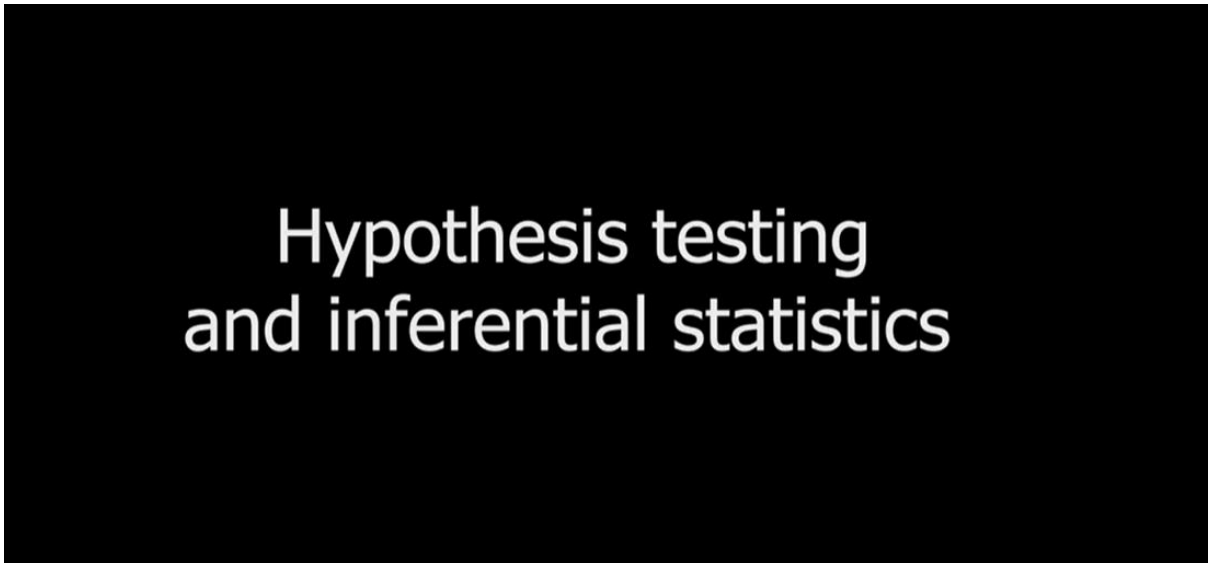


Hypothesis Testing

Definition:

Hypothesis testing is a statistical method used to make decisions using data from a sample. The goal is to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population.

Steps in Hypothesis Testing:

1. **State the Hypotheses:**
    - **Null Hypothesis (H0):** There is no effect or difference, and it represents a statement of "no change" or "no difference."
    - **Alternative Hypothesis (H1 or Ha):** Contradicts the null hypothesis and represents a statement of "there is an effect" or "there is a difference."

Steps in Hypothesis Testing

1. State the Hypotheses
2. Choose a Significance Level (α)
3. Collect Data and Compute a Test Statistic
4. Determine the P-value
5. Make a Decision

Types of Tests

1. Z-test → large sample sizes
2. T-test → smaller
3. ANOVA → 3 or more
4. Chi-square Test → Categorical var.

Scenario:

A basketball coach claims that his team scores more than 50 points on average per game over a season. We want to test this claim based on the scores from 5 randomly selected games.

Hypothesis:

- **Null Hypothesis (H0):** The team's average score per game is 50 points.
- **Alternative Hypothesis (H1):** The team's average score per game is more than 50 points.

Sample Data:

Let's assume the team scored the following points in 5 games:

$53, 47, 52, 55, 49$

Steps in Hypothesis Testing:

1. **Calculate the Sample Mean:**
   $\bar{x} = \frac{(53+47+52+55+49)}{5} = 51.2$

2. **State the Hypotheses:**

   - H0: $\mu = 50$
   - H1: $\mu > 50$

3. **Comparison:**

   - Compare the sample mean (51.2) with the hypothesized mean (50).

4. **Make a Decision:**

   - If the sample mean is greater than the hypothesized mean, we lean towards rejecting the null hypothesis in favor of the alternative. In this simplified example, we're not using a formal test statistic or calculating a p-value.

Decision:

The sample mean of 51.2 is greater than 50. Based on this simple comparison (and ignoring the statistical rigor of error probabilities), we might conclude that the evidence supports the coach's claim that the team scores more than 50 points on average per game.

1. A/B Testing (Marketing and Product Development)

Example:

An e-commerce site may test two different homepage designs (A and B) to see which one leads to higher sales. The null hypothesis (H0) might state that there is no difference in sales between the two designs, while the alternative hypothesis (H1) might state that design B increases sales compared to design A.

2. Medicine and Healthcare

Example:

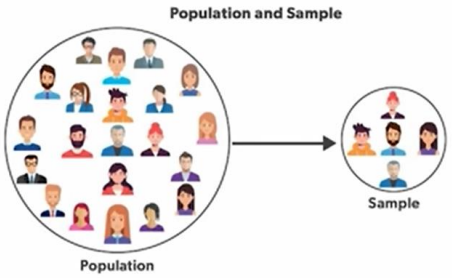
A new drug is developed to lower blood pressure. Researchers conduct a clinical trial where one group receives the new drug and another receives a placebo. The null hypothesis could be that the new drug has no effect on blood pressure, with the alternative being that it does have an effect.

Inferential Statistics

Key Concepts of Inferential Statistics

1. **Population and Sample**:

   - **Population**: The entire group that you want to draw conclusions about.
   - **Sample**: A subset of the population that is used to represent the entire group.



Estimation:

- **Point Estimation**: Involves using sample data to estimate an unknown population parameter (e.g., population mean or proportion).
- **Interval Estimation (Confidence Intervals)**: Instead of a single point estimate, an interval estimation provides a range within which the parameter is expected to lie, with a certain degree of confidence (typically 95% or 99%).

Common Point Estimators:

- **Sample Mean** ($\bar{x}$) as an estimator for the Population Mean (μ).
- **Sample Proportion** ($\hat{p}$) as an estimator for the Population Proportion (p).
- **Sample Variance** (s^2) as an estimator for the Population Variance (σ^2).

Example:

If you have a sample of students' scores from a class test, and you calculate the average score from this sample, that average (sample mean) is a point estimate of the population mean score, assuming the sample is representative of the entire class.

Components of Confidence Intervals:

- **Confidence Level:** Typically expressed as a percentage (e.g., 95% or 99%), the confidence level indicates the probability that the confidence interval computed from a sample contains the true population parameter.
- **Margin of Error:** The range above and below the point estimate that the actual parameter is expected to fall within.

Formula for Confidence Interval:

For a population mean, the confidence interval is typically calculated as:

$\bar{x} \pm z \left(\frac{s}{\sqrt{n}} \right)$

where:

- $\bar{x}$ is the sample mean.
- z is the z-score corresponding to the desired confidence level (e.g., 1.96 for 95% confidence).
- s is the sample standard deviation.
- n is the sample size.