

From last time:  $B = \bar{y}/\bar{x}$

Stat 576

4-15-25

$$\hat{B} = \bar{y}/\bar{x}$$

①

$$E[\hat{B}] = B + \frac{1}{\bar{x}^2} [B V(\bar{x}) - \text{Cov}(\bar{x}, \bar{y})]$$

$$V[\hat{B}] = \frac{1}{\bar{x}^2} [B^2 V(\bar{x}) + V(\bar{y}) - 2B \text{Cov}(\bar{x}, \bar{y})]$$

We know  $V(\bar{x})$ ,  $V(\bar{y})$ , but what is  $\text{Cov}(\bar{x}, \bar{y})$ ?

---

Let  $U_j = X_j + Y_j$ ,  $u_i = x_i + y_i$ ,  $\bar{u} = \bar{x} + \bar{y}$

Find  $V(\bar{u})$  2 different ways.

First,  $V(\bar{u}) = V(\bar{x} + \bar{y})$

②

$$= V(\bar{x}) + V(\bar{y}) + 2\text{Cov}(\bar{x}, \bar{y})$$

Second,  $V(\bar{u}) = \frac{1}{n} S_u^2 (1 - \frac{1}{n})$

$$S_u^2 = \frac{1}{n-1} \sum_{j=1}^n (U_j - \bar{u})^2$$

$$= \frac{1}{n-1} \sum_{j=1}^n \left( X_j + Y_j - (\bar{x} + \bar{y}) \right)^2$$

$$\begin{aligned}
 &= \frac{1}{N-1} \left[ \sum_{i=1}^N (X_i - \bar{X})^2 + \sum_{i=1}^N (Y_i - \bar{Y})^2 \right. \\
 &\quad \left. + 2 \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \right] \quad (3) \\
 &= S_x^2 + S_y^2 + 2 \underbrace{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}_{S_{xy}}
 \end{aligned}$$

Equate the 2 results

$$V(\bar{X}) + V(\bar{Y}) + 2 \text{Cov}(\bar{X}, \bar{Y}) \stackrel{\text{set}}{=} \left( S_x^2 + S_y^2 + 2 S_{xy} \right) \frac{1}{n} \left( 1 - \frac{n}{N} \right)$$

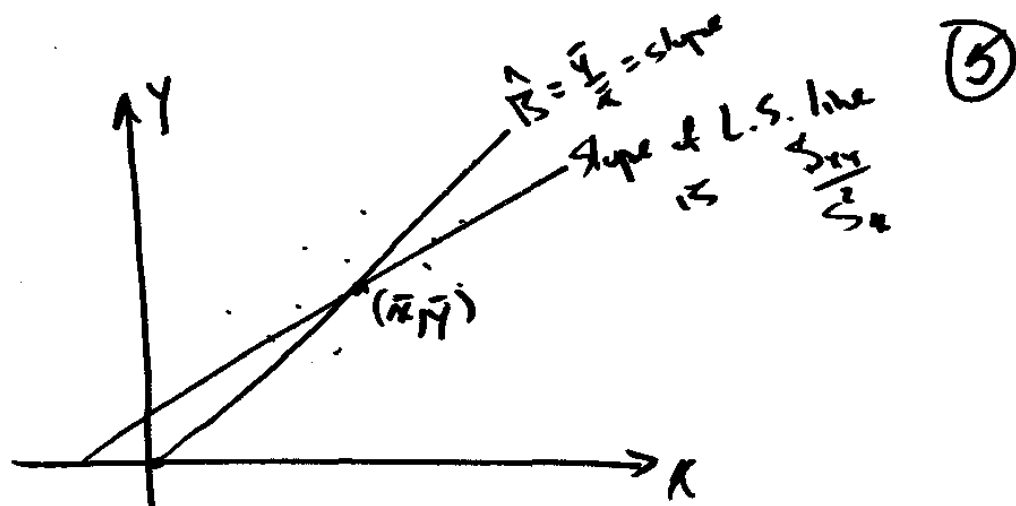
$$\therefore \text{Cov}(\bar{X}, \bar{Y}) = \frac{1}{n} S_{xy} \left( 1 - \frac{n}{N} \right) \quad (4)$$

---


$$\begin{aligned}
 \text{Now } E(\hat{B}) &= B + \frac{1}{X^2} \left[ B \frac{1}{n} \left( 1 - \frac{n}{N} \right) S_x^2 - \frac{1}{n} \left( 1 - \frac{n}{N} \right) S_{xy} \right] \\
 &= B + \frac{1}{X^2} \frac{1}{n} \left( 1 - \frac{n}{N} \right) \left[ B S_x^2 - S_{xy} \right]
 \end{aligned}$$

Note: as  $n \rightarrow N$ ,  $\text{Bias}(\hat{B}) \rightarrow 0$

Also note: If  $B = \frac{S_{xy}}{S_x^2}$ , the bias term vanishes



that is, if the least-squares line passes through the origin, then  $\hat{B}$  will be an unbiased estimator of  $B$ .

$$V(\hat{B}) = \frac{1}{\bar{X}^2} \left[ B^2 \frac{1}{n} S_x^2 + \frac{1}{n} S_y^2 - 2B \frac{1}{n} S_{xy} \right] \quad (6)$$

$$= \frac{1}{n} \left( 1 - \frac{1}{N} \right) \frac{1}{\bar{X}^2} \left[ B^2 S_x^2 + S_y^2 - 2B S_{xy} \right]$$

Note:  $V(\hat{B}) \rightarrow 0$  as  $n \rightarrow N$

So  $\hat{B}$  is a consistent estimator of  $B$

$$\hat{V}(\hat{B}) = \frac{1}{n} \left( 1 - \frac{1}{N} \right) \frac{1}{\bar{X}^2} \left[ \hat{B}^2 S_x^2 + S_y^2 - 2\hat{B} S_{xy} \right]$$

$$\text{let } s_e^2 = \hat{B}^2 s_x^2 + s_y^2 - 2\hat{B}s_{xy}$$

(7)

$$\text{So } \hat{V}(\hat{B}) = \frac{1}{n} \left(1 - \frac{1}{N}\right) \frac{1}{K^2} s_e^2$$

A confidence interval for B is

$$\hat{B} \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{B})}, \text{ or}$$

$$\bar{y}_{\hat{K}} \pm z_{\alpha/2} \sqrt{\frac{1}{n} \left(1 - \frac{1}{N}\right) \frac{1}{K^2} s_e^2}$$

But our real goal was to estimate  $\bar{Y}$

(8)

Primitive method: use  $\bar{y}$

Ratio method:  $B = \frac{\bar{y}}{\bar{x}}$  so  $\bar{Y} = B\bar{X}$

$$\text{let } \bar{y}_{\text{rat}} = \hat{B}\bar{X} = \frac{\bar{y}}{\bar{x}} \bar{X}$$

Which is better?

$$E[\bar{y}_{\text{rat}}] = E[\hat{B}\bar{X}] = \bar{X} E[\hat{B}]$$

$$\approx \bar{X} \left[ B + \frac{1}{n} \text{tr} \frac{1}{K^2} (B s_x^2 - s_{xy}) \right]$$

$$= \bar{Y} + \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{\bar{X}} (B S_x^2 - S_{xy}) \quad (9)$$

$\bar{Y}_{rat}$  is Approx unbiased

As  $n \rightarrow N$  the Bias  $\rightarrow 0$

$$\begin{aligned} V[\bar{Y}_{rat}] &= V[\hat{B} \bar{X}] = \bar{X}^2 V[\hat{B}] \\ &= \bar{X}^2 \left[ \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{\bar{X}^2} [B^2 S_x^2 + S_y^2 - 2B S_{xy}] \right] \end{aligned}$$

Compare this to  $V[\bar{Y}] = \frac{1}{n} \left(1 - \frac{n}{N}\right) S_y^2$

When is  $V[\bar{Y}_{rat}] < V[\bar{Y}]$  ?

(10)

True when  $B^2 S_x^2 - 2B S_{xy} < 0$

$$\begin{aligned} \text{Case 1: } B > 0 \quad B S_x^2 &< 2 S_{xy} \\ B &< 2 \frac{S_{xy}}{S_x^2} \end{aligned}$$

$$\begin{aligned} \text{Case 2: } B < 0 \quad B S_x^2 - 2 S_{xy} &> 0 \\ B &> 2 \frac{S_{xy}}{S_x^2} \end{aligned}$$

That is, if  $B$  is between  $\pm 2$  L.S. Slope, the ratio estimator beats  $\bar{Y}$

(11)

## Regression & difference estimators

$$\text{Let } \bar{y}_{\text{reg}} = \bar{y} - B(\bar{\kappa} - \bar{x})$$

↑  
any fixed constant

$$\begin{aligned} E[\bar{y}_{\text{reg}}] &= E[\bar{y}] - B E(\bar{\kappa} - \bar{x}) \\ &= \bar{y} \end{aligned}$$

Special case  $B=0$  :  $\bar{y}_{\text{reg}} = \bar{y}$

Special case  $B=1$  :  $\bar{y}_{\text{diff}} = \bar{y} - (\bar{\kappa} - \bar{x})$

(12)

$$V[\bar{y}_{\text{reg}}] = V[\bar{y} - B(\bar{\kappa} - \bar{x})]$$

$$= V[\bar{y}] + B^2 V(\bar{\kappa}) - 2B \text{Cov}(\bar{y}, \bar{\kappa})$$

$$= \frac{1}{n} \sum_y t_{pc} + B^2 \frac{1}{n} \sum_{\kappa} t_{pc} - 2B \frac{1}{n} \sum_{\kappa y} t_{pc}$$

$$= \frac{1}{n} \left(1 - \frac{1}{N}\right) \underbrace{\left[ \sum_y^2 + B^2 \sum_{\kappa}^2 - 2B \sum_{\kappa y} \right]}_{\sum_e^2}$$

For the difference estimator,

$$V[\bar{y}_{\text{diff}}] = \frac{1}{n} \left(1 - \frac{1}{N}\right) \left[ \sum_y^2 + \sum_{\kappa}^2 - 2 \sum_{\kappa y} \right]$$

When is  $\bar{y}_{reg}$  minimized?

(13)

Occurs when  $B = \frac{S_{xy}}{S_x^2}$

When this value is used, what is the minimum variance obtained?

$$\begin{aligned} V[\bar{y}_{reg}] &= \frac{1}{n} \left(1 - \frac{1}{n}\right) \left[ S_y^2 + \frac{S_{xy}^2}{S_x^4} S_x^2 - 2 \frac{S_{xy}}{S_x} S_{xy} \right] \\ &= \frac{1}{n} \left(1 - \frac{1}{n}\right) \left[ S_y^2 - \frac{S_{xy}^2}{S_x^2} \right] \end{aligned}$$

$$= \frac{1}{n} \left(1 - \frac{1}{n}\right) S_y^2 \left[ 1 - \frac{S_{xy}^2}{S_x^2 S_y^2} \right]$$

(14)

$$V[\bar{y}_{reg}] = \frac{1}{n} \left(1 - \frac{1}{n}\right) S_y^2 [1 - R^2] < \frac{1}{n} \left(1 - \frac{1}{n}\right) S_y^2$$

as long as  $R^2 > 0$