PSCI2300 The Study of Politics
Sampling 1

Tetsuya Matsubayashi

University of North Texas

March 8, 2011
Sampling

- So far we are through research designs, hypotheses, variables, and the conduct of literature review

- Our plan
  - Sampling
  - Survey data
  - Univariate data analysis
  - Bivariate data analysis
  - Multivariate data analysis
Assume you test a hypothesis using a cross-sectional design.

We can collect data of all of 30 OECD nations, 50 states, 3,141 counties, etc.

But how about the nearly 300 million people living in the United States?

How can we make claims about what people in general think about politics?

We usually rely on a sample of people for collecting data.
Two Key Questions

- What are samples and how are they collected?
- How much confidence can we place in statements about a population given observations derived from a very few of its members?
Population or Sample

- **Population** – any well-defined set of units of analysis
- **Sample** – any subset of units collected in some manner from the population

The sample size and how its members are chosen determine the quality of inferences about the whole population.

Example:
- Population: All students at UNT
- Sample: Randomly selected students

Population or sample? – determined by practical reasons
If the sample is selected properly, the information it produces may be used to make inferences about the whole population.

Sampling arises whenever a researcher takes measurements on a subset of a population, however defined, covered by the hypothesis being investigated.

The purpose of sampling is to collect data that provide an accurate estimate of a population parameter.

Population parameter – Certain characteristics of population (e.g., averages).

Sample statistic – the estimated value of a population characteristic that is calculated from sample data.
Distribution of partisanship in the United States in March, 2011.

We want to know unknown proportions $D$ (Democrats), $R$ (Republicans), and $I$ (Independents) in the population. (Note that $D + R + I = 1$.)

A statistic based on a properly collected sample provides an estimated value of $\hat{D}$, $\hat{R}$, $\hat{I}$.

Our goal is to make a good unbiased guess about the true value (= population parameter) of $D$, $R$, $I$. 
Example of Incomplete Sampling Frame

- *Literary Digest Poll* of 1936 on Alf Landon VS FDR.
- The poll predicted Landon’s victory.
- The poll relied on a huge sample (including about 100,000 respondents) drawn from telephone directories and automobile registration lists.
- What went wrong?
- Today, what would be the problems of the use of telephone directories?
- *Sample bias* will occur if a sampling frame is incomplete or in appropriate.
Types of Samples

**Probability Sample:** A sample for which each element in the total population has a known probability of being included in the sample.

- Simple random samples
- Systematic samples
- Stratified samples – proportionate and disproportionate
- Cluster samples
Nonprobability Sample: A sample for which each element in the total population has an unknown probability of being selected. This sampling framework rules out the use of statistical theory to make inferences.

- Purposive samples
- Convenience samples
- Quota samples
- Snowball samples
Each element has an *equal* chance of being selected.

A list of all the elements in the population must be available.

1. Assign a number to each element in the sample frame, and then use a random numbers table (list of random numbers).
2. By-the-lot method.
Systematic Samples

- Elements are selected from a list at predetermined intervals systematically.
- A list of the target population is required.
- Example: Suppose we want to draw a sample of 100 names from a list of all political science students.
  1. Calculate the *sampling interval* = (# of elements in the list)/(# of sample size).
  2. Systematically go through the list and select every ?th students.
  3. To determine where on the list to begin, we need to make a random start (select a number at random using random numbers table).
- Systematic sampling may result in a biased sample!
  - Elements on the list have been ranked according to a characteristic.
Stratified Samples

- Elements sharing one or more characteristics are grouped, and elements are selected from each group in proportion to the group’s representation in the total population.

- Proportionate Sample – Each stratum is represented in proportion to its size in the population.

- Example: Average GPA of UNT students

<table>
<thead>
<tr>
<th></th>
<th>Pol Sci</th>
<th>Journalism</th>
<th>Business</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>500</td>
<td>100</td>
<td>200</td>
<td>800</td>
</tr>
<tr>
<td>Proportion (weight)</td>
<td>0.625</td>
<td>0.125</td>
<td>0.250</td>
<td>1</td>
</tr>
<tr>
<td>Proportionate sample</td>
<td>62</td>
<td>12</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>Disproportionate sample</td>
<td>34</td>
<td>33</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td>Sample mean GPA</td>
<td>3.3</td>
<td>3.0</td>
<td>2.7</td>
<td>?</td>
</tr>
</tbody>
</table>
Disproportionate Sample – Use a stratified sample in which elements sharing a characteristic are underrepresented or overrepresented in our sample.

Use of *weighting factor* – a mathematical factor used to make a disproportionate sample representative.

It is essential to select characteristics on which to stratify a list...
Cluster Samples

- The sampling frame initially consists of clusters of elements
- Cluster sampling is used to address the problem of having NO list of the elements in the target population
Note that even though we did not know the # of households ahead of time, each household has an *equal* chance of being selected.

Question – What’s the probability that each household is being selected?

**National opinion poll**

1. List and sample States
2. List and sample Counties within each states
3. List and sample municipalities within counties
4. List and sample census tracts within municipalities
5. List and sample blocks within census tracts
6. List and sample households within each block