

Reg. No.

Question Paper Code

11542

B.E./B.Tech. - DEGREE EXAMINATIONS, NOV/DEC 2022

Fifth Semester

Information Technology

20ITPW501 - STATISTICAL ANALYSIS USING R PROGRAMMING WITH

LABORATORY

(Regulations 2020)

Duration: 3 Hours

Max. Marks: 100

Answer ALL Questions

PART - A (10 × 2 = 20 Marks)

- | | <i>Marks,
K-Level, CO</i> |
|---|-------------------------------|
| 1. List out the methods for treating the missed values in R. | 2,K2,CO1 |
| 2. Differentiate between Vector, List, Matrix, and Data frame. | 2,K2,CO1 |
| 3. Differentiate between Sampling and Population. | 2,K2,CO2 |
| 4. Define Histograms. | 2,K2,CO2 |
| 5. What is One-Sample T Test? | 2,K1,CO3 |
| 6. Define Strip Charts. | 2,K2,CO3 |
| 7. Can you solve the Multiclass Classification problems using Logistic Regression? If Yes then How? | 2,K2,CO4 |
| 8. Regression analysis was applied between \$ sales (y) and \$ advertising (x) across all the branches of a major international corporation. The following regression function was obtained. $Y = 5000 + 7.25x$
If the advertising budgets of two branches of the corporation differ by \$30,000, then what will be the predicted difference in their sales? | 2,K2,CO4 |
| 9. Mention how you can produce Co-Relations and Co-Variations. | 2,K2,CO5 |
| 10. Differentiate Over Fitting and Under Fitting concepts in Regression. | 2,K2,CO5 |

Answer ALL Questions

PART - B (5 × 13 = 65 Marks)

11. a) Explain about Lists and Data Frames in R with suitable example. 13, K2,CO1
- OR**
- b) Explain about Flow control in R with suitable example. 13, K2,CO1
12. a) Simulate how coin-tossing can be done using rbinom function with appropriate example instead of sampling. How exactly would you do that? 13, K3,CO2

OR

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create

11542

- b) A rule of thumb is that 5% of the normal distribution lies outside an Interval approximately $\pm 2s$ about the mean. To what extent is this true? Illustrate how Quantiles and Normal Distribution methods are used to find the limits corresponding to 1%, 0.5%, and 0.1%? What is the position of the quartiles measured in standard deviation units? 13, K3, CO2

13. a) Explain about Dot Charts and Pie Charts with suitable example. 13, K3, CO3

OR

- b) In the data set vitcap, use a Two Sample T Test to compare the vital capacity for the two groups. Calculate a 99% confidence interval for the difference. The result of this comparison may be misleading. Why? 13, K3, CO3

Vital Capacity Data Set

Group = c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L)

Age = c(39, 40, 41, 41, 45, 49, 52, 47, 61, 65, 58, 59, 27, 25, 24, 32, 23, 25, 32, 18, 19, 26, 33, 27)

vital.capacity = c(4.62, 5.29, 5.52, 3.71, 4.02, 5.09, 2.7, 4.31, 2.7, 3.03, 2.73, 3.67, 5.29, 3.67, 5.82, 4.77, 5.71, 4.47, 4.55, 4.61, 5.86, 5.2, 4.44, 5.52)

14. a) Categorize the various type of regression with suitable example. 13, K2, CO4

OR

- b) Explain about displaying multivariate data and write an R program to plotting Multivariate data. 13, K2, CO4

15. a) Explain how Polynomial Regression can be implemented in R and Steps involved with suitable example. 13, K2, CO6

OR

- b) Explain One way ANOVA test with suitable example. 13, K2, CO6

PART C (1 × 15 = 15 Marks)

16. a) Analyze a null and alternate hypothesis for the following claim “A generic brand of the anti-histamine Diphenhydramine markets a Capsule with a 50 milligram dose. The manufacturer is worried that the machine that fills the capsules has come out of calibration and is no longer creating capsules with the appropriate dosage” and Justify using One sampled T test. Use the following collected sample of 15 capsules. Drug.Dose(in Milligrams) = c(28,29,35,37,32,26,37,39,22,29,36,38,42,38,41) 15, K3, CO5

OR

- b) Explain the different kinds of Lines, Point Plots and Graphs in R with suitable example. Vary the plot symbol, line type, line width, and color in examples. 15, K3, CO5