# Principles of Statistical Inference

Rosario Barone

Tor Vergata University of Rome

Statistical tools for decision making

Undergraduate Degree in Global Governance

A.Y. 2023/2024

# Statistical Inference

**Definition**

Statistical inference is the process of drawing conclusions about a population based on a sample from that population.

- It involves using probability theory and statistical techniques to analyze data, estimate parameters, and test hypotheses.
- Statistical inference plays a crucial role in scientific research, decision-making, and data-driven decision-making in various fields.

# Population and Sample

## Population

The population is the entire group or set of individuals, objects, or events of interest in a statistical study. It is usually large and may be infinite.

## Sample

A sample is a subset of the population that is selected and studied to draw inferences about the population as a whole. It is often smaller and more manageable than the entire population.

# Sampling Methods

- Simple Random Sampling: Each member of the population has an equal chance of being selected.
- Stratified Sampling: The population is divided into distinct groups (strata), and samples are taken from each group in proportion to their representation in the population.
- Cluster Sampling: The population is divided into clusters, and a random sample of clusters is selected. All members within the chosen clusters are included in the sample.
- Systematic Sampling: A starting point is selected randomly, and every $n$-th member of the population is selected.

## Point Estimation

Point estimation involves estimating a population parameter using a single value or point estimate based on sample data.

## Interval Estimation

Interval estimation provides a range or interval of values that is likely to contain the population parameter with a certain level of confidence.

## Hypothesis Testing

Hypothesis testing is used to make decisions or draw conclusions about a population based on sample data.

# Estimator

## Estimator

Let $\Theta$ be the parameter space associated with a statistical model with probability density (or mass) function $f(x;\theta)$, and let $x_1, x_2, \ldots, x_n$ be a random sample from the population, where $\theta \in \Theta$. An estimator, denoted by $\hat{\theta}$, is a measurable function of the sample:

$$\hat{\theta} = g(x_1, x_2, \ldots, x_n).$$

- A (point) estimator is a function that associates with each possible sample a value of the parameter $\theta$ to be estimated. It is a function of a sample of data randomly drawn from a population.
- A point estimator is, therefore, a random variable function of the sample, taking values in the parameter space (i.e., in the set of possible values of the parameter).

# Estimate

**Estimate**

An estimate is a specific numerical value calculated from observed data using an estimator. It serves as a point approximation for an unknown parameter in a statistical model.

- The estimate provides a single-value guess or prediction for the true parameter value based on the available sample information.
- An estimate is subject to variability due to the randomness in the sampling process.

## Population Mean ($\mu$)

The population mean, denoted by $\mu$, represents the average value of all observations in a population.

## Sample Mean Estimator ($\bar{X}_n$)

The sample mean estimator, denoted by $\bar{X}_n$, is an estimator that represents the average value of observations in a sample:

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

,where $n$ is the sample size.

- **Population Mean:** $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$, where $N$ is the population size.
- Population mean is a **fixed value** describing the entire population.
- The Sample mean is an **estimate** of the population mean based on a $(X_1, \ldots, X_n)$.

# Asymptotic Theory

- Asymptotic theory provides a powerful framework for understanding the behavior of statistical estimators as the sample size grows ($n \to \infty$).

- Asymptotic theory is based on the principles of limit theorems, such as the law of large numbers and the central limit theorem.

# Convergence of Random Variables

- Convergence in Distribution: A sequence of random variables $X_1, X_2, X_3, \ldots$ converges in distribution to a random variable $X$ if

  $$\lim_{n \to \infty} F_{X_n}(x) = F_X(x) \text{ for all } x \text{ where } F_X \text{ is continuous.}$$

  It is denoted as $X_n \xrightarrow{d} X$

- Convergence in Probability: A sequence of random variables $X_1, X_2, X_3, \ldots$ converges in probability to a random variable $X$ if

  $$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0 \text{ for all } \varepsilon > 0$$

  It is denoted as $X_n \xrightarrow{P} X$.

- Almost Sure Convergence: A sequence of random variables $X_1, X_2, X_3, \ldots$ converges almost surely (or almost everywhere) to a random variable $X$ if

  $$P(\lim_{n \to \infty} X_n = X) = 1.$$

  it is denoted as $X_n \xrightarrow{\text{a.s.}} X$.

# The strong law of large numbers

- As the sample size increases, the sample mean converges **almost surely** to the population mean.
- It provides the foundation for estimating population parameters using sample statistics.

### Theorem

**Strong Law of Large Numbers (SLLN):** *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with mean $\mu$. Then, almost surely,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu.$$

# The weak law of large numbers

- As the sample size increases, the sample mean converges **in probability** to the population mean.

**Theorem**

**Weak Law of Large Numbers (WLLN):** *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with mean $\mu$ and variance $\sigma^2$. Define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ as the sample mean. Then, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0.$$

# The Central Limit Theorem

- The central limit theorem states that as the sample size increases, the distribution of the sample mean approaches a normal distribution.
- It is a fundamental result that allows us to make inferences about population parameters using the properties of the normal distribution.

### Theorem

*Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with mean $\mu$ and variance $\sigma^2$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean, and $\bar{Z}_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ as the standardized sample mean. The central limit theorem states that $\bar{Z}_n$ converges in distribution to the standard normal distribution as $n \to \infty$.*