SAMPLING METHODS AND APPROACHES IN RESEARCH

Dr. Yasmeen Sultana Farooqi,
Associate Professor, Department of Media Science, Karachi

Sampling is the practice of choosing a random sample of individuals from a larger group in order to ascertain its characteristics. Each individual variable quantifies one or more qualities of observable entities that can be categorized as autonomous objects or persons (for example, weight, position, and color). In survey sampling, weights can be applied to data to account for the sample design; this is especially true in stratified sampling. Practice is guided by probability and statistical theory. To extract information on a population, sampling is widely used in all fields of study.

There are several phases to the sampling procedure:

- Identifying and defining the vulnerable population
- Creating a sampling frame (a sampling frame is a collection of items/a list or map containing all the units from which a sample is obtained)
- Using a sampling approach to choose items or events from the frame
- Sample size calculation
- Placement of the example approach into action
- Sampling and data collecting
- Data that can be selected

In statistics, a simple random sample is a subset of people (a sample) selected from a larger collection (a population). Each person is chosen at random and entirely by chance, implying that each individual has the same chance of being chosen at any moment throughout the sampling process, and that each subset of k persons has the same chance of being chosen for the sample as any other subset of k people. Simple random sampling, not systematic random sampling, is the name given to this approach and method. A simple random sample is a way of surveying that is completely objective. Saul McLeod (2014) Because it can be used in conjunction with more complicated sample procedures, simple random sampling is a fundamental type of sampling. The assumption behind simple random sampling is that each basic unit has an equal chance of being chosen. Consider the following scenario: N university students want to attend a play, but there are only X N tickets available, and they want to create a fair procedure for allocating tickets. Then, for each participant, a number between 0 and N-1 is allocated, and random numbers are generated either electronically or from a table of random numbers (which is always included at the back of any normal mathematical/statistical table). Any numbers outside of the range 0 to N-1, as well as any previously selected numbers, are ignored. The first X digits would indicate who would get a gate pass. Such sampling is frequently done "without replacement" in small and often large populations, meaning that no member of the population is chosen more than once. While simple random sampling with replacement is conceivable, it is less common and is frequently characterized more fully as simple random sample with 33 replacements. Without replacement, sampling fails to be independent, but still fulfils exchangeability, and many conclusions remain valid. Furthermore, sampling without replacement for a small sample from a large population is basically similar to sampling with replacement for a small sample from a large population, because the likelihood of selecting the same person twice is low. If a large number of samples are picked, it is vital to choose people at random so that the average sample accurately represents the population. This isn't to say that a sample is a perfect representation of the entire population. The sample can be utilized to derive externally valid inferences about the entire population using simple random sampling. In terms of idea, simple random sampling is the simplest probability sampling approach. It demands the creation of a thorough sample frame, which may be difficult or impossible to achieve for large populations. Even when a whole frame is provided, when more relevant information about the population units is available, more efficient solutions may be possible. It has the advantages of being error-free and requiring little prior knowledge of the population other than the frame. Its simplicity also makes the data collected in this manner very easy to comprehend. For these reasons, simple random sampling is best used when there is minimal information about the population and data collection can be done efficiently with randomly dispersed objects, or when the sampling cost is low enough that efficiency takes a second place to simplicity. If none of these conditions are met, stratified or cluster sampling may be a better choice.

Table of Random Numbers: A random number table is created specifically for the purpose of collecting random samples. To utilize this table, all population units are initially numbered as follows:

- 01-99, if the population has less than a hundred units.
- If the value is less than 1,000, the range is 001-999.
- 0001-9999, if the value is less than 10,000, and so on.

The table is then used to determine the quantity of units required. Assume that a random sample of five kids from a school of 99 is needed. We begin by obtaining a list of all 99 pupils in the school. Due to the fact that the school population is fewer than 100, we will number the pupils as follows: 01, 02, 03..., 99. Then we take Run down or cross any two columns or rows of the random number table. Every time we come across a unique number between 01 and 99, we make a note of it. The students whose numbers match to the first five distinct numbers were drawn in this manner from our five-person random sample.

Frame: A frame is a collection of all the population units from which sample units are drawn. These units are referred to as sample units. The frame in the preceding example of five pupils is the school list. Take note of the distinction between samples and sampling units. The former refers to items included inside a sample, while the latter refers to things contained within a population. Typical instances of frame are as follows:
1) A list of schools, if a sample of schools is to be formed.
2) A list of all the pupils at a school if the school's enrollment has to be determined chosen to create the sample.
3) A list of all tax payers in a certain region, and so forth.
Sampling in research can be classified into different types:

1. Survey Sampling

Survey sampling is a statistical term that refers to the practise of choosing a random sample of components from a target population in order to conduct a survey. The word "survey" may apply to a wide variety of observational methods or procedures. In most cases, survey sampling entails the use of a questionnaire to ascertain individuals' traits and/or opinions. The topic of survey data collecting is the many methods for contacting members of a sample after they have been chosen. The aim of sampling is to minimise the expense and/or effort associated with surveying the whole target population. A census or full enumeration is a survey that surveys the whole target population.

Probability samples and non-probability samples are the two main categories of survey samples. Probability-based samples use a sampling strategy that incorporates defined probability (perhaps adapted probabilities specified by an adaptive procedure). Probability-based sampling enables inference about the target population based on the design. The conclusions are based on a given objective probability distribution. Inferences drawn from probability-based surveys may nevertheless be subject to a variety of different kinds of bias. Concerning survey sample, the following things should be noted:

- Surveys that are not conducted using probability sampling have a harder time quantifying their bias or sample error. Surveys conducted using non-probability samples often fail to accurately reflect the intended population.
- Probabilistic sampling is a de facto norm in academic and government survey research.

Surveys must be carried out:

a. Selection of samples using widely recognised statistical techniques (e.g., probabilistic sampling)
b. Techniques that may be used to calculate the sampling error). Any use of non-probability
c. Statistical justifications for sampling techniques (e.g., cut-off or model-based samples) must be provided.
d. Capability to quantify estimate inaccuracy.

In addition to random sampling and design-based inference, additional techniques are used.

a. Statistical techniques, including model-assisted and model-based sampling.

For instance, many surveys have a high rate of non-response. While the units are originally selected with known probability, the reasons behind their non-response are unclear. For surveys with a high rate of non-response, statisticians have suggested the following:

b. Statistical models, which are used to evaluate data sets.

The model-based method is much more often employed in statistical inference than the design-based approach, which is mostly utilized for survey sampling. All assumptions are effectively incorporated in the model when using a model-based approach.

2. Systematic Sampling:

The term "systematic sampling" refers to a statistical technique that entails selecting items from an ordered sample frame. The most often used technique of systematic sampling is equal-probability sampling. In this method, the list is traversed in a cyclical fashion, with a return to the top whenever the list's end is reached. The sampling procedure begins with a random selection of one element from the list, followed by a selection of every kth element in the frame, where k is the sampling interval (sometimes referred to as the skip): this is computed as: Where n is the sample size and N is the population size.

Each element in the population has a known and equal chance of selection using this method. As a result, systematic sampling is functionally equivalent to random sampling. It is, nevertheless, much more efficient (if variance within systematic sample is more than variance of population).

Because systematic sample units are evenly distributed across the population, systematic sampling should be used only if the supplied population is conceptually homogenous. The researcher must verify that the sample interval he or she chooses does not conceal a trend. Any pattern would put randomization in jeopardy.

For instance, if a supermarket wishes to investigate its customers' purchasing patterns, they may use systematic sampling to choose every tenth or fifteenth client entering the store and perform the research on this sample.

This is a system-assisted random sampling. A random beginning point is selected from the sampling frame, and further selections are made at regular intervals. Consider the following scenario: you want to sample eight homes from a street of 120. 120/8=15, which means that after a random beginning point between 1 and 15, the 15th house is selected. If the randomly chosen starting point is 11, the homes are 11, 26, 41, 56, 71, 86, 101, and 116. As an aside, if every 15th home was a "corner house," this corner pattern would obliterate the population's unpredictability.

If, as is increasingly often the case, the population is not equally divided (sample 8 homes out of 125, where 125/8=15.625), should you sample every 15th or every 16th house?

If every sixteenth home is selected, 8*16=128, there is a chance that the last house chosen does not exist. On the other side, if every 15th home is chosen, 8 multiplied by 15 is 120, which means that the last five houses will never be chosen. To ensure that each house has an equal chance of being selected, the random starting point should be changed to a non-integer between 0 and 15.625 (inclusive on one endpoint only); the interval should now be non-integral (15.625); and each non-integer selected should be rounded up to the next integer. If the random starting point is 3.6, the houses are 4, 20, 35, 50, 66, 82, 98, and 113, with three 15-cycle intervals and four 16-cycle intervals.

To demonstrate the risk of systematic skip hiding a pattern, consider sampling a planned community with 10 homes per block on each street. This puts homes No. 1, 10, 11, 20, 21, 30... on block corners; corner blocks may be less desirable since a greater portion of their size is taken up by street frontage and other non-buildable space. If we then sample every tenth home, our sample will either consist entirely of corner houses (if we begin at 1 or 10) or will be entirely devoid of corner buildings (if we begin at any other point); in either case, our sample will be unrepresentative.

Systematic sampling may also be employed with selection probabilities that are not equal. Rather of simply counting the population's components and choosing the kth unit, we assign each element a space along a number line based on its selection probability. Then, using a uniform distribution between 0 and 1, we create a random start and go down the number line in increments of one.

For instance, we have a population of five units (A to E). We wish to assign a 20% chance of selection to unit A, 40% to unit B, and 100% to unit E. Assuming that the alphabetical order is maintained, we assign each unit to the following interval:

A: between 0 and 0.2
B: between 0.2 and 0.6 (= 0.2 + 0.4)
C: between 0.6 and 1.2 (= 0.6 + 0.6)
D: between 1.2 and 2.0 (= 1.2 + 0.8)
E: between 2.0 and 3.0 (= 2.0 + 1.0)

If our random start value was 0.156, we would choose the unit whose interval includes this value first (i.e., A). Following that, we'd choose the interval containing 1.156 (element C), and then 2.156. (Element E).

If our random start value were 0.350, we would choose amongst the points 0.350 (B), 1.350 (D), and 2.350. (E).

Distinction between a systematic random sample and a simple random sample

Consider a school of 1000 pupils, equally split between boys and girls, and assume a researcher wishes to examine 100 of them further. All of their names could be placed in a bucket and then 100 names may be drawn. Not only does each individual have an equal chance of being picked, but we can also simply calculate the probability P of a particular individual being chosen if we know the sample size (n) and the population (N):

1. In the event that a certain individual may only be chosen once (i.e., after selection a person is removed from the selection pool)

2. If every chosen individual is returned to the selection pool (i.e., may be picked more than once), this implies that every student in the school has a roughly 1 in 10 chance of being selected using this technique. Additionally, all 100-student combinations have the same chance of selection.

When a systematic pattern is included into random sampling, the term "systematic (random) sampling" is used. For instance, suppose all pupils at the school had a number assigned to their names ranging from 0001 to 1000, and we chose a random beginning point, e.g. 0533, and then randomly selected every tenth name afterwards to create our sample of 100. (starting over with 0003 after reaching 0993). In this regard, this method is comparable to cluster sampling, since the first unit determines the remainder. This is no longer pure random sampling, since some combinations of 100 students have a higher chance of selection than others — for example, 3, 13, 23,..., 993 have a 1/10 chance of being chosen, while 1, 2, 3,..., 100 are not.

3. Stratified Sampling

Stratified sampling is a technique for selecting individuals from a population. When subpopulations within a larger population differ, it is beneficial to sample each subpopulation (stratum) separately via stratification. Prior to sampling, stratification is the process of separating the population into homogenous groupings. The strata should be mutually exclusive: each member of the population should belong to a single stratum. Additionally, the strata should be comprehensive collectively: no demographic element should be omitted. Then, within each stratum, simple random sampling or systematic sampling is used.

This often enhances the sample's representativeness by decreasing sampling error. It can provide a more stable weighted mean than the arithmetic mean of a simple random sampling of the population.

Stratified sampling is a variance reduction technique used in computational statistics where Monte Carlo techniques are employed to estimate population statistics from a known population.

Strategies involved in Stratified sampling

1. Proportionate allocation employs a sample fraction proportionate to the entire population in each of the strata. For example, if the population X contains m men and f females (m + f = X), the relative sizes of the two samples (x1 = m/X males, x2 = f/X females) should represent this proportion.
2. Optimal allocation (or disproportionate allocation) - Each stratum has a standard deviation proportional to the standard deviation of the variable's distribution. To produce the least sampling variance feasible, larger samples are collected in strata with the highest variability.

Stratified sampling guarantees that at least one observation is chosen from each stratum, even if the chance of selecting it is much less than one. As a result, if there are thin strata, the population's statistical characteristics may be compromised. A general guideline is that the population should have no more than six strata, although this may vary depending on the circumstances - for example, if there are 100 strata each with one million observations, it is entirely acceptable to do a 10% stratified sample on them.

A real-world example of stratified sampling in action would be a political poll. If, however, To ensure that respondents reflected the population's variety, the researcher would explicitly aim to include members of different minority groups such as race or religion, depending on their proportionality to the overall population. Thus, a stratified survey may claim to be more representative of the population than a random or systematic sample survey. The following are some of the benefits and drawbacks of stratified sampling.

Advantages of Stratified sampling over other sampling methods

If an area's population density changes significantly, stratified sampling ensures that estimates may be made with similar precision in various sections of the territory and that comparisons of subregions can be performed with equal statistical power. For instance, in Ontario, a survey conducted throughout the province may use a higher sampling fraction in the less populated north, because the population disparity between north and south is so great that a sampling fraction based on the provincial sample as a whole may result in the collection of only a few data points from the north. Additionally, randomization may be utilised to increase the representativeness of a study's population.

Disadvantages of Stratified Sampling

When the population cannot be thoroughly partitioned into separate subgroups, stratified sampling is ineffective. Making subgroup sample sizes proportionate to the quantity of data available from the subgroups, rather than scaling sample sizes to subgroup sizes, would be a misapplication of the method (or to their variances, if known to vary significantly e.g. by means of an F Test). The stratified sampling issue in the presence of unknown class priors (ratios of subpopulations within the total population) may have a detrimental impact on the performance of any study on the dataset, for example, classification. Minimax sampling ratios may be employed in this case to ensure that the dataset is resilient to the uncertainty inherent in the underlying data generation process.

Practical example

In general, the sample size in each stratum is proportional to the stratum's size.
This is what is referred to as proportionate allocation. Assume that a business employs the following personnel:
90-year-old man, full-time
18-year-old guy, part-time
9 females who work full-time
63-year-old female, part-time

180 in total

If we are needed to sample 40 employees stratified by the aforementioned categories, the first step is to determine the total number of employees (180) and the percentage of each category.
% male, full-time = 90 180 = 50%
% male, part-time = 18 180 = 10%
% female, full-time = 9 180 = 5%
% female, part-time = 63 180 = 35%
This indicates that out of our 40-person sample,

Males should make up 50% of the workforce on a full-time basis.
10% should be part-time male employees.
5% of the workforce should be female and full-time.
35% should be part-time female employees.
50% of 40 equals twenty.
10% of 40 equals four.
5% of 40 equals 2.
35% of 40 is 14.

Another simple method that eliminates the need to compute percentages is to multiply the size of each group by the sample size and divide by the overall population size (size of whole staff):
Full-time male = 90 (40 180) = 20
Part-time male = 18 (40 180) = 4
full-time female = 9 (40 180) = 2
63 (40 180) = 14 female, part-time

4. Cluster Sampling:

In many instances, population units are classified as members of a natural group. For example, the school is a natural grouping of school children if the population of school children is of interest, while markets are a natural grouping of food prices if food prices are of interest.

Clusters are natural groupings composed of the components of interest. Cluster sampling is the process of selecting a sample of clusters from all accessible clusters of interest and then studying all the units inside the chosen clusters.

5. Multi-Stage Sampling:

Assume a research is conducted that includes all of the country's states. The investigator may first choose six states at random (1st stage), and then 10 local government areas at random from each of the six states (second stage). This is a two-stage sampling procedure. He may choose for a three-stage system by visiting zones. Take note that each level is contained inside a cluster.

6. Some Non-Random Sampling Method

Any sampling procedure where chance devices are not used to select the required sample is said to be non-random.

Examples of non-random sampling scheme are:
Systematic Sampling, Quota Sampling and Haphazard Sampling.

I. Systematic Sampling:

As previously stated, this sampling method selects sample units at specified intervals from the frame. For instance, rather of randomly choosing 10 pupils from a school, one may choose a systematic sample consisting of every ninth name on the school's list of 90 names.

The primary drawback is that one may wind up with units that are identical to one another due to an unforeseen periodic fluctuation in the frame. In reality, randomization may be introduced here by randomly determining where on the list the systematic selection will begin.

II. Quota Sampling:

Here, one tries to depict the many classes that may exist within a population. It is often used in surveys of public opinion and market research. In such surveys, interviewers are obliged to guarantee that the sample includes a certain number of units in different categories such as age, sex, economic level, and geographic region. The primary distinction between this and stratified sampling is the absence of a well-defined selection criterion in quota sampling. Apart from random selection from each stratum, the units to be included in the sample are known before to sampling in stratified sampling. These units are unknown in advance when quota sampling is used.

III. Haphazard Sampling:

The selector believes he is making a random pick in this case. A good illustration is the kind of selection that often occurs in public areas by:

1. Agents of press
2. Entrepreneurs

Who interviews individuals in the first place, seeking their perspectives on a subject of interest to them?

However, no random mechanism is utilized in this kind of selection. Numerous biases may be introduced into such random choosing.

References

1. Saul McLeod (2014) Sampling Methods http://www.simplypsychology.org/sampling.html Accessed 25th February, 2015
2. Sampling techniques. http://www.rgs.org/OurWork/Schools/Fieldwork+and+local+learning/Fieldwork+techniques/Sampling+techniques.htm
3. Simple Random Sampling and other Sampling Methods. Statistical Concepts and Reasoning https://onlinecourses.science.psu.edu/stat100/node/18
4. James Lani.(2013) Dissertation workshop. http://www.statisticssolutions.com Accessed 3 Rd March,2014.
5. Research Methodology : An Introduction Google Accessed
6. Howell, K. E. (2013) Introduction to the Philosophy of Methodology. London: Sage Publications
7. Irny, S.I. and Rose, A.A. (2005) “Designing a Strategic Information Systems Planning
8. Methodology for Malaysian Institutes of Higher Learning (isp- ipta), Issues in Information System, Volume VI, No. 1, 2005.
9. Katsicas, Sokratis K. (2009,). "Chapter 35". In Vacca, John. Computer and Information Security Handbook. Morgan Kaufmann Publications. Elsevier Inc. p. 605. ISBN 978-0-12-374354-1.
10. Lodico, Marguerite G.; Spaulding, Dean T.; Voegtle, Katherine H. (2010). Methods in Educational Research: From Theory to Practice. Wiley. ISBN 978-0-470-58869-7.
11. Igwenagu C.M (2006). Collection of Data. Basic Statistics and Probability 147-160. First Ed. Prince Press and Communications ISBN 978 052 761 34
12. Igwenagu C.M (2006). Statistical Estimation. Basic Statistics and Probability 131. First Ed. Prince Press and Communications ISBN 978 052 761 34
13. Igbokwe Philomemna(2009) Introduction to Research and Research Methodology.
14. Research Methodology and Grant writing pg1-10 published by Rex Charles & Patrick Ltd. Anambra State, Nigeria ISBN:978-978-49233-0-9
15. Shabbir, T., M Nadeemullah, & Saeed Memon. (2020). Uses and Impact of 'Open Data' Technology for Developing Social Sector in Pakistan. Pakistan Journal of Multidisciplinary Research, 1(1), 50-64. Retrieved from https://www.pjmr.org/pjmr/article/view/24