BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Sampling Theory

Course Code: 23ST03CC

Unit-I

SAMPLING - INTRODUCTION

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. N. Saranya
Guest Faculty
Department of Statistics

SAMPLING THEORY- 23ST03CC

UNIT - I

Sampling theory is a branch of statistics that focuses on selecting a sample from a population to estimate the population's characteristics:

- **Purpose :** Sampling theory is used to estimate population parameters, such as the mean, variance, and proportion of a specific unit.
- **When it's useful :** Sampling theory is especially useful when it's impractical or impossible to observe the entire population.
- **How it works :** Sampling theory provides methodologies for selecting a sample that accurately reflects the population.
- **Benefits :** Sampling is less expensive and faster than collecting data from the entire population.
- **Applications :** Sampling is used in business, medical research, and survey methodology.

Some key ideas in sampling theory include:

- **Stratification**: Dividing the population into groups so that the variability between groups is greater than within groups.
- **Clustering**: Grouping the population into identifiable clumps for easier sampling.
- **Sample size**: As the sample size increases, the sample's properties become more similar to the population's.

Some key points and definitions in sampling theory includes:

Population: A population is the complete set group of individuals, whether that group comprises a nation or a group of people with a common characteristic. It is denotedby (N).

Sample: It is the subset of a population that has been collected through data collection. It is denoted by (n).

Statistic: Any function of the sample values is called statistic.

Event: Event sampling is used to sample behavior in observational research. It is where an observer records the number of times a certain behavior occurs**.**

Estimator: An estimator is a random sample and may take different values from sample to sample.

Parameter: A parameter is a number describing a whole population (e.g., population mean), while a statistic is a number describing a sample (e.g., sample mean). The goal of quantitative research is to understand characteristics of populations by finding parameters**.**

Standard Error: The positive square root of variance is called standard error of the estimator.

Finite Population: A population said to be finite if it has countable number of values. For example: number of students in the Bharathidasan University.

Infinite Population: A population said to be infinite if it has uncountable number of values. For example: Number of stars in the sky.

Sampling Unit: The contribution of population which all the individuals to the sample from the population that cannot be further subdivided for the purpose of sampling at time is called sampling unit. For example, to know that the average income of a family, the head of familyis called sampling unit.

Sampling Frame: we need to frame the structure of the survey for adapting any sampling procedure, it is essential to have a list or map to identify each sampling unit by a number, such a list or map is called sampling frame. For example, If list of voters in a particular place.

Need for Sampling

The sampling methods have been extensively used for a variety of purposes and in great diversity of situations. In practice it may not be possible to collect the information on all unitsof a population due to various reasons such as

- ❖ Lack of resources in terms of money, personnel and equipment.
- ❖ The experimentation may be destructive in nature. Eg- finding out the germination percentage of seed material or in evaluating the efficiency of an insecticide. This experimentation is destructive.
- ❖ The data may be was useful if they are not collected within a time limit. The census survey will take longer time as compared to the sample survey. Hence for getting quick results sampling is preferred. Moreover a sample survey will be less costly than complete enumeration.
- ❖ Sampling remains the only way when population contains infinitely many number of units.
- ❖ Greater accuracy.

CENSUS AND SAMPLE SURVEY:

The total count of all units of the population for a certain characteristic is known as complete enumeration or census survey. The money, man power and time required forcarrying out complete enumeration is generally large and there are many situations with limited means where the complete enumeration will not be possible. When only a part of the population is selected denoted as sample and examine it is called sample enumeration or sample survey.

Limitations of Sampling vs. Census:

- **Less time:** There is considerable saving in time and labor since only a part of the population has to be examined. The sampling results can we obtain more muchrapidly and it can analysis such faster since relatively and process.

- **Reduced cost of the survey:** Sampling usually results in reduction of cost, in terms of money and in terms time. All though the amount of labor and expenses are invalid in collecting information. All generally greater per unit of the sample then complete enumeration the total cost of the sample survey is the expected to be much smaller than that of complete census of since in most of the cases our resources or limited in terms of money and the time with in which the results of the survey should be obtain it is usually imperative to the sampling rather than complete to sampling rather than complete enumerations.

- **Greater accuracy of results:** The results of the sample are usually much morereliable than those obtained from a complete census due to the following reasons.

    - It is always possible to determine the extent of the sampling errors.

    - Non-sampling errors due to factors such as training of the field workers measuring and recording, observations, location of units incompetents of returns biased due to interviews etc. There are likely to be of a serious naturein complete census than in a sample survey. Non sampling errors can be controlled more effectively by employing more qualified and better trained personal better supervision and better equipment for processing and analysis of relatively limited data. Moreover, it is easier to guard against incomplete and inaccurate returns.

- **Greater scope:** Sample survey generally has greater scope as compared with complete census the complete enumeration is impracticable rather inconceivable if the survey requires highly trained personal and more sophisticated equipment for the collection and analysis of the data. Since sample survey saves in time and money it is possible to have a through and intensive enquiry because detailed information can be obtained from a small group of respondents.

- If the population is too large, for example of trees in a jungle we are left with no way but to resort to sampling.

- If testing is destructive i.e., if the quality of an article can be determined only of an article in the process of testing as for example:

    - ✓ Testing the quality of milk or chemical salt by analysis.

    - ✓ Testing the breaking strength of chalks.

    - ✓ Testing of crackers and explosives.

    - ✓ Testing the life of an electric tube or bulb etc.

Complete enumeration is impractical and sampling techniques is the only method tobe used in such cases.

- If the population is hypothetical for example while tossing a coin, the process may continue indefinitely. Sampling method is the only scientific method of estimating parameters of the universe/population.

PRINCIPLE STEPS IN A SAMPLE SURVEY

- **Objectives of the survey:** The first step is to define clear and concrete terms of the objectives of the survey. The sponsors of the survey should take care that these objectives are commensurate with the available resources in terms of money, man power and the time limit required for theavailability of the results of the survey.

- **Defining the population to be sampled:** The population that is the aggregate of objects from which sample is chosen should bedefined in clear and unambiguous terms.

- **The frame and sampling units:** The population must be capable of division into sampling units for purpose of the sample selection. The sampling units must cover the entire population and must be distinct unambiguous and not overlapping in the sense that every element of the population belongsto one and only one sampling units. In order to cover the population decided upon there should be some list map or other acceptable material called the frame to serve as a guide for the population to be covered.

- **Data to be collected:** The data should be collected keeping in view the objectives of the survey. We should not have the tendency to collect too many data some of which are never subsequentlyexamined and analyzed.

- **The questionnaire or schedule:** Having decided about the type of the data to be collected the next important part of the sample survey is the construction of the questionnaire or schedule of the enquires which requires skill special technique as well as familiarity with the subject matter under study. The question should be clear, brief collaborative non offending courteous in tone unambiguous and to the point, so that not much scope of guessing is the left on the part of the respondent orinterviewer suitable and detailed instruction for filling up the questionnaire or scheduleshould also prepared.

- **Method of collecting information:**
  - *Interview method:* In this method the investigator goes from house to house and interviews the individuals personally. He asks the questions one by one and fills up the schedule on the basis of the information guess by the individuals.
  - *Mailed questionnaire method:* In this method the questionnaire is mailed to the individuals are required to theindividuals who are required to fill up and return it duly completed.

- **Non-respondent:** Quite often the data cannot be collected for the sampled units. This incompleteness is called non response which obviously tends to change the results. In such cases of response should be handled with caution in order to draw unbiased and valid conclusions.

- **Selection of proper sampling designs:** The size of the sample (n) the procedure of selection and the estimation of thepopulation parameters along with their margins of uncertainty are some of the important statistical problems that should receive that careful attention. A number of designs for the selection of a sample are available and a judicious selection will guarantee good and available estimation.

- **Organization of field work:** It is absolutely essential that the person should be thoroughly trained in locating the sample units, recording the measurements the methods of the collection of required data before starting the field work. The success of a survey to a great extent depends upon the reliable field work. It is very necessary to make provision for adequate supervisory staff for inspection after field work.

- **Summary and analysis of the data:**
  - ***Scrutiny and editing of the data:*** An initial quality check should be carried out by the supervisory staff whilethe investigators are in the field.
  - ***Tabulation of the data****:* Before carrying out the tabulation of the data we must decide about theprocedure for the quality of the data. For the large scale survey, mission tabulationwill obliviously be much quicker and economical.
  - ***Statistical analysis:*** Statistical analysis should be made only after the data has been properly scrutinized, edited and tabulated. Different method of estimation may be available for the same data, appropriate formulae should be used to provide final estimates of the required information.
  - ***Reporting and conclusions:*** Finally, the report incorporating detailed statement of the different stages ofthe survey should be prepared.

- **Information gained for future surveys:** Any complete survey is helpful in providing a note of caution and taking lesson fromit for designing future surveys. The information gained from any completed sample in the form of the data regarding means, standard deviation and the nature of the variability of the principle of the measurements tougher sampling.

PRINCIPLE OF SAMPLING SURVEY

- **Principle of statistical regularity:** This principle as its origin in the mathematical theory of probability. According to the king, the law of statistical regularity lays down moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of the large principle stress the desirability and the importance of selecting the sample at random so that each every unit in the population has an equal chances of being selecting the sample.

- **Principle of validity***:* By the validity of a sample design we mean that it should enable us to obtain validtest and estimates about the parameters of the population. The samples obtain by the techniques of the probability sampling, satisfy this principle.

- **Principle of the optimization***:* The principle improves upon obtaining optimum results in terms of efficiency and the cost of the design with the resources at our disposal. The reciprocal of the sampling variance of an estimate provides a measure of its efficiency while a measure of the cost of the design is provided by the total expenses incurred in terms money and man hour.
  - Achieving a given level of efficiency at minimum cost 1 and
  - Obtaining maximum possible efficiency with given level of cost.

LIMITATION OF SAMPLING

Advantages:

- The sampling units are drawn in scientific manner.
- Appropriate sampling techniques are used and
- The sample size is adequate.

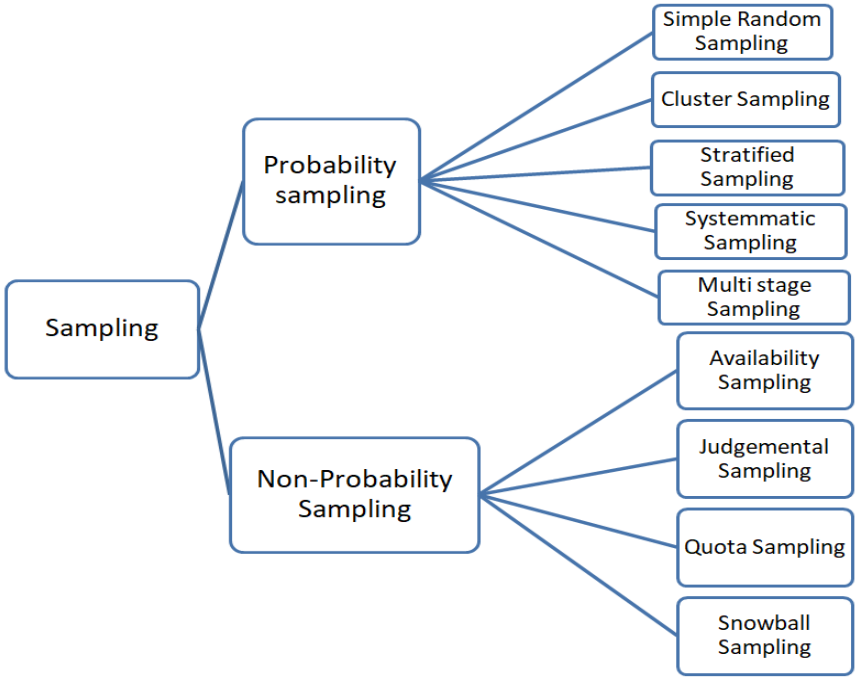
Disadvantages:

- Proper care should be taken in the planning and execution of the sample, survey otherwise the results obtained might be inaccurate and misleading.
- Sampling theory requires services of the trained and qualified person and sophisticated equipment for its planning execution and analysis. In the absence of these results, the sample survey are not trust worthy.

However, if the information required about each and every unit of the universe there is no way but to resort to complete enumeration. If time and money are not important factors or if the universe is not too large, a complete census maybe better than any sampling.

TYPES OF SAMPLING

The technique or method of selection is fundamental importance in the theory of samplings and usually depends upon the nature of the data and the types of the enquiry the procedure a selecting a sample may be broadly classified under the following three heads.



- **Simple Random Sampling**: In this type of sampling, members are chosen randomly from the population, merely by chance. This can be done by either putting chits in a bowl like a lottery system or spinning the wheel. The advantage of simple random sampling it that it is easy cost-efficient, reliable and represents the whole population.
- **Cluster Sampling**: In this type of sampling, the whole population is divided into some groups or clusters. Units with similar characteristics are kept in one cluster. For example, People can be grouped according to their age or country.

- **Stratified Sampling**: In this type of sampling, the population is divided into strata according to some similar characteristics. Then members are chosen from each stratum to make the final sample. The advantage of this type of sampling is that we can have all types of samples from all the groups.
- **Systematic Sampling**: In systematic sampling, every nth unit from the population is taken. That means a sample from the population is selected at every regular interval. The starting point is selected randomly and after that, every nth element is selected. In the below figure, n=3, so every 3rd element is selected.

The types of Non-Probability Sampling are:

- **Availability Sampling:** This is also known as **convenience sampling.** This occurs when the researcher selects the samples based on availability. For example: If a student wants to do research on how many college students are using the canteen for lunch. He will select his own college and nearby colleges to do the survey.
- **Judgmental Sampling:** It is also called **purposive sampling**. In this samples are selected on basis of the researcher's own knowledge, experience and intuition. The researcher selects this technique when they feel that other sampling techniques are time-consuming and he is confident about his knowledge.
- **Quota Sampling:** In this type of sampling, the researcher divides the population into some quotas according to some characteristic and select the members from each quota.
- **Snowball Sampling:** This is also known as **chain-referral sampling**. In this, reference from existing samples is taken to collect the samples.

Probability sampling vs Non-probability Sampling Methods

The below table shows a few differences between probability sampling methods and non-probability sampling methods.

Probability Sampling Methods	Non-probability Sampling Methods
Probability Sampling is a sampling technique in which samples taken from a larger population are chosen based on probability theory.	Non-probability sampling method is a technique in which the researcher chooses samples based on subjective judgment, preferably random selection.
These are also known as Random sampling methods.	These are also called non-random sampling methods.
These are used for research which is conclusive	These are used for research which is exploratory.
These involve a long time to get the data.	These are easy ways to collect the data quickly.
There is an underlying hypothesis in probability sampling before the study starts. Also, the objective of this method is to validate the defined hypothesis.	The hypothesis is derived later by conducting the research study in the case of non-probability sampling.

QUESTIONNAIRE

Questionnaire consists of a list questions parading to the enquiry is preferred to have a blank space for answer. This questionnaire is sent respondence who are expected to writethe answers in the blank space. A covering letter is also sent along with the questionnaire requesting the respondence to extent their full cooperation by giving the correct replace and retaining the questionnaire duly field in time.

Merits

a. Questionnaire method is economical.
b. It can we widely used when the area of investing is large.
c. It saves money labor and time.
d. Error in the investing is very small because the information is explained directlyto the respondence.

Demerits

i. In this method there is no direct connection between the investigator and the respondent. Therefore we can sure about the accuracy and reliability of the information.
ii. This method is suitable only for literate people in many countries illiterate people cannot and replay the questionnaire.
iii. There is long delay in receiving questionnaire duly field in time.
iv. People may not give correct answer, thus one is lead to false conclusion.
v. Sometimes the information may not will to gives the return answer.

SCHEDULE

It is most widely used method of collection of primary data. The number of enumeration are selected and trained, they all provide with standard questionnaire specific training are instructions are given to them for filling of schedule each enumerates will be charge of a certain area the investigator goes to respondents along with the questionnaire and gets answer to the question in schedule and records their answers he explain clearly theadjective and purpose of the enquiry.

Merits

i. This method is very useful in extensive enquires.
ii. It yield reliable accurate result because the enumerates or educated and trained.
iii. The scope of the enquiry can also be greatly enlarged.
iv. Even if the respondence it liberate this technique can be widely used.

As the enumerators personally obtain the information there is less Choice for the non-respondence.

Demerits

i. This method is an expensive.
ii. This method is time consuming because the enumerators go personally to obtaine the information.
iii. Personal bias of the enumerators may lead to false conclusion.
iv. The quality of the collected data depends upon the personal qualities of the enumerators.

SAMPLING AND NON-SAMPLING ERRORS

The errors involved in the collection, processing and analysis data in a survey may be classified as Sampling error and Non-sampling errors.

Sampling error: The error which arises due to only a sample being to estimate the population being used to estimate the population parameter is tripped sample error or sampling fluctuations. The error is inherent and available in any and every sampling scheme. A sample with the smallest sampling error will always be consisted a good reprehensive of the population. This error can be reduced by increasing the size of the sample unfair the degree of sampling error is inversely proportional to the square root of the sample size.

Non-sampling error: Besides sampling error, the sample estimate may be subject to the other errors which group together is term non sampling error. The main sources of non-sampling errors are,

- Failure to measure some of the units in the selected sample.
- Observational errors due to defective measurements technique.
- Errors intuition editing, coding and tabulation of the results.
- In practice the census survey results many suffer from non-sampling error. Although these may be free from sampling error. The non-sampling error is likely to increase with increase in sample size while sampling error decrease with increase in sample size.

Sampling errors

Sources and types of sampling and non-sampling errors: The non-sampling errors occur at any one are more of this stages of the survey planning field work and tabulation of the survey data. These errors are badly classify errors follows:

Type 1: Non-Response Error : Errors resulting from inadequate preparation.
Type 2: Responses Error : Error resulting in the stage of the collection or taking observations.
Type 3: Tabulation Error : Errors resulting from data processing.

Type 1: Non response error : These errors may be assigned mainly,

- Due to the use of faulty frame of the sampling units,
- Biased method of the selection of units.
- In adequate schedule,

If the sampling frame is not abraded or old frame is use on account of economic or time saving device it may lead to bias as the targeted population is not enumerated. These a use of such frames may lead either to inclusion of some units not belonging to the population or to omission of units belonging to the population. Such procedure mat brings unknown bias. In some situation apart of sampled units may refuses to respond to the question or may be not at home at the time of interview. It may also lead to this type of error. It can be seen that the method will provide biased estimate. Some of the main sources assigned to these errors may be as follows:

- Omission on duplication of units due to ambiguous defection of Local units or wrong identify of units and or in accurate and inconsistent o objectives.
- In accurate methods of interview or inappropriate schedule.
- Difficulty arising arrivers and Sourness on the part of respondents or faulty method of enumeration data collection.

Type 2: Response Error

These errors refer in general to the difference between the individual true value and the corresponding sample value irrespective of the resend of discrepancy. Sometimes there may be interaction between both of them and it may be inflate these errors. The measurements device or technique mat is defective and may cause observational errors may be assigned as under,

- In adequate supervision and inspection of field staff.
- Inadequate drained and experience field staffs.
- Problems involved in data collection and other types of errors on the part of respondence.

Types 3: Tabulation Error

These errors can we assigned number of defective method, number of coding punching, tabulation etc., these method may referred according to the techniques' employed and eccumpents available for the data processing. To these errors bias due to estimation procedure may also include. This bias may be consisted as part of tabulation errors. The main sources to these errors may be assigned as follows.

i. In adequate scrutiny Pasic data.
ii. Errors in data processing abrasion such coding, punching, listing, verification,etc.
iii. Other errors committed admitted during publication presentation of results.

Non-Response and its Effects:

Non-response prevents the researcher from collecting data for all units in the sample. It is a common source of error, particularly in survey research.

Causes of non response include:

- Poor survey design or errors in data collection
- Wrong target audience (e.g., asking residents of an elderly home about participation in extreme sports)
- Asking questions likely to be skipped (e.g., sensitive questions about drugs, sexual behavior, or infidelity)
- Inability to contact potential respondents (e.g., when your sample includes individuals who don’t have a steady home address)
- Conducting multiple waves of data collection (e.g., asking the same respondents to fill in the same survey at different points in time)
- Not taking into account linguistic or technical difficulties (e.g., a language barrier)

Types of non-response

Usually, a distinction is made between two types of non response:

- **Unit non response** encompasses instances where all data for a sampled unit is missing—i.e., a number of respondents didn't complete the survey at all (missing data).
- **Item non response** occurs where only part of the data could not be obtained—i.e., a number of respondents selectively skipped the same survey question.

It is important to keep in mind that non response bias is **always** associated with a specific variable (like manager workload in the previous example). Respondents and non respondents differ with respect to that variable (workload) specifically. Because managers' decision to participate or not in the survey relates to their workload, the data is not randomized, leading respondents and non respondents to differ in a way that is significant to the research.

National Sample Survey Office (NSSO)

The National Sample Survey Office (NSSO) headed by a Director General is responsible for conduct of large scale sample surveys in diverse fields on All India basis. Primarily data are collected through nation-wide household surveys on various socio-economic subjects, Annual Survey of Industries (ASI), etc. Besides these surveys, NSSO collects data on rural and urban prices and plays a significant role in the improvement of crop statistics through supervision of the area enumeration and crop estimation surveys of the State agencies. It also maintains a frame of urban area units for use in sample surveys in urban areas.

The NSSO has four Divisions:

- Survey Design and Research Division (SDRD): This Division, located at Kolkata, is responsible for technical planning of surveys, formulation of concepts and definitions, sampling design, designing of inquiry schedules, drawing up of tabulation plan, analysis and presentation of survey results.

- Field Operations Division (FOD): The Division, with its headquarters at Delhi/Faridabad and a network of six Zonal Offices, 52 Regional Offices and 117 Sub-Regional Offices spread throughout the country, is responsible for the collection of primary data for the surveys undertaken by NSS.

- Data Processing Division (DPD): The Division, with its headquarters at Kolkata and 5 other Data Processing Centers at various places, is responsible for sample selection, software development, processing, validation and tabulation of the data collected through surveys. Price and Wages in Rural India collected through schedule 3.01(R) is being processed at DPC Giridih. In addition, DPD is also processing the data of Periodic Labour Force Survey (PLFS). Industrial Statistics Wing (IS Wing), DPD, NSS, Kolkata is responsible for sample selection, data processing, validation and tabulation of the Annual Survey of Industries(ASI) data collected through a dedicated web-portal.

- Survey Coordination Division (SCD): This Division, located at New Delhi, coordinates all the activities of different Divisions of NSS. It also brings out the bi-annual journal of NSS, titled "Sarvekshana", and organizes National Seminars on the results of various Socio-economic surveys undertaken by NSS.

Central Statistics Office (CSO)

Earlier known as the Central Statistics Organization of India, CSO is responsible for the coordination of statistical activities in India, and evolving and maintaining statistical standards.

- The CSO is headed by the Director-General who is assisted by Five additional Director-Generals and four Deputy Director-Generals, six Joint Directors, seven special task officers, thirty deputy directors, 48 assistant directors, and other supporting staff.
- It has a well-equipped Graphical Unit. The CSO is located in Delhi.
- Its activities include:
  - National Account Statistics (NAS)
  - Conduct of Annual Survey of Industries (ASI), Economic Censuses and its follow up surveys,
  - Index of Industrial Production Compilation, as well as Consumer Price Indices for Urban Non-Manual Employees. (Read about Index of Industrial Production (IIP) from the linked article.)
  - Human Development Statistics
  - Gender Statistics
  - Imparting training in Official Statistics
  - Five Year Plan work relating to the Development of Statistics in the States and Union Territories
  - Broadcasting statistical information, energy, construction, and environment statistics, work relating to trade revision of National Industrial Classification, etc.

...oooOooo...