

STATISTICS

Paper-I

Time Allowed : Three Hours

Full Marks : 300

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

All questions carry equal marks.

SECTION—A

1. Attempt any FIVE of the following subparts :—

(a) Let $\Omega = \{1, 2, \dots, n\}$ and $\mathcal{F}$ be the family of all subsets of Ω with $P(\{i\}) = \alpha_i$, $i = 1, 2, \dots, n$. Extend the function P to a finitely additive probability measure defined on the whole field $\mathcal{F}$. What conditions have to be imposed on the number α_i ?

(b) Ten people are randomly seated at a round table. What is the probability that a particular couple will sit next to each other ?

(c) Let $\{X_n\}$ be a sequence of independent identically distributed random variables

following Uniform distribution on $[0, \theta]$ ($\theta > 0$) and $M_n = \max. (X_1, X_2, \dots, X_n)$. Verify whether $\{M_n\}$ converges in distribution to some random variable or not ?

- (d) Find the characteristic function of a random variable X having distribution with density :

$$f_x(x) = \begin{cases} 1-|x| & \text{if } |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

- (e) State Gauss-Markov theorem and explain its application in linear estimation.
- (f) What are principal components ? Obtain their relations with eigenvalues and eigenvectors of dispersion matrix. 5×12=60 Marks
2. (a) Show that $|P(A) - P(B)| \leq P(A \Delta B)$ for any two events A and B and a probability measure $P(\cdot)$.
- (b) A fair die is tossed. Let X denote 1 or 3 according as an odd or even number appears. Find the distribution, expectation and standard deviation of X .
- (c) Prove that the following conditions are equivalent :
- C1 : A, B, C are independent events.
- C2 : $\Omega-A, B, C$ are independent events.

C3 : Ω -A, Ω -B, C are independent events.

C4 : Ω -A, Ω -B, Ω -C are independent events.

3×20=60 marks

3. (a) Define probability generating function (pgf) of a random variable. If $P_x(t)$ is the pgf of X then obtain the pgf of $X + 2$.
- (b) State any form of the central limit theorem (CLT). Using CLT, estimate the probability that the number of the heads in 10,000 independent tosses of a fair coin differs by less than 1% from 5000.
- (c) What is exponential family of distributions ? Verify whether the following family belongs to the exponential family :
- (i) $X \sim b(n, p)$
 - (ii) $X \sim U [\alpha, \beta]$,
 - (iii) $X \sim C(1, \theta)$.

3×20=60 marks

4. (a) Based on a data in a study on savings and loan company, using a multiple regression model with two independent variables, where y = profit margin, x_1 = net revenue, x_2 = number of branch offices. The model equation suggested is,

$$y = 1.565 + 0.237x_1 - 0.0002x_2 + e.$$

- (i) Interpret the values of the regression coefficients.

- (ii) If the number of branch offices remain fixed and net revenue increases by 1, what is the average change in profit margin ?
- (iii) What would the mean profit margin be with a net revenue of 4.0 and 6500 branch offices ?
- (b) In an experiment to investigate the performance of four different brands of spark plugs for use on a 125-cc motorcycle, five plugs of each brand were tested and the number of miles (at a constant speed) until failure was observed. A partially completed ANOVA table is given. Fill in the missing entries and test the relevant hypotheses using a 0.05 level of significance :

Source of Variation	d.f.	Sum of squares	Mean square	F	Table Value of F
Treatments	-	-	-	-	3.24
Error	-	2,35,419.04	-		
Total	-	3,10,500.76			

- (c) Let $X'AX$ and $X'BX$ be two quadratic forms of a multivariate normal vector X , where A and B are real symmetric matrices. Obtain a necessary and sufficient condition for the independence of these two quadratic forms. $3 \times 20 = 60$ Marks

SECTION—B

Answer any FIVE sub-parts.

5. (a) What is a statistical decision problem ? Explain how point estimation and hypothesis-testing can be viewed as decision problems. 4+8=12

- (b) Given the distribution :

$$f(x) = {}^4C_x p^x q^{4-x}$$

$x = 0, 1, \dots, 4$ for testing

$$H_0 : p = \frac{1}{4} \text{ against } H_1 : p = \frac{1}{3},$$

critical region is $x > 1$. Calculate power of the test. 12

- (c) Explain sufficient statistic. Find sufficient statistics for θ in the following :

(i) $f(x, \theta) = \theta x^{\theta-1} \quad 0 < x < 1$

(ii) $f(x, \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp(-x^2/2\theta) \quad 0 \leq x \leq \theta.$

4+4+4=12

- (d) In cluster sampling, if sizes of cluster are not

same, show that the sample mean $\bar{y}_n = \frac{1}{n} \sum_{i=1}^n \bar{y}_i$

is a biased estimator of $\bar{Y}$. Derive the expression for its variance. 4+8=12

- (e) Define ratio method of estimation and give practical situations where it may be used. 6+6=12

(f) Show that for a BIBD with parameters b, v, r, k, λ

(i) $bk = vr$.

(ii) $\lambda(v-1) = r(k-1)$. 4+8=12

6. (a) Define Uniformly Most Powerful (UMP) test and unbiased test. Prove that an UMP test is unbiased but the converse is not always true. 4+8=12

(b) (i) Find the maximum likelihood estimates of α and β of the following distribution :

$$f(x; \alpha, \beta) = \beta e^{-\beta(x-\alpha)}; \alpha \leq x < \infty, \beta > 0.$$

(ii) If $x_1, x_2, \dots, x_n$ be a random sample from the exponential distribution :

$$f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}; x \geq 0, \theta > 0.$$

Estimate θ by the method of moments.

6+6=12

(c) Obtain the average sample size (A.S.N.) for testing :

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1.$$

for normal population with variance σ^2 . 12

(d) State and prove Neyman-Pearson Lemma in the generalized form and bring out its usefulness.

12

(e) (i) Explain the suitable test to test the null

hypothesis H_0 that the set of observations is random. Discuss the procedure to test such hypothesis. 3+3=6

- (ii) Explain a uni-parametric test for testing that the median of a continuous distribution is θ_0 , stating the assumptions if any. 5+1=6

7. (a) A random sample of size n is drawn from a population with p.d.f.

$$f(x) = \frac{1}{\pi[1+(x-\mu)^2]} ; -\infty \leq x \leq \infty$$

Prove that sample median is a consistent estimator of μ . 12

- (b) What is a likelihood ratio test ? State its important properties. Give an example of LR test which is uniformly most powerful unbiased. 4+2+6=12

- (c) Discuss how the prior information and the sample information are used to get posterior distribution. 12

- (d) State and prove Chapman Robbins Inequality. 12

- (e) Prove the following, for $S_N = Z_1 + Z_2 + \dots + Z_N$,

(i) $E(S_N) = E(N) \cdot E(Z)$

(ii) If $E(Z) = 0$, then $E(S_N^2) = E(N) \cdot V(Z)$. 6+6

8. (a) Define Horvitz-Thompson Estimator for the population mean and obtain its variance. 4+8=12

- (b) Prove that the variance of the mean $\bar{y}_{sy}$ of systematic is :

$$V(\bar{y}_{sy}) = \frac{nk-1}{nk} \cdot \frac{S^2}{n} [1 + (n+1)\rho]$$

where $nk = N$ and ρ is the intra class correlation coefficient between the units of the same systematic sample. 12

- (c) In a two-stage sampling procedure with unequal first stage units, find the estimates of the population total and their mean square errors. 6+6=12

- (d) Give the method of analysis of a 5×5 Latin Square design and explain the testing of difference of two treatment means in this design. 10+2=12

- (e) What are the advantages of factorial experiments over single factor experiment ? Write down the main effects and interaction effects of 2^3 factorial experiment and give its ANOVA table. 4+2+6=12