Sampling: Design and Procedures

Agenda

1) Overview
2) Sample or Census
3) The Sampling Design Process
   i. Define the Target Population
   ii. Determine the Sampling Frame
   iii. Select a Sampling Technique
   iv. Determine the Sample Size
   v. Execute the Sampling Process
4) A Classification of Sampling Techniques
   i. Nonprobability Sampling Techniques
      a. Convenience Sampling
      b. Judgmental Sampling
      c. Quota Sampling
      d. Snowball Sampling
   ii. Probability Sampling Techniques
      a. Simple Random Sampling
      b. Systematic Sampling
      c. Stratified Sampling
      d. Cluster Sampling
      e. Other Probability Sampling Techniques

5. Choosing Nonprobability versus Probability Sampling

6. Uses of Nonprobability versus Probability Sampling
Sample or Census?

Population

The aggregate of all elements, sharing some characteristics, that comprises the universe for the purpose of the marketing research problem

Census

A complete enumeration of the elements of a population or study objects

Sample

A subgroup of the elements of the population selected for participation in the study
## Sample vs. Census

<table>
<thead>
<tr>
<th>Conditions favoring the use of</th>
<th>SAMPLE</th>
<th>CENSUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Budget</td>
<td>small</td>
<td>high</td>
</tr>
<tr>
<td>2. Time available</td>
<td>short</td>
<td>Long</td>
</tr>
<tr>
<td>3. Population size</td>
<td>large</td>
<td>small</td>
</tr>
</tbody>
</table>

---

**The Sample Plan is the process used to select units from the population to be used in the sample**

---

*Practitioner Viewpoint*

"If a sample is not correctly chosen, the research may produce misleading conclusions. Always and everywhere, careful attention to the sample plan helps researchers reach their objectives."

Larry Peterson, Principal & President, Peterson & Partners, Inc.

Sample Sampling, Inc.
The Sampling Design Process

1. Define the Population
2. Determine the Sampling Frame
3. Select Sampling Technique(s)
4. Determine the Sample Size
5. Execute the Sampling Process

Population: Aggregate of all elements, defined prior to selection of the sample
Target Population: Aggregate of elements from which the sample is drawn
Sampling Frame: List of all the sampling units available for selection at the stage of the sampling process
Sample: Elements selected for analysis
Define the Target Population

The target population is the collection of elements or objects that possess the information sought by the researcher and about which inferences are to be made.

The target population should be defined in terms of elements, sampling units, extent, and time.

Target Population

- An element is the object about which or from which the information is desired: respondents, products, stores, companies, families,…
- A sampling unit is an element, or a unit containing the element, that is available for selection at some stage of the sampling process.
- Extent refers to the geographical boundaries.
- Time is the time period under consideration.
**Target Population (example)**

- **Element:** Male or female head of the household responsible for most of the shopping at department stores
- **Sampling units:** Households; then male or female head of the household.
- **Extend:** Porto
- **Time:** November, 2004

**Target Population (example)**

- **Element:** Male or female head of households
- **Sampling units:** Working telephone numbers; then male or female head of households
- **Extend:** Porto
- **Time:** November, 2004
**Target Population (example)**

- **Element**: adults meeting three qualifications
  - Age 25 or older
  - Live in Algarve at least seven months of the year
  - Have a driver's license

- **Sampling units**: household with a telephone number; then adults meeting the defined qualifications

- **Extend**: Porto

- **Time**: Period of the survey

---

**Target Population (example)**

- **Element**: Our product

- **Sampling units**: Supermarkets, drugstores; then our product

- **Extend**: Porto

- **Time**: Period of the survey
**Target Population (example)**

- **Element**: Chemical engineers
- **Sampling units**: Companies purchasing over 300 000€ of chemicals per year; then chemical engineers
- **Extend**: Europe
- **Time**: 2003

**Target Population (example)**

- **Element**: Females 18-50
- **Sampling units**: Females 18-50
- **Extend**: Porto
- **Time**: November, 2004
Suppose that Revlon wanted to assess consumer response to a new line of lipsticks and wanted to sample females over 18 years of age.

- a) it may be possible to sample females over 18 directly, in which case a sampling unit would be the same as an element;
- b) alternatively, the sampling unit might be households; in the later case, households would be sampled and all females over 18 in each selected household would be interviewed; here, the sampling unit and the element are different.

- Target Population (example)

- Sampling Frame

  A representation of the elements of the target population. It consists of a list or set of directions for identifying the target population

  Examples
  - telephone book
  - association directory listing the firms in an industry
  - a mailing list purchased from a commercial organization
  - a map
Define the Target Population

Important qualitative factors in determining the sample size

- the importance of the decision
- the nature of the research
- the number of variables
- the nature of the analysis
- sample sizes used in similar studies
- incidence rates
- completion rates
- resource constraints

Sample Sizes Used in Marketing Research Studies

<table>
<thead>
<tr>
<th>Type of Study</th>
<th>Minimum Size</th>
<th>Typical Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem identification research (e.g. market potential)</td>
<td>500</td>
<td>1,000-2,500</td>
</tr>
<tr>
<td>Problem-solving research (e.g. pricing)</td>
<td>200</td>
<td>300-500</td>
</tr>
<tr>
<td>Product tests</td>
<td>200</td>
<td>300-500</td>
</tr>
<tr>
<td>Test marketing studies</td>
<td>200</td>
<td>300-500</td>
</tr>
<tr>
<td>TV, radio, or print advertising (per commercial or ad tested)</td>
<td>150</td>
<td>200-300</td>
</tr>
<tr>
<td>Test-market audits</td>
<td>10 stores</td>
<td>10-20 stores</td>
</tr>
<tr>
<td>Focus groups</td>
<td>2 groups</td>
<td>4-12 groups</td>
</tr>
</tbody>
</table>
Classification of Sampling Techniques

- **Probability Sampling**
  - a sampling procedure in which each element of the population has a fixed probabilistic chance of being selected for the sample

- **Nonprobability Sampling**
  - Sampling techniques that do not use chance selection procedures. Rather, they rely on the personal judgment of the researcher
Probability Sampling Techniques

Simple Random Sampling

- Each element in the population has a known and equal probability of selection.
- Each possible sample of a given size \( n \) has a known and equal probability of being the sample actually selected.
- This implies that every element is selected independently of every other element.
- The sample is drawn from a random procedure from a sampling frame.
### Simple Random Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population identified uniquely by number</td>
<td>Selection by random number</td>
<td>Every member of the population has an equal chance of being selected into the sample</td>
</tr>
</tbody>
</table>

### Procedures for Drawing Probability Samples

1. Select a suitable sampling frame
2. Each element is assigned a number from 1 to N (pop. size)
3. Generate n (sample size) different random numbers between 1 and N
4. The numbers generated denote the elements that should be included in the sample
Systematic Sampling

- The sample is chosen by selecting a random starting point and then picking every nth element in succession from the sampling frame.
- The sampling interval, k, is determined by dividing the population size N by the sample size n and rounding to the nearest integer.
- When the ordering of the elements is related to the characteristic of interest, systematic sampling increases the representativeness of the sample.
- If the ordering of the elements produces a cyclical pattern, systematic sampling may decrease the representativeness of the sample.

For example, there are 100,000 elements in the population and a sample of 1,000 is desired. In this case the sampling interval, k, is 100. A random number between 1 and 100 is selected. If, for example, this number is 23, the sample consists of elements 23, 123, 223, 323, 423, 523, and so on.

Tennis' Systematic Sampling Returns a Smash

Tennis magazine conducted a mail survey of its subscribers to gain a better understanding of its market. Systematic sampling was employed to select a sample of 1,472 subscribers from the publication's domestic circulation list. If we assume that the subscriber list had 1,472,000 names, the sampling interval would be 1,000 (1,472,000/1,472). A number from 1 to 1,000 was drawn at random. Beginning with that number, every 1,000th subscriber was selected.

A brand-new dollar bill was included with the questionnaire as an incentive to respondents. An alert postcard was mailed one week before the survey. A second, follow-up, questionnaire was sent to the whole sample ten days after the initial questionnaire. There were 76 post office returns, so the net effective mailing was 1,396. Six weeks after the first mailing, 778 completed questionnaires were returned, yielding a response rate of 56%.
Example

- Suppose that you want to select a random sample of 250 names from the white pages of a phone book. Let's also say that there are 55,000 names listed in the white pages. A systematic sample provides a convenient way to choose the sample.
Systematic Sampling

1. Select a suitable sampling frame
2. Each element is assigned a number from 1 to N (pop. size)
3. Determine the sampling interval \(i = \frac{N}{n}\). If \(i\) is a fraction, round to the nearest integer
4. Select a random number, \(r\), between 1 and \(i\), as explained in simple random sampling
5. The elements with the following numbers will comprise the systematic random sample: \(r, r+i, r+2i, r+3i, r+4i, \ldots, r+(n-1)i\)

Stratified Sampling

- A two-step process in which the population is partitioned into subpopulations, or strata.
- The strata should be mutually exclusive and collectively exhaustive in that every population element should be assigned to one and only one stratum and no population elements should be omitted.
- Next, elements are selected from each stratum by a random procedure, usually SRS.
- A major objective of stratified sampling is to increase precision without increasing cost.
Stratified Sampling

- The elements within a stratum should be as homogeneous as possible, but the elements in different strata should be as heterogeneous as possible.
- The stratification variables should also be closely related to the characteristic of interest.
- Finally, the variables should decrease the cost of the stratification process by being easy to measure and apply.
- In proportionate stratified sampling, the size of the sample drawn from each stratum is proportionate to the relative size of that stratum in the total population.
- In disproportionate stratified sampling, the size of the sample from each stratum is proportionate to the relative size of that stratum and to the standard deviation of the distribution of the characteristic of interest among all the elements in that stratum.

### Stratified Random Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population is separated into (e.g.) two subgroups (strata)</td>
<td>Random selection of a proportional number of stratum members from each stratum</td>
<td>Every member of each stratum (or not) in the population has an equal chance of being selected into the sample (proportional sampling)</td>
</tr>
</tbody>
</table>
Procedures for Drawing Probability Samples

1. Select a suitable frame
2. Select the stratification variable(s) and the number of strata, $H$
3. Divide the entire population into $H$ strata. Based on the classification variable, each element of the population is assigned to one of the $H$ strata
4. In each stratum, number the elements from 1 to $N_h$ (the pop. size of stratum $h$)
5. Determine the sample size of each stratum, $n_h$, based on proportionate or disproportionate stratified sampling, where

$$
\sum_{h=1}^{H} n_h = n
$$

6. In each stratum, select a simple random sample of size $n_h$

Stratified Sample: Example

- A magazine aims to conduct a satisfaction study of its clients. For this purpose, a stratified sample is drawn, based on two stratification variables.
  - Client antiquity (<1 year; between 1 and 3 years; > 3 years;
  - Residence region (North, Center, South and Islands)
- The firm will survey 800 clients and decides to use a proportionate stratified sampling.
- $N = 6000$ subscribers

? Calculate the size of the sample in each stratum
Stratified Sample: Example

<table>
<thead>
<tr>
<th>Region</th>
<th>Population &lt; 1 year</th>
<th>Sample &lt; 1 year</th>
<th>Population 1-3 years</th>
<th>Sample 1-3 years</th>
<th>Population ≥ 3 years</th>
<th>Sample ≥ 3 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>500</td>
<td>400</td>
<td></td>
<td></td>
<td>350</td>
<td></td>
</tr>
<tr>
<td>Centre</td>
<td>1100</td>
<td>800</td>
<td></td>
<td></td>
<td>550</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>600</td>
<td>500</td>
<td></td>
<td></td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Islands</td>
<td>500</td>
<td>350</td>
<td></td>
<td></td>
<td>150</td>
<td></td>
</tr>
</tbody>
</table>

Cluster Sampling

- The target population is first divided into mutually exclusive and collectively exhaustive subpopulations, or clusters.
- Then a random sample of clusters is selected, based on a probability sampling technique such as SRS.
- For each selected cluster, either all the elements are included in the sample (one-stage) or a sample of elements is drawn probabilistically (two-stage).
- Elements within a cluster should be as heterogeneous as possible, but clusters themselves should be as homogeneous as possible. Ideally, each cluster should be a small-scale representation of the population.
- In probability proportionate to size sampling, the clusters are sampled with probability proportional to size. In the second stage, the probability of selecting a sampling unit in a selected cluster varies inversely with the size of the cluster.
### Cluster Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population is groups (Clusters)</td>
<td>Random selection of 2 clusters with random selection of members of these clusters (2-stage)</td>
<td>Every cluster (A, B, C, D, or E) in the population has an equal chance of being selected into the sample, and every cluster member has an equal chance of being selected from that cluster.</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A two-step area cluster sample (sampling several clusters) is preferable to a one-step (selecting only one cluster) sample unless the clusters are homogeneous.
Types of Cluster Sampling

Cluster Sampling

- One-Stage Sampling
- Two-Stage Sampling
- Multistage Sampling

Simple Cluster Sampling
- Probability Proportionate to Size Sampling
Cluster Sampling

1. Assign a number from 1 to N to each element in the population
2. Divide the population into C clusters of which c will be included in the sample
3. Calculate the sampling interval i, i=N/c (round to nearest integer)
4. Select a random number r between 1 and i, as explained in simple random sampling
5. Identify elements with the following numbers:
   r, r+i, r+2i, ..., r+(c-1)i
6. Select the clusters that contain the identified elements
7. Select sampling units within each selected cluster based on SRS or systematic sampling
8. Remove clusters exceeding sampling interval i. Calculate new population size N*, number of clusters to be selected C* = C-1, and new sampling interval i*.

Repeat the process until each of the remaining clusters has a population less than the sampling interval. If b clusters have been selected with certainty, select the remaining c-b clusters according to steps 1 through 7. The fraction of units to be sampled with certainty is the overall sampling fraction = n/N. Thus, for clusters selected with certainty, we would select n_s=(n/N)(N_1+N_2+...+N_b) units. The units selected from clusters selected under PPS sampling will therefore be n*=n- n_s.
Convenience Sampling

Convenience sampling attempts to obtain a sample of convenient elements. Often, respondents are selected because they happen to be in the right place at the right time.

- use of students, and members of social organizations
- mall intercept interviews without qualifying the respondents
- department stores using charge account lists
- “people on the street” interviews
Judgmental Sampling

Judgmental sampling is a form of convenience sampling in which the population elements are selected based on the judgment of the researcher.

- test markets
- purchase engineers selected in industrial marketing research
- bellwether precincts selected in voting behavior research
- expert witnesses used in court

Quota Sampling

Quota sampling may be viewed as two-stage restricted judgmental sampling. The first stage consists of developing control categories, or quotas, of population elements. In the second stage, sample elements are selected based on convenience or judgment.

<table>
<thead>
<tr>
<th>Control Characteristic</th>
<th>Population composition</th>
<th>Sample composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Percentage</td>
<td>Percentage</td>
</tr>
<tr>
<td>Male</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Female</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
### Quota sample example

<table>
<thead>
<tr>
<th>Propensity to donate</th>
<th>Have a flag</th>
<th>No flag</th>
<th>Have a flag</th>
<th>No flag</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td><strong>48%</strong></td>
<td><strong>52%</strong></td>
<td><strong>48%</strong></td>
<td><strong>52%</strong></td>
<td></td>
</tr>
<tr>
<td>18 to 30</td>
<td>30</td>
<td>30</td>
<td>33</td>
<td>32</td>
<td>125</td>
</tr>
<tr>
<td>31 to 45</td>
<td>48</td>
<td>48</td>
<td>52</td>
<td>52</td>
<td>200</td>
</tr>
<tr>
<td>46 to 60</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>75</td>
</tr>
<tr>
<td>Over 60</td>
<td>24</td>
<td>24</td>
<td>26</td>
<td>26</td>
<td>100</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>120</strong></td>
<td><strong>120</strong></td>
<td><strong>130</strong></td>
<td><strong>130</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>240</strong></td>
<td><strong>260</strong></td>
<td></td>
<td></td>
<td><strong>500</strong></td>
</tr>
</tbody>
</table>
### Snowball Sampling

In **snowball sampling**, an initial group of respondents is selected, usually at random.

- After being interviewed, these respondents are asked to identify others who belong to the target population of interest.
- Subsequent respondents are selected based on the referrals.

### Strengths and Weaknesses of Basic Sampling Techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonprobability Sampling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convenience sampling</td>
<td>Least expensive, least time-consuming, most</td>
<td>Selection bias, sample not representative, not</td>
</tr>
<tr>
<td></td>
<td>convenient</td>
<td>recommended for descriptive or causal research.</td>
</tr>
<tr>
<td>Judgmental sampling</td>
<td>Low cost, convenient, not time-consuming</td>
<td>Does not allow generalization, subjective</td>
</tr>
<tr>
<td>Quota sampling</td>
<td>Sample can be controlled for certain</td>
<td>Selection bias, no assurance of representativeness</td>
</tr>
<tr>
<td></td>
<td>characteristics</td>
<td></td>
</tr>
<tr>
<td>Snowball sampling</td>
<td>Can estimate rare characteristics</td>
<td>Time-consuming</td>
</tr>
<tr>
<td><strong>Probability sampling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple random sampling (SRS)</td>
<td>Easily understood, results projectable</td>
<td>Difficult to construct sampling frame, expensive,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lower precision, no assurance of representativeness.</td>
</tr>
<tr>
<td>Systematic sampling</td>
<td>Can increase representativeness, easier to</td>
<td>Can decrease representativeness</td>
</tr>
<tr>
<td></td>
<td>implement than SRS, sampling frame not</td>
<td></td>
</tr>
<tr>
<td></td>
<td>necessary</td>
<td></td>
</tr>
<tr>
<td>Stratified sampling</td>
<td>Include all important subpopulations, precision</td>
<td>Difficult to select relevant stratification</td>
</tr>
<tr>
<td></td>
<td></td>
<td>variables, not feasible to stratify on many</td>
</tr>
<tr>
<td></td>
<td></td>
<td>variables, expensive</td>
</tr>
<tr>
<td>Cluster sampling</td>
<td>Easy to implement, cost effective</td>
<td>Imprecise, difficult to compute and interpret</td>
</tr>
<tr>
<td></td>
<td></td>
<td>results</td>
</tr>
</tbody>
</table>
## Choosing Nonprobability vs. Probability Sampling

<table>
<thead>
<tr>
<th>Factors</th>
<th>Conditions Favoring the Use of Nonprobability sampling</th>
<th>Probability sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of research</td>
<td>Exploratory</td>
<td>Conclusive</td>
</tr>
<tr>
<td>Relative magnitude of sampling and nonsampling errors</td>
<td>Nonsampling errors are larger</td>
<td>Sampling errors are larger</td>
</tr>
<tr>
<td>Variability in the population</td>
<td>Homogeneous (low)</td>
<td>Heterogeneous (high)</td>
</tr>
<tr>
<td>Statistical considerations</td>
<td>Unfavorable</td>
<td>Favorable</td>
</tr>
<tr>
<td>Operational considerations</td>
<td>Favorable</td>
<td>Unfavorable</td>
</tr>
</tbody>
</table>