Kunthavai Naachiyaar Govt. Arts College (W) Autonomous, Thanjavur

M.Stat

CC - XI : SAMPLING THEORY 18KP3S11

Unit – I

Hours : 6
Credit: 5

Simple random sampling with and without replacement. Simple random sampling for proportions. Properties of estimates of mean and variance-confidence limits-Estimation of sample size for proportions, Estimation of sample size.

Unit – II

Stratified random sampling – methods of allocation – Relative precision of stratified random sampling with simple random sampling – Estimation of gain in precision due to stratification – stratified sampling for proportions – Estimation of sample size.

Unit – III

Systematic random sampling –Linear systematic sampling – Circular systematic sampling Estimation of the variance – comparison of systematic sampling with SRS and stratified sampling – Concept of ratio and regression estimators.

Unit – IV

Cluster sampling – Equal cluster sampling – Estimator of mean and its variance – relative efficiency of cluster sampling optimum cluster size – Multi – Stage sampling – Two stage sampling with equal first stage units – Estimator of mean and its variance. Two – stage sampling with unequal first stage units Estimators of mean and its variance.

Unit – V

Multistage sampling – Double sampling for stratification – Optimal allocation – Double sampling for difference estimator – Double sampling for ratio estimator – Double sampling for regression estimator.

Text Book:

Daroga Singh and F.S.Chaudhary, Theory and analysis of Sampling Survey Design, New Age International (P) Ltd., Chennai.

Books for study Reference:

1. Moorthy, M.N (1967)- Sampling Theory and Methods, Statistical Publishing Society, Calcutta.
2. Cochran, W.G. (1994) – Sampling Techniques, Wiley Easter Lt

Simple Random Sampling :-

The simplest and common most method of sampling is simple Random sampling in which the sample is drawn unit by unit at each draw therefore Simple Random Sampling in a method of selecting in 'n' units out of a population of size N by giving equal Probability to all units or a Sampling Procedure in which all possible Combination of n units that may be formed from the population of size N units have the same Random Sampling.

Difference between SRSWR and SRSWOR :-

If the units are selected to drawn one by one in such a way to ~~back~~ a unit drawn at a time is replace back to the population before the subsequent drawn. It is known as Simple random Sampling with replace.

Example :-

In this type of Sampling with replacement from a population of Size the Probability of a Selection of a unit at each draw. remaining 1/N. digits number may be obtained by Co.

SRSWOR :-

In Simple Random Sampling without Replacement a unit can be included more than one in a Sample. Therefore if the requiste Sample Size is in the effective sample size is Some times less than n due to the effective Sample Size is inclusion of one or more units more than once with the idea than effective Sample Size should idea be to the Simple random Sampling without replacement.

Example:-

In this method a unit selected once is not included in the population at any subsequent drawn. Hence the probability of drawing a unit from a population of N units at r^{th} draw is $1/(n-r+1)$

In Simple random Sampling the probability of Selecting of any Sample of units 'n' from a population Consisting of 'N' unit remain Same $1/N\ Cn$

Hence NCN is the number of all Possible Sample.

Theorem : 1

Statement

The probability that a specified unit of the population being select at any given drawn is equal to the probability of being select at the first drawn.

Proof :

$$P(E_r) = (1 - 1/N)\left[1 - 1/_{N-1}\right]\left[1 - 1/_{N-2}\right]$$

$$\left[1 - \frac{1}{N-(r-1)}\right]\left[1 - \frac{1}{N-(r-1)}\right]$$

$$= \prod_{h=1}^{r-1} P\left[\text{That the specified unit is non selected in the } r^{th} \text{ drawn}\right]$$

$$P\left[\text{That the specified unit is selected in the } r^{th} \text{ drawn}\right]$$

$$= \prod_{h=1}^{r-1}\left[1 - \frac{1}{N-(h-1)}\right]\left[1 - \frac{1}{N-(r-1)}\right]$$

$$= \prod_{h=1}^{r-1}\left[\frac{N-(h-1)-1}{N-(h-1)}\right]\left[\frac{N-(r-1)-1}{N-r-1}\right]$$

$$P[E_r] = \prod_{h=1}^{r-1}\left[\frac{N-h+1-1}{N-(h-1)}\right]\left[\frac{N-r+1-1}{N-r-1}\right]$$

$$= \prod_{h=1}^{r-1}\left[\frac{N-h}{N-(h-1)}\right]\left[\frac{N-r}{N-(r-1)}\right]$$

$$P[E_r] = \left[\frac{N-1}{N-(1-1)}\right]\left[\frac{N-2}{N-(2-1)}\right]\cdot\left[\frac{N-3}{N-(3-1)}\right]\cdots$$

$$\frac{N-(r-1)}{N-[(r-1)-1]} \left[\frac{N-r}{N-(r-1)} \right]$$

$$= \frac{N-1}{N-1} \cdot \frac{N-2}{N-2} \cdot \frac{N-3}{N-3} \cdots \frac{N-(r-1)}{N-r} \cdot \frac{N-r}{N-(r-1)}$$

$$P[E_r] = \frac{1}{N}$$

$$= P(E_1)$$

$$P[E_r] = \frac{1}{N} = P[E_1]$$

Theorem : 2

Statement :

The probability of a specified unit being induced in the sample is equal to n/N.

Proof:

The probability of each case is $P(E_r) = 1/N$

$(x_1, x_2, \ldots x_n)$ Thus the required probability is n/N

Corrollary 1: The probability of a specified sample of "n" units ignoring order is $\frac{1}{N_{C_n}}$

corrollary 2:

If in the population of N units m added so that probability of selection units specified drawn is

$$\frac{1}{N-m+m} = (N-m+m)^{-1}$$

Procedure of selecting a random sample.

Since the theory of sampling is based on the assumption of random sampling the technique of R S is of basic significance some of the procedure used for selecting a R S are as follows.

(i) Lottery method

(ii) Random Number table

(i) Lottery method:

In practice a ticket chit may be associated with each unit of the population Thus each sampling unit has its identification is possible before, each draw, Draws of tickets a simple required size is obtained It may be rather shutting in Practice. Human bias and prejudice may also creep in this method.

ii) Uses the random numbers

A random number to be arrangement of digits 0 to 9 in either a linear or rectangular pattern where each position is digits.

A random numbers all 0, 1, 2, ... 9 appear independent of each other some randome common use

(i) Tippets random number tables

(ii) Fisher and rates tables

iii) Kendel and Smith ...

Scanned with CamScanner

To as certain whether these series of random number are really random the applie

(i) Frequency test

ii) serial test

iii) group test

iv) Poker test

A Practical method of selecting a random Sample is to choose having the same frequen similarly three or more

The simplest way of selecting a sample of required size selecting a random number N and taking the rejections number greater than N

(i) Remainder Approach

ii) Quotient Approach

iii) Independent choice of Digits.

Methods of sampling :-

the technique of selecting a simpl is of fundamental important in sampling theory and usually depend upon the nature broadly classified under following heads

(i) Random Sampling (or) probability Sampling

(ii) Non - Random Sampling (or) Non - Probability Sampling method

OPPO A5 2020

Scanned with CamScanner

i) Random Sampling

This is the method of selecting Sample according to certain case of prob in which each unit of population selecten in the sample two types.

A. Simple Random Sampling

B. Resticted Random sampling It is the three types

(i) stratified Sampling

(ii) systematic sampling

(iii) multi - stage sampling

ii) Non - Random Sampling (or) non - probability Sampling method.

A sample of elementary units basis of personal such meant is called a Non - R S (or) Non - probability sampling. five types

(i) purposive sampling

(ii) cluster sampling

(iii) Cauoter sampling

(iv) convenient sampling

(v) sequential sampling

Sampling Units:

The consistent of a population which are the Individuals to be sampled from the popula and cannot be further subdivided for the purpose of sampling at a time are called Sampling Units.

Sampling Frame:

For adopting any sampling procedure it is exenstical to have a list or map identifying each sampling unit by a number such a list or map is called sampling frame

Finite population correlation

random sample of size "n" infinite population known that the variation mean σ^2/n. Change in this result population introduction factor $(N-\frac{n}{N})$. The factor $(N-n/N)$ for the variance and $\sqrt{N-n/N}$ for standard error called the finite population correlation.

Merits and limitations of simple Random sampling:

Merits: the sample unit selected at random giving each unit equal chance of being selected. The element of subjectivity of personal eliminated As such as SRS more repesentative purpose sampling

Limitations :-

(i) The selecting of SRS requires an up to data frame. (i.e) completly cataqud population from which samples are to be drawn.

ii) A simple random may result in the selection of the sampling units which are widely speed geographically collecting the data may be much in terms of money

iii) For a give precision simple random sampling usually requires large sample size as compared to stratified random sampling.

Simple Random sampling of Attributes:

An attributes is a qualitative characteristic which cannot be measured qualitatively

Example, honesty, beauty.

(i) Defective items in a large consignment of such items

ii) The literates or the bread linear in a town

iii) The educated unemployed persons in a city and so on.

Example:-

The N population units denoted by $U_1, U_2 \ldots U_N$ Sample unit by $U_1 \ldots U_n$. let $Y_h \Rightarrow y \, (h = 1, 2 \ldots N)$ be the value of character for the h^{th} unit selected

... lation and $y_h = y \, (h = 1.2 \ldots n)$ the value of character for the h^{th} unit selected,

$$Y = Y_1 + Y_2 + \cdots + Y_n = \sum_{h=1}^{N} Y_n$$

$$y = y_1 + y_2 + \cdots + y_n = \sum_{h=1}^{n} y_n$$

Population:-

$$\bar{Y} = \bar{Y}_N = \frac{1}{N} \sum_{h=1}^{N} Y_N$$

Sample mean:-

$$\bar{y} = \bar{y}_n = \frac{1}{n} \sum_{h=1}^{n} y_n$$

Sample mean $\bar{y}_n$ may also returned alternatively

$$\bar{y} = \bar{y}_n = \frac{1}{n} \sum_{n=1}^{N} a_h Y_n$$

Where,

$$a_h = \begin{cases} 1 & \text{if } k^{th} \text{ unit is included in a sample} \\ 0 & \text{if } h^{th} \text{ unit is not included in sample.} \end{cases}$$

Population mean Square:-

$$S^2 = \frac{1}{N-1} \sum_{h=1}^{N} (Y_n - \bar{Y}_n)^2$$

Here,

$$S^2 = \frac{1}{N-1} \sum_{n=1}^{N} \left[Y_n^2 + \bar{Y}_N^2 - 2 Y_n \bar{Y}_n \right]$$

$$S^2 = \frac{1}{N-1} \left[\sum_{n=1}^{N} Y_n^2 + N \bar{Y}_N^2 - 2 \sum_{n=1}^{N} Y_n \bar{Y}_n \right]$$

$$= \frac{1}{N-1}\left[\sum_{n=1}^{N} Y_n^2 + N\bar{Y}_N^2 - 2N\bar{Y}_N\bar{Y}_N\right]$$

Since,

$$\bar{Y}_N = \frac{1}{N}\sum_{n=1}^{N} Y_n \Rightarrow N\bar{Y}_N = \sum_{n=1}^{N} Y_n$$

$$S^2 = \frac{1}{N-1}\left[\sum_{n=1}^{N} Y_n^2 + N\bar{Y}_N^2 - 2N\bar{Y}_N^2\right]$$

$$S^2 = \frac{1}{N-1}\left[\sum_{n=1}^{N} Y_n^2 - N\bar{Y}_N^2\right]$$

Sample mean Square:

$$s^2 = \frac{1}{n-1}\sum_{n=1}^{n}(y_n - \bar{y}_n)^2$$

Here,
$$s^2 = \frac{1}{n-1}\left[\sum_{n=1}^{n} y_n^2 + n\bar{y}_n^2 - 2\sum_{n=1}^{n} y_n\bar{y}_n\right]$$

$$= \frac{1}{n-1}\left[\sum_{n=1}^{n} y_n^2 + n\bar{y}_n^2 - 2n\bar{y}_n\bar{y}_n\right]$$

$$= \frac{1}{n-1}\left[\sum_{n=1}^{n} y_n^2 + n\bar{y}_n^2 - 2n\bar{y}_n^2\right]$$

$$s^2 = \frac{1}{n-1}\left[\sum_{n=1}^{n} y_n^2 - n\bar{y}_n^2\right]$$

Simple Random Sampling for without Replacement:

Theorem : 3

Statement:

In Simple Random Sampling without Replacement (SRSWOR), the Sample mean is an unbiased estimate of the population mean [or]

$$E(\bar{y}_n) = \bar{Y}_N$$

Proof:-

$$\bar{y}_n = 1/n \sum_{i=1}^{N} a_i y_i$$

$$E(\bar{y}_n) = E[1/n \sum_{i=1}^{N} a_i y_i]$$

$$= \frac{1}{n} \sum_{i=1}^{N} E(a_i) y_i \rightarrow (1).$$

Since a_i takes only two values,

$$E(a_i) = 1 \cdot P(a_i = 1) + 0 \cdot P(a_i = 0)$$

$$= 1 \cdot P + 0 \cdot P$$

$$= P$$

$$E(a_i) = 1 \cdot n/N + 0 \, (1 - n/N)$$

$$E(a_i) = n/N \rightarrow (2)$$

equation (2) Subsuiting in (1)

$$E(\bar{y}_n) = 1/n \sum_{i=1}^{N} n/N \, Y_i$$

$$= 1/N \sum_{i=1}^{N} Y_i$$

A5 2020

$E(\bar{y}_n) = \bar{Y}_N$

Hence the Proof

Theorem 4 :

Statement :-

In SRSWOR the Sample mean square is an unbiased estimate of the population mean square (or)

$$E(s^2) = S^2 \rightarrow (1)$$

Proof:-

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left(\frac{1}{n}\sum y_i\right)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i^2\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n} y_i^2 + \sum_{i\neq j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{n-1}\left[\left(1-\frac{1}{n}\right)\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i\neq j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{n-1}\left(1-\frac{1}{n}\right)\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j]$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j \rightarrow (2)$$

$$E(s^2) = \frac{1}{n} E\left(\sum_{i=1}^{n} y_i^2\right) - \frac{1}{n(n-1)} E\left(\sum_{i\neq j=1}^{n} y_i y_j\right)$$

we have,

$$E\left(\sum_{i=1}^{n} y_i^2\right) = E\left(\sum_{i=1}^{N} a_i Y_i^2\right) = \sum_{i=1}^{N} E(a_i) Y_i^2$$

Since a_i takes only two value.

$$E(a_i) = 1 \cdot P(a_i = 1) + 0 \cdot p(a_i = 1)$$

$$= 1 \cdot p + 0 \cdot p$$

$$= \cancel{\phi}\, 1 \cdot n/N + 0\,[1 - n/N]$$

$$E(a_i) = n/N$$

$$E\left[\sum_{i=1}^{n} y_i^2\right] = n/N \sum_{i=1}^{N} Y_i^2 \rightarrow (4)$$

$$E\left[\sum_{i \neq j=1}^{n} y_i y_j\right] = E\left[\sum_{i \neq j=1}^{N} a_i a_j Y_i Y_j\right]$$

$$= \sum_{i \neq j=1}^{N} E(a_i a_j) Y_i Y_j$$

Now,

$$E(a_i a_j) = 1 \cdot P(a_i a_j = 1) + 0 \cdot P \cdot (a_i a_j = 1)$$

$$= P(a_i = 1) \cap \cdot P(a_j = 1)$$

$$= P(a_j = 1) / P(a_i = 1)$$

$$E(a_i a_j) = \frac{n(n-1)}{N(N-1)}$$

∵ Because Probability of $(a_i = 1) = P[i^{th}$ unit is included in the sample of size $n = n/N$ and

$P(a_j = 1) / P(a_i = 1) = P$ (j^{th} unit is included in the Sample given that i^{th} unit is included in Sample).

$$E\left(\sum_{i \neq j = 1}^{n} y_i y_j\right) = \frac{n(n-1)}{N(N-1)} \sum_{i \neq j = 1}^{N} Y_i Y_j \rightarrow 5$$

Subsciting from (4) and (5) in (3) we get

$$E(s^2) = 1/n \cdot n/N \sum_{i=1}^{N} Y_i^2 - \frac{1}{n(n-1)} - \left(\frac{n(n-1)}{N(N-1)}\right) \sum_{i \neq j = 1}^{N} Y_i Y_j$$

$$E(S^2) = \frac{1}{N} \sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)} \sum_{i \neq j = 1}^{N} Y_i Y_j$$

$$E(S^2) = \frac{1}{N-1} \left[\sum_{i=1}^{N} Y_i^2 - N\bar{Y}_N^2\right]$$

Theorem 5 :-

Statement :-

In SRSWOR the variance of the sample mean is given by

$$Var(\bar{y}_n) = \frac{N-n}{N} \cdot \frac{S^2}{n}$$

Proof :-

$$Var(\bar{y}_n) = \frac{N-n}{N} \cdot \frac{S^2}{N} \rightarrow (1)$$

$$Var(\bar{y}_n) = E(\bar{y}_n^2) - [E(\bar{y}_n)]^2$$

$$= E(\bar{y}_n^2) - \bar{Y}_n^2 \rightarrow (2)$$

$$E(\bar{y}_n^2) = E\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2$$

$$E(\bar{y}_n^2) = \frac{1}{n^2} E\left[\sum_{i=1}^{n} y_i^2 + \sum_{i\neq j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{n^2}\left[E\left(\sum_{i=1}^{n} y_i^2\right) + E\left(\sum_{i\neq j=1}^{n} y_i y_j\right)\right] \rightarrow ($$

$$(1) \Rightarrow E\left(\sum_{i=1}^{n} y_i^2\right) = \sum_{i=1}^{N} E(a_i) Y_i$$

$$E\left(\sum_{i=1}^{n} y_i^2\right) = n/N \sum_{i=1}^{N} E(a_i) Y_i$$

$$E\left(\sum_{i=1}^{n} y_i^2\right) = n/N \sum_{i=1}^{N} Y_i^2$$

But :

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N} (Y_i - \bar{Y}_n)^2$$

$$(N-1)S^2 = \sum_{i=1}^{N} Y_i^2 - N\bar{Y}_N^2$$

$$(N-1)S^2 + N\bar{Y}_N^2 = \sum_{i=1}^{N} Y_i^2$$

$$E\left(\sum_{i=1}^{n} y_i^2\right) = n/N\left[(N-1)S^2 + N\bar{Y}_N^2\right]$$

$$= n\left[\frac{(N-1)}{N} S^2 + \frac{N\bar{Y}_N^2}{N}\right]$$

$$E\left(\sum_{i=1}^{n} y_i^2\right) = n\left[\frac{(N-1)}{N} S^2 + \bar{Y}_N^2\right] \rightarrow (4)$$

$$(2) \Rightarrow E\left(\sum_{i\neq j=1}^{n} y_i y_j\right) = \sum_{i\neq j=1}^{n} E(a_i a_j) Y_i Y_j$$

$$= \frac{n(n-1)}{N(N-1)} \sum_{i\neq j=1}^{N} Y_i Y_j$$

OPPO A5 2020

$$= \frac{n(n-1)}{N(N-1)} \left[\left[\sum_{i=1}^{N} Y_i\right]^2 - \sum_{i=1}^{N} Y_i^2 \right]$$

$$E\left[\sum_{i\neq j=1}^{n} y_i y_j\right] = \frac{n(n-1)}{N(N-1)} \left[N^2 \bar{Y}_N^2 - (N-1)S^2 - N\bar{Y}_N^2 \right]$$

$$= \frac{n(n-1)}{N(N-1)} \left[N(N-1)\bar{Y}_N^2 - (N-1)S^2 \right]$$

$$E\left(\sum_{i\neq j=1}^{n} y_i y_j\right) = n(n-1)\left[\bar{Y}_N^2 - \frac{S^2}{n}\right]$$

$$\bar{Y}_N = 1/N \sum_{i=1}^{N} Y_i$$

$$N\bar{Y}_N^2 = \left(\sum_{i=1}^{N} Y_i\right)^2$$

$$= N^2 \bar{Y}_N^2$$

Subsuting from equation number ④ and ⑤ in ③ we get.

$$E(\bar{y}_n)^2 = \frac{1}{n^2}\left[E\left(\sum_{i=1}^{n} y_i^2\right) + E\left[\sum_{i\neq j=1}^{n} y_i y_j\right]\right]$$

$$= 1/n^2 \cdot n\left[\frac{(N-1)S^2}{N} + \bar{Y}_N^2\right] + \frac{1}{n^2}\left[n(n-1)\left[\bar{Y}_N^2 - S^2/N\right]\right]$$

$$= 1/n\left[\frac{(N-1)S^2}{N} + \bar{Y}_N^2\right] + \frac{(n-1)}{n}\left[\bar{Y}_N^2 - \frac{S^2}{N}\right]$$

$$= 1/n\left[\left(1 - n/N\right)S^2 + \bar{Y}_N^2\right) + (1 - 1/n)$$

$$\left[\bar{Y}_N^2 - S^2/N\right]$$

$$= \frac{S^2}{n} - \cancel{\frac{S^2}{nN}} + \cancel{\frac{\bar{Y}_N^2}{n}} + \bar{Y}_N^2 - \cancel{\frac{\bar{Y}_N^2}{n}} - \frac{S^2}{N} + \cancel{\frac{S^2}{Nn}}$$

$$= \bar{Y}_N^2 + \frac{S^2}{n} - \frac{S^2}{N}$$

$$E(\bar{y}_n^2) = \bar{Y}_N^2 + \left(1/n - 1/N\right)S^2 \rightarrow (6)$$

Subsituting from equation number.

(6) in (2) we get.

$$Var(\bar{y}_n) = E(\bar{y}_n^2) - \bar{Y}_n^2$$

$$Var(\bar{y}_n) = \bar{Y}_N^2 + (1/n - 1/N) S^2 - \bar{Y}_N^2$$

$$Var(\bar{y}_n) = \frac{N-n}{N} \cdot \frac{S^2}{N}$$

Hence its proved.

Simple Random Sampling for with Replacement.

Theorem : 6

Statement :-

The variance for sample mean and SRSWR is given by

$$Var(\bar{y}) = \frac{N-1}{N} \cdot \frac{S^2}{n}.$$

Proof :-

In SRSWR the Sample units are independent (i.e) each and every units are the Sample are identically with variance σ^2

$$Var(\bar{y}) = Var\left(1/n \sum_{n=1}^{n} y_n\right)$$

$$= 1/n^2 \sum_{i=1}^{n} Var(y_n)$$

∴ Because the co-variance term vanish.

Scanned with CamScanner

$$\text{Var}(\bar{y}) = \frac{1}{n^2} \sum_{n=1}^{n} \sigma^2$$

$$= \frac{1}{n^2} \cdot n\sigma^2$$

$$\text{Var}(\bar{y}) = \frac{\sigma^2}{n} \rightarrow (1)$$

But

$$S^2 = \frac{1}{N-1} \left[\sum_{n=1}^{N} Y_n^2 - N \bar{Y}_N^2 \right]$$

$$\sigma^2 = \frac{1}{N} \left[\sum_{n=1}^{N} Y_n^2 - N \bar{Y}_N^2 \right]$$

$$(N-1) S^2 = N \sigma^2$$

$$\sigma^2 = \left(\frac{N-1}{N} \right) S^2 \rightarrow (2)$$

Sub equa (2) in (1) we get,

$$\frac{\sigma^2}{n} = \left(\frac{N-1}{N} \right) \frac{S^2}{n}$$

SRSWOR Vs SRSWR is given.

$$\text{var}(\bar{y})\ \text{SRSWOR} = \left(N - n/N \right) \frac{S^2}{n}$$

$$\text{var}(\bar{y})\ \text{SRSWR} = \left(N - \frac{1}{N} \right) \frac{S^2}{n}$$

$$\text{var}(\bar{y})\ \text{SRSWOR} > \text{var}(\bar{y})\ \text{SRSWOR}$$

$$\left(N - \frac{1}{N} \right) \frac{S^2}{n} > \left(N - n/N \right) S^2/n$$

Hence the Proof.

Simple Random Sampling for with Replacement

Theorem : 6

The variance for sample mean of SRSWR is given by

$$\text{var}(\bar{y}) = \frac{N-1}{N} \cdot \frac{S^2}{n}$$

Proof!

$$var(\bar{y}) = Var\left(1/n \sum_{n=1}^{n} y_n\right)$$

$$= \frac{1}{n^2} \sum_{n=1}^{n} var(y_n)$$

Because the co-var term vanish

$$var(\bar{y}) = 1/n^2 \sum_{n=1}^{n} \sigma^2$$

$$= 1/n^2 \, n \, \sigma^2$$

$$var(\bar{y}) = \sigma^2/n \longrightarrow (1)$$

$$S^2 = \frac{1}{N-1}\left[\sum_{n=1}^{N} Y_n^2 - N\bar{Y}_N^2\right]$$

$$\sigma^2 = \frac{1}{N}\left[\sum_{n=1}^{N} Y_N^2 - N\bar{Y}_N^2\right]$$

$$(N-1) S^2 = N\sigma^2$$

$$\sigma^2 = \left(\frac{N-1}{N}\right) S^2 \longrightarrow (2)$$

Sub equ (2) in (1)

$$\sigma^2/n = \left(\frac{N-1}{N}\right) S^2/n$$

$$var(\bar{y})\ SRSWR > var(\bar{y})\ SRSWOR$$

$$(N - 1/N)\frac{S^2}{n} > \left(N - n/N\right) S^2/n$$

Hence the Proof.

Theorem: 7

$U \cap I$

$$U_n = x_n + y_n \quad = x_n + y_n$$

$$\bar{U} = \frac{1}{N}\sum_{n=1}^{N} U_n = (\bar{X} + \bar{Y})$$

$$\bar{u} = \frac{1}{n}\sum_{n=1}^{n} U_n = (\bar{x} + \bar{y})$$

$$(U_n - \bar{U}) = (X_n + Y_n) - (\bar{X} + \bar{Y})$$

$$= X_n + Y_n - \bar{X} - \bar{Y} \Rightarrow X_n - \bar{X} + Y_n - \bar{Y}$$

$$var(\bar{u}) = E(\bar{u} - \bar{U})^2$$

$$= \frac{N-n}{N} \cdot \frac{Su^2}{n}$$

$$var(\bar{u}) = \left(\frac{1}{n} - \frac{1}{N}\right) Su^2 \rightarrow (1)$$

But,

$$Su^2 = \frac{1}{N-1}\sum_{n=1}^{N}(U_n - \bar{U})^2$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}\left[(X_n - \bar{X}) + (Y_n - \bar{Y})\right]^2$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}\left[(X_n - \bar{X})^2 + (Y_n - \bar{Y})^2 + 2(X_n - \bar{X})(Y_n - \bar{Y})\right]$$

$$= \frac{1}{N-1}\left[\sum_{n=1}^{N}(X_n - \bar{X})^2 + \sum_{n=1}^{N}(Y_n - \bar{Y})^2 + 2\sum_{n=1}^{N}(X_n - \bar{X})(Y_n - \bar{Y})\right] \rightarrow (2)$$

Subtiting equ (1) in (2) we get

$$var(\bar{u}) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N-1}\left[\sum_{n=1}^{N}(X_n - \bar{X})^2 + \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N-1}\sum_{n=1}^{N}(Y_n - \bar{Y})^2 + 2\left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})(Y_n\right.$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) Sx^2 + \left(\frac{1}{n} - \frac{1}{N}\right) Sy^2 + 2\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{N-1}\right)\sum_{n=1}^{N}(X_n - \bar{X})(Y_n - \bar{Y})$$

$$= var(\bar{x}) + var(\bar{y}) + 2\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{N-1}\right)\sum_{n=1}^{N}(X_n - \bar{X})(Y_n - \bar{Y}) \rightarrow (3)$$

Theorem : 8

Sample proportion $p = a/n$ is an unbiased

$P = A/N$ $E(p) = P$ sample proportion is an unbiased estimator.

Proof:

$$p = \frac{1}{n} \sum_{n=1}^{n} y_n$$

$$E(p) = E\left[\frac{1}{n} \sum_{n=1}^{n} y_n\right]$$

$$= \frac{1}{n} \sum_{n=1}^{N} E(a_n) y_n$$

$$= \frac{1}{n} \sum_{n=1}^{N} n/N \; y_n$$

$$E(p) = \frac{1}{N} \sum_{n=1}^{N} y_n .$$

Unbiased estimator.

$$E(p) = P \qquad NP = A$$

Theorem : 9

Sample proportion in SRSWOR the variance

$$\text{var}(p) = E(p-P) = \frac{N-n}{n} PQ/(N-1)$$

Proof:

Simpling Random sampling

$$\text{Var}(y) = \frac{N-n}{n} \frac{S^2}{n}$$

$$= \frac{N-n}{N} \frac{\frac{NPQ}{N-1}}{n}$$

$$= \frac{N-n}{N} \cdot \frac{\frac{NPQ}{N-1}}{n}$$

$$\text{Var}(\bar{y}) = N-n/n \cdot \frac{PQ}{N-1}$$

OPPO A5 2020

Estimation of sample size for properties

We know that the sample properties is an unbiased estimator of population proportion consider the following equation Sample size

$$P\left[|p - P| < \alpha\right] = 1 - \alpha \rightarrow ①$$

marginal error for the estimating sample size with specified precision α.

$$z = \frac{p - P}{SE(p)} \sim N(0,1)$$

$$P = \left[\left|\frac{p - P}{S.E(p)}\right| < z_\alpha\right] = 1 - \alpha \rightarrow ②$$

z_α is the value by

$$P\left[|z| \geq z_\alpha\right] = \alpha$$

$$Var(p) = \frac{N-n}{N-1} \frac{PQ}{n}$$

equal ② implies

$$P\left[\frac{p - P}{\sqrt{N - n/_{N-1} \ PQ/_n}} < z_\alpha\right] = 1 - \alpha \rightarrow (3)$$

Comparing ① & ③ we have

$$d = z_\alpha \sqrt{\frac{N-n}{N-1} \frac{PQ}{n}}$$

We know that

$$P\left[|z| < z_\alpha\right] = 1-\alpha$$

where

$z \sim N(0,1)$ for large sample

(or)

$$P\left[|t| \leq t_\alpha\right] = 1-\alpha$$

when $t \sim t(n-1)$ for small sample

$$P\left[\frac{\bar{y}-\bar{Y}}{S.E(\bar{y})} \leq z_\alpha\right] = 1-\alpha$$

$$P\left[|\bar{y}-\bar{Y}| \leq z_\alpha\, S.E(\bar{y})\right] = 1-\alpha$$

$$P\left[-z_\alpha\, S.E(\bar{y}) \leq (\bar{y}-\bar{Y}) \leq z_\alpha \cdot S.E(\bar{y})\right] = 1-\alpha$$

$$P\left[\bar{y}-z_\alpha\, S.E(\bar{y}) \leq \bar{Y} \leq \bar{y}+z_\alpha\, S.E(\bar{y})\right] = 1-\alpha$$

The confident interval for $\bar{y}$ is

$$\bar{y} \pm z_\alpha \sqrt{1-F}\ \frac{S}{\sqrt{n}}$$

Since

$$\text{var}(\bar{y}) = \frac{N-n}{N} \cdot \frac{S^2}{n}$$

$$= (1-n/N)\ S^2/n$$

$$\text{Var}(\bar{y}) = (1-F)\ S^2/n$$

$$E(y) = \sqrt{\text{var}(\bar{y})}$$

$$S.E(\bar{y}) = \sqrt{1-F}\ S/\sqrt{n}$$

where

$$F = n/N$$

Scanned with CamScanner

UNIT - II

Stratified Random Sampling.

In SRSWOR we obtained

$$Var(\bar{y}_n) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$$

This implies that the variance of the sample estimate of the population mean is

(i) inversely proportional to the sample size &

(ii) directly proportional to the variability of the sampling units in the population.

Stratification means division into layers. Auxiliary information (i) units within cash group are homogeneous as possible and (ii) the group means are widly different as possible. Population consisting of N sampling units is divided into K

$N_1, N_2 \ldots, N_n$, $N = \sum_{i=1}^{k} N_i$. If a simple random Sample of size n_i, $(i = 1, 2, \ldots k)$ the stratum respectively $n = \sum_{i=1}^{k} n_i$, the sample is termed as stratified Random sample of size n and the technique of drawing such a sample is called the stratified random Sampling.

Stratified

Principal Advantages of Statistifdd Random Sampling.

1. More Representative: In an unstratified ran sample some strata may be over represen other may be under-representative some may be exculeded altogther. stratified sampling ensures

2. Greater Accuracy:

stratified sampling provide estim with increased precision. Moreever, statified sampling enables us to obtain the result of known precision for each of the stratum.

3. Administrative convenience.

Compared with simple random sample, the stratified sample would be more concentrated geographically, Accordingly, the time collection the data & interviewing greater case and convenience.

4. Sometimes the sampling problems may differ markedly in different of the population.

Notations and terminalogy

Estimate of population Mean and its Variance

Theorem 7.6: y_{st} is an unbiased estimate of the population mean $\bar{Y}_N$,

$$E(\bar{y}_{st}) = \bar{Y}_N$$

$$E(\bar{y}_n) = \frac{1}{N} \sum_{i=1}^{k} N_i \cdot E(\bar{y}_{ni}) = \frac{1}{N} \sum_{i=1}^{k} N_t \bar{Y}_{Ni} = \bar{Y}_N$$

Theorem 7.7

$$\text{var}(\bar{y}_{st}) = \frac{1}{N^2} \sum_{i=1}^{k} N_i(N_i - n_i) \frac{S_i^2}{n_i} =$$

$$\sum_{i=1}^{k} p_i^2 \left(\frac{1}{n_t} - \frac{1}{N_t}\right) S_t^2$$

Proof:

since the sample in each stratum is simple random sample without replacement

$$\text{Var}(\bar{y}_n) = \left(\frac{1}{n_t} - \frac{1}{N_t}\right) S_t^2$$

where S_i^2 is as defined in

$$\text{Var}(\bar{y}_n) = \text{Var}\left(\sum_{i=1}^{k} p(\bar{y}_{ni}) =\right)$$

$$\sum_{i=1}^{k} p_i^2 \, \text{Var}(\bar{y}_{ni}),$$

the co-var terms vanish since the sample from different independent

$$\text{var}(\bar{y}_n) = \sum_{i=1}^{k} p_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2,$$

$$= \frac{1}{N^2} \sum_{i=1}^{k} N_i(N_i - n_i) \frac{S_i^2}{n_i}$$

we that $\text{Var}(\bar{y}_{st})$ depends on S_i^2, If S_i^2, are small

Allocation of sample size

The allocation of the sample size for different strata is done in the following two ways

(a) proportional allocation

(b) optimum allocation

Proportional Allocation

Allocation of n_i's to various strata is called Proportional. the sample fraction is constant for each stratum

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_k}{N_k} = \frac{\Sigma n_i}{\Sigma N_i} = \frac{n}{N} = C$$

$$\frac{n_i}{N_i} = C = \frac{n}{N} \Rightarrow n_i \propto N_i \ (i=1,2,\cdots k)$$

Proportional allocation each stratum is represented according

In proportional allocation $Var(\bar{y}_{st})$

$$Var(y_{st})_{prop} = \frac{1}{N^2} \sum_{i=1}^{k} N_i [N_i - n_i] \frac{S_i^2}{n_i}$$

$$= \sum_{i=1}^{k} \frac{N_i}{N^2} \left[\frac{N_i}{n_i} - 1 \right] S_i^2$$

$$= \sum_{i=1}^{k} \frac{P_i}{N} \left(\frac{N}{n} - 1 \right) S_i^2,$$

$$= \left(\frac{1}{n} - \frac{1}{N} \right) \sum_{i=1}^{k} P_i S_i^2$$

Optimum Allocation: Another guiding principal in the determination of the n_i's is to choise them so as to

(a) minimise the variance, of the estimate for

(i) fixed size n (ii) fixed cost.

(b) minimise the total cost for fixed desired precision

The optimum Allocation n_i's are to be obtained such that

(1) $Var(\bar{y}_{st})$ is minimum for fixed n.

(2) $Var(\bar{y}_{st})$ is minimum for fixed total cost c.

(3) total cost c is minimum for fixed value of $Var(y_{st}) = V_0$.

Cost function:

In any sample survey, the value of Information on the experimental units must always balanced from the cost function c in stratified

$$c = a + \sum_{i=1}^{k} c_i n_i$$

a is the overhead cost and c_i is the cost per unit in the i^{th} stratum.

Theorem 7.8. $Var(\bar{y}_{st})$ is minimum for fixed total size of the sample if $\propto N_i S_i$.

Proof: Here the problem is to minimise

$$Var(\bar{y}_{st}) = \frac{1}{N^2} \sum_{i=1}^{k} N_i(N_i - n_i) \frac{S_i^2}{n_i}$$

Subject to the condition

$$\sum_{i=1}^{k} n_i = n$$

This is equalent to minimising

$$\phi = Var(\bar{y}_{st}) + \lambda \left(\sum_{i=1}^{k} n_i - n\right)$$

$$= \frac{1}{N^2} \sum_{i=1}^{k} N_i (N_i - n_i) \frac{S_t^2}{n_i} + \lambda \left(\sum_{i=1}^{k} n_i - n\right)$$

$$\frac{\partial \phi}{\partial n_i} = -\frac{N_i^2 S_i^2}{N^2 n_i^2} + \lambda = 0$$

$$n_i = \frac{N_t S_t}{N\sqrt{\lambda}}$$

also $$\frac{\partial^2 \phi}{\partial n_i^2} = \frac{2 N_t^2 S_t^2}{N^2 n_i^2} > 0$$

Comparison of stratified Random Sampling with Simple Random Sampling without Stratification.

$$Var(\bar{y}_{st})_P = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{k} p_i S_i^2$$

$$Var(\bar{y}_n)_R = \left(\frac{1}{n} - \frac{1}{N}\right) S^2$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{k} \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_N)^2$$

S^2 is termed as S_t^2.

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{k} \sum_{j=1}^{N_i} (Y_{ij} - Y_{Ni} + Y_{Ni} - \bar{Y}_N)^2$$

$$(N-1)S^2 = \sum_i \left[\sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_{Ni})^2\right] + \sum_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$+ 2 \sum_i \left[(\bar{Y}_{Ni} - \bar{Y}_N) \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_{Nt})\right]$$

$$= \sum_{i=1}^{k} (N_i - 1) S_t^2 + \sum_{i=1}^{k} N_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

The product term vanish since $\sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_{Ni})$, from mean is zero.

$$N_i - 1 = N_i \text{ and } N-1 \simeq N.$$

$$NS^2 \simeq \sum_{i=1}^{k} N_i S_i^2 + \sum_{i=1}^{k} N_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$S^2 \simeq \sum_{i=1}^{k} P_i S_i^2 + \sum_{i=1}^{k} P_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$Var(\bar{y}_n)_R \simeq \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{k} P_i S_i^2$$

$$+ \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{k} P_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$\simeq Var(\bar{y}_{st})_P + \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{k} P_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$Var(\bar{y}_n)_R \geq Var(\bar{y}_{st})_P$$

Neyman's Allocation Vs Proportional Allocation.

$Var(\bar{y}_{st})_P$ and $Var(\bar{y}_{st})_N$ for the variance of estimate mean is stratified sampling with proportional allo and Neyman's optim alloc respectively.

$$Var(\bar{y}_{st})_P - Var(\bar{y}_{st})_N = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{k} P_i S_i^2$$

$$- \left(\frac{1}{n} (\Sigma P_i S_i)^2 - \frac{1}{N} \Sigma P_i S_i^2\right)$$

$$= \frac{1}{n} \left[\sum_{i=1}^{k} P_i S_i^2 - \left(\sum_{i=1}^{k} P_i S_i\right)^2\right]$$

$$= \frac{1}{n} \sum_{i=1}^{k} P_i (S_i - S)^2$$

$S = \sum_{i=1}^{k} P_i S_i^2 = \frac{1}{N} \sum_{i=1}^{k} N_i S_i$ is the weighted mean of the stratum standard deviation,

$$Var(\bar{y}_{st})_P \geq Var(\bar{y}_{st})_N$$

Sutituting for $Var(\bar{y}_{st})_P$ from

$$Var(\bar{y}_n)_R = Var(\bar{y}_{st})_N + \frac{1}{n}\sum_{i=1}^{k} P_i(S_i-\bar{S})^2 + \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{k} P_i(\bar{Y}_{N_i}-\bar{Y}_N)^2$$

$$Var(\bar{y}_n)_R \geq Var(\bar{y}_{st})_N$$

$$Var(\bar{y}_n)_R \geq Var(\bar{y}_{st})_P \geq Var(\bar{y}_{st})_N$$

Remarks on optimum & Proportional Allocation.

1. The most serious limitation of "optimum allocation" Knowledge of S_i's in advance.

$$n' \geq \frac{\left(\sum_{i=1}^{k} P_i S_i\right)^2 - \sum_{i=1}^{k} P_i S_i^2}{2\sum_{i=1}^{k} P_i(S_i-\bar{S})^2}$$

2. Neyman's optimum allocation of the sample to different strata on the basis of one characteristic result in loss of precision on other charaters as compared to the method of proportional allocation.

Scanned with CamScanner