

# 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

**All questions of this section are compulsory.**

**Q1.** (a) Explain sampling and non-sampling errors. State the sources of non-sampling errors. 10

(b) Consider the three-variable regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

where  $X_1$  and  $X_2$  are two independent variables,  $Y$  is the dependent variable,  $\beta$ 's are the coefficients and  $\varepsilon$  is the random component,  $\varepsilon \sim N(0, \sigma^2)$ . Obtain the estimates of parameters using OLS method of estimation. 15



- (c) The following figures give the production of a commodity of a country from 2001 to 2007 :

Year	Production (in thousand metric tonnes)
2001	4.7
2002	5.9
2003	7.2
2004	7.9
2005	7.6
2006	7.0
2007	7.9

Obtain the trend values by a four-year moving average method.

15

- Q2.** (a) The following data show the cost of living indices for different groups along with their weights (% of expenditure), for middle-class people of some city in 2010. Obtain the general cost of living index number. Mr. X was getting a salary of ₹ 25,000 in 2010 and ₹ 82,900 in 2016. Calculate how much he ought to have received as extra allowance in 2016 to maintain his same standard of living.

10

Group	Base year group index	Weight
Food	396	71.3
Clothing	550	2.7
Fuel and Light	360	9.3
House rent	110	6.7
Miscellaneous	290	10

- (b) A simple random sample of 25 households was drawn from a city area containing 13745 households. The number of persons per household in the sample were as follows :

5, 6, 3, 4, 2, 3, 3, 3, 4, 4, 3, 2, 7, 4, 7, 6, 4, 4, 1, 4, 3, 4, 5, 6, 3.

Estimate the total number of people in the area and also compute the variance of the estimate.

15

- (c) Discuss the concept of heteroscedasticity. Explain this by graphical representation. Show that OLS estimators are unbiased even under the condition of heteroscedasticity. Obtain its variance.

15



## SECTION B

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

- Q3.** (a) A finite population of size 98 is divided into three strata. It is given that  $N_1 = 2N_2 = 4N_3$  and  $S_1 = 2S_2 = 4S_3$ . If a sample of 21 is to be selected from the population, obtain the number of units to be selected from each stratum under Neyman allocation. 10
- (b) Define linear systematic sampling scheme with  $N = nk$ . Obtain the expression for the variance of the sample mean in this case. When will this be more advantageous than SRSWOR scheme? 15
- (c) Define Horvitz-Thompson unordered estimator  $\hat{Y}_{HT}$  for the population total  $Y$ . Obtain an expression for  $V(\hat{Y}_{HT})$ . Comment on the estimator. 15
- Q4.** (a) Explain generalized least squares method of estimation. Write its model along with assumptions. Obtain the estimates of parameters and their variances. 15
- (b) Discuss the meaning of Autocorrelation. Give a comparison between autocorrelation and serial correlation with an example. Give example of variance-covariance matrix of size 3 to show
- (i) heteroscedasticity with no autocorrelation
- (ii) homoscedasticity with autocorrelation
- (iii) heteroscedasticity with autocorrelation.
- Prove that mean of autocorrelated  $U$ 's turns out to be zero. 15
- (c) For general linear model, show that least squares estimator  $\hat{\beta}$  is the best linear unbiased estimate of  $\beta$ . 10
- Q5.** (a) Two stationary time series  $u_t$  and  $u'_t$  are added together and a new series is formed by  $v_t = u_t + u'_t$ . If  $u_t$  and  $u'_t$  are independent, show that the autocorrelation of lag  $\kappa$  of  $v_t$  is given by
- $$\frac{\rho_\kappa \text{Var}(u_t) + \rho'_\kappa \text{Var}(u'_t)}{\text{Var}(u_t) + \text{Var}(u'_t)}$$
- where  $\rho_\kappa$  and  $\rho'_\kappa$  refer to the autocorrelation of  $u_t$  and  $u'_t$  respectively. 15



- (b) State the law of demand.

Consider the following econometric model :

$$q_i \text{ (quantity demanded)} = a_0 + b_0 p_i + u_i \quad (b_0 < 0)$$

$$q_i \text{ (quantity supplied)} = a_1 + b_1 p_i + v_i \quad (b_1 > 0)$$

the errors  $u_i$  and  $v_i$  being uncorrelated.

Find the slope of the regression line  $q_i$  on  $p_i$ .

15

- (c) What is periodogram ?

A time series  $u_t$  is composed of one periodic with period  $\lambda$  and amplitude  $a$  and the other term an irregular component ( $b_t$ )

$$u_t = a \sin \frac{2\pi t}{\lambda} + b_t$$

Describe a method for finding the true cyclical period  $\lambda$ .

Indicate how you will modify the method if the cyclical component is composed of several periodic terms  $\lambda_1, \lambda_2, \dots, \lambda_k$ .

10

- Q6.** (a) Explain double sampling scheme. Give an example. Obtain the ratio estimator for the population mean of the study variate when the mean of the auxiliary variate is unknown.

10

- (b) In a two-stage sampling, the first and the second stage units are drawn by adopting SRSWOR. Suggest an unbiased estimator for the population mean. Derive an expression for its variance.

(You may assume variance of sample mean under SRSWOR).

15

- (c) State the linear regression estimate for population mean. Obtain the estimate of the regression coefficient by least square method. Show that this estimate is unbiased. Under what condition is the estimate unbiased ?

15

- Q7.** (a) What is the need of simultaneous equation system in economic theory ? Discuss Keynesian model as a structural equation model of a national economy. Obtain its reduced form equations and hence reduced form disturbance term.

15

- (b) Let  $Y_1$  and  $Y_2$  be two endogenous variables. Let further  $X_1, X_2$  and  $X_3$  be three predetermined variables. A simultaneous equation is expressed as

$$Y_1 = \gamma_{21} Y_2 + \beta_{11} X_1 + \varepsilon_1$$

$$Y_2 = \gamma_{12} Y_1 + \beta_{22} X_2 + \beta_{32} X_3 + \varepsilon_2$$

Discuss a method to identify the model.

15

- (c) Discuss full information maximum likelihood method of estimation. Discuss only how to find the normal equations.

10



Q8. (a) Define a stationary time series.

For the stationary AR(2) process

$$X_t = \frac{5}{6} X_{t-1} - \frac{1}{6} X_{t-2} + e_t$$

where  $e_t \sim N(0, \sigma_e^2)$  denotes the noise.

Find the autocorrelation function of lags 1 and 2.

10

(b) Describe the different schemes for explaining the oscillations in a stationary time series. Explain the use of correlograms for discriminating between the above schemes.

15

(c) Define Gini coefficient.

The following table presents data on distribution of personal income by decile groups of households for each group separately for both rural and urban sectors of a country for the year 2010.

On a graph paper, draw the Lorenz curve for each sector.

Also, compute the Gini coefficient for each sector.

Hence compare the income inequality of two sectors.

15

Household	Percentage share in total income	
	Rural	Urban
0 – 10	3.0	2.3
10 – 20	4.4	3.3
20 – 30	5.4	4.1
30 – 40	6.4	4.8
40 – 50	7.4	5.8
50 – 60	8.3	7.0
60 – 70	9.6	8.7
70 – 80	10.9	10.0
80 – 90	13.8	14.0
90 – 100	30.8	40.0

