Research Methods in Linguistics
1302740

Lecture (3)
Population Samples
LEARNING OBJECTIVES

- Learn the reasons for sampling
- Develop an understanding about different sampling methods
- Distinguish between probability & non probability sampling
- Discuss the relative advantages & disadvantages of each sampling methods
Population

- the group you are ultimately interested in knowing more about their linguistic behaviour
- On the basis of sample study we can predict and generalize the behavior of mass phenomena.
- “entire aggregation of cases that meets a designated set of criteria".
A sample is “a smaller (but hopefully representative) collection of units from a population used to determine truths about that population” (Field, 2005)
Sample vs. Census

- **Census**: an accounting of the complete population
- A census study occurs if the entire population is *very small* or it is reasonable to include the entire population (for other reasons).

- It is called a census sample because data is gathered on every member of the population.
**Why sample?**

- The population of interest is usually too large to attempt to survey all of its members.
- Resources (time, money) and workload

**So...**

- A carefully chosen sample can be used to represent the population.
- The sample reflects the characteristics of the population from which it is drawn.
- Gives results with known accuracy that can be calculated mathematically
If all members of a population were identical, the population is considered to be *homogenous*.

That is, the characteristics of any one individual in the population would be the same as the characteristics of any other individual (little or no variation among individuals).
When individual members of a population are different from each other, the population is considered to be *heterogeneous* (having significant variation among individuals).
Using data to say something (*make an inference*) with confidence, about a whole (population) based on the study of only a few (sample).
Who do you want to generalize to?

What population can you get access to?

How can you get access to them?

Who is in your study?

The Theoretical Population

The Study Population

The Sampling Frame

The Sample
WHAT IS SAMPLING?

- Sampling is the process of selecting observations (a sample) to provide an adequate description and robust inferences of the population.
  - The sample is *representative* of the population.

What exactly IS a “sample”? 
TYPES OF SAMPLING?

- There are 2 types of sampling:
  - Non-Probability sampling
  - Probability sampling
**PROBABILITY VERSUS NONPROBABILITY**

- **Probability Samples**: each member of the population has a known non-zero probability of being selected
  - Methods include *random sampling*, *systematic sampling*, and *stratified sampling*.

- **Nonprobability Samples**: members are selected from the population in some nonrandom manner
  - Methods include *convenience sampling*, *judgment sampling*, *quota sampling*, and *snowball sampling*.
Simple Random Sampling

Non-Probability Sampling
Probability Samples: each member of the population has a known non-zero probability of being selected

Methods include
1. (simple) random sampling
2. systematic sampling
3. stratified sampling

“...it would be nice to just observe ALL of the people...”
Random sampling is the purest form of probability sampling.

- Each member of the population has an equal and known chance of being selected.

- When there are very large populations, it is often ‘difficult’ to identify every member of the population, so the pool of available subjects becomes biased.

  You can use software to generate random numbers or to draw directly from the columns of random numbers.
1. (SIMPLE) RANDOM SAMPLING

Simple Random Sampling
1. (SIMPLE) RANDOM SAMPLING

Lottery method

Random number tables

Part of a Table of Random Numbers

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>61424</td>
<td>20419</td>
<td>86546</td>
<td>00517</td>
<td></td>
</tr>
<tr>
<td>90222</td>
<td>27993</td>
<td>04952</td>
<td>66762</td>
<td></td>
</tr>
<tr>
<td>50349</td>
<td>71146</td>
<td>97668</td>
<td>86523</td>
<td></td>
</tr>
<tr>
<td>85676</td>
<td>10005</td>
<td>08216</td>
<td>25906</td>
<td></td>
</tr>
<tr>
<td>02429</td>
<td>19761</td>
<td>15370</td>
<td>43882</td>
<td></td>
</tr>
<tr>
<td>90519</td>
<td>61988</td>
<td>40164</td>
<td>15815</td>
<td></td>
</tr>
<tr>
<td>20631</td>
<td>88967</td>
<td>19660</td>
<td>89624</td>
<td></td>
</tr>
<tr>
<td>89990</td>
<td>78733</td>
<td>16447</td>
<td>27932</td>
<td></td>
</tr>
</tbody>
</table>
1. **(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>
1. (Simple) Random Sampling

Advantages...
- ...easy to conduct
- ...strategy requires minimum knowledge of the population to be sampled

Disadvantages...
- ...need names of all population members
- ...may over-represent or under-estimate sample members
- ...there is difficulty in reaching all selected in the sample
2. SYSTEMATIC SAMPLING

- **Systematic sampling** is often used instead of random sampling. It is also called an *Nth* name selection technique.

- After the required sample size has been calculated, every *Nth* record is selected from a list of population members.

- As long as the list does not contain any hidden order, this sampling method is as good as the random sampling method.

- Its only advantage over the random sampling technique is *simplicity* (and possibly cost effectiveness).
2. SYSTEMATIC SAMPLING
**2. SYSTEMATIC SAMPLING**

Procedure

- Number units in population from 1 to N.
- Decide on the n that you want or need.
- $N/n = k$ the interval size.
- Randomly select a number from 1 to k.
- Take every $k^{th}$ unit.
## 2. Systematic Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directory of the population (sample frame)</td>
<td>Selection via skip interval with a random starting point</td>
<td>Every member of the sample frame has an equal chance of being selected into the sample</td>
</tr>
</tbody>
</table>
2. SYSTEMATIC SAMPLING

advantages…
• …sample selection is simple
• may be more precise than simple random sample.

disadvantages…
• …all members of the population do not have an equal chance of being selected
• …the Kth person may be related to a periodical order in the population list, producing unrepresentativeness in the sample
3. STRATIFIED SAMPLING

- **Stratified sampling** is commonly used probability method that is superior to random sampling because it reduces sampling error.

- Sometimes called "proportional" or "quota" random sampling.

- A *stratum* is a subset of the population that share at least one common characteristic; such as *males* and *females*.
  
  ◦ Identify relevant strata and their actual representation in the population.

  ◦ Random sampling is then used to select a *sufficient* number of subjects from each stratum.

  ◦ Stratified sampling is often used when one or more of the strata in the population have a low incidence relative to the other strata.
3. STRATIFIED SAMPLING

Stratified Random Sampling
3. STRATIFIED 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 (I or II) in the population has an equal chance of being selected into the sample (proportional sampling)</td>
</tr>
</tbody>
</table>
3. STRATIFIED SAMPLING

advantages...
- ...more precise sample
- ...can be used for both proportions and stratification sampling
- ...sample represents the desired strata

disadvantages
- ...need names of all population members
- ...there is difficulty in reaching all selected in the sample
Nonprobability Samples: “Members are selected from the population in some nonrandom manner” (Barreiro, 2009)

Methods include
1. convenience sampling
2. judgment sampling
3. quota sampling
4. snowball sampling
1. **Convenience Sampling**

- **Convenience sampling** is used in exploratory research where the researcher is interested in getting an inexpensive approximation.
- The sample is selected because they are convenient (to the researcher).
- It is a nonprobability method.
  - Often used during *preliminary research* (pilot studies) efforts to get an estimate without incurring the cost or time required to select a random sample.
1. CONVENIENCE SAMPLING

Convenience Sampling
1. **CONVENIENCE SAMPLING**

- Exploratory research
- Inexpensive approximation
  - Ex: preliminary research efforts to attain the number of \( L_1, L_2, \ldots, L_n \) speakers at university
- Saves time and money
  - selected because they are willing and available
1. **CONVENIENCE SAMPLING**

- Convenience samples: samples drawn at the convenience of the *interviewer*. People tend to make the selection at familiar locations and to choose respondents who are like themselves.

- Error occurs
  1. in the form of members of the population who are infrequent or nonusers of that location
  1. who are not typical in the population
1. CONVENIENCE SAMPLING

advantages…
• useful in pilot studies.

disadvantages…
• …difficulty in determining how much of the effect (dependent variable) results from the cause (independent variable)
2. JUDGMENT SAMPLING

- **Judgment (Purposive) sampling** is a common nonprobability method.

- The sample is selected based upon judgment.
  - an extension of convenience sampling

- Researcher's knowledge is used to *hand pick* the cases to be included in the sample

- When using this method, the researcher must be confident that the chosen sample is truly *representative* of the entire population.
2. JUDGMENT SAMPLING
2. JUDGMENT SAMPLING

- Subjective judgment

- “The person who is selecting the sample is who tries to make the sample representative, depending on his opinion or purpose, thus being the representation subject” (Barreiro, 2009)

- Requires researcher *confidence* that the sample truly represents an entire population
2. JUDGMENT SAMPLING

advantages...
• Small no. of sampling units
• Study unknown traits/case sampling

disadvantages...
• ...potential for inaccuracy in the researcher’s criteria and resulting sample selections
• Personal prejudice & bias
• No objective way of evaluating reliability of results
3. QUOTA SAMPLING

- Quota sampling is the nonprobability equivalent of stratified sampling.
  - First identify the strata and their proportions as they are represented in the population.
  - Then convenience or judgment sampling is used to select the required number of subjects from each stratum.
3. QUOTA SAMPLING

Quota Sampling

Quota:
Male, Above 50
3. QUOTA SAMPLING

• ...people who are less accessible (more difficult to contact, more reluctant to participate) are under-represented

disadvantages...
Snowball sampling is a special nonprobability method used when the desired sample characteristic is rare. It may be extremely difficult or cost prohibitive to locate respondents in these situations. This technique relies on referrals from initial subjects to generate additional subjects (friend-of-friend). It lowers search costs; however, it introduces bias because the technique itself reduces the likelihood that the sample will represent a good cross section from the population.
4. SNOWBALL SAMPLING

SNOWBALL SAMPLING

Diagram showing a network of people where each person can recruit others, expanding the sample progressively.
ADVANTAGES OF NON-PROBABILITY SAMPLING

advantages...
• access to difficult to reach populations (other methods may not yield any results).
• Convenient
• Economical

disadvantages...
• not representative of the population and will result in a biased sample as it is self-selecting.
The more heterogeneous a population is, the larger the sample needs to be.

Depends on topic – frequently it occurs?

For probability sampling, the larger the sample size, the better.

With nonprobability samples, not generalizable regardless – still consider stability of results.
Sample size depends on:

◦ How much sampling error can be tolerated—levels of precision

◦ Size of the population—sample size matters with small populations

◦ Variation within the population with respect to the characteristic of interest—what you are investigating

◦ Smallest subgroup within the sample for which estimates are needed

◦ Sample needs to be big enough to properly estimate the smallest subgroup
Rule of thumb: “the larger the sample size, the more closely your sample data will match that from the population” (Birchall, 2009)

Key factors to consider:
- How accurate you wish to be
- How confident you are in the results
- What budget you have available