KUMARAGURU
college of technology
character is life

MCA DEGREE EXAMINATIONS: NOV/DEC 2024

(Regulation 2020)

First Semester

MASTER OF COMPUTER APPLICATIONS

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

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

   b)   12   d)   5

2. Arrange the following steps involved in box plot technique.   CO1 [K$_1$]

   i) Determine the lower and upper fences:

   ii) Draw vertical lines at Q_1, M, Q_3. Enclose these vertical lines in a box.

   iii) Label the lower and upper fences.

   iv) Draw whiskers.

   v) Identify the outliers

   a)   i-iii-iv-iii-v   b)   i-iii-ii-iv-v

   b)   iv-i-iii-ii-v   d)   ii-i-iii-v-iv

3. Match the List I with List II   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

4. If r=0.8 and b_{xy}=1.2 SD$_x$=2.4 then the value of σ_y is   CO2 [K$_3$]

   a)   1.6   b) -1.6

   c)   1.5   d) -1.2

5. Examine the two statements carefully and select the answer using the codes given below.   CO3   [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 R is the correct explanation of A

   b) A is true and R is false

   c) Both A and R are individually true 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   [K3]

   a) $\dfrac{1}{27}$

   b) $\dfrac{1}{27}$

   c) $\dfrac{2}{27}$

   d) $\dfrac{2}{27}$

7. 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 σ_1^2 and σ_2^2 respectively. The test statistic is   CO4   [K1]

   a) $\dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \sim N(0,1)$

   b) $\dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}}} \sim N(0,1)$

   c) $\dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$

   d) $\dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{1}{\sigma_1^2} - \dfrac{1}{\sigma_2^2}}} \sim N(0,1)$

8. In testing of hypothesis about mean, the 95% confidence interval for μ is given by   CO4   [K1]

   a) $\left(\overline{x} - 2.58\dfrac{\sigma}{\sqrt{n}}, \overline{x} + 2.58\dfrac{\sigma}{\sqrt{n}}\right)$

   b) $\left(\overline{x} - 1.96\dfrac{\sigma}{\sqrt{n}}, \overline{x} + 1.96\dfrac{\sigma}{\sqrt{n}}\right)$

   c) $\left(\mu - 1.96\dfrac{\sigma}{\sqrt{n}}, \overline{x} + \mu\dfrac{\sigma}{\sqrt{n}}\right)$

   d) $\left(\overline{x} - \mu\dfrac{\sigma}{\sqrt{n}}, \overline{x} + \mu\dfrac{\sigma}{\sqrt{n}}\right)$

9. Which of the following statements are not true in ANOVA?   CO5   [K2]

   i) The observations are dependent.

   ii) The parent population is normal

   iii) Various treatment and environmental effects are additive in nature.

   iv) The samples have not been randomly selected from the population.

   a) i), iv)

   b) i), ii), iii)

   c) ii), iii), iv)

   d) iii) , iv)

10. If X is a random variable with density function $f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$ then E(X) is   CO3   [K3]

    a) 0.5

    b) 1

    c) 0.3

    d) −1

PART B (10 x 2 = 20 Marks)

11. Draw a Pie diagram to represent the following data of proposed expenditure by a state government for the year 2025-2026. — CO1 [K3]

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

12. The wheat production (in Kg) of 20 acres is given below. — CO1 [K3]
1120,1240, 1320, 1040, 1080, 1200, 1440, 1360, 1680, 1730, 1785, 1342, 1960, 1880, 1755, 1720, 1600, 1470, 1750, 1885. Determine the quartile deviation.

13. Using scatter diagram explain perfect positive correlation, perfect negative correlation, no correlation, strong positive correlation and strong negative correlation. — CO2 [K2]

14. If r=0.8 and b_{xy}=1.2 SD_x=2.4, compute the value of b_{yx} — CO2 [K3]

15. A discrete random variable X has the following probability distribution. Determine the value of 'a' — CO3 [K3]

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

16. If X and Y are independent random variables with variance 2 & 3 respectively then calculate $Var(3X + 4Y)$ — CO3 [K3]

17. List the uses of χ^2- distribution. — CO4 [K2]

18. Define type I error and type II error in the concept of hypothesis test. — CO4 [K2]

19. State the basic principles of design of experiments. — CO5 [K1]

20. State any two differences between Randomized Block Design & Latin Square Design. — CO5 [K2]

PART C (6 x 5 = 30 Marks)

21. Calculate the mean, median and mode of the following data. — 5 CO1 [K5]

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. — 5 CO1 [K5]

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. Calculate the rank correlation coefficient between marks in the selection test (X) and the proficiency test (Y) of 9 recruits. — 5 CO2 [K3]

X	10	15	12	17	13	16	24	14	22
Y	30	42	45	46	33	34	40	35	39

24. When a die is thrown, X denotes the number that throws up. Determine the expectation of X — 5 CO3 [K3]

25. An intelligence test given to two groups of students gave the following information — 5 CO4 [K3]

	Mean score	S.D.	Number
Group 1	75	10	50
Group 2	70	12	100

Is the difference in the mean scores of boys and girls statistically significant?

26. A completely randomized design experiment with ten plots and three treatments gave the results given below. Analyze the results for the effects of treatments. — 5 CO5 [K3]

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 FOUR Questions
PART D (4 x 10 = 40 Marks)

27. The prices of two commodities A & B over 10 weeks are given below. Which price shows less variation?   10   CO1   [K_3]

A	54	55	53	56	52	52	58	49	50	51
B	108	107	105	106	105	103	102	104	104	101

28. You are given the following information about advertising and sales of a company in a particular year.   10   CO2   [K_3]
    i)Estimate the likely sales when advertisement expenditure is Rs. 15 lakhs and
    ii)What should be the advertisement expenditure if the company wants to attain sales target of Rs. 120 lakhs?

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

29. 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 2040hrs and SD of 60 hrs. Estimate the number of bulbs likely to burn for (i) more than 2150 hrs (ii) less than 1950hrs and (iii) more than 1920 hrs but less than 2160hrs.   10   CO3   [K_3]

30. 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.   10   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

31. 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_5]

	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

****