SLIDES 62 and 63

(a)

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

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

$\bar{x} = \frac{1}{T}\sum x_i \Rightarrow T\bar{x} = \sum_i x_i$ and same for $\bar{y}$: $T\bar{y} = \sum y_i$

$$= \sum x_i y_i - T\bar{x}\bar{y} = \sum x_i y_i - \bar{x}\sum y_i = \sum y_i (x_i - \bar{x}) \qquad \underline{A}$$

Slide 47 this is the numerator of $\hat{\beta}$

$$\sum_i (x_i-\bar{x})^2 = \sum_i \left(x_i^2 - 2x_i\bar{x} + (\bar{x})^2\right) = \sum_i x_i^2 - 2\bar{x}\sum_i x_i + T(\bar{x})^2$$

$$= \sum_i x_i^2 - 2\bar{x}(T\bar{x}) + T(\bar{x})^2 = \sum_i x_i^2 - 2T(\bar{x})^2 + T(\bar{x})^2$$

$$= \sum_i x_i^2 - T(\bar{x})^2 \qquad \underline{B}$$

Slide 47, this is the denominator of $\hat{\beta}$

as of slide 47 — from A, B

$$\hat{\beta} = \frac{A}{B} = \frac{\sum_i x_i y_i - T\bar{x}\bar{y}}{\sum_i x_i^2 - T\bar{x}^2} = \frac{\sum_i y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

two equivalent ways of writing $\hat{\beta}$

equation (7) on slide 62

$y_i = \beta_0 + \beta_1 x_i + u_i$ by definition for simple linear models (Slide 38)

$$\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\underbrace{\sum_i (\underbrace{x_i - \bar{x}}_{\text{"Variation"}})^2}_{\text{"Squared Variation"}}} \quad \underline{C}$$

"Total Squared variation"

$\leftarrow SST_x$ "Sample squared total Variation"

note that β_0, β_1 are the "true" population parameters (unknown, of the Population regression function PRF)

Focus on the numerator $\underline{C}$: $\sum_i (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)$

$$= \sum_i \left[\beta_0 (x_i - \bar{x}) + \beta_1 x_i (x_i - \bar{x}) + u_i (x_i - \bar{x})\right] = \beta_0 \sum_i (x_i - \bar{x}) \ \underline{D} + \beta_1 \sum_i (x_i - \bar{x}) x_i \ \underline{E} + \sum_i (x_i - \bar{x}) u_i$$

$\underline{D}$ $\beta_0 \sum_i (x_i - \bar{x}) = (\sum x_i - T\bar{x})\beta_0$

$$= \left(\sum_i x_i - \cancel{T}\frac{1}{\cancel{T}}\sum_i x_i\right)\beta_0 = 0$$

Definition of $\bar{x} = \frac{1}{T}\sum_i x_i$

"the sum of deviations from the sample mean is always zero"!

$\Rightarrow$ $\underline{D}$ is zero

$\rightarrow$ for the numerator $\underline{C}$ we have

$$\sum_i (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i) = \sum_i (x_i - \bar{x}) u_i + \beta_1 \sum_i (x_i - \bar{x}) x_i$$

going back to $\hat{\beta}$ $\left(=\frac{A}{B}\right)$

$$\hat{\beta}_1 = \frac{\sum_i (x_i-\bar{x})(\beta_0+\beta_1 x_i + u_i)}{SST_x} = \frac{1}{SST_x}\left[\beta_1 \underbrace{\sum_i (x_i-\bar{x})x_i}_{E} + \sum_i (x_i-\bar{x})u_i\right]$$

$T\bar{x} = \sum_i x_i$ *

$$E : \sum_i (x_i-\bar{x})x_i = \sum_i [x_i^2 - \bar{x}x_i] = \sum_i x_i^2 - \bar{x}\sum x_i = \sum_i x_i^2 - \bar{x}\,T\bar{x}$$

$$= \sum_i x_i^2 - T\bar{x}^2 = \sum_i x_i^2 \underbrace{- 2T\bar{x}^2 + T\bar{x}^2}_{= -T\bar{x}^2}$$

$$= \sum_i x_i^2 - 2T\bar{x}\bar{x} + T\bar{x}^2$$

$$= \sum_i x_i^2 - 2\bar{x}\sum_i x_i + T\bar{x}^2$$

$$= \sum_i \left[x_i^2 - 2\bar{x}x_i + \bar{x}^2\right]$$ this is the square of $(x_i-\bar{x})$

$$= \sum_i \left((x_i-\bar{x})^2\right) \overset{\text{By definition of } SST_x}{=} SST_x$$

$$\hat{\beta}_1 = \frac{1}{SST_x}\left[\beta_1 SST_x + \sum_i (x_i-\bar{x})u_i\right] = \beta_1 + \frac{\sum_i \overbrace{(x_i-\bar{x})}^{d_i} u_i}{SST_x}$$

$$= \beta_1 + \frac{1}{SST_x}\sum_i d_i u_i$$

u_i : random error for observation i

d_i are deterministic: Based entirely on the observed values of the x_i, $i=1,\dots,T$

interpretation: $\hat{\beta}$ Deviates from β_1 (the true unknow population parameter)

by some random term that is entirely due to the errors in the sample

the "real" unobserved ones, not the observed residuals after $\hat{\beta}_1$ and $\hat{\beta}_0$ are estimated!

$E(\hat{\beta}_1) = E\left\{ \beta_1 + \frac{1}{SST_x} \sum_i d_i u_i \right\} = (\beta_1) + \frac{1}{SST_x} \sum_i E(d_i u_i)$ Slide 66

linearity of expectation (applyied twice)

- $E(a+b) = E(a) + E(b)$ that is
- $E(\sum_i a_i) = \sum_i E(a_i)$

d_i are deterministic, not random

$= \beta_1 + \frac{1}{SST_x} \underbrace{\sum_i d_i E(u_i)}_{=0} = \beta_1 \Rightarrow E(\hat{\beta}_1) = \beta_1$

β_1 is a real number, not a random variable $E(\beta_1) = \beta_1$

this prooves that $\hat{\beta}_1$ is unbiased for β_1

why is $\sum d_i E(u_i) = 0$?

Assumption 4 : $E(u_i | x_i) = 0$

Recall the law of total expectation $E(x) = E(E(x|y))$

"The expectation of x is the expectation of $\underline{E(x|y)}$ for some conditioning variable y"

Random Variable

$z = E[x|y]$ is a random variable!

eg.

$z_1 = E[x | y = y_1]$

$z_2 = E[x | y = y_2]$

$\vdots$

$z_n = E[x | y = y_m]$

depending on what's the outcome of y you get a different outcome of $z = x|y$

So z is a random variable and makes sence to take $E[E(x|y)]$

$E[E(u_i|x_i)] = E[0] = 0$

$\Rightarrow \sum_i d_i E(u_i) = 0$

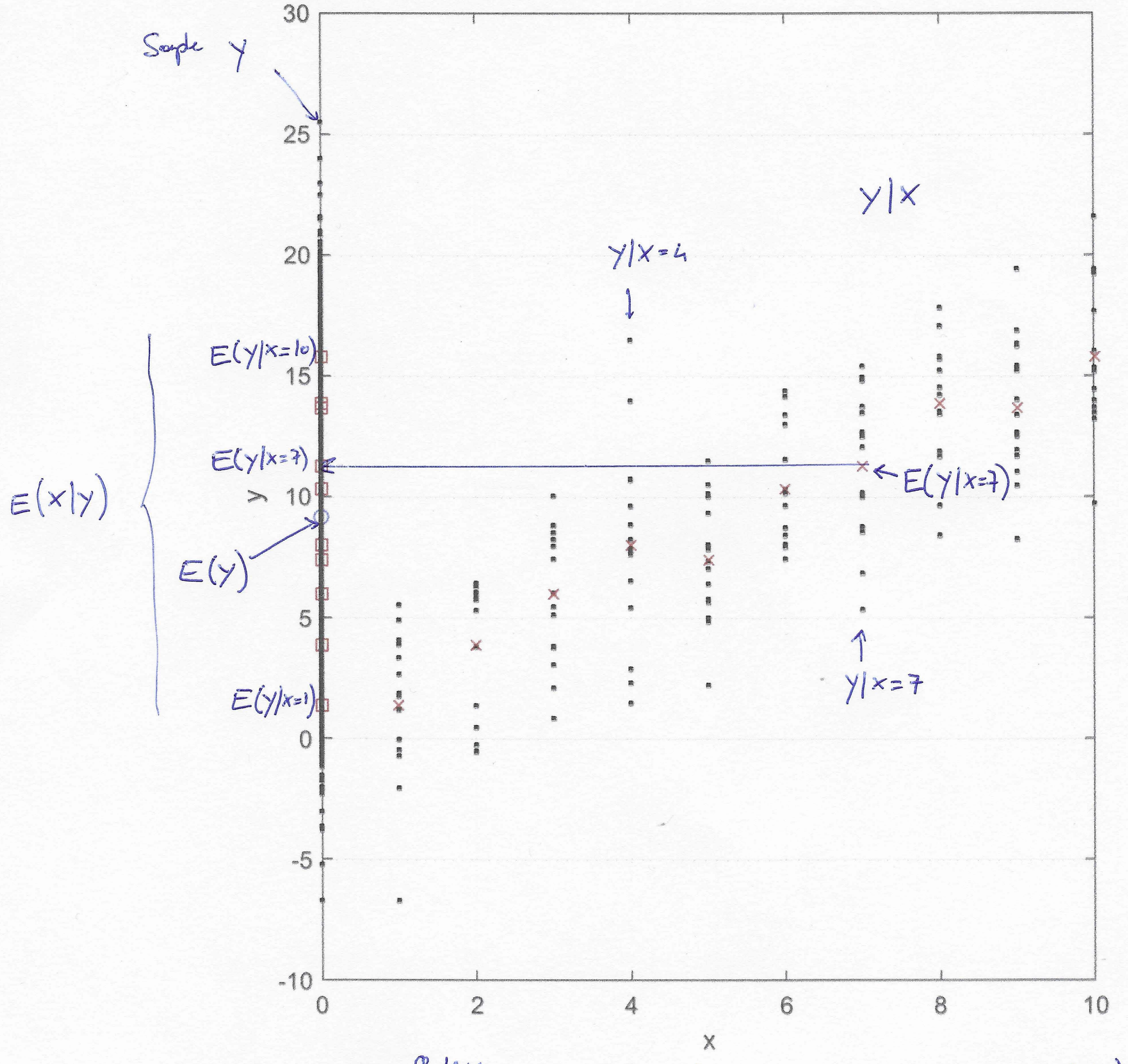
$\forall i \; E(u_i) = 0$

Note: if Assumption 4 fails, $E(u_i) \neq 0$

So $E(\hat{\beta}_1) = \beta_1 + \text{Something} \neq 0 \neq \beta_1$

→ the estimator is Biased!

Low of total expectation



9,1616

(1.39, 3.88, 5.99, ..., 15.81)

$$E(Y) = E\left[E(Y|X)\right]$$

the value "○" is the Average of the "□"

where each "□" is the Average of $Y|X=x_i$

$$\mathbb{E}(\hat{\beta}_0) \underset{\substack{\uparrow \\ \text{OLS definition/solution} \\ \text{for } \hat{\beta}_0}}{=} \mathbb{E}(\bar{y} - \hat{\beta}_1\bar{x}) = \mathbb{E}(\underbrace{\beta_0 + \beta_1\bar{x} + \bar{u}}_{\bar{y} = \beta_0 + \beta_1\bar{x} + \bar{u}} - \hat{\beta}_1\bar{x})$$

$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u} \rightarrow$ See Notes on slide 67

$$= \mathbb{E}(\beta_0 + \bar{x}(\beta_1 - \hat{\beta}_1) + \bar{u})$$

$$\underset{\text{linearity of } \mathbb{E}}{=} \mathbb{E}(\beta_0) + \underset{\text{non random}}{\bar{x}}\,\mathbb{E}(\beta_1 - \hat{\beta}_1) + \mathbb{E}(\bar{u})$$

$$= \beta_0 + \bar{x}\{\beta_1 - \mathbb{E}(\hat{\beta}_1)\} + \mathbb{E}\left[\frac{1}{T}\sum_i u_i\right]$$

$$= \beta_0 + \cancel{\bar{x}\,0} + \underbrace{\frac{1}{T}\cancel{\sum_i \mathbb{E}(u_i)}}_{=0} = \beta_0$$

$\mathbb{E}(\hat{\beta}_1) = \beta_1$ Because we showed that the estimator is unbiased

$=0$ Law of total expectation + Assumption 4 $(\mathbb{E}(u_i|x_i) = 0 \ \forall i)$

$\Rightarrow \mathbb{E}(\hat{\beta}_0) = \beta_0$, Also the OLS estimator $\hat{\beta}_0$ is unbiased

Slide 71

Recall: $Var(x) = \mathbb{E}(x^2) - (\mathbb{E}(x))^2$ *

$$\sigma^2 \underset{\substack{\uparrow \\ \text{by definition, Assuption 5}}}{=} V(u|x) \overset{*}{=} \mathbb{E}[u^2|x] - \underbrace{(\underset{\substack{\uparrow = 0 \\ \text{Assumption 4}}}{\mathbb{E}(u|x)})^2}_{=0} = \mathbb{E}(u^2|x)$$

So $\sigma^2 = V(u|x)$ *'

$\quad\;\; = \mathbb{E}(u^2|x)$

"Law of total Variance": $V(x) = \mathbb{E}(V(x|y)) + V(\mathbb{E}(x|y))$

In our case:

$$V(u) = \underbrace{\mathbb{E}(V(u|x))}_{=\sigma^2(*')} + \underbrace{V(\mathbb{E}(u|x))}_{=0,\ Ass.4}$$

$$= \mathbb{E}(\underset{\substack{\uparrow \\ \text{Some Real number}}}{\sigma^2}) = \sigma^2$$

$\rightarrow$ not only $\sigma^2 = V(u|x)$ but also $\sigma^2 = V(u)$

σ^2 is the conditional error Variance and σ^2 is the unconditional error Variance

Some More notes:

Notes SLIDE 67

(1) $\bar{y} = a + b\bar{x} + \bar{u}$ this because $y_i = a + bx_i + u_i$ (Assumption 1)

$$\Sigma_i y_i = \Sigma_i (a + bx_i + u_i)$$

$$\Sigma_i y_i = \Sigma_i a + b\Sigma_i x_i + \Sigma u_i$$ divide all by T

$$\frac{1}{T}\Sigma_i y_i = \frac{1}{\cancel{T}}\cancel{T}a + b\frac{1}{T}\Sigma_i x_i + \frac{1}{T}u_i$$ by definition of Sample Average

$$\bar{y} = a + b\bar{x} + \bar{u}$$

(2) Assuption 1 : $y_i = a + bx_i + u_i$

~~Notes~~ SLIDE 72

a take $E(y_i|x_i)$ this is $E(a + bx_i + u_i|x_i)$

$$= a + bx + \underbrace{E(u_i|x_i)}_{=0 \text{ Ass. 4}}$$

So writing Ass. 4 $E(y_i|x_i) = 0$ is the same as writing
$$E(y_i|x_i) = a + bx$$

b take $V(y_i|x_i)$ this is $V(a + bx_i + u_i|x_i)$

$$= V(\cancel{a}|x_i)^{=0} + V(\cancel{bx_i}|x_i)^{=0} + V(u_i|x_i)$$ [there is no covariance as there is only 1 Random Variable in the expression: In general: $Var(x+y) = Var(x) + Var(y) + \underline{\underline{2cov(x,y)}}$]

not random, these are just real numbers
Var (Real number) = 0

$$= V(u_i|x_i)$$

So Writing Ass. 5 $V(y_i|x_i) = \sigma^2$ is the same as writing
$$V(y_i|x_i) = \sigma^2$$

On Slide 63 we obtained

Slide 74 (a)

$$\hat{\beta}_1 = \beta_1 + \frac{1}{SST_x} \sum_i d_i u_i$$

Now, let's look at the Variance of $\hat{\beta}_1$

$$V(\hat{\beta}_1) = V\left(\beta_1 + \frac{1}{SST_x} \sum d_i u_i\right)$$

$$= \frac{1}{SST_x^2} Var\left(\sum d_i u_i\right) \overset{A}{\underset{*}{=}} \frac{1}{SST_x^2} \sum d_i^2 \, Var(u_i)$$

- $Var(\text{constant}) = 0$ i.e. $Var(\beta_1) = 0$
- $Var(aX) = a^2 Var(x)$ with a being a constant

* In general $Var(x+y) = Var(x) + Var(y) + \underset{\uparrow \text{ Covariance}}{2\,Cov(x,y)}$ and for summation

$$\left.\begin{aligned} Var\left(\sum_i x_i\right) &= \sum Var(x_i) + 2\sum_{i<j} Cov(x_i, x_j) \\ &= \sum Var(x_i) + \sum_{i \neq j} Cov(x_i, x_j) \end{aligned}\right\} \text{equivalent}$$

So $Var\left(\sum_i d_i u_i\right) = \sum Var(d_i u_i) + \sum_{i \neq j} Cov(d_i u_i, d_j u_j)$

$$= \sum_i d_i^2 \, Var(u_i) + \sum_{i \neq j} d_i d_j \, Cov(u_i, u_j)$$

- $Var(aX) = a^2 Var(X)$
- $Cov(aX, bY) = ab\, Cov(X, Y)$

Why is $Cov(u_i, u_j) = 0$?

Also related to slides 93, 94

→ Assumption 2: Random Sampling

Step 1 (x_i, y_i) (x_j, x_j) are independent → y_i and y_j are independent

→ the randomness in y_i and y_j is entirely due to u_i and u_j

→ u_i and u_j are independent

Step 2 Now show that $Cov(X, Y) = 0$ if X and Y are independent

$$Cov(X,Y) \underset{\text{By definition}}{=} E\big((X - E(X))(Y - E(Y))\big) = E\left[XY - XE(Y) - YE(X) + E(X)E(Y)\right]$$

$$= E[XY] - E(X)E(Y) - \cancel{E(X)E(Y)} + \cancel{E(X)}E(Y)$$

$$= E[XY] - E[X]E(Y)$$

So $Cov(u_i, u_j) = E[u_i u_j) - E(u_i)E(u_j)$

Recall from chapter 1

- X, Y are independent if $F_{XY}(x,y) = F_X(x)\, F_Y(y)$
  and analogously for densities $f_{XY}(x,y) = f_X(x)\, f_Y(y)$
- $\mathbb{E}(X) \overset{\text{Def.}}{=}_{*_1} \int_{\mathbb{R}} x f_X(x)\, dx$ for a simple real-valued R.V. X with support in $\mathbb{R}$

$$\mathbb{E}(XY) = \int_{\mathbb{R}\times\mathbb{R}} xy\, f_{XY}(x,y)\, dx\, dy = \int_{\mathbb{R}}\int_{\mathbb{R}} xy\, f_{XY}(x,y)\, dx\, dy$$

Now for u_i, u_j

$$\mathbb{E}(u_i u_j) = \int_{\mathbb{R}\times\mathbb{R}} u_i u_j\, f_{U_iU_j}(u_i u_j)\, du_i\, du_j \overset{\text{Independence } u_i, u_j}{=} \int_{\mathbb{R}\times\mathbb{R}} u_i u_j\, f_{U_i}(u_i)\, f_{U_j}(u_j)\, du_i\, du_j$$

$$= \int_{\mathbb{R}} u_i f_{U_i}(u_i)\, du_i \int_{\mathbb{R}} u_j f_{U_j}(u_j)\, du_j \overset{*_1}{=} \mathbb{E}(u_i)\, \mathbb{E}(u_j)$$

Break the integral over $\mathbb{R}^2$

$\Rightarrow$ **Note:** if x and y are indep then $Cov(x,y)$ is always zero:

$$Cov(x,y) = E(XY) - E(X)E(Y) \overset{\text{independence}}{=} \cancel{E(X)E(Y)} - \cancel{E(X)E(Y)} = 0$$

(no matter what $E(X)$ and $E(Y)$ are!)

$\Rightarrow Cov(u_i, u_j) = 0$

So $Var\left(\sum_i d_i u_i\right) = \sum_i d_i^2 Var(u_i) + \underbrace{\cancel{\sum_{ij} d_i d_j Cov(u_i, u_j)}}_{=0}$

$$= \sum_i d_i^2 \sigma^2$$

Ass 5.

$$\Rightarrow Var(\hat{\beta}_1) \overset{*}{\underset{A}{=}} \frac{1}{SST_x^2} \sum_i d_i^2 Var(u_i) = \frac{1}{SST_x^2} \sigma^2 \sum_i d_i^2$$

$\sum_i (x_i - \bar{x})^2 = SST_x$

$$= \frac{1}{SST_x^2} \cdot \sigma^2 \cdot \cancel{SST_x} = \frac{\sigma^2}{SST_x}$$

Related to Slide 93

<u>Note 1</u>: a $Cov(u_i, u_j) = 0$ Because of Random Sampling (Ass. 2) [and Ass. 4]

$\rightarrow$ this leads (with Ass. 5) to $Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$

b alternatively you can Drop Ass. 2 and Replace it with Ass 2 (BIS): $Cov(u_i, u_j) = 0 \quad \forall i \neq j$

Here $Var(\hat{\beta}_1) = \sum_i d_i^2 Var(u_i) + \underbrace{\cancel{\sum_{i\neq j} d_i d_j Cov(u_i, u_j)}}_{=0} = \sum_i d_i^2 Var(u_i)$

holds by Assumption 2 Bis, And you don't need most of the proofs above

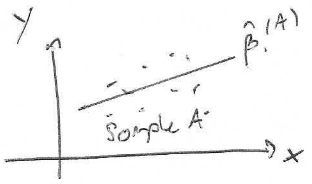
Ass 2 : Random Sampling
Ass 2 bis : $Cov(u_i, u_j) = 0 \quad \forall i \neq j$

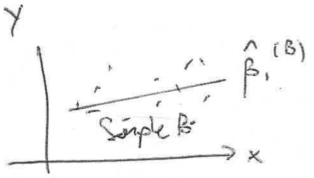
} Are alternatives that Both lead to $Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$ Slide 74 (c) along with the other hypotheses

→ depending on the context you can find Ass 2 (statistics) or Ass 2 bis (time-series)

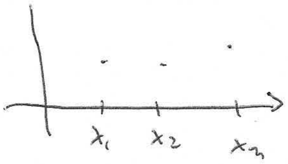
<u>Note 2</u>: we proved that independence implies covariance = 0
⇒ Ass 2 → Ass 2 bis , i.e. Ass 2 is stronger

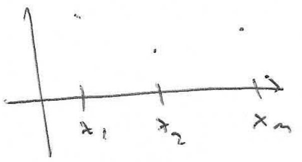
<u>Note 3</u>: theorem on slide 73
" where these are conditional on the sample values $\{x_1, \dots x_n\}$ "
→ what does it mean?



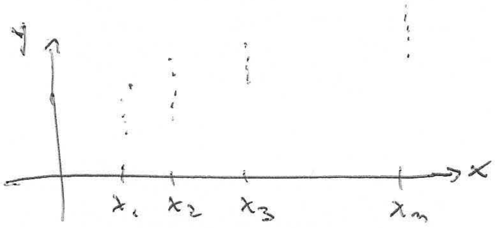


Random Sampling of (x, y) $*$

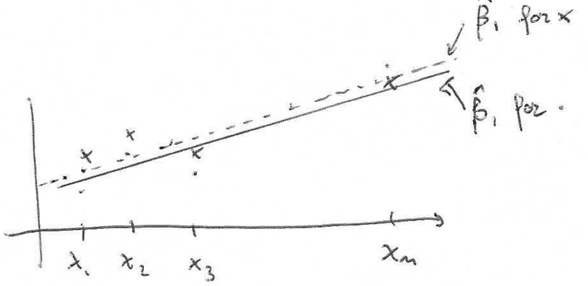




Repeated Sampling of Y given $x_1, \dots, x_n$



We discuss the variance of $\hat{\beta}_1$ in this setting, $*'$
assuming $x_1, x_2, \dots, x_n$ fixed



We look at the Variance of $\hat{\beta}_1$ for repeated samples of y with $x_1, x_2, \dots, x_n$ held fixed

→ $V(\hat{\beta}_1 | x_1, \dots, x_n)$ $*'$ opposed to $V(\hat{\beta}_1)$ $*$

So what we computed is $V(\hat{\beta}_1 | x_1, -, x_n)$ where $x_1, -, x_n$ are fixed and treated as non-random

take Again $V(\sum_i d_i u_i)$ and Assume there is no conditioning on $x_1, -, x_n$

$$V(\underset{\substack{\uparrow \\ \text{Random}}}{\sum_i d_i u_i}) = V(\sum_i \underset{\substack{\uparrow \\ \text{Random}}}{(x_i - \bar{x})} u_i)$$

these are products of R.V => $d_i u_i$ is a New R.V. of which we know nothing about and we have no Hypotheses. Noting to do, conditioning is necessary!

$$= \sum_i Var(d_i u_i) + \sum_{i \neq j} Cov(d_i u_i, d_j u_j)$$

=> everything is much more complex and the earlier result does not apply : $V(\hat{\beta}_1 | x_1, -, x_n) \neq V(\hat{\beta}_1)$

the two are quite different

- conceptually
- on the paper, practically

Recall the OLS solution for $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$Var(\hat{\beta}_0) = V(\bar{y} - \hat{\beta}_1 \bar{x}) = V(\bar{y}) + \bar{x}^2 Var(\hat{\beta}_1) - 2\cancel{cov(\bar{y}, \hat{\beta}_1)}\,\bar{x}$$

$=0$ to proove

Variance of Sample mean = Ch. 1 slide 116

$$= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{SST_x}$$

just proved

$$= \frac{\sigma^2}{n}\frac{SST_x}{SST_x} + \bar{x}^2 \frac{\sigma^2}{SST_x}\frac{n}{n} = \frac{(SST_x + n\bar{x}^2)\,\sigma^2}{n\,SST_x} \overset{*}{=}$$

$=1$ ← does nothing → $=1$

$$SST_x = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - 2\bar{x}\sum_i x_i + \sum_i \bar{x}^2$$

$$= \sum_i x_i^2 - 2n\bar{x}\bar{x} + n\bar{x}^2$$

$$= \sum_i x_i^2 - n\bar{x}^2 \quad \rightarrow \quad SST_x + n\bar{x}^2 = \sum_i x_i^2$$

$$\overset{*}{=} \frac{\sigma^2 \sum_i x_i^2}{n\,SST_x} = \frac{\sigma^2 \sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2}$$

Now show that $cov(\bar{y}, \hat{\beta}_1) = 0$

$$cov(\bar{y}, \hat{\beta}_1) = Cov\left(\frac{1}{n}\sum_i y_i,\ \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_j (x_i - \bar{x})^2}\right)$$

$$= \frac{1}{n}\frac{1}{\sum_i (x_i - \bar{x})^2} Cov\left\{\sum_i y_i,\ \sum_j (x_j - \bar{x}) y_j\right\}$$

- $cov(a x, y) = a\,Cov(x,y)$
- $Cov(\sum_i x_i, y) = \sum_i Cov(x_i, y_i)$

$$= \frac{1}{n}\frac{1}{\sum_i (x_i - \bar{x})^2} \sum_j (x_j - \bar{x}) \sum_i Cov(y_i, y_j)$$

$= 0$ if $i \neq j$ (Ass. 2)

$= \sigma^2$ if $i = j$ [$Cov(x_i, x_i) = Var(x_i)$]

$$= \frac{1}{n}\frac{1}{\sum_i (x_i - \bar{x})^2} \cancel{\sum_i (x_i - \bar{x})}\,\sigma^2 = 0$$

$=0$

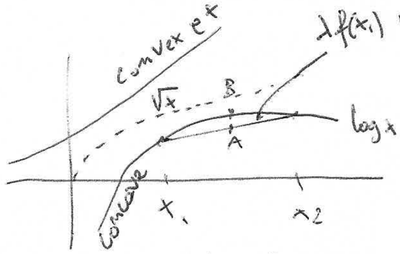
$\sum$ of Deviations from the Sample mean $= 0$

(Notes on slide 62 and 63 (a))

Jensen Inequality: In general for a r.v. X:

Slide 81 $0 < \lambda < 1$

$\lambda f(x_1) + (1-\lambda) f(x_2)$



if φ is convex $\varphi(E(x)) \leq E(\varphi(x))$

if φ is concave $\varphi(E(x)) \geq E(\varphi(x))$

A : $\frac{1}{2}f(x_1) + \frac{1}{2}f(x_2)$

B : $f(\frac{1}{2}x_1 + \frac{1}{2}x_2)$

$A \leq B$ ↑ Def. of concavity!

Show that $\hat{\sigma}$ is biased for $\sigma = \sqrt{\sigma^2}$

Now:

$\varphi(\cdot) = \sqrt{\cdot}$ (square root is concave)

$\sqrt{E(x)} \geq E(\sqrt{x})$ Replace x with the estimator for σ^2, that is $\hat{\sigma}^2$

$\sqrt{E(\hat{\sigma}^2)} \geq E(\sqrt{\hat{\sigma}^2})$

$\sqrt{\sigma^2} \geq E(\hat{\sigma})$ $\Rightarrow$ $\sigma \geq E(\hat{\sigma})$

Because $E(\hat{\sigma}^2) = \sigma^2$ that is $\hat{\sigma}^2$ is unbiased for σ (Slide 81)

$\Rightarrow$ $\sigma - E(\hat{\sigma}) \geq 0$ So $\hat{\sigma}$ is biased for σ

$(\text{or } \sigma \neq E(\hat{\sigma}))$

Show that $\hat{\sigma}$ is consistent

$$\hat{\sigma} = \sqrt{\frac{1}{n-2} \sum_i \hat{u}_i^2} = \left(\frac{1}{n-2}\right)^{\frac{1}{2}} \left(\sum_i \hat{u}_i^2\right)^{\frac{1}{2}}$$

$$Var(\hat{\sigma}) = \frac{1}{n-2} Var\left(\sqrt{\sum_i \hat{u}_i^2}\right)$$

Heuristically as $n \to \infty$, $\frac{1}{n-2} Var(\sqrt{\sum_i \hat{u}_i^2}) \to 0$