

## Mini Project: Descriptive and Inferential Statistics

### Problem Statement:

This project is designed to develop foundational understanding in **Descriptive and Inferential Statistics**. The assignment covers key concepts such as **Measures of Central Tendency** (Mean, Median, Mode), **Measures of Dispersion** (Range, Variance, Standard Deviation), and **key inferential methods** like **Confidence Intervals** and **Hypothesis Testing**. By completing this project, students will gain both theoretical knowledge and practical skills in interpreting and analyzing data.

### Guidelines for Students:

#### 1. Foundational Knowledge:

- Understand the key statistical definitions, data types, and the concept of population vs sample.
- Learn how to calculate measures of central tendency and dispersion.
- Understand how confidence intervals and hypothesis testing are used to infer characteristics about a population from sample data.

#### 2. Hands-on Learning:

- Manually calculate measures such as mean, median, mode, variance, and standard deviation.
- Interpret statistical results, including the concept of confidence intervals and hypothesis testing.

#### 3. Model Evaluation:

- Practice interpreting the results of statistical tests and applying them to make data-driven decisions.

## Step-by-Step Project Outline:

### Q1. Key Statistical Definitions

**Objective:** Understand foundational statistical terms.

**Problem:**

Write short definitions (2-3 lines each) for the following:

- a) Population and Sample
- b) Descriptive Statistics and Inferential Statistics
- c) Parameter and Statistic
- d) Qualitative and Quantitative Data

### Q2. Measures of Central Tendency - Definitions

**Objective:** Learn basic concepts of data centering.

**Problem:**

Define the following terms with one example each:

- a) Mean
- b) Median
- c) Mode

### Q3. Manual Calculation of Mean, Median, and Mode

**Objective:** Apply manual formulas to real data.

**Problem:**

Given the dataset:

12, 18, 14, 16, 18, 20, 18, 15, 12, 18, 14, 16, 18, 20, 18, 15

Calculate:

- a) Mean

- b) Median
- c) Mode

#### **Q4. Levels of Measurement**

**Objective:** Understand classification of data types.

**Problem:**

Define and give one example for each level of measurement:

- a) Nominal
- b) Ordinal
- c) Interval
- d) Ratio

#### **Q5. Variance and Standard Deviation - Theory**

**Objective:** Understand spread/variability in data.

**Problem:**

- a) Define Variance and Standard Deviation.
- b) Explain why Standard Deviation is more interpretable than Variance.

#### **Q6. Manual Calculation - Variance and Standard Deviation**

**Objective:** Practice computing data spread.

**Problem:**

Given the data:

8, 10, 12, 14, 16, 10, 12, 14, 16

Calculate:

- a) Sample Variance
- b) Sample Standard Deviation

### Q7. Range and Interquartile Range (IQR)

**Objective:** Use position-based dispersion metrics.

**Problem:**

Given the dataset:

22, 29, 25, 31, 35, 40, 45, 48, 50

- a) Arrange data in ascending order
- b) Calculate the Range
- c) Find Q1, Q3, and IQR

### Q8. Five-number Summary and Boxplot Concept

**Objective:** Summarize distribution of data.

**Problem:**

- Define the Five-number Summary and explain each component:
  - Minimum
  - Q1 (First Quartile)
  - Median
  - Q3 (Third Quartile)
  - Maximum
- Describe how boxplots help in detecting outliers.

### Q9. Confidence Interval for the Mean

**Objective:** Estimate population means using sample data.

**Problem:**

A sample of 36 students has an average height of 162 cm with a standard deviation of 6 cm. Calculate the 95% Confidence Interval for the population mean.  
(Hint: Use  $Z = 1.96$  for 95% confidence)

## **Q10. Hypothesis Testing - One Sample Z-Test**

**Objective:** Make decisions using statistical testing.

**Problem:**

The average salary in a city is ₹30,000. A random sample of 49 employees has an average salary of ₹31,000 with a standard deviation of ₹4,900.

Test the hypothesis at the 5% level of significance to determine if the average salary has increased.

- a) State the null and alternative hypothesis:
- b) Calculate the Z-score:
- c) Conclude the result using critical value ( $\pm 1.96$ )

## **Dataset for the Project:**

You can use synthetic datasets or refer to the following sources for real-world data:

- **Kaggle Datasets:**
  - [Kaggle Datasets](#)

## **Expected Outcomes:**

- Students will gain proficiency in computing and interpreting key statistical measures.
- They will develop skills to perform confidence intervals and hypothesis testing.
- This project will allow students to apply descriptive and inferential statistical methods to real-world datasets.