MedLife Clinics

ISSN: 2689-5943

Review Article

A Basic Approach in Sampling Methodology and Sample Size Calculation

Ilker Etikan[*] and Ogunjesa Babatope

Department of Biostatistics, Near East University, Cyprus

Abstract

A major concept in clinical and experimental studies is the selection of subjects or units for the conduct of studies. This is essentially in order to make acceptable inferences that can be generalized to the population. Failure to select a sample correctly from the population could result in errors that can produce a misleading conclusion. Therefore, this study introduces us to basic concepts in sampling methodology and various methods used in sample selection. In conclusion, method on how to calculate sample size for a survey is discussed.

Keywords: Population; Sample; Survey; Sample size; Sampling; Sampling error; Randomization

Introduction

Making decision especially in medical studies is contingent on the subjects of such studies. If subjects for any clinical studies are not rightly chosen or numbers are inadequate relative to the population, a conclusion made from such studies could be erroneous and hence invalid. Due to the limited knowledge of many physicians and clinicians in quantitative and qualitative techniques, errors from experimental studies as a result of incorrect subjects' selections are not uncommon [1,2].

In order to understand background knowledge of clinical and experimental studies, some basic concepts in sampling or survey methodology need to be explained.

Basic Concepts in Sampling Methodology

Population

This is defined as a group of individuals, objects, items or entities with common characteristics or attributes. Population in the statistical concept does not only entail human subjects [3]. It could mean other inanimate units once it is well defined such as a number of stethoscopes produced by medical device plants. Mathematically, it is stated as "***N***". In medical research, the population could be individuals suffering from Cardio Vascular Diseases (CVD) or people undergoing chemotherapies for cancer treatment. It could be the total number of child deliveries recorded in a city. Since it is practically impossible most time to obtain information from everyone e.g., all cancer patients in a specific setting such as a country or a continent, the need to obtain information from certain numbers of such patients for inference is necessary. A population can either be finite or infinite [4].

Citation: Etikan I, Babatope O. A Basic Approach in Sampling Methodology and Sample Size Calculation. Med Life Clin. 2019; 1(2): 1006.

Publisher Name: Medtext Publications LLC

Manuscript compiled: October 10th, 2019

***Corresponding author:** Ilker Etikan, Department of Biostatistics, Near East University, Faculty of Medicine, Nicosia-TRNC, Cyprus, E-mail: ietikan@gmail.com; ilker.etikan@neu.edu.tr

Finite population: A population in which its subjects or items are exhaustible and countable is called a finite population. The total number of nurses or medical doctor in a hospital is an example of a finite population since they can all be counted.

Infinite population: A population in which the number of items or subjects cannot be exhaustively counted is called an infinite population. An example of this could be cancer cells or red blood cells in the body.

Sample

Sample could be termed a subset of a population. They are representatives of the population [5]. Through sample, the conclusion is generalized to the population. Informed decision making in medicine and public health is through samples that are considered true reflection and representatives of the population. Sample is mathematically represented as "***n***". The method of selecting a sample from the population is called sampling. There are major factors that can influence sample representations in research studies. Some of the factors are the size of the population, the sampling techniques, and the responses.

Reason for selecting sample: There are quite a number of reasons for selecting samples from the population. Examples include:

Time: Some clinical studies are time-bound and therefore exigent. Trying to obtain information from all concerned subjects (population) could span beyond the allocated time frame. Thus, the need to consider sample selection.

Finance: Another reason to consider sample selection for decision making in clinical studies is the advantage of cost-effectiveness. More financial resources are often required if all members of the population are required.

Human resources challenge: For many public health kinds of research, there could be limited human resources having the requisite understanding of those studies entails. Hence the available workforce could only be limited to cover a number of subjects optimally.

Location challenge: In order to confront the problem of limited access to the population of interest as a result of geographical constraints sample selection becomes essential.

Population size: Often time, getting to obtain information from all subjects or items of interest could be cumbersome and drudgery especially if the population is so large. Hence, obtaining a sample from such population is considered the best alternative.

Sampling frame

It is a list consisting of all population elements from which sample can be drawn. A list of older people in an aged support home, register list of deaths or live births are all examples of the sample frame.

Sampling techniques

The method through which a sample is selected from the population is called sampling. Sampling techniques play a critical role in clinical and any other experimental studies [6]. There are two broad divisions of sampling techniques and they are probability sampling and non-probability sampling [7,8].

Probability sampling

This is a sampling method in which every unit or subject in the population has an equal chance or probability of being selected. Samples selections are done randomly [9]. Probability sampling is considered the best form of sampling due to the equal representativeness of the sampling units [10,11]. Examples of probability sampling techniques include: Simple random sampling, systematic sampling, cluster sampling, and stratified sampling (Figure 1).

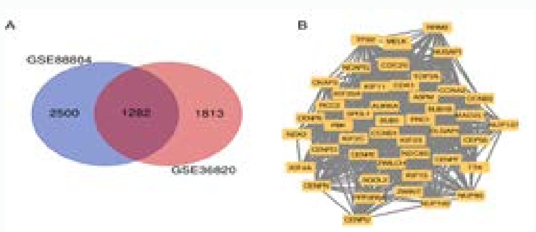

Figure 1: Probability Sampling.

Simple random sampling (SRS): This is the simplest form of a probability random sampling method. When selecting a sample, each observational unit/subject has an equal chance of being selected. This method is most suitable when the population is homogenous with fairly small population size. SRS can be done through a lottery system or allotment of random numbers to population units from a random number table or a computer-generated numbers algorithm to experimental units after which samples that matched some random numbers are selected [12]. A simple description of SRS is given below:

Suppose a hospital has 200 Nurses and the management board is interested to sponsor 20 to an international conference.

- **First Step:** A name list of all the 200 Nurses is needed to be made. The whole 200 Nurses is the population (N)
- **Second Step:** A sequence numbering is assigned to each of the 200 Nurses (say from 1,2, 3…..200). This list is called the sampling frame.
- **Third Step:** Having confirmed that 20 Nurses are needed, i.e. sample size (n=20). By using a random number software generator, 20 random numbers can be generated that are found between 1 through 200. The 20 random numbers generated that matches the assigned numbers in the second step becomes the sample to be selected.
- The merit of this method is that bias is almost non-existent and the researcher might not necessarily have the knowledge about the subjects/units. Also, it is quite basic with no knowledge of high quantitative computations. It is quite easy to get a sample size from it since the population is large.

A major demerit of this method is that it could be strenuous with high financial implications especially if the population is so large. This method can only be carried out provided that there is a sampling frame. Without a sampling frame, it is quite impossible. It is a fair method of selection but however, this selection could be deemed unfit especially if the eventual selected subjects or units do not meet the envisaged objectives. An example is the random selection of Nurses who do not have requisite and qualified knowledge that ought to be a base requirement for entry into an advanced training technique course leaving those qualified behind in the pool.

Stratified sampling: If a population consists of various distinct groups, they can be grouped into sub-categories called Strata. After which a random selection of a sample from each stratum (singular form of Strata) can be carried out. Thus, each stratum is an independent sub-population cohort of the generation population with each consisting of unique or homogenous group classification. Each element in the stratum thus has an equal chance of being selected. As an example, Medical officers in a hospital can be divided into categories relative to years of experience, sex, race, departments or types of institution graduated from (private or public medical schools).

In terms of years of experience stratification, those having 0 years to 3 years of experience can be the first category (first stratum), while those with 4 years to 7 years of experience could be the second stratum and those with above 7 years could be grouped in a third category (third stratum). The strata thus become more homogenous in nature based on years of experience. Therefore, a random sample selection of medical officers will be done from each of the three strata formed.

The advantage of a stratified sampling is that it ensures a fair and even selection of units that vary across different groups or classification thus reducing sampling bias. However, the drawback for this method is that it requires a sampling frame for each stratum and it could also be difficult when selecting the criteria for stratification.

Systematic sampling: This is a probability sampling technique in which every k'th element in a population list is selected sequentially [12]. To mitigate bias, it is advised that the first element selection should be at random while subsequent selection will follow every k'th element ordering. Given that an NGO volunteer register list consists of 2000 individuals, and 100 participants are required to be selected for a program participation, the first person can be selected randomly (say individual on 9th position on the list), then every 10th individual from the first selected individual will be selected till the 100 participants are completed. In the example above, after the 9th individual on the list was selected, the second selection will be the individual in the 19th position, followed by the individual on 29th position and so on till the 100 participants limit is reached.

The advantage of this method is that is simple and selection procedure is easy. Sample selection is evenly distributed over the population. The disadvantage is that if there is a hidden pattern in the list, there could be a problem of overrepresentation and sample selection can be compromised. Example of this hidden pattern could be admission of nursing students into an academic institution. If the school policies favor admission of students from the locality before admission consideration for an outsider, the school register lists could be filled with local students before others on the register list. Using a systematic sampling on such a register list for a research study that requires student perception of the school admission policy could be sentimental and compromised.

Cluster sampling: This is one of the most popular sampling techniques used in epidemiological researches. However, it is often confused with the method of stratified sampling but there is a unique difference between the two sampling methods. In Cluster sampling, there is a division of the population into groups, and these groups are called clusters. Then, some clusters are randomly selected. Each unit or subject found in the randomly selected clusters is then totally included in the sample needed. Whereas in stratified sampling, there is only a random selection of elements from each of the strata to form the required sample.

Example: Given that a town is selected for an immunization intervention program for children aged 5 years and below and the town is made of ten (10) districts. The ten (10) districts translate to ten (10) clusters. Given that three (3) districts are randomly selected out of the ten (10) districts, the selected three (3) districts are the clusters to be considered for the program and all children aged 5 years and below in these three (3) selected districts are included in the sample for the intervention program.

Cluster sampling becomes more effective if the clusters constituents are heterogeneous in nature unlike in stratified sampling in which the strata are homogenous. A major advantage of cluster sampling is cost-effectiveness and timeliness but drawbacks include a high margin of sampling error.

Non-probability sampling

This is a sampling technique in which every object or experimental does not have an equal chance or probability of being selected [13]. As a result of the skewed selection of subjects/units, they are often highly susceptible to bias and other forms of selection errors. This sampling technique is based on Researcher preference and discretion. Examples of non-probability techniques are quota sampling, purposive sampling, convenience sampling, and snowball sampling (Figure 2).

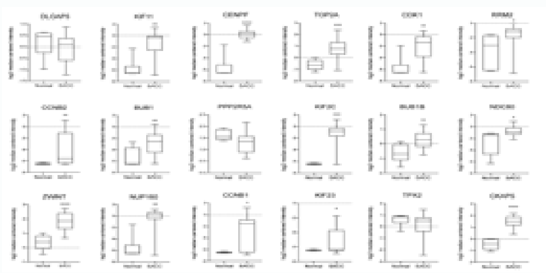

Figure 2: Non- Probability Sampling.

Quota sampling: This method is widely used in polls and considered the non-probability version of stratified sampling. The population is divided into unique distinct groups of units/objects and a researcher makes selections from each of the distinct groups based on their selection criteria, proportion or preference. Suppose a researcher is expected to obtain the opinion of five hundred (500) women on the use of contraception relative to the socio-economic level, the researcher can decided to go to a low income community and select 250 women based on his judgment and equally replicate such exercise in a high-income neighborhood by selecting another 250 women.

Quota sampling is useful where there is a time limitation, low budget and a sampling frame is not available. However, it is very prone to selection bias since each selected subject does not have an equal chance of being chosen.

Purposive sampling: This is also called a judgmental or subjective sampling method. This is a method of non-probability sampling based on the knowledge and understanding of a researcher in selecting the needed sample from a population for a study. A researcher in this study reach out to participants he felt can meet up with the objective of a study or subject of interest the study is investigating. Individuals with no knowledge of study objectives are not selected. This method is commonly used in qualitative research studies and focus group discussions in which experts in the subject of interest are selected based on the expert experiences and knowledge inclinations.

The advantages of this sampling technique are that it is time and cost-effective. It makes it easier to narrow to subjects of interest. The disadvantages of this method are that the researcher is prone to make a subjective judgment and thus increase the bias on sample selection. It is not always reliable and the result of the study might not be generalized to the population.

Convenience sampling: This is a sampling selection based on the accessibility of respondents within reach. It can also be called Accidental sampling. Subjects or units are merely selected because they can easily be found around and the researcher has regular access to them. Examples of this sampling method could entail surveying friends, neighbors or families, moving across the corners of the streets to ask for volunteers for their opinion, online polls and so on. Convenience sampling is often used as a pilot study in order to gain insight before a full-fledge research activity takes place. It is a very simple and easy sampling technique. When used as a pilot study, it enables researchers to reframe the area of research questions that lacks clarity and add up other insightful questions in a bid to generate effective responses. It is time-efficient and requires low financial expenditure to carry out. However, it is dangerous because subjects or respondents that are actually fitting for a research objective might not be the set of respondents that were eventually selected for the study. Hence, results from such studies can be highly flawed with errors and deviation from the research objectives.

Snow ball sampling: This sampling method is commonly used in very sensitive culturally studies and also used in situations of the rarity of experimental subjects/respondents or when they are very hard to reach. It is based on a referral mechanism in which identified subjects help to reach out to other unknown subjects. Examples of this sampling technique could be studied on public health issues such as research on the consumption of illicit drugs, HIV/AIDs infected individuals, or victims of sexual abuse.

This method lowers the research cost of selecting a study's sample but could be the plague with a lot of bias as a selected sample may not be true representatives across the cross-section of participants required.

Sampling and Non- Sampling Errors

Errors need to be taking into account when conducting research studies. This is essential especially when such studies are imperative for the general population. There are two major types of errors in study and they are sampling and non-sampling errors.

Sampling error

The error that occurs due to the use of a sample rather than the whole population is called Sampling Error. This error is a deviation from the actual measure, trait or attributes of true entity. The major cause of sampling error is that subjects are drawn out as sample from the population and these drawn-out subjects could as well have their own individual inherent variability. Aside from this, the use of a defective or wrong sampling method could result in this type of error. Two major factors that influence the extent of sampling errors are the sample size and sampling technique. In quantitative studies, the confidence limits, standard error, p-values, co-efficient of variances are used as measures of sampling errors. As a result of sampling error, researchers are always advised to use the most suitable technique in sample selection that fits the study objectives and also makes use of sufficient sample size in order to reduce the magnitude of sampling error. Likewise, randomization during sampling is encouraged.

Non-sampling error

All other forms of errors that occur apart from sampling error are called non-sampling errors. These errors could occur as a result of several factors which could stem out from the instrument used in data collection, the subject/respondents or the individuals that collect, collate, sort, analyze and presents the data output. In terms of the instrument, the instrument might not be well specified. If the instrument is a questionnaire, the questionnaire might not consist of appropriate or clear questions that align with the research objectives. The questions and instructions might also be difficult for respondents to understand thus leading to inappropriate responses.

Errors from respondents could be as a result of mood/feeling, non-response, deliberate answer omissions, false responses, sentiments, fatigue and so on. Errors can also occur when collating or sorting data by research staff. There could be wrong data imputations during analysis and usage of wrong statistical or mathematical methods. Results interpretation could be wrongly typed or organized appropriately (Figure 3).

Sample size calculation

Many studies are being conducted without taking cognizant of the sample size effect. Most researchers/investigators with no quantitative background often find this as a major barrier in their research studies. After setting a research objective and identifying the appropriate respondents, a major step to consider again is getting an appropriate sample size for the study.

Getting appropriate and adequate sample sizes of respondents who are randomly selected helps reduce sampling error or biases in researches. The adequacy of the sample size is not just about the proportion to the population but also takes into consideration the selected sample in at last to the diversity existing in the population, the objectives of the investigators as well as the statistical modeling techniques to be employed [14]. Though it is established that larger sample size that nears the population diminish sampling errors and increases the result validity [15], however, there is a tendency that seeking to get more samples that happen to exceeds an appropriate computed size could overstrain resources of investigators [16].

There are several sample size calculating methods relative to the objectives of the studies and study design. However, a common formula for calculating sample size in survey studies from a finite population (countable population) is given below.

$$n = \frac{N * X}{X + N - 1}$$

Where

$$X = (Z_{\alpha} / 2^2 * P(1 - P)) / MOE^2$$

n= Sample size; p= Proportion of sample; MOE= Margin of error; N= Population size

Z-(α/2) =the critical value of the normal distribution at a α/2 (for a confidence interval level of 95%, α is 0.05 and the critical value is 1.96). This entails the required level of confidence for the estimate.

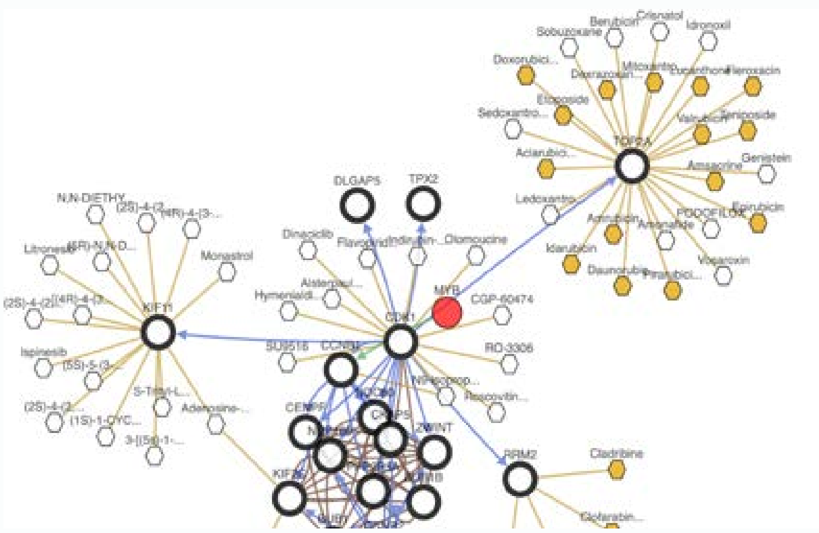

Figure 3: Example of Non-Sampling Errors.

MOE indicated the amount of error allowed. A 5% margin of error is usually selected. P is the proportion or percentage of the sample having a definite attribute of interest. For instance, if 60% is the percentage of people suffering from diarrhea, (100-P) which translates to 40% is the proportion or percentage of people not suffering from diarrhea. In getting the proportion prior to a study, [17] posited that investigators make use of 50% (0.50) as the proportion value in a bid to maximize variability change and optimization of an estimated sample.

Example: Suppose a Town consisted of 3510 children that are aged 5years and below. Estimate the sample size required to conduct a program for children less than 5 years old aimed at testing vaccine effectiveness.

Solution:

Thus, N=3510, p= 0.5, MOE= 0.05, and z_ (α/2) = 1.96

$$n = \frac{N * X}{X + N - 1}$$

$$X = (Z_{\alpha} / 2^2 * P(1-P)) / MOE^2$$

$$X = \frac{1.96^2 * 0.5(1-0.5)}{0.05^2}$$

X= 384.16

$$n = \frac{3510 * 384.16}{384.16 + 3510 - 1}$$

$$n = \frac{1348401.6}{3893.16}$$

n=386

Thus 346 children are the minimum required sample size for this study.

Conclusion

In conclusion, the importance of sampling in carrying out research cannot be overemphasized. The need to identify the most suitable sampling method is important as this plays an effect on the level of error recorded. A large sample size that nears the population size can provide a better estimate that can be generalized to the population. In reducing the risks of non-sampling errors, researchers are urged to take into account modalities that can effectively ensure valid studies from tested and review instruments of data collection to the use of qualified enumerators and other workers involved in data analysis and result preparations.

References

1. Taylor AA, Byrne-Davis LM. Clinician Numeracy: The Development of an Assessment Measure for Doctors. Numeracy. 2016;9(1):5.
2. Fuji KT, Galt KA. Research skills training for the Doctor of Pharmacy in US Schools of Pharmacy: a descriptive study. Int J Pharm Pract. 2009;17(2):115-21.
3. Banerjee A, Chaudhury S. Statistics without tears: Populations and samples. Ind Psychiatry J. 2010;19(1):60-5.
4. Kozak M. Finite and Infinite Populations in Biological Statistics: Should We Distinguish Them? J Am Sci. 2008;4(1):59-62.
5. Hanlon B, Larget B. Samples and Populations. 2011.
6. Suresh K, Thomas SV, Suresh G. Design, data analysis and sampling techniques for clinical research. Ann Ind Acad Neurol. 2011;14(4):287-90.
7. Elfil M, Negida A. Sampling methods in Clinical Research; an Educational Review. Emerg (Tehran). 2017;5(1):e52.
8. Shorten A, Moorley C. Selecting the sample. Evid Based Nurs. 2014:17(2):32-3.
9. Fowler FJ. Sampling. Survey Research Methods. 4th ed. Thousand Oaks: Sage Publications; 2009. p. 19-47.
10. Thompson SK. Sampling. 3rd ed. Hoboken NJ: John Wiley & Sons Inc; 2012.
11. Curtin R, Presser S, Singer E. Changes in telephone survey non response over the past quarter century. Public Opinion Quarterly. 2005;69(1):87-98.
12. Babbie E. The Logic of Sampling. The Practice of Social Research. 10th ed. Belmont: Hadsworth/Thomson Learning; 2004. p. 178-217.
13. Etikan I, SulaimanAbubakar M, Rukayya Sunusi A. Comparison of Convenience Sampling and Purposive Sampling. Am J Theor Appl Stat. 2016;5(1):1-4.
14. Taherdoost H. Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. Int J Acad Res Manage. 2016;5(2):18-27.
15. Philip S. What is sampling error?. BMJ. 2012;344:e4285.
16. Gill J, Johnson P. Research Methods for Managers. 4th ed. Thousand Oaks: Sage Publications; 2010.
17. Bartlett JE, Kotrlik JW, Higgins CC. Organizational Research: Determining Appropriate Sample Size in Survey Research. Inf Technol Learn Perform J. 2001;19:43-50.