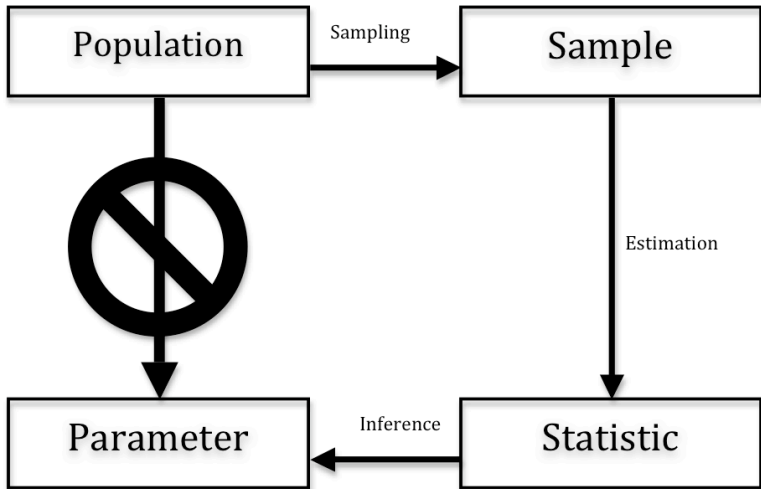


PLSC 308: Introduction to Political Research

Christopher Zorn

February 23, 2016

Inference: The Idea



Sampling: Terms

- **Units of Analysis:** The things we want to learn about (a/k/a “elements”)
- **Population:** *All* of the units of analysis (a/k/a the **sampling frame**)
- **Stratum** (plural: strata): A subgroup of the population sharing a common trait or traits
- **Sample:** The subset of the population selected for examination
- **Primary sampling units:** The unit that is sampled (often, but not always, the unit of analysis)
- **Sample size:** The *number* of units of analysis in the sample (always denoted N)

Two Problems With Samples

Bias

- *Systematic* differences between the sample and the population.
- Usually due to the sampling (or research) design.

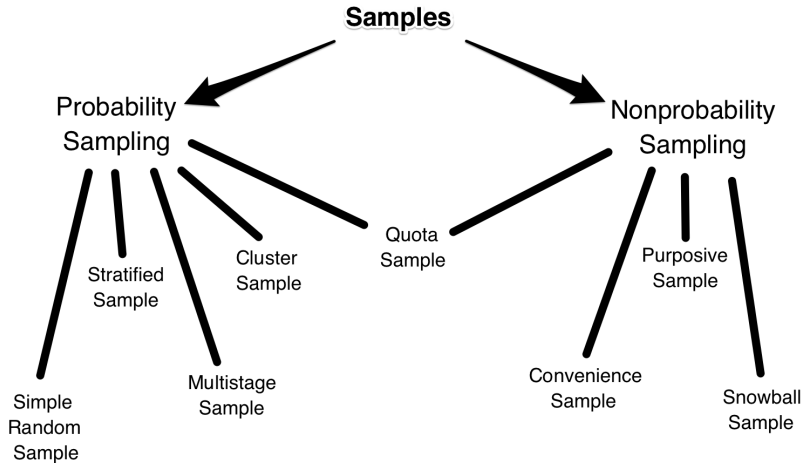
Sampling Error

- Differences between the sample and the population that are *nonsystematic*.
- Due to the randomness inherent to the sampling design.

In general:

**Bias is a much bigger problem
than sampling error.**

Sample Types



Simple Random Sampling

Any sampling design where the probability of any given unit being selected into the sample on any given draw is $1/N$.

or

Any sampling design where the probability of any given unit being selected into the sample is the same as any other unit in the population.

Simple Random Sampling

The Good:

- Mathematically easy to understand and implement
- Leads to the simplest / most straightforward methods of inference

The Bad:

- Difficult to define and draw... Must
 - *Know* every unit in the population, and
 - Be able to *include* all selected units in the sample drawn.
- Can yield poor results for small subpopulations / strata.

Stratified Sampling

Steps:

1. Divide the sample into strata based on predefined characteristics.
2. Conduct simple random sampling within each stratum.

For two groups (strata) A and B :

- Proportional stratified sample = the proportion of A in the sample is the same as in the population.
- Oversampling = the proportion of A in the sample is *greater than* in the population.
- Undersampling = the proportion of A in the sample is *less than* in the population.

Cluster Sampling

Steps:

1. Divide the sample into clusters based on predefined characteristics.
2. Draw a simple random sample of the clusters.
3. Include all units in each selected cluster in the final sample.

Cluster sampling:

- Changes the PSU from the unit of analysis to the cluster...
- Makes $\Pr(\text{sample unit } i)$ nonconstant / undefined
- *Most major media polls are done via cluster sampling*

Multistage Sampling

Steps:

1. Select a “cluster,” identify subclusters of units within the cluster, etc. until we get to the “lowest” level cluster.
2. Select – randomly or in a stratified way – some number of top-level clusters.
3. Within each selected cluster, select – again, randomly or stratifying – some number of subclusters.
4. Within subclusters, select sub-subclusters, etc.
5. At the “lowest” subcluster level, select some number of units from each sub-cluster.

Multistage Sampling

Example (Agresti and Finlay): sample survey respondents by first selecting blocks, then selecting houses within blocks, then selecting residents within each (selected) house.

- Blocks are *clusters*, houses are *subclusters*, and the individuals are the units finally sampled.
- Allows for probability samples without knowing identities of every unit sampled, via sampling rules (e.g., “select one person from among those in each house with equal probability.”)
- *Most large, national surveys are conducted using multistage sampling.*

Nonprobability Samples

A sample where probability that every unit is in the sample is not (or cannot be) known.

Flavors:

- **Convenience Sampling:** What the same suggests.
- **Purposive Sampling:** The researcher selects units on the basis of whether s/he believes they ought to be in the sample.
- **Snowball Sampling:** Selects a unit, and then sample other units with some relationship to that first unit.

Quota Sampling

Researcher samples units within strata up to some quota, and then stops.

- E.g., a survey researcher might question 100 men and 100 women.
- Used a great deal in pre-WWII studies.
- Combined with (say) convenience sampling → nonprobability sample.
- Combined with probability (e.g., stratified) sampling → better.

- Probability samples yield sampling error.
 - Smallest = (generally) simple random sampling
 - Stratified *can* be smaller
 - Multi-stage = complex...
- Nonprobability samples can lead to bias; also have (complex) sampling error.