

## INTRODUCTION: Purpose, Scope, and Plan

Hüseyin Taştan<sup>1</sup>

<sup>1</sup>Yıldız Technical University  
Department of Economics

## Introduction to Econometrics I

- ▶ The purpose of this course is to teach fundamental methods of analysis in econometrics at an undergraduate level.
- ▶ Econometrics basically deals with developing statistical methods for estimating and testing economic relationships and theories. As we will discuss throughout this course, econometrics draws heavily on economic theory, statistical methods, and mathematics.
- ▶ In this course we will follow Jeffrey M. Wooldridge's *Introductory Econometrics: A Modern Approach* as our main textbook. (6th Edition, 2016, Cengage Learning)
- ▶ The same text is used in the second part of this course (Econometrics II).

## Introduction to Econometrics I

- ▶ The division is based primarily on the type of data: Econometrics I predominantly builds the fundamentals using cross-sectional data framework whereas Econometrics II deals with time series analysis and extensions.
- ▶ Please read the syllabus for the lecture plan of this course
- ▶ The syllabus of Econometrics is available at:  
<https://github.com/htastan/Econometrics-I>
- ▶ This address also includes class materials.
- ▶ Applications will be carried out using the R program.

## Introduction to Econometrics I

- ▶ Evaluation is based on two midterms and a final. (Midterms 60% + Final 40%)
- ▶ The exact midterm dates and times are available in the syllabus. The exams will be administered online.
- ▶ Final exam: following the last day of classes (to be announced)
- ▶ Grading policy: here is a typical grade distribution. Note that you need at least 40/100 to pass this course with a grade of DC.

0-29	FF
30-39	DD
40-49	DC
50-59	CC
60-69	CB
70-79	BB
80-89	BA
90-100	AA

## What is econometrics?

- ▶ Literal meaning: **economic measurement**: “econo-metrics”.
- ▶ But the scope of econometrics is much wider.
- ▶ Two popular definitions of econometrics:
- ▶ *“Econometrics may be defined as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena”*. (A.S. Goldberger, 1964).
- ▶ *“...econometrics may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of science”*. (P. Samuelson)

## Scope of Econometrics

- ▶ Developing statistical methods for the estimation of economic relationships,
- ▶ Testing economic theories and hypotheses,
- ▶ Evaluating and applying economic policies,
- ▶ Prediction/Forecasting,
- ▶ Collecting and analyzing nonexperimental or observational data.

## Why Do We Need Econometrics?

- ▶ We learned statistical methods so why do we need a separate discipline?
- ▶ The reason is as follows: econometrics focuses on the analysis of nonexperimental economic data.
- ▶ **Nonexperimental data** (or **observational data**) are not obtained through controlled experiments on economic agents (consumers, firms, households, sectors, countries, etc.)
- ▶ **Experimental data** are collected in laboratory environments in natural sciences.
- ▶ Although some social experiments can be devised it is usually difficult (costly) or impossible to conduct economic experiments.
- ▶ Unlike statistical methods employed in natural sciences, econometrics develops special methods to handle nonexperimental/observational data.

## Econometrics is about measuring causal relationships

In your economic theory classes, you've learned that economics suggests certain type of relationships among variables. In many cases, these involve important policy implications.

- ▶ For example, what happens to employment if the minimum wage is raised by, say, 10%?
- ▶ What is the price elasticity of water? Is it inelastic? What is the quantitative effect of raising the unit price of water by 5%?
- ▶ What is the impact of a 1 percentage point increase in the Central Bank policy rate on the growth of income (GDP) or unemployment?
- ▶ What is the quantitative impact of another year of education on lifetime earnings? (a related question: why are you here?)

## Components of Econometrics

- ▶ Econometric inputs:
  - ▶ Economic Theory
  - ▶ Mathematics
  - ▶ Statistical Theory
  - ▶ Data
  - ▶ Computers (CPU power)
  - ▶ Interpretation
- ▶ Econometric outputs:
  - ▶ Estimation - Measurement
  - ▶ Inference - Hypothesis testing
  - ▶ Forecasting - Prediction
  - ▶ Evaluation

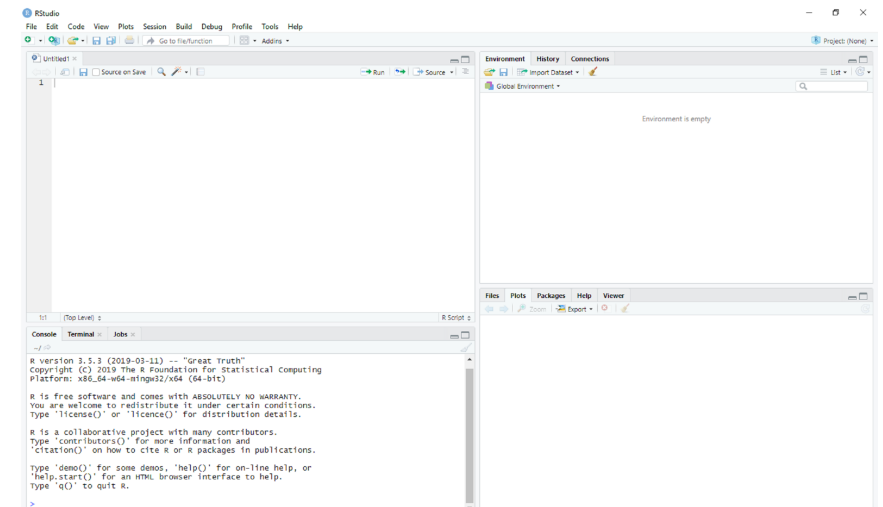
## Big Data, Machine Learning, and Econometrics

- ▶ **Big Data:** large data sets generated from various sources including online/mobile activities, social media, business transactions, etc. Can be so big that the standard PCs cannot be used to store and analyze such data sets,
- ▶ **Data Science:** an emergent interdisciplinary field combining computer science, statistics, data mining, and machine learning.
- ▶ **Data Mining:** summarize big data to find patterns and insights
- ▶ **Machine Learning:** is concerned primarily with prediction. Given a set of attributes (variables), what is the best model to predict a variable? Less (or no) emphasis on causation. Machine = computer, learning = estimation
- ▶ **Econometrics:** is concerned with both causal relationships and prediction.

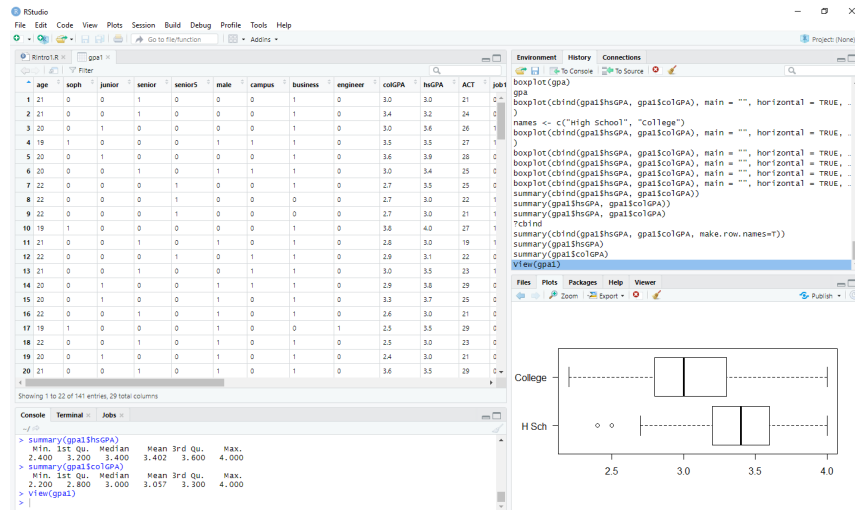
## Econometric Software

- ▶ There are several econometric software platforms that we can use in applied research
- ▶ Some of these are commercial, e.g., Stata, Eviews, SPSS, SAS, RATS, Gauss
- ▶ There are also open-source software: Gretl, R, among others.
- ▶ Gretl can be downloaded from <http://gretl.sourceforge.net/index.html> and <http://gretl.sourceforge.net/win32/>
- ▶ For more information on R go to <https://www.r-project.org/>
- ▶ RStudio may be used as an integrated development environment for R: <https://www.rstudio.com/products/RStudio/>
- ▶ Let us review R and RStudio, and Gretl.

## RStudio (version 1.2.1335)

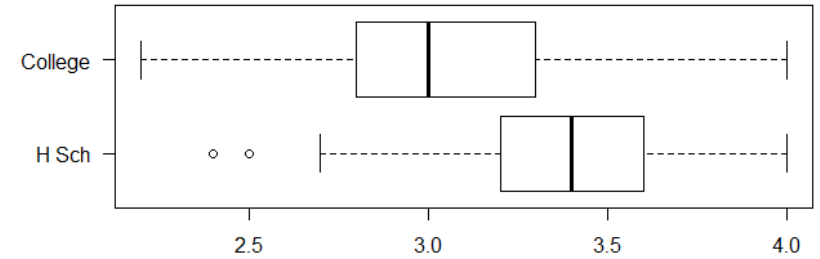


## Wooldridge data set: gpa1



## Wooldridge data set: gpa1

Summary statistics and box plots of high school GPA and college GPA

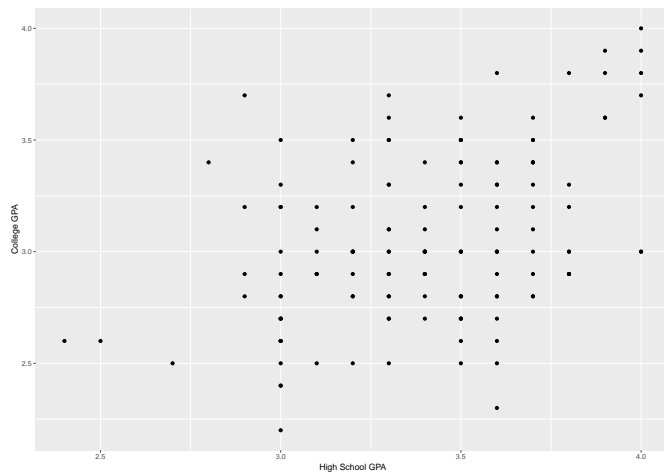


```

> summary(gpa1$hsGPA)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.400   3.200   3.400   3.402   3.600   4.000
> summary(gpa1$colGPA)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.200   2.800   3.000   3.057   3.300   4.000
  
```

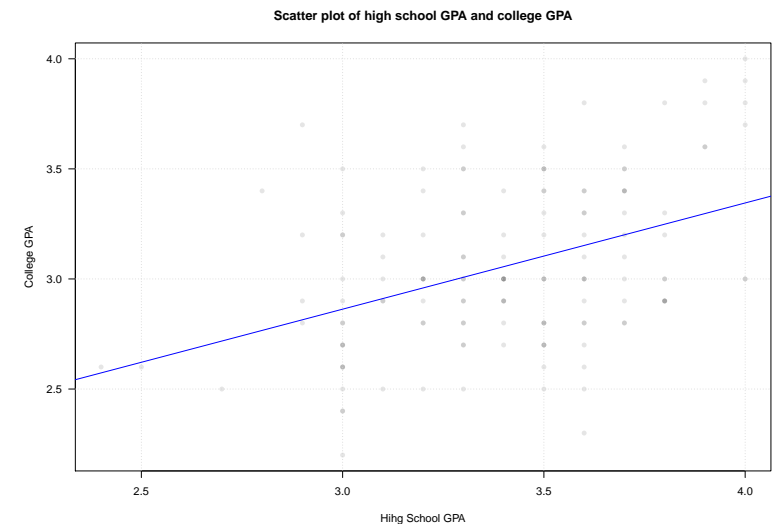
## Wooldridge data set: gpa1

Let us try to predict the success of a student in college given the past academic success (high school GPA). For a student with a high school GPA of 3, what is your prediction for the college GPA?



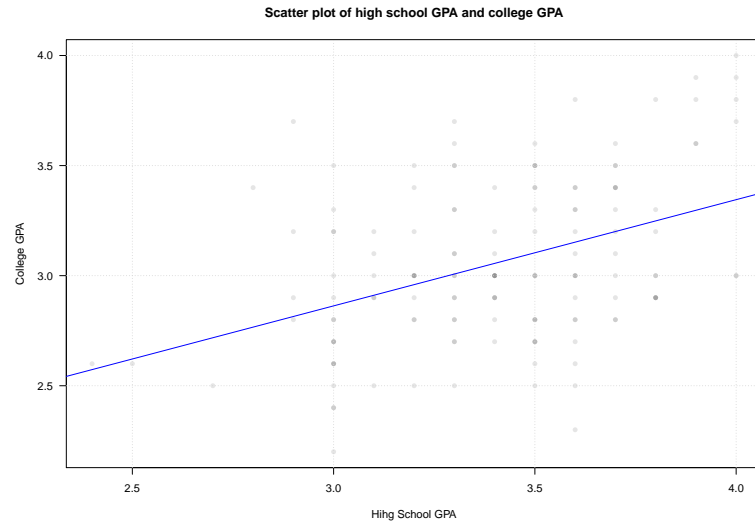
## Wooldridge data set: gpa1

A linear function is fitted such that  $colGPA = a + b \times hsGPA$  where  $a$  and  $b$  are unknowns that are estimated from data



## Wooldridge data set: gpa1

$$\text{predicted colGPA} = 1.42 + 0.48 \text{ hsGPA}$$



18

## Econometric Software

- ▶ R provides a powerful computational environment. It is widely used in almost every scientific discipline. It has become especially popular in machine learning applications and big data analysis.
- ▶ We will use R in our lab and in-class applications. But it is also a good idea to be informed about other software packages.
- ▶ Gretl is an open source (free and open to modifications) software which is menu-driven but also programmable.