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 in **TWO** Sections.

Candidate has to attempt **FIVE** questions in all.

Both the 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)* From the following data, calculate price index numbers for the current year by *(i)* Laspeyres' method, *(ii)* Paasche's method, *(iii)* Marshall-Edgeworth method and *(iv)* Fisher's ideal method :

Items	*Base year*		*Current year*	
	Quantity	*Price*	*Quantity*	*Price*
A	15	4	10	6
B	20	3	25	4
C	10	6	20	5
D	30	5	25	5

10

(b) Consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ E(\varepsilon_i) = 0,\ V(\varepsilon_i) = \sigma_i^2,\ i = 1, 2, \ldots, n$$

Obtain the variance of ordinary least squares estimate (OLSE) and generalized least squares estimate (GLSE). Show that var (GLSE) ≤ var (OLSE).

Also, show that the efficiency of OLSE and GLSE depends on the correlation between $\sigma_i(x_i - \bar{x})$ and $\dfrac{x_i - \bar{x}}{\sigma_i}$.

Discuss briefly the estimation procedure in the two cases when—

(i) $\sigma_i^2,\ i = 1, \ldots, n$ are known;

(ii) $\sigma_i^2,\ i = 1, \ldots, n$ are partially unknown (viz., $\sigma_i^2 \propto \sigma^2 X_i^{2\lambda}$). 15

(c) With two strata, a sampler would like to have $n_1 = n_2$ for administrative convenience, instead of using values given by Neyman allocation. If $V(\bar{y}_{st})$, $V_N(\bar{y}_{st})$ denote the variances given by $n_1 = n_2$ and Neyman allocation respectively ignoring fpc, show that the fractional increase in variance

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

where $r = \frac{n_1}{n_2}$ as given by Neyman allocation. Samples of size n_1 and n_2 have been drawn by simple random sampling without replacement (SRSWOR) from two strata. 15

2. *(a)* How can one overcome the difficulty of changing expectation with each draw in sampling with varying probability and without replacement? Explain with examples. 10

(b) Consider AR(2) given by $X_t = X_{t-1} - 0 \cdot 5X_{t-2} + w_t$, where w_t is white noise $(0, \sigma^2)$.

(i) Is the process weak (covariance) stationary? Give reasons for your answer.

(ii) Calculate corr (X_1, X_2) and corr (X_5, X_7).

If the coefficient of X_{t-1} is 1·1 and that of X_{t-2} is – 0·18, is the AR(2) process still covariance stationary? 15

(c) What are autoregressive and distributed lag models? Explain Koyck's approach of estimating distributed lag models. 15

SECTION—B

3. *(a)* In case of ratio estimator, when does the concept of two ratios arise? Explain with the help of an example. Derive the expression for covariance between two estimates of these ratios which are correlated and have different denominators. Also, write down the estimate of this covariance. 10

(b) Define product and regression estimators of the population mean. Derive their mean squared errors to the first order of approximation and compare their efficiencies. 15

(c) In a village, there are 8 orchards serially numbered from 1 to 8 with 350, 440, 200, 400, 500, 250, 300, 260 trees respectively. Their yields (in 10 kg) are respectively 400, 500, 220, 440, 600, 300, 350, 300. A sample of 2 orchards with PPSWOR was drawn. The orchards bearing serial numbers 8 and 3 were selected in the sample. Calculate Des Raj ordered estimator of population total and estimate its standard error. 15

4. (a) Explain the Durbin-Watson (D-W) test for autocorrelation and give the conclusions. Also, mention the basic assumptions underlying the D-W d-statistic. 10

(b) Following is the data relating to the regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + e_i$$

where Y = per capita consumption expenditure, X_2 = per capita disposable income and X_3 = time.

$N = 15,\ \bar{Y} = 1942 \cdot 333,\ \bar{X}_2 = 2126 \cdot 333,\ \bar{X}_3 = 8 \cdot 0$

$\Sigma(Y_i - \bar{Y})^2 = 830121 \cdot 333,\quad \Sigma(X_{2i} - \bar{X}_2)^2 = 1103111 \cdot 333,\quad \Sigma(X_{3i} - \bar{X}_3)^2 = 280$

$$(X'X)^{-1} = \begin{bmatrix} 37 \cdot 237491 & -0 \cdot 0225079 & 1 \cdot 3366965 \\ -0 \cdot 0225079 & 0 \cdot 0000137 & -0 \cdot 0008319 \\ 1 \cdot 3366965 & -0 \cdot 0008319 & 0 \cdot 054034 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 29135 \\ 62905821 \\ 247934 \end{bmatrix}$$

Find the least square estimate of β say $\hat{\beta}$, variance-covariance matrix of $\hat{\beta}$, R^2 and adjusted R^2 ($\bar{R}^2$). 15

(c) Discuss the identifiability of each of the following simultaneous equations models with the structural equations :

(i) $y_1 = \beta_{13} y_3 + \gamma_{11} x_1 + \gamma_{13} x_3 + \varepsilon_1$

(ii) $y_1 = \gamma_{21} x_1 + \gamma_{23} x_3 + \varepsilon_2$

(iii) $y_2 = \beta_{33} y_3 + \gamma_{31} x_1 + \gamma_{32} x_2 + \varepsilon_3$

15

5. *(a)* Calculate the Fisher's ideal price index from the following data and show that it satisfies both time reversal test and factor reversal test :

Commodity	*Base year*		*Current year*	
	Price	*Total value*	*Price*	*Total value*
A	6	300	10	560
B	2	200	2	240
C	4	240	6	360
D	10	300	12	288
E	8	320	12	432 .

10

(b) Explain price elasticity of demand. Give reasons for the negative values of price elasticity of demand in most situations.

Suppose the price elasticity of demand for a commodity is – 0·3. If there is a 5% increase in the price of the commodity, by what percentage will the demand for the commodity go down? 15

(c) Show that the two MA(1) processes

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

$$Y_t = \eta_t + \frac{1}{\theta}\eta_{t-1}$$

have the same autocorrelation function. Here $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, $\eta_t \sim WN(0, \sigma_\eta^2)$, ε_t and η_t are independent. 15

6. *(a)* Explain briefly the objectives of forming NSSO and its current functions. 10

(b) In two-stage sampling with first stage units of equal size M, if n first stage units out of N and then m subunits from each of the first stage units are selected by simple random sampling without replacement, show that sample mean is an unbiased estimator of population mean. Find the variance of sample mean and its unbiased estimate. 15

(c) Explain cumulative total method for the selection of a PPS sample. Obtain a BLUE of population total in a PPS sample of size n. 15

7. *(a)* Briefly mention about limited-information maximum likelihood method for simultaneous equations models. Write down some important relationships between limited-information maximum likelihood method and two-stage least squares method. 10

(b) Consider the following regression model :

$$y = X\beta + u, \quad u \sim N(0, \sigma^2 I_n)$$

where $y(n \times 1)$, $X(n \times p)$, $\beta(p \times 1)$ and $u(n \times 1)$. Further cov $(X, u) \neq 0$.

(i) Show that OLSE is inconsistent.

(ii) Suppose that $Z(n \times p, n > p)$ is an instrumental variable (IV) matrix. Z is independent of u but is highly correlated with X. Using Z, derive an appropriate estimator of β, so that the estimator is consistent for β and asymptotically $(n \to \infty)$ normal. 15

(c) *(i)* In a simple regression model justify whether the OLS estimators in the presence of perfect multicollinearity are BLUE or not. 5

(ii) In a simple regression model does a high value of R^2 indicate a definite evidence of high collinearity? Explain. 5

(iii) Explain any two remedial measures of multicollinearity in a simple regression model. 5

8. *(a)* If the ratio between Laspeyres' and Paasche's index numbers is 28/27, find the missing figure (x) in the following table :

Commodity	*Base year*		*Current year*	
	Price	*Quantity*	*Price*	*Quantity*
A	1	10	2	5
B	1	5	x	2

10

(b) Consider ARIMA (1, 1) process given by

$$X_t = 0 \cdot 5 X_{t-1} + W_t - 0 \cdot 3 W_{t-1}$$

where W_t is the white noise $(0, \sigma^2)$. Since the process is weakly stationary, it has the Wold representation (in terms of usual notations $\psi_0, \psi_1, \psi_2, \ldots$). Calculate the first four coefficients in the representation. 15

(c) Explain briefly :

(i) Engel curves

(ii) Lorenz curves

(iii) Gini coefficient 15

★ ★ ★