Sampling weights and estimation overview

Week 2 (2.1 and 2.4)

Stat 260, St. Clair

Sampling design

- Section 2.2. "small example" looked at 3 different sampling designs that resulted in different sampling probabilities for each possible *sample* of $n = 2$ units.

Definition: The probability that **a sampling unit** is included in the sample of n units is its **inclusion probability.**

$$\pi_i = P(\text{unit } i \text{ is included in the sample})$$

Sampling weights

Definition: The **sampling weight** of unit i is equal to the inverse of its inclusion probability:

$$w_i = \frac{1}{\pi_i}$$

- loosely: tells us the number of units in the population that are represented by sampling unit i

2.2 small example

Design 3

S	$P(S)$
$\{1,2\}$	$1/3$
$\{1,3\}$	$1/3$
$\{2,3\}$	$1/3$

unit 1 - D3

$\pi_1 = P(\text{lake 1 included})$

$= P(\{1,2\} \text{ OR } \{1,3\})$

$= \frac{1}{3} + \frac{1}{3} = 2/3$

$w_1 = \frac{3}{2} = 1.5$

Lake 1 represents 1.5 lakes in the population.

Unbiased estimation

When sampling **without replacement**, the following estimator of **population total** is always unbiased regardless of sampling design:

$$\boxed{\hat{t}_{HT} = \sum_{\text{sampled units}} w_i y_i} = \underline{\sum_{\text{sampled units}} \frac{y_i}{\pi_i}}$$

- This estimator is known as the **Horvitz-Thompson** estimator

Overview of estimation

- The Horvitz-Thompson estimator is the basis for estimation for *many* sampling designs.
- up next:
  - Design: Simple Random Sample estimation story
  - SRS example
  - Intro to the `survey` package