I.E.S / ISS - 09

381

C-JTT-J-TC

STATISTICS – III

Time Allowed : Three Hours | *Maximum Marks : 200*

INSTRUCTIONS

Candidates should attempt FIVE questions in ALL including Questions No. 1 and 5 which are compulsory. The remaining THREE questions should be answered by choosing at least ONE question each from Section A and Section B.

The number of marks carried by each question is indicated against each.

Answers must be written only in ENGLISH.

(Symbols and abbreviations are as usual)

If any data/value is to be assumed for answering a question, the same must be mentioned clearly.

SECTION A

1. Answer any ***five*** parts : *5×8=40*

(a) What do you mean by sampling frame ? Why do you need it ? Give reasons for preferring the sampling to complete enumeration.

(b) Describe the procedure for drawing a simple random sample of size n without replacement from a finite population of size N. Obtain the probabilities of drawing a specified sample in simple random sampling with and without replacement.

(c) Distinguish between two-stage and two-phase sampling schemes. Give suitable examples.

(d) Give the layout and analysis of a completely randomized design. Which of the basic principles are satisfied in this design and how ?

(e) Explain the term 'connected' in the context of block designs. Show that a balanced incomplete block (BIB) design with usual parameters v, b, r, k (≥ 2) and λ, is always connected.

(f) What do you mean by confounding in a factorial experiment ? Explain how you will obtain the confounded interactions from a key block.

2. (a) Obtain an unbiased estimator of population variance in case of simple random sampling without replacement.

(b) Let $y_1, y_2, \ldots, y_n$ be the measurements with respect to a characteristic Y on the units of a simple random sample drawn without replacement from a population of size N.

Show that

$$\text{covariance } (y_i, y_j) = -\frac{\sigma^2}{N-1}$$

where

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2$$

(c) Explain the principle of stratification. How would you determine the best mode of stratification ?

With two strata a sampler would like to have $n_1 = n_2$ for administrative convenience instead of using the values given by the Neyman allocation. If $V(\bar{y}_{st})$ and $V_{opt}(\bar{y}_{st})$ denote the variances given by the $n_1 = n_2$ and the Neyman allocations respectively show that

$$\frac{V(\bar{y}_{st}) - V_{opt}(\bar{y}_{st})}{V_{opt}(\bar{y}_{st})} = \left(\frac{r-1}{r+1}\right)^2$$

where $r = \frac{n_1}{n_2}$ as given by Neyman allocation.

(d) Explain cluster sampling. Show that the efficiency of cluster sampling relative to simple random sampling is approximately equal to

$$\frac{1}{1+(M-1)\rho}$$

where the symbols have their usual meaning. Interpret this result. *4×10=40*

3. (a) Obtain an estimate of a missing observation in a Latin square design. How does the subsequent analysis differ from the usual case ?

(b) Explain the uniformity trials. Discuss their role in picking up a proper design.

(c) Describe the analysis of a BIB design with the recovery of interblock information.

(d) Following are the block contents of one of the blocks of a replicate of a 2^5 experiment in factors A, B, C, D and E :

a, bc, cd, be, de, abd, ace, abcde.

Identify the confounded effects. Write down the remaining blocks of the replicate. *4×10=40*

4. (a) In a two-stage sampling scheme, the population consists of N first-stage units and each first stage unit has M second stage units. A sample of n first stage units and m second stage units from each selected first stage unit is drawn with equal probability at each stage. Obtain the expression for the variance of the unbiased estimator of the population mean.

(b) Explain the ratio method of estimation of the mean of a population on the basis of a simple random sample without replacement. Obtain the standard error of this estimator when the sample size is large.

(c) What is the use of analysis of covariance ? Give the general procedure of analysis of covariance for a RBD, stating the necessary assumptions.

(d) What is a split plot design ? Give an example of a situation where this is a natural choice. Outline the analysis of data from a split-plot experiment where the main plots form the randomized block. *4×10=40*

SECTION B

5. Answer any ***five*** parts : *5×8=40*

(a) Distinguish between fixed base and chain base methods for the construction of index numbers. Discuss their relative merits.

(b) Explain ratio to moving average method for calculating seasonal indices.

(c) Describe the Yule – Slutsky effect of moving average operation on the random component of a time series.

(d) Discuss the problems involved in the statistical estimation of market demand function.

(e) If the demand curve is of the form

$$p = a\, e^{-kx}$$

where p is the price and x is the demand, prove that the elasticity of demand is $\frac{1}{kx}$. Hence deduce the elasticity of demand for $p = 10\, e^{-x/2}$.

(f) State the problem of measurement errors in the explanatory variables of a linear regression model and explain its consequences.

6. (a) What is a periodogram ? Describe the method of periodogram analysis for determining the hidden-periodicities in a time-series.

(b) Explain the time-reversal and factor-reversal tests for an ideal index number. Give an example of index number which satisfies both of these tests.

(c) Describe, briefly, the Engel's law and Engel's curve. What are the different forms of Engel's curve usually employed for fitting to the family budget data ?

(d) State the probability density function of the Pareto distribution and give its cumulative form. Interpret the constants involved. *4×10=40*

7. (a) What are the assumptions in the classical linear model ? Show that the least squares estimator for the vector of coefficients is the best linear unbiased estimator.

(b) Discuss the possibility of consistent estimation of parameters in an equation whose disturbances are serially correlated.

(c) Describe the two-stage least squares method of estimation for the simultaneous equations model. Show that two-stage least squares estimators are consistent.

(d) State the problem of identification in the simultaneous equations model. State the order and rank conditions for identifiability. Illustrate your answer with an example. *4×10=40*

8. (a) Describe different schemes for explaining the oscillations in a stationary time series. Explain the use of correlogram for discriminating among the above schemes.

(b) Define Marshall – Edgeworth index number and examine whether it satisfies Fisher's tests for index numbers. Also show that this index number will lie between the Laspeyres' and Paasche's index numbers.

(c) Discuss Goldfeld – Quandt test for detecting the heteroscedasticity of the disturbance terms.

(d) Define multicollinearity. What are the causes of multicollinearity ? How will you detect the same and solve it ? *4×10=40*