UNIT 1:

CENSUS AND SAMPLING SURVEY:

In a census, data about all individual units (e.g. people or households) are collected in the population.

In a survey, data are only collected for a sub-part of the population; this part is called a sample.

These data are then used to estimate the characteristics of the whole population.

PRINCIPAL STEPS IN A SAMPLE SURVEY:

Surveys vary greatly in their complexity.

To take a sample from 5000 cards, neatly arranged and numbered in a file, is an easy task.

It is another matter to sample the inhabitants of a region where transport is by water through the forests, where there are no maps, where 15 different dialects are spoken, and where the inhabitants are very suspicious of an inquisitive stranger.

Problems that are baffling in one survey may be trivial or non-existent in another.

The principle steps in a survey are grouped somewhat arbitrarily under 11 headings.

- Objectives of survey
- Population to be sampled
- Data to be collected
- Degree of precision Desired
- Methods of Measurement
- The Frame
- Selection of the sample
- The Pre-test
- Organization of the field work
- Summary and Analysis of the data
- Information gained for future surveys

PILOT SURVEY:

A pilot survey is a strategy used to test the questionnaire using a smaller sample compared to the planned sample size. In this phase of conducting a survey, the questionnaire is administered to a percentage of the total sample population, or in more informal cases just to a convenience sample.

Pilot experiments are frequently carried out before large-scale quantitative research, in an attempt to avoid time and money being used on an inadequately designed project. A pilot study is usually carried out on members of the relevant population. A pilot study is often used to test the design of the full-scale experiment which then can be adjusted. It is a potentially valuable insight and, should anything be missing in the pilot study, it can be added to the full-scale (and more expensive) experiment to improve the chances of a clear outcome.

Non-sampling error :

- A **non-sampling error** is a statistical term that refers to an **error** that results during data collection, causing the data to differ from the true values.
- Non-sampling error is the error that arises in a data collection process as a result of factors other than taking a sample.
- Non-sampling errors have the potential to cause bias in polls, surveys or samples.
- There are many different types of non-sampling errors and the names used to describe them are not consistent. Examples of non-sampling errors are generally more useful than using names to describe them.

Sampling error :

A **sampling error** is limited to any differences between **sample** values and universe values that arise because the **sample** size was limited.

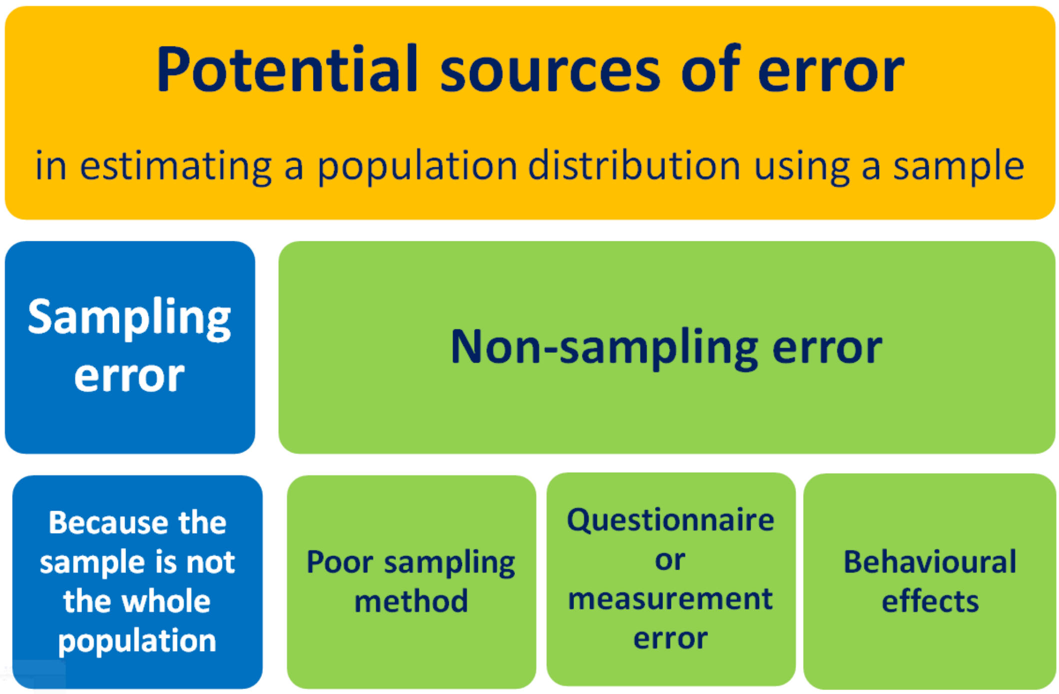
Sampling error is the difference between a population parameter and a **sample** statistic used to estimate it. For example, the difference between a population mean and a **sample** mean is **sampling error**.

Sampling error is the error that arises in a data collection process as a result of taking a sample from a population rather than using the whole

population.
Sampling error is one of two reasons for the difference between an estimate of a population parameter and the true, but unknown, value of the population parameter.

The other reason is non-sampling error.

Even if a sampling process has no non-sampling errors then estimates from different random samples (of the same size) will vary from sample to sample, and each estimate is likely to be different from the true value of the population parameter.
The sampling error for a given sample is unknown but when the sampling is random, for some estimates (for example, sample mean, sample proportion) theoretical methods may be used to measure the extent of the variation caused by sampling error.



9/02/2021.

Estimation of Mean and it Variance Notations :-

let 'i' - denotes the i^{th} stratum

'j' - denotes the j^{th} sampling units in the stratum

'N_i' - denotes the no. of units in the i^{th} stratum

'n_i' - denotes the no. of sample units in the selected from the i^{th} stratum.

'W_i' - The Weight of the i^{th} stratum ($W_i = N_i/N$).

'f_i' $= \frac{n_i}{N_i}$ - Sampling fraction in the i^{th} stratum

'y_{ij}' - the value of 'j' unit i^{th} stratum

'$\bar{Y}_i$' $= \frac{\sum_{j=1}^{N_i} y_{ij}}{N_i}$ - The stratum mean of the i^{th} stratum

'$\bar{y}_i$' $= \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$, sampling mean of i^{th} Stratum.

'S_i^2' $= \frac{\sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2}{N_i - 1}$, i^{th} stratum variance

Suppose a population of 'N' units, no. of stratum.

The population mean per unit can be written as,

$$\bar{Y} = \sum_{i=1}^{k} \sum_{j=1}^{N_i} y_{ij} / N \qquad \left[\bar{Y}_i = \frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i} \right.$$

$$= \sum_{i=1}^{k} N_i \bar{Y}_i / N \qquad \Leftarrow \qquad \left. \sum_{j=1}^{N_i} Y_{ij} = \bar{Y}_i N_i \right]$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_i .$$

Consider an estimator $\bar{y}_{st}$ for the population mean $\bar{Y}$

$$\bar{y}_{st} = \sum_{i=1}^{k} N_i \bar{y}_i / N$$

Which is difference from overall sample mean,

$$\bar{y} = \sum_{i=1}^{n} n_i \bar{y}_i / n.$$

Theorem:

If, in every stratum the sample estimator is unbiased and sample is drawn independent in difference stratas Then $\bar{y}_{st}$ is unbiased estimator of population mean and its sampling variance is given by,

$$V(\bar{y}_{st}) = \sum_{i=1}^{k} N_i^2 V(\bar{y}_i) / N^2$$

$$= \sum_{i=1}^{k} W_i^2 V(\bar{y}_i)$$

Pf: W.K.T

$$\bar{y}_{st} = \sum_{i=1}^{k} N_i \bar{y}_i / N$$

Taking expectation on both sides,

$$\bar{y}_{st} = \sum_{i=1}^{k} N_i \bar{y}_i / N$$

$$E(\bar{y}_{st}) = E\left[\sum_{i=1}^{k} N_i \bar{y}_i / N\right]$$

$$= \sum_{i=1}^{k} N_i E(\bar{y}_i) / N$$

$$= \sum_{i=1}^{n} N_i \bar{Y}_i / N \qquad \left[\bar{Y}_i = \frac{\sum_{i=1}^{n} Y_i}{N}\right.$$

$$= \sum_{i=1}^{n} Y_i / N \qquad \left. N\bar{Y}_i = \sum_{i=1}^{n} Y_i\right]$$

$$= \bar{Y}$$

Hence, '$\bar{y}_{st}$' - is unbiased estimator for population mean.

To find out sampling variance of $\bar{y}_{st}$. Since sampling is done independently in each stratum.

$$V(\bar{y}_{st}) = V\left(\sum_{i=1}^{k} N_i \bar{y}_i / N\right)$$

$$= \sum_{i=1}^{k} N_i^2 \frac{V(\bar{y}_i)}{N^2}$$

$$= \sum_{i=1}^{k} W_i^2 V(\bar{y}_i)$$

$$\left[\text{Where } W_i = N_i/N\right].$$

Theorem:

If in every stratum the sample estimator is unbiased and sample is drawn independently in difference strata, Then '$\bar{y}_{st}$' is unbiased estimator of population mean and it sampling variance is given by,

$$V(\bar{y}_{st}) = \sum_{i=1}^{k} N_i^2 \, V(y_i) / N^2$$

$$= \sum_{i=1}^{k} W_i^2 \, V(\bar{y}_i)$$

Pf:-

W.k.T

$$\bar{y}_{st} = \sum_{i=1}^{k} N_i \, \bar{y}_i / N$$

Taking expectation on both sides,

$$E(\bar{y}_{st}) = E\left[\sum_{i=1}^{k} N_i \, \bar{y}_i / N\right]$$

$$= \sum_{i=1}^{k} N_i \, E(\bar{y}_i) / N$$

$$= \sum_{i=1}^{k} N_i \, \bar{Y}_i / N.$$

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

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

Hence, 'y_{st}' is unbiased estimator for population mean.

To find out sampling variance of $\bar{y}_{st}$. Since sampling is done independently in each stratum.

$$V(\bar{y}_{st}) = V\left(\sum_{i=1}^{k} N_i \, \bar{y}_i / N\right)$$

$$= \sum_{i=1}^{k} N_i^2 \, V(\bar{y}_i) / N^2$$

$$= \sum_{i=1}^{k} W_i^2 \, V(\bar{y}_i)$$

Theorem:

For statisfied SRSWOR sampline estimator 'y_{st}' is unbiased and its sampling variance is given by,

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

$$= \sum_{i=1}^{k} (1 - f_i) W_i^2 S_i^2 / n_i$$

Pf:

Since, in each strata, a simple random sampling is taking '$\bar{y}_i$' is an unbiased estimator for '$\bar{Y}_i$' and hence by previous theorem it can be show that

i.e., $E(\hat{\bar{y}}_{st}) = \bar{Y}$

To find out the sampling variance of $(\bar{y}_{st})$ without replacement (~~SRW~~ SRSWOR)

In SRSWOR we have,

$$V(\bar{y}) = (N - n) S^2 / N \cdot n.$$

Now for i^{th} stratum we can write

$$V(\bar{y}_i) = (N_i - n_i) S_i^2 / N_i n_i \longrightarrow (1)$$

We have,

$$V(\bar{y}_{st}) = \sum_{i=1}^{k} N_i^2 V(y_i) / N^2 \longrightarrow (2)$$

Sub/: (1) in (2)

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

$$\left[V(\bar{y}_{st}) = \sum \frac{N_i (N_i - n_i) S_i^2}{N^2 n_i} \right.$$

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

$$= \sum_{i=1}^{k} \left[N_i (N_i - n_i) S_i^2 / N^2 n_i \right] = W_i^2 (1 - f_i) S_i^2$$

//

$$= \sum_{i=1}^{k} (1-f_i) W_i^2 \cdot S_i^2/n_i$$

Systematic Sample:

Estimation of population mean and its sampling Variance:-

Theorem:

In systematic sampling with sampling interval 'k', the sample mean ($\bar{y}_{sy}$) is an unbiased estimator (UBE) of '$\bar{y}$' population mean. It's sampling variance is given by

$$V(\bar{y}_{sy}) = \frac{k-1}{k} S_c^2$$

$$\text{Where } S_c^2 = \frac{\sum_{i}^{k} (\bar{y}_i - \bar{y})^2}{k-1}$$

$\downarrow$

Mean sum of square between column means in the population.

Pf:

Since, the probability of selection of the i^{th} systematic sample is ($1/k$),

$$E(\bar{y}_{sy}) = 1/k \sum_{i=1}^{k} \bar{y}_i = \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} y_{ij}}{nk}.$$

Hence, ($\bar{y}_{sy}$) is an UBE of '$\bar{Y}$'

Total sum of square due to column is given by,

$$\sum_{i}^{k} (\bar{y}_i - \bar{Y})^2 = (k-1) S_c^2$$

Hence, Variance

$$V(\bar{y}_{sy}) = E(\bar{y}_i - \bar{Y})^2$$

$$= \frac{\sum_{i=1}^{k} (\bar{y}_i - \bar{Y})^2}{k}$$

It is informed that variance of systematic sample mean will decrease when sample size increase. That is systematic sample in delicate device and should be used carefully

Theorem:

The Sampling variance of sample mean '$\bar{y}_{sy}$' is given by,

$$V(\bar{y}_{sy}) = \frac{(N-1)}{N} S^2 - \frac{(n-1)k}{n} S^2_w$$

$$= \sigma^2 - \sigma_w^2$$

$$\text{where } S_w^2 = \frac{\sum\sum (y_{ij} - \bar{y}_i)^2}{k(n-1)} \rightarrow (1)$$

with $(n-1) S_w^2 = n \sigma_w^2$ and σ^2 its usual meaning.

$k = \frac{N}{n}$ popn size, Sample size, Sample Interval

Proof:

W.k.T the total sum of square is

$$TSS = (N-1) S^2$$

$$= \sum\sum (y_{ij} - \bar{Y})^2$$

pg no: T.77. fundamental Applied statistics book.

$$TSS = \sum\sum (y_{ij} - \bar{y}_i)^2 + n\sum(\bar{y}_i - \bar{Y})^2 \rightarrow (2)$$

Also We know that, Variance of $(\bar{y}_{sy})$ is

$$V(\bar{y}_{sy}) = \frac{\sum(\bar{y}_i - \bar{Y})^2}{k} \rightarrow (3) \quad (TSS)$$

$$(N-1) S^2 = k(n-1) S_w^2 + nk\, V(\bar{y}_{sy})$$

$$V(\bar{y}_{sy}) = \frac{(N-1)}{N} S^2 - \frac{(n-1)}{n} S^2_w$$

$$nk\,V(\bar{y}_{sys}) = (N-1)S^2 - k(n-1)S^2\omega$$

$k = N/n.$

$\therefore nk = N.$

$$V(\bar{y}_{sys}) = \frac{(N-1)S^2}{nk} - \frac{k(n-1)}{nk}S^2\omega.$$

$$= \frac{(N-1)S^2}{N} - \frac{(n-1)}{n}S^2\omega$$

$$V(\bar{y}_{sys}) = \sigma^2 - \sigma^2\omega.$$

Hence Proved.

Simple random Sampling: (*)

It is tecquine of join sampling in such way that each unit of population as an equal and independent chance of beinging include sample.

In SRS, from population of N units in the probability of drawing any unit at the first draw is $1/N$, the prob. of drawing any unit at the second unit from among the available N-1 units is $1/_{N-1}$ and so on.

Selection of a SRS:- (*) 5m

Random sample can be obtained by any of the following methods.

(i) Lottery method

(ii) Random Nos. method.

(i) Lottery method:

The simpliest method of selecting a random sample is lottery method which is illustrated as follows:

Let a sample of size 25 be required from a population of size 200. Prepare 200 cards or pieces of papers of same size and colour, write on each of them the name or other distinguishing mark of one unit of the population. Fold them uniformly and shuffle them well 25 are to be selected from the lot.

(ii) Random number method: 5m

When the size of the population is large, the method provides an easier procedure. Tippett's table is most popular and it consist of 41600 digits given as 10,400 four figured number.

In this method the difficulty of preparing cards is avoided. The units of the population are first assigned number for identification. Consiqutive nos. are noted from the table of random nos. Those units of the population constitute the sample.

(iii) Population mean square $= S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y}_N)^2$

$$= \frac{1}{N-1} \left[\sum_{i=1}^{N} y_i^2 - N \bar{Y}_N^2 \right]$$

(iv) Sample mean

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

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

Simple Random Sampling without replacement: (SRSWOR)

Important results:

1. $E(\bar{y}_n) = \bar{Y}_N$ (Imp) sm

10 main
2. $E(s^2) = S^2$
3. $V(\bar{y}_n) = \frac{N-n}{N} \frac{S^2}{n}$

Theorem:

Prove that, in SRSWOR, the sample mean is an unbiased estimate of the population mean ie, $E(\bar{y}_n) = \bar{Y}_N$

Proof:

we know that $\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$.

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

Where $a_i = \begin{cases} 1, & \text{if } i^{th} \text{ unit is included in the sample} \\ 0, & \text{" " " is not " " "} \end{cases}$

taking expections on both sides in (1) we get

$$E(\bar{y}_n) = E\left[\frac{1}{n}\sum_{i=1}^{N} a_i Y_i\right]$$

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

Since a_i takes only two values, 1 and 0,

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

$= 1\cdot P(i^{th}$ unit is included in a sample of size of size $n)$

$+ 0\, P(i^{th}$ unit is not included in a sample of size $n)$

$$= 1\cdot\frac{n}{N} + 0\left(1 - \frac{n}{N}\right)$$

$$= \frac{n}{N} \Rightarrow (3)$$

Put (3) in (2), we get

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

$$= \frac{1}{N} \sum_{i=1}^{N} y_i = \bar{Y}_N$$

Remark:

The Prob. that the unit is included in the sample

$$= \sum_{i=1}^{n} \frac{1}{N} = \frac{n}{N}$$

Theorem 2:

Prove that in SRSWOR the sample mean square is an unbiased estimate of the population mean square, ie, $E(s^2) = S^2$

Proof:

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

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

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

$$\boxed{\begin{aligned}(y_1+y_2)^2 &= y_1^2+y_2^2+2y_1y_2\\ &= \sum_{i=1}^{2} y_i^2 + \sum_{i\neq j=1}^{2} y_i y_j\end{aligned}}$$

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

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

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

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

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

$$\boxed{\begin{aligned} i_0 &= y_1^2+y_2^2 + y_1y_2+y_2y_1 \\ &= y_1^2+y_2^2 + 2y_1y_2 \end{aligned}}$$

Taking expectation on both sides we get,

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

Consider

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

$$=\frac{1}{n} E\left[\sum_{i=1}^{n} a_i y_i^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n} E(a_i) y_i^2$$

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

$$=1\cdot\frac{n}{N} + 0\cdot\left(1-\frac{n}{N}\right) = \frac{n}{N}$$

$$E\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2\right] = \frac{1}{n}\sum_{i=1}^{n}\frac{n}{N} y_i^2$$

$$=\frac{1}{N}\sum_{i=1}^{N} y_i^2 \Rightarrow (2)$$

Consider $E\left[\frac{1}{n(n-1)} \sum_{i \neq j=1}^{n} y_i y_j\right] = E\left[\frac{1}{n(n-1)} \sum_{i \neq j=1}^{N} a_i a_j y_i y_j\right]$

$$= \frac{1}{n(n-1)} \sum_{i \neq j=1}^{N} E(a_i a_j) y_i y_j$$

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

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

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

$$= \frac{n}{N} \cdot \frac{n-1}{N-1}$$ $\because$ The even are independent i.e. $P(A \cap B) = P(A) \, P(B/A)$

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

$$= \frac{1}{N(N-1)} \sum_{i \neq j=1}^{N} y_i y_j \Rightarrow (3)$$

Substitute (2) & (3) in (1), we get.

$$E(s^2) = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i^2 - \frac{1}{N(N-1)} \cdot \sum_{i \neq j=1}^{N} y_i y_j \Rightarrow (4)$$

we know that

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

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

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

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

$$\frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{N-1}{N}s^2 = \underline{\bar{Y}_N^2}$$

$$\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{N-1}{N}s^2 = \left[\frac{1}{N}\sum_{i=1}^{N} y_i\right]^2$$

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

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

$$\Rightarrow \sum_{i \neq j = 1} y_i y_j = N\sum_{i=1}^{N} y_i^2 - N(N-1)s^2 - \sum_{i=1}^{N} y_i^2 \Rightarrow (5)$$

Put (5) in (4) we get

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

$$= \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}\sum_{i=1}^{N} y_i^2 + s^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N} y_i^2$$

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

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

P.T.O

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

$$= S^2$$

$$\therefore E(s^2) = S^2$$

Therom 3:

P.T in SRSWOR $V(\bar{y}_n) = \frac{N-n}{n} \frac{S^2}{n}$

Proof:

We know that $V(\bar{y}) = E(\bar{y}_n)^2 - [E(\bar{y}_n)]^2$

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

$\because$ Sample mean is an unbaised estimate of the Population mean.

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

$$\therefore V(\bar{y}_n) = E\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right]^2 - \bar{Y}_N^2$$

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

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

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

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

$$= \sum_{i=1}^{n} \frac{n}{N} y_i^2$$

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

$= 1 \frac{n}{N} - 0 \cdot \left[1 - \frac{n}{N}\right]$

$= \frac{n}{N}$

$$= \sum_{i=1}^{n} \frac{n}{N} y_i^2$$

$$= \frac{n}{N} \sum_{i=1}^{N} y_i^2 \Rightarrow (2)$$

But we know that,

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

$$(N-1) S^2 = \sum_{i=1}^{n} y_i^2 - N \bar{y}_N^2$$

$$(N-1) S^n + N \bar{Y}_N^2 = \sum_{i=1}^{N} y_i^2 \Rightarrow (3)$$

Subtitution (3) in (1) we get,

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

$$= \frac{n(N-1)}{N} S^2 + n \bar{Y}_N^2 \Rightarrow (4)$$

Consider:

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

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

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

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

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

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

$$= \sum_{i \neq j = 1}^{N} \frac{n}{N} y_i y_j = \frac{n(n-1)}{N(N-1)} \sum_{i \neq j = 1}^{n} y_i y_j \Rightarrow (5)$$

But $\left(\sum_{i=1}^{N} y_i\right)^2 = \sum_{i=1}^{} y_i^2 + \sum_{i \neq j = 1}^{N} y_i y_j$

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

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

$$(N \bar{Y}_N)^2 - (N-1)S^2 - N \bar{Y}_N^2 = \sum_{i \neq j = 1}^{N} y_i y_j$$

$$\Rightarrow \sum_{i \neq j = 1}^{N} y_i y_j = N^2 \bar{Y}_N^2 - N \bar{Y}_N^2 - (N-1)S^2$$

$$= \bar{Y}_N^2 N(N-1) - (N-1)S^2 \Rightarrow (6)$$

Substitute (6) in (5), we get.

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

$= n(n-1)\bar{Y}_N^2 - \frac{n(n-1)}{N} S^2 \Rightarrow (7)$

Substitute (4) & (7), we get.

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

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

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

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

$$= \frac{S^2}{n} - \frac{S^2}{N}$$

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

Remark:

1. We know that variance of $V(\bar{y}_n) = \frac{N-n}{N}\frac{S^2}{n}$

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

$$= (1-f)\frac{S^2}{n}$$

Where $f = \frac{n}{N}$ is called sampling fraction and the factor $(1-f)$ is called finite population Correlation (F.P.C).

If the population size N is very large or if n is sample compare with N then

$$f = \frac{n}{N} \to 0 \ \& \ F.P.C \Rightarrow 1$$

If be f.p.c $\to$ 1 then $V(\bar{y}_n) = \frac{S^2}{n}$

2. Standard eorror of the sampling distribution of $S.E(\bar{y}_n) = \sqrt{\text{variance}}$

$$= \sqrt{\frac{N-n}{N}\frac{S^2}{n}} = \sqrt{1-\frac{n}{N}}\ \frac{S}{\sqrt{n}}$$

$$= \sqrt{1-F}\ \frac{S}{\sqrt{n}}.$$

$\therefore \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$

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

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

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

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

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

Simple random sampling with replacement:
[SRSWR]

35 4|310
35
20
20
10

15
10
2

Important results:

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

2. $V(\bar{y}_n) = \frac{\sigma^2}{n}$ (or) $\frac{N-1}{N} \frac{S^2}{n}$

3. $E(s^2) = \sigma^2$

Theorem 1:

In SRSWR Prove that $E(\bar{y}_n) = \bar{Y}_N$

Proof:
The Proof in SRSWOR can be reproduced here.

Theorem 2:

In SRSWR, Prove that

$$V(\bar{y}_n) = \frac{\sigma^2}{n} \text{ (or) } \frac{N-1}{N} \frac{S^2}{n}$$

Proof:
We know that, $V(\bar{y}_n) = V\left[\frac{1}{n} \sum_{i=1}^{n} y_i\right]$

$V(ax) = a^2 V(x)$

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

$$= \frac{1}{n^2} \sum_{i=1}^{n} V(y_i)$$

Here each draws are identically independently distributed we get the common variance i.e.,

$V(y_i) = \sigma^2$

$$\therefore V(\bar{y}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2$$

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

$$= \frac{\sigma^2}{n} \Rightarrow (1)$$

But $\Rightarrow$ Population variance

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y}_N^2)$$

$$N\sigma^2 = \sum_{i=1}^{N} [y_i - \bar{Y}_N^2] \Rightarrow (2)$$

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

$$(N-1)S^2 = \sum_{i=1}^{N} (y_i - \bar{Y}_N^2) \Rightarrow (3)$$

Comparing (2) & (3), we get

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

$$= \sigma^2 = \frac{(N-1)}{N} S^2 \Rightarrow (4)$$

Put (4) in (1), we get

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

Theorm 3:

In SRSWR, Prove that

$$E(s^2) = \sigma^2$$

Proof:

We know that,

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

Add and subtract $\bar{Y}$, we get

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

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

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

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

$$\because \sum (y_i - \bar{Y}) = \sum y_i - n\bar{Y} = n\bar{y} - n\bar{Y} = n(\bar{y} - \bar{Y})$$

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

Taking Expectation on both sides,

$$E(s^2) = \frac{1}{n-1} \left[\sum_{i=1}^{n} E(y_i - \bar{Y})^2 - nE(\bar{y} - \bar{Y})^2 \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} \sigma_i^2 - \cancel{n} \frac{\sigma^2}{\cancel{n}} \right]$$

$$\because E(\bar{y} - \bar{Y})^2 = V(\bar{y}) = \frac{\sigma^2}{n}$$

$$= \frac{1}{n-1} \left[n\sigma^2 - \sigma^2 \right]$$

$$= \frac{1}{\cancel{n-1}} \times \sigma^2 \cancel{(n-1)}$$

$$= \sigma^2 //$$

SRSWOR vs SRSWR.

In SRSWOR, we know that,

$$V(\bar{y}_n) = \frac{N-n}{N} \frac{S^2}{n} \Rightarrow (1)$$

If we consider SRSWR from a population with variance σ^2, then

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

$$= \frac{1}{n^2} \sum_{i=1}^{n} V(y_i)$$

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

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

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

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

$$\therefore V(\bar{y}_n) = \frac{N-1}{N} \frac{S^2}{n} \Rightarrow (2)$$

Comparing (1) and (2), we see that the variance of the sample mean is more in SRSWR as compared with its variance in SRSWOR i.e., $V(SRSWR) > V(SRSWOR)$.

Hence, SRSWOR provides more efficient estimator of $\bar{Y}_N$ relative to SRSWR

STRATIFIED RANDOM SAMPLING:

Definition:

In simple random sampling it has been seen that the precision of the standard estimator of the population total depends on two aspects namely the sample size n and variability S^2 of the character under study. In SRSWOR we obtain $V(\bar{y}_n) = \frac{N-n}{N} \frac{S^2}{n}$

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

This implies that variance of $\bar{y}_n$ is.

(i) Inversely proportional to the sample size.

ii) Directly proportional to the variable of the sampling units in the population.

∴ In order to get the estimator with increased precision one can increase the sample space. However consideration of caused limits the size of the sample.

The other possible way to estimate the population total with greater precisions is divide the population into several groups each of which is more homogenous then the entire population and drop the sample of pre-determined size from each of these groups. The groups into which the population is divided are called strata and drawing sample from each of the strata is stratified sampling.

Stratification:

Stratification means division into layers. Auxiliary information related to the character under study may be used to divide the population into various groups.

Stratified random sampling:

The population consisting of N sampling units is divide into k relatively homogeneous mutually disjoint subgroups termed as strata, of sizes $N_1, N_2, \ldots N_k$ such that ΣN_i. If a simple random sample of size n_i, $(i = 1, 2, \ldots, k)$ is drawn from each of the stratum respectively such that $n = \Sigma n$, the sample is termed as stratified random sample of size n and the technique of drawing such a sample sample is called stratified random sampling.

Notations and Terminology:

In population there are N sample units and they are divided into k no. of strata.

N_i = The no. of sampling units in the i^{th} stratum $(i = 1, 2, \ldots k)$

$$N = \sum_{i=1}^{k} N_i$$

n_i = The no. of sampling units selected with SRSWOR from the i^{th} stratum.

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

N_1	N_2	N_3	....	N_i ..	N_k
↓	↓	↓			

$n_1 + n_2 + n_3$ ...... $+ n_i + \cdots + n_k = n$

Population:

Let Y_{ij}, ($j = 1, 2, \ldots, N_i$, $i = 1, 2, \ldots, k$) be the value of the j^{th} unit in the i^{th} stratum.

$\overline{Y}_{N_i}$ = Population mean of i^{th} stratum

$$= \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}$$

$\overline{Y}_N$ = Population mean

$$= \frac{1}{N} \sum_i \sum_j Y_{ij}$$

$$= \frac{1}{N} \sum_{i=1}^{k} N_i \overline{Y}_{N_i}$$

$$= \sum_{i=1}^{k} P_i \overline{Y}_{N_i}$$

Where $P_i = \frac{N_i}{N}$ = weight of i^{th} stratum

S_i^2 = Population mean square of the i^{th} stratum

$$= \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (Y_{ij} - \overline{Y}_{N_i})^2, \quad (i = 1, 2, \ldots k).$$