1. Introduction

Welcome to the series of E-learning modules on Methods of selecting a simple random sample. In this module we are going to cover the basic principle of Simple random sample, Lottery method, Random numbers method, Computer method, Selection from ungrouped population, Grouped population and Contingency table.

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

- Explain the basic principle of Simple random sample
- Explain Lottery method
- Describe the method of random numbers
- Explain the generation using computers
- Describe the selection from ungrouped, grouped populations and contingency tables

Simple Random Sample (SRS) is the most widely known Random Sample. This is characterized by the fact that the probability of selection is the same for every case in the population.

Simple random sampling is a method of selecting ‘n’ units from a population of size N such that every possible sample of size ‘n’ has equal chance of being drawn.

An example may make this easier to understand. Imagine you want to carry out a survey of hundred voters in a small town with a population of one thousand eligible voters. With a town of this size, there are "old-fashioned" ways to draw a sample.

We could write the names of all voters on a piece of paper, put all pieces of paper into a box and draw hundred tickets at random.

You shake the box, draw a piece of paper and set it aside, shake again, draw another, set it aside, etc. Draw until we had hundred slips of paper. These hundred slips of paper form our sample. And this sample would be drawn through a simple random sampling procedure - at each draw, every name in the box had the same probability of being chosen.

In the real-world social research, it is difficult to obtain the designs that employ simple random sampling.
We can imagine some situations where it might be possible – you want to interview a sample of doctors in a hospital about work conditions.

So you get a list of all the physicians that work in the hospital, write their names on a piece of paper, put those pieces of paper in the box, shake and draw. But in most real-world instances it is impossible to list everything on a piece of paper and put it in a box. In this case we have to randomly draw numbers until desired sample size is reached.

2. Selection of Simple Random Sample

Selection of Simple Random Sample:
There are many methods to proceed with simple random sampling. The most primitive and mechanical would be the lottery method.

1. Lottery Method: This is an old classical method but it is a powerful technique. Modern methods of selection are very close to this method. Each member of the population is assigned a unique number. All the units of the population are numbered from 1 to N. This is called sampling frame. These numbers are written on the small slips of paper or the small round metallic balls. The paper slips or the metallic balls should be of the same size otherwise the selected sample will not be truly random.

Each number is placed in a bowl or a hat and mixed thoroughly. The blind-folded researcher then picks numbered tags from the hat or a bowl. All the individuals bearing the numbers picked by the researcher are the subjects for the study.

Again the population of slips is mixed and the next unit is selected. In this manner, the number of slips equal to the sample size n is selected. The units of the population which appear on the selected slips make the simple random sample. This method of selection is commonly used when size of the population is small. For a large population there is a big heap of paper slips and it is difficult to mix.

In a simple random sample, one person must take a random sample from a population, which does not have any order in which one chooses the specific individual. Let us assume you had a school with thousand students, divided equally into boys and girls, and you wanted to select 100 of them for further study. You might put all their names in a bucket and then pull hundred names out. Not only does each person have an equal chance of being selected, we can also easily calculate the probability of a given person being chosen, since we know the sample size (n) and the population (N).

Such a procedure is obviously impractical; if not altogether it is impossible to complex problems of sampling. In fact the practical utility of such a method is very much limited. The lottery method is quite time consuming and cumbersome to use if the population is sufficiently large.

2. “Mechanical randomization” or “Random Numbers “Method. The Lottery Method of selecting a sample becomes tedious when the population is large or when the large sample is needed. Fortunately we can take a random sample in a relatively easier way without taking the trouble of enlisting all possible samples on paper slips as explained above. In such cases we use random numbers to draw samples. Random numbers are the numbers obtained by random sampling and recorded by the Statisticians. Here, we use Fisher and Yates Table.

3. Procedures for Selecting Random Samples

We have three procedures for selecting random samples for many populations.

1) Ungrouped Data
2) Grouped Data (Frequency Distribution)
3) Contingency Table

Ungrouped data

All the units of the population are numbered from 1 to N or from 0 to N-1. We consult the random number table to take a simple random sample. Suppose the size of the population is eighty and we have to select a random sample of eight units.

The units of the population are numbered from zero one to eighty. We read two-digit numbers from the table of random numbers. We can take a start from any columns or rows of the table. Let us consult random number table given in this content. Two-digit numbers are taken from the table. Any number above eighty will be ignored and if any number is repeated, we shall not record it if sampling is done without replacement.

Let us read the first two columns of the Fisher and Yates table. The random number from the table is ten, thirty-seven, zero eight, twelve, sixty-six, thirty-one, sixty-three and seventy-three. The two numbers ninety-nine and eighty-five have not been recorded because the population does not contain these numbers. The units of the population whose numbers have been selected constitute the simple random sample.

Let us suppose that the size of the population is hundred. If the units are numbered from zero zero one to hundred, we shall have to read 3-digit random numbers. From the random number table, the random numbers are hundred, three seventy five, zero eight four, nine nine zero and one two eight and so on.

We find that most of the numbers are above 100 and we are wasting our time while reading the table. We can avoid it by numbering the units of the population from zero zero to ninety-nine. In this way, we shall read 2-digit numbers from the table. Thus, if N is hundred, thousand or ten thousand, the numbering is done from zero zero to ninety-nine, zero zero zero to nine nine nine or zero zero zero zero to nine nine nine nine.

The method of drawing random sample consists in the following steps:

(i) Identify the N units in the population with the numbers from 1 to N
(ii) Select at random any page of the random number tables and pick up the numbers in any row or column or diagonal at random
(iii) The population units corresponding to the numbers selected in step (ii) constitute a random sampleHere are the different sets of random numbers commonly used in practice. The numbers in these tables have been subjected to statistical tests for randomness of a series and their randomness has been well established for all practical purposes.

Trippet's (nineteen twenty seven) random numbers tables. (Tracts for computers, No. Fifteen, Cambridge University Press).

a) Trippet Number Table - 10,400 4 digited random numbers, giving in all 41,600 digits selected at random from the British Census Reports.

b) Fisher and Yates (nineteen thirty eight) tables (in Statistical tables for biological, agricultural and medical research) comprise fifteen thousand digits arranged in two's. Fisher and Yates obtained these tables by drawing numbers at random from the tenth to nineteenth digits of A. S. Thomson's twenty figure logarithmic tables

c) Kendoll and Babington Smith's (nineteen thirty nine random tables consists of one lakh digits grouped into twenty five thousand sets of 4 digited random numbers
d) Random Corporation (nineteen fifty five) (Free Press, Illinois) random number table consists of one million random digits.

Suppose we want to draw a random sample of size fifty from the population having five thousand five hundred fifty five units.

We take any column of random numbers from Fisher & Yates Table.

Go on copying 4 digited random numbers (since the population size five thousand five hundred fifty five has 4 digits) leaving zero and numbers greater than five thousand five hundred fifty five. We copy such fifty numbers.

If we have numbers which are already copied then the scheme is called with Replacement (WR) scheme. If we do not copy the numbers which are already copied then the scheme is called Without Replacement (WOR) scheme. If there is a large percent of rejection, to avoid the draw we slightly modify our procedure. We divide the random numbers selected by the given population i.e., population size and make the remainder to be selected.

4. Grouped Data

Grouped data (Frequency Distribution)
We follow the following procedure for the selection of the sample from grouped data or frequency distribution.

1) We first cumulate the given frequencies.
2) Corresponding to the cumulative frequencies the inclusive class intervals are formed.
3) The required sample is obtained by drawing random numbers from Fisher & Yates Table.
4) Same procedure is applied as that of ungrouped data.

Example:
Draw a Random sample of size fifteen from the following frequency distribution

Figure 1

C.I	0-5	5-10	10-15	15-20	20-25	25-30	30-35
Frequency	10	12	13	8	14	16	7

Solution is:
Draw a table consisting of following columns.

Figure 2

C. I	Frequency	L.T.C.F	Inclusive C.I's	Tally Marks	Frequency
0-5	10	10	00-10	///	3
5-10	12	22	11-22	/	1
10-15	13	35	23-35		0
15-20	8	43	36-43	///	3
20-25	14	57	44-57	//	2
25-30	16	73	58-73	~~////~~	5
30-35	7	80 = N	74-80	/	1

Column1: consists of class intervals or the variable values given in the population.
Column 2: consists of frequencies or number of observations taking particular values of the variables in the population.
Column 3: Obtain the cumulative frequencies. As in this case they are ten, twenty two, thirty five, forty three, fifty seven, seventy three and eighty.
Column 4: Obtain the Inclusive Class intervals for the cumulative frequencies obtained in the third column like zero zero to ten, eleven to twenty two, twenty three to thirty five, thirty six to forty three, forty four to fifty seven, fifty eight to seventy three and seventy four to eighty. All limits are two digited because total frequency N equals to eighty is a two digited number.
Column 5: Therefore, we select two digited fifteen random numbers leaving the numbers zero zero and the numbers greater than eighty under WOR scheme from the Fisher & Yates table The random number are:
Zero Three, forty nine, forty three, thirty six, fifty six, seventy, sixty eight, fifty nine, thirteen, forty, seventy one , seventy eight, Zero nine, fifty eight and ten
Select first random number zero three put a tally mark against the interval zero zero to ten. Like this continue the process until you reach a fifteenth random number ten which falls in the interval zero zero to ten.
Column 6: Count the total number of tally marks of each intervals. Observe the total of the column 6 must be equal to fifteen, a required sample size.

Hence the required sample is:

Figure 3

C.I	0-5	5-10	10-15	15-20	20-25	25-30	30-35
Frequency	3	1	0	3	2	5	1

5. Selection of Sample from Contingency Table

Selection of the sample from Contingency Table:

1) Cumulate the frequencies of the first, second, ……, last Rows.
2) Inclusive class Intervals corresponding to the cumulative frequencies are formed.
3) Apply same procedure as that of Ungrouped data
4) Required sample is obtained by drawing random numbers from Fisher & Yates Table.

Example:
Draw a Random sample of size 10 from the following contingency Table.

Figure 4

Sex Class	**Male**	**Female**
A	123	153
B	145	303

Figure 5

Sex Class	**Male**	**Female**
A	123 000-123 /	153 124-276 ~~////~~
B	145 277-421 /	303 422-724 ///

Cumulative Frequencies, Inclusive Class Intervals and tally marks are obtained as follows:
Cumulative frequencies are one twenty three, two seventy six, four twenty one and seven

twenty four.
Since total frequency is seven twenty four we select three digited ten random numbers leaving the numbers zero zero zero and numbers greater than seven hundred and twenty four under WOR scheme from Fisher & Yates table from page one thirty four, Column one.
The random numbers are:
Zero three four, one sixty seven, one twenty five, five fifty five, one sixty two, six thirty, three thirty two, five seventy six, one eighty one and two sixty six.
Procedure is same as ungrouped data. In this case for each cell of the table inclusive class intervals are constructed and then mark tally marks.

Therefore the required sample is:

Figure 6

Sex Class	Male	Female
A	1	5
B	1	3

Using the Computer
Another way would be to let a computer to do a random selection from your population. For populations with a small number of members, it is advisable to use the first method but if the population has many members, a computer-aided random selection is preferred.
The facility of selecting a simple random sample is available on the computers. The computer is used for selecting a sample of prize-bond winners, a sample of Hajj applicants, and a sample of applicants for residential plots and for various other purposes.

A ***simple random sample*** is similar to putting the names of all students in a hat, and then drawing two hundred tickets **without replacement**. More practical procedure is to create a **frame***,* or a list, of all WMU students and have a computer randomly select two hundred names from the frame.

On a smaller scale, one may use a **table of random numbers to** select subjects from the frame. The method of simple random sampling has ideal statistical properties in the sense that each element in the frame has an equal chance of being selected.

However many target populations have no convenient frames available. Examples are the populations of Kalamazoo residents, customers of Meijer Stores, potential credit card customers, and U.S. registered voters. In these cases, more convenient frames that simply approximate target population may be used, like phone books and mailing lists.

One may note that it is easy to draw random samples from finite population with the aid of random number tables only when the lists are available and the items are readily numbered. But in some situations it is often impossible to proceed in the way we have narrated above.

For example: If we want to estimate the mean height of trees in a forest, it would not be possible to number the trees and choose random numbers to select random samples. In such situations what we should do is to select some trees for the sample using appropriate sample technique and should treat the sample as a random sample for study purpose.

Random sample from a infinite universe:
So far we have talked about random sampling keeping in view only the finite populations. But what about random sampling in context of infinite populations?

It is relatively difficult to explain the concept of random sample from an infinite population. However few examples will show the basic characteristic of such a sample. Suppose we consider the twenty throws of a fair dice as a sample from the hypothetically infinite population which consists of the results of all possible throws of the dice. If the probability of getting a particular number say 1, is the same for each throw and the twenty throws are all independent then we say that the sample is random.

Similarly, it would be said to be sampling from an infinite population if we sample with replacement from a finite population. In addition to that our sample would be considered as a random sample if in each draw all elements of the population have the same probability of being selected and successive draws happen to be independent. In brief one can say that selection of each item in a random sample from an infinite population is controlled by the same probabilities and that successive selections are independent of one another.

There are many reasons why one would choose a different type of probability sample in practice.
Although simple random sampling is the ideal for social science. Most of the statistics used are based on assumptions of SRS, in practice, SRS are rarely seen.

It can be terribly inefficient, and particularly difficult when large samples are needed. Other probability methods are more common. Yet SRS is essential, both as a method and as an easy-to-understand method of selecting a sample.

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

- Basic principle of simple random sample
- Lottery method
- Method of random numbers
- Selection using computers
- Selection of samples from Ungrouped, grouped population and contingency table