ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 04

Evaluation of Sampling Error and Non-Sampling Error in Research

Jyoti
Teacher
Kasturba Gandhi Balika Vidyalay
Lucknow

Abstract

Errors in data collection and analysis are inevitable in any research process, but understanding their nature is crucial for ensuring reliability and validity. Two broad categories of errors affect research quality: sampling error and non-sampling error. Sampling error arises because only a subset of the population is observed, leading to natural variability between the sample and the population. Non-sampling error, on the other hand, encompasses a wide range of issues, including non-response, measurement bias, coverage error, and data processing mistakes, which often pose greater threats to research validity. This paper evaluates both types of errors in terms of their definitions, causes, measurement, and impact. Using illustrative examples from surveys, censuses, and opinion polls, the study compares the magnitude of sampling and non-sampling errors. The analysis reveals that while sampling error is quantifiable and decreases with larger sample sizes, non-sampling error is more difficult to detect and can remain substantial even in large surveys. Results also highlight how inappropriate research designs, respondent behavior, and data handling processes exacerbate non-sampling errors. Strategies such as improved training of enumerators, pilot studies, statistical adjustments, and use of digital technologies are discussed as ways of minimizing both error types. Ultimately, the evaluation underscores that accurate research depends not only on selecting appropriate sample sizes but also on minimizing non-sampling errors that may otherwise bias findings beyond correction.

Keywords: Sampling error, non-sampling error, Bias, Data quality, Research methodology, Statistical evaluation

Introduction

Every research process, whether social, economic, or scientific, is subject to imperfections in data collection and measurement. Errors undermine the quality of research outcomes and can mislead policy decisions, academic conclusions, or business strategies. Among the different types of errors, sampling error and non-sampling error constitute two fundamental categories that require careful evaluation.

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 04

Sampling error is a statistical phenomenon that occurs when information is derived from a sample rather than the entire population. Since no sample is a perfect mirror of the population, differences naturally arise. These discrepancies—though random—can be measured through statistical tools such as the standard error, confidence intervals, and margins of error. For instance, if we estimate the average income of households using a random sample of 1,000 families from a city of 1 million households, the resulting mean will likely differ from the true population mean. This difference represents sampling error.

Non-sampling error, however, is far broader in scope and often more problematic. It refers to all other types of inaccuracies that are not related to sampling variability. These include non-response error (when some respondents fail to participate), measurement error (when questions are misunderstood or poorly designed), coverage error (when parts of the population are excluded), and processing error (when mistakes occur during coding, entry, or analysis). Unlike sampling error, non-sampling error does not decrease automatically with larger samples. In fact, increasing the sample size without addressing non-sampling issues may amplify the problem.

The significance of evaluating both error types lies in their combined effect on research reliability. While sampling error is generally well-understood and mathematically estimable, non-sampling error often goes undetected until results are critically reviewed. Both errors reduce the validity of research outcomes, but they require different strategies for prevention and correction.

This paper provides a systematic evaluation of sampling error and non-sampling error. It reviews relevant literature, outlines methodological frameworks for measurement, presents illustrative data tables comparing their effects, and discusses implications for researchers and policymakers.

Review of Related Literature

The issue of sampling error has long been studied in statistics and survey methodology. Cochran (1977) provided foundational frameworks for measuring sampling error using probability theory, introducing concepts of simple random sampling, stratification, and cluster sampling. More recent works (Lohr, 2019; Särndal, 2011) have extended these principles by applying them to large-scale national surveys and complex survey designs.

Studies emphasize that the magnitude of sampling error decreases with larger sample sizes, but its distribution depends on design effects. Kish (2015) highlighted that poorly designed samples can increase variance even when the sample size is large.

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 04

In contrast, non-sampling error has gained increasing attention in the past two decades. Groves et al. (2011) categorized non-sampling errors into four major groups: coverage, measurement, non-response, and processing errors. They argue that non-sampling errors often dominate the total survey error, particularly in social research where respondent behaviors and interviewer effects distort results.

Examples from real-world studies further illustrate this. The U.S. Census Bureau (2017) reported that while sampling error was small in large censuses, non-sampling errors such as undercounting minorities and non-response bias significantly distorted demographic estimates. Similarly, Indian National Sample Survey reports (NSSO, 2018) highlighted issues of underreporting household expenditures due to recall bias, a clear case of non-sampling error.

Election polling has also revealed the importance of non-sampling errors. The 2016 U.S. presidential election polls (Kennedy et al., 2017) showed that while sampling error margins were within ±3%, non-sampling errors like social desirability bias and misreporting led to inaccurate forecasts.

Overall, literature suggests that while sampling error is scientifically manageable, non-sampling error remains a persistent challenge in research methodology.

Objectives of the Study

1. To evaluate the nature, measurement, and implications of sampling error in research.
2. To identify major sources of non-sampling error and their impact on data quality.
3. To compare the relative significance of sampling and non-sampling errors.
4. To suggest strategies for minimizing both error types in research practice.

Methodology

This study adopts a secondary research approach, analyzing existing statistical theories, survey reports, and case studies to evaluate sampling and non-sampling errors. To illustrate their comparative effects, hypothetical survey data have been generated, and statistical formulas are applied to estimate sampling errors. Non-sampling errors are assessed based on reported rates from published surveys (e.g., NSSO, Census Bureau, Gallup polls).

Measurement of Sampling Error

Sampling error is typically measured using the formula for the standard error of the mean (SEM):

$$SE = \sigma/\sqrt{n}$$

Where:

- σ = population standard deviation
- n = sample size

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 04

Confidence intervals (CI) are then constructed to estimate the population parameter within a given range.

Measurement of Non-Sampling Error

Unlike sampling error, non-sampling error cannot be expressed by a single formula. Instead, it is assessed through survey audits, re-interviews, validation studies, and comparison with administrative records.

For this study, non-sampling error categories include:

- **Non-response error**: measured by response rates and demographic analysis.
- **Measurement error**: detected by comparing survey responses with external benchmarks.
- **Coverage error**: assessed by comparing sampling frames with population lists.
- **Processing error**: identified through data entry validation checks.

Results and Discussion

Table 1: Sampling Error Estimates (Hypothetical Data)

Sample Size (n)	Population Mean	Sample Mean	Standard Error	95% CI Range	Sampling Error (%)
100	50	48.5	2.1	46.4 – 50.6	3.0%
500	50	49.3	0.95	48.4 – 50.2	1.4%
1000	50	49.7	0.67	49.0 – 50.4	0.6%

Interpretation:

As sample size increases, the standard error decreases, and the sample mean converges toward the population mean. Thus, sampling error is predictable and controllable through proper sample design.

Table 2: Sources of Non-Sampling Error (Based on Survey Studies)

Error Type	Description	Estimated Impact (%)	Example from Studies
Non-response Error	Respondents refusing or unavailable	5–10%	NSSO surveys (2018)
Response Bias	Misreporting/false answers	3–7%	U.S. election polls
Processing Error	Data entry/coding mistakes	1–4%	Census (2017)
Coverage Error	Excluded groups from sampling frame	2–6%	Migrant population

Interpretation:

Unlike sampling error, non-sampling errors are not automatically reduced by larger samples. For example, even a survey of 50,000 households may suffer from high non-response bias if marginalized groups are systematically underrepresented.

Comparative Evaluation

Quantifiability

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 04

- Sampling error is mathematically measurable because it is based on probability theory. Using formulas for standard error and confidence intervals, researchers can estimate the extent of variability caused by observing only a sample rather than the entire population.
- Non-sampling error, however, is often hidden and unquantifiable. Bias due to poor questionnaire design, respondent misreporting, or non-response cannot be captured by simple formulas and often requires independent validation studies.

Impact of Sample Size

- In theory, as the sample size increases, sampling error decreases because estimates become more stable and closer to the population parameter. For instance, doubling the sample size reduces the standard error roughly by the square root of two.
- Non-sampling error does not automatically decline with larger samples. In fact, larger surveys may face higher risks of interviewer errors, processing mistakes, and coverage gaps, which can magnify non-sampling biases.

Correctability

- Sampling error can be corrected or reduced through statistical adjustments such as weighting, stratification, or using advanced estimators.
- Non-sampling error usually requires improvements in research practice—for example, better fieldwork training, clearer survey instruments, pilot testing, or quality control measures. Unlike sampling error, it cannot be solved by mathematics alone.

Overall Risk

- Sampling error tends to be random in nature, which means that across repeated samples, positive and negative deviations tend to cancel out.
- Non-sampling error is usually systematic and directional, meaning it consistently biases results in one direction. As a result, it is more damaging to the credibility of research outcomes and often more difficult to detect.

Conclusion

This paper evaluated sampling error and non-sampling error in research, highlighting their definitions, causes, and impacts. The analysis demonstrated that sampling error is quantifiable, predictable, and reducible through appropriate sample size and design. Non-sampling error, however, poses a greater challenge due to its hidden, systematic, and often unpredictable nature. Errors such as non-response bias, coverage error, and data processing mistakes can significantly distort findings, even in large and well-designed surveys.

ISSN 2584-041X(P)
ISSN 3048-7730(E)
VOL.06 ISSUE 04

The evaluation suggests that while statistical methods can control sampling error, minimizing non-sampling error requires organizational, procedural, and technological interventions. Training of enumerators, rigorous pre-testing of instruments, use of digital data collection, validation checks, and triangulation with administrative records are essential strategies.

Ultimately, reliable research depends on recognizing the dual threats of sampling and non-sampling errors and adopting comprehensive quality assurance measures. By integrating statistical rigor with fieldwork discipline, researchers can minimize errors and produce findings that truly reflect reality.

References

1. **Cochran, W. G. (1977).** *Sampling techniques*. John Wiley & Sons.
2. **Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011).** *Survey methodology*. John Wiley & Sons.
3. **Kennedy, C., Blumenthal, M., Clement, S., Clinton, J., Durand, C., Franklin, C., ... & Witt, E. (2017).** An evaluation of 2016 election polls in the United States. *Public Opinion Quarterly, 81*(1), 1-30.
4. **Kish, L. (2015).** *Survey sampling*. John Wiley & Sons.
5. **Lohr, S. L. (2019).** *Sampling: Design and analysis*. Chapman & Hall/CRC.
6. **National Sample Survey Office (NSSO). (2018).** *Household Consumer Expenditure Report*. Government of India.
7. **Särndal, C. E. (2011).** The 3Ps: Prediction, prospective, and perspective. *Survey Methodology, 37*(2), 151-163.
8. **U.S. Census Bureau. (2017).** *Coverage measurement in the 2010 census*. U.S. Department of Commerce.