

# **Econometrics II**

## **Review of Probability and Statistics**

Lasha Chochua

2026

# Welcome to Econometrics II

- **Instructor:** Lasha Chochua
  - Email: [lasha.chochua@iset.ge](mailto:lasha.chochua@iset.ge)
- **Teaching Assistants:**
  - **Khatia Javakhishvili** - [khatia.javakhishvili@iset.ge](mailto:khatia.javakhishvili@iset.ge)
  - **Mariam Lobjanidze** - [m.lobjanidze@iset.ge](mailto:m.lobjanidze@iset.ge)
  - **Sopio Mkervalidze** - [sopio.mkervalidze@iset.ge](mailto:sopio.mkervalidze@iset.ge)
- **Duration:** 14 weeks, not including midterm and final exam weeks
- **Platform:** Moodle for resources and submissions
- **Requirement:** Bring your laptops for practice sessions and, occasionally, for lectures (this will be communicated beforehand).

# Volunteer Teaching Assistants

## Note

The course is supported by the following **Volunteer Teaching Assistants**:

- **Salome Lobjanidze** — [salome.lobjanidze@iset.ge](mailto:salome.lobjanidze@iset.ge)
- **Durmishkhan Gigineishvili** — [durmishkhan.gigineishvili@iset.ge](mailto:durmishkhan.gigineishvili@iset.ge)
- **Tea Pavliashvili** — [tea.pavliashvili@iset.ge](mailto:tea.pavliashvili@iset.ge)
- **Tinatin Tutberidze** — [tinatin.tutberidze@iset.ge](mailto:tinatin.tutberidze@iset.ge)
- **Nika Gogichashvili** — [nika.gogichashvili@iset.ge](mailto:nika.gogichashvili@iset.ge)

# Volunteer Teaching Assistants

**!** Important

## Main areas of support

- **R-related questions and coding support:**  
**Salome Lobjanidze and Durmishkhan Gigineishvili**
- **Theory-related questions and recitations:**  
**Tea Pavliashvili, Tinatin Tutberidze, and Nika Gogichashvili**

# Goals of the Course

- Build on foundational econometrics to develop advanced skills.
  - Focus on multiple regression analysis and its applications.
  - Diagnose and correct common econometric issues (e.g., heteroskedasticity, omitted variable bias).
  - Introduce **instrumental variables, panel data regression, and binary response models**.
  - Gain hands-on experience using **R** for empirical research.

## Main Texts

- 1 Stock, J. H., & Watson, M. W. (2019) - *Introduction to Econometrics*.
  - 2 Wooldridge, J. M. (2019) - *Introductory Econometrics: A Modern Approach*.
  - 3 Heiss, F. (2020) - *Using R for Introductory Econometrics*.
- **Additional Reading:**
    - Lecture notes and materials on Moodle.

# Topics Covered

- 1 Review of Statistics and Probability
- 2 Multiple Regression Analysis
- 3 Hypothesis Tests and Confidence Intervals in Multiple Regression
- 4 Assessing Studies Based on Multiple Regression
- 5 Instrumental Variables (IV) Regression
- 6 Panel Data Regression
- 7 Binary Response Models

# Course Structure

- **Lectures:** 21 hours (1.5h per week over 14 weeks)
- **Seminars:** 21 hours (1.5h per week over 14 weeks)
- **Student Work:** 83 hours, including preparation for exams and assignments.

# Assessment Structure

- **Assignments & Participation:** 10 points
- **Group Project:** 10 points
  - A group of 3-5 students (randomly assigned) must prepare a project that applies econometric methods learned in class using **R** on real-world data.
  - The project should demonstrate **practical implementation, data analysis, and interpretation of results**, showcasing proficiency in econometric techniques.
- **Midterm Exam:** 40 points
- **Final Exam:** 40 points

# Course Expectations: Clearing Common Misconceptions

## 1 Work Weekly – Not Before the Exam

- This course cannot be mastered in one week.
- Concepts build on each other.
- Regular problem solving is essential for success.

## 2 Attend Classes and Recitations

- Lectures develop intuition.
- Recitations train problem-solving skills.
- Missing classes means missing structure.

# Course Expectations: Clearing Common Misconceptions

## 3 Exams Test Understanding – Not Memorization

- Do **not** expect previously solved problems with changed numbers.
- Exams test your ability to apply concepts to new settings.

## 4 Ask Questions – Early and Often

- If something is unclear, ask the TAs, volunteer TAs, or me.
- Do not hesitate to ask questions.
- Confusion that lasts for weeks becomes costly.

## 5 Grades Reflect Performance

- Please do not request grade changes.
- Every exam is graded carefully and consistently.
- The best way to improve your grade is preparation - not negotiation.

# Get Ready!

- Course materials available on **Moodle**.
- First two weeks: **Review of Probability and Statistics**.
- One more time — **Stay engaged, practice regularly, and ask questions!**

# Questions?

Feel free to reach out!

- **Lasha Chochua** - [lasha.chochua@iset.ge](mailto:lasha.chochua@iset.ge)
- **Khatia Javakhishvili** - [khatia.javakhishvili@iset.ge](mailto:khatia.javakhishvili@iset.ge)
- **Mariam Lobjanidze** - [m.lobjanidze@iset.ge](mailto:m.lobjanidze@iset.ge)
- **Sopio Mkervalidze** - [sopio.mkervalidze@iset.ge](mailto:sopio.mkervalidze@iset.ge)

# Review of Probability

# Content

- In this part, we will **review**:
  - Random variables and probabilities, probability distribution, and probability density
  - Expected value, standard deviation, and variance of a random variable
  - Joint probability and conditional probability of two random variables
  - Random sampling, Law of Large Numbers (LLNs), and the Central Limit Theorem (CLT).

## Basic Definitions

- **Random Experiment:** A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes:** The mutually exclusive potential results of a random process.
- **Sample Space:** The set  $S$  of the possible outcomes of an experiment.
- **Event:** A subset of the sample space,  $E \subseteq S$ .
- **Random Variable:** Function from set  $S$  to a real number.

## Basic Definitions

- **Random Experiment:** A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes:** The mutually exclusive potential results of a random process.
- **Sample Space:** The set  $S$  of the possible outcomes of an experiment.
- **Event:** A subset of the sample space,  $E \subseteq S$ .
- **Random Variable:** Function from set  $S$  to a real number.
- **Probability:** A mapping from all subsets of the sample space  $S$  to  $[0, 1]$  with these properties:
  - $\Pr(S) = 1$
  - $0 \leq \Pr(E) \leq 1$  for all  $E \subseteq S$
  - for disjoint events,  $\Pr(E_1 \cup E_2 \cup \dots) = \Pr(E_1) + \Pr(E_2) + \dots$

## Examples

- **Random Experiment:** Flipping a coin.
- **Sample Space:**

## Examples

- **Random Experiment:** Flipping a coin.
- **Sample Space:**

$$S = \{H, T\}.$$

- **Random Variable:**

## Examples

- **Random Experiment:** Flipping a coin.
- **Sample Space:**

$$S = \{H, T\}.$$

- **Random Variable:**
- Equivalent representations:

$$X = \begin{cases} 1 & \text{if } H \\ 0 & \text{if } T \end{cases}$$

or

$$X = \begin{cases} 1 & \text{if } T \\ 0 & \text{if } H \end{cases}$$

## Examples

- **Random Experiment:** Flipping a coin.

- **Sample Space:**

$$S = \{H, T\}.$$

- **Random Variable:**

- Equivalent representations:

$$X = \begin{cases} 1 & \text{if } H \\ 0 & \text{if } T \end{cases}$$

or

$$X = \begin{cases} 1 & \text{if } T \\ 0 & \text{if } H \end{cases}$$

- Other examples: Rolling a die, rolling two dice and observing the sums, drawing colored balls from an urn.

# Discrete Random Variables

- A random variable is a function from the sample space  $S$  to a real number.
- A random variable is **discrete** if it takes countably many distinct values:  
 $X \in \{x_1, \dots, x_n\}$ .
- **Probability Mass Function (PMF):**

$$p = f(x) = \Pr(X = x)$$

- **Cumulative Distribution Function (CDF):**

$$F(x) = \Pr(X \leq x) = \sum_{i=1}^n p_i 1\{x_i \leq x\}$$

where  $1\{x_i \leq x\}$  is an indicator variable equal to 1 if the condition is satisfied; 0 otherwise.

# Continuous Random Variables

- A **continuous random variable** is one which takes an uncountably infinite number of possible values, each with vanishingly small probability, where the distance between values is usually meaningful.
- **Examples:** A real value taken between 0 and 1; a random point on a plane; measurement of lengths, weights, temperatures.
- The **cdf** of a continuous random variable is continuous and differentiable—its **pdf** may have jumps, but commonly used distributions can be represented by a continuous, differentiable probability density function.

## Continuous Random Variables (Cont.)

- **Probability Density Function:** The derivative of the Cumulative Distribution Function:

$$f(x) = \frac{dF}{dx}(x)$$

$$F(x) = \int_{-\infty}^x f(t) dt$$

- The lower bound of the support of the random variable could be strictly greater than  $-\infty$ , of course, a common case being the positive real numbers:

$$\int_0^x f(t) dt$$

## Conditional Probability

- **Joint Probability:** Probability of event  $A$  and event  $B$  occurring together:

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event  $A$  given that event  $B$  has occurred:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- **Marginal Probability:** Unconditional Probability of event  $A$  irrespective of whether event  $B$  has occurred or not—“an absolute” determined in the universe.

## Conditional Probability (Cont.)

- $A$  is **independent** of  $B$  iff  $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$
- $A$  independent of  $B$  is sometimes denoted  $A \perp\!\!\!\perp B$ .

- $A \perp\!\!\!\perp B$  implies:

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If  $A$  and  $B$  are independent, knowing whether  $B$  occurred carries no information about  $A$ .

# Bayes' Rule

- **Bayes' Rule:**

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}$$

- **Example:** Interpreting the results of screening tests. The test is not perfect—false positives and false negatives randomly occur.
- A positive test is therefore only a presumption of sickness, not an absolute certainty.

## Bayes' Rule (Cont.)

- Let  $\Pr(\text{sick})$  be the unconditional probability of being sick. Bayes' rule gives the probability of being sick conditional on testing positive,  $\Pr(\text{sick}|\text{positive})$ :

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{positive}|\text{sick}) \cdot \Pr(\text{sick})}{\Pr(\text{positive})}$$

- $\Pr(\text{sick}|\text{positive})$  is larger than  $\Pr(\text{sick})$ .
- $\Pr(\text{sick}|\text{positive})$  and  $\Pr(\text{positive}|\text{sick})$  are commonly confused, but they can be very different.

## Bivariate Distributions: Discrete Case

- **Discrete Bivariate Distribution:**

- **Joint pmf:**

$$f_{X,Y}(x, y) = \Pr(X = x, Y = y)$$

- **Marginal pmf:**

$$f_X(x) = \Pr(X = x) = \sum_{y \in Y} \Pr(X = x, Y = y)$$

- **Conditional pmf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

## Bivariate Distributions: Continuous Case

- **Continuous Bivariate Distribution:**

- **Joint cdf:**

$$F_{X,Y}(x, y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, t) du dt$$

- **Joint pdf:**

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y)$$

- **Marginal cdf:**

$$F_X(x) = \Pr(X \leq x) = F_{X,Y}(x, \infty)$$

- **Marginal pdf:**

$$f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

- **Conditional pdf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

# Independence

- Two random variables  $X, Y$  are independent if:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for all } x, y$$

- This definition holds for both discrete and continuous variables.
- $X \perp\!\!\!\perp Y$  implies:

$$f_{X|Y}(x|y) = f_X(x)$$

# Expectation

- **Expectation of Discrete Random Variable:**

$$E[X] = \sum_{i=1}^n x_i f_X(x_i)$$

- **Expectation of Continuous Random Variable:**

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

- The expectation of a random variable is also known as “expected value,” “first moment,” and more casually as “average.”

## Properties of Expectations

- The expectation operator is linear:

$$E[aX + bY] = aE[X] + bE[Y]$$

- The expectation of a composition is:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

for some real-valued function  $g$  with appropriate support.

# Variance

- The expectation and variance are special cases of the “moments” of a random variable. The expectation is the first moment—the variance is the second “central” moment. Moments of order  $k$  are defined as follows:

- $k$ th moment of  $X$ :

$$E[X^k]$$

- $k$ th central moment of  $X$ :

$$E[(X - E[X])^k]$$

- **Variance as Second Central Moment:**

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

# Standard Deviation

- **Standard deviation:**

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

- The standard deviation is expressed in the same units as the expected value, whereas the variance is expressed in squared units, which may not be so easily interpreted.
- While the expectation operator is linear, the variance is quadratic in the following sense:

$$\text{Var}(a + bX) = b^2\text{Var}(X)$$

- The standard deviation satisfies:

$$\sigma(a + bX) = b\sigma(X)$$

- This is not the same as “linearity” since the constant term  $a$  has vanished.

# Covariance

- **Covariance:**

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

- **Properties:**

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$\text{Cov}(X, Y) = E[(X - E[X])Y] = E[X(Y - E[Y])]$$

$$\text{Cov}(a + bX + cY, \alpha + \beta X + \gamma Y) = b\beta \text{Var}(X) + c\gamma \text{Var}(Y) + (b\gamma + c\beta) \text{Cov}(X, Y)$$

- **Bivariate Variance:**

$$\text{Var}(X) = \text{Cov}(X, X)$$

$$\text{Var}(a + bX + cY) = b^2 \text{Var}(X) + c^2 \text{Var}(Y) + 2bc \text{Cov}(X, Y)$$

$$\text{Var}\left(\sum_{i=1}^N b_i X_i\right) = \sum_{i=1}^N \sum_{j=1}^N b_i b_j \text{Cov}(X_i, X_j)$$

## Variance of Sums

- Variance of sums:

$$\begin{aligned}\text{Var}(a + bX + cY) &= E \left[ (a + bX + cY - E[a + bX + cY])^2 \right] \\ &= E \left[ (a + bX + cY - (a + bE[X] + cE[Y]))^2 \right] \\ &= E \left[ (b(X - E[X]) + c(Y - E[Y]))^2 \right] \\ &= E \left[ b^2(X - E[X])^2 + c^2(Y - E[Y])^2 + 2bc(X - E[X])(Y - E[Y]) \right] \\ &= b^2 E \left[ (X - E[X])^2 \right] + c^2 E \left[ (Y - E[Y])^2 \right] + 2bc E \left[ (X - E[X])(Y - E[Y]) \right] \\ &= b^2 \text{Var}(X) + c^2 \text{Var}(Y) + 2bc \text{Cov}(X, Y)\end{aligned}$$

# Correlation

- **Correlation:**

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$$

- **Cauchy-Schwartz Inequality:**

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}$$

- **Properties:**

- $-1 \leq \text{Corr}(X, Y) \leq 1$
- $\text{Corr}(X, Y) = +1$  if  $Y = a + bX$ , with  $b > 0$
- $\text{Corr}(X, Y) = -1$  if  $Y = a + bX$ , with  $b < 0$
- $\text{Corr}(X, Y) = 0$  if  $X$  and  $Y$  are uncorrelated.

# Conditional Expectation

- **Conditional Expectation of Discrete Random Variable:**

$$E[Y|X = x] = \sum_{i=1}^n y_i \Pr(y_i|X = x)$$

- **Conditional Expectation of Continuous Random Variable:**

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|X = x) dy$$

# Law of Iterated Expectation

- Law of Iterated Expectation:

$$E[Y] = E[E(Y|X)]$$

- The computation involves two steps:
  - The **inner expectation** on the right-hand side of the equation is computed using the **conditional distribution** of  $Y$  given  $X$ .
  - The **outer expectation** is computed using the **marginal distribution** of  $X$ .
- If  $X$  has sample space  $x_1, x_2, \dots, x_n$ , the law can be written explicitly as:

$$E[Y] = \sum_{i=1}^n E(Y|X = x_i) \Pr(X = x_i)$$

## Law of Iterated Expectation (Cont.)

- Example: You choose an integer  $X$  at random between 1 and 3. Then you choose an integer  $Y$  at random between 1 and  $X = x$ . Calculate the expected value of  $Y$ :

## Law of Iterated Expectation (Cont.)

- Example: You choose an integer  $X$  at random between 1 and 3. Then you choose an integer  $Y$  at random between 1 and  $X = x$ . Calculate the expected value of  $Y$ :

$$\begin{aligned} E[Y] &= E[E(Y|X)] = \sum_{i=1}^3 E[Y|X = x_i] \Pr(X = x_i) \\ &= \frac{1}{3} \cdot \left( 1 + \frac{1+2}{2} + \frac{1+2+3}{3} \right) \\ &= 1.5 \end{aligned}$$

# Conditional Variance

- **Conditional Variance:**

$$\text{Var}(Y|X) = E [(Y - E[Y|X])^2|X]$$

## **i** Note

Conditional variance simply asks:

*“On average, how large are my squared prediction errors – for a given value of  $X$ ?”*

# Normal Distribution

- **Normal Distribution:** A continuous distribution with a symmetric bell-shaped probability density function.
- The normal density with mean  $\mu$  and variance  $\sigma^2$  is symmetric around  $\mu$  and has 95% of its probability between  $\mu - 1.96\sigma$  and  $\mu + 1.96\sigma$ .
- A normally distributed random variable is uniquely defined by its mean and variance and is denoted:

$$Y \sim N(\mu, \sigma^2)$$

- Two parameters  $\mu$  and  $\sigma^2$  are sufficient to completely describe any normal distribution.

## Standard Normal Distribution

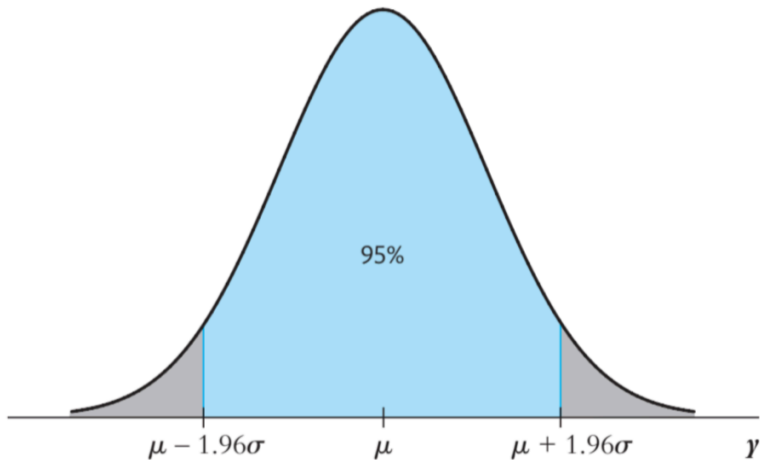
- **Standard Normal Distribution:** The special case with mean zero and unit variance,  $N(0, 1)$ . The standard normal cumulative distribution is often denoted  $\Phi$ , that is, for a fixed value  $z$ :

$$\Pr(Z \leq z) = \Phi(z)$$

- All normal distributions can be converted to the standard normal by standardization:

$$Y \sim N(\mu, \sigma^2) \implies Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

# Normal Distribution



The Normal distribution has "thin tails": very little probability lies in the tails, and so few outliers are expected to occur.

## Chi-Square Distribution

- **Chi-Square Distribution:** Distribution of the sum of  $m$  squared independent standard normal random variables. This distribution is parameterized by the “degrees of freedom”  $m$ .
- Named after the Greek letter “chi”  $\chi$ .
- Let  $Z_1, Z_2, Z_3$  be independent standard normal random variables. Then  $Z_1^2 + Z_2^2 + Z_3^2$  has a chi-squared distribution with 3 degrees of freedom:

$$Z_1, Z_2, Z_3 \sim iid N(0, 1) \implies Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2$$

- The 95th percentile of the  $\chi_3^2$  distribution is 7.81, so:

$$\Pr(Z_1^2 + Z_2^2 + Z_3^2 \leq 7.81) = 0.95$$

## Student-t Distribution

- Named after Statistician **William Sealy Gosset**, who used the pseudonym “Student.”
- Ideally you would standardize like this:

$$Z = \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$$

- But you **never know**  $\sigma$  in reality. So you replace it with  $s$  – your sample estimate:

$$t = \frac{\bar{X} - \mu}{s} \sim t_{n-1}$$

- The problem:  $s$  is random. It wobbles around the true  $\sigma$ . That extra wobble makes the distribution **fatter-tailed** than a normal. Here enters student-t distribution.

## Student-t Distribution (Cont.)

- Let  $Z$  be a standard normal random variable, let  $W$  be a chi-squared random variable with  $m$  degrees of freedom, with  $Z$  and  $W$  independently distributed, then:

$$Z \sim N(0, 1), \quad W \sim \chi^2(m) \implies \frac{Z}{\sqrt{W/m}} \sim t(m)$$

- The Student-t distribution has a bell shape similar to that of the normal distribution, but with “fatter” tails. As the degrees of freedom are increased, the Student-t distribution tends to the standard normal distribution. For values of  $m$  larger than 20, there is little practical difference between the two distributions.

## Fisher's F Distribution

- **F Distribution:** The distribution of the ratio of two independently distributed chi-squared random variables. Used to conduct F-tests, particularly in the context of Analysis of Variance (ANOVA) and when comparing the fits of different linear regression models, e.g., Chow test.
- Named after Statistician **Ronald Aylmer Fisher**.
- Let  $V$  be a chi-squared random variable with  $n$  degrees of freedom, let  $W$  be a chi-squared random variable with  $m$  degrees of freedom, with  $V$  and  $W$  independently distributed, then:

$$W \sim \chi^2(m), \quad V \sim \chi^2(n), \quad W \perp\!\!\!\perp V \implies F = \frac{W/m}{V/n} \sim F(m, n)$$

# The Law of Large Numbers (LLNs) and the Central Limit Theorem (CLT)

## Sampling - Basics

- Let's recall the concept of random sampling.
- A collection of variables  $\{X_1, X_2, \dots, X_n\}$  is a **sample** from the distribution  $F$  if they are identically distributed with distribution  $F$ .
- A collection of variables  $\{X_1, X_2, \dots, X_n\}$  is a **random sample** if they are mutually independent and identically distributed across  $i = 1, \dots, n$ .

**Note:** When observations are both independent and identically distributed, we refer to them as **iid random variables** or a **random sample**.

# Introduction to Asymptotics

- Let  $\{X_n\}$  be a sequence of random variables and  $\bar{X}_n$  be the sample mean of the first  $n$  terms of the sequence:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

- The **Law of Large Numbers** is a proposition stating a set of conditions that are sufficient to guarantee the convergence of the **sample mean** to the **population mean**, as the sample size  $n$  increases. It is called:
  - The **Weak** Law of Large Numbers (WLLN) if the sequence  $\{\bar{X}_n\}$  converges in probability;

## Convergence in Probability - Intuition

- Two random variables are “close to each other” if there is a high probability that their difference is very small.
- Let  $\{X_n\}$  be a sequence of random variables defined on a sample space. Let  $X$  be a random variable and  $\epsilon$  a strictly positive number. Consider the probability:

$$P(|X_n - X| > \epsilon)$$

- $X_n$  is considered far from  $X$  when  $|X_n - X| > \epsilon$ , so this probability measures how likely it is that  $X_n$  is far from  $X$ .
- If  $\{X_n\}$  converges to  $X$ , then  $P(|X_n - X| > \epsilon)$  should decrease as  $n$  increases.

## Convergence in Probability - Formal Definition

**Definition:** A sequence of random variables,  $X_1, X_2, \dots$ , converges in probability to a random variable  $X$  if and only if:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0 \quad \text{for any } \epsilon > 0.$$

- Here,  $X$  is called the probability limit of the sequence, and convergence is indicated by:

$$X_n \xrightarrow{P} X \quad \text{or by} \quad \text{plim } X_n = X \quad \text{as } n \rightarrow \infty.$$

### **i** Note

Convergence in probability is a probabilistic promise, not a guarantee about any single realization. You are controlling the odds – but nature still rolls the dice.

## Convergence in Probability - Example

- Let  $X$  be a discrete random variable with support  $R_X = \{0, 1\}$  and probability mass function:

$$p_X(x) = \begin{cases} \frac{1}{3}, & \text{if } x = 1, \\ \frac{2}{3}, & \text{if } x = 0, \\ 0, & \text{otherwise.} \end{cases}$$

- Consider a sequence of random variables  $\{X_n\}$  whose generic term is:

$$X_n = \left(1 + \frac{1}{n}\right) X$$

- Does  $\{X_n\}$  converge in probability to  $X$ ?

## Convergence in Probability - Example (cont.)

- Take any  $\epsilon > 0$ . Note that:

$$|X_n - X| = \left(1 + \frac{1}{n}\right)X - X = \frac{1}{n}X$$

- Consider the following cases:
  - **Case 1:** When  $X = 0$ , which happens with  $\frac{2}{3}$  probability:

$$|X_n - X| = \frac{1}{n} \times 0 = 0 \quad \text{so} \quad |X_n - X| \leq \epsilon$$

- **Case 2:** When  $X = 1$ , which happens with  $\frac{1}{3}$  probability:

$$|X_n - X| = \frac{1}{n} \times 1 = \frac{1}{n}$$

- $|X_n - X| \leq \epsilon$  if and only if  $\frac{1}{n} \leq \epsilon$  (i.e.,  $n \geq \frac{1}{\epsilon}$ ).

## Convergence in Probability - Example (cont.)

- Therefore:

$$P(|X_n - X| \leq \epsilon) = \begin{cases} \frac{2}{3}, & \text{if } n < \frac{1}{\epsilon} \\ 1, & \text{if } n \geq \frac{1}{\epsilon} \end{cases}$$

- And:

$$P(|X_n - X| > \epsilon) = \begin{cases} \frac{1}{3}, & \text{if } n < \frac{1}{\epsilon} \\ 0, & \text{if } n \geq \frac{1}{\epsilon} \end{cases}$$

- Thus,  $P(|X_n - X| > \epsilon)$  converges to 0 as  $n$  increases.

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0 \quad \text{for any } \epsilon > 0.$$

# WLLN

**Theorem:** Let  $X_1, X_2, \dots$  be iid random variables with  $\mathbb{E}[X_i] = \mu$  and  $\text{Var}(X_i) = \sigma^2 < \infty$ .

Define  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Then, for every  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| < \epsilon) = 1,$$

that is,  $\bar{X}_n$  converges in probability to  $\mu$ .

- The WLLN asserts that, under general conditions, the sample mean converges to the population mean as  $n \rightarrow \infty$ .
- More general versions of the WLLN require only that the mean is finite.

## WLLN - Proof

- The proof is straightforward and uses Chebyshev's Inequality. For every  $\epsilon > 0$ :

$$P(|\bar{X}_n - \mu| \geq \epsilon) = P((\bar{X}_n - \mu)^2 \geq \epsilon^2) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Thus:

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

- Hence, the probability that  $\bar{X}_n$  deviates from  $\mu$  by more than  $\epsilon$  approaches 0 as  $n \rightarrow \infty$ :

$$P(|\bar{X}_n - \mu| < \epsilon) = 1 - P(|\bar{X}_n - \mu| \geq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2} \rightarrow 1 \text{ as } n \rightarrow \infty.$$

# Demonstration of the Law of Large Numbers

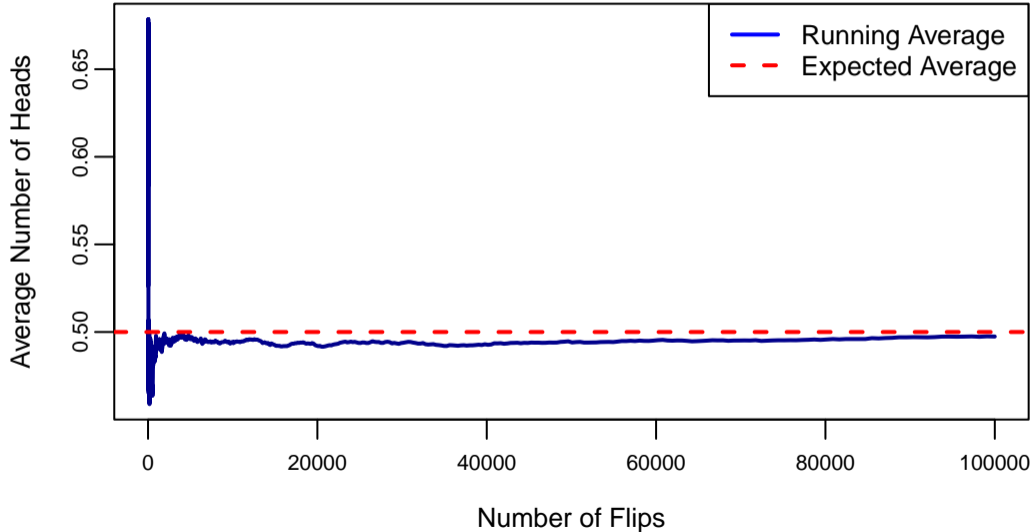
- Ever wondered what happens if you keep flipping a coin infinitely?
- **Warning:** Do not attempt the strategy shown in the picture—unless you're really, really into flipping coins!



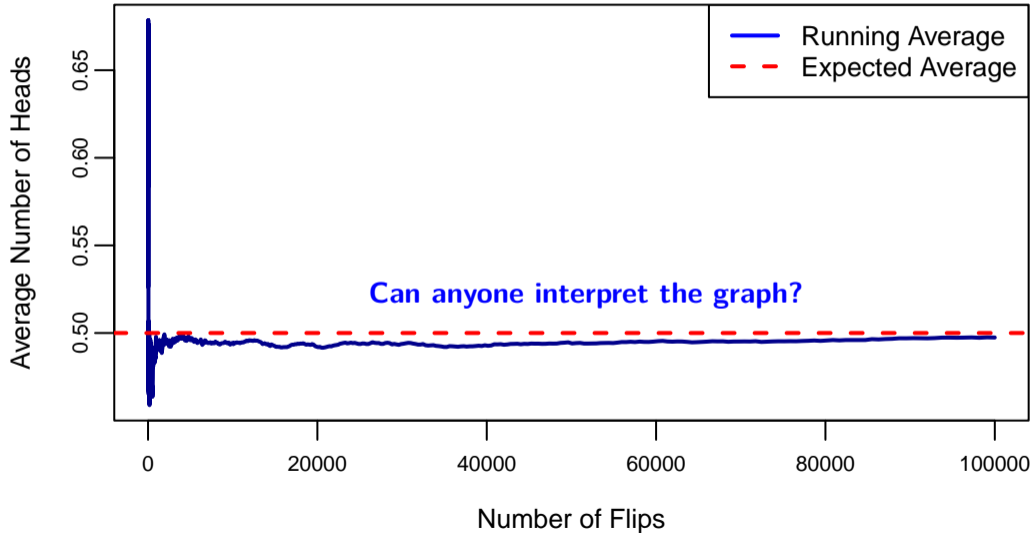
## R Code for Coin Flip Simulation

```
# Set the seed for reproducibility
set.seed(123)
# Set the maximum number of flips
max_flips <- 100000
# Generate a sequence of flip sizes
flip_sizes <- unique(round(exp(seq(log(10),
log(max_flips), length.out = 1000))))
# Simulate the coin flips
flips <- rbinom(max_flips, 1, 0.5)
# Initialize a vector to store the running averages
running_avg <- numeric(length(flip_sizes))
# Calculate the running averages for each flip size
for (i in seq_along(flip_sizes)) {n <- flip_sizes[i]
  running_avg[i] <- mean(flips[1:n])
}
```

# Results of the Coin Flip Simulation



# Results of the Coin Flip Simulation



## A Remarkable Idea

- The **Law of Large Numbers** captures something almost philosophical.
- The world around us is full of randomness.  
Individual outcomes are unpredictable.
- But when we **repeat events many times**, something surprising happens:
  - patterns begin to appear,
  - averages stabilize,
  - randomness reveals an underlying order.
- Humanity slowly discovered this powerful idea:  
*Randomness itself becomes predictable when we look at it in the aggregate.*
- This is one of the most beautiful insights in probability.

## Why This Matters

- This idea is not just mathematics. It is one of the reasons **modern civilization works**.
- Entire systems depend on the stability of averages:
  - insurance,
  - banking,
  - public health,
  - opinion polling,
  - economic measurement.
- If averages did **not** stabilize, we could not price risk, and manage uncertainty.
- Econometrics builds on this insight: **by observing many outcomes, we learn something reliable about the world**.
- The mathematician **Jacob Bernoulli** spent almost **20 years** proving this result.
- He called it his “**Golden Theorem**” because he understood how profound it was.

# Richard Feynman on Beauty and Science

## Note

“I have a friend who’s an artist and has sometimes taken a view which I don’t agree with very well. He’ll hold up a flower and say, ‘Look how beautiful it is,’ and I’ll agree. Then he says, ‘I, as an artist, can see how beautiful this is, but you as a scientist take this all apart and it becomes a dull thing.’

I think he’s kind of nutty.

First of all, the beauty that he sees is available to other people—and to me, too, I believe. Although I may not be quite as refined aesthetically as he is, I can appreciate the beauty of a flower.

At the same time, I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside which also have a beauty. I mean, it’s not just beauty at this dimension of one centimeter; there’s also beauty at a smaller dimension, the inner structure. Also the processes, the fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting; it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic? All kinds of interesting questions which the science knowledge only adds to the excitement and mystery and awe of a flower. It only adds. I don’t understand how it subtracts.” Richard Feynman, *The Pleasure of Finding Things Out* (1981)

# Understanding Convergence in Distribution

- Let  $\{X_n\}$  be a sequence of random variables.
- Define the **cumulative distribution function (CDF)** of  $X_n$  as:

$$F_n(x) = P(X_n \leq x).$$

- Once we fix  $x$ , the value  $F_n(x)$  is a **real number**.
- The sequence  $\{F_n(x)\}$  is a sequence of real numbers.

## Definition - Convergence in Distribution

**Definition:** Let  $\{X_n\}$  be a *sequence of random variables*. Denote by  $F_n(x)$  the distribution function of  $X_n$ .

We say that  $\{X_n\}$  is **convergent in distribution**, or **convergent in law**, if and only if there exists a distribution function  $F_X(x)$  such that the sequence  $\{F_n(x)\}$  converges to  $F_X(x)$  for all points  $x \in \mathbb{R}$  where  $F_X(x)$  is continuous.

$$F_n(x) \rightarrow F_X(x) \quad \text{for all } x \in \mathbb{R}, \text{ where } F_X(x) \text{ is continuous.}$$

If a random variable  $X$  has distribution function  $F_X(x)$ , then  $X$  is called the **limit in distribution** (or **limit in law**) of the sequence, and convergence is indicated by:

$$X_n \xrightarrow{d} X.$$

## Convergence in Distribution - Example

- Let  $\{X_n\}$  be a sequence of random variables having distribution functions:

$$F_n(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - (1 - x)^n & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

- Find the **limit in distribution** (if it exists) of the sequence  $\{X_n\}$ .

## Solution

- If  $x = 0$ :

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} [1 - (1 - x)^n] = 1 - \lim_{n \rightarrow \infty} (1 - 0)^n = 1 - 1 = 0.$$

- If  $0 < x \leq 1$ :

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} [1 - (1 - x)^n] = 1 - \lim_{n \rightarrow \infty} (1 - x)^n = 1 - 0 = 1.$$

- Therefore, the limit function:

$$G_X(x) = \lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

is **not a proper distribution function**, as it is not right-continuous at  $x = 0$ .

## Solution (Cont.)

- However, defining:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

results in a proper distribution function, meaning  $\{X_n\}$  **converges in distribution** to a random variable  $X$  with this distribution.

## Central Limit Theorem (CLT)

- Let  $\{X_n\}$  be a sequence of random variables, and let  $\bar{X}_n$  be the sample mean of the first  $n$  terms:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

- The CLT states that, under certain conditions, the sample mean  $\bar{X}_n$  converges to a normal distribution as  $n$  increases:

$$\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow{d} Z$$

- $Z$  is a standard normal random variable,  $\mu$  and  $\sigma$  are constants.

## Understanding the CLT

- Without  $\sqrt{n}$ ,  $\frac{\bar{X}_n - \mu}{\sigma}$  would converge to zero or explode by the Law of Large Numbers.
- When the CLT conditions hold, the transformed variable follows a normal distribution:

$$\sqrt{n} (\bar{X}_n - \mu) / \sigma \sim N(0, 1).$$

## Practical usage

- 1 Observe a sample of  $n$  observations  $X_1, X_2, \dots, X_n$ .
- 2 If  $n$  is large, the distribution of  $\sqrt{n}(\bar{X}_n - \mu)/\sigma$  is approximately normal.
- 3 Therefore, we assume:

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0, 1).$$

- 4 Consequently, the distribution of the sample mean follows:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

## Lindeberg-Lévy Central Limit Theorem

**Proposition - Lindeberg-Lévy CLT:** Let  $\{X_n\}$  be an *i.i.d.* sequence of random variables such that:

$$\mathbb{E}[X_n] = \mu < \infty, \quad \forall n \in \mathbb{N}$$

$$\text{Var}[X_n] = \sigma^2 < \infty, \quad \forall n \in \mathbb{N}$$

where  $\sigma^2 > 0$ . Then, a Central Limit Theorem applies to the sample mean  $\bar{X}_n$ :

$$\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow{d} Z$$

where  $Z$  is a standard normal random variable, and  $\xrightarrow{d}$  denotes convergence in distribution.

**Note:** The proof will be available in Moodle, but I prefer that you remember the idea behind the proposition.

## CLT - Example

- Let  $\{X_n\}$  be a sequence of independent Bernoulli random variables with parameter  $p = \frac{1}{2}$ .
- A generic term  $X_n$  of the sequence has support:

$$R_{X_n} = \{0, 1\}$$

- The probability mass function of  $X_n$  is given by:

$$p_{X_n}(x) = \begin{cases} 1/2, & \text{if } x = 1 \\ 1/2, & \text{if } x = 0 \\ 0, & \text{if } x \notin R_{X_n} \end{cases}$$

- Use a Central Limit Theorem to derive an approximate distribution for the mean of the first 100 terms of the sequence.

## Solution

- The sequence  $\{X_n\}$  is an IID sequence. The mean of a generic term is:

$$\mathbb{E}[X_n] = \sum_{x \in R_{X_n}} xp_{X_n}(x) = 1 \cdot p_{X_n}(1) + 0 \cdot p_{X_n}(0) = \frac{1}{2} < \infty$$

- The variance of a generic term is:

$$\mathbb{E}[X_n^2] = \sum_{x \in R_{X_n}} x^2 p_{X_n}(x) = 1^2 \cdot p_{X_n}(1) + 0^2 \cdot p_{X_n}(0) = \frac{1}{2}$$

$$\text{Var}[X_n] = \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4} < \infty$$

## Solution (Cont.)

- The sequence  $\{X_n\}$  satisfies the conditions of the **Lindeberg-Lévy Central Limit Theorem** (IID, finite mean, finite variance).
- The mean of the first 100 terms of the sequence is:

$$\bar{X}_{100} = \frac{1}{100} \sum_{i=1}^{100} X_i$$

- Applying the Central Limit Theorem:

$$\bar{X}_n \sim N \left( \mathbb{E}[X_n], \frac{\text{Var}[X_n]}{n} \right)$$

$$\bar{X}_{100} \sim N \left( \frac{1}{2}, \frac{1}{400} \right)$$

## Alternative Derivation

- By the **Central Limit Theorem**, for large  $n$ :

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

- Plugging in the values:

$$\sqrt{100} \left( \bar{X}_{100} - \frac{1}{2} \right) \xrightarrow{d} N\left(0, \frac{1}{4}\right)$$

- Thus, the approximate distribution of  $\bar{X}_{100}$  is:

$$\bar{X}_{100} \sim N\left(\frac{1}{2}, \frac{1}{400}\right).$$

## Convergence of transformations - Useful Rules

# Continuous Mapping Theorem

**Proposition** Let  $\{X_n\}$  be a sequence of  $K$ -dimensional random vectors. Let  $g : \mathbb{R}^K \rightarrow \mathbb{R}^L$  be a *continuous function*. Then:

$$X_n \xrightarrow{P} X \quad \Longrightarrow \quad g(X_n) \xrightarrow{P} g(X)$$

$$X_n \xrightarrow{d} X \quad \Longrightarrow \quad g(X_n) \xrightarrow{d} g(X)$$

# Convergence in Probability of Sums and Products

**Proposition** *If  $X_n \xrightarrow{P} X$  and  $Y_n \xrightarrow{P} Y$ . Then:*

$$X_n + Y_n \xrightarrow{P} X + Y$$

$$X_n Y_n \xrightarrow{P} XY$$

# Convergence in Distribution of Sums and Products

**Proposition** *If*

$$[X_n \ Y_n] \xrightarrow{d} [X \ Y]$$

*then:*

$$X_n + Y_n \xrightarrow{d} X + Y$$

$$X_n Y_n \xrightarrow{d} XY$$

# Slutsky's Theorem

- Slutsky's theorem concerns the convergence in distribution of the product of two sequences, one converging in distribution and the other converging in probability to a constant.

**Proposition** If  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{P} c$ , where  $c$  is a constant, then:

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} cX$$

## Convergence of Ratios

**Proposition** *If two sequences of random variables  $\{X_n\}$  and  $\{Y_n\}$  converge to  $X$  and  $Y$  respectively, then:*

$$X_n/Y_n \rightarrow X/Y$$

*provided  $Y$  is almost surely different from 0. Convergence can be in probability, almost surely, or in distribution (but the latter requires joint convergence in distribution of  $\{X_n\}$  and  $\{Y_n\}$ ).*

# Summary

## 1 Probability Distributions:

- Random variable probabilities are summarized by:
  - Cumulative distribution function (CDF)
  - Probability distribution function (PDF) for discrete variables
  - Probability density function (PDF) for continuous variables

## 2 Expected Value:

- The expected value of  $Y$ ,  $\mathbb{E}(Y)$ , is the probability-weighted average value.
- Variance of  $Y$ :  $\sigma_Y^2 = \mathbb{E}[(Y - \mu_Y)^2]$ .
- Standard deviation is the square root of variance.

## 3 Joint and Conditional Probabilities:

- Joint probabilities for  $X$  and  $Y$  are described by their joint probability distribution.
- Conditional probability distribution of  $Y$  given  $X = x$  describes  $Y$  values when  $X = x$ .

## 4 Normal Distribution:

- A normally distributed random variable has a symmetric, bell-shaped density.

## Summary (Cont.)

### 5 Random Sampling:

- Simple random sampling produces  $n$  i.i.d. observations:  $Y_1, Y_2, \dots, Y_n$ .

### 6 Sample Average $\bar{Y}$ :

- $\bar{Y}$  is a random variable with:
  - **Sampling Distribution:** Mean  $\mu_Y$  and variance  $\sigma_Y^2/n$ .
  - **Law of Large Numbers (LLN):**  $\bar{Y}$  converges in probability to  $\mu_Y$  as  $n$  increases.
  - **Central Limit Theorem (CLT):** For large  $n$ :

$$\frac{\bar{Y} - \mu_Y}{\sigma_Y/\sqrt{n}} \sim N(0, 1)$$

or equivalently:

$$\bar{Y} \sim N(\mu_Y, \sigma_Y^2/n)$$

# Statistics

# Content

- In this part, we will **review**:
  - Estimators of the population mean, the sample mean, the sample variance, the sample standard deviation, the standard error, the sample covariance, and the sample correlation
  - The null hypothesis and alternatives
  - The  $p$ -value and the  $t$ -statistic
  - How to construct confidence intervals for the population mean
  - How to compare means from different populations

# Statistics: The Science of Learning from Data

- **Definition:** Using data to understand the world.
- **Objective:** Answer questions about distributions in populations.
- **Example Questions:**
  - What is the mean earnings of recent graduates?
  - Do mean earnings differ between men and women?

# The Challenge of Comprehensive Surveys

- **Comprehensive Surveys:**

- Exhaustive population surveys are costly and rare.
- Examples:
  - **Censuses:** Conducted periodically (e.g., every 10 years).
- Costs of modern censuses:
  - UK 2011 Census: £482 million.
  - India 2011 Census: 2200 crore (~US\$320 million).

# The Statistical Approach

- **Key Insight:** Use random sampling to study population distributions.
- **Random Sampling Example:**
  - Survey 1000 people from China's population of 1.4 billion (2018).
- **Statistical Inference:** Draw conclusions about the full population from the sample.

# Core Statistical Methods in Econometrics

## 1 Estimation:

- Compute a “best guess” for population characteristics (e.g., mean).

## 2 Hypothesis Testing:

- Test hypotheses using sample evidence.

## 3 Confidence Intervals:

- Estimate a range for an unknown population characteristic.

# Estimating the Population Mean

- **Estimator:** A function of a sample of data drawn randomly from a population. An **estimate** is a numerical value using data from a specific sample. An estimator is a random variable, an estimate is not.
- The sample mean  $\bar{Y}$  is a natural estimate of  $\mu_Y$ .
- It is not the only possible estimator. There are many useful estimators of  $\mu_Y$ .
- **Can you think of other natural estimators?**

## Notation

- An estimator of  $\mu_Y$  is typically denoted  $\hat{\mu}_Y$ .
- We might denote two estimators  $\hat{\mu}_Y$  and  $\tilde{\mu}_Y$ . The sample mean  $\bar{Y}$  is one such estimator.
- If we deal with a single random variable  $Y$ , we sometimes write just  $\mu$  and  $\hat{\mu}$ .
- When our focus is on the sample size  $n$ , we often denote it for emphasis, e.g.,  $\bar{Y}_n$ .
- We may denote a particular estimate with a lower-case letter, e.g.,  $\bar{y}$ .
- We may denote a particular value such as the population mean with a 0 index, e.g.,  $\mu_0$ .

# Properties of Estimators

- **What are desirable characteristics of the sampling distribution of an estimator?**
  - Close to the unknown value ... in an “average” sense!
  - Stabilizes arbitrarily closely to the true value as the sample size is increased.
  - Uses sample data to avoid erratic fluctuations caused by small samples — estimates do not go “all over the place.”
- Some of these properties may be traded off, in particular centrality and dispersion, e.g., bias or consistency vs. efficiency.
- Some estimators may perform well for very large samples, but not so well for small samples.
- To fairly compare estimators, you must be clear about your objectives. For instance, an estimator that is equal to a constant, say 0, for any sample, could have a very large bias, but it is absolutely efficient!

## Properties of Estimators (Cont.)

- **Desirable statistical properties of estimators:**

- 1 **Unbiased:** The estimator  $\hat{\mu}_Y$  is unbiased if:

$$E[\hat{\mu}_Y] = \mu_Y$$

(The estimator is close on average.)

- 2 **Consistent:** The estimator is consistent if the event that  $\hat{\mu}_Y$  is arbitrarily close to the true value  $\mu_Y$  becomes certain as the sample size  $n$  increases:

$$\Pr(|\hat{\mu}_Y - \mu_Y| < \epsilon) \xrightarrow{n \rightarrow \infty} 1$$

A consistent estimator “converges in probability” to the true mean:  $\hat{\mu}_Y \xrightarrow{p} \mu_Y$ .

- 3 **Efficient:**  $\hat{\mu}_Y$  is efficient if it has a smaller variance:

$$\text{Var}(\hat{\mu}_Y) < \text{Var}(\tilde{\mu}_Y)$$

(The more efficient estimator has the tighter sampling distribution.)

# Understanding Sample Averages and the Null Hypothesis

- The sample average,  $\bar{Y}$ , will rarely match the hypothesized value,  $\mu_{Y,0}$ , exactly.
  - Differences arise because:
    - The null hypothesis is false, meaning the true mean does not equal  $\mu_{Y,0}$ .
    - The null hypothesis is true, but random sampling causes  $\bar{Y}$  to differ from  $\mu_{Y,0}$ .
  - It is impossible to distinguish these scenarios with certainty.
- While a sample cannot provide conclusive evidence about the null hypothesis, a probabilistic calculation allows testing the null hypothesis while accounting for sampling uncertainty.

# Hypothesis Testing

- **Null Hypothesis:**

$$H_0 : E[Y] = \mu_{Y,0}$$

where  $\mu_{Y,0}$  is a specific value.

- Example of null: On average, in the population, college graduates earn \$20 per hour.

- **Alternative Hypothesis:** Specifies what is true if the null hypothesis is not.

- **Two-Sided Alternative:**

$$H_1 : E[Y] \neq \mu_{Y,0}$$

- **One-Sided Alternative:**

$$H_A : E[Y] > \mu_{Y,0}$$

(or in the other direction:  $E[Y] < \mu_{Y,0}$ )

- **Statistical Hypothesis Testing:** Either reject the null hypothesis or fail to reject it.

# The $p$ -Value and Its Interpretation

- The  $p$ -value (or significance probability) measures the probability of observing a statistic as extreme or more extreme than the one computed from the sample, assuming the null hypothesis is true.
  - Example:
    - Sample average wage:  $\bar{Y} = 22.64$ .
    - Null hypothesis population mean:  $\mu_{Y,0} = 20$ .
- $p$ -value: Probability of observing  $\bar{Y}$  as far or farther from 20 as 22.64 due to random sampling, assuming  $H_0$  is true.
  - **Small  $p$ -value (e.g., 0.1%)**: Strong evidence to reject  $H_0$ .
  - **Large  $p$ -value (e.g., 40%)**: Weak evidence against  $H_0$ , so do not reject  $H_0$ .

## Calculating the $p$ -Value When $\sigma_{\bar{Y}}$ Is Known

- **Sampling Distribution:**

- If the sample size is large, under the null hypothesis:

$$\bar{Y} \sim N(\mu_{Y,0}, \sigma_{\bar{Y}}^2), \quad \text{where } \sigma_{\bar{Y}}^2 = \sigma^2/n.$$

- **Standardized Version:**

- The standardized  $\bar{Y}$  follows a standard normal distribution:

$$\frac{\bar{Y} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \sim N(0, 1).$$

## Calculating the $p$ -Value When $\sigma_{\bar{Y}}$ Is Known (Cont.)

- **Definition of  $p$ -Value:**

- The  $p$ -value is the probability of obtaining a value of  $\bar{Y}$  farther from  $\mu_{Y,0}$  than  $\bar{Y}_{\text{act}}$  under the null hypothesis:

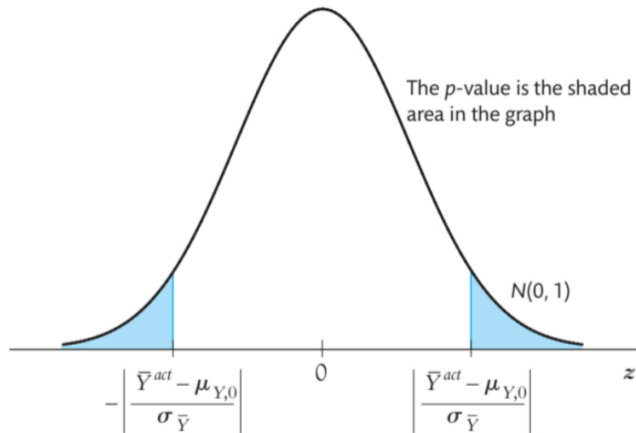
$$p\text{-value} = Pr_{H_0} \left( \left| \frac{\bar{Y} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \right| \geq \left| \frac{\bar{Y}_{\text{act}} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \right| \right).$$

$$p\text{-value} = Pr_{H_0} \left( \frac{\bar{Y} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \leq - \left| \frac{\bar{Y}_{\text{act}} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \right| \right) + Pr_{H_0} \left( \frac{\bar{Y} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \geq \left| \frac{\bar{Y}_{\text{act}} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \right| \right).$$

$$p\text{-value} = 2\Phi \left( - \left| \frac{\bar{Y}_{\text{act}} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \right| \right).$$

- The formula depends on the variance of the population distribution,  $\sigma^2$ .
- In practice,  $\sigma^2$  is typically unknown and must be estimated.

# p-value



The  $p$ -value is the probability of drawing a value of  $\bar{Y}$  that differs from  $\mu_{Y,0}$  by at least as much as the actual sample mean  $\bar{Y}^{act}$ . In large samples,  $\bar{Y}$  is distributed  $N(\mu_{Y,0}, \sigma_Y^2/n)$  under the null hypothesis. The  $p$ -value is the shaded standard normal tail probability.

## Sample Variance and Standard Deviation

- **Sample Variance:** An estimator of the population variance  $\sigma_Y^2$ , denoted  $s_Y^2$  or  $\hat{\sigma}_Y^2$ .
- **Sample Standard Deviation:** An estimator of the population standard deviation  $\sigma_Y$ , denoted  $s_Y$  or  $\hat{\sigma}_Y$ .
  - The sample standard deviation is the square root of the sample variance.
- **Population Variance:**

$$\sigma_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu_Y)^2$$

## Sample Variance and Standard Deviation (Cont.)

- **Sample Variance:**

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

- **Degrees of Freedom Adjustment:** The sample variance divides by  $n - 1$  instead of  $n$  to correct for bias.
- **Consistency:** The sample variance is a consistent estimator of the population variance:

$$s_Y^2 \xrightarrow{p} \sigma_Y^2$$

## Standard Error

- **Standard Error:** An estimator of the standard deviation of the sampling distribution of  $\bar{Y}$ , denoted  $SE(\bar{Y})$ .
  - The standard deviation of  $\bar{Y}$  is  $\sigma_Y/\sqrt{n}$ , so  $s_Y/\sqrt{n}$  is used as an estimator of  $\sigma_Y$ .
  - $s_Y/\sqrt{n}$  is a common and practical estimator.
- If  $Y_1, \dots, Y_n$  are i.i.d., the standard error is:  $SE(\bar{Y}) = \frac{s_Y}{\sqrt{n}}$
- For Bernoulli distribution with success probability  $p$ :  $SE(\bar{Y}) = \sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}$

## $p$ -Value for Unknown $\sigma_Y^2$

- Replace  $\sigma_Y$  with the standard error  $\hat{\sigma}_Y$  to compute the  $p$ -value.
- If  $Y_1, \dots, Y_n$  are i.i.d., the  $p$ -value is:  $p\text{-value} = 2\Phi\left(-\left|\frac{\bar{Y}^{\text{act}} - \mu_{Y,0}}{\text{SE}(Y)}\right|\right)$
- **$t$ -Statistic:**
  - For large  $n$ , the estimator  $s_Y^2$  is close to  $\sigma_Y^2$ .
  - The distribution of the  $t$ -statistic is approximately:  $t \sim N(0, 1)$
- The  $p$ -value can also be expressed using the  $t$ -statistic:  $p\text{-value} = 2\Phi(-|t^{\text{act}}|)$

## Example

- A sample of size  $n = 200$  recent college graduates is used to test the null hypothesis that the mean wage  $\mathbb{E}[Y]$  is \$20 per hour. The sample average is  $\bar{Y}^{\text{act}} = \$22.64$  and the sample standard deviation is  $s_Y = \$18.14$ .
- The standard error of  $\bar{Y}$  is:

$$\text{SE}(\bar{Y}) = \frac{s_Y}{\sqrt{n}} = \frac{18.14}{\sqrt{200}} = 1.28$$

- The  $t$ -statistic is:

$$t^{\text{act}} = \frac{\bar{Y}^{\text{act}} - 20}{1.28} = \frac{22.64 - 20}{1.28} = 2.06$$

- The corresponding  $p$ -value is:  $p\text{-value} = 2\Phi(-2.06) = 0.039$
- Under the null hypothesis, the probability of drawing a sample mean no less different from the null as the one actually computed is **3.9%**.

## One-Sided Alternative

- **One-Sided Alternative:**

$$H_A : \mathbb{E}[Y] > \mu_{Y,0} \quad \implies \quad p\text{-value} = \Pr(Z > t^{\text{act}} | H_0) = 1 - \Phi(t^{\text{act}})$$

- Because the distribution is symmetric, one-sided alternatives can be found with the same critical value in absolute value — only the sign of the critical value differs between left-sided and right-sided alternatives.
- The critical value for the rejection region in a test against a one-sided alternative is  $t_{\text{crit}} = 1.64$  — smaller than the critical value in the case of a two-sided alternative  $t_{\text{crit}} = 1.96$ . This is because the tail probability is located exclusively on one side, rather than divided evenly.

## Confidence Intervals

- **Confidence Set:** Contains the true population mean  $\mu_Y$  with a pre-specified probability.
- **Confidence Level:** Probability that  $\mu_Y$  is within the confidence set.
- **Confidence Interval:** Range of possible values for  $\mu_Y$  between a lower and an upper limit.
- **Coverage Probability:** Likelihood that a confidence interval contains the true population mean.
- A **95% two-sided confidence interval** for  $\mu_Y$  is constructed to contain  $\mu_Y$  in 95% of random samples:

$$90\% \text{ CI: } \bar{Y} \pm 1.64\text{SE}(\bar{Y})$$

$$95\% \text{ CI: } \bar{Y} \pm 1.96\text{SE}(\bar{Y})$$

$$99\% \text{ CI: } \bar{Y} \pm 2.58\text{SE}(\bar{Y})$$

## Test for Difference Between Two Means

- Consider the null hypothesis that two groups  $w$  and  $m$  differ by a quantity  $d_0$ :

$$H_0 : \mu_w - \mu_m = d_0$$

$$H_1 : \mu_w - \mu_m \neq d_0$$

- Example:** Test the hypothesis that the earnings of women and men are no different on average, i.e.,  $d_0 = 0$ :

$$H_0 : \mu_{\text{women}} = \mu_{\text{men}}$$

$$H_1 : \mu_{\text{women}} \neq \mu_{\text{men}}$$

- To test the hypothesis, collect two samples of size  $n_w$  and  $n_m$  and compute the sample means  $\bar{Y}_w$  and  $\bar{Y}_m$ . A natural estimator for this test is  $\bar{Y}_m - \bar{Y}_w$ .

## Sample Distribution of $\bar{Y}_m - \bar{Y}_w$

- Sample Distributions:

$$\bar{Y}_w \sim N\left(\mu_w, \frac{\sigma_w^2}{n_w}\right), \quad \bar{Y}_m \sim N\left(\mu_m, \frac{\sigma_m^2}{n_m}\right)$$

$$\bar{Y}_m \perp \bar{Y}_w$$

- Difference Distribution:

$$\bar{Y}_m - \bar{Y}_w \sim N\left(\mu_m - \mu_w, \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w}\right)$$

## Sample Distribution of $\bar{Y}_m - \bar{Y}_w$ (Cont.)

- **Standard Error:**

$$SE(\bar{Y}_m - \bar{Y}_w) = \sqrt{\frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w}}$$

- ***t*-Statistic for Comparing Two Means:**

$$t = \frac{\bar{Y}_m - \bar{Y}_w - d_0}{SE(\bar{Y}_m - \bar{Y}_w)}$$

# Summary

## 1 The Sample Mean $\bar{Y}$ :

- $\bar{Y}$  is an estimator of the population mean  $\mu_Y$ . If the sample observations  $Y_1, \dots, Y_n$  are i.i.d.:
  - The sampling distribution of  $\bar{Y}$  has mean  $\mu_Y$  and variance  $\sigma_Y^2/n$ .
  - $\bar{Y}$  is **unbiased**.
  - **Law of Large Numbers (LLN)**: As  $n$  increases,  $\bar{Y}$  tends to  $\mu_Y$  —  $\bar{Y}$  is a consistent estimator of  $\mu_Y$ .
  - **Central Limit Theorem (CLT)**: For large  $n$ ,  $\bar{Y}$  has an approximately normal sampling distribution.

## 2 $t$ -Statistic:

- Used to test the null hypothesis that the population mean equals a particular value. For large  $n$ , the  $t$ -statistic follows a standard normal distribution under  $H_0$ .

## Summary (Cont.)

### 3 $p$ -Value:

- The  $t$ -statistic calculates the  $p$ -value associated with  $H_0$ . The  $p$ -value is the probability of observing a  $t$ -statistic as extreme as the one computed, assuming  $H_0$  is true. A small  $p$ -value suggests rejecting  $H_0$ .

### 4 Confidence Intervals:

- A 95% confidence interval for  $\bar{Y}$  contains  $\mu_Y$  in 95% of samples.

### 5 Comparing Means:

- Hypothesis tests and confidence intervals for differences in means of two populations are conceptually similar to those for a single mean.

# Reading

- Stock & Watson (2019) - Chapters 2 and 3