SAMPLE

AND

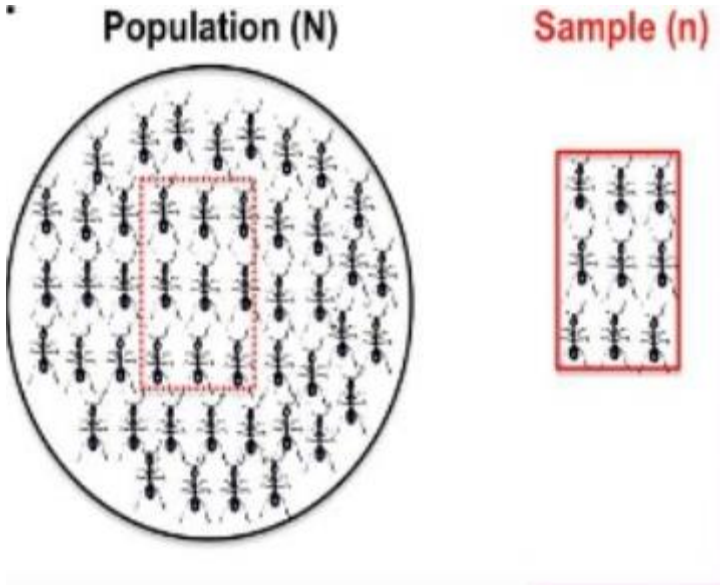
SAMPLING

CHETAN
KUMAR

SAMPLE ?

- A sample is the small collection of the population selected with the objective to show or tell properties of the concerned population.

e.g.



SAMPLING ?

- It is a tool which helps to know the characteristics of unit or population by examining a small part of it.

SAMPLING UNIT?

It is a single element or group of elements subject to selection in a group.

SAMPLE SIZE ?

The number of frequency of the sample unit is called the sample size.

To get the accurate result, the size of the sample should not be too small or too large.

- Example

Examining a handful of grains to ascertain the quality of the entire lot.

Examining 2 or 3 grains of boiling rice to know whether pot of rice is ready or not.

Evaluating a drop of blood to know the blood constitution of the whole body.

PURPOSE OF SAMPLING

- To provide **various types of statistical information of a qualitative or quantitative nature about the whole** by examining a few selected units.

What are the assumptions of probability sampling ?

When sampling method is adopted by the researcher, the basic assumption is that the samples so selected out of the population are the best representative of the population under study. Thus good samples are those who accurately represent the population.

PROCESS OF SAMPLING

- It is the procedure required right from defining a population to the actual selection of sample elements. There are seven steps involved in this process.

1. Define the population.
2. Identify the sampling frame.
3. Specify the sampling unit.
4. Specify the sampling method.
5. Determine the sample size.
6. Specify the sampling plan.
7. Select the sample.

Define the population.

- It is the aggregate of all the elements defined prior to selection of the sample.

E.g.

If we were to conduct a survey on the consumption of tea in Gujarat, then these specifications might be as follows

(i) Element: Housewives

(ii) Sampling units: Households, then housewives

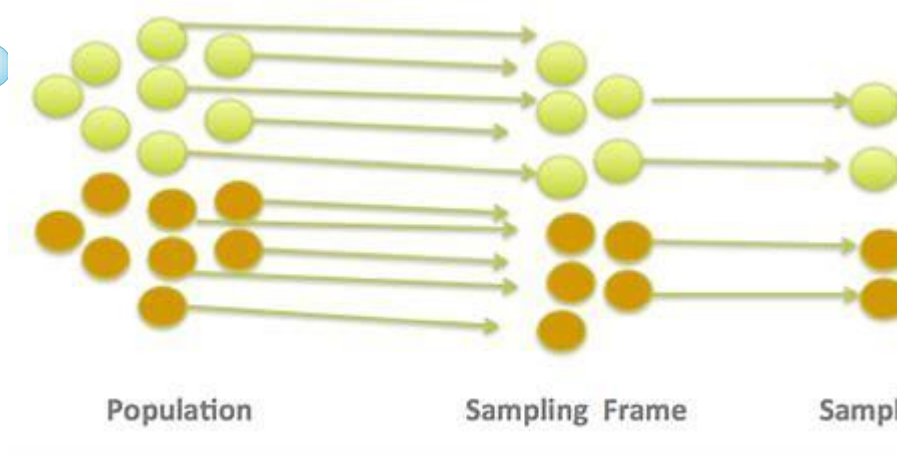
(iii) Extent Gujarat State

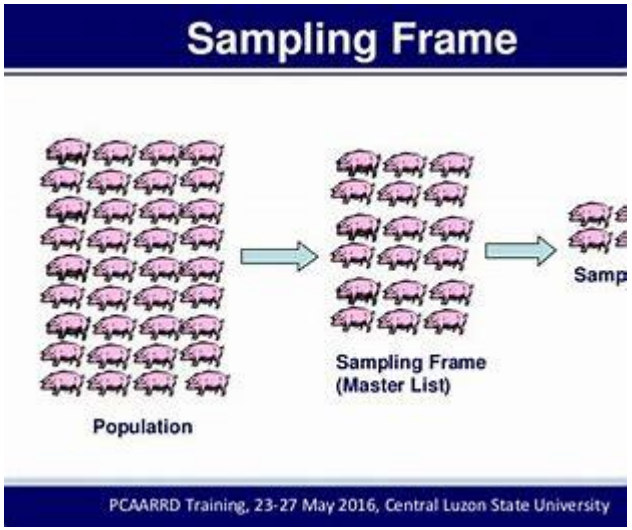
(iv) Time January 1-10, 1999

Identify the sampling frame.

Sampling Frame

- A list of elements from which the sample may be drawn
- The sampling frame is also called the working population, because it provides the list that can be operationally worked with. E.g. UPM CIMB customers' telephone list. If a complete lists is not available, can use maps as sampling frame.
- Sampling frame error - **Sampling frame error** occurs when certain elements are excluded (e.g. those without telephone was excluded) or when the entire population is not accurately represented in the sample frame (e.g. oversea travel survey – include those can't afford to travel overseas – wrong sampling frame).





Specify the sampling unit

- The sampling unit is the basic unit containing the elements/characteristics of the target population. The sampling unit may be different from the individual.

  E.g.

  if one wanted a sample of housewives, it might be possible to have access to such a sample directly. However, it is easier to select households as the sampling unit and then interview housewives in each of the households.

Specify the sampling method

- It indicates how the sample units are selected. One of the most important decisions in this regard is to determine from the two
  - Probability
  - Non-probability sample

Determine the sample size

- In other words, one has to decide how many elements of the target population are to be chosen.

Specify the sampling plan

- This means that one should indicate how decisions made so far are to be implemented.

  For example,

  if a survey of households is to be conducted, a sampling plan should define a household, contain instructions to the interviewer as to how he should take a systematic sample of households, advise him on what he should do when no one is available on his visit to the household, and so on.

Select the sample

PRINCIPLE/LAW OF SAMPLING

- Two main principles on which sampling is based:
  1. Principles of Statistical Regularity
  2. Principle of Inertia of Large Numbers

Principles of Statistical Regularity

- If a sample is taken at random from a population, it is likely to possess almost the **same characteristics** as that of the population.

  For example,

  Man varies to a limited extent in color, height, weight, etc. but it can always be identified as man.

  But **results** derived from the sample data may be slightly different from that of the population.

  e.g. rohu fish of pond is 15 cm, whereas it may be 14.5 cm or 15.5 cm for the sample size.

Principle of Inertia of Large Numbers

- It says that other things being equal, larger the size of the sample, more accurate the results are likely to be.

  (the size of the sample is the number of sampling units which are selected from a population for investigation)

  - Larger the size of the sample more stable it will be.
  - When large numbers are considered the variations in the component parts tend to balance each other and, therefore, the variation in the aggregate is insignificant.

  E.g.

  To calculate the chance of baby boy/girl in a delivery if small sample is taken e.g. 10 women are observe then there are chances that ration will not be 50-50% but it can be 9-1, 8-2 or 7-3, etc. so with a large sample it has a more chance of 50-50% boys and girls in a delivery.

ESSENTIALS OF SAMPLING

- For useful sample results, it is necessary that a sample possesses the following essentials:
  1. Representativeness
  2. Adequacy
  3. Independence
  4. Homogeneity

Representativeness

- It means the selected sample from the population should be representative of the population i.e. have the same characteristics as that of the population.

  To ensure this random method of selection should be used.

Adequacy

The size of the sample should be adequate to make the result more reliable.

Independence

- The individual items composing the sample should be selected independently of each other and all items of the universe should have same chance of being selected in the sample.

HOMOGENEITY

Two samples taken from the universe should give approximately the same result.

METHODS OF SAMPLING

- All methods can be grouped into two-
  1. **Random Sampling (Probability Sampling)**
  2. **Non-Random Sampling**

Random Sampling

- It is also known as **probability sampling** because, every element has the same chance of selection and thus laws of probability can be applied.

Random sampling basically means selection of element/unit should be made without deliberate discrimination.

For scientific research purposes, random sampling is only applied.

Two types:

1. **Simple Random Sampling**
2. **Restricted Random Sampling**

Simple Random Sampling

- It refers to the sampling technique in which each and every item of the population is given an equal chance of being included in the sample.

Properties

- Free from personal bias
- Investigator does not exercise his discretion or preference in the choice of items
- Selection of item depend entirely on chance.
- Not always used as a primary sampling procedure but necessary to introduce an element of randomness in the final selection.
- If the sample chosen at random has a sufficient large size it will represent all groups in the universe.

So this method is also known as method of chance selection /representative sampling.

Demertis

Difficult and expensive

METHODS OF SIMPLE RANDOM SAMPLE

- Two methods:
  1. **Lottery method**
  2. **Table of random numbers**

Lottery Method

- Simplest and the most popular.

  Process includes, numbering/naming the all items of the universe/population on the separate slips of paper of **identical shape and size and color** and then are folded and mixed up in a container or drum. A blindfold selection is then made of the number of slips required to constitute the desired size of the sample. The selection of items thus depends entirely on chance.

Example-

To choose a sample of 100 patients from 1000 patients, name of all the patients were written on separate slips of paper, fold these slips, mix them and then make blindfold selection of 100 slips.

Limitation

Difficult if the population size is very large.

TABLE OF RANDOM NUMBERS

- A random number table is a **series of digits (0 to 9) arranged randomly in rows and columns.**

There are three such tables available:

Tippet's table of random number

Fisher and Yate's number

Kendall and Babington Smith numbers

The most commonly used is the Tippet's table

It is consists of 41,600 digits, arranged into four-figure numbers.

2952	6641	3992	9792	7969	5911	3170	5624
4167	9524	1542	1396	7203	5356	1300	2693
2370	7483	3408	2762	3563	1089	6913	7691
0560	5246	1112	6107	6008	8126	4233	8776
2754	9143	1405	9025	7002	6111	8816	6446

- It is important that the starting point in the table of random numbers be selected in some random fashion so that every unit has an equal chance of being selected.

Example-

Suppose we have to select 20 rats out of 6,000 kept in a big case. The procedure is to number all the 6000 items from 1 to 6,000. A page from Tippett's table may then be selected and the first twenty numbers upto 6,000 noted down. Rats bearing those numbers will be included in the sample.

This method cannot be followed in case of qualitative data.

MERITS OF RANDOM SAMPLING

- **LESS TIME** Time is saved not only in collecting data but also in processing it.

LESS COST less costly because few items are studied in sampling.

Best for infinite population

More reliable results two reasons for it

1. Possible to determine the extent of sampling error.
2. Errors in other types of survey are more serious.

The destructive nature of certain tests

as toxicity tests can not be applied on entire population so a sample is required for it.

LIMITATIONS

- Small size of population
- Result may be false, inaccurate and misleading if the sample has not been drawn properly.
- Personal biasness may be present there due to choice of technique and drawing of sampling units.
- Complicated sampling plan that may require more time, labour and money than a complete count.
- If information is required for each and every unit in the domains of study, a complete enumeration survey is required.

RESTRICTED RANDOM SAMPLING

- Three types
  1. **Systematic sampling**
  2. **Stratified sampling**
  3. **Cluster sampling**

Systematic sampling

- In this method every K^{th} item of the list is selected for sampling.

$$K = \frac{size\ of\ the\ universe}{sample\ size}$$

This method is used in those cases where a complete list of the population from which sampling is to be drawn is available.

Example-

If a pond consists of 1000 fishes of a species and if we want to draw a sample of 200 fishes this means we must take every fifth item (i.e., K=5)

$$K = \frac{1000}{200} = 5$$

This mean the first item between one and five shall be selected at random. Suppose it comes out to be 3. Now we shall go on adding five and obtain numbers of the desired sample. Thus, the second item would be the 8^{th} fish; the third, 13^{th} ; the fourth, 18^{th} fish; and so on.

MERITS

- It is more convenient to adopt than the simple random sampling.

The time and labour involved by this sampling are relatively smaller.

The results obtained are also found to be generally satisfactory.

If populations are sufficiently large, homogeneous and each unit is numbered, then, this method can give accurate results.

LIMITATIONS

It can become less representative if we are dealing with populations having hidden periodicities.

STRATIFIED SAMPLING

- In this method population is first divided in to homogenous groups or classes called strata. Then the sample is drawn from each stratum by simple random method in proportion to its size. The resulting sample is called as the stratified sample.

This method is followed when population is not homogeneous.

This method is more representative than simple random sampling in a given large population.

TYPES

1. Proportionate Stratified Sampling
2. Disproportionate Stratified Sampling

Proportionate Stratified Sampling

- In this method the number of items drawn for the sample making is directly proportional to the ratio of the item in population.

Example-

If population is divided into four stratas, their respective sizes being 15, 10, 20 and 55% of the population and a sample of 1000 is to be drawn, the desired proportion sample may be obtained in the following manner:

From stratum one 1000 (0.15) = 150 items

From stratum two 1000 (0.10) = 100 items

From stratum three 1000 (0.20) = 200 items

From stratum four 1000 (0.55) = 550 items

Sample size = 1000

Disproportionate Stratified Sampling

- In this method the number of items drawn for the sample making is not directly proportional to the ratio of the item in population.

Example-

If population is divided into four stratas, their respective sizes being 15, 10, 20 and 55% of the population and a sample of 1000 is to be drawn, the desired **disproportion** sample may be obtained in the following manner:

From stratum one = 250 items

From stratum two = 250 items

From stratum three = 250 items

From stratum four = 250 items

Sample size = 1000

- Proportional stratification yields a sample that represents the universe with respect to the proportion in each stratum in the population.

  This procedure is satisfactory if there is no great difference in variation from stratum to stratum.

  But it is not efficient procedure if there is considerable variation in different stratas.

Merits

- Gives greater accuracy

Stratified samples can be more concentrated geographically.

LIMITATIONS

Utmost care must be exercised in dividing the population into various strata. Failing the point the results may not be reliable.

CLUSTER SAMPLING

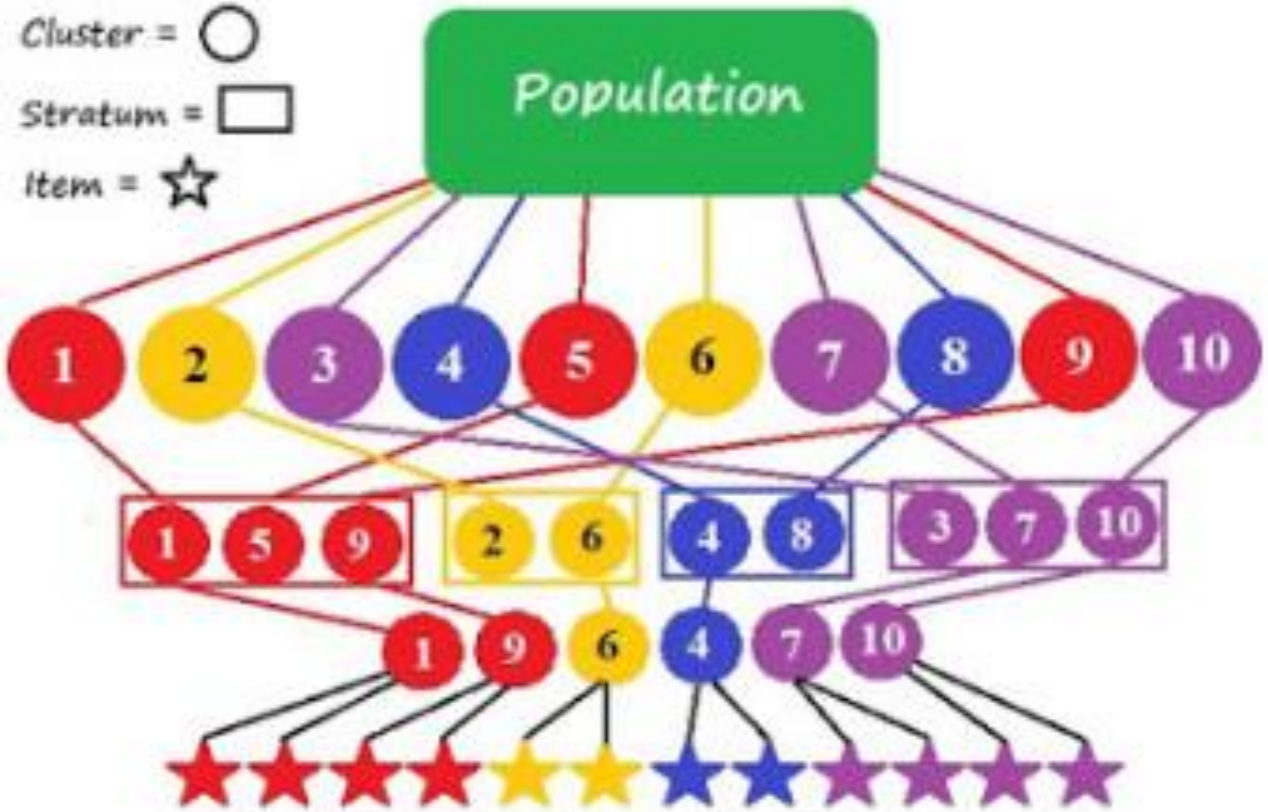
- A cluster is randomly selected group. This method is used when units of population are natural groups such as hospital ward, slums of a town, etc.

CLUSTER SAMPLING

- The population is divided into subgroups (clusters) like families.
- A simple random sample is taken from each cluster

MULTISTAGE SAMPLING

- It is a complex form of cluster sampling, sometimes, also known as multistage cluster sampling. During this sampling method, significant clusters of the selected people are split into **sub-groups** at various stages to make it simpler for primary data collection.
- Variety of sampling methods can be used here.
- Most **surveys conducted by professional** polling organization use some combination of stratified and cluster sampling as well as simple random sampling.



Process

- In this method, the whole population is divided first stage sampling unit from which random sampling are selected.

The selected first stage is then subdivided into second stage units from which another sample is selected.

Third and fourth stage sampling is done in same manner if necessary.

Types

- Two types

Multistage cluster sampling:

Multistage cluster sampling is a complex type of cluster sampling. The researcher divides the population into groups at various stages for better data collection, management, and interpretation. These groups are called clusters.

Multistage random sampling:

The concept of multistage random sampling technique is similar to multistage cluster sampling. But in this case, the researcher chooses the samples randomly at each stage. Here, the researcher does not create clusters, but he/she narrows down the sample by applying random sampling.

- Here's an example of a multistage design.
  - Let's consider the sample location as the USA. The research goal is to assess the online spending trends of people in the US through an online questionnaire. Researchers can form their sample group comprising 200 households in the following manner:
  - Firstly, choose the number of states using simple random sampling (or any other probability sampling). For example, select ten states.
  - Secondly, choose five districts within each state using the systematic sampling method (or any other probability sampling).
  - Thirdly, choose four households from each district using the systematic sampling or simple random sampling method. You will end up with 200 houses that you can include in the sample group for research.

Applications of multistage sampling

1. Applied to a multistage design where the **population is too vast** and researching every individual is impossible.
2. To gather student perceptions from students belonging to various colleges, studying different courses and located throughout the country.
3. To survey employees of a multinational company belonging to multiple locations in multiple countries.
4. Government bureaus use this method all the time to draw inferences from the population.

Advantage

1. It allows researchers to **apply cluster or random sampling** after determining the groups.
2. Researchers can apply multistage sampling to make clusters and sub-clusters until the researcher **reaches the desired size** or type of group.
3. Researchers can divide the population into groups without restrictions. It allows flexibility to the researchers to choose the sample carefully.
4. It is **useful while collecting primary data** from a geographically dispersed population.
5. **Cost-effective and time-effective** because this method helps cut down the population into smaller groups.
6. Finding the right survey sample becomes very convenient for researchers.
7. It does not need a complete list of all the members of the target population, dramatically reducing sample preparation cost.

Disadvantage

1. Each stage in sampling introduces sampling error- the more stages there are, the more error there tends to be.

NON-RANDOM SAMPLING

CHETAN KUMAR

JUDGEMENT SAMPLING

- In this sampling technique the members are chosen only on the basis of the researcher's knowledge and judgment.
  - There are chances that the results obtained will be highly accurate with a minimum margin of error.
  - Judgmental sampling is most effective in situations where there are only a restricted number of people in a population who own qualities that a researcher expects from the target population.

For instance, in situations where a researcher conducts convenience sampling to gather feedback from professors about their university but the fact that there are high chances of the results to be skewed, researchers prefer judgmental sampling to select those professors who will provide 100% feedback about the university.

Judgement Sampling Advantage

1. **Consumes minimum time for execution:** In this sampling approach, researcher expertise is important and there are no other barriers involved due to which selecting a sample becomes extremely convenient.
2. **Allows researchers to approach their target market directly:** There are no criteria involved in selecting a sample except for the researcher's preferences. Due to this, he/she can communicate directly with the target audience of their choice and produce desired results.
3. **Almost real-time results:** A quick poll or survey can be conducted with the sample using judgmental sampling since the members of the sample will possess appropriate knowledge and understanding of the subject.

Disadvantage

1. Selecting each individual of the sample
2. researcher may or may not have the appropriate proficiency to conduct an effective sampling process.

QUOTA SAMPLING

- Quota sampling is defined as a non-probability sampling method in which researchers create a sample involving individuals that represent a population. Researchers choose these individuals according to specific traits or qualities.

  For example, a cigarette company wants to find out what age group prefers what brand of cigarettes in a particular city. He/she applies quotas on the age groups of 21-30, 31-40, 41-50, and 51+. From this information, the researcher gauges the smoking trend among the population of the city.

Types

- **Controlled quota sampling:**

  Controlled quota sampling imposes restrictions on the researcher's choice of samples. Here, the researcher is limited to the selection of samples.

  Uncontrolled quota sampling:

  Uncontrolled quota sampling does not impose any restrictions on the researcher's choice of samples. Here, the researcher chooses sample members at will.

Advantages of quota sampling

- Here are the top four advantages of quota sampling

Saves time: Because of the involvement of a quota for sample creation, this sampling process is quick and straightforward.

Research convenience: By using quota sampling and appropriate research questions, interpreting information and responses to the survey is a much convenient process for a researcher.

Accurate representation of the population of interest: Researchers effectively represent a population using this sampling technique. There is **no room for over-representation** as this sampling technique helps researchers to study the population using specific quotas.

Saves money: The budget required for executing this sampling method is minimalistic.

Disadvantage

- Not as representative of the population as a whole as other sampling methods.

  Because the sample is non-random it is impossible to assess the possible sampling error.

SAMPLING ERROR

- The error which arises due to the drawing inferences about the population on the basis of few observations (sample) is termed sampling error.

The errors mainly arising at the stages of ascertainment and processing of data which are termed non-sampling errors are common in both complete enumeration and sample survey.

Types

1. **Biased errors**
2. **Unbiased errors**

Biased errors

Theses errors arise from any bias in selection, estimation, etc.

Unbiased errors

These errors arise due to chance differences between the members of population inluded in the sample and those not included.

Thus total sampling error is made up of error due to bias, if any, and the random sampling error.

Cause

1. Faulty process of selection
2. Faulty work during the collection of information
3. Faulty methods of analysis

Avoidance of Bias/Reducing Sampling Errors

1. Sample to be drawn entirely at random
2. Specific problem selection
3. Systematic documentation of related research
4. Effective enumeration
5. Effective pretesting
6. Controlling methodological bias
7. Selection of appropriate sampling techniques

Non-Sampling Error

- Non-sampling errors refers to biases and mistakes in selection of sample.

- **CAUSES FOR NON-SAMPLING ERRORS**
  - Sampling operations
  - Inadequate of response
  - Misunderstanding the concept
  - Lack of knowledge
  - Concealment of the truth.
  - Loaded questions
  - Processing errors
  - Sample size