Populations and samples: Often misused concepts in Social Science researches

Saman Asawapoom
Graduate School, Sisaket Rajabhat University, Thailand
samubon@hotmail.com

ABSTRACT

Research populations and samples are basic concepts that research learners and practitioners must know and apply them correctly. Research report investigations revealed that understanding and application of the two concepts were incorrect in many cases. Consequently, serious mistakes are the applications of inferential statistics. With misunderstanding on research populations and samples, it leads to misuse inferential statistics in hypothesis testing as well. Based on the state and problems, the writer reckons the importance of revision and presenting these concepts in all dimensions, from the problem background, understanding research population, research samples, referential statistical concepts, and finally, the conclusion and implication to research conducting. The writer hopes that this article would clarify both the concepts and their applications.

Keywords
Research samples, Research population, Missed concepts, Social Science

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

Being asked which is better between "Conducting the research using the population and the sample"? Without hesitation, the answer would be "using the population" because the ultimate goals of most researches are to find out the characters of the parameters. However, in reality, it is hardly possible because most populations under investigation are usually comprised of large numbers of subjects. So, most of researches have been done by using the samples drawn or selected from the target populations. However, the ultimate goal of the research is still to obtain the characteristics of the population under study. In order to do so, the proper technique of sample sampling and inferential statistics come to play important parts. At the same time, the techniques cause a lot of problems, too. This article is trying to review and propose the implication of the concepts.

Understanding Research Populations

The notation of research population is a common concept among researchers and academics. However, the word populations in lay languages might mean different things. In research language, the population could be just about anything under study, such as people, events, insects, and any others. Whatever the population of interest for research might be, it could be under one of the three following concepts: universe, target population, or treatment population (Runyon, Haber, Pittenger, and Coleman, 1996; Fraenkel and Wallen, 2006; Bordens and Abbott, 2011; Bryman, 2012).

Universe or ultimate population refers to all subjects or elements that compile as the population of a kind. The problem is that sometime, the population is too large to count and locate all the members. Or sometime, not all members of the population could not be located and registered and be counted as practical subjects of the research population. So, when a researcher conducts a research, he or she needs to define a population boundary for that research as a target population.

Target population, or simply research population, refers to an accessible list of all the subjects of the interested population under study. If all subjects of the research population can be listed, then, it is called the finite population and a probability research sample could be drawn by using a probability sampling technique. But if all subjects of the research population are hardly defined, located, and listed, it is called the infinite population. In case of infinite populations, non-probability sampling techniques are applied. Either kinds of sampling demand different techniques of inferential statistics which can be generalized to the target population. In some cases, the researchers might conceptualize the target population by assuming a subdivided target population, compiled of the research subjects' common characters. In Abbott's example (Bordens and Abbott, 2011), a researcher wants to study the effect of a new computer-based teaching technique on eighth- grade students. The entire population of all eighth-grade students would be too large to define and access. So, the researcher would assume that eighth-grade students enrolling in a certain group of high schools under the jurisdiction of an educational area as the target research population, and then a sample would be drawn from that defined population.

Treatment population, as Runyon, Haber, Pittenger, and Coleman (1996) pointed out, is a hypothetical entity that does not really exist except when the experiment is conducted. The treatment population enables the research to draw a study sample and then conduct random assignment to study group samples. Treatment population may also be recognized as hypothetical population since it occurs only as the population when experiment research is conducted based on a hypothesis or some hypotheses. This definition of research population could be very useful for experiment

researches which are most problematic techniques of this type of researches in Thailand.

Based on the concepts discussed above, defining the population frame is a crucial task of research design because the finite population provides probability of a probability sample which enables the research's possibility to generalize the sample statistic results to the population parameters by using inferential statistics. In case of infinite population members, the research has to choose either nonparametric sampling technique or just report the descriptive statistics, or graphs or content descriptions. However, if the researcher decides to select the sample by using nonparametric sampling, it is necessary to use nonparametric inferential statistics to test the hypothesis. On the other hand, if probability sampling is used, it will enable the researcher to test the research hypothesis by using parametric inferential statistics to generalize the research findings to the population understudied.

Understanding Research Samples

As mentioned above, samples are used in conducting researches more often than the whole population although the study from the populations is far better than from the samples. A sample is simply a small subgroup chosen from the larger population as the best representative subjects for the study. The problem is how we can obtain those best representative subjects. In practice, it is not always possible to get the best representative subjects for the study. However, if the appropriate sampling technique is applied in obtaining the sample, it is likely that the best representative sample is obtained for the study. Reading books on research, many sampling methods or techniques are suggested. In summary, there are two main methods of obtaining research samples, Random Sampling Method and Nonrandom Sampling Method (Fraenkel and Wallen, 2006).

Random Sampling.

Random Sampling, also known as Parametric Sampling, is a process of obtaining research sampling subjects based on probability theory, meaning every member of the population presumably has an equal chance of being selected. Although, there are different methods of drawing a random sample, but all methods have the same objective, that is to select a representative sample from the population. Reviewing most textbooks, four sampling methods come to highlight (Cresswell, 2008; Alder and Clark, 2011, Fraenkel and Wallen, 2006, Bryman, 2012; Mertens, 2010), but since random assignment is also treated as random sampling as well (Runyon, Haber, Pittenger, and Coleman, 1996), so in this article, the writer proposes 5 methods of random sampling methods as fallows.

Simple Random **Sampling**.

Simple Random Sampling, as the name suggested, it simply means that every research sample subject is drawn directly from all population members. Two main simple random sampling techniques are replacement Sampling and Non-replacement Sampling. Replacement Sampling Technique is done by returning the previous drawn sample subject back to population every time the subject has been drawn before drawing the new subject while Non-replacement Sampling technique will not return the drawn subject. By returning the previous drawn sample subject, it is believed that every subject has equal chance or probability of being drawn as the research sample. Similarly, though not the same, it is nearly equal chance of being drawn as the research sample when Non-Replacement Technique is applied. However, either case, it is qualified as Random Sampling or Probability Sampling.

Stratified Random Sampling.

Another random sampling method that is often used by researchers, especially in the case of the population could be subdivided into subgroups or stratums, which are the sources of its name. The first step of using Stratified Random Sampling, the researcher must identify stratums of the target population that will be used main sample subgroups of the sample. The main stratums refer to subgroups that could cover all other subgroups used as research independent variables or studied variables. Example, when a researcher is planning to study about administrators' attitude towards professional royalty. School sizes where the administrators work could be used as the main stratum of the target population because each school size or stratum embeds all other characters of the subjects such as genders, education qualifications, and terms of service in the position.

After the researcher identifies the subgroup variables to use as the main stratum, the next step is to draw the research sample subjects, proportionately or disproportionately, from each stratum by using Simple Random Method. So, when the researcher uses this method to draw the research sample, he or she does not need to state both Stratified Random Sampling and Simple Random Sampling because Simple Random Sampling is a built-in technique of Stratified Random Sampling already. However, there are two types of Stratified Random Sampling, proportionate and disproportionate stratified random sampling, as mentioned above. With proportionate random technique, the number of elements or subjects from each stratum in relation to its proportion of total population while disproportionate stratified random technique, the proportion of the sample subjects is not conformed to its population or reversed to its proportion of total population, meaning the less numbers of the population members the more sample subjects would be drawn. Disproportionate stratified random technique, usually, is applied in cases of the members of each stratum varies a lot. For example, a numbers of large size schools are 10 schools, while the medium size schools are 50, and the small size schools are over a hundred.

Cluster Random Sampling

Cluster Random Sampling is applied, usually, when the population is large and difficult to locate sample units. So, the researcher has to identify and divided the population into groups and each group proposes the similar characteristics, called clusters. Clusters are not the same as stratums, mentioned in Stratified Random Sampling, in the

sense of compositions of their members. Members of the stratum consist of similar characters of the subjects while members of the cluster consist of representative members of all characters of that population. For example, if the members of the population include 100 apples, 150 oranges, and 300 pears. The members of each cluster must include some apples, some oranges, and some pears. The first and second cluster, each might consist of 33 apples, 50 oranges, and 100 pears, and the third cluster consists of 34 apples, 50 oranges, and 100 pears. Any of these three clusters could be drawn as a research sample and whatever cluster obtained will be the representative sample of the population under study

Multi-stage Random Sampling.

Multi-stage Random Sampling comes to use in case of a large population as well, but the clusters of the population can be subdivided into many levels, for example, national level, regional level, province level, and district level. When a researcher comes across the situation like this, the researcher should apply this technique by firstly draw the sample of national level, then regional, provincial, and finally district level. From the district level, the research will consider it as target population for the study and then apply either Simple Random Sampling, or Stratified Radom Sampling as the next step to draw the research sample. To illustrate the whole process of using Multistage Random Sampling as a sampling method to draw the research sample, see how it is done in the following example.

Suppose that a research wants to conduct a survey on how all secondary school administrators in Thailand perceive attitude towards professional certification. The researcher, firstly, files the list of all secondary school administers in Thailand, grouped into national level, regional level, provincial level, and district level. Let's say that there are five regions in Thailand, and then researcher might draw 20 percent of these five regions as an initial target population. The initial target population is "one region", and from this region, the researcher draws 10 percent of all the provinces in the region as the next target population. Assumingly, the numbers are three provinces. Next step, the researcher draws 5 percent of all districts in those three provinces. From these five percent of districts drawn in the last step, the research come to numbers of all secondary schools in those districts, then the research might apply Simple Random Sampling, or Stratified Random Sampling to draw the research sample for that study, finally.

Random Assignment.

Radom Assignment is a sampling method applicable mostly for experimental researches since it is likely impossible to draw a lot of sampling subjects for experimental research. At the same time, the researcher wants to generalize the findings to interested target population. The concept of Random Assignment as Runyon, Haber, Pittenger, and Coleman wrote in their book (1996), it could be summarized that Random Assignment is a process by which research subjects are assigned to experience a level of the independent variable. The goal of Random Assignment is to make sure that the experiment groups are comparable before the experiment begins. The final step of Random Assignment is that the researcher may arrange research subjects into experiment group and control group by using Simple Random Sampling one by one, or drawn as a cluster. By doing this, random assignment is considered as Parametric Sampling as well. In case of one experiment group, all the subjects would be used as experiment group. However, the more crucial step in obtaining the experiment subjects is not how to conduct random assignment, the writer thinks, but how the population is defined.

As mentioned early in this paper on the concept of population, it could be concluded that there are three main types of populations. The first type is the population that could be defined and all the subjects could be listed, the second type is the population that could be roughly defined but not all the subjects could be identified or accessible, and the third type is treatment or hypothetical population. In conducting experimental research, it is hard to drawn the sample subjects from the first two types of population. However, the best that the researcher could do is to conduct random assignment, so that there would be no questions of generalization and the use of probability inferential statistics. Therefore, the experimental research should be conducted the experiment on the treatment or hypothesis population.

Runyon, Haber, Pittenger, and Coleman (1996) pointed out that theoretically we could toss coins infinite numbers of times and count those as the population. But would anyone, be crazy enough, to do that? When we want to test the possibility of tossing coins, we would use 10 or twenty tossing coin events to estimate the population events of coin tossing. It is theoretically true as well when the researcher is interested in using AIDS patients as subjects of experiments. No one could ever imagine how to define all AIDS patient subjects in the population, so he or she uses treatment or hypothesis population, draws a sample and then random assign the subjects to experiment groups. For example, the researcher is interested to test a teaching method on improving a first year of primary students' learning. First, the researcher would reason that the first year of primary students in her school, or in her school district, are treatment population of similar first year primary school students elsewhere. After that she or he would draw a sample from all students in her or his school, or school district. Finally, random assignment would be used to assign the subject to the experimental group or groups, depending on how the research is designed. This is how random assignment is useful in experiment studies.

Nonrandom Sampling

Nonrandom sampling is also known as non-probability sampling, as pointed out in Bryman's book (2012) is essentially an umbrella term to capture all forms of sampling techniques that are not conducted according to the canons of probability sampling. Some of these methods include Selecting Informants or Criterion Sampling, Quota Sampling, Snowball Sampling, and Convenience Sampling (Babbie, 2011; Cresswell, 2008; Alder and Clark 2011; Mertens, 2010; Bryman, 2012; and Fraenkel and Wallen, 2006). These methods would be summarized as follow.

Qualification-based Informants

Qualification-based Informants is a method of selecting research informants based on their qualifications who are appropriated for the purpose or purposes of the research. This method is also known as "Criterion Sampling" because the informants would be selected according to some criterion suitable for those researches. For example, the purpose of the research is to investigate how mobs were set up and the effective of demonstration during the certain time. The criterion might include (1) mob leaders, (2) key activists who were injured during the demonstration, (3) former government officials during the event, and (4) field media at the time. The researcher might select 5-7 informants from each group for interviewing or focus group meeting to obtain information related to the topic of interests.

Quota Sampling

Quota Sampling is the intension of obtaining the research subjects included in the sample as equal number of each character dimension needed for the research. For example, the sample subjects might be the equal numbers of males and females, or equal number of parents from each economic level class. After the quota is set, what the researcher needs to do is to obtain research subjects of each group of quota to provide information or characters needed for the research.

Snow Ball

Snow Ball Sampling Method is very useful in the situation where research subjects are not readily available and hard to identify. The researcher might have to start with an initial contact on the research area of interest to identify other potential participants. For example, in case of the research would like to investigate the original and development of "Mor Lum Sing" (Modern Isarn Folk Music and Singing) the researcher might start with one of Mor Lum Sing teacher, of whom the researcher classified to be one of the expert of the kind. After that the research would ask the first informant to recommend other potential participants for the research. The process would go on and on like this until the researcher finds that the list starts circling around the same names of informants or at least the researcher has sufficient information for the research topic. Then the Snow Ball Sampling Process ends.

Convenience Sampling

Convenience Sampling is often misunderstood that the researcher might just select any research subjects to include in the study. In fact, the researcher should start with logical reasoning why such a group of subjects is suitable and meaningful for the research. For example, the researcher might gather information from any customer coming to shop at a convenient store. But first of all, the research might have to set up a quota sampling based on every other day and different periods of time to obtain the information so that the subjects sampled would be representatives of the population as much as possible although the population might be infinite. This method might be called "Quota-Convenience Sampling." Another example might be the case of a professor decides to use the students who enroll in his class as subjects of his research. This seems to be convenient and easy enough to choose the research subjects. But in fact, if the professor wants his study be generalized to wider population, he needs to reason logically that the group of students is much similar to other classes at his university or university in his town. Moreover, the researcher has to use random assignment technique to assign his subjects to the experiment research design he uses in that study in order to signify study as true experiment study. So, Convenience Sampling is not just convenient as it sounds, but another very sophisticate method of sampling as well.

Last, but not least, on this subject is that some novice researchers misunderstand between "research sample" and "research subject". For example, when the researcher wants to study the administrators' attitude towards administrational profession, the population might be 550 school administrators in a certain region. A sample used in this study might include 226 school administrators in that region, drawn by means of Simple Random Method. The whole group of representative school administrators used in this study is "the research sample" while each school administrator is "the research subject". Most researches are conducted by using a single sample, and the sample might include one subject or many subjects. So, when the researcher mentions about the sample, it is a singular noun except for conducting the research of two or more target populations. In such case, there would be more than "one sample", of course.

Data Analysis and Utility

The writer includes data analysis and utility in this article, although it is not a type of sampling, but it is related to samplings. Mainly, the concepts of populations and samplings mentioned above are more applicable to quantitative researches than qualitative ones. Data analysis and its applications, here by followed, would be on quantitative researches as such. For the purpose of this article, method and calculation of statistical analysis will not be included, but focused on application of the analysis.

According to my investigation on statistical methods used in researches, I found most text books written on descriptive statistics, probability and nonprobability inferential statistical methods and usages, but rarely include the whole thing together (Jackson, 2009; Sapsford and Jupp, 2008; Glass and Stanley, 1970; Black, 2011; Brase and Brase, 1987; Runyon,Haber, Pittenger, and Coleman, 1996; Anderson, Sweeney, and Williams,1996; Haeussler, Paul, and Wood, 2011; and Siegel, and Castellan, 1988). However, it could be concluded into a basic rule that if the researcher want to generalize the research findings based on the sample to its population, the right kind of inferential statistic must be applied.

Probability Inferential statistics can be applied only when the sample is drawn by using probability sampling techniques. And it is equally true that nonprobability inferential statistics would be more applicable in the case that the sample is drawn by using nonprobability sampling.

However, for descriptive statistical analysis and report, the descriptive statistics could be used with any types of sampling, probability or nonprobability samplings, followed by appropriate scales of measurement (Nominal, Ordinal, Interval, or Ratio Scale) and the purpose of data analysis. Details of the concepts and methods could be reviewed from most research and statistic textbooks. The purpose of mentioning about data analysis and utility is to point out the caution of misuses of statistics with data obtained from different types of sampling.

One of the most serious mistakes in using statistic techniques on data analysis is that some researchers apply "testing hypothesis statistics" on data obtained from the population. Once the researcher studies from the population, the data obtained are parameters. So, there is no need to use inferential statistics to generalize the findings to the population.

Conclusions and Implications

Although population and sampling are the two concepts commonly introduced to the new research learners and emphasized in most textbooks on the subject, but numbers of researchers often make mistakes applying them. The most common mistake is misunderstanding the sampling concepts, but the more serious mistake is to use the wrong statistical technique to analyze obtained data from certain types of samplings, as mentioned above. Some researchers mixed up concept of Purposive Sampling with Qualification-based Sampling or Convenience Sampling. More serious mistake is that when researchers use probability inferential statistics on the data obtained from purposive sampling. As seriously mistaken as what has been mentioned above is applying inferential statistics on data obtained from the population. The writer hopes that this article could clarify how research populations are defined, the different types of samplings are distinguished, and the most of all is the correct use of statistic for analyzing different types of sampling. However, readers need to review and learn more of technical concepts and procedures to apply the concepts presenting in this article more practically.

References

[1] Alder, E.S. & Clark, R. (2011). An invitation to social research: How it's done (4th ed). Australia: Wadsworth Cengage Learning.

[2] Anderson, D.R., Sweeney, D.J., & Williams, T.A. (1996). Statistics for business and economics (6th ed.). Minneapolis/St Paul: West Publishing Company.

[3] Babbie, E. (2011). The Practice of social research (11th ed.). Australia: Thomson.

[4] Black. K. (2011). Applied business statistics: Making better business decisions (6th ed.). Asia: John Wiley & Sons (Asia).

[5] Brase, C.H. & Brase, C.P. (1983). Understandable statistics: Concepts and methods (3rd ed.). Lexington: D.C. Heath and Company.

[6] Bordens, K.S. & Abbott, B.B. (2002). Research design and methods: A process approach. (8th ed.). The United States of America: McGraw-Hill Companies.

[7] Bryman, A. Social research methods. (4th ed.). (2012). Oxford: Oxford University Press.

[8] Creswell, J.W. (2008). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (3rd ed.). New Jersey: Edwards Brothers.

[9] Fraenkel, J.R. & Wallen, N.E. (2006). How to design and evaluate research in education (6th ed.). Boston: McGraw-Hill.

[10] Glass, G.V. & Stanley, J.C. (1970). Statistical methods in education and psychology. New Jersey: Prentice-Hall.

[11] Haeussler, E.F., Paul, R.S. & Wood, R.J. (2011). Introductory mathematical analysis: For business, economics and life and social sciences (3rd ed.). Boston: Prentice Hall.

[12] Jackson, S.L. (2009). Research methods and statistics: A critical thinking approach.

[13] Australia: Wadsworth Cengage Learning.

[14] Mertens, D.M. (2010). Research and evaluation in educational and psychology: integrating diversity with quantitative, qualitative, and mixed methods. Los Angeles: Sage Publication.

[15] Runyon, R.P., Haber, A., Pittenger, D.J., & Coleman, K.A. (1996). Fundamentals of behavioral statistics (8th ed.). Boston: McGraw-Hill.

[16] Siegel, S. & Castellan, N.J. (1988). Nonparamettric statistics for the

behavioral sciences (2nd ed.). New York: McGraw-Hill.

[17] Sapsford, R. & Jupp, V. (ed.) (2008). Data collection and analysis. Los Angeles: Sage Publication.