(Pages : 6) P – 1312

Reg. No. : .....

Name : .....

Second Semester B.Sc. Degree Examination, September 2022

First Degree Programme under CBCSS

Statistics

Foundation Course

ST 1221 : STATISTICAL METHODS II

(2020 Admission Onwards)

Time : 3 Hours Max. Marks : 80

Instructions: Use of calculator is permitted.

SECTION – A

(Very short answer)

Answer **all** the questions. Each question carries **1** mark.

1. What are the limits for Karl Pearson correlation coefficient?

2. What is the effect of change of origin and scale on correlation coefficient?

3. Who introduced the term regression?

4. With two attributes, what is the total number of ultimate class frequencies?

5. If the regression coefficients are –0.9 and –0.4, what is correlation coefficient?

6. In the case of perfect correlation, two regression lines are

7. What is knowledge discovery in data base?

8. Define neural network

9. What is c function in R?

10. Define data warehouse.

(10 × 1 = 10 Marks)

SECTION – B

(Short answer)

Answer any **eight** questions. Each question carries 2 marks.

11. What are regression coefficients?

12. Define Spearman's rank correlation coefficient.

13. Indicate whether the following statements are true or false:

    (a) Karl Pearson correlation coefficient measures the degree of linear relationship.

    (b) Correlation between price and demand is a positive correlation.

    (c) Regression coefficients are independent of change of origin and scale.

    (d) Uncorrelated variables are independent.

14. Write short notes on principle of least squares.

15. Define rank correlation. What is its importance?

16. With the usual notations if $r_{12} = 0.7, r_{23} = r_{13} = 0.5$, find $r_{13.2}$.

17. State the properties of multiple correlation coefficient.

18. Show that two independent variables are uncorrelated.

19. Define correlation ratio. What is its use?

20. How to compute median using R?

21. What are decision trees in data mining?

22. What is the importance of logistic regression in data mining techniques?

23. Define predictive data mining.

24. How to create data frame in R?

25. How to draw a pie diagram in R?

26. Mention the usage of time series in data mining.

(8 × 2 = 16 Marks)

SECTION – C

(Short Essay questions)

Answer any **six** questions. Each question carries **4** marks.

27. Explain scatter diagram method of studying correlation.

28. Find mean values of X and Y and correlation coefficient between them from the following regression equations:
$2Y - X - 50 = 0$
$3Y - 2X - 10 = 0$

29. Define attributes. What do you mean by independence of attributes? Give a criterion of independence for attributes A and B.

30. Explain the method of fitting of an exponential curve $y = ab^x$ by principle of least squares.

31. Find the angle between two lines of regression in a bivariate distribution.

32. Distinguish between positive and negative correlations with examples.

33. Define standard error and probable error of correlation coefficient. Point out the uses of probable error of correlation coefficient.

34. Explain the procedure for obtaining quartiles and percentiles using R software.

35. Describe briefly the applications of data mining techniques.

36. Write down the R commands to plot the scatter diagram and to compute Karl Pearson coefficient of correlation for a given bivariate data.

37. Explain briefly the key features of online analytical processing.

38. Explain why excel is used for statistical purposes.

(6 × 4 = 24 Marks)

SECTION – D

(Essay questions)

Answer any **two** questions. **Each** question carries **15** marks.

39. (a) 'Define Karl Pearson coefficient of correlation, Give its interpretations.

(b) Calculate coefficient of correlation between X and Y series from the following data and calculate its probable error. **(5 + 10 = 15)**

X	78	89	96	69	59	79	68	61
Y	125	137	156	112	107	136	123	108

40. (a) Explain the concept of regression. Write down the equations of regression lines in a bivariate distribution. What are its uses? Why there are two lines of regression? When do the two-lines coincide?

(b) The ranking of ten students in Mathematics and Statistics are as follows:

Mathematics	6	5	3	10	2	4	9	7	8	1
Statistics	3	8	4	9	1	6	10	7	5	2

Calculate Spearman's rank correlation coefficient. **(8 + 7 = 15)**

41. (a) Describe the method of fitting a straight line $y = a + bx$ by the principle of least squares.

(b) Explain partial and multiple correlations with examples. Write down the formula for multiple correlations in terms of total and partial correlations. With the usual notations if $r_{12} = 0.77, r_{23} = 0.52$ and $r_{13} = 0.72$, find $R_{1.23}$.

(6 + 9 = 15)

42. (a) Define data mining and OLAP. Explain the role of data mining in data ware housing.

(b) What is clustering? What are the different clustering techniques?

(c) Explain link analysis in data mining techniques. **(7 + 4 + 4 = 15)**

43. Explain the importance of R software in data analysis. Describe the computations of various measures of dispersion, skewness and kurtosis using R software.

(3 + 12 = 15)

44. Write short notes on the following: **(5 × 3 = 15)**

(a) Classification in data mining

(b) Nearest neighbor technique

(c) Yule's coefficient of association

(d) Computation of variance in excel

(e) Assignment operators in R

(2 × 15 = 30 Marks)