

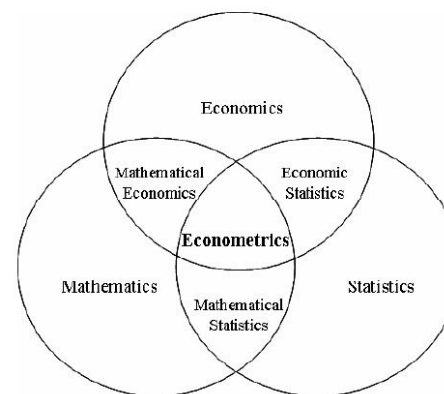
# EE 325 Introduction to Econometrics

Week1: Introduction and  
Nature of Regression Analysis



1

**Econometrics is a mixture of  
economics, mathematics, and statistics**



2

- Economics suggests interesting relations, often with policy implications, but most of the time theory does not provide quantitative magnitudes of causal effects.
  - What is the price elasticity of cigarettes?
  - What is the effect of reducing class size on student achievement?
  - What is the effect on earnings of a year of education?
  - What is the effect on output growth of a 1 percentage point increase in interest rates by the central bank?

3

## Controlled experiments

- The focus of this course is the use of statistical and econometric methods to quantify causal effects.
- Ideally, we would like to conduct a controlled experiment with cigarette prices, class size, returns to education, central bank interest rates, etc.
- But most of the time, we cannot do so and must use observational (non-experimental) data.

4

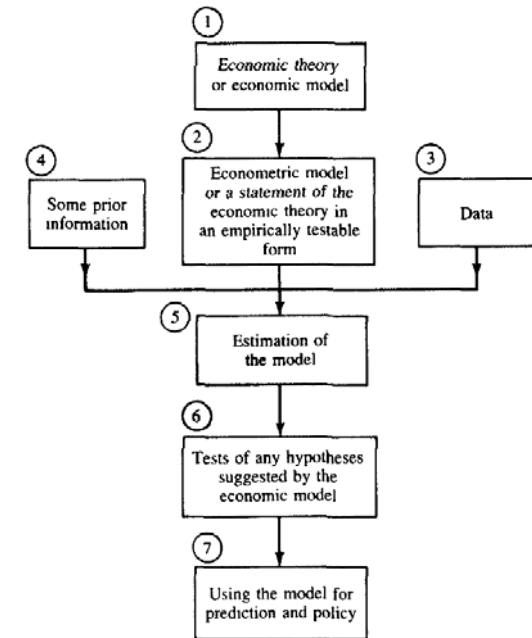
# Methodology of Econometrics

Broadly speaking, traditional econometric methodology proceeds along the following lines:

1. Statement of economic theory;
2. Specification of the mathematical model of the theory;
3. Specification of the econometric model of the theory;
4. Obtaining the data;
5. Estimation of the parameters of the econometric model;
6. Hypothesis testing;
7. Forecasting or prediction;
8. Using the model for control or policy purposes.

5

## Stages of applied econometric research



6

# Types of Data

Types of Data for empirical analysis:

1. Cross-Section
2. Time Series
3. Pooled
4. Panel, Longitudinal, or Micropanel Data

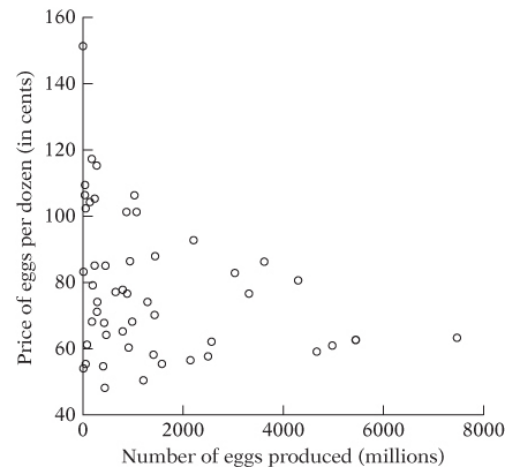
7

# Cross Section Data

- Cross-sectional data: Data collected at the same point in time
- Consists of a sample of individuals, households, firms, cities, states, countries, or a variety of other units, taken at a given point in time.
- Often obtained by random sampling from the underlying population
- The analysis of cross-sectional data is closely aligned with the applied economics fields, such as labor economics, health economics, industrial organization, urban economics, and local public finance

8

## Relationship between eggs produced and prices, 1990



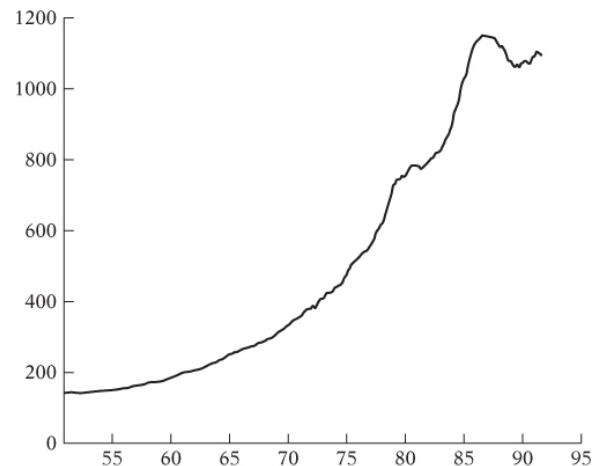
9

## Time Series Data

- Time series data consists of observations on a variable or several variables over time
- Data may be collected at regular time intervals such as
  - Daily – Stock prices or Weather reports
  - Weekly - Money supply in the U.S. economy
  - Monthly – Unemployment rate or CPI
  - Quarterly - GDP
  - Annually – Government budgets

10

## Money supply: U.S., Jan 1959-Sept 1999



11

## Pooled Data

- Data are elements of both time series and cross-section data
  - Two cross-sectional household surveys are taken in the U.S., one in 1985 and one in 1990

12

## Panel or Longitudinal Data

- Consists of a time series for each cross-sectional member in the data set
- E.g. we have wage, education, and employment history for a set of individuals followed over a ten-year period
- Key feature of panel data that distinguishes them from a pooled cross section is that the same cross-sectional units are followed over a given time period

13

## The Sources of Data

- The data used in empirical analysis may be collected by government agency, an international agency (e.g., IMF or the World Bank), a private organization (e.g., the Standard & Poor's Corporation)
- Experimental or Nonexperimental data

14

## Web-Links

### INTERNATIONAL ORGANIZATIONS

- [The Organization for Economic Cooperation and Development \(OECD\)](http://www.oecd.org) (<http://www.oecd.org>).  
The Organization for Economic Cooperation and Development (OECD) reports statistics on labor market conditions in many advanced economies.
- [IZA](http://www.iza.org) (<http://www.iza.org>).  
IZA is a Bonn-based research institute that conducts labor research. Their discussion paper series provides up-to-date research on labor issues in many countries.
- **IMF**  
(<http://www.imf.org/external/data.htm>)
- **WORLDBANK**  
(<http://data.worldbank.org/>)

15

- **THAILAND MINISTRY OF LABOUR** (<http://www.mol.go.th/en>)
- **INTERNATIONAL LABOUR ORGANIZATION (ILO)**  
(<http://www.ilo.org>)
- **BUREAU OF LABOR STATISTICS (BLS)**  
The Bureau of Labor Statistics (BLS) is the government agency responsible for calculating the monthly unemployment rate as well as the Consumer Price Index. Their website contains a lot of information on many aspects of the U.S. labor market, as well as comparable international statistics. The BLS also publishes a detailed description of how it defines and measures the concepts of labor force and unemployment. The BLS is officially a part of the U.S. Department of Labor, which also collects detailed information on various aspects of the state-run unemployment insurance program.

16

## ■ NATIONAL STATISTICAL OFFICE THAILAND

(<http://www.nso.go.th>)

## ■ BUREAU OF THE CENSUS

The Bureau of the Census reports detailed demographic and labor market information, including historical data by race and gender, mobility/migration patterns, and health insurance coverage. The Bureau of the Census also publishes annually the Statistical Abstract of the United States, which contains data on a wide variety of topics including many aspects of the U.S. economy.

17

## IMMIGRATION

- [Bureau of Citizenship and Immigration Services \(BCIS\)](http://www.uscis.gov/graphics/index.htm) (<http://www.uscis.gov/graphics/index.htm>). The website of the Bureau of Citizenship and Immigration Services (BCIS) contains information on U.S. immigration policy.
- [Department of Homeland Security](http://www.dhs.gov/ximgtn/statistics/) (<http://www.dhs.gov/ximgtn/statistics/>). The Department of Homeland Security reports detailed statistics on immigration to the United States.
- [International Trade Administration](http://www.ita.doc.gov) (<http://www.ita.doc.gov>). The International Trade Administration publishes detailed information on trade patterns and regulations.
- [Citizen and Immigration Canada](http://www.cic.gc.ca/english/index.asp) (<http://www.cic.gc.ca/english/index.asp>) The website of Citizen and Immigration Canada has the "test" that allows a potential applicant to determine if he or she qualifies for a visa.
- [Australia's Department of Immigration and Multicultural and Indigenous Affairs](http://www.immi.gov.au/) (<http://www.immi.gov.au/>) The website of Australia's Department of Immigration and Multicultural and Indigenous Affairs has the similar test required by Australian authorities.

18

## EDUCATION

### ■ [The U.S. Department of Education](http://www.ed.gov/index.jsp)

(<http://www.ed.gov/index.jsp>). The U.S. Department of Education's website provides an essential introduction to many of the education programs in the United States.

### ■ [General Education Development \(GED\)](http://www.acenet.edu//AM/Template.cfm?Section=Home)

(<http://www.acenet.edu//AM/Template.cfm?Section=Home>). The American Council on Education gives useful information to prospective takers of the General Education Development (GED) exam.

19

## POVERTY / MINIMUM WAGES / DEVELOPMENT

- [The Green Book](http://waysandmeans.house.gov/Documents.asp?section-5813) (<http://waysandmeans.house.gov/Documents.asp?section-5813>). The *Green Book*, also known as the *Overview of Entitlement Programs*, is published by the Committee on Ways and Means, U.S. House of Representatives. Each chapter describes a particular type of public assistance program.
- [The Employment Policies Institute](http://www.epionline.org) (<http://www.epionline.org>)
- [The United Nations Development Programme](http://www.undp.org/) (<http://www.undp.org/>). The United Nations Development Programme maintains an extensive database describing income inequality in many countries.
- [OECD](http://www.oecd.org/) (<http://www.oecd.org/>) The OECD reports unemployment statistics for many advanced economies and frequently publishes reports comparing the unemployment situation in different countries.
- [Population Resource Center](http://www.prcdc.org/) (<http://www.prcdc.org/>) Many organizations are concerned with the politically charged issues of population growth and control. The Population Resource Center is an organization that limits itself to promoting the use of accurate data for objective analysis.

20

## NON-PROFIT ORGANIZATIONS

- [The National Bureau of Economic Research \(NBER\)](http://www.nber.org) (<http://www.nber.org>). The National Bureau of Economic Research (NBER) publishes a working paper series that represents the frontier of empirical research in economics. Their website also contains a number of widely used data sets. The working papers and data can be accessed and downloaded by students and faculty at many universities.

21

## The Accuracy of Data

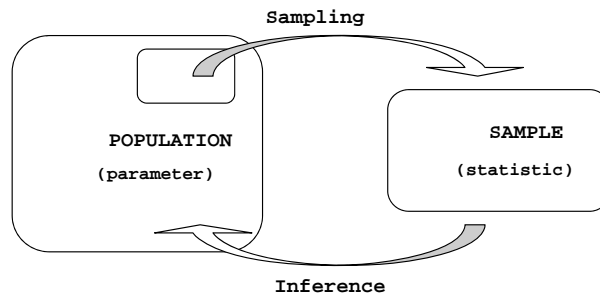
The researcher should always keep in mind that the results of research are only as good as the quality of the data

22

## Review of Inferential Statistics

**Inferential statistics** are methods used to make conclusions or inferences concerning some unknown aspect of a **population** based on data from a **sample**. Statistical inference is the process of making an estimate, prediction, or decision about a population based on a sample.

A **population** is a collection of all possible individuals, objects, or measurements of interest; A **parameter** is a numerical measurement describing some characteristic of a population.

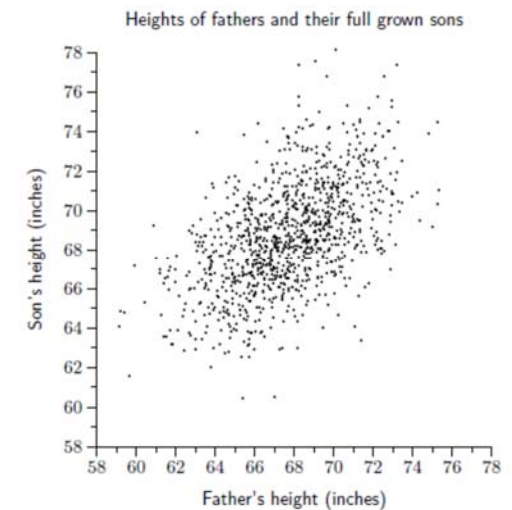


A **sample** is a portion, or part, of the population of interest; A **statistic** is a numerical measurement describing some characteristic of a sample.

23

## Pearson's sample

- The following scatter diagram shows the heights of 1,078 fathers and their full-grown sons, in England, circa 1900.
- There is one dot for each father-son pair.

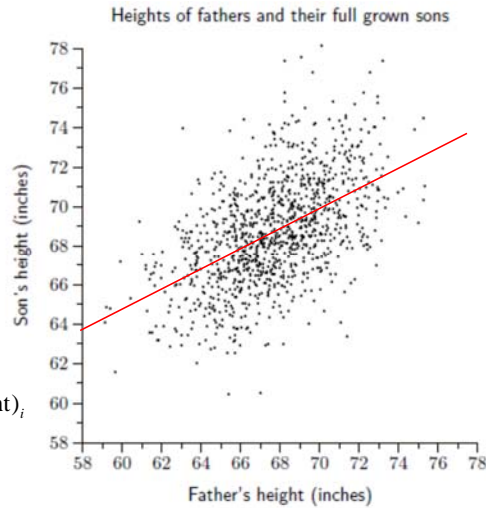


24

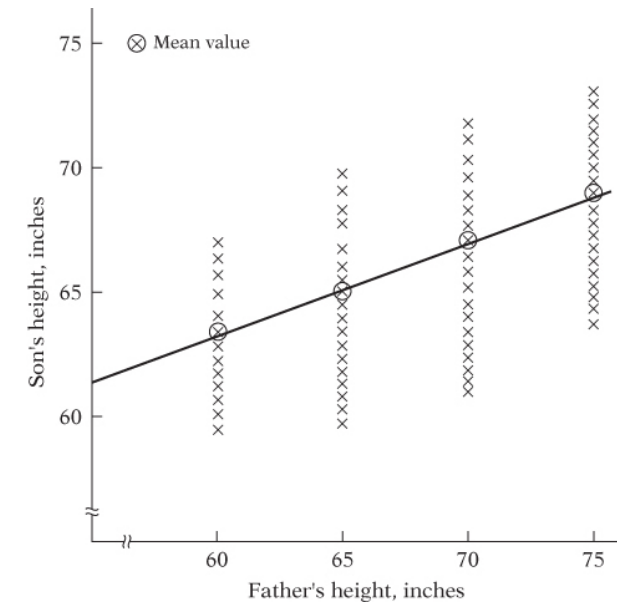
## So how can we draw a line?

- Using the ordinary least squares (OLS) method, we can get the sample regression line:

$$\widehat{\text{son's height}}_i = \hat{\alpha} + \hat{\beta}(\text{father's height})_i$$



25



26

## Regression

- Regression analysis** is concerned with the study of the dependence of one variable, the dependent variable, on one or more other variables, the explanatory variables, with a view to estimating the population mean of the former in terms of the known repeating sampling values of the latter.

27

## The Simple Regression Model

- We begin our study with an extensively investigation of the estimation of a linear relationship between two variables,  $Y_i$  and  $X_i$ , of the form:

$$Y_i = \alpha + \beta X_i + u_i$$

for  $i = 1, 2, \dots, n$

Y	X
<b>Dependent variable</b>	<b>Independent variable</b>
Explained variable	Explanatory variable
Predictand / Predicted	Predictor
Regressand	Regressor
Outcome	Covariate
Response	Stimulus
Controlled variable	Control variable

28

## The Population Regression Function (PRF)

- Imagine we wish to find the relationship between father's height and son's height.
- Imagine we had data on ALL the world's father's and son's heights.
- Then we can get the **Population Regression Function!**

$$\text{son's height}_i = \alpha + \beta(\text{father's height})_i + u_i$$

- But we can't observe the heights of ALL fathers and sons in the world throughout time.
- So we have to do with a sample!

29

## The Population and Sample Regression Lines

- In general, the simple population regression line can be written as:

$$E(Y | X_i) = \alpha + \beta X_i \quad \text{or} \quad Y_i = \alpha + \beta X_i + u_i$$

- Taking a sample, the estimated regression line can be written as either:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \quad \text{or} \quad Y_i = \hat{\alpha} + \hat{\beta} X_i + \hat{u}_i$$

30

## The Stochastic Term ( $u_i$ )

- 1. Randomness of human behavior
- 2. Unavailability of data
- 3. Omission of variables from the function
- 4. Wrong or imperfect specification of the functional form
- 5. Errors of aggregation
- 6. Errors of measurement

31

## Source

Gujarati, D.N. (2009) Basic Econometrics. 5th ed. Singapore, McGraw-Hill.

32