

A-HDR/HRR-N-TUC

STATISTICS

Paper—III

Time Allowed : Three Hours

Maximum Marks : 200

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.

Question no. 1 and 5 are compulsory and out of the remaining, THREE are to be attempted choosing at least ONE from each Section.

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

Unless other-wise mentioned, symbols and notations have their usual standard meanings.

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

Candidates should attempt questions/parts as per the instructions given in the Section.

All parts and sub-parts of a question are to be attempted together in the answer book.

Attempts of questions shall be counted in chronological 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 answer book must be clearly struck off.

Answers must be written in ENGLISH only.

SECTION—A

1. Answer any **FIVE** parts of the following : $8 \times 5 = 40$
- (a) You are asked to plan a sample survey to find out the incidence of educated unemployment in your State. Give details of your plan including an outline of a suitable questionnaire for this purpose. 8
- (b) In stratified population with two strata, the values of W_h , S_h and C_h are given. If the cost function is of the form $C = c_1 n_1 + c_2 n_2$, obtain expression for $\frac{n_h}{n}$, $h = 1, 2$; that minimize C for given value of $V(\bar{y}_{st})$; where the symbols have their usual meanings. 8
- (c) 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 and the standard error of your estimate. 8
- (d) What is meant by the design of experiment ? Describe the situations in which the three basic designs, namely CRD, RBD and LSD are appropriate, and state the rule or rules by which treatments are allocated to units in these designs. 8
- (e) What do you understand by a split-plot design ? When will such a design be adopted ? Distinguish it from a usual confounded factorial experiment. 8
- (f) Obtain necessary conditions for a symmetric BIBD with an even number of treatments to exist. 8

2. (a) From the following analysis of variance table of a cluster sample estimate its efficiency relative to simple random sampling. Assume that the clusters are of equal size and the number of clusters in the population is very large :

Source	S.S.	d.f.	
Between clusters	40	5	
Within clusters	96	24	10

- (b) What are the situations in which ratio estimate is used ? Obtain an unbiased ratio type estimator for the population mean. 10
- (c) Suppose a simple random sample of size 2 is drawn from a finite population (y_1, y_2, y_3) . Corresponding to the three possible samples $S_1 = (y_1, y_2)$, $S_2 = (y_2, y_3)$ and $S_3 = (y_1, y_3)$ a linear estimator $l(S)$ for estimating the population mean is defined as follows :

$$l(S_1) = \frac{2}{3} y_1 + \frac{1}{2} y_2, \quad l(S_2) = \frac{1}{2} y_2 + \frac{1}{2} y_3,$$

$$l(S_3) = \frac{1}{3} y_1 + \frac{1}{2} y_3.$$

- (i) Show that the estimator $l(S)$ is unbiased estimator of population mean and obtain its variance.
- (ii) If $y_1 = 1$, $y_2 = 2$ and $y_3 = 3$ show that $V(l(S)) < V(\bar{y})$; where \bar{y} is the mean of sample of size 2. 10
- (d) A population consists of 6 units with sizes of the auxiliary variable as 12, 33, 45, 84, 16, 10. Select a PPS sample of size 2, given that the two selected random numbers from 1 to 200 are 27 and 186. Also state unbiased estimator of the population total and derive its variance based on the selected random sample. 10.

3. (a) What is meant by analysis of covariance (ANCOVA) ? Discuss the test procedure to test the necessity of the use of ANCOVA in one way and two-way classifications. 10
- (b) Let N be the incident matrix of a BIBD with parameters v, b, r, k, λ .
- (i) Show that NN' is non invertible matrix if $k = v$.
- (ii) Show that the characteristic roots of NN' are rk and $r - \lambda$ of multiplicative 1 and $v - 1$ respectively. 10
- (c) Construct balanced group of sets for 2^4 factorial, blocks of 4 units in which CD, ABC, ABD are confounded. 10
- (d) Explain clearly the purposes of replications, randomizations and local control in design of experiments. 10
4. (a) Define linear systematic sampling. Does it give equal probability of selection to all the population units ? State the advantages and limitations of systematic sampling. Can a systematic sample be drawn without a physical list of population units ? Justify your claim. 10
- (b) Consider a population $U = \{1, 2, 3\}$ with the following design $P(\{1, 2\}) = 1/2, P(\{1, 3\}) = 1/4 = P(\{2, 3\})$. Give the first order inclusive probability. Hence check whether the design is fixed size design. Give the variance-covariance matrix of indicator variables for inclusion in the sample. 10

- (c) What is an incomplete block design ? When is it said to be balanced ? Explain how interblock information can be recovered in a balanced incomplete block design. 10
- (d) Define a Latin square design. Show how you split up the total sum of squares into components for this design. Give the analysis of variance table. 10

SECTION—B

5. Answer any **FIVE** parts of the following : $8 \times 5 = 40$
- (a) Define price elasticity of demand and income elasticity of demand. Point out their uses in economic analysis. 8
- (b) Let d_1 and d_2 represent the demand of a commodity for two strata of a population. If η_1 and η_2 be the elasticities of demand with respect to national income for the two strata. Show that the corresponding elasticity η for the two strata combined would be given by :
- $$\eta = (\eta_1 d_1 + \eta_2 d_2) / (d_1 + d_2). \quad 8$$
- (c) Define Laspeyres' and Paasche's index numbers. It is sometimes stated that Laspeyres' price index number tends to overestimate price changes while Paasche's price index number tends to underestimate them. Justify or refute the above statement giving reasons. 8

- (d) Describe General Linear Model in K independent variables. Justify the inclusion of a random error in this model. 8
- (e) Explain the types of multicollinearity that arise in econometric investigations. What are the implications of this problem (i) for prediction of dependent variable and (ii) for the estimation of regression coefficients ? 8
- (f) Discuss how by pooling the time series and cross section data, multicollinearity can be resolved. 8
6. (a) What are the two mathematical models employed for time series analysis ? Can one model be considered as a particular type of the other one ? Which one of the two models is considered to be more useful and why ? Discuss each of the above aspects. 10
- (b) Define oscillatory times series. State the difference between cyclical time series and oscillatory time series. State the reasons which are responsible to produce oscillatory time series. Can oscillatory time-series be said to be stationary time series ? 10
- (c) Discuss how you will proceed for constructing a cost of living index number for a given expenditure group in a city. 10
- (d) What is periodogram ? Discuss how the periodogram becomes helpful to determine the periodicity hidden in the time series. 10

7. (a) In usual notation, consider the standard linear model : $Y = X\beta + U$, $U \sim N(0, \sigma^2 I)$. Show that the MLE $\hat{\beta}$ of β have the distribution $N(\beta, \sigma^2(X'X)^{-1})$, assuming $X'X$ to be an invertible matrix. 10
- (b) Discuss the problem of heteroscedasticity in GLM. Describe Glejser test for detecting heteroscedasticity. What are the difficulties in using Glejser test ? How do you overcome these difficulties ? 10
- (c) Explain the term 'autocorrelation'. What are the consequences of autocorrelation ? Explain how the Durbin-Watson d statistic is used to detect the presence of autocorrelation. 10
- (d) Describe the problem of 'multicollinearity' in econometrics and explain how will you detect it. 10
8. (a) Define Fisher index number. Show why Fisher index number is said to be ideal index number. Also show why Laspeyres' and Paasche's index number are not ideal one. 10
- (b) For an infinite series generated by the moving average of a random component with equal weights, the correlogram is given by :

$$r_k = \begin{cases} 1 - \frac{k}{m}, & \text{for } k \leq m \\ 0, & \text{for } k > m. \end{cases} \quad 10$$

(7)

(Contd.)

- (c) What is Ridge regression ? Show that Ridge regression estimator can not only resolve the deadlock of multicollinearity but it can also give an efficient estimator even though it is a biased estimator. 10
- (d) After listing the underlying conditions, show that the indirect least squares method of estimation and the 2-SLS method of estimation are equivalent. 10