**Serial No.** 0006811     A–LVV–O–UVC

# 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 from the following : 5×8=40

   (a) Show that for SRSWOR, the sample proportion is an unbiased estimator of the population proportion of members possessing a certain character. Hence obtain the variance of the estimator.      8

   (b) Explain the principles of forming strata and clusters. What are the differences between stratified sampling and cluster sampling ?      8

   (c) (i) Explain why the variance of the estimate of population mean based on a single systematic sample is not unbiasedly estimable.

       (ii) If two independent systematic samples of size four each from a population of 36 units have y-values of a characteristic y as

           24, 26, 22, 28

       and 27, 25, 23, 29

       respectively, calculate an unbiased estimate of population mean of y and an unbiased estimate of its variance.      8

   (d) Explain briefly the main principles of design of experiments and their analogy in sample surveys with the corresponding concepts.      8

(e) Discuss how you would proceed with the analysis if data on one plot is missing in a Latin Square Design. **8**

(f) Suppose an experimenter wants to conduct a $3^3$ experiment in blocks of size 9 plots. Write down the treatments to 3 blocks with $AB^2$ as the confounding contrast. **8**

2. (a) Consider a population $U = \{u_1, u_2, u_3, u_4\}$. The values of the study variable are $y(u_i) = i, i = 1(1)4$.

(i) A sample of size 2 is drawn from the population with (i) SRSWOR (ii) SRSWR. Calculate $V_1 = \text{Var}(\bar{y}|\text{SRSWOR})$ and $V_2 = \text{Var}(\bar{y}|\text{SRSWR})$ where $\bar{y}$ denotes the sample mean.

(ii) A linear systematic sample of size 2 is drawn when the units of the population are arranged as follows :

Arrangement 1 : $u_1, u_4, u_2, u_3$
Arrangement 2 : $u_1, u_2, u_4, u_3$
Calculate $\text{Var}(\bar{y}|\text{arrangement } 1) = v_3$ and $\text{Var}(\bar{y}|\text{arrangement } 2) = v_4$.

(iii) Show that $v_4 < v_1 < v_2 < v_3$. **10**

3

(Contd.)

(b) (i) For any sampling design, describe the ratio method of estimating the population total Y of a study variate y, when auxiliary information on a related variable x is available with a known population total X. Denote this estimator by $\hat{Y}_R$ and derive its approximate bias.

(ii) Consider two ratio estimators $r_1$ and $r_2$ for a parameter $\theta$ based on a sample of size n with biases $B(r_1) = b_1$ and $B(r_2) = b_2 = nb_1$. Using this information, obtain an unbiased estimator of $\theta$ and its variance expression. 10

(c) Consider a population $U = \{1, 2, 3, 4, 5\}$. For a sampling design, probabilities of samples drawn are $P_r\{(1, 2)\} = 0.2$, $P_r\{(2, 3, 4)\} = 0.1$, $P_r\{(1, 4, 5)\} = 0.3$, $P_r\{(2, 4, 5)\} = 0.2$, $P_r\{(1, 2, 3, 4)\} = 0.2$. Calculate the first-order inclusion probabilities. Obtain estimates $\sum_{i \in S} \dfrac{l_i y_i}{\pi_i}$ of the parametric function $\sum_{i=1}^{N} l_i y_i$ where $l_i$'s are known real numbers, not all zero. 10

(Contd.)

(d) For a two-stage sampling design, let n villages be selected with probabilities proportional to a given size and with replacement from a population of N villages. From the $i^{th}$ selected village consisting of $M_i$ households (hh), $m_i$ households are selected by simple random sampling without replacement. Write down an unbiased estimator

of $Y = \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{M_i} Y_{ij}$, where $Y_{ij}$ is the value of a

study variate for the $j^{th}$ hh of the $i^{th}$ village.

Also write down an unbiased estimator of Var $(\hat{Y})$.

10

3. (a) Describe how uniformity trials in experimental designs and pilot studies in sample surveys are conducted, mentioning their uses. 10

(b) Find the C-matrix for the following design and obtain independent estimable treatment contrasts :

$B_1 = (1, 2, 2)$, $B_2 = (2, 4, 4)$, $B_3 = (1, 1, 2)$,
$B_4 = (3, 4, 4)$. 10

(c) Explain what is meant by a split-plot design. Suppose that factor A has p levels which are arranged in a Randomized Block Design having r replicates. Let factor B have q levels which are

(Contd.)

applied to plots of a block after subdividing each plot into q sub plots. Write down the model, clearly explaining the notations and assumptions and present a blank ANOVA table with sources of variation and degrees of freedom. 10

(d) State the conditions to be satisfied by a Partially Balanced Incomplete Block (PBIB) design.

Prove the following restrictions of a PBIBD :

(i) $\sum_{j=1}^{m} n_j = v - 1;$

(ii) $r(K-1) = \sum_{j=1}^{m} n_j \lambda_j$ 10

4. (a) What are the advantages of stratified sampling ? Consider the allocation of sample size $n_i$ to strata given by $n_i \propto N_i \sigma_i^{\delta}$, where $N_i$ is the stratum size, $\sigma_i$ is the within stratum standard deviation of the i th stratum, (i = 1, 2, ...., k) and $\delta$ is real. Write down an expression for the $Var(\hat{\bar{Y}}_{st.})$ where $\hat{\bar{Y}}_{st.}$ stands for the unbiased estimator of the population mean $\bar{Y}$ of the study variate y based on Simple Random Sampling without replacement in each stratum. 10

6

(b) Consider a population $U = \{u_1, u_2, \ldots, u_N\}$. Observations of study variable (y) and auxiliary variable (x) are $y(u_i) = y_i$, $x(u_i) = x_i$, $i = 1, 2, \ldots, N$.

Let $p_i = \dfrac{X_i}{X}$ where $X = \displaystyle\sum_{i=1}^{N} X_i$. A sample of size n is drawn with PPSWR. Show that for population total $T = \displaystyle\sum_{1}^{N} y_i$, the estimator $\hat{Y}_{pps} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \dfrac{y_i}{p_i}$ is an unbiased estimator. Obtain the variance of the estimator of the population total. **10**

(c) Verify whether the following arrangement is a symmetrical BIB design. Consider each set a block :

(12, 0, 2, 8), (6, 7, 9, 2), (4, 5, 7, 0), (5, 6, 8, 1), (11, 12, 1, 7), (3, 4, 6, 12), (10, 11, 0, 6), (9, 10, 12, 5), (2, 3, 5, 11), (1, 2, 4, 10), (8, 9, 11, 4), (7, 8, 10, 3), (0, 1, 3, 9).

Is it a connected BIBD ? **10**

(d) 32 plots are arranged in the form of a $4 \times 8$ rectangle. Give a design for $2^5$ factorial experiment with factors A, B, C, D and E in these plots confounding the effects ABC, CDE, ABDE with the rows and AB, CD, ABCD, EDE, ADE, BCE, ACE with the columns. **10**

(Contd.)

5. Answer any **FIVE** parts from the following : 5×8=40

   (a) List the main components of a time series. Explain the method of link relative for measurement of seasonal fluctuations of a time series. 8

   (b) Discuss the considerations in the choice of (i) base period and (ii) the formula (method), for constructing price index numbers. 8

   (c) Describe exact and near multicollinearity in a regression model. Discuss, along with justification, the effect of these situations on (i) sampling variance and (ii) prediction. 8

   (d) Describe the problem of heteroscedasticity in linear regression models. Outline **any one** method for overcoming this problem. 8

   (e) Explain the use of Pareto curve in the study of income distribution. 8

   (f) Describe in detail Indirect Least Squares method for estimating structural parameters. Is the estimator unbiased and/or consistent ? 8

6. (a) State the problem of autocorrelation in a general linear model (GLM). Why does this problem arise ? Demonstrate its effect on ordinary least squares estimator (OLSE). 10

(Contd.)

(b) What do you mean by forecasting in economic models ? Define best linear unbiased predictor and obtain the same in a simple linear regression model satisfying all the basic assumptions. 10

(c) Explain variate-difference method for trend analysis. How is the appropriate order of differencing determined ? 10

(d) In the construction of price index numbers, explain the terms 'time reversal test' and 'factor reversal test'. Show that both these tests are satisfied by the Fisher index. 10

7. (a) Explain the identification problem in a system of simultaneous equations. State, without proof, the rank and order conditions for identifiability of an equation.

Examine for identifiability of the following model :

Demand : $M_t = \beta_0 + \beta_1 Y_t - \beta_2 R_t + \beta_3 P_t + U_{1t}$

Supply : $M_t = \alpha_0 + \alpha_1 Y_t + U_{2t}$

where M = money, Y = income,

R = rate of interest, P = price.

Assume that R and P are predetermined. 10

(Contd )

(b) Describe the two-stage least squares (2SLS) procedure for structural estimation in a simultaneous equations model. Show that it coincides with indirect least squares method when the equation is exactly identified. 10

(c) Discuss how the Engel curve is constructed on the basis of family budget data. Explain and interpret the Engel law in this context. How are variations in household size and composition handled ? 10

(d) Write down the auto correlation function of order K. for an AR(1) model $X_t = 0.7X_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ is a white noise process. Show that this model can be expressed as a moving average process of infinite order. Check the model for stationarity.

10

8. (a) The following information concerns changes in price and consumption (quantity) of certain major components of the consumption-basket of the labour class :

| Sl. No. | Item | Unit | Year 2000 | | Year 2014 | |
|---|---|---|---|---|---|---|
| | | | Price (Rs.) | Consumption | Price (Rs.) | Consumption |
| 1 | Rice | Quintal | 500 | 16 | 640 | 20 |
| 2 | Wheat | Quintal | 240 | 12 | 320 | 10 |
| 3 | Cloth | Metre | 16 | 50 | 20 | 35 |

(Contd.)

Compute price index using :

(i)   Fisher's method;

(ii)  Marshall-Edgeworth method.

Also interpret the results.                    10

(b)  Discuss the main aspects of the sampling design used by the NSSO for collection of data on household consumer expenditure. If you were to improve upon this design, explain what modifications you would suggest, with a clear justification.                    10

(c)  Define periodogram. State the connection between the periodogram and the autocovariance function of a time series. Show that h(w) defined by

$$h(w) = \left(\frac{1}{2\pi}\right) \sum_{t=-\infty}^{\infty} K(\tau) \exp(-iw\tau)$$

is of period $2\pi$, where K(.) is the auto covariance function.                    10

(d)  Define partial auto correlation function (PACF). Determine this function for the model :

$$X_t = 0.7 \, X_{t-1} + 0.3 X_{t-2} + \in_t$$

where $\{\in_t\}$ is a sequence of uncorrelated random variables with mean zero and variance $\sigma_\in^2$.                    10