



B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Third Semester

COMMON TO ALL

U18MAR0201: 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.

Time: Three Hours

Maximum Marks: 100

Answer all the Questions: -

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

- | | | | |
|-----|---|-----|-------------------|
| 1. | What do you understand by Qualitative data and quantitative data? | CO1 | [K ₁] |
| 2. | Distinguish between primary data and secondary data. | CO1 | [K ₁] |
| 3. | The mean marks of 100 students were found to be 40. Later it was discovered that a score of 53 was misread as 83. Find the corrected mean corresponding to the corrected score. | CO2 | [K ₂] |
| 4. | If the regression coefficient of X on Y is 0.8 and the regression coefficient of Y on X is 0.2, then find the correlation coefficient. | CO2 | [K ₂] |
| 5. | Define Type I Error & Type II Error. | CO3 | [K ₁] |
| 6. | State any two applications of χ^2 (Chi-square)-test. | CO3 | [K ₁] |
| 7. | State the basic principles of design of Experiments | CO4 | [K ₁] |
| 8. | Write any two differences between RBD and CRD. | CO4 | [K ₁] |
| 9. | Define random vector. | CO5 | [K ₁] |
| 10. | Write down the formula of Variance–Covariance. | CO5 | [K ₁] |

Answer any FIVE Questions: -

PART B (5 x 16 = 80 Marks)

(Answer not more than 400 words)

- | | | | | | |
|-----|----|---|---|-----|-------------------|
| 11. | a) | Explain various types of Measurement Scales with an example. | 8 | CO1 | [K ₂] |
| | | The following table shows the time taken (in minutes) by 100 students to travel to school on a particular day | 8 | CO1 | [K ₂] |

b)

Time	0-5	5-10	10-15	15-20	20-25
No. of Students	5	25	40	17	13

Draw the histogram. Also find:

- (i) The number of students who travel to school within 15 minutes.
(ii) Number of students whose travelling time is more than 20 minutes.

12. a) A study of 100 engineering companies gives the following information. Calculate the standard deviation of the profit earned. 8 CO2 [K₂]

Profit (in crore)	0-10	10-20	20-30	30-40	40-50	50-60
No. of companies	8	12	20	30	20	10

- The following data gives the heights (in inches) of father (X) and his eldest son (Y). Compute the correlation coefficient between the heights of fathers and sons using Karl Pearson's method. 8 CO2 [K₂]

- b) X : 65 66 67 67 68 69 70 72
Y : 67 68 65 68 72 72 69 71

13. a) A company producing LED bulbs finds that mean life span of the population of its bulbs is 2000 hours with a standard derivation of 150 hours. A sample of 100 bulbs randomly chosen is found to have the mean life span of 1950 hours. Test, at 5% level of significance, whether the mean life span of the bulbs is significantly different from 2000 hours. 8 CO3 [K₃]

- b) A company gave an intensive training to its salesmen to increase the sales. A random sample of 10 salesmen was selected and the value (in lakhs of Rupees) of their sales per month, made before and after the training is recorded in the following table. Test whether there is any increase in mean sales at 5% level of significance. 8 CO3 [K₃]

Before	15	22	6	17	12	20	18	14	10	16
After	17	23	16	20	14	21	18	20	10	11

14. a) Two random samples gave the following data: 8 CO3 [K₃]

Sample	Size	Mean	Variance
I	8	9.6	1.2
II	11	16.5	2.5

Can we conclude that the two samples have been drawn from the same normal population?

- b) The following table gives the classification of 100 workers according to the sex and nature of work. Test whether nature of work is independent of the sex of the worker. 8 CO3 [K₃]

Sex / Nature of Work	Skilled	Unskilled	Total
-------------------------	---------	-----------	-------

Male	40	20	60
Female	10	30	40
Total	50	50	100

15. a) A reputed marketing agency in India has three different training programs for its salesmen. The three programs are Method – A, B, C. To assess the success of the programs, 4 salesmen from each of the programs were sent to the field. Their performances in terms of sales are given in the following table. Test whether there is significant difference among methods and among salesmen. 16 CO4 [K₃]

Salesman	Methods		
	A	B	C
1	4	6	2
2	6	10	6
3	5	7	4
4	7	5	4

16. a) Compute the principal component to the following variance covariance matrix 12 CO5 [K₃]

$$\Sigma = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix}$$
- b) Explain the principal components and population principal components 4 CO5 [K₁]
