

## STATISTICS (Paper II)

**Time Allowed : Three Hours**

**Maximum Marks : 300**

### QUESTION PAPER SPECIFIC INSTRUCTIONS

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

There are TWENTY questions divided in five Sections.

Candidate has to select any THREE Sections and attempt FIVE questions in all from them.

All questions carry equal marks.

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. (a) What are the measures of the qualities of an estimator in the case of sampling from finite populations? Explain them. Also show that sampling without replacement is better than sampling with replacement in the case of simple random sampling.  
(b) What is probability proportional to size (PPS) sampling? In what situations is PPS sampling recommended? Also explain a method of selecting a PPS sample without replacement. Further obtain an unbiased estimator of the population mean and the variance of this estimator using a PPS sample.  
(c) In the case of stratified random sampling, assuming a simple linear cost function, obtain the optimal strata sample sizes which minimize the variance of the unbiased estimator of the population proportion when the total cost is fixed. Also obtain the expression for the total sample size.
2. (a) Distinguish between two phase and two stage sampling. For the case of two stage sampling with equal sized first stage units, obtain an unbiased estimator of the population mean and find its variance. Also estimate this variance unbiasedly.  
(b) In the context of estimating a population percentage, show that  $V_{st} : V_{sy} : V_s \approx 1/n : 1/n : n$  if the population has a linear trend, where  $V$  stands for the variance of the unbiased estimator of the population percentage and st, sy and sr respectively stand for stratified random, systematic and simple random sampling.  
(c) With an example explain ratio estimator. In the case of simple random sampling without replacement obtain an expression for the bias of the ratio estimator of the population mean.
3. (a) Explain the need for and the benefits due to the use of randomisation, replication and local control techniques in designing statistical experiments.  
(b) Explaining the rationale behind the missing plot technique, derive the formula for estimating a single missing value in a Latin square design. In such a situation explain how you will test for (i) the equality of the treatment means and (ii) the difference between two treatment means one of which relates to the missing value.  
(c) For a  $2^4$  factorial experiment with 3 replications, each containing 4 blocks, obtain a design such that no main effect is confounded and any interaction is confounded not more than once. Also obtain the variance of the estimate of a main effect and that of a confounded interaction. Further explain how you will calculate the sum of squares due to a confounded interaction.
4. (a) What is a split plot design? When is it recommended? Explain the analysis of data arising from such a design where the main unit treatments are arranged in a latin square. Also explain how you will compare the effects of (i) two sub-unit treatments within a main unit,

- (ii) two main unit treatments, and (iii) two sub-unit treatments in two different main units.
- (b) In what situations is the use of incomplete designs recommended? In the usual notations, show that  $B \geq t + r - k$  for a BIBD.
- (c) What is a simple lattice design? Explain the analysis of data arising from such a design. Also explain a method of testing whether a linear contrast of the treatment effects is zero or not.

### SECTION 'B'

5. (a) Explain the construction and use of  $\bar{X}$  and R charts when the process standards are (i) given and (ii) not given. How will you estimate the process standard deviation from the chart?
- (b) Distinguish between assignable and chance causes of variation in the quality of a manufactured item. When is a manufacturing process said to be in statistical control? How will you find out whether a manufacturing process is in statistical control or not?
- (c) Describe CUSUM charts, stating their advantages and limitations. Describe the decision interval scheme and the V-mask procedure used in CUSUM charts. Also show that these schemes are equivalent.
6. (a) Distinguish between type A and type B OC curves with reference to a single sample attribute inspection plan. How are they constructed?
- (b) Obtain a variable sampling plan with double specification limits and with specified values of  $\alpha$ ,  $\beta$ ,  $p_1$  and  $p_2$  (usual notations) when the quality characteristic is normally distributed with a known standard deviation.
- (c) What is a sequential sampling plan? Also obtain the formula for the ATE for an item by item SPR plan when rejected lots are inspected one hundred percent.
7. (a) The life time of the  $i^{th}$  component of an component system is uniformly distributed over the interval (0,1). Find the reliability of the system when the components are connected (i) in series and (ii) in parallel.
- (b) What is a 'k out of n' system? Derive an expression for the reliability of such a system in terms of the reliability of the components.
- (c) Show that the Weibull distribution
- $$f(t) = a\alpha t^{\alpha-1} e^{-at^\alpha} \quad (a > 0, \alpha > 0)$$
- has increasing failure rate when  $\alpha > 1$ , constant failure rate if  $\alpha = 1$  and decreasing failure rate if  $\alpha < 1$ .
8. (a) What are type I and type II censored life testing experiments? If the life time distribution is exponential, explain how you will estimate the parameter of the distribution under type II censoring.
- (b) Explain the main features of a bath tub model failure curve and its suitability in describing real life situations, Illustrate by a suitable example.
- (c) Explain the different types of redundancies and their role in improving the reliability of a system.

### SECTION 'C'

9. (a) With reference to a homogeneous Markov chain, define a recurrent state. Also prove that a state i is recurrent if and only if the series  $\sum p_{ii}^{(n)}$  is divergent, where  $(p_{ii}^{(n)})$  is the matrix of n-step transition probabilities of the Markov chain.
- (b) If  $X(t)$ ,  $t > 0$ , is a Poisson process then obtain the distribution of  $X(s)$  given  $X(t) = n$ ,  $0 < s < t$

- (c) Explain the  $M|M|1$  queueing system. For this system obtain the steady state distribution of the number of persons in the system and the average waiting time of a customer in steady state.
10. (a) Show that the objective function of a linear programming problem (LPP) will assume its minimum value at an extreme point of the convex set of feasible solutions of the LPP. If the objective function assumes the minimum value at more than one point then show that it will assume this minimum value at infinite number of points.
- (b) Obtain a necessary and sufficient condition for a basic feasible solution to a LPP to be optimal.
- (c) Discuss the economic interpretation of the primal and the dual problems with suitable examples.
11. (a) Explain the transportation problem. Show that a finite minimum feasible solution always exists in this case.
- (b) Describe and justify the algorithm used to find the optimal solution for a transportation problem by the modified stepping stone method.
- (c) Describe the Moran model for a dam with a finite capacity (discrete case). If the inputs follow the geometric distribution and there is unit release, obtain the stationary distribution of the dam content.
12. (a) What are machine, assembly and high level languages? State their merits and limitations,
- (b) (i) Distinguish between a FUNCTION sub-programme and a SUBROUTINE sub-programme. Also state the rules governing their use. (ii) Explain the uses of 'Computed GO TO' and 'Assigned GO TO' statements in FORTRAN. Also state the rules governing their use.
- (c) Write a FORTRAN programme to calculate the correlation coefficient on the basis of a set of  $n$  pairs of observations  $(X_i, Y_i)$ ,  $i=1,2,\dots,n$ , to be read by the programme.

### SECTION 'D'

13. (a) With a suitable illustration explain the notion of seasonal effects in a time-series. Also explain any one method of measuring the seasonal effects in a time series. Point out the merits and limitations of the method suggested.
- (b) Explain how you will fit the model  $Y = ae^{bX+u}$ , where  $a$  and  $b$  are constants to be estimated and  $u$  is a random variable with mean zero, and variance  $\sigma^2$ . The data consist of  $n$  pairs of observations on  $(X, Y)$ . How will you estimate  $\sigma^2$ ?
- (c) Explain, briefly, a method for studying the random component in a time series, stating the assumptions made,
14. (a) What are index numbers? State their main uses. Also distinguish between chain base and fixed base index numbers and explain how one can be obtained from the other in the case of a single item.
- (b) Define Marshall-Edgeworth index number and examine whether it satisfies the index number tests. Also show that this index number will lie between the Laspeyre's and Paasche's index numbers.
- (c) Explain the term 'demand elasticity'. What do you understand by family budget data? How will you use this data for estimating the income elasticity of demand of selected items?
15. (a) Define autocorrelation and explain how this gets generated in a linear econometric model. Also explain a test for detecting the presence of autocorrelation among the error terms in a model.

- (b) What do you understand by heteroscedastic error term in a model? What are the sources of such disturbances? Explain any one method of estimating the parameters in the presence of such heteroscedastic disturbance terms.
- (c) Explain the following terms, with examples (i) Predetermined variables (ii) Exogenous variables (iii) Endogenous variables
- 16. (a) Explain the identification problem in the context of simultaneous equation models. Illustrate with examples.
- (b) State the rank and order conditions for a model to be just identified and justify the same.
- (c) Explain the indirect least squares method of estimating the parameters of an identified equation. Also show that this method is not applicable if the equation is under identified.

### SECTION 'E'

- 17. (a) What are the sources of demographic data in India? Explain and compare the population register and population census as sources of demographic data.
- (b) Define and compare infant death rate and infant mortality rate (IMR). Also explain how IMR is determined from death statistics.
- (c) What is an abridged life table? Explain the nature, need and King's method of construction of such tables.
- 18. (a) Stating the underlying assumptions, derive the Logistic growth model. Also state the main features of the model.
- (b) Define Gross and Net reproduction rates. Explain their uses and how to compute them.
- (c) What do you understand by stable and quasi-stable populations? Derive Lotka's equation linking the intrinsic rate of growth with the age structure of a stable population. Hence, show that under neutral change in mortality by age to age distribution of a stable population does not change.
- 19. (a) Explain the following terms (i)  $\sigma$ -scores (ii) standard scores, and (iii) T-scores. Also explain how to get T-scores if the test scores are in the form of a frequency table.
- (b) Define reliability of a test. Also explain the methods of determining the reliability of tests.
- (c) What do you understand by intelligence quotient? How is it measured? How is the validity of an intelligence test generally determined?
- 20. Write explanatory notes on any three of the following:
  - (a) Morbidity and its measures
  - (b) Health Surveys
  - (c) Standardisation of vital rates
  - (d) Force of mortality and its evaluation