HOW PROBABILITY SAMPLING DIFFERENTIATE WITH NON-PROBABILITY SAMPLING

Satish*

ABSTRACT

Almost any type of sample has some utility when estimating population quantities. The focus in this paper is to indicate what type or combination of types of sampling can be used in various situations to differentiate with probability sampling and non-probability sampling. Several of these methods have little or no utility in the scientific area but even in the best of circumstances, particularly complex ones, both probabilistic and non-probabilistic procedures have to be used because of lack of knowledge and cost. Researchers collect information by a wide variety of methods, ranging from the experimental designs used in the physical sciences through to the surveys more common in the social sciences. Many of these methods of gathering information involve a choice of experimental subject. For example, we may want to choose the patients to be examined in a medical study, or the respondents to be interviewed in a survey.

This choice can be made using probability-based methods, where the choice is by some "mechanical" procedure involving lists of random numbers, or the equivalent. Alternatively, the choice may be made by other methods, invoking some element of judgement. Methods involving judgement are sometimes referred to as purposive selection, judgement selection, or non-probability selection.

***Key Words**: probability-based samples, quality assessment; Non-sampling errors*

*Extension lecturer, govt P.G College, Jind

INTRODUCTION

The word Probability derives from probity, a measure of the authority of a witness in a legal case in Europe, and often correlated with the witness's nobility. In a sense, this differs much from the modern meaning of probability, which, in contrast, is used as a measure of the weight of empirical evidence, and is arrived at from inductive reasoning and statistical inference.

HISTORY OF PROBABILITY THEORY

The branch of mathematics known as probability theory was inspired by gambling problems. The earliest work was performed by Girolamo Cardano (1501-1576) an Italian mathematician, physician, and gambler. In his manual Liber de Ludo Aleae, Cardano discusses many of the basic concepts of probability complete with a systematic analysis of gambling problems. Unfortunately, Cardano's work had little effect on the development of probability because his manual, which did not appeared in print until 1663, received little attention.

In 1654, another gambler named Chevalier de Méré created a dice proposition which he believed would make money. He would bet even money that he could roll at least one 12 in 24 rolls of two dice. However, when the Chevalier began losing money, he asked his mathematician friend Blaise Pascal (1623-1662) to analyze the proposition. Pascal determined that this proposition will lose about 51% of the time. Inspired by this proposition, Pascal began studying more of these types of problems. He discussed them with another famous mathematician, Pierre de Fermat (1601-1665) and together they laid the foundation of probability theory.

PROBABILITY THEORY:

Probability sampling means that everyone in a given population has an equal chance of being surveyed for a particular piece of research. Let’s say we want to know how many people would choose blue as their favorite color. If we wanted to answer that question in the context of the average American, that would mean that everyone in the United States would have an equal chance of being sampled for the study. The same holds true for sub-segments of the population.

For example, if you wanted the opinions of pregnant moms, a probability sample would mean that every pregnant mom would have an equal chance of participating in the research.

In probability sampling every element in the population has a *known nonzero probability* of selection. The simple random is the best known probability sample, in which each member of the population has an equal probability of being selected.

Probability sampling designs are used when the representativeness of the sample is of importance in the interest of wider generalisability. When time or other factors, rather than generalisability, become critical, non-probability sampling is generally use In probability-based sampling, the first step is to decide on the population of interest, that is, the population we want the results about. This could be, for example, all persons aged 18 years or over who are resident in private households in New Zealand.

We then establish a frame - a listing, at least in principle - of all the units of that population. For our example of the persons in private households, we might use a geographic frame. Private dwellings would be listed according to the geographic area they are in, and people listed inside dwellings.

We select a sample from this frame using a probabilistic algorithm. It is important that every element of the frame has a known chance of being selected, and that we can calculate the probability of selecting the sample we end up with.

The sample might well be selected in several stages. In our example geographical areas might first be selected, then dwellings inside these areas. Finally, people might be selected inside the dwellings.

In saying that we use a probabilistic algorithm to select the sample, one important feature is that interviewers will have no choice about who they are to interview. The algorithm specifies who is to be in the sample.

To produce our results, we combine the responses from the sample in a way which takes account of the selection probabilities. Our aim is that, if the sampling were to be repeated many times, the expected value of the results from the repeated samples would be the same as the result we would get if we surveyed the whole population.

Because we know the probability of getting each sample we select, we can also calculate a sampling error for the results. The sampling error tells us the amount of variation in the results due to the sampling alone. It gives a measure of the quality of the sample design, and of the survey results.

BIG PROBLEMS IN PROBABILITY THEORY ?

Most branches of mathematics have big, sexy famous open problems. Number theory has the Riemann hypothesis and the Langlands program, among many others. Geometry had the Poincaré conjecture for a long time, and currently has the classification of 4-manifolds. PDE theory has the Navier-Stokes equation to deal with.

So what are the big problems in probability theory and stochastic analysis? I'm a grad student working in the field, but I can't name any major unsolved conjectures or open problems which are driving research. I've heard that stochastic Löwner evolutions are a big field of study these days, but I don't know what the conjectures or problems relating to them are.

NON PROBABILITY SAMPLING:

Non-probability sampling comes in various shapes and sizes, but the essence of it is that a bias exists in the group of people you are surveying. Let's think about it in the context of our fictional color preference survey. If I asked the question to all of my friends, the results are not representative of anything other than the opinion of my friends and, specifically, those friends to whom I decided to send the survey. Another example of non-probability sampling would occur if I were to send you the survey and then ask you to pass the survey onto a friend. This effect, called snowballing, creates a biased sample wherein not everyone has an equal chance of being sampled.

The selection of units in non-probability sampling is quite arbitrary, as researchers rely heavily on personal judgment. It should be noted that there are no appropriate statistical techniques for measuring random sampling error from a non-probability sample. Thus projecting the data beyond the sample is statistically inappropriate. Nevertheless, there are occasions when non-probability samples are best suited for the researcher's purpose

Any sampling procedure where the final samples' is not obtained by means of "real life probability sampling" will be classified here as a non-probability sampling procedure.

Typically, there are no known inclusion probabilities, and sometimes there is not even an exactly defined population. A few well-known examples of non-probability selection procedures are the following:

Self-selected respondents. An invitation to answer a number of questions is given to a large (and often not well-defined) group of people, for example by announcement in a newspaper or on an internet site. Anyone who reads the invitation is allowed to answer, but usually only a very small fraction do in fact respond.

Respondents selected by interviewers. The interviewers have the freedom to select people to interview, for example, in the street or in a shopping mall. Sometimes the freedom is restricted by quota rules, saying that there has to be, for example, the same number of men and women.

Respondents selected by experts. An expert or a researcher contacts a number of persons who happen to be available and who are believed in some sense to be “typical members” of the larger group of people that one is really interested in.

PROBLEM IN NON-PROBABILITY SAMPLING

This is more biased, because the individuals chosen are not at random. They also might not represent what another population thinks

CONTRAST OF PROBABILITY AND NON-PROBABILITY SAMPLING:

Probability sampling procedures, in contrast to the remaining class of non-probability sampling procedures. Under ideal text-book conditions, probability sampling is usually described along the following lines:

(1) A sample of units is to be selected from the population, using some known randomization mechanism.

(2) It is possible (at least in principle) to list all the samples that can be obtained using this procedure.

(3) We know the probability of each possible sample, when this procedure is used.

(4) For each unit in the population we can find its inclusion probability, that is, the probability that it will be selected. The inclusion probability can be obtained, for example, by adding the probabilities of all the possible samples that contain this special unit.

(5) Each unit in the population is to have a strictly positive inclusion probability. This is a necessary and sufficient condition for the existence of an unbiased estimator of the population total.

Probability Sampling	Non-probability Sampling
You have a complete sampling frame. You have contact information for the entire population	**Used when there isn't an exhaustive population list available**. Some units are unable to be selected, therefore you have no way of knowing the size and effect of sampling error (missed persons, unequal representation, etc.).
You can select a random sample from your population. Since all persons (or "units") have an equal chance of being selected for your survey, you can randomly select participants without missing entire portions of your audience.	Not random.
You can generalize your results from a random sample. With this data collection method and a decent response rate, you can extrapolate your results to the entire population.	**Can be effective when trying to generate ideas and getting feedback**, but you cannot generalize your results to an entire population with a high level of confidence. Quota samples (males and females, etc.) are an example
Can be more expensive and time-consuming than convenience or purposive sampling.	**More convenient and less costly**, but doesn't hold up to expectations of probability theory

CONCLUSION:

Probability samples are expensive but provide results that can be extrapolated to a wide population.

Non-probability samples are less expensive but are limited in extrapolating results. Interpretation of results should be viewed with caution.

REFERENCES:

1. Gregoire, T. G.: 1998, Design-based and model-based inference in survey sampling: Appreciating the difference, Canadian J. For. Res 28, 1429–1447.
2. Smith TFM (1983), "On the validity of Inferences from Non-random Samples", Journal of the Royal Statistical Society, Vol. 146, pp 394-403.
3. Stephan FP and McCarthy PJ (1958), "Sampling Opinions", John Wiley and Sons, New York.
4. Deming WE (1960), "Sample Design in Business Research", John Wiley and Sons, New York