Total number of printed pages–7

3 (Sem–3 /CBCS) STA HC 2

2021

(Held in 2022)

STATISTICS

(Honours)

Paper : STA–HC–3026

(Sampling and Indian Official Statistics)

Full Marks : 60

Time : Three hours

The figures in the margin indicate full marks for the questions.

1. Answer the following as directed : 1×7=7

    (a) The number of possible samples of size n from a population of N units drawn without replacement is

    (i) $N!$

    (ii) n^2

    (iii) N^n

    (iv) $^N C_n$

    (Choose the correct option)

(b) ———— sampling is the most efficient of all methods for estimating the effect of linear trend.

(Fill in the blank)

(c) Inverse of the sampling fraction is called ———— factor.

(Fill in the blank)

(d) Systematic sampling is more precise than simple random sampling only if units within the sample are ————.

(Fill in the blank)

(e) If ρ is the intraclass correlation between units of the same systematic sample, what conclusion will you draw if $\rho < -\dfrac{1}{(nk-1)}$?

(f) State the condition under which the regression estimator reduces to the ratio estimator.

(g) Write down the range in which the intracluster correlation co-efficient between elements within cluster (ρ) lies.

2. Answer the following questions briefly : 2×4=8

(a) What is probability sampling ?

(b) Name the *three* principles of sampling theory.

(c) Suppose in a population, population size $N=11$ and sample size $n=4$. Taking $k=3$, write down all possible samples using systematic sampling.

(d) Define two-stage sampling.

3. Answer *any three* out of the following questions : 5×3=15

(a) Deduce an expression for the variance of the estimate of population mean in case of simple random sampling without replacement.

(b) Write a short note on non-sampling errors.

(c) State the principles that are to be followed is stratifying a population.

(d) Describe the procedure of the linear systematic sampling scheme.

(e) What is a regression estimator ? When is it preferred over the ratio estimator ?

$$3+2=5$$

4. Answer either (i) or (ii) of the following questions :

$$10 \times 3 = 30$$

(a) (i) In stratified random sampling with the cost function

$$C = a + \sum_{i=1}^{k} n_i C_i$$

where the overhead cost a is constant and C_i is the average cost of sampling one unit in the i th stratum. Prove that

$$n_i = \frac{n N_i S_i / \sqrt{C_i}}{\sum_{i=1}^{k} \left(N_i S_i / \sqrt{C_i} \right)}$$

where S_i denotes the stratum variance. From the above relation, state the conditions under which a larger sample needs to be taken.

$$7+3=10$$

(ii) Describe the methods of collection of official statistics in India. In this context, discuss the role of Ministry of Statistics and Program Implementation (MoSPI). $6+4=10$

(b) (i) What is probability proportional to size sampling ? State the basic difference between simple random sampling and pps sampling.

Show that in pps sampling with replacement, an unbiased estimator of the population total y is given by $2+2+6=10$

$$\hat{Y}_{pps} = \frac{1}{n} \sum_{i=1}^{n} (y_i / p_i)$$

with sampling variance

$$V\left(\hat{Y}_{pps}\right) = \frac{1}{n} \sum_{i=1}^{n} p_i \left(\frac{y_i}{p_i} - Y \right)^2$$

Or

(ii) With usual notation, show that

$$V_{opt} \leq V_{prop} \leq V_{rand}$$ 10

(c) (i) In simple random sampling without replacement, for large n, show that

$$v(\hat{R}) = \frac{(1-f)R^2}{n}\left[C_y^2 + C_x^2 - 2\rho C_x C_y\right]$$

where $\hat{R}$ denotes the ratio estimator, ρ is the correlation co-efficient between the study variate y and auxiliary variate x,

$$C_y = \frac{S_y}{\overline{Y}_N} \text{ and } C_x = \frac{S_x}{\overline{X}_N}$$ 10

Or

(ii) Show that in simple random sampling without replacement of n clusters containing M elements from a population of N clusters, the sample mean $\overline{y}_n$ is an unbiased estimator of $\overline{Y}$ and its variance is given by

$$V(\overline{y}_n) \cong \frac{(1-f)}{nM}S^2\left[1 + (M-1)\rho\right] \text{ for large } N$$

where ρ is the intracluster correlation co-efficient.   3+7=10

———