Sampling Techniques and Errors

Sampling techniques are essential for collecting data efficiently when surveying an entire population is impractical or too costly. Random sampling methods aim to ensure that samples represent the population by reducing bias and allowing for accurate generalizations. Nonrandom sampling methods, while simpler, often introduce bias and limit the reliability of conclusions. This section explores common sampling techniques, errors that can arise during sampling, and methods to mitigate these issues.

Sampling Techniques

Sampling is used when gathering data from an entire population is impractical or too expensive. A sample should represent the population's characteristics, and random sampling methods help achieve this.

What are the Common Sampling Techniques?

These are not an exhaustive list, just the most common types you are likely to encounter.

Random Sampling Techniques

- **Simple Random Sampling:** Every individual in the population has an equal chance of being selected, ensuring that all samples of the same size are equally likely.

- **Stratified Sampling:** Divide the population into groups (strata) and take a proportionate random sample from each group. For instance, sample students from various departments to represent a college population.

- **Cluster Sampling:** Divide the population into clusters, then randomly select some clusters and include all members from those clusters in the sample. For example, randomly select departments in a college and survey all their students.

- **Systematic Sampling:** Randomly select a starting point and pick every nth member from a population list. For example, choose every 50th name in a phone book for a survey

Non-Random Sampling Techniques

- **Convenience Sampling:** Individuals are chosen based on accessibility and ease rather than randomness. For example, Surveying the first 20 people entering a grocery store about their shopping habits.

Nonrandom sampling methods, such as **convenience sampling**, should generally be avoided as they can lead to biased results. For example, surveying the first 20 customers entering a coffee shop might overrepresent certain groups, such as early risers or individuals with flexible schedules, while excluding others, like night-shift workers.

Random sampling methods aim to reduce bias and ensure representativeness by giving all individuals an equal chance of selection. However, even random sampling is subject to **random sampling error**, where chance variations may cause the sample to differ from the population. Larger sample sizes reduce this error, improving the reliability of the results. Researchers must also address **nonrandom sampling errors** like undercoverage, nonresponse, and volunteer response bias, as these can further skew results. Well will discussing these sampling errors shortly.

For now, we want to focus on identifying the common types of sampling techniques.

Example 19

A study intends to determine the average tuition paid by Tennessee Technological University undergraduate students per semester. Identify the sampling method used in each scenario:

- **Part A:** Organize students by year (e.g., first-year, sophomore), then select 25 from each.

- **Part B:** Use a random number generator to select one student, then pick every 50th student until 75 students are included.

- **Part C:** Select 75 students completely at random, with equal probability for all.

- **Part D:** Randomly pick two years (e.g., first-year and senior), and survey all students in those years.

- **Part E:** Survey the first 100 students encountered in front of the library on a specific day.

Solution: Identify Sampling Techniques

- **Part A: Stratified Sampling:** Students are grouped by year and sampled proportionally.

- **Part B: Systematic Sampling:** A starting point is randomly selected, and every 50th student is chosen.

- **Part C: Simple Random Sampling:** All students have an equal probability of selection.

- **Part D: Cluster Sampling:** Entire years are selected, and all students in those years are included.

- **Part E: Convenience Sampling:** Students are chosen based on availability, leading to potential bias.

■

Sampling Errors

What is a Random Sampling Error?

A **random sampling error** occurs when there is a discrepancy between a sample result and the true population result. This type of error arises purely due to chance, as a sample is only a subset of the population and may not perfectly represent it. For example, if a random sample of 50 students is drawn from a school of 1,000, the sample mean test score might differ slightly from the population mean simply by chance.

Key Point: Random sampling error is unavoidable in any sampling process but can be minimized by increasing the sample size.

What is a Nonrandom Sampling Error?

There are three important types of nonrandom sampling errors: **undercoverage**, **nonresponse**, and **volunteer response**.

- **What is Undercoverage?**
  Undercoverage occurs when some groups in the population are systematically excluded from the sampling process, making the sample unrepresentative of the entire population.
  Example
  A survey conducted by randomly calling landline phone numbers will exclude people who only use cell phones, such as younger adults, leading to biased results.

- **What is Nonresponse?**
  Nonresponse occurs when a portion of the selected sample does not participate in the study, and their responses may differ systematically from those who do respond.
  Example
  In a mail survey asking about personal finances, people who feel uncomfortable sharing financial information may choose not to respond, skewing the results toward those who are more open about their finances.

- **What is a Volunteer Response?**
  Volunteer response occurs when participants self-select into the sample rather than being randomly chosen, often leading to an overrepresentation of individuals with strong opinions.
  Example
  An online poll about a controversial policy is likely to attract participants who feel strongly for or against the policy, while those who are neutral or indifferent are less likely to respond.

Example 20

A university is conducting a study to understand student preferences for campus dining options. The following scenarios describe how data was collected. Identify the type of nonrandom sampling error for each part and explain your reasoning.

- **Part A:** The university posts a survey link on its social media pages and encourages students to share their opinions. Most responses come from students who are either very satisfied or very dissatisfied with campus dining options.

- **Part B:** The university randomly emails 1,000 students asking them to complete the survey. Only 250 students respond, and many of the responses are from students who frequently use campus dining services.

- **Part C:** The university conducts the survey at a single campus dining hall during lunchtime. Students who do not use the dining hall or eat lunch on campus are excluded from the sample.

Solution

- **Part A: Volunteer Response**

The survey link was shared publicly, allowing students to self-select into the sample. This leads to volunteer response bias, as students with strong opinions are more likely to participate, while those with neutral opinions are underrepresented.

- **Part B: Nonresponse**
  Out of the 1,000 students contacted, only 250 responded, creating nonresponse bias. The opinions of nonrespondents may differ systematically from those of respondents, particularly since frequent diners are overrepresented in the responses.

- **Part C: Undercoverage**
  The survey was conducted at a single dining hall during lunchtime, excluding students who do not use the dining hall or eat lunch on campus. This undercoverage results in a sample that is not representative of the entire student population.

■

What is a Non-sampling Error?

A **non-sampling error** is caused by human error or flaws in the data collection process. These errors can occur regardless of how the sample is chosen. Examples include mistyping data into a computer, misinterpreting survey questions, or using faulty measuring instruments.

Key Point: Non-sampling errors can often be reduced by careful planning, training, and implementing quality control measures during data collection and entry.

Conclusion

Effective sampling is key to obtaining reliable and meaningful data. Random sampling methods, like stratified and systematic sampling, reduce bias and provide representative samples, while nonrandom methods, such as convenience sampling, often produce unreliable results. Identifying and addressing common sampling errors, including undercoverage and nonresponse, further improves the quality of data collection. By carefully selecting sampling techniques and minimizing errors, researchers can ensure their findings are accurate and applicable.