

Sampling methods and techniques



**ASSIGN
BUSTER**

Chapter-4: SAMPLING METHODS AND TECHNIQUES

4. 1: INTRODUCTION:

Statistics in general deals with a large number of figures. It does not deal with a single figure. All the items under considerations in any field of enquiry constitute a 'universe' or 'population'. The term population is referred to any collection of individuals or of their attributes or of results of operations which can be numerically specified. Thus, there may be population of weights of individuals, heights of trees, prices of wheat, number of plants in a field, number of students in an institution/university etc. A population with finite number of individuals or members is called a 'finite' population. For instance, the population of ages of twenty boys in a class is an example of finite population. A population with infinite number of members is known as infinite population. The population of pressures at various points in the atmosphere is an example of 'infinite' population. For any statistical investigation with large population size, complete enumeration (or census) of the population is impracticable, for example, estimation of average monthly income of the individuals in the entire country. Further, in some cases, if the population is infinite, then the complete enumeration is impossible. As an illustration, to know the total amount of timber available in the forest, the entire forest can not be cut to know how much timber is available there. The analysis of the entire population in the study is called as 'census' method of collecting data.

In practice, on the other hand, it so happens that it is not possible to examine or consider all the items of a population. Again in maximum cases consideration of all the items of a population is also not necessary.

Sometimes it is possible to obtain sufficiently accurate results by studying a part/a segment of the population. Thus the few items are selected from the population in such a way that they are the representative of the universe and these representatives in research are called as 'sample'. The process of selecting the representatives from the population is called 'sampling'. Thus sampling is simply the process of learning about population on the basis of sample drawn from it. Under this method a small group of universe is taken as the representative of the whole mass and the results are drawn. It is the method to make social/business investigation practicable and easy. For example, only 20 students are selected from a universe of 120 students who are perusing MBA degree from a particular institute situated at Puna or 50 households are selected from a village of 250 households. For determining the population characteristic, instead of enumerating all the units in the population, the units in the sample only are observed and the parameters of the population are estimated accordingly. Sampling is, therefore, resorted to when either it is impossible to enumerate all the units in the whole population or when it is too costly to enumerate in terms of time and money or when the uncertainty inherent in sampling is more than compensated by the possibilities of errors in complete enumeration.

4. 2 SAMPLING DESIGN:

A sample design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure the researcher would adopt in selecting items for the sample. Sample design also leads to a procedure to tell the number of items to be included in the sample i. e., the size of the sample. Hence, sample design is determined before the collection

of data. Among various types of sample design technique, the researcher should choose that samples which are reliable and appropriate for his research study.

Steps in sample design

There are various steps which the researcher should follow. Those are,

(i) Type of universe:

In the first step the researcher should clarify and should be expert in the study of universe. The universe may be finite (no of items are know) or Infinite (numbers of items are not know).

(ii) Sampling unit:

A decision has to be taken concerning a sampling unit before selecting a sample. Sampling unit may be a geographical one such as state, district, village etc., or construction unit such as house, flat, etc., or it may be a social unit such as family, club, school etc., or it may be an individual.

(iii) Source list:

Source list is known as ' sampling frame' from which sample is to be drawn. It consists the names of all items of a universe. Such a list would be comprehensive, correct, reliable and appropriate and the source list should be a representative of the population.

(iv) Size of sample:

Size of sample refers to the number of items to be selected from the universe to constitute a sample. Selection of sample size is a headache to the researcher. The size should not be too large or too small rather it should be optimum. An optimum sample is one which fulfills the requirements of

efficiency, representativeness, reliability and flexibility. The parameters of interest in a research study must be kept in view, while deciding the size of the sample. Cost factor i. e., budgetary conditions should also be taken into consideration (For more detail analysis on the determination of size of sample please refer to section 4. 5 of this chapter).

(v) Sampling procedure:

In the final step of the sample design, a researcher must decide the type of the sample s/he will use i. e., s/he must decide about the techniques to be used in selecting the items for the sample.

Criteria for selecting a sample procedure

While selecting samples a researcher must remember that the procedure of sampling analysis involves two costs viz., (i) the cost of collecting the data and (ii) the cost of an incorrect inferences resulting from the data. So, far as the cost of collecting data is concerned, it completely depends on the researcher to reduce it and to some extent it is within the control of the researcher. But the real problem arises while taking into account about the cost of incorrect inferences which is again of two types,

1. Systematic bias and

2. Sampling error.

1). Systematic bias results from errors in the sampling procedures, and it cannot be reduced or eliminated by increasing the sample size. It can be eliminated by eliminating and correcting the causes which are responsible for its occurrence. Following are some causes of the occurrence of systematic bias which requires concern to the researcher.

<https://assignbuster.com/sampling-methods-and-techniques/>

i. Inappropriate sampling frame:

If the sampling frame is inappropriate i. e., a biased representation of the universe, then it will result in a systematic bias.

ii. Defective measuring device:

The second cause of occurrence of systematic bias is the selection of defective measuring devices. The measuring devices may be the interviewers; the questionnaire or other instrument used to collect data or may be physical measuring devices. If the questionnaire or the interviewer is biased and/or if the physical measuring device is defective this will lead to the occurrence of systematic bias.

iii. Non-respondents:

If the researcher is unable to sample all the individuals initially included in the sample, there may arise a systematic bias. The reason is that in such a situation the likelihood of establishing correct or receiving a response from an individual is often corrected with the measure of what is to be estimated.

iv. Natural bias in the reporting of data:

There is usually a downward bias in the individual income data collected by the income tax department where as an upward bias is found in the income data collected by some social organizations. People give less income data when asked for income tax but they overstate when asked for social status.

v. Indeterminacy principle:

Same times a researcher finds that individuals act differently when kept under observations than what they do when kept in non-observed situation.

2). Sampling errors on the other hand, is the random variations in the sample estimated around the true population parameters. Since they occur randomly and are equally likely to be in either direction, their nature happens to be of compensatory type and the expected value of such errors happens to be equal to zero. Sampling error decreases with the increase in sample size and it happens to be a smaller magnitude in case where the population is characterized as homogeneous. Sampling error can be measured for a given sampling design and size which is called as ' a precision of the sampling plan'. If the sample size is increased, the precision can be improved but increase in sample size causes limitations like cost of collecting data, and also increases the systematic bias. Thus the effective way to increase the precision is usually to select a better sampling design which has a smaller sampling error for a given sample size at a given cost. Therefore, it shows that while selecting a sampling procedure the researcher must ensure that the procedure causes a relatively small sampling error and helps to control the systematic bias in a better way.

Characteristic of a good sample design:

From the above analysis, we can list down the characteristics of a good sample as follows,

Sample design must result in a truly representative sample,

Sample design must be such which results in a small sampling error,

Sampling design must be viable in the context of funds available for the research study,

Sample design must be such that systematic bias can be controlled in a better way, and

Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

4.3 SCOPE OF SAMPLING METHOD:

As discussed earlier in this chapter, the census method is one where all the units of the population under investigation are selected and information are gathered from all of them for drawing inferences. This method of data collection is rarely used by few researchers for some specific studies. The suitability of the census method requires very few conditions like- (i) where the area of study is very limited and is within the reach of the researcher, and/or (ii) the researcher is having enough time for covering the entire population, and/or (iii) the study is funded with adequate and sufficient finances to meet the expenses needed, and/or (iv) all the units of the population are behaving homogeneous characteristics, and/or (v) the study sought for adopting this specific method only. However, except these few characteristics, in all other cases it is better to use the sampling method to collect the data. It is again can be said as, it is not so easy to fulfill all the above requirements for adopting the census method. Following are some scope of sampling method of data collection.

(a) Objectives of Sampling Method:

The prime objectives of the sample survey are to obtain accurate and reliable information about the universe under study with minimization of cost, time, and energy. For example let that one want to estimate the monthly expenditure behaviour of 80 management students living in Apeejay
<https://assignbuster.com/sampling-methods-and-techniques/>

Institute of Technology, boys and girls hostel. Then census method will be appropriate one as there are around 900 borders who are residing in both the hostels. But if one wants to estimate the students living in entire hostels of management colleges (more than 45) situated in Greater Noida then sampling method may be the right option.

(b) Characteristics of Good Sample:

Out of the experiences, the researchers have opined various features/characteristics of a good sample. A good sample is one which satisfies all or few of the following conditions:

Representativeness: When sampling method is adopted by the researcher, the basic assumption is that the samples so selected out of the population are the best representative of the population under study. Thus good samples are those who accurately represent the population. Probability sampling technique yield representative samples. On measurement terms, the sample must be valid. The validity of a sample depends upon its accuracy.

Accuracy: Accuracy is defined as the degree to which bias is absent from the sample. An accurate (unbiased) sample is one which exactly represents the population. It is free from any influence that causes any differences between sample value and population value.

Size: A good sample must be adequate in size and reliable. The sample size should be such that the inferences drawn from the sample are accurate to a given level of confidence to represent the entire population under study.

(c) Merits of Sampling Method:

There are some advantages of choosing sampling method. They are

- * The volume of data in case of sampling method is small, which can be collected and analyzed quickly. Hence one can get the results urgently if s/he desires.
- * Some times census method is impossible to be employed. For example: the list of manufactures of 'saries' in India. In such a case a sample is tested to represent the entire population.
- * Since the sample is small in size hence, detailed information from the respondents can be collected
- * Qualified personnel as investigating authorities can be appointed
- * Sampling method seems to be more economical than that of the census method of data collection
- * Cross checking in case of any error may be possible. If required then data can be collected again and again from the same respondents.

(d) Demerits of Sampling Method:

Some demerits of sampling method are:

- * There is possibility that the results obtained may be false, inaccurate and misleading if the sample might not have been drawn properly out of the population under study

* Possibilities of sample errors are comparatively more. The investigator may have personal bias especially with regard to choice of techniques and drawing sampling units.

* The size of sample may not be sufficient to represent the entire universe.

* When the universe is small one then it is not advisable to go for sampling technique of data collection.

4. 4. LAWS OF SAMPLING:

Aggarwal and Diwan in their study have mentioned about two fundamental principles on which the sampling theory rests on:

1. The law of statistical regularity and
2. The law of inertia of large numbers.

Both the above theories are discussed as follows:

1. Law of Statistical Regularity

The law states that if a moderately large number of items are selected at random from a given population, the characteristics of those items will reflect, to a fairly accurate degree, the characteristic of the entire population. For example, if 300 employees are picked from a company at random and the average height is found out, the result will be nearly the same as will be found if all the employees of the company are picked up and measured.

The reliability of the law depends on the two factors viz., (i) the size of the sample which says that the larger the sample, the more reliable are its

indicators. The reliability of the sample is proportional to the square root of the number of items it contains and larger the samples the more representative and stable, and (ii) the sample must be chosen at random.

There are various characteristics on which the applicability of the law is based on. The first one is that with the use of this law, a part of the universe can be chosen. Thus, when census method for collecting information is not possible because of the constraints, then with the help of this law and by using the method of random sampling, researchers can determine the sample units. The second one is that, when selection is made at random, then by using this law, all good, bad and average units of the entire population have equal chance of being selected and third characteristics is that with the help of this law, inferences drawn from a particular inquiry for different time and place can be used for all other places with little adjustments.

2. Law of inertia of large numbers:

The law of inertia of large numbers is a corollary of the law of statistical regularity and lays down that 'in large masses of data abnormalities will occur, but in all probability, exceptional items will offset each other, leaving the average unchanged subject, where the element of time enters, to the general trend of data'. According to King, 'the law of inertia of large numbers asserts that large aggregates are the results of the movements of its separate parts, and it is impossible that the latter will all be moving in the same direction at the same time. Consequently, their movements will tend to compensate one another, and the larger the number involved, the more complete will this compensation be'. To summarize the above definitions it can

be found that the larger the number of items one chose from a universe, the greater is the possibility of accuracy. Hence, the law is based on the fact that if one part of a large group varies in one direction, the probability that another equal part of the same group would vary in the opposite direction, so that the total change would be insignificant.

4. 5 DETERMINATION OF SAMPLE SIZE:

One of the important characteristics of a good sample is that it must be adequate in size in relation to the population. What should be appropriate sample size? The Air University sampling and surveying handbook answers this question by developing three different formulas for determining appropriate sample size based on three different situations as derived below.

Formula-1: If the nature of the study is such that the researcher has to report the results as percentages (proportions) of the sample responding, then the formula for calculating sample size is:

where n = sample size required, N = Number of people in the population, P = estimated percentage of the population possessing attribute of interest, A = Accuracy desired, expressed as a decimal (i. e., 0. 01, 0. 02, 0. 03, 0. 04, 0. 05 etc.) and Z = number of standard deviation units of the sampling distribution corresponding to the desired confidence level (see Appendix-I for Z values).

Formula-2: If the nature of the study is such that the researcher has to report the results of the study as means (averages) of the sample responding, then the formula will be:

where, n = sample size required, N = Number of people in the population, P = estimated standard deviation of the attribute of interest in the population, A = Accuracy desired, expressed as a decimal (i. e., 0. 01, 0. 02, 0. 03, 0. 04,

0.05 etc.) and Z = number of standard deviation units of the sampling distribution corresponding to the desired confidence level.

Formula-3: If, on the other hand, the nature of the study is such that the researcher is planning to report the results in a variety of ways, or if, the researcher is getting difficulty in estimating percentage or standard deviation of the attribute of interest, then following formula may be more suitable for use:

where, n = sample size required, N = total population size (either known or estimated), d = precision level (usually 0.05 or 0.10) and Z = number of standard deviation units of the sampling distribution corresponding to the desired confidence level.

The above formula can be clearer with the below derived example. Let that the total population (N) = 10000 and the researcher decided to consider this study at 95% confidence level and ± 5 percent precision level ($d = 0.05$, $Z = 1.96$), the sample size ' n ' will be:

So, a representative sample of 370 (369.98 rounded up) would be sufficient to satisfy the risk level. An analysis of the formula shows that the required sample size will increase most rapidly if: (i) the confidence level (Z factor) is increased, or (ii) the precision level (d) is made smaller.

In case the nature of the study is such that the population is stratified into more than one group, the size of each group will be its proportion (percentage) in the population times the total sample size as computed above. To illustrate, recall the example as discussed above of four stratified groups. Using the ' n ' of 370 calculated above, each of these strata should have the following sample sizes:

* Business community, male $370 * 0.455 = 168.35 = 168$

* Business community, female $370 * 0.195 = 72.15 = 72$

* Government officer, male $370 * 0.245 = 90.65 = 91$

* Government officer, female $370 * 0.105 = 38.85 = 39$

Factors Affecting the Size of Sample:

The size of sample depends on number of factors. Some important among them are:

Homogeneity or Heterogeneity of the universe:

Selection of sample depends on the nature of the universe. It says that if the nature of universe is homogeneous then a small sample will represent the behaviour of entire universe. This will lead to selection of small sample size rather than a large one. On the other hand, if the universe is heterogeneous in nature then samples are to be chosen as from each heterogeneous unit.

Number of classes proposed:

If a large number of class intervals to be made then the size of sample should be more because it has to represent the entire universe. In case of small samples there is the possibility that some samples may not be included.

Nature of study:

The size of sample also depends on the nature of study. For an intensive study which may be for a long time, large samples are to be chosen.

Similarly, in case of general studies large number of respondents may be appropriate one but if the study is of technical in nature then the selection of

large number of respondents may cause difficulty while gathering information.

4. Practical considerations:

Practical considerations are the availability of finance, time for study along with the availability of the trained and experienced experts. These factors weight a lot in the process of selecting sample size.

Geographic area of the study:

If the area covered by a survey is very large (a country or a state) and the size of the population is quite large, then the size of sample should be large. But if the area and the size of the population are small, than relatively small sample could be enough.

4. 6 TECHNIQUES OF SAMPLING:

Equally important to the size of the sample is the determination of the type of sampling techniques to be followed. Different types of sampling techniques are used for drawing a sample plan. The techniques of sampling are classified into two broad categories as derived below:

Probability sampling and

Non-probability sampling

1. Probability Sampling:

It provides a scientific technique of drawing samples from the universe. In such a case each unit has some defined pre-assigned probability of being chosen in the sample. Different types of probability sampling techniques are:

(i) Random sampling

(ii) Systematic sampling

(iii) Stratified sampling

(iv) Cluster sampling and

(v) Multi-stage sampling

(i) Random Sampling:

A random sampling is one where each item in the universe has an equal or known opportunity of being selected. In addition, the selection of one member should in no way influence the selection of another. According to W. M. Harper, ' a random sample is a sample selected in such a way that every item in the population has an equal chance of being included'. This type of sampling is more suitable comparatively in large samples and when population is homogeneous, that is, one composed of members who all possess the same attribute that the researcher are interested in measuring. The simple random sample requires less knowledge about the population than other techniques of probability sampling, but it does have two major drawbacks. One is if the population is large, a great deal of time must be spent listing and numbering the members. The other is the fact that a simple random sample will not adequately represent many population attributes (characteristics) unless the sample is relatively large. In identifying the population to be surveyed, homogeneity can be determined by asking the question, ' what is (are) the common characteristic(s) that are of interest?' These may include such characteristics as age, sex, rank/grade, position, income, religion or political affiliation, etc., whatever is the base of the research study that the researcher interested in measuring. One of the

greatest advantage is that random sampling always produces the smallest possible sampling error. In a very real sense, the size of the sampling error in a random sample is affected only by random chance. Because a random sample contains the least amount of sampling error, it can be said that it is an unbiased sample. Note that this does not mean that this sampling technique contains no error, but rather the minimum possible amount of error.

Process of Selecting Random Samples:

There are four methods of drawing out a sample on random basis. They are:

Ø Lottery Method:

Under this method the various units of the universe are numbered on small and identical slips of papers which are folded and mixed together in a drum or in a flat container. A blindfold selection is then made from the number of slips required to constitute the desired size of sample.

Ø Use of random number tables:

The most practical and economical method of selecting a random sample consists in the use of random numbers table which have been so constructed that each of the digits from 0, 1, 2... 9 appears with approximately the same frequency and independently with each other. The best way to choose a sample is to use a random number table (or let a computer generate a series of random numbers automatically). In either case, the researcher would assign each member of the population a unique number (or perhaps use a number already assigned to them such as telephone number, zip code, etc.). The members of the population chosen for the sample will be those whose numbers are identical to the ones extracted from the random number table

(or computer) in succession until the desired sample size is reached (an example of a random number table and instructions for its use appear in Appendix-II attached at the end of this book). Many statistical texts or mathematical tables treat random number generation. A less rigorous procedure for determining randomness is to write the name of each member of the population on a separate card, and with continuous mixing, draw out cards until the desired sample size is reached.

Ø Selecting from sequential list:

Under this the names of the respondents/items are first arranged serially according to alphabetical, geographical or simply in serial order. Then out of this every 10th number or any such number that is determined by the researcher based on the cases may be taken up.

Ø Grid system:

According to this method a map of the entire area under study is prepared. Then a screen with sequence is placed upon the map and the areas falling within the selected area are considered as samples.

It is however, drawing a random sample calls for the following precaution:

Ø Populations to be sampled must be clearly defined.

Ø Different units should approximately of equal size.

Ø The unit must be independent of each other.

Ø Each unit should be accessible. Unit once selected should not be ignored or replaced by any other unit.

Merits of Random Sample Method:

- Ø It is more scientific method of taking out samples from the universe since it minimizes personal bias

- Ø Less possibility of sampling error

- Ø No advance knowledge of the characteristic of the population is necessary under this method

- Ø It is assumed that the samples drawn under this method are true representative of the universe and

- Ø This method provides us most reliable and maximum information at the least cost which save time, money and also labour.

Demerits of Random Sample Method:

Random sample method of data collection is having some practical difficulties. Some important one's are as follows:

- Ø This method requires complete list of the universe. But in real life such information is not available in much research studies which restricts the use of this method freely

- Ø In field research where the area of coverage is fairly large then the units selected under this method are expected to be scattered in widely geographical area and thus, may be time consuming

- Ø The selected sample may not be a true representative of the universe and

- Ø Some times this method gives such results whose probability is very small or negligible.

(ii) Stratified Random Sampling:

This method is used when the population is heterogeneous rather than homogeneous (or as discussed above, when the researcher wants to obtain a representative sample across many population attributes). A heterogeneous population is composed of unlike elements; such as, officers of different ranks, different levels of management personnel, civilians and military personnel, or the patrons of a discount store (differing by gender or age). A stratified random sample is defined as a combination of independent samples selected in proper proportions from homogeneous groups within a heterogeneous population. The procedure calls for categorizing the heterogeneous population into groups that are homogeneous in themselves. If one group is proportionally larger than the other, its sample size should also be proportionally larger. The number of groups to be considered is determined by the characteristics of the population. For example, if one is comparing Business community and governmental officer segments on a self determined base, each of these will be a separate group. After dividing the population into groups, then each homogeneous group is to be sampled by using any techniques of probability sampling, of course as per the requirement. Finally, the sample statistics are to be calculated for each group to determine how many members are needed from each subgroup. Two separate cases derived in Box-4. 1 and Box-4. 2 is enough for the readers to clear their fundamental on applicability of stratified random sampling

Box-4. 1: Selecting Samples by Using Stratified Sampling Technique

Let's say that the researcher wants to draw a random sample from a population of a village to assess their opinions on some issue related to

<https://assignbuster.com/sampling-methods-and-techniques/>

income inequality. In addition, s/he would like to determine if the opinions differ by government officials and business community and also by gender of the individuals surveyed. It is recognized that the sample s/he wants to draw is heterogeneous in respect of the two attributes of interest to the researcher. So, four homogeneous subgroups are created: like (i) Business community, male; (ii) Business community, female; (iii) Government Officials, male and (iv) Government officials, female

Now, each group is homogeneous on both attributes. To ensure each subgroup in the sample will represent its counterpart subgroup in the population, the researcher must ensure each subgroup represented in the sample in the same proportion to the other subgroups as they are in the population. Let's assume that it is known (or can be estimated) that the population of the selected village which is to be distributed as