KUMARAGURU
college of technology
character is life

Register Number:……..…

MCA DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2020)

First Semester

MASTER OF COMPUTER APPLICATIONS

P20MAT1101**:** Probability and Statistics for Data Analysis

(Statistical table can be permitted)

COURSE OUTCOMES

CO1: Understand about data collection, represent data graphically using bar chart and pie chart and compute various measures of central tendency and dispersion for analysis of data.

CO2: Interpret the correlation between variables and predict unknown values using regression.

CO3: Explore random variables and predict probabilities for situations following normal distribution.

CO4: Perform hypothesis testing using large sample tests and Chi square test and interpret the results which will form the basis for data analysis.

CO5: Understand the principles of design of experiments and perform analysis of variance.

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. If the mean of five observations x, $x + 4$, $x + 6$, $x + 8$ and $x + 12$ is 16 then the value of x is — CO1 [K_2]

   a) 6 b) 10

   c) 12 d) 5

2. The profits (in '000 Rs.) of a company for the last 8 years are given below. The coefficient of range is — CO1 [K_3]

Year	1975	1976	1977	1978	1979	1980	1981	1982
Profit	40	30	80	100	120	90	200	230

   a) 0.77 b) 0.87

   b) 0.88 d) 0.74

3. If r=0.8 and b_{xy}=1.2 SD_x=2.4 then the value of σ_y is — CO2 [K3]

   a) 1.6 b) -1.6

   c) 1.5 d) -1.2

4. Match the List I with List II with multiple choice code — CO2 [K_1]

List I (Type of correlation)	List II (Value of correlation coefficient 'r')
A) Strong positive correlation	i) $0 \leq r < 0.3$
B) Moderate positive correlation	ii) $0.8 \leq r \leq 1$
C) Weak positive correlation	iii) $0.3 \leq r < 0.5$
D) No correlation	iv) $0.5 \leq r < 0.8$

   a) A-iii, B-iv, C-ii, D-i b) A-ii, B-iii, C-iv, D-i

   c) A-ii, B-iv, C-iii, D-i d) A-iii, B-ii, C-iii, D-i

5. Examine the two statements carefully and select the answer using the codes given below. — CO3 [K_4]

Assertion (A) : The cumulative distribution function of a discrete random variable is always non-decreasing.
Reason (R) : The cumulative distribution function of a discrete random variable gives the probability of a value of the variable being less than or equal to a certain value.

a) Both A and R are individually true and R is the correct explanation of A
b) A is true and R is false
c) Both A and R are individually true and R is not the correct explanation of A
d) A is false and R is true

6. A continuous random variable X that can assume any value between $x = 2$ and $x = 5$ has the density function $f(x) = k(1 + x)$. The value of k is — CO3 [K_3]

a) $\frac{1}{27}$
b) 27
c) $\frac{2}{27}$
d) $\frac{27}{2}$

7. Which of the following statements are true under χ^2 distribution? — CO4 [K_2]

i) The shape of the distribution depends on n.
ii) The mean and variance are n and n^2 respectively.
iii) As $n \to \infty$, the χ^2-distribution approaches a normal distribution.
iv) The sum of independent χ^2-variates is not a χ^2-variate

a) i) & ii)
b) i) & iii)
c) i) & iv)
d) ii) & iv

8. Let $\overline{x_1}$ and $\overline{x_2}$ be the means of two independent large samples of size n_1 and n_2 drawn from two normal populations with means μ_1 and μ_2 and variances ${\sigma_1}^2$ and ${\sigma_2}^2$ respectively. The test statistic is — CO4 [K_L]

a) $\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \sim N(0,1)$
b) $\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}} \sim N(0,1)$
c) $\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} \sim N(0,1)$
d) $\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}}} \sim N(0,1)$

9. The Main aim of design of experiment is to control the _____ — CO4 [K_1]

a) Extraneous variables
b) Experiment variables
c) Local Control
d) Blocks

10. Replication means _____ — CO1 [K_1]

a) Grouping
b) Control
c) Repetition
d) Balancing

PART B (10 x 2 = 20 Marks)

11. The wheat production (in Kg) of 20 acres is given below. — CO1 [K_3]
1120,1240, 1320, 1040, 1080, 1200, 1440, 1360, 1680, 1730, 1785, 1342, 1960, 1880, 1755, 1720, 1600, 1470, 1750, 1885. Determine the quartile deviation of the data.

12. Draw a Pie diagram to represent the following data of proposed expenditure by a state government for the year 2024-2025. — CO1 [K_3]

Items	Agriculture & Rural Development	Industry & Urban Development	Health & Education	Miscellaneous
Expenditure (in million Rs.)	4200	1500	1000	500

13. Using scatter diagram explain perfect positive correlation, perfect negative correlation, no correlation, strong positive correlation and strong negative correlation. CO2 [K_2]

14. In a partially destroyed laboratory record of an analysis of a correlation data, the following results only are legible. Regression equations are 8x – 10y +66 = 0 ; 40x – 18y =214 Determine the mean values of X and Y. CO2 [K_3]

15. A discrete random variable X has the following probability distribution. CO3 [K_3]

X	0	1	2	3	4	5	6	7	8
P(X)	a	3a	5a	7a	9a	11a	13a	15a	17a

Determine the value of 'a' and hence derive the cumulative distributive function.

16. The mileage in thousands of miles in which car owners get with a certain kind of tyre is a random variable having p.d.f $f(x) = \frac{1}{20}e^{-x/20}$, for $x \geq 0$. What is the probability that one of these tyres will last atmost 10,000 miles? CO3 [K_3]

17. What do you mean by type I error and type II error in the concept of hypothesis test? CO4 [K_2]

18. What are the uses of χ^2- distribution? CO4 [K_1]

19. State any two differences between Randomized Block Design & Latin Square Design. CO5 [K_2]

20. Is it possible to analyze 2 × 2 Latin square design? Justify your answer. CO5 [K_2]

PART C (10 x 5 = 50 Marks)

21. Calculate the mean, median and mode of the following data. CO1 [K_5]

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	7	7	14	17	15	8	2

22. The number of finished articles turned out per day by different number of workers in a factory is given in the following table. Calculate the standard deviation of the daily output of finished articles. CO1 [K_5]

No. of articles	18	19	20	21	22	23	24	25	26	27
No. of workers	3	7	11	14	18	17	13	8	5	4

23. Find Karl Pearson' correlation coefficient for the following heights in inches of fathers (X) and their sons (Y) CO2 [K_3]

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

24. You are given the following information about advertising and sales of a company in a particular year. Estimate the likely sales when advertisement expenditure is Rs. 15 lakhs CO2 [K_5]

	Advertisement Expenses (Rs.lakhs) (X)	Sales (Rs.lakhs) (Y)
Average Price	10	90
S.D. of price	03	12
Correlation coefficient is 0.8		

25. When a die is thrown, X denotes the number that throws up. Determine the expectation of X, variance of X and standard deviation of X. CO3 [K_3]

26. In a test of 2000 electric bulbs it was found that the life of a particular make was normally distributed with an average life of 2040 hours and SD of 60 hours. Estimate the number of bulbs likely to burn for more than 1920 hrs but less than 2160 hrs. CO3 [K_3]

27. The mean breaking strength of the cables supplies by a manufacturer is 1800 with an SD of 100. By a new technique in the manufacturing process, it is claimed that the breaking strength of the cable has increased. To test this claim a sample of 50 cables is tested and it is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance. CO4 [K_3]

28. Two researchers adopted different sampling techniques while investigating the same group of students to find the number of students falling into different intelligence level. The results are given in the following table. Using Chi square test, verify whether the sampling techniques adopted by the two researchers are significantly different. CO4 [K_3]

Researcher	Below average	Average	Above average	Excellent	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

29. Explain the basic principles involved in design of experiments and the conditions for performing ANOVA test. CO5 [K_2]

30. A Randomized Design experiment with ten plots and three treatments gave the results given below. Analyze the results for the effects of treatments. CO5 [K_3]

Plot No.	1	2	3	4	5	6	7	8	9	10
Treatment	A	B	C	A	C	C	A	B	A	B
Yield	5	4	3	7	5	1	3	4	1	7

Answer any TWO Questions
PART D (2 x 10 = 20 Marks)

31. An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the following results: (10) CO1 [K_3]

	Firm A	Firm B
No. of workers	500	600
Average daily wages (Rs.)	186	175
Variance of distribution of wages	81	100

i) Which firm , A or B has a larger wage bill?
ii) In which firm, A or B, is there greater variability in individual wages?
iii) Calculate the average daily wage of all workers.
iv) the variance of the distribution of wages of all the workers in the firms A and B taken together.

32. Ten competitors in a music competition are ranked by three judges in the following order:Using rank correlation coefficient, determine which pair of judges have common taste in music. (10) CO2 [K_3]

Competitor	1	2	3	4	5	6	7	8	9	10
Judge A	1	6	5	10	3	2	4	9	7	8
Judge B	3	5	8	4	7	10	2	1	6	9
Judge C	6	4	9	8	1	2	3	10	5	7

33. The figures in the following 5x5 Latin Square are the numbers of minutes, the engines E_1, E_2, E_3, E_4 & E_5 tuned up by mechanics M_1, M_2, M_3, M_4 & M_5 ran with gallon of fuel A, B, C, D and E. Use Analysis of Variance technique to check whether there is a significant difference between the performance of the mechanics , engines and fuels.. Use the level of significance $\alpha = 1$ % to test. (10) CO5 [K_3]

	E_1	E_2	E_3	E_4	E_5
M_1	A31	B24	C20	D20	E18
M_2	B21	C27	D23	E25	A31
M_3	C21	D27	E25	A29	B21
M_4	D21	E25	A33	B25	C22
M_5	E21	A37	B24	C24	D20

****