Sampling:

Why do you need sampling?

Description of any statistical tool starts with "Let $x_1, x_2,\ldots, x_n$ be a random sample from population…."

Based on this sample, the statistical analysis is conducted.

As a matter of fact, statistics has utility only because it can provide statistical inferences for the entire population using the sample data.

Sampling:

How to obtain these "$x_1, x_2,\ldots, x_n$" ?

If these "$x_1, x_2,\ldots, x_n$" are good, we get good inferences.

If these "$x_1, x_2,\ldots, x_n$" are bad, we get bad inferences.

Entire success of statistical tools depends upon the outcomes and the outcome depends upon the quality of sample used in the analysis.

Sampling:

Sampling theory helps.

It provides methodologies for choosing "$x_1, x_2,\ldots, x_n$" .

The methodologies ensures that the "$x_1, x_2,\ldots, x_n$" are "good" and as per the requirements of the statistical tools to be used.

Sampling:

Sampling theory provides the tools and techniques for data collection keeping in mind the objectives to be fulfilled and nature of population.

These are two ways of obtaining the information

1. Sample surveys

2. Complete enumeration or census

Sampling:

- **Sample surveys collect information on a fraction of total population whereas**
- **the information on whole population is collected in census.**

Some surveys are conducted regularly like economic surveys, agricultural surveys etc.

Some surveys are need based and are conducted when some need arise, e.g., consumer satisfaction surveys at a newly opened shopping mall to see the satisfaction level with the amenities provided in the mall .

Sampling Unit:

An element or a group of elements on which observations can be taken is called a sampling unit.

The objective of the survey helps in determining the definition of sampling unit.

Definition of sampling unit depends and varies as per the objective of the survey.

Sampling Unit:

Example:

Objective: To determine the total income of all the persons in the household.

Sampling unit: Household.

Example:

Objective: To determine the income of any particular person in the household.

Sampling unit: Income of the particular person in the household.

Sampling Unit:

Example:

Objective: To study the health conditions.

Sampling unit: The person on whom the readings on the blood sugar level, blood pressure and other factors will be obtained. These values will together classify the person as healthy or unhealthy.

Sampling Unit:

Example:

Objective: A fish food increases the weight of the fish or not.

Sampling unit: What is sampling unit?

Weight of fish or weight of aquarium?

Population:

Collection of all the sampling units in a given region at a particular point of time or a particular period is called population.

Example:

Objective: Medical facilities in a hospital are to be surveyed through the patients.

Population: Total number of patients registered in the hospital during the time period of survey.

Population:

Example:

Objective: To study the production of wheat in a district.

Population: All the fields cultivating wheat in that district.

Population:

Population size: Total number of sampling units in the population.

Denoted generally by *N*.

The population size can be finite or infinite (*N* is large).

Census:

Complete count of population is called census.

The observations on all the sampling units in the population are collected in a census.

For example, in India, the census is conducted at every tenth year in which observations on all the persons staying in India is collected.

Sample:

Collection of One or more sampling units selected from the population according to some specified procedure.

A sample consists only of a portion of the population units.

Representative Sample:

All salient features of population are present in the sample.

Every sample has to be a representative sample.

For example, if a population has 30% male and 70% female, then we also expect the sample to have nearly 30% male and 70% females.

Representative Sample:

In another example, if we take out a handful of wheat from a 100 Kg. bag of wheat, we expect the same quality of wheat in hand as inside the bag.

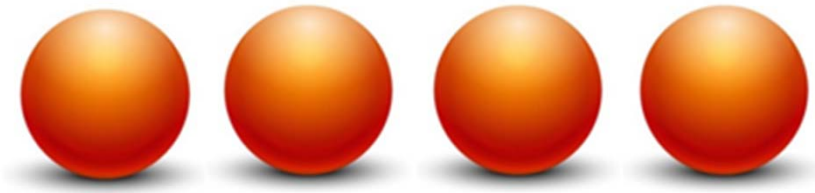
It is expected that a drop of blood will give the same information as all the blood in the body.

Population and Sample:

Population of balls of size 10

Sample 1 of size 3

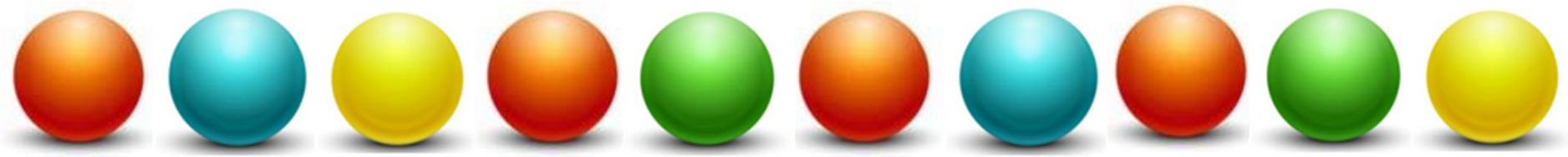
Sample 1 of size 4



Sample 1 of size 5

Population and Sample:

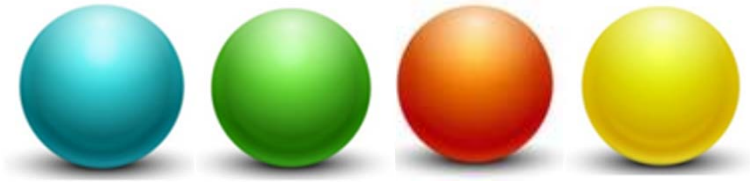
Population of balls of size 10



Sample 1 of size 3
(Green ball is missing)

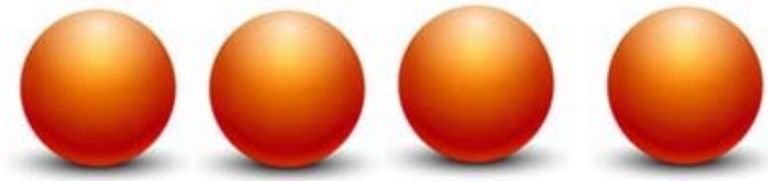
Sample 2 of size 4

(All colour balls are present)



Sample 3 of size 5
(Blue ball is repeated)

Sample 4 of size 4
(Only red balls are selected)

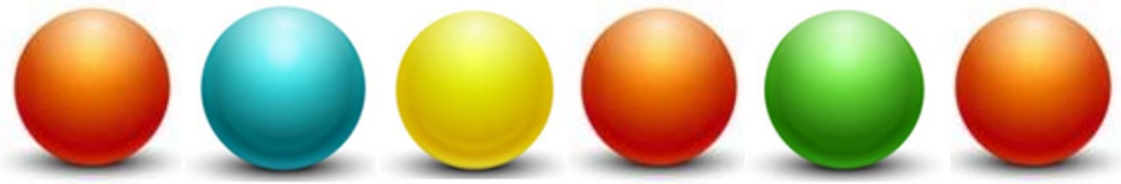


Census and Sample:

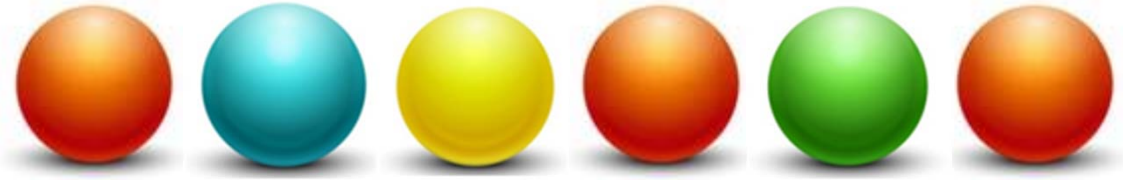
In the context of sample surveys, a collection of units like people, cities, countries etc. is called a finite population.

A census is a 100% sample and it is a complete count of the population.

Population



Sample



This is census.

Sampling Frame:

List of all the units of the population to be surveyed constitutes the sampling frame.

All the sampling units in the sampling frame have identification particulars.

For example, all the students in a particular university listed along with their roll numbers constitutes the sampling frame.

Similarly, the list of households with the name of head of family or house address constitutes the sampling frame.

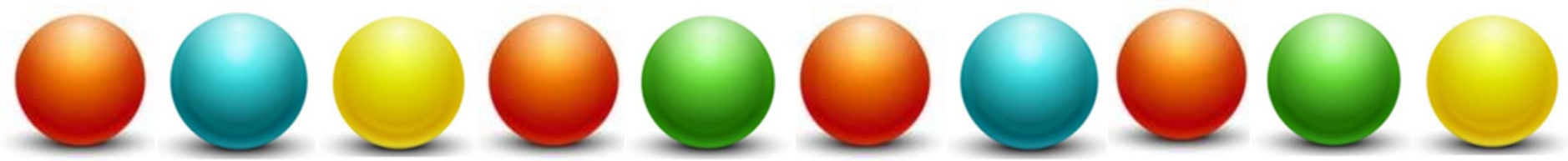
Simple Random Sampling:

Population of balls of size 10

Sampling frame

Simple Random Sampling :

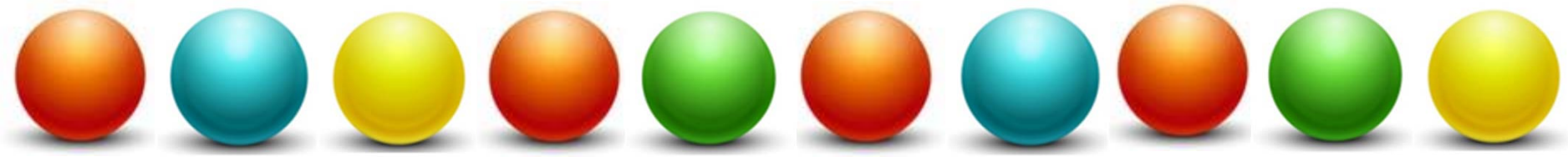
Population of balls of size 10



Sampling frame

Representative Sample:

Population:



Sample 1: Is this representative?
(Green and blues balls are missing)

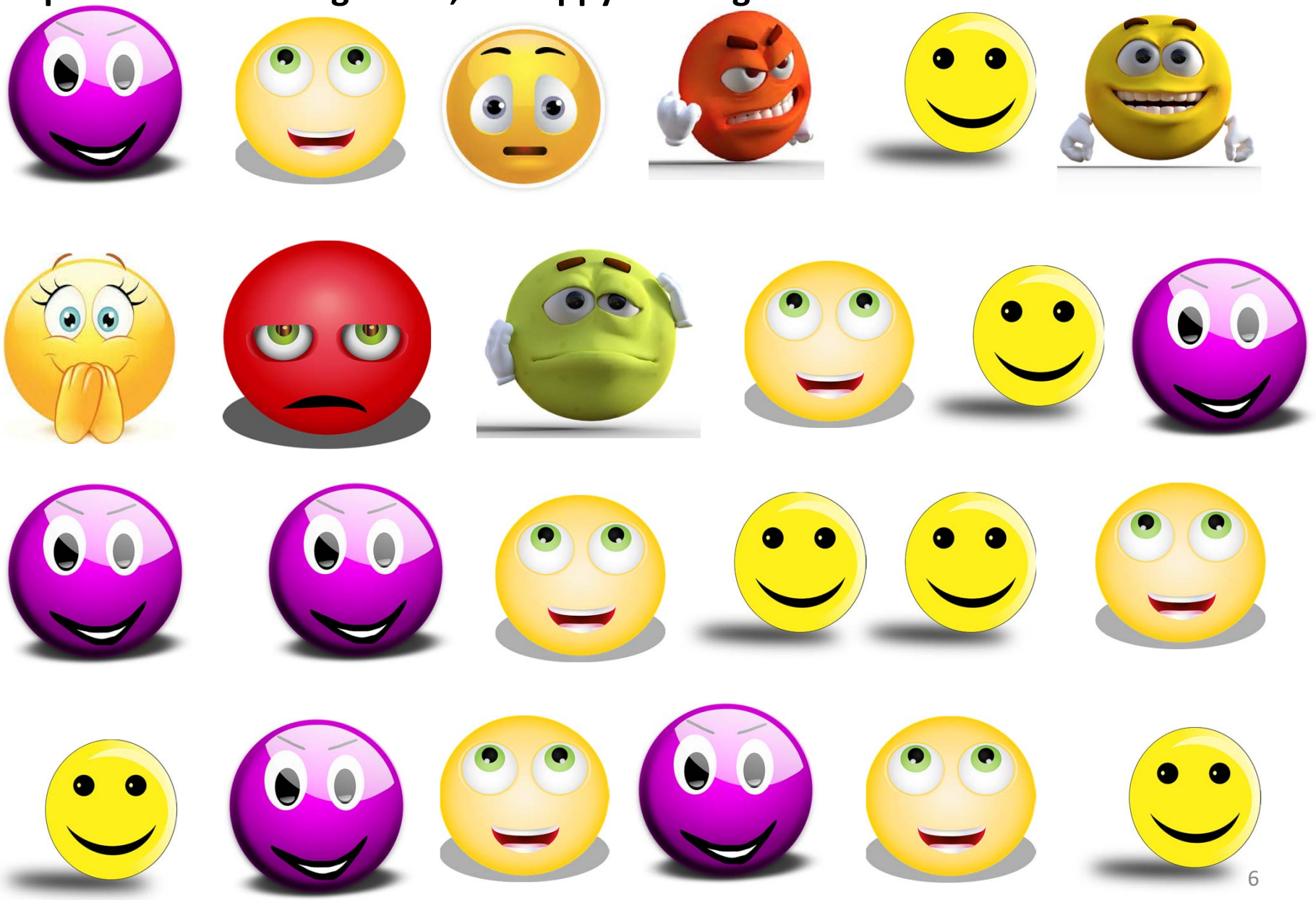
Sample 2: Is this representative?
(Red ball is missing)

Sample 3: Is this representative?
(All balls are proportionally present)

Sample 4: Is this representative?
(Red ball is over represented)

Representative Sample:

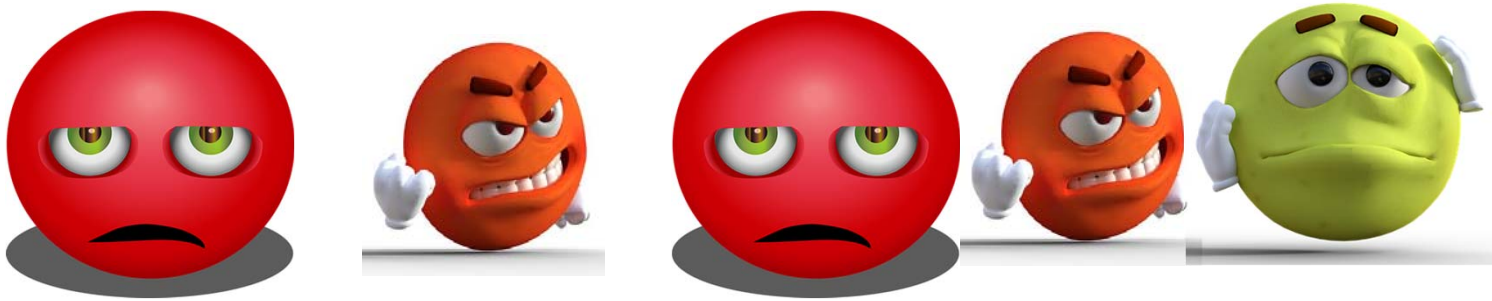
Population: Gives in general, a "Happy“ feeling



Representative Sample:

Is this representative? Appears as if mostly people are not happy.

Is this representative? Appears as if mostly people are Annoyed.



Is this representative? Appears as if mostly people are Happy.

Ensuring representativeness:

There are two possible ways to ensure that the selected sample is representative.

1. **Random sample or Probability sample**
2. **Non-random sample or Purposive sample**
3. **Quota Sample**

Ensuring representativeness:

1. Random Sample or Probability Sample:

The selection of units in the sample from a population is governed by the laws of chance or probability.

The probability of selection of a unit can be equal as well as unequal.

Ensuring representativeness:

2. Non-Random Sample or Purposive Sample:

The selection of units in the sample from population is not governed by the probability laws.

It is the sample based on non-random laws.

Examples:

- **Units are selected on the basis of personal judgment of the surveyor.**
- **The persons volunteering to take some medical test.**
- **The persons volunteering to drink a new type of coffee.**

Ensuring representativeness:

3. Quota Sample:

The survey in this case is continued until a predetermined number of units with the characteristic under study are picked up.

For example, in order to conduct an experiment for rare type of disease, the survey is continued till the required number of patients with disease are collected.

Advantages of Sampling Over Complete Enumeration

- **Reduced cost and enlarged scope**

Sampling involves the collection of data on smaller number of units in comparison to complete enumeration, so the cost involved in the collection of information is reduced.

Additional information can be obtained at little cost in comparison to conducting another survey.

Advantages of Sampling Over Complete Enumeration

- **Reduced cost and enlarged scope**

For example, when an interviewer is collecting information on health conditions then he/she can also ask some questions on health practices. This will provide additional information on health practices and the cost involved will be much less than conducting an entirely new survey on health practices.

Advantages of Sampling Over Complete Enumeration

- **Organization of work**

It is easier to manage the organization of collection of smaller number of units than all the units in a census.

For example, to draw a representative sample from a state, it is easier to manage to draw small samples from every city than from the whole state. This results in more accuracy in the statistical inferences because better organization provides better data and in turn improved statistical inferences are obtained

Advantages of Sampling Over Complete Enumeration

- **Greater accuracy**

The persons involved in the collection of data are trained and greater accuracy in the data will be achieved if lesser number of units are to be selected.

Advantages of Sampling Over Complete Enumeration

- **Urgent information required**

The data from a sample can be quickly summarized.

For example, the forecasting of the crop production can be done quickly on the basis of a sample of data then collecting first all the observation.

Advantages of Sampling Over Complete Enumeration

- **Feasibility**

Conducting the experiment on smaller number of units, particularly when the units are destroyed, is more feasible.

For example, in determining the life of bulbs, it is more feasible to fuse minimum number of bulbs.

Similarly, in any medical experiment, it is more feasible to use less number of animals.

Type of Surveys:

1. Demographic Surveys

These surveys are conducted to collect the demographic data, e.g., household surveys, family size, number of males in families, etc.

Such surveys are useful in the policy formulation for any city, state or country for the welfare of the people.

Type of Surveys:

2. Educational Surveys

These surveys are conducted to collect the educational data, e.g., how many children go to school, how many persons are graduate, etc.

Such surveys are conducted to examine the educational programs in school and colleges.

Generally, schools are selected first and then the students from each school constitue the sample.

Type of Surveys:

3. Economic Surveys

These surveys are conducted to collect the economic data, e.g., data related to export / import of goods, industrial production, consumer expenditure etc.

Such data is helpful in constructing the indexes indicating the growth in a particular sector of economy or even the overall economic growth of the country.

Type of Surveys:

4. Employment Surveys

These surveys are conducted to collect the employment related data, e.g., employment rate, labour conditions, wages, etc. in a city, state or country.

Such data helps in constructing various indices to know the employment conditions among the people.

Type of Surveys:

5. Health and Nutrition Surveys

These surveys are conducted to collect the data related to health and nutrition issues, e.g., number of visits to doctors, food given to children, nutritional value etc.

Such surveys are conducted in cities, states as well as countries by national and international organizations like UNICEF, WHO etc.

Type of Surveys:

6. Agricultural Surveys

These surveys are conducted to collect the agriculture related data to estimate, e.g., the acreage and production of crops, livestock numbers, use of fertilizers, use of pesticides and other related topics.

The government bases its planning related to the food issues for the people based on such surveys.

Type of Surveys:

7. Marketing Surveys

These surveys are conducted to collect the data related to marketing. They are conducted by major companies, manufacturers or those who provide services to consumer etc.

Such data is used for knowing the satisfaction and opinion of consumers as well as in developing the sales, purchase and, promotional activities etc.

Type of Surveys:

8. Election Surveys

These surveys are conducted to study the outcome of an election or a poll. For example, such polls are conducted in democratic countries to have the opinions of people about any candidate who is contesting the election.

Type of Surveys:

9. Public Polls And Surveys

These surveys are conducted to collect the public opinion on any particular issue. Such surveys are generally conducted by news media and agencies which conduct polls and surveys on current topics of interest to public.

Type of Surveys:

10. Campus Surveys

These surveys are conducted on the students of any educational institution to study about the educational programs, living facilities, dining facilities, sports activities, etc.

Principal Steps in Conducting a Survey:

1. Define objective of survey.

The objective of the survey has to be clearly defined and well understood by the person planning to conduct it.

It is expected from the statistician to be well versed with the issues to be addressed in consultation with the person who wants to get the survey conducted.

In complex surveys, sometimes the objective is forgotten and data is collected on those issues which are far away from the objectives.

Principal Steps in Conducting a Survey:

2. Decide and choose population to be sampled.

Based on the objectives of the survey, decide the population from which the information can be obtained.

For example, population of farmers is to be sampled for an agricultural survey whereas the population of patients has to be sampled for determining the medical facilities in a hospital.

Principal Steps in Conducting a Survey:

3. Decide the Data to be collected.

It is important to decide that which data is relevant for fulfilling the objectives of the survey and to note that no essential data is omitted.

Sometimes too many questions are asked and some of their outcomes are never utilized. This lowers the quality of the responses and in turn results in lower efficiency in statistical inferences.

Principal Steps in Conducting a Survey:

4. Decide the degree of precision required.

The results of any sample survey are always subjected to some uncertainty. Such uncertainty can be reduced by taking larger samples or using superior instruments.

This involves more cost and more time. So it is very important to decide about the required degree of precision in data. This needs to be conveyed to the surveyor also.

Principal Steps in Conducting a Survey:

5. Decide the method of measurement.

The choice of measuring instrument and method to measure the data from the population needs to be specified clearly.

For example, the data has to be collected through interview, questionnaire, personal visit, combination of any of these approaches, etc.

The forms in which the data is to be recorded so that the data can be transferred to mechanical equipment for easily creating the data summary etc. is also needed to be prepared accordingly.

Principal Steps in Conducting a Survey:

6. Decide the sampling frame.

The sampling frame has to be clearly specified.

The population is divided into sampling units such that the units cover the whole population and every sampling unit is tagged with identification. The list of all sampling units is called the frame.

The frame must cover the whole population and the units must not overlap with each other in the sense that every element in the population must belong to one and only one unit.

For example, the sampling unit can be an individual member in the family or the whole family.

Principal Steps in Conducting a Survey:

7. Decide the scheme of selection of sample.

The size of the sample needs to be specified for the given sampling plan.

This helps in determining and comparing the relative cost and time of different sampling plans.

The method and plan adopted for drawing a representative sample should also be detailed.

Principal Steps in Conducting a Survey:

8. Conduct the Pre-test.

It is advised to try the questionnaire and field methods on a small scale .

This may reveal some troubles and problems beforehand which the surveyor may face in field in large scale surveys .

Principal Steps in Conducting a Survey:

9. Organization of the field work

How to conduct the survey, how to handle business administrative issues, providing proper training to surveyors, procedures, plans for handling the nonresponse and missing observations etc. are some of the issues which need to be addressed for organizing the survey work in the fields.

The procedure for early checking of the quality of return should be prescribed. How to handle the situation when the respondent is not available should be clarified.

Principal Steps in Conducting a Survey:

10. Organization of the field work

It is to be noted that based on the objectives of the data, the suitable statistical tool is decided which can answers the relevant questions.

In order to use the statistical tool, a valid data set is required and this dictates the choice of responses to be obtained for the questions in the questionnaire, e.g., the data has to be qualitative, quantitative, nominal, ordinal etc.

Principal Steps in Conducting a Survey:

10. Organization of the field work

After getting the completed questionnaire back, it needs to be edited in the sense to amend the recording errors and delete the erroneous data.

The

- **tabulating procedures,**
- **methods of estimation and**
- **tolerable amount of error in the estimation**

needs to be decided before the start of survey.

Principal Steps in Conducting a Survey:

10. Organization of the field work

Different methods of estimation may be available to get the answer of the same query from the same data set.

So the data needs to be collected which is compatible with the chosen estimation procedure.

Principal Steps in Conducting a Survey:

11. Decide how to present the summary and analysis of data

The completed surveys work as guide for improved sample surveys in future.

They also supply various types of prior information required for using in various statistical tools, e.g., mean, variance, nature of variability, cost involved etc.

Any completed sample survey act as potential guide for the surveys to be conducted in the future.

Principal Steps in Conducting a Survey:

11. Decide how to present the summary and analysis of data

It is generally seen that the things always do not go in the same way in any complex survey as planned earlier.

Such precautions and alerts helps in avoiding the mistakes in the execution of future surveys.

Methods of Data Collection

1. Physical observations and measurements

The surveyor contacts the respondent personally through meeting.

The surveyor observes the sampling unit and records the data.

The surveyor can always use his prior experience to collect the data in a better way.

For example, a young man telling his age as 60 years can easily be observed and corrected by the surveyor.

Methods of Data Collection

2. Personal interview

The surveyor is supplied with a well prepared questionnaire.

The surveyor goes to the respondents and asks the same questions mentioned in the questionnaire.

The data in the questionnaire is then filled up accordingly based on the responses from the respondents.

Methods of Data Collection

3. Mail enquiry

The well prepared questionnaire is sent to the respondents through postal mail, e-mail, etc.

The respondents are requested to fill up the questionnaires and send it back.

In case of postal mail, many times the questionnaires are accompanied by a self addressed envelope with postage stamps to avoid any non-response due to the cost of postage.

Methods of Data Collection

4. Web based survey

The survey is conducted online through internet based web pages.

There are various websites which provide such facility.

The questionnaires are to be in their formats and the link is sent to the respondents through email.

Methods of Data Collection

4. Web based survey

By clicking on the link, the erespondent is brought to the concerned website and the answers are to b given online.

These answers are recorded and responses as well as their statistics is sent to the surveyor.

The respondents should have internet connection to support the data collection with this procedure.

Methods of Data Collection

5. Registration

The respondent is required to register the data at some designated place.

For example, the number of births and deaths along with the details are recorded at city municipal office which are provided by the family members.

Methods of Data Collection

6. Transcription from records

The sample of data is collected from the already recorded information .

For example, the details of the number of persons in different families or number of births/deaths in a city can be obtained from the city municipal office directly.

The methods in (1) to (5) provide primary data which means collecting the data directly from the source.

The method in (6) provides the secondary data which means getting the data from the primary sources.

Methods of Data Collection

7. Online forms, e.g., Google forms etc.

Variability Control in Sample Surveys

The variability control is an important issue any statistical analysis.

A general objective is to draw statistical inferences with minimum variability.

There are various types of sampling schemes which are adopted in different conditions.

These schemes help is controlling the variability at different stages.

Such sampling schemes can be classified in the following way.

Variability Control in Sample Surveys

1. Before selection of sampling units

- **Stratified sampling**
- **Cluster sampling**
- **Two stage sampling**
- **Double sampling etc.**

Variability Control in Sample Surveys

2. At the time of selection of sampling units

- **Systematic sampling**
- **Varying probability sampling**

Variability Control in Sample Surveys

3. After the selection of sampling units

- **Ratio method of estimation**
- **Regression method of estimation**

Note that the ratio and regression methods are the methods of estimation and not the methods of drawing samples.