1. Introduction

Welcome to the series of e-learning modules on Stratified Random Sampling with Simple Random Sampling With Replacement and Simple Random Sampling Without Replacement. In this module we are going cover the basic principle of stratified random sampling, the need for stratification, merits and demerits of the method, stratified random sampling with Simple Random Sampling With Replacement and Simple Random Sampling Without Replacement.

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

- Stratified Random Sampling
- Principal stages of the method
- Merits and demerits of the technique
- Stratified sampling with Simple Random Sampling With Replacement and Simple Random Sampling Without Replacement

Any result produced from a sample can be used to estimate the corresponding result for the population.
It is absolutely essential that the sample taken is as representative as possible of that population. Common sense rightly suggests that, the larger the sample the more representative it is likely to be and also the more expensive it is to take and analyze.

A random sample is ideal for statistical analysis but, for various reasons, other methods also have been devised for when this sampling is not feasible.
Stratified Random Sampling:
Very often the population that is investigated is not homogeneous. In order to make the sample more representative in such situations, we would also like to consider different sections of the population which are homogeneous within themselves. In these cases the stratified random sampling procedure is used.

The population is divided into a number of sections called strata based upon one or more classification criteria.
Within each stratum, a simple random sample is taken independently from the members of the subpopulation.
The number to be sampled from each stratum depends on its size relative to the population.

Since a random sample is taken from each stratum the whole population is adequately represented.
Here, the basic idea is that the heterogeneous population is divided into several sectors each of which is more or less homogeneous.
Thus the variability in each stratum that exists in the population is adequately represented in the sample as well.

Here, if a classification system results in 3 subgroups or strata, and Group A has 50% of the population, and Group B and Group C have 25% each, the sample we draw must conform to the same relative sizes (half of the sample from A, a quarter each from B and C).
The samples taken from each strata are then pooled together to form the overall sample.

The precision of an estimate depends on its sampling variance which can be achieved by increasing the sample size or sampling fraction n by N.
Only other way of increasing the precision of sample mean is to frame a sampling technique which will decrease s square.
One such technique is known as Stratified Sampling.
Under simple random sampling without replacement, variance of the sample mean depends on ‘n’ and s square

If s square is large, Variance of y bar is large. To have smaller variance we can have larger sample ‘n’.
But it is not always possible to increase the sample size. Hence, to have smaller variance, we go for another new technique called Stratified Random Sampling.
Stratification means division into layers.

Strata may defined as any group:

Credit card users versus noncredit card users, by gender, age, industry, purchasers Vs non-purchasers, current customers versus past customers, etc.

Once the strata are identified, a simple random sample is drawn within each stratum.

Once the survey is completed, the strata are then weighted back to the population proportions.

2. Characteristic & Requirements of Stratified Sampling

Stratified sampling involves classifying the population into categories and then choosing a sample which consists of participants from each category in the same proportions as they are in the population.

For Example: If you wanted to carry out a stratified sample of students from a college you might decide that important variables are sex, 1st or 2nd years, age, have a part-time job and so on.

You could then identify how many participants there are in each of these categories and choose the same proportion of participants in these categories for your study.

The strength of stratified sampling is therefore that your sample should be representative of the population.

Stratified sampling can be very time consuming as the categories have to be identified and calculated.
As with random sampling, if you do not have details of all the people in your target population you would struggle to conduct a stratified sample.
If the sample is not randomly selected from the categories it is then called a quota sample.

Stratified Random sampling is a restricted random method, which, by using the available information concerning the population, attempts to design a more efficient sample than obtained by the simple random procedure.

We shall now have a look at the procedure followed while applying stratified random sampling technique.
Step 1: The universe to be sampled is subdivided into groups which are mutually exclusive and includes all items in the universe
Step 2: A simple random sample is then chosen independently from each group

This sampling procedure differs from simple random sampling in that in simple random sampling the sample items are chosen at random from the entire universe, and the distribution of the sample among strata is left entirely to chance.

While in stratified random sampling the sampling is designed so that a designated number of items is chosen from each stratum.

3. Requirements involved in setting up a Stratified Random Sample

Some of the requirements involved in setting up a stratified random sample are:
One, Base of Stratification
Two, Number of strata and
Three, Sample size within strata.
We shall look into the details of the first requirement: Base of Stratification.
Here the question arises as to what characteristic should be used to subdivide the universe into different strata?
As a general rule, strata are created on the basis of a variable known to be correlated with the variable of interest and for which information on each universe element is known.
Strata should be constructed in a way which will minimize the differences among sampling units within the strata and maximize the difference among strata.

- For example: Suppose we are interested in studying the consumption pattern of the people of Delhi, the city of Delhi may be divided into various parts, such as zones or wards, and from each part a sample may be taken at random.
- Before deciding about stratification we must have a knowledge of the traits of the population.
- Such knowledge may be based upon expert judgment, past data and preliminary observations from pilot studies etc.

The purpose of stratification is to increase the efficiency of sampling by dividing a heterogeneous universe in a way such that:

- There is as much homogeneity as possible within each stratum, and
- A marked difference is possible between the strata.

The second requirement is in finding the number of strata to be constructed.

The practical considerations limit the number of strata that is feasible. The costs of adding more strata may soon outrun benefits.

In general, more than six strata may be undesirable.

The third issue is the sample size within the strata. The question we need to ask here is ‘How many observations should be taken from each stratum?

When deciding on this question we can consider either a proportional or a disproportional allocation.

In proportional allocation one sample from each stratum, in-proportion to its relative weight is considered. In disproportional allocation this is not the case.

With either approach, stratified sampling guarantees that every element in the population has a chance of being selected.

It may be pointed out that, proportional allocation approach is simple and if all that we know about each stratum is the number of items in that stratum, it is generally the preferred procedure.
In disproportional sampling the different strata are sampled at different rates.
As a general rule when variability among observations within a stratum is high, we sample that stratum at a higher rate than for strata with less internal variation.

Stratification is most useful when the stratifying variables are:

- Simple to work with
- Easy to observe, and
- Closely related to the topic of survey

Stratified sampling is appropriate when the population is already divided into groups of different sizes and we wish to acknowledge this fact.

Suppose, a physician's patients are divided into four groups according to age as shown in the table below.

Figure 1

	Age group	%ge of total
Composition of patients by age	Birth -19 years	30
	20-39 years	40
	40-59 years	20
	60 years and older	10

We have the composition of patients by age. The age groups we classify are: Birth to 19 years, and corresponding percentage of total is 30, 20 to 29 years, 40 percent, 40 – 59 years is 20 percent and older than 60 years is 10 percent.
The physician wants to find out how many hours his patients sleep.

To obtain an estimate of this characteristic of the population he could take a random sample from each of the four age groups and give weight to the samples according to the percentage of patients in that group. This would be an example for stratified sample.

A non-homogeneous population is sub-divided into various non-overlapping sub-populations called strata. When these have been determined, a sample is drawn from each, with a separate draw for each of the different strata.

A sample obtained by combining all these units is called a Stratified

Random Sample, and the technique is known as Stratified Random Sampling.

Stratified survey sampling enables you to focus on specific groups, say for example: women or rural people, ensuring that they will be represented in the sample.

Although random survey sampling, if done correctly, will give the researcher roughly proportional samples of all groups, disproportional stratified sampling will guarantee that a certain group is adequately represented.

Stratified random sampling gives more precise information than simple random sampling for a given sample size.
So, if information on all members of the population that divides them into strata that seems relevant is available, stratified sampling will usually be used.
One of the simplest extensions of simple random sampling involves the use of stratification of the population of interest.

4. Need for Creating a Strata & Merits of Stratified Random Sampling

Now we shall look at the need for creating a strata.

There are primarily four reasons for using stratification.

The first reason is that the sub-populations that form the strata are the domains of study or interest and thus, we need to get separate estimates for each stratum.

For example: We may take separate samples for men and women because we are interested in making separate inferences for the said sub-populations, as well as for comparing them.

Secondly, The subpopulations or strata are such that different methods or sampling schemes may be suitable for different strata.

For example: If we are estimating the number of words in a text book we might take different samples from different sections, say the cover page text, introduction text, the main text, the references, the index etc.

Another need to create a strata is due to the fact that, the population is geographically diverse and the stratification is more suitable for groups of common characteristics .Thus, stratification is a convenient way of organizing sampling and data collection.

Lastly, by employing stratification in a careful manner, we can get improved precision for our estimates of population quantities.
This is the idea behind the efficiency gain obtained with stratification. If you create strata within which the units share similar characteristics, for example income, and are considerably different from units in other strata, example: occupation, type of dwelling, then you would only need a small sample from each stratum to get a precise estimate of total income for that stratum.

You could then combine these estimates to get a precise estimate of total income for the whole population.

At the same time, if you were to use a simple random sampling approach in the whole population without stratification, the sample would need to be larger than the total of all stratum samples to get an estimate of total income with the same level of precision.

The advantages of Stratified random sampling:

- This kind of sampling offers a more representative sample
- There is greater precision in the results
- It is convenient from an administration perspective
- The geographical concentration of such samples is greater.

Next, we shall look into the merits of this kind of sampling in detail.

1. More Representative sample:
By this, we mean that:
In unstratified random sampling method, some strata may be over represented, others may be underrepresented while some may be excluded altogether.
Stratified sampling ensures the desired representation in the sample of the various strata in the population.
It excludes the possibility of any essential group of population being completely excluded in the sample.

A properly constructed and executed stratified random sampling plan overcomes the drawback of purposive sampling and simple random sampling and still enjoys the virtues of both these methods.
It divides the given universe into a number of homogeneous sub-groups with respect to purposive characteristic and then uses the technique of random sampling in drawing samples from each stratum.

A stratified random sample gives adequate representation to each strata or important section of the population, eliminating the possibility of any important group of the population being completely ignored.
The stratified random sampling provides a more representative sample of the population and accordingly results in less variability as compared to other sampling designs.

2. <u>Greater Precision:</u>

As a consequence of reduction in variability within each stratum, stratified random sampling provides more efficient estimates as compared to simple random sampling.

For example: the sample estimate of a population mean is more efficient in both proportional and Neyman's allocation of the samples to different strata in stratified random sampling as compared to the corresponding estimates obtained by simple random sampling.

Sometimes it is desired to achieve different degrees of accuracy for different segments of the population.
Stratified random sampling is the only sampling plan which enables us to obtain the results of known precision for each of the stratum.

3. Administrative Convenience

The division of the population into relatively homogeneous sub-groups brings administrative convenience.

Unlike random samples, the stratified samples are expected to be localized geographically. This ultimately results in reduced costs and time saving in terms of collection of the data, interviewing the respondents and supervision of the field work.

4. Geographical Concentration

As compared to simple random sampling, the stratified samples would be more concentrated geographically. That is, the units from different strata may be selected in such a way that all of them are localized in one geographical area.

Accordingly time and money involved in collecting the data and interviewing the individuals may be considerably reduced and the supervision of the field work could be allotted with greater ease and convenience.

Sometimes the sampling problems may differ significantly in different parts of the population.

In such situations the problem can be tackled effectively through stratified sampling by regarding each segment of the population as a different strata and approaching upon them independently during sampling.

In such cases we can deal with the problem through Stratified Random Sampling.

5. Demerits of Stratified Random Sampling

Demerits of Stratified Random Sampling:

One. If stratification is faulty the results will be biased. The error due to wrong stratification cannot be compensated even by taking large samples

As already pointed out the success of the stratified sampling depends on:

- Effective stratification of the universe into homogeneous strata
- Appropriate size of the samples to be drawn from each stratum

Demerits of Stratified Random Sampling:

One. If stratification is faulty the results will be biased. The error due to wrong stratification cannot be compensated even by taking large samples.

If the strata are overlapping, unsuitable or disproportionate, the selection of the samples may be biased.

Such errors cannot be compensated even by taking large samples.

The second demerit is that:

The allocation of the sample sizes to different strata requires an accurate knowledge of the population size in each stratum ‘N-h’.

Neyman’s principle of Optimum allocation, ‘n-h’ proportional to ‘N-h,S-h’ requires additional knowledge of the variability or standard deviation of each strata.

‘N-h’ and ‘S-h’ usually being unknown, is a serious limitation for the effective use of stratified random sampling.

Thirdly, it is a very difficult task to divide the universe into homogeneous strata, and

Finally, disproportionate stratification requires weighting, which again introduces selective

factor in the sample and under weighting makes the sample unrepresentative.

Using stratified sampling, the population is divided into homogeneous, mutually exclusive groups called strata, and then, independent samples are selected from each stratum.
When Simple Random Sampling With Replacement is used to select the sample within each stratum, the sample design is called Stratified Random Sampling With SRSWR.

For example,
Suppose we want an estimate of the number of high school students who have part-time jobs at the national level and also in each province.
If we were to select a Simple Random Sampling With Replacement sample of 25,000 people from a list of all high school students in Canada (assuming such a list was available for selection), we would end up on an average with just a little over 100 people from Prince Edward Island, since they account for less than half of a percent of the whole Canadian population.

Stratifying your list by province, again assuming that this information is available, and then selecting a sample size for each province would allow you to decide on the exact sample size needed for that specific province.
Then, selecting the required samples from each province by Simple Random Sampling With Replacement will give us a more representative sample.

Thus, in order to get a good representation of Prince Edward Island, you would use a larger

sample than the one allotted to it by the Simple Random Sampling approach directly.

Stratified Random Samples Without Replacement are obtained by separating the population into mutually exclusive sets, or stratas, and then drawing simple random samples without replacement from each stratum.

Here is an example:
An Ontario school board wanted to assess the students' opinion on dropping grade from the secondary school program.
They decided to survey students from Elmsview High School.

To ensure a representative sample of students from all grade levels, the school board used a stratified sampling technique.
In this case, the strata were the 5 grade levels, grades 9 to 13. The school board then selected a sample within each stratum.

The students selected in this sample were extracted using Stratified Random Samples

Without Replacement making up a total sample of 100 students.

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

- Illustrated principle of stratified random sampling
- The need for stratification
- Merits and demerits of the method
- Stratified sampling with SRSWR and SRSWOR