

Introduction to Econometrics

ECON-UN3412 - Recitation 1

Matthew Davis

September 21, 2021



Outline

Recitation 1 (today)

- ▶ Introduction
- ▶ Organization of the course
- ▶ Why R?



Outline

Recitation 2 (next week)

- ▶ Submitting problem sets
- ▶ Recitation notes
- ▶ Getting started with R/Stata
- ▶ Preparing for the first problem set

Recitation 3+ (rest of semester)

- ▶ Practice problems
- ▶ Relevant R and Stata commands
- ▶ Additional notes: applications or points of emphasis



About me

- ▶ From Southeast Asia
- ▶ Research: environmental + development
- ▶ Recitations: Tue. 2:00-3:00pm, Knox 104
- ▶ Office Hours: Mon. 12:30-2:30pm, Zoom
- ▶ Email: wad2113@columbia.edu
 - ▶ Grad school, advice, research, etc.
 - ▶ Questions about course material: Ed Discussion or office hours



Logistics

- ▶ Recitations
 - ▶ R/Stata instruction
 - ▶ Ungraded questions in the problem set
 - ▶ Uploads in Courseworks : Files : Recitations : Matthew Davis
- ▶ Office hours
 - ▶ Please come if you need help!



Teachers (Lop). The table below shows the sample means and standard deviations from her sample.

Variable	Sample Mean	Sample Standard Deviation
TestScore	780.3	89.9
RPM	0.40	8.38
TEAP	0.2	3.8

After standardizing RPM and TEAP and subtracting the sample mean from TestScore, she estimates the following regression:

$$\widehat{\text{TestScore}} = -48.7 \times \text{RPM} + 8.7 \times \text{TEAP} + 44.9$$

a. You are interested in using the estimated regression to predict average test scores for an out-of-sample school with RPM = 0.52 and TEAP = 0.11.

1. Compute the transformed (standardized) values of RPM and TEAP for this school; that is, compute the X^{std} values from the X^{raw} values, as discussed preceding Equation (14.2).
2. Compute the predicted value of average test scores for this school.
3. The actual average test score for the school is 775.3. Compute the error for your prediction.
4. The regression shown above was estimated using the standardized regressors and the demeaned value of TestScore. Suppose the regression

Following questions will not be graded, they are for you to practice and will be discussed at the recitation.

1. [ungraded] SW Exercise (not Empirical Exercises) 14.3
2. [ungraded] SW Exercise (not Empirical Exercises) 14.3
3. [ungraded] SW Exercise (not Empirical Exercises) 14.3
4. [ungraded] SW Exercise (not Empirical Exercises) 14.3
5. [ungraded] We will use California School Data for this exercise: CEdashboard

Variables in use: school dataset

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_1 + \hat{\beta}_2 \hat{x}_2 \Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_1 + \hat{\beta}_2 \hat{x}_2$$
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \left(\frac{x_1 - \bar{x}_1}{s_{x_1}} \right) + \hat{\beta}_2 \left(\frac{x_2 - \bar{x}_2}{s_{x_2}} \right)$$
$$\hat{y} - \bar{y} = \hat{\beta}_1 \left(\frac{x_1 - \bar{x}_1}{s_{x_1}} \right) + \hat{\beta}_2 \left(\frac{x_2 - \bar{x}_2}{s_{x_2}} \right)$$
$$\Rightarrow \hat{y} = \bar{y} + \underbrace{\hat{\beta}_1 \left(\frac{x_1 - \bar{x}_1}{s_{x_1}} \right)}_{\hat{\beta}_1} + \underbrace{\hat{\beta}_2 \left(\frac{x_2 - \bar{x}_2}{s_{x_2}} \right)}_{\hat{\beta}_2} + \underbrace{\left(\frac{\hat{\beta}_1}{s_{x_1}} \right) x_1}_{\hat{\beta}_1} + \underbrace{\left(\frac{\hat{\beta}_2}{s_{x_2}} \right) x_2}_{\hat{\beta}_2}$$

Results

```

-48.7125, D=0.079, B=0.04, S=0.00021, 1
780.3119
+-----+
+ 81.498, 79.048, 2.00146, 79.048, 3031
+ 4.1012
+-----+

```

Variables in use: school dataset



Stata vs. R

Stata

- ▶ Used by most academic econs, especially seniors (and then passed down to their students and RAs)
- ▶ Lecture slides will feature Stata commands
- ▶ Commands for regressions usually very simple one-liners
- ▶ No need to think about installing packages, all functions are in-built
- ▶ All eight TAs are familiar with Stata, only three with R

R

- ▶ Used by most statisticians, data scientists, younger econs
- ▶ Free (\$0/lifetime < \$90/year) and open source (very Google-able help)
- ▶ Stress-free pset and exam submission
 - ▶ Automatic and sleek pdfs
 - ▶ Stata's log files and graphs have proven problematic for students
- ▶ A flexible programming language, not just a statistical software
 - ▶ Handles multiple datasets and types of data
 - ▶ Better for data science, computer science, data viz, GIS, web design...
 - ▶ Complements Python, Julia, Matlab



Past student feedback

"I quickly just wanted to add a quick personal note: throughout the entire first half of the semester your recitations have been incredible, informative, clear, and a much-appreciated alternative to having to learn Stata. I wanted to personally say thank you for basically designing an entire parallel curriculum and for always responding to our questions and giving us feedback."

"Just wanted to say thank you for helping me and the rest of the students in Econometrics this semester. I know it definitely was not an easy (nor typical) year, but I'm so happy to have been able to do the class in R and I could not have done it without your help!"



Recitations and office hours, R support underlined

Recitations

- ▶ Tuesday
 - ▶ Umar 8:00-9:00am
 - ▶ Matt 2:00-3:00pm
- ▶ Wednesday
 - ▶ Tushar 3:00-4:00pm
- ▶ Thursday
- ▶ Friday
 - ▶ Natalie 1:00-2:00
 - ▶ Xiao 1:00-2:00
- ▶ Saturday
 - ▶ Nadia 1:00-2:00
- ▶ Monday
 - ▶ Seung-hun 9:00-10:00am

Office Hours

- ▶ Tuesday
 - ▶ Umar 12:00-2:00pm
 - ▶ Katya 4:15-5:15pm
- ▶ Wednesday
 - ▶ Tushar 10:15-11:15am
 - ▶ Tushar 5:00-6:00pm
- ▶ Thursday
 - ▶ Natalie 10:00am-12:00pm
 - ▶ Katya 4:15-5:15pm
- ▶ Friday
 - ▶ Nadia 8:00-10:00am
 - ▶ Xiao 2:00-4:00pm
- ▶ Monday
 - ▶ Seung-hun 10:30am-12:30pm
 - ▶ Matt 12:30-2:30pm



Miscellaneous

- ▶ Potential grad-student strike
- ▶ Can help facilitate a group chat among R users so you can help one another



“Homework” for next week

- ▶ Download R:
<https://r-project.org>
- ▶ Download RStudio:
<https://rstudio.com/products/rstudio>
- ▶ Optional: bring your laptop

