Provide a summary of earnings by marital status. Create a set of indicator variables for marital status based on **marstatus**. Create a set of variables that interact these marital status indicators with earnings.

Exercise 4

Exercise 4

Consider the dataset in **section 2.6.** Create a box-and-whisker plot of **earnings** (in levels) for all the data and for each year of educational attainment (use variable **education**). Create a histogram of **earnings** (in levels) using 100 bins and a kernel density estimate. Do earnings in levels appear to be right skewed? Create a scatterplot of **earnings** against **education**. Provide a single figure that uses **scatterplot**, **lfit**, and **lowess** of **earnings** against **education**. Add titles for the axes and graph heading.

Exercise 5

Exercise 5

Consider the dataset in **section 2.6.** Create kernel density plots for lnearns using the kernel(epan2) option with kernel

$$K(z) = \frac{3}{5}\left(1 - \frac{z^2}{5}\right)$$

for $|z| < 1$ and using the **kernel(rectangle)** option with kernel $K(z) = 1/2$ for $|z| < 1$. Repeat with the bandwidth increased from the default to 0.3. What makes a bigger difference, choice of kernel or choice of bandwidth? The comparison is easier if the four graphs are saved using the **saving()** option and then combined using the **graph combine** command.

Exercise 6

Exercise 6

Consider the dataset in **section 2.6.** For each of the available kernels that can be used with the **kdensity** command, obtain a kernel density plot for **lnearns** using the default bandwidth, and save the graph using the **saving()** option. Then, combine all graphs on one page using the **graph combine** command and options such as **rows(4) ysize(8) xsize(5)**. Comment on the relative smoothness of the various graphs.

Exercise 7

Exercise 7

Consider the dataset in **section 2.6.** Perform lowess regression of **lnearns** on **hours** using the default bandwidth and using bandwidth of 0.01. Does the bandwidth make a difference? A moving average y of after data are sorted by x is a simple case of nonparametric regression of y on x . Sort the data by **hours**. Create a centered 25-period moving average of **lnearns** with ith observation

$$yma_i = \frac{1}{25}\sum_{j=-12}^{j=12} y_{i+j}$$

This is easiest using **forvalues**. Plot this moving average against **hours** using the **twoway connected** graph command. Compare with the lowess plot.

References I


A. Colin Cameron, P. K. T. (2022a).
Microeconometrics Using Stata, Second Edition, Volume I: Cross-Sectional and Panel Regression Models.
Stata Press, 2 edition.


A. Colin Cameron, P. K. T. (2022b).
Microeconometrics Using Stata, Second Edition, Volume II: Nonlinear Models and Casual Inference Methods.
Stata Press, 2 edition.

Microeconometrics Using Stata

Data management and graphics: Exercises

Jhon R. Ordoñez [1] [2] [3]

[1] National University of San Cristóbal de Huamanga
[2] Faculty of Economic, Administrative and Accounting Sciences
[3] Professional School of Economics

Research Assistant

September 20, 2024