C.S.E.-Mains-1996

STATISTICS

PAPER - I

Time allowed: 3 Hours **Maximum Marks : 100**

Candidates should attempt any five questions choosing at least one but not more than two from each Section.
Assume suitable data if considered necessary and indicate the same clearly.
The number of marks carried by each question is indicated at the end of the question.
Parts of the same question must be answered together and must not be interposed between answers to other questions.

SECTION I

(Probability)

1. (a) (i) For n events $A_1, A_2, \ldots, A_n$ show that

$$\sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) \le P \bigcup_{i=1}^{n} A_i \le \sum_{i=1}^{n} P(A_i)$$

(ii) Let $\{A_n\}$ be a non-decreasing sequence of events. State that

$$\lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right)$$

(b) An urn contains R red and W white balls. Balls are drawn one by one without replacement. If A_k is the event that a red ball is drawn for the first time on the k^{th} draw, then determine $P(A_k)$. Let p be the proportion of red balls in the urn before the first thaw. Show that

$$P(A_k) \to p(1-p)^{k-1} \text{ as } R + W \to \infty$$

(c) Explain clearly why there must be a mistake in each of the following statements:

(i) The probability that a student will get an 'A' in a statistics course is 0.32 and the probability that he or she will get either an 'A' or a 'B' is 0.27.

(ii) A company is working on the construction of two shopping centers. The probability that the larger one will be completed on time is 0.35 and the probability that both will be completed on time is 0.42.

26 + 20 + 14

2. (a) State and prove Bayes' theorem.

(b) If two random variables have the joint density

$$f(x_1, x_2) = \begin{cases} x_1 x_2 & \text{for } 0 < x_1 < 1,\ 0 < x_2 < 2 \\ 0 & \text{elsewhere} \end{cases}$$

find the probabilities that

(i) both random variables will take on values less than 1.

(ii) the sum of the values taken on by both will be less than 1.

(c) Show that under suitable conditions the binomial distribution approaches the Poisson distribution.

20+20+20

3. (a) State and prove Markov's theorem for WLLN.

(b) Verify the following:

(i) $X_n \xrightarrow{p} X, X_n \xrightarrow{p} Y \Rightarrow P[X = Y] = 1$

(ii) $X_n \xrightarrow{r} X,$ for some $r > 0 \Rightarrow X_n \xrightarrow{p} X$

(c) (i) State Kolmogorov's SLLN for a sequence of independent random variables.

(ii) Verify whether the following sequence of independent random variables obey the SLLN :

$$P X_k = \frac{1}{2^k} = P \ X_k = -\frac{1}{2^k} = \frac{1}{2}, k = 1, 2, \ldots$$

20+20+20

4. (a) Define the probability generating function of a random variable. Find the pgf of a geometric distribution and hence find its mean and variance.

(b) State the properties of a characteristic function and show that it is uniformly continuous.

(c) State and prove the Lindeberg-Levy central limit theorem.

20+15+25

SECTION II

(Statistical Inference)

5. (a) State and prove Rao-Blackwell theorem. State its importance.

(b) Let (T_n) be a sequence of estimators such that $E(T_n) \rightarrow \Theta$ and Var $(T_n) \rightarrow 0$ as $n \rightarrow \infty$. Show that T_n is a consistent estimator for Θ.

(c) Let be an MVU estimator of $g(\Theta)$ and T_1 an unbiased estimator with efficiency $e_\theta < 1$. Then can any unbiased linear function of T_0 and T_1 be an MVU estimator of $g(\Theta)$? Justify your answer.

25+15+20

6. (a) Let Θ be a real parameter and the random variable x have probability density with MLR in T(x). Determine the size α UMP test of $H_0 : \Theta \le \Theta_0$ against $H_1 : \Theta > \Theta_0$.

(b) Let X_i (i = 1, 2, …, n) be a random sample from a rectangular distribution $r(0, \Theta)$. Show that for testing $H_0 : \Theta = \Theta_0$ against $H_1 : \Theta > \Theta_0$, any critical region W for which $P_{\theta 0}$ (W) = α and which necessarily includes the sample point if max $(x_1,\ldots, x_n) > \Theta_0$ is UMP at level α.

30+30

7. (a) Let X be a N (μ, σ^2) variable, where both the parameters are unknown. Determine the LR test of $H_0 : \mu=\mu_0$ against $H_1 : \mu \ne \mu_0$ based on a sample of size n. Determine the power function of the test.

(b) Let the random variable X have the pmf $f_\theta(x) = \Theta^x(1 - \Theta)^{1-x}$, x = 0, 1 and $0 < \Theta < 1$. Derive the SPR test of H_0: $\Theta = \Theta_0$ against H_1 : = 0,. Determine the OC function of the test.

30+30

8. (a) Define the empirical distribution function and find its expectation and variance. Hence show that it is a consistent estimator of the population distribution function.

(b) Describe the Kolmogorov test for goodness of fit. Compare this test with the chi-square test for goodness of fit.

(c) Obtain a $100(1-\alpha)\%$ confidence interval for the population median based on the sign test.

20+20+20

SECTION III

(Linear Inference and Multivariate Analysis)

9. (a) With the usual Gauss-Markov linear model, define the estimation space and the error space. Obtain expressions for the SS due to error and the SS due to all possible BLUEs of parametric functions. Find the expectation of the former.

(b) Let $l_1'\hat{\beta}$ and $l_2'\hat{\beta}$ be the LS estimators of two estimable functions $l_1'\beta$ and $l_2'\beta$ respectively. Obtain the variances and covariances of the estimators.

30+30

10. (a) Given two regression lines

$$Y_{ki} = \beta_k X_i + e_{ki} (k = 1, 2; i = 1, 2, ..., n)$$

show that the F-statistic for testing H : $\beta_1 = \beta_2$ can be put in the form

$$F = \frac{\left(\hat{\beta}_1 - \hat{\beta}_2\right)^2}{2S^2\left(\sum_1^n x_i^2\right)^{-1}}$$

Write the expression of S^2.

(b) In the multiple linear regression model with $(p-1)$ independent variables, derive a test to find out whether the regression on the regressor variables is significant or not. Show that the test statistic can be expressed in terms of the sample multiple correlation coefficient.

(c) Write down the model of a complete three-way classified data with equal number of observations per cell. Write down the corresponding ANOVA table.

20+25+15

11. (a) Let $Y = (Y_1, Y_2, ..., Y_n)'$ be a vector of random variables with

pdf $K^{-1} \exp\left[-\frac{1}{2}(y-\Theta)^t \Sigma^{-1}(y-\Theta)\right]$.

(i) Show that $K = (2\pi)^{n/2} |\Sigma|^{1/2}$ and

(ii) Determine E[Y] and D[Y].

(b) Show that under usual normal population regression model the joint distribution of the partial regression coefficients is multivariate normal.

30+30

12. (a) Define Mahalanobis' distance between two populations as estimated from the sample on the basis of p characters. State how the D^2-statistic is used to test the hypothesis of no difference in mean values of the p characters.

(b) Define Fisher's discriminant function between two p-variate populations. Indicate the connection between the discriminant function coefficients and the D^2-statistic. Describe a test to judge the adequacy of a pressigned discriminant function.

30+30

STATISTICS

PAPER - II

Time allowed: 3 Hours **Maximum Marks : 100**

Candidates should select any three Section and attempt any five questions from the selected Sections. choosing at least one but not more than two questions from each of the selected Sections. Assume suitable data if considered necessary and indicate the same clearly. All questions carry equal marks. Parts of the same question must be answered together and must not be interposed between answers to other questions.

SECTION I

(Sampling Theory and Design of Experiments)

1. (a) Define random sampling. Why are random samples distinctly superior to non-random samples ? Describe linear systematic sampling with interval k and show that it is an equal probability sampling method.

   (b) State the Horvitz-Thompson estimator of a population total in PPS sampling without replacement. Show that it is unbiased.

   (c) State the common reasons for stratifying a population before sampling.

2. (a) Outline the ratio and regression methods of estimating a population total. When and how are these methods useful?

   (b) Explain the use of double sampling for stratification.

   (c) Write a brief note on "non-sampling errors".

3. (a) State the basic principles of designed experiments and their role in such experiments.

   (b) Give the linear model and outline the method of analysis of a RBD.

   (c) Define a BIBD. Show that the parameters of a BIBD satisfy the relations

   (i) $bk = vr$

   (ii) $\lambda(v-1) = r(k-1)$

4. (a) Analyze completely the following data from an experiment which used a Latin Square design.

Order of assembly	Operator			
	I	II	III	IV
1	10 C	14 D	7 A	8 B
2	7 B	18 C	11 D	8 A
3	5 A	10 B	11 C	9 D
4	10 D	10 A	12 B	14 C

The letters A, B, C and D represent four assembly methods. The relevant table value off for a = 0.05 is 4.76.

(b) What are factorial experiments? Illustrate the concept of total confounding in such experiments.

(c) Write an explanatory note on simple lattice designs.

SECTION II

(Engineering Statistics)

5. (a) Explain the significance of the following concepts in the construction of control charts:

   (i) Assignable causes of quality variation

   (ii) Rational subgroups

   (b) Describe briefly the construction and working of a CUSUM chart for a variable characteristic.

   (c) When and how is a np-chart useful ? What is the statistical basis of this chart ?

6. (a) Describe the working of a double sampling plan for product control. What are the advantages of this plan as compared to a single sampling plan?

   (b) Define the concepts of AOQ and ATI in rectifying inspection. Derive expressions for these under a double sampling plan.

   (c) Devise a known sigma variables sampling plan - when there is a two-sided specification on tile quality characteristic.

7. (a) Define a parallel system. Obtain its reliability function and the mean life of the system if all the units are equally reliable and reliability law is exponential.

   (b) Suggest a procedure for estimating the parameters of a two parameter exponential distribution under type II censoring.

   (c) Describe a bath-tub failure model and its applications.

8. (a) Explain the use of redundancy in reliability improvement.

   (b) Define the terms:

   (i) Hazard function

   (ii) IFR and DFR distributions

   Provide examples.

   (c) Illustrate the concepts of

   (i) availability, and

   (ii) maintainability.

SECTION III

(Operational Research)

9. (a) Explain the scope of Operational Research.

   (b) Derive the steady-state results (probabilities, mean numbers and mean waiting times) for an M/M/s queue.

   (c) What is an imbedded Markov chain? For the M/G/l queue identify the imbedded Markov chain.

10. (a) Define (i) basic feasible solution, (ii) slack and surplus variables and (iii) artificial variables in the context of a linear programming problem (LPP).

(b) Show that if an optimal solution exists for a LPP then it corresponds to an extreme point of the feasible region.

(c) Describe the simplex algorithm for solving a LPP.

How is this procedure modified when some of the variables are unrestricted?

11. (a) Define a transportation problem with m sources and n destinations. Show that in such a problem only (m + n – 1) restrictions are linearly independent.

(b) Prove that a transportation problem can never have unbounded solutions.

(c) Outline the Hungarian method of solving an assignment problem.

12. (a) Specify a few replacement policies.

(b) Explain the following control statements in FORTRAN language:

(i) Unconditional GO TO

(ii) Computed GO TO

(iii) Arithmetic IF

(c) Write a FORTRAN program to compute the coefficient of variation from a data set having n ungrouped observations.

SECTION IV

(Quantitative Economics)

13. (a) Specify the additive and multiplicative models for a time series. How are seasonal indices constructed ?

(b) State the time and factor reversal tests for an index number. Verify whether the price index numbers due to Paasche and Marshall-Edgeworth satisfy these two tests.

(c) Write a note on variate difference method.

14. (a) Formulate a general linear model (GLM), clearly stating the assumptions. Derive the MLE of the parameter vector when the errors in the GLM follow a multivariate normal distribution.

(b) Explain the procedure for testing the hypothesis $H_0 : \beta_1 + \beta_2 = c$ (some given constant) in the model $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$. State the underlying assumptions.

(c) Illustrate a technique which may be used to represent factors like temporal effects and qualitative variables in a GLM.

15. (a) Mention the main sources that cause serially correlated disturbances. Outline the Cochrane-Orcutt procedure to estimate the parameters of a linear model when the errors are serially correlated.

(b) Show that measurement error in the explanatory variables poses a serious estimation problem in the set up of a GLM. Outline an estimation method based on instrumental variables in this context.

(c) Explain the multi-collinearity problem. Interpret it geometrically and algebraically. Suggest methods to overcome the problem.

16. (a) Compare the ILS and 2 SLS methods of estimation in simultaneous equation models.

(b) Examine for identifiability the two-equation system

$$\beta_{11} y_{1t} + \beta_{12} y_{2t} + \gamma_{11} y_{1t} + \gamma_{12} x_{2t} = u_{1t}$$

$$\beta_{21} y_{1t} + \beta_{22} y_{2t} + \gamma_{21} y_{1t} + \gamma_{22} x_{2t} = u_{2t}$$

(i) as it stands, and

(ii) if $\gamma_{12} = 0$, $\gamma_{21} = 0$.

(c) Write a note on short term economic forecasting.

SECTION V

(Demography and Psychometry)

17. (a) Define the following and explain their uses:

(i) Density and proximity of population

(ii) Infant mortality rate

(iii) Net reproduction rate

(b) Stating the underlying assumptions, outline the construction of a life-table.

(c) In the usual notation show

$$q_x = \frac{2m_x}{2 + m_x}$$

in the setup of a life-table.

18. (a) Examine Gompertz curve as a growth model.

(b) Describe stable and quasi-stable population models. How are these models useful for estimating demographic parameters?

(c) Explain the direct method of standardising death rates.

19. (a) What is test reliability ? How is this measured ? Describe one method fully.

(b) Examine the effect of lengthening and repeating a test on test reliability.

(c) A test with 40 items has a reliability coefficient of 0.68. What is the reliability of a comparable test with (i) 80 items, (ii) 100 items ?

20. Write explanatory notes on any three of the following:

(a) National sample surveys

(b) IQ tests

(c) Logistic curve

(d) Balancing equation in population studies.