Collection of Data
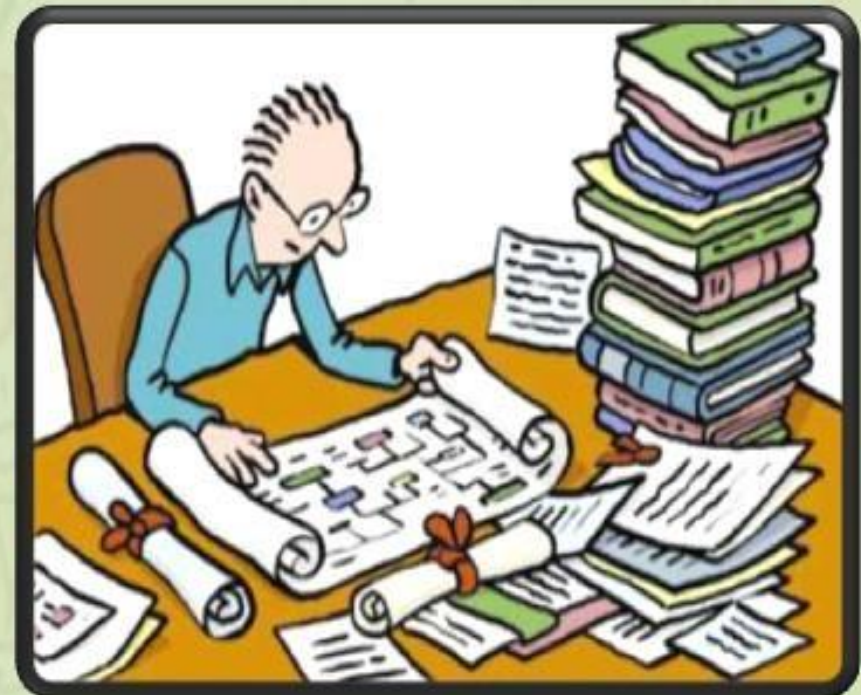
Collection of Data:

Statistical Process involves Collection and Analysis of Data and drawing meaningful conclusion from the Data.

Statistical Process starts with Collection of the Data.

This Data is either collected by Researcher first time from its Original source or it can be the Data collected by someone else and used by he Researcher.

Collection of Data:

The Data collected for the first time is called **Primary Data**.

The source from where Primary Data is collected is called **Primary Source** of Data. Such research is called **Primary Research**.

The Data collected by someone else is called **Secondary Data**.

The sources from where Secondary Data is collected is called **Secondary Source** of Data. Such research is called **Secondary Research**.

Also called **"Desk Research"**.

Investigator (Researcher)	Enumerator
The person who plans and conducts the statistical Investigation or Enquiry.	The person hired by the Investigator who goes out in the field and collects the actual Data. Investigator can also be Enumerator.

Some important Terms:

Respondent (Informant):

The person who answers the questions in the Questionnaire or Schedule.

Some important Terms:

Population (Universe):

Population is subject of Study or Investigation in statistical enquiry.

It is the Entire Group or Aggregate of Units from which Information is to be collected by the Investigator.
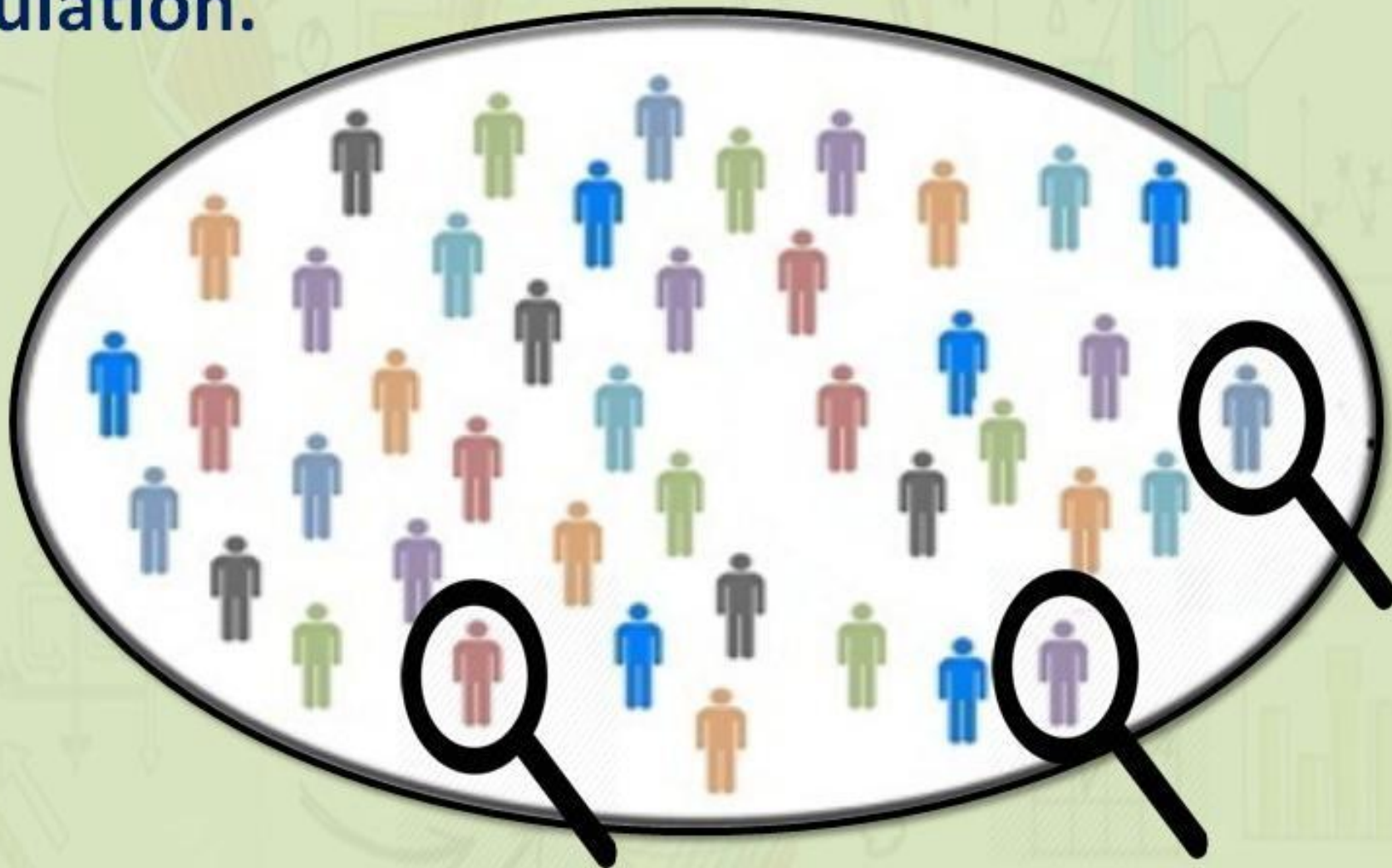
It is the total items under study.

Some important Terms:

Element of the Population (Universe):

One single **Item** or **Unit** or **Individual** from the Population.

Primary Source of Data	Secondary Source of Data
Primary Source means Data Collection from its Original Source.	Secondary Source means Data already collected by others.
The Data collected is first hand data (Original Data) (Primary Data).	The Data collected is second hand data (Past Data) (Secondary Data).
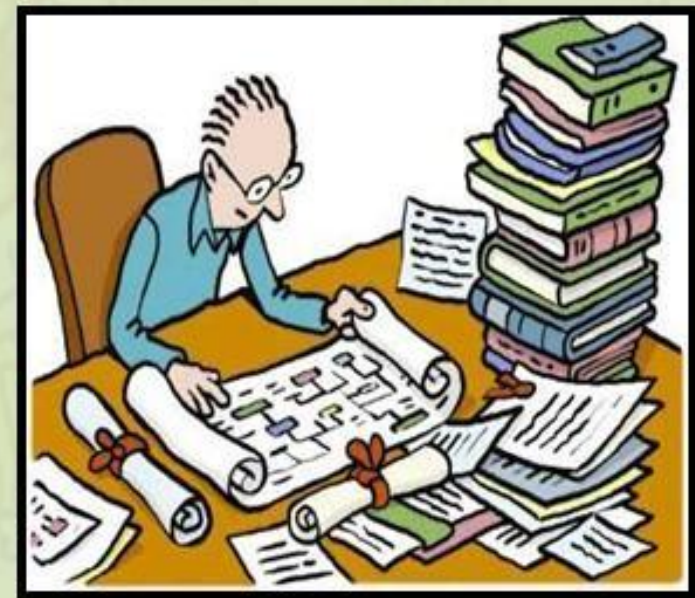
Primary Data	Secondary Data
Meaning:	
Primary Data is the first hand (fresh) data collected by the Investigator (Researcher) himself.	Secondary Data is the data collected earlier by someone else.
Specific:	
Primary Data is as per the specific requirements and objectives of the investigator (Researcher).	Secondary Data may not be as per the specific requirements of the investigators because it is collected by someone else.

Primary Data	Secondary Data
Cost:	
Primary Data collection is costly because the Investigator (Researcher) himself has to go to the area of investigation and collect the data.	Secondary Data collection is not costly because investigator is using the already collected data.
Time Consuming:	
Primary Data collection is time consuming for the same reason mentioned above.	Secondary Data is less time consuming.

Primary Data	Secondary Data
Accuracy:	
Primary Data is more accurate.	Secondary Data is less accurate.
Reliability:	
Primary Data is more reliable.	Secondary Data is less reliable.
Originality:	
Primary Data is original data. It is also called Real Time Data.	Secondary Data is not original but second hand data. It is also called Past Data.

Collecting Primary Data:

Data is collected by Businessmen to find out the response of the customers for a new product, to collect information form making improvements in the existing product.

Data is collected by Govt. to find out income levels of the Population and the type of Employment (Salaries or Self-Employed.

For collecting these data, a "**Survey**" is conducted which contains different questions. This is called "**Questionnaire**".

Questionnaire	Schedule
In Questionnaire all the information is filled by the informant himself.	In Schedule, information is supplied by informant and filled by Enumerator.

Survey: (Statistical Survey)

A Survey is a structured inquiry to obtain Aggregate Data about the Subject of Study.

Survey is conducted to gather Information (Data) from a number of people so that the specific and general opinions can be found out about a particular Issue or Topic of Enquiry or a Product.

Questionnaire:

A Questionnaire is an instrument consisting of a series of questions for the purpose of collecting Data (Information) from a group of People.
It is a kind of a written interview.

Qualities of a Good Questionnaire:

Proper Instructions:

Questionnaire should start with proper instructions on how to fill or answer the Questions.

If the Respondent notices instructions after the questionnaire is filled then there are changes of wrong responses.

Qualities of a Good Questionnaire:

Limited Number of Questions:

Questionnaire should not be very long. There should be minimum possible number of questions.

Qualities of a Good Questionnaire:

Easy to Understand (Simplicity):

Questions should be easy to understand. There should be no ambiguity in the questions. Use of difficult words should be avoided in the questionnaire. Respondent should be able to answer question quickly.

Qualities of a Good Questionnaire:

Questions should be in Order (Properly arranged):

Questions should be arranged in order so that the person answering questions feels comfortable.

For Example:
All question asking Personal Information of the Respondent should be together.
All questions regarding Price of the Product should be together.
All Questions regarding Quality of the Product should be together.

Qualities of a Good Questionnaire:

General Questions to Specific Questions:

Questionnaire should start with General Questions and Specific questions should be asked later.

For Example:

All question asking Personal Information of the Respondent should be asked first.

Then Questions can be asked about a Product.

Then Questions can be asked about the particular Brand of the Product.

Qualities of a Good Questionnaire:

Questions should be unbiased:

Questions should not encourage the respondent to give a particular answer (Leading Questions).

Example:

Poor Question:

How do you like the taste of this high quality Product.
(This Question prompts the respondent to give a positive answer)

Good Question:

How do you like the taste of this Product.

Qualities of a Good Questionnaire:

Questions should be Non-controversial:

Questions should not be related to controversial topics because people may not like to answer such questions. Personal questions should be avoided.

Qualities of a Good Questionnaire:

Questions should avoid calculations:

Questions should be such that the respondent can easily think of the answer. If it includes calculations or is based on complicated data it might be difficult for respondent to answer or he may not answer the question.

Qualities of a Good Questionnaire:

Questions may be added which help in Cross verification:

Such questions may be asked which help in cross verification.

Types of Questions:

Closed-Ended Question (Structured Questions):

Closed-Ended Questions are questions that can be answered by choosing from a limited number of options. Examples are: True/False, Yes/No/ MCQs (Multiple Choice Questions), Rating Scale etc.

Closed-Ended Questions are very quick to answer. However it may not have the Respondents' option. For this problem one option of "Any Other" may be added with space to write answer.

Types of Questions:

Open-Ended Question:

These are the questions that cannot be answered with Yes or No response or very limited options. Individualised (Varied) responses are received and all responses can be different from the other.

Respondent are free to given any answer of Open-Ended questions. However answering them may be time consuming.

Closed-Ended Questions	Open-Ended Questions
Do you thing Globalisation is Good?	What do you think about Globalisation?
Would you like to do job after college?	What would you like to do after college?
Do you like the Product?	What do you like about the Product?

Modes of Data Collection:

After knowing about the Survey and the Questionnaire and how a Good Questionnaire is prepared, Lets find out how do we reach the Target audience for our Data Collection.

There are three Basic Ways of Data Collection:

1. Personal Interview

2. Mailing Questionnaire (Survey)

3. Telephone Interview

Modes of Data Collection:

1. Personal Interview:

(Direct Personal Investigation)

Under this method, Researcher personally (face-to-face) collects information (Data) from the Informant (Respondent).

Suitability:

The **field** of study is small or **limited**.

Researcher has **access** to all the members.

Direct contact with informant is required.

Accuracy is more important.

Greater **originality** of Data is required.

Information is sensitive and **secrecy** is to be maintained.

Modes of Data Collection:

1. Personal Interview:
(Direct Personal Investigation)

Merits:

Data accuracy is high.

Data collected is Original.

Response Rate is high.

Data is reliable as collected by investigator himself.

Doubts of informant can be cleared quickly. No scope of Misinterpretation and misunderstanding.

Flexibility is there as Data or Questions can be changed quickly as per informant's feedback.

Extra (Supplementary) information can be gathered when investigator is in direct contact with the informant.

Modes of Data Collection:

1. Personal Interview:
(Direct Personal Investigation)

Demerits:

This method is not suitable if area if investigation is large.

This method is Time consuming as the investigator has to go to each and every person.

This method is costly (Expensive).

Presence of Researcher may prevent the respondent from saying what they really think.

Data may be affected by the personal bias of the investigator. It affects the credibility and reliability of the Data.

Modes of Data Collection:

2. Mailing Questionnaire (Survey)

Under this method, the Questionnaire is mailed to the Informant who is requested to complete it and send it back within a fixed period of time.

Suitability:

Area of investigation is very large and it is not possible for researcher to personally go to every informant.

The Respondent is educated (Literate).

Modes of Data Collection:

2. Mailing Questionnaire (Survey)

Merits:

This method is Economical (less expensive).

Wide are can be covered. People in remote areas can be accessed who might be inaccessible in person or by Telephone.

Informant has sufficient time to fill the questionnaire.

Data collected is original.

Investigator cannot influence the Informant.

Data is reliable because of being original.

Good method for sensitive questions.

Modes of Data Collection:

2. Mailing Questionnaire (Survey)

Demerits:

Generally Informants do not take interest in filling mailed questionnaire.

Not suitable when the informants are not educated.

Response rate is long.

Response rate is low because some Informants ignore it, some do not fill it completely and some may be lost during the mail transit.

Doubts of Informants cannot be cleared. This may result in wrong answers by informants.

Reactions cannot be watched.

Modes of Data Collection:

3. Telephone Interview

In Telephonic Interview, the Investigator asks questions on Telephone.

Suitability:

The **field** of study is large.

Accuracy is more important.

Greater **originality** of Data is required.

Modes of Data Collection:

3. Telephone Interview

Merits:

This method is Economical (Cheaper).

It is less time consuming as information can be gathered very quickly through phone.

Doubts of the Respondent can be cleared quickly.

This method is suitable if the respondent is reluctant to answer in Personal Interview.

Less influence of Investigator on Informant.

Modes of Data Collection:

3. Telephone Interview

Demerits:

This method not suitable if Respondents do not have Telephone.

Response rate may be low because many a times people do not pick phone or they do not like to answer questions on phone if Questionnaire is long.

Possibility of influence by Investigator.

Reactions cannot be watched.

Pilot Survey: (Pre-Testing)

After preparing Questionnaire and understanding different methods of how data can be collected (Personal Interview, Mail, Telephone),
Now its time to actually decided upon conducting the Investigation on the Ground.

However it is better to conduct the Investigation on a small group of people so that mistakes can be found and improvements can be done for the final large scale Survey.

This small initial survey is called "Pilot Survey" or "Pre-Testing".

Pilot Survey: (Pre-Testing)

Benefits:

It gives a preliminary idea about the final level full scale survey.

It helps in testing the questionnaire, if some questions needs to be changed or removed then it can be done.

More instructions can be given in the questionnaire if Respondents needs more clarifications about some questions.

Performance of the enumerators can be assessed.

Time and Cost involved in Final full scale survey can be assessed.

After conducting the Pilot Survey, Its time to actually conduct the large scale Survey.

Does this large scale Survey include each and everyone Or only some of them?

Lets find out?

Census Method (Complete Enumeration) and Sample Method of Survey:

Census Survey (Complete Enumeration):

A method which includes every Item (Element) (Unit) of the Population (Universe) is called Census Method.

Population Census is the best example of Census Method of Statistical Investigation.

The previous Census was conducted in 2011 and next Census year is 2021.
(Census is conducted every 10 years)
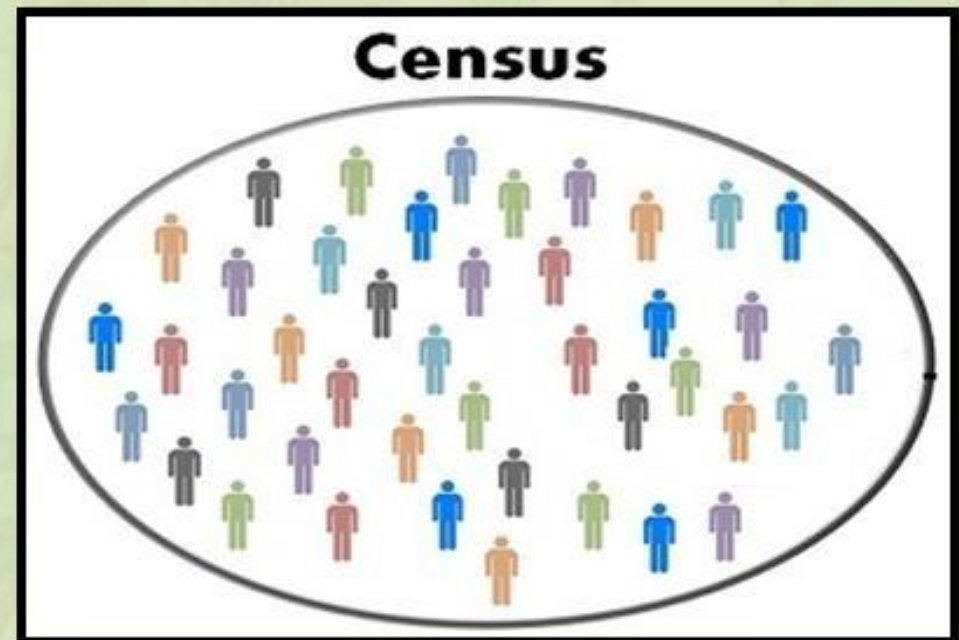
Census

Census Method (Complete Enumeration) and Sample Method of Survey:

Census Survey (Complete Enumeration):

Suitability:

- When the Population size (field of enquiry) is small.

- Items are of diverse characteristics and all have to be investigated to find out the fair and accurate Data.

- All the items need to be thoroughly (intensely) examined.

- High degree of accuracy is required.
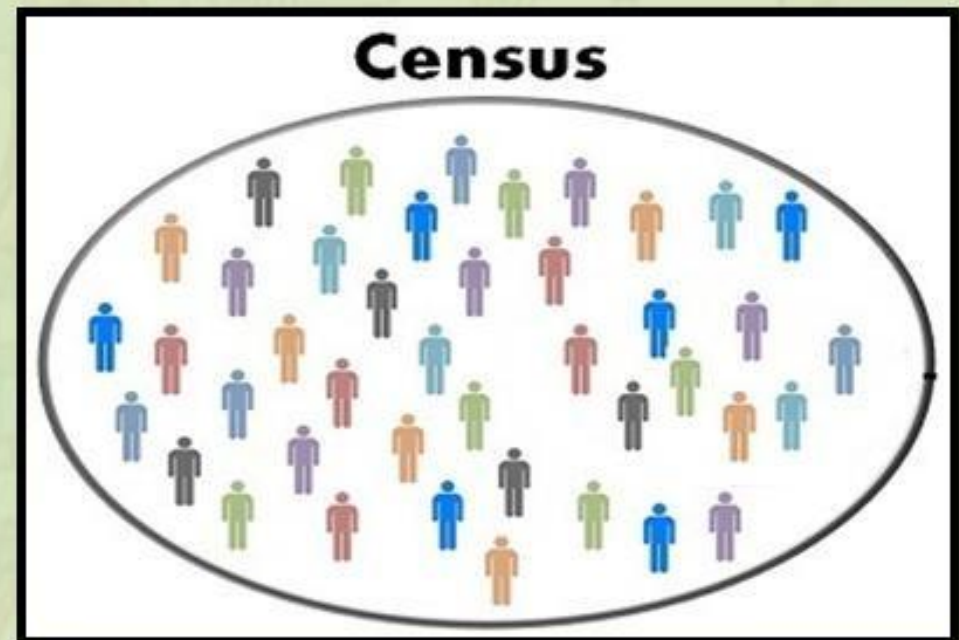
Census Method (Complete Enumeration) and Sample Method of Survey:

Census Survey (Complete Enumeration):

Merits:

Data collected is reliable and accurate as each and every unit is investigated.

Removes any possibility of bias in Sample selection.

Extensive (Comprehensive information is received from Census method.

Complex nature of Data can be effectively studied by Census method.
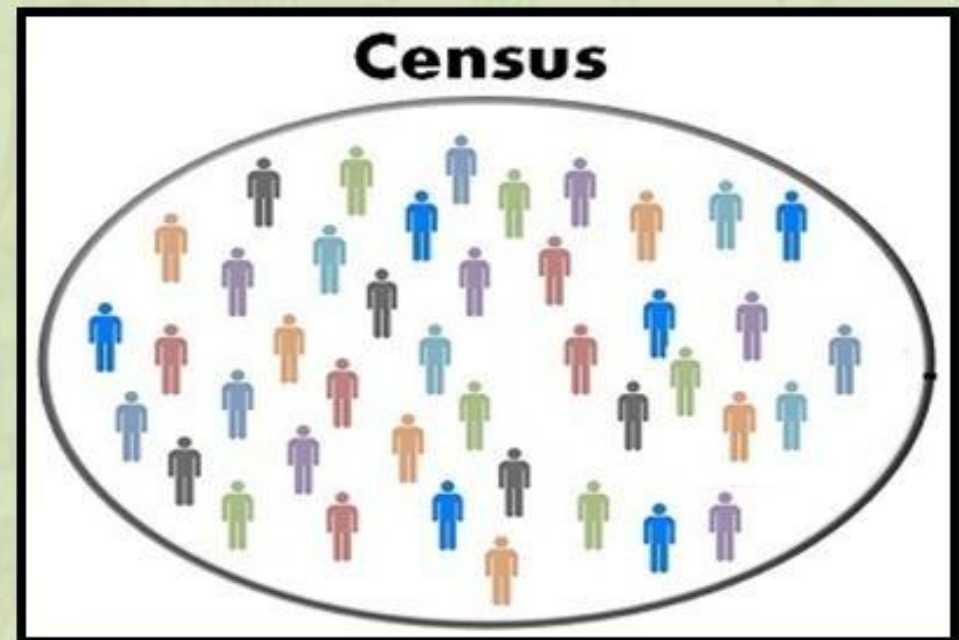
Census Method (Complete Enumeration) and Sample Method of Survey:

Census Survey (Complete Enumeration):

Demerits:

It is costly method as each and every unit is to be investigated.

It is time consuming method as each and every unit is to be investigated.

Large number of enumerators are required to cover all the area or units of the Population. Training the Enumerators also costly and time consuming.

Census method is not suitable for large Investigations overing wide area.
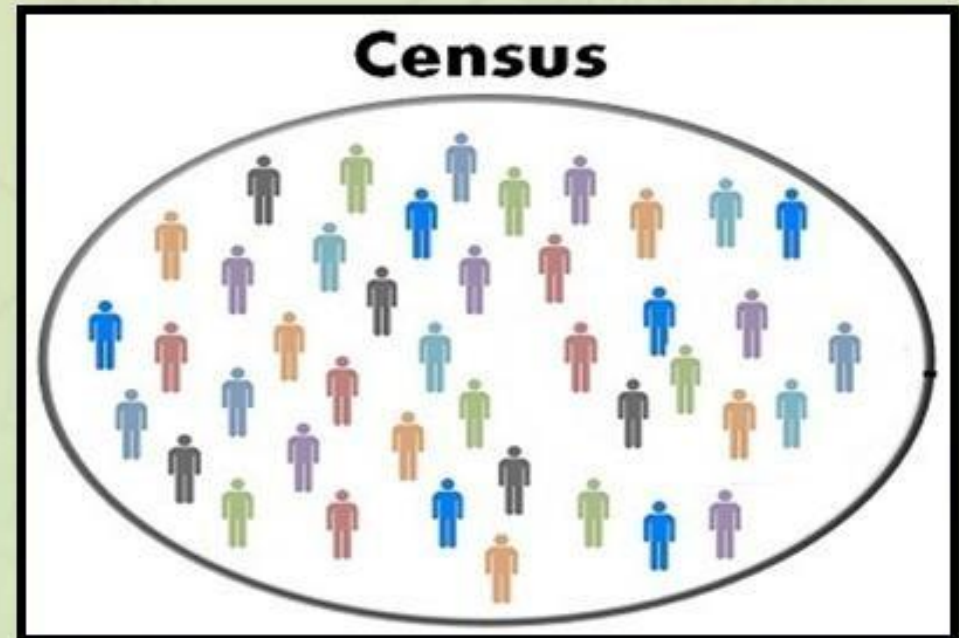
Census

Census Method (Complete Enumeration) and Sample Method of Survey:

Sample Survey:

If the Researcher (Investigator) thinks that Census Survey is not possible or feasible then he may decide to select a "**Representative Sample**".

Representative Sample is that sample of the whole Population (Universe) which fairly represents the whole Population.

Selecting the right Representative Sample is very important to make the Survey reasonably accurate.
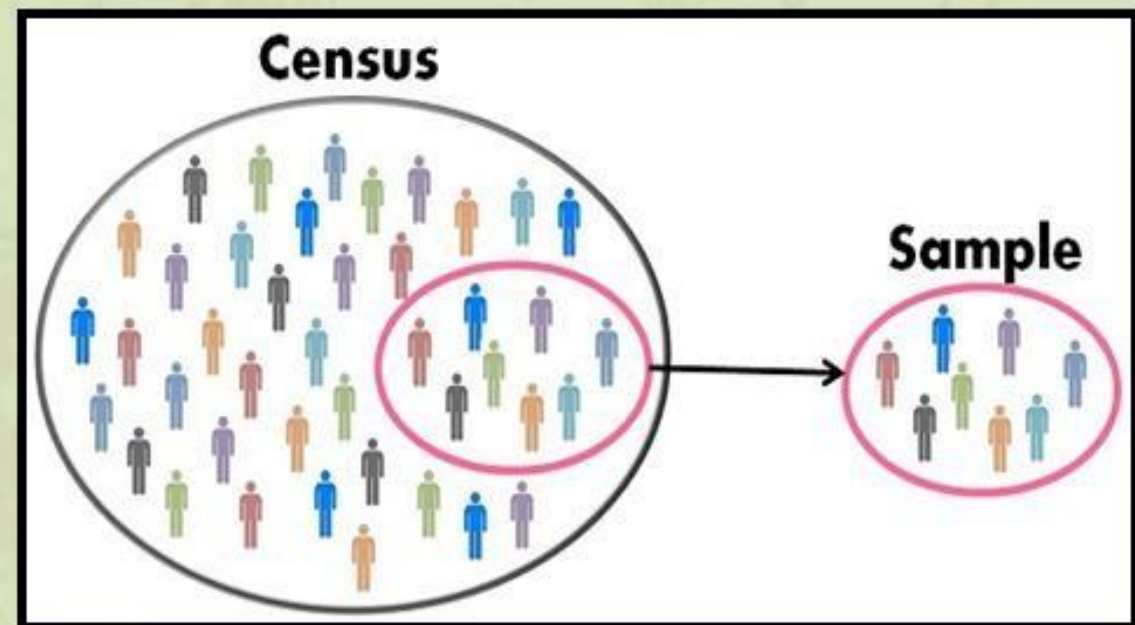
Do You Know?

Most of the Surveys are Sample Surveys.
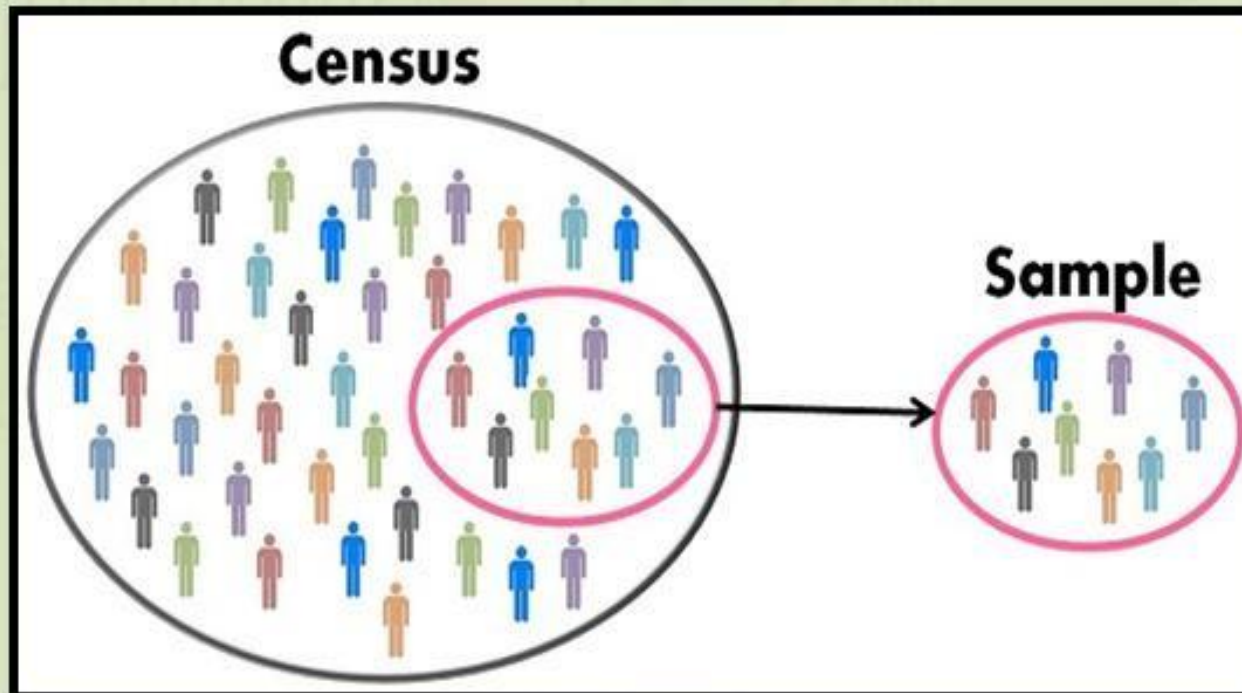
Note:

Representative Sample

It is a Sample which fairly accurately represents the whole Population (Universe).

Finding a Representative Sample is not an easy Task and requires Specialised Knowledge of different Techniques of Sample Selection.
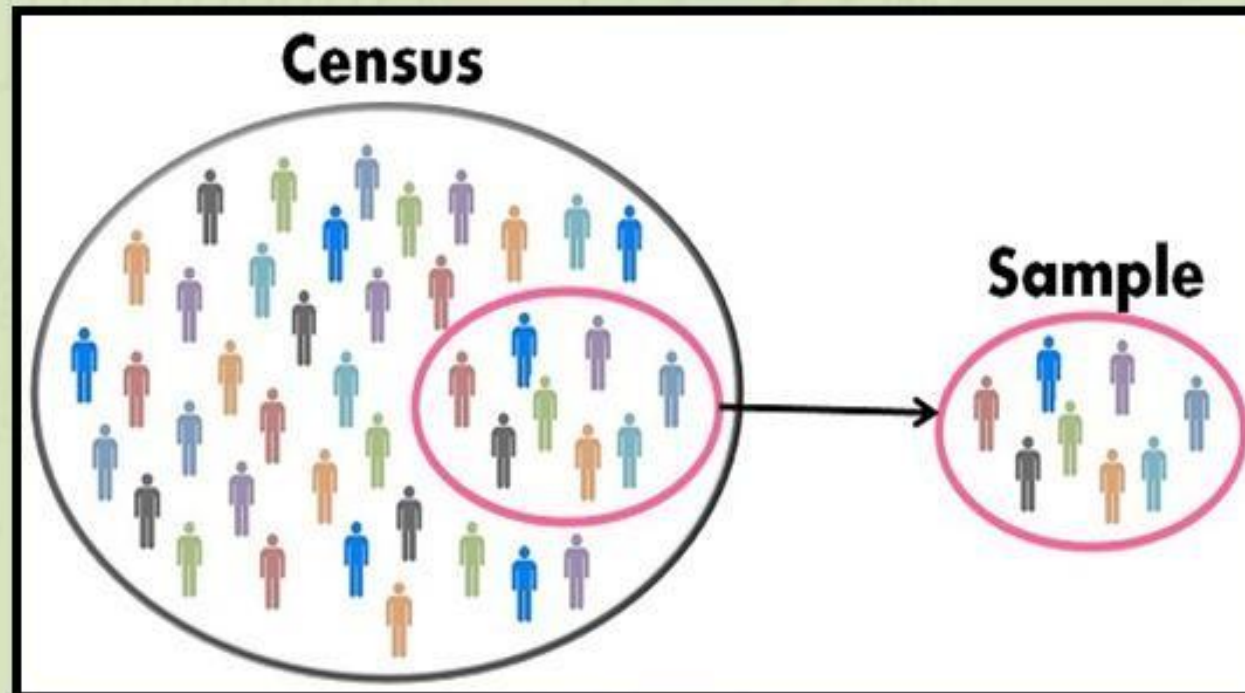
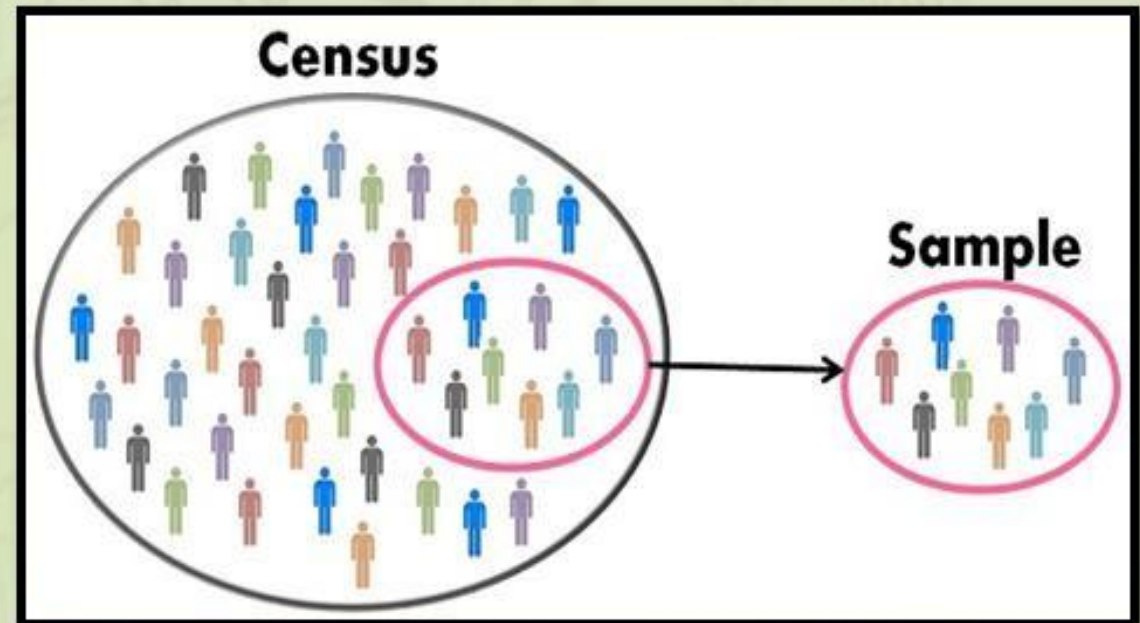# Census Method (Complete Enumeration) and Sample Method of Survey:

Sample Survey:

Suitability:

When the size of the Population (Field of Study) is fairly large.

High Degree of Accuracy is not required.

The Units are not diverse and one can fairly understand the population by studying few units.

There is no need to intensively examine every unit of Population.

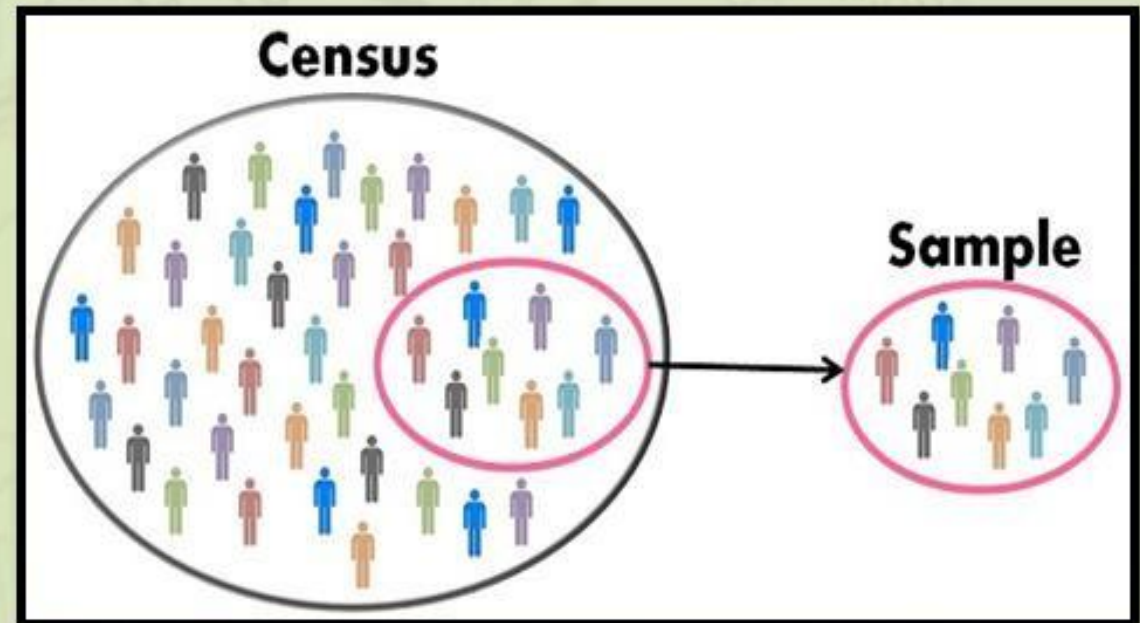# Census Method (Complete Enumeration) and Sample Method of Survey:

Sample Survey:

Merits:

It is Economical because only some of the units called Sample, have to be examined.

It is less time consuming as only Sample units not all units are studied.

Units can be studied more intensively because less units are to be studied.

It is effective for large Investigations.

Census Method (Complete Enumeration) and Sample Method of Survey:
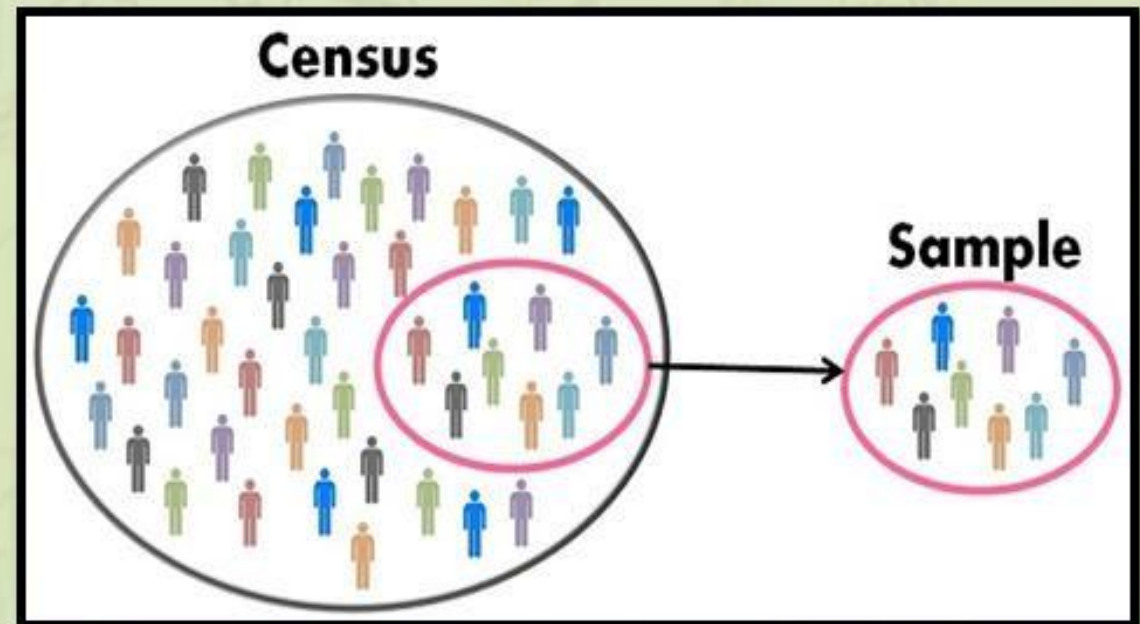
Sample Survey:

Demerits:

Results may not be 100% correct as only a part of the Population is studies.

If Sample selection is not right then results will not be reliable.

Investigator may be biased in Sample selection.

Finding a Representative Sample is a difficult task.

Note:

Sampling Frame

It is the list of the Items or Individuals forming a Population from which a Sample is taken.

Methods of Sampling

Random Sampling

Non-Random Sampling

Random Sampling:
(Lottery Method)

In Random Sampling, every Item of the Population has an equal change of being selected.
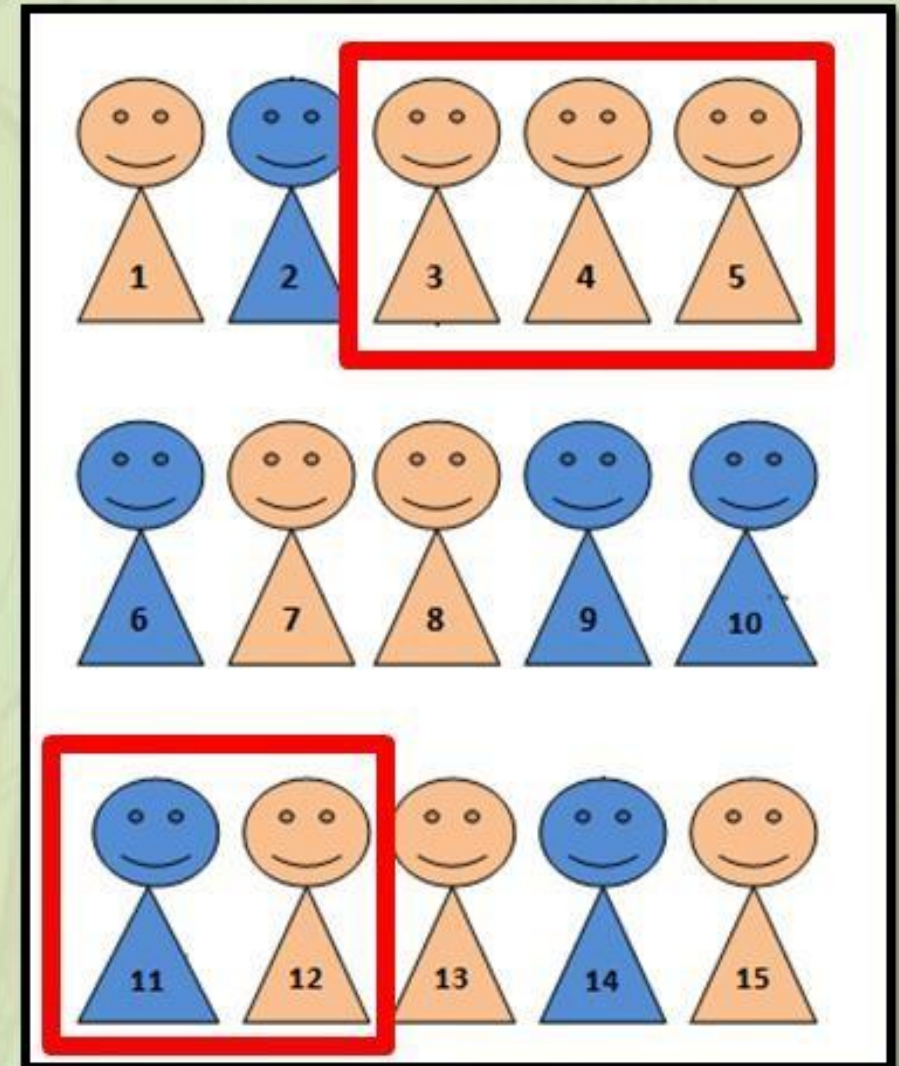
Non-Random Sampling:

In Non-Random Sampling, every Item of the Population does not have equal change of being selected.

Sample units depend on the convenience or Judgement of the Investigator.

Non-Random Sampling is mainly based on Judgement, Purpose, Convenience and Quota.

Sampling Errors and Non-Sampling Errors:

Sampling Errors:

Sampling Error is the difference between the result based on Sample Estimates and the result based on Census Estimate.

For Example:

Average of five Numbers 3,9,7,1,5 is **5**.

However if Sample is taken as **3** and **1**, Then out of this Sample Average is **2** [(3+1)/2].

Sampling Error =

Census Average – Sample Average

= **5 – 2 = 3**.

Sampling Errors can be minimised by taking larger Samples.

Note:

Sampling Errors
are
not possible
in
Census Method.

Sampling Errors and Non-Sampling Errors:

Non-Sampling Errors:

These include:

Sampling Bias,

Non-Response Errors,

Errors in Data Acquisition.

Sampling Errors and Non-Sampling Errors:

Non-Sampling Errors: Sampling Bias:

Its occurs where some elements of the population cannot be included in the sample.

Sampling bias results form not selecting a truly Random Sample which is a Representative Sample of the larger Population.

For Example the Head of the Village asks Surveyors to include all children of his home, in a study of Some Disease on Rural Children.

This might affect the results of the Sample as the Children of the Head of the Village must be more fit and healthier and less likely to have that particular Disease.

Sampling Errors and Non-Sampling Errors:

Non-Sampling Errors: Sampling Bias:

For Example:

A News Channel conducts an Exit Poll for elections where Party A and Party B are contesting.

Party A is more popular in Urban Areas (with less Population) and Party B is more popular in Rural Areas (with more Population). News Channel conducts an Exit Poll of 100 people which include 80 from Urban an 20 from Rural. This Exit Poll will show that party A is winning election but actually on ground, Party B wins the elections. The problem in the Sample of News Channel is that it is not a Representative Sample. Representative Sample should have included more people from Rural Areas than Urban Areas.

Sampling Errors and Non-Sampling Errors:

Non-Sampling Errors: Non-Response Errors:

Non-Response error happens when the Investigator is:

Either not able to reach a person chosen in the Sample,

Or the person chosen refused to respond.

In this case the final Sample may not be truly or fairly Representative Sample.

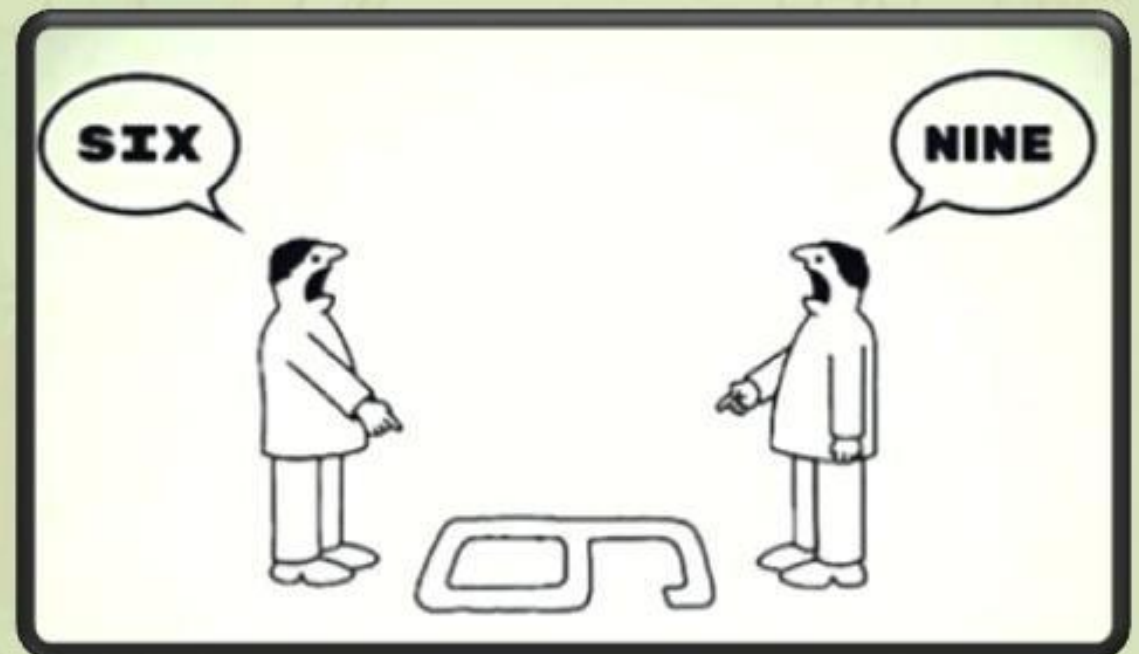# Sampling Errors and Non-Sampling Errors:

Non-Sampling Errors: Data Acquisition Errors:

It includes the errors of:

Recording the wrong data such as height of the students have been measured differently by different interviewers and

Data has been recorded wrongly such as 29 has been recorded as 92.

Note:

**Non-Sampling Errors
Can exist
in both
Census Method
And
Sample Method.**

Census of India

Agencies involved in the collection of Census of India includes:
Census of India,
National Sample Survey (NSSO),
Central Statistical Office (CSO),
Registrar General of India (RGI),
Directorate General of Commercial Intelligence and Statistics (DGCIS),
Labour Bureau etc.

Census of India

Census of India:

(Website: censusofindia.gov.in)

Census of India provides most comprehensive Demographic records of the Population.

Census is being conducted in India since 1881.

Census includes information related to different aspects of the Population such as:

Size of Population,

Population Density,

Sex Ratio,

Literacy Rate,

Migration of Population,

Rural Urban Population Ratio.

Census of India

National Sample Survey:

NSS is a Govt organisation with the objective of conducting National level Surveys on Socio-Economic issues.

NNS publishes quarterly reports called "**Sarvekshana**".

NSS provides information regarding:

Literacy rates,

School Enrolments,

Employment and Unemployment,

Morbidity, Maternity, Child Care,

Public Distribution System,

Manufacturing and

Service Sector Enterprises,

Industrial Activities, Retails Prices.

Do You Know?

Census of the year **2021** will be a Paperless Census.
It will use Digital Methods of Data Collection.