INDIAN STATISTICAL INSTITUTE

POST GRADUATE DIPLOMA IN BUSINESS ANALYTICS - FIRST YEAR
SEMESTER EXAMINATION
FIRST SEMESTER, 2025-26

Date : 10.12.2025

STATISTICAL STRUCTURES IN DATA : PART A

TIME: $1\frac{1}{2}$ HOURS FULL MARKS: 40

Answers should be brief and to the point. If you use any result proved in class, clearly state the result before using it. This question paper carries 44 points. Answer as many as you can. The maximum you can score in 40.

1. (a) Consider a data set containing 11 observations $0, 1, 2, 3 \ldots, 10$. Is it possible to add two new observations to this data set such that both the mean and the variance remain unchanged? If you think it is possible, find the two observations. If you think it is not possible, justify your answer. [6]

   (b) If the dispersion matrix of a three-dimensional random vector is given by

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix},$$

   show that ρ cannot be smaller than -0.5. [4]

2. Prove or disprove the following statements. [$4 \times 6 = 24$]

   (a) The moment based measure of kurtosis $\beta_2 = m_4/m_2^2$ (where m_r denotes the r-th central moment) always lies between 0 and 1.

   (b) If the product moment correlation coefficient computed from two data sets are r_1 and r_2, respectively, then the same computed from the pooled data set must lie between r_1 and r_2.

   (c) In a simple linear regression problem, it is not possible to have the two regression lines (y on x and x on y) as (i) $2x + 3y = 7$ and $4x - 6y = 2$ or (ii) $2x + 3y = 7$ and $4x + 6y = 15$.

   (d) If $x_i = 2^{-i}$ and $y_i = 2i - 1$ for $i = 1, 2, \ldots, 10$, the function $\psi(\theta) = \sum_{i=1}^{10} |y_i - \theta x_i|$ has a unique minimizer at $\theta = 2$.

   (e) In a data set containing 100 bivariate observations, if X takes only two distinct values 1 and 2, the two regression models $Y = \beta_0 + \beta_1 X$ and $Y = \alpha_0 X + \alpha_1 X^2$ will lead to the same residual sum of squares.

   (f) Consider a problem of minimizing a strictly convex objective function. If the majorization-minimization algorithm is used for this purpose, the observed sequence of minimizers obtained during iterations will always converge to the global minima.

3. Suppose that $\mathbf{X} = (X_1, X_2, X_3)$ follows a multivariate normal distribution with the mean $(1, 2, 3)$ and the dispersion matrix Σ, whose all diagonal elements are 4. prove the following results.

   (a) If the variance of the first principal component is 10, show that the determinant of the dispersion matrix cannot be bigger than 10. [4]

   (b) If the off-diagonal elements are also equal and positive, find the distribution of the second principal component. [6]