- Cross-sectional data – Sample at a time instant.
- Time series data

- Sample isn't necessarily perfectly indicative of what's going on in popⁿ
  — SAMPLING ERROR

- Causal effect / Reversed causal effect – Eg. Income of people after Military service. It could be that they earn less, but it could also be that the people who have low earning capability actually go to military

LI – Life time income, LIP – Life time income potential

$LI = f(M, LIP)$

— Selection bias: Difference betⁿ the individuals, groups (test & control) under comparison, Eg. LIP betⁿ Civilian & M. Ideally, here, one would pick random individuals and send to M or C. and then analyse.

SAMPLING ERROR:

sample 1 wage w education E

$$\left.\begin{aligned} w_1 &= \alpha + \beta_1 E \\ w_2 &= \alpha + \beta_2 E \\ &\vdots \\ w_{1000} &= \alpha + \beta_{1000} E \end{aligned}\right\} \text{Average } w = \alpha^s + \beta^s E$$

In actual population let's say that $w^P = \alpha^P + \beta^P E$

$\beta^s \neq \beta^P \Rightarrow$ due to sampling error.

β^s: estimate of average effect of educⁿ on wage

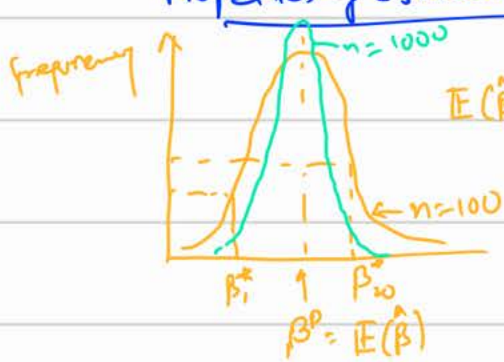
ESTIMATORS: Get popⁿ parameter from SAMPLE

sample data $\longrightarrow \hat{\beta} \longrightarrow \beta^*$

$\hat{\beta}$: estimator (a math function)

β^*: point estimate of popⁿ parameter

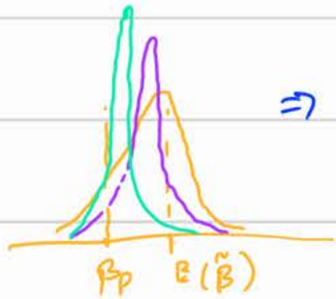
Properties of Estimators:



$\mathbb{E}(\hat{\beta}) = \beta^P$ – unbiased ① on average we get popⁿ parameter

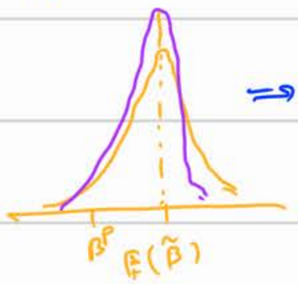
② as $n \uparrow$, β^* is closer to β^P

$n \to \infty$ $\hat{\beta} \to \beta^P$ – consistency



$\Rightarrow$ consistent, but biased

as $\uparrow$ inc. sample size, $\hat{\beta}$ gets closer of β^P



$\Rightarrow$ Biased & unconsistent

increase sample size gives same value

Unbiased vs. consistent:

Mean Height Pop^n: μ Sample $= \tilde{X} = \frac{1}{N-1}\sum_{i=1}^{N} x_i$

Is $\tilde{X}$ a biased or unbiased estimator of μ?

Assume: Pop^n individuals height $x_i = \mu + \varepsilon_i$ ε_i - with mean zero error

$$\mathbb{E}[\tilde{X}] = \frac{1}{N-1}\sum_{i=1}^{N}\mathbb{E}[X_i] = \frac{1}{N-1}N\cdot\mu = \frac{N\cdot\mu}{N-1} \neq \mu$$

just for demonstration it should be N

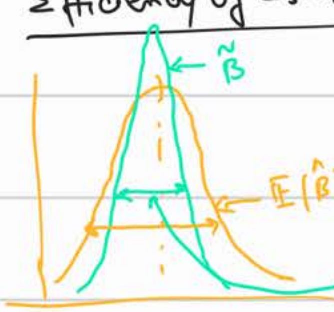
For same N, Biased Estimator

But as $N \to \infty$

$\mathbb{E}[\tilde{X}] \to \mu$

so, eventhough the estimator is biased, it's consistent.

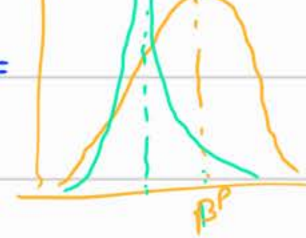
Efficiency Of Estimator:



$\longrightarrow$ $\tilde{\beta}$ is more efficent than $\hat{\beta}$

$\mathbb{E}(\hat{\beta}) = \beta^P$ unbiased

smaller range of value



efficient but biased

Tradeoff bet^n efficiency & biasedness $\Leftarrow$

Overall

1. Unbiased $\mathbb{E}(\hat{\beta}) = \beta^P$
2. consistency $n \to \infty \Rightarrow \hat{\beta} \to \beta^P$
3. Efficiency
4. Linear in parameters - $\hat{\beta} \sim$ linear function of parameters

Lines of Best Fit:

$$S = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

squaring puts more weight on the deviation

$$S' = \sum_{i=1}^{N} (y_i - \hat{y}_i)^4 \rightarrow \text{still more weight}$$

resulting in blue line

Least square as a Good Estimator:

B L U E — Estimator

B — Best (efficient)

L — Linear

U — Unbiased

BLUE - under Least square assumption.

$$S = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

$$\frac{\partial S}{\partial \hat{\alpha}} = 0 = -2 \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i) \quad \text{— (1)}$$

$$\frac{\partial S}{\partial \hat{\beta}} = 0 = -2 \sum_{i=1}^{N} x_i (y_i - \hat{\alpha} - \hat{\beta} x_i) \quad \text{— (2)}$$

$$\frac{1}{N} \sum_{i=1}^{N} x_i = \bar{x}$$

$$\frac{1}{N} \sum_{i=1}^{N} y_i = \bar{y}$$

$$\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \sum (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})$$

$$= \sum x_i y_i - \bar{y} \sum x_i - \bar{x} \cdot \sum (y_i) + N \bar{x}\bar{y}$$

$$= \sum x_i y_i - N \bar{y}\bar{x} - N \bar{x}\bar{y} + N \bar{x}\bar{y}$$

$$= \sum x_i y_i - N \bar{y}\bar{x} = \sum x_i y_i - \bar{x} \sum y_i$$

$$= \sum (y_i x_i - y_i \bar{x})$$

$$\boxed{\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum y_i (x_i - \bar{x})}$$

by symmetry $= \sum x_i (y_i - \bar{y})$

From (1)

$$\sum_{i=1}^{N} y_i = \hat{\alpha} \sum_{i=1}^{N} 1 + \hat{\beta} \sum x_i$$

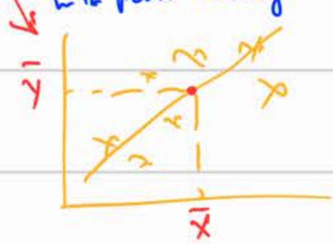
$$N \bar{y} = \hat{\alpha} N + \hat{\beta} N \bar{x}$$

$$\boxed{\bar{y} = \hat{\alpha} + \hat{\beta} \bar{x}} \quad \text{— (3)}$$

This means that the line

$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

will pass through sample mean



From (2)

$$\sum x_i y_i = \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2$$

$$= \hat{\alpha} N \bar{x} + \hat{\beta} \sum x_i^2$$

Using (3) $\sum x_i y_i = (\bar{y} - \hat{\beta} \bar{x}) N \bar{x} + \hat{\beta} \sum x_i^2$

$$\sum x_i y_i = N\bar{x}\bar{y} - N\hat{\beta}\bar{x}^2 + \hat{\beta}\sum x_i^2$$

$$\sum x_i y_i - N\bar{x}\bar{y} = \hat{\beta}\left(\sum x_i^2 - N\bar{x}^2\right)$$

$$\hat{\beta} = \frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sum x_i^2 - N\bar{x}^2}$$

$$\hat{\beta} = \frac{\sum (x_i y_i - \bar{x} y_i)}{\sum (x_i^2 - \bar{x} x_i)} \qquad N\bar{y} = \sum_{i=1}^{N} y_i$$

from previous

$$\boxed{\sum (x_i-\bar{x})(y_i-\bar{y}) = \sum y_i (x_i - \bar{x})}$$

by symmetry $= \sum x_i (y_i - \bar{y})$

$$\hat{\beta} = \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i^2 - \bar{x} x_i)}$$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\boxed{\hat{\beta} = \frac{\operatorname{cov}(x_i, y_i)}{\operatorname{Var}(x_i)}}$$