| Course Code: ITA04 | Course Name: Statistics with R Programming |
|---|---|

### CLASS TEST-3    Total Marks :50        Date:12-02-25

| S.No | TEST 1 | MARKS | COs | BT LEVEL |
|---|---|---|---|---|
| 1 | A meteorological department maintains a data frame containing weather records for different cities, including temperature, humidity, wind speed, and precipitation levels. The analyst needs to analyze this data to extract useful insights. First, they need to create a data frame that includes weather details for five cities. Then, they must filter out cities where the temperature exceeds 30°C. Finally, they should display the extracted subset of cities. <br> **Dataset:** <br> Cities: "New York", "Los Angeles", "Chicago", "Houston", "Miami" <br> Temperature (°C): 22, 28, 15, 32, 35 <br> Humidity (%): 65, 70, 55, 80, 85 <br> Wind Speed (km/h): 10, 12, 14, 8, 7 <br> Precipitation (mm): 5, 0, 10, 2, 20 <br> **Questions:** <br> 1.Write an R script to create a data frame containing weather details, including cities, temperature, humidity, wind speed, and precipitation. Then, display the complete data frame. <br> 2.Modify the script to filter and display only the cities where the temperature exceeds 30°C using the subset() function. <br> 3.Use a conditional statement to identify cities with precipitation above 10mm and print their names along with a message indicating that they have high rainfall. | 10 | C03 | BL3 |
| 2 | A city's traffic department monitors accident-prone zones to enhance road safety. They collect data on accident occurrences, severity levels, and affected areas over five days. An analyst needs to process this data for further insights. First, they create a data frame containing accident records along with their respective dates. To ensure proper analysis, they convert the date column into Date format using the as.Date() function. Next, they inspect the structure of the data frame using the str() function to confirm the data types. This helps identify trends in accident occurrences and severity levels. Finally, the processed data is used for implementing road safety measures and policy planning. <br><br> **Tasks:** <br> 1.Create a data frame containing daily accident records for five days. The data frame should include: <br> Dates: "2024-03-01", "2024-03-02", "2024-03-03", "2024-03-04", "2024-03-05" <br> Accidents Reported: 5, 8, 3, 10, 7 <br> Severity Levels: "Moderate", "High", "Low", "High", "Moderate" <br> Affected Zones: "Zone A", "Zone B", "Zone C", "Zone B", "Zone A" <br> 2.Convert the Dates column into Date format using the as.Date() function. <br> 3.Display the structure of the data frame using the str() function to inspect its data types. <br> **Questions:** <br> 1.Write an R script to create an accident zone data frame and convert the date column into Date format using the as.Date() function. Then, display the complete data frame. <br> 2.Modify the script to use the str() function to check and display the structure of the accident zone data frame. Explain the output. <br> 3.Use a conditional statement to identify days where more than 7 accidents were reported and print the date along with a message indicating a high accident rate. | 10 | C03 | BL3 |

| | | | | |
|---|---|---|---|---|
| | 4.Filter and display only the days where the severity level was marked as "High" using the subset() function. | | | |
| 3 | A technology research firm tracks the growth of artificial intelligence (AI) across various industries. They collect data on AI adoption rates, investment levels, and the number of AI startups emerging each year. An analyst needs to process this data to gain insights into AI expansion. First, they create a data frame containing AI growth statistics for five years. They then add a new column representing the percentage increase in AI adoption. Additionally, they add a new row with projected data for the next year. Finally, they merge two related data frames using the rbind() function to consolidate information. This structured data helps in forecasting AI trends and investment opportunities.<br>**Tasks:**<br>1.Create a data frame containing AI growth data for five years, including:<br>Years: "2019", "2020", "2021", "2022", "2023"<br>AI Adoption Rate (%): 25, 30, 40, 55, 65<br>Investment in AI (Billion USD): 10, 15, 22, 30, 45<br>Number of AI Startups: 150, 200, 280, 350, 500<br>2.Add a new column calculating the percentage increase in AI adoption compared to the previous year.<br>3.Add a new row with projected data for 2024:<br>AI Adoption Rate (%): 75<br>Investment in AI (Billion USD): 60<br>Number of AI Startups: 650<br>Percentage Increase: Based on 2023 data<br>4.Merge another data frame containing additional insights (e.g., AI patents filed each year) using the rbind() function.<br>5.Display the final data frame.<br>**Questions:**<br>1.Write an R script to create an AI growth data frame and display it.<br>2.Modify the script to add a new column representing the percentage increase in AI adoption each year.<br>3.Use rbind() to add a new row containing projected AI data for 2024. Display the updated data frame.<br>4.Merge another data frame containing the number of AI patents filed per year using rbind(), then display the final result. | 10 | C03 | BL2 |
| 4. | A school administrator is analyzing student performance in a recent examination. They have a dataset containing student names, scores, the number of attempts, and whether the student qualified. The administrator needs to perform several tasks to refine and manipulate the data for better insights.<br>First, they extract specific rows and columns to focus on particular students. Then, they add a new column indicating each student's country of origin. Next, they update the dataset by adding new student records. To better organize the data, they sort it based on student names and scores. Finally, they save the processed data into a file for future reference and display its contents to verify accuracy. This structured approach helps in student performance analysis and decision-making regarding academic support.<br> **Tasks:**<br>Create below data frame<br>exam_data = data.frame (name = c('Anastasia', 'Dima', 'Katherine', 'James','Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'),<br>score = c (12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),<br>attempts = c (1, 3, 2, 3, 2, 3, 1, 1, 2, 1),<br>qualify = c ('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes'))<br>**Questions:**<br>a. Write a R program to extract 3rd and 5th rows with 1st and 3rd columns from a given data frame b. Write a R program to add a new column named country in a given data frame<br>b. Country <- c ("USA", | 10 | C03 | BL3 |

| | | | | |
|---|---|---|---|---|
| | "USA","USA","USA","UK","USA","USA","India","USA","USA")<br>c. Write a R program to add new row(s) to an existing data frame<br>new_exam_data = data. frame(name = c('Robert', 'Sophia'),score = c(10.5,9), attempts = c(1, 3),qualify = c('yes', 'no'))<br>d. Write a R program to sort a given data frame by name and score<br>Write a R program to save the information of a data frame in a file and display the information of the file. | | | |
| 5. | A global shipping company tracks cargo movements to improve logistics and efficiency. They collect data on vessel names, cargo weight, departure and arrival ports, travel duration, and delays. An analyst needs to process this data for operational insights. First, they create a data frame containing cargo details for five different shipments. Then, they save this data as an **Excel file (XLS/XLSX)** for record-keeping. They use **getwd()** to check the current working directory and **setwd()** to modify it if necessary before saving the file. Finally, they read the Excel file back into R and analyze it to identify trends in cargo weight and shipping delays. This structured approach helps improve logistics planning and reduce transit inefficiencies.<br> **Tasks:**<br>   1. **Create a data frame** containing ship cargo details, including:<br>      o Vessel Name: "Titan", "Poseidon", "Neptune", "Atlantis", "Odyssey"<br>      o Cargo Weight (Tons): 5000, 7200, 6100, 4500, 8000<br>      o Departure Port: "Shanghai", "Los Angeles", "Hamburg", "Singapore", "Dubai"<br>      o Arrival Port: "New York", "Tokyo", "Sydney", "London", "Mumbai"<br>      o Travel Duration (Days): 15, 20, 18, 12, 25<br>      o Delay (Hours): 5, 0, 10, 3, 8<br>   2. **Save the data frame** to an **Excel file (XLSX format)** using the write.xlsx() function.<br>   3. **Use getwd()** to check the current working directory and **setwd()** to change it if needed before saving the file.<br>   4. **Read the Excel file back** into R using read.xlsx() and display its contents.<br>   5. **Analyze the cargo data** by filtering delayed shipments and calculating the average cargo weight per shipment.<br>**Questions:**<br>   1. Write an R script to create a ship cargo data frame and display it.<br>   2. Modify the script to save the data frame as an **Excel file (XLSX)** and verify its storage using getwd() and setwd().<br>   3. Write an R program to read the stored **Excel file (XLSX)** into a data frame and display its contents.<br>   4. Analyze the ship cargo data by filtering delayed shipments and calculating the average cargo weight per shipment. | 10 | C03 | BL4 |