LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.A. DEGREE EXAMINATION – **ECONOMICS**

THIRD SEMESTER – **NOVEMBER 2022**

PEC 3301 – DATA ANALYTICS FOR ECONOMISTS

Date: 30-11-2022 Dept. No. Max. : 100 Marks

Time: 09:00 AM - 12:00 NOON

PART- A

Answer any FIVE questions in 75 words each. **(5 X 4=20 Marks)**

1. Distinguish between predictive and descriptive models.
2. What are the social implications of Data Mining?
3. What is CLV? How is it determined?
4. Define a data frame in R and give the syntax to create and display a data frame.
5. What are the types of operators in R Language?
6. State the various data types in R.
7. Differentiate between parameter and statistic.

PART- B

Answer any FOUR questions in 300 words each. **(4X10=40 Marks)**

8. Examine the key steps involved in the Knowledge Discovery in Databases (KDD) process.
9. Differentiate between Logistic Regression and Simple Linear Regression.
10. Explain the scope of visualization techniques and its various types.
11. Provide a diagrammatic overview of the analytics process model.
12. Write the R code to create a data frame and name it 'student data', with six variables, dept.no., name, age, marks in sem1, marks in sem. 2, marks in sem. 3, also enter ten values for each variable, find the average marks in the three semesters and categorize the students as having obtained distinction (≥ 75), first class (≥60 but < 75), second class(≥ 50 but less than 60), third class(≥40 but less than 50) and failed (<40) and report the classification as the seventh column with column name 'class'.
13. State the relational operators in R, describe and give an example for each one.
14. What is statistical inference? Explain the concept of hypothesis testing and the steps involved in testing a statistical hypothesis.

PART- C

Answer any TWO questions in 1200 words each. **(2 X 20 = 40 marks)**

15. The data on monthly profits (in lakhs of INR) of 12 shops on a road before and after closure of the road for Metrorail construction are given below:

Before	12.50	18.25	16.40	20.80	8.25	16.85	17.50	13.75	15.60	17.20	14.85	19.40
After	10.25	16.45	10.58	22.15	7.05	16.40	13.63	11.90	16.40	11.90	12.48	13.38

The Metrorail constructing company claims there is no impact due to the construction activity while the merchants claim that their average monthly profits have gone down significantly after the work started. You are asked to investigate these claims. Formulate the appropriate null and alternative hypotheses explaining the notations.

(a) Formulate the null and alternate hypotheses

(b) Write the code to get the simple linear regression model equation.

(c) Write the simple linear regression model equation.

(d) Give the interpretation of the coefficients in the R-Output.

(e) Give the clear interpretations of the p-values in the output, and R^2.

```
Call:
lm(formula = After ~ Before)

Coefficients:
(Intercept)       Before
    -1.0080       0.9128

> print(summary(relation))

Call:
lm(formula = After ~ Before)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3821 -1.7002  0.1448  1.1062  4.1716

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.0080     3.7284  -0.270  0.79238
Before        0.9128     0.2292   3.982  0.00259 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.543 on 10 degrees of freedom
Multiple R-squared:  0.6132,    Adjusted R-squared:  0.5746
F-statistic: 15.86 on 1 and 10 DF,  p-value: 0.002593
```

16. Explain the concept of Simple linear regression and multiple linear regression analysis in statistical modeling.
17. Illustrate the use of decision trees with the help of a suitable example.
18. Give a detailed account of various implementation issues associated with data mining.

&&&&&&&&&&&&