Reg. No…… Name…… **25U664**

B. Sc DEGREE END SEMESTER EXAMINATION - MARCH 2025

SEMESTER- 6: **STATISTICS FOR COMPUTER APPLICATION**

COURSE: **19U6CRCST7- COMPUTER AIDED DATA ANALYSIS USING EXCEL AND R**

(For Regular 2022 Admission and 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. What are some key features that have contributed to the popularity of R programming language in the data science community?
2. Discuss the functionality of the cbind() function in R. Give an example illustrating the usage of cbind() to combine vectors or matrices.
3. Write various steps involved in multiple correlation analysis in MS Excel.
4. Write a short note based on different types of operators in R program. Give an example for each.
5. How to perform two sample t test in MS Excel?
6. Write an R code to calculate mean and standard deviation of a dataset using relevant R functions.
7. Write an R code to obtain the geometric mean for a set of n natural numbers.
8. Explain how to customize titles, axis labels and legend in a ggplot2 scatterplot. Provide sample code to demonstrate changing plot aesthetics.

PART – B

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

9. Suppose you're tasked with analysing a dataset containing information about customer transactions for a retail company. The dataset is provided in a CSV file named "transactions.csv". The dataset includes columns such as "CustomerID", "TransactionID", "ProductID", "Quantity", "Price", and "TransactionDate". As a data analyst, outline the steps you would take to explore and understand the dataset using R. Provide examples of R code snippets to perform basic data handling tasks such as loading the dataset, summarizing key statistics.
10. A random sample of 15 light bulbs from manufacturer A gave a mean lifetime of 1450 hours with a standard deviation of 90 hours. Another random sample of 12 light bulbs from manufacturer B gave a mean lifetime of 1520 hours with a standard deviation of 100 hours. At $\alpha = 0.05$, is there sufficient evidence to conclude that the mean lifetimes of the light bulbs made by the two manufactures are different (MS Excel).
11. Given the following exam scores for a class of 25 students:

    71, 62, 85, 56, 90, 83, 67, 55, 77, 81, 92, 69, 57, 78, 82, 75, 63, 87, 88, 61, 79, 59, 73, 89, 86.

    Create this data in MS Excel and import it into R and then find the mean, median and mode. Comment on what each measures tells us about the data.

12. A dietitian wanted to determine whether a new weight loss program was effective. She recruited 10 overweight volunteers and recorded their weights before and after participating in the program. The data is given below:

Subject	: 1	2	3	4	5	6	7	8	9	10
Weights Before (lbs)	: 180	165	175	188	192	204	215	210	197	189
Weights After (lbs)	: 176	162	171	182	180	195	203	201	191	180

Using a significance level α= 0.05, is there sufficient evidence to conclude that the new weight loss program reduces weight (using MS Excel).

13. A beverage company claims their 16oz sodas are filled on average to 15.8oz. A quality control team randomly selects 75 sodas and measures the amount filled in each. The sample mean fill amount is 15.7 oz and the sample standard deviation is 0.6oz. Test the claim at 5% significance level that the true average fill amount is less than 15.8oz (oz stands for fluid ounces) (using R program).

14. A quality control manager wants to determine if there is a difference in variability between two new manufacturing processes A and B. Samples were taken from each process and the results as shown below:

Process A: 2.1, 2.0, 1.8, 2.2, 1.9

Process B: 2.8, 3.1, 2.7, 3.0, 2.9

Conduct an F- test at the 95% confidence level in Excel to determine if there is evidence that the variance of process B is different than process A.

15. The following data shows the annual revenue generated by a company over the past 8 years:

Year	2014	2015	2016	2017	2018	2019	2020	2021
Revenue (in thousand $)	165	187	211	235	268	292	315	333

(a) Plot the data on a scatterplot with year on the x- axis and revenue on the y- axis.

(b) Fit a linear trend line to the data.

(c) Determine the equation of the trend line. What does the slope and y- intercept represent in terms of the revenue data (using R program).

16. A dataset contains the math test scores and reading test scores for 15 students. The data is given below:

Math Test Score	Reading Test Score
68	78
81	63
77	69
86	77
73	82
95	88
62	58
79	72
85	81
91	83
74	89
89	86
71	75
82	84
76	78

Use this data to perform a linear regression analysis in Excel to determine if a student's math test score can predict their reading test score. Create a scatterplot of the data. Use the data to build a linear regression model to predict reading score based on math score. Also, based on the regression output, what is the predicted reading score for a student who scored 80 on the math test?