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.*

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

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.

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

*Answers must be written in **ENGLISH** only.*

SECTION A

Both the questions are compulsory.

Q1. (a) For SRSWOR (N, n), show that the sample proportion p is unbiased for the population proportion P. Also derive the sampling variance of this estimator. *10*

(b) What is the problem in estimating a linear regression model in presence of multicollinearity ? How is multicollinearity detected ? Explain how ridge estimation tackles this issue. *15*

(c) Consider the MA(1) process $X_n = e_n + \beta e_{n-1}$, where $e_n \sim N(0, 1)$.

For a data set it is noted that autocovariances are $\hat{\gamma}_0 = 1$ and $\hat{\gamma}_1 = -0{\cdot}25$.

(i) Estimate β. Which value of the estimate do you think we should choose and why ?

(ii) What problem do we have if $\hat{\gamma}_1 = -0{\cdot}5$? How would the variance of the error have affected the change ? *15*

Q2. (a) Given below are the figures on production (in thousand metric tons) of a cooperative sugar factory :

Year :	2010	2011	2012	2013	2014	2015	2016
Production :	77	88	84	85	91	98	90

(i) Fit a linear trend by least squares method. Tabulate the trend values.

(ii) Compute the monthly estimated increase in production during the period. *10*

(b) If, in every stratum, the simple estimator $\bar{y}_h$ is unbiased, then show that

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$$

is unbiased for population mean $\bar{y}$, where W_h is the proportion of population units in the strata and L denotes the total number of strata in the population.

Derive the sampling variance of $\bar{y}_{st}$ and state how you would unbiasedly estimate the same. *15*

(c) In the context of a finitely distributed lag model, discuss the problem of OLS estimation and suggest how to obtain good (consistent) estimates of the parameters in such a model by bringing in some restrictions on lag weights. *15*

SECTION B

Answer any *three* questions of the six questions given below.

Q3. (a) Explain and illustrate the following :

(i) Two-stage sampling

(ii) Two-phase sampling

Pinpoint the difference between the two types of sampling schemes. *10*

(b) Write briefly on

(i) Sample size determination in surveys;

(ii) Cumulative total method for PPSWR sampling;

(iii) Rao-Hartley-Cochran Scheme. *15*

(c) Discuss the following allocations of the sample size in stratified random sampling :

(i) Proportional allocation

(ii) Neyman allocation

(iii) Optimum allocation with a linear cost function

Explain the practical implications of these methods. *15*

Q4. (a) Discuss Koyck approach to an infinitely distributed lag model and obtain the mean lag for Koyck's model. What are the basic features of Koyck's transformed model ? *10*

(b) For the following linear regression model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad i = 1, ..., n$$

$$E(u_i) = 0,\ Cov(u_i, u_j) = 0 \quad i \neq j,\ E(u_i^2) = \sigma_i^2,$$

obtain the OLS estimates of β_1 and β_2 when σ_i^2, i = 1, ..., n are known. Discuss what steps could be taken when σ_i^2, i = 1, ..., n are unknown. Revise the least squares estimates of β_1 and β_2 when $\sigma_i^2 = \sigma^2 E^2(Y_i)$ where σ^2 is unknown. *15*

(c) Discuss the problem of estimating parameters by OLS in the presence of serial correlation in the following model :

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \; -1 < \rho < 1, \; \rho \text{ is known.}$$

$$E(\varepsilon_t) = 0, \; V(\varepsilon_t) = \sigma^2, \; Cov(\varepsilon_t, \varepsilon_{t+s}) = 0 \;\; s \neq 0.$$

Propose suitable estimates of β_1 and β_2. Also calculate the variance of the estimate of β_2. How can this estimate be modified when ρ is unknown ? *15*

Q5. (a) Demand and supply functions of a certain commodity are respectively

$$x_d = 240 + 10 \frac{dp}{dt} - 4p \quad \text{kg per month;}$$

$$x_s = 100 \frac{dp}{dt} + 6p - 60 \quad \text{kg per month,}$$

where p is price of the commodity at time t.

Find the time path of p for dynamic equilibrium if the initial price is to be ₹ 72 per kg. *10*

(b) Explain briefly the methods of computing price index numbers

(i) by simple average of price relatives;

(ii) by simple aggregate of prices; and

(iii) by weighted aggregate of prices. *15*

(c) Discuss the different forms of the Engel curve that are usually employed for fitting to family-budget data. In such fitting, how would you tackle the following complications ? *15*

(i) Household expenditure on a particular item depends, besides depending on income, on the number of persons per family.

(ii) Consumption of families of the same size differs because of varying age and sex consumption.

Q6. (a) Give an illustration for linear systematic sampling. Show that, under this method, a positive correlation between units in the same sample inflates the sampling variance of the estimator of population total. *10*

(b) Consider a population of N = 6 units with values 1, 2, 3, 4, 5 and 6.

(i) Write down all possible samples of size 2 drawn by SRSWOR scheme. Verify that the sample mean is unbiased for the population mean.

(ii) Also compute the sampling variance of the sample mean. *15*

(c) Explain the ratio method of estimation for estimating a population total. Show that it is generally biased. Evaluate the mean squared error of the estimator to the first order of approximation. Assume SRSWOR of n units from the population. *15*

Q7. (a) Using standard notations, briefly explain the instrumental variable technique in the context of estimating the coefficients in a linear regression model. State the situations when this technique is applicable. *10*

(b) State the rank and order conditions for identifiability of parameters in a system of structural equations. Which one of these two conditions is sufficient for identifiability ? Establish this condition mathematically. *15*

(c) Discuss the estimation of parameters of an equation appearing in a simultaneous equation system by Limited Information Maximum Likelihood method. State whether the estimator (if it exists) is unique. (An outline of the approach is adequate) *15*

Q8. (a) Show that the relationship $X_t = 0{\cdot}7X_{t-1} + 0{\cdot}3X_{t-2} + \varepsilon_t + 0{\cdot}7\varepsilon_{t-1}$ (where ε_t denotes white noise) defines ARIMA(1, 1, 1) model. *10*

(b) What do you understand by the seasonal variations in a time series ? Give example.

Explain the method of link relatives of computing the seasonal indices. *15*

(c) Define correlogram.

For an infinite series generated by the average of a random component with equal weights, show that the correlogram is

$$\rho_k = \begin{cases} 1 - \dfrac{k}{m} & \text{for } k \leq m \\ 0 & \text{for } k > m \end{cases}$$ *15*