

Econometrics

Part A

October 2022 – December 2022

Prerequisites

Basic knowledge of statistics and probability:

- ▶ Discrete and continuous random variables.
- ▶ Probability distributions: especially the Gaussian.
- ▶ Descriptive statistics: sample mean, variance and confidence intervals.

What is econometrics?

- ▶ Stock and Watson state that:

Econometrics is the science and art of using economic theory and statistical techniques to analyze economic data.

What is econometrics?

- ▶ Stock and Watson state that:

Econometrics is the science and art of using economic theory and statistical techniques to analyze economic data.

- ▶ In this course we will learn econometric techniques used to answer economic questions utilizing data on individuals, firms, municipalities, states or countries observed at one or multiple points in time.

Population and samples

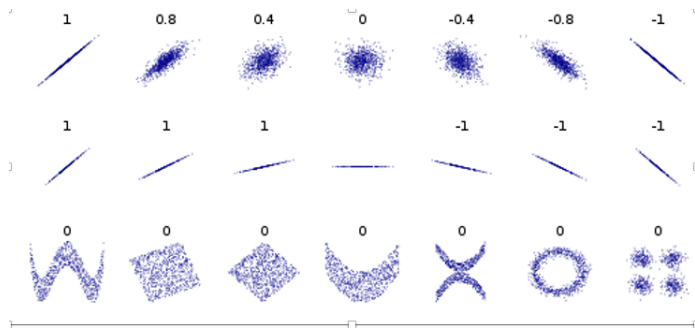
- ▶ Let X and Y two variables for which we seek their relationship (association).
- ▶ First thing to check is their **correlation**:

$$\rho = \frac{\text{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

- ▶ But a census is time consuming and usually costly in money.
- ▶ Instead of a census we can draw random samples and calculate a **sample estimator** r for ρ :

$$r = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}} = \frac{S_{xy}}{S_x S_y}$$

Visualization of r



Econometric Questions: Does reducing class size improve test scores?

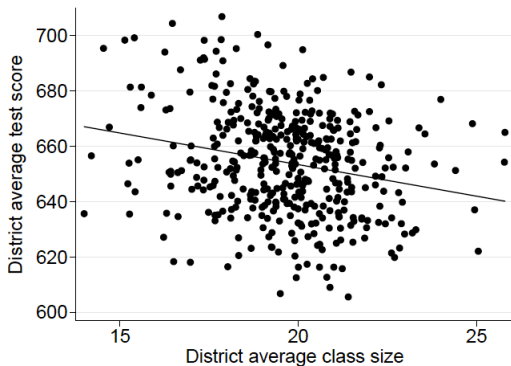


Figure: Using data on 420 California school districts with information on class size and test scores, we will analyze whether reducing class size improves students test scores.

Econometric Questions: What are the returns to education?

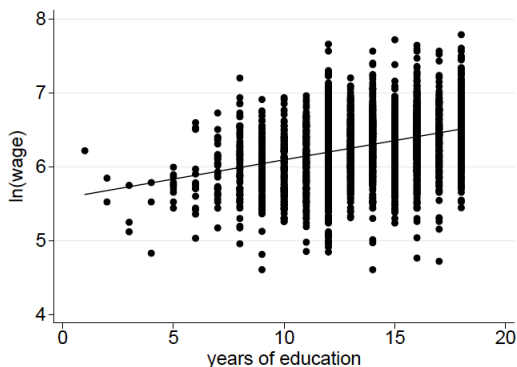


Figure: Using data on 3,010 full-time working men in the US we will analyze whether obtaining more years of education increase wages.

Econometric Questions: Does increasing the tax on beer reduce traffic fatalities?

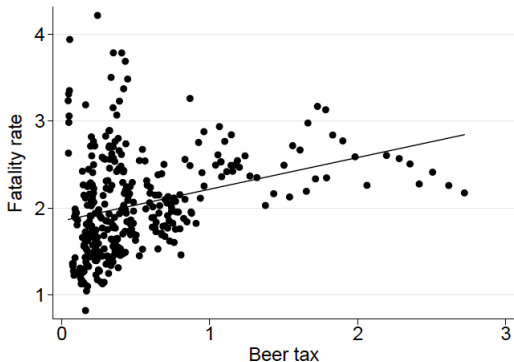


Figure: Using data on 48 U.S. states for the years 1982-1988, we will analyze whether there is an effect of the tax on beer on the traffic fatality rate.

An important question beyond correlation

- ▶ What is the effect on Y if we reduce X by one? Or can we explain the variations in Y from those in X ?

An important question beyond correlation

- ▶ What is the effect on Y if we reduce X by one? Or can we explain the variations in Y from those in X ?
- ▶ We answer by drawing the **population regression** line shown in the Figures:

$$y = \beta_0 + \beta_1 x := E(y|x) \quad (1)$$

- ▶ y : response of dependent variable
- ▶ x : explanatory or independent variables or covariates
- ▶ β_1 is the slope of the line and give the change on y for a unit change on x .

An important question beyond correlation

- ▶ What is the effect on Y if we reduce X by one? Or can we explain the variations in Y from those in X ?
- ▶ We answer by drawing the **population regression** line shown in the Figures:

$$y = \beta_0 + \beta_1 x := E(y|x) \quad (1)$$

- ▶ y : response of dependent variable
 - ▶ x : explanatory or independent variables or covariates
 - ▶ β_1 is the slope of the line and give the change on y for a unit change on x .
- ▶ (1) describes a perfect/deterministic/population relation between x and y .

An important question beyond correlation

- ▶ What is the effect on Y if we reduce X by one? Or can we explain the variations in Y from those in X ?
- ▶ We answer by drawing the **population regression** line shown in the Figures:

$$y = \beta_0 + \beta_1 x := E(y|x) \quad (1)$$

- ▶ y : response of dependent variable
 - ▶ x : explanatory or independent variables or covariates
 - ▶ β_1 is the slope of the line and give the change on y for a unit change on x .
- ▶ (1) describes a perfect/deterministic/population relation between x and y .
 - ▶ We usually rely on random samples and thus we also add an “error” term:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \text{Dist}(\theta). \quad (2)$$

An important question beyond correlation

- ▶ What is the effect on Y if we reduce X by one? Or can we explain the variations in Y from those in X ?
- ▶ We answer by drawing the **population regression** line shown in the Figures:

$$y = \beta_0 + \beta_1 x := E(y|x) \quad (1)$$

- ▶ y : response of dependent variable
 - ▶ x : explanatory or independent variables or covariates
 - ▶ β_1 is the slope of the line and give the change on y for a unit change on x .
- ▶ (1) describes a perfect/deterministic/population relation between x and y .
 - ▶ We usually rely on random samples and thus we also add an “error” term:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \text{Dist}(\theta). \quad (2)$$

- ▶ (2) is the simple linear regression model.

Simple linear regression model: Assumptions

- A.1 Linear relationship between y and x .
- A.2 $E(\epsilon|x) = 0$: x does not carry information for the error term ϵ .
- A.3 $\{x_i, y_i\}_{i=1}^n$ are i.i.d

OLS estimation

- ▶ Derivation of OLS estimators
- ▶ Goodness of fit
- ▶ Statistical properties