Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/57/20241324

Use of proper sampling techniques to research studies

Xinyuan Wang

Statistics, Jilin University of Finance and Economics, ChangChun, 130117, China

3119233778@qq.com

Abstract. This article extensively examines several sampling strategies, emphasising probability and non-probability sampling approaches. Sampling is essential in research as it influences the degree to which research findings accurately represent and can be applied to a larger population. This article provides a definition of probability sampling and explores its different approaches. The text highlights the benefits of probability sampling, such as its capacity to guarantee impartial selection and facilitate statistical inference. Furthermore, this article delves into the constraints and factors to be taken into account regarding probability sampling, including the need for resources and the possibility of mistakes in sample frames. Subsequently, this paper examined non-probability sampling strategies and investigated methodologies such as convenience sampling, purposive sampling, and quota sampling. This article examines the fundamental concepts of non-probability sampling and highlights its specific applicability in investigating groups that are rare or challenging to access. A comprehensive examination was undertaken to assess the benefits and drawbacks of non probability sampling, with a particular focus on problems such as researcher bias and the absence of statistical generalizability. This article offers a thorough examination of probability and non-probability sampling approaches, allowing scholars to acquire a more profound comprehension of the benefits, constraints, and suitable uses of each approach. It acknowledges the significance of choosing the most suitable sampling method according to research goals, available resources, and the attributes of the target population. In conclusion, this work offers significant resources for researchers who wish to make well-informed choices on sampling procedures in their research.

Keywords: Sampling techniques, Non-probability theory, Probability theory, purposive sampling, quota sampling.

1. Introduction

Sampling is a crucial component of research methodology that allows researchers to make significant inferences about bigger populations by analysing smaller samples. Opting for suitable sample strategies is essential to guarantee the validity and reliability of study findings. Comprehending numerous sample procedures and their applications is vital for researchers across many fields[1].

This article aims to offer a thorough review and comparative study of probability and non-probability sampling approaches. Through an analysis of the benefits, constraints, and uses of each approach, researchers will get a more profound comprehension of the aspects that impact the choice of sampling methods. The purpose of this page is to help researchers choose the most appropriate sampling method

DOI: 10.54254/2755-2721/57/20241324

for their particular research goals, features of the population being studied, and constraints in available resources.

This article will present researchers with a range of sampling design options by examining probability sampling techniques like simple random sampling, cluster sampling, and systematic sampling, as well as non-probability sampling techniques such as convenience sampling, purposive sampling, and quota sampling.

Furthermore, this essay will examine the pragmatic importance of implementing these sampling strategies in actual research situations. Through the examination of case studies and the evaluation of the merits and drawbacks of various sampling procedures, researchers will acquire a more profound comprehension of the influence of sampling approaches on study outcomes.

This article aims to enhance the existing knowledge on sampling procedures, empowering researchers to make well-informed choices regarding the selection and utilisation of sample methods in their own research endeavours.

2. Non-probability theory

2.1. Convenience sampling

Convenience sampling is a non-probabilistic sampling technique where researchers choose individuals who are readily available and willing to take part in the study. This technology is commonly employed in scenarios when obtaining representative samples from the target population is challenging due to constraints such as time, resources, or other limiting factors.

In convenience sampling, researchers typically recruit participants based on their proximity to the study site, likelihood and willingness to participate, or their connection to the researcher or research topic. Examples of convenience sampling include surveys conducted on university campuses, recruiting online survey respondents through social media, or collecting data from individuals who happen to pass through a research laboratory.

Although convenient sampling may seem like a simple and cost-effective way to obtain data, it has several limitations. One of the main drawbacks of convenient sampling is that it may not accurately represent the entire population being studied. Participants who are easily accessible and willing to participate may not be able to represent a larger population, resulting in biased or unrepresentative outcomes. In addition, the small sample size typically associated with convenient sampling may limit the generalizability of research results.

Despite limitations, convenient sampling may be a useful method in certain research environments, such as exploratory studies or pilot projects, where the main objective is to collect preliminary data before conducting larger scale studies. However, researchers must be cautious when using convenience sampling and be aware of the potential biases and limitations of this method.

2.2. Purposive sampling

Targeted sampling is the process of selecting "typical" or "representative" units from the population to form samples based on individual knowledge, experience and judgment.

The advantage of this sampling method is that it can give play to the subjective initiative of relevant personnel and make use of existing relevant information. The disadvantage is that there is no uniform standard for the best method of selecting representative units or how to determine which units are most representative.

The deliberate selection of primary studies for inclusion in the qualitative evidence synthesis offered both advantages and disadvantages. We were able to attain a sufficiently extensive geographical distribution of primary research while restricting the number of studies included in the synthesis. This allowed us to incorporate studies with comprehensive data and research that closely aligned with the synthesis aims. Nevertheless, it is possible that we have failed to consider primary studies that did not match the sample criteria but would have been valuable for the synthesis. In addition, this study

DOI: 10.54254/2755-2721/57/20241324

employed a thematic approach to synthesise qualitative evidence. Various synthesis methodologies may have guided us towards distinct sample methods or have shown disparate discoveries.

In the future, it is necessary to do research on purposive sampling for qualitative evidence synthesis in order to evaluate the strength and reliability of various sampling frameworks. Further investigation is required to determine the most effective method for evaluating the level of data richness in qualitative primary investigations.The user's text is empty.

2.3. Quota sampling

The sampler obtains a sample that is roughly proportional[2] to the variety show structure in certain characteristics according to the prescribed quota. It first classifies the population according to certain criteria and allocates a quota of units to be investigated for each category in proportion. Then, the sampler conducts a judgment sampling for each category.

But it has the same problem the purposive sampling has.

3. Probability theory

3.1. Simple random sampling

Simple random sampling is a probabilistic sampling approach employed in research methods. Simple random sampling is the most fundamental form of sampling, in which every member of a population has an equal probability of being selected for the sample. Additionally, all feasible samples of a specific size have an equal chance of being chosen. The user's text is enclosed in tags.This approach guarantees that the sample accurately reflects the entire population, minimising the possibility of bias and enhancing the capacity to apply the findings to a wider context.

In order to conduct a basic random sample, researchers must initially establish the specific group they intend to investigate. Subsequently, it is imperative for them to choose an appropriate sample size, which is contingent upon various aspects including the research inquiry, the desired level of accuracy, and the accessible resources. After determining the sample size, researchers employ a randomization technique, such as a random number generator or a table of random numbers, to choose participants from the population.

Simple random sampling offers the key benefit of guaranteeing equal opportunity for every person of the population to be chosen, hence minimising bias and enhancing the representativeness of the sample. Additionally, it enables statistical inference, wherein the conclusions drawn from the sample can be extrapolated to the broader population with a certain degree of certainty.

Nevertheless, there are constraints associated with employing basic random sampling. It can be both laborious and expensive, particularly when managing extensive populations. Moreover, if the sampling frame is inaccurate or outdated, the sample may not accurately reflect the population[3].

3.2. Cluster sampling

Cluster sampling is a sampling technique commonly used in statistics where the entire population is divided into groups, or "clusters," and a random sample of these clusters is chosen for analysis. [4]Ideally, each cluster should be a small representative of the entire population. Cluster sampling is usually used when it is impractical to carry out simple random sampling for the whole group due to scale, cost or logistics considerations.

Adaptive cluster sampling (ACS) is extended to the case when the primary sampling units consist of circular field plots. When conducting field work for environmental monitoring, circular field plots are often preferred as they are easily set up by field workers. ACS was developed by tessellating the area frame into square plots. By using a two-phase sampling procedure, a first-phase sample of circular field plots can be established as the primary sampling units, from which ACS can be performed.[5]

Next is an overview of the advantages and disadvantages of cluster analysis.

DOI: 10.54254/2755-2721/57/20241324

3.2.1. Advantages: First, it is cost-effective and time saving, especially for a large and extensive population. Second, it is easier to manage than other complex sampling technologies. Third, it is feasible when the list of all individuals in the population does not exist but the list of clusters is available.

3.2.2. Disadvantages: First, compared with simple random sampling, the sampling error is higher, which may reduce the sampling accuracy. Second, if clusters cannot accurately represent the population, the selection and composition of clusters may introduce bias. Third, analysis can be complex, especially in multistage cluster sampling.

3.3. Systematic sampling

Systematic sampling is another important sampling method used in statistics and research. It offers an alternative to simple random sampling, especially in situations where a complete list of the population is available.

Here are the advantages disadvantages.

3.3.1. Advantages: First and foremost, the key advantages are simplicity and ease of execution.Furthermore, it is more efficient and economical in terms of time and money when compared to basic random sampling.Thirdly, it guarantees a comprehensive representation of the sample throughout the entire population.

Furthermore, the LPMxy exhibited a high degree of similarity to the systematic design with relation to the majority of target variables. The relative efficiencies (REs) of the LPM designs incorporating auxiliary data, as compared to the systematic design, ranged from 0.74 to 1.18, depending on the target variable under study. The SRS estimator for variance exhibited a predictable bias and shown a tendency to be overly cautious in its estimation. In the case of LPMxy, the Grafström-Schelin estimator yielded overestimated results. The user's text is "[6]".

3.3.2. Disadvantages: Potential for periodicity bias: If there is a hidden pattern in the population that coincides with the sampling interval, it can lead to biased results.

Not as random as simple random sampling, as only the starting point is chosen randomly.

Inflexibility once the interval is set.

4. Conclusion

To summarise, this work has analysed the various approaches to sampling, including probability and non-probability methods, each with their distinct benefits and limitations that are suitable for different research situations. Probability sampling approaches, such as simple random sampling, stratified sampling, cluster sampling, and systematic sampling, are notable for their ability to provide samples that accurately represent populations. These methods guarantee that every individual in the population has a recognised and typically equal probability of being chosen, enabling the findings to be extrapolated to a broader population with measurable levels of confidence.

Alternatively, in cases where the research context prohibits random selection or when the objective is to examine specific attributes of the population, non-probabilistic sampling approaches such as convenience sampling, judgement sampling, and quota sampling become essential. These technologies are typically more economical and efficient in terms of time, but they have limitations in terms of scalability and potential bias.

The choice of sampling methods relies on the study objectives, characteristics of the research population, resource availability, and desired level of accuracy. If the conditions allow, probability sampling is the preferred method for research that aims to achieve high external validity. Conversely, non-probabilistic sampling is beneficial when exploratory research or practice restricts the stated method.

Finally, the advantages and limitations of each sampling technique are recognized so that researchers can make wise decisions to improve the credibility and reliability of their research results. Researchers

must carefully document their sampling process, address potential biases, and discuss the extent to which research results are disseminated. Therefore, judicious selection and application of sampling methods is still the cornerstone of robust and effective research across all disciplines.

References

[1] Li Jinchang. Applied Sampling Techniques (Third Edition) [M]. Science Press, 2015
[2] Encyclopedia of Quality of Life and Well-Being Research Quota Sampling ISBN: 978-94-007-0752-8 Deden Rukmana
[3] West, PW. Simple random sampling of individual items in the absence of a sampling frame that lists the individuals[J]. New Zealand Journal of Forestry Science volume 46
[4] Ames H, Glenton C, Lewin S. Purposive sampling in a qualitative evidence synthesis: A worked example from a synthesis on parental perceptions of vaccination communication[J]. Bmc Medical Research Methodology, 2019, 19(1). DOI: 10.1186/s12874-019-0665-4.
[5] Two-phase adaptive cluster sampling with circular field plots Wilmer Prentius, Anton Grafström
[6] Comparison of the local pivotal method and systematic sampling for national forest inventories Minna Räty, Mikko Kuronen, Mari Myllymäki, Annika Kangas, Kai Mäkisara & Juha Heikkinen,Forest Ecosystems 7, Article number: 54 (2020)