SAMPLING

Shahbaz Baig

Abstract: The process of drawing sample (a part) from whole set of population or universe is called sampling. From population we draw sample. There are varieties of option available to a researcher fore selection of a sample. Inappropriate sample of sampling technique will surely introduce bias.

Key words: Sampling , Randomised Sampling, Bias, Population

INTRODUCTION

Since it is generally impossible to study an entire population (every individual in a country, all college students, every geographic area,all etc.), researchers typically rely on sampling to acquire a section of the population to perform an experiment or observational study.

It is important that the group selected be representative of the population, and not biased in a systematic manner. For example, a group comprised of the wealthiest individuals in a given area probably would not accurately reflect the opinions of the entire population in that area. For this reason, randomization is typically employed to achieve an unbiased sample. The most common sampling designs are simple random sampling, stratified random sampling, and multistage random sampling.

Most survey work involves sampling from finite populations. There are Paris to any sampling strategy (design).

First, there is a selection procedure, the manner in which sampling units are selected from a population.

Second, there is an estimation procedure that prescribed how inferences are to be drawn from sample to the population.

SAMPLING

The process of drawing sample (a part) from whole set of population or universe is called sampling. (Sampling is a subset of a population)

POPULATION OR UNIVERSE

Is a set objects or persons having a common characteristic. From population we draw sample.

Article Citation: Baig S, Sampling. Indep Rev Jul-Sep 2017;19(7-9): 115-126.

Date received: 14/06/2017
Date Accepted: 30/06/2017

Correspondence Address:
Dr. Shahbaz Baig, MBBS, MPH
Assistant Professor of Community Medicine
Independent Medical College, Faisalabad.

1. Dr, Shahbaz Baig, MBBS, MPH
Assitant Professor Community Medicine,
Independent Medical College, Faisalabad.

SAMPLING FRAME

Is the actual list of all persons / units from which the sample is selected.

ADVANTAGES OF A SAMPLING

One may ask question "why not to study the entire population?" or to go for a census? Following are the main reasons:

REDUCE COST:

If data are secured from only a small fraction of the aggregate , expenditure are smaller than if a complete sences is attempted .

GREATER SPEED:

For the same reason, the data can be collected and summarized more quickly with a sample than with a complete count.

GREATER SCOPE:

In certain type of inquiry highly traind perfessional of specialized equipment limmeted availability, must be used to obtainet the data.A complete sences is impracticable the choice lies between obtaining the information by sampling or not at all.Thus surveys that rely on sampling have more scope flexibility.

GREATER ACCURACY:

A sample may produce more accurate results than the kind of complete enumeration that can b taken.

(SAMPLING TECHNIQUES BY WILLIAM G. COCHRAN , THIRD EDITION)

1. one can't having accessibility to each and every person in population and sample provides us pportunity to select a subset of population (accessibility)
2. It is cost effective (saves lot of money)
3. Less resources are required in terms of manpower, time, material (stationary), etc.
4. It is sometime more feasible on the basis of a sample study (quick)
5. Generalization is enhanced if random / probability sampling is carried out.

DISADVANTAGE

1. There is always a chance sampling error
2. It may create a feeling of dissatisfaction and discrimination in people who are not included in the study.
3. It is also unethical and morally unacceptable, if some benefit is being offered to the sampled population. For e.g benefit may be in the shape of food items like milk, biscuits etc.

USES OF SAMPLING

1) On a smaller scale local gournment city, stat, and country are making increasd use of sample surveys to obtain information needed for future planning and for meeting pressing problems.
2) Market research is havely dependent on the sampling approch.
3) Busennes and industry have many uses for sampling in attempting to increase the efficiency of there internal operation.

TYPES OF SAMPLING

There are varieties of option available to a researcher fore selection of a sample. Inappropriate sample of sampling technique will surely introduce bias. Mainly two types of sampling techniques are used

1) Random or probability sampling
2) Non Random or Non probability sampling

RANDOM OR PROBABILITY SAMPLING

In probability sampling we select sample randomly and every person from population has an equal chance to be included or excluded from the sample.

Following are types of probability sampling

SIMPLE RANDOM SAMPLING

In this type of sampling, the first step is to draw up a list of all the individuals in a population and to number them. This is called sampling frame.

In second step the required number of individual is selected, each has the same and equal chance of being chosen, this could be done through:-

a. Balloting / Lottery
b. Table of random numbers.
c. Computer generated random number
d. Flipping fair a coin
e. Rolling a dice.

Advantages and disadvantages of simple random sampling

The advantages and disadvantages of simple random sampling are explained below. Many of these are similar to other types of probability sampling technique, but with some exceptions. Whilst simple random sampling is one of the 'gold standards' of sampling techniques, it presents many challenges for students conducting dissertation research at the undergraduate and master's level. If you are still contemplating whether to use simple random sampling for your dissertation after reading these advantages and disadvantages, you may find the following article, Making sensible sampling choices[coming soon], helpful.

Advantages of simple random sampling.

The aim of the simple random sample is to reduce the potential for human bias in the selection of cases to be included in the sample. As a result, the simple random sample provides us with a sample that is highly representative of the population being studied, assuming that there is limitedmissing data.

Since the units selected for inclusion in the sample are chosen using probabilistic methods, simple random sampling allows us to make generalisations (i.e. statistical inferences) from the sample to the population. This is a major advantage because such generalisations are more likely to be considered to have external validity.

Disadvantages of simple random sampling

A simple random sample can only be carried out if the list of the population is available and complete. Sometimes, a list of the population you are interested in will simply not exist.

Where a list of the population does exist, attaining a complete list can be difficult for a number of reasons:

1. Even if a list is readily available, it may be challenging to gain access to that list. The list may be protected by privacy policies or require a lengthy process to attain permissions.
2. There may be no single list detailing the population you are interested in. As a result, it may be difficult and time consuming to bring together numerous sub-lists to create a final list from which you want to select your sample. As an undergraduate and master's level dissertation student, you may simply not have sufficient time to do this.
3. Many lists will not be in the public domain and their purchase may be expensive; at least in terms of the research funds of a typical undergraduate or master's level dissertation student.

4. In terms of human populations (as opposed to other types of populations; see the article: Sampling: The basics), some of these populations will be expensive and time consuming to contact, even where a list is available. Assuming that your list has all the contact details of potential participants in the first instance, managing the different ways (e.g. postal, telephone, email) that may be required to contact your sample may be challenging, not forgetting the fact that your sample may also be geographical scattered.

In the case of human populations, to avoid potential bias in your sample, you will also need to try and ensure that an adequate proportion of your sample takes part in the research. This may require re-contacting non-respondents, can be very time consuming, or reaching out to new respondents.

If you are an undergraduate or master's level dissertation student considering using simple random sampling, you may also want to read more about how to put together your sampling strategy [see the section: Sampling Strategy].

SYSTEMATIC SAMPLING

In this type of sampling individuals are selected at regular intervals.

Systematic sampling involves three steps:
First, determine the sampling interval, which is symbolized by "k," (it is the population size divided by the desired sample size).

Second, randomly select a number between 1 and k, and include that person in your sample.

Third, also include each kth element in your sample. For example if k is 10 and your randomly selected number between 1 and 10 was 5, then you will select persons 5, 15, 25, 35, 45, etc.

When you get to the end of your sampling frame you will have all the people to be included in your sample.

One potential (but rarely occurring) problem is called periodicity (i.e., there is a cyclical pattern in the sampling frame). It could occur when you attach several ordered lists to one another (e.g., if you had took lists from multiple teachers who had all ordered their lists on some variable such as IQ). On the other hand, stratification within one overall list is not a problem at all (e.g., if you have one list and have it ordered by gender, or by IQ). Basically, if you are attaching multiple lists to one another, there could be a problem. It would be better to reorganize the lists into one overall list (i.e., sampling frame).

For example to select a 5% or 1 in 20, sample of population, the starting point is chosen from numbers 1 to 20, and the n every 20th person on the list is taken.

Suppose 8 is the random number selected, then the sample would comprise individuals, having no:
8, 28, 48, 68, 88, 108, and so on.

The sample is equally spaced throughout the area or population to be sampled.

If 'n' is the desired sample size e.g 10 and total population is
N = 50 then
K = N/n = 50/10 = 5

We will randomly select a number from 1 to

5 and then every 5th number will be taken. Say for example at random number 4 is selected. Now we will take number 4,9,14,19,24,29,34,39,44 and 49

it will complete sample of 10 person from 50 indiviuals.

ADVANTAGES OF SYSTEMATIC SAMPLING[3]

The main advantage of using systematic sampling over simple random sampling is its simplicity. It allows the researcher to add a degree of system or process into the random selection of subjects.
It is easier to draw a sample and often easier to exicute with out mistake. This is a particular advantage when the drawing is done in the field.

(SAMPLING TECHNIQUES BY WILLIAM G. COCHRAN, THIRD EDITION)

DISADVANTAGE OF SYSTEMATIC SAMPLING

The process of selection can interact with a hidden periodic trait within the population. If the sampling technique coincides with the periodicity of the trait, the sampling technique will no longer be random and representativeness of the sample is compromised.

NOTES

Since systematic random sampling is a type of probability sampling, the researcher must ensure that all the members of the population have equal chances of being selected as the starting point or the initial subject.

The researcher must be certain that the chosen constant interval between subjects do not reflect a certain pattern of traits present in the population. If a pattern in the population exists and it coincides with the interval set by the researcher, randomness of the sampling technique is compromised.

STRATIFIED SAMPLING

This type of sampling is recommended when population consist of various sub groups (strata), which differ with respect to the feature understudy e.g is age, sex, ethic groups etc.

The population is stratified according to the features and persons are selected with the help of simple random sampling, from each stratum to ensure that they all are adequately represented.

The second option for selection of sample from various strata is proportional allocation according to size of strata.

For example there are 2 strata, strata A & strata B. strata A consist of 12 persons and strata B has 8 persons.
Strata A = 1,2,3,4,5,6,7,8,9,10,11,12
Strata B = 1,2,3,4,5,6,7,8

We want 25% sample i.e. out of 12 + 8 = 20 we require 5 persons.
n1 = number of persons from strata A
= n * (N1/N) = 5 * 12/20 = 3
n = desired number is sample = 5
N1= number is strata A = 12
N = overall total = 20
Number of persons from strata B = n2
= n * (N2/N)
= 5 * 8/20
= 2

STRATIFIED SAMPLING STRATEGIES

1. PROPORTIONATE ALLOCATION

Uses a sampling fraction in each of the strata that is proportional to that of the

total population. For instance, if the population consists of 60% in the male stratum and 40% in the female stratum, then the relative size of the two samples (three males, two females) should reflect this proportion.

2. **OPTIMUM ALLOCATION**
(or Disproportionate allocation) - Each stratum is proportionate to the standard deviation of the distribution of the variable. Larger samples are taken in the strata with the greatest variability to generate the least possible sampling variance.
3. **EQUAL ALLOCATION:**
The allocation is called is equal when from each straum, equal number of sampling units is selected i.e. the total sample size n is distributed equally among all the k strata.

(SAMPLING TECHNIQUES BY WILLIAM G. COCHRAN, THIRD EDITION)

4. **NAYMAN ALLOCATION**
This method of allocation was porposed by j.Nayman (1894-1981) in 1934 and it consists of finding n which minimizes the variance of the stratified sampling mean for a fixed total sample size n, assuming the costs of the survying the units , to be the same in all strata.
(SHER MUHAMMAD CHUDHRY)

ADVANTAGES

A real-world example of using stratified sampling would be for a political survey. If the respondents needed to reflect the diversity of the population, the researcher would specifically seek to include participants of various minority groups such as race or religion, based on their proportionality to the total population as mentioned above. A stratified survey could thus claim to be more representative of the population than a survey of simple random sampling or systematic sampling.

Similarly, if population density varies greatly within a region, stratified sampling will ensure that estimates can be made with equal accuracy in different parts of the region, and that comparisons of sub-regions can be made with equal statistical power. For example, inOntario a survey taken throughout the province might use a larger sampling fraction in the less populated north, since the disparity in population between north and south is so great that a sampling fraction based on the provincial sample as a whole might result in the collection of only a handful of data from the north.

Randomized stratification can also be used to improve population representativeness in a study.

DISADVANTAGES

Stratified sampling is not useful when the population cannot be exhaustively partitioned into disjoint subgroups. It would be a misapplication of the technique to make subgroups' sample sizes proportional to the amount of data available from the subgroups, rather than scaling sample sizes to subgroup sizes (or to their variances, if known to vary significantly e.g. by means of an F Test). Data representing each subgroup are taken to be of equal importance if suspected variation among them warrants stratified sampling. If, on the other hand, the very variances vary so much, among subgroups, that the data need to be stratified by variance, there is no way to make the subgroup sample sizes proportional (at the same time) to the subgroups' sizes within the total population.

(What is the most efficient way to partition sampling resources among groups that vary in both their means and their variances?)

5) Multistage Sampling7

In multistage sampling selection is done in stage until sampling units (houses, school, persons, etc) are arrived. The procedure may contain two or more stages.

A multistage random sample is constructed by taking a series of simple random samples in stages. This type of sampling is often more practical than simple random sampling for studies requiring "on location" analysis, such as door-to-door surveys.

In a multistage random sample, a large area, such as a country, is first divided into smaller regions (such as states), and a random sample of these regions is collected.

In the second stage, a random sample of smaller areas (such as counties) is taken from within each of the regions chosen in the first stage.

Then, in the third stage, a random sample of even smaller areas (such as neighborhoods) is taken from within each of the areas chosen in the second stage. If these areas are sufficiently small for the purposes of the study, then the researcher might stop at the third stage. If not, he or she may continue to sample from the areas chosen in the third stage, etc., until appropriately small areas have been chosen.

Advantages

1. resources can be concentrated in limited number of places.
2. Cost effective
3. Less time consuming
4. Sampling frame is not needed for whole population.

Disadvantages

1. precision (reliability is comprised if we compare it with simple random sampling.
2. Sampling error is increased

CLUSTER SAMPLING[5,6]

Cluster Sampling is a sampling technique used when "natural" groupings are evident in a statistical population. In this technique, the total population is divided into the groups (or clusters) and a sample of the groups is selected. Then the required information is collected from the elements within each selected group. This may be done for every element in these groups or a subsample of elements may be selected within each of these groups.

A common motivation for cluster sampling is to reduce the average cost per interview.

Given a fixed budget, this can allow an increased sample size.

Assuming a fixed sample size, the technique gives more accurate results when most of the variation in the population is within the groups, not between them.

Cluster Elements

Elements within a cluster should ideally be as heterogeneous as possible, but there should be homogeneity between cluster means.

Each cluster should be a small scale representation of the total population.

The clusters should be mutually exclusive and collectively exhaustive.

A random sampling technique is then used on any relevant clusters to choose which clusters to include in the study.

In single-stage cluster sampling, all the elements from each of the selected clusters are used.

In two-stage cluster sampling, a random sampling technique is applied to the elements from each of the selected clusters.

Difference Between Cluster Sampling And Stratified Sampling

The main difference between cluster sampling and stratified sampling is that in cluster sampling the cluster is treated as the sampling unit so analysis is done on a population of clusters (at least in the first stage).

In stratified sampling, the analysis is done on elements within strata.

In stratified sampling, a random sample is drawn from each of the strata, whereas in cluster sampling only the selected clusters are studied.

The main objective of cluster sampling is to reduce costs by increasing sampling efficiency. This contrasts with stratified sampling where the main objective is to increase precision.

There also exists multistage sampling, where more than two steps are taken in selecting clusters from clusters.

Aspects Of Cluster Sampling

One version of cluster sampling is area sampling or geographical cluster sampling. Clusters consist of geographical areas. Because a geographically dispersed population can be expensive to survey, greater economy than simple random sampling can be achieved by treating several respondents within a local area as a cluster. It is usually necessary to increase the total sample size to achieve equivalent precision in the estimators, but cost savings may make that feasible.

In some situations, cluster analysis is only appropriate when the clusters are approximately the same size. This can be achieved by combining clusters. If this is not possible, probability proportionate to size sampling is used. In this method, the probability of selecting any cluster varies with the size of the cluster, giving larger clusters a greater probability of selection and smaller clusters a lower probability. However, if clusters are selected with probability proportionate to size, the same number of interviews should be carried out in each sampled cluster so that each unit sampled has the same probability of selection.

Cluster sampling is used to estimate high mortalities in cases such as wars, famines and natural disasters.[5]

Advantages

Can be cheaper than other methods - e.g. fewer travel expenses, administration costs

Disadvantages

Higher sampling error, which can be expressed in the so-called "design effect", the ratio between the number of subjects in the cluster study and the number of subjects in an equally reliable, randomly sampled unclustered study.[6]

NON PROBABILITY SAMPLING (GRAB)

Every individual don't have equal chance of selection or rejection. Generalization of these types of sampling is always compromised.

We can divide nonprobability sampling methods into two broad types:

1) Accidental / Convenience
2) Purposive.

Most sampling methods are purposive in nature because we usually approach the sampling problem with a specific plan in mind. The most important distinctions among these types of sampling methods are the ones between the different types of purposive sampling approaches.

ACCIDENTAL, HAPHAZARD OR CONVENIENCE SAMPLING

One of the most common methods of sampling goes under the various titles listed here. I would include in this category the traditional "man on the street" (of course, now it's probably the "person on the street") interviews conducted frequently by television news programs to get a quick (although nonrepresentative) reading of public opinion. I would also argue that the typical use of college students in much psychological research is primarily a matter of convenience. (You don't really believe that psychologists use college students because they believe they're representative of the population at large, do you?). In clinical practice,we might use clients who are available to us as our sample. In many research contexts, we sample simply by asking for volunteers. Clearly, the problem with all of these types of samples is that we have no evidence that they are representative of the populations we're interested in generalizing to -- and in many cases we would clearly suspect that they are not.

PURPOSIVE SAMPLING

In purposive sampling, we sample with a purpose in mind. We usually would have one or more specific predefined groups we are seeking. For instance, have you ever run into people in a mall or on the street who are carrying a clipboard and who are stopping various people and asking if they could interview them? Most likely they are conducting a purposive sample (and most likely they are engaged in market research). They might be looking for Caucasian females between 30-40 years old. They size up the people passing by and anyone who looks to be in that category they stop to ask if they will participate. One of the first things they're likely to do is verify that the respondent does in fact meet the criteria for being in the sample. Purposive sampling can be very useful for situations where you need to reach a targeted sample quickly and where sampling for proportionality is not the primary concern. With a purposive sample, you are likely to get the opinions of your target population, but you are also likely to overweight subgroups in your population that are more readily accessible.

All of the methods that follow can be considered subcategories of purposive sampling methods. We might sample for specific groups or types of people as in modal instance, expert, or quota sampling. We might sample for diversity as in heterogeneity sampling. Or, we might capitalize on informal social networks to identify specific respondents who are hard to locate otherwise, as in snowball sampling. In all of these methods we know what we want -- we are sampling with a purpose.

JUDGEMENT SAMPLING:

This the type of sampling in which the decision of which unit included in the sample is left to the judgement of the enumerator of the survey because the claim is that he knows the units in the populations so he is the best qualified to determine which units should be included in the sample so as to be represented of the population.
(FAQIR MUHAMMAD Chudhry)

QUOTA SAMPLING

In quota sampling, you select people nonrandomly according to some fixed quota. There are two types of quota sampling: proportional and non proportional. In proportional quota sampling you want to represent the major characteristics of the population by sampling a proportional amount of each. For instance, if you know the population has 40% women and 60% men, and that you want a total sample size of 100, you will continue sampling until you get those percentages and then you will stop. So, if you've already got the 40 women for your sample, but not the sixty men, you will continue to sample men but even if legitimate women respondents come along, you will not sample them because you have already "met your quota." The problem here (as in much purposive sampling) is that you have to decide the specific characteristics on which you will base the quota. Will it be by gender, age, education race, religion, etc.?

Nonproportional quota sampling

is a bit less restrictive. In this method, you specify the minimum number of sampled units you want in each category. here, you're not concerned with having numbers that match the proportions in the population. Instead, you simply want to have enough to assure that you will be able to talk about even small groups in the population. This method is the nonprobabilistic analogue of stratified random sampling in that it is typically used to assure that smaller groups are adequately represented in your sample.

Heterogeneity Sampling

We sample for heterogeneity when we want to include all opinions or views, and we aren't concerned about representing these views proportionately. Another term for this is sampling fordiversity. In many brainstorming or nominal group processes (including concept mapping), we would use some form of heterogeneity sampling because our primary interest is in getting broad spectrum of ideas, not identifying the "average" or "modal instance" ones. In effect, what we would like to be sampling is not people, but ideas. We imagine that there is a universe of all possible ideas relevant to some topic and that we want to sample this population, not the population of people who have the ideas. Clearly, in order to get all of the ideas, and especially the "outlier" or unusual ones, we have to include a broad and diverse range of participants. Heterogeneity sampling is, in this sense, almost the opposite of modal instance sampling.

Snowball Sampling

In snowball sampling, you begin by identifying someone who meets the criteria for inclusion in your study. You then ask them to recommend others who they may know who also meet the criteria. Although this method would hardly lead to representative samples, there are times when it may be the best method available. Snowball sampling is especially useful when you are trying to reach populations that are inaccessible or hard to find. For instance, if you are studying the homeless, you are not likely to be able

to find good lists of homeless people within a specific geographical area. However, if you go to that area and identify one or two, you may find that they know very well who the other homeless people in their vicinity are and how you can find them.

SAMPLING AND NON SAMPLING ERROR

Sampling Error (random error)

When one draws a sample from sampling frame (population), the possibility always exists that the sample selected is not representative of the target population, this is due to sampling error.

We can reduce the likelihood of happening of sampling error by:

a) Random selection
b) Increasing the size of sample

Sampling error and non sampling error can occur due to

- Observer Variations
- Inter observer variations
- Intra observer variations

Technical errors

- Defective instruments
- Erroneous calibrations
- Faulty reagents

Confounding Factor;

It is a factor which is associated with both (exposure and disease) and is distributerd unequally in study and control groups. This factor distorts the study results. For example age could be biggest confounding factor in many studies.

Biological Variation in the subjects

SAMPLING VARIATION

If we take different samples from the same population, because of chance, different samples give different results. This is called sampling variation.

SAMPLING WITH OR WITHOUT REPLACEMENT

Sampling with replacement

It is providing second chance to a sampling unit depending on the size of the units. For example if a school has doubled the number of students as compared to other school, it should have double the chance of being selected. Sampling with replacement is the answer. It another chance of being selected

Sampling without Replacement

In sampling without replacement, the element is not returned once it has been chosen.

SAMPLE SIZE ESTIMATION

There are so many formulas for estimation of sample size. The eventual sample size is a compromise between feasibility in terms of resources and desirable sample size. One of the formula for sample size estimation is given below.

$$n = \frac{N \times Z2 \times P \times (100 - P)}{d2\,(N - 1) + Z2 \times P \times (100 - P)}$$

n = sample size
N= size of population
Z = level of confidence i.e., 95% = 1.96
P = presumed prevalence in % age
D = acceptable margin of error usually 5%

Refferences

1.

2. http://www.southalabama.edu/coe/bset/johnson/lectures/lec7.htm

3. Joan Joseph Castillo (2009). Systematic Sampling. Retrieved 28 Feb. 2012 from Experiment Resources: http://www.ex-periment-resources.com/systematic-sampling.html

4. http://www.coventry.ac.uk/ec/~nhunt/meths/strati.htmlAccessed 2008/01/27.

5. David Brown, Study Claims Iraq's 'Excess'Death Toll Has Reached 655,000, Washington Post, Wednesday, October 11, 2006. Retrieved September 14, 2010.

6. Kerry and Bland (1998). Statistics notes: The intracluster correlation coefficient in cluster randomisation. British Medical Journal, 316,1455-1460.

7 (Definition and example taken from Valerie J. Easton and John H. McColl's Statistics Glossary v1.1) http://www.stat.yale.edu/Courses/1997-98/101/sample.htm

8. ©2006, William M.K. Trochim, Research Methods Knowledge Base