Sampling Theory

Introduction

Before giving the notion of sampling we will first define population. In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group.

Population:- The group of individuals under study is called Population.

Thus in Statistics, population is an aggregate of objects, animate or inanimate under study. The population may be finite or infinite.

It is obvious that for any statistical investigation complete enumeration of the population is rather impracticable. e.g. if we want to have an idea of the average per capita(monthly) income of the people in India, we will have to enumerate all the earning individuals in the country, which is rather a very difficult task.

If the population is infinite, complete enumeration is not possible. Also if the units are destroyed in the course of inspection 100% inspection, though possible, is not at all desirable. But even if the population is finite or the inspection is not destructive; 100 % inspection is not taken recourse to because of multiplicity of causes viz, administrative and financial implications, time factor etc. and we take the help of Sampling.

Sample:- A finite subset of statistical individuals in a population is called a sample

Sample size:- The number of individuals in a sample is called the sample size.

For the purpose of determining population characteristics, instead of enumerating entire population, the individuals in the sample only are observed. Then the sample characteristics are utilized to approximately determine or estimate the population.

Sampling is quite often used in our day-to-day practical life. e.g. in a shop we assess the quality of sugar, wheat or any other commodity by taking a handful of it from the bag and then decide to purchase it or not. A housewife normally tests the cooked products to find if they are properly cooked and contain the proper quantity of salt.

Parameter:- The statistical constants viz; mean, variance of the population is called as parameter.

Statistic:- Statistical measures computed from the sample observations e.g. mean, variance is called as statistic.

In practice parameter values are not known and their estimates based on the sample values are generally used. Thus statistic may be regarded as an estimate of the parameter, obtained from sample is a function of sample values only. It may be pointed out that a statistic as it is based on sample values and as there are multiple choices of the samples that can be drawn from a population varies from sample to sample.

Census Survey

In this method is to be followed each and every member of the population is approached and information is collected about and from him. The term population or universe denotes the totality of the objects of study. e.g. if we are going to study the economic conditions of primary teachers in Maharashtra is the population for this study. If we are going to study, the wages of the land labourer in a district and if there are 5 lakh workers then totality of all these 5 lakh workers is the population for this study. The method of collecting data from entire population is called the census survey. If the census method is to be followed in the above examples, then we have to collect about the economic conditions of every primary teacher in Maharashtra and the wages of every land labourer of the districts.

Sample Survey

If instead of studying the entire population, a part of it is studied it is called the sample survey. From the study of the sample we draw the inference about the entire population. it is assumed that if the sample is unbiased and sufficiently large in number then the characteristics of the population do not differ widely from those of the samples chosen properly. Thus the sampling method is to be used in the above example we would study the economic conditions of few properly selected primary teachers and would estimate the results for all the teachers.

Advantages of sampling method over census method

The main advantages of sampling method over the census method may be as follows

1) **Less Time**- There is considerable saving in time and labour since only part of the population has to be examined. The sampling results can be obtained more rapidly and the data can be analysed much faster since relatively fewer data have to be collected and processed.

2) **Reduced cost of the Survey**- Sampling usually results in reduction in cost in terms of money and in terms of man hours. The total cost of sample survey is expected to be much smaller than the complete census. Since in most of the cases our resources are limited in terms of money and the time within which the results of the survey should be obtained, it is usually imperative to resort to sampling rather than complete enumeration.

3) **Greater Accuracy of results**-The results of sample survey are usually much more reliable than those obtained from a complete census due to the following reasons.

i) It is always possible to determine the extent of the sampling error.

ii) The non-sampling errors due to number of factors such as training of field workers, measuring and recording observations, location of units, incompleteness of returns, etc. are likely to be of a serious nature in complete census than in a sample survey.

4) **Greater Scope**- Sample survey has generally greater scope as compared with complete census. The complete enumeration is impracticable, rather inconceivable if the survey requires a highly trained personnel and more sophisticated equipment for the collection and analysis of the data.

5) If the population is too large, as e.g. trees in jungle, we are left with no way but to resort to sampling.

6) If the testing is destructive i.e. if the quality of an article can be determined only by destroying the article in the process of testing. e. g.

i) Testing of quality of milk or chemical salt by analysis

ii) Testing of breaking strength of chalks.

iii) Testing of crackers and explosives

iv) Testing of life of an electric tube or bulbs etc.

Complete enumeration is impracticable and sampling technique is the only method to be used in such cases.

7) If the population is hypothetical e. g. in coin tossing problem where the process may continue indefinitely (any number of times) sampling method is the only scientific method of estimating the parameters of the universe.

Methods of 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 types of enquiry. The procedures of selecting a sample may be broadly classified under the following three methods.

1) **Deliberate (Purposive) Sampling**

2) **Probability Sampling**

3) **Mixed Sampling**

1) Deliberate (Purposive) 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 judgment of the investigator. This sampling suffers from drawbacks of favoritism and nepotism depending upon the beliefs and prejudices of the investigator and thus does not give a representative sample of the population. e.g. if an investigator wants to give the picture that the standard of living has increased in the city of New Delhi, he may take individuals in the sample from the posh localities like Defense Colony, South Extension, Golf Link, etc. and ignore the localities where low income group and middle class families live.

This sampling method is seldom used and cannot be recommended for general use since it is often biased due to element of subjectiveness on the part of the investigator. However, if the investigator is experienced and skilled and this sampling is carefully applied, then judgment samples may yield valuable results.

2) Probability Sampling-

Probability sampling 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. The different types of probability sampling are

i) Where each unit has an equal chance being selected.

ii) Sampling units have different probabilities of being selected.

iii) Probability of selection of a unit is proportional to the sample size.

3) 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.

Simple Random Sampling

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.

In this method an equal probability of selection is assigned to each unit of the population at the first draw. It also implies an equal probability of selecting any unit from the available units at subsequent draws.

Thus in Simple Random Sampling from a population of N units, the probability of drawing any unit at the first draw is $\frac{1}{N}$, the probability of drawing any unit in the second draw from among the available (N-1) units is $\frac{1}{N-1}$ and so on.

Let E_r be the event that any specified unit is selected at the r^{th} draw. Then

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

$$= \prod_{i=1}^{r-1} \text{P[ith unit is selected]} \times \text{[It is selected at rth darw given that it is not selected at the previous(r} - 1\text{)draws]}$$

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

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

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

This leads to a very interesting and important property of Simple Random sampling Without Replacement **(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.

There are two types of Simple Random sampling

1) Simple Random Sampling with Replacement (SRSWR)

2) Simple Random Sampling without Replacement (SRSWOR)

1) Simple Random Sampling with Replacement (SRSWR)

In this method; first element is selected at random from the population. It is recorded or studied completely and the replaced back in the population. Afterwards second element is selected. This process is continued till a sample of required size is selected. In this method population size remains the same at every draw. This method of sampling is called as Simple Random Sampling with Replacement.

One of the serious drawbacks of this method is that, the same element may be selected more than once in the sample.

2) Simple Random Sampling without Replacement (SRSWOR)

There is another procedure of selecting elements in which, elements are selected at random but those are not replaced back in the population. This method of selecting sample is called as Simple Random Sampling without Replacement. In this method population size goes on decreasing at each draw. The drawback of getting the same element selected more than once is overcome in SRSWOR.

Illustrations of Simple Random Sampling

i) Suppose a lot of 500 articles is submitted for inspection to determine the proportion of defective articles one can use SRSWOR.

ii) In order to conduct a socio-economic survey of a certain village we can take SRSWOR and find per capita income of a village.

iii) In order to test average petrol consumption of a lot of scooter manufactured a SRSWOR or SRSWR can be used.

iv) Testing human blood by taking few drops out an individual's body is a SRSWOR.

v) In order to find average life of a bulb we take SRSWOR from a manufactured lot.

Notations and Terminology

Let us consider a (finite) population of N units and 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. Let Y_i (i= 1,2,3,……..N) be the value of the character for the i^{th} unit in the population and corresponding small letters y_i (i=1,2,3,…..n) denote the value of the character for the i^{th} unit selected in the sample.

	Population	Sample
1]	**N- Population size**	**n- Sample size**
2]	**Y_i - i^{th} unit in the population**	**y_i - i^{th} unit in the sample**
	Where i= 1,2,3,……..N	Where i= 1,2,3,……..n

3] **Population mean**

$$\overline{Y_N} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

Sample mean

$$\overline{y_n} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Population Total

$$\sum_{i=1}^{N} Y_i = N\overline{Y_N}$$

Sample Total

$$\sum_{i=1}^{n} y_i = N\overline{y_n}$$

Alternatively

$$Sample\ mean = \overline{y_n} = \frac{1}{n}\sum_{i=1}^{N} \alpha_i Y_i$$

α_i = 0 if i^{th} unit is not included in the sample.

= 1 if i^{th} unit is included in the sample.

4] **Population variance**

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \overline{Y_N})^2$$

sample variance

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

5] **population mean square**

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

sample mean square

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

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

Theorem -1- In Simple Random Sampling without Replacement (SRSWOR) the sample mean is an unbiased estimate of the population mean

i. e. $\boldsymbol{E}(\bar{y}_n) = \bar{Y}_N$

Proof- consider

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

Let us define

$$\alpha_i = \begin{cases} 1\ if\ the\ ith\ unit\ is\ included\ in\ the\ sample \\ 0\ if\ \ ith\ unit\ is\ not\ included\ in\ the\ sample \end{cases}$$

$$\overline{y_n} = \frac{1}{n}\sum_{i=1}^{N} \alpha_i Y_i$$

Taking expectation on the both sides

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

$$E(\alpha_i) = 1 \times \text{P[ith unit is included in the sample]} +$$
$$0 \times \text{P[ith unit is not included in the sample]}$$

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

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

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

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

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

Theorem-2 In SRSWOR sample mean square is unbiased estimate of population mean square.

i. e. $E(s^2) = S^2$

Proof - By the definition of sample mean square

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

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

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

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

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

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

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

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

Let us define

$$\alpha_i = \begin{cases} 1 \; if \; the\; ith\; unit\; is\; included\; in\; the\; sample \\ 0 \; if \;\; ith\; unit\; is\; not\; included\; in\; the\; sample \end{cases}$$

$$\alpha_j = \begin{cases} 1 \; if \; the\; jth\; unit\; is\; included\; in\; the\; sample \\ 0 \; if \;\; jth\; unit\; is\; not\; included\; in\; the\; sample \end{cases}$$

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

Taking expectation on both sides

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

$$E(\alpha_i) = 1 \times \text{P[ith unit is included in the sample]} + 0 \times \text{P[ith unit is not included in the sample]}$$

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

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

Consider

$$E(\alpha_i \alpha_j) = 1 \times \text{P}[\alpha_i \alpha_j = 1] + 0 \times \text{P}[\alpha_i \alpha_j = 0]$$

$$= \text{P}[\alpha_i = 1, \alpha_j = 1]$$

$$= \text{P}[\alpha_i = 1] \times \text{P}\left[\alpha_j = 1 \big/ \alpha_i = 1\right]$$

$$= \frac{n}{N} \times P[\textit{jth unit is included in the sample provided that ith unit is included in the sample}]$$

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

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

Substituting the values of $E(\alpha_i)$ and $E(\alpha_i\ \alpha_j)$we get

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

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

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

$$E(s^2) = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N} Y_i^2$$

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

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

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

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

⇒ Sample mean square is unbiased estimate of population mean square.

Theorem 3- In SRSWOR the variance of sample mean is

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

Proof:-

By the definition of variance we have

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

$$V(\bar{y}_n) = E(\bar{y}_n{}^2) - \bar{Y}_N^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1) \qquad \because E(\bar{y}_n) = \bar{Y}_N$$

Consider

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

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

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

Let us define

$$\alpha_i = \begin{cases} 1 \; if \; the \; ith \; unit \; is \; included \; in \; the \; sample \\ 0 \; if \;\; ith \; unit \; is \; not \; included \; in \; the \; sample \end{cases}$$

$$\alpha_j = \begin{cases} 1 \; if \; the \; jth \; unit \; is \; included \; in \; the \; sample \\ 0 \; if \;\; jth \; unit \; is \; not \; included \; in \; the \; sample \end{cases}$$

$$E(\alpha_i) = 1 \times \text{P[ith unit is included in the sample]} +$$
$$0 \times \text{P[ith unit is not included in the sample]}$$
$$= 1 \times \frac{n}{N} + 0 \times \left(1 - \frac{n}{N}\right)$$
$$= \frac{n}{N}$$

Consider

$$E(\alpha_i\alpha_j) = 1 \times \mathrm{P}[\alpha_i\alpha_j = 1] + 0 \times \mathrm{P}[\alpha_i\alpha_j = 0]$$
$$= \mathrm{P}[\alpha_i = 1, \alpha_j = 1]$$
$$= \mathrm{P}[\alpha_i = 1] \times \mathrm{P}\left[\alpha_j = 1 / \alpha_i = 1\right]$$
$$= \frac{n}{N} \times P[jth \; unit \; is \; included \; in \; the \; sample \; provided$$
$$that \; ith \; unit \; is \; included \; in \; the \; sample]$$

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

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

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

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

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

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

$$E(\bar{y}_n^2) = \frac{1}{Nn}\left[\sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{(N-1)}\left[\left(\sum_{i=1}^{N} Y_i\right)^2 - \sum_{i=1}^{N} Y_i^2\right]\right]$$

$$E(\bar{y}_n^2) = \frac{1}{Nn}\left[\sum_{i=1}^{N} Y_i^2 - \frac{(n-1)}{(N-1)} \sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2\right]$$

$$E\left(\bar{y}_n^{2\,2}\right) = \frac{1}{Nn}\left[\left(1 - \frac{n-1}{N-1}\right) \sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{(N-1)} N^2 \bar{Y}_N^2\right]$$

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

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

$$E(\bar{y}_n^2) = \frac{N-n}{Nn(N-1)} \sum_{i=1}^{N} Y_i^2 + \frac{N(n-1)}{n(N-1)} \bar{Y}_N^2 \text{ --------------- } (2)$$

Substituting equation (2) in equation (1) we get

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

$$V(\bar{y}_n) = \frac{N-n}{Nn(N-1)} \sum_{i=1}^{N} Y_i^2 + \frac{N(n-1)}{n(N-1)} \bar{Y}_N^2 - \bar{Y}_N^2$$

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

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

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

$$V(\bar{y}_n) = \frac{N-n}{Nn(N-1)}\sum_{i=1}^{N} Y_i^2 - \frac{N-n}{n(N-1)}\bar{Y}_N^2$$

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

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

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

Theorem-4 Variance of sample total is given by

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

Proof:- Consider

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

But

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

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

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

Theorem-5 In SRSWR, the sample mean is unbiased estimate of population mean

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

Proof:- Here method is SRSWR, we define random variable t_i as number of times ith unit is included in the sample. Here t_i takes 0,1,2,3,……n values and the probability that ith unit is selected at any draw in case of with replacement is $p = \frac{1}{N}$

Hence t_i be variable follows Binomial distribution in n draws with probability *p*.

$t_i \sim B(n,p)$ and $E(t_i) = np = \frac{n}{N}$ $V(ti) = npq = \frac{n}{N}\left(1 - \frac{1}{N}\right) = \frac{n(N-1)}{N^2}$

Now

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

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

Now for E(t_i) we have if $t_i \sim B(n,p)$ then E(t_i) = np $=\frac{n}{N}$

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

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

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

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

Theorem-6 In case of SRSWR show that

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

By the definition of variance we have

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

$$V(\bar{y}_n) = E(\bar{y}_n{}^2) - \overline{Y_N}^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1) \qquad \because E(\overline{y_n}) = \overline{Y_N}$$

Consider

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

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

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

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

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

Now for $E(t_i^2)$

$$V(t_i) = E(t_i^2) - [E(t_i)]^2$$

$$E(t_i^2) = V(t_i) + [E(t_i)]^2$$

$$E(t_i^2) = \frac{n(N-1)}{N^2} + \frac{n^2}{N^2}$$

$$E(t_i^2) = \frac{n}{N^2}[N + n - 1]$$

For $E(t_i t_j)$ we have

$$Cov(t_i t_j) = E(t_i t_j) - E(t_i)E(t_j)$$

$$E(t_i t_j) = Cov(t_i t_j) + E(t_i)E(t_j)$$

Here jointly $(t_i t_j)$ are distributed as multinomial distribution having co-variance

$$Cov(t_i t_j) = -np_i p_j = -n\frac{1}{N}\frac{1}{N} = -\frac{n}{N^2}$$

$$E(t_i t_j) = Cov(t_i t_j) + E(t_i)E(t_j)$$

$$E(t_i t_j) = -\frac{n}{N^2} + \frac{n}{N} \times \frac{n}{N}$$

$$E(t_i t_j) = -\frac{n}{N^2} + \frac{n^2}{N^2}$$

$$E(t_i t_j) = \frac{n(n-1)}{N^2}$$

Therefore equation (2) becomes

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

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

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

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

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

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

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

$$E(\bar{y}_n{}^2) = \frac{n}{n^2 N}\sum_{i=1}^{N} Y_i^2 + \frac{n(n-1)}{n^2}\bar{Y}_N{}^2$$

$$E(\bar{y}_n{}^2) = \frac{n}{n^2 N}\sum_{i=1}^{N} Y_i^2 + \frac{n(n-1)}{n^2}\bar{Y}_N{}^2$$

$$E(\bar{y}_n{}^2) = \frac{1}{nN}\sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{n}\bar{Y}_N{}^2$$

Substituting this value in equation (1) we get

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

$$V(\bar{y}_n) = \frac{1}{nN}\sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{n}\bar{Y}_N{}^2 - \overline{Y_N}^2$$

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

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

$$V(\bar{y}_n) = \frac{1}{nN}\sum_{i=1}^{N} Y_i^2 - \frac{1}{n}\bar{Y}_N{}^2$$

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

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

But we know that

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

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

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