UNIT 2 SIMPLE RANDOM SAMPLING

Structure

2.1 Introduction
Objectives
2.2 Methods of Selection of a Sample
Lottery Method
Random Number Method
Computer Random Number Generation Method
2.3 Properties of Simple Random Sampling
Merits and Demerits of Simple Random Sampling
2.4 Simple Random Sampling of Attributes
2.5 Sample Size for Specific Precision
2.6 Summary
2.7 Solutions/Answers

2.1 INTRODUCTION

Simple random sampling refers to the sampling technique in which each and every item of the population is having an equal chance of being included in the sample. The selection is thus free from any personal bias because the investigator does not make any preference in the choice of items. Since selection of items in the sample, depends entirely on chance, this method is also known as the method of probability sampling.

Random sampling is sometimes referred to a representative sampling. If the sample is chosen at random and the size of the sample is sufficiently large, it will represent all groups in the population. An element of randomness is necessary to be introduced in the final selection of the item. If that is not introduced, bias is likely to enter and make the sample unrepresentative.

Methods of selection of a simple random sample are explained in Section 2.2. In Section 2.3 the properties of simple random sampling are described. Simple random sampling of attributes is introduced in Section 2.4 whereas in Section 2.5 the sample size determination for specific precision is described briefly.

Objectives

After studying this unit, you would be able to

- describe the simple random sampling;
- explain the method of SRSWR and SRSWOR;
- explain and derive the properties of simple random sampling;
- calculate the variance of the estimate of the sample mean;
- describe the SRS for attribute and its properties; and
- describe and draw a appropriate sample size for specific precision.

2.2 METHODS OF SELECTION OF A SAMPLE

The random sample obtained by a method of selection in which every item has an equal chance to be selected in the sample. The random sample depends not only on method of selection but also on the sample size and nature of population.

Simple Random Sampling

If a sample of n units is selected randomly from a population of size N, this method is known as simple random sampling. As the name suggest, simple random sampling is a method in which the required number of elements /units are selected simply by random method from the target population. One can select a simple random sample by either of these two methods with replacement method and without replacement method.

Simple Random Sampling with Replacement (SRSWR)

When simple random samples are selected in the way that units which has been selected as sample unit is remixed or replaced in the population before the selection of the next unit in the sample then the method is known as simple random sampling with replacement.

Simple Random Sampling without Replacement (SRSWOR)

When simple random sample are selected in the way that a unit is selected as sample unit is not mixed or replaced in the population before the selection of the next unit. This method is known as simple random sampling without replacement i.e. once a unit is selected in the sample will never be selected again in the sample.

Some procedures are simple for small population and is not so for the large population. Proper care has to be taken to ensure that selected sample is random. Random sample can be obtained by any of the following methods:

1. By Lottery method;
2. By 'Mechanical Randomization' or 'Random Numbers' method; and
3. By Computer Random Number Generation method.

2.2.1 Lottery Method

This is very popular method of selecting a random sample under which all items of the population are numbered or named on separate slips of paper. These slips of paper should be of identical size, color and shape. These slips are then folded and mixed up in a container or box or drum. A blind fold selection is then made of the number of slips required to constitute the desired size of sample. The selection of items thus depends entirely on chance. For example if we want to select n candidates out of N. We assign the numbers 1 to N. One number to each candidate and write these numbers (1 to N) on N slips which are made as homogeneous as possible. These slips are then put in bag and thoroughly shuffled and then n slips are drawn one by one. Then the n candidates corresponding to numbers on the slip drawn will constitute a random sample.

If we draw a slip and note down the number written on the slip and then again replace the slip in the bag and then draw the next. These number of slips constitute a sample of required size is called a sample of SRSWR. If we do not replace the first slip which has already been drawn in the bag for the subsequent draws then it is called SRSWOR.

The above method is very popular in lottery draws where a decision about prizes is to be made. However, while adopting lottery method, it is absolutely essential to see that the slips are as homogeneous as possible in terms of size, shape and color, otherwise there is a lot of possibility of personal prejudice and bias affects in the result.

2.2.2 Random Number Method

The lottery method, discussed above, become quite cumbersome to use if the size of the population is very large. An alternative method of random selection is that of using the table of random numbers. The most practical and inexpensive method of selecting a random sample consists in the use of 'Random number tables' which have been so constructed that each of the digits 0, 1, 2, … , 9 appears with approximately the same frequency and independently.

If we have to select a sample from a population of size N ($\leq$ 99) then the numbers can be combined two by two to give pairs from 00 to 99. The method of drawing the random sample consists in the following steps:

1. Identify the N units in the population with the numbers from 1 to N;
2. Select at random, any page of the random number tables and pick up the numbers in any row or column or diagonal at random and discart the number which is greater than N; and
3. The population units corresponding to the numbers selected in step-2, constitute the random sample.

The sample may be selected with replacement or without replacement. In the case of sampling with replacement, a number occurring more than once is accepted. A unit is repeated as many times as a random number occurs. But in the case of sampling without replacement, if a number in random number table or remainder occurs more than once is omitted at any sequent stage. In the above selection procedure numbering of units from 00 onwards and making use of remainders has an advantage, as no random numbers is being wasted during the selection procedure. This saves time and labor. For example a population consists of 20 units and a sample of size 5 is to be selected from this population. Since 20 is a two digit figure, unit are numbered as 00, 01… 19. Five random numbers are obtained from a two digit random number table. They are given as 85, 63, 52, 34, and 46. On dividing 85 by 20 the remainder is 5, hence select the unit on serial no. 5. Similarly, dividing 63, 52, 34 and 46 by 20, the respectively remainder are 3, 12, 14 and 6. Hence selected units are at serial numbers 05, 03, 12, 14 and 06. These selected units constitute the sample.

A number of random number tables are available such as:

1. **Tippet's Random Number Tables**
   These tables consist of 10,400 four digits numbers, giving in all 10,400 $\times$ 4 i.e. 41,600 digits.

2. **Fisher and Yate's Random Number Tables**
   These tables consist of 15000 digits arranged in two digits numbers.

3. **Kendall and Babington Smith's Random Tables**
   These tables consist of 1,00,000 digits grouped into 25,000 sets of 4 digits random numbers.

4. **Rand Corporation Random Number Tables**
   These tables consist of one million random digits consisting of 2,00,000 random numbers of 5 digits each.

2.2.3 Computer Random Number Generation Method

The main disadvantage of random number method is that if we want to draw a sample of large size from the target population then it will take a long time to draw random numbers from the random number table. Thus to save time and energy in generation of large numbers of a random numbers one can opt computer random numbers generation method. Many kinds of methods have been used for generation of random number from computer but we shall not discuss all of them here and describe only linear congruential generation method.

Linear Congruential Generation Method

Linear congruential method can take many different forms but the most commonly used form is defined by

$z_i = (az_{i-1} + c) \bmod m$ for i =1, 2, …

where, a, c and z_i are to be in the range (0, 1, 2,…, m - 1) and integers

a - Multiplier Integer

c - Shift or Increment Integer

m - Modulus

Here, (n) mod m means remainder term when n is divided by m.

Suppose we want to draw a sample of size 4 from the population of size 8.

Take m = 8, a = 5, c =7 and $z_0 = 4$ the resulting sequence of random numbers is calculated as

$z_1 = (5 \times 4 + 7) \bmod 8 = 3$

$z_2 = (5 \times 3 + 7) \bmod 8 = 6$

$z_3 = (5 \times 6 + 7) \bmod 8 = 5$

$z_4 = (5 \times 5 + 7) \bmod 8 = 0$

For above calculation we may also make a computer program which is beyond this course. We are, therefore, not going to discuss it here.

2.3 PROPERTIES OF SIMPLE RANDOM SAMPLING

Terminologies

N = Population size

n = Sample size

X_i = Value of the character under study for the i^{th} unit in the population

x_i = Value of the character under study for the i^{th} unit in the sample

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \text{Population mean}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \text{Sample mean}$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 = \text{Population mean square}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \text{Sample mean square}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 = \text{Population variance}$$

Theorem 1: Prove that 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: In simple random sampling method an equal probability of selection is assigned to each unit of the population at the first draw.

Thus, in SRS 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 $(N-1)$ is $1/(N-1)$ and so on.

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

P (E_r) = Prob.{A specific unit is not selected at any one of previous (r - 1) draws and then selected at the r^{th} draw}

$$= \sum_{i=1}^{r-1} P\text{ (It is not selected at } i^{th}\text{ draw)}$$

$\times$ P (It is selected at r^{th} draw that is not selected at the previous (r - 1) draw)

$$P(E_r) = \sum_{i=1}^{r-1}\left[1 - \frac{1}{N-(i-1)}\right] \times \frac{1}{N-(r-1)}$$

$$P(E_r) = \sum_{i=1}^{r-1}\left[\frac{N-i}{N-(i-1)}\right] \times \frac{1}{N-(r-1)}$$

$$P(E_r) = \frac{(N-1)}{N} \times \frac{(N-2)}{(N-1)} \times \frac{(N-3)}{(N-2)} \times \ldots \times \frac{(N-r+1)}{(N-r+2)} \times \frac{1}{(N-r+1)}$$

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

That means

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

Theorem 2: The probability that a specified unit is selected in the sample of size n is $\frac{n}{N}$

Proof: Since a specified unit can be selected in the sample of size n in n mutually exclusive ways, viz. it can be selected in the sample at the r^{th} draw (r =1, 2,…, n) and since the probability that it is selected at r^{th} draw is

$$P(E_r) = \frac{1}{N} \qquad ; r = 1, 2, 3, ..., n$$

Therefore, the probability that a specified unit is included in the sample would be the sum of the probabilities of inclusion in the sample at 1^{st} draw, 2^{nd} draw, …, n^{th} draw. Thus, by addition theorem of probability, we get

$$P\left(\bigcup_{r=1}^{n} E_r\right) = \sum_{r=1}^{n} \frac{1}{N} = \frac{n}{N}$$

Theorem 3: The possible numbers of sample of size n from a population of size N if sampling is done with replacement is N^n.

Proof: The first unit can be drawn from N units in N ways. Similarly, second unit can also be drawn in N ways because the first selected unit again mixed with the population. So on up to the selection of n^{th} unit.

Thus, the total number of ways are

$${}^{N}C_1 . {}^{N}C_1 . {}^{N}C_1 ... {}^{N}C_1 \text{ (n times)}$$

$$\Rightarrow \left({}^{N}C_1\right)^n = N^n$$

Theorem 4: In SRSWOR the sample mean $\bar{x}$ is an unbiased estimator of population mean $\bar{X}$.

Proof: We have

$$E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = E\left[\frac{1}{n}\sum_{i=1}^{N} a_i X_i\right]$$

where, a_i = 1 if i^{th} unit is included in the sample
0 if i^{th} unit is not included in the sample

Since, a_i takes only two values 1 and 0

$$E(a_i) = 1.P(a_i = 1) + 0.P(a_i = 0)$$

= 1.P (i^{th} unit is included in a sample of size n) + 0.P (i^{th} unit is not included in the sample)

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

Hence,

$$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{N} \frac{n}{N} X_i = \frac{1}{N}\sum_{i=1}^{N} X_i = \bar{X}$$

Theorem 5: In SRSWR, the sample mean $\bar{x}$ is an unbiased estimator of population mean $\bar{X}$.

Proof: We have

$$E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]$$

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

$$= \frac{1}{n}\sum_{i=1}^{n} \bar{X}$$

$$= \frac{1}{n} n.\bar{X} = \bar{X}$$

Theorem 6: In SRSWOR, the sample mean square is an unbiased estimate of the population mean square, i.e.

$$E\left(s^2\right) = S^2$$

Proof: We have

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

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i\neq j=1}^{n} x_i\, x_j\right]$$

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

$$= \frac{1}{n-1} \times \frac{(n-1)}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} x_i x_j$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} x_i\, x_j$$

$$E\,(s)^2 = \frac{1}{n}E\left[\sum_{i=1}^{n} x_i^2\right] - \frac{1}{n(n-1)}E\left[\sum_{i\neq j=1}^{n} x_i\, x_j\right] \qquad \ldots (1)$$

We have

$$E\left[\sum_{i=1}^{n} x_1^2\right] = E\left[\sum_{i=1}^{N} a_i\, X_i^{\,2}\right]$$

$$= \sum_{i=1}^{N} E\left(a_i\right) X_i^{\,2} \qquad \ldots (2)$$

where, a_i = 1 if i^{th} unit is included in the sample
0 if i^{th} unit is not included in the sample … (3)

Therefore,

$$E\left[\sum_{i=1}^{n} x_i^2\right] = \frac{n}{N}\sum_{i=1}^{N} X_i^2 \qquad \text{… (4)}$$

and

$$E\left[\sum_{i\neq j=1}^{n} x_i\, x_j\right] = E\left[\sum_{i\neq j=1}^{N} a_i a_j X_i\, X_j\right]$$

$$= \left[\sum_{i\neq j=1}^{N} E\left(a_i a_j\right) X_i\, X_j\right] \qquad \text{… (5)}$$

where, a_i and a_j are defined in equation (3)

Therefore,

$$E\left(a_i\, a_j\right) = 1.P\left(a_i\, a_j = 1\right) + 0.P\left(a_i\, a_j = 0\right)$$

$$\Rightarrow P\left[\left(a_i = 1\right)\cap\left(a_j = 1\right)\right]$$

$$\Rightarrow P\left(a_i = 1\right).P\left(a_{j=1} \big/ a_i = 1\right)$$

$$\Rightarrow \frac{n(n-1)}{N(N-1)} \qquad \text{… (6)}$$

Because

$$E\left(a_i = 1\right) = P\left[i^{th}\text{ unit is included in the sample}\right] = \frac{n}{N}$$

and

$$P\left(a_j = 1 \big/ a_i = 1\right) = P\left[\begin{array}{l}j^{th}\text{ unit is included in the sample given}\\ \text{that } i^{th}\text{ unit is included the sample}\end{array}\right]$$

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

Substituting in equation (5), we get

$$E\left[\sum_{i\neq j=1}^{n} x_i\, x_j\right] = \sum_{i\neq j=1}^{N} \frac{n(n-1)}{N(N-1)} X_i\, X_j \qquad \text{… (7)}$$

Substituting from equations (4) and (7) in equation (1), we get

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

$$= \frac{1}{N-1}\left[\frac{N-1}{N}\sum_{i=1}^{N} X_i^2 - \frac{1}{N}\sum_{i\neq j=1}^{N} X_i X_j\right]$$

$$= \frac{1}{N-1}\left[\left(1-\frac{1}{N}\right)\sum_{i=1}^{N} X_i^{\,2} - \frac{1}{N}\sum_{i\neq j=1}^{N} X_i\, X_j\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} X_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} X_i^2 + \sum_{i\neq j=1}^{N} X_i X_j\right)\right]$$

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

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right] = S^2$$

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

Theorem 7: In SRSWOR, the variance of the sample mean is given by

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

Proof: We have

$$\mathrm{Var}(\bar{x}) = E\left[\bar{x} - E(\bar{x})\right]^2$$

$$= E(\bar{x})^2 - \bar{X}^2 \qquad \ldots (8)$$

Now

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

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

$$= \frac{1}{n^2}\left[E\left(\sum_{i=1}^{n} x_i^2\right) + E\left(\sum_{i=1}^{n} x_j\right)\right] \qquad \ldots (9)$$

From equation (4), we have

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

But $\displaystyle\sum_{i=1}^{N}(X_i - \bar{X})^2 = \sum_{i=1}^{N} X_i^2 - N\bar{X}^2$

$$\Rightarrow \sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N}(X_i - \bar{X})^2 + N\bar{X}^2$$

$$\Rightarrow \sum_{i=1}^{N} X_i^2 = (N-1)S^2 + N\bar{X}^2$$

Therefore,

$$E\left(\sum_{i=1}^{n} x_i^2\right) = n\left[\left(\frac{N-1}{N}\right) S^2 + \bar{X}^2\right] \qquad \ldots (10)$$

Also from equation (7), we have

$$E\left(\sum_{i \neq j=1}^{n} x_i x_j\right) = \frac{n(n-1)}{N(N-1)}\sum_{i \neq j=1}^{N} X_i X_j$$

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

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

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

$$= n(n-1)\left[\bar{X}^2 - \frac{S^2}{N}\right] \qquad \ldots (11)$$

Substituting from equations (10) and (11) in equation (9), we get

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

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

$$= \bar{X}^2 + \left(\frac{1}{n}-\frac{1}{N}\right)S^2 \qquad \ldots (12)$$

Substituting from equation (12) in equation (8), we get

$$\mathrm{Var}\left(\bar{x}\right) = \bar{X}^2 + \left(\frac{1}{n}-\frac{1}{N}\right)S^2 - \bar{X}^2$$

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

Theorem 8: In SRSWR, variance of sample mean is given by

$$\mathrm{Var}\left(\bar{x}\right) = \frac{(N-1)}{nN}S^2$$

Proof: We have

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Again,

$$\mathrm{Var}\left(\bar{x}\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(x_i)$$

Since in case of SRSWR each observation is independent, therefore

$$\mathrm{Var}\left(\bar{x}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2$$

$$= \frac{1}{n^2}n.\sigma^2 = \frac{1}{n}\sigma^2 \qquad \ldots (13)$$

But

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

$$\sigma^2 = \frac{N-1}{N}S^2 \qquad \ldots (14)$$

Substituting from equation (14) in equation (13) we get

$$\mathrm{Var}\left(\bar{x}\right) = \frac{(N-1)}{nN}S^2$$

Theorem 9: The variance of the sample mean is more in SRSWR in comparison to its variance in SRSWOR, i .e.

$$\mathrm{Var}_{SRSWR}\left(\bar{x}\right) > \mathrm{Var}_{SRSWOR}\left(\bar{x}\right)$$

Proof: We have

$$\mathrm{Var}_{SRSWR}\left(\bar{x}\right) = \frac{N-1}{nN}S^2$$

and $$\mathrm{Var}_{SRSWOR}(\bar{x}) = \frac{N-n}{nN}S^2$$

Therefore,

$$\mathrm{Var}_{SRSWR}(\bar{x}) - \mathrm{Var}_{SRSWOR}(\bar{x}) = \frac{(N-1)}{nN}S^2 - \frac{(N-n)}{nN}S^2$$

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

$$= \frac{1}{N\,n}S^2\left[n-1\right]$$

$$= \left(\frac{n-1}{n\,N}\right)S^2 > 0$$

That implies $\mathrm{Var}_{SRSWR}(\bar{x}) > \mathrm{Var}_{SRSWOR}(\bar{x})$

That means variance of the sample mean is more in SRSWR as compared with its variance in the case of SRSWOR. In other words SRSWOR provides a more efficient estimate of sample mean relative to SRSWR.

2.3.1 Merits and Demerits of Simple Random Sampling

Merits

Simple random sampling has the following merits:

1. In simple random sampling each unit of the population has equal chance to be included in the sample; and
2. Efficiency of the estimates can be found out in simple random sampling because all the estimates are calculated by using the probability theory.

Demerits

Despite merits, simple random sampling has some demerits too viz.

1. An up-to-date frame of population is required in simple random sampling;
2. Some administrative inconvenience arises in simple random sampling if some of the units are spreaded in a wide area. So collecting information from these related units may be problem; and
3. SRS required larger sample size than any other sampling for a fix level of precision.

Example 1: A population have 7 units 1, 2, 3, 4, 5, 6, 7. Write down all possible samples of size 2 (without replacement) which can be drawn from the given population and verify that sample mean is an unbiased estimate of the population mean. Also calculate its sample variance and verify that

$$\mathrm{Var}_{SRSWR}(\bar{x}) > \mathrm{Var}_{SRSWOR}(\bar{x})$$

Solution: We have

$$X = 1, 2, 3, 4, 5, 6, 7$$

$$\bar{X} = \frac{1+2+3+4+5+6+7}{7} = 4$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$$

$$=\frac{1}{6}\times(9+4+1+0+1+4+9)$$

$$=\frac{28}{6}=4.666$$

All possible samples of size 2 are as follows:

Sample No.	Sample values	Sample Mean $(\bar{x})$	$\bar{x}-\bar{X}$	$(\bar{x}-\bar{X})^2$
1	(1,2)	1.5	-2.5	6.25
2	(1,3)	2.0	-2.0	4.00
3	(1,4)	2.5	-1.5	2.25
4	(1,5)	3.0	-1.0	1.00
5	(1,6)	3.5	-0.5	0.25
6	(1,7)	4.0	0	0
7	(2,3)	2.5	-1.5	2.25
8	(2,4)	3.0	-1.0	1.00
9	(2,5)	3.5	-0.5	0.25
10	(2,6)	4.0	0	0
11	(2,7)	4.5	+0.5	0.25
12	(3,4)	3.5	-0.5	0.25
13	(3,5)	4.0	0	0
14	(3,6)	4.5	+0.5	0.25
15	(3,7)	5.0	+1.0	1.00
16	(4,5)	4.5	+0.5	0.25
17	(4,6)	5.0	+1.0	1.00
18	(4,7)	5.5	+1.5	2.25
19	(5,6)	5.5	+1.5	2.25
20	(5,7)	6.0	+2.0	4.00
21	(6,7)	6.5	+2.5	6.25
Total		84.0		35.00

From the table, we have

$$\sum \bar{x}_i = 84.0 \text{ and } \sum_{i=1}^{k} (\bar{x}_i-\bar{X})^2 = 35.00$$

$$E(\bar{x}) = \frac{\sum_{i=1}^{^N C_n} \bar{x}_i}{^N C_n} = \frac{84}{21} = 4 = \bar{X}$$

$$Var(\bar{x}) = \frac{1}{^N C_n}\sum_{i=1}^{^N C_n} (\bar{x}_i-\bar{X})^2 = \frac{1}{21}\times 35.00 = 1.667$$

$$\sigma^2 = \frac{N-1}{N}S^2 = \frac{6}{7}\times 4.667 = 4.0008$$

Verification: In SRSWOR the variance of sample mean is given by

$$\mathrm{Var}(\bar{x}) = \frac{N-n}{Nn}S^2 = \frac{7-2}{7\times 2}4.667 = 1.667$$

In SRSWR the variance of sample mean is given by

$$\mathrm{Var}(\bar{x}) = \frac{\sigma^2}{n} = \frac{4.0008}{2} = 2.0004$$

Hence, $\mathrm{Var}(\bar{x})_{SRSWR} > \mathrm{Var}(\bar{x})_{SRSWOR}$

E1) Draw all possible samples of size 2 from the population {2, 3, 4} and verify that $E(\bar{x}) = \bar{X}$ also find variance.

E2) How many random samples of size 5 can be drawn from a population of size 10 if sampling is done with replacement?

E3) From a population of 50 units, a random sample of size 10 is drawn without replacement. From the sample following result are obtained.

$$\sum_{i=1}^{n} x_i = 48\ ,\ \sum_{i=1}^{n}(x_i - \bar{x})^2 = 36$$

Calculate the sample mean and its variance.

E4) Draw all possible samples of size 2 from the population {8, 12, 16} and verify that

$$E(\bar{x}) = \bar{X}$$

and find variance of estimate of the population mean.

E5) From a population of size N=100, a random sample of size 10 is drawn without replacement. From the sample following results are obtained.

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 45$$

Calculate the variance of sample mean.

2.4 SIMPLE RANDOM SAMPLING OF ATTRIBUTES

A qualitative characteristic which cannot be measured numerically is known as an attribute i.e. honesty, intelligence, beauty, etc. In many situations, it is not possible to measure the characteristic under study but possible to classify the population into various classes according to the attributes under study. For example, we can divide a population of a colony into two classes only say literate and illiterate with respect to attribute literacy. Hence the units in the population can be distributed in these two classes accordingly as it possesses or does not possess the given attribute. After taking a sample of size n, we may be interested in estimating the total number of proportion of the defined attribute.

Notations and Terminologies

Let us suppose that a population having N units $X_1, X_2, \ldots, X_N$ is classified into k mutually disjoint and exhaustive classes. Then

π = The proportion of units processing the given attribute in population A/N

π' = The proportion of units not processing the given attribute in population $A'/N = 1 - \pi$

Let us consider SRSWOR sample of size n. From this population if 'a' is the number of units in a sample possessing the given attribute then

p = proportion of sampled units possessing the given attribute = a / n

q = proportion of sampled units not possessing the given attribute = a'/n

Let X_i be the i^{th} unit of the population, where i = 1, 2, …, N.

Then, $X_i = 1$ if i^{th} unit possesses the given attribute

= 0 if it does not possess the given attribute

Similarly, x_i denote i^{th} unit in the sample

Then, $x_i = 1$, if i^{th} sampled unit possesses the given attribute

$x_i = 0$, if i^{th} sampled unit does not possess the given attribute

The $\sum_{i=1}^{N} X_i$ = A, the number of units in the population possessing the given attribute.

and $\sum_{i=1}^{n} x_i$ = a, the number of sampled units possessing the given attributes.

Thus,

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{A}{N} = \pi$$

and

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{a}{n} = p$$

Similarly,

$$\sum_{i=1}^{N} X_i^2 = A = N\pi$$

and

$$\sum_{i=1}^{n} x_i^2 = a = np$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} X_i^2 - N\overline{X}^2\right]$$

$$= \frac{1}{N-1}\left[N\pi - N\pi^2\right] = \frac{N\pi(1-\pi)}{N-1}$$

Similarly,

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

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

Theorem 10: Sample proportion p is an unbiased estimate of population proportion π, i.e.

$$E(p) = \pi$$

Proof: We have

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{a}{n} = p$$

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{A}{N} = \pi$$

We know that in simple random sampling the sample mean provides on unbiased estimate of the population mean

$$E(\bar{x}) = \bar{X}$$

Therefore $E(p) = \pi$

Theorem 11: In SRSWOR, show that Var (p) = $\frac{N-n}{N-1}.\frac{\pi(1-\pi)}{n}$

Proof: We have, $\text{Var}(p) = \text{Var}(\bar{x})$

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

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

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

2.5 SAMPLE SIZE FOR SPECIFIC PRECISION

A very first problem faced by a statistician in any sample survey is to determine the sample size so that the population parameters may be estimated with a specified precision. The degree of precision can be determined in terms of

1. The level of significance in the estimate; and
2. The confidence interval within which this estimate lie with respect to given level of significance.

Let us consider the parameter $\bar{X}$ the population mean of the population of size N. We know that $\bar{x}$ sample mean based on n units is unbiased estimate of $\bar{X}$. Let the difference between estimate value $\bar{x}$ and the population mean $\bar{X}$ is d and level of confidence is $(1-\alpha)$, then the sample size is determined by the equation.

$$P\left[\left|\bar{x}-\bar{X}\right|<d\right]=1-\alpha \qquad \ldots (15)$$

or
$$P\left[\left|\bar{x}-\bar{X}\right|\geq d\right]=\alpha \qquad \ldots (16)$$

where, α is very small preassinged probability and is nown as the level of significance.·

If n is sufficiently large and we consider SRSWOR, then the statistic

$$Z = \frac{\bar{x}-E(\bar{x})}{SE(\bar{x})} = \frac{\bar{x}-\bar{X}}{\sqrt{Var(\bar{x})}} = \frac{\bar{x}-\bar{X}}{S\sqrt{\frac{1}{n}-\frac{1}{N}}} \qquad \ldots (17)$$

where, Z is a standard normal variate.

Accordingly, if we take $\alpha = 0.05$, then we have

$$P\left[\left|Z\right|\geq 1.96\right]=0.05$$

$$P\left[\left|\frac{\bar{x}-\bar{X}}{S\sqrt{\frac{1}{n}-\frac{1}{N}}}\right|\geq 1.96\right]=0.05$$

$$P\left[\left|\bar{x}-\bar{X}\right|\geq 1.96\times S\sqrt{\frac{1}{n}-\frac{1}{N}}\right]=0.05$$

Comparing with equation (16), we get

$$d=1.96\times S\sqrt{\frac{1}{n}-\frac{1}{N}}$$

$$\frac{d^2}{S^2(1.96)^2}=\frac{1}{n}-\frac{1}{N}$$

$$n=\frac{NS^2(1.96)^2}{Nd^2+S^2(1.96)^2}=\frac{3.84NS^2}{3.84S^2+Nd^2} \qquad \ldots (18)$$

This formula gives the sample size in SRSWOR for estimating population mean with confidence level 95 % and margin of error d, provided n is large.

Similarly, if n is small the statistics z follows the student's t distribution with (n-1) degree of freedom is given by

$$t = \frac{\bar{x}-\bar{X}}{S\sqrt{\left(\frac{1}{n}-\frac{1}{N}\right)}}$$

If t_α is the critical value of t for (n−1) df and at α level of significance then n is given by the equation

$$P\left[\left|\bar{x}-\bar{X}\right| \geq S\sqrt{\frac{1}{n}-\frac{1}{N}}.\, t_\alpha\right] = \alpha \qquad \ldots (19)$$

Comparing with equation (16) we get

$$d = S\sqrt{\left(\frac{1}{n}-\frac{1}{N}\right)}.\, t_\alpha$$

$$\frac{d^2}{S^2\, t_\alpha{}^2} = \frac{1}{n}-\frac{1}{N}$$

$$n = \frac{NS^2\, t_\alpha^2}{Nd^2 + S^2 t_\alpha^2} = \frac{S^2 t_\alpha^2}{d^2 + \left(S^2\, t_\alpha^2 / N\right)}$$

2.6 SUMMARY

In this unit, we have discussed:

1. The simple random sampling;
2. The method of SRSWR and SRSWOR;
3. The properties of simple random sampling;
4. Method of finding the variance of the estimate of the sample mean;
5. The simple random sampling for attribute and its properties; and
6. The sample size determination for specific precision.

2.7 SOLUTIONS / ANSWERS

E1) In SRSWOR the number of sample is

$$^{N}C_n = {}^{3}C_2 = 3$$

The samples with their means are as follows:

Sr. No	**Sample**	**Mean $(\bar{x})$**
1	2,3	2.5
2	2,4	3
3	3,4	3.5
		$\sum_{i=1}^{3}\bar{x}_i = 9$

$$E(\bar{x}) = \frac{1}{{}^{N}C_{n}} \sum_{i=1}^{{}^{N}C_{n}} \bar{x}_i$$

$$= \frac{9}{3} = 3$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$= \frac{1}{3}(2+3+4)$$

$$= \frac{9}{3} = 3$$

Therefore,

$$E(\bar{x}) = \bar{X}$$

Again $$V(\bar{x}) = \frac{N-n}{N.n} S^2$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

$$= \frac{1}{2} \left\{(2-3)^2 + (3-3)^2 + (4-3)^2\right\}$$

$$=1$$

Therefore,

$$\mathrm{Var}(\bar{x}) = \frac{3-2}{3 \times 2} \times 1$$

$$= \frac{1}{6} = 0.166$$

E2) The first unit can be drawn from 10 units in ${}^{10}C_1 = 10$ ways. Since sampling is done with replacement so the second unit can be drawn in ${}^{10}C_1 = 10$ ways … so on upto the selection of 5^{th} Unit. Thus the total ways are 10.10.10.10.10 = 10^5 ways.

E3) We have

$$\sum_{i=1}^{n} x_i = 48$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{10} \times 48 = 4.8$$

So

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

$$s^2 = \frac{1}{9} \times 36 = 4$$

which is the estimate value of S^2.

Therefore,

$$\text{Variance}(\bar{x}) = \frac{N-n}{Nn}S^2 = \frac{50-10}{50\times 10}\times 4$$
$$= \frac{16}{50} = 0.32$$

E4) In SRSWOR the number of samples is ${}^{N}C_{n} = {}^{3}C_{2} = 3$ and samples with their means are

S. No.	Sample	Mean $(\bar{x}_i)$
1	(8, 12)	10
2	(8, 16)	12
3	(12, 16)	14
Total		36

$$\sum_{i=1}^{3}\bar{x}_i = 36\ ,$$

Therefore,

$$E(\bar{x}) = \frac{1}{{}^{N}C_{n}}\sum_{i=1}^{{}^{N}C_{n}}\bar{x}_i$$
$$= \frac{1}{3}\left(\sum_{i=1}^{3}\bar{x}_i\right)$$
$$= \frac{1}{3}\times 36 = 12$$

Again $\quad \bar{X} = \frac{8+12+16}{3} = 12$

Therefore,

$$E(\bar{x}) = \bar{X}$$

Again estimator of population mean is sample mean and so its variance

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

where,

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i-\bar{X}\right)^2$$
$$= \frac{1}{3-1}\left[(8-12)^2+(12-12)^2+(16-12)^2\right]$$
$$= \frac{1}{2}\left[16+0+16\right] = 16$$

Therefore,

$$\text{Var}(\bar{x}) = \left(\frac{1}{2}-\frac{1}{3}\right)\times 16$$

$$= \left(\frac{3-2}{6}\right)16 = \frac{16}{6}$$

$$= \frac{8}{3} = 2.66$$

E5) We have,

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

and

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

$$= \frac{1}{9} \times 45 = 5$$

Which is the estimate value of S^2.

Therefore,

$$\text{Var}(\bar{x}) = \left(\frac{1}{10} - \frac{1}{100}\right) \times 5$$

$$= \frac{9}{100} \times 5 = 0.45$$