Subject	ECONOMICS
Paper No and Title	Paper 2: Quantitative Methods II (Statistical Methods)
Module No and Title	Module 11: Types of Sampling
Module Tag	ECO_P2_M11

Principal Investigator	Co- Principal Investigator	Co- Principal Investigator	Co- Principal Investigator and Technical Coordinator
Prof H C Pokhriyal Executive Director School of Open Learning University of Delhi Delhi-110007	**Dr Jaswinder Singh** Principal SGTB Khalsa College University of Delhi Delhi-110007	**Dr Jatinder Bir Singh** Principal Sri Guru Gobind Singh College of Commerce University of Delhi	**Dr Vimal Rarh** Deputy Director, Centre for e-Learning and Assistant Professor, Department of Chemistry, SGTB Khalsa College, University of Delhi *Specialised in : e-Learning and Educational Technologies*

Paper Coordinator	Content Writer	Reviewer
Dr Harish Dhawan Associate Professor Ram Lal Anand (Eve.) University of Delhi	**Ms. Mandeep Kaur** Assistant Professor Maitreyi College University of Delhi	**Prof. Gurmail Singh** Professor Department of Economics Panjab University Chandigarh

Anchor Institute : SGTB Khalsa College, University of Delhi

TABLE OF CONTENTS

1. Learning Outcomes

After studying this module, you shall be able to

- Know methods of random and non-random sampling.
- Learn the conditions under which techniques of random and non-random sampling can be applied.
- Identify the conditions (of the population and the requirements of the sample) under which each technique within random sampling and nonrandom can be applied.
- Evaluate the comparative benefits of each type of sampling techniques.

2. Introduction

In order to make an assessment of the population parameter of our interest, we can interview each and every member of the population. For example, in order to assess the per capita income of people living in a city, we can interview each and every member of the population (that is all the residents of the city) about their income. This method, called the complete enumeration method (or for short Census method) can be very time consuming and resource using.
A method which is manageable and less resource consuming is to take a sample from the population and calculate an estimator of the population parameter (more on estimators in the next module). Taking a sample out of the population to calculate an estimator may lead to bias in calculating the estimate. Hence, to avoid this bias, we aim to take a representative sample of the population.
This module takes a tour of two kinds of sample designs. The design of a sample refers to the methods used to choose a sample from the population.
There are two kinds of sample designs:

3. Random Sampling or Probability Sampling:

In this design of sample, we use chance to choose a sample. The sample is chosen by chance, with prior information available on what samples can be chosen and what are the probabilities of each sample getting chosen. Every item of the population has a known chance of getting selected for the sample.
Specifically, random sampling has the following mathematical properties[1]:
a) Distinct samples from the population can be defined, which means that we can clearly state the items that belong to a particular sample of the population.
b) Each sample has a known probability of selection.
c) Each sample is selected by a random process. It may have equal or unequal probability of getting selected.
d) The method for computing the estimate from the sample must be stated and lead to unique estimates for a specific sample. So for example, it can be stated that the estimate is the average of the measurements on the individual items of the sample.

[1] Cochran William G., *Sampling Techniques*, Third Edition, (John Wiley and Sons, 1977).

This sampling procedure is amenable to the calculation of frequency distribution of the estimates (for each sample, when repeated sampling is done). We know the number of times a particular sample will be selected and thereafter the estimate from the sample. Thus, a well-defined sampling theory can be developed for such procedures.
Also, for this procedure, it was realized that by the use of sampling theory and normal distribution, the amount of error to be expected in the estimates made from the sample can be approximately predicted.
There are various ways to obtain samples that represent the population. Some of the methods of sampling are the following:

3.1 Simple Random Sampling[2]:

It is a method of obtaining a sample of size 'n' from a population size of 'N' units such that each of the C^N_n samples chosen have an equal probability of getting chosen[3]. To be precise, the random variables $X_1, X_2 \ldots X_n$ are said to form a simple random sample of size n if the following two conditions are met:
a) The X_i's are independent random variables.
b) Every X_i has equal probability.
The X_is are then termed as *independent and identically distributed (iid).* Random sampling is then possible if sampling is with replacement or is from an infinite population (in which case the probability of X_is is equal and X_is become independent).
There are various ways by which we obtain random samples. One is the lottery method, in which individual units of the population are allotted a number, which are then put onto slips of paper. These slips are then shuffled and a random draw of required numbers (which constitute the sample size) is done. This constitutes a random sample. The other method is that of using random sampling numbers[4].
The most important virtue of this sampling method is that it is easiest to derive a probability distribution of the sample statistic than for any other sampling method.
In simple random sampling, each X_i of the sample has equal probability of p= n/N. Here *all* the units of the sample are selected using a random mechanism[5].
For example, a national fast food chain wants to randomly select 5 out of 50 states to sample taste of its consumers. A simple random sample will ensure that the C^{50}_5= 2118760 samples of size 5 will have the same likelihood of being used in the study.[6]

3.2. Systematic Random Sampling[7]:

It is useful when the population units are ordered or listed in a random fashion[8]. For example, in the case, where houses are randomly arranged in rows.
Suppose a few houses at random are to be selected from a given city. Here systematic sampling can be used because the houses are usually arranged in rows and are thus numbered. The first house is selected at random and then the next 10^{th} or 15^{th} house is selected systematically. This is

---

[2] Devore Jay L., *Probability and Statistics for Engineers* (Cengage Learning, 2008).
[3] Cochran William G., *Sampling Techniques*, Third Edition, (John Wiley and Sons, 1977).
[4] Nagar A.L and Das R.K, *Basic Statistics*, (Oxford University Press, 1983).
[5] ILO, "Chapter 5- Sampling" in Consumer Price Index Manual: Theory and Practice, 69- 79.
[6] Webster Allen L, *Applied Statistics for Business and Economics* (Tata McGraw Hill, Third Edition).
[7] Nagar and Das, *op.cit*.
[8] Webster, *op.cit*.

called systematic random sampling. In systematic sampling, the sampling units are selected at equal distances from each other in the frame, with random selection of only the first unit.
Another population in which systematic sampling is used is when selecting items from a production or assembly line for quality testing. A manufacturer may select the first item on the production line randomly, after which every 20th item is selected for the sample.
Another example is when a market researcher may select the first person who enters a store after which every 10th person is selected for the purpose of the sample.
A sampling interval is calculated by dividing the population size by the desired sample size (N/n). For example, a university has 10,000 students and a systematic random sample of 500 students is to be taken. A sampling interval is calculated by dividing 10,000 by 500 (which is equal to 20). Then, all students are ordered in a sequential number. The first student is selected at random from 1 to 20 and thereafter every 20th student is selected.
Both simple random sampling and systematic sampling techniques are usually recommended when all the population units are relatively homogeneous.

3.3 Stratified Random Sampling:

This method is adopted when the population from which a sample has to be drawn is heterogeneous. We then divide the heterogeneous groups into homogeneous groups called *strata* and then draw a random sample from each stratum[9]. The heterogeneous groups or subpopulations must be non-overlapping and adding up all the units of the strata must give us the total number of units of the population.
In other words, stratified sampling divides the population of N units into subpopulations of N_1, N_2,… N_l units, respectively. These subpopulations are no overlapping, and together they form the entire population, so that

$N_1 + N_2 + \ldots + N_l = N$

The subpopulations are called strata. A random sample is then drawn from each stratum, the drawings being made independently in different strata. The sample sizes within the strata are denoted by n_1, n_2,…, n_l, respectively.
We can use one of the two approaches. Either we select at random from each stratum a specified number of elements *corresponding to the proportion of the stratum in the population* as a whole or we draw an *equal number of elements from each stratum and give weight to the results according to the stratum's proportion of total population*. With either approach, stratified sampling guarantees that every element in the population has a chance of being selected.
For example, in order to select a sample of 10 joinees from a population of 100 joinees, where there are 4 locations (North, East, and West& South) with the proportion as North 40%, East 10%, West 20% and South 30%. The sample can be selected as 4 from North, 1 from East, 2 from West and 3 from South OR select equal sample from all location and give weight to sample from North 40%, East 10%, West 20% and South 30%. The sample will be selected in the manner as in case of any of the random sampling techniques described above.

This procedure is described as stratified random sampling. Stratified random sampling can be used for various reasons, some of which are the following:

[9] Nagar and Das, *op.cit*.

i) Administrative convenience may facilitate the use of stratification; for example, the agency conducting the survey may have field offices in different geographical stratum, each of which can supervise the survey for a part of the population.
ii) Sampling problems differ significantly for different parts of the population. For example, people living in institutions (e.g., hotels, hospitals, prisons) are often placed in a different stratum from people living in ordinary homes because a different approach to the sampling is appropriate for the two situations. Population of people can be divided into different strata of income groups-for example high income group, upper middle income group, lower middle income group and low income group which can greatly enhance the analytical capacity of the sample statistic.
iii) Stratification may enhance the precision in the estimates of the characteristics (as compared to simple random sample) of the whole population by dividing a heterogeneous population into subpopulations, each of which is internally homogeneous. The estimates of each strata can be combined into precise estimate for the whole population[10].

3.4. Multi Stage Random Sampling:

We use the following example to illustrate multi stage sampling. Suppose we have to estimate the per acre yield of wheat in Punjab. We might obtain the entire list of villages in Punjab and select a given number of villages at random according to random sampling procedure. Then from out of selected villages we might select a given number of fields at random. This is a two stage random sampling procedure. In another situation one might first select some districts at random, then from out of the selected districts select a given number of villages at random and then finally select a few plots on the selected fields. This procedure is called multi stage random sampling procedure[11].

3.5.Cluster Sampling[12]:

It refers to sampling surveys in which the sampling unit consists of groups or clusters of smaller units called as elements. It is used in the following cases:
a) Cases in which a reliable list of individual elements in the population might not be available and it can be prohibitively expensive to construct such a list.
b) Cases in which population units are widely distributed in any large geographic region or natural clusters and drawing a sample from such widely dispersed geographical regions is difficult. From maps of the region, however, it can be divided into areal units such as blocks in the cities and segments of land with readily identifiable boundaries. Cluster sampling thus, solves the problem of constructing a list of sampling units.
For example, we want to estimate health insurance coverage in a city. We can take a random sample of say 100 households, which means we need a sampling list of the city's households. If the list is not available, we have to conduct a census of the households, which can be very expensive. Also, because the sample size is small compared to the number of total households in the city, we have to sample only few (say one or two) in each block of the city.
Alternatively, we can select 5 blocks out of say, 200 blocks (into which the city has been divided) and in each block interview 20 households. This implies that a listing frame of only 5 blocks needs to be prepared which means that less time and cost will be required.

[10] Cochran, *op.cit.*
[11] Nagar and Das, *op.cit.*
[12] Cochran, *op.cit.*

The blocks are called the Primary Sampling Units or the clusters. The households are called the Secondary Sampling Units.[13]
Another example is that of taking a sample of children in an elementary school. We could take a simple random sample. An alternative is to take a random sample of classes and then measure *all students* in the selected classes[14].
It should be noted that cluster sampling is different from stratified random sampling. While in stratified random sampling a random sample is taken from *each* of the stratum, cluster sampling treats each cluster as a sampling unit and hence, *all observations* in the *selected* clusters are included in the sample.[15]
Every cluster unit may contain the same number of elements. However, cluster units may also contain different numbers of elements.

3.6. Probability proportional to size (PPS) Sampling[16] :

Sampling is a kind of cluster sampling in which the probability that a sampling unit gets selected in the sample depends on the size of the population of the sampling unit, giving large clusters (that is those having larger population) a higher probability of getting selected than small clusters. It is useful when the sampling units vary considerably in size because it assures that those in larger sites have the same probability of getting into the sample as those in smaller sites, and vice versa.
An example of systematic PPS sampling is provided below:
The example[17] shows how a sample of 3 outlets can be drawn from 10. In this case, we have the number of employees as our size measure. We look at the list, where we have included the cumulative sizes and the inclusion intervals. We take the total number of our size measure which is 90 in this case, and divide it by the sample size, 3. This gives us a sampling interval (SI) of 30. We next choose a random number between 1 and 30. Say we take 25 (This is called Random Start[18]). The sample will then consist of the outlets whose inclusion intervals cover the numbers 25, 25+30 and 25+2*30. This implies that outlets 4, 6 and 10 are selected.

Table. 1 pps sampling

Outlet	Number of Employees	Cumulative x	Inclusion interval	Included when starting point is 25
1	13	13	1- 13	
2	2	15	14- 15	
3	5	20	16- 20	
4	9	29	21- 29	#
5	1	30	30	

[13] Saifudin Ahmed, *Methods in Sample Surveys* http://ocw.jhsph.edu/courses/statmethodsforsamplesurveys/PDFs/Lecture5.pdf
[14] http://www.stat.purdue.edu/~jennings/stat522/notes/topic5.pdf
[15] Webster, *op.cit.*
[16] ILO, Chapter 5, Sampling, Consumer Price Index- Theory and Practice.
[17] ILO, "Chapter 5- Sampling" in Consumer Price Index Manual: Theory and Practice, 69- 79.
[18] RHRC CONSORTIUM MONITORING AND EVALUATION TOOLKIT SURVEY SAMPLING TECHNIQUE EXAMPLE, Prepared by Therese McGinn Heilbrunn Department of Population and Family Health Mailman School of Public Health, Columbia University

6	25	55	31-55	#
7	10	65	56- 65	
8	6	71	66- 71	
9	11	82	72- 82	
10	8	90	83- 90	#

4. Non random sampling

Every item in a population has a known chance of being included in a sample in random sampling. However, in non-random sampling, this is not the case. Here, not all outcomes have a known chance of occurring. In fact, there are some items which do not zero probability of getting selected. It is hence not the most representative of the population. This introduces significant biases in the estimate of the estimator. There are three common non random sampling techniques:

4.1 Quota Sampling:

It is similar to stratified sampling, with the difference being that the selection from a particular stratum is non random. The quota of a particular stratum is fixed by the researcher, independent of the proportions of that strata in the population, so as to ensure the inclusion of that particular stratum in the sample.

For example, in a study of views of minorities on death penalty for the State of Punjab, a random sample of people will for sure miss the Muslims (since there are not many Muslims in that state). To ensure their inclusion in the sample, a fixed 1% quota in sample is fixed by the researcher, even though it is not representative of the population. The quota, however, serves to make comparisons across all the minority communities.

4.2 Convenience Sampling:

This method does not produce representative sample, because only those items are selected for sampling which can be accessed most easily.

4.3 Volunteer Sampling:

This is a method in which those people are part of the sample who volunteers to participate in a particular sample survey. There is as such no design of the sample. For example, a radio station may ask its audience to give their opinion on a certain issue, with only two possible opinions- yes or no. A mobile phone number is offered to the audience for a SMS facility of Y for Yes and N for No. A time period is offered to the audience to text their opinion after which the phone number facility is closed and it’s on this basis that a conclusion is formed.

4.4 Purposive sampling

This method is also called judgmental sampling, in which the sampling units are so selected as to meet the broad objective of the survey, by taking a sample that represents the control variables of the study and for which the results of the entire population are already known. The sample so selected would thus, have strong correlation with other variables of our interest.

The method requires prior information of the characteristics of the sampling units, before being selected as an item of the sample. Hence, a huge amount of discretion and judgment on the part of the sampler is required. This is unlike the methods of random sampling, in which each sampling units is seen as interchangeable.

There are different kinds of purposive sampling- for example maximum variation sampling, homogeneous sampling, typical case sampling, extreme case sampling, critical case sampling and expert sampling. All these sampling methods define the control variable- for example, expert sampling omits out amateurs, typical case sampling omits out outliers from the sampling frame etc.

An example for purposive sampling is the following: to determine the overall performance (that includes academic and other pursuits) of average student of a particular grade, we shall take a sample of 10 students out the 100 students of that grade, such that they are just average. The control variable here is their marks and its possible correlations with other variables of overall performance (like sports, arts, music etc.). It also requires prior information on the part of the researcher about the sampling unit- the students in this case. This is the case of typical sampling.

These methods, thus, do not ensure good representation of the entire population, though under right conditions these methods can yield useful results.

5. Summary

In conclusion random sampling entails the following:

- Distinct samples from the population can be defined.
- Each sample has a known chance of getting selected.
- Each sample is selected randomly, with equal or unequal probability of getting selected.
- This procedure is amenable to the calculation of frequency distribution of the estimates and hence a well-defined sampling theory can be developed for such procedures.
- With the use of sampling theory and the assumption of normal distribution, these methods also enable calculation and prediction of amount of error expected from the estimates of the population parameter.

There are many methods that can be employed to assure that a sample is close to randomness. Simple random sampling, systematic sampling, stratified sampling, cluster sampling, multi stage sampling and PPS sampling are some of the methods to ensure random samples. Each method is relevant to different population kinds.

On the other hand, non-random samples have the following features:

- It is a sampling methodology in which all items do not have a known chance of getting selected. In fact, some items have zero probability of getting selected.

- There is a high chance of selecting biased samples by this method, which does not represent the population.

There are number of methods by which we can take a non-random sample, some of which includes quota sampling, convenience sampling, volunteer sampling and purposive sampling.