**DAY -2**

|  |  |
| --- | --- |
| **Ex.1**  # Create a matrix for the data  data <- matrix(c(18, 2, 20, 22, 28, 10, 20, 40, 40), nrow = 3, byrow = TRUE)  rownames(data) <- c("5-6 years", "7-8 years", "9-10 years")  colnames(data) <- c("A", "B", "C")  # Calculate covariance between B and C  cov\_bc <- cov(data[, "B"], data[, "C"])  print(paste("Covariance between B and C:", cov\_bc))  # Calculate sample covariance matrix for the preferences  cov\_matrix <- cov(data)  print("Sample Covariance Matrix:")  print(cov\_matrix)  # Calculate correlation between B and C  cor\_bc <- cor(data[, "B"], data[, "C"])  print(paste("Correlation between B and C:", cor\_bc))  # Calculate sample correlation matrix for the preferences  cor\_matrix <- cor(data)  print("Sample Correlation Matrix:")  print(cor\_matrix) | **Ex.2**  # Prices data  prices <- c(1, 1, 5, 5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18, 18, 18, 18, 20, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 25, 25, 25, 25, 25, 28, 28, 30)  # Equal-frequency partitioning with bin equal to 3  num\_bins <- 3  bin\_labels <- cut(prices, breaks = num\_bins, labels = FALSE)  # Data smoothing using bin boundaries  bin\_boundaries <- cut(prices, breaks = num\_bins)  bin\_boundaries\_clean <- as.numeric(as.character(bin\_boundaries))  # Plot histogram for frequency division  hist(prices, breaks = num\_bins, main = "Histogram of Prices", xlab = "Price", col = "lightblue") |
| **Ex.3**  # Data for Class A and Class B  class\_a <- c(76, 35, 47, 64, 95, 66, 89, 36, 84)  class\_b <- c(51, 56, 84, 60, 59, 70, 63, 66, 50)  # Calculate mean, median, and range for each class  mean\_a <- mean(class\_a)  mean\_b <- mean(class\_b)  median\_a <- median(class\_a)  median\_b <- median(class\_b)  range\_a <- max(class\_a) - min(class\_a)  range\_b <- max(class\_b) - min(class\_b)  # Determine which class scored higher mean, median, and range  mean\_comparison <- ifelse(mean\_a > mean\_b, "Class A", "Class B")  median\_comparison <- ifelse(median\_a > median\_b, "Class A", "Class B")  range\_comparison <- ifelse(range\_a > range\_b, "Class A", "Class B")  # Print the results  cat("Mean:", mean\_comparison, "had a higher mean.\n")  cat("Median:", median\_comparison, "had a higher median.\n")  cat("Range:", range\_comparison, "had a higher range.\n")  # Plot boxplot  boxplot(class\_a, class\_b, names = c("Class A", "Class B"), col = c("skyblue", "lightgreen"),  main = "Comparison of Exam Scores for Class A and Class B",  ylab = "Scores", xlab = "Class") | **Ex.4**  # Define the given data  data <- c(200, 300, 400, 600, 1000)  # Min-max normalization by setting min = 0 and max = 1  min\_max\_normalized <- (data - min(data)) / (max(data) - min(data))  # Z-score normalization  z\_score\_normalized <- (data - mean(data)) / sd(data)  # Print the normalized values  cat("Min-max normalized data:", min\_max\_normalized, "\n")  cat("Z-score normalized data:", z\_score\_normalized, "\n") |
| **Ex.5**  # Load the AirPassengers dataset  data("AirPassengers")  # Create a histogram with specified bins  hist(AirPassengers, breaks = seq(100, 700, by = 150), xlim = c(100, 700),  main = "Histogram of AirPassengers dataset",  xlab = "Passenger Count", ylab = "Frequency") | **Ex.6**  # Load the mtcars dataset  data(mtcars)  # Plot the first line  plot(mtcars$mpg, type = "l", col = "blue", xlab = "Index", ylab = "mpg and qsec")  # Add the second line to the plot  lines(mtcars$qsec, type = "l", col = "red")  # Add a legend  legend("topright", legend = c("mpg", "qsec"), col = c("blue", "red"), lty = 1) |
| **Ex.7**  # Read the CSV file  data <- read.csv("D:/phd/Dataset/water\_potability.csv")  # Check the data types of columns  str(data)  # Convert Potability column to factor  data$Potability <- as.factor(data$Potability)  # Convert Potability factor levels to numeric  data$Potability <- as.numeric(data$Potability)  # Remove rows with missing or infinite values in Hardness  data <- data[is.finite(data$Hardness), ]  # Replace NA values with mean of Hardness  data$Hardness[is.na(data$Hardness)] <- mean(data$Hardness, na.rm = TRUE)  # Replace NA values with mean of Potability  data$Potability[is.na(data$Potability)] <- mean(data$Potability, na.rm = TRUE)  # Plotting Potability vs. Hardness  plot(data$Hardness, data$Potability, xlab = "Hardness", ylab = "Potability", main = "Scatter plot of Potability vs. Hardness")  # Fit linear regression model  model <- lm(Potability ~ Hardness, data = data)  # Add the regression line to the plot  abline(model, col = "red")  # Predict Potability for Hardness=88  predict\_value <- predict(model, newdata = data.frame(Hardness = 88))  cat("Predicted Potability for Hardness=88:", predict\_value, "\n")  # Summary of the linear regression model  summary(model) | **Ex.8**  # Load the mtcars dataset  data("mtcars")  # Create a boxplot for mpg vs. cyl  boxplot(mpg ~ cyl, data = mtcars,  xlab = "Number of Cylinders",  ylab = "Miles per Gallon",  main = "Boxplot of Miles per Gallon vs. Number of Cylinders") |
| **Ex.9**  # Sample data  set.seed(42)  points <- c(50, 60, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290, 295, 300, 350, 400)  # Add a few outliers  outliers <- c(20, 25, 30, 35, 400, 450, 500)  points <- c(points, outliers)  # Create a boxplot  boxplot(points,  main = "Boxplot of Points Scored by Tennis Team Players",  ylab = "Points",  ylim = c(0, 500),  notch = TRUE, # Add a notch to the box  outline = TRUE, # Show outliers as individual points  col = "lightblue" # Box color  ) | **Ex.10**  # Load the dataset  diabetes <- read.csv("D:/phd/Dataset/diabetes\_dataset.csv")  # Scatterplot  plot(diabetes$Age, diabetes$BloodPressure,  xlab = "Age", ylab = "Blood Pressure",  main = "Scatterplot of Blood Pressure vs. Age")  # Create age groups  age\_groups <- cut(diabetes$Age, breaks = 4)  # Calculate average BloodPressure for each age group  avg\_bp <- tapply(diabetes$BloodPressure, age\_groups, mean)  # Bar chart  barplot(avg\_bp,  xlab = "Age Group", ylab = "Average Blood Pressure",  main = "Average Blood Pressure by Age Group",  col = "skyblue") |
| **# Load built-in datasets**  **data(iris)**  **data(mtcars)**  **# Scatter plot**  **plot(iris$Sepal.Length, iris$Sepal.Width, main = "Scatter plot of Sepal Length vs. Width", xlab = "Sepal Length", ylab = "Sepal Width")**  **# Line chart**  **plot(mtcars$mpg, type = "l", main = "Line chart of MPG", xlab = "Car Index", ylab = "Miles per Gallon")**  **# Bar chart**  **barplot(table(iris$Species), main = "Bar chart of Iris Species", xlab = "Species", ylab = "Frequency")**  **# Histogram**  **hist(mtcars$mpg, main = "Histogram of MPG", xlab = "Miles per Gallon", ylab = "Frequency")**  **# Box plot**  **boxplot(mtcars$mpg, main = "Boxplot of MPG")**  **# Pie chart**  **pie(table(iris$Species), main = "Pie chart of Iris Species")**  **# Heatmap**  **heatmap(cor(mtcars), main = "Heatmap of mtcars dataset")**  **# Scatter plot matrix**  **pairs(iris[, 1:4], main = "Scatterplot Matrix of Iris dataset")**  **install.packages("ggplot2")**  **# Area chart (requires 'ggplot2' package)**  **library(ggplot2)**  **ggplot(mtcars, aes(x = factor(cyl), fill = factor(gear))) +**  **geom\_bar(position = "fill") +**  **labs(title = "Stacked Bar Chart of Cars by Cylinder and Gear")**  **# Violin plot (requires 'ggplot2' package)**  **ggplot(mtcars, aes(x = factor(cyl), y = mpg, fill = factor(cyl))) +**  **geom\_violin() +**  **labs(title = "Violin Plot of MPG by Cylinder")** |  |