

SNS COLLEGE OF TECHNOLOGY

Kurumbapalayam (Po), Coimbatore – 641 035

An Autonomous Institution

Accredited by NAAC – UGC with ‘A++’ Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

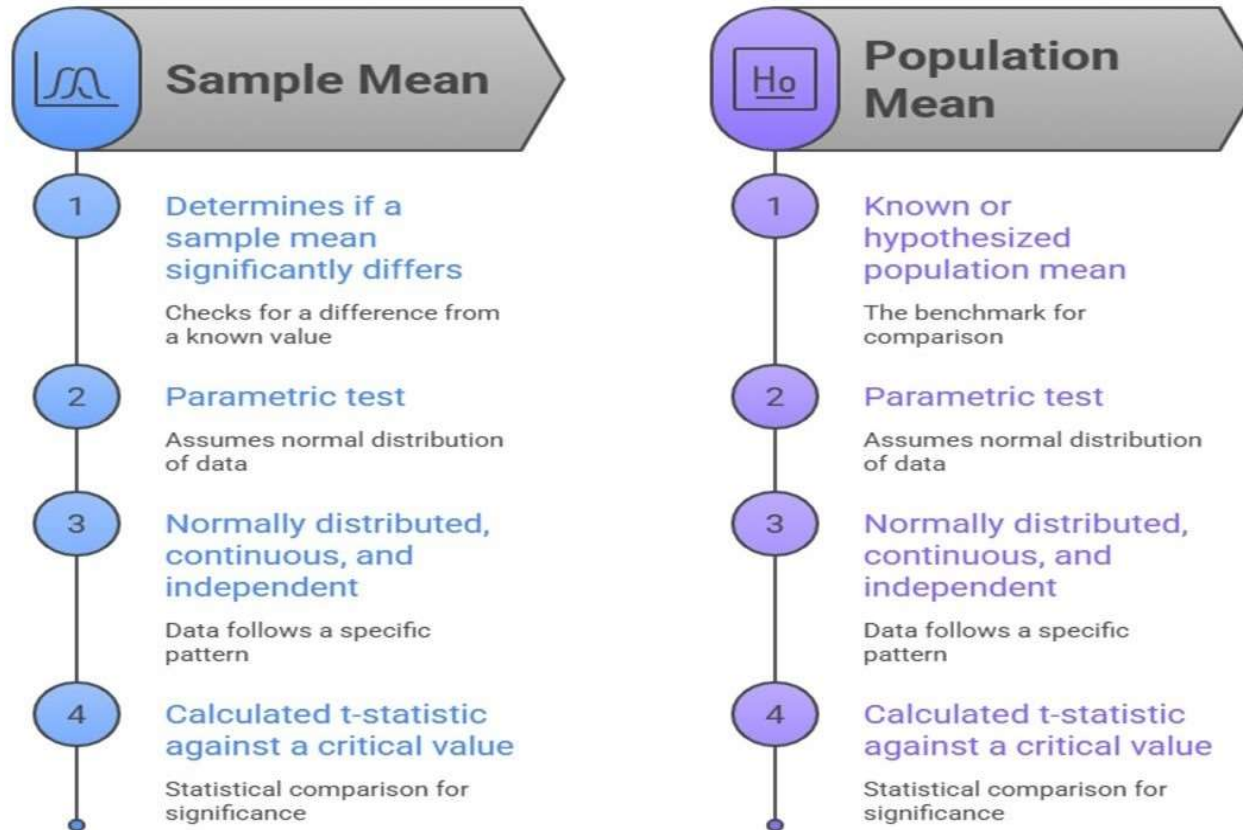
**COURSE NAME: 23ADO201- DATA SCIENCE FUNDAMENTALS
(OPEN ELECTIVE)**

III YEAR / VI SEMESTER

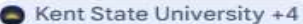
Unit 4 - ANALYSIS OF VARIANCE

Topic : T-test for one sample

Empathy - Sample Mean vs. Hypothesized Population Mean?

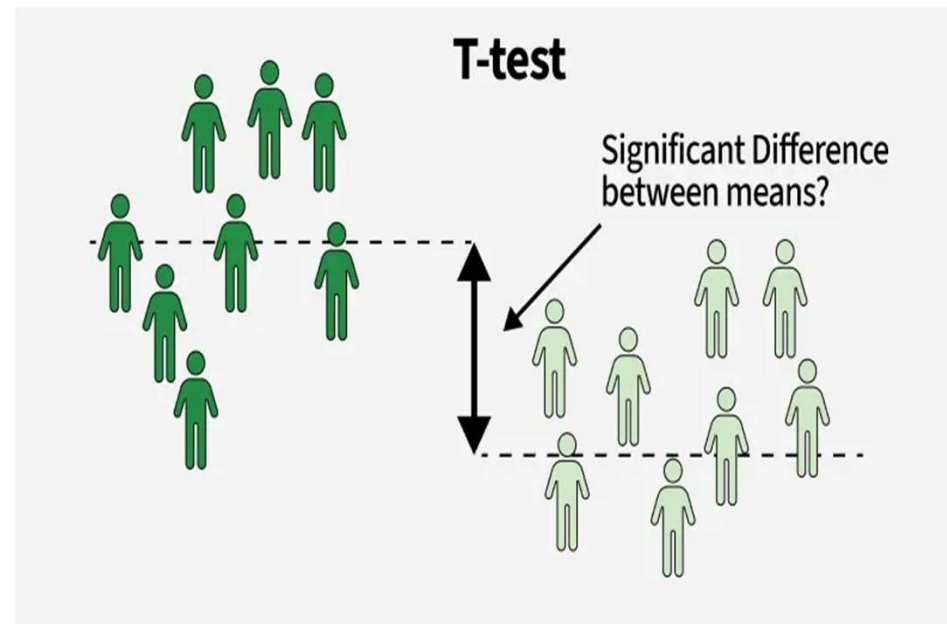


Key Concepts

- **Purpose:** To test if the mean of a sample is significantly different from a specific value.
- **Hypotheses:**
 - **Null Hypothesis (H_0):** Sample mean equals the population mean ($\bar{x} = \mu_0$).
 - **Alternative Hypothesis (H_a):** Sample mean differs from the population mean ($\bar{x} \neq \mu_0$).
- **Formula:** $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
 - \bar{x} : Sample mean
 - μ_0 : Hypothesized population mean
 - s : Sample standard deviation
 - n : Sample size.
- **Assumptions:** The data should be roughly normally distributed and continuous, without significant outliers. 

The t-test is used to compare the averages of two groups to see if they are significantly different from each other. Suppose you want to compare the test scores of two groups of students:

- Group 1:** 30 students who studied with Method A.
- Group 2:** 30 students who studied with Method B.



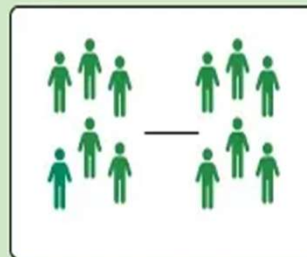
Types of t test

One sample t-test



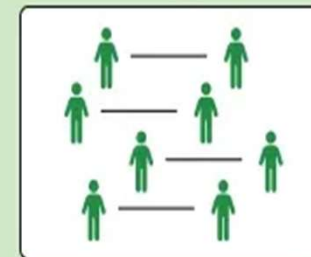
Is there a difference between a group and the population.

Independent samples t-test



Is there a difference between two groups

Paired samples t-test



Is there a difference in a group between two points in time

1. One sample T-test

One sample t-test is used for comparison of the sample mean of the data to a particularly given value. We can use this when the sample size is small. (under 30) data is collected randomly and it is approximately normally distributed. It can be calculated as:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Mean of the sample

Reference Value

Standard deviation

Number of cases

Example: The weights of 25 obese people were taken before enrolling them into the nutrition camp. The population mean weight is found to be 45 kg before starting the camp. After finishing the camp for the same 25 people the sample mean was found to be 75 with a standard deviation of 25. Did the fitness camp work?

Python Implementation

Before its implementation we should have some basic knowledge about [numpy](#) and [scipy](#).

```
import scipy.stats as stats
import numpy as np

popu_mean = 45
s_mean = 75
s_std = 25
s_size = 25

t_stat = (s_mean - popu_mean) / (s_std / np.sqrt(s_size))
df = s_size - 1
alpha = 0.05
cr_t = stats.t.ppf(1 - alpha, df)

p_v = 1 - stats.t.cdf(t_stat, df)

print("T-Statistic:", t_stat)
print("Critical t-value:", cr_t)
print("P-Value:", p_v)

print('With T-value :')
if t_stat > cr_t:
    print("Significant difference. Camp had effect.")
else:
    print("No significant difference. Camp had no effect.")

print('With P-value :')
if p_v > alpha:
    print("Significant difference. Camp had effect.")
else:
    print("No significant difference. Camp had no effect.")
```

Output:

T-Statistic: 6.0

Critical t-value: 1.7108820799094275

P-Value: 1.703654035845048e-06

With T-value :

Significant difference. Camp had effect.

With P-value :

No significant difference. Camp had no effect.

2. Independent sample T-test

An Independent sample t-test commonly known as an unpaired sample t-test is used to find out if the differences found between two groups is actually significant or just a random occurrence. We can use this when:

- the population mean or standard deviation is unknown. (information about the population is unknown)
- the two samples are separate/independent. For i.e. boys and girls (the two are independent of each other)

It can be calculated using:

The diagram shows the formula for an independent sample t-test with green arrows pointing to specific parts of the equation:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Annotations in the diagram:

- Mean of the sample 1**: points to \bar{X}_1
- Mean of the sample 2**: points to \bar{X}_2
- Standard deviation sample 1 and 2**: points to S_1^2 and S_2^2
- Number of cases sample 1 and 2**: points to n_1 and n_2

3. Paired Two-sample T-test

Paired sample t-test also known as dependent sample t-test is used to find out if the difference in the mean of two samples is 0. The test is done on dependent samples usually focusing on a particular group of people or things. In this each entity is measured twice resulting in a pair of observations.

We can use this when:

- Two similar samples are given. [i.e. Scores obtained in English and Math (both subjects)]
- The dependent variable data is continuous.
- The observations are independent of one another.
- The dependent variable is approximately normally distributed.

Mean of the difference

$$t = \frac{\bar{X}_d - 0}{\frac{s}{\sqrt{n}}}$$

Standard deviation

Number of cases

Activity



Activity: One-Sample T-Test (Real-Life Example)

Objective

To test whether the average value of a sample differs from a known population mean.

Scenario

A teacher claims that the **average marks** of students in a class is **75**.

You collect marks of **10 students**:

Sample Data:

72, 78, 74, 70, 76, 80, 68, 77, 73, 75

Step 1: State Hypotheses

•Null Hypothesis (H_0): $\mu = 75$ (no difference)

•Alternative Hypothesis (H_1): $\mu \neq 75$ (there is a difference)

Step 2: Use the T-Test Formula

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Where:

- \bar{x} = sample mean
- μ = population mean
- s = sample standard deviation
- n = sample size

Step 3: Calculate Values

- Sample mean (\bar{x}) = 74.3
- Population mean (μ) = 75
- Standard deviation (s) \approx 3.7
- Sample size (n) = 10

Now calculate:

$$t = (74.3 - 75) / (3.7 / \sqrt{10})$$

$$t \approx -0.60$$

Step 4: Decision Rule

- Degrees of freedom (df) = $n - 1 = 9$
- At 5% significance level, critical $t \approx \pm 2.262$

Step 5: Conclusion

- Calculated $t = -0.60$
- Since **-0.60 lies within -2.262 and +2.262**,

Fail to reject H_0

MCQ



One-Sample T-Test – MCQs

1. What is the main purpose of a one-sample t-test?

- A) Compare two sample means
- B) Compare sample mean with population mean
- C) Compare variances
- D) Test correlation

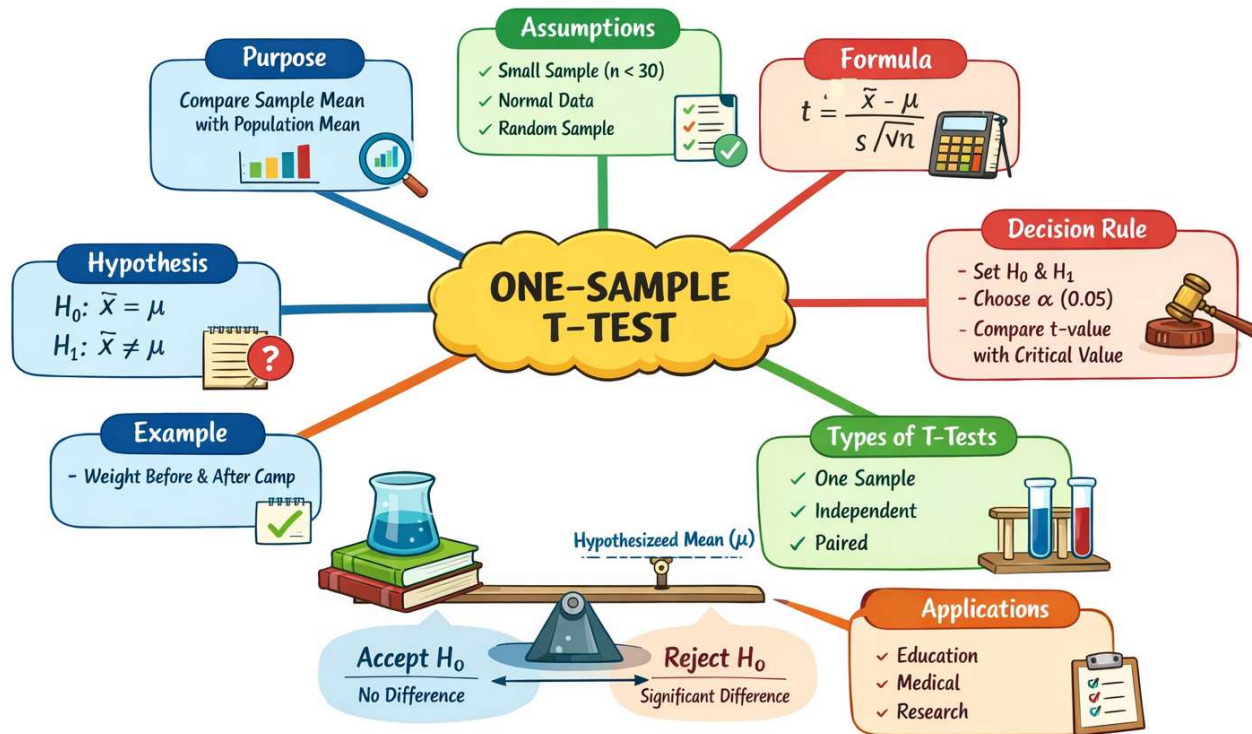
Answer: B

2. When do we use a one-sample t-test?

- A) When population standard deviation is known
- B) When sample size is very large
- C) When population standard deviation is unknown
- D) When data is categorical

Answer: C

MINDMAP



TEXT BOOKS	
1.	Srinivasan Desikan and Gopaldaswamy Ramesh, —Software Testing – Principles and Practices, Pearson Education, 2006.
REFERENCES	
1.	Ilene Burnstein, —Practical Software Testing, Springer International Edition, 2003
2.	Edward Kit, Software Testing in the Real World – Improving the Process, Pearson Education, 1995
3.	Boris Beizer, Software Testing Techniques – 2nd Edition, Van Nostrand Reinhold, New York, 1990.
4.	Aditya P. Mathur, —Foundations of Software Testing _ Fundamental Algorithms and Techniques, Dorling Kindersley (India) Pvt. Ltd., Pearson Education, 2008.

THANK YOU!

15