

Non-probabilistic samples are just as valid as probabilistic samples. We use them for a number of reasons. However, since they are **NOT** probabilistic, they do pose one great restriction on your ability to generalize research findings: You **CANNOT generalize statistically** from a non-probability sample. Otherwise, you can use a non-probability sample in all the same ways that you can use probability samples. You **can** perform statistical analyses. For example, you could use the t-test to determine if there are differences between treatment and comparison groups. However, you could not then extend your findings beyond the sample -- you could not generalize your findings statistically. You could, of course, generalize theoretically.

Theoretical & Accessible Population

Just as is true for probability samples, you must define the theoretical and the accessible population prior to selecting a non-probability sample. The accessible population must be a reasonable choice in terms of how well it represents or reflects the theoretical population with respect to the variables of interest in a study. Let's assume that you are examining the relationship between educational achievement and attitudes about capital punishment. Gainesville, Florida, would probably not be a good choice for the accessible population for this study. The population in Gainesville just does not reflect the state's population as a whole in terms of educational achievement, one of the two key variables in this study. Tampa, Orlando, Jacksonville or even Ocala would probably be fine as accessible populations. You have to justify how well the accessible population reflects or represents the theoretical population **in terms of key variables of interest**, no matter whether you use probability or non-probability sampling.

Sampling Frame

Most kinds of non-probability samples are greatly strengthened if you also define a sampling frame. Again, this is the same procedure that you must use to select a probability sample. You have some kind of list or roll or other way of identifying all of the members of the accessible population that you can actually locate. You then use this "frame" to select the actual sample subjects.

General Types of Non-Probability Samples

Some authors would probably divide the non-probability samples into somewhat different groups. Unfortunately, not as much attention has been paid to non-probability as to probability sampling. There is, therefore, much less literature about these kinds of samples. In general, I have been able to identify five basic reasons for using a non-probability sample. (1) The subjects (or objects) may be scarce or hard to locate. (2) The researcher may want to be able to compare some subset of the population to the larger population of interest. (3) The researcher may want to compare groups, but not use a stratified probability sample. (4) The researcher may need to rely on volunteers. (5) The researcher may want to establish specific criteria used to select participants.

Referral Samples

Referral sampling is sometimes called snowball sampling. However, the more correct term is referral sampling. This sampling approach is often necessary when the subjects are scarce or hard to find. This is one case in which it is usually impossible to develop a sampling frame. By definition, we often cannot identify the members of the accessible population *a priori*. The basic procedure is to locate initial members of the population, often through consultation with key informants. The researcher then asks these initially identified members to identify additional members of the accessible population, who then identify yet more members, and so on. Sampling is usually discontinued when one of two conditions are met. (1) The researcher determines *a priori* the number of subjects

that he/she must include in the study and ceases sampling when that number is reached. (2) The researcher ceases sampling when no or very few new members are identified; e.g., when the referrals are to individuals who have already been identified.

Example. The theoretical population is homeless people. The accessible population is homeless people in X County. The researcher contacts key informants who deal regularly with the homeless, such as people who run shelters or other kinds of service providers. Note that the homeless individuals identified by the key informants are probably not very representative of the homeless population in X County as a whole. The mere fact that they interact with service providers may well mean that they differ from homeless people who have moved further out of the mainstream of society. Therefore, the researcher continues the process by referral from those individuals initially identified. The procedure ends when the condition set by the researcher are met (no new names appear or the required number of subjects have been located).

Quota Samples

Quota samples are used under two very different sets of conditions.

(1) This is the less common use of quota samples. Quota samples allow the researcher to purposefully **over-represent** certain subgroups in a larger population of interest. In this use, quota sampling occurs **within the framework of a probability sample**. The quota refers to the number of subjects in specific subsets that will be recruited, not to the sample size as a whole. In this usage, the additional subjects included in the quota sample are **above and beyond** the minimal sample size determined by probability sampling. E.g., the additional subjects are added to the initial sample size. When quotas are used within the framework of a probability sample, the researcher cannot use the sample as a whole (minimum probability sample + quota sample) to make statistical generalizations. He/she can make statistical generalizations based on the probability sample alone. The quota sample is **not** chosen randomly, but rather by sequential recruitment, in most cases (e.g., take every member of the sub-group that appears on the sampling frame, in order, until the quota is met).

Example. The objective of the study is to compare the attitudes of university students and non-university students with regard to the war in Iraq. UF is chosen as the accessible population for the university students (and maybe Marion County for the non-university students -- certainly **not** Gainesville or Alachua County). However, the researcher knows that international students at UF may well have attitudes about this war that are different than those of UF students as a whole. He/she would like to know if this is true, and if so, how the sub-group differs from the accessible population as a whole. He/she determines that a sample of 323 students is needed (alpha 0.05). A random sample of 323 students is selected. However, since international students comprise only 4% of UF students, the researcher knows *a priori* that the random sample of 323 will probably only include about 13 international students. This number is too small to perform most statistical analyses. He/she therefore decides to include every international student whose name appears on UF's rolls and who is willing to participate in the study until a total of 40 international students are included.

(2) Very commonly researchers use a quota sample as a substitute for a stratified probabilistic sample. That is the researcher divides the population into mutually exclusive sub-groups (like men and women). He/she then determines how many subjects are needed in each group, based on the total sample size required or desired. Every willing subject in each group is included in the sample until the required number for each subgroup is met. Note that this is not a probability sample at all. It is not possible to calculate a confidence interval or sampling error for the sample. The use of

quota sampling as a substitute for a stratified random sample (a probability sample) is highly debated. Supporters argue that it is efficient and inexpensive and permits researchers to conduct studies that might otherwise be impossible. Critics, including most statisticians, argue that the results are not very meaningful because they cannot be statistically generalized.

Example. A researcher wants to know whether people who buy organic food differ from those who do not in regard to educational level, income, race, gender and age. He/she wants to be able to compare the two groups statistically. He/she decides to administer a self-completion questionnaire to shoppers outside supermarkets in Jacksonville. The researcher expects that non-organic buyers will far outnumber organic buyers, based on national data about food purchasing trends. He/she decides to include 100 participants in each group. The researcher will simply stand outside the stores and ask individuals who emerge if they are willing to answer some questions. The first question will be: "Do you regularly purchase organic food?" At first, it does not matter if an individual says "yes" or "no" because the researcher wants 100 in each group. After a few days, however, the researcher has the 100 non-organic buyers. He/she then politely tells people who answer "no" to the first question, "Thank you for your time. That's all I needed to know." He/she continues to recruit organic buyers until the quota of 100 is met. Note that the researcher can perform statistical tests to compare these two groups in terms of age, gender, etc. **but he/she cannot statistically generalize the results.** In other words, the researcher may find that the average age of the organic buyer group is 27 while the average age of the non-organic buyers is 38. If this difference is statistically significant, the researcher can conclude that the **organic and non-organic buyers in this study differed significantly in age.** But the researcher **cannot say** that the average age of organic food buyers in general is 27, or that there is a difference in age between organic and non-organic food buyers in the population as a whole.

Volunteer Samples

Volunteer samples are used under two very different conditions. The first involves research where there are ethical issues involved, usually where there is some sort of intervention that affects the participants and that may pose some risk to the participants. The second is usually justified simply because it was "easier" to recruit volunteers than to select a probability sample. This is often called a "convenience" sample. This is not, in my opinion, a valid reason for relying on volunteers. In fact, I am of the personal opinion that a convenience sample is not a valid sample in most cases. Volunteer samples are often selected from a sampling frame. The main problem with volunteer samples of any sort is that the participants **probably are not** representative of either the theoretical or the accessible population. The fact that someone volunteers may well mean that the individual differs from other non-volunteers in ways that have impacts on the outcomes or results of the study. The degree to which results from volunteer samples can be statistically generalized is debated.

Example. A school board wants to test a remedial math teaching program for 9th grade students in the district. Participating in the program poses some threat to the students because they will be placed in a "special" class, which may cause other students to taunt them, damage their self esteem, isolate them from their friends, etc. Therefore, the researcher must rely on volunteers. He/she develops a sampling frame that consists of all 9th grade students performing below grade level in mathematics. He/she then asks the parents of these students to volunteer for their children to participate in the remedial program. The treatment group consists of the volunteers and the control or comparison group of the other "poor performers" who do not volunteer. The problem is that the volunteers (or at least their parents) may be more motivated than the non-volunteers. Perhaps they volunteer because they are more concerned about their performance, want to do better, etc. Perhaps the non-volunteers just do not much care. Now the sample selection process may affect the outcome -- improvement in performance. Internal validity is compromised by the sample selection process.

Criteria Samples

Criteria samples are widely used in true experiments and quasi-experiments. In fact, the basis of the quasi-experiment is often that some criteria were used to select the treatment participants that did not apply to the control or comparison group participants. Many true experiments in the medical fields routinely depend for sample selection on criteria-based screening. Participants may be rejected who do not meet criteria like age, presence of other illnesses, smoking, chronic conditions, gender, etc. As we will see in the coming weeks, there are specific experimental and quasi-experimental designs that are used to overcome or at least reduce the effects of using criteria to select participants in these two groups of research designs.

Example. A school board wants to test a new, highly touted math teaching program for 9th grade students in the district. The board members are most concerned about those students who are performing below grade level. Therefore, the researcher uses a design in which the below grade level students are assigned to the treatment group (get the new program) and the other students are assigned to the comparison group (continued with the traditional program). The researcher then compares **change in performance** between the two groups over the course of the first half of the school year as the outcome variable. This is an acceptable design, but the researcher could overcome the problem of the criteria sample by using a switching treatments design. E.g., during the second half of the school year the below grade level students return to the traditional program (become the control or comparison group) and the other students now get the new program. This now eliminates the criteria effect on results.

Researchers also often use criteria samples because they are especially interested in certain subjects or cases. Most often, this is because the researcher wants to generalize theoretically and is little, if at all, concerned with theoretical generalization. Criteria samples will often permit the researcher to select cases that will provide the best basis for theoretical generalization. There are many types of criteria samples that can be used in these circumstances.

Maximum Variation Samples

Probability samples will produce a sample that approximates the standard bell curve in most cases. We have seen this in our exercises and discussion. This means that most of the subjects taken in a probability sample will lie somewhere near the central tendency in regard to the variables of interest. In layman's terms, they will be "pretty average." However, these cases often do not provide much contrast. In theory building research, the objective is often to discover the characteristics or features of subjects or cases that **differ** in order to be able to better understand the factor or variables that affect the outcome -- to better understand why the subjects or cases are different. In these circumstances, the researcher may want to use a maximum variation sample. This means selecting, on purpose, subjects or cases that lie outside the norm in regard to the variables of interest.

Example. A researcher wants to understand what causes high school students to engage in violent acts. He/she is trying to discover some previously unidentified factors in the students' lives that lead them to engage in this behavior. He/she selects two types of students for the study, some who have never been in any trouble at all in high school and others who have repeatedly suffered disciplinary action for violence at school. This permits the researcher to better understand the **differences** in the family and community environments of "violent" and "non-violent" high school students and identify factors for further research.

Homogeneous Samples

Sometimes the researcher wants to understand how one variable, factor or group of variables, among many, influences the outcome. Use this approach when you know that there are many interacting factors, including things like socio-economic status, gender and ethnicity, that you know

a priori will probably affect the outcome, but when you do **NOT** want to study the effects of these complex factors. Taking a probability sample would in most cases give you a wide range -- the full range -- of variability for these factors unless you could find some way to reasonably screen for them so that you could narrow the theoretical population. However, that would be very hard to do *a priori* because variables like “ethnicity” are in and of themselves complex and difficult to operationalize. The homogeneous sample allows you to select subjects that are as nearly alike as possible in terms of the variables or factors that you do not want to study so that you can home in on the effects of the variable or factors that are of interest.

Example. A researcher wants to understand how familial support affects an individual’s success in reducing weight, in sticking to a diet and reaching desired weight loss goals. He/she knows, based on previous research, that many factors affect this outcome, including, for example, ethnicity, gender, age, and socio-economic status. He/she gets a sampling frame through a local weigh loss clinic. Note that this frame is itself probably comprised of a volunteer sample because it is unlikely that everyone who goes to the clinic will be interested in participating in a weight loss study. Let’s say that 150 people say they would be interested in participating in the study. The researcher conducts eliminates all men to try to eliminate the effects of gender -- although sex and gender are not really the same thing. Now there are 100 people left. He/she also eliminates people who are under 30 and over 50 to reduce the effects of age. Now there are 80 people left. This is just like taking a criteria sample. Now comes the tough part. Looking at the list of remaining potential candidates, the researcher decides to limit the final sample to “anglo” subjects because 60 of the remaining 80 potential subjects define themselves as “white.” However, the researcher will have to go further in defining this variable because he/she will want to eliminate, for example, people who check the “white” box for race, but are of Hispanic/Latino heritage. He/she will also probably want to eliminate first-generation immigrants. At any rate, after doing this very carefully, the researcher is now down to 40 potential subjects. He/she interviews those people to learn about their socio-economic status. Finally, he/she derives a socio-economic profile of each potential subject and arrives at a definition -- for the purposes of this study -- of “people of lower-middle socio-economic status” because once again this is the largest remaining group. This leaves the researcher with 28 subjects for the study. Note that in a probability sample or a criteria-based sample, all of these decisions would have to be made *a priori*. In this case, except for gender and age, the researcher is defining the criteria that are used for final sample selection *post hoc*, that is, after the sample selection has begun.

Matched Samples

Matched samples are often used with quasi-experiments and case study designs. Use this approach when you have a treatment group that is pre-determined. They share some characteristic of interest for your research. For example, they may all be part of a drug rehabilitation program. You want to know whether this characteristics of interest -- a predictor or independent variable -- affects the outcome variable. The problem is that you know that lots of other things may affect the outcome variable too. Again, as in the case of the homogeneous sample, these are complex factors like gender, socio-economic status and ethnicity. In a matched sample, you select the comparison (or quasi-control) group to **match** the treatment group as closely as possible in terms of these factors that are **NOT** of interest to you.

Example. One of my students wanted to know whether participation in a training program offered by a non-governmental organization (NGO or non-profit) affected the decisions of farmers in a certain region of Brazil in terms of using agroforestry systems on their farms. She specifically wanted to know if the NGO training increased their knowledge, improved their attitude, and increased their self-confidence in their ability to manage agroforestry systems. There were 32 farmers involved in the training program. The problem was how to select the comparison or control group. She knew that factors like land tenure, size of farm, years of farming experience,

and cropping system on the farm would be important based on previous research. She therefore first collected the data from the treatment group, the participants in the training program. Then she selected 30 other farmers as the comparison group based on the four variables I just mentioned. They all owned their farms, just as the people in the treatment group did. The range of farm size in the treatment (NGO) group was something like 5 to 15 hectares. She selected comparison farmers whose farms fell within that range. The NGO farmers had something like 2 to 10 years of experience as farmers, so she selected other farmers with the same range of years of farming experience -- and so on. She **matched** the control or comparison to the treatment group based on these variables of no interest to her.

Critical Samples

Researchers use this non-probability sampling approach when they want to understand why or how a phenomenon behaves or occurs when it happens under a set of conditions that differs from the circumstances where the phenomenon “normally” occurs. Use this approach when you want to study why or how something happens under unique, new, or unusual conditions -- again, based on the research. Medical researchers often use critical cases, for example, when well-tested, successful treatments for illness and disease fail under circumstances in which one would expect them to succeed.

Example. For many years school violence was believed to be a phenomenon primarily in “poor” or “inner-city” schools. The tacit, if not explicit, assumption was that violence at school was a reflection of the community from which the students came -- that the root causes of school violence were the more general social conditions of poverty, stress, risk, etc. Then we got examples of spectacular school violence in perfectly middle class communities. These incidents gained nationwide, indeed international, publicity. It was immediately clear that low socio-economic status or coming from “poor” neighborhoods was not the only, perhaps not even the most important, factor creating violence on the part of youth in the school system. Researchers therefore intensively studied several of these “critical” cases to try to discover what characteristics of the youth and the community and school setting were different from those reported in previous studies of “poor” schools -- and what characteristics were held in common.

These are just a few of the more common kinds of judgmental samples that can be used. There are many other valid ways of selecting judgmental samples. ***It was the most convenient way to sample is not one of them.*** And always explain how you went about selecting the sample.