Linear Regression

Fall 2025

Simple Linear Regression Model

↳ "one predictor & one response"

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, \ldots, n$$

Notation:

- y_i = i^{th} observed value of the response variable Y $\boxed{\text{RANDOM}}$

- x_i = i^{th} observed value of the predictor variable X $\boxed{\text{FIXED}}$

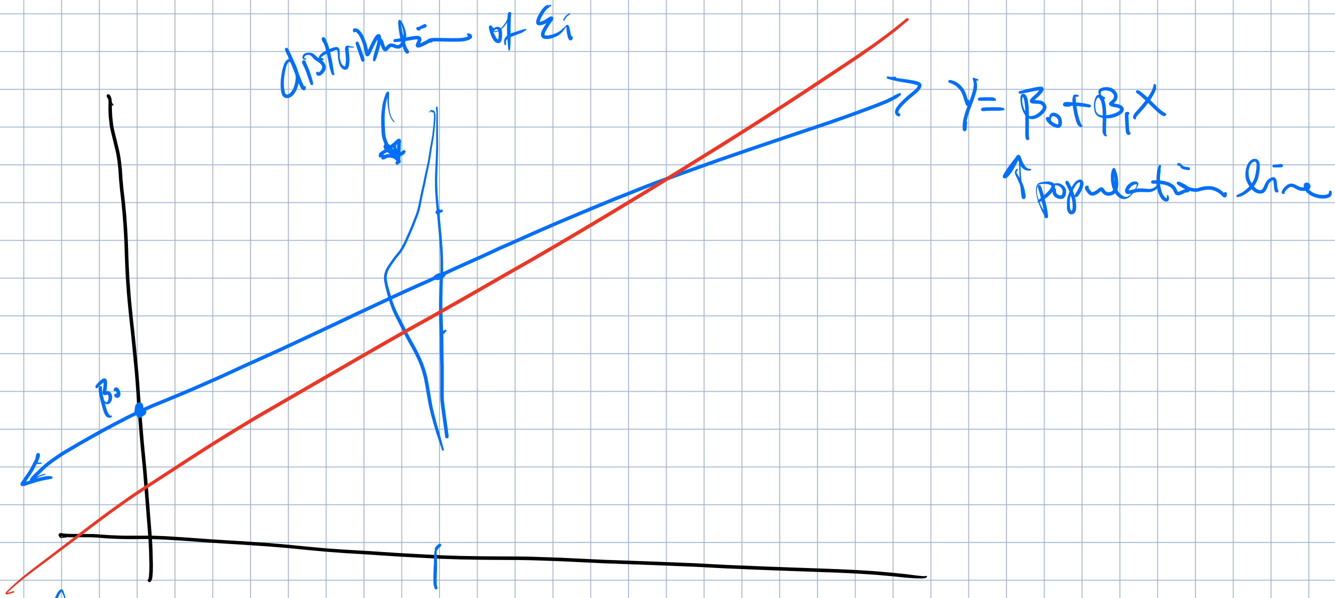
$$\{(x_i, y_i)\}_{i=1,\ldots,n}$$

- n: sample size $\boxed{\text{FIXED}}$

- β_0: unknown intercept parameter $\boxed{\text{FIXED}}$

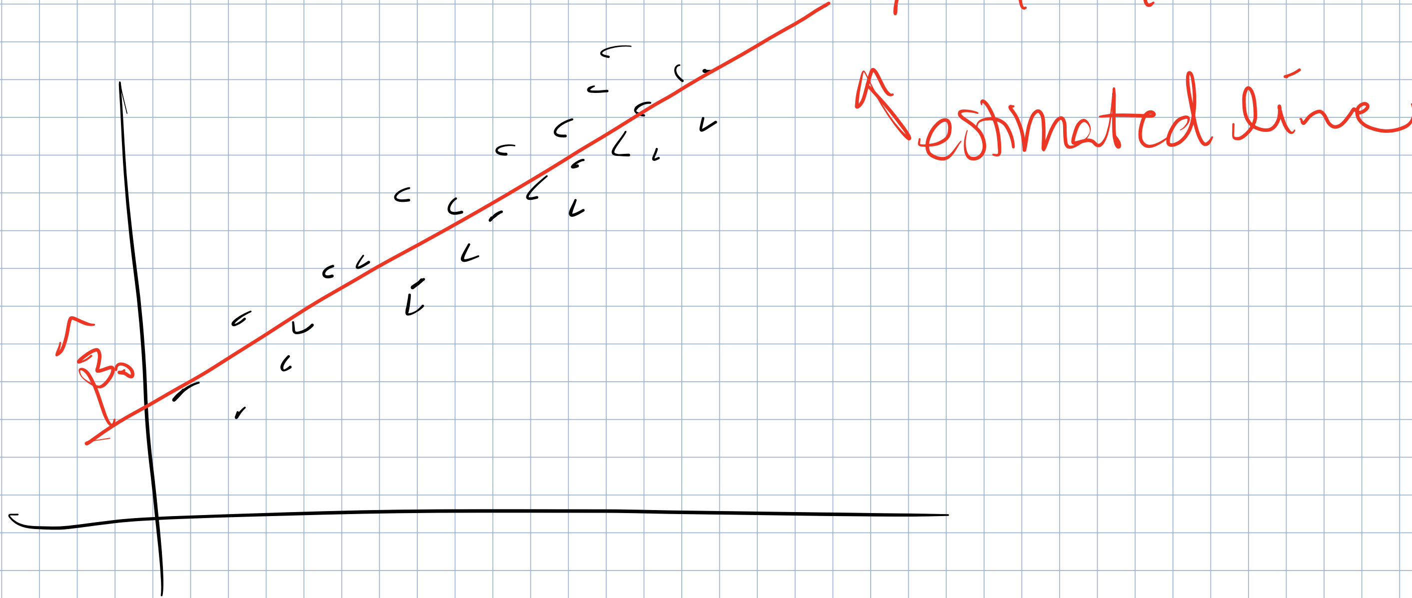
- β_1: unknown slope parameter $\boxed{\text{FIXED}}$

- ε_i: random error term $\boxed{\text{RANDOM}}$



$$\{(x_i, y_i)\}_{i=1,\dots,n}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

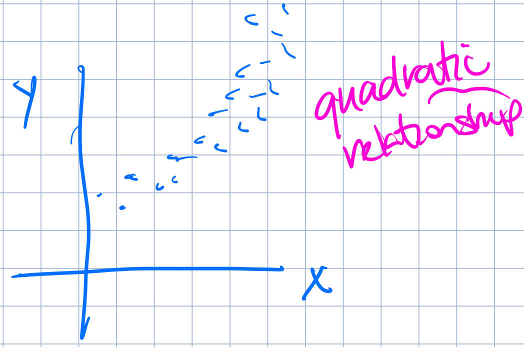


Observations:

① simple: only 1 predictor var

② linear: y_i is a linear fn of its parameters β_0 & β_1

Ex:



(A) $Y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i$

Let's let $z_i = x_i^2$

$\Rightarrow Y_i = \beta_0 + \beta_1 z_i + \varepsilon_i$ ✓ linear

(B) $Y_i = \beta_0 + \beta_1 \log x_i + \varepsilon_i$ ✓ linear

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

Model Assumptions

- ε_i's are RANDOM variables that satisfy

BASIC
(i) $E(\varepsilon_i) = 0$
(ii) $Var(\varepsilon_i) = \sigma^2 \leftarrow$ constant w/ respect to X
(iii) $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$
↳ uncorrelated errors

CLASSICAL $\left[\text{(iv)} \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2) \right.$

Useful for inference

- x_i's are fixed

(If we take x_i's as random, I need a random effects model.)

Regression Function

The regression function is

$$g(x) = E(Y \mid X = x)$$

linear

$$= \beta_0 + \beta_1 X$$

Idea: Always estimating g as our target...

Under some assumptions the form of g changes.

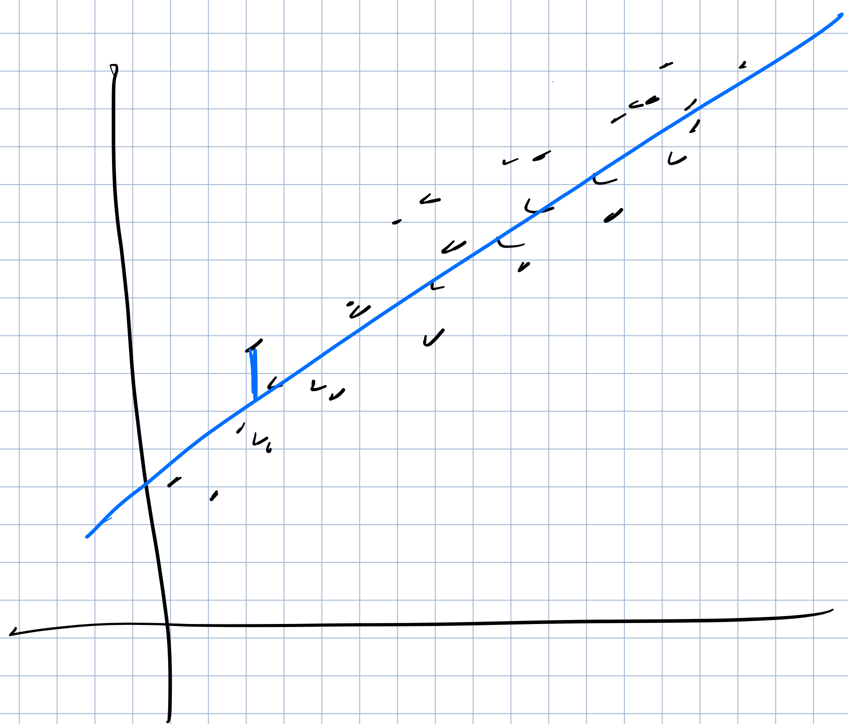
Here g boils down to two parameters: β_0 & β_1.

This is equivalent (in SLR case) to estimating

① β_0 & ② β_1.

The Least Squares Principle

is how we try to solve that problem.

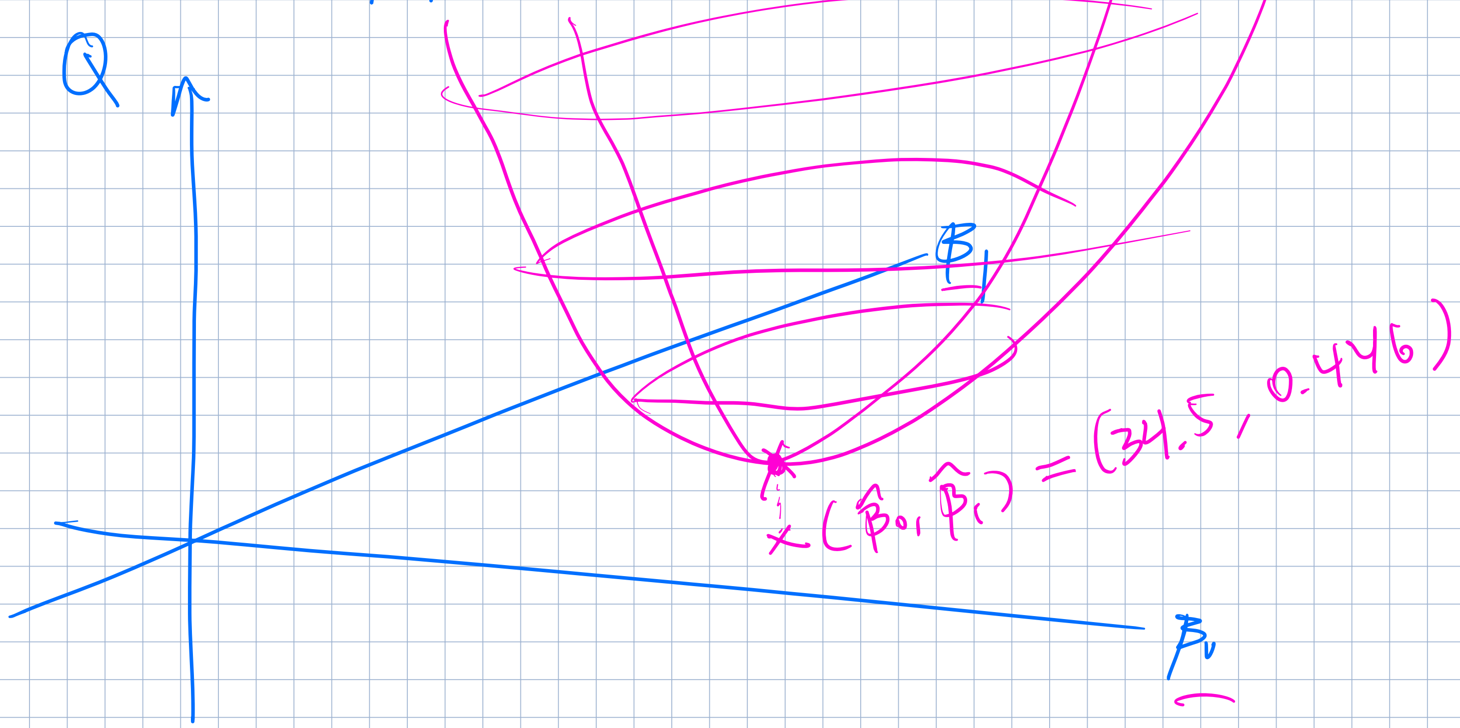


I want to find the line that minimizes total squared error:

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

$$= \sum_{i=1}^{n} \varepsilon_i^2$$

$$= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$



In other words,

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

So the estimates $(\hat{\beta}_0, \hat{\beta}_1)$ are the choices of β_0 & β_1 values that satisfy:

① $$\frac{\partial Q}{\partial \beta_0} \overset{!}{=} 0$$

② $$\frac{\partial Q}{\partial \beta_1} \overset{!}{=} 0$$

① $$\frac{\partial Q}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_0} (\underbrace{y_i - \beta_0 - \beta_1 x_i})^2$$

$$= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(0 - 1 - 0)$$

$$= \boxed{-2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \overset{!}{=} 0}^{A} \quad //$$

② $$\frac{\partial Q}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(0 - 0 - x_i)$$

$$\boxed{= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i \overset{!}{=} 0}^{B}$$

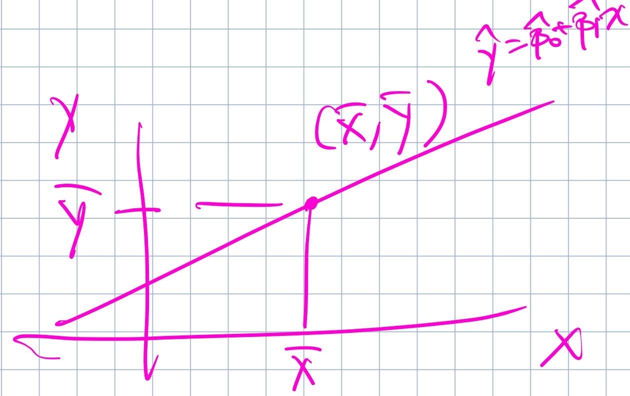
A
$$-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)=0$$

$$\bar y=\frac{\sum_i y_i}{n}\iff \sum_i y_i=n\bar y$$

$$\Rightarrow \sum_{i=1}^{n} y_i-\sum_{i=1}^{n}\beta_0-\sum_{i=1}^{n}\beta_1 x_i=0$$

$$n\bar y-n\beta_0-\beta_1(n\bar x)=0$$

$$n\bar y=n\beta_0+n\beta_1\bar x$$



Solutions $(\hat\beta_0,\hat\beta_1)$ must satisfy condition 1

$$\boxed{\bar y=\hat\beta_0+\hat\beta_1\bar x}$$

(B)
$$-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)x_i=0$$

plug into here

$$\Rightarrow \sum_{i=1}^{n} x_i y_i-\beta_0\sum_{i=1}^{n}x_i-\beta_1\sum_{i=1}^{n}x_i^2=0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i-(\bar y-\beta_1\bar x)\underbrace{\sum_{i=1}^{n}x_i}_{=n\bar x}-\beta_1\sum_i x_i^2=0$$

Solve for β_1:

$$\sum_{i=1}^{n} x_i y_i-(n\bar x\bar y-\beta_1 n\bar x^2)-\beta_1\sum_{i=1}^{n}x_i^2=0$$

$$\sum_{i=1}^{n} x_i y_i-n\bar x\bar y+\beta_1 n\bar x^2-\beta_1\sum_{i=1}^{n}x_i^2=0$$

$$\Rightarrow \quad \sum_i x_i y_i - n\bar{x}\bar{y} = \beta_1 \sum_{i=1}^{n} x_i^2 - \beta_1 n\bar{x}^2$$

$$= \beta_1 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right)$$

Solution:

$$\Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\vdots$$

HW

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\widehat{\text{cov}}(x, y)}{\widehat{\text{Var}}(x)}$$