Register Number:……..

KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2022

(Regulation 2018)

Third Semester

Common To All Branches Except CSE/ IT/ ISE/ AI&DS

U18MAR0202 – Statistical Foundations for Data Science

COURSE OUTCOMES

CO1:	Understand about data collection, represent data graphically using bar chart and pie chart.
CO2:	Compute various measures of central tendency and dispersion for analysis of data. Interpret the correlation between variables and predict unknown values using regression.
CO3:	Perform hypothesis testing using large sample tests and Chi square test and interpret the results, which will form the basis for data analysis.
CO4:	Understand the principles of design of experiments and perform analysis of variance.
CO5:	Learn and apply multivariate analysis necessary for Principal Component Analysis.
CO6:	Use R software to solve problems in the above topics.

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

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

1.	Explain about data collection and its types	CO1	[K_1]
2.	Compare primary and secondary data	CO1	[K_1]
3.	There are two branches of a company employing 280 and 320 persons respectively. If the A.M. of the salaries of the two branches are Rs. 750 and Rs.937.5 respectively, find the A.M. of the salaries of the employees of the company as a whole.	CO2	[K_2]
4.	Explain about Spearman's rank correlation coefficient.	CO2	[K_2]
5.	State type I and type II error.	CO3	[K_1]
6.	Write the test statistics for testing of hypothesis about difference between two means.	CO3	[K_2]
7.	List the basic principles of Experimental design.	CO4	[K_1]
8.	Write the advantages of completely randomized design.	CO4	[K_1]
9.	Define Correlation matrix.	CO5	[K_2]
10.	Define Random vector.	CO5	[K_2]

Answer any FIVE Questions: -
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11.	a)	Represent the expenditures on various items of a family by a pie chart. **Items** / **Expenditure (in 100 rupees)** Food — 50 Clothing — 30 Rent — 20 Fuel — 15 Misc. — 35	8	CO1	[K_3]
	b)	Temperature of 20 winter days in Pakistan is recorded below: 24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27. Construct frequency distribution.	8	CO1	[K_3]
12.	a)	Calculate mean, median and mode for the following data: Less than: 10 20 30 40 50 60 70 Frequency: 7 14 28 45 60 68 70	8	CO2	[K_4]
	b)	Find Karl Pearson' correlation coefficient for the following heights in inches of fathers (x) and their sons (y): X: 65 66 67 67 68 69 70 72 Y: 67 68 65 68 72 72 69 71	8	CO2	[K_4]
13.	a)	(i) Ten oil cans are taken at random from an automatic filling machine. The mean weight of the tins is 15.8 kg and standard deviation is 0.5 kg. Does the sample mean differ significantly from the intended weight of 16 kg?	8	CO3	[K_3]
	b)	Two random samples gave the following results: Sample 1: Size 10, Sample mean 15, Sum of the square of deviations from the mean 90 Sample 2: Size 12, Sample mean 14, Sum of the square of deviations from the mean 108 Examine whether the samples come from the same normal population at 5% level of significance.	8	CO3	[K_3]
14		The following data represents the number of units of productions per day turned out by different workers using four different types of machines. Machine Type	16	CO4	[K_4]

			A	B	C	D			
		1	44	38	47	36			
	Workers	2	46	40	52	43			
		3	34	36	44	32			
		4	43	38	46	33			
		5	38	42	49	39			

Test whether the 5 men differ with respect to mean productivity and test whether the mean productivity is the same for the four different machine types.

15.		Let X1 and X2 have the joint probability mass function $P(x_1, x_2) = \frac{x_1+2x_2}{18}$, $x_1 = 1,2$, $x_2 = 1,2$. Find (i) Marginal probability mass functions of x_1 and x_2 (ii) Mean Vector (iii) Variance-Covariance matrix (iv) population correlation matrix (v) Standard deviation $V^{1/2}$.	16	CO5	[K_3]
16.	a)	(i) The weights of 20 randomly selected juniors at MSHS are recorded below:	8	CO2	[K_4]

121	126	130	132	143	137	141	144	148	205
125	128	131	133	135	139	141	147	153	213

a) Construct a boxplot of the data

b) Determine if there are any mild or extreme outliers.

	b)	The mean lifetime of a sample of 100 light tubes produced by a company is found to be 1580 hours with standard deviation of 90 hours. Test the hypothesis that the mean lifetime of the tubes produced by the company is 1600 hours.	8	CO3	[K_3]

****