Sample Survey and Sampling Methods

Course unit ECON 53115
MA/MSSc in Economics

Prof. WM Semasinghe
Dept. of Economics

What is a sample survey?

Sample survey is a study which focus to obtain data and information to calculate statistics over the population parameters.

Important concepts of sample survey

There are several important concepts relating to the sample survey and sampling methods.

Population:

• *Population* includes all members of a defined group that are studying or information that collecting for data driven decisions.

• In statistics, population means *the whole of the information which comes under the purview of statistical investigation.*

• A population can be defined as the collection of all individuals or items with the characteristics one wishes to understand.

• Population may be ***finite*** or ***infinite*** as the number of observations or individuals or items in it is finite or infinite.

eg.

• In a study on households average income in Sri Lanka, population includes all households in the country – population is finite.

• In an investigation on the quality of a batch of a product of a certain firm, population includes all items produced by the firm in the given period of time – population is finite.

• In the sampling process whether a population is finite or infinite depends on the ability to count all the individuals or objects separately practically.

- In a study on the fishes in the Indian Ocean, population includes all fishes in the ocean - population is infinite.

- **Population may be hypothetical; eg. Outcomes of tossing a coin.**

'A set of all possible or hypothetically possible outcomes of a statistical investigation'.

- The number of individuals/items of a population is denoted by *'N'*, and

- If the value of a certain characteristic of i^{th} individual/item *is* y_i

- then $N = y_1, y_2, \ldots\ldots, y_N$

Sample

- A finite sub-set of individuals or items in a population is called a *sample*.

- Sample is always finite.

- A sample is a scientifically drawn group that actually *possesses the same characteristics* of the population – if it is drawn randomly.

- A sample drawn from a population provides valuable information about the parent population.

Sample size

• Number of individuals or items included in a sample is called ‘size of the sample’ and denoted by ‘*n*’.

• If the value of a certain characteristic of the i^{th} individual/item of the sample is y_i, then,

$$n = y_1, y_2, \ldots, y_n$$

Census vs. Sample survey

Examination of each and every individuals or items of a population is called *Census* or *Complete enumeration*.

eg. Census of population and housing – DCS

Census of public and semi-government sector employment - DCS

- However, in many statistical investigations complete enumeration is impossible and impracticable.

- The cost of a census in terms of time and money is very high.

- When the population is too large

Ex. Estimating the average monthly income of the people in India. (population is finite/ 100% inspection is impracticable)

- When the population is infinite

eg. Estimating the average height of a tree in the Sinharaja forest. (population is infinite/ 100% inspection is impossible)

eg. Testing the quality of water in a lake. (population is infinite/complete enumeration is impossible)

- When the items are destroyed.

eg. - Estimating the average length of life of a electric bulb.

(population is finite;100% inspection is impracticable)

- Blood testing

- Examination of a sample or part of a population to determine the population characteristics is called ***sample survey.***

eg. Labor Force Survey – DCS

Survey of construction industries - DCS

Parameters and Statistics

- Various statistical measures such as mean, median, mode, variation, standard deviation are estimated in statistical investigations.

- When these measures defined for a population are called ***parameters*** and denoted usually by **θ**.

- Any statistical measure defined for a sample is called ***statistic*** and usually denoted by $\hat{\theta}$.

- Parameters are statistical measures based on population data whereas statistics are statistical measures based on sample data.

Statistical measure	Population (θ)	Sample $(\hat{\theta})$
Mean	μ	$\bar{X}$
Variation	σ^2	S^2
Standard deviation	σ	s
Proportion	P	p
Size	N	n

Why sample surveys?

Principle Advantages of Sample Survey

The principle advantages of sample surveys against the complete enumeration can be listed as follows:

- **Greater speed**

Since only a part of the population has to be examined, data can be collected and analyzed within short period of time compared to the census. Therefore, result can be obtain more rapidly.

- **Reduced cost**

Cost of the sample survey is less than the complete enumeration. Although the cost per unit is generally greater of sample survey, total cost is expected to be much smaller than that of complete enumeration.

- **Greater Accuracy**

The results of a sample survey are much more reliable than those obtain from a census because skilled and experienced personals, sophisticated equipment and techniques etc. which are important in a survey can be used sufficiently and careful supervision can be made in collecting and analyzing data.

Greater scope

Sample survey has generally greater scope as compared with the census. In certain types of inquiry, highly trained personal and/or specialized equipment must be used to collect the data.

In complete enumerations, it is difficult to afford such personals/equipments sufficiently.

Moreover, in sample survey, it is possible to have a thorough and intensive enquiry because a more detailed information can be obtained from a small number of items/individuals.

In addition to these advantages conducting sample survey is unavoidable in some instances such as:

i. When the population is too large or infinite. For example study on the trees in a jungle, fishes in an Ocean, sand in a beach etc.

ii. When testing is destructive. For example, testing the quality of milk, breaking strength of loran bars, explosives, length of life of an electric bulb etc.

iii. If the population is hypothetical. For example in coin tossing problem where the process may continue indefinitely, sampling is the only scientific method of estimating the parameters of the population.

Limitations of Sampling

The advantages of sampling over complete enumeration mentioned above can be attained only if,

i. Appropriate sampling techniques is used,

ii. The sampling units are drawn in a scientific manner, and

iii. The sample size is adequate.

Sampling theory has its own limitations and problems which may be briefly outlined as follows:

1. Proper care should be given in the planning and execution of the sample survey, otherwise the result obtained might be inaccurate and misleading.

2. Sampling theory requires the services of trained and qualified personnel and sophisticated equipment for its planning, execution and analysis. In the absence of these, the results of the sample survey are not trustworthy.

3. If the information is required about each and every unit of the population, complete enumeration is compulsory.

4. If time and money are not important factors or if the population is not too large, a complete enumeration is better than sample survey.

The Principle Steps in a Sample Survey

The main steps involved in the planning and execution of a sample survey may be cited somewhat arbitrary under the following headings:

i. Objectives of the survey

The objectives of the survey should be defined in clear and concrete terms. This is very important in planning the survey.

- What you want to learn?
- Whom you will survey and what you will ask them.
- What data should be collected
- When the data is collected. All these depend on the objective/s of the study.

- When the goal is unclear, the results will probably be unclear.

ii. Defining the population to be sampled

The population from which sample is chosen should be defined in clear and unambiguous terms.

Defining of the population is not difficult when the sampling units are homogenous.

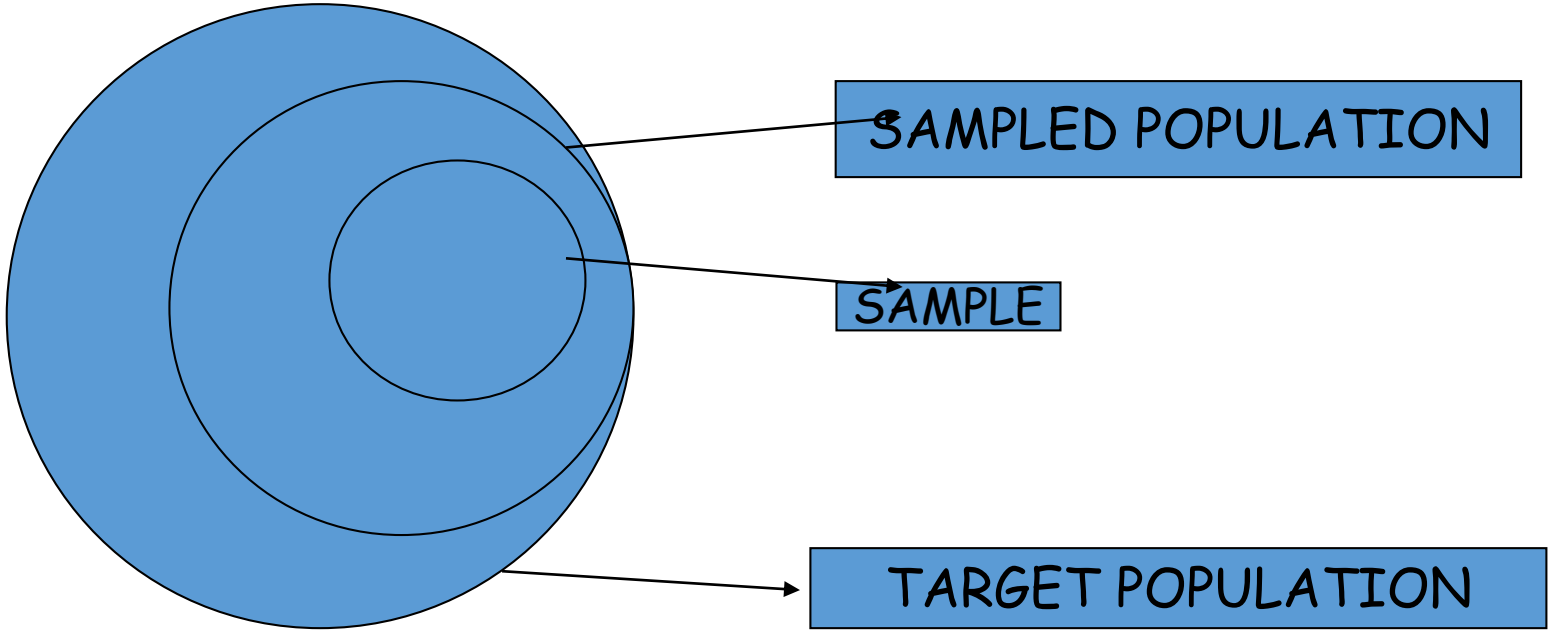
For example, defining of population is not difficult in a study focuses on estimating the average length of life of a bulb of a certain brand because the population is homogenous.

However, sampling from a heterogeneous population is not easy.

The rules must be set up clearly to define sampling units. The enumerator must be able to decide in the field, without hesitation, whether or not to include a given unit in the population.

The population to be sampled (*sampled population)* should coincide with the population about which information is required (*target population).*

Some times for reasons of practicability or convenience, sampled population is restricted than the targeted population.



Such a situation, it should remember that conclusions drawn from the survey applicable only for the sampled population.

iii. Data to be Collected

The data to be collected depends on the objectives of the study.

- Which data to be collected
- How to collect data
- Who will collect the data
- When to collect the data

It is important to verify that all data relevant to the objectives are obtained and no essential data are omitted.

All so, it should ensure that no irrelevant information are collected.

The practical method to ensure these is to produce an outline covering all information required to accomplish the objectives.

Types of Data

Primary data: The data collected by researcher for his/her specific purpose and for his/her specific use. As they are collected for the first time they are original in nature.

Secondary data: The data collected by other researchers or institutions for their purposes and available for the use of others. They have already been used and analyzed by someone else.

Experimental data

Observational data

Cross section data

Time series data

From among these different type of data in sample surveys collect and used largely the primary and secondary type data.

Methods of Collecting Primary Data

There are various measuring instruments and methods of collecting primary data from a human population.

- Survey ⟹
  - Questionnaire method
  - Interview method
- Observation method
- Focus group discussions (FGD)
- Case studies
- KIIs
- Web survey

The researcher should chose the most appropriate method among these alternatives

Accuracy aimed at and the costs involve are the important factors that should be taken into account.

All these methods have certain advantages as well as disadvantages.

Secondary Data Sources

v. The Frame

Before selecting the sample, the population must be divided into parts that are called *sampling units*. The list, map, chart or other acceptable material which construct including all sampling units are called '*the frame*'.

The frame must,

- cover the whole of population
- not overlap, in the sense that every element in the population belongs to one and only one unit.
- complete, in the sense that it includes all units in the population,
- include the accurate information of the relevant units, up-to-date

vi. Selection of the Sample

There are variety of plans by which the sample may be selected. Sampling design is very important for the accuracy and reliability of the estimates.

Selection of a proper sampling design is the responsibility of the researcher. In addition to the nature of the population, the relative cost and time involved should also be considered before making a final selection of the sampling design.

For each sampling design, rough estimate of the sample size (n) can be obtain for the desired degree of precision.

vii. The Pretest

Before start the main field work, it is useful to try out the questionnaire and/or the field methods on a small scale.

It will help and allow,

- to improve the questionnaire or the data collecting method,
- to plan the fieldwork,
- to identify the problems arising in various steps,
- to train the enumerates,
- to get an idea about the cost and time required for the main field work etc.

viii. Organization of Field Work

- The success of a survey to a great extent depends upon the reliability of the field work. Thus, it is essential to trained the enumerators in locating the sample units, recording the measurements, the methods of collection of required data etc. before starting the field work.

- It is very necessary to assign adequate supervisory staff for the inspection of field work.

- Plans must be made for handling non-response errors i.e. the failure of the enumerator to obtain information from certain sample units.

ix. Summery and Analysis of the Data

The analysis of data may be broadly considered under the few categories.

1. Scrutiny and editing of the data: Supervisory staff should be carried out the quality checking scrutiny in the hope of amending recording errors or at least of deleting data that are obviously erroneous and inconsistent.

2. Tabulation of Data: The data should be tabulated employing most appropriate method. However, before the tabulation it should decide the procedure for tabulation of the data which are incomplete due to non-response to the certain items in the questionnaire and where certain questions are deleted in the editing process.

3. Statistical Analysis:

After the scrutinizing, editing and tabulating, the computations that lead to the statistical estimates are performed. Different methods of estimation may be available for the same data.

4. Reporting and Conclusions:

Finally, a report incorporating detailed statement of the different stages of the survey should be prepared. In the presentation of the result, it is good practice to report the amount of error to be expected in the most important estimates.

x. Information Gained for Future Surveys

More information that we have initially about a population, it will make easier to work out a sample that will give accurate estimates.

Thus, information gained from any completed survey over the population in the form of data such as the means, standard deviations etc. and the nature of variability of the population, with the cost involved in obtaining the data serves as a potential guide for improve the future surveys.

Methods of Sampling

A sample can be selected from a population in various ways. Different situation call for different methods of sampling. There are two methods of sampling:

Random *or* Probability Sampling

Random or probability sampling is the scientific technique of drawing samples from the population according to some laws of chance in which each unit in the population has some definite pre-assigned probability of being selected in the sample. The main probability sampling methods are:

i. Simple Random Sampling
ii. Stratified Random Sampling
iii. Cluster Sampling
iv. Systematic Sampling

Non-Random *or* Non-Probability Sampling

The methods that sampling units being selected on the basis of personal judgment is called non-probability sampling. In this method, personal knowledge and opinion are used to identify the individuals/items from the population.

It does not involve probability of selection.

The population may not be well defined.

There are several non-probability sampling methods. Followings are the mostly used methods:

- Judgment Sampling
- Quota Sampling
- Convenience Sampling
- Snowball Sampling

Advantages of Probability Sampling

- It gives a representative sample even if the population is heterogeneous.
- Statistical measures can be estimated and evaluated by sample statistic in terms of certain degree of precision.
- Since the estimates are unbiased, they can be generalized to the population.
- It is used to draw statistical inferences.
- Mathematical statistics and probability can be applied to analyze and interpret the data

Disadvantages of Probability Sampling

- Cost of sampling in terms of money and time is high compared to the non-probability sampling.

- Non-response error is significant.

- When sampling frame is not sufficient, complete and up-to-date, the sample may not representative and the statistical measures may not accurate.

Advantages of Non-probability Sampling

- Sampling frame is not required.
- Scientific knowledge is not required.
- Results can be taken within less time period.
  Thus, this technique is most popular in market researches.
- Cost is less than to the probability sampling
- Non-response error does not arise.

Disadvantages of Non-probability Sampling

- Since the selection of sample units depends entirely on the discretion and judgment of the investigator, the sample may suffers from the drawbacks of favoritism and nepotism depending upon the beliefs and prejudices.

  As such, the sample may not be a representative for the population.

- Statistical measures are not valid.

- Sampling theory cannot be used to test the statistical reliability of the estimates.

- Findings cannot be generalized

Sampling and Non-sampling Errors

There can be discrepancies in the statistical measures of population, *i.e. parameter* and the statistical measures of sample drawn from the same population, *i.e. statistic*. These discrepancies are known as **Errors in Sampling**.

Errors in Sampling are of two types:

i Sampling errors

ii. Non-sampling errors

Sampling Errors

Sampling errors have their origin in sampling and arise due to the fact that only a part of population, i.e. sample has been used to estimate population parameters and draw inferences about the population.

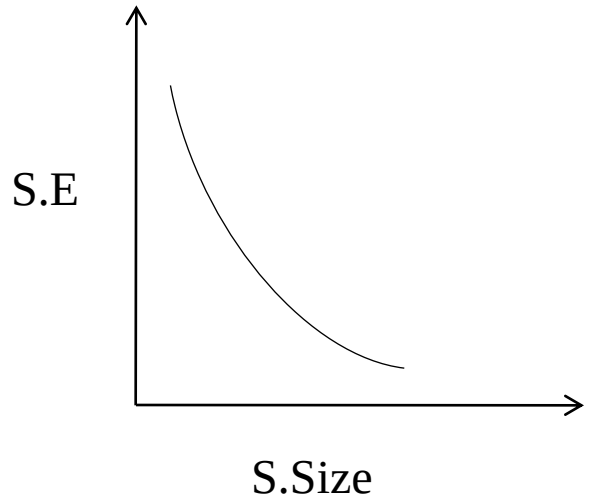
Sources of sampling errors:

1. Faulty selection of the sample.

2. Faulty demarcation of sampling units.

3. Substitution

- Sampling errors is inversely proportional to the sample size.



Non-sampling errors

Non-sampling errors preliminarily arise at the stages of observation, collecting, processing, tabulating and analyzing of data and publishing the result. Thus, non-sampling errors may arise both in sample surveys and complete enumeration.

Sources of non-sampling errors

1. Faulty planning or definition
2. Response errors

i. Response errors may be accidental.
ii. Prestige bias
iii. Self-interest
iv. Bias due to interviewer
v. Failure of respondent's memory

3. Due to the negligence and carelessness on the part of investigator.

4. Non-response biases.

5. Compiling errors.

6. Publication errors.

- Non-sampling errors are likely to be more serious in a complete census as compared to a sample survey.
- Non-sampling error tends to increase as the sample size increases.
- Quite often, the non-sampling error in a complete census is grater than both the sampling and non-sampling errors taken together in a sample survey.

Simple Random Sampling (SRS)

SRS is a technique of drawing a sample in such a way that *each unit of the population has an equal and independent chance of being included in the sample*.

When there are *N* units of the population, each unit has *1/N* probability of being selected to the sample.

Let us suppose that a sample of size n is drawn from a population of size N. There are ${}_{N}C_{n}$ possible samples.

SRC is the technique of selecting a sample in such a way that each of ${}_{N}C_{n}$ samples has an equal chance or probability [$p = 1/{}_{N}C_{n}$] of being selected.

Selection of a Simple Random Sample

Two methods of drawing a Simple Random Sample:

i. Lottery system

ii. Using a random number table

iii. Computer based selection

Available random number tables are:

i. Trippet's Random Number Series,
ii. Fisher's and Yale's Random Number Series,
iii. Kendall and Badington Random Number Series,
iv. Rand Corporation Random Number Series,

Definitions and Notation

Characteristics

The properties that attempt to measure and record for every unit comes into the sample is called *characteristics*.

	Population	Sample
Total	$Y = \sum_{i=1}^{N} y_i = y_1 + y_2 + ... + y_N$	$\hat{Y} = \sum_{i=1}^{n} y_i = y_1 + y_2 + ...y_n$ $\hat{Y} = N\bar{y}$ is an unbiased estimater for Y
Mean	$\overline{Y} = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{y_1 + y_2 + ... + y_n}{N}$	$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{y_1 + y_2 + ... + y_n}{n}$

Estimation of the sample size

Principle steps involved in the choice of sample size:

1. There must be some statement concerning what is expected of the sample. This statement may be,
   - in terms of desired limits of error or
   - the expected accuracy of the result obtain.

2. Available resources:

The chosen value of n must be appraised to see whether it is consistent with the resources available to take the sample.

.

$$n = \frac{\frac{t^2 PQ}{d^2}}{1 + \frac{1}{N}\left(\frac{t^2 PQ}{d^2} - 1\right)} \qquad (1)$$

We can substitute p and q for P and Q

When N is large we can defined

$$n_0 = \frac{t^2 pq}{d^2} = \frac{pq}{V} \qquad where\ V = \frac{d^2}{t^2}$$

Substituting n_0 into (1)

$$n = \frac{n_0}{1 + (n_0 - 1)/N} \quad or \quad \frac{n_0}{1 + \frac{n_0}{N}}$$

Sample Size Determination Using Krejcie and Morgan Table

Sample Size Formula for Finite Population

$$S = \frac{X^2 NP(1-P)}{d^2(N-1) + X^2 P(1-P)}$$

Where:

S = Required Sample size

χ^2 = Table value of χ^2 for 1 degrees of freedom at the desired confidence level (e.g. 95% - 1.96 x 1.96 = 3.841)

N = Population Size

P = Population proportion (expressed as decimal) (assumed to be 0.5 (50%)

d = Degree of accuracy (5%), expressed as a proportion (.05); It is margin of error

https://home.kku.ac.th/sompong/guest_speaker/KrejcieandMorgan_article.pdf

Sample Size Formula for Infinite Population

$$n = \frac{Z^2 \times p\,(1-p)}{M^2}$$

Where:

n = Sample Size for infinite population

Z = Z value (e.g. 1.96 for 95% confidence level)

P = population proportion (expressed as decimal) (assumed to be 0.5 (50%)

M = Margin of Error at 5% (0.05)

Merits of SRS

Merits of the SRS over the other sampling methods, apart from the broad advantages of probability sampling methods are:

1. Since the sample units are selected at random giving each unit an equal chance of being selected, personal bias is completely eliminated. Thus, SRS is more representative of the population as compared to the judgment or purposive sampling.

2. Sampling procedure is simple compared to the other probability sampling methods.

3. Cost in terms of money and time is low compared to the other probability sampling methods.

Drawbacks of SRS

1. The selection of a simple random sample requires an up-to-date frame.

2. Administrative inconvenience.

3. If the population is heterogeneous the sample may not be a representative sample.

4. For a given precession, simple random sampling usually requires large sample size as compared to stratified random sampling.

Stratified Random Sampling

Population of N units is divided into subpopulations of N_1, N_2, ….N_L units considering a characteristic of the population. These subpopulations are non-overlapping (**mutually exclusive/disjoint**), and together they comprise the whole of the population (e**xhaustive**).

$$N_1 + N_2 + ... + N_L = N$$

- The subpopulations are called *strata*.
- After determining the strata, a sample is drawn from each strata independently.
- The sample size drawn from the population is denoted by *n*, and the sample sizes within the strata are denoted by n_1, n_2, …, n_L.
- When a simple random sample is taken in each stratum, the whole method is called *stratified random sampling*.

- In a study on the career goals of the students in a university, the stratification can be done based on the faculties, subjects, departments, etc.

- In a survey on the business firms, the firms can separate into sub-populations based on the turnover, number of employees, etc. as small, medium, and large.

- In a study on the households in a city, the households can be stratified taking into account their income levels as high, middle and low income.

- In a study on the agricultural farms, the stratification factor may be the size of the farms.

To create a stratified random sample, there are seven steps:

(a) defining the population;

(b) listing the population (sampling frame);

(c) choosing the relevant stratification;

(d) listing the population according to the chosen stratification

(e) choosing your sample size; .

(f) calculating a proportionate stratification; and

(g) using a simple random to select the sample

Determination of sample size and allocation among strata

In stratified sampling

i. Decide the size of the sample

ii. Planning to allocation among each strata

Two methods of allocation:

i. Optimum allocation

ii. Proportional allocation

Optimum Allocation

Under the optimum allocation the determination of sample size and the allocation among the strata are guided by two principles. i.e. so as to:

1. Minimize the $V(\bar{y}_{st})$ for a specified cost of taking the sample,

2. Minimize the cost of taking the sample for a specified value of $V(\bar{y}_{st})$

The simplest cost function is of the form

$$C = c_0 + \sum c_h n_h$$

Where, c_0 = overhead cost
c_h = a unit cost of the stratum h
n_h = number of units taken from h

Estimation of size of n_h

In stratified sampling with a linear cost function in the above form, the size of n_h for the minimization of $V(\bar{y}_{st})$ is:

$$n_h = \frac{W_h S_h / \sqrt{c_h}}{\sum (W_h S_h / \sqrt{c_h})} \times n = \frac{N_h S_h / \sqrt{c_h}}{\sum (N_h S_h / \sqrt{c_h})} \times n \quad \cdots\cdots\cdots \quad (1)$$

Estimation of Size of n

When the cost of sampling is given, above equation (1) must be substituted into the cost function in order to compute the sample size (n):

$$n = \frac{(C - c_0)\sum (N_h S_h / \sqrt{c_h})}{\sum (N_h S_h \sqrt{c_h})}$$

If $V(\bar{y}_{st})$ is given as $V(\bar{y}_{st}) = V$ sample size n is

$$n = \frac{(\sum W_h S_h \sqrt{c_h}) \sum W_h S_h / \sqrt{c_h}}{V + (1/N) \sum W_h S_h^2}$$

Proportional Allocation

When the number of sampling units drawn from a stratum is proportional to the size of the stratum is called proportional allocation.

Number of units selected from stratum h is $n_h = \frac{N_h}{N}(n)$

- Thus, in proportional allocation each stratum is represented according to its size.

eg. If N = 10,000, N_h= 300, and n = 1000, then

$$n_h = \frac{300}{10{,}000} \times 1000 = 30$$

Advantages of Stratifies Random Sampling

1. More representative: In an un-stratified random sample some strata may be over-represented, others may under represented while some may be excluded altogether. Stratified sampling ensures any desired representation in the sample of the various strata in the population.

2. Since the **units** are selected using **probabilistic methods**, stratified random sampling allows us to make statistical measures from the data collected that will be considered to be statistically valid.

3. Greater accuracy: Stratified sampling provides estimates with increased precision than SRS.

4. It enables us to obtain the results of known precision for each of the stratum.

5. This is an effective sampling technique for studying how a trend or issue might differ across subgroups.

6. Administrative convenience: as compared to the simple random sample, in the stratified random sampling, the supervision of the field work could be greater ease and convenience.

7. Sometimes the sampling problems may differ markedly in different parts of the population, e.g. a population under study might be consisting of,

(i) literates and illiterates, or

(ii) people living in institution such as hotels, prisons, hospitals, refugee camps, and those living in ordinary homes, or

(iii) people living in hill areas and plain areas.

In such cases, we can take this different parts as different strata.

Disadvantages

- A stratified random sample can only be carried out only if a **complete list of the population** is **available**.
- Stratification, in some cases is not easy. Comprehensive knowledge over the population is required.
- Non-response error may be higher. This may require re-contacting non-respondents.
- It is more complex to organize and analyze the results compared to simple random sampling.
- Time consuming
- Cost is higher

Cluster Sampling

❖ A random sampling method

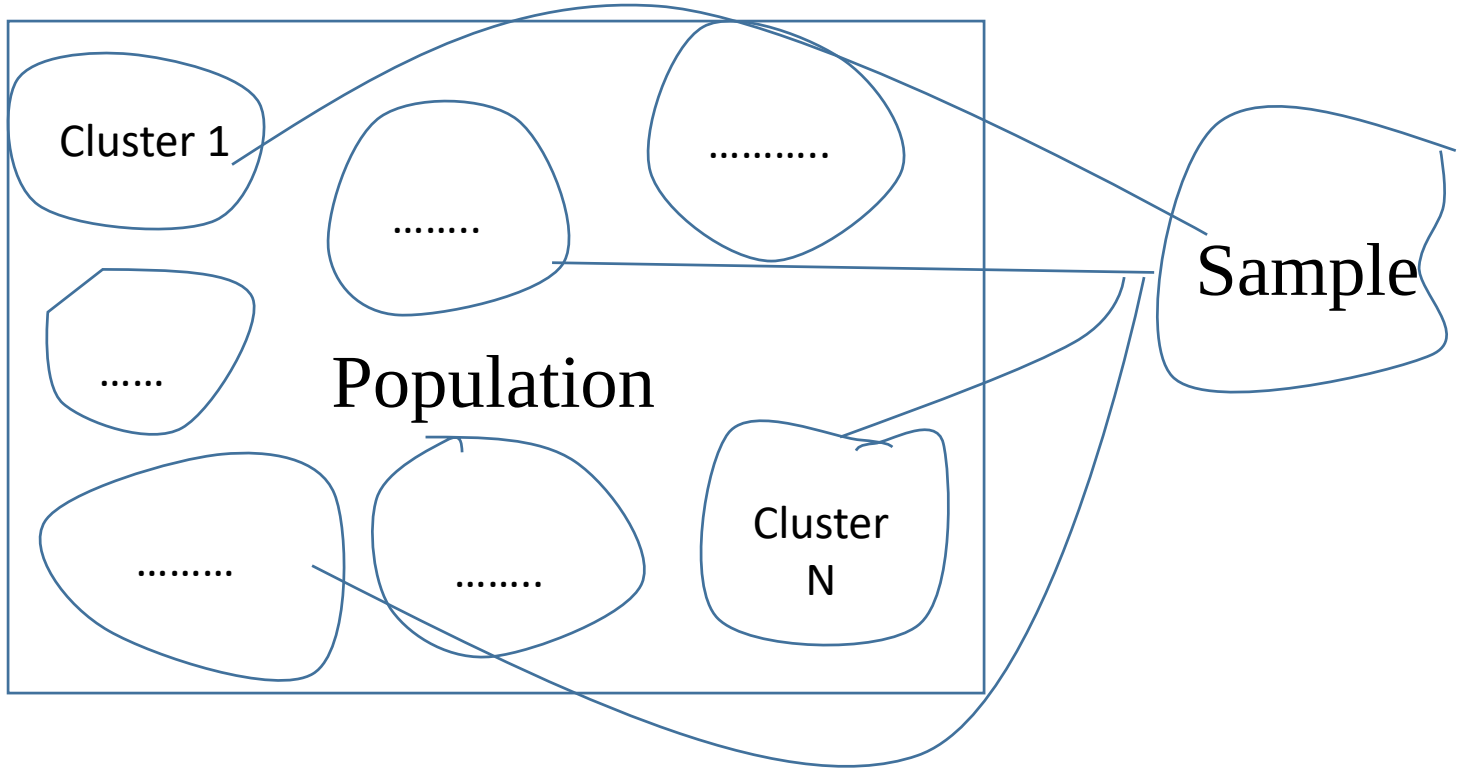
- When there are no proper sampling frame other random sampling methods cannot be applied.
- Some times forming a sampling frame is difficult and costly when the population is dispersed throughout the large geographical area.
- When the population is infinite, it is impossible.
- If the entire population is unclear or unknown, sampling frame cannot be formed.
- If the sample clusters are geographically convenient
- The clusters are 'natural' in a population

In the cluster sampling,

- The population is divided into *N* groups, called **clusters**.
- The researcher randomly selects *n* clusters to include in the sample.
- The number of observations within each cluster M_i is known, then $M = M_1 + M_2 + M_3 + ... + M_{N-1} + M_N$.
- Each element of the population can be assigned to one, and only one, cluster.

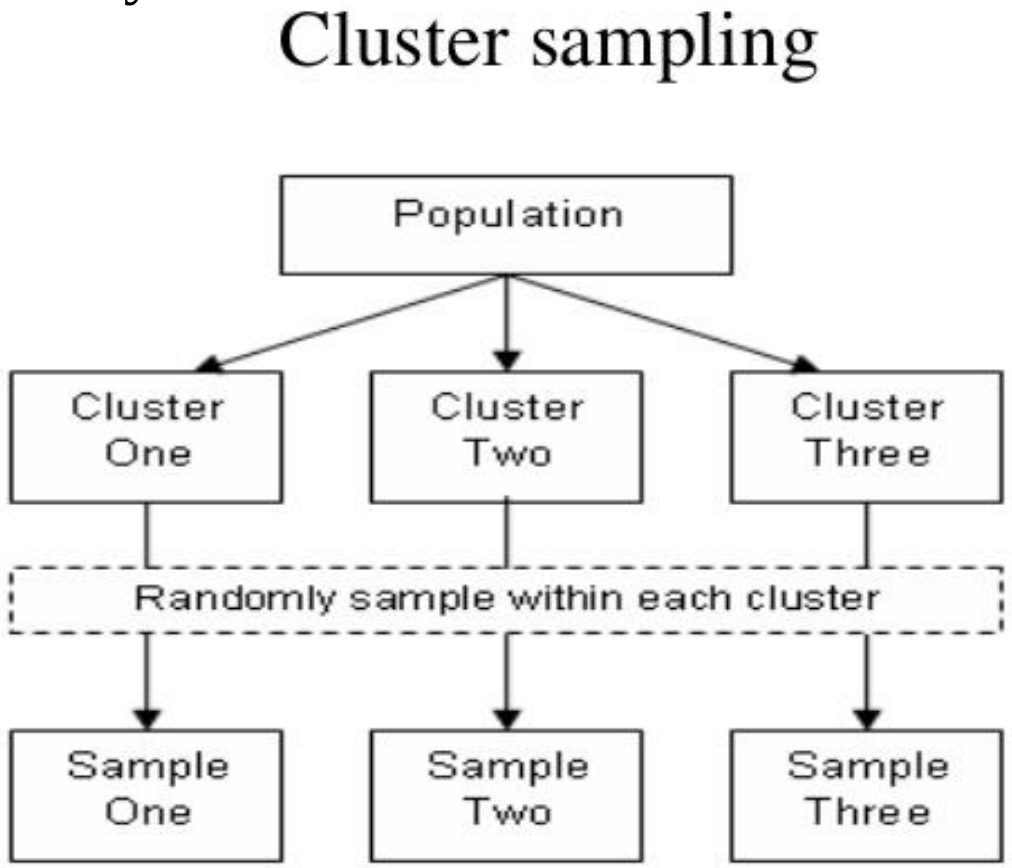
One-stage sampling

- First select few clusters randomly
- Second, all of the elements in the selected clusters are included in the sample
- Data are collected from every units in the selected clusters.



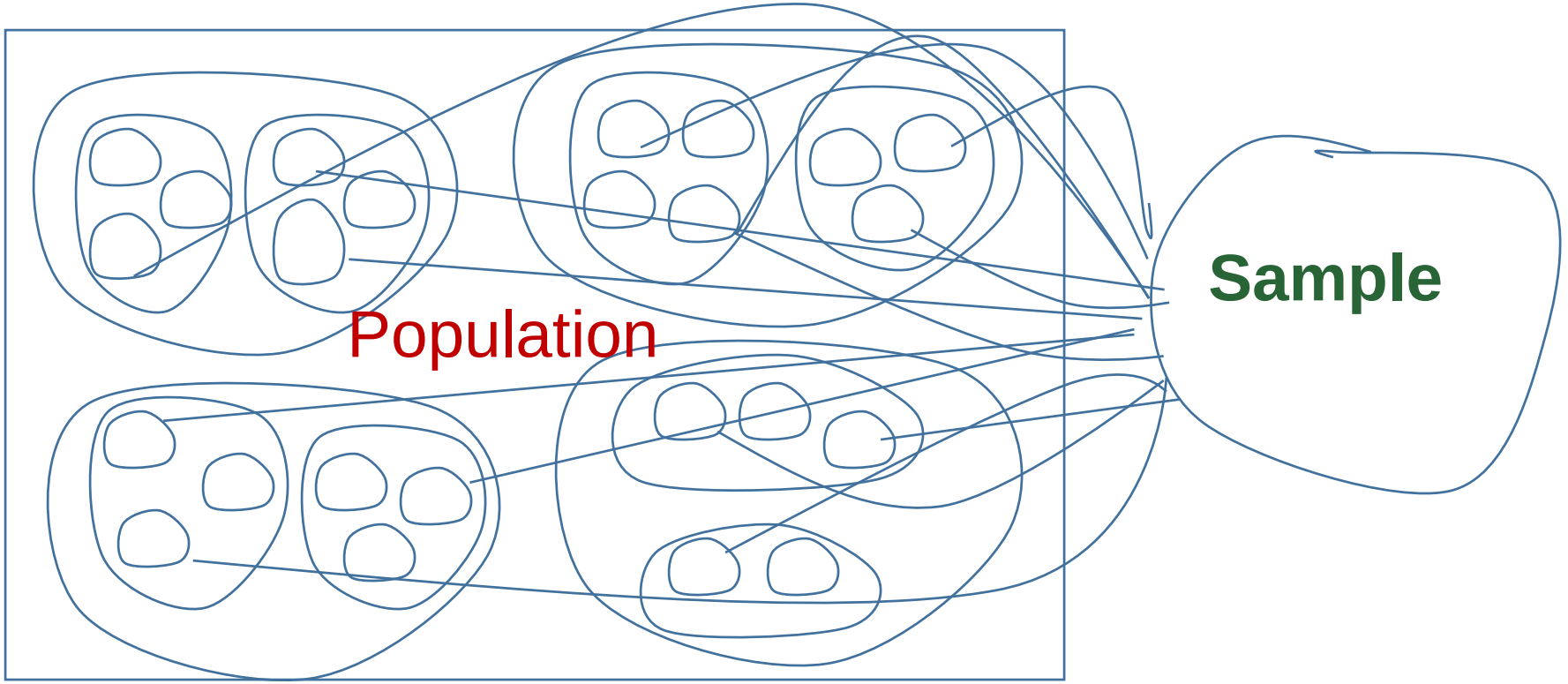
Two-stage sampling

- First, a few clusters are selected randomly
- Second, a subset of elements within selected clusters are randomly drawn to include in the sample.
- Data are collected from the every units selected from each cluster.



Multi-stage Cluster sampling

- Elements are drawn randomly in several steps from selected clusters for inclusion in the sample.
- This can be done easily when subgroups are large and there can be identified subgroups within the subgroups.



Advantages

1. It is the most time-efficient and cost-efficient probability design for large geographical areas
2. This method is easy to be used from practical viewpoint
3. Can easily apply when clusters are naturally formed

Disadvantages

1. Requires group-level information to be known
2. Commonly has higher sampling error than alternative probability sampling methods
3. Precision of measures may be less than the methods of other alternative probability sampling methods
4. May be less representative

Systematic Sampling

- Systematic sampling is a type of probability sampling method.
- Sample members from a larger population are selected according to a random starting point and a fixed periodic interval.
- This interval, called the **sampling interval (k)**, is calculated by dividing the population size (N) by the desired sample size (n).

$$k = \frac{N}{n}$$

If first element: 1, 1+k, 1+2k, 1+3k, 1+4k,...

If second element: 2, 2+k, 2+2k, 2+3k, 2+4k,...

k-th element : k, 2k, 3k, 4k,... nk.

Steps of a systematic sampling

1. Create a list (sampling frame in which numbered from 1 to N).

2. Decide the sample size (n).

3. Decide the sampling interval (k)

4. Select a beginning number (randomly from first k numbers).

5. Gather a list (n) based on the interval number.

Advantages

- A simple and easy technique that produces a random sample that is free from bias.

- Can be used in an office/ laboratory as well as in a natural setting
- Precession is higher at least the precession of a stratified sample which takes 1 unit from one strata.
- Cost-effective and Time-effective when compared with stratified random sampling

Disadvantages

- When there are regular fluctuations or a systematic array in the population, systematic sampling method creates bias sample.

For example, it could be possible that every tenth person in a array of racially diverse population could be a Chines. In such a case, the systematic sample would be biased because it would be composed of mostly (or all) Chines people, rather than reflecting the racial diversity of the total population.

- When N is not a complete multiplication of n, among the possible systematic samples number of elements in one sample is less in one elements than others.

- Entire sample depend on the element taken from among the first k units

- Proper sampling frame is essential