Ratio estimation - the details

Week 4 (4.1)

Stat 260, St. Clair

Ratio estimation: for a ratio parameter

- For each sampling unit, we measure two variables: x_i and y_i
- **Ratio Parameter** Suppose our parameter of interest looks like

$$B = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i} = \frac{t_y}{t_x} = \frac{\bar{y}_{\mathcal{U}}}{\bar{x}_{\mathcal{U}}}$$

- **Estimator** with a SRS of units:

$$\hat{B} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{\bar{y}}{\bar{x}}$$

- SE and bias: now!

Ratio estimation: for a ratio parameter

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

$$SE(\hat{B}) \approx \sqrt{\left(1 - \frac{n}{N}\right) \frac{s_e^2}{n\bar{x}^2}}$$

where s_e is the sample SD of the "errors" $e_i = y_i - \hat{B}x_i$

- Computational "by hand" short cut:

$$s_e^2 = s_y^2 + \hat{B}^2 s_x^2 - 2\hat{B}\hat{R}s_y s_x$$

  - s_x, s_y: sample SD's of x/y measurements
  - $\hat{R}$: correlation coefficient of the sampled x/y measurements

Example 1: revisit petition

Signatures to a petition were collected on 676 sheets of paper.

- You take a SRS of 50 sheets and find 1515 signatures, 771 of which are from registered voters.
- The estimated proportion signatures belonging to registered voters is

$$\hat{p} = \frac{771}{1515} = 0.5089$$

Compute the SE for this estimate.

Example 1: revisit petition

- The standard deviation for the number of signatures per sheet is 2.929 and the standard deviation for the number of registered voter signatures per sheet is 3.909.
- The sample correlation between the total number of signatures and number of registered voter signatures per sheet is 0.4497.

Ratio estimation for a mean or total

- Suppose we **know a population mean or total** for an **auxiliary** variable x

$$t_x \text{ and/or } \bar{x}_U \text{ are known}$$

- **Population total estimate** for response y can be written

$$\hat{t}_{y,r} = \hat{B} t_x$$

- **Population mean estimate** for response y can be written

$$\hat{\bar{y}}_r = \hat{B} \bar{x}_{\mathcal{U}}$$

- SE and bias: Now!

Ratio estimation for a mean or total

- SE: when n is "large" (or bias of $\hat{B}$ about 0)

$$SE(\hat{t}_{y,r}) \approx \sqrt{\left(1 - \frac{n}{N}\right)\left(\frac{t_x}{\bar{x}}\right)^2 \frac{s_e^2}{n}}$$

$$SE(\hat{\bar{y}}_r) \approx \sqrt{\left(1 - \frac{n}{N}\right)\left(\frac{\bar{x}_{\mathcal{U}}}{\bar{x}}\right)^2 \frac{s_e^2}{n}}$$

where s_e is the sample SD of the "errors" $e_i = y_i - \hat{B} x_i$

- Computational "by hand" short cut:

$$s_e^2 = s_y^2 + \hat{B}^2 s_x^2 - 2\hat{B}\hat{R} s_y s_x$$

Example 4: a classic

what?	1920 (x)	1925 (y)
population total	$t_x = 22,919$	$t_y = ??$
sample total	$\sum_{i=1}^{49} x_i = 5054$	$\sum_{i=1}^{49} y_i = 6262$
sample variance	$s_x^2 = 10,900.4$	$s_y^2 = 15,158.8$
sample correlation	$\hat{R} = 0.9817$	

$$\hat{t}_{y,ratio} = 28,397 \quad SE(\hat{t}_{y,ratio}) = ??$$

Ratio estimation for a mean or total vs. SRS estimation

When is

$$V(\hat{\bar{y}}_r) \overset{?}{<} V(\bar{y}_{srs})$$

Theory: Bias approximation

$$B = \frac{t_y}{t_x} \quad \hat{B} = \frac{\hat{t}_y}{\hat{t}_x}$$

$$Bias(\hat{B}) = E(\hat{B}) - B = -\frac{Cov(\hat{B}, \bar{x})}{\bar{x}_{\mathcal{U}}}$$

Theory: Bias approximation

What is bias if $y_i \approx cx_i$ for all i where c is some fixed value?

Theory: Bias approximation

What is the maximum value of bias (approximately)?