

**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 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 Booklet must be clearly struck off.*

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

## SECTION A

**Both the questions are compulsory.**

- Q1.** (a) Compare Simple Random Sampling Without Replacement (SRSWOR) and Simple Random Sampling With Replacement (SRSWR) and find the value of  $n$  such that variance of the sample mean in SRSWOR is exactly half of the variance of the sample mean in SRSWR of the same size. 10
- (b) Consider the simple linear regression model :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n;$$

$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  are independent and identically distributed with mean zero and constant variance  $\sigma^2$ . Show that least square estimates  $\hat{\beta}_0$  and  $\hat{\beta}_1$  are linear functions of  $y_1, y_2, \dots, y_n$  and also compute

variance-covariance matrix of  $\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$  and its determinant. 15

- (c) Let  $Y_t = Z_t + \theta Z_{t-1}$ , where  $\{Z_t\}$  be a sequence of iid random variables with mean zero and variance  $\sigma_Z^2$ . Show that a real valued function on  $Z$ , defined as :

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

is an autocovariance function if  $|\rho| < \frac{1}{2}$ . 15

- Q2.** (a) Find the spectral density function  $f(t)$  of a continuous parameter process, having correlation function  $\rho(t) = e^{-t^2}$ ,  $-\infty < t < \infty$ . Also find the value of  $f(\sqrt{\log 16})$ . 10
- (b) From bivariate population of  $N$  units, a simple random sample  $(x_i, y_i); i = 1, 2, \dots, n$  is drawn without replacement with corresponding means  $(\bar{x}_n, \bar{y}_n)$ . Show that  $\text{Cov}(\bar{x}_n, \bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{xy}$ . 15
- (c) Explain the term exponential smoothing. When is exponential smoothing most useful? Interpret the smoothing constant  $\alpha$ , what is its range? How is  $\alpha$  related to degree of smoothing? 15



## SECTION B

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

- Q3.** (a) A simple random sample of  $n$  clusters, each containing  $M$  elements, is drawn from the  $N$  clusters in the population. Then show that the sample mean per element  $\bar{y}$  is an unbiased estimate of  $\bar{Y}$  with variance

$$V(\bar{y}) = \frac{1-f}{nM} S^2 [1 + (M-1)\rho],$$

where  $f$  = sampling fraction and  $\rho$  is the intracluster correlation coefficient. 10

- (b) Using first approximation to variance of ratio estimator  $R_n$ , show that

$$\frac{|\text{Bias in } R_n|}{\sqrt{\text{Var}(R_n)}} \leq \frac{\sqrt{V(x_n)}}{\bar{x}_N}. \quad 15$$

- (c) A village has five orchards, containing 15, 30, 25, 10 and 20 trees respectively. If the yields (in 10 kg) of these 5 orchards are 18, 35, 29, 12, and 24 respectively and selecting sample of two units at 2<sup>nd</sup> and 4<sup>th</sup> position, estimate the total production of five orchards along with standard error using Horvitz-Thompson estimator. 15

- Q4.** (a) Let  $(X, Y)$  have the joint pdf given by

$$f(x, y) = \begin{cases} 1 & \text{if } |y| < x, \quad 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the regression of  $Y$  on  $X$  is linear but regression of  $X$  on  $Y$  is not linear. 15

- (b) Given  $X'X = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 10 \end{pmatrix}$  and  $X'Y = \begin{pmatrix} 10 & 20 \\ 20 & 10 \\ 30 & 20 \end{pmatrix}$ , estimate the model

$$y_{1t} = \beta_{12} y_{2t} + \gamma_{11} x_{1t} + \gamma_{12} x_{2t} + u_{1t}$$

$$y_{2t} = \beta_{21} y_{1t} + \gamma_{23} x_{3t} + u_{2t}$$

using 2SLS method. 15

- (c) State the methods of detecting presence of Heteroscedasticity. Discuss any one of them. 10

- Q5.** (a) Define Laspeyres' index number and Paasche's index number. If  $L(p)$  and  $P(q)$  respectively represents Laspeyres' index number for price and Paasche's index number for quantity, then show that

$$L(p)/L(q) = P(p)/P(q). \quad 15$$

- (b) Define autoregression series of order  $k$ . Consider the autoregression process  $U_t = a\xi + \varepsilon_t$ , where  $-\infty < t < \infty$ ,  $\varepsilon_t, \varepsilon_{t+1}, \dots$  and  $\xi$  be independent variables with zero mean and unit variance. Show that the process is stationary with correlation  $\rho_1 = \rho_2 = \dots = \frac{a^2}{1+a^2}$ .

15

- (c) Explain Time Series model and its components. 10

- Q6.** (a) Explain Stratified Random Sampling method and the problem associated with stratification. Also write down the advantages of Stratified Random Sampling. 15

- (b) Explain double sampling plan for attributes and derive the expression of OC-curve in double sampling plan. 15

- (c) What is multicollinearity? Discuss the effect of multicollinearity using 3-variate linear regression model. 10

- Q7.** (a) For a Markov process

$$u_t = \rho u_{t-1} + v_t \text{ with } |\rho| < 1$$

and random variables  $v_t$  are such that  $E(v_t) = 0$ ,  $\text{Var}(v_t) = \sigma_v^2$  and  $\text{Cov}(v_t, v_s) = 0$  for  $t \neq s$ .

Show that :

15

(i) 
$$u_t = \sum_{r=0}^{\infty} \rho^r v_{t-r}$$

(ii) 
$$\text{Var}(u_t) = \frac{\sigma_v^2}{1-\rho^2}$$

(iii) 
$$\text{Cov}(u_t, u_{t-s}) = \frac{\rho^s}{(1-\rho^2)} \sigma_v^2$$

- (b) Discuss Durbin-Watson test for Autocorrelation. The data of the following table are the OLS residuals of a consumption function :

$$\hat{C}_t = -3.02 + 0.93 Y_t$$

Calculate Durbin-Watson d-statistic. Write your conclusion.

10

Year	$e_t$
1994	0.6
1995	1.9
1996	-1.8
1997	-2.7
1998	-2.9
1999	1.4
2000	3.3
2001	0.3
2002	0.8
2003	2.3
2004	-1.4
2005	-1.1

(Table for d-statistic significance points is attached)



For Q.No. 7(b)

The Durbin-Watson  $d$ -Statistic  
Significance Points of  $d_L$  and  $d_U$  5%

$n$	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	1.09	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.47	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

**Note :**  $k'$  = Number of explanatory variables excluding the constant term.  
 $n$  = Number of observations.

- (c) State the conditions of identification for structural form of the system of simultaneous equations.

Discuss the identification of the following model, assuming Y's as endogenous and X's as predetermined variables :

15

$$Y_1 = \alpha_{10} + \alpha_{12} Y_2 + \alpha_{13} Y_3 + \beta_{11} X_1 + u_1$$

$$Y_2 = \alpha_{20} + \alpha_{23} Y_3 + \beta_{21} X_1 + \beta_{22} X_2 + u_2$$

$$Y_3 = \alpha_{30} + \alpha_{31} Y_1 + \beta_{31} X_1 + \beta_{32} X_2 + u_3$$

$$Y_4 = \alpha_{40} + \alpha_{41} Y_1 + \alpha_{42} Y_2 + \beta_{43} X_3 + u_4$$

- Q8.** (a) Explain AR (p), MA (q), ARMA (p, q) and ARIMA (p, d, q) processes. How would you find out the appropriate values of p, d and q while modelling the given time series ? State the procedure to estimate the parameters of the ARIMA model.

15

- (b) Define Dickey-Fuller (DF) test. How would you use DF test to find out if the given time series contains a unit root ? If a unit root exists, how would you characterize such a time series ?

15

- (c) Explain autocovariance and autocorrelation functions. If  $y_1, y_2, \dots, y_n$  are n observations made at n successive time points of a stationary process, then in usual notations define autocovariance and autocorrelation matrix of order n. For  $n = 3$ , show that

$$-1 \leq \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \leq 1.$$

10