CSM – 68 / 15
Statistics
Paper – I

Time : 3 hours

Full Marks : 300

The figures in the right-hand margin indicate marks.

*Candidates should attempt Q. No. 1 from Section – A and Q. No. 5 from Section – B which are compulsory and **three** of the remaining questions, selecting at least **one** from each Section.*

Section – A

1. Attempt any **five** of the following sub-parts : 12×5 = 60

(a) Show that Poisson distribution can be obtained as an approximation to Binomial distribution.

(b) If $p(x) = \begin{cases} \frac{x}{15}, & x = 1, 2, 3, 4, 5 \\ 0 & \text{elsewhere} \end{cases}$

find (i) $P(X = 1 \text{ or } X = 2)$

(ii) $P\left(\frac{1}{2} < X < \frac{5}{2} \mid X > 1\right)$

Sketch the distribution function F(x).

(c) Given the joint pdf of (X, Y)

$$f(x, y) = \begin{cases} 8xy & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Obtain E(X), E(Y) and Cor (X, Y).

(d) Derive the characteristic function of the standard normal distribution.

(e) In a simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$, obtain the least squares estimation of β_0 and β_1. Stating the assumptions, derive the procedure to test for the significance of β_1.

(f) Define Hotelling's T^2-Statistic and Mahalanobis D^2-statistics. State the relationship between them. Mention any two applications of T^2-statistic.

2. (a) A bag contains 6 red, 5 white and 4 black balls. If two balls are drawn at random, find the probability that :

   (i) None of them is red

   (ii) One is white and one is black

(b) If X and Y are independent rvs show that :

(i) $E(cX + dY) = cE(X) + dE(Y)$, $c, d \in R$.

(ii) $V(aX - bY) = a^2 v(X) + b^2 v(Y)$, $a, b \in R$.

(c) Derive the moment generating function of a binomial distribution $B(n, p)$. Hence, derive its mean and variance. 20×3 = 60

3. (a) The joint pdf $f(x, y)$ is given by $f(x, y) = x^2 + \frac{xy}{3}$, $0 < x < 1$, $0 < y < 2$, find the marginal density function of X and Y, conditional density function of X given Y = y, $E(X \mid y)$ and $V(X \mid y)$.

(b) If $X_n \xrightarrow{P} X$ and g is a continuous function in red line R, then show that $g(X_n) \xrightarrow{P} g(X)$.

(c) Let $\{X_n\}$ be a sequence of independent rvs with $P(X_n = n^\lambda) = P(X_n = -n^\lambda) = \frac{1}{2}$. Examine for what values of λ :

(i) The central limit theorem holds

(ii) WLLN holds 20×3 = 60

4. (a) State the multiple linear regression model with the assumptions. Explain a procedure to estimate the parameters of the model. Define the coefficient of determination R^2 for this model.

(b) Given the simple correlations among the three variables $r_{12} = 0.77$, $r_{13} = 0.72$ and $r_{23} = 0.52$, find the partial correlation coefficient $r_{12.3}$ and the multiple correlation coefficient $R_{1.23}$.

(c) Derive the Bayesian Classification rule to classify an observation into one of the two multivariate normal population with equal covariance matrices. 20×3 = 60

Section – B

5. Answer any **three** of the following :

(a) (i) State Neyman Factorisation Theorem. Obtain the sufficient statistics for the parameters of the following distributions :

$$P(\lambda), \text{ and } N(\mu, \sigma^2)$$

(ii) Based on a random sample of size n from $f(x, \theta) = \theta e^{-\theta x}$, $x > 0$, derive the UMVUEs of $\frac{1}{\theta}$ and $\frac{1}{\theta^2}$. 10+10 = 20

(b) (i) Define monotone likelihood ratio property. Examine whether the following family possess this property :

$$N(\mu_0, \sigma^2), \sigma^2 > 0$$

(ii) Explain Wald's SPRT and describe the test proceduce for binomial proportion.

8+12 = 20

(c) (i) Explain a procedure for the sample size determination, for estimating the population mean under SRSWR.

(ii) Discuss the selection procedure for drawing a sample of n units from N units under PPSWR. 10+10 = 20

(d) (i) Explain fixed effects, random effects and mixed effects models with examples.

(ii) Explain the three basic principles and their importance in the design of experiments. 8+12 = 20

6. (a) Define complete statistics and show that ΣX_i is a complete statistics for the Poisson family of distributions.

(b) State Cramer-Rao inequality. Obtain the Cramer-Rao lower bound for the variance of an unbiased estimator of μ based on a sample of size n from $N(\mu, 1)$.

(c) Define a most powerful (MP) test. Construct an MP test of $H_0 : \theta = 1$ against $H_1 : \theta = 2$ based on a sample of size n from $N(\theta, 1)$.

(d) State Basu's Theorem on ancillary statistics.

Show that Y_n and $\frac{Y_1}{Y_n}$ are independent rvs, based on a random sample of size n from $V(0, \theta)$, where Y_r is the r^{th} order statistics.

(e) Explain the method of scoring for computing maximum likelihood estimators. 12×5 = 60

7. (a) State a Latin Square Design (LSD) with the assumptions. Describe the analysis of this model to test the significance of the relevant hypothesis. 12

(b) Describe the following non-parametric tests : 12

(i) Kolmogorov-Smirnov Test (One sample)

(ii) Mann-Whitney U test

(c) Let $X \sim N|\mu, \sigma^2)$. Construct the likelihood ratio test to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, when σ^2 is unknown. 12

(d) Distinguish between complete confounding and partial confounding in factorial experiment. Illustrate the layout of these designs in a 2^3 factorial experiment. 12

(e) (i) Explain Warner's randomised response technique.

(ii) Define Horvitz-Thompson estimator of the population total Y and derive its variance. 4+8 = 12

8. (a) Explain the role of auxiliary variables in ratio and regression methods of estimation. Show that regression estimator is more efficient than ratio estimator for estimating the population mean. 12

(b) (i) Distinguish between stratified sampling, cluster sampling and two-stage sampling.

(ii) In a cluster sampling with equal cluster size, suggest an unbiased estimator for the population mean and derive variance of the estimator. 6+6 = 12

(c) Explain the missing plot technique in design of experiment. Obtain an estimator for a single missing observation in a RBD and derive its variance. 12

(d) What are factorial experiments ? Mention the advantages. State the orthogonal contrasts for the main effects and the interaction effects in a 3^2 factorial experiment. Set up the ANOVA table for testing the relevant hypothesis. 12

(e) (i) Define BIBD and give a layout of this design. Show that BIBD is a connected design.

(ii) Define split plot experiment with whole plot treatments and split plot treatments within a whole plot, with r being number of replications. Outline the analysis of this design. 4+8 = 12