The Classical Simple Linear Regression Model

Data: $\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$

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

Simple Linear Regression Model:

$$y_i = m(x_i) + \varepsilon_i \quad \&$$

$$\mu(x) = \beta_0 + \beta_1 x$$

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

Assumptions:

1. The relationship b/w x & y is actually linear.

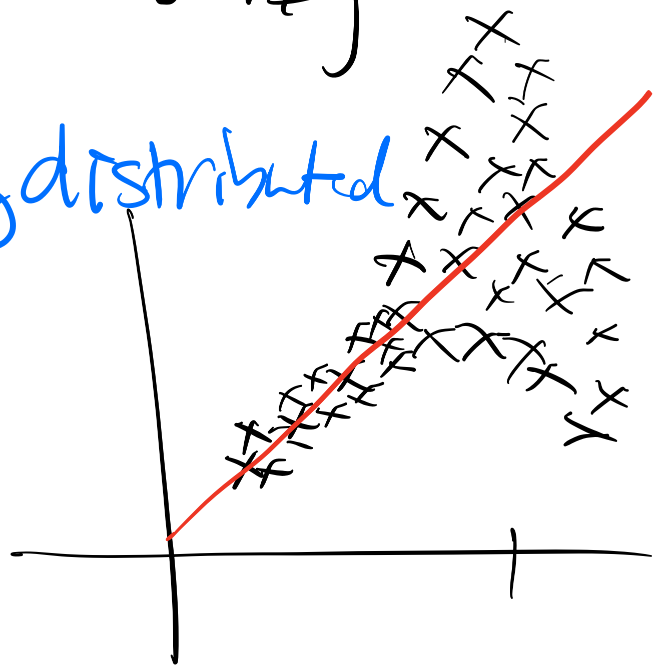
2. i. $E(\varepsilon_i) = 0$

ii. $Var(\varepsilon_i) = \sigma^2 \perp\!\!\!\perp x$

iii. $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall\, i \neq j$

(classical)

iv. ε_i are Normally distributed



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

$\varepsilon_i \overset{ind}{\sim} ?(0, \sigma^2)$

③ The predictor x_i has a fixed effect on Y_i.

i. The x_i's are Treated as constants

ii. We don't need to consider the

distribution of the x_i's.

iii. If you want to allow x_i's to have a random effect on y_i's $\Rightarrow$ you need a mixed-effects model.

Exercises:

1. Calculate the expected value of y_i for a fixed i.

2. Calculate the variance of y_i for a fixed i.

3. Calculate the marginal dist. of y_i for a fixed i.

4. Calculate the joint dist. $y_1, \cdots, y_n$.

① $E(y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i)$

$= E(\underbrace{\beta_0 + \beta_1 x_i}_{\text{const.}}) + E(\varepsilon_i)$

$= \beta_0 + \beta_1 x_i + 0$

$= \underbrace{\beta_0 + \beta_1 x_i}$

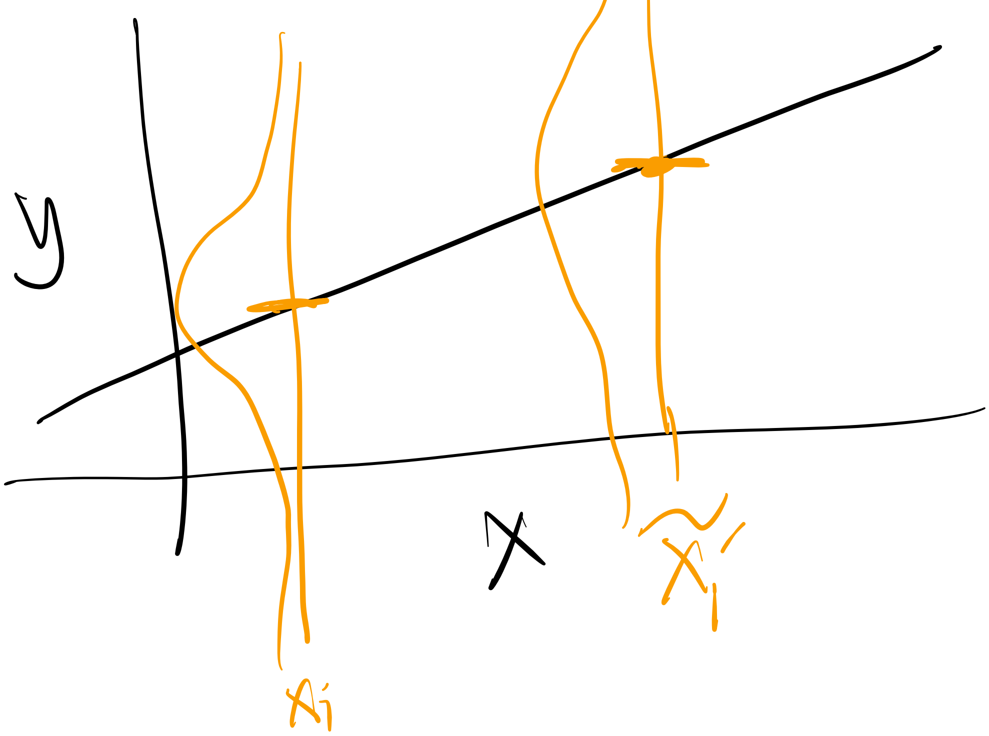
② $Var(y_i) = Var(\underbrace{\beta_0 + \beta_1 x_i} + \varepsilon_i)$

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

③ $y_i = \beta_0 + \beta_1 x_i + \underline{\varepsilon_i}$

$\varepsilon_i \sim N(0, \sigma^2)$

$\underbrace{\beta_0 + \beta_1 x_i}_{y_i} + \varepsilon_i \sim \boxed{N(\beta_0 + \beta_1 x_i, \; \sigma^2)}$



4. $y_1, \ldots, y_n$ mutually $\perp\!\!\!\perp$ indep.

$$f_{Y_1,\dots,Y_n}(y_1,y_2,\dots,y_n) = \prod_{i=1}^{n} f_{Y_i}(y_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\frac{(y_1-\beta_0-\beta_1 x_1)^2}{2\sigma^2}} \times e^{-(y_2-\beta_0-\beta_1 x_2)^2/2\sigma^2} \times \dots \times$$

$$= \underbrace{(2\pi\sigma^2)^{-n/2} e^{-\frac{\boxed{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2}}{2\sigma^2}}}$$

likelihood $= L(\beta_0, \beta_1)$

maximizing the likelihood $\Leftrightarrow$

minimizing sum of squared residuals

Estimating β_0 & β_1

"criteria"
loss function
objective function

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

We want β_0, β_1 which minimize $Q(\beta_0, \beta_1)$.

$$\left.\begin{array}{ll} ① & \partial Q / \partial \beta_0 \overset{!}{=} 0 \\ ② & \partial Q / \partial \beta_1 \overset{!}{=} 0 \end{array}\right] \text{system of eqs} \Rightarrow \hat{\beta}_0, \hat{\beta}_1$$

$$Q(a,b) = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

$$① \frac{\partial Q}{\partial a} = \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

$$= \sum_{i=1}^{n} 2(y_i - a - bx_i)(-1)$$

$$= -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\Rightarrow \sum_i y_i - na - b \sum_{i=1}^{n} x_i = 0$$

$$\frac{\sum_i y_i}{n} = \frac{na}{n} + \frac{b \sum_{i=1}^{n} x_i}{n} \quad ⊛$$

$$\bar{y} = a + b\bar{x} \Rightarrow a = \bar{y} - b\bar{x}$$

② $\frac{\partial Q}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$

$= 2\sum_{i=1}^{n} (y_i - a - bx_i)(-x_i) \overset{!}{=} 0$

$\Rightarrow -\cancel{2} \sum_{i=1}^{n} (x_i y_i - a x_i - b x_i^2) \overset{!}{=} 0$

$\Rightarrow \sum_{i=1}^{n} x_i y_i - a\sum_i x_i - b \sum_{i=1}^{n} x_i^2 = 0$

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

$\sum_{i=1}^{n} x_i y_i - (n\bar{x}\bar{y} - bn\bar{x}^2) = b \sum_i x_i^2$

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

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

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

$$a = \hat{\beta}_0 = \bar{y} - b\bar{x}$$

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

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