Political Science 104

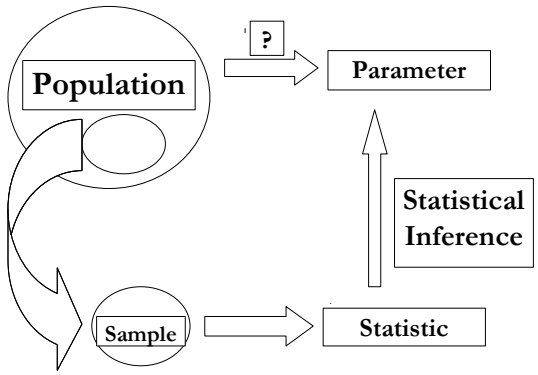
Lecture 12:
Sampling and Statistical Inference

Statistical Inference

Using statistical information we can observe to make inferences about some things we can't observe

Statistical Inference

- A *population* is the universe of things we are interested in studying. For causal inference usually very broad (i.e., all countries that exist now and will ever exist).
- We are specifically interested in some characteristics of the population. These characteristics are called *parameters*.
- We can almost never observe the complete population. Instead, we draw a *sample* from the population.
- We calculate *sample statistics* using the sample.
- We then make an *inference* about the population parameters using these sample statistics.



Sampling from a Population

- Obviously, statistical inference will only work if we draw a "good" sample from the population.
- *Sample selection*: is the sample representative of the population?
- *Sample size*: is the sample large enough to draw reliable inferences about the population?

Sample Selection

- Our goal is to draw a *random sample*.
- In a random sample, every element in the population has an equal chance to be in the sample.
- It is very hard to draw truly random samples.
  - Random digit dialing (RDD), but what about people with no phones?
  - Web surveys, but what about people with no internet access?
- Biased samples will lead to biased inferences about the population.

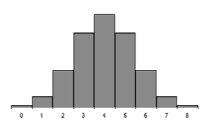
Bad Sample Selection

Sample Size

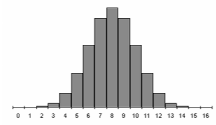
- Larger samples will lead to more accurate inferences than smaller samples.
- The size of the population *does not* change the required sample size for a given level of accuracy.
- There are diminishing returns to accuracy as sample size increases.
- Most samples trade off accuracy for reduced cost. Example: national public opinion polls usually interview 1000-1500 people.

Sampling Distributions as Sample Size Increases

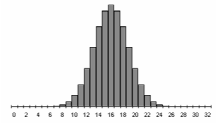
- 8 coin flips



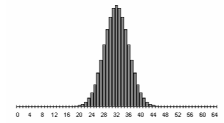
- 16 coin flips



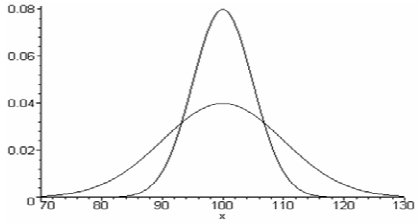
- 32 coin flips



- 64 coin flips



Changes in Sample Size

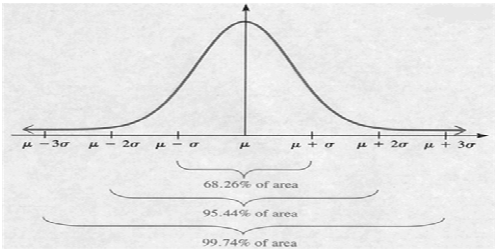


- Increasing sample size reduces the variance of the sampling distribution, making our estimates of the population parameter more accurate.

Sampling Distributions

- These distributions are known as *sampling distributions*. A sampling distribution is the distribution of a sample statistic under repeated sampling.
- **The Central Limit Theorem**: The sample statistics from random samples of a population will be normally distributed around the population parameter with variance σ^2/n.

The Normal Distribution



- About 68% of the time our sample statistic will be within 1 standard deviation of the true population parameter.
- About 95% of the time our sample statistic will be within 2 standard deviations of the true population parameter.

Sampling Distributions and Hypothesis Testing

- We now see that nearly all sampling distributions (the distribution of sample statistics we would estimate under repeated sampling) are normally distributed.
- How can we take advantage of this fact to test our hypotheses? We'll do that next lecture …