

Sampling and Error: Does the Bad Apple Spoil the Whole Bunch?

Sampling can be defined as the process whereby a subset of items is picked from a set, and done so using a systematic process. In other words, sampling in the context of social science means using a randomization technique to pick respondents from a larger population, and through that technique removing selection and other biases.

Principles

The principle of sampling is built upon several truisms. First, the researcher needs to pick a sample of a larger population because time, money, or lack of access to potential respondents prohibits a census. Thus, the researcher needs to be able to get data from some but not all potential respondents. Second, all individuals have intrinsic biases that will either consciously or sub-consciously cause them to influence their selection of people to take part in a study. Hence, if the researcher does not apply a randomization technique that eliminates her ability to bias the selection of participants, the results will be biased and potentially meaningless. Third, for quantitative studies sampling is intended to allow the researcher to extrapolate statistically from a sample to the population being studied. To put it another way, a social scientist may want to study the demographics of all college students in Massachusetts, but due to lack of money, time or a list of names must be satisfied with sampling a portion of those students, then using statistics to ‘guesstimate’ the accuracy of that sample. In qualitative studies, sampling becomes important as a tool to remove the bias of the researcher in picking whom to interview, it becomes essential in insuring replicable results, and necessary to assure accuracy and validity of responses. Thus, sampling is an integral part of estimating the characteristics of a larger population and removing bias in the selection process. Similarly, in qualitative research, sampling becomes a method to account for and remove bias as well as gaining valid and accurate data.

Types of Samples

The methods of sampling are as varied as they are different, but the core issue remains: Does the sampling technique allow the researcher to systematically pick participants in a social science research project, and does that technique minimize or eliminate bias? That being said, some of the more common techniques are included below.

Quantitative Sampling Methods

Simple Random Sample (SRS)

A simple random sample occurs when every unit in the population is known is accessible, and has an equal probability of being selected. Indeed, SRS is the simplest and least complicated sampling technique to administer. Conversely, it is the most difficult to achieve. The simplicity of SRS comes from the known attributes of the population as well as the ability to access it. In SRS, a statistically valid sample is drawn using a chance mechanism to select from an entire and complete list of the study population. In

other words, a SRS of the members of a fraternity would involve putting each of the names of its members into a hat and drawing randomly from the population to determine who will be studied. The allocation of winning numbers in many state lotteries works much the same way. The point is that the entire population is subjected to a chance mechanism that eliminates researcher selection bias while randomly choosing the sample.

Interestingly, the strengths of SRS also are the source of its weaknesses. Often times all the units of the population are either not known or are unable to be accessed, thus rendering a SRS impossible because the probability of selection cannot be allocated equally among the potential participants. An example would be a proposed study of sexual orientation among Americans. A list of the entire population is not available, and if it were the unwieldy size of the database would cause problems for even the most sophisticated researcher. Likewise, even if the list was available all the units of the population may not be accessible, e.g. homeless transients. Thus, SRS would not be the appropriate sampling model.

Systematic Sampling (SS)

SS is a valuable sampling technique in the absence of known population parameters. Whereas SRS requires that each unit have an equal chance of being selected, SS assumes that this is not the case and merely seeks to randomize selection within the population. Furthermore, SS is widely considered to be among the easiest sampling techniques because it is proportional to the population being studied, and is allocated proportionately. The first step in SS is gaining an estimate of the population size. Then, a statistically valid sample is selected. Next, the population is divided by the sample size to select the person to be studied. Finally, a chance mechanism is applied to the selection to minimize several biases implicit in SS. Importantly, SS is administered sequentially, that is, to each 'nth' person in a string of people.

For example, a sponsor desires a study of education levels among people attending a rodeo. The entire population is neither known nor accessible. Hence, the method of choice might be SS. The sponsor estimates that 10,000 people will attend the rodeo, and after consulting a statistics book it is determined that the sample size should be 200. Biases like selection bias and periodicity can confound a systematic sample, so a chance mechanism should be used to decide whom to interview. Thus, a coin would be flipped, and 'heads' means the selected person is interviewed while 'tails' means that the selected person is passed over in favor of the next person in the sequence. Since a coin-based chance mechanism offers a 50:50 chance of selection, the sample is effectively doubled. Hence, to interview 200 people at the rodeo, 400 would have to be selected; of those 400 the chance mechanism would successfully allocate 200 for interviews. Therefore, every 25th person who leaves the rodeo arena would be selected, a coin would be flipped, and all those who are 'heads' would be interviewed for the study.

Stratified Sampling (STS)

Whereas SRS relies upon the simplest allocation of chance to the most difficult population to define, and SS applies proportional sampling to an estimated population, STS implements a layered sample to an estimated population. In other words, STS applies a sampling technique that stratifies the estimated population into sub-groups, then attempts to sample among those sub-groups. STS is usually applicable in instances where each subdivision demonstrates less variability than the whole, or where the researcher wants to study the relationships among sub-groups.

An example might be provided by a proposed study of autistic children. The exact parameters of the population might not be known, so SRS would not be used. Similarly, the literature indicates that distinct differences exist within the autistic population based upon gender and ethnicity. Hence, SS would be inappropriate because it wouldn't account for those strata. Thus, STS would be a good choice. It would allow for each stratum, in this case gender and ethnicity, to be studied independently and then compared with each other.

In STS, the sampling mechanism is often times very similar to SS. The population of each sub-group is estimated, then a statistically valid sample size is determined and allocated by systematic sampling within each sub-group. As is evident by the multiply layered approach, STS is far more complex than either SRS or SS. However, a benefit of STS is that it allows for far more nuanced studies and sophisticated analyses.

Qualitative Sampling Methods

Reference Sampling (RS)

Reference sampling is a technique where the group being studied self-defines the sampling technique. Sometimes referred to as snowballing, the sampling technique clearly is not intended to be representative of any larger population. Indeed, it doesn't even claim to remove self-selection bias. Instead, RS actually seeks out self-selection in an effort to provide depth of meaning among like-minded or similar people.

The purpose of RS is two fold: to allow a frame to self-define and thus provide context and depth, and secondly to remove researcher selection bias. Often times in qualitative research, the question is not whether a sample is representative, but whether it accurately reflects the meanings and lives of those studied. On the one hand, quantitative methods such as SRS attempt to say that by studying 'x' number of people, the researcher can then say that all people feel a certain way. On the other hand, qualitative methods like RS attempt to say that by selecting and researching participants in a meaningful way, the data that are obtained accurately reflect the participants alone. Hence, in RS, the goal is to gather data that sheds light on the highly personalized and contextualized meanings of participants.

In other words, think of RS as an attempt to go deep into a well of meaning among similar people by letting them do the work of lowering you down into the well. Of course the sample is highly biased toward the views of those who self-selected, but those biases are readily accounted for by triangulating RS. In triangulating RS, the researcher purposively picks several different types of people, and applies RS to each separate frame. In the process, the other frames cover each frame's intrinsic bias, and thus an accurate picture of meaning is provided.

The actual technique of reference sampling is very simple. After identifying the frames, the researcher interviews a single person inside that frame. At the conclusion of the interview, the respondent is asked for two people who would be willing to talk (references). Likewise, the respondent is asked if he can provide a reference for the researcher by way of calling the other respondents, or letting the researcher use the original respondent's name when setting up future interviews. The problem of sample size is solved when either the frame begins providing all of the same references or the respondents all begin giving the same answers.

As an example, imagine that a sponsor wants a study of feelings about sexual orientation among students on campus. RS involves first deciding upon your frames, in this case gays, lesbians, bisexuals and heterosexuals. Then, the researcher establishes contact with a respondent in each frame who is willing to be interviewed. After the interview, each respondent recommends 2 more people to be interviewed. Then the researcher asks the respondent if he can use the respondent's name when contacting the next respondents. As the name implies, this provides a reference and makes getting subsequent interviews much easier. In this way, the sample snowballs, growing larger as each successive reference provides more references. When each of the frames either begins giving the same references over and over, or when the respondents in each frame provide the same answers, sampling is concluded in that frame. In this way, the researcher can be confident that accurate depth of meaning and context has been achieved.

Purposive Sampling (PS)

Purposive sampling is exactly that: the sample is chosen by the researcher with explicit attributes in mind, with the intent of providing in-depth analysis of those attributes and how they relate to the selected individual. Purposive sampling is not generally statistically valid, nor is it free from bias, thus it is unacceptable for quantitative studies. Nonetheless, it can be used successfully to provide depth and meaning to a qualitative study.

In PS, the research simply decides upon the selection attributes and applies the sample. A sponsor who wants to study the intelligence level of blonde women might provide a grotesque but demonstrative example. In this case, the criteria of selection would be blonde hair. Obviously, the researcher is not only bigoted and idiosyncratic in his use of stereotypes, but the very application of biased selection criteria precludes the results. Nevertheless, useful meaning might be gained about how blondes feel about being blonde. In other words, if the removal of researcher or sponsor bias is desired, purposive

sampling is completely unacceptable. If gaining both hypothesized and unanticipated meanings are the goal, then it is useful.

Convenience Sampling (CS)

Convenience sampling is a technique where the researcher basically talks to whomever will talk to her. It's convenient, hence the name! Often times in IQPs, students' methodologies either are poorly planned, poorly implemented, or conditions outside their control make valid sampling impossible. Hence, they are unable to randomize their sampling technique. In the panic that besets an IQP in such dire straits, the students turn to whoever will give them data. In other words, students and researchers fall into the mistaken assumption that *something* is better than *nothing*. The problem lay in the fact that the results of convenience sampling are neither representative nor valid. Since no chance mechanism was used to allocate the sample, no extrapolation to a larger population is possible. Likewise, since biases were not controlled for, the results of CS are basically meaningless. The data from such studies are confounded by various and unidentified biases and therefore the data cannot be interpreted with any accuracy.

Summary

The goal of sampling is first to provide a systematized method whereby participants in a social science project can be selected. Likewise, sampling intends to remove biases and therefore provide accurate results. Lastly, proper sampling allows either the extrapolation of results from a sample to the entire population in quantitative studies or for the accurate collection and recollection of qualitative data from respondents. Most important, proper sampling method can be replicated by future researchers.

In quantitative studies, proper sampling helps to eliminate bias and provides accurate answers, by providing a chance mechanism in which the participants were randomly selected. In qualitative studies, proper sampling helps to remove researcher selection bias, allows for accurate interpolation of meaning, and allows respondents to provide depth of context and nuanced answers that no survey could discover.

References

- Berg, B. (1998). *Qualitative Research Methods for the Social Sciences*. Boston: Allyn and Bacon.
- Blau, P. (1964). *Exchange and Power in Social Life*. New York: John Wiley and Sons.
- Dillman, D. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley and Sons
- Mendenhall, W. (1998). *Introduction to Probability and Statistics*. Pacific Grove: Brooks/Cole.

Ott, L. and W. Mendenhall. (1987). *Statistics; Tools for Social Sciences*. Boston: PWS.

Salant, P. and D. Dillman. (1994). *How To Conduct Your Own Survey*. New York: John Wiley and Sons.

Scheaffer, R., W. Mendenhall, and L. Ott. (1995). *Elementary Survey Sampling*. Boston: PWS.