C.S.E. Mains 2000

STATISTICS

PAPER – I

Time Allowed: 3 Hours *Maximum Marks: 300*

SECTION A

(Probability and Statistical Inference)

1. Attempt any five sub-pans:

(a) Let $P(S_1, S_2) = \exp[-\lambda_1 - \lambda_2 - \mu + \lambda_1 S + \lambda_2 S_2 + \mu S_1 S_2], (\lambda_1, \lambda_2, \mu) > 0$

be bivariate probability generating function (PGF) of (X, Y).

Find the marginal PGF of X and Y. Find also the PGF of X + Y. Examine whether X and Y ale independent.

12

(b) Apply a suitable inversion theorem to obtain the probability density function of the random variable whose characteristic function is given by

$\phi(t) = 1 - |t| \leq 1$

12

(c) State and prove Chapman-Kolmogorov equation.

12

(d) Slate the method of minimum Chi-square for estimating a parameter and show that for large sample, the minimum Chi-square equations and likelihood equations are identical, providing identical estimates.

12

(e) Bring out the relation between uniformly most accurate (UMA) confidence sets of confidence coefficient 1 — α and uniformly most powerful non-randomized tests of level α.

(f) Develop the Mann-Whitnev-Wilcoxon test and obtain the mean and variance of the test statistic. How is the test carried out for large samples?

12

2. (a) (i) Define expectation of a random variable (r.v.). Show that if X is a non-negative r.v. with distribution function F then

$$E(X) = \int_0^\infty (1 - F(x))\,dx,$$

provided the expectation exists.

8

(ii) Define convergence in probability (P), convergence in r^{th} mean (L_r) and convergence in distribution (D). Show that convergence in e mean implies convergence in probability and convergence in probability implies convergence in distribution.

12

(b) (i) State a necessary and sufficient condition for a function $\phi(t)$ on a real line to be the characteristic function of a random variable. Prove that it is necessary

12

(ii) Explain the concept of strong and weak laws of large numbers. State and prove Khintchin's WLLN.

12

(c) Explain the concept of Central Limit Theorems (CLT).

Let $\{X_n\}$ be a sequence of uniformly distributed random variables over $\left(-\beta n^{\lambda}, \beta n^{\lambda}\right), \beta > 0, \lambda > 0$. Test whether SLLN. WLLN and CLT hold for $\{X_n\}$.

3. (a) (i) If T be a sufficient statistic for an unknown parameter θ and U be any unbiased estimator of θ, then prove that ψ(T) of T such that $\psi(t) = E(U|T = t)$ is also an unbiased and has smaller variance than that of U.

15

(ii) Let $X_1, X_2, \ldots\ldots\ldots\ldots X_n$ be a random sample from a Poisson distribution with p.m.f.

$$f_\theta(x) = \begin{cases} \dfrac{e^{-\theta}\theta^x}{x!}, & x = 0, 1, 2\ldots\ldots \\ 0 & otherwise \end{cases}$$

where the parameter θ is unknown (θ > 0). Find the Minimum Variance Unbiased Estimator of $\dfrac{e^{-\theta}\theta^x}{k!}$.

12

(b) (i) Define a consistent estimator for

$\theta = (\theta_1, \theta_2, \ldots\ldots \theta_k)$

Show that T_n is a consistent estimator of θ only if $E(T_n) \to \theta$ and $V(T_n) \to 0$ as $n \to \infty$

15

(ii) Given a random sample of size n from a Cauchy population with location parameter θ show that the sample median is a consistent estimator of θ.

10

(c) A discrete random variable X has a distribution dependent on the specification of the values of a parameter θ as indicated in the following table. Find the MLE of θ when x = 2 is an observation of X

8

x	0	1	2	3	4
P(x) θ = 0	0.1	0.2	0.3	0	0.4
P(x) θ = 1	0.2	0.3	0.2	0.1	0.2
P(x) θ = 2	0.4	0.3	0.2	0.1	0

4. (a) (i) Given a random sample of size n from a Bernoulli (1, θ) distribution, find the UMP size-α test for $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$. Derive the power function P(θ) of the test.

10

(ii) Let φ(x) be a test function such that

$$\phi(x) = \begin{cases} 1 \text{ if } & x > k, \\ \gamma & x = k, \\ 0 & x < k, \end{cases}$$

where $0 \le \gamma \le 1$ and if X has density f(x, θ) which has monotone likelihood ratio in x where θ is real parameter, then show that the power function of the test φ(x) is increasing in θ.

10

(b) (i) Let $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_n)$ be independent random samples from the continuous distribution functions F(x) and F(x – θ), respectively, where F is unknown. It is required to test $H_0 : \theta = \theta$. Write down the median test statistic T and its asymptotic distribution under H_0.

10

(ii) Suppose $x_1, x_2, \ldots x_n$ are i.i.d. $N(\mu, \sigma^2)$ observations where μ is known. Obtain the uniformly most accurate (UMA) sets for σ^2.

10

(c) Let $X_1, X_2, \ldots$ be i.i.d. N(θ, 1) random variables and suppose that we want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 (> \theta_0)$. Determine the size n of the fixed sample of the most powerful test of size-α and power 1 - β Determine also average sample number under $H_0 E_{\theta_0}(N)$. of the SPRT of the same problem with strength (α, β)

SECTION B

(Linear Inference, Multivariate Analysis, Sampling Theory and Design of Experiments)

5. Attempt any five sub-pads:

(a) In connection with the Gauss-Markov linear models, what do you mean by Design matrix and Normal equations?

Discuss the least squares principle and use it to derive the normal equations.

12

(b) Identify, giving reasons, which of the following models are linear? Wherever possible, suggest a transformation $Z_i = g(Y_i)$ that will result in a linear model for Z_i:

(i) $Y_i = \beta_1 + \beta_2 x_i^2 + e_i$

(ii) $Y_i = \beta_1 + \beta_2 x_i + e_i^2$

(c) Define the Horvitz-Thompson estimator of the population total based on a sample of n distinct units from a finite population of N units with given positive inclusion probabilities of first and second orders. Show that the estimator is unbiased and derive the expression for its variance.

12

(d) Discuss the advantages of stratified sampling in comparison to unstratified sampling.

12

(e) What is missing plot technique? Describe, in detail, the analysis of a RHO with one missing observation. How will the analysis be modified if there are more than one missing observations?

12

(f) Establish the parametric relations of a BIHD with parameters v, b, r, n, λ. Show that in a symmetric BIBD, the number of treatments common between any two blocks is λ.

12

6. (a) Consider the following linear model

$y_1 = \theta_1 + \theta_4 + \theta_5 + e_1$

$y_2 = \theta_2 - \theta_4 + e_2$

$y_3 = \theta_3 + 2\theta_5 + e_3$

$y_4 = \theta_1 + \theta_2 + 2\theta_5 + e_4$

$y_5 = \theta_1 + \theta_3 + \theta_4 - e_5$

Where ei's are i.i.d. N(0, σ^2) variables

(i) Show that $\sum_{i=1}^{5} l_i \theta_i$ is estimable if and only if $l_4 = l_1 - l_2$ and $l_5 = l_1 + l_2 - 2l_3$.

(ii) Obtain the BLUe of $\theta_1 + \theta_2 + \theta_3$ and $\theta_3 - 2\theta_5$ b3W

10

(b) (i) Define principal components and explain their significance in statistical practice.

10

(ii) Let $X \sim N_p(\mu, \Sigma)$ unknown and $\mu = (\mu_1, \mu_2, \ldots, \mu_p)'$. Suggest a test for $H_0 : \mu_1 = \mu_2 = \ldots \mu_p$ against H_1 : all μ s are not equal.

15

(c) Let $X \sim N_p(\mu, \Sigma)$. Let G(q × p) be a matrix of rank q. show that $GX \sim N_p(G\mu_p, G\Sigma G')$

15

7. (a) (i) Define firs and second order inclusion probabilities. Evaluate them for SRSWR and SRSWOR.

8

(ii) Define a ratio estimator for population mean on the basis of a sample of size n selected with SRSWOR. Derive the expressions for its bias and mean square error in large sample.

12

(b) (i) Describe a method for selection of n units from a finite population with (1) PPSWR and (2) PPSWOR.

10

(ii) Define Hansen-Hurwitz Estimator (HHE) for the population mean and obtain its variance.

10

(c) Suppose a population consists of N first stage units (fsu) each of which contains M second-stage units (ssu). An SRSWOR of size n is taken from the N fsu and then from each selected fsu m ssu are selected at random without replacement. On the basis of the mn ssu suggest an unbiased estimator for the population mean. Find its variance. How will you determine a and m for a cost function

$c = a + nc_1 + nmc_2,$

where c = total cost, a = overhead cost, c_1 = cost per fsu, c_2 = cost per ssu, with minimum variance.

20

8. (a) When is a two-way ANOVA model said to be mixed? How will you formulate the different hypotheses of this model? Describe how the hypotheses are tested.

20

(b) (i) Distinguish between elementary contrasts and orthogonal contrasts in the context of a set of v treatments.

6

(ii) Construct a non-randomized layout of a $(3^3, 3)$ partially confounded factorial experiment in four replicates, confounding second order interaction effects only in each replicate. What property does this design exhibit in respect of the loss of precision of the confounded effects?

16

(c) Why do we need to develop split plot design? Write down the model of the design and give the estimates of the parameters involved in the model. Write down the ANOVA table and give the expressions of the estimates of standard errors for the different types of comparisons.

18

STATISTICS

PAPER - II

Time Allowed: 3 Hours ***Maximum Marks: 300***

SECTION A

1. Attempt any five sub-parts:

(a) Explain how you will detect lack of statistical control in a manufacturing process using a control chart.

12

(b) Given p_1, $1 - \alpha$, p_2 and β (usual notations), obtain the item by item sequential sampling plan for attributes.

12

(c) Define the terms reliability, maintainability and availability. Also obtain the reliability of an item whose life time distribution is given by

$$f(t) = a\alpha t^{\alpha-1} e^{-at^{\alpha}}; a\alpha < 0$$

12

(d) Show that the inter-occurrence time follows the exponential distribution in the case of a Poisson process.

12

(e) What do you understand by sensitivity analysis with reference to a Linear Programming Problem? Explain.

12

(f) Explain the graphical method of solving a rectangular game.

12

2. (a) Explain the terms AOQ and ATI. For a double Sampling attributes plan with rectifying inspection derive the expressions for the AOQ and the ATI.

20

(b) A system consists of n components and it will function if and only if at least $k(1 \le k \le n)$ of then components function. If the life time of the i^{th} component is exponential with parameter a find the reliability of the system.

12

(c) Distinguish between the p- and the np-charts. Also explain the construction and operation of the latter.

15

3. (a) Obtain the rules for choosing the incoming variable and the outgoing variable in the simplex method of solving a linear programming problem.

25

(b) For a finite irreducible aperiodic Markov chain with n-step transition probabilities $p_{ij}^{(n)}$ show that $\lim_{n \to \infty} p_{ij}^{(n)}$ exists and is independent of the initial state.

20

(c) A baking company sells its cakes by weight. It makes a profit of Rs. 0.95 for even kg (kilogram) of cake sold on the day it is baked. It disposes off all cakes not sold on the day they-are baked at a loss of Rs. 1. 15 per kg. If the demand for the cake is known to have the probability density f(d) = 0.03 - 0.0000d, 0 ≤ d ≤ c, find the optimum amount of cake the company should bake daily.

15

4. (a) What are $\overline{X}$ and R charts? When are they used? Explain how they are constructed and analyzed. Also explain how you would estimate the process standards using these charts.

15

(b) Write a brief note on MIL-STD-414 tables.

15

(c) Define the dual of a linear programming problem (LPP). Also explain the relationships between the solutions of the primal LPP and its dual. If the primal problem has an unbounded solution what can you say 4bout the solution of the dual?

15

(d) Show that every zero-sum two person game problem can be formulated as a LPP.

15

SECTION B

5. Attempt any five sub-parts:

(a) Explain Box-Jenkins methodology in analyzing a time series.

12

(b) State Laspeyres' and Paasche's index numbers for prices and examine whether they satisfy the various tests for index numbers.

12

(c) What organizations are responsible for collecting population statistics in India? Discuss the types of data collected by them.

12

(d) A population grows in such a manner that $\frac{1}{P(t)}\frac{dP(t)}{dt}$ is a linear function of log P(t) where P(t) is the population size at time t. Obtain the explicit expression for P(t) and state its properties.

12

(e) Explain the problems in developing rates for morbidity. Also define crude and specific morbidity prevalence rates.

12

(f) What is intelligence quotient? How is it constructed?

12

6. (a) What is a stationary time series? How do you test the stationary of a given time series? 20

(b) Explain the generalized least squares (GLS) method of estimation. Also obtain the GLS estimate of β in the model

$y_t = \beta x_t + u_t, t = 1, 2, \ldots n$ with

$E(u_t) = 0, E(u_t^2) = \sigma^2 x_t$ and $E(u_t u_t) = 0 \; if \;\; t \neq t'$

25

(c) Explain the identification problem and state the rank and order conditions for indenting a model equation. Are these conditions necessary and sufficient? 15

7. (a) What is a complete life table? Explain how such a table is constructed. 25

(b) Explain how the infant mortality rate is constructed. 10

(c) What are stable and quasi-stable populations? Derive Lotka's equation in this context and explain its utility. 25

8. (a) What are the agencies engaged in the collection and publication of price statistics in India? Also discuss the type of statistics published, their reliability and limitations. 20

(b) Explain the functions of any two of the organizations engaged in the collection of statistics of industrial production, both under the organized and the unorganized sectors. Where are these statistics published? 10

(c) Explain the utility of σ-scores and standard scores in analyzing psychological data. State the assumptions under which they are obtained. Under what conditions will these scores be equal? 15

(d) Explain factor analysis and its uses in psychological investigations. 15