INTRODUCTION

When we obtain the information of a large group or population from only a part of it, the method or process is known as Sample Survey. The small part or unit of the entire population is known as the SAMPLE. We use sampling in our day to day life. For eg: When the doctor ask us to get the blood test done the lab assistant pricks out few drops of blood from our body and those drops are tested in the laboratory. The information or the test result of those few blood drops tells us about the blood of our entire body. So those few blood drops is known as SAMPLE and the process is known as SAMPLING.

BASIC TERMINOLOGIES:

1. **POPULATION**: In the Statistical terminology, all the items which fall within the purview of enquiry are known as POPULATION or UNIVERSE. In other words , the Population is a complete set of all possible observations of the type which is to be investigated. For example:
   a) The blood of our body.
   b) All the students of Thakur College.
   c) Total number of shops in Thakur Village.

2. **COMPLETE ENUMERATION** : When each and every unit , item or person of a population is examined or investigated , then the method is known as COMPLETE ENUMERATION or CENSUS. For example : After every 10 years the entire population of our country is surveyed , going from door to door. This type of survey is known as COMPLETE ENUMERATION or Census Survey.

3. **Population Unit** : Each person or each item of the Population is called POPULATION UNIT.

4. **Finite Population** : If a population consists of finite/countable number of units then the population is said to be FINITE POPULATION. For example : workers in a factory , students of a particular college.

5. **Infinite Population** : A population is said to be Infinite if the number of units of the population under study has infinite number of units. For example: the number of stars in the sky, the number of people watching a particular television program.

6. **Sampling** : The procedure of selecting SOME units from the population is known as Sampling. It is estimated that the samples give correct information about the concerned population. The decision is based on the study of the units known as samples and not the entire population. Sampling is used when it is not possible to study each and every unit of the Population. In Statistics the word POPULATION does not refer only to the people but to all items or units that have been chosen for study.

7. **Sample** : The selected portion of the entire population that is to be studied to give the information of the entire population is known as SAMPLE. It is always believed that the samples give correct idea about the population. Most of the decision of the population are

based on the study of the samples and the conclusion received after the sample survey. It is a finite subset of units defined in a population. The number of units in a sample is called sample size.

8. Sampling unit : The constituent of a population which are individuals to be sampled from the population ad cannot be further subdivided for the purpose of the sampling at a time are called sampling units. For exampling : To know the average age of the students of a particular class, if the age of any five students are taken for study then these five students are the sampling units.

9. Sample frame : It is essential to have a list to identify each sampling unit by a number to adopt any sampling procedure. Such a list or map is called sampling frame. For example : the list of voters, a list of students of a college, a list of workers of a factory, a list of farmers of a particular village, etc.

10. **Parameters and Statistic** : There are certain measures like mean, median , mode and standard deviation which describe the sample and the population . When they describe the characteristics of a population , they are called PARAMETER. When they describe the characteristics of a sample , they are called STATISTIC. So we can say that a Parameter is a characteristic of a population and a Statistic is a characteristic of a sample.
    When the parameters are unknown , they are estimated from the values of the statistic since samples are subsets of population. For population we use N ,μ,S as the standard symbols and for sample we use n, $\bar{x}$, s as the standard symbol to depict size, mean and standard deviation respectively.

11. **Estimator**: Estimator is the statistic whose calculated value is used to estimate a population parameter. Sample mean $\bar{y}$ is an estimator of population parameter μ. An estimator is always a random variable. The numerical value of the estimator is known as Estimate.

12. **Bias**: The difference between the expected value and the true value of the parameter being estimated is known as BIAS. Thus Bias B=E(t)-θ , where t is an estimator and θ is the parameter. Bias can be positive or negative.

INTRODUCTIONS (…..contd)

BASIC TERMINOLOGIES:

13. **Unbiased Estimator**: If E(t)=θ, then the estimator t is said to be an Unbiased Estimator of the parameter θ.

14. **Mean Square Error**: The average squared difference between the estimated values and what is estimated is known as Mean Square Error. The mean square error (MSE) of an estimator t of a parameter θ is the function of θ defined by $E\ (t-\theta\ )^2$.

15. **Standard Error** : Standard Error is the standard deviation of sampling distribution of statistic and is represented as S.E where S.E = $\sigma/\sqrt{n}$ where σ is the standard deviation and n is the number units in the sample or the size of the sample.

Methods of Collecting Data:

Information on population can be collected in two ways -**Census Method** and **Sample Method**.

1. **Census Method** or **Complete Enumeration Method**: When every element or unit of the Population is included in the investigation , then the method is known as CENSUS METHOD or COMPLETE ENUMERATION METHOD i.e. the data is collected from each and every unit of the Population. For example : if we want to study the pass percentage of a particular college , then the information of all the students of that college will be studied. In this method no student will be left out.
   Population Census of India : The population census of our country is taken after every 10 years. The last census was taken in 2010. The first census was taken in 1871-72.
   Merits and limitations of Census Method:
   1. **Merits** :
      a) The results are more accurate and reliable as the data are collected from each and every unit of the population and are studied.
      b) It is possible to do intensive study.
      c) The data collected may be used for various surveys and analysis.
   2. **Limitations** :
      a) It is a costly method and it requires a large number of enumerators.
      b) More resources such as money, labour, time, energy etc. are required in this type of survey.
      c) If the population/universe is infinite , then this type of survey is not possible.
      d) Sometimes this method is not feasible if it leads to destruction of the population. For example: if we want to know the average life of electric bulb of a particular company XYZ, then the Census Survey Method will destroy the entire population (of the bulbs).
2. **Sample Survey:** When a part or a fraction of the entire population is studied, then the method is known as Sample Survey. In this method some units of the population are selected and studied. These units are known as samples. The result obtained from these samples will be taken for the entire population.

Some surveys , for example: economic surveys, agricultural surveys etc. are conducted regularly. Some are conducted when the need arises, for example: consumer satisfaction surveys at a newly opened shopping mall to see the satisfaction level with the amenities provided in the mall.

Reasons for selecting a sample:

a) If the population is infinite, the method of complete enumeration is impossible.
b) If the result is required in short time the method of sample survey is used.
c) When the area of survey is wide.
d) When the resources for survey i.e .the money and trained persons are limited.
e) When the item or unit is destroyed during investigation.

Principles of Sampling:

1) Principle of Statistical regularity: A moderately large number of units chosen at random from a large population are almost sure on the average to possess the characteristic of the large group/population.
2) Principle of Inertia of large numbers: As the sample size increases, the result tend to be more accurate and reliable.
3) Principle of Validity: This states that the sampling methods provide valid estimates about the population units(parameter).
4) Principle of Optimization: This principle takes into account the desirability of obtaining a sampling design which gives optimum results.

The most important purpose of sampling is to get maximum information about the population under consideration at minimum cost, time and human resource. This is best achieved when the sample contains all the properties of the population.

Principles of Sampling

1. Principle of Statistical regularity: A moderately large number of units chosen at random from a large population are almost sure on the average to possess the characteristic of the large group/population.
2. Principle of Inertia of large numbers: As the sample size increases, the result tend to be more accurate and reliable.
3. Principle of Validity: This states that the sampling methods provide valid estimates about the population units(parameter).
4. Principle of Optimization: This principle takes into account the desirability of obtaining a sampling design which gives optimum results.

The most important purpose of sampling is to get maximum information about the population under consideration at minimum cost, time and human resource. This is best achieved when the sample contains all the properties of the population.

Principal Steps in a Sample Survey

The Principal steps in a Sample Survey are:

(1) **State objectives clearly**: The objectives of the survey has to be clearly defined and well understood by the person who is planning to conduct it.

(2) **Decide on the population to be sampled**: Decision should be taken on the population that is to be sampled that coincide with the information that is needed. Based on the objectives of the survey, the decision of the population from which the information can be obtained should be taken. For example: The population of the patients have to be sampled for determining the medical facilities in a hospital. The population of the students have to be sampled for determining the facilities and opportunities available in a college/school.

(3) **Decide on methods of measurement**: We have to decide on the method adopted for data collection. Data can be collected by one of the following methods:
a) **Physical observations and measurements**: The surveyor contacts the respondents personally through meeting. He observes the sampling unit and records the data.
b) **Personal Interview**: In this method the surveyor goes to the respondents and ask some questions to them. Those questions are provided to them and these set of questions are known as Questionnaire. Then according to the response from the respondents the data in the questionnaire is filled accordingly.
c) **Telephonic interview**: If the investigator ask questions to the respondents over the telephone then the method is known as Telephonic Interview. The advantage of this method is that it is cheaper than the personal interviews and can be conducted in a shorter period of time. They allow the researcher to assist the respondent by clarifying the questions. When the respondent is hesitant in answering some personal questions, then this method is better.
d) **Mail enquiry**: When the well prepared questionnaire is sent to the respondents through postal mail, e-mail etc. then this method is known as Mail Enquiry. In this method

the respondents are requested to fill up the questionnaire and send it back to the researcher.

e) **Web Based Enquiry**: In this method the survey is conducted Online through internet based web pages. This facility is provided by various websites. The questionnaires are prepared by the surveyor/researcher in the proper format and the link is sent to the respondents through E-mail or Whatsapp. By clicking on the link the respondent's answers are recorded. For example: google forms.

f) **Recorded Information**: The sample of data is collected from the already recorded information. The advantage of this method is that time, money and labour is saved to a great extent because the data is ready for the survey.

(4) **Decide on type of sampling that needs to be done:** Here the first step is to divide the entire population into parts. The sampling units are to be identified.

(5) **Select sample relevant to cost and time:** 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 noted.

(6) **Pre-test Questionnaire :** The questionnaire is tested in advance with a small sample of respondents to eliminate troubles. This method of pre – testing is known as Pilot Survey. Pilot survey helps in assessing the suitability of questions, clarity of instructions, performance of enumerators and the cost and time involved in the actual survey. So this method may reveal some troubles and problems beforehand which the surveyor may face in field in large scale surveys.

(7) **Organise fieldwork:** It means that surveyors are given proper training regarding procedures, plans for handling the non-responses and missing observations etc.

(8) **Analyse and Summarize the data:** Based on the objectives of the survey ad the data, the suitable statistical tool is decided which can answer 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.

For example: the data has to be qualitative, quantitative, nominal, ordinal, etc. After getting the questionnaire back, it reads to be edited to rectify the recording error 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 survey is started. Different methods of estimation may be available to get the answer of the same query from the same set of data.

(9) Write report of the finding:

Prepare the report based on the findings and analysis using different pictorial representation of data for better understanding

Principal Steps in a Sample Survey

The Principal steps in a Sample Survey are:

1. **Pre-test Questionnaire :** The questionnaire is tested in advance with a small sample of respondents to eliminate troubles. This method of pre – testing is known as Pilot Survey. Pilot survey helps in assessing the suitability of questions, clarity of instructions, performance of enumerators and the cost and time involved in the actual survey. So this method may reveal some troubles and problems beforehand which the surveyor may face in field in large scale surveys.

2. **Organise fieldwork:** It means that surveyors are given proper training regarding procedures, plans for handling the non-responses and missing observations etc.

3. **Analyse and Summarize the data:** Based on the objectives of the survey ad the data, the suitable statistical tool is decided which can answer 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.

   For example: the data has to be qualitative, quantitative, nominal, ordinal, etc. After getting the questionnaire back, it reads to be edited to rectify the recording error 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 survey is started. Different methods of estimation may be available to get the answer of the same query from the same set of data.

4. **Write report of the finding:**

   Prepare the report based on the findings and analysis using different pictorial representation of data for better understanding

Characteristics of a good Questionnaire

(a) Number of questions should be minimum.

(b) Questions should be in logical order, moving from easy to more difficult questions or the level of difficulty should increase chronologically.

(c) Questions should be short and simple. Technical terms and vague expressions which are capable of different interpretations should be avoided.

(d) Questions fetching YES or NO answers are preferable. There may be some multiple choice questions requiring lengthy answers are to be avoided.

(e) The questions which are personal or which require memory power and calculations should be avoided.

(f) Questions should be enabled to be cross checked. Deliberate or unconscious mistakes can be detected to an extent.

(g) Questions should be carefully framed so as to cover the entire scope of the survey.

(h) The working used in the questionnaire should be such that they don't hurt the feelings of the respondent or arouse resentment.

(i) Confidential information should not be asked in the questionnaire as far as possible.

(j) Physical appearance should be attractive and sufficient space should be provided for answering each of the questions.

Difference between CENSUS and SAMPLE SURVEY

The major difference between Census and Sample Survey are as follows:

(1) The Census is the systematic method of collecting and recording data about all the members or units of the population. Sampling is defined as the subset of the population selected to represent the entire population having all the characteristics of the population.

(2) The Census is alternately known as Complete Enumeration Survey Method. In contrast, Sampling is also known as a Partial Enumeration Survey Method.

(3) In the Census Method, each and every unit/member of the population is researched. But in Sample Survey only a handful of units/members are selected from the population for analysis and research. For example: If the researcher wants to study the percentage of Covid 19 patients of Maharashtra, then a sample can be 10% of the total population of Maharashtra.

(4) Census is a very time-consuming method of survey whereas in sampling method the survey does not take much time because the number of units which are to be studied are less than the number of units in the population.

(5) The Census Method need high capital investment as it involves the collection and research of all the values of the population. But Sampling is an economic method because a part of the population is studied.

(6) The results drawn by conducting a Census is accurate and reliable whereas there are chances of errors in the result drawn from the Sample Survey Method as each and every unit/member of the population is not studied. For example: During election time the television network provide election coverage. They also try to predict results. This is done through exit polls wherein a random sample of voters who exit the polling booths are asked whom they voted for. From the data of the sample of voters, the prediction is made. So exit poll predictions are not always correct.

(7) The size of the sample determines the probability of error in the outcomes, i.e, larger the size of the sample less are the chances of errors and smaller the size , higher are the chances of errors. This is not possible with Census Method as all the units of the population are taken into consideration.

(8) When population units are destructed while taking observations about the characteristic under study, then Census Survey Method is not feasible. Sample Survey Method is possible in all situations.

(9) If the population is of heterogeneous nature then Census Method is best suited. But Sampling Method is appropriate for homogeneous nature of the population.

Sampling Errors and **Non Sampling Errors**

The two types of errors in a Sample Survey are **Sampling Errors** and **Non Sampling Errors**.

1) **Sampling Errors:** Although a sample is a part of the population, it cannot be expected generally to supply full information about the population. So there may be in most cases some difference between Statistic and Parameter. The discrepancy between parameter and its estimate due to sampling process is known as Sampling Error.
2) **Non-Sampling Errors:** When error occur during collection of actual information, these errors are known as Non-Sampling Errors. Sampling Error is one which occur due to un-representativeness of the sample collected for observation. On the other hand , non-sampling error is an error which arises from human error such as in problem identification, method or procedure used etc.

Significant differences between **Sampling** and **Non- Sampling Errors**

1) Sampling Error is a statistical error which happens in case when the sample selected does not perfectly represent the population of interest. Non-Sampling Error occurs due to sources while conducting survey activities.
2) Sampling Error arises because of the variation between the true mean value of the sample and the population. On the other hand, the Non-Sampling Error arises because of deficiency and inappropriate analysis of data.
3) Non-Sampling Error can be random or non-random whereas Sampling Error occurs in the random sampling only.
4) Sample Error arises only when the sample is taken as a representative of the population. But Non-Sampling Error arise both in sampling and complete enumeration or census method.
5) Sampling Error is mainly dependent on sample size, i.e., if the sample size increases, the possibility of error decreases. But Non-Sampling Error is not dependent on the sample size. So with the increase in sample size , it will not be reduced.

TYPES OF SAMPLING

The fundamental importance in Sampling Theory is the technique of selecting a sample and it depends upon the nature of investigation. The sampling procedures which are commonly used may be classified as:

1. Probability Sampling
2. Non-Probability Sampling

Advantages and Limitations of Sampling

1) **Advantages**: There are many advantages of sampling methods over census method. These are as follows:
   (a) Sampling saves time and labour.
   (b) It results in reduction of cost in terms of money and man-hour.
   (c) Sampling ends up with greater accuracy of results.
   (d) It has greater scope.
   (e) It has greater adaptability.
   (f) If the population is large, hypothetical or destroyable, sampling is the only method to be used.
2) **Limitations**: The limitations of sample are as follows:
   (a) Sampling is to be done by qualified and experienced persons. Otherwise the results may be wrong .
   (b) Sample method may give extreme values sometimes instead of the mixed values.
   (c) There is the possibility of sampling errors. Census Survey is free from sampling errors.

Types Of Sampling

The fundamental importance in Sampling Theory is the technique of selecting a sample and it depends upon the nature of investigation. The sampling procedures which are commonly used may be classified as:

(1) Probability Sampling
(2) Non-Probability Sampling

(1) **Probability Sampling** (Random Sampling): A probability sampling is one where the selection of units from the population is made according to known probabilities (finite population).The commonly used methods of this type of sampling are :
   (i) Simple Random Sampling
   (ii) Stratified Random Sampling
   (iii) Systematic Random Sampling
   (iv) Cluster Sampling
   (v) Two Stage Sampling
   (vi) Probability Proportional To Sample Size (PPS)

Probability Sampling (Random Sampling)

A probability sampling is one where the selection of units from the population is made according to known probabilities (finite population).The commonly used methods of this type of sampling are :

(i) Simple Random Sampling
(ii) Stratified Random Sampling
(iii) Systematic Random Sampling
(iv) Cluster Sampling
(v) Two Stage Sampling
(vi) Probability Proportional To Sample Size (PPS)

Advantages and **Disadvantages** of Probability Sampling:

ADVANTAGES:
(a) This process is both cost and time effective.
(b) Probability Sampling is an easy way of sampling as it does not involve a complicated process.
(c) This method of sampling does not require any technical knowledge because of the simplicity with which this can be done.

DISADVANTAGES:
(a) This may not be feasible for cost, time or personal reasons.
(b) Cannot be used when complete population list is not available.
(c) More complex than Non-Probability Sampling.

Non-Probability Sampling

It is the one where discretion of the enumerator or surveyor is used to select 'representative' units from the population (or) to infer that a sample is a representative of the population. This method is also called Judgement Sampling or Purposive Sampling.

This method is mainly used for opinion surveys, a common type of judgement sample used in surveys in quota sample. This method is not used in general because of prejudice and bias of the enumerator.

However, if the enumerator is experienced or expert, then this method may yield valuable results. For example: in the market research survey of the performance of their new car, the sample will be all new car purchases.

Non-Probability Sampling types commonly used are:

(i) Purposive Sampling
(ii) Convenient Sampling
(iii) Quota Sampling

Advantages of Non-Probability Sampling:

(i) Non-Probability sampling is a more conducive and practical method for researchers deploying survey in the real world although statisticians prefer probability sampling as it gives unbiased result. However, if done correctly non – probability sampling can yield similar if not the same quality of results.

(ii) Getting responses using non-probability sampling is faster and more cost-effective as compared to the probability sampling because sample is known to the researcher. They are motivated to respond quickly as compared to people who are randomly selected.

Disadvantages of Non-Probability Sampling:

(i) In non-probability sampling , the researcher needs to think through potential reasons for biases. It is important to have a sample that represents closely the population.

(ii) While choosing a sample in non-probability sampling, researchers need to be careful about recruits distorting data. At the end of the day, research is carried out to obtain meaningful insights and useful data.

Significant Differences between Probability & Non-Probability Sampling

(i) The sampling technique , in which the units of the population get an equal opportunity of being selected as a representative sample is known as Probability Sampling. A sampling method in which it is known that which individual or unit from the population will be chosen as a sample is called Non-Probability Sampling.

(ii) In Probability Sampling , the enumerator chooses the representative to be part of the sample randomly, whereas, in Non-Probability Sampling, the units are selected arbitrarily by the researcher.

(iii) The basis of probability Sampling is randomization or chance , so it is also known as Random Sampling. On the other hand, in Non-Probability Sampling, randomization technique is not used for selecting a sample. Hence it is considered as Non-Probability Sampling.

(iv) As the units are selected randomly by the researchers in Probability Sampling , so the extent to which it represents the whole population is higher as compared to the Non-Probability Sampling.

(v) Probability Sampling is used when the research is conclusive in nature. On the other hand, when the research is exploration, Non-Probability Sampling should be used.

(vi) The chances of selection in Probability Sampling is fixed and known as opposed to Non-Probability Sampling, the selection probability is zero , i.e, it is neither specified not known.

(vii) The results generated by Probability Sampling are free from bias while the results of Non-Probability Sampling are more or less biased.

National Statistical Office

The Ministry of Statistics and Programme Implementation came into existence as an Independent Ministry on 15-10-1999 after the merger of the Department of Statistics and the Department of Programme Implementation. The Ministry has two wings, one related to Statistics and the other Programme Implementation. The Statistics Wing called the **National Statistical Office**(**NSO**) consists of the **Central Statistical Office** (**CSO**), The Computer Centre and the National Sample Survey Office (**NSSO**).

National Statistical Office (**NSO**) has the following **responsibilities**:

1) Acts as the nodal agency for planned development of the statistical system in the country, lays down and maintains norms and standards in the field of statistics, involving concepts and definitions, methodology of data collection, processing of data and dissemination of results.
2) Co-ordinates the statistical work in respect of the Ministries/Departments of the Government of India on statistical methodology and on statistical analysis of data.
3) Prepares national accounts as well as publishes annual estimates of national product, government and private consumption of fixed capital, as also the state level gross capital formation of Supra-Regional sectors and prepares comparable estimates of State Domestic Product(**SDP**) at current prices.
4) Maintains liaison with international statistical organisations such as the United Nations Statistical Division(**UNSD**), the Economic and Social Commission for Asia and the Pacific(**ESCAP**), the Statistical Institute for Asia and the Pacific(**SIAP**), the International Monetary Fund(**IMF**), the Asian Development Bank(**ADB**), the Food and Agriculture Organisations(**FAO**), the International Labour Organisations(**ILO**)etc.
5) Compiles and releases the Index of Industrial Production(**IIP**) every month in the form of quick estimates; conducts the Annual Survey of Industries(**ASI**) and provides statistical information to assess and evaluate the changes in growth, composition and structure of organised manufacturing sector.

Central Statistical Office(**CSO**)

The main responsibility assigned to the CSO is to bring about co-ordination of statistical activities among various statistical agencies in the Central Government and of Statistical Bureaus of State Governments which were set up for similar co-ordination of activities of statistical agencies of the state level.

The **CSO** has the following **6 Divisions**:

1) National Accounts Division(**NAD**): This Division is responsible for the preparation of national accounts which includes Gross Domestic Product, Government and Private Final Consumption Expenditure, Fixed Capital Formation and other macro-economic aggregates.
2) Social Statistic Division(**SSD**): This Division is entrusted with Statistical monitoring on the Millennium Development Goals; Environmental Economic Accounting; Grant-in-aid for research; workshops/seminars/conferences in Official/Applied Statistics, National/ International awards for Statisticians; National Data Bank (**NDB**) on socio-religious categories, Basic Statistics for Local Level Development(**BSLLD**) Pilot Scheme; Time-use survey and release of regular and ad-hoc statistical publications.
3) Economic Statistics Division(**ESD**): This Division conducts Economic Censuses and Annual Surveys of Industries (**ASI**), compiles All India Index of Industrial Production (**IIP**), Energy Statistics and Infrastructure Statistics etc.
4) Training Division: This Division is primarily responsible for the training of man-power in theoretical and applied statistics to tackle the emerging challenges of data collection, analysis and dissemination required for evidence based policy making as also for planning, monitoring and evaluation.

5) Co-ordination and Publication Division (**CAP**): The Division looks after co-ordination work within CSO as well as with the Ministries and State/UT Government in Statistical matters, organizes Conference of Central and State Statistical Organizations (**COCSSO**) and 'Statistics Day' every year, prepares Results Framework Document(**RFD**), Citizens'/ Clients' Charter and Annual Action Plan and Outcome of Budget and Annual Plan of the Ministry.
6) Data Storage and Data Dissemination (**DSDD**) including Computer Centre: It handles the data processing jobs of the MOSPI, provides training to statistical personal on software, maintains the MOSPI's website and the National Data Warehouse of Official Statistics.

National Sample Survey Office (**NSSO**)

Formerly known as National Sample Survey Organization, **NSSO** is responsible for conduct of large scale surveys in diverse fields in All India basis. Primary data are collected through nation-wide household surveys on various socio-economic subjects, Annual Survey of Industries (**ASI**) etc. It also 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.

Central Statistical Office(CSO)

The main responsibility assigned to the CSO is to bring about co-ordination of statistical activities among various statistical agencies in the Central Government and of Statistical Bureaus of State Governments which were set up for similar co-ordination of activities of statistical agencies of the state level.

The **CSO** has the following **6 Divisions**:

1) National Accounts Division(**NAD**): This Division is responsible for the preparation of national accounts which includes Gross Domestic Product, Government and Private Final Consumption Expenditure, Fixed Capital Formation and other macro-economic aggregates.
2) Social Statistic Division(**SSD**): This Division is entrusted with Statistical monitoring on the Millennium Development Goals; Environmental Economic Accounting; Grant-in-aid for research; workshops/seminars/conferences in Official/Applied Statistics, National/ International awards for Statisticians; National Data Bank (**NDB**) on socio-religious categories, Basic Statistics for Local Level Development(**BSLLD**) Pilot Scheme; Time-use survey and release of regular and ad-hoc statistical publications.
3) Economic Statistics Division(**ESD**): This Division conducts Economic Censuses and Annual Surveys of Industries (**ASI**), compiles All India Index of Industrial Production (**IIP**), Energy Statistics and Infrastructure Statistics etc.
4) Training Division: This Division is primarily responsible for the training of man-power in theoretical and applied statistics to tackle the emerging challenges of data collection, analysis and dissemination required for evidence based policy making as also for planning, monitoring and evaluation.
5) Co-ordination and Publication Division (**CAP**): The Division looks after co-ordination work within CSO as well as with the Ministries and State/UT Government in Statistical matters, organizes Conference of Central and State Statistical Organizations (**COCSSO**) and 'Statistics Day' every year, prepares Results Framework Document(**RFD**), Citizens'/ Clients' Charter and Annual Action Plan and Outcome of Budget and Annual Plan of the Ministry.
6) Data Storage and Data Dissemination (**DSDD**) including Computer Centre: It handles the data processing jobs of the MOSPI, provides training to statistical personal on software, maintains the MOSPI's website and the National Data Warehouse of Official Statistics.

National Sample Survey Office (NSSO)

Formerly known as National Sample Survey Organization, **NSSO** is responsible for conduct of large scale surveys in diverse fields in All India basis. Primary data are collected through nation-wide household surveys on various socio-economic subjects, Annual Survey of Industries (**ASI**) etc. It also 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.

< 08-07-2020 >

National Sample Survey Office (NSSO)

Formerly known as National Sample Survey Organization, **NSSO** is responsible for conduct of large scale surveys in diverse fields in All India basis. Primary data are collected through nation-wide household surveys on various socio-economic subjects, Annual Survey of Industries (**ASI**) etc. It also 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

Four divisions of the NSSO:

1) **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 plans, analysis and presentation of survey results.
2) **Field Operations Division** (FOD): This Division has its headquarters at Delhi/ Faridabad and a network of **six** Zonal Offices, **49** Regional Offices and **118** Sub-Regional Offices spread throughout the country. It is responsible for the collection of primary data for the surveys undertaken by **NSSO**.
3) **Data Processing Division** (DPD): The Division, with its headquarters at Kolkata and **6** other Data Processing Centres at various, is responsible for sample selection, software development, processing, validation and tabulation of the data collected through surveys.
4) **Co-ordination of Publication Division** (CPD): This Division, located at New Delhi, co-ordinates all the activities of different Divisions of **NSSO**. It also brings out the **bi-annual** journal of **NSSO**, titled ''**Sarvekshana**'', and organises National Seminars on the results of various Socio-economic surveys undertaken by **NSSO**.

< 10-07-2020 >

Simple Random Sampling

Definition: Simple Random Sampling is a sampling technique in which each and every unit of the population has an equal chance of being included/selected in the sample. Thus the selection is free from any personal bias as the investigator does not make any preference in the choice of units. Since selection of items/units in the sample are entirely dependent on chance, so this method is Probability Sampling. If the size of the sample is large then it will represent the population.

SIMPLE RANDOM SAMPLE (SRS)

SRS is a method of selection of a sample comprising of 'n' number of sampling units out of the population having 'N' number of total sampling units such that every unit has an equal chance of being selected/chosen.

SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT (SRSWOR)

In this method the population units can appear in the sample at most once, i.e. , the unit once selected is not returned in the population before the next selection.

SIMPLE RANDOM SAMPLING WITH REPLACEMENT (SRSWR)

In this method the population units may appear in the sample more than once, i.e., the unit once selected is returned to the population before the next selection.

Methods of Selection of a Simple Random Sample

The commonly used methods of selection of a Simple Random Sample are:

1) **Lottery Method**
2) **Random Numbers Table**
3) **Calculators or Computers**

1) **Lottery Method**

This is the most popular and simplest method of Simple Random Sampling. In this method all the items of the population are numbered and are written on separate slips of paper of same size, shape and colour. They are folded and mixed up in a container. The required numbers of slips are selected at random for the desire sample size. If SRSWOR is to be drawn then the slips are selected one after the other without putting them back into the container. But if SRSWR is to be drawn then after noting the number of the selected slip, the slip is replaced in the container before next draw/selection.

Drawbacks:

1) The process of writing N number of slips is cumbersome and shuffling a large number of slips, where the population size is very large , is difficult.

2) This method is inapplicable for infinite population.
3) Human bias may enter while selecting the slips. Hence the other alternative method, i.e., use of random number tables can be used.

2) Table of Random Numbers

When the population is large, this method is used. There are several standard tables of random numbers. Some of them are as follows:

(i) Tippett's Table
(ii) Fisher and Yates' Table
(iii) Kendall and Smith's Table

Example 1. Ten random numbers are given below:

34, 80, 50, 02, 90, 61, 24, 85, 13, 49.

Select 6 units out of 40 units under:

(i) With Replacement,
(ii) Without Replacement

Solution: The population units are numbered 1 to 40.

Sr. No.	Random Number(RN)	Remainder (RN-40)	Sampled Unit	SRSWR Sampled unit	SRSWOR Sampled unit
1	34	34	34	34	34
2	80	40	40	40	40
3	50	10	10	10	10
4	02	02	02	02	02
5	90	10	10	10	Reject/ Ignore
6	61	21	21	21	21
7	24	24	24	____(sample of size 6)	24

3) Calculators or Computers

Random number can be generated through scientific calculators or computers. For each press of the RND command we get new random numbers. The ways of selection of sample is similar to that of using random number table.

PROBABILITY OF DRAWING A SAMPLE

Case 1: SRSWOR

Let the population contains N units and we have to select a sample of size n.

If n units are selected by SRSWOR, the total number of possible samples is $1/{}^{N}C_{n}$.

Case 2: SRSWR

When n units are selected with SRSWR, the total number of possible samples is N^{n}.

Thus the probability of drawing a sample is $1/N^{n}$

PROBABILITY OF DRAWING AN UNIT

Case 1: SRSWOR

Let A_j denote an event that a particular unit u_i is not selected at the jth draw. Then the probability of selecting, say, ith unit at kth draw is

P (selecting of u_i at kth draw) $= P(A_1 \cap A_2 \cap A_3 \ldots\ldots \bar{A}_k)$

$= P(A_1)\, P(A_2/A_1) \ldots\ldots\, P(A_{k-1}/A_1A_2 \ldots A_{k-2})\, P(\bar{A}_k/A_1A_2 \ldots A_{k-1})$

$=(1-1/N)\,(1-1/N-1) \ldots\ldots (1-1/N-k+2)\,(1/N-k+1)$

$=(N-1/N)\,(N-1-1/N-1) \ldots\ldots (N-k+2-1/N-k+2)\,(1/N-k+1)$

$=(N-1/N)\,(N-2/N-1) \ldots\ldots (N-k+1/N-k+2)\,(1/N-k+1)$

$=1/N$

Case 2: SRSWR

In this method the population size remains the same before each selection/draw. Hence probability of selection of any unit in the sample at any draw is 1/N.

P (selecting of ith unit u_i at kth draw) = 1/N

Example 2: How many random samples of size 2 can be drawn from a population of size 10 if sampling is done (a) with replacement, (b) without replacement?

Solution: Given: N=10, n=2

(a) We know that when n units are selected from population of size N with SRSWR , the total number of possible samples is $N^n = 10^2 = 100$.

(b) We know that when n units are selected from population of size N with SRSWOR, the total number of possible samples is $({}^{N}C_n) = {}^{10}C_2 = 45$

PROBABILITY OF DRAWING A SAMPLE

Case 1: SRSWOR

Let the population contains N units and we have to select a sample of size n.

If n units are selected by SRSWOR, the total number of possible samples is $1/{}^{N}C_{n}$.

Case 2: SRSWR

When n units are selected with SRSWR, the total number of possible samples is N^{n}.

Thus the probability of drawing a sample is $1/N^{n}$

PROBABILITY OF DRAWING AN UNIT

Case 1: SRSWOR

Let A_j denote an event that a particular unit u_i is not selected at the jth draw. Then the probability of selecting, say, ith unit at kth draw is

P (selecting of u_i at kth draw) $= P(A_1 \cap A_2 \cap A_3 \ldots\ldots \bar{A}_k)$

$= P(A_1)\, P(A_2/A_1) \ldots\ldots P(A_{k-1}/A_1A_2\ldots\ldots A_{k-2})\, P(\bar{A}_k/A_1A_2\ldots\ldots A_{k-1})$

$= (1-1/N)\,(1-1/N-1) \ldots\ldots (1-1/N-k+2)\,(1/N-k+1)$

$= (N-1/N)\,(N-1-1/N-1) \ldots\ldots (N-k+2-1/N-k+2)\,(1/N-k+1)$

$= (N-1/N)\,(N-2/N-1) \ldots\ldots (N-k+1/N-k+2)\,(1/N-k+1)$

$= 1/N$

Case 2: SRSWR

In this method the population size remains the same before each selection/draw. Hence probability of selection of any unit in the sample at any draw is 1/N.

P (selecting of ith unit u_i at kth draw) = 1/N

Example 2: How many random samples of size 2 can be drawn from a population of size 10 if sampling is done (a) with replacement, (b) without replacement?

Solution: Given: N=10, n=2

(c) We know that when n units are selected from population of size N with SRSWR , the total number of possible samples is $N^n = 10^2 = 100$.
(d) We know that when n units are selected from population of size N with SRSWOR, the total number of possible samples is $({}^{N}C_{n}) = {}^{10}C_{2} = 45$

SIMPLE RANDOM SAMPLING FOR VARIABLES

PROBABILITY OF DRAWING AN UNIT

Case 1: SRSWOR

Let A_j denote an event that a particular unit u_i is not selected at the jth draw. Then the probability of selecting, say, ith unit at kth draw is

P (selecting of u_i at kth draw) $= P(A_1 \cap A_2 \cap A_3 \ldots\ldots.. \bar{A}_k)$

$= P(A_1)\, P(A_2/A_1) \ldots\ldots\ldots P(A_{k-1}/A_1A_2 \ldots\ldots A_{k-2})\, P(\bar{A}_k/A_1A_2 \ldots\ldots A_{k-1})$

$=(1-1/N)\,(1-1/N-1) \ldots\ldots\ldots (1-1/N-k+2)\,(1/N-k+1)$

$=(N-1/N)\,(N-1-1/N-1) \ldots\ldots.. (N-k+2-1/N-k+2)\,(1/N-k+1)$

$=(N-1/N)\,(N-2/N-1) \ldots\ldots\ldots.. (N-k+1/N-k+2)\,(1/N-k+1)$

$=1/N$

Case 2: SRSWR

In this method the population size remains the same before each selection/draw. Hence probability of selection of any unit in the sample at any draw is 1/N.

P (selecting of ith unit u_i at kth draw) = 1/N

Example 2: How many random samples of size 2 can be drawn from a population of size 10 if sampling is done (a) with replacement, (b) without replacement?

Solution: Given: N=10, n=2

(e) We know that when n units are selected from population of size N with SRSWR , the total number of possible samples is $N^n = 10^2 = 100$.

(f) We know that when n units are selected from population of size N with SRSWOR, the total number of possible samples is $({}^{N}C_{n}) = {}^{10}C_{2} = 45$

SIMPLE RANDOM SAMPLING FOR VARIABLES

When the data values are variables we select the sample units and the characteristic under study is noted for selected units.

Notations:

N= Number of units in the population (population size)

N= Number of units in the sample (sample size)

Y_i = Value of the characteristic for the i^{th} unit of the population i = 1, 2, 3,………., N

$Y = \Sigma Y_i$ (i=1, 2, 3,…., N) = Population total (i=1,2,3,….,N)

$\bar{Y} = 1/N\, \Sigma Y_i$ (i= 1, 2, 3,…., N) = Population Mean (i=1,2,3,……N)

$\sigma^2 = 1/N\ \Sigma\ (Y_i - \bar{Y})^2$ = Population variance (i=1,2,3…..,N)

$S^2 = 1/N\text{-}1\ \Sigma\ (Y_i - \bar{Y})^2$ = Population mean square error. (i=1,2,3,…..N)

Let y_i denotes the i^{th} observation from the sample. Then

$\bar{y} = 1/n\ \Sigma\ y_i$ = sample mean (i=1,2,3,……,n)

$s^2 = 1/n\text{-}1\ \Sigma\ (y_i - \bar{y})^2$ = sample mean square error (i=1,2,3,…..,n)

Theorem 1: In SRSWOR, sample mean is an unbiased estimate of the population mean.

Proof: Let $Y_1, Y_2, \ldots, Y_N$ be the population values and $y_1, y_2, \ldots, y_n$ be the sample observations when a sample of size n is selected from a population of size N by the method of SRSWOR. Then,

Population mean = $\bar{Y} = 1/N \, \Sigma Y_i, \quad i=1, 2, 3, \ldots N$

Sample mean = $\bar{y} = 1/n \, \Sigma y_i, \quad i=1, 2, 3, \ldots n$

Which can be written as

$\bar{y} = 1/n \, \Sigma y_i \quad I = 1, 2, \ldots n$

$=1/n \, \Sigma a_i Y_i \quad i = 1, 2, \ldots N$

Where $a_i = 1$, if i^{th} unit from the population is included in the sample.

= 0, if i^{th} unit from the population is not included in the

Sample.

Then $E(a_i) = 1 * P(i^{th}$ unit is selected in the sample$) + 0 * P(i^{th}$ unit is not included in the sample$)$

$= 1 * n/N + 0 * (1 – n/N)$

$= n/N$

Let $E(\bar{Y}) = 1/n \, \Sigma E(a_i) Y_i \quad i=1, 2, \ldots N$

$=1/n \, \Sigma n/N \, Y_i \quad i=1, 2, \ldots N$

$=1/N \, \Sigma Y_i \quad i=1, 2, \ldots N$

$=\bar{Y}$

Thus sample mean is an unbiased estimate of population mean. **Proved**.

Theorem 2: In SRSWOR, sample mean square is an unbiased estimate of the population mean square.

Proof: Let $Y_1, Y_2, \ldots, Y_N$ be the population values and let $y_1, y_2, \ldots, y_n$ be the sample observations when a sample of size n is selected from population of size N by the method of SRSWOR. Then,

Population mean = $\bar{Y} = 1/N \, \Sigma Y_i$ $\quad$ i=1, 2,…..N

Population mean square = $S^2 = i/N\text{-}1 \, \Sigma (Y_i - \bar{Y})^2$, $\quad$ i=1, 2,…….,N

Sample mean = $\bar{y} = 1/n \, \Sigma y_i$, $\quad$ i=1, 2, ….., n

Now, $\bar{y} = 1/n \, \Sigma y_i = 1/n \, \Sigma a_i Y_i$

Where, $a_i = 1$, if i^{th} unit from the population is included in the sample.

$= 0$, if i^{th} unit from the population is not included in the sample.

Then, $E(a_i) = 1 * P$ (i^{th} unit is selected in the sample) + $0 * P$ (i^{th} unit is not included in the sample)

$= 1 * n/N + 0 * (1 - n/N)$

$= n/N$

$E(a_i^2) = 1 * P$ (i^{th} unit is selected in the sample) + $0 * P$ (i^{th} unit is not included in the sample)

$=1 * n/N + 0 * (1 - n/N)$

$=n/N$

$E(a_i a_j) = 1 * P(a_i a_j = 1) + 0 * P(a_i a_j = 0)$

$= 1 * P(a_i = 1, a_j = 1)$

$= 1 * P(a_i = 1) \, P(a_j = 1 / a_i = 1)$

$= 1 * n/N * (n-1)/(N-1)$

$= n(n-1)/N(N-1)$

We have to show that

$E(s2) = S2$

Let $s2 = 1/n\text{-}1 \, \Sigma (yi - \bar{y})2$

$= 1/n\text{-}1 \, \Sigma (y_i^2 - 2y_i\bar{y} + \bar{y}^2)$

$= 1/n\text{-}1 \, (\Sigma y_i^2 - 2\bar{y}\Sigma y_i + n\bar{y}^2)$

$= 1/n\text{-}1 \, (\Sigma y_i^2 - 2\bar{y}n\bar{y} + n\bar{y}^2)$

$= 1/n-1 \, (\Sigma y_i^2 - 2n\bar{y}^2 + n\bar{y}^2)$

$= 1/n-1 \, (\Sigma yi^2 - n\bar{y}^2)$……………………………..*

$= 1/n-1 \, [\Sigma y_i^2 - n * 1/n^2 (\Sigma y_i)^2]$ $i=1, 2, \ldots\ldots n$

$= 1/n-1 \, [\Sigma y_i^2 - 1/n (\Sigma y_i)^2]$ $i=1, 2, \ldots\ldots n$

$= 1/n-1 \, [\Sigma y_i^2 - 1/n (\Sigma y_i^2 + \Sigma y_i y_j)]$ $i \neq j=1, 2, \ldots, n$

$= 1/n-1 \, [\Sigma y_i^2 - 1/n \, \Sigma y_i^2 - 1/n \, \Sigma y_i y_j]$ $i \neq j=1, 2, \ldots n$

$= 1/n-1 \, (1 - 1/n) \, \Sigma y_i^2 - 1/n(n-1) \, \Sigma y_i y_j$ $i \neq j=1, 2, \ldots n$

$= 1/n-1 \, (n-1/n) \, \Sigma y_i^2 - 1/n(n-1) \, \Sigma y_i y_j$

$= 1/n \, \Sigma y_i^2 - 1/n(n-1) \, \Sigma y_i y_j$

$E(s^2) = 1/n \, \Sigma \, E(y_i^2) - 1/n(n-1) \, \Sigma \, E(y_i y_j)$………………(i)

Let,

$\Sigma y_i^2 = \Sigma a_i^2 Y_i^2$ $i=1, 2, 3, \ldots\ldots N$

$\Sigma \, E (y_i^2) = \Sigma \, E (a_i^2) Y_i^2$

$= \Sigma \, n/N \, (Y_i^2)$

$= n/N \, \Sigma \, Y_i^2$…………………………………(ii)

And $\Sigma y_i y_j = \Sigma a_i a_j Y_i Y_j$ $i \neq j=1, 2, \ldots\ldots N$

$\Sigma \, E(y_i y_j) = \Sigma \, E(a_i a_j) Y_i Y_j$

$= \Sigma \, n(n-1)/N(N-1) \, Y_i Y_j$

$= n(n-1)/N(N-1) \, \Sigma \, Y_i Y_j$…………………(iii)

Using (ii) & (iii) in (i) we get,

$E (s^2) = 1/n * n/N \, \Sigma Y_i^2 - 1/n(n-1) * n(n-1)/N(N-1) \, \Sigma Y_i Y_j$ $i \neq j=1, 2, \ldots N$

$= 1/N \, \Sigma Y_i^2 - 1/N(N-1) \, \Sigma Y_i Y_j$

$= 1/N \, \Sigma Y_i^2 - 1/N(N-1) \, [(\Sigma Y_i)^2 - \Sigma Y_i^2]$

$= 1/N \, \Sigma Y_i^2 - 1/N(N-1) * (\Sigma Y_i)^2 + 1/N(N-1) * \Sigma Y_i^2$

$= 1/N \, \Sigma Y_i^2 - 1/N(N-1) * (N\bar{Y})^2 + 1/N(N-1) * \Sigma Y_i^2$

$= 1/N \, \Sigma Y_i^2 \, (1 - 1/N-1) - (N^2 \bar{Y}^2)/N(N-1)$

$= [\Sigma Y_i^2/N \, (N-1+1)/N-1] - N\bar{Y}^2/N-1$

$= \Sigma Y_i^2/N * N/N-1 - N\bar{Y}^2/N-1$

$= 1/N-1 \, (\Sigma Y_i^2 - N\bar{Y}^2)$

$= S^2$

$E(s^2) = S^2$ ……… **proved.**

Theorem 3: In SRSWOR, the variance of sample mean is given by

$V(\bar{y}) = (1/n – 1/N) S^2 = (N-n/Nn) S^2$

Proof: We have,

$V(\bar{y}) = E [\bar{y} - E(\bar{y})]^2$

$= E [\bar{y}^2 - 2\bar{y}E(\bar{y}) + \{E(\bar{y})^2 \}]$

$=E[\bar{y}^2 - 2\bar{y}\bar{Y} + \bar{Y}^2]$

$=E (\bar{y}^2) - 2\bar{Y}E(\bar{y}) + \bar{Y}^2$

$=E (\bar{y}^2) - 2\bar{Y}*\bar{Y} + \bar{Y}^2$

$=E (\bar{y}^2) - \bar{Y}^2$

$E(\bar{y}^2) = E (1/n \Sigma y_i)^2$

$= 1/n^2 E(\Sigma y_i)^2$

$= 1/n^2 E(\Sigma y_i^2 + \Sigma y_i y_j)$ $i \ne j=1, 2,........n$

$= 1/n^2 E(\Sigma y_i)^2 + 1/n^2 E(\Sigma y_i y_j)$

$E(\bar{y}^2) = 1/n^2 E(\Sigma y_i)^2 + 1/n^2 E(\Sigma y_i yj)$……………….(*)

Consider ,

$E(\Sigma y_i)^2 = n/N \Sigma Y_i^2$

$= n/N[(Y_i - \bar{Y})^2 + N\bar{Y}^2]$ [Since $\Sigma y_i^2 = \Sigma a_i Y_i^2$, so $\Sigma E(y_i^2) = \Sigma(a_i^2)Y_i^2$

$=\Sigma n/N*Y_i^2 =n/N \Sigma Y_i^2$]

$= n/N [(N-1)S^2 + N\bar{Y}^2]$ [Since $S^2=1/N-1*\Sigma(Y_i-\bar{Y})^2$ or, $(N-1)S^2=\Sigma(Y_i-\bar{Y})^2$]

$E (\Sigma Y_i Y_j) = n(n-1)/N(N-1) * \Sigma Y_i Y_j$ $i \ne j=1, 2,…….,N$

$= n(n-1)/N(N-1) [(\Sigma Y_i)^2 – \Sigma Y_i^2]$

$= n(n=1)/N(N-1) [(N\bar{Y})^2 – \{(N-1)S^2 + N\bar{Y}^2\}]$

$= n(n-1)/N(N-1) [N^2\bar{Y}^2 – NS^2 + S^2 - N\bar{Y}^2]$

$= n(n-1)/N(N-1) [\bar{Y}^2 N(N-1) – (N-1)S^2]$

$= n(n-1)/N(N-1) * (N-1) [N\bar{Y}^2 – S^2]$

$= n(n-1)/N * (N\bar{Y}^2 – NS^2/N)$

$= n(n-1)/N * N(\bar{Y}^2 – S^2/N)$

$= n(n-1) (\bar{Y}^2 – S^2/N)$

Hence, $E(\bar{y}^2) = 1/n^2\, E(\Sigma y_i)^2 + 1/n^2\, E(\Sigma y_i y_j)$ $\quad i \neq j=1,2,\ldots\ldots n$

$= 1/n^2\, [n/N\{(N-1)S^2 + N\bar{Y}^2\}] + 1/n^2[n(n-1)(\bar{Y}^2 - S^2/N)$

$= N\bar{Y}^2/nN + (n-1)/n*\bar{Y}^2 + (N-1)/Nn*S^2 - (n-1)/Nn*S^2$

$= \bar{Y}^2(1/n+1-1/n) + (N-1-n+1)S^2/Nn$

$= \bar{Y}^2 + (N-n/Nn)S^2$

$= \bar{Y}^2 + (1/n - 1/N)S^2$

Hence, $V(\bar{y}) = E[\bar{y} - E(\bar{y})]^2$

$= E[\bar{y}^2 - 2\bar{y}E(\bar{y}) + \{E(\bar{y})\}^2]$

$= E[\bar{y}^2 - 2\bar{y}\bar{Y} + \bar{Y}^2]$

$= E(\bar{y}^2) - 2\bar{Y}E(\bar{y}) + \bar{Y}^2$

$= \bar{Y}^2 + (1/n - 1/N)S^2 - \bar{Y}^2$

$= (1/n - 1/N)S^2$

$= (N-n/Nn)S^2$

Note: $f = n/N$ = sampling fraction

Hence, $V(\bar{y})_{WOR} = (1/n-1/N)S^2$

$= (1-n/N)S^2/n$

$= (1-f)S^2/n$

$(1 - f)$ is known as finite population correction. When N is large, f is small. So

$(1 - f) \rightarrow 1$.

Then $V(\bar{y}) = S^2/n$

Estimation of variance:

$V(\bar{y})_{WOR} = (1/n - 1/N)S^2$

$E(s^2) = S^2$

$\hat{S}^2 = s^2$

Hence, $\hat{V}(\bar{y})_{WOR} = (1/n - 1/N)\hat{S}^2$

$= (1/n - 1/N)s^2$

Estimation of variance:

$$V(\bar{y})_{WOR} = (1/n - 1/N)S^2$$

$$E(s^2) = S^2$$

$$\hat{S}^2 = s^2$$

Hence, $\hat{V}(\bar{y})_{WOR} = (1/n - 1/N)\hat{S}^2$

$$= (1/n - 1/N)s^2$$

Confidence interval for population mean:

When a sample of size n is selected from a population of size N by the method of SRSWOR, we know the results:

$E(\bar{y}) = \bar{Y}$

$V(\bar{y}) = (N\text{-}n/Nn)S^2$

Assuming the Normal Distribution for the population, sample mean $\bar{y}$ follows $N\{\bar{Y}, (N\text{-}n/Nn)S^2\}$.

From Standard Normal tables for given value of α we can read the ordinate $Z_{\alpha/2}$ such that

$P(\{|\bar{y}-\bar{Y}|/\sqrt{V(\bar{y})}\} < Z_{\alpha/2}) = 1-\alpha$

$P(-Z_{\alpha/2} < (\bar{y} - \bar{Y})/\sqrt{V(\bar{y})} < Z_{\alpha/2}) = 1-\alpha$

$P(\bar{y} - Z_{\alpha/2}\sqrt{V(\bar{y})} < \bar{Y} < \bar{y} + Z_{\alpha/2}\sqrt{V(\bar{y})}) = 1 - \alpha$

Also, we have estimate of $V(\bar{y}) = \hat{v}(\bar{y}) = (N\text{-}n/Nn)s2$

Hence, 100 (1-α)% confidence interval for population mean is

$\{ \bar{y} - \sqrt{V(\hat{\bar{y}})}\, Z_{\alpha/2} , \bar{y} + \sqrt{V(\hat{\bar{y}})}\, Z_{\alpha/2} \}$

$= \{ \bar{y} - \sqrt{(N\text{-}n/Nn)} * s * Z_{\alpha/2}, \bar{y} + \sqrt{(N\text{-}n/Nn)} * s * Z_{\alpha/2} \}$

Estimation of Population Total and its Variance:

We know that sample mean is an unbiased estimate of population mean.

- ⇨ $E(\bar{y}) = \bar{Y}$
- ⇨ $\hat{\bar{Y}} = \bar{y}$

And population total = $Y = N\bar{Y}$

⇨ $\hat{Y} = N\hat{\bar{Y}} = N\bar{y}$

Hence, estimation of population total is $N\bar{y}$.

Also, we have,

$V(\bar{y})_{WOR} = (1/n - 1/N)\, S^2$

⇨ $V(N\bar{y})_{WOR} = N^2\, V(\bar{y})$
$= N^2\,(1/n - 1/N)\, S^2$
$= N^2\,(N - n / Nn)\, S^2$
$= N\,(N - n)/n\, S^2$

⇨ $V(\hat{Y})_{WOR} = V(N\bar{y}) = \{N(N-1)/n\}S^2$

And an estimate of variance is

$V(\hat{Y})_{WOR} = \{N(N - 1)/n\}\,\hat{S}^2 = \{N(N - 1)/n\}\, s^2$

Confidence Interval for population total:

We know,

$E(N\bar{y}) = Y$ = Population Total

$V(N\bar{y}) = N^2\,(N - n/Nn)\, S^2$

$= N\,(N - n/n)\, S^2$

Estimate of $V(N\bar{y}) = v(N\bar{y}) = \{N(N-n)/n\}\, s^2$

Hence, 100(1-α)% confidence interval for population total is

$\{ N\bar{y} - \surd V(N\hat{\bar{y}})\, Z_{\alpha/2}\, ,\ N\bar{y} + \surd V(N\hat{\bar{y}})\, Z_{\alpha/2} \}$

$= \{ N\bar{y} - \surd N(N - n)/n * s * Z_{\alpha/2}\, ,\ N\bar{y} + \surd N(N - n)/n * s * Z_{\alpha/2} \}$

SIMPLE RANDOM SAMPLING WITH REPLACEMENT (SRSWR):

Let $Y_1, Y_2, \ldots, Y_N$ be the population values and $y_1, y_2, \ldots, y_n$ be the sample observations when a sample of size n is selected from a population of size N by the method of Simple Random Sampling With Replacement (SRSWR). Since sample units are drawn independently we can treat y_i's as independently identically distributed random variables with mean $\bar{Y}$ and variance σ^2.

Population mean = $\bar{Y} = 1/N\ \Sigma\ Y_i$, $i=1, 2, \ldots, N$

Population variance = $\sigma^2 = 1/N\ \Sigma\ (Y_i - \bar{Y})^2$, $i=1, 2, \ldots, N$

Population mean square = $S^2 = 1/N-1\ \Sigma\ (Y_i - \bar{Y})^2$, $i=1, 2, \ldots, N$

Sample mean = $\bar{y} = 1/n\ \Sigma\ y_i$, $i=1, 2, \ldots, n$

Sample mean square = $s^2 = 1/n-1\ \Sigma\ (y_i - \bar{y})^2$, $i=1, 2, \ldots, n$

Estimation of population Mean and its Variance

Theorem 4:

In Simple Random Sampling With Replacement,

(i) Sample mean is an unbiased estimate of the population mean.
(ii) Sample mean square is an unbiased estimate of the population variance σ^2.

Proof: Let $Y_1, Y_2,\ldots, Y_N$ be the population values and $y_1, y_2,\ldots, y_n$ be the sample observations when a sample of size n is selected from a population of size N by the method of SRSWR. Then,

Population mean = $\bar{Y} = 1/N\ \Sigma\ Y_i \quad i=1, 2,\ldots,N$

Sample mean = $\bar{y} = 1/n\ \Sigma\ y_i \quad i=1, 2,\ldots,n$

Let a random variable a_i denote the number of times the i^{th} unit from the population appears in the sample, i=1, 2,…..,N. So a_i takes values 0,1,2,….,n.

The joint distribution of $a_1, a_2,\ldots,a_N$ is the multinomial distribution given by

$P(a_i a_j \ldots a_N) = \{ n!/(a_1!\ a_2! \ldots a_N!)\} * 1/N^n$ where $\Sigma\ a_i = N$, i=1,2,…..N

We know for multinomial distribution,

$E(a_i) = n/N$, $V(a_i) = n(N-1)/N$, for i=1, 2,….,N

$Cov(a_i, a_j) = -n/N^2$, for $i \neq j = 1, 2,\ldots,N$

(i) We can write sample mean as

$\bar{y} = 1/n\ \Sigma\ a_i Y_i$ for i=1, 2,…..,N

Consider $E(\bar{y}) = E(1/n\ \Sigma\ a_i Y_i)$

$= 1/n\ \Sigma\ E(a_i) Y_i$

$= 1/n\ \Sigma\ (n/N) Y_i$

$= 1/N\ \Sigma\ Y_i$

$= \bar{Y}$

(ii) Consider, $s^2 = (1/n-1)\ \Sigma\ (y_i - \bar{y})^2$

$= (1/n-1)(\Sigma\ y_i^2 - n\bar{y}^2)$

$= (1/n-1)[\Sigma\ y_i^2 - n(1/n\ \Sigma\ y_i)^2]$

$= (1/n-1)[\Sigma\ y_i^2 - 1/n(\Sigma\ y_i)^2]$

$= (1/n\text{-}1) [\Sigma y_i^2 - 1/n (\Sigma y_i^2 + \Sigma y_i y_j)]$

$= (1/n\text{-}1) \Sigma y_i^2 - \{ 1/n (n\text{-}1) \} \Sigma y_i^2 \}$

$- \{1/n(n\text{-}1)\} \Sigma y_i y_j$ for $i \neq j=1, 2,\ldots..n$

$= \Sigma y_i^2 \{1/n\text{-}1 - 1/n(n\text{-}1)\} - \{1/n(n\text{-}1)\}\Sigma y_i y_j$

$= \Sigma y_i^2 \{n\text{-}1/n(n\text{-}1)\} - \{1/n(n\text{-}1)\} \Sigma y_i y_j$

$= 1/n \Sigma y_i^2 - \{1/n(n\text{-}1)\} \Sigma y_i y_j$

⇨ $E (s^2) = 1/n \Sigma E(y_i^2) - \{1/n(n\text{-}1)\} \Sigma E (y_i y_j)$ ……….(i)

Consider ,

$E (\Sigma y_i^2) = \Sigma E (a_i^2) Y_i^2$, for i=1, 2,……,n

$= \Sigma n/N Y_i^2$

$= n/N \Sigma Y_i^2$…………………(ii)

And,

$E (\Sigma y_i y_j) = \Sigma E (a_i a_j) Y_i Y_j$, for $i \neq j=1, 2,\ldots,N$

$= \Sigma [Cov (a_i,a_j) + E (a_i) E (a_j)] Y_i Y_j$

$= \Sigma (-n/N^2 + n*n/N*N) Y_i Y_j$

$= n(n\text{-}1)/N^2 \Sigma Y_i Y_j$…………………(iii)

Putting (ii) and (iii) in (i) we get,

$E(s^2) = 1/n*n/N \Sigma Y_i^2 - 1/n(n\text{-}1) [n(n\text{-}1)/N^2 \Sigma Y_i Y_j]$

$= 1/N\Sigma Y_i^2 - 1/N^2 \Sigma Y_i Y_j$

$= 1/N \Sigma Y_i^2 - 1/N^2 [(\Sigma Y_i)^2 - \Sigma Y_i^2]$

$= 1/N \Sigma Y_i^2 - 1/N^2 [(N\bar{Y})^2 - \Sigma Y_i^2]$

$= 1/N \Sigma Y_i^2 - \bar{Y}^2 + 1/N^2 \Sigma Y_i^2$

$= 1/N \Sigma Y_i^2 - \bar{Y}^2$ [As N^2 is very large, $1/N^2 \rightarrow 0$]

Hence, $E (s^2) = 1/N \Sigma Y_i^2 - (\bar{Y})^2$

$= 1/N (\Sigma Y_i^2 - N\bar{Y}^2)$

$= 1/N \Sigma (Y_i - \bar{Y})^2$

$= \sigma^2$

Hence proved.

Theorem 5: In Simple Random Sampling With Replacement (SRSWR), the variance of the sample mean is

$$V(\bar{y})_{WR} = (N - 1/nN)S^2 = \sigma^2/n$$

Proof: Let $Y_1, Y_2, \ldots, Y_N$ be the population values and $y_1, y_2, \ldots, y_n$ be the sample observations when a sample of size n is selected from a population of size N by the method of SRSWR. Then,

Population mean = $\bar{Y} = 1/N \, \Sigma Y_i$, for i = 1, 2,……,N

Sample mean = $\bar{y} = 1/n \, \Sigma y_i$, for i = 1, 2,…….,n

Let a random variable a_i denote the number of times the i^{th} unit from the population appears in the sample, i = 1, 2,…….N. So a_i takes the values 0, 1, 2,…….,n.

The joint distribution of $a_1, a_2, \ldots, a_N$ is the multinomial distribution given by

$$P(a_i a_j \ldots a_N) = (n!/a_1!a_2!\ldots a_N!) * 1/N^n \quad \text{where } \Sigma a_i = N,\ i=1,2,\ldots N$$

We know for multinomial distribution,

$$E(a_i) = n/N, \quad V(a_i) = \{n(N-1)\}/N^2 \quad \text{for } i = 1, 2, \ldots N$$

$$Cov(a_i, a_j) = -n/N^2, \quad \text{for } i \neq j = 1, 2, \ldots, N$$

We can write sample mean as

$$\bar{y} = (1/n)\, \Sigma a_i Y_i$$

$$E(\bar{y}) = E[(1/n)\, \Sigma a_i Y_i]$$

$$= (1/n)\, \Sigma E(ai) Y_i$$

$$= (1/n) * (n/N)\, \Sigma Y_i$$

$$= (1/N)\, \Sigma Y_i$$

$$= \bar{Y}$$

$$V(\bar{y}) = (1/n^2) V(\Sigma a_i Y_i)$$

$$= (1/n^2)[\Sigma V(a_i) Y_i^2 + \Sigma Cov(a_i a_j) Y_i Y_j], \quad \text{for } i \neq j = 1,2,\ldots N$$

$$= (1/n^2)[\Sigma \{n(N-1)/N^2\} Y_i^2 + \Sigma (-n/N^2) Y_i Y_j]$$

$$= \{(N-1)/nN^2\}\, \Sigma Y_i^2 - (1/nN^2)\, \Sigma Y_i Y_j$$

$$= (1/nN^2)[(N-1)\, \Sigma Y_i^2 - \Sigma Y_i Y_j]$$

$$= (1/nN^2)[(N-1)\, \Sigma Y_i^2 - \{(\Sigma Y_i)^2 - \Sigma Y_i^2\}]$$

$$= (1/nN^2)[(N-1)\, \Sigma Y_i^2 - (\Sigma Y_i)^2 + \Sigma Y_i^2]$$

$= (1/nN^2) [\Sigma Y_i^2 (N - 1 + 1) - (\Sigma Y_i)^2]$

$= (1/nN^2) [N \Sigma Y_i^2 - (\Sigma Y_i)^2]$

$= (1/nN^2) [N \Sigma Y_i^2 - (N\bar{Y})^2]$

$= (1/nN^2) * N \Sigma Y_i^2 - (1/nN^2) * (N^2\bar{Y}^2)$

$= (1/nN) [\Sigma Y_i^2 - N\bar{Y}^2]$

$= (1/nN) (N-1) S^2$

$= (1/nN) n\sigma^2$

$= \sigma^2/n$

Hence proved.

Confidence Interval (C I) for population Mean

When a sample of size n is selected from a population of size N by the method of SRSWR , we know that

$E(\bar{y}) = \bar{Y}$, $V(\bar{y}) = \sigma^2/n = (N\text{-}1/Nn)S^2$

Estimate of $\hat{V}(\bar{y}) = v(\bar{y}) = (N\text{-}1/Nn)S^2$

Assuming the Normal Distribution for the population , sample mean $\bar{y}$ follows $N(\bar{Y}, \sigma^2/n)$. From Standard Normal tables for given value α we can write the ordinate $Z_{\alpha/2}$ as

$P(|\bar{y} - \bar{Y}|/\surd V(\bar{y}) < Z_{\alpha/2}) = 1 - \alpha$

- ⇨ $P[-Z_{\alpha/2} < (\bar{y} - \bar{Y})/\surd V(\bar{y}) < Z_{\alpha/2}] = 1 - \alpha$
- ⇨ $P[\bar{y} - Z_{\alpha/2} \surd V(\bar{y}) < \bar{Y} < \bar{y} + Z_{\alpha/2} \surd V(\bar{y})] = 1 - \alpha$
- ⇨ $P[\bar{y} - \surd \hat{V}(\bar{y}) Z_{\alpha/2} < \bar{Y} < \bar{y} + \surd \hat{V}(\bar{y}) Z_{\alpha/2}] = 1 - \alpha$

Hence, 100(1 – α)% C.I for population mean is

$\{ \bar{y} - Z_{\alpha/2}\surd \hat{V}(\bar{y}) , \bar{y} + Z_{\alpha/2} \surd \hat{V}(\bar{y}) \}$

Estimation of population total and its variance:

We know that sample mean is an unbiased estimate of the population mean.

- ⇨ $E(\bar{y}) = \bar{Y}$
- ⇨ $\hat{\bar{Y}} = \bar{y}$

And population total is

$\Sigma Y_i = N\bar{Y}$, for i = 1, 2,…….,N

- ⇨ $\hat{Y} = N\hat{\bar{Y}} = N\bar{y}$

So estimate of population total is $N\bar{y}$.

Now, we will find the variance of the estimate of population total.

We know,

$V(\bar{y})_{WR} = (N - 1/Nn) S^2$

- ⇨ $V(N\bar{y})_{WR} = N^2 V(\bar{y})$
- ⇨ $V(N\bar{y})_{WR} = N^2 (\sigma^2/n)$
- ⇨ $V(N\bar{y})_{WR} = [N(N\text{-}1)/Nn] S^2$

Thus, the variance of the estimate of the population total is

$V(\hat{\bar{Y}})_{WR} = V(N\bar{y})_{WR} = N^2 (\sigma^2/n) = \{N(N\text{-}1)/Nn\}S^2$

And an estimate of variance of the estimate of the population total is

$V(\hat{Y})_{WR} = (N^2/n)\,\hat{\sigma}^2 = (N^2/n)s^2$

Confidence Interval (C.I) for population total:

100(1 – α)% C.I for population total is

$\{ N\bar{y} - Z_{\alpha/2} \sqrt{\hat{V}(N\bar{y})} , N\bar{y} + Z_{\alpha/2} \sqrt{\hat{V}(N\bar{y})} \}$

$= \{ N\bar{y} - Z_{\alpha/2} Ns/\sqrt{n} , N\bar{y} + Z_{\alpha/2} Ns/\sqrt{n} \}$

Comparison between SRSWOR and SRSWR:

We know :

$V(\bar{y})_{WR} = \{(N-1)/Nn\}S^2$ and

$V(\bar{y})_{WOR} = \{(N-n)/Nn\}S^2$

$n > 1$

$N – 1 > N – n$

- ⇨ $\{ (N – 1)/Nn \}S^2 > \{ (N – n)/Nn \}S^2$
- ⇨ $V(\bar{y})_{WR} > V(\bar{y})_{WOR}$

Thus, we see that $V(\bar{y})_{WR}$ is more than $V(\bar{y})_{WOR}$. So the estimate obtained by SRSWOR is more efficient than the estimate obtained by SRSWR.