1511E1235 Candidate's Seat No :________

M.Sc. Sem.-3 Examination

503

Statistics

Time : 2-30 Hours] **November-2025** **[Max. Marks : 70**

Note: Attempt all questions.

Q.1

(i) Define multivariate normal distribution and obtain its characteristic function. [7]

(ii) If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{C}$: $m \times p$ be any arbitrary matrix then show that

$$\mathbf{Y}_{m\times 1} = \mathbf{C}_{m\times p}\mathbf{X}_{p\times 1} \sim N_m(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}') .$$ [7]

OR

(i) In usual notations, let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\mathbf{X}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ partition as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}\begin{matrix} r \\ s \end{matrix} \quad , \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}\begin{matrix} r \\ s \end{matrix} \quad \text{and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

$r + s = p$.

Show that (a) $\mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2$ and $\mathbf{X}_2$ are independently distributed.

(b) $\mathbf{X}_2 \sim N_s(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. [7]

(ii) Show that if $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then any sub-vector of $\mathbf{X}$ is also normally distributed with mean equal to the corresponding sub-vector of $\boldsymbol{\mu}$ and with variance-covariance matrix equal to the corresponding sub-matrix of $\boldsymbol{\Sigma}$. [7]

Q.2

(i) What do you understand by canonical correlation? Also write it's application. [7]

(ii)In usual notations, show that (a) $1 - R_{1.23}^2 = (1 - r_{12}^2)(1 - r_{13.2}^2)$ (b) $R_{1.23}^2 = r_{12}^2 + r_{13}^2$, if $r_{23} = 0$. [7]

OR

(i) If all the correlation coefficients of zero order in a set of p-variates are equal to ρ, show that (a) every partial correlation of s'th order is $\frac{\rho}{1 + s\rho}$ (b) the coefficient of multiple correlation R of a variate with the other $(p-1)$ variates is given by

$$1 - R^2 = (1 - \rho)\left[\frac{1 + (p-1)\rho}{1 + (p-2)\rho}\right].$$ [7]

(ii) Describe canonical correlation coefficients and canonical variates. In usual notation show that canonical correlation are solution of the determinant equation

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & -\lambda\Sigma_{22} \end{vmatrix} = 0 .$$ [7]

P.T.O

E-1235-2

Q.3

(i) Define Wishart distribution. How is a Wishart-distributed matrix constructed from multivariate normal variables? [7]

(ii) In usual notations, If $D \sim W_p(D \mid n \mid \Sigma)$, then for any arbitrary matrix $H : m \times p$, show that $HDH' \sim W_m(HDH' \mid n \mid H\Sigma H')$. [7]

OR

(i) If $V_1 \sim W_p(V_1 \mid n_1 \mid \Sigma)$ and $V_2 \sim W_p(V_2 \mid n_2 \mid \Sigma)$ are independently distributed then show that $V_1 + V_2 \sim W_p(V_1 + V_2 \mid n_1 + n_2 \mid \Sigma)$. [7]

(ii) If $D \sim W_p(D, n, \Sigma)$, then obtain the distribution of $|D|/|\Sigma|$. Also obtain $E|D|^h$. [7]

Q.4

(i) Define principal components. When should you use principal component analysis? Also discuss limitations of principal component analysis. [7]

(ii) Let $\mathbf{\Sigma}$ be the covariance matrix associated with the random vector $X' = \lfloor X_1, X_2, \cdots, X_p \rfloor$. Let $\mathbf{\Sigma}$ have the eigenvalue-eigenvector pairs $(\lambda_1, e_1,), (\lambda_2, e_2), \cdots, (\lambda_p, e_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. The *i*th principal component is given by $Y_i = e_i' X = e_{1i} X_1 + e_{2i} X_2 + \cdots + e_{pi} X_p$, $i = 1, 2, \cdots, p$ .Then show that Var $(Y_i) = e_i' \Sigma e_i = \lambda_i$, and Cov $(Y_i, Y_k) = e_i' \Sigma e_k = 0,\ i \neq k$. [7]

OR

(i) Explain how Hotelling T^2 is used for testing equality of two mean vectors from two different multivariate normal populations with unknown but equal covariance matrices. [7]

(ii) What do you understand by discriminant analysis and classification? What are the key assumptions of discriminant analysis? Give some real-world applications of discriminant analysis. [7]

Q.5 Answer any seven: [14]

(i) Define population correlation matrix.

(ii) For a multivariate normal random vector, the variance-covariance matrix is always________.
(A) square matrix (B) non-negative definite (C) symmetric (D) all of these

(iii) Multivariate normal distribution is most commonly used in multiple statistical methods because:
(A) it is easier than other distributions
(B) it's assumptions holds due to multiple central limit theorem in all cases
(C) this distribution holds for all types of variables
(D) it converges to unity

E1235·3

(iv) The quantity $(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})$ involved in multivariate normal density function represents:
(A) multivariate normal density
(B) dispersion matrix
(C) exponential series
(D) Mahalanobis squared distance

(v) What is the primary goal of canonical correlation analysis?
(A) To predict a single dependent variable from multiple independent variables
(B) To find the correlation between one variable and a set of others
(C) To determine the relationships between two sets of multiple variables
(D) To identify factors within a single set of varaibles

(vi) Which of the following is a key output of canonical analysis?
(A) A single regression equation
(B) Correlation coefficients between individual variables
(C) A series of canonical variates, which are linear combinations of the original variables
(D) The mean and standard deviation of each variable

(vii) The _______ distribution is a multivariate generalization of chi-square distribution.
(A) Multivariate Normal (B) Hotelling's T^2
(C) Wishart distribution (D) none of these

(viii) The Wishart distribution is a family of distributions for _______ matrices
(A) symmetric positive definite (B) asymmetric positive definite
(C) symmetric negative definite (D) asymmetric negative definite

(ix) Give one limitation of Wishart distribution.

(x) How are the principal components in principal component analysis determined?
(A) By selecting the features with the lowest variance
(B) By finding the directions of maximum variance in the data
(C) Through a supervised learning process that labels each component
(D) By randomly selecting a subset of the original features

(xi) What is the purpose of eigenvalues in principal component analysis?
(A) They represent the variance explained by each principal component
(B) They determine the number of principal components to retain
(C) They are used for data visualization
(D) They measure the similarity between data points

(xii) Which of the following is a primary objective of discriminant analysis?
(A) To maximize the variance within each group
(B) To explain the variance in a continuous dependent variable using independent variables
(C) To develop linear combinations of predictor variables that best discriminate between groups
(D) To predict the values of continuous dependent variables

***