1. Introduction

Welcome to the series of e-learning modules on Simple Random Sampling With and Without Replacement and properties. In this module we are going cover the basic concepts of simple random sampling, two types of simple random sampling, probability of selection of samples, its merits and demerits and related results.

By the end of this session, you will be able to explain:

- Simple Random Sampling (SRS) with and without replacement
- Number of possible samples
- Probability of selection of samples
- Properties and related results
- Merits and demerits of SRS

In statistics, simple random sample is a subset of *individuals* (a sample) chosen from a *larger set* (a population).

Each individual is chosen *randomly* and entirely by chance, such that each individual has the same *probability* of being chosen at any stage during the sampling process, and each subset of n individuals has the same probability of being chosen for the sample as any other subset of n individuals.

In other words,
Simple random sampling is a method of probability sampling in which every unit has an equal non-zero chance of being selected.
Here, we should note that this process should not be confused with random sampling.

Now let's get a basic idea about Simple Random Sampling.

- "Random" refers to the method of selecting a sample rather than to the particular sample selected
- Refers to the process rather than the outcome of the process
- Random selection process determines the selection probability
- The inverse of the selection probability plays a key role in linking sample data to the population quantities

Simple random sampling refers to a sampling method that has the following properties.

- The population considered for a simple random selection consists of N objects, and the sample consists of n objects
- All possible samples of n objects are equally likely to occur
- Simple random sampling is a basic type of sampling, since it can be a component of other more complex sampling methods
- The principle of simple random sampling is that every object has the same possibility to be chosen

2. Ways to Draw a Sample

There are two ways to draw a sample.

1) Sampling with Replacement and
2) Sampling without Replacement

Sampling with replacement means that once a person is selected to be in a sample, that person is placed back in the population to possibly be sampled again.
In short, in this case duplicated selection is allowed.
Although simple random sampling can be conducted with replacement instead, this is less common and would normally be described more fully as simple random sampling '*with replacement*'.

Here, without replacement means that once an individual is sampled, that person is not placed back in the population for re-sampling.
In other words, here duplicated selection is not allowed.
In small populations and often in large ones, such sampling is typically done '*without replacement*'. One deliberately avoids choosing any member of the population more than once. Sampling done without replacement is no longer independent, but still satisfies exchangeability, and hence many results still hold.

Further, for a small sample from a large population, sampling without replacement is approximately the same as sampling with replacement, since the odds of choosing the same individual twice is low.
Two definitions of simple random sample:

1) Each element has an equal chance of being selected
2) Each possible sample has an equal chance of being selected
   The first definition appears to be valid in with replacement sampling.

But is it valid for without replacement sampling?

Now we shall look at an example where we have 'N' college students want to get a ticket for a basketball game, but there are not enough tickets 'X' for them, so they decide to have a fair way to see who gets to go. Then, everybody is given a number (0 to N-1), and random numbers are generated, either electronically or from a table of random numbers. Non-existent numbers are ignored, as are any numbers previously selected. The first X numbers would be the lucky ticket winners.

First, we shall consider a case of taking Without Replacement Sample of n is equal to 2 from a population N is equal to 3 - a, b and c.
The probability of selecting any element in the first draw is 1/3.
What is the probability of selecting any element in the second draw?

You would think that it is 1 by 2 or say fifty percent, because one is already selected. You have to realize that 1 by 2 is a conditional probability since a second draw of any particular element is possible only when that element is not selected in the first draw.
The probability of not selecting any particular element in the first draw is one minus one by three which is equal to two by three, when you multiply one by two to two by three, we get

one by three, suggesting that the probability of selecting any element at each draw is the same.

Now, let’s find out if the selection probability is the same with both of sampling procedures?

Now in the next example, consider the case of selection two of three samples a, b and c.

With the without replacement sampling there are three possible samples. They are (a , b), (a , c) and (b , c)

Note that each is included two times in three possible samples.

The probability that any element is included in the sample is two by three. In general it is n by N, a sample ratio.

Now, in the ‘with replacement’ sampling there are 9 possible samples.

These are (a , a) (a , b) (a , c) (b , a) (b , b) (b , c) (c , a)(c , b) and (c , c). Here, note that each is included 6 times in 9 possible samples. The probability that any element is included in the sample is 6 by 9 which is equal to 2 by 3.

Here too, the sampling ratio is n/N.

‘With replacement’ method is equivalent to ‘without replacement’ method when we take samples from a large population.

3. Examples

Here is another example where we need to make a selection of 3 addicts from a population of 9. Since there are 3 persons in the sample, the selection procedure has three steps.

Step 1: Selection of the first sampled subject

Step 2: Selection of the second sampled object

Step 3: Selection of the third sample object

In sampling without replacement the selection process is the same as at step 1. That is each addict in the population has the same probability of being selected.

At step 2, however, the situation changes.

Once the first addict is chosen, he is not placed back in the population.

Thus at step 2, the second addict to be sampled comes from the remaining 8 addicts in the population, all of who have the same probability of being selected, that is, 1 in 8.

At the third step, the selection is derived from a population of 7 addicts, with each addict having a probability of 1 in 7 of being selected. Once the steps are completed, the sample contains 3 different addicts.

Unfortunately, the reduced selection probability from the first to the third step is at odds with statistical theory for deriving the variance of the sample mean.

Such theory assumes the sample was selected with replacement. Yet in practice, most simple random samples are drawn without replacement, since we want to avoid the strange assumption of one person being tallied as two or more. To resolve this disparity between statistical theory and practice, the variance formulas used in simple random sampling are somewhat changed.

When drawing a sample from a population, there are many combinations of people that could be selected.

- To calculate the number of possible samples that can be drawn without replacement, disregarding order, Combination N.C.N is equal to N factorial divided by N minus n factorial multiplied by n factorial.

Where, N is the number of people in the population, n is the number of sampled persons, and the exclamatory symbol is the factorial notation for the sequential multiplication of N times a number minus 1, continuing until it reaches 1. That is, "N factorial" is N times N minus one times N minus two and the like with the last number being one.

In our example, we are selecting without replacement and disregarding the order from a sample of three addicts from a population of nine addicts.
Using Formula we find there are:
9 factorial by nine minus three factorial into three factorial which is equal to 84. That is 84 possible samples.

In the realistic world of sampling, subjects are typically not included in the sample more than once. Also, the order in which subjects are selected for a survey is not important. That is, Roy-Sam-Ben is considered the same as Sam-Ben-Roy.
Hence in most surveys, samples are selected disregarding the order and without replacement.

In sampling with replacement all 9 addicts have the same probability of being selected at steps 1, 2 and 3, since the selected addict is placed back into the population before each step.
With this form of sampling, the same person could be sampled multiple times. In the extreme, the sample of 3 addicts could be one person selected three times.

To calculate the number of possible samples that can be drawn with replacement – Permutation, N to the power n, where N is the number in the total population and n is the number of units being sampled.

For example when selecting three persons from the population of nine addicts the sample could have been Joe-Jon-Hall, or Sam-Bob-Nat, or Roy-Sam-Ben, or any of the many other combinations. To be exact, in sampling with replacement from the population there are N to the power n which is equal to nine to the power three equals to seven hundred and twenty nine. That is, seven hundred and twenty nine different combinations of three addicts that could have been selected.

In Simple Random Sampling, if the units selected at previous stage are replaced back into the population before drawing a next unit then, such a technique is called Simple Random Sampling With Replacement.
Here we have to let the population size be N and the Sample size be n. Under SRSWR we can have N to the power n possible samples. The probability of selecting a sample is one by N to the power n.

4. Theorems and Properties of Simple Random Sampling

Here is a theorem which we shall prove.
The theorem states that: In Simple Random Sampling With Replacement, the probability of selection of a unit at any stage in a sample is given by one by N.

Here we shall look at the proof for the theorem.
Let the population size be N. The probability of selecting a unit in the first draw is one by N. We replace the unit back to the population before the second draw hence there will again be N units. Therefore the probability of the unit selected at the second draw is one by N. Hence for every draw there will be N units in the population as the units are replaced. Therefore, the probability that a unit being selected at any stage is one by N.

In Simple Random Sampling if the units selected at previous stage are not replaced back into the population before drawing a next stage unit then such a technique is called Simple Random Sampling Without Replacement.
Here, let the population size be N and the Sample size be n. Under Simple Random Sampling Without Replacement we can have N.C.N possible samples.
The probability of selecting a sample is one by N.C.N.

The next theorem states that: In a Simple Random Sampling Without Replacement, the probability of selection of a unit at any stage in a sample is given by 1 by N.

Here is the proof. Now, let the population size be N. The probability of selecting a unit in the first draw is one by N.
Probability of unit being selected in the second draw equals to the probability of unit not selected in the first draw and probability of unit selected in the second draw, given that it is not selected in the first draw
Which is equal to (one minus one by N) multiplied by (one by N minus one) Which is equal to (N minus one by N) multiplied by (one by N minus one)
Equals to one by N.

Similarly, the probability of unit being selected in the third draw equals to the probability of unit not selected in the first draw multiplied by probability of unit not selected at the second draw given that it is not selected in the first draw multiplied by probability of unit being selected in the third draw given that it is not selected in the first and second draw.
This Equals to, one minus one by N multiplied by one minus one by N minus one multiplied by one by N minus two.
Equals to N minus one by N multiplied by N minus two by N minus one multiplied by one by N minus two.
Equals to one by N
Therefore, the probability of a unit being selected in any stage equals to one by N.

Here are the properties of Simple Random Sampling

- Sample mean is an unbiased estimate of the population mean under both Simple Random Sampling With Replacement and Simple Random Sampling Without Replacement.
- Sample Mean Square is unbiased for the population variance under Simple Random Sampling With Replacement
- Sample Mean Square is unbiased for the population Mean Square under Simple Random Sampling Without Replacement
- In case of sampling from a population containing attributes, sample proportion is an unbiased estimate of population proportion under Simple Random Sampling With Replacement and Simple Random Sampling Without Replacement

5. Advantages and Disadvantages of Simple Random Sampling

Now we shall look at the advantages and disadvantages of Simple Random Sampling.

1. The selection of items in the sample depends entirely on chance. So, there is no possibility of personal bias affecting the results.
2. As compared to non-random sampling, Simple Random Sampling represents the universe in a better way. As the size of the sample increases it becomes increasingly representative of the population.

3. Analyst can easily assess the accuracy of this estimate because sampling errors follow the principle of chance.

4. This method enables the analyst to provide the most reliable information at the least cost.

5. It is free of classification error, and it requires minimum advance knowledge of the population other than the frame.

6. Its simplicity makes it relatively easy to interpret the data collected.

For these reasons, Simple Random Sampling best suits situations where not much information is available about the population and data collection can be efficiently conducted on randomly distributed items, or where the cost of sampling is small enough to make efficiency less important than simplicity.

The limitations of Simple Random Sampling includes that:

1. It is often difficult for the investigator to have an up-to-date list of all the items of the population to be sampled. Hence, there is a restriction to the use of this method in economics and business data where very often we have to employ restricted random sampling designs.
2. The size of the samples required to ensure statistical reliability is usually larger under Simple Random Sampling.

3. In case of field surveys, using Simple Random Sampling tends to be too widely dispersed geographically and the time and cost of In case of field surveys, using Simple Random Sampling tends to be too widely dispersed geographically and the time and cost of collecting data becomes too large.

4. Random sampling may produce the most non random looking results**.**

To conclude let's quickly take a look at some important points to note.

Simple random samples are characterized by the way in which they are selected. Conceptually, simple random sampling is the simplest of the probability sampling techniques. It requires a complete sampling frame, which may not be available or feasible to construct for large populations.

Even if a complete frame is available, more efficient approaches may be possible if other useful information is available about the units in the population.

An unbiased random selection of individuals is important so that in the long run, the sample represents the population. However, this does not guarantee that a particular sample is a perfect representation of the population.

Simple random sampling merely allows one to draw externally valid conclusions about the entire population based on the sample.

Here's a summary of our learning in this session:

- Introduction to Simple Random sampling
- Types of Simple Random sampling
- Number of possible samples under SRSWR and SRSWOR
- Related results & Properties
- Merits and demerits of SRS