Lesson 07 - Generalization

Objectives: 1. Sampling from finite populations
2. Random Sample
3. Bias

Previously, observed sample from a random process.

Today, sample from a population.

Overall Objective: How do we select the sample so the characteristics of the sample are representative of the larger population.

Population (N) - Everyone. All observational units.

Sample (n) - some subset to observe → "Statistic"

Convenience Sample - not random. It's Lazy. (voluntary response, mail-in survey...)
- Biased sampling method, ↳ sampling statistic consistently ↑ or ↓ the population parameter of interest
- Doesn't represent the larger population

Simple Random Sample - Everyone equally likley to be chosen from population

Sampling Distribution

- Central limit theorem says that the distribution of sample proportions is approx normal if validity conditions met

- mean → π
- $SD(\hat{p}) \rightarrow \sqrt{\frac{\pi(1-\pi)}{n}}$

If validity conditions met
- ≥10 successes & failures (allows us to estimate p-value) (normal)
- $N > 20 * n$ for a finite population
  - allows us to estimate SD

Does larger n solve a biased sampling method? - no
We want n large though.

Objective: make inferences about population proportions based on sample proportion

Process:
1. random sample selected from a population
2. use simulation or theory based approach to get p-value
3. make inferences about the population