IES / ISS Exam. 2019

STATISTICS

PAPER—III

Time Allowed : Three Hours	Maximum Marks : 200

QUESTION PAPER SPECIFIC INSTRUCTIONS

Please read each of the following instructions carefully before attempting questions

There are EIGHT questions divided under TWO Sections.

Candidate has to attempt FIVE questions in all.

Both the TWO questions in Section—A are compulsory.

Out of the SIX questions in Section—B, any THREE questions are to be attempted.

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

Unless otherwise mentioned, symbols and notations have their usual standard meanings.

Assume suitable data, if necessary, and indicate the same clearly.

Attempts of questions shall be counted in sequential order. Unless struck off, attempt of a question shall be counted even if attempted partly.

Any page or portion of the page left blank in the Question-cum-Answer (QCA) Booklet must be clearly struck off.

Answers must be written in ENGLISH only.

SECTION—A

1. *(a)* Describe cumulative total and Lahiri's methods of selecting a sample. Show that both the methods provide probability proportional to sample size. 10

(b) For a multiple linear regression model satisfying all the basic assumptions, show that ordinary least squares (OLS) estimators of regression coefficient vector β and estimator of error variance σ^2 based on OLSE of β are unbiased, jointly sufficient and efficient estimators of β and σ^2. 15

(c) Given $X_t = \beta_1 + \beta_2 t + w_t$, $w_t \sim \text{iid}(0, \sigma^2)$, where β_1 and β_2 are constants.

(i) Determine whether X_t is stationary.

(ii) Show that $Y_t = X_t - X_{t-1}$ is stationary.

(iii) Show that the mean of moving average $v_t = \frac{1}{2q+1} \sum_{i=-q}^{q} X_{t-i}$ is $\beta_1 + \beta_2 t$. 15

2. *(a)* Distinguish clearly between sampling and non-sampling errors. In this context, describe why non-response errors arise. Explain the steps of Hansen-Hurwitz technique of sub-sampling of non-respondents for estimating the population mean. Obtain the variance of the suggested estimator. 15

(b) Define best linear unbiased predictor (BLUP). Obtain the expression for BLUP in case of a simple linear regression model satisfying all the basic assumptions. Show that BLUP can be written in terms of BLUE of the regression coefficients. 15

(c) The demand and supply functions for three commodities are given below :

$$d_A = 64 - 9p_A - 3p_B - 3p_C;\ S_A = 10 + 6p_A$$

$$d_B = 12 + 3p_A - 6p_B + 3p_C;\ S_B = 15 + 6p_B$$

$$d_C = 21 - 3p_A + 3p_B - 5p_C;\ S_C = 3 + p_C$$

Find the equilibrium values for prices and supplies for all the three commodities. 10

SECTION—B

3. *(a)* Differentiate between PPS and πPS sampling schemes. What is Midzuno-Sen sampling scheme? Obtain the probability of selecting a sample of size n from a population of size N under this scheme. Hence prove that the usual ratio estimator becomes unbiased under this scheme. 10

(b) In what sense does cluster sampling differ from simple random sampling? A sample of n clusters is drawn at random without replacement from a population consisting of N clusters each of size M. Obtain the best linear unbiased estimator of the mean per unit in the population and show that its approximate variance is

$$\frac{N-n}{Nn}\cdot\frac{S^2}{M}\{1+(M-1)\rho\}$$

where N is large enough, and ρ and S^2 respectively denote the intra-class correlation coefficient and population mean square. 15

(c) What is Horvitz-Thompson estimator of population total? Give the Yates-Grundy expression of its variance and obtain an unbiased estimator of this variance. Show that this estimator of the variance will never be negative under Midzuno-Sen sampling scheme. 15

4. *(a)* Explain multicollinearity problem in a regression model. What are its consequences? State different indicators of multicollinearity and explain. 10

(b) Define ridge estimator. Obtain its mean squared error. Derive the region for the ridge parameter such that ridge estimator dominates OLSE under mean squared error criterion. 15

(c) Describe a heteroscedastic regression model. Explain how to estimate the parameters of heteroscedastic regression model with $D(\varepsilon)=\sigma^2 V$, when *(i)* V is completely known and *(ii)* V is partially known. 15

5. *(a)* *(i)* Let the price index numbers computed by Laspeyres' and Paasche's methods be denoted respectively by $L(p)$ and $P(p)$. Let similar quantities for quantity indices be defined and denoted as $L(q)$ and $P(q)$. Prove that if the values of any three of the above are known, the remaining one can be calculated.

(ii) If Bowley's and Fisher's price indices are denoted by x and y respectively, then show that Laspeyres' index (L) and Paasche's index (P) are given by

$$\text{either}\quad L=x+\sqrt{x^2-y^2}\quad\text{and}\quad P=x-\sqrt{x^2-y^2}$$

$$\text{or}\quad L=x-\sqrt{x^2-y^2}\quad\text{and}\quad P=x+\sqrt{x^2-y^2}$$ 10

(b) (i) What are the advantages and disadvantages of fixed-base and chain-base indices?

(ii) What are the points to be considered in selecting a base period for the purpose of calculating an index number? 15

(c) Show that the real-valued function

$$r(h) = \begin{cases} 1, & \text{if } h = 0 \\ \rho, & \text{if } h = \pm 1 \\ 0, & \text{otherwise} \end{cases}$$

is an auto-covariance function, if $|\rho| < \frac{1}{2}$. 15

6. (a) (i) Show how you would find the sample size n_i $(i = 1, 2, \cdots, k)$ to be selected from the ith stratum under optimum allocation if the total cost of the survey is a fixed quantity.

(ii) The ratios of the sizes and the standard deviations of two strata are 11 and 9 respectively. The ratio of the sample sizes for a general allocation is n_1 / n_2 and that for the Neyman allocation is n_1' / n_2', and the ratio of the two ratios is 2. If the objective is to estimate the population mean by taking with replacement random samples from within strata, compare the precision of the two allocations. 15

(b) Define precisely the simple random sampling with and without replacement. If an SRSWOR sample of size n is selected from a population of size N, then show that inclusion probabilities π_i and π_{ij} $(i \neq j)$ are respectively given by n / N and $n(n-1) / N(N-1)$.

A random sample of 150 boys from 6505 boys in an area showed that 31 had some kind of deficiency. Estimate the proportion of deficient boys in the population and find the standard error of your estimate. 10

(c) Why is the method of systematic sampling widely used in practice? Illustrate your answer with some practical situations. Show that the systematic sampling is more precise than simple random sampling if the variance within the systematic sample is larger than the population variance as a whole. What happens if the units in the same systematic sample have a positive correlation among them? 15

7. (a) Define instrumental variable estimator. Using instrumental variable method, obtain the estimate of the regression coefficient in a first-order auto-correlated regression model containing lagged endogenous variable Y_{t-1}. 15

(b) Derive rank condition for identification of a structural equation when there are zero restrictions on the structural parameters. 10

(c) Describe 2-stage least squares (2SLS) method of estimation. Show that 2SLS estimators are identical to indirect least squares (ILS) estimators if the equation is exactly identified. 15

8. (a) For a moving average process with weights $\{a_1, a_2, \cdots, a_m\}$ of random components $\{e_i, i=1, 2, \cdots\}$, where e_i's are iid $N(0, \sigma^2)$, obtain the correlogram function. Find its form, when all the weights are equal and their sum is 1. 15

(b) Show that if $X(t)$ is a real process, then its spectrum $h(\omega)$ can be put in the form

$$h(\omega) = \frac{\sigma_X^2}{2\pi} + \frac{1}{\pi} \sum_{t=1}^{\infty} k(t) \cos(\omega t)$$

where $k(\cdot)$ is the auto-covariance function and $\sigma_X^2 = k(0) = \text{var}(X(t))$. 10

(c) A sequence of price indices was constructed with 2001 as base year and from 2004, the base year was changed to 2003. Using the data given below, splice the first series (with base 2001) to second series (with base 2003) and then calculate new indices by shifting the base to 2004 : 15

Year	2001	2002	2003	2004	2005	2006
IN_{2001}	100	125	118·75	—	—	—
IN_{2003}	—	—	100	115	124·2	130·41

★ ★ ★