Reg. No……………………………………….   Name …………………………………………… **26U667**

B.Sc. DEGREE END SEMESTER EXAMINATION - MARCH 2026

SEMESTER **6: STATISTICS FOR COMPUTER APPLICATION**

COURSE: **19U6CRCST7 - COMPUTER AIDED DATA ANALYSIS USING EXCEL AND R**
(For Regular 2023 Admission and 2022/2021/2020/2019 Admissions)

Time : Three Hours   Max Marks: 75

PART- A

(Each question carries five marks, maximum marks from this part is 25)

1. Mention any Three advantages of R programming over MS excel

2. Explain the difference between the $c()$ and $scan()$ functions for inputting data into R with examples.

3. You have a sample size of $n = 45$, but the population standard deviation (σ) is unknown. Should you use a Z-test or a t-test in Excel? Justify.

4. Distinguish between **simple and multiple correlation** analysis in the context of Excel.

5. Distinguish between a **Histogram** and a **Bar Diagram** in terms of the types of data (continuous vs. discrete) they represent. Provide the R function name for each.

6. How do you calculate the **Geometric Mean** of a dataset using R? Provide the necessary function or formula.

7. Explain the purpose of the **Chi-square test** in inferential statistics and how it is implemented in R software

8. Define **Dispersion** and list the functions used to calculate **Variance** in R.

PART - B

(Each question carries ten marks, maximum marks from this part is 50)

9. You are given a dataset "weather_data.csv" containing columns "Date", "Temperature", "Humidity", and "Rainfall". As a data analyst, write the R code to:
    - Load the dataset into the R environment.
    - Display the first 10 rows.
    - Calculate the Arithmetic Mean and Median for "Temperature".
    - Identify any missing values in the "Rainfall" column.

10. A manufacturing company wants to compare the weight of products from two different machines.

    Machine A: 50.2, 50.5, 49.8, 51.0, 50.1

    Machine B: 51.5, 51.8, 51.2, 52.0, 51.1

    Perform a Two-sample t-test in MS Excel to determine if there is a significant difference between the two machines at significance level = 0.01.

11. Standard R does not have a built-in $geometric.mean()$ function in the base package.

    (a) Write a custom R function to calculate the Geometric Mean for a numeric vector x.

    (b) Demonstrate how to call this function using the vector $v <- c(2, 8, 32)$

12. A company tracks its monthly advertising spend and corresponding sales revenue.

    Spend (in Rs): 10, 15, 20, 25, 30, 35

    Sales (in Rs): 100, 145, 190, 250, 310, 360

    In MS Excel, perform a Simple Regression analysis to find the relationship between Spend and Sales. What is the predicted sale if the spend is Rs 40?

13. Use R software to perform a One-sample t-test to determine if the mean weight of a sample of 20 items (mean = 48g, SD = 4g) differs significantly from the population mean of 50g at a 5% significance level.

14. A biologist wants to visualize the growth of a plant over 10 days.

    a. Days: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

    b. Height: 2, 3.5, 5, 7, 9, 12, 15, 18, 22, 25

    c. Write R code to:

    • Create an XY Plot of the data.

    • Add a Trend line to the plot.

    • Customize the plot with a blue line and circular points.

15. Explain the procedure for conducting an F-test for variance in MS Excel to compare the consistency of two different investment portfolios.

16. The following data represents the population of a town (in thousands) over 6 years:

    Year: 2018, 2019, 2020, 2021, 2022, 2023

    Pop: 50, 54, 59, 63, 68, 72

    Using R, fit a linear trend line, determine the equation of the line, and interpret the Correlation Coefficient (r).