Stat 576
10-10-17
①

From last time: $B = \bar{Y}/\bar{X}$

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

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

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

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

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.

②

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

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

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

$$S_u^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left(U_j - \bar{U}\right)^2$$

$$= \frac{1}{N-1}\sum_{j=1}^{N}\left(X_j + Y_j - (\bar{X} + \bar{Y})\right)^2$$

(3)

$$= \frac{1}{N-1}\left[\sum_{j=1}^{N}(X_j-\bar{X})^2 + \sum(Y_j-\bar{Y})^2 + 2\sum(X_j-\bar{X})(Y_j-\bar{Y})\right]$$

$$= S_x^2 + S_y^2 + 2\underbrace{\frac{1}{N-1}\sum_{j=1}^{N}(X_j-\bar{X})(Y_j-\bar{Y})}_{S_{xy}}$$

<u>Equate the 2 results</u>

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

(4)

$$\therefore\ \mathrm{Cov}(\bar{x},\bar{y}) = \frac{1}{n}S_{xy}\left(1-\frac{n}{N}\right)$$

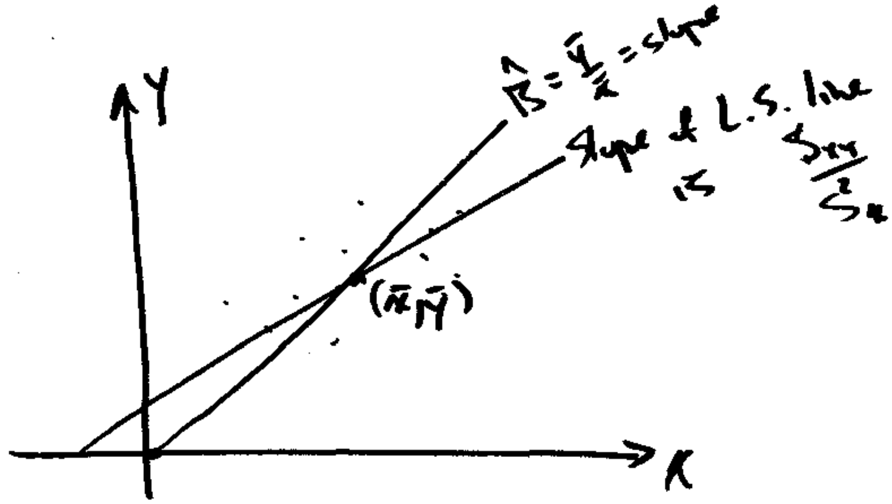
Now $E(\hat{B}) = B + \frac{1}{\bar{X}^2}\left[B\frac{1}{n}\left(1-\frac{n}{N}\right)S_x^2 - \frac{1}{n}\left(1-\frac{n}{N}\right)S_{xy}\right]$

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

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

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

5



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

6

$$V(\hat{B}) = \frac{1}{\bar{X}^2}\Big[B^2 \frac{1}{n} S_x^2 \,\text{fpc} + \frac{1}{n} S_y^2 \,\text{fpc} - 2B\frac{1}{n} S_{xy}\,\text{fpc}\Big]$$

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

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

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

$$\hat{V}(\hat{B}) = \frac{1}{n}\left(1-\frac{n}{N}\right)\frac{1}{\bar{x}^2}\Big[\hat{B}^2 s_x^2 + s_y^2 - 2\hat{B} s_{xy}\Big]$$

(7)

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

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

A confidence interval for B is

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

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

(8)

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

Primitive method: use $\bar{y}$

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

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

Which is better?

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

$$\approx \bar{X}\left[B + \frac{1}{n}\text{fpc}\,\frac{1}{\bar{X}^2}\left(B^2 S_x^2 - S_{xy}\right)\right]$$

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

$\bar{y}_{rat}$ is approx unbiased

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

$$V[\bar{y}_{rat}] = V[\hat{B}\,\bar{x}] = \bar{X}^2 V[\hat{B}]$$

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

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

10

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

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

Case 1: $B > 0$ $\quad B S_x^2 < 2 S_{xy}$

$$B < 2\frac{S_{xy}}{S_x^2}$$

Case 2: $B < 0$ $\quad B S_x^2 - 2 S_{xy} > 0$

$$B > 2\frac{S_{xy}}{S_x^2}$$

That is, if B is between ± 2 L.S. slope, the ratio estimator beats $\bar{y}$

(11)

Regression & difference estimators

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

any fixed constant

$$E[\bar{y}_{reg}] = E[\bar{y}] - B\,E(\bar{x} - \bar{X})$$
$$= \bar{Y}$$

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

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

(12)

$$V[\bar{y}_{reg}] = V[\bar{y} - B(\bar{x} - \bar{X})]$$
$$= V[\bar{y}] + B^2 V(\bar{x}) - 2B\,Cov(\bar{y}, \bar{x})$$
$$= \frac{1}{n} S_Y^2 \, fpc + B^2 \frac{1}{n} S_x^2 \, fpc - 2B \frac{1}{n} S_{xy} \, fpc$$
$$= \frac{1}{n}\left(1 - \frac{n}{N}\right)\underbrace{\left[S_Y^2 + B^2 S_x^2 - 2B S_{xy}\right]}_{S_e^2}$$

For the difference estimator,

$$V[\bar{y}_{diff}] = \frac{1}{n}\left(1 - \frac{n}{N}\right)\left[S_Y^2 + S_x^2 - S_{xy}\right]$$

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

Occurs when $B = \dfrac{S_{XY}}{S_X^2}$

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

$$V[\bar{y}_{reg}] = \frac{1}{n}\left(1-\frac{n}{N}\right)\left[S_Y^2 + \frac{S_{XY}^2}{S_X^4}S_X^2 - 2\frac{S_{XY}}{S_X}S_{XY}\right]$$

$$= \frac{1}{n}\left(1-\frac{n}{N}\right)\left[S_Y^2 - \frac{S_{XY}^2}{S_X^2}\right]$$

14

$$= \frac{1}{n}\left(1-\frac{n}{N}\right)S_Y^2\left[1 - \frac{S_{XY}^2}{S_X^2 S_Y^2}\right]$$

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

as long as $R^2 > 0$