SNS COLLEGE OF TECHNOLOGY

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade

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

Department of Computer Science and Engineering

Course Code & Title : 23ADO201 - Data Science Fundamentals

III YEAR / VI SEMESTER - ECE

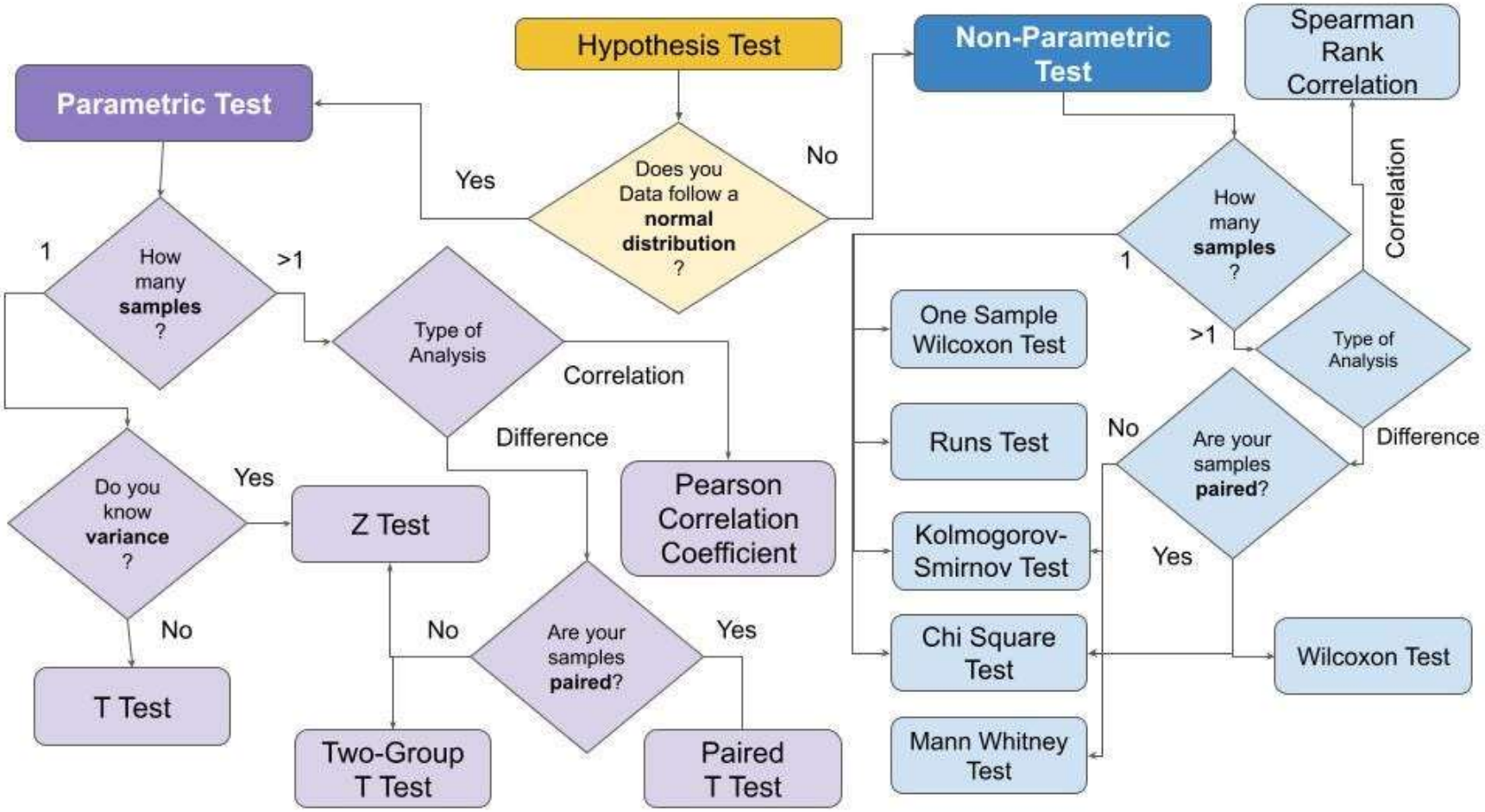
Unit 3 - INFERENTIAL STATISTICS

Topic : Hypothesis testing , z-test and z-test procedure

K.KARTHIKEYAN AP /CSE

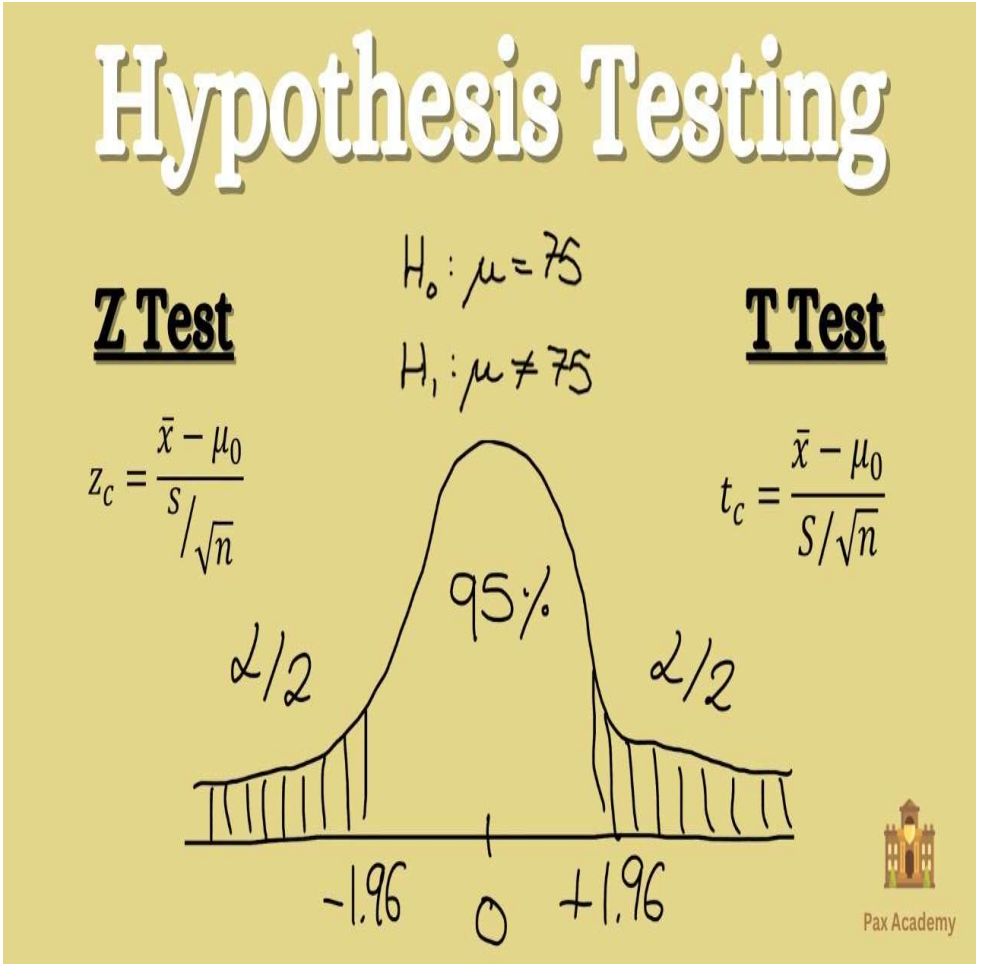
Empathy

www.snsgroups.com



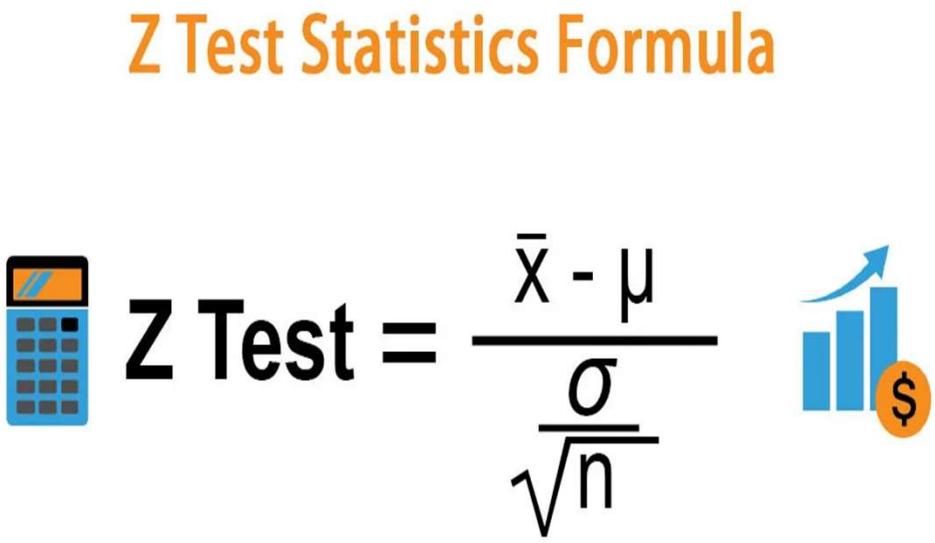
Hypothesis testing

- Hypothesis testing is a statistical method for making inferences about a population parameter using sample data, typically by comparing a calculated test statistic to a critical value to accept or reject a null hypothesis (H0)

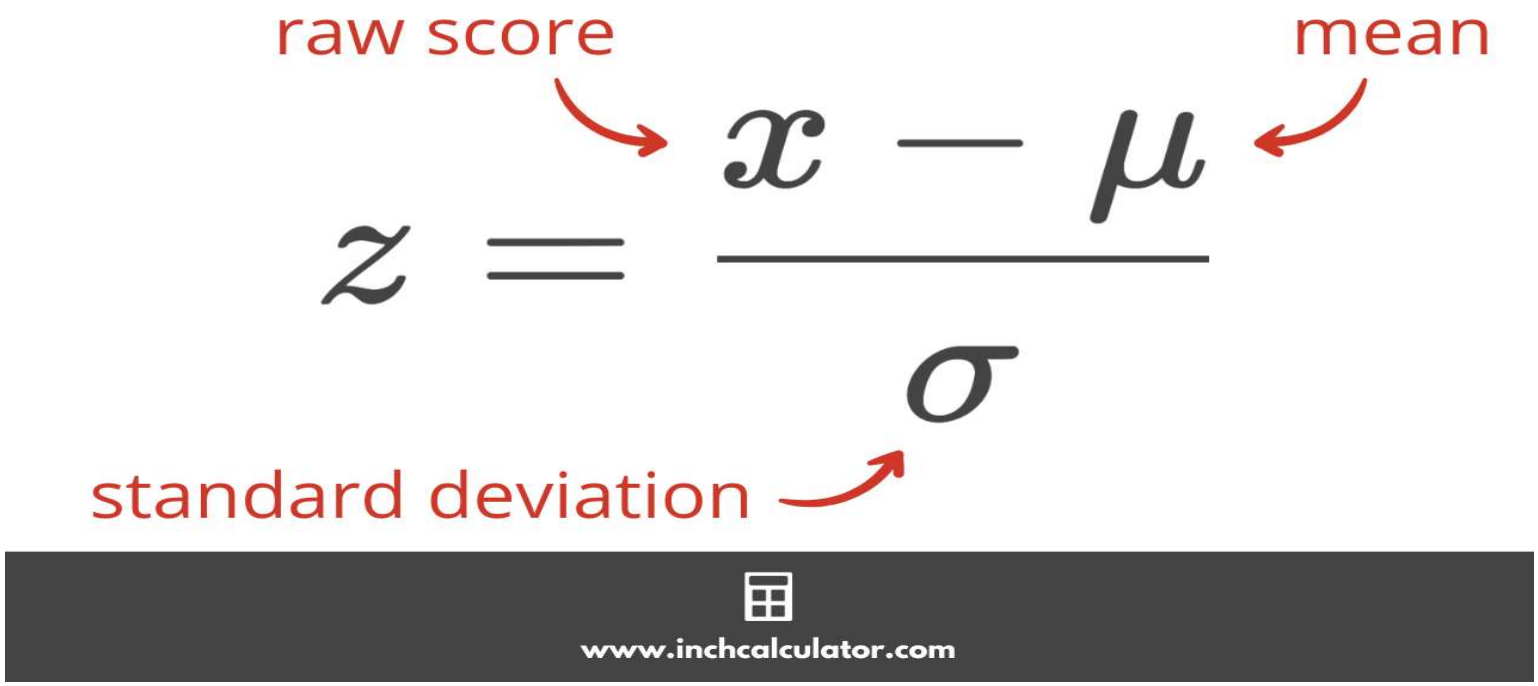


What is a Z-Test?

A z-test is an inferential tool that determines if the difference between a sample mean and the population mean is statistically significant. It operates under the assumption that the data follows a normal distribution and the population variance is known z-test



Z-Score Formula



where:

x : mean of the sample.

μ : mean of the population.

σ : Standard deviation of the population

Problem -1

Let's understand with the help of example The average family annual income in India is 200k with a standard deviation of 5k and the average family annual income in Delhi is 300k. Then Z-Score for Delhi will be.

$$\begin{aligned}\text{Z-Score} &= \frac{\bar{x} - \mu}{\sigma} \\ &= \frac{300 - 200}{5} \\ &= 20\end{aligned}$$

Type of Z-test

There are mainly two types of Z-tests. Let's understand them one by one:

1. One-Sample Z test

A one-sample Z-test is used to determine if the mean of a single sample is significantly different from a known population mean. Use it when:

- The population standard deviation is known.
- The sample size is large (usually $n>30$).
- The data is approximately normally distributed.

One-Sample z-test for p

State Hypothesis

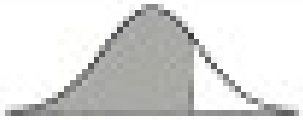
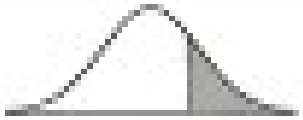
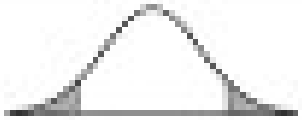
$H_0: p = P$ $H_a: p < P$	$H_0: p = P$ $H_a: p > P$	$H_0: p = P$ $H_a: p \neq P$

Find Standardized Test Statistic

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{(P)(1-P)}{n}}}$$

Z = standardized test statistic
$\hat{p}$ = sample proportion
P = H_0 proportion

Find the p-value

		
$H_a: p < P$	$H_a: p > P$	$H_a: p \neq P$
On z-table $P(x < z)$	On z-table $P(x > z)$	On z-table $2 \times P(x > \lvert z \rvert)$

Conclusion

p-value< α	Reject H_0
p-value> α	Fail to Reject H_0

Suppose a company claims that their new smartphone has an average battery life of 12 hours. A consumer group tests 100 phones and finds an average battery life of 11.8 hours with a known population standard deviation of 0.5 hours.

Step 1: Hypotheses

- $H_0: \mu = 12$:
- $H_1: \mu \neq 12$

Step 2: Calculate the Z-Score

We can calculate Z-score using the formula:

$$z = \frac{x - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Where: $\bar{x} = 11.8$, $\mu = 12$,$\sigma = 0.5$ and $n = 100$

After putting the value we get:

$$z = \frac{11.8 - 12}{\frac{0.5}{\sqrt{100}}} = -4$$

Step 3: Decision

Since $|Z| = 4 > 1.96$) critical value for$\alpha = 0.05$) we reject H_0 indicate significant evidence against the company's claim.

```
import numpy as np
from statsmodels.stats.weightstats import ztest

data = [11.8] * 100
population_mean = 12
population_std_dev = 0.5

z_statistic, p_value = ztest(data, value=population_mean)

print(f"Z-Statistic: {z_statistic:.4f}")
print(f"P-Value: {p_value:.4f}")

alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: The average battery
life is different from 12 hours.")
else:
    print("Fail to reject the null hypothesis: The average
battery life is not significantly different from 12 hours.")
```

Output:

Z-Statistic: -5601281313739 70.2500
P-Value: 0.0000
Reject the null hypothesis: The average battery life is different from 12 hour

Activity

Activity 1: Hypothesis Testing Concept Activity

Materials Needed

•Board / Projector

•Calculator

•Sample data sheet

Scenario

A college claims that the **average score of students in statistics is 70 marks**. A teacher collects scores from **36 students** and finds the **sample mean is 74** with a **standard deviation of 12**.

Student Task

Students must determine whether the college claim is correct using **hypothesis testing**.

Steps for Students

1.State the Hypotheses

•H_0 : $\mu = 70$ (Average mark is 70)

•H_1 : $\mu \neq 70$ (Average mark is not 70)

2.Identify Given Values

•Sample mean ($\bar{x}$) = 74

•Population mean (μ) = 70

•Standard deviation (σ) = 12

•Sample size (n) = 36

MCQ

MCQ on Hypothesis Testing and Z-Test

1. Hypothesis testing is used to

A. Collect data

B. Analyze population parameters using sample data

C. Draw graphs

D. Store data

Answer: B

2. The statement that assumes no difference or no effect is called

A. Alternative hypothesis

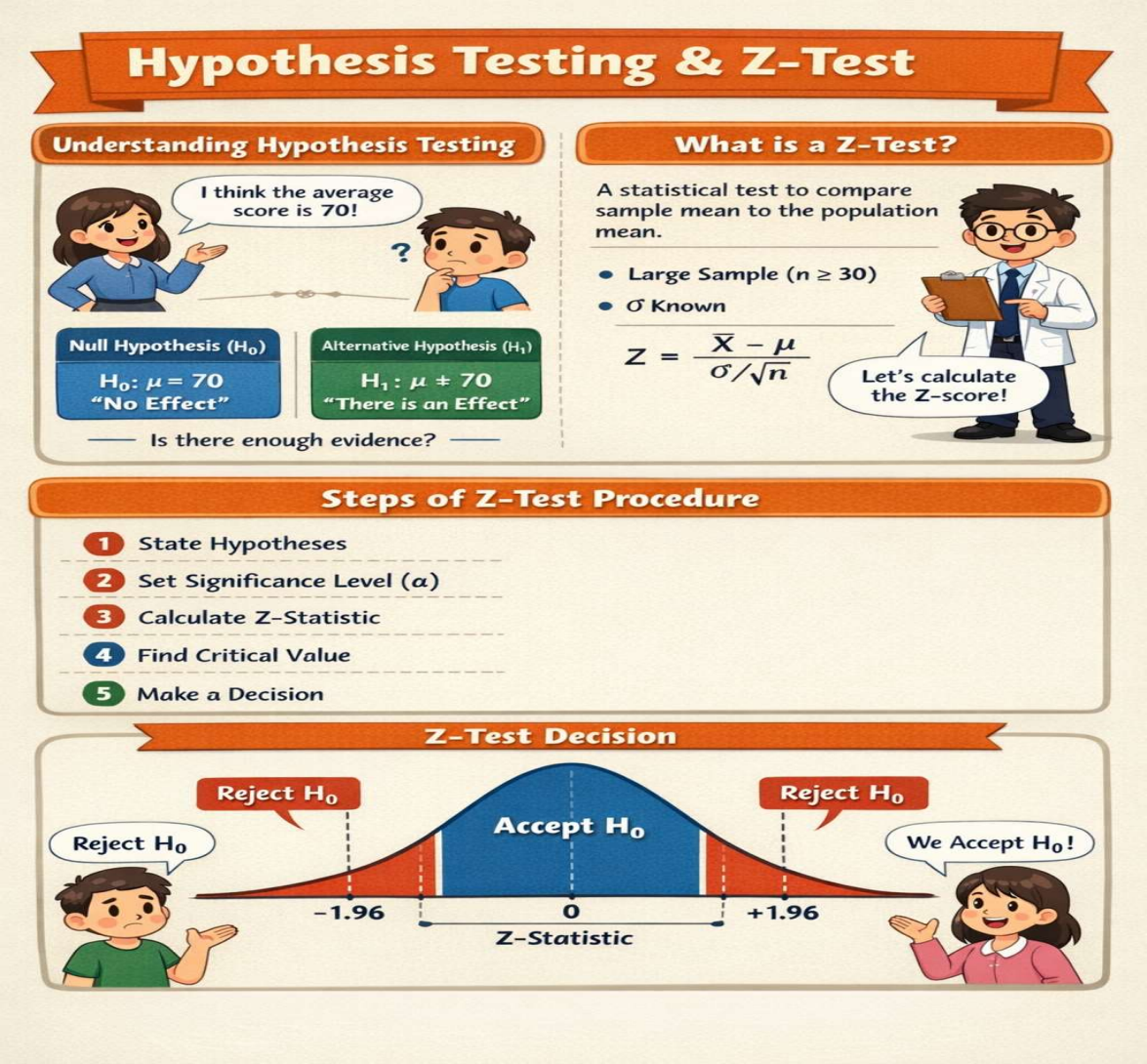
B. Research hypothesis

C. Null hypothesis

D. Statistical hypothesis

Answer: C

MIND MAP


Hypothesis Testing & Z-Test
Understanding Hypothesis Testing
I think the average score is 70!
?
Null Hypothesis (H_0)
$H_0: \mu = 70$
"No Effect"
Alternative Hypothesis (H_1)
$H_1: \mu \neq 70$
"There is an Effect"
Is there enough evidence?
What is a Z-Test?
A statistical test to compare sample mean to the population mean.
Large Sample ($n \geq 30$)
σ Known
$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$
Let's calculate the Z-score!
Steps of Z-Test Procedure
1 State Hypotheses
2 Set Significance Level (α)
3 Calculate Z-Statistic
4 Find Critical Value
5 Make a Decision
Z-Test Decision
Reject H_0
Reject H_0
Accept H_0
Reject H_0
We Accept H_0!
-1.96
0
+1.96
Z-Statistic

REFERENCE BOOKS

1.Allen B. Downey, “Think Stats: Exploratory Data Analysis in Python”, Green Tea Press, 2014.
2.Sanjeev J. Wagh, Manisha S. Bhende, Anuradha D. Thakare, “Fundamentals of Data Science”, CRC Press, 2022.
3.Chirag Shah, “A Hands-On Introduction to Data Science”, Cambridge University Press, 2020.
4.Vineet Raina, Srinath Krishnamurthy, “Building an Effective Data Science Practice: A Framework to Bootstrap and Manage a Successful Data Science Practice”, A press, 2021.

www.snsgroups.com