# B. Sc. DEGREE END SEMESTER EXAMINATION - MARCH 2024

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

COURSE: **19U6CRCST07 - COMPUTER AIDED DATA ANALYSIS USING EXCEL AND R**

*(For Regular 2021 Admission and Supplementary 2020/2019 Admissions)*

Time: Three Hours                                                                                    Max. Marks: 75

## PART A

### *(Each question carries 5 marks, maximum marks from this part is 25)*

1. How can we import a CSV file into R and create a data frame?
2. How can we create a scatter plot and bar chart in Excel to visualize the relationship between two variables.
3. Write a function in R to get all prime numbers up to a given number.
4. Explain how to interpret the correlation coefficient and its significance in Excel.
5. How can we perform a Two sample t test for means in Excel.
6. Explain the differences between vectors, matrices and data frame in R.
7. What are the potential limitations of using Excel for statistical analysis?
8. Describe different kinds of data objects/ structures and their operations in R.

## PART B

### *(Each question carries 10 marks, maximum marks from this part is 50)*

9. The following data were collected on 15 obese girls:

| Weight in Kg (Y) | Lean Body Weight (X1) | Mean Daily Caloric Intake (X2) |
|---|---|---|
| 79.2 | 54.3 | 2670 |
| 64.0 | 44.3 | 820 |
| 67.0 | 47.8 | 1210 |
| 78.4 | 53.9 | 2678 |
| 66.0 | 47.5 | 1205 |
| 63.0 | 43.0 | 815 |
| 65.9 | 47.1 | 1200 |
| 63.1 | 44.0 | 1180 |
| 73.2 | 44.1 | 1850 |
| 66.5 | 48.3 | 1260 |
| 61.9 | 43.5 | 1170 |
| 72.5 | 43.3 | 1852 |
| 101.1 | 66.4 | 1790 |
| 66.2 | 47.5 | 1250 |
| 99.9 | 66.1 | 1789 |

Find the multiple correlation coefficient and test it for significance using MS Excel.

10. The following are the pulmonary blood flow (PBF) and pulmonary blood volume (PBV) values recorded for 16 infants and children with congenital heart disease.

| X | Y |
|---|---|
| 168 | 4.31 |
| 280 | 3.4 |
| 391 | 6.2 |
| 420 | 17.3 |
| 303 | 12.3 |
| 429 | 13.99 |
| 605 | 8.73 |
| 522 | 8.9 |
| 224 | 5.87 |
| 291 | 5.0 |
| 233 | 3.51 |
| 370 | 4.24 |
| 531 | 19.41 |
| 516 | 16.61 |
| 211 | 7.21 |
| 439 | 11.6 |

Find the regression equation describing the linear relationship between the two variables. Compute $r^2$ and adjusted $r^2$ using R program.

11. A concurrent cohort study was carried out to measure the effect of birth weight on the subsequent growth of children. 300 children with birth weight 2 Kg to 2.5 Kg were followed till age one, when anthropometric measurements were made to assess the nutritional status. A similar number of children born during the same period with birth weight greater than 2.5 Kg also followed up and data is given below:

| Type | Low Birth Weight | Normal Birth Weight |
|---|---|---|
| No. of children studied | 300 | 300 |
| No. of malnourished at age one | 120 | 60 |

Carry out a $\chi^2$ test to assess the difference in malnourishment between two groups of children using an R program.

12. (a) Distribution of blood groups of students in a university are given below. Construct pie-chart for males and females using R functions.

| Blood Group | Male | Female |
|---|---|---|
| A | 427 | 317 |
| B | 559 | 412 |
| O | 521 | 367 |
| AB | 122 | 85 |

(b) The systolic blood pressure (in mm mercury) of a sample of 100 individuals from the Honolulu Hear study population of 7683 persons were given below. Construct a Histogram using R program.

| 102 | 116 | 118 | 118 | 104 | 118 | 114 | 116 | 122 | 156 |
| 138 | 152 | 134 | 138 | 112 | 128 | 114 | 154 | 94 | 132 |
| 190 | 134 | 178 | 134 | 140 | 116 | 98 | 126 | 122 | 140 |
| 122 | 132 | 134 | 108 | 102 | 134 | 128 | 140 | 172 | 150 |
| 128 | 130 | 162 | 96 | 142 | 108 | 130 | 122 | 100 | 130 |
| 112 | 118 | 162 | 142 | 146 | 134 | 122 | 154 | 150 | 118 |
| 128 | 136 | 120 | 122 | 92 | 124 | 112 | 140 | 154 | 162 |
| 116 | 108 | 98 | 146 | 112 | 124 | 106 | 120 | 170 | 128 |
| 134 | 108 | 141 | 126 | 152 | 114 | 128 | 140 | 140 | 130 |
| 101 | 128 | 118 | 176 | 116 | 154 | 128 | 114 | 144 | 208 |

13. A manufacturing company produces three different types of widgets: Type A, Type B and Type C. The company is concerned about the consistency of the diameters of these widgets across different production runs. To investigate this, the company takes a random sample of 10 widgets from each type and measures their diameters. The diameters (in millimetres) are as follows:

**Type A**: 12.1, 11.9, 12.0, 11.8, 12.2, 12.8, 12.5, 12.2, 12.1, 12.6

**Type B**: 11.8, 11.7, 11.9, 12.0, 11.5, 11.6, 11.2, 11.9, 11.3, 12.0

**Type C**: 12.0, 12.1, 12.2, 12.3, 12.4, 12.5, 12.6, 12.7, 12.8, 12.9

Using the provided data, conduct an F- test for equality of variances at a significance level of 0.05 (using MS Excel).

14. Suppose you are interested in comparing the average monthly electricity bills of two different types of households: households with solar panels installed (Group A) and households without solar panels (Group B).

**Group A (Households with solar panels):**

Monthly electric bills (in dollars): $120, $110, $115, $125, $130, $132

**Group B (Households without solar panels):**

Monthly electricity bills (in dollars): $150, $160, $155, $165, $170, $178

Using the provided data, conduct a two-sample Z-test for the mean to determine whether there is a significant difference between the average monthly electricity bills of households with and without solar panels. Assume a significance level of 0.05 (using MS Excel).

15. (a) The heights of 10 students are given. Write R program to obtain the mean and median for the following data.

    150   172   160   169   180   173   158   165   182   163

    (b) The weights of 10 persons are given below. Write R program to find the SD and Variance for the given data.

    45   70   62   65   81   70   52   58   78   59

    (c) Draw a bar chart using Excel for the data given below:

| Country | Production of rice in 2017- 18(in metric tons) |
|---|---|
| China | 149 |
| India | 113 |
| Indonesia | 37 |
| Bangladesh | 33 |
| Vietnam | 28 |
| Thailand | 20 |

16. We are conducting a study to analyse the effectiveness of a new teaching method designed to improve students' math skills. The study involves comparing the performance of students befor and after the implementation of the new teaching method. You have collected data from two groups: Group A (students who underwent the new teaching method) and Group B (students who received traditional teaching)

    **Group A (New Teaching Method):**

    Before: 80, 85, 78, 82, 79

    After: 85, 88, 82, 87, 84

    **Group B (Traditional Teaching):**

    Before: 75, 78, 72, 77, 74

    After: 78, 80, 75, 79, 76

    (a) Conduct a one-sample t-test to determine if there is a significant improvement in the math scores of students in Group A after undergoing the new teaching method. Use a significance level of 0.05 (using R software).

    (b) Perform a two-sample t-test to compare the mean math scores between Group A and Group B after the teaching interventions. Use a significance level of 0.05 (using R software).

    **********