Sampling Theory

(for III BA/BSc Statistics Students)

By

K. Ashok

Depart of Statistics

PR Govt. College (A)

Kakinada

Sampling Theory

(Design of sample surveys)

Introduction:

Already we learned some basics of sampling in the previous year. In this year we are going to discuss some other points and sampling designs. Even some basics are presented here.

There are TWO types of surveys to collect the data.

- Complete Enumeration (Census)
- Sample Survey

In the former one, the data are collected from each and every unit of Population (aggregate of objects under study).

In the later one, the data are collected from a portion of population (called as sample).

In many practical issues, sample survey is advantageous over Census. In particular, in testing and especially in case of destructive it is impossible to conduct Census, and sample survey is only the way.

Parameters: The characteristics of the population are called parameters.

Eg: Population Mean(μ), Population SD(σ), Population Proportion(P),etc

Statistic: The function or a measure of Sample observations is called a sample.

Eg: Sample mean($\bar{x}$), Sample SD(s), Sample proportion(p), etc

Sampling distribution: The distribution of Statistic is called a sampling distribution.

- The average of the distribution of statistic is called Expectation.
- The standard deviation of the distribution of the statistic is called the **Standard Error(S.E)** of the statistic.

Eg:

- S.E of $\bar{x}$ = $(\sigma/\sqrt{n}\ )\ (\frac{N-n}{N-1})$
- S.E of p = $(\sqrt{PQ/n}\ )\ (\frac{N-n}{N-1})$

Where N=Population Size, n= sample size

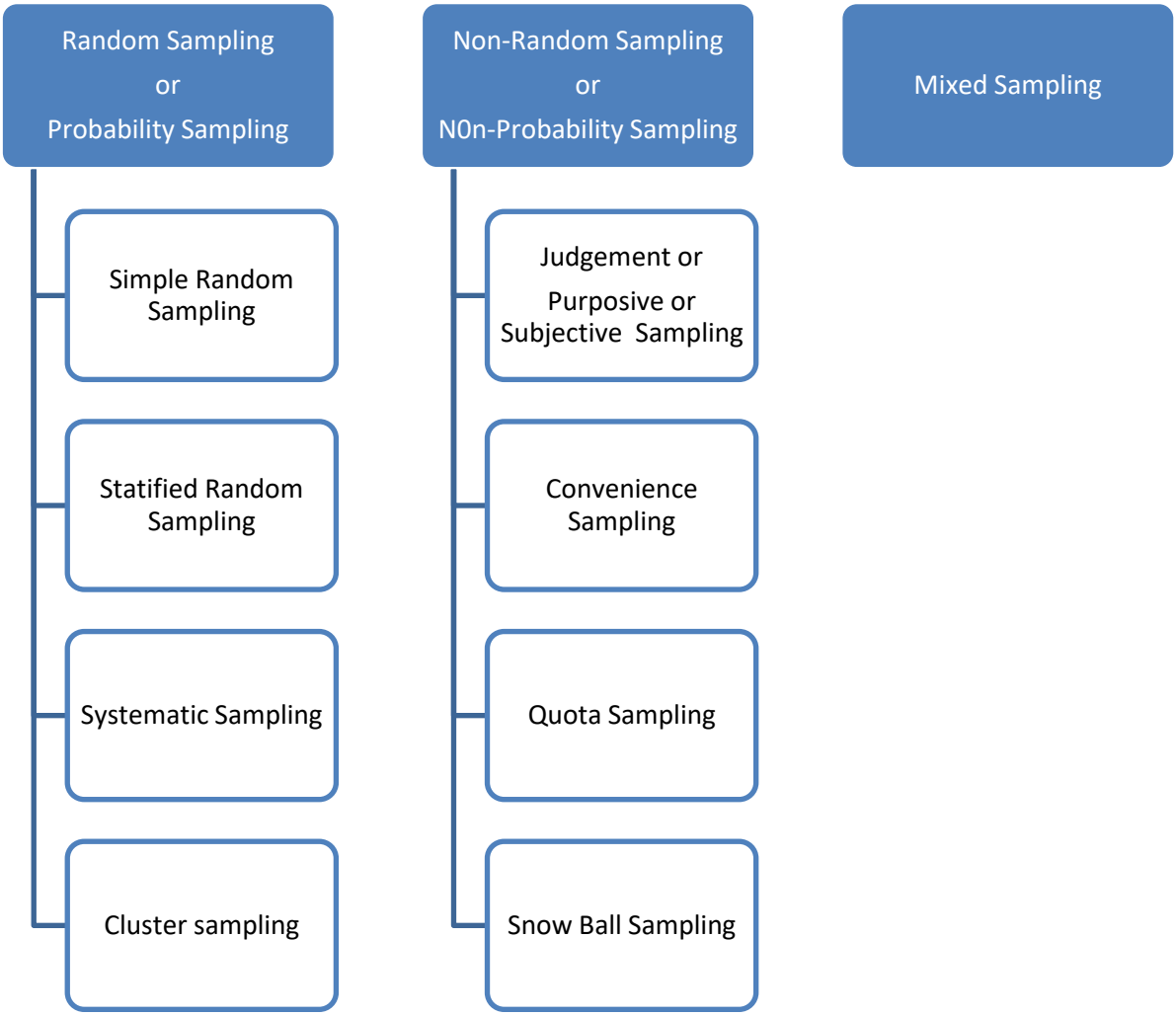
Here $\frac{N-n}{N-1}$ is called finite population correction factor.

It can be neglected for large populations.

Sampling:

The process of selecting sample from a population is called sampling. There are different methods of Sampling.

Methods of Sampling:



In this paper, we shall discuss first 3 random sampling methods.

Principal Steps in a Sample Survey:

The following are main steps involved in planning and execution of a sample survey.

1. **Objectives of the Survey**: The first and foremost step is to define the objectives of the survey in clear and concrete terms. The sponsoring agency of survey should take care that these objectives are corresponding with the available resources in terms of money, man power and time limit.
2. **Defining the Sampled Population**: The population from which sample is chosen should be defined in clear and unambiguous terms. Some practical difficulties in handling certain segments of population (specially boarder cases) may point to their elimination from the scope of survey. Thus for reasons of practicability or convenience the population to be sampled (Sampled Population) may be different than the population for which results are wanted (Target Population).
3. **Sampling Units and the Frame**: The units of sampled population are called sampling units. These must cover the entire sampled population and they must be distinct, non-overlapping and unambiguous so that every element of the population belons to only one sampling unit. For Eg, In socio-economic survey for selecting people in a city, the sampling unit might be an individual person, a family or a block in a locality.

   The list of sampling units is called a Frame. It may be in form of a map or any other acceptable material. The frame serves as a guide to the population to be covered. This frame should be accurate and dynamic.

4. **Selection of proper sampling design:** A number of designs (Probability sampling methods and Non-probability sampling methods) for selection of sample are available. A proper design should be selected considering efficiency, cost and time.
5. **Selection of method of collecting information**: A appropriate method should be selected for collecting information keeping accuracy, cost and time. A proper care should be taken in case of non-respondents.

   Some important methods are, Interview method and Questionnaire method.
6. **Data to be collecte**d: The data should be collected keeping view the objectives of the survey. A practical method is to chalk out an outline of the tables that the survey should produce. This would help in elimination the collection of irrelevant and too may information and ensure that no essential data are omitted.
7. **Organization of Field work**: It is absolutely essential that the personnel should be well trained in locating sample units, recording the measurements and the methods of collection of required data before starting field work. The success of a survey to a great extent depends upon the reliable field work. It is necessary to inspect the field work after completion by supervisory staff.
8. **The Pre-test**: From practical point of view a small pre-test should be conducted. Here Pre-test means trying out the questionnaire and field methods on a small scale. It always helps to decide upon effective method of asking questions and disclose certain problems, otherwise it will be quite serious on a large scale sample survey.

9. **Summary and Analysis of the data**: The analysis of the data may be classified as,

   i) **Scrutiny and editing of data:** An initial quality check should be carried out by the supervisory staff while the investigators are in the field. The scrutiny and editing of completed schedules and questionnaires will help in eliminating the erroneous and inconsistent data.

   ii) **Tabulation of data:** Before carrying out the tabulation of data, we must decide the procedure of tabulation of the data which are incomplete due to non-response to certain items in questionnaire. For a large scale sample survey, advanced software tools will help to tabulate the data and these need coding for qualitative variables.

   Iii) **Statistical Analysis**: After the data has been properly scrutinized, edited and tabulate, a very careful statistical analysis is to be made. Appropriate formulae should be used to provide unbiased and accurate results.

10. **Reporting and conclusions**: Finally a report incorporating detailed statement of the different stages of the survey should be prepared. In the presentation of the results, it is good practice to report the technical aspect of the design, i.e., the types of the estimators used along with expected error.

11. **Information gained for future surveys:** The information gained from any complicated survey and sample in the form of the data regarding means, SD's ,costs, time, etc serve as a potential guide for future surveys. It serves as a lesson to the organisers for future surveys in recognising and rectifying the mistakes committed in the execution of the survey.

Principles of Sampling:

The theory of sampling based on the following important principles.

a) **Principle of Statistical Regularity**: The principle has its origin in the theory of probability. According to King " the law of statistical regularity lays down that a moderately **large numbe**r of items chosen at **random** from a large group are almost sure on the average to process the characteristics of the large group".

   An immediate derivation from the principle of statistical regularity is the **Principle of Inertia of Large Numbers**.

b) **Principle of Validity**: The principle of validity states that the sampling design enable us to obtain valid tests and estimates about the parameters of the population. The samples obtained by the techniques of probability sampling satisfy this principle.

c) **Principle of Optimization**: The principal of optimization states that the results of the design in terms of efficiency and cost with available resources is optimum. The principle of optimization consists in

   - Achieving a given level of efficiency at minimum cost
   - Obtaining maximum possible efficiency with a given level of cost.

Scope of Errors (Sampling and Non-sampling Errors):

The errors involved in the collection, processing and analysis of a data can be classified into two heads.

i) Sampling Errors, and ii) Non-sampling Errors

i) Sampling Errors:

The sampling errors arise due to the fact that only a portion of the population has been used to estimate the parameters and draw inferences about the population. Thus these errors are present in sample survey and absent in census survey.

Sampling errors are due to the following reasons:

1. **Faulty selection of the sample**: Some of bias is introduced by the use of improper sampling technique for the selection of sample. In most of situations, expect few, judgement sampling provides biased results. It can be overcome by using a random sampling.
2. **Substitution**: In general the investigators substitute a convenient unit in place the sampling unit which is difficult to investigate. This leads to some bias since the characteristics may be different from unit to unit.
3. **Faulty demarcation of sampling units:** There will be some bias due to defective demarcation of sampling units. It will be happen in most area surveys dealing with border lines.
4. **Improper choice of statistic:** a constant error due to improper selection of statistic for estimating population parameters.

   For example, we know that sampling variance s^2 is biased for estimating population variance σ^2 where as $ns^2/(n-1)$ is unbiased.

**** The sampling errors can be reduced by increasing the sample size, since the S.E is inversely proportional to the square root of sample size.**

ii) Non-sampling Errors:

The non-sampling errors arise at the stages present in both the census and sample surveys. Non-sampling errors can occur at every stage of planning and execution of census and sample survey. These arise from the following factors.

1. **Faulty planning or Definitions**: The planning of the survey consists of definitions various objectives. Here the non-sampling errors can be due to:
   a) Errors due to location of units, errors in recording measurements, errors due to ill-designed questionnaire, etc.
   b) Data specification being inadequate and inconsistent
   c) Lack of trained and qualified personnel.
2. **Response Errors**: These errors resulting from responses may be due to the following reasons.
   - May be accidental
   - Due to the prestige of informant (it will upgrade or downgrade the response)
   - Self interest of informant
   - Bias due to interviewer
3. **Non-response Bias**: It occurs if full information is not obtained on all the selected sampling units
4. **Compiling Errors**: The various operations of data such as editing, coding, tabulation, etc are potential sources of error. These can be control through verification, consistency check, etc.

5. **Publication Errors**: The publication errors (errors committed during presentation and printing of results) due to the two sources – mechanics of publication and failure of survey organisation to point out the limitations of statistics.

** **The non-sampling errors can reduced by assigning trained, experienced and skilled personnel.**

** **The data obtained in a complete census, although free from sampling errors, would still be subject to non-sampling errors where as data obtained in a sample survey would be subject to both sampling and non-sampling errors**.

Advantages of Sampling over Census:

1. **Less Time**: There is considerable saving in time and labour since only a portion of population has to be examined. At the same time, results can be obtained rapidly and analyzed much faster.
2. **Reduction in cost**: Sampling usually results in reduction in cost in terms of money and man hours. Since in most of cases our resources are limited in terms of money and the time, sampling is more advantageous than census.
3. **Greater Accuracy of results**: The results of a sample survey are usually much more reliable than those obtained from a complete census due to the following reasons:
    - It is possible to determine the extent of the sampling error
    - Scope of non-sampling errors is less in sampling compared with census
4. **Greater Scope**: The complete census is impracticable if the survey requires a highly trained personnel and more sophisticated equipment for collection and analysis of data. It is possible to have a thorough and intensive enquiry because a more detailed information can be obtained from a small group of respondents.
5. If Population is too large, if testing is destructive, if population is hypothetical, sampling only the way.

Self Assessment Questions:

Multiple Choice Questions (MCQ):

1. If information is collected from each and every unit of population then the survey is called ( )
   a) sample survey b) census survey c) pilot survey d) none
2. In sample survey the information is collected from ( )
   a) every unit of the population b) few selected units of the population
   c) both d) none
3. Sampling frame is a list of ( )
   a) Voters b) Random numbers c) Sampling units of population d) None
4. Which of the following method is used to collect data ( )
   a) Interview method b) Questionnaire method
   c) schedule method d) all
5. Which of the following is not a principal step of sample survey ( )
   a) preparation of frame b) summary and analysis of data
   c) validity of sampling design d) information gained for future survey
6. Which of the following is the principle of sample survey ( )
   a) principle of statistical regularity b) principle of validity
   c) principle of optimization d) all
7. Sampling errors occurred in ( )
   a) sample survey b) census survey c) both d) none
8. Non-sampling errors occurred in ( )
   a) sample survey b) census survey c) both d) none
9. The reason of sampling error is ( )
   a) improper selection of sample
   b) improper selection of statistics for estimation
   c) faulty identification of sampling units d) all
10. Which of the following are the factors of non-sampling errors ( )
    a) respondents b) non-respondents c) compiling of data d) all
11. Sampling errors can be reduced by ( )
    a) increasing sample size b) decreasing sample size
    c) controlling response errors d) controlling errors in publishing
12. Non-Sampling errors can be reduced by ( )
    a) increasing sample size b) decreasing sample size

c) controlling response errors

d) assigning qualified, trained and experienced personal

13. Which of the following is advantage of sampling over census ( )

a) less time and cost b) accuracy of results

c) greater scope d) all

14. Which of the following is not non-probability sampling ( )

a) purposive sampling b) quota sampling

c) area sampling d) snowball sampling

15. Which of the following is a probability sampling ( )

a) random sampling b) systematic sampling

c) cluster sampling d) all

Descriptive Questions:

SHORT QUESTIONS:

1. Explain the principles of sample survey.
2. Write the difference between sampling versus census.
3. Explain the types of sampling.

ESSAY QUESTIONS:

4. Explain the principal steps involved in conducting a large scale sample survey.
5. Explain sampling errors and non sampling errors.