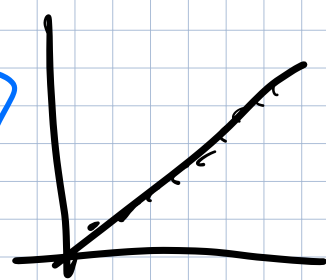
Regression

Fall 2024

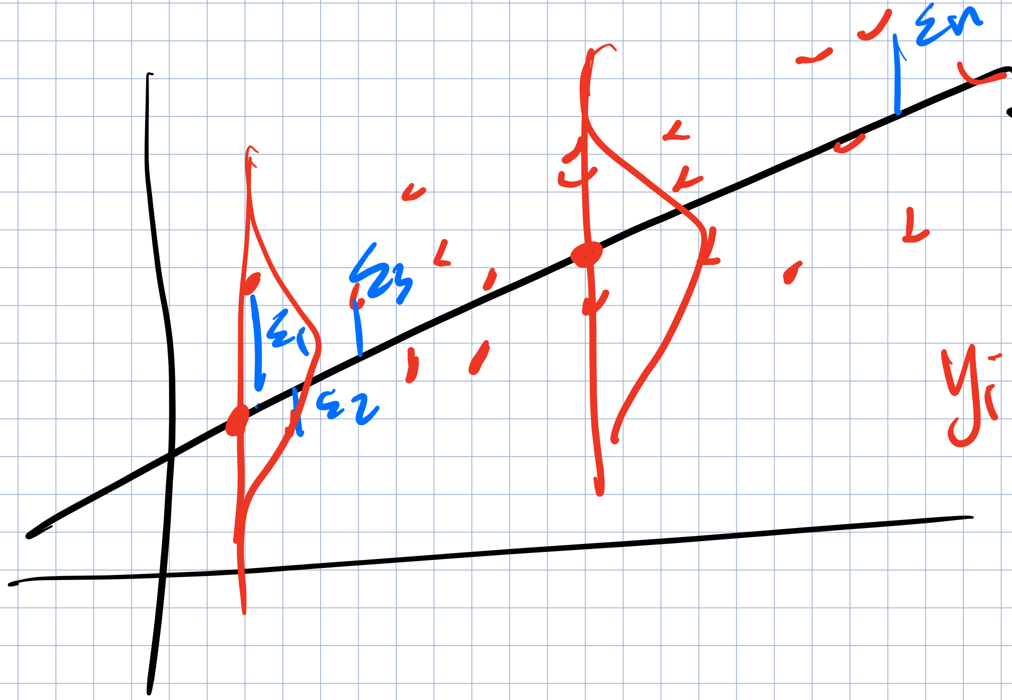
I. Simple Linear Regression Model

$$\boxed{y_i = \beta_0 + \beta_1 x_i + \varepsilon_i}$$

$i = 1, \ldots, n$



- y_i = i^{th} observed value of the (random) response variable Y
- x_i = i^{th} observed value of the (fixed) predictor X
- n = sample size (fixed)
- β_0 = unknown intercept parameter (fixed)
- β_1 = unknown slope parameter (fixed)
- ε_i = random error term for the i^{th} obs. (random)



$Y_i = \beta_0 + \beta_1 X_i$

$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

↑ "① Classic ② Simple ③ Linear Regression model"

① classic: model assumptions are "standard"

② simple: only one predictor x

③ linear: y_i is linear in the parameters

Ex:

(I) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (lin)

(II) $y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i$ (lin)

(III) $y_i = \beta_0 + \beta_1 \log x_i + \varepsilon_i$ (lin)

(X) $y_i = \beta_0 + \dfrac{\beta_1 x_i}{\beta_2 + x_i} + \varepsilon_i$ (not linear)

$$\boxed{y_i = \beta_0 + \beta_1 z_i + \varepsilon} \qquad z_i = x_i^2$$

② Model Assumptions for SLR

- ε_i's are random error terms that satisfy:

BASIC

(i) $E(\varepsilon_i) = 0$

(ii) $Var(\varepsilon_i) = \sigma^2$

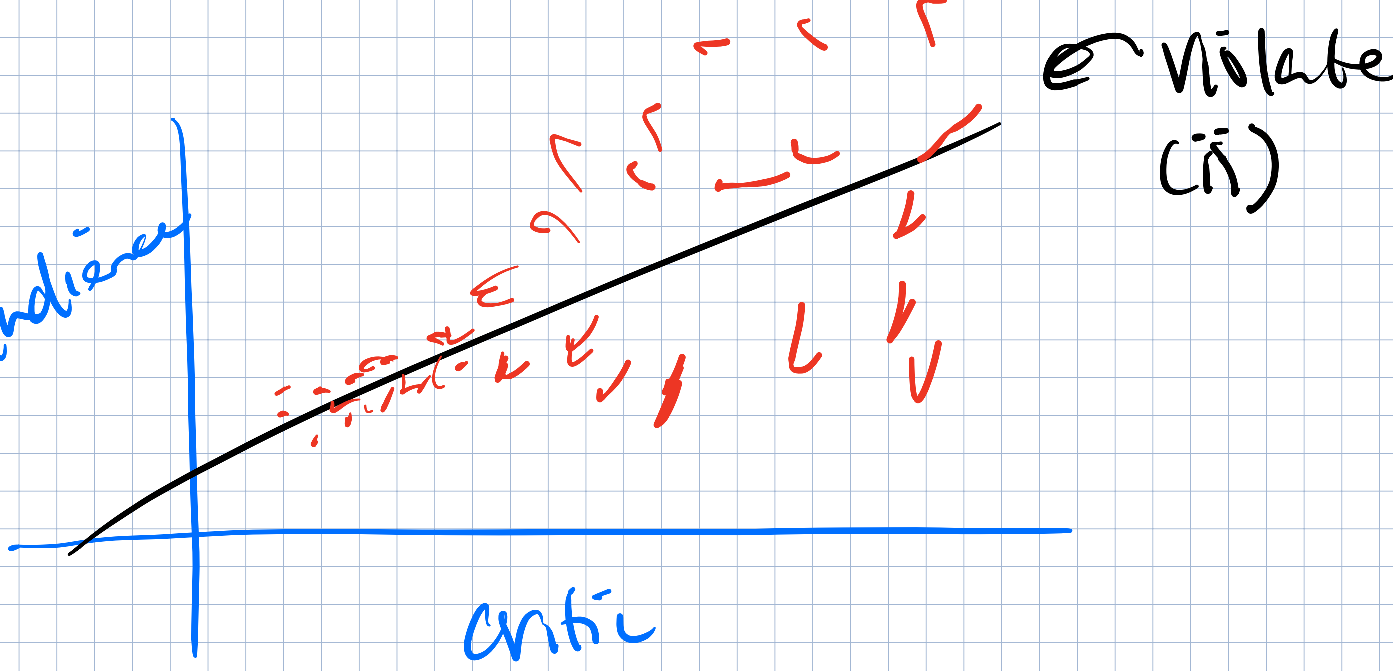
(iii) $Cov(\varepsilon_i, \varepsilon_j) = 0$ – no correlations exist b/w errors

CLASSICAL

(iv) $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

iid – independent & identically distributed

this assumption is needed for inference but not estimation



$Var(\varepsilon_i) = \sigma^2(x_i)$

- x_i's are fixed

i. x_is are treated as indiv. constants

ii. we don't have to worry the dist. of x_i

(If we allow random X's then we need "random effects" model)

3. Regression Function

The regression function is

$$g(x) = E(Y \mid X = x) = \beta_0 + \beta_1 X$$

population qty

The goal of SLR is to estimate

$$g(x) = \beta_0 + \beta_1 X$$

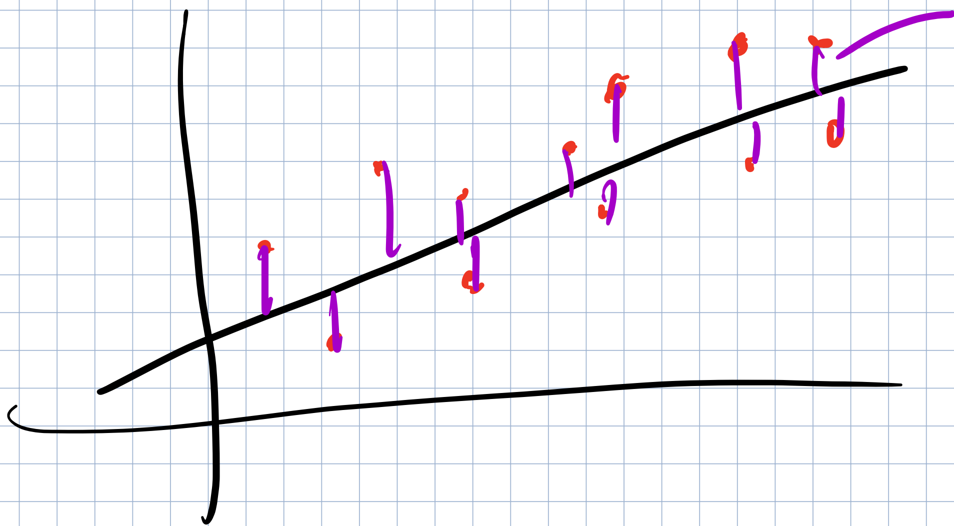
$\downarrow$

① Need to Estimate β_0

② Need to Estimate β_1

Guiding Principle for Estimation:

Least Squares Principle



minimize the errors, squared.

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_i \varepsilon_i^2$$

recall:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$y_i - \beta_0 - \beta_1 x_i = \varepsilon_i$$

$$= \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$$

Task:

This quantity is called the "sum of squared errors"

Minimize $Q(\beta_0, \beta_1) = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$

with respect to β_0 & β_1

$$\frac{\partial Q}{\partial \beta_0} = 2 \sum_i (y_i - \beta_0 - \beta_1 x_i)(-1) \overset{\text{set}}{=} 0$$

$$\Rightarrow \sum_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\Rightarrow \sum_i y_i - \underline{\underline{\sum_i \beta_0}} - \beta_1 \sum_i x_i = 0$$

$$\Rightarrow \sum_i y_i - \beta_1 \sum x_i = n\beta_0$$

$$\Rightarrow \bar{y} - \beta_1 \bar{x} = \underline{\underline{\beta_0}}$$ (1st cond)

$$\frac{\partial Q}{\partial \beta_1} = 2\sum_i (y_i - \beta_0 - \beta_1 x_i)(-x_i)$$

$$= -2\sum_i (x_i y_i - \beta_0 x_i - \beta_1 x_i^2)$$

$$\stackrel{!}{=} 0$$

$$\Rightarrow \sum_i x_i y_i - \beta_0 \sum_i x_i - \beta_1 \sum_i x_i^2 = 0$$

$$\beta_1 \sum_i x_i^2 = \sum_i x_i y_i - \beta_0 \sum_i x_i$$

$$\Rightarrow \beta_1 \sum_i x_i^2 = \sum_i x_i y_i - (\bar{y} - \beta_1 \bar{x})\, n\bar{x}$$

$$\Rightarrow \beta_1 \sum_i x_i^2 = \sum_i x_i y_i - n\bar{x}\bar{y} + \beta_1 n\bar{x}^2$$

$$\Rightarrow \beta_1 \left(\sum_i x_i^2 - n\bar{x}^2\right) = \sum_i x_i y_i - n\bar{x}\bar{y}$$

$$\Rightarrow \beta_1 = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} \quad \left(\begin{smallmatrix}\text{2nd}\\ \text{cond}\end{smallmatrix}\right)$$

I just showed (almost) that the minimizers

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}}\, Q(\beta_0, \beta_1)$$

must satisfy:

① $\bar{y} - \hat{\beta}_1 \bar{x} = \hat{\beta}_0$

② $\hat{\beta}_1 = \dfrac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2}$

⋮

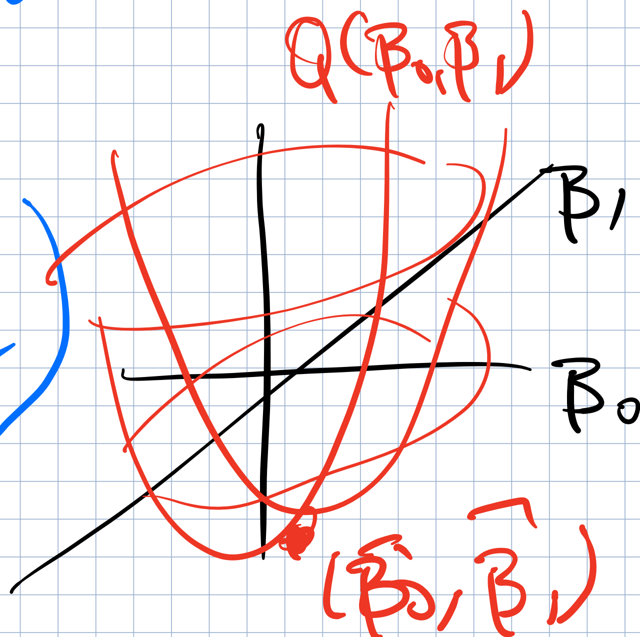
$$= \frac{\text{sample Cov of } (x_i, y_i)}{\text{Sample variance of } (x_i)}$$

$$= \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X, X)}$$

$$= \frac{\cancel{\frac{1}{n-1}} \sum_i (x_i - \bar{x})(y_i - \bar{y})}{\cancel{\frac{1}{n-1}} \sum_i (x_i - \bar{x})^2}$$

(Technically to rigorously argue that $\hat{\beta}_0$ & $\hat{\beta}_1$ actually are minimizes we should show that

$$H = \begin{pmatrix} \frac{\partial^2 Q}{\partial \beta_0^2} & \frac{\partial Q^2}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 Q}{\partial \beta_0 \partial \beta_1} & \frac{\partial Q^2}{\partial \beta_1^2} \end{pmatrix}$$ is pos. def.)



Turns out that:

$$H = \begin{pmatrix} 2n & 2n\bar{x} \\ 2n\bar{x} & 2\sum_i x_i^2 \end{pmatrix}$$

for any $a \neq 0$,

$$a^T H a > 0$$