CHAPTER-2

POPULATION AND SAMPLE

2.1 Introduction:-

While studying any phenomenon we come across types of characteristics constant and variable

(i) constant:-The characteristic which does not change its value (or nature) is considered as **constant.** For example: Height of a person after 25 years of age altitude of a certain place from sea level, etc.

(ii) **variable:-** The characteristics which are qualitative or quantitative in nature and change their values (or nature) is called variable. For example: Examination result of a candidate can be recorded as pass or fail which is a qualitative variable characteristics, where as we can express a candidate's performance as percentage of marks which is a quantitative variable.

2.1 Types of Characteristic: Attributes and variables

2.1.1 Attribute:- A qualitative characteristic is called an attribute. For example: sex, nationality, religion, grade in examination , blood group, beauty, defectiveness of an article produced by a machine.

Steven S.S. introduced four types of scales of measurements viz. the nominal, ordinal, interval and ratio scales. Attributes are measured using nominal and ordinal scale.

2.1.2 Nominal Scale:- Nominal scale consists of two or more named categories into which the objects are classified. For example: Classification of students in various divisions of the same standard, Classification of individuals using Gender, caste, nationality, Classification of house numbers, survey numbers pin code numbers etc.

Remarks : (i) In nominal scale if numbers are used, then those are allotted in purely arbitrary manner. Those numbers are just for identification purpose used in places of labels. (ii) Those numbers are interchangeable.

2.1.3 Ordinal Scale: Ordinal scale of measurement gives numbers to groups of objects using some quantifiable characteristics, therefore ordered arrangement of groups are possible. For example:

(a) Groups of individuals according to income such as poor, middle class, rich (b) Groups of students according to grades in examination , such as fail , second class, first class, first class with distinction

(c) Groups of individuals as dull or intelligent; soft or hard etc.

Remark : (i) In the ordinal scale, numbers given to groups as lables , serve the purpose of ranks. Hence those lables are not interchangeable. In the ordinal scale, the groups are ordered according to some characteristic, suppose individuals A, B,C are given ranks 1,2,3 respectively according to their height. A is the shortest and C is the tallest. In this case heights of A, B, C may not be equispaced, however, they possess equispaced ranks.

(ii) Rank of individual B is 2, however, height of B is not double the height of A or height of C is not three times the height of A. Here we note that the heights of B, C are not exact integer multiples of height of A, however, height of B ,C are not exact integer multiples of height of A, however, they possess ranks 1,2,3 which are integer multiples of rank of A. Here we only men that B is taller than A and C is taller than B.

2.1.4 Variable:- A quantitative characteristic which changes its value is clled a variable. For example: weight of person , examination marks, population of a country, profit of a salesman.

Note :- Variable are measured using interval scale and ratio scale.

The drawback of ordinal scale that the units are not equispaced is overcome in interval scale.

2.1.5 Interval scale: Interval scale of measurement has equal units of measurement, however, the Zero point is arbitrary.

The classic example of interval scale in our day-to day life is the Centigrade or Fahrenheit scale of temperature measurement. In both the

scales zero is arbitrary, it does not mean absence of heat. Moreover 60^0 C does not contain exactly double the heat that 30^0C has. However , the difference in temperature between 10^0C to 20^0C is same as that between 50^0C to 60^0C (or that between similar pair)

Drawback : In interval scale zero is arbit rary, it is chosen as per convenience, therefore we can add (or subtract) a constant in the readings on interval sale without affecting form, of scale. However , we cannot multiply or divide the readings by constant.

Use: Inspite of the drawback in interval scale, it is used for convenience in behavioural sciences to study mental and social variables and traits.

All the drawbacks existing in the earlier three scales of measurement viz. nominal ordinal, interval scales are overcome in ratio scale. It is the best scale of measurement. It is used almost all places.

2.1.6 Ratio scale: Ratio scale of measurement has equal units of measurement and those are taken from a true zero.

All the measurements of type height (cm), time (hours) etc. are the examples of ratio scales. In this scale 60 kg weight is exactly double heavy as compared to 30 kg weight.

It can be clearly noticed that variables can be measured by numbers.

Further the variables can be divided into two categories:

(i) Discrete variable:-

Definition : A variable taking only particular values is called as discrete variable.

For example : Number of students in a class, number of articles produced by a machine, population of a country, number of workers in a factory etc. are discrete variables. Most of the discrete variables have integral values.

(ii) Continuous variable:-

Definition: A variable taking all possible values in a certain range is called as continuous variable.

For example : Weight of a person , length of a screw produced by a machine, temperature at a certain place, agricultural production , electricity consumption of a family, speed of a vehicle are the examples of continuous variable.

It is observed that many continuous variables such as marks , income, weight of a person etc. look like discrete variables after the measurement. This is mainly due to the limitations of the measuring instruments. Using better instruments one can have accurate measurement and overcome this difficulty.

2.1.7 Directional Data or Circular Scale:-

Some variables are cyclic or rhythmic in nature according to time. For example, blood pressure, reproductive cycles, body temperature, mental alertness, sleep-wake cycles, hormonal plasticity. Such variable are called biological rhythms control characteristics. The corresponding data is considered directional data or circular data. Moreover direction of wind, direction of earth magnetic pole, direction of birds movement, direction of river of river flow are the examples of directional data.

Definition:-The variables measured in angles i.e. on circular scale rather that linear scale are called as **directional data or circular data.**

The statistical tools such mean, variance used in usual manner do not remain meaningful and suitable. The statistical methods to be used in these instances are entirely different.

Directional data is also observed in circular movements of automobiles parts, oceanography, travel of ships etc..

2.2 Types of Data

Data Set: A set of measurement obtained on some variable is called a data set.

For example observation of the patient's heart rate, the data set might be 60, 70, 64, 55, 70, 74, 51, 80.

Collection of data is a very important work and needs to be done carefully. One has to decide the objectives clearly before collecting the data. In order to determine dependable and reliable results, proper data should be collected in a proper way. The data according to the method of collection are of two types viz., (a) Primary data (b) Secondary data.

2.2.1 (a) Primary Data:

Primary data means original data (i.e. facts and figures) obtained by an investigator himself. Primary data may be a result of a survey or enquiry conducted. This may be regarded as first-hand information. Population census results, is a classical example of primary data. Primary data are also called as raw data. No doubt, primary data are more reliable than any other type put are expensive and time consuming. Primary data are collected by the following methods.

(i) Direct personal investigation or interview:

In this method, the investigator meets concerned persons known as 'informants' and collects necessary information by the process of interview. Investigator should be thorough in handling problems of investigation. This will result into reliable data. Investigator has to go upto the source of original information. For example, if he wants to know the amount of production , in a particular industry, he should collect the figures by visiting the machine floor, rather than from office or bulletin. This is the best method of collecting primary data. However, the investigator has to take certain precautions.

(ii) Indirect oral investigation.

(iii) Investigation through questionnaire.

2.2.1 (b) Secondary Data:

Data taken from sources like office records, bulletins, reports etc. which are already collected by some other agency is called 'secondary data'.

The data which are already collected may be tabulated, classified, ordered etc. Hence, it is called processed or finished data. Thus, secondary data can also be called finished product.

'Secondary data' is a relative term. For example, if 'A' collects original data, then it becomes primary data for him, where as if the same data is used by B, then its becomes secondary data for B. In this case, the only difference is that the user of secondary data may not have thorough understanding of the background as the user of primary data has.

2.2.2 Difference between primary and secondary data:

[i] The main difference lies in the method of collection.

[ii] Primary data are original in nature. Hence those are more accurate than secondary data.

[iii] Collection of primary data is expensive as well as time- consuming.

[iv] Primary data can be elicited in accordance with the objectives of a study. Secondary data may fail in this regard.

The investigator has to decide whether primary data are to be collected or secondary data to be collected by considering advantages and disadvantages of each. Moreover he has to consider nature, scope of enquiry, funds and time available , accuracy etc. to decide which type of data are to be collected.

2.2.3 The following are the main sources of secondary data:

[i] Publications of the state and central Government and public sector organizations such as, Reserve bank of India Bulletin, periodical reports of Central Statistical Organization (CSO), National Sample Survey (NSS), demographical institutes, research institutes, meteorological institutes etc.

[ii] Publications of international organizations such as UNO,WHO, UNESCO.

[iii] Reports of municipal corporations, Zilla Parishad and similar other organization.

[iv] Reports of different commissions and organizations like chambers of commerce, stock exchanges and UGC etc.

[v] News papers, Journals, Periodicals such as Economic Times, Financial express, India Year Book, Manorama Year Book etc. provide required data.

[vi] Publications of research institutes , universities also provide data to prepare an exhaustive list of all publications and sources is difficult task,. Apart from these sources, there are many other sources such as records of private concerns, unpublished reports prepared by consultants, research workers etc.

After deciding whether to collect primary or secondary data, the investigator has to fix up the method of collecting the data keeping in view the funds , time and other resources at his disposal. There are two ways of collecting data (i) Sampling (ii) Complete enumeration.Thus the methods of data collection are (i) surveys (ii) laboratory experiments, (iii) simulation.

Surveys: with the help of sample surveys or complete enumeration primary or secondary data can be collected.

Laboratory experiments: The observations generated in laboratory experiments will be a method of data collection.

2.2.4 The other Types of Data:

There is yet another of looking at data. Earlier we have considered the way of collection. However, the type of data exists due to the nature of data and some other characteristics. If we consider the data when it was collected. Thus , we introduce the time characteristics. It gives rise to the data specially termed as time series data. Sometimes at a fixed time moment we collect data, where time is considered but hold constant. Such data are referred to as cross-sectional data. The specific definitions are as follows:

2.2.5 Time series data: The data arranged in the chronological order (as per the order of occurrence are called as time series data.

For example:

(i) Daily sales of a departmental store.

(ii) Daily electricity consumption of a town.

(iii) Prime of gold recorded daily.

2.2.6 Cross-sectional data: The values of variables observed at a particular time at several places or on several objects are called as cross-sectional data.

For example:

(i) Sales on a specific day of several departmental store is a cross-sectional data. However, daily sales of a specific departmental store constitute time series data.

(ii) Electricity consumption on a specific day for several towns constitution cross- sectional data. However, daily electricity consumption of a specific town is a time series data.

2.3 Population and Sample:-

Introduction:-

In order to study a group of large number of items we require drawing sample. We use technique of sampling several times in everyday life. For example, while purchasing food grains we inspect only handful of grains and draw conclusion about the whole sack. Similarly while examining blood of an individual few drops are enough for diagnosis. Quality of milk is tested with the help of a small quantity of milk taken out of can, instead of entire milk in the can. Sampling is a well-accepted means of collecting information. Moreover it is believed to be scientific and objective procedure of selecting items. Sampling plays very important role in statistical inference.

2.3.1 Population:

In the technical language of statistics the work population is used in some what a wider sense. It does not mean only a human population. For example,

(i) In the study of industrial development, all the industries under consideration is the population.

(ii) In the study of socio economic conditions of a particular village, all families or houses in the village will be a population.

(iii) In the study of agricultural yield , all the cultivated farms together will be a populations.

(iv) In titration experiment solution in beaker is a population. Thus population may be a group of employees; collection of books, total industrial production, a group of a person's suffering from a particular disease, collection of explosives, group of students etc.

2.3.2 Definition: An aggregate of objects or individuals under study is called population or universe. Population may contain finite or infinite elements.

2.3.3 Limitations of census method:

(1) Census method provides reliable results; but due to voluminous work it is expensive and time consuming. It requires a large amount of manpower.

(2) There are some situations where census is possible but impracticable. For example , testing blood of an individual . In this case, entire blood cannot be tested. Thus census cannot be used here. Similarly, in testing explosives, testing of average life of bulb produced in a lot, testing strength of construction material, census method cannot be used.

(3) If the population is infinite, census cannot be used.

2.3.4 Sampling: It is the process to select the sample is called as sampling method.

Sample:-Any part of population under study is called a sample.

Illustrations: (i) while purchasing food grains, we inspect only a handful of grains and draw conclusions about the quality of the whole lot. In this case, handful of grains is a sample and the whole lot is population.

(ii) While examining blood of an individual, a few drops are taken out of human body for diagnosis. These drops form a sample whereas entire blood

in the body is a population. In this case, conclusions based on sample are accepted for population without any doubt as far as the method is concerned. In this case, census is impracticable.

(iii) For testing quality of milk, a small quantity of milk is tested instead of entire bulk.

(iv) A housewife confirms whether the food is properly cooked or not with the help of few particles taken out of the container. Clearly, the food in the container is a population , whereas food taken out of container for inspection is a sample.

Note:-

(a) Sampling is a well-accepted means of collecting information.

(b) It is believed to be scientific and objective procedure of selecting items. Thus, sampling plays important role in further statistical analysis.

As the sampling methods are used to study population , the samples should be chosen carefully. A natural requirement would be that the sample should be representative of concerned population. There are several methods of sampling in practice. We shall deal with some of these in later sections.

2.3.5 Advantages of Sampling Over Census:-

(i) Reduced Time: As compared to census, sample consists of a less number of elements. Hence , there is a considerable reduction in processing time. The results can be obtained quickly due to time saved in data collection and further analysis.

(ii) Reduced Cost: - There is reduction in cost, both in terms of time and manpower, in sampling, only a part of population is under consideration. Therefore, expenses incurred in collection of data and its analysis is always less than those in census, thus sampling is economical.

(iii) Greater Accuracy:- As compared to census, only limited number of elements are to be processed. Therefore, sophisticated machinery, well trained staff can be used and accuracy can be increased. Due to the reduced

volume of work, it can be completed efficiently and without fatigue. Moreover, elements will be free from non-sampling errors such as incompleteness of returns, biases due to interviews, inaccurate returns etc.

(iv) Greater Scope: (a) If population is infinite or too large or cost per unit is too high, census is impracticable. (b) If the testing is destructive i.e. elements gets destroyed in the analysis , sampling is the only alternative available to us; for example, testing the life of a bulb, testing strength of building construction material, testing fat percentage of milk, testing human blood etc. (c) Suppose a company manufactures a remedial medicine for a certain disease. All the patients suffering from the particular disease man not be ready to try the newly manufactured medicine. In this case sampling has larger scope than census.

2.3.6 Sampling Unit:-

Members or elements of populations are called sampling units. In the sampling process, population is divided into small units which are called the sampling units. For example, in a socio-economic survey, a family is a sampling unit; whereas in a health survey, an individual will be a sampling unit. Sampling units must be distinct and unambiguous in nature. Sampling units together must cover the entire population. In other words, sampling unit is the smallest part of the population which cannot be further subdivided for the said purpose.

2.3.7 Sampling Frame:

It is an exhaustive list of all members or elements of population Sampling frame gives guidelines to cover the entire population. The frame should be up to date and suitable for the purpose of survey or enquiry. In a socio-economic survey frame man be determined from the records at Gram panchayat or ration cards. To prepare a good frame is a difficult job.

Defective frame does affect the result of the survey.Samples can be selected in two ways:

2.3.8 Deliberate Selection of Items or Non-random Sampling: In this method investigator selects elements in any manner which is suitable to him. For example he may select elements on first come first served basis.This method is unscientific. It may produce unreliable results. There is likelihood of a partial view in this method. The figures collected in this way do not obey statistical laws or laws of probability. Hence, such data will not be useful for further analysis and interpretation, to avoid such problems, another method is used.

To select candidates for debating competition, certainly deliberate selection of suitable candidates will be done. It is a purposive sampling. It is non-random in nature. In the advertisement campaign for cosmetics, certainly a sample of youngsters will be selected. It is a non-random sample.

2.3.9 Random Sampling:

In this method the selection of units in the sample is done impartially. Personal or any kind of bias in selection is avoided in random sampling. Further, if each unit has an equal chance of selection the sample is called as simple.

Random Sample:-

(i) Sample is selected with a view to study the concerned population. Therefore, sample should be selected in such a way that it will represent all important characteristics of the population. This may be achieved if elements in sample are selected at random .Thus , sample is a miniature of population .

(ii) Sampling units should be independent.

(iii) Sampling should be evenly spread over the population. It can be achieved by dividing population in homogeneous sub-groups and selecting a random sample from each sub-group.

Limitations of sampling:

(i) Sample will be a perfect representative of population only if the population is homogeneous. For example, in case of testing human blood, a few drops are enough as a sample.

(ii) If sampling units (i.e. elements in sample) are drawn in a scientific method, then the sample is a good representative of population.

(iii) Sampling is reliable if the sample size is adequate.

(iv) Sampling furnishes meaningful results if proper care is taken in the planning and execution of the sample survey. Moreover, properly trained personnel and sophisticated equipments are required to collect the data and analyse it. Otherwise , results may turn out to be inaccurate.

(v) If population is not too large and there are no restrictions over resources in terms of money or time, census would give better results.

2.3.10 Methods of Sampling:-

A success of sampling method mainly depends upon proper selection of sampling method. Different sampling methods are in practice. A sampling method which suits the purpose is selected. Sampling methods are mainly classified into two classes viz. (i) non-random sampling and (ii) random sampling (or probability sampling). In the earlier discussion we have studied the importance of random sampling.

We discuss some popularly used random sampling methods.

(1) Simple random sampling (2) Stratified random sampling (3) Systematic sampling (4) Cluster sampling (5) Two stage sampling.

[1] Simple Random Sampling (SRS):

It is the easiest and the most commonly used method of sampling . In this method each element of population is given same chance of getting selected in the sample. If population consists of N elements than probability of selecting any element at any draw is $\frac{1}{N}$. Further, there are two types of simple random sampling.

[i] Simple Random Sampling with Replacement (SRSWR):

In this method, first element is selected at random from the population. It is recorded or studied completely and then replaced back in the population. After wards second element is selected similarly. This process is continued still a sample of required size is selected similarly. This process is continued till a sample of required size is selected. In this method population size remains the same at every draw. This method of sampling is called as simple random sampling with replacement.

One of the serious drawbacks of this method is that , the same element may be selected more than once in the sample.

[ii] Simple Random Sampling without Replacement (SRSWOR):

In this method first elements is selected at random but those are not replaced back in the population. This method of selecting sample is called as simple random sampling without replacement . In this method population size goes on decreasing at each draw. The drawback of getting the same element selected more than once is overcome in SRSWOR.

Illustrations of Simple Random Sampling:

(i) Suppose a lot of 500 articles is submitted for inspection to determine the proportion of defective articles one can use SRSWOR.

(ii) In order to conduct a socio-economic survey of a certain village we can use SRSWOR and find per capita income of village.

(iii) In order to test average petrol consumption of a lot of scooters manufactured SRSWOR or SRSWR can be used.

(iv) To find diameter of a rod, generally we take reading at few points on a rod and then find the average of readings. These readings form a SRSWOR. This is practiced for physical measurements of articles.

(v) Testing human blood taking few drops out an individual's body is a SRSWOR.

(vi) In order to find average life of a bulb we take SRSWOR from manufactured lot.

Simple random sampling suffers form some drawbacks such as, it may not be proper representative when population is heterogeneous, widely spread etc. Some part of population may not be represented in simple at all. In order to avoid these problems some other sampling methods used.

[2] Stratified Random Sampling:

If population is not homogeneous, SRS is not very effective. Therefore the entire population is divided into several homogeneous groups called as strata (singular stratum). **A simple random sample of a suitable size is selected from each stratum and then combining these sampled observations we can form a sample.** Then this process is called as stratified random sample.

A properly designed stratified random sampling gives better results than simple random sampling. Moreover this method is more suitable from administrative point of view.

Illustrations of Stratified Random Sampling

(i) To estimate annual income per family we divide the population into homogeneous groups such as families with yearly income below Rs.20,000; between Rs. 20,000 – Rs. 50,000 ; between Rs. 50,000 Rs. 1 Lakh and above. Rs. 1 Lakh . Afterwards we used stratified random sampling taking above groups as strata.

(ii) Suppose the proportion of defective articles is to be estimated in a manufacturing process. Then we can use stratified random sampling by taking strata as production in the different shifts.

(iii) In order to estimate crop yield we can divide the field under cultivation in plots, which are equally fertile considered as strata.

(iv) To conduct health survey in a college we can use stratified random sampling by considering strata as the faculties or classes or gender etc.

[3] Systematic Sampling:

In this method population is in the form of serial numbers. Therefore, entire population is divided into some equal parts according to serial numbers. We select a random sample from first group then automatically remaining random samples selected from the remaining groups and therefore, total number of random samples is equal to total number of groups.

For example: [i] Suppose we have total population N**.** To draw a systematic sample, sampling units are numbered 1 to N. In this method we divide population into n equal parts according to serial numbers. Suppose each of the part includes K units (we assume N = nk).

Note that 1^{st} group will contain units bearing serial numbers 1 to k, 2^{nd} group will contain units bearing serial numbers k + 1 to 2 k and so on. Then we select a number at random from 1 to k. Suppose this is j then j^{th} unit in serial order from each group is taken. Thus it will form, a sample of size n which is called as systematic sample. If j^{th} unit is selected then systematic sample will include j^{th}, $(j + k)^{th}$, $(j + 2k)^{th}$ …, $[j+(n-1)k]^{th}$ observations from original list.

[ii] If a sample of size 15 from 150 units is to be drawn, we need to make 15 groups, each of size 10. Thus, here N = 150, K = 10, n = 15. We need to select one unit form first group. Suppose 3^{rd} unit gets selected at random, then other 14 units will be automatically selected. Those will be bearing serial number 13, 23, 33,….., 143.

Entire sample can be selected by selecting every k^{th} after the unit selected from the first group. Thus only one unit drawn at random from first group determines entire sample.

Illustration of Systematic Sampling:

[iii] To select houses for a survey we can use house numbers, in this case systematic sample is prepared.

[iv] Suppose a shopkeeper wants to study customers purchasing habits, he may use bill book. He can choose a systematic sample using the numbers on bill he has.

[v] Farms can be selected by taking systematic sample using survey numbers.

[vi] Suppose a committee of n = 6 students is to be selected from a class of N =60 students then we can make a groups each of k = 10 students using their roll numbers. **We select a student at random from first group.** If 7^{th} students selected , then from each of the next groups we select 7^{th} student. Thus a systematic sample will include students with roll numbers 7, 17, 27, 37, 47, 57.

Population

1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	17	1	1	20
1	2	3	4	5	6		8	9	
2	2	2	2	2	2	27	2	2	30
1	2	3	4	5	6		8	9	
3	3	3	3	3	3	37	3	3	40
1	2	3	4	5	6		8	9	
4	4	4	4	4	4	47	4	4	50
1	2	3	4	5	6		8	9	
5	5	5	5	5	5	57	5	5	60
1	2	3	4	5	6		8	9	

Systematic Sample

Drawbacks:

[i] Systematic sample may not be proper representative if population has hidden periodicities. For example, suppose sales during a year are available. If we take sales of every seventh day in a sample, then sample may contain all Sundays, on which sales might be high.

[ii] If $N \neq nk$, sample ,size does not remain fixed.

[4] Cluster Sampling:

Sometimes population consists of several groups, which are convenient to use as sampling units. These groups are called as clusters.

In cluster sampling, we consider cluster as sampling unit and choose simple random sample of clusters.

In the village list of all houses may be available, however list of individuals in each house may not be available in the village Panchayat office.

For example: In order to conduct a survey of village we can take family as a unit and then we collect information of each member in the family, rather than collecting information of individuals. In this case family plays a role of cluster.

Illustrations of Cluster Sampling:

(i) Suppose a consignment of screws is available in packets. Then we will select few packets in a sample and then inspect each screw from a sampled packet.

(ii) In health survey some villages may be selected from a state and then each individual in the sampled village may be considered.

(iii) To conduct sales campaign, area under consideration may be divided in Zones. A sample of zones may be selected first and then each shop in the selected zone may be contacted.

Note:

(i) Smaller the size of clusters better will be the results.

(ii) If the area under population is subdivided in small area segments as clusters then the same sampling procedure is called as area sampling instead of cluster sampling.

If cluster consists of a widely scattered units, it gives better results, however, cost of surveying increases as compared to the clusters containing neighboring units.

Comparison between Cluster Sampling and Stratified Sampling:

(i) In stratified sampling a sample is taken from each stratum, whereas in cluster sampling entire cluster is included in the sample.

(ii) In stratified sampling each stratum is represented in sample while in cluster sampling it is not so.

(iii) In stratified sampling strata should be homogeneous (within variation small) to produce better results whereas in cluster sampling heterogeneous clusters produce better results.

(iv) In stratified sampling between strata variation should be a large as possible whereas in cluster variation should be as small as possible.

(v) In stratified sampling each stratum should be completely catalogued whereas in cluster sampling only the clusters selected in the sample are required to the catalogued.

[5] Two-Stage Sampling:

Whenever population consists of clusters (or groups), we choose a sample taking **cluster as a sampling unit.** Further, from each of the selected clusters we draw a sample of elements (in the clusters) instead of including

the entire cluster. This procedure of sampling is called as two-stage sampling or sub-sampling.

In this procedure clusters are sampling units at the first stage or primary units and these units become population for second stage of sampling. Elements within the first stage units become sampling units at second stage, these units are referred to as sub-units (or second stage units , or secondary units).

Illustrations of two –stage sampling:

(i) For conducting a health survey, we may consider districts as first stage units. We choose some districts in this stage of sampling, latter a sample of village from each district.

(ii) In the crop survey we can divide the entire field under consideration in various regions. We take a sample of regions first and then we select farms from selected regions.

(iii) Suppose a consignment containing packets of screws is submitted for inspection. First a sample of packets is selected. Later, screws from each of these selected packets are sampled. In this manner a two stage sample is selected.

Remarks:

(i) The sampling methods used at the two stages may be different or same.

(ii) If two-stage procedure is generalized further, it is called as multi-stage sampling.

(iii) Two stage sampling is easy to implement.

(iv) In simple random sampling we need the entire population to be catalogued. However, in two stage sampling only the sampled groups (primary units) are required to be catalogued.

(v) Two stage sampling is convenient from administrative point of view.

Objective type questions:

(A) Fill in the blanks

[1] Sampling is ……in many situations.

Answer: **Necessary**

[2] All sampling units are present in……..population

Answer: **Real**

[3] If the number of units in a population are limited, it is known as ……population

Answer: **Finite**

[4] A population consisting of an unlimited number of units is called an …….population.

Answer: **Infinite**

[5] If all the units of a population are surveyed it is called …….

Answer: **Complete enumeration**

[6] The errors other than sampling errors are termed as…..

Answer: **Non sampling error**

[7] The discrepancy between a parameter and its estimate due to sampling process is known as ……

Answer: **sampling error**

[8] Any population constants is called a ……

Answer: **Parameter**

[9] The list of all the items of a population is known as ……

Answer: **Sampling frame**

[10] Another name of population is…..

Answer: **Universe**

[11] The sampling procedure in the population is first divided into homogeneous groups and then a sample is drawn from each group is called ……..

Answer: **Stratified sampling**

[12] Stratified sampling is appropriate when population is…..

Answer: **Heterogeneous**

[13] Stratification is done in respect of certain. …..

Answer: **Character**

[14] Deciding the sample size for each stratum is known as …..

Answer: **Allocation**

[15] If the sample size of each stratum is in proportion to stratum size, it is called …..

Answer: **Proportional allocation**

[16] Stratified sampling comes under the category of ……..sampling

Answer: **Restricted**

[17] More heterogeneous is the population,….. is the sample size

Answer: **Larger**

[18] The quantity $\frac{N-n}{N}$ in usual notation is called …..

Answer: **Finite population correction**

[19] The quantity $\frac{n}{N}$ in usual notation is called.

Answer: **Sampling frame**

[20] When the population consists of units arranged in a sequence then …..sampling is used.

Answer: **Systematic**

[21] The main advantage of systematic sample is …..and ……

Answer: **Simple, Cheap**

[22] The main disadvantage of systematic sample is that…….formula for estimating the standard error of sample mean is available.

Answer: **No single**

[23] When the population size N is a multiple of sample size n, ….. systematic sampling is appropriate.

Answer: **Linear**

[24] When the population size N is not divisible by the sample size n, ….. systematic sampling is appropriate..

Answer: **Circular**

(B) Multiple choice questions: Choose the correct alternative from the following

[1] The data originally collected from an investigation are known as…..

(a) Primary data (b) Secondary data

(c) Both (a) & (b) (d) None of these

Answer: (a) **Primary data**

[2] Which of the following sampling techniques is an equal probability selection method (i.e., EPSEM) in which every individual in the population has an equal chance of being selected?

(a) Simple random sampling

(b) Systematic sampling

(c) Proportional stratified sampling

(d) Cluster sampling using the PPS technique

Answer: **(a) Simple random sampling**

[3] Which of the following will give a more "accurate" representation of the

population from which sample has been taken?

(a) A large sample based on the convenience sampling technique

(b) A small sample based on simple random sampling

(c) A large sample based on simple random sampling

(d) A small cluster sample

Answer: **(c) A large sample based on simple random sampling**

[4] Sampling in qualitative research is similar to which type of sampling in quantitative research?

(a) Simple random sampling

(b) Systematic sampling

(c) Quota sampling

(d) Purposive sampling

Answer: (d) **Purposive sampling**

[5] Which of the following would generally require the largest sample size?

(a) Cluster sampling

(b) Simple random sampling

(c) Systematic sampling

(d) Proportional stratified sampling

Answer: **(a) Cluster sampling**

[6] How often does the Census Bureau take a complete population count?

(a) Every year (b) Every five years

(c) Every ten years (d) Twice a year

Answer: **(c) Every ten years**

[7] A number calculated with complete population data and quantifies a characteristic of the population is called

(a) A datum (b) A statistic

(c) A parameter (d) A population

Answer: **(c) A parameter**

[8] The type of sampling in which each member of the population selected

for the sample is returned to the population before the next member is selected is called…

(a) Sampling without replacement

(b) Sampling with replacement

(c) Simple random sampling

(d) Systematic sampling

Answer: **(b) Sampling with replacement**

[9] Which of the following is a type of random sampling?

(a) Cluster sampling (b) Convenience sampling

(c) Quota sampling (d) Purposive sampling

Answer: **(a) Cluster sampling**

[10] Which of the following would usually require the smallest sample size because of its efficiency?

(a) One stage cluster sampling

(b) Simple random sampling

(c) Two stage cluster sampling

(d) Quota sampling

Answer: **(b) Simple random sampling**

[11] A technique used when selecting clusters of different sizes is called.

(a) Cluster sampling

(b) One-stage sampling

(c) Two-stage sampling

(d) Probability proportional to size or PPS

Answer: **(d) Probability proportional to size or PPS**

[12] The process of drawing a sample from a population is known as.

(a) Sampling (b) Census

(c) Survey research (d) None of the above

Answer: **(a) Sampling**

[13] It is recommended to use the whole population rather than a sample

when the population size is of what size?

(a) 500 or less (b) 100 or less

(c) 1000 or less (d) NOTA

Answer: (**b) 100 or less**

[14] Determining the sample interval (represented by k), randomly selecting a number between 1 and k, and including each kth element in your sample are the steps for which form of sampling?

(a) Simple Random Sampling

(b) Stratified Random Sampling

(c) Systematic Sampling

(d) Cluster sampling

Answer: **(c) Systematic Sampling**

[15] The non-random sampling type that involves selecting a convenience sample from a population with a specific set of characteristics for your research study is called…

(a) Convenience sampling (b) Quota sampling

(c) Purposive sampling (d) Snowball sampling

Answer: **(c) Purposive sampling**

[16] What is the standard deviation of a sampling distribution called?

(a) Sampling error (b) Sample error

(c) Standard error (d) Simple error

Answer: **(c) Standard error**

[17] Sample is a subset of a……

(a) Population (b) Parameter

(c) Statistic (d) NOTA

Answer: **(a) Population**

[C] Numerical Problems:

[1] If a population consists of 50 observations then how many

(a) SRSWOR each of size 10 can be selected

Answer: ${}^{50}C_{10}$

(b) SRSWR each of size 10 can be selected

Answer: 50^{10}

(c) Systematic samples each of size 10 can be selected.

Answer: 5

[2] Suggest appropriate sampling methods, giving area, in each of the following situations

(a) To estimate the average price of books in a library a sample of 500 books is to be selected from 10,000 books having accession numbers.

Answer: Systematic sampling, Stratified sampling

(b) In order to estimate average pocket money spent by the students in a Shri Chhatrapati Shivaji Mahavidyalaya having 3000 students, a sample of 400 students is to be selected.

Answer: Stratified sampling

(c) A market surveyor wants to select a sample of 1000 persons using telephone directory.

Answer: Systematic sampling, stratified sampling

(d) To find the daily total requirement of electricity consumption in township containing 3000 houses, 500 offices, 600 shops, 100 factories; a sample of 1000 units is to be selected.

Answer: Stratified sampling

(e) To find the daily total requirement of petrol for two wheelers in a certain city a sample of 5% of two wheelers using RTO registers is to be selected.

Answer: Systematic sampling, Stratified sampling

(f) In a socio-economic survey a sample of 1000 families is to be selected from a certain village.

Answer: SRSWOR, Stratified sampling, Systematic sampling

(g) To find the average house tax paid by citizens a sample of 500 families is to be selected using municipal corporation records.

Answer: SRSWOR, Stratified sampling, Systematic sampling

(h) To find the average income of employee, in a Shri Chhatrapati Shivaji Mahavidyalaya, Shrigonda employing various categories such as Professors, Associate Professors, Assistant Professors, Office Superindent, clerks, Peons

Answer: Stratified sampling

(i) In an industrial survey a sample of size 50 is to be selected. The area under consideration includes 100 small scale, 200 medium scale and 50 large scale industries.

Answer: Stratified sampling

[3] Identify the sampling scheme used in the following situations:

(a) For a science exhibition, 5 students are selected from each class to work as volunteers.

Answer: Stratified sampling

(b) A teacher distributed hand-outs to the students in the first row only.

Answer: Cluster sampling

(c) An examination question paper contains 10 questions of which any 5 are to be attempted. Dr. Manmath selected questions bearing even serial numbers.

Answer: Systematic sampling

(d) Suppose there are 10 divisions of F.Y.B.Sc, named as A, B, …. in a Shri Chhatrapati Shivaji Mahavidyalaya.

(i) A sample of two divisions is selected first and then 10 students from each of the selected divisions are chosen for managing sports activity.

Answer: Two-stage sampling

(ii) A sample of two divisions is selected and all the students from those divisions are engaged to conduct a survey.

Answer: Cluster sampling

Exercise: Theory Questions:

[1] Explain the different methods of classification briefly.

[2] Explain the following terms with illustrations:

(i) attribute (ii) discrete variable (iii) continuous variable.

[3] Explain the following terms:

(i) class limits (ii) class boundaries (iii) class width (iv) class frequency (v) less than type cumulative frequency (vi) more than type cumulative frequency etc. (vii) relative frequency (vii) open end class.

[4] Describe the four different scales of measurement.

[5] Explain the terms : Variable, attribute.

[6] Distinguish between variable and attribute.

[7] Write a short note on: nominal scale, ordinal scale, ratio scale, interval scale.

[8] Distinguish between primary data and secondary data.

[9] Explain the terms: time series data, cross-sectional data, failure data and directional data.

[10] Explain the terms: Population, sample, sampling unit.

[11] State the advantages and the limitations of sampling over census.

[12] What are the requirements of a good sample?

[13] Explain what mean by random sample . Why random sample is preferable? Explain the various methods of achieving randomness.

[14] Explain the procedure of drawing:

(a) SRSWR (b) SRSWOR (c) Stratified random sampling (d) systematic sampling (e) cluster sampling (f) two stage sampling.

[15] State the advantages of simple random sampling and drawbacks of the same. Also, explain how these drawbacks can be overcome.

[16] State advantages and limitations of stratified sampling and systematic sampling.

[17] How does SRSWR differ from SRSWOR?

[18] Make critical comparison between (a) Sampling and census. (b) Stratified random sampling and cluster sampling. (c) Cluster sampling and two stage sampling. (d) Random sampling and non-random sampling.

[19] Give illustrations of each of the following sampling methods:

(a) SRSWR (b) SRSWOR (c) Stratified sampling (d) Systematic sampling (e) Cluster sampling (f) Two-stage sampling.

[20] Explain the situation where sampling has larger scope as compared to census.