Econ 318 – Econometrics

Richard Schwinn

Spring 2015
MW 4:15-5:30 p.m.
Section 1

Text: *A Guide to Basic Econometric Techniques* by Elia Kacapyr

What is Econometrics?

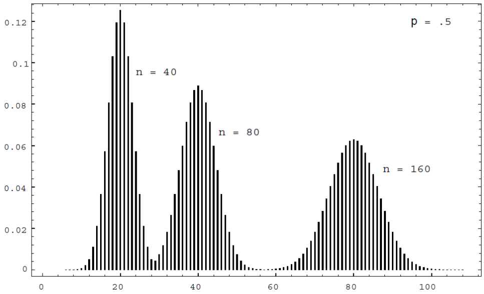
It means *economic measurement*.

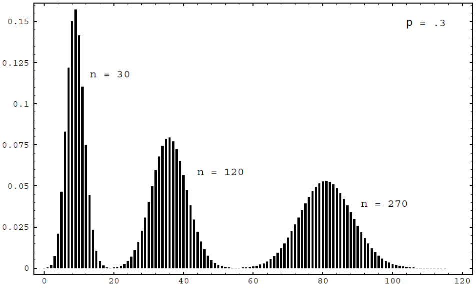
Terminology

1. e.g. (exempli gratis) $\equiv$ *example*
2. i.e. (id est) $\equiv$ *that is*

Bernoulli Trials

Consider a Bernoulli trials process with probability p for success on each trial. Let $X_i = 1$ or 0 according as the ith outcome is a success or failure. Then $S_n = X_1 + X_2 + \cdots + X_n$ counts the number of successes in n trials.





Some observations:

- The maximum values appeared near the expected value np, which drags the graph to the right.
- These maximum values approach 0 as n increased, which causes the spikes graphs to flatten out.
- However, if we properly scale S_n, something interesting will happen.

Standardized Sums

The *standardized sum* of S_n is given by

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}. \tag{1}$$

S_n^* always has expected value 0 and variance 1.

We rescale the distribution histogram of S_n for $n = 270$ and $p = .3$ and compare it with standard normal density.

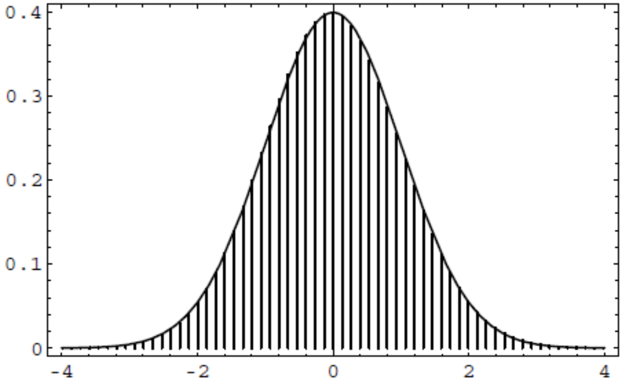


Figure 1: Distribution histogram of S_n^* compared with standard normal density

Observations

- From the previous plot, we observe that the distribution of S_n^* is approximately $N(0,1)$.
- Instead of standardizing S_n, if we standardize the average $A_n = (X_1 + \cdots + X_n)/n$, we will have a similar result, namely the standardized average A_n^* is approximately $N(0,1)$.
- Roughly speaking, if we standardize the sum (or the average) of iid random variables, when n is large, the standardized sum (or average) will be approximately standard normal.

Quick Review

1. Former basketball player Michael Jordan is 77 inches tall. Assuming that heights follow approximately a normal distribution with mean 70 and standard deviation $\sigma = 3$,
   1.1 What is his corresponding z-score?
   1.2 What proportion of men are taller than him?
2. For each problem below draw a picture of the normal curve and shade the area you have to find. Let Z represent a variable following a standard normal distribution.
   2.1 Find the proportion that is less than $z = 2.00$.
   2.2 Find the proportion that is between$z = .13$ and $z = 1.75$.
   2.3 Find the proportion that is greater than $z = 1.86$.
   2.4 Find the z-score for the 64^{th} percentile.
   2.5 Find the z-scores that bound the middle 50% of all data
   2.6 Find the z-score for the 24^{th} percentile.

Principles of Data Reduction

An experimenter uses information in a sample $X_1, ..., X_n$ to make inferences about an unknown parameter θ. If n is large then the sample may be hard to interpret.

- Any statistic $\beta_0(X_1, , X_n)$, defines a form of data reduction.
- If one uses only the observed value of the statistic $\beta_0(X_1, ..., X_n)$, rather than the whole observed sample $(X_1, ..., X_n)$ then one will treat any two samples $(x_1, ..., x_n)$ and $(y_1, ..., y_n)$, that satisfy $\beta_0(x_1, ..., x_n) = \beta_0(y, ..., y_n)$, as equal
- even though the actual sample values may be different in some ways.

Note that

- Capital letter are used to denote random variables,
- Lower case letters are used for sample data,
- Greek letters are typically used for parameters and their estimates.

The **Scientific Method** is the systematic use

- of observation, measurement, and experiment
- in the formulation testing, and modification of hypotheses.
- It is different from other methods in that scientists seek to let reality speak for itself.
- Econometrics is a specific form of the scientific method
  - First we state an economic theory. For example, *When income goes up, consumer spending goes up, but not by as much*. (i.e. people save some of their increased income)
  - Then we test the theoretical predictions empirically (i.e. using data)

Keynsian Law of Consumption

- In words: When income increases, consumer spending goes up, but not by as much.
- Mathematically

$$Consumption = Autonomous\ Consumption + mpc * Income$$

or for short

$$C_i = C_a + mpc * Y_i$$

Keynsian Law of Consumption

$$C_i = C_a + mpc * Y_i$$

Recall that Greek is typically used for parameters we will be estimating. ϵ_i is used to represent all other unidentified factors important to consumption.

$$C_i = \beta_0 + \beta_1 * Y_i + \epsilon_i$$

What does Keyness law imply about β_1?

$$0 < \beta_1 < 1$$

Three Data Formats

Cross-section Data

Indiv.	C	Y
1	76k	83k
2	22k	23k
3	56k	57k
4	45k	36k

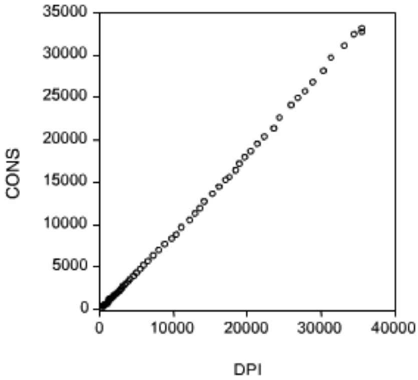
Time Series Data

Year	C	Y
2002	24.6k	26.2k
2003	25k	26.5k
2004	25.8k	27.2k
2005	25.4k	27.3k

Panel Data

Year	Indiv.	C	Y
2002	1	76k	83k
2002	1	77k	82k
2003	2	22k	23k
2003	2	29k	35k

- This is a random sample of incomes and levels of consumption taken in 2010.
- What kind of dataset is displayed?

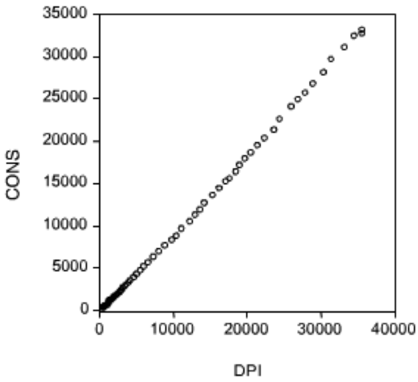


$$C_i = \beta_0 + \beta_1 * Y_i$$

Let's consider possible estimates of the model parameters. Suppose we think that people spend 100% of their income and that they have no other sources of wealth.

What would our null hypothesis, H_0, (the theory to be tested) look like?

$$H_0 : C_i = 0 + 1.0 * Y_i$$



$$H_0 : C_i = 0 + 1.0 * Y_i$$

This implies that

$$\beta_0 = 0$$

and

$$\beta_1 = 1.$$

In order to test the H_0, I've estimated the model using the crown jewel of econometrics: **Operation of Least Squares (OLS)**. OLS is also known as regression.

$$C_i = -0.16 + 0.93Y + e_i$$

The coefficients are only estimates and not necessarily true so we use *hats* to denote estimated parameters.

$$\widehat{\beta}_0 = -0.16$$

and

$$\widehat{\beta}_1 = 0.93.$$

$$H_0 : C_i = 0 + 1.0 * Y_i$$

We estimated

$$C_i = -0.16 + 0.93Y + e_i$$

which implies

$$\widehat{\beta}_0 = -0.16 \neq 0$$

and

$$\widehat{\beta}_1 = 0.93 \neq 1.$$

- For now we don't have the statistical tools to test whether these estimates offer evidence contradicting the H_0.
- Instead, we on to the next use of modeling techniques: Forecasting

Forecasting

$$C_i = -0.16 + 0.93Y + e_i$$

- If I earned 40k last year, what is my expected spending?
- Simply plug 40k into the estimated model.

$$C_i = \widehat{\beta}_0 + \widehat{\beta}_1(40k) + e_i$$

$$C_i = -0.16 + 0.93(40k) + e_i$$

$$C_i = 37.2$$

Problem

Problem

Hours	GPA
2	2
2	3
3	4

Plot the points and draw a line through the dots so the vertical distance between the dots and the line equals zero. Can we do this with other lines?

Jobs

- Moneyball (Statisticians taking jobs from scouts)
- Marketing Analytics (Statisticians taking jobs from marketing professionals)
- Risk Analysis (Statisticians replacing bank loan officers)

Terminology

1. **Cross-sectional data** – Observations on multiple entities at one point in time.
2. **Econometrics** – Testing economic hypotheses with statistical techniques.
3. **Econometric model** – A mathematical expression of the relationship between variables.
4. **Ordinary least-squares** – A technique for fitting lines to scattergrams.
5. **Panel data** – Observations on multiple entities over time.
6. **Scattergram** – A graph showing corresponding pairs of values for two variables.
7. **Time-series data** – Observations on an entity over time

- *A Guide to Basic Econometric Techniques* by Elia Kacapyr