2021

STATISTICS — HONOURS

Paper : CC-14

(Multivariate Analysis and Non-Parametric Methods)

Full Marks : 50

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

1. Answer ***any five*** questions : 2×5

(a) Define a multivariate data set. Give an example.

(b) If $\underset{\sim}{x}_1,...,\underset{\sim}{x}_n$ denote n sample observations on a p-vector $\underset{\sim}{x}=\left(x_1,...,x_p\right)'$, where $x_1, x_2,...,x_p$ are p univariates, define sample mean vector $\bar{\underset{\sim}{x}}$. Show that it is the vector of sample means of $x_1, x_2,...,x_p$.

(c) What do you mean by a random p-vector $\underset{\sim}{X}=\left(X_1,...,X_p\right)'$?

(d) What is the main objective of Principal Component Analysis (PCA)?

(e) Mention a situation with justification when the dispersion matrix of a random vector is positive semi definite (p.s.d.).

(f) Discuss a real life application of multinomial distribution.

(g) Suppose a non-parametric test is proposed for testing the null hypothesis $H_0 : \theta = 10$, where θ denotes population median. Is H_0 a simple hypothesis or a composite one? Justify.

(h) Mention a merit and a demerit of Wilcoxon Signed Rank test over Ordinary Sign test.

Answer ***any two*** questions from questions **2–4**.

2. Show that a collection $\underset{\sim}{X}=\left(X_1,...,X_p\right)$, defined on ($\Omega$, $\mathcal{F}$); where Ω denotes reference set and $\mathcal{F}$, a sigma field from Ω, such that $\underset{\sim}{X} \in \mathbb{R}^p$, is a random p-vector if and only if $X_1, ..., X_p$ are p random univariate. 5

3. (a) Suppose $\sum_{p\times p} = Disp(\underset{\sim}{X})$, $\underset{\sim}{X}$ being a random p-vector. Prove that $\text{Disp}\left(A'\underset{\sim}{X}\right)=A'\sum A$, where $\underset{\text{A}}{\text{p×p}}$ is a fixed matrix with real elements.

(b) If $\sum$ and R denote the dispersion and correlation martix of a random p-vector, prove that $\sum$, positive definite $\Leftrightarrow R$, positive definite. 5

Please Turn Over

4. Discuss the mechanism of choosing principal components in a PCA. Explain why this mechanism helps in reducing the number of components of a random vector. 5

Answer ***any three*** questions from questions **5-9**.

5. (a) Discuss Wald-Wolfowitz Run test for testing whether two independent random samples are drawn from the same population or not.

(b) Based on a sample of distribution of digital ration cards to n families, describe briefly a test procedure for testing whether the cards have been distributed randomly among privileged and under-privileged families. 10

6. In the context of a random p-vector $\underset{\sim}{X} = \left(X_1, X_2, ..., X_p\right)', p \geq 2$, suppose $\eta_{1\cdot 23...p}$ denotes the true regression of X_1 on $X_2,...X_p$. Then show that

(a) $\text{cov}\left(X_j, \eta_{1\cdot 23...p}\right) = \sigma_{1j} = \text{cov}\left(X_1, X_j\right) \forall j = 2,3,...,p.$

(b) $\text{cov}\left(X_1, \eta_{1\cdot 23...p}\right) = V\left(\eta_{1\cdot 23...p}\right).$

(c) $\rho^2_{1\cdot 23...p} \leq \xi^2_{1\cdot 23...p}$, where $\rho_{1\cdot 23...p}$ denotes the multiple correlation of X_1 on $X_2,...,X_p$, and $\xi_{1\cdot 23...p} = \text{corr}\left(X_1, \eta_{1\cdot 23...p}\right).$ 10

7. If a random p-vector $\underset{\sim}{X} \sim N_p\left(\underset{\sim}{\mu}, \Sigma\right), \Sigma, p.d,$, then

(a) prove that $\left(\underset{\sim}{X} - \underset{\sim}{\mu}\right)' \Sigma^{-1} \left(\underset{\sim}{X} - \underset{\sim}{\mu}\right) \sim \chi^2_p$

(b) find the distribution of $\underset{\sim}{X}^{(2)} - \overset{s\times r}{A}\, \underset{\sim}{X}^{(1)}$ where $\underset{\sim}{X}_{p\times 1} = \left(\underset{\sim}{X}^{(1)'}_{r\times 1}, \underset{\sim}{X}^{(2)'}_{s\times 1}\right)'$, $\underset{\sim}{X}^{(1)}_{r\times 1}$ and $\underset{\sim}{X}^{(2)}_{s\times 1}$ being two partitions of $\underset{\sim}{X}_{p\times 1}, r + s = p, r, s \geq 1$, and $\overset{s\times r}{A}$ is a fixed matrix of rank s. 10

8. Suppose $\underset{\sim}{Y} = \left(Y_1, ..., Y_k\right) \sim \text{multinomial}\left(m; p_1, ..., p_k\right); \sum_{i=1}^{k} p_i < 1$

(a) Find the partial correlation coefficient $\rho_{12\cdot 34...k}$ between Y_1 and Y_2 eliminating the effects of $Y_3, ..., Y_k$.

(b) Get the distribution of $\sum_{i=1}^{k} Y_i$. 10

9. Suppose yield of paddy (X_1) is suggested to predict based on a linear function of amount of rainfall (X_2), amount of temperature (X_3), amount of irrigation (X_4) applied and amount of fertilizers (X_5) and pesticides (X_6) used. Assuming normality suggest a suitable test procedure, on the basis of n sample observations on (X_1, X_2,..., X_6), for testing.

(a) Whether the linear prediction function is useful or not for the purpose of prediction, and

(b) Whether the amount of irrigation (X_4) can be dropped from the linear prediction formula without making any significant change in the usefulness of the prediction formula. 10

———————