Experiment No.5

Q1. You have a dataset of individuals with information on their age, height, weight, and BMI

(Body Mass Index). Analyze the data to:

o Calculate the mean, median, and standard deviation of each variable.

o Determine if there is a correlation between height and weight.

o Visualize the distribution of age using a histogram and check for any outliers

using a boxplot.

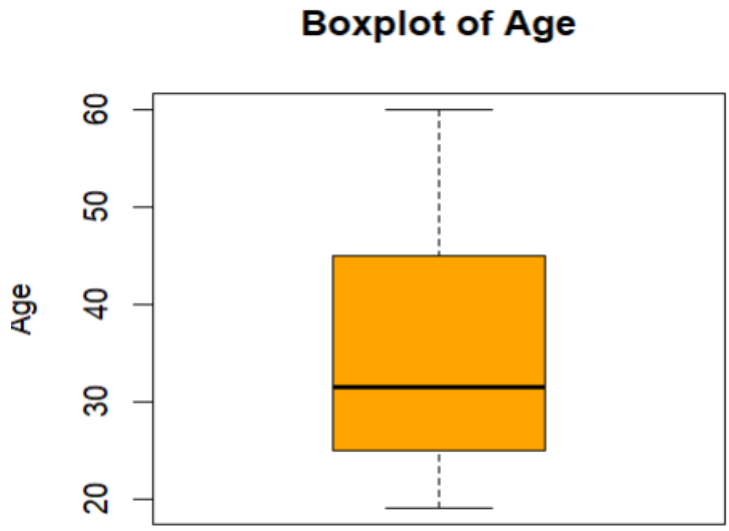
Program:-

```
data <- data.frame(
  age = c(25, 30, 22, 45, 55, 33, 41, 19, 60, 28),
  height = c(170, 165, 160, 175, 180, 168, 172, 158, 177, 169),
  weight = c(70, 60, 55, 80, 90, 65, 75, 50, 85, 62),
  bmi = c(24.2, 22.0, 21.5, 26.1, 27.8, 23.0, 25.4, 20.0, 27.1, 21.7)
)
# 1. Calculate mean, median, and standard deviation of each variable
mean_age <- mean(data$age)
median_age <- median(data$age)
sd_age <- sd(data$age)
mean_height <- mean(data$height)
median_height <- median(data$height)
sd_height <- sd(data$height)
mean_weight <- mean(data$weight)
median_weight <- median(data$weight)
sd_weight <- sd(data$weight)
mean_bmi <- mean(data$bmi)
median_bmi <- median(data$bmi)
sd_bmi <- sd(data$bmi)
```

```
# Print results
cat("Age - Mean:", mean_age, "Median:", median_age, "SD:", sd_age, "\n")
cat("Height - Mean:", mean_height, "Median:", median_height, "SD:", sd_height, "\n")
cat("Weight - Mean:", mean_weight, "Median:", median_weight, "SD:", sd_weight, "\n")
cat("BMI - Mean:", mean_bmi, "Median:", median_bmi, "SD:", sd_bmi, "\n")
# 2. Correlation between height and weight
correlation <- cor(data$height, data$weight)
cat("Correlation between height and weight:", correlation, "\n")
# 3. Visualize age distribution with a histogram
hist(data$age, main = "Age Distribution", xlab = "Age", col = "skyblue", border = "black")
boxplot(data$age, main = "Boxplot of Age", ylab = "Age", col = "orange")
```

Output :-

```
# Print results
Age - Mean: 35.8 Median: 31.5 SD: 13.97458
Height - Mean: 169.4 Median: 169.5 SD: 7.058486
Weight - Mean: 69.2 Median: 67.5 SD: 13.15548
BMI - Mean: 23.88 Median: 23.6 SD: 2.642726
# 2. Correlation between height and weight
Correlation between height and weight: 0.9790361
 # 3. Visualize age distribution with a histogram
```



Q2. A company provides sales data for its products, including variables like product price,

units sold, and profit. Your task is to:

o Calculate the average sales, median price, and total profit.

o Analyze the variability in sales across products using standard deviation.

o Visualize the sales and profit data using scatter plots and draw conclusions about

the relationship between price and profit.

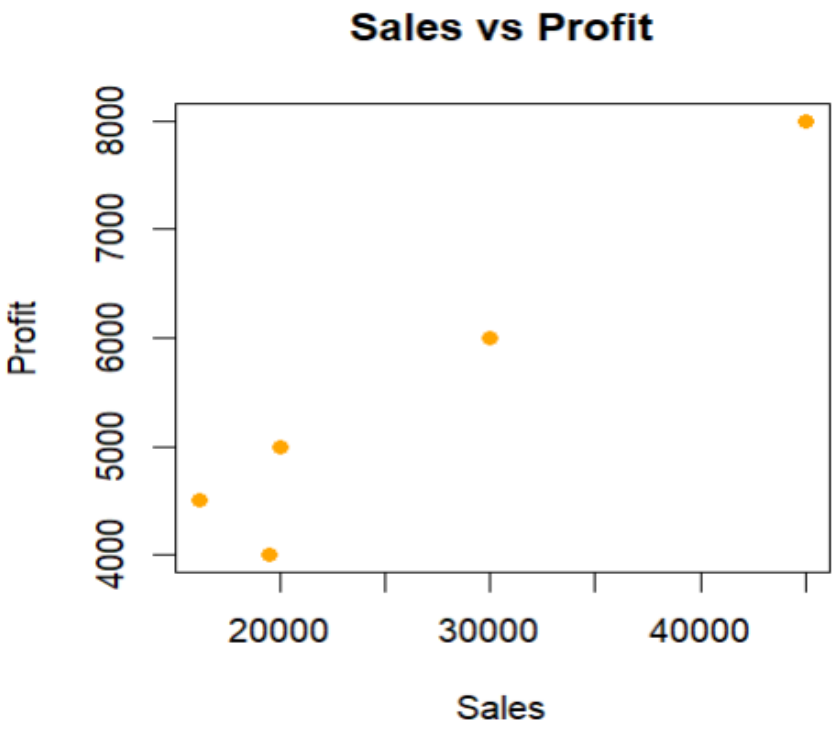
Program:-

```
sales_data <- data.frame(
  product = c("A", "B", "C", "D", "E"),
  price = c(100, 150, 120, 130, 90),
  units_sold = c(200, 300, 250, 150, 180),
  profit = c(5000, 8000, 6000, 4000, 4500)
)
# 1. Calculate average sales, median price, and total profit
sales_data$sales <- sales_data$price * sales_data$units_sold  # Calculate total sales for each product
average_sales <- mean(sales_data$sales)  # Average sales
median_price <- median(sales_data$price)  # Median product price
total_profit <- sum(sales_data$profit)  # Total profit
# Print results
cat("Average Sales:", average_sales, "\n")
cat("Median Price:", median_price, "\n")
cat("Total Profit:", total_profit, "\n")
# 2. Analyze variability in sales using standard deviation
sales_sd <- sd(sales_data$sales)
cat("Standard Deviation of Sales:", sales_sd, "\n")
# 3. Visualize sales and profit data using scatter plots
plot(sales_data$price, sales_data$profit,
     main = "Price vs Profit",
     xlab = "Price",
```

```
    ylab = "Profit",
    pch = 19, col = "blue")
plot(sales_data$sales, sales_data$profit,
    main = "Sales vs Profit",
    xlab = "Sales",
    ylab = "Profit",
    pch = 19, col = "orange")
```

Output:-

```
# Print results
Average Sales: 26140
Median Price: 120
Total Profit: 27500
 # 2. Analyze variability in sales using standard deviation
Standard Deviation of Sales: 11738.74
 # 3. Visualize sales and profit data using scatter plots
```



Q3. You have a dataset containing the exam scores of students in three subjects: Mathematics, Science, and English. Analyze the scores to:

o Compute the mean, median, and mode for each subject.

o Identify the subject with the highest variability in scores (using standard deviation).

o Create a boxplot for each subject and analyze the spread and outliers.

o Visualize the relationship between Mathematics and Science scores using a scatter plot and determine if there is a trend.

Program:-

```
scores <- data.frame(
  student = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
  mathematics = c(90, 85, 78, 92, 88, 84, 76, 91, 89, 77),
  science = c(88, 79, 85, 95, 91, 80, 83, 92, 87, 78),
  english = c(85, 82, 90, 86, 80, 88, 84, 87, 91, 79)
)
# 1. Compute the mean, median, and mode for each subject
mean_math <- mean(scores$mathematics)
median_math <- median(scores$mathematics)
mode_math <- as.numeric(names(sort(table(scores$mathematics), decreasing = TRUE)[1]))
mean_sci <- mean(scores$science)
median_sci <- median(scores$science)
mode_sci <- as.numeric(names(sort(table(scores$science), decreasing = TRUE)[1]))
mean_eng <- mean(scores$english)
median_eng <- median(scores$english)
mode_eng <- as.numeric(names(sort(table(scores$english), decreasing = TRUE)[1]))
# Print results for each subject
cat("Mathematics - Mean:", mean_math, "Median:", median_math, "Mode:", mode_math, "\n")
cat("Science - Mean:", mean_sci, "Median:", median_sci, "Mode:", mode_sci, "\n")
cat("English - Mean:", mean_eng, "Median:", median_eng, "Mode:", mode_eng, "\n")
# 2. Identify the subject with the highest variability using standard deviation
```

```
sd_math <- sd(scores$mathematics)
sd_sci <- sd(scores$science)
sd_eng <- sd(scores$english)
cat("Standard Deviation - Mathematics:", sd_math, "\n")
cat("Standard Deviation - Science:", sd_sci, "\n")
cat("Standard Deviation - English:", sd_eng, "\n")
max_sd <- max(sd_math, sd_sci, sd_eng)
if (max_sd == sd_math) {
  cat("Subject with highest variability: Mathematics\n")
} else if (max_sd == sd_sci) {
  cat("Subject with highest variability: Science\n")
} else {
  cat("Subject with highest variability: English\n")
}
# 3. Create boxplots for each subject
boxplot(scores$mathematics, main = "Boxplot for Mathematics", ylab = "Scores", col =
"lightblue")
boxplot(scores$science, main = "Boxplot for Science", ylab = "Scores", col = "lightgreen")
boxplot(scores$english, main = "Boxplot for English", ylab = "Scores", col = "lightcoral")
# 4. Visualize the relationship between Mathematics and Science scores using a scatter plot
plot(scores$mathematics, scores$science,
     main = "Mathematics vs Science Scores",
     xlab = "Mathematics Scores",
     ylab = "Science Scores",
     pch = 19, col = "blue")
abline(lm(scores$science ~ scores$mathematics), col = "red")
```

Output:-

```
# Print results for each subject
Mathematics - Mean: 85 Median: 86.5 Mode: 76
Science - Mean: 85.8 Median: 86 Mode: 78
English - Mean: 85.2 Median: 85.5 Mode: 79
```

Standard Deviation - Mathematics: 6.055301

Standard Deviation - Science: 5.82714

Subject with highest variability: Mathematics

3. Create boxplots for each subject