

Reg. No.:



Name:

University of Kerala

U8895

Second Semester FYUGP Degree Examination, April 2025

Discipline Specific Core Course

STATISTICS

UK2DSCSTA102 - QUANTITATIVE DATA ANALYTICS II

Academic Level: 100-199

Time: 2 Hours(120 Mins)

Max. Marks: 56

Part A.6 Marks:Time 5 Minutes.(Cognitive Level :Remember(RE)/Understand(UN)) Objective Type.1 mark each, Answer all questions

Qn No.	Question	CL	CO
1	When is a data said to be inconsistent in statistics of attributes?	RE	4
2	State any two elementary properties of probability.	RE	3
3	Which discrete probability distribution involves a fixed number of independent trials?	UN	3
4	Explain the meaning of zero correlation between the variables.	UN	1
5	What type of association exists when one attribute's presence reduces the occurrence of another? Options : A)Positive association B)Perfect association C)Negative association D)Independence	UN	4
6	State True or False – “Regression analysis can be used to predict the value of a dependent variable based on one or more independent variables.”	UN	2

Part B.10 Marks.Time:20 Minutes (Cognitive Level:Understand(UN)/Apply(AP))Two-three sentences.2 marks each.Answer all questions

Qn No.	Question	CL	CO																
7	How is a scatter diagram constructed?	UN	1																
8	Explain how Yule's coefficient of association can be used to test the independence between two attributes. What value of the coefficient indicates independence?	UN	4																
9	Write down the probability mass function of a binomial distribution with mean 3 and variance 2.	AP	3																
10	The regression lines are $2x + y = 1$ and $x + 3y = 6$. Estimate the value of y when $x = 4$.	AP	2																
11	Let X be a random variable with the following probability mass function. Determine the value of 'a'. Also obtain $P(X = 2)$. <table border="1"><tr><td>x</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>P(X=x)</td><td>a</td><td>3a</td><td>5a</td><td>7a</td><td>9a</td><td>11a</td><td>13a</td></tr></table>	x	0	1	2	3	4	5	6	P(X=x)	a	3a	5a	7a	9a	11a	13a	AP	3
x	0	1	2	3	4	5	6												
P(X=x)	a	3a	5a	7a	9a	11a	13a												

Part C.16 Marks.Time:35 Minutes.(Cognitive Level :Apply(AP)/Analyse(AN))Short Answer.4 marks each, Answer all 4 questions,choosing among options * within each question

Qn No.	Question	CL	CO
12	A) A box contains 10 balls, of which 6 are red and 4 are blue. A sample of 5 balls are selected at random. Find the probability that the sample contain (i) 3 red and 2 blue balls (ii) 1 blue and 4 red balls. OR B) At a job fair, 80 candidates apply for different job roles. Among them, 30 applied for technical roles, 25 applied for marketing roles and 25 applied for administrative roles. Each candidate applied for only one type of job. If one candidate is selected at random, what is the probability that (i) the person applied for a technical or marketing role? (ii) the person applied for a marketing or administrative role?	AP	3, 3
13	A) Given that the correlation coefficient between variables X and Y is $r = 0.75$, apply the properties of correlation to determine the correlation coefficient in the following cases: (i) Between 2X and Y (ii) Between X and 5Y (iii) Between 3X and 4Y OR B) Given a dataset of 20 students, the relationship between study hours (X) and exam scores (Y) is summarized as follows: $\sum X = 100$, $\sum Y = 1600$, $\sum X^2 = 580$, $\sum Y^2 = 130,000$, $\sum XY = 8400$. Apply the formula for Pearson's correlation coefficient to calculate the strength and direction of the relationship between study hours and exam scores. Interpret the result so obtained.	AP	1, 1
14	A) Compare and contrast Yule's coefficient of association and coefficient of colligation. OR B) In a survey of 2,980 people, 230 individuals were smokers and had respiratory issues, 260 were non-smokers with respiratory issues, and 2,340 were non-smokers without respiratory issues, the rest were smokers without respiratory issues. Analyze the relationship between smoking and respiratory issues by (i) identifying the marginal and joint frequencies and (ii) completing the 2×2 contingency table of attributes.	AN	4, 4
15	A) Can $3x + 4y = 20$ and $5x - 7y = 10$ be the two regression lines? Justify your answer. OR B) A company studies the relationship between the number of advertisements (X) and the number of units sold (Y) for 8 different products. The following summary statistics were obtained: $\sum x = 48$, $\sum y = 80$, $\sum xy = 640$, $\sum x^2 = 520$, $\sum y^2 = 800$. Using simple linear regression, estimate the number of units sold when the number of advertisements is 12.	AN	2, 2

Part D.24 Marks.Time: 60 Minutes.(Cognitive Level :Analyse(AN)/Evaluate(EV)/Create(CR)) Long Answer 6 Marks each.Answer all 4 questions choosing among options * within each question

Qn No.	Question	CL	CO																											
16	A) The mean and variance of a normal distribution are 30 and 25 respectively. Analyze the probability distribution of X by determining the probabilities of (i) $X > 35$, (ii) $20 \leq X \leq 32$ OR B) A group consists of 4 men and 3 women. A selection of 3 people is to be made from this group. Analyse the selection probabilities by calculating the probability of (i) exactly two women (ii) no women (iii) at least one woman (iv) at least 2 women (v) at the most 2 women.	AN	3, 3																											
17	A) Out of 1,000 citizens, 600 attended a health awareness seminar. Among them, 150 got a preventive health check-up later. Among those who didn't attend, only 50 got a check-up. Assess the association between seminar attendance and health check-ups using Yule's coefficient of association. Based on the calculated coefficient, interpret whether attending the seminar has a positive, negative, or no significant effect on the likelihood of getting a health check-up. OR B) In a locality, 1,482 individuals were exposed to smallpox. Among them, 343 had been vaccinated and only 35 of the vaccinated individuals were attacked. In total, 368 individuals were attacked. Based on this information, assess whether vaccination can be considered an effective preventive measure against smallpox.	EV	4, 4																											
18	A) A career advisor ranked 8 students based on leadership skills (X) and final placement outcomes (Y). (i) Calculate the Spearman's rank correlation coefficient. (ii) Evaluate the relationship between leadership ranking and placement outcomes. What practical conclusions can you draw from your findings? <table border="1"><thead><tr><th>Student</th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th><th>H</th></tr></thead><tbody><tr><td>X</td><td>1</td><td>4</td><td>3</td><td>6</td><td>2</td><td>7</td><td>5</td><td>8</td></tr><tr><td>Y</td><td>2</td><td>1</td><td>5</td><td>4</td><td>3</td><td>6</td><td>7</td><td>8</td></tr></tbody></table> OR B)	Student	A	B	C	D	E	F	G	H	X	1	4	3	6	2	7	5	8	Y	2	1	5	4	3	6	7	8	EV	1, 1
Student	A	B	C	D	E	F	G	H																						
X	1	4	3	6	2	7	5	8																						
Y	2	1	5	4	3	6	7	8																						

Qn No.	Question	CL	CO																
	A company wants to analyze whether increasing its advertising budget leads to higher sales revenue. The marketing team collects data over four months, tracking the amount spent on advertising (X, in thousands of dollars) and the corresponding sales revenue (Y, in thousands of dollars). <table><tr><td>X</td><td>10</td><td>20</td><td>30</td><td>40</td></tr><tr><td>Y</td><td>15</td><td>25</td><td>35</td><td>45</td></tr></table> (i) Compute Karl Pearson’s correlation coefficient for the given data. (ii) Critically evaluate the implications of the correlation value. Can we conclude that increasing the advertising budget will always lead to higher sales revenue?	X	10	20	30	40	Y	15	25	35	45								
X	10	20	30	40															
Y	15	25	35	45															
19	A) For the following set of observations: (2, 5), (3, 7), (4, 9), (5, 10), (6, 12), (7, 14), (8, 16), (9, 18). (i) Develop the two regression lines for the given data. (ii) Using the appropriate regression line, estimate the corresponding y value when x = 12. OR B) An event organizer wants to predict the number of attendees (Y, in hundreds) based on the number of advertisements published (X). Using the provided event data, (i) Construct a simple linear regression model to predict the number of attendees based on advertisements. (ii) Forecast the number of attendees if 20 advertisements are published. <table><tr><td>X</td><td>2</td><td>4</td><td>6</td><td>8</td><td>10</td><td>12</td><td>14</td></tr><tr><td>Y</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr></table>	X	2	4	6	8	10	12	14	Y	1	2	3	4	5	6	7	CR	2, 2
X	2	4	6	8	10	12	14												
Y	1	2	3	4	5	6	7												