UNIT 2

SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT

Structure

2.1 INTRODUCTION

The previous unit provided a detailed study of the most fundamental and simplest type of sampling scheme, namely, 'Simple Random Sampling' (SRS). Section 1.6 in the Unit 1 of the present block provides an exclusive definition of Simple Random Sampling. Further, in the Sub-section 1.6.2, it was described that Simple Random Sampling scheme might be either 'With Replacement' or 'Without Replacement' of units selected at different draws, in the population before subsequent draws. Accordingly, we have two Simple Random Sampling schemes, first, 'Simple Random Sampling with Replacement' (SRSWR) and second, 'Simple Random Sampling Without Replacement' (SRSWOR). However, due to limited space in the Unit 1, we mainly confined ourselves to the study of many of the concepts, theories and salient properties related to Simple Random Sampling with Replacement scheme.

We have seen that Simple Random Sampling with Replacement scheme is the sampling scheme in which each and every unit of the population has equal chance of being selected in the sample at each draw. This equal probability selection of each unit is possible because a unit selected at any draw is replaced in the population again before the next draw. Accordingly, such a scheme is called 'Simple Random Sampling with Replacement' (SRSWR).

Another method of selection of different units in the sample may be followed. If in the selection of a simple random sample is made without replacing the selected units in the population after subsequent draws, it is termed as 'Simple Random Sampling without Replacement' (SRSWOR). Thus, the basic difference between Simple Random Sampling with Replacement and Simple Random Sampling without Replacement schemes is only due to the replacement or non-replacement of selected units before subsequent draws. However, this difference produces such results which make Simple Random Sampling without Replacement better than Simple Random Sampling with Replacement as far as the performance of the estimators of the parameters is concerned.

Since, use of different techniques involved in selecting a random sample from the given population, schemes Simple Random Sampling with Replacement and Simple Random Sampling without Replacement require separate studies as far as changes in a number of concepts and derived results are concerned. However, many of the basic fundamentals, which are discussed in the context of Simple Random Sampling with Replacement, are also applicable in the case of Simple Random Sampling without Replacement as well. This facilitates us not to discuss and describe those concepts and results again in this unit.

In this unit, Section 2.2 discusses about the methods of selection of a Simple Random Sampling without Replacement sample. It provides (i) Lottery Method (or, Chit Method) and (ii) use of Random Number Tables. Section 2.3 is devoted to describing and providing proof of some fundamental properties of Simple Random Sampling without Replacement which distinguish it from Simple Random Sampling with Replacement scheme. Section 2.4 discusses the problem of estimating some of the population parameters of importance using a Simple Random Sampling without Replacement sample, such as, population mean, population total, and population variance. Derivations of the results are presented in the form of theorems. Section 2.5 presents the derivation of the variance of the sample mean and estimation of the Variance of sample mean. Section 2.6 presents an extensive comparison of Simple Random Sampling without Replacement scheme with Simple Random Sampling with Replacement scheme and establishes the advantages of Simple Random Sampling without Replacement scheme over Simple Random Sampling with Replacement scheme.

Expected Learning Outcomes

After studying this unit, you should be able to:

- define "Simple Random Sampling without Replacement (SRSWOR);
- describe the methods of drawing a Simple Random Sampling without Replacement sample from a given population;
- discuss the fundamental properties of Simple Random Sampling without Replacement and differentiate Simple Random Sampling without Replacement from Simple Random Sampling with Replacement;
- determine the estimators of population mean, population total, and population variance along with their properties;
- derive the expressions of variances of the sample mean estimator; and
- obtain the estimator of the variance of the sample mean estimator.

2.2 METHODS OF SELECTING A SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT SAMPLE

In the previous unit, we have already mentioned how a 'With Replacement Random Sample', in general and Simple Random Sampling with Replacement Sample, in particular can be selected from a given population of finite size. You know that in order to select a random sample from any population, any chance mechanism is used which assures that each unit of the population has some pre-assigned non-zero probability of selection, say p, such that the conditions:

$$0 < p_i < 1 \quad \text{and} \quad \sum_{i=1}^{N} p_i = 1$$

are satisfied, where p_i stands for the selection probability assigned to the i^{th} unit of the population. For different types of situations and populations, different types of chance mechanisms might be used.

As far as the SRS scheme is concerned the definition of it, as given under sub-section 1.6.1 of the previous unit, states that the selection probabilities which are assigned to the units belonging to the population must be equal for each unit. Thus, for a population of size N, the selection probability p_i must be 1/N for i = 1, 2, 3, …, N. We have seen that two chance mechanisms, namely, (i) Lottery (Chit) method and (ii) Random Number Table method fulfil the condition of equal probability selection for each and every unit of the population. Accordingly, in the Sub-sections 1.6.2 and 1.6.3 in Unit 1, we described how these methods can be used in order to select a Simple Random Sampling with Replacement sample.

It is clear from the definition and the Theorem 1, stated and derived in the previous unit, that Simple Random Sampling with Replacement scheme; being simple random sampling scheme is an Equal Probability Selection Method (EPSEM), implying that the probability of selection of units of the population in the sample is same for each unit belonging to the population, which is, 1/N, i.e., inverse of the population size.

However, till now, for SRSWOR scheme, neither we have seen any indication from anywhere nor we have proved any result showing that the selection probabilities of all the units of the population in the sample at any draw is 1/N, similar to the case of Simple Random Sampling with Replacement scheme. Unless we prove it, SRSWOR scheme cannot be claimed to be EPSEM, like SRSWR scheme. However, afterwards in a separate section, we shall show that under Simple Random Sampling without Replacement scheme also, each of the population units have equal probability of selection in the sample at any draw and it is equal to 1/N.

We shall now describe the procedure of selecting a **Simple Random Sampling without Replacement sample** using these methods.

2.2.1 Lottery Method

As mentioned earlier, this method of selecting a random sample can be used particularly, for small and moderate size populations. The method would be quite cumbersome in case the population size is large enough, since, preparation of chits of paper, identical in all respects and writing labels on these chits would be very much time consuming and costly too. The method consists of the following steps:

(i) Let the population size be N. Then, obviously labels of units will be first N integers, so we have sampling frame as {1, 2, 3…, i,…, N}.

(ii) Prepare N chits of papers of same colour, thickness, material and size; so that all the chits must be identical in all respects and no chit could be distinguished from other chits.

(iii) Write numbers 1 to N on these chits such that each chit has a single number.

(iv) Fold these chits in same manner in such a way that numbers written on them are not visible without unfolding them.

(v) Put all the folded chits in a box or vessel and mix them properly.

(vi) Let the sample size be n. Then, we have to select n chits from the box or vessel one by one in a number of draws in order to constitute the random sample of size n.

(vii) Select one chit from the vessel, open it and note down the number it consists of. Say, it is k. Then k^{th} unit, U_k $(1 \le k \le N)$ of the population is assumed to be selected in the sample in the first draw. Since, there are N identical chits in the vessel and one chit is randomly drawn out of it, the chance mechanism used assures that probability of selection of the first unit would be equal to 1/N.

(viii) Now, do not replace this selected chit in the population again. Then, in the population there remains (N – 1) units after the first draw which are:

$$U: \{U_1, U_2, \ldots, U_{k-1}, U_{k+1}, \ldots, U_N\}$$

and the frame consists of (N –1) numbers i = 1, 2, 3, …, N; $i \ne k$. Out of the remaining (N –1) chits in the vessel, we have to select one chit randomly at the second draw. Obviously then, each of the remaining units in the population will have conditional probability of selection in the sample equal to $1/(N-1)$, given that the unit U_k is already selected at the first draw.

Select again one chit randomly from the vessel in the second draw and note down the number written on it. Say it is r; then r^{th} unit of the population, U_r ($1 \le r \le N$; $r \ne k$) is assumed to be selected for the sample at the second draw.

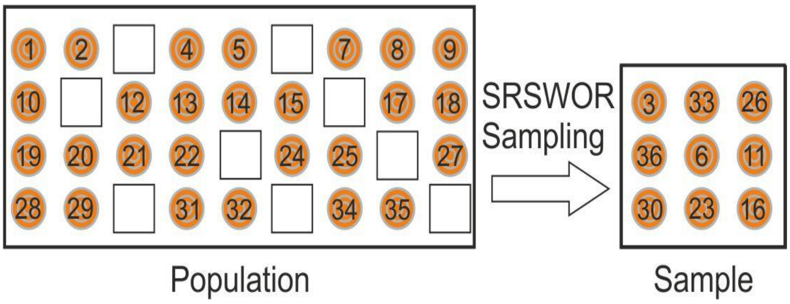

Fig. 2.1: Simple Random Sampling without Replacement Scheme

(ix) Do not replace the chit drawn at the second draw in the vessel, so that the vessel now consists of only $(N-2)$ chits. This means that units U_k and U_r are not included in the population before next draw. Obviously, then each unit of the population will have conditional probability of selection in the sample equal to $1/(N-2)$, given that units U_k and U_r are already selected, respectively, at the first and second draws.

Now, select one chit randomly from the vessel in the third draw and note down the number written on it. Say it is t; then t^{th} unit of the population, U_t ($1 \le t \le N$; $t \ne k, r$) is assumed to be selected for the sample at the third draw.

(x) Repeat this process a number of times until we get the n^{th} unit in the sample. The Simple Random Sampling without Replacement sample will then consist of labels {k, r, t, …} such that $k \ne r \ne t \ne$…., and the units selected in the sample will be $\{U_k, U_r, U_t, ……\}$.

(xi) Note down the corresponding values of the study variable as measured/reported, which would be Y_k, Y_r, Y_t, ….

Remark 2.1: Since, the sample size is n, you can see that at the n^{th} draw a unit will be selected with the conditional probability of being selected equal to $1/(N-n+1)$, given that in previous $(n-1)$ draws some other $(n-1)$ units of the population were selected.

Remark 2.2: Since, in Simple Random Sampling without Replacement method of selection, population size reduces by one after each draw, the maximum size of the sample cannot exceed N, which is not the case in Simple Random Sampling with Replacement.

Now, let us illustrate the process through the following example:

Example 1: Consider a population **U** of 10 units of a product of a Toys Manufacturing Company. Explain the method of selection of a simple random sampling without replacement sample of size 4 from the population using Lottery method.

Solution: Let us illustrate the selection of a simple random sampling without replacement sample of size 4 from the population of size 10.

1. Using the Lottery method, first of all we prepare the set of chits consist of numbers 1, 2, 3, …., 10 written on them. Let 10 number of chits be prepared of same colour, thickness, material, and size, i.e., identical in all respects, and numbers 1 to 10 written on these chits such that each chit has a single number.
2. We put the folded chits in a Box and mix them properly.
3. Then we select one chit from the Box, open it and note down the number it consists of. Say it is 5. Then the 5th unit U_5 $(1 \le 5 \le 10)$ of the population is assumed to be selected in the sample in the first draw.
4. Since we are using the without replacement method, we do not replace the selected chit in Box before the next draw. Now the total number of remaining chits in the box are 9.
5. Now, we select one chit again randomly from the well shuffled box in the second draw and note down the number written on it. Say, it is 7. Then the 7th unit U_7 $(1 \le 7 \le 9)$ of the population is assumed to be selected in the sample in the second draw.
6. Following the same method, we do not replace the 7th unit again in the population before the 3rd draw. Therefore, the remaining chits in the box are 8.
7. In the third draw we select again one chit randomly and note down the number written on it. Say it is 10. Then 10th unit is assumed to be selected in the sample in the third draw.
8. We again do not replace the 10th unit in the population and select again one chit from the remaining 7 chits in the Box and note down the number written on it. Say it is 1. The 1st unit of the population U_1 is assumed to be selected in the sample.
9. Therefore, the sample of 4 consists of 1st, 5th, 7th, and 10th unit as shown in the given Fig. 2.2.

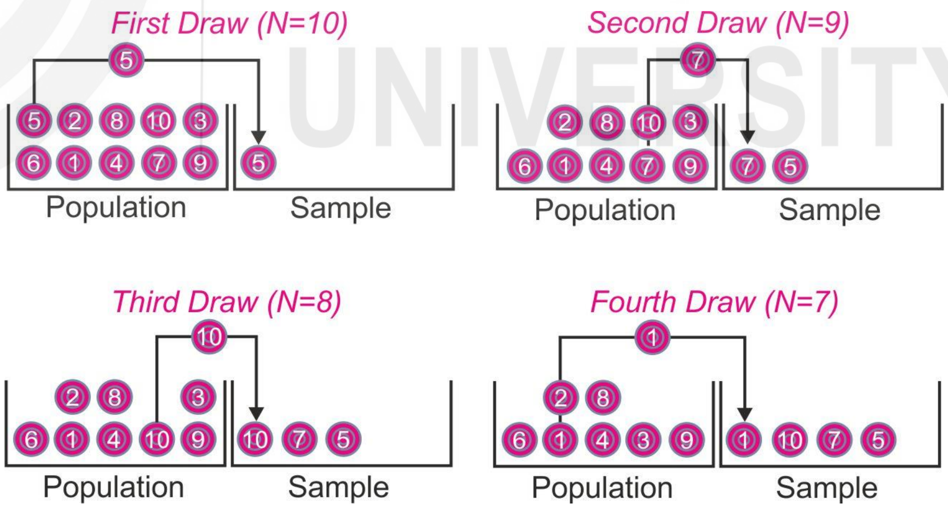

Fig. 2.2: Simple Random Sampling without Replacement Scheme

2.2.2 Use of Random Number Tables

Since Random Number Tables are prepared in such a manner that each random number in the table assures selection of units in the sample with equal probability selection method; for large size populations one can use them instead of using time consuming and costly Lottery method. The selection of a random sample with Use of Random Number Tables is done using following steps:

(1) Let us first assign labels to each of the units in the population. If population size is N, we assign numbers 1 to N to the population units, such that each unit is assigned a single number ranging from 1 to N.

During assigning the numbers to the units of the population, it is not necessary to arrange them any specific order. They can be arranged in a sequence following any method, the only purpose is to distinguish these units on the basis of numbers assigned to them. We know that such numbers assigned to the units are known as 'labels' of units and the complete and exhaustive set of labels provide us *'sampling frame'*, which is used during the selection procedure in order to visualize which units are included in the sample.

(2) Observe what is the maximum label assigned and how many digits it consists of. Use a random number table which consists of random numbers with more number of digits than the maximum label. This is necessary so that all the labels and the largest label also can be selected, if it is drawn at any draw. Generally, most of the random number tables have six-digit random numbers in order to admit maximum population size of 999999, which rarely happens in any survey.

Further, on the basis of the label, decide how many digits of the random numbers have to be used. For example, if random numbers are six-digit numbers and the labels are ranging from 1 to 850, we may consider either the leftmost three digits or the right most three digits of a random number during the selection process. Thus, if random number 347109 appears in the table, choosing the leftmost three digits **<u>347</u>**109 we say that label 347 is selected randomly and if the number **<u>089</u>**002 appears, we say that label 89 is selected.

(3) To start with the process of selection of random numbers, one is free to start with a number lying at the cross-section of a particular row and a particular column. Moreover, he/she is free to move further either row-wise or column-wise for selection of other random numbers. In this way, he/she may move to any number of rows or columns till the process of selection of required number of units is complete.

Following this rule, draw a random number and if the number is smaller than or equal to the largest label, it is assumed to be selected at the first draw for including it into the sample, otherwise, neglect this number and move to the next random number.

(4) Since, we have to select units in the sample using Simple Random Sampling without Replacement scheme, the label once selected at any draw, will not be considered in any next draw. For selecting the next unit, either move row-wise or column-wise, as decided, and go through random numbers one by one, unless a number less than or equal to largest label is obtained. If such a number is found, select it as the unit selected at the second draw, provided it is not the same unit as selected at the first draw.

(5) Repeat this process of selection of units one by one, avoiding the repetition of already selected labels, until the required sample size is reached.

(6) Using the labels selected in the sample, locate the units of the population which are included in the sample.

(7) Finally, enumerate the selected units for getting the measurements on the study variable.

Example 2: Consider a population consisting of 800 employees working in a large manufacturing company. The aim is to select a simple random sample of size 50 using without replacement method in order to estimate the average income per month in the population. You are provided the income per month (in rupees) of all the 800 employees. Choose a Simple Random Sampling without Replacement sample of employees.

Solution: Since the population size N is 800, employees should be allotted labels ranging from 1 to 800. For allotting labels, we may use either the attendance records of the employees, if maintained centrally in the record section or any other official record maintained in different sections of the company. Obviously, labelling each and every employee with a single number, our sampling frame would be like L: {1, 2, 3,…., 799, 800}. Since the largest label is of 3-digit, we can use the random number table consisting of 4-digit numbers, given in the **Appendix - A** of the Unit 1. Let us start the selection procedure randomly from the beginning of the second column of the table and moving column-wise. Further, let us decide to consider only the last three digits of every random number, since the largest label is a 3-digit number.

We observe that the first random number in the second column is 6**833**. The last three digits **833** is not to be selected as, it is a number greater than the largest label, 800. So, we move to the next number in the same column which is 4**154.** Since 154 belongs to the set of labels, it indicates that the **154th unit** of the population is selected in the sample **at the first draw**.

Similarly, next number, 3**890** does not provide us any new selection for the obvious reason. Next number is 5**692** which indicates that **692nd unit** of the population is assumed to be selected in the sample **at the second draw**.

You can now see that next five labels which are to be selected consecutively in the sample in the **third, fourth, fifth, sixth** and **seventh draws** are 516, 486, 271, 765 and 364, indicating these draws provide **516th unit**, **486th unit**, **271st unit**, **765th unit** and **364th unit** in the sample.

Continuing in the same manner, we can select one by one the other 43 units in the sample; keeping in mind that repetition of units once selected at any draw is not allowed in further draws. If at any draw a label seems to be repeated, we skip that number and move further as if did not appear.

After noting down all the selected labels, we can have the list of 50 selected employees of the company and then, their monthly income can be enumerated by contacting them personally. The sampled monthly income provides us necessary data for estimating the average monthly income in the population.

2.2.3 Modified Approaches for Simple Random Sampling without Replacement Scheme

Since many of the random numbers become ineffective in the sense that they do not provide us any new unit of the population in the sample, the number of draws in Simple Random Sampling without Replacement scheme would be more than the size of the sample. It is, in fact, a drawback of selection methods which is related with both Simple Random Sampling with Replacement and Simple Random Sampling without Replacement methods. As for instance, we observed that within the first seven draws in the above example, there were 3 ineffective numbers, which are 6**833**, 3**890** and 0**924**. However, there are two approaches through which we can utilize fruitfully the ineffective random numbers also to generate labels which can be thought of being selected in the sample. The two approaches are as follows:

(i) Remainder Approach

In this method there is no need to reject any random number because of the reason mentioned above; rather such random number can also be fruitfully utilized to generate a label which can be included in the sample. Since this method is equally applicable to Simple Random Sampling with Replacement and Simple Random Sampling without Replacement methods, below we shall consider both the schemes simultaneously. The steps are as follows:

(1) Let N be an r-digit number and let its r-digit highest multiple be N^*. We can divide the range of numbers 01 to N^* in two ranges: 01 to N and N+1 to N^*. Ignore all other r-digit numbers greater than N^*. Select a random number from the table, say k, from 1 to N^*.

(2) If the selected r-digit number k is in the range 01 to N, the number is directly selected; but if the number k is in the range N+1 to N^*, divide k by N and find the reminder. Then select the unit of the population in the sample whose label is equal to the reminder. If k is greater than N^*, ignore the number and proceed for the next random number.

(3) If the reminder is zero, select the last unit of the population.

Let us consider the following example:

Example 3: Consider the population **U** of 40 labours working in a construction company in which wage per day (in rupees) is the characteristic under study:

U = {263, 398, 209, 331, 170, 125, 339, 181, 331, 170, 140, 100, 550, 75, 277, 255, 225, 415, 180, 440, 220, 25, 395, 480, 365, 178, 135, 150, 280, 390, 515, 565, 295, 20, 260, 300, 75, 222, 55, 312}.

Select a sample of size n = 12 using random number table method.

Solution: Here, N = 40, a two-digit number; so, r = 2. Then, N^* = 80; which is the highest 2-digit multiple of 40. We shall, therefore, consider random numbers from the table only from 01 to 80 ignoring 2-digit numbers from 81 to 99. If the random number selected is in the range 01 to 40, then the selected number (label) is directly included in the sample, but if the number selected is in the range 41 to 80; divide the number by 40 and use the reminder for selecting a label equal to the reminder. Skip all other 2-digit numbers beyond

80. If the reminder at any draw is equal to a label already selected in some previous draws, avoid it in case of Simple Random Sampling without Replacement selection method and include it again in the sample in case of Simple Random Sampling with Replacement selection method.

Considering the random number table given in **Appendix - A** in the previous unit, let us start with 4-digit random number **34**36, lying in the first row and first column and move row-wise. Here we decided to take the leftmost two digits of random numbers. Since 34 < 40 (=N), it is selected in the sample at the first draw in case of both the sampling schemes. Next 4-digit random number is **68**33. Since 40 <**68**< 80, we divide 68 by 40. Reminder is 28, so label 28 is selected in the sample in both the cases at the second draw. Next 4-digit number is **58**09, so we divide 58 by 40; the remainder is 18 which is selected in the sample as the label selected in the third draw. Next 4-digit number is **91**69. Since 91 > 80, skip this number and proceed for the next 4-digit number. Proceeding similarly further, it can be seen that the 12 labels which are finally included in the Simple Random Sampling without Replacement sample are {34, 28, 18, 10, 16, 25, 13, 21, 1, 26, 35, 36}; whereas the labels which are finally included in the Simple Random Sampling with Replacement sample are {34, 28, 18, 10, 16, 25, 28, 13, 21, 1, 26, 35}.

(ii) Quotient Approach

Let N be an r-digit number and let its r-digit highest multiple be N^* such that $N^*/N = t$. A random number k is chosen from 0 to (N^*-1). Dividing k by t, the quotient r is obtained and the unit bearing the label is (r-1) is selected in the sample. Here, in the given example, N = 40, N^* = 80 and t = 80/40 = 2. Let the 4-digit random number selected is **68**33 lying between 0 to 79. Dividing 68 by 2, the quotient is 34, hence, the unit bearing label 33 is selected in the sample. Let another random number selected be **13**32. Dividing 13 by 2 we get quotient 6.5. The integer part of it is 6. Thus, 5^{th}unit is selected in the sample.

It is seen that both the above-mentioned approaches are equally applicable for both Simple Random Sampling with Replacement and Simple Random Sampling without Replacement schemes.

Now you may try to answer the following Self-Assessment Question:

SAQ 1

What do you mean by Simple Random Sampling without Replacement (SRSWOR) scheme? Explain what is meant by the term **'Without Replacement'** here.

2.3 FUNDAMENTAL PROPERTIES OF SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT SCHEME

In this section, we shall describe some of the fundamental properties of Simple Random Sampling without Replacement scheme, which make it different from Simple Random Sampling with Replacement scheme; though both are Simple Random Sampling schemes. Wherever necessary we shall provide proofs of these results in terms of theorems. Throughout in this section N stands for the population size and n stands for the sample size.

Theorem 1: In Simple Random Sampling without Replacement scheme, the probability of a specified unit of the population being selected at any draw is equal to the probability of its being selected at the first draw, which is 1/N.

Proof: Let the specified unit of the population be r^{th} unit.

Consider the first draw. Since the population size is N; there would be N units in the population at the moment the first draw is made. Therefore, probability of selection of the r^{th} unit of the population at the first draw would be 1/N.

Now, consider the second draw. Since under this scheme, the r^{th} unit drawn at the first draw is not to be replaced in the population, there would be (N –1) units remaining in the population after the first draw. Therefore, if the r^{th} unit is not selected at the first draw, it may be selected at any future draws.

Therefore, the probability that it is selected at the second draw will be given by the product of two probabilities:

(i) the probability of the event that it is not selected in the first draw; and

(ii) the probability of the event that it is selected at the second draw given that it is not selected at the first draw.

Clearly, the second probability is a conditional probability of the type P(B|A), where B and A are two events. Since, draws are independent to each other, we can apply the multiplication theorem of probability in order to find the unconditional probability of selection of the r^{th} unit at the second draw.

Therefore, we have

$$P\begin{pmatrix}r^{th}\text{unit is}\\ \text{selected at the}\\ \text{second draw}\end{pmatrix} = P\begin{pmatrix}r^{th}\text{unit is not}\\ \text{selected at the}\\ \text{first draw}\end{pmatrix} \times P\begin{pmatrix}\text{it is selected at the}\\ \text{second draw given}\\ \text{that it is not selected}\\ \text{at the first draw}\end{pmatrix};$$

Clearly,

$$P\begin{pmatrix}r^{th}\text{unit is not}\\ \text{selected at the}\\ \text{first draw}\end{pmatrix} = P\left(1-\text{it is selected at the first draw}\right) = \left(1-\frac{1}{N}\right);$$

and the conditional probability,

$$P\begin{pmatrix}\text{it is selected at the second draw given that it is not}\\ \text{selected at the first draw}\end{pmatrix} = \frac{1}{N-1};$$

since at the first draw some other unit is selected with selection probability $\frac{1}{N}$.

Therefore,

$$P\begin{pmatrix}\text{it is selected at the second draw and it is not}\\ \text{selected at the first draw}\end{pmatrix} = \left(1-\frac{1}{N}\right)\frac{1}{N-1} = \frac{1}{N};$$

indicating that the unconditional or absolute probability of selection of the r^{th} unit at the second draw is also $\frac{1}{N}$.

Now, consider the third draw. Obviously, after the second draw, the size of the population will be reduced to (N – 2). Therefore,

$$P\begin{pmatrix}r^{th}\text{unit is}\\ \text{selected at}\\ \text{the third draw}\end{pmatrix}=P\begin{pmatrix}\text{It is not}\\ \text{selected at}\\ \text{the first draw}\end{pmatrix}\times P\begin{pmatrix}\text{It is not}\\ \text{selected at the}\\ \text{second draw}\end{pmatrix}\times P\begin{pmatrix}\text{It is selected}\\ \text{at the third}\\ \text{draw}\end{pmatrix};$$

$$P\begin{pmatrix}r^{th}\text{unit is selected}\\ \text{at the third draw}\end{pmatrix}=\left(1-\frac{1}{N}\right)\times\left(1-\frac{1}{N-1}\right)\times\left(\frac{1}{N-2}\right)=\frac{1}{N};$$

Similarly, consider the r^{th} draw.

We have, with the same arguments, that the

$$P\begin{pmatrix}r^{th}\text{unit is}\\ \text{selected at}\\ \text{the } r^{th}\text{draw}\end{pmatrix}=P\begin{pmatrix}\text{It is not}\\ \text{selected at}\\ \text{the } 1^{st}\text{draw}\end{pmatrix}\times P\begin{pmatrix}\text{It is not}\\ \text{selected at}\\ \text{the } 2^{nd}\text{draw}\end{pmatrix}\times\ldots\ldots$$

$$\times P\begin{pmatrix}\text{It is not}\\ \text{selected at the}\\ (r-1)^{th}\text{ draw}\end{pmatrix}\times P\begin{pmatrix}\text{It is}\\ \text{selected at}\\ \text{the } r^{th}\text{draw}\end{pmatrix};$$

$$=\left(1-\frac{1}{N}\right)\left(1-\frac{1}{N-1}\right)\ldots\left(1-\frac{1}{N-r+2}\right)\left(\frac{1}{N-r+1}\right)=\frac{1}{N}.$$

This shows that the probability of selecting any specified unit at any draw is $\frac{1}{N}$.

We, therefore, conclude that the probability of selection of all the units under Simple Random Sampling without Replacement is 1/N.

Remark 2.3: The **Theorem 1** shows that SRSWOR is also an Equal Probability of Selection Method (EPSEM), similar to Simple Random Sampling with Replacement, that is, we can define the Simple Random Sampling without Replacement as a sampling scheme in which the probability of selection of any unit of the population at any draw is 1/N; implying that it also follows the rule of Equal Probability of Selection Method (EPSEM).

Remark 2.4: As per the definition of Simple Random Sampling provided in previous unit, it is the method of selection of units from a given population one by one in a number of independent draws, following some chance mechanism, in such a way that each unit of the population has equal chance of being selected in the sample.

After the discussion made separately on Simple Random Sampling with Replacement and Simple Random Sampling without Replacement, it is now clear that as for as the Simple Random sampling is concerned, the above definition gives the true nature of the SRS scheme, and it does not change according to the schemes “With Replacement” (WR) or “Without Replacement” (WOR).

Theorem 2: In Simple Random Sampling without Replacement scheme, the probability that a specified unit of the population is included in the sample of size n is n/N.

Proof: In Simple Random Sampling without Replacement scheme also, we have seen in **Theorem 1** above that probability of selecting a particular unit of the population at any draw is **1/N**. Since all the draws are independent to each other and the unit may be selected independently either at the first draw, second draw, third draw, …, or at the n^{th} draw with same probability of selection; the probability $P(E_r)$, denoting the probability that the specified unit is selected in the sample at the r^{th} draw will be given

$$P\left(E_r\right)=\frac{1}{N}; \text{ for } r = 1, 2, 3,\ldots, n;$$

Thus, due to the additive law of probability we have

$$\sum_{r=1}^{n} P\left(E_r\right)=P(E_1)+P(E_2)+....+P(E_n)=\frac{1}{N}+\frac{1}{N}+........+\frac{1}{N}$$

$$P\left(\bigcup_{r=1}^{n}E_r\right)=\sum_{r=1}^{n}\frac{1}{N}=\frac{n}{N} \qquad \ldots (2.1)$$

Theorem 3: Total number of samples of size n which may be selected from a population of size N under Simple Random Sampling without Replacement scheme is ${}^{N}C_n$.

Proof: Since, in this case a unit selected once is not again replaced in the population; therefore, in each draw a different unit from the population has to be selected in the sample.

In other words, a sample of size n consists of n distinct units of the population. Clearly, it is equivalent to the problem of combination in which n places available in the sample have to be filled by N units belonging to the population without the repetition of units in the sample. If order of selection of units in the sample is immaterial, we know that it would be ${}^{N}C_n$.

Thus, the total number of Simple Random Sampling without Replacement samples would be ${}^{N}C_n$, which is also denoted as $\binom{N}{n}$.

Theorem 4: The probability of selection of a Simple Random Sampling without Replacement sample is given by

$$\frac{1}{{}^{N}C_n}=\left({}^{N}C_n\right)^{-1}=\binom{N}{n}^{-1}.$$

Proof: Since, in Simple Random Sampling without Replacement, the total number of simple random sampling without replacement samples are ${}^{N}C_n$ and since each sample out of possible ${}^{N}C_n$ samples has an equal probability of selection; the probability of selection of a Simple Random Sampling without Replacement sample will be $\frac{1}{{}^{N}C_n}$ which is also denoted as $\binom{N}{n}^{-1}$.

Now you may try to answer the following Self-Assessment Question:

SAQ 2

In Simple Random Sampling without Replacement scheme, show that the probability of a specified unit of the population being selected at any draw is equal to the probability of its being selected at the first draw, which is 1/N.

2.4 ESTIMATION OF POPULATION MEAN, TOTAL AND VARIANCE IN SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT SCHEME

After the discussions on the nature of Simple Random Sampling without Replacement scheme, methods of selection of a sample following the rules of Simple Random Sampling without Replacement scheme and basic results of this scheme; now we are able to go through the second aspect of the scheme, that is, the 'estimation procedure' under the scheme. As before, we shall describe here how the Population Mean, Population Total and Population Variance can be estimated through using a Simple Random Sampling without Replacement sample. Let us discuss the respective estimators and their salient properties one by one.

2.4.1 Estimation of Population Mean

We know that population mean is given by

$$\bar{Y} = \frac{1}{N}\left[Y_1 + Y_2 + Y_3 + ... + Y_N\right] = \frac{1}{N}\sum_{i=1}^{N} Y_i \qquad \text{... (2.2)}$$

The Estimation Theory of Statistics tells us that besides the sample mean, there are other measures of central tendency, like, median, mode, quartiles, etc., which may also be used to estimate the population mean; but, due to a number of good properties it possesses, sample mean is considered to be the best estimator of population mean. Also, it is a direct and immediate estimator of population mean, as, its nature and computation method are same as that of population mean.

Therefore, we shall consider here the sample mean as an estimator of population mean $\bar{Y}$. The sample mean of a sample of size n would be

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

We have the following property of sample mean:

Theorem 5: Under the scheme Simple Random Sampling without Replacement, sample mean, $\bar{y}$ is an unbiased estimator of the population mean $\bar{Y}$.

Proof: We have population mean as

$$\bar{Y} = \frac{1}{N}\left[Y_1 + Y_2 + Y_3 + ... + Y_N\right] = \frac{1}{N}\sum_{i=1}^{N} Y_i \qquad \text{... (2.3)}$$

and sample mean as

$$\bar{y} = \frac{1}{n}\sum_{i}^{n} y_i \qquad \text{... (2.4)}$$

To show the unbiasedness property of the estimator $\bar{y}$, we are required to show that $E(\bar{y})$ must be equal to $\bar{Y}$. We have

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

$$= \frac{1}{n}\left[\sum_{i}^{n} E(y_i)\right] = \frac{1}{n}\left[E(y_1) + E(y_2) + \ldots + E(y_n)\right] \qquad \ldots (2.5)$$

using the results of expectation.

We have seen that whether it is Simple Random Sampling with Replacement (SRSWR) or Simple Random Sampling without Replacement (SRSWOR) scheme, the probability of selection of a particular unit of the population in the sample at each draw is 1/N (see the **Theorem 1** under Section 2.3 above).

Therefore, each y_i, i= 1, 2, …, n is a random variable assuming values Y_1, Y_2, …, Y_N with selection probability 1/N and, hence, we have

$$E(y_i) = \sum_{i=1}^{N} Y_i\left(\frac{1}{N}\right) = \frac{1}{N}\sum_{i=1}^{N} Y_i = \bar{Y} \qquad \text{for } i = 1, 2, \ldots, n.$$

Thus, equation (2.5) reduces to

$$E(\bar{y}) = \frac{1}{n}\left[E(y_1) + E(y_2) + \ldots + E(y_n)\right] = \frac{1}{n}\left[\bar{Y} + \bar{Y} + \ldots + \bar{Y}\right];$$

where inside the bracket we have n terms all equal to $\bar{Y}$.

Therefore,

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

showing that sample mean $\bar{y}$ is an unbiased estimator of the population mean $\bar{Y}$ under (SRSWOR) scheme. This establishes the Theorem.

Remark 2.5: Combining the **Theorem 5** of the Unit 1 of this Block and of the present unit; we conclude that as far as the estimation of the population mean is concerned, the immediate estimator, the sample mean $\bar{y}$, possesses the property of unbiasedness either it is (SRSWR) or (SRSWOR) scheme. You are familiar of the fact that unbiasedness is a desirable property of any estimator.

Remark 2.6: The alternative proof of the theorem 5, presented in the previous unit given for (SRSWR) scheme is applicable as well for the (SRSWOR) scheme also, since, in both cases the selection probability of a unit of the population in the sample at any draw is same. Therefore, in taking expectation of the estimator $\bar{y}$, no difference occurs.

2.4.2 Estimation of Population Total

The total of the population values is also a desired parameter in some cases rather than average of the population values. We know the relation between population total and the population mean. It is given in notations as

$$\text{Population Mean } (\bar{Y}) = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{\text{Total of the Population Values}}{\text{Population Size}}$$

$$\Rightarrow \sum_{i=1}^{N} Y = N \times \bar{Y}$$

$$\Rightarrow \text{Population Total} = \text{Population Size} \times \text{ Population Mean}$$

This relation helps us in quickly finding the unbiased estimator of population total using the estimator of population mean. We have the result in the following theorem of sample total:

Theorem 6: An unbiased estimator of population total is given by $N\bar{y}$.

Proof: Since the population mean is

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

this implies that population total $\sum_{i=1}^{N} Y_i = N\bar{Y}$

Now, consider the sample mean given by

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

which indicates that sample total $\sum_{i}^{n} y_i = n\bar{y}$

We know that $E(\bar{y}) = \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

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

$$\Rightarrow E(N\bar{y}) = \sum_{i=1}^{N} Y_i \quad (\text{the population total})$$

This shows that for estimating the population total, an unbiased estimator is obtained as $N\bar{y}$. So, the sample mean $\bar{y}$, when multiplied by population size, i.e., $N\bar{y}$ provides the unbiased estimator of population total. This establishes the property mentioned in the above theorem.

2.4.3 Estimation of Population Variance

We know that population variability may be measured either by population variance,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 \qquad \ldots (2.6)$$

or by population mean square

$$S^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2. \qquad \ldots (2.7)$$

Also, from equation (2.6) and (2.7) we see that

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

Therefore, if we find an estimator of any of the two measures i.e., Population variance and population mean squares, we can find easily the estimator of the other.

As described earlier in the previous unit, sample mean square, s^2, is an unbiased estimator of population variance, σ^2, when sampling scheme was Simple Random Sampling with Replacement. We shall observe here whether we get the same result in the case of Simple Random Sampling without Replacement also or we have a different result in SRSWOR scheme. In fact, we get a different result in case of Simple Random Sampling without Replacement scheme. The following theorem shows that sample mean square, s^2, is an unbiased estimator of population mean square, S^2, and not of the population variance, σ^2.

Theorem 7: Under Simple Random Sampling without Replacement scheme, s^2 is an unbiased estimator of population mean square, S^2.

Proof: The sample mean square,

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

Therefore, we have

$$E\left(s^2\right) = E\left\{\frac{1}{n-1}\sum_i^n y_i^2 + \frac{n}{n-1}\bar{y}^2 - \frac{2n}{n-1}\bar{y}^2\right\},$$

(since, $\frac{1}{n-1}\sum_i^n 2y_i\bar{y} = \frac{2\bar{y}}{n-1}\sum_i^n y_i = \frac{2n}{n-1}\bar{y}^2$)

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

$$= \frac{1}{n-1}\sum_i^n E(y_i^2) - \frac{n}{n-1}E\left(\bar{y}^2\right) \qquad \ldots (2.8)$$

Now, as per definition of expectation, we have

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

Also, we have

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

Therefore, substituting these values in (2.8), we have (2.8) as

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

$$= \frac{n}{N(n-1)}\sum_{i=1}^N Y_i^2 - \frac{1}{n(n-1)}E\left(\sum_i^n y_i^2 + \sum_i^n\sum_{j\neq i}^n y_iy_j\right)$$

{since Square of Sums = Sum of Squares + Product terms}

$$= \frac{n}{N(n-1)}\sum_{i=1}^N Y_i^2 - \frac{1}{n(n-1)}\left(\sum_i^n E\left(y_i^2\right)\right) - \frac{1}{n(n-1)}\left(\sum_i^n\sum_{j\neq i}^n E\left(y_iy_j\right)\right)$$

…(2.9)

Now, consider the term $\sum_i^n\sum_{j\neq i}^n E\left(y_iy_j\right)$ in the last bracket in the equation (2.9).

Now, in order to evaluate the term $\sum_i^n\sum_{j\neq i}^n E\left(y_iy_j\right)$ in case of Simple Random Sampling without Replacement scheme, we shall consider the selection procedure of a sample under this scheme. Since, in Simple Random Sampling without Replacement scheme, a unit drawn in the sample at any draw is not replaced in the population again before the next draw, the population size decreases by one after every draw and, hence, the conditional probability of selecting a different unit in the next draw is not same as it is in the first draw (see **Theorem 1** under section 2.3 for reference). In this sense, the draws cannot be considered to be independent to each other as it was in the case of Simple Random Sampling with Replacement scheme.

This fact changes the value of the expression $E\left(\sum_i^n \sum_{j\neq i}^n y_i y_j\right)$ in the equation (2.9), other things remaining same. We then have

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

since, there would be n (n–1) product terms and

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

Substituting this value in the expression (2.9), we get

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

We know that

Square of Sums = Sum of Squares + Product terms;

that is, $\left(\sum_{i=1}^N Y_i\right)^2 = \sum_{i=1}^N Y_i^2 + \sum_{j\neq i=1}^N Y_i Y_j$;

therefore, we have

$$\sum_{j\neq i=1}^N Y_i Y_j = \left(\sum_{i=1}^N Y_i\right)^2 - \sum_{i=1}^N Y_i^2 = N^2\bar{Y}^2 - \sum_{i=1}^N Y_i^2$$

and hence, we have

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

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

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

This completes the proof of the property mentioned in the above theorem.

Remark 2.7: From this result, we can find the unbiased estimator of population variance σ^2 also. We observe that,

$$E\left(s^2\right) = S^2 = \frac{N}{(N-1)}\sigma^2 \quad \Rightarrow E\left(\frac{N-1}{N}s^2\right) = \sigma^2.$$

Therefore, an unbiased estimator of population variance, σ^2, is $\left(\frac{N-1}{N}s^2\right)$.

Remark 2.8: We can now see the difference in the results of **Theorem *7*** of the Unit 1 and **Theorem *7*** of the present unit. While in Simple Random

Sampling without Replacement scheme, $E(s^2) = \sigma^2$; in Simple Random Sampling without Replacement scheme, we have $E(s^2) = S^2$.

Although, for large populations, (N – 1) is almost equal to N in the expression,

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

so, practically we see that $\sigma^2 \cong S^2$. But for small size populations, $\sigma^2 \neq S^2$.

Now you may try to answer the following Self-Assessment Question:

SAQ 3

Find an unbiased estimator of population variance if the sample be drawn under Simple Random Sampling without Replacement.

2.5 VARIANCE OF THE SAMPLE MEAN

The estimator of population mean in Simple Random Sampling without Replacement scheme is sample mean $\bar{y}$. Since, sample mean varies over sample to sample, it is a random variable which assumes ${}^NC_n$ values with equal probability $\frac{1}{{}^NC_n}$ (see the Theorem 4 of this unit). Obviously, being a random variable, it possesses some variability amongst the values and, therefore, it has some variance. Let us find the variance of this estimator under the Simple Random Sampling without Replacement scheme in the form of the theorem of estimator $\bar{y}$, as:

Theorem 8: The variance of the estimator $\bar{y}$, under Simple Random Sampling without Replacement scheme, is given by

$$V(\bar{y}) = \frac{N-n}{Nn}S^2 \qquad \ldots (2.10)$$

Proof: We know that

$$V(\hat{\theta}) = E\left[\hat{\theta} - E(\hat{\theta})\right]^2;$$ where $\hat{\theta}$ is the sample mean estimator.

So, we get,

$$V(\bar{y}) = E\left[\bar{y} - E(\bar{y})\right]^2 = E(\bar{y}^2) - [E(\bar{y})]^2; \qquad \ldots (2.11)$$

after expanding the expression $E\left[\bar{y} - E(\bar{y})\right]^2$ and simplifying it.

We see that $[E(\bar{y})]^2 = \bar{Y}^2$ since, $E(\bar{y}) = \bar{Y}$.

Let us consider the expression $E(\bar{y}^2)$. We have

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

$$= \frac{1}{n^2}\sum_{i}^{n}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i^2\right) + \frac{n(n-1)}{n^2}E(y_i y_j) \qquad \ldots (2.12)$$

Under this scheme, we have seen that,

$$E(y_i y_j) = \frac{1}{N(N-1)} \sum_{i \neq j=1}^{N} Y_i Y_j. \qquad , (i \neq j).$$

Therefore, we have

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

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

Therefore,

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

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

Collecting terms containing $\sum_{i=1}^{N} Y_i^2$ and $\bar{Y}^2$ separately and solving for their multipliers, we get

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

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

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

This completes the proof of the theorem.

Remark 2.9: In the previous unit, we observed that in case of SRSWR scheme, the sampling variance of the estimator of population mean, denoted by $V(\bar{y})_{WR}$ is obtained as

$$V(\bar{y})_{WR} = \left(\frac{N-1}{N}\right).\frac{S^2}{n} = \frac{\sigma^2}{n}; \qquad \text{… (2.13)}$$

whereas, the sampling variance of the same estimator, $\bar{y}$, in case of SRSWOR scheme is given by

$$V(\bar{y})_{WOR} = \left(\frac{N-n}{N}\right).\frac{S^2}{n}. \qquad \text{… (2.14)}$$

We shall explain why the expressions come in this way. We know that in the distribution theory of Statistics, for an infinite population, the sampling variance of the sample mean estimator is given by $\frac{\sigma^2}{n}$ where σ^2 represents the variance of the population. For example, in the case of a normal distribution $N(\mu, \sigma^2)$; the sample mean, $\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. The central limit theorem states that for infinitely large sample, the sample mean always have variance equal to

$$V(\bar{y}) = \frac{\text{Population Variance}}{\text{Sample Size}} = \frac{\sigma^2}{n} .$$

Therefore, the expression (2.13) indicates that, although we deal with a finite population in sampling theory, theoretically under 'with replacement' case, the finite population behaves like an infinite population. This is, in fact, due to the reason that under this selection scheme, replacement of units in the population after each draw makes it possible to select a sample of infinite size, if we wish, implying that the finite population is equivalent to a population of infinite size.

Contrary to this, in Simple Random Sampling without Replacement scheme, we observe that sampling variance of the mean estimator, $\bar{y}$, is

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

implying that in case of Simple Random Sampling without Replacement scheme, the population does not behave like an infinite population. This is because of the reason that under this scheme, samples of sizes larger than the population size are not possible to select, since the size of the population decreases by one after each draw and finally, population diminishes after the N^{th} draw.

In fact, the term $\left(\frac{1}{n}-\frac{1}{N}\right)$ in the variance expression is called the "finite population correction" (f.p.c).

2.5.1 Estimation of the Variance of Sample Mean

We can find an unbiased estimator of the sampling variance of the mean estimator in Simple Random Sampling without Replacement so that whenever the numerical value of the parameter S^2 is not known, we can use its estimator, as obtained from the sample values. We have the following theorem regarding the estimate of the sampling variance:

Theorem 9: An unbiased estimator of the sampling variance, is given by

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

Proof: We have, from the Theorem 7

$$V(\bar{y}) = \frac{N-1}{Nn}S^2 = \frac{N-1}{Nn}E\left(s^2\right) \qquad \text{Since } E\left(s^2\right) = S^2$$

$$\Rightarrow V(\bar{y}) = E\left(\frac{N-1}{Nn}s^2\right)$$

then

$$E\left[\frac{(N-1)}{Nn}s^2\right] = \frac{N-1}{n\,N}S^2 = V(\bar{y})$$

This shows that an unbiased estimator of $V(\bar{y})$ is given by

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

$$\Rightarrow \text{Est.}\{V(\bar{y})\} = \frac{N-n}{nN}s^2$$

where, s^2, being sample mean square, can be computed with the help of sample values.

We may use the notation $\hat{V}(\bar{y})$ or Est.$(V(\bar{y}))$ to denote the estimator of $V(\bar{y})$, therefore, we express the result as

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

Now you may try to answer the following Self-Assessment Question:

SAQ 4

Let us consider the finite population given by

U : {5, 45, 60, 35, 10, 22}.

On the basis of a Simple Random Sampling without Replacement sample of size 3, find an estimate of the population mean. Also, find the sampling variance of the concerned estimator.

2.6 COMPARISON OF SIMPLE RANDOM SAMPLING WITH REPLACEMENT AND SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT SCHEMES

We have already mentioned in the previous unit that the main purpose of various sampling schemes is to develop efficient estimation techniques on the basis of selected sample for estimating some population parameters. Here the term *"Efficiency"* of an estimator has the same meaning as defined in the Theory of Estimation in the subject of Statistics. It is reproduced below:

'Given two estimators θ_1 and θ_2 of a parameter, the estimator θ_1 is said to be more efficient than θ_2 if the Variance or Mean Square Error of θ_1 is less than the Variance or Mean Square Error of θ_2'.

In this context, we can observe which one of the methods Simple Random Sampling with Replacement or Simple Random Sampling without Replacement is more efficient than the other.

Since,

$$V(\bar{y})_{WR} = \left(\frac{N-1}{N}\right).\frac{S^2}{n} = \left(1-\frac{1}{N}\right).\frac{S^2}{n}$$

and

$$V(\bar{y})_{WOR} = \left(\frac{N-n}{N}\right).\frac{S^2}{n} = \left(1-\frac{n}{N}\right).\frac{S^2}{n}$$

It is clear that

$$V(\bar{y})_{WOR} < V(\bar{y})_{WR} \text{ i.e., Since } \left(1-\frac{n}{N}\right) < \left(1-\frac{1}{N}\right), \text{ as } n > 1 \text{ always.}$$

Due to this reason, whenever one wishes to use Simple Random Sampling scheme for estimating a population parameter, Simple Random Sampling without Replacement is always preferred over Simple Random Sampling with Replacement scheme. Henceforth, when in the further texts we talk about SRS scheme, we always mean Simple Random Sampling without Replacement.

Example 4: Explain by considering a sample of size n = 2 from a population consisting of five elements 2, 3, 6, 8, 11 that Simple Random Sampling without Replacement gives a better estimate of population mean than Simple Random Sampling with Replacement.

Solution: The five elements of the population are 2, 3, 6, 8 and 11.

$\therefore$ Population Mean, $\bar{Y} = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$

and $\sum_{i=1}^{5} Y_i^2 = (2)^2 + (3)^2 + (6)^2 + (8)^2 + (11)^2 = 4+9+36+64+121 = 234$

Let us find the value of variance

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

$$= \frac{1}{4}\left[234 - 5\times(6)^2\right] = \frac{1}{4}\left[234-180\right] = \frac{54}{4} = 13.5$$

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

$$= \frac{1}{5}\left[234 - 5\times(6)^2\right] = \frac{1}{5}\left[234-180\right] = \frac{54}{5} = 10.8$$

Therefore,

$$V(\bar{y})_{WOR} = \frac{(N-n)}{N.n}S^2 = \frac{(5-2)}{5\times 2}\times 13.5$$

$$= \frac{3}{10}\times 13.5 = \frac{40.5}{10} = 4.05$$

and

$$V(\bar{y})_{WR} = \frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4$$

Since $V(\bar{y})_{WOR} \le V(\bar{y})_{WR}$;

Simple Radnom Sampling without Replacement gives a better estimator of population mean than Simple Random Sampling with Replacement.

Example 5: In a population of size N = 5, the values of the population characteristic are 1, 3, 5, 7, 9, and a sample of size 2 is drawn using SRS without replacement scheme. Prove that sample mean $\bar{y}$ is an unbiased estimator of population mean $(\bar{Y})$. Also calculate the variance of the sample mean $\bar{y}$

Solution: We are given that $N=5$ and units are 1, 3, 5, 7 and 9.

Population Mean, $\bar{Y}=\frac{1+3+5+7+9}{5}=5$

The number of possible samples in case of without replacement are ${}^5C_2=10$. The possible samples are as follows:

Sample No	Sample	$\bar{y}$	$\bar{y}-\bar{Y}$	$(\bar{y}-\bar{Y})^2$
1	(1, 3)	2	–3	9
2	(1, 5)	3	–2	4
3	(1, 7)	4	–1	1
4	(1, 9)	5	0	0
5	(3, 5)	4	–1	1
6	(3, 7)	5	0	0
7	(3, 9)	6	1	1
8	(5, 7)	6	1	1
9	(5, 9)	7	2	4
10	(7, 9)	8	3	9
		$\sum_i^{10}\bar{y}_i=50$		$\sum_i^{10}(\bar{y}_i-\bar{Y})^2=30$

The sample mean is:

$$E(\bar{y})=\frac{\sum_i^{10}\bar{y}_i}{{}^NC_n}=\frac{50}{10}=5=\bar{Y}$$

Thus, sample mean is an unbiased estimate of population mean. Now,

$$\sum Y_i^2=1^2+3^2+5^2+7^2+9^2=165$$

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

$$=\frac{1}{4}\left[165-5\times(5)^2\right]=\frac{1}{4}[165-125]=\frac{40}{4}=10$$

Therefore,

$$V(\bar{y})_{WOR}=\frac{(N-n)}{N.n}S^2=\frac{(5-2)}{5\times 2}\times 10=3$$

Also by definition,

$$Var(\bar{y})=\frac{\sum_i^{{}^NC_n}(\bar{y}_i-\bar{Y})^2}{{}^NC_n}=\frac{30}{10}=3$$

Example 6: A random sample of 10 has been taken from population of 100 students without replacement, where heights are given below:

165, 160, 165, 170, 172, 160, 165, 175, 164, 168

Estimate the mean height of the population and the standard error of mean.

Solution: We are given that $N = 100$ and $n = 10$

Sample mean, $\bar{y} = \dfrac{165+160+165+170+172+160+165+175+164+168}{10}$

$$= \frac{1664}{10} = 166.4\,\text{cm}$$

For estimating Standard Error of mean, let us calculate Sample mean square,

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

$$= \frac{1}{9}\begin{bmatrix}(165-166.4)^2 + (160-166.4)^2 + (165-166.4)^2 + (170-166.4)^2 \\ +(170-166.4)^2 + (160-166.4)^2 + (165-166.4)^2 + (175-166.4)^2 \\ +(164.166.4)^2 + (168-166.4)^2\end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix}(-1.4)^2 + (-6.4)^2 + (-1.4)^2 + (3.6)^2 + (5.6)^2 \\ +(-6.4)^2 + (-1.4)^2 + (8.6)^2 + (-2.4)^2 + (1.6)^2\end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix}(1.96) + (40.96) + (1.96) + (12.96) + (31.36) \\ +(40.96) + (1.96) + (73.96) + (5.76) + (2.56)\end{bmatrix}$$

$$= \frac{2144}{9} = 23.82$$

Therefore, the estimate of Standard Error of mean is,

$$S.E.(\bar{y}) = \sqrt{\frac{(N-n)}{N}\left(\frac{s^2}{n}\right)} = \sqrt{\frac{(100-10)}{100} \times \frac{23.82}{10}} = \sqrt{\frac{2143.8}{1000}} = 1.46$$

Now you may try to answer the following Self-Assessment Question:

SAQ 5

A simple random sample with replacement of 30 households was drawn from a city area consisting of 1500 households. The number of persons per household in the sample were as follows:

5, 6, 3, 3, 2, 3, 3, 3, 4, 4, 3, 2, 7, 4, 3, 5, 4, 4, 3, 3, 4, 3, 3, 1, 2, 4, 3, 4, 2, 4

Estimate the average number of people in the area and standard error of the sample mean.

2.7 SUMMARY

In this unit, we have discussed:

- The concept of "Simple Random Sampling Without Replacement" (SRSWOR) scheme in contrast to "Simple Random Sampling with Replacement" (SRSWR) scheme, discussed in detail in the previous unit.
- The problem of selecting a Simple Random Sampling without Replacement sample from the given population with example.
- The most popular methods of selection, namely, Lottery method (chit method) and Use of Random Number Tables using some examples.

- Some fundamental properties possessed by Simple Random Sampling without Replacement scheme.
- The estimation of parameters, like, population mean, population total and population variance on the basis of Simple Random Sampling without Replacement sample, along with the properties of estimators of these parameters.
- The sampling variance of the estimator of population mean.
- An unbiased estimator of the sampling variance of the mean estimator.
- A comparative study between Simple Random Sampling with Replacement and Simple Random Sampling without Replacement schemes in respect to their efficiency
- Simple Random Sampling without Replacement scheme provides estimator which is more efficient than that under Simple Random Sampling with Replacement.

2.8 TERMINAL QUESTIONS

1. Describe the different steps which are used in the selection process of a Simple Random Sampling without Replacement sample.

2. Let a finite population consist of marks of 20 students in Physics out of 100 marks in the annual examination. In order to estimate the average marks of these students in Physics, it was decided to select a random sample of marks of size 7 using Simple Random Sampling without Replacement scheme. Using the random number tables, explain the method of selection of 7 units in the sample.

3. How many Simple Random Sampling without Replacement samples of size 3 can be selected from a population of size 6? Let the population be U: {A, B, C, D, E, F}, then write all the Simple Random Sampling without Replacement samples of size 3.

4. The weights of 5 students are given below:

S. No	1	2	3	4	5
Weight (kg)	56	49	66	64	55

(a) Calculate population mean, population mean square and population variance.

(b) Enumerate all possible samples of size 2 without replacement and show that

(i) Sample mean $(\bar{y})$ is unbiased estimate for population mean $\bar{Y}$

(ii) Sample mean square (s^2) is unbiased estimate of population mean square (S^2).

(c) Calculate the variance of sample mean $(\bar{y})$.

2.9 ANSWERS / SOLUTIONS

Self-Assessment Questions (SAQs)

1. **Hint:** See the Section 2.1 for your answer.
2. **Hint:** See the **Theorem 1** of the unit for the answer to the question.
3. **Hint:** The answer to this question is based on the **Theorem 7** and **Remark 2.8**.
4. Population is U: {5, 45, 60, 35, 10, 22}. Here N = 6 and n = 3.

   We shall select a SRSWOR sample of size 3 out of this population. Starting from the first row, moving column-wise and taking only the unit place digit of the random numbers, we see that random numbers 3436, 6133, 9853 are selected, but since digit 3 is repeating in the last two numbers, we ignore the last number 9853 and move further for the next random number, which is 5807, but since 7 >N (=6), so this number is also avoided.

   The next random number in the sequence is 6291, so label 1 is selected. Thus, the selected labels are {6, 3, 1}.

   Therefore, the values of these units will be considered for finding the estimate of the population mean. The selected values of the variable are observed as {22, 60, 5}.

   The sample mean is

$$\bar{y} = \frac{(22+60+5)}{3} = 29$$

   The population mean $\bar{Y}$ is computed as

$$\bar{Y} = \frac{(5+45+60+35+10+22)}{6} = \frac{177}{6} = 29.6$$

   The estimated value, therefore, seems to be very close to the actual value of the parameter.

   The variance of the estimate will be equal to $\frac{N-1}{Nn}S^2$. Computing,

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{6} Y_i^2 - \frac{N}{N-1}\bar{Y}^2$$

$$= \frac{1}{5}\left[\left(5^2+45^2+60^2+35^2+10^2+22^2\right) - 6\times(29.5)^2\right]$$

$$= \frac{1}{5}\left[7459 - 5221.5\right] = \frac{2237.5}{5} = 447.5$$

   and substituting it in the expression

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

$$= \frac{(6-1)}{6\times 3}\times 447.5 = 124.30$$

   the variance of the estimate can easily be obtained.

5. We are given that n = 30 and N = 1500

Sample mean, $\bar{y} = \frac{1}{n}\sum_{i}^{30} y_i = \frac{1}{30} \times 104 = 3.467$

For calculating the standard error, we need to calculate

$$\sum_{i}^{30} y_i^2 = \begin{matrix} 25+36+9+9+4+9+9+9+16+16+9+4 \\ 49+16+9+25+16+16+9+9+16+9+9+1 \\ +4+16+9+16+4+16 \end{matrix}$$

$$\sum_{i}^{30} y_i^2 = 404$$

Then, sample mean square:

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

$$= \frac{1}{30-1}\left[404 - 30 \times (3.467)^2\right]$$

$$= \frac{1}{29}\left[404 - 360.60\right] = \frac{43.397}{29} = 1.496$$

Therefore,

$$S.E.(\bar{y}) = \sqrt{\frac{(N-n)}{N}\frac{s^2}{n}}$$

$$= \sqrt{\frac{1470 \times 1.496}{45000}} = \sqrt{0.0489} = 0.2211$$

Terminal Questions (TQs)

1. **Hint:** For the answer of the first part of the question, see the sub-sections 2.2.1 or 2.2.2.

2. In the second part, the population size is 20 and sample size is 7. Using the Random Number Table given in the Appendix – A of the previous unit, let us select a SRSWOR sample. Let us start with the fourth row of the table and move row-wise. Since, the population size is a two-digit number, we consider the right-most two digits of each random number. We see that random numbers consisting right-most two digits in the fourth row and subsequent rows, which are less than or equal to 20 are 5807, 5001, 0004, 9314, 8803, 9516 and 2010. Therefore, we observe that units with labels {**7**, **1**, **4**, **14**, **3**, **16**, and **10**} will be included. We also observe that no unit is repeated in the sample, so it is a SRSWOR sample.

3. We know that total number of samples in SRSWOR with population size N and sample size n is $^{N}C_{n}$.

In the question, N = 6 and n = 3; therefore, there will be $^{6}C_{3}$ = 20 samples. Since, the population is U : {A, B, C, D, E, F}, the samples of size 3 will be as follows:

{A, B, C}; {A, B, D}; {A, B, E}; {A, B, F}; {A, C, D}; {A, C, E}; {A, C, F};{A, D, E}; {A, D, F}; {A, E, F}; {B, C, D}; {B, C, E}; {B, C, F}; {B, D, E}; {B, D, F}; {B, E, F}; {C, D, E}; {C, D, F}; {C, E, F};{D, E, F}.

4. (a) The population mean of all the 5 students is given by

$$\bar{Y} = \frac{56+49+66+64+55}{5} = 58\,\text{kg}$$

The population mean square of the weights of the students is given by

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

$$= \frac{1}{4}\left[\sum_{i=1}^{N} Y_i^2 - 5\bar{Y}^2\right]$$

$$= \frac{1}{4}\left[\left\{(56)^2 + (49)^2 + (66)^2 + (64)^2 + (55)^2\right\} - 5\times(58)^2\right]$$

$$= \frac{1}{4}\left[\left\{3136 + 2401 + 4356 + 4096 + 3025\right\} - 5\times 3364\right]$$

$$S^2 = \frac{1}{4}\left[17014 - 16820\right] = \frac{194}{4} = 48.5\,\text{kg}$$

The population variance of the weights of the students is given by

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

$$= \frac{1}{5}\left[\sum_{i=1}^{N} Y_i^2 - 5\bar{Y}^2\right] = \frac{1}{5}\left[17014 - 16820\right] = \frac{194}{5} = 38.8\,\text{kg}$$

(b) Let us enumerate possible samples of size 2, i.e., ${}^5C_2 = 10$ sample in case of with replacement:

S. No	Sample	Sample Mean $(\bar{y})$	Sample Mean Square	$(\bar{y})^2$
1	(56, 49)	52.5	$(56-52.5)^2 + (49-52.5)^2 = 24.50$	2756.25
2	(56, 66)	61.0	50.00	3721.00
3	(56, 64)	60.0	32.00	3600.00
4	(56, 55)	55.5	0.50	3080.25
5	(49, 66)	57.5	144.50	3306.25
6	(49, 64)	56.5	112.50	3192.25
7	(49, 55)	52.0	18.00	2704.00
8	(66, 64)	65.0	02.00	4225.00
9	(66, 55)	60.5	60.50	3660.25
10	(64, 55)	59.5	40.50	3540.25
	Total	580.0	485.00	33785.50

(i) The mean of all possible sample means

$$E(\bar{y}) = \frac{1}{{}^5C_2}\sum_{i}^{10}\bar{y}_i = \frac{580}{10} = 58 = \bar{Y}$$

Thus, sample mean is an unbiased estimate of population mean.

(ii) The average of 10 sample mean squares

$$s^2 = \frac{1}{10}\sum_{i}^{10}\left(y_i - \bar{y}_i\right)^2 = \frac{485}{10} = 48.5\,\text{kg}$$

Thus, sample mean square is also an unbiased estimate of population mean square.

$$E\left(s^2\right) = S^2 = 48.5$$

(c) Therefore,

$$V(\bar{y})_{WOR} = \frac{(N-n)}{N.n}S^2 = \frac{(5-2)}{5\times 2}\times 48.5 = 14.55$$

Also, by definition,

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

$$= \frac{33785.50}{10} - 3364 = 3378.55 - 3364 = 14.55$$

and

$$V(\bar{y})_{WR} = \frac{\sigma^2}{n} = \frac{38.8}{2} = 19.4$$

Since, $V(\bar{y})_{WOR} \leq V(\bar{y})_{WR}$

Simple Random Sampling without Replacement gives a better estimator of population mean than Simple Random Sampling with Replacement.