CHAPTER 7

7. Sampling Techniques

7.1. Basic Concepts

Population: A population is the totality of all subjects, measurements or individuals possessing certain common characteristics that are being studied. The population represents the target of an investigation, and the objective of the investigation is to draw conclusions about the population hence we sometimes call it ***target population.*** The population could be finite or infinite (an imaginary collection of units)

There are two ways of investigation: Census and sample survey.

Census: a complete enumeration of the population. But in most real problems it cannot be realized, hence we take sample.

Sample: A sample from a population is the set of measurements that are actually collected in the course of an investigation. It should be selected using some predefined sampling technique in such a way that they represent the population very well.

The process of collecting information from all the elements of a large population may be expensive, time consuming and difficult. Thus, we study a sample and generalize the properties of the sample to the entire population.

Parameter: Characteristic or measure obtained from a population.

Statistic: Characteristic or measure obtained from a sample.

Sampling: The process or method of sample selection from the population.

Sampling unit: the ultimate unit to be sampled or elements of the population to be sampled.

Examples:

If somebody studies Scio-economic status of the households, households is the sampling unit. If one studies performance of freshman students in some college, the student is the sampling unit.

Sampling frame: is the list of all elements in a population.

Examples: List of households.

List of students in the registrar office.

Errors in sample survey:

There are **two types** of errors

a) Sampling error:

- Is the discrepancy between the population value and sample value.

- May arise due to in appropriate sampling techniques applied

b) Non sampling errors: are errors due to procedure bias such as:

- Due to incorrect responses

- Measurement

- Errors at different stages in processing the data.

7.2. Reasons for Sampling

Reduced cost

Greater speed

Greater accuracy

Greater scope

Avoids destructive test

The only option when the population is infinite

Because of the above consideration, in practice we take sample and make conclusion about the population values such as population mean and population variance, known as parameters of the population.

Sometimes taking a census makes more sense than using a sample. Some of the reasons include:

Universality

Qualitativeness

Detailedness

Non-representativeness

Exercises:

1. Describe briefly the difference between

Population and sample

Parameter and statistic

Census and sample survey

Sampling and non-sampling error

Sampling frame and sampling units

2. Why do researchers usually select sample elements from a given population?

3. Mention some of the disadvantage of sampling.

4. An insurance company has insured 300,000 cars over the last six years. The company would like to know the number of cars involved in one or more accidents over this period. The manger selected 1000 cars from the files and made a record of cars that were involved in one or more accidents.

a. What is the population?

b. What is the sample?

c. What is the variable of interest to the insurance company?

7.3. Types of sampling techniques

- There are two types of sampling techniques.

A. **Random Sampling or probability sampling.**

- Is a method of sampling in which all elements in the population have a pre-assigned non zero probability to be included in to the sample.

Examples:

- Simple random sampling
- Stratified random sampling
- Cluster sampling
- Systematic sampling
- Multi-stage sampling

1. **Simple Random Sampling***:*

- Is a method of selecting items from a population such that every possible sample of specific size has an equal chance of being selected. In this case, sampling may be with or without replacement. Or

- All elements in the population have the same pre-assigned non zero probability to be included in to the sample.

- Simple random sampling can be done either using the lottery method or table of random numbers.

Table of Random Numbers

Table of random numbers are tables of the digits 0, 1, 2,…,, 9, each digit having an equal chance of selection at any draw. For convenience, the numbers are put in blocks of five. In using these tables to select a simple random sample, the steps are:

i. Number the units in the population from 1 to N (prepare frame of the population).

ii. Then proceed in the following way

If the first digit of N is a number between 5 and 9 inclusively, the following method of selection is adequate. Suppose N=528 and we want n=10.

Select three columns from the table of random numbers, say columns 25 to 27. Go down the three columns selecting the first 10 distinct numbers between 001 & 528. These are 36, 509, 364, 417, 348, 127, 149, 186, 439, and 329. Then the units with these roll numbers are our samples.

Note: If sampling is without replacement, reject all the numbers that comes more than once.

2. **Stratified Random Sampling**:
   - The population will be divided in to non-overlapping but exhaustive groups called strata.
   - Simple random samples will be chosen from each stratum.
   - Elements in the same strata should be more or less homogeneous while different in different strata.
   - It is applied if the population is heterogeneous.
   - Some of the criteria for dividing a population into strata are: Sex (male, female); Age (under 18, 18 to 28, 29 to 39); Occupation (blue-collar, professional, other).

3. **Cluster Sampling**:
   - The population is divided in to non-overlapping groups called clusters.
   - A simple random sample of groups or cluster of elements is chosen and all the sampling units in the selected clusters will be surveyed.
   - Clusters are formed in a way that elements within a cluster are heterogeneous,

i.e. observations in each cluster should be more or less dissimilar.

   - Cluster sampling is useful when it is difficult or costly to generate a simple random sample. For example, to estimate the average annual household income in a large city we use cluster sampling, because to use simple random sampling we need a complete list of households in the city from which to sample. To use stratified random sampling, we would again need the list of households. A less expensive way is to let each block within the city represent a cluster. A sample of clusters could then be randomly selected, and every household within these clusters could be interviewed to find the average annual household income.

4. **Systematic Sampling**:
   - A complete list of all elements within the population (sampling frame) is required.
   - The procedure starts in determining the first element to be included in the sample.
   - Then the technique is to take the kth item from the sampling frame.
   - Let N = population size, n = sample size, k = N/n = sampling interval.
   - Choose any number between 1 and k. suppose it is j $(1 \leq j \leq k)$.

- The jth unit is selected at first and then $(j + k)^{th}$, $(j + 2k)^{th}$, …, etc until the required sample size is selected.

5. **Multi-stage sampling**: Selection is done in stages until the final sampling units (for example, households or persons) are arrived at. The primary sampling unit (PSU) is the sampling unit (usually large size) in the first sampling stage. The secondary sampling unit (SSU) is the sampling unit in the second sampling stage, etc. for example – The PSUs could be kebeles and the SSUs could be households.

B. Non Random Sampling or non-probability sampling

It is a sampling technique in which every unit of the population has no the same probability of being selected in to the sample. It is to select elements from the population by any mechanism that does not involve a random selection process.

Examples: Judgment sampling, Convenience sampling, Quota Sampling.

i) Judgment Sampling

This involves choosing purely by the judgment of the investigator. The investigator chooses a sample of individuals that are thought to be representative of the population.

ii) Quota Sampling

This sampling method has some aspects in common with stratified sampling, but has no randomization. The population is divided into strata as in stratified sampling, but a judgment sample is chosen from each stratum.

iii) Convenience Sampling

It is a non-random sampling technique in which items for the sample are selected for the convenience of the investigator. That is a convenience sample is obtained by taking individuals that are easy to find.

o No judgment is used here

o Convenience samples are very cheap but naturally exposed to bias.

o Not representative to the population

o It is used when conducting pilot survey.

The Central Limit Theorem

Let $\bar{X}$ denote the mean of the observations in a random sample of size n from a population having a mean μ and standard deviation σ. Denote the mean of the $\bar{X}$ distribution by $\mu_{\bar{X}}$ and the standard deviation of the $\bar{X}$ distribution by $\sigma_{\bar{X}}$.

The central limit theorem turns out that if our samples are independent and identically distributed, the variable that the sample is from has a finite population mean and variance and n is sufficiently large , then:

➢ The sample mean has the same expected value as the population mean, $\mu_{\bar{X}} = \mu$.

➢ The standard error of the sample mean is, $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$.

➢ The sample mean is normally distributed, regardless of the distribution of the original variable.