Analysis of Economics Data
Chapter 1: Analysis of Economics Data

© A. Colin Cameron
Univ. of Calif. Davis

November 2022

CHAPTER 1: Analysis of Economics Data

- This book provides an introduction to econometrics.
- This uses a subset of statistical methods
  - most notably regression analysis
  - an outcome y varies with one or more variables.
- The book emphasizes economic interpretation of economics-related data.

Chapter Outline

1. Statistical Methods
2. Types of Data
3. Regression Analysis
4. Overview

1.1 Statistical Methods

- There are two aspects to statistical analysis of data
  - Descriptive analysis
    - ★ mean, median, standard deviation, ...
    - ★ graphs and charts such as histograms and bar charts
  - Statistical inference
    - ★ extrapolate from the sample to the population
    - ★ often using confidence intervals and/or hypothesis tests
    - ★ this is more challenging than data summary
- Much of this book entails statistical inference.

1.2 Types of Data

- There are broad types of data:
  - Numerical data that are continuous
    - ★ e.g. GDP, earnings.
  - Numerical data that are discrete.
    - ★ e.g. number of doctor visits by an individual in one year
  - Categorical data
    - ★ e.g. employed, unemployed or out of the labor force.
- The book focuses on continuous numerical data
  - this is the data type usually analyzed in economics
  - more advanced courses adapt the methods of this book to the other types of data.

Observational Data

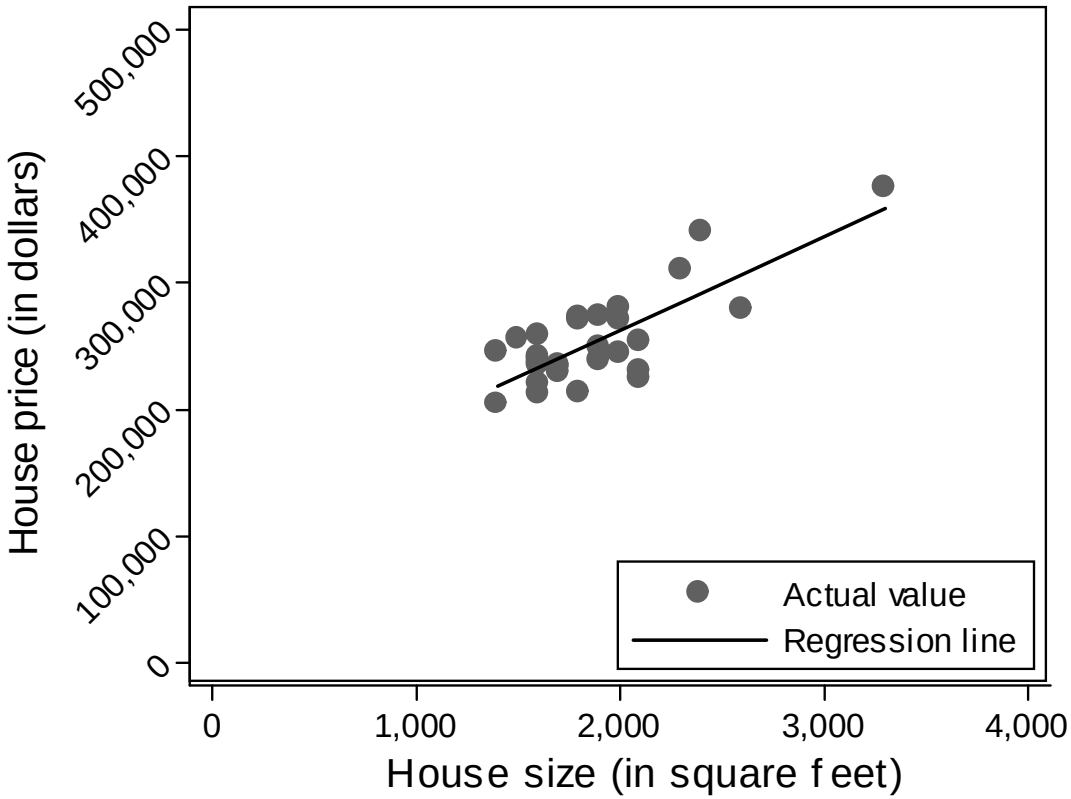
- Observational data
  - based on observed behavior in an uncontrolled environment
  - economics data are most often observational.
- Experimental data
  - observations on the results of experiments that can be controlled by the investigator.
- It is difficult to establish causal effects using observational data
  - e.g. in determining the causal effect of a college degree on earnings we need to control for individual self-selection into college
  - advanced econometrics research seeks to estimate causal relationships even with observational data.
- The book focuses on measuring association (not causation) using observational data
  - causal methods are presented in ch. 17 and in some case studies in ch. 13.

Three Types of Data Collection

- Distinguish between three types of data collection:
  - cross-section
    - ★ individuals (people, firms, countries, ...) at a point in time
    - ★ denoted by subscript $i = 1, ..., n$, e.g. x_i
  - time series
    - ★ over time for the same individual (stock price, US GDP, ...)
    - ★ denoted by subscript $t = 1, ..., T$, e.g. x_t
  - panel data (or longitudinal data)
    - ★ individuals over time
    - ★ denoted by subscripts i and t, e.g. x_{it}.
- The same basic statistical methods apply in all cases
  - but each has its own special considerations for statistical inference
    - ★ notably computing standard errors (the precision of estimates)
  - and has its own special considerations for model specification.
- We focus on cross-section data
  - this is the simplest and most common case.

1.3 Regression Analysis

- Economic data analysis focuses on regression analysis.
- Example in chapters 5-7 is relationship between house price (y) and house size in square feet (x) for 29 sales
  - slope is 74 so one more square foot associated with \$74 higher price



Book Outline

- Univariate data (chapters 2-4)
  - single series x
  - covered in introductory statistics.
- Bivariate data (chapters 5-9)
  - two series y and x
  - regression line is $y = b_1 + b_2 x$
- Multivariate data (chapters 10-15)
  - many series
  - regression line is $y = b_1 + b_2 x_2 + b_3 x_3 + \cdots + b_k x_k$
- Further Topics (chapters 16-17).

Background

- Summation notation is used throughout
  - $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$
  - e.g. $\sum_{i=1}^{3}(2 + 3/i) = (2 + 3/1) + (2 + 3/2) + (2 + 3/3) = 11.5$.
- Calculus is used occasionally but is not essential
  - Let Δy denote the change in y and Δx denote the change in x
  - Then $\Delta y / \Delta x$ is the change in y when x changes by one unit.
  - The derivative dy/dx equals $\Delta y / \Delta x$ as $\Delta x \to 0$.
- Natural logarithms and exponentials are used (Chapter 9).
- Expected values are used (Chapter 3). In particular
  - Population mean $\mu = \mathrm{E}[X]$
  - Population variance $\sigma^2 = \mathrm{E}[(X - \mu)^2]$

Key Learning Tool

- Learning-by-doing.
  - Do data examples using an econometrics or statistical package
  - Do chapter exercises and course assignments.