**Descriptive statistics refers to a set of statistical methods used to summarize and present data in a clear and understandable form**

**1. Measures of Central Tendency**

**📌 What?**  
These are numbers that represent the center or average of a dataset.

**Types:**

* **Mean (Average)**
* **Median (Middle value)**
* **Mode (Most frequent value)**

**📆 When to Use?**  
When you want a single value that describes the typical or average data point in your dataset.

**🎯 Why?**  
It helps in summarizing the data with one representative number, especially useful in comparison or quick insights.

# Sample data

data <- c(10, 20, 20, 30, 40, 50)

# Mean

mean\_value <- mean(data)

# Median

median\_value <- median(data)

# Mode (custom function as R has no built-in mode)

get\_mode <- function(v) {

uniqv <- unique(v)

uniqv[which.max(tabulate(match(v, uniqv)))]

}

mode\_value <- get\_mode(data)

cat("Mean:", mean\_value, "\n")

cat("Median:", median\_value, "\n")

cat("Mode:", mode\_value, "\n")

**Measures of Dispersion**

**📌 What?**  
These tell you **how spread out** the data values are.

**Types:**

* **Range (Max - Min)**
* **Variance (how far each number is from the mean)**
* **Standard Deviation (square root of variance)**

**📆 When to Use?**  
When you want to know how consistent or variable your dataset is.

**🎯 Why?**  
Averages alone can be misleading. Dispersion shows whether data points are close to the mean or scattered widely.

# Sample data

data <- c(10, 20, 20, 30, 40, 50)

# Range

range\_value <- range(data)

range\_diff <- diff(range\_value)

# Variance

variance\_value <- var(data)

# Standard Deviation

sd\_value <- sd(data)

cat("Range:", range\_value, " | Difference:", range\_diff, "\n")

cat("Variance:", variance\_value, "\n")

cat("Standard Deviation:", sd\_value, "\n")

## ****3. Frequency Distributions****

**📌 What?**  
Shows how often each value (or range of values) appears in the dataset.

**📆 When to Use?**  
When you want to see the pattern of repetition, or group data into intervals (like histograms).

**🎯 Why?**  
Great for visualizing trends, outliers, and understanding where values are concentrated.

# Sample data

data <- c(10, 20, 20, 30, 30, 30, 40, 50)

# Frequency Table

freq\_table <- table(data)

print(freq\_table)

# Plotting Histogram

hist(data, main = "Histogram of Data", xlab = "Values", col = "lightblue", border = "black")

**. Correlations**

**📌 What?**  
Measures the relationship between two variables (e.g., height & weight).

**Types:**

* **Positive Correlation:** both increase together
* **Negative Correlation:** one increases, other decreases
* **No Correlation**

**📆 When to Use?**  
When analyzing the strength and direction of relationships between variables.

**🎯 Why?**  
Useful for predicting values or finding linked behaviors.

# Sample data

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

# Correlation Coefficient (Pearson)

correlation <- cor(x, y, method = "pearson")

cat("Correlation Coefficient:", correlation, "\n")

# Scatter plot

plot(x, y, main = "Scatter Plot", xlab = "X", ylab = "Y", col = "blue", pch = 19)

abline(lm(y ~ x), col = "red") # Add regression line

| **Concept** | **R Function(s)** | **Output Example** |
| --- | --- | --- |
| **Mean** | mean(data) | 28.3 |
| **Median** | median(data) | 25 |
| **Mode** | Custom get\_mode() | 30 |
| **Range** | range(data) / diff() | 10 50 / 40 |
| **Variance** | var(data) | 200 |
| **SD** | sd(data) | 14.14 |
| **Frequency** | table(data) | 10:1, 20:2, etc. |
| **Correlation** | cor(x, y) | 1.0 (perfect positive) |

Csv file data

StudentID,Math,Science,English

1,78,85,88

2,62,70,72

3,90,95,92

4,55,60,65

5,80,82,85

6,85,88,90

# Load CSV File

data <- read.csv("student\_scores.csv")

# View dataset

print("Raw Data:")

print(data)

# 1. Central Tendency Measures

cat("\n--- Measures of Central Tendency ---\n")

cat("Mean (Math):", mean(data$Math), "\n")

cat("Median (Math):", median(data$Math), "\n")

get\_mode <- function(v) {

uniqv <- unique(v)

uniqv[which.max(tabulate(match(v, uniqv)))]

}

cat("Mode (Math):", get\_mode(data$Math), "\n")

# 2. Measures of Dispersion

cat("\n--- Measures of Dispersion ---\n")

cat("Range (Math):", range(data$Math), " | Difference:", diff(range(data$Math)), "\n")

cat("Variance (Math):", var(data$Math), "\n")

cat("Standard Deviation (Math):", sd(data$Math), "\n")

# 3. Frequency Distribution (for Math scores)

cat("\n--- Frequency Distribution ---\n")

print(table(data$Math))

hist(data$Math, main="Math Scores Histogram", xlab="Scores", col="lightgreen", border="black")

# 4. Correlation (Math vs Science)

cat("\n--- Correlation (Math vs Science) ---\n")

correlation <- cor(data$Math, data$Science)

cat("Correlation Coefficient:", correlation, "\n")

plot(data$Math, data$Science, main="Math vs Science",

xlab="Math Scores", ylab="Science Scores", pch=19, col="blue")

abline(lm(data$Science ~ data$Math), col="red")