Estimation in Domains

Week 5 (4.3)

Stat 260, St. Clair

Domains

Domain: subpopulations of interest $\mathcal{U}_d$

- We want to estimate a domain parameter: $t_d, \bar{y}_{\mathcal{U}_d}, p_d$

Design

We take a SRS of size n from a population $\mathcal{U}$

- **Problem:** n_d, the number of respondents in the domain, varies from sample to sample

Domain estimation: a ratio estimator

- Domain indicator:

$$x_i = \begin{cases} 1 & \text{if unit } i \text{ is in the domain} \\ 0 & \text{if unit } i \text{ is not in the domain} \end{cases}$$

- Domain sample size:

$$n_d = \sum_{i=1}^{n} x_i$$

Domain estimation: a ratio estimator

- Domain responses:

$$u_i = x_i y_i = \begin{cases} y_i & \text{if unit } i \text{ is in the domain} \\ 0 & \text{if unit } i \text{ is not in the domain} \end{cases}$$

- Domain sample mean:

$$\bar{y}_d = \frac{\sum_{i \in \text{domain}} y_i}{n_d} = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} x_i}$$

Domain estimation for a mean

- **Domain Parameter** mean $\bar{y}_{\mathcal{U}_d}$
- **Estimator** with a SRS of units: a ratio estimator

$$\bar{y}_d = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} x_i}$$

- **SE**: when n is "large" (or bias about 0)

$$SE(\bar{y}_d) \approx \sqrt{\left(1 - \frac{n}{N}\right) \frac{s_e^2}{n\bar{x}^2}} = \sqrt{\left(1 - \frac{n}{N}\right) \left(\frac{n}{n-1}\right) \left(\frac{n_d - 1}{n_d}\right) \frac{s_d^2}{n_d}}$$

where s_d is the sample SD of the measurements y_i in the sampled domain.

Domain estimation for a proportion

- **Domain Parameter** proportion p_d
- Response y_i is an indicator of "success"
- **Estimator** with a SRS of units: a ratio estimator

$$\hat{p}_d = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i}$$

- **SE**: same as the mean except $s_d^2 = \dfrac{n_d}{n_d - 1}\hat{p}_d(1 - \hat{p}_d)$

Example 1

An economist wants to estimate the average weekly amount spent on food by households containing children in a small town.

A complete list of all 2500 households in the county is available, but identifying those households with children is impossible.

So the economist selects a SRS of 500 households and observes 420 that contain children.

Of the 420 households with children, he records an average of $120.35 spent on food during a week and a sample SD of $42.20.

Example 1(a):

Estimate the average weekly amount of money spent on food by all households with children in the county and compute the standard error of your estimate.

Domain estimation for a total

- **Domain Parameter** total t_d

Two scenarios:

1. If N_d, population size of the domain, is **known**

$$\hat{t}_d = N_d \bar{y}_d, \quad SE(\hat{t}_d) = N_d SE(\bar{y}_d)$$

Domain estimation for a total

- **Domain Parameter** total t_d

Two scenarios:

2. If N_d, population size of the domain, is **unknown**

- **Estimator**: SRS total estimate with u_i as the response

$$\hat{t}_d = N\bar{u}$$

- **SE**:

$$SE(\hat{t}_d) = NSE(\bar{u}) = N\sqrt{\left(1 - \frac{n}{N}\right)\frac{s_u^2}{n}}$$

where s_u is the sample SD of the measurements u_i in the sample and

$$s_u^2 = \frac{1}{n-1}\left[(n_d - 1)s_d^2 + n_d \bar{y}_d^2 \left(1 - \frac{n_d}{n}\right)\right]$$

Example 1(b):

Estimate the total weekly amount of money spent on food by households with children and compute the standard error of your estimate.

Bias of domain mean estimator

Small when n is large or n_d/n is close to 1

Proof

$$SE(\bar{y}_d) \approx \sqrt{\left(1-\frac{n}{N}\right)\frac{s_e^2}{n\bar{x}^2}} = \sqrt{\left(1-\frac{n}{N}\right)\left(\frac{n}{n-1}\right)\left(\frac{n_d-1}{n_d}\right)\frac{s_d^2}{n_d}}$$

Proof

$$s_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(u_i-\bar{u})^2 = \frac{1}{n-1}\left[(n_d-1)s_d^2 + n_d\bar{y}_d^2\left(1-\frac{n_d}{n}\right)\right]$$