

STATISTICS

(Paper I)

Time Allowed : Three Hours
Maximum Marks : 300

QUESTION PAPER SPECIFIC INSTRUCTIONS

(Please read each of the following instructions carefully before attempting questions)

There are EIGHT questions divided in two Sections.

Candidate has to attempt FIVE questions in all.

Question Nos. 1 and 5 are compulsory and out of the remaining, any THREE are to be attempted choosing at least ONE question from each Section.

The number of marks carried by a question/part is indicated against it.

Wherever any assumptions are made for answering a question, they must be clearly indicated.

Word limit in questions, wherever specified, should be adhered to.

SECTION 'A'

1. Answer any five parts of the following:

12×5=60

- (a) In every box of a certain cereal is a picture of a football player. A full set is represented by m players. Find the probability that n ($> m$) boxes of cereal must be purchased to obtain a full set of pictures.

- (b) A random variable X has the following distribution:

x	$p(x)$
0	0
1	k
2	$2k$
3	$2k$
4	$3k$
5	k^2
6	$2k^2$
7	$7k^2+k$

Find k and determine the distribution function of X .

- (c) Explain, with one illustration of each, convergence in distribution and convergence in probability for a sequence of random variables. State the relation between these two.
- (d) Explain the difference between classical estimation and sequential estimation. Describe SPRT
- (e) Given a random sample of observation from the rectangular distribution with p.d.f.

$$f(x) = \begin{cases} 1/\theta & 0 \leq x \leq \theta \\ 0 & \text{outside} \end{cases}$$

find the m.l.e. of θ .

- (f) Briefly describe the likelihood ratio test for equality of variances of k univariate normal populations.

2. (a) For the geometric distribution

$$P(X=x) = pq^k, \quad k=0,1,2,3,\dots, \quad 0 < p < 1, \quad q=1-p$$

find the p.g.f. Hence find the mean and variance.

20

- (b) Examine if WLLN and CLT hold for the sequence of mutually independent random variables $\{X_k\}$, where

$$P(X_k=K) = P(X_k=-K) = 1/2\sqrt{K}$$

20

$$P(X_k=0)=1-1/\sqrt{K}$$

- (c) Let (X, Y) be a two-dimensional non-negative continuous random variable having the joint density

$$f(x,y)=4xy e^{-(x^2+y^2)}, x,y \geq 0$$

Derive the density function of $U=\sqrt{(x^2+y^2)}$ 20

3. (a) Define completeness of a statistic. Show that if a sufficient statistic is complete, then it is minimal sufficient. 20
- (b) Show that a necessary and sufficient condition for a distribution to admit an unbiased estimator with variance attaining the Cramer-Rao lower bound is that it belongs to the one-parameter exponential family. 20
- (c) Let X have the p.m.f. given by $P(X=-1)=\theta$, $P(X=x)=(1-\theta)^2\theta^x$ $x=0,1,2,\dots$ and $\theta \in (0,1)$. Obtain the m.v.u.e. of $(1-\theta)^2$. 20
4. (a) Define a monotone likelihood ratio family of distributions. Prove that the hypergeometric distribution is a member of the MLR family. 20
- (b) Let $X (i=1,2,\dots,n)$ be a random sample from the rectangular $(0,\theta)$ population. Find a UMP test of the hypothesis $H_0:\theta=\theta_0$ (given) against all alternatives. 20
- (c) Define a family of UMA confidence sets of an unknown parameter θ at confidence level $1-\alpha$. State a result giving a one-one correspondence between the family of UMA confidence sets and the family of acceptance regions of relevant UMP tests at level α . 20

SECTION 'B'

5. Answer any five parts of the following:

12×5=60

- (a) What are outliers in the context of regression analysis? How can you use Cook's distance to detect outliers?
- (b) Explain briefly the method of factor analysis and its uses.
- (c) With reference to a completely balanced two-way random effects model, find unbiased estimators of the different variance components.
- (d) Let $Y_1, Y_2, \dots, Y_n$, ($n \geq p$) be i.i.d. random vectors, each having $N_p(0, \Sigma)$ distribution. Define $A = \sum_{\alpha} y_{\alpha} y'_{\alpha}$. Show that for $n > p+1$

$$E(A) = 1/(n-p-1)\Sigma - 1$$

assuming Σ to be positive definite.

- (e) When do you use a split plot design? Give the blank ANOVA table for a split-plot experiment.
- (f) Describe Horvitz-Thompson estimator of a finite population total and give an estimate of its standard error.
6. (a) Let $y^{n \times 1}$ be a random vector distributed as $N_n(0, \sigma^2 I_n)$. Derive necessary and sufficient conditions for the quadratic form $y' A y$ to be distributed as χ^2 . 20
- (b) Briefly explain the development and use of linear discriminant functions in the context of two multivariate populations. 20
- (c) In a Gauss-Markoff set-up, define an estimable parametric function. Show that if $r(X)=p$, then every parametric function is estimable where we consider the set-up $(y^{n \times 1}, X^{n \times p} \beta^{p \times 1}, \sigma^2 I_n)$ 20
7. (a) Consider PPS sampling without replacement, based on positive size measures. Define the Des Raj estimator of the population total and prove its unbiasedness. 20

- (b) Assume that the population size is an integral multiple of the sample size. Compare the efficiencies of the sample mean based on stratified random sample with one unit per stratum and systematic sample when there is a linear trend in the population. **20**
- (c) Define the regression estimator of a finite population mean and work out its approximate mean square error, assuming that the second sample is a sub-sample of the first sample. **20**
8. (a) Explain the concept of confounding in a factorial experiment. **20**
- (b) In a Latin Square design, two observations from the same row are missing. Derive the analysis of the design. **20**
- (c) When is a BIB design said to be resolvable? Show that for a resolvable BIBD, $b \geq v + r - 1$ (in usual notations). **20**