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 lik

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

- **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:

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

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

intro: r

$-1 \le \hat{R} \le 1$

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!

$$\rightarrow SE(\hat{t}_{y,r}) = SE(\hat{B} t_x) = t_x \, SE(\hat{B})$$

Ratio estimation for a mean or total

$\bar{x}$ = SRS mean of x.

- 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:

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

Ratio estimation for a mean or total vs. SRS estimation

When is

ratio $\overset{?}{<}$ SRS

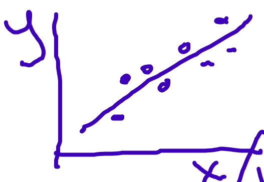
$$V(\hat{\bar{y}}_r) \overset{?}{<} V(\bar{y}_{srs})$$

$$\left(1-\frac{n}{N}\right)\left(\frac{\bar{x}_u}{\bar{x}}\right)\frac{s_e^2}{n} \overset{?}{<} \left(1-\frac{n}{N}\right)\frac{s_y^2}{n}$$

Assume: $\left(\frac{\bar{x}_u}{\bar{x}}\right) \approx 1$

$$\left(1-\frac{n}{N}\right)\frac{s_e^2}{n} \overset{?}{<} \left(1-\frac{n}{N}\right)\frac{s_y^2}{n}$$

$$\boxed{s_e^2 \overset{?}{<} s_y^2}$$



linearly

X + Y correlated

condition:
$\hat{R} \geq \frac{1}{2}$ then
SE ratio < SE SRS

$y_i \approx Bx_i$ $\quad \frac{y_i}{x_i} \approx B \Rightarrow e_i = y_i - \hat{B}x_i \approx 0 \Rightarrow s_e^2 \approx 0$

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.
  x
- The estimated proportion signatures belongin to registered voters is

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

Compute the SE for this estimate.

sample SD (per sheet) of y / x

y: register voter

x: # sig.

corr. between y + x in sample

Example 1: revisit petition

S_x

- The standard deviation for the number of signatures per sheet is 2.929 and the standard deviation for the number of ~~female~~ reg. voter signatures per sheet is 3.909. $\rightarrow S_y$
- The sample correlation between the total number of signatures and number of ~~female~~ reg. voter signatures per sheet is 0.4497. $\rightarrow \hat{R} = .4497$

$$s_e^2 = 3.909^2 + (.5089)^2 \, 2.929^2 - 2(.5089)(.4497)(3.909)(2.929)$$

$$= 11.51$$

$$SE(\hat{B}) = \sqrt{\left(1 - \frac{50}{676}\right)\frac{11.51}{50(1515/50)^2}} \approx .015$$

Example 4: a classic

$n = 49$ $N = 196$

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	$\hat{R} = 0.9817$	

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

$$s_e^2 = 621.85$$

$$SE(\hat{t}_{y,r}) = \sqrt{\left(1 - \frac{49}{196}\right)\left(\frac{22,919}{5054/49}\right)^2 \frac{621.85}{49}}$$

$$\approx 685.5$$

Theory: Bias approximation

Cov = covariance

= $\hat{B}$ & $\bar{x}$ co-vary together

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

SRS mean of x

$$Bias(\hat{B}) = \underbrace{E(\hat{B})}_{\text{SRS}} - B = -\frac{Cov(\hat{B}, \bar{x})}{\bar{x}_U}$$

$\bar{x}_U \longrightarrow$ pop. mean x

$\boxed{\text{Proof}}$ A.2

$$Cov(\hat{B}, \bar{x}) = E[\hat{B}\bar{x}] - E(\hat{B})E(\bar{x})$$

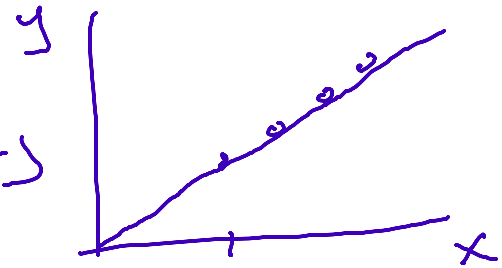
$$= E\left[\frac{\bar{y}}{\cancel{\bar{x}}} \cdot \cancel{\bar{x}}\right] - E(\hat{B})E(\bar{x})$$

$$= \bar{y}_U - E(\hat{B})\bar{x}_U$$

* SRS $\bar{y}, \bar{x}$ are unbiased est.

$$E(\hat{B}) = \frac{\bar{y}_U - Cov(\hat{B}, \bar{x})}{\bar{x}_U} = \underbrace{\boxed{\frac{\bar{y}_U}{\bar{x}_U}}}_{=B} - \frac{Cov(\hat{B}, \bar{x})}{\bar{x}_U}$$

Theory: Bias approximation



What is bias if $y_i \approx \overset{c}{\cancel{\beta}} x_i$ for all i?

$$\hat{B} = \frac{\bar{y}}{\bar{x}} \approx \frac{c\,\bar{x}}{\bar{x}} = c \rightarrow \text{constant value doesn't vary sample to sample}$$

$$\mathrm{Cov}(\hat{B}, \bar{x}) \approx \mathrm{Cov}(c, \bar{x}) = \underline{0}$$

$$\underline{\text{Bias} \approx 0}$$

strong linear assoc. x + y

Theory: Bias approximation

What is the maximum value of bias (approximately)?

$-1, +1$

$$\underset{R}{\mathrm{Cor}} = \frac{\mathrm{Cov}}{SE_1 \, SE_2}$$

$$0 \le \left|\mathrm{Bias}(\hat{B})\right| = \left| - \frac{\mathrm{Cov}(\hat{B}, \bar{X})}{\bar{X}_u} \right| = \left| \frac{\mathrm{Cor}(\hat{B}, \bar{X}) \cdot SE(\hat{B}) \, SE(\bar{X})}{\bar{X}_u} \right|$$

$$0 \le \left|\mathrm{Bias}(\hat{B})\right| \le \left| \frac{1 \cdot SE(\hat{B}) \, SE(\bar{X})}{\bar{X}_u} \right|$$

upper bound on bias $= \dfrac{SE(\hat{B}) \, SE(\bar{X})}{|\bar{X}_u|}$

$SE(\bar{X}) \rightarrow$ SRS SE for $\bar{X}$

$\hookrightarrow$ bias small when n is big