Properties of OLS Estimators, Part I

Q: Suppose we are going to ① gather data $\{(x_i, y_i), i = 1 \ldots n\}$

② Compute $(\hat{\beta}_0, \hat{\beta}_1)$, OLS estimators

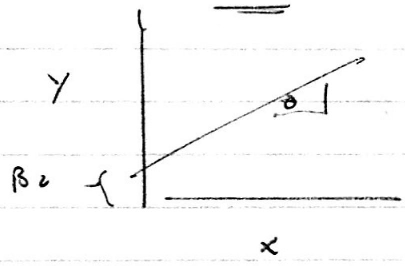
How close will $\{\hat{\beta}_0 + \hat{\beta}_1 x\}$ be to $E[Y \mid X = x]$?

Two Scenarios: Linear Model correct – $E[Y \mid X = x] = \beta_0 + \beta_1 x$ for some unknowns (β_0, β_1)

Linear Mode incorrect $E[Y \mid X = x]$ not linear in x.

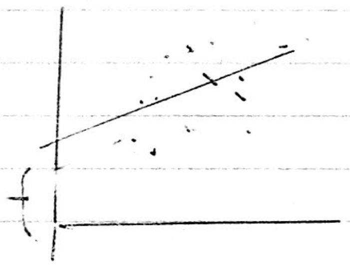
We first consider the case where the model is correct.

Truth

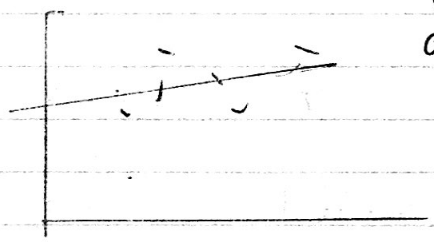


$\tan^{-1}\theta = \beta_1$

Dataset 1
OLS Line 1



Dataset 2
OLS Line 2



Etc

The values of $(\hat{\beta}_0, \hat{\beta}_1)$ depend on the outcome of your experiment or study.

If your experiment has random error or noise, or your study is a random sample, then $(\hat{\beta}_0, \hat{\beta}_1)$ are random too!

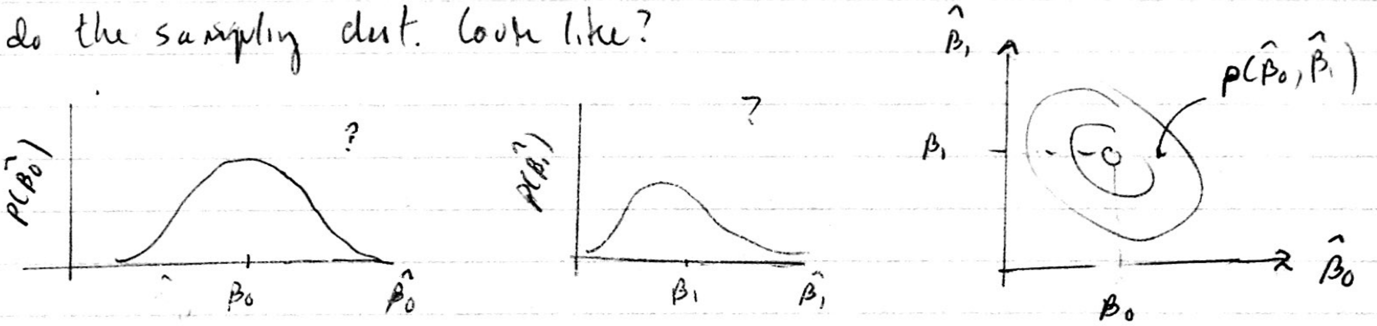
Truth	Experimental result
β_0, β_1	$(\hat\beta_0, \hat\beta_1)_1$
	$(\hat\beta_0, \hat\beta_1)_2$
	$\vdots$
	$(\hat\beta_0, \hat\beta_1)_K$

$\left.\begin{matrix} (\hat\beta_0, \hat\beta_1)_1 \\ \vdots \\ (\hat\beta_0, \hat\beta_1)_K \end{matrix}\right\}$ — different possible experimental results.

The variability of $(\hat\beta_0, \hat\beta_1)$ accross possible experimental outcomes is the sampling variability of $(\hat\beta_0, \hat\beta_1)$

The probability dist of $(\hat\beta_0, \hat\beta_1)$ accross possible outcomes is the sampling dist. of $(\hat\beta_0, \hat\beta_1)$.

What do the sampling dist. look like?



Result #1: If $E[Y \mid X = x] = \beta_0 + \beta_1 x$, then

$$E[\hat\beta_0] = \beta_0$$
$$E[\hat\beta_1] = \beta_1$$

, i.e. $(\hat\beta_0, \hat\beta_1)$ are unbiased estimators of (β_0, β_1).

Proof: Recall $\hat\beta_1 = \dfrac{SXY}{SXX} = \dfrac{\Sigma(x_i - \bar x)(y_i - \bar y)}{\Sigma(x_i - \bar x)^2}$

$$E[\hat\beta_1 \mid \underset{\sim}{x}] = \frac{\Sigma(x_i - \bar x)\, E[(Y_i - \bar Y) \mid \underset{\sim}{x}]}{\Sigma(x_i - \bar x)^2}$$

$$E[\,Y_i - \bar{Y} \mid \underset{\sim}{x}\,] = \;?$$

$$= E[Y_i \mid \underset{\sim}{x}] - E[\bar{Y} \mid \underset{\sim}{x}]$$

$$= (\beta_0 + \beta_1 x_i) - E[\,\tfrac{1}{n}\textstyle\sum Y_i \mid \underset{\sim}{x}\,]$$

$$= (\beta_0 + \beta_1 x_i) - \tfrac{1}{n}\textstyle\sum E[Y_i \mid \underset{\sim}{x}]$$

$$= (\beta_0 + \beta_1 x_i) - \tfrac{1}{n}\textstyle\sum \beta_0 + \beta_1 x_i$$

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

So $$E[\hat{\beta}_1 \mid \underset{\sim}{x}] = \frac{\sum (x_i - \bar{x})\, \beta_1 (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \beta_1 \left(\frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) = \underline{\underline{\beta_1}}$$

$\underline{\beta_0}$: $$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

$$E[\hat{\beta}_0 \mid \underset{\sim}{x}] = \beta_0 + \beta_1 \bar{x} - E[\hat{\beta}_1 \bar{x}]$$

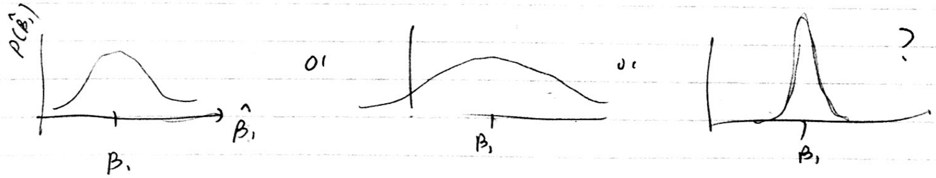
$$= \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0$$

$\underline{\underline{\text{So}}}$: If $E[Y \mid X = x] = \beta_0 + \beta_1 x$ for some β_0, β_1,

then $E[\hat{\beta}_0, \hat{\beta}_1] = (\beta_0, \beta_1)$.

$=$

This says the sampling distributions of $(\hat{\beta}_0, \hat{\beta}_1)$ are centered around the correct values, but it doesn't say how <u>concentrated</u> these distributions are.



<u>Concentration</u> of $(\hat{\beta}_0, \hat{\beta}_1)$ around (β_0, β_1) can be measured with variance and covariance.

Result #2 if ① $E[Y_i | X = x_i] = \beta_0 + \beta_1 x_i$ (linear model)

② $Var(Y_i | X = x_i) = \sigma^2$ (constant variance)

③ $Cov(Y_i, Y_{i'} | X_i, X_{i'}) = 0$ (observations are uncorrelated)

then $E[(\hat\beta_0, \hat\beta_1) | \underset{\sim}{x}] = (\beta_0, \beta_1)$

and

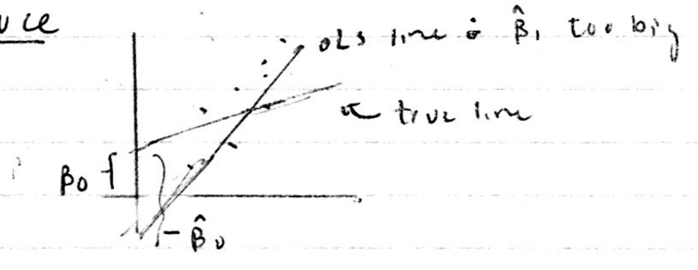
$$
\left.\begin{aligned}
Var(\hat\beta_1 | \underset{\sim}{x}) &= \sigma^2 \times \frac{1}{Sxx} \\
Var(\hat\beta_0 | \underset{\sim}{x}) &= \sigma^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{Sxx}\right) \\
Cov(\hat\beta_0, \hat\beta_1 | \underset{\sim}{x}) &= -\sigma^2 \left(\frac{\bar{x}}{Sxx}\right)
\end{aligned}\right\}
$$

will prove when we get to multiple linear regression.

why is this negative? what does it mean?

Interpretation: if $\hat\beta_1$ is higher than β_1, we expect $\hat\beta_0$ to be lower than β_0.

Picture



Math: $Cov(\hat\beta_0, \hat\beta_1 | \underset{\sim}{x}] = Cov(\bar{y} - \hat\beta_1 \bar{x}, \hat\beta_1 | \underset{\sim}{x})$

$= Cov(\bar{y}, \hat\beta_1 | \underset{\sim}{x}) - Cov(\hat\beta_1 \bar{x}, \hat\beta_1 | \underset{\sim}{x})$

(why 0?)

$= 0 - \bar{x}\, Cov(\hat\beta_1, \hat\beta_1 | \underset{\sim}{x})$

$= - \bar{x}\, Var(\hat\beta_1 | \underset{\sim}{x})$

$= \dfrac{-\bar{x}\,\sigma^2}{Sxx}$