# BU551V - Applied Econometrics
# Lecture Notes

Rodrigo Miguel

February 8, 2023

# Contents

# Chapter 1

# Introduction to Applied Econometrics

## 1.1 What is statistics?

Statistics is the science of uncertainty.

- What *could* be;
- What *might* be;
- What *probably* is.

Statistics can be divided into two:

- **Descriptive statistics:** Used to summarize information (how many), describe in a concise and easy to understand way.
- **Inferential statistics:** Asks about what we can learn from the population in the sample.

### 1.1.1 Example

If an airline is selling tickets to a flight with only 100 seats. A safe solution is to accept only 100 reservations. However, due to the human nature, cancellations and no-show-ups might happen. What would be the optimal number of reservations the airline should make?

## 1.2 Statistical Inference

### 1.2.1 Statistical Inference

Is used when we want to learn about a population given a sample/subset.

### 1.2.2 Steps to Make a Statistical Inference

- Identify the population of interest;
- Specify a model for the population relationship of interest that contains unknown parameters.
- Obtain a **random** sample from the population.
- Estimate the parameters using the sample.

### 1.2.3 Random Sample

A **Random Sample** is a set of random variables (RVs) that follow a probability density function (pdf) $f(y, \theta)$.
When the RVs are from a random sample with $f(y, \theta)$ density, they are said to be **independent and identically distributed**.
Random samples can be assumed to be drawn from a normal distribution. Its population can be characterized by two parameters:

- Mean $\mu$;
- Variance $\sigma^2$.

Usually we are only interested in $\mu$, however, in some cases, we also need to know $\sigma^2$.

### 1.2.4 Estimators

An estimator is a rule that assigns a value of $\theta$ to the sample.
For an actual realized sample, the estimate is just the average in the sample:

$$\overline{y} = \frac{(y_1 + y_2 + ... + y_n)}{n}$$

### 1.2.5 General Expression for Estimators

An estimator $W$ of a parameter $\theta$ can be expressed as an abstract formula:

$$W = h(Y_1, Y_2, ..., Y_n)$$

, where $h$ is an unknown function.
When we plug a set of actual observations $y_1, y_2, ...y_n$ into the function h, we obtain an estimate of $\theta$.

## 1.3 Choosing an estimator

To choose an estimator, criteria needs to be followed so that an estimator with "desirable" properties is correctly chosen. Here's the several properties that are used as choosing criteria:

- Unbiasedness;
- Efficiency;
- Consistency.

## 1.4 Unbiasedness

An estimator of $\theta$ is **unbiased** if:

$$E(W) = \theta$$

This does, however, **not** mean that an estimate in a particular sample is equal to $\theta$.
It means that:

1. We draw random samples from the population many times;
2. Compute an estimate each time;
3. Average these estimates over all random samples;
4. Then we obtain the true parameter $\theta$.

### 1.4.1 Bias of an estimator

An estimator of $\theta$, W, is a biased estimator if:

$$E(W) \neq \theta$$

Its bias is defined as:

$$Bias(W) \equiv E(W) - \theta$$

- An estimator being unbiased does not necessarily mean it's a good estimator.

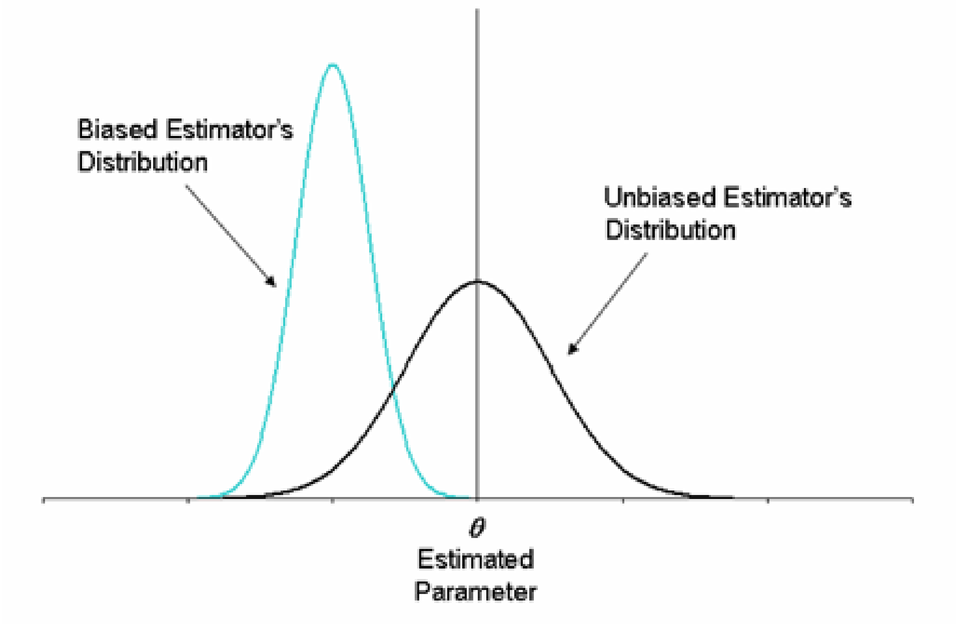


Figure 1.1: An unbiased estimator W, and an estimator with positive bias, $W_2$.

## 1.5 Efficiency

### 1.5.1 Sampling Variance of Estimators

- Unbiasedness ensures that the distribution of an estimator has a mean value which is equal to the true parameter $\theta$.
- It is a good property but not enough.
- We want an estimator that can show us a mean value equal to $\theta$ and centred around $\theta$ as tightly as possible.

To measure how spread out a distribution is, we use the variance, Var(W). To calculate it, we use the sample average as an estimator $w = \overline{Y}$.

**Summary** If $Y_1, ..., Y_n$ is a random sample from a population with mean $\mu$ and variance $\sigma^2$:

- $\overline{Y}$ has a mean $\mu$ and variance $\frac{\sigma^2}{n}$

### 1.5.2 Which estimator is better?

Among unbiased estimators, we prefer the estimator with the smallest variance.

## 1.6 Consistency

Consistency, is a property that captures how estimators improve as the sample size, $n$, increases.

### 1.6.1 Example

- An estimator $\overline{Y}$ has $Var(\overline{Y}) = \frac{\sigma^2}{n}$, meaning that as $n$ gets larger, the estimator improves.
- An estimator $Y_1$ has $Var(Y_1) = \sigma^2$, meaning no improvement even if $n$ gets larger.

### 1.6.2 Importance of Consistency

Consistency is a minimum requirement of an estimator used in econometrics.

### 1.6.3 Law of Large Number

The result that $\overline{Y}$ is consistent for $\mu$ is known as the **law of large numbers** (LLN):

$$plim(\overline{Y}) = \mu$$

# Chapter 2

# Point Estimation

## 2.1 Terminology and Notation

### 2.1.1 Estimator vs. Estimate

- Estimator is a formula;
- Estimate is a value.
- Suppose $Y_1$, ..., $Y_n$ is a random sample from the population with mean $\mu$.
  - An example of an **estimator** is the sample average $\overline{Y}$:

    $$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

  - Using a sample, the actual value of $\overline{Y}$ can be calculated.
  - Using this,

    $$\overline{y} = \frac{1}{3}(16 + 8 + 21) = 15$$

    , *15* is the **estimate**.

### 2.1.2 Notations

- Upper case is used to denote an estimator.
  - Which itself is a random variable.
  - Example: $W$, $\overline{Y}$, $S_X$, $S_{XY}$.
- After obtaining a sample, you can calculate an estimate.
- Lower case is used to denote an estimate.
  - Example: $w$, $overliney$, $s_X$, $s_{XY}$.

## 2.2 Maximum Likelihood and Least Squares

### 2.2.1 Maximum likelihood

- Suppose $Y_1$, ... $Y_n$ is a random smample from the population distribution, f(Y; $\theta$).
- As Yi is a random sample from f(y;$\theta$), the joint distribution of $Y_1$, ..., $Y_n$ is:

$$f(Y_1;\theta) \times f(Y_2;\theta) \times ... \times f(Y_n;\theta)$$

- Using this, define the likelihood function as:

$$L(\theta; Y_1, ..., Y_n) = f(Y_1;\theta) \times f(Y_2;\theta) \times ... \times f(Y_n;\theta)$$

- The maximum likelihood estimator (MLE) of $\theta$, call it W, chooses the value of $\theta$ that maximises the likelihood function.
- For a practical reason, it is more convenient to work with the log-likelihood function.
- The MLE is known to be:
  - Consistent and sometimes unbiases;
  - Asymptotically efficient.
- We will need maximum likelihood to estimate the parameters of advanced econometric models.

### 2.2.2 Least Squares

- The second estimator is called **least squares estimator**;
- Suppose $Y_1$, ..., $Y_n$ is a random sample from the population iwth mean $\mu$;
- Objective - We want to estimate $\mu$ using the sample;
- How? - The least squares estimator choose the value of $m$ that makes the sum of squared deviations as small as possible;
- Solving this for **m** gives:

$$m = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}$$

  - whch is simply the sample average.

- For many distributions, such as the normal distribution, the sample average $\overline{Y}$ is also the MLE of $\mu$;
- Thus, the principles of least squares and maximum likelihood often result in the same estimator.

## 2.3 Method of Moments

- Usually, the parameter to be estimated, $\theta$, is shown to be related to some expected value in the distribution of $Y$, such as $E(Y)$. Using this, the method of moments estimation proceeds as follows.
- Suppose that $\theta$ is related to the population mean as:

$$\theta = g(\mu)$$

  for some function g.
- To estimate $\theta$, it is natural to replace $\mu$ with $\overline{Y}$, which gives us the estimator of $\mu$:

$$g(\overline{Y})$$

- Summary - What we are doing is replacing the population moment, $\mu$, with its sample counterpart, $\bar(Y)$.

### 2.3.1 Sample Covariance

- The covariance between two random variables, X and Y, is defined as:

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

- Method of moments suggests replacing population moments with their sample counterparts:

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})(Y_i - \overline{Y})}{n} = \frac{1}{n} \times \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

  - This is a consistent estimator of $\sigma_{XY}$.
- It can be shown (even though we don't) that replacing $n$ with $n - 1$ yields an unbiased (and consistent) estimator of $\sigma_{XY}$:

$$S_{XY} = \frac{1}{n-1} \times \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

### 2.3.2 Sample correlation coefficient

- The population correlation is defined as:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- The method of moments suggests estimating $\rho_{XY}$ by replacing population moments with their sample counterparts;
  - Called "sample correlation coefficient".