About Sampling
# TABLE OF CONTENTS

- About Sampling ............................................................................................................ 1
  - Why is SAMPLING important? .................................................................................. 1
- Common Language ........................................................................................................ 1
- Probability & Nonprobability Sampling ..................................................................... 1
- Sampling Methods ...................................................................................................... 2
  - Simple Random Sample (SRS) ................................................................................ 2
  - Stratified Random Sampling ................................................................................... 2
  - Cluster Sampling ..................................................................................................... 2
  - Systematic Sampling ............................................................................................... 2
  - Availability Sampling ............................................................................................... 3
  - Snowball Sampling .................................................................................................. 3
  - Quota Sampling ....................................................................................................... 3
  - Dimensional Sampling ............................................................................................. 3
- Difficulties in Sampling ............................................................................................... 4
  - Using the Wrong Sampling Frame .......................................................................... 4
  - Not reaching the individuals selected ...................................................................... 4
  - Having a low response rate ...................................................................................... 4
  - Problems with generalizability ............................................................................... 4
- Estimating Sample Size ............................................................................................... 5
  - Research Hypothesis ............................................................................................... 5
  - Precision ................................................................................................................ 5
  - Population homogeneity ......................................................................................... 5
  - Sampling fraction .................................................................................................... 6
  - Sampling technique ............................................................................................... 6
About Sampling

Why is SAMPLING important?

- With the majority of studies, we are trying to make inferences about the population as a whole. A classic example would be linking smoking to lung cancer. However, the task to collect information about the population as a whole is often unrealistic, so instead we take a sample of the population of interest and infer results of this sample to the population.

Common Language

Probability & Nonprobability Sampling

- Probability sampling is a method of ensuring that each member of the population has approximately equal chance of being included in the sample.

  Using probability sampling, we calculate sampling error, which estimates the difference between results drawn from the sample and actual values within the population.

- Nonprobability sampling occurs when the investigator is not concerned with balanced sample representation.

  This form of sampling is meant to measure a possible relationship between the independent and dependent variable, without generalizing to the larger population. It may also be used to uncover "hidden" populations, not accessible by probability sampling.
Sampling Methods

Simple Random Sample (SRS)

- Simple random sampling is a form of probability sampling, with two requirements:

  1) a list of units in the population
  2) a source of random number generation

Stratified Random Sampling

- A stratified random sample collects information by dividing the population into groups (strata), such as breaking a province into counties, and then performing an SRS on each strata.

  Stratified random sampling is advantageous in that it is a form of probability sampling which reduces variability common to a typical SRS.

Cluster Sampling

- Often confused with stratified random sampling, cluster sampling divides the population into groups, called clusters. It is different than its stratified cousin, in that a random sample of clusters is selected, as opposed to random sampling being performed within each cluster.

  One advantage to this technique lies in the reduced volume in dealing with a list of clusters, instead of every individual within the sample.

Systematic Sampling

- In systematic sampling, the list would be divided into as many consecutive segments as needed, with the same randomly chosen starting point in each segment.
Availability Sampling

- We now come to our first form of nonprobability sampling.

Availability sampling involves the researcher using subjects that are easily available. This is used when a complete sampling frame is extraordinarily difficult to develop, as with learning disabled individuals. These samples are often used in experimental or quasi-experimental research, and characteristically lack generalizability.

Snowball Sampling

- In snowball sampling, as we find new cases, those continually lead us to more cases. This is akin to a snowball gaining in size as it rolls down a hill. Due to the relation between cases, this sampling method is often used in studying subcultures. Unfortunately, it may miss those isolated from particular social networks (e.g. not all homeless people use homeless shelters).

Quota Sampling

- Like previous methods, quota sampling involves breaking the population into multiple categories. From these categories, a quota is set on the number of members permitted to each category. This method is differentiated from stratified sampling, in that it depends on availability of members, and is ideal where quick collection of data is more important than representativeness of that data.

Dimensional Sampling

- Dimensional sampling is used when the researcher wishes to increase the representativeness of their sample. This involves two steps:

  1) specify all important dimensions/variables
  2) choose a sample that includes at least one case where each possible combination of dimensions has occurred
Difficulties in Sampling

Using the Wrong Sampling Frame

- This error occurs when wanted members are overlooked, or unwanted members are included. An example of a wrong sampling frame might include a phone book, which excludes a large subset of the population with unlisted numbers. The primary remedy is to change to a more inclusive frame.

Not reaching the individuals selected

- Bias can be introduced into the sample if the researcher easily gives up on individuals not reached initially. This highlights the value of putting time and effort into a smaller sample.

Having a low response rate

- Response rate is beyond the researcher’s control, and should be reported in research summaries. With lower response rates, generalizability to the larger population is impeded, as the differences between the respondents and non-respondents may be drastic.

Problems with generalizability

- This issue often occurs with nonprobability sampling. Results from these samples can only be applied to one particular setting and group, without generalizing to others.
Estimating Sample Size

- The key concern in determining your sample is that it is representative of your population. Generally, the larger your sample, the more representative it is; however, where this is financially infeasible, the following factors may apply:

  1) Research hypothesis
  2) Level of precision
  3) Homogeneity of the population
  4) Sampling fraction
  5) Sampling technique

Research Hypothesis

- Researchers require a number of cases to analyze in testing their hypotheses. There is no accepted minimum for a sample size, but a number of researchers set 100 as their absolute minimum. Below this threshold, the statistical procedures used may be questioned.

Precision

- This is representative of how much sampling error the researcher is willing to allow. Directly related to sample size, precision increases as the sample size grows. There is always the chance that a sample drawn is unrepresentative of the target population, so researchers use confidence intervals to indicate levels of precision in sampling.

Population homogeneity

- This is a measure indicative of the variability of a sample. Unfortunately, homogeneity is a crude measurement, as it is difficult for researchers to know or estimate the homogeneity of their target population beforehand.
**Sampling fraction**

- Is the number of subjects in the sample relative to the number of subjects in the population. At times, the fraction may be larger than reasonably required. In this case, the following equation is used for adjusting the sample size:

\[ n' = \frac{n}{1 + (n/N)} \]

\( n' \) = adjusted sample size  
\( n \) = estimated sample size ignoring the sampling fraction  
\( N \) = population size

**Sampling technique**

- Whereas the previous factors discussed were for simple random sampling, changing the sampling technique often leads to a change in the required sample size.
**Glossary**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability sampling:</td>
<td>involves the researcher using subjects that are easily available.</td>
</tr>
<tr>
<td>Census:</td>
<td>Survey which measures the entire population.</td>
</tr>
<tr>
<td>Cluster sampling:</td>
<td>divides the population into groups, called clusters, then a random sample of clusters is selected, as opposed to a SRS within each cluster.</td>
</tr>
<tr>
<td>Dimensional sampling:</td>
<td>involves identifying important dimensions/variables and drawing a sample which includes at least one subject with each possible combination of dimensions.</td>
</tr>
<tr>
<td>Nonprobability Sampling:</td>
<td>The researcher is unaware of the probability for each subject in the population to be included in the sample.</td>
</tr>
<tr>
<td>Population:</td>
<td>The membership of the group under investigation, as a whole.</td>
</tr>
<tr>
<td>Probability Sampling:</td>
<td>Sample which allows each subject in a population some chance at being selected.</td>
</tr>
<tr>
<td>Quota sampling:</td>
<td>involves breaking the population into multiple categories. From these categories, a quota is set on the number of members permitted to each category.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Representative sample:</td>
<td>A sample that accurately reflects the target population.</td>
</tr>
<tr>
<td>Sample:</td>
<td>A subset of the population, who are actually measured.</td>
</tr>
<tr>
<td>Sample survey:</td>
<td>Measurements drawn on the sample chosen.</td>
</tr>
<tr>
<td>Sampling frame:</td>
<td>List from which the sample is drawn, with the hopes that is representative of the entire population.</td>
</tr>
<tr>
<td>Simple random sampling:</td>
<td>a form of probability sampling, in which units within the population are randomly selected for the sample.</td>
</tr>
<tr>
<td>Snowball sampling:</td>
<td>new cases lead to new cases, which lead to new cases, etc.</td>
</tr>
<tr>
<td>Stratified random sampling:</td>
<td>collects information by dividing the population into groups (strata), and then performing an SRS on each strata.</td>
</tr>
<tr>
<td>Systematic sampling:</td>
<td>the list is divided into as many consecutive segments as needed, with the same randomly chosen starting point in each segment.</td>
</tr>
<tr>
<td>Unit:</td>
<td>One member of the group of persons or objects under investigation.</td>
</tr>
</tbody>
</table>
References
