UNIT - II

SIMPLE RANDOM SAMPLING

The technique or Method of Selecting a Sample is of fundamental importance in the theory of Sampling and usually depends upon the nature of the data and type of enquiry.

The procedures of Selecting a Sample may be broadly classified under the following three heads:

(i) Subjective (or) judgement Sampling

(ii) Probability Sampling

(iii) Mixed Sampling.

Subjective or Purposive or Judgement Sampling

In this Sampling the Sample is selected with definite purpose in view and the choice of the Sampling units depends entirely on the discretion and judgement of the investigator.

Probability Sampling

It is the scientific method of selecting samples according to some laws of chance in which each unit in the population has some definite pre-assigned probability of being selected in the sample.

Mixed Sampling

If the samples are selected partly according to some laws of chance and partly according to a fixed Sampling rule (no assignment of probabilities) they are termed as mixed samples and the technique of selecting

Such samples is known as mixed sampling.

The different types of sampling as given above have a no of variations. Some of which may be listed below.

1. Simple Random sampling
2. Stratified Random Sampling
3. Systematic Sampling
4. Multistage Sampling
5. Quasi Random sampling
6. Area Sampling
7. Simple Cluster Sampling
8. Multistage cluster sampling
9. Quota sampling.

Simple Random Sampling (SRS)

It is the technique of drawing a sample in such a way that each unit of the population has an equal and independent chance of being included in the sample.

SRS can also be defined equivalently as follows. Let us suppose that a sample of size n is drawn from a population of size N. There are $\binom{N}{n}$ possible samples. SRS is the technique of selecting the sample in such a way that each of the $\binom{N}{n}$ samples has an equal chance or probability $p = \frac{1}{\binom{N}{n}}$ of being selected.

Simple Random Sampling with Replacement (SRSWR)

In SRSWR, the sample selected at first draw is replaced, before the second sample is drawn.

Suppose if our population consists of N units, the probability of selecting any unit at first draw is $\frac{1}{N}$, the probability of selecting any unit at second draw is also $\frac{1}{N}$. Here a same unit may be selected more than one times.

Simple Random Sampling without Replacement (SRSWOR)

In SRSWOR, the sample selected a 1^{st} draw is not replaced, ~~before~~ the second sample is drawn. The probability of selecting any unit in the first draw is $\frac{1}{N}$, the probability of selecting a another unit in the 2^{nd} draw is $\frac{1}{N-1}$. Here there is no chance for the same unit to be selected for more than one time.

Theorem 1 :

Statement

In SRSWOR, the probability of selecting a specified unit of the population at any given draw is equal to the probability of its being selected at the first draw.

Proof :

Consider the population consisting of N units.

Observation of selecting any unit at the 1^{st} draw $- \frac{1}{N}$

" 2^{nd} " $- \frac{1}{N-1}$

" r^{th} draw $-$ $\frac{1}{\ }$

Let E_r $(r = 1, 2 \ldots n)$ denote the event of selecting any specified unit at r^{th} draw

$P(E_r) =$ Prob { that the specified unit is not selected in any one of the previous $(r-1)$ draws and then selected at the r^{th} draw }

$$= \left(1 - \frac{1}{N}\right) \times \left(1 - \frac{1}{N-1}\right) \cdots \left(1 - \frac{1}{N-(r-2)}\right) \times \left(\frac{1}{N-(r-1)}\right)$$

$$= \frac{N-\cancel{1}}{N} \times \frac{N-\cancel{2}}{N-\cancel{1}} \times \frac{N-3}{N-\cancel{2}} \cdots \frac{N-\cancel{r+1}}{N-r+2} \times \frac{1}{N-\cancel{r+1}}$$

$$P(E_r) = \frac{1}{N} = P(E_1)$$

Thus, the probability of selecting the any specified unit at any given draw is equal to the probablity of selecting it at the first draw.

Theorem : 2

S.T the probability of a specified unit being included in the sample is $\frac{n}{N}$

Proof :

The prob/ of a speified unit selected at r^{th} draw is $P(E_r) = \frac{1}{N}$ $(r = 1, 2 \ldots n$

Since a speified unit can be included in a sample of size n is 'n' mutually exclusive ways.

$\therefore$ by addition theorem, the prob/ of specified unit is included in the sample is $\frac{1}{N} + \frac{1}{N} + \cdots + \frac{1}{N}$ (n times) (ie) $= \frac{n}{N}$.

Notations and Terminology :

Let us consider a finite population of N units and let y be the character under consideration. The capital letters are used to describe the characteristics of the population whereas small letters refer to sample observations.

Thus, for example the N population units may be denoted by $U_1, U_2 \ldots U_N$ and the sample units will be denoted by $u_1, u_2 \ldots u_n$.

Let Y_i $(i = 1, 2 \ldots N)$ be the value of the character for the i^{th} unit in the population and the corresponding small letters y_i $(i = 1, 2 \ldots n)$ denote the value of the character for i^{th} unit selected in the sample. Then we define

$$\text{Population Mean} = \bar{Y}_N = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

$$\text{Sample Mean} = \bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Sample mean $\bar{y}_n$ may also be written alternatively as

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^{N} a_i Y_i$$

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

S^2 = Mean square for the population

= Population Mean square

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

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

s^2 = Mean square for the sample

= Sample Mean square

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

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

Theorem : 3

In SRS WOR , the sample mean is an unbiased estimate of the population Mean

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

Proof:

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

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

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

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

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

Hence

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

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

$\therefore E(\bar{y}_n) = \bar{Y}_N$ Hence the proof.

Theorem : 4

In SRSWOR, the sample mean square is an unbiased estimate of the population Mean square

ie) $E(s^2) = S^2$

Proof :

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

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

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

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

$$\because \sum y_i^2 = \sum y_i^2 + \sum y_i y_j$$

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

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

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

$$s^2 = \frac{1}{n}\sum_{i=1}^{N} a_i Y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{N} a_i a_j Y_i Y_j$$

$$\because \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{N} a_i Y_i^2, \quad \sum_{i=1}^{n} y_i y_j = \sum_{i=1}^{N} a_i a_j Y_i Y_j$$

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

$$\therefore E(s^2) = \frac{1}{n}\sum_{i=1}^{N} E(a_i)\, y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{N} E(a_i a_j)\, y_i y_j$$

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

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

$$\Rightarrow E(a_i) = \frac{n}{N}$$

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

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

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

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

$\because P(A \cap B) = P(A) \cdot P(B/A)$

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

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

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

$$\sum y_i y_j = \left(\sum y_i\right)^2 - \sum y_i^2$$

$$= (N\bar{y}_N)^2 - \sum y_i^2$$

$\because \frac{1}{N}\sum y_i = \bar{y}_N$

$$= N^2 \bar{y}_N^2 - \sum y_i^2$$

$\sum y_i = N\bar{y}_N$

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

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

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

$$= \frac{1}{\cancel{N}} \cdot \frac{\cancel{N}}{N-1}\sum_{i=1}^{N} y_i^2 - \frac{N}{N-1}\bar{y}_N^2$$

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

(ie) $E(s^2) = S^2$

$\Rightarrow$ The sample mean square is an unbiased estimator of population mean square.

Theorem : 5

In SRSWOR, the variance of the sample mean is given by $V(\bar{y}_n) = \frac{S^2}{n} \cdot \frac{N-n}{N} = (1-f)\frac{S^2}{n}$

Proof :

We have $\bar{y}_n = \sum_{i=1}^{n} \frac{y_i}{n}$

$$V(\bar{y}_n) = E(\bar{y}_n^2) - [E(\bar{y}_n)]^2 \quad \because V(x) = E(x^2) - (E(x))^2$$

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

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

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

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

$$\left(\because \sum_{i=1}^{n} y_i^2 = \sum y_i^2 + \sum_{i\neq j=1}^{n} y_i y_j\right)$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{N} a_i Y_i^2 + \sum_{i\neq j=1}^{N} a_i a_j Y_i Y_j\right]$$

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

$$\text{ie) } E(\bar{y}_n)^2 = \frac{1}{n^2}\left[\sum_{i=1}^{N} E(a_i) Y_i^2 + \sum_{i\neq j=1}^{N} E(a_i a_j) Y_i Y_j\right]$$

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

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

$$= \frac{1}{n.N}\sum_{i=1}^{N}y_i^2 + \frac{(n-1)}{n.N(N-1)}\sum_{i\neq j=1}^{N}y_i y_j \quad - \quad *$$

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

$$\Rightarrow \sum y_i^2 = (N-1)s^2 + N\bar{Y}_N^2 \rightarrow (1)$$

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

$$= (N\bar{Y}_N)^2 - \left[(N-1)s^2 + N\bar{Y}_N^2\right] \text{ using } s^2 \text{ formula}$$

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

using (1) & (2), (*) becomes

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

$$= \frac{(N-1)s^2}{n.N} + \frac{\cancel{N}\bar{Y}_N^2}{n.\cancel{N}} + \frac{n-1}{n.\cancel{N}(N-1)}N^{\cancel{2}}\bar{Y}_N^2 - \frac{(n-1)\cancel{(N-1)}s^2}{n.N\cancel{(N-1)}} - \frac{(n-1)\cancel{N}\bar{Y}_N^2}{n.\cancel{N}(N-1)}$$

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

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

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

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

$$\therefore V(\bar{y}_n) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) + \bar{Y}_N^2 - \bar{Y}_N^2$$

$$V(\bar{y}_n) = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

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

$$\therefore V(\bar{y}_n) = \frac{s^2}{n}(1-f) \quad \because \frac{n}{N} = f$$

Hence the Proof.

Variance of the estimate of the population Total

(ie) $V(\hat{Y}_N) = \frac{N^2 s^2}{n}(1-f)$

Proof :

we have $V(\bar{Y}_N) = \frac{s^2}{n}(1-f)$

$$V(\hat{Y}_N) = V(N\bar{Y}_N)$$

$$= N^2 V(\bar{Y}_N)$$

$$\therefore V(\hat{Y}_N) = \frac{N^2 s^2}{n}(1-f) \text{ using } V(\bar{y}_n)$$

Remarks

$f = \frac{n}{N}$ is called the sampling fraction and

Consequently

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

$$= (1-f)\, s^2/n$$

The factor $(1-f)$ is called the finite population correction (f.p.c). If the pop. size N is very large or n is small compared with N then $f = \frac{n}{N} \to 0$ and consequently f.p.c $\to 1$.

Selection of a simple Random Sample :

Some procedure which is simple and good for small population is not so for the large population. Generally the method of selection should be independent of the properties of sampled population. Proper care to be taken to ensure that selected sample is random.

Methods of selecting simple Random Sampling are

1. Lottery Method
2. Mechanical Randomization or Random Numbers

Lottery Method

In this method all the items in the population of size N are numbered from 1 to N. These numbers are written on some slips which are identical in shape and size. Then those slips are put in a box and thoroughly mixed. Take one slip at random and note that number. The unit corresponds to this number will be included in the sample. Repeat the procedure n times to get the desire sample of size n.

Table of Random Numbers (or) Mechanical Randomization

When the population size increases, it becomes more and more difficult to follow lottery method. In a table of random numbers there is an equal probability for any digit from 0 to 9 appear in any Particular position.

The following steps involved in selection of the sample using Random Numbers tables.

1. Assign the serial number for each member in the Population say 1 to N. If we want to select a sample from the population of size $N (\leq 99)$. Then the numbers are combined two by two to give Pairs from 0 to 99. Similarly if $N (\leq 999)$, $N (\leq 9999)$ and so on; then Combine the digits are three by three, four by four and so on respectively.
2. The starting point may be selected at random the table can be read either horizontally or vertically.
3. Read the first number, if it is less than the Population size, unit in the population corresponding to this number is included in the sample. If the number is greater than the pop size or repeated then we select the next number by the similar way.

Simple Random Sampling for proportions (or) Attributes

Sometimes units in the Populations may be classified into two groups namely possessing a Particular characteristic and not having the characteristic For ex: the persons are classified into employed and unemployed (graduate and ungraduate) Smokers and non-smokers etc., Similarly the villages are classified as large and small (according to the population) factories are classified in to large and small (by the amount invested as well as the no of employers) Thus we want to estimate the proportion or percentage of total no. of units in the population possessing the characteristics.

Notations

Let us suppose that the pop consisting of N units and it is classified in to two classes c (possessing the attribute) and c' (not possessing the attributes)

Let us define the variable

$$y_i = \begin{cases} 1, & \text{if } i^{th} \text{ unit falls in } c \text{ (sampling unit possess the attributes)} \\ 0, & \text{if } i^{th} \text{ unit } \sout{\text{does not pass}} \text{ falls in } c' \text{ (sampling unit doesnot possess the attributes)} \end{cases}$$

Let a and A denote the total no. of units having the characteristic in the sample and population respectively

$$a = \sum_{i=1}^{n} y_i \qquad A = \sum_{i=1}^{N} Y_i$$

$$a + a' = n \qquad A + A' = N$$

Sample proportion

$$p = \frac{a}{n} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$$

$$q = 1 - p$$

Population proportion

$$P = \frac{A}{N} = \frac{\sum_{i=1}^{N} Y_i}{N} = \overline{Y}$$

$Q = 1 - P$

Population mean square :

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

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

$$= \frac{1}{N-1} \left[A - NP^2 \right] \qquad \because \sum_{i=1}^{N} Y_i^2 = A = NP$$

$$= \frac{1}{N-1} \left[NP - NP^2 \right]$$

$$= \frac{NP(1-P)}{N-1} = \frac{NPQ}{N-1}$$

Sample Mean square

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

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

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

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

$$= \frac{np(1-p)}{n-1} = \frac{npq}{n-1}$$

Theorem: 1

S.T the sample proportion is an unbiased estimate of the population proportion (ie) $E(p) = P$

Proof

consider $p = \frac{a}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$

Taking expectations on both sides we get

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

$$= \frac{1}{n} \sum_{i=1}^{n} \bar{Y} \qquad \because E(y_i) = \bar{Y}$$

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

$$= \frac{1}{n} \cdot nP = P$$

$$\therefore E(p) = P$$

Thus the sample proportion is an unbiased estimate of the population proportion.

Note

An estimate of A is given by

$$\hat{A} = NP \qquad E(p) = P$$

$$E(NP) = NP = A$$

Theorem : 2

In SRSWOR, the variance of the sample proportion p is given by $V(p) = \left(\frac{N-n}{N-1}\right) \frac{PQ}{n}$

Proof :

In SRSWOR, the variance of sample mean is

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

In SRS for proportions, we have

$$\bar{y} = p \quad \text{and} \quad S^2 = \frac{NPQ}{N-1}$$

using the above, we can write the eqn (1) as

$$V(p) = \frac{N-n}{N} \frac{NPQ}{N-1} \times \frac{1}{n}$$

$$= \left(\frac{N-n}{N-1}\right) \frac{PQ}{n}$$

Note

The variance of $\hat{A}$ is given by

$$V(\hat{A}) = N^2 V(p)$$

$$= N^2 \frac{N-n}{N-1} \cdot \frac{PQ}{n}$$

Theorem : 3

The unbiased estimate of variance $V(p)$ is given by

$$V(\hat{p}) = v(p) = \frac{N-n}{N} \frac{pq}{n-1}$$

Proof

In SRSWOR, an unbiased estimate of $V(\bar{y})$ is given by

$$\hat{V}(\bar{y}) = v(\bar{y}) = \frac{N-n}{N} \frac{s^2}{n}$$

In SRS for proportion p takes the place of $\bar{y}$

and $s^2 = \frac{npq}{n-1}$

$$\therefore v(p) = \frac{N-n}{N} \frac{npq}{n(n-1)}$$

$$v(p) = \frac{N-n}{N} \frac{pq}{n-1}$$

Finite population correction

When we draw a random sample of size n from an infinite population the variance of the sample mean is σ^2/n. On the other hand if we draw a r.s of size n from the finite population, the variance of the sample mean is

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

Where $f = \frac{n}{N}$ = sampling fraction

Introduction of the factor $(1-f)$ in the variance formula is known as finite population correction factor (f.p.c). If the pop size N is very large or if n is small compared with N then $f = \frac{n}{N} \to 0$ and consequently f.p.c $\to 1$

In practice f.p.c is ignored whenever the sampling fraction is less than 5%.

Problem :

Consider a pop of 6 units with values 1, 2, 3, 4, 5, 6 write down all possible samples of 2 (without replacement) from this pop and verify that sample mean is an unbiased estimate of the population mean. Also calculate its sampling variance and verify that

(i) It agrees with the formula for the variance of the sample mean and

(ii) This variance is less than the variance obtained from sampling with replacement.

Soln :

$Y =$ 1 2 3 4 5 6

$y^2 =$ 1 4 9 16 25 36

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{21}{6} = 3.5$$

$$S^2 = \frac{1}{N-1} \Sigma (Y - \bar{Y})^2$$

$$= \frac{1}{N-1} [\Sigma y^2 - N\bar{y}^2]$$

$$= \frac{1}{5}\left[91 - 6 \times 12.25 \right] = 3.5$$

$$\sigma^2 = \frac{N-1}{N} s^2 = \frac{5}{6} \times 3.5 = 2.917$$

The total no. of samples of size $n=2$ from a pop of $N=6$ units is ${}^6C_2 = 15$, as enumerated in the following table

Sample NO	Sample values y	$\bar{y}$	$\bar{y}-\bar{Y}$	$(\bar{y}-\bar{Y})^2$
1	(1,2)	1.5	-2	4
2	(1,3)	2	-1.5	2.25
3	(1,4)	2.5	-1	1
4	(1,5)	3	-0.5	0.25
5	(1,6)	3.5	0	0
6	(2,3)	2.5	-0.1	1
7	(2,4)	3	-0.5	0.25
8	(2,5)	3.5	0	0
9	(2,6)	4	0.5	0.25
10	(3,4)	3.5	0	0
11	(3,5)	4	0.5	0.25
12	(3,6)	4.5	1.0	1
13	(4,5)	4.5	1.0	1
14	(4,6)	5	1.5	2.25
15	(5,6)	5.5	2	4
Total		52.5	0	17.5

$$E(\bar{y}) = \frac{1}{N_{C_n}} \sum_{i=1}^{15} \bar{y}_i$$

$$= \frac{52.5}{15} = 3.5 = \bar{Y}$$

which implies that sample mean $\bar{y}$ is an unbiased estimate of Pop mean $\bar{Y}$

$$V(\bar{y}) = \frac{1}{15} \sum_{i=1}^{15} (\bar{y}_i - \bar{Y})^2$$

$$= \frac{1}{15} (17.5) = 1.167$$

(i) $$V(\bar{y}_{SRSWOR}) = \frac{N-n}{Nn} S^2$$

$$= \frac{6-2}{6 \times 2} 3.5 = 1.167$$

$$V(\bar{y}) = V(\bar{y}_{SRSWOR})$$

(ii) $$V(\bar{y}_{SRSWR}) = \frac{\sigma^2}{n} = \frac{2.917}{2} = 1.458$$

$$V(\bar{y}_{SRSWR}) > V(\bar{y}_{SRSWOR})$$

2. A Population consists of values 4, 6, 9, 10 and 11. Draw all possible samples of size 2 under SRSWOR and verify that $E(s^2) = S^2$

Soln:

Sample No	Sample values	$\bar{y}$	$(\bar{y}-\bar{Y})$	$(\bar{y}-\bar{Y})^2$
1	(4, 6)	5	-3	9
2	(4, 9)	6.5	-1.5	2.25
3	(4, 10)	7	-1	1
4	(4, 11)	7.5	-0.5	0.25
5	(6, 9)	7.5	-0.5	0.25

6	(6, 10)	8	0	0
7	(6, 11)	8.5	0.5	0.25
8	(9, 10)	9.5	1.5	2.25
9	(9, 11)	10	2	4
10	(10, 11)	10.5	2.5	6.25
				25.5

$$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{4+9+6+10+11}{5} = 8$$

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

Y_i	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
4	-4	16
6	-2	4
9	1	1
10	2	4
11	3	9
		34

$$S^2 = \frac{1}{4}(34) = 8.5$$

$$E(s^2) = \frac{N-n}{Nn} S^2 \Rightarrow$$

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

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

$$= \frac{1}{10}(25.5) = 2.55$$

$$\frac{Nn}{N-n} s^2 = \frac{5 \times 2}{5-2} (2.55)$$

$$= \frac{10}{3} 2.55 = 8.5$$

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

Determination of Sample Size

The first step of a Statistician in a sample survey is to decide how large the sample should be. There is no hard and fast rule for deciding the sample size. However we have rational method to solve this problem.

When we take a small sample, the estimate obtained is inaccurate and less useful. On the other hand if we take the larger sample, the accuracy of the estimate is higher than what we required and also the cost involved is more. Consequently the first step is to decide how large an error we can tolerate the estimate. The next step is to express the allowable error in terms of confidence limits.

For large n, the sample mean is assumed to be normally distributed. Then $100(1-\alpha)\%$ confidence Interval is given by

$$\bar{y} \pm Z_\alpha \, \sigma/\sqrt{n}$$

where Z_α is the critical value from Standard normal table at $\alpha\%$ l.o.s.

Here we have ignore the finite population correction

Let $L = Z_\alpha \dfrac{\sigma}{\sqrt{n}}$

$$\sqrt{n} = \frac{Z_\alpha \sigma}{L}$$

$$\therefore n = \frac{Z_\alpha^2 \sigma^2}{L^2} \longrightarrow (1)$$

To find n we must have estimate of σ usually the values of σ taken from the results of previous samplings of these population or of similar population

In eqn (1) we ignored f.p.c This formula suits for majority of applications. If f.p.c is more than 10% a revised value of n say n' should be consider by taking into account f.p.c

The value of n' is given by

$$n' = \frac{n}{1+f} \quad \text{where } f = \frac{n}{N}$$

S.T $\hat{Y} = N\bar{y}$ is an unbiased estimate of the population total (Y). Also obtain its variance.

Proof:

Consider $\hat{Y} = N\bar{y}$

Taking expectation on both sides, we get

$$E(\hat{Y}) = E(N\bar{y})$$

$$= N\,E(\bar{y}) \qquad \because E(\bar{y}) = \bar{Y}$$

$$= N\bar{Y} \qquad \because \bar{Y} = \frac{Y}{N}$$

$$\therefore E(\hat{Y}) = Y$$

Theorem

In SRSWR the variance of the sample mean is given by

$$V(\bar{y}) = \frac{N-n}{N}\,\frac{s^2}{n}$$

Proof

Consider $V(\bar{y}) = E(\bar{y} - \bar{Y})^2$

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

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

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

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

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

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

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

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

Consider

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

$$= \sigma^2 \rightarrow (2)$$

Since the draws are independent we have

$$E(y_i - \bar{Y})(y_j - \bar{Y}) = 0 \rightarrow (3)$$

using (2) & (3), (1) can be written as

$$V(\bar{y}) = \frac{1}{n^2}\left[\sum_{i=1}^{n} \sigma^2 + 0\right]$$

$$= \frac{1}{n^{\cancel{2}}} \cancel{n} \cdot \sigma^2 = \sigma^2/n \qquad \because \sigma^2 = \frac{N-1}{N} S^2$$

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

Hence the proof.

Note :

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

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

$$\therefore V(\bar{y})_{SRSWOR} < V(\bar{y})_{SRSWR}$$

Theorem

In SRSWR, The sample mean square is always an unbiased estimate of the population Variance

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

Proof

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

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

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

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

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

$$\because \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

Taking expectation on both sides

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

Consider

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

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

$$= \sigma^2/n \longrightarrow (3)$$

using (2) & (3), (1) can be written as

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

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

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

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

Hence the proof.