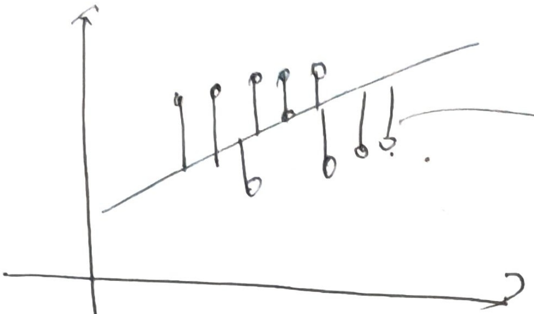
독립변수 1개) $y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i=1,\dots,n$



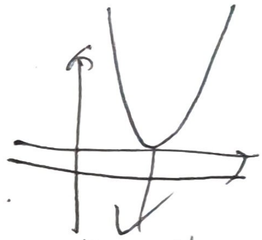
$\epsilon_i = (y_i - \beta_0 - \beta_1 x_i).$

Least square Estimation.

$$S(\beta_0,\beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

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

β_0, β_1에 대한 Convex ftn.



미분시 0이 되는곳이 최소.

① $$\frac{\partial S(\beta_0,\beta_1)}{\partial \beta_0} = 2 \cdot \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)(-1) = 0.$$

② $$\frac{\partial S(\beta_0,\beta_1)}{\partial \beta_1} = 2 \cdot \sum_{i=1}^{N} \left(y_i - (\beta_0 + \beta_1 x_i)\right)(-x_i) = 0.$$

최소제곱법은 오차의 총량을 줄일수 있는 회귀 계수 추출은 회귀계수.

$$\sum_{i=1}^{n} y_i - \beta_0 \cdot n \cdot \beta_1 - \sum_{i=1}^{n} x_i = 0. \quad ① \times \frac{1}{n} = ①'$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \frac{\sum x_i^2}{\quad} = 0. \quad ② \times \frac{1}{\sum x_i} = ②'$$

②

$\cancel{③ = -②' + ①' \Rightarrow \beta_1 \left(\frac{\sum x_i^2}{\sum x_i} - \right.}$

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

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

$\frac{\sum y_i}{n} - \beta_0 - \beta_1 \frac{\sum x_i}{n} = 0$ … $① \times \frac{1}{n} \Rightarrow ①'$

$\frac{\sum x_i y_i}{\sum x_i} - \beta_0 - \beta_1 \frac{\sum x_i^2}{\sum x_i} = 0$. $② \times \frac{1}{\sum x_i} \Rightarrow ②'$

$③ = -②' + ①' \Rightarrow \beta_1 \left(\frac{\sum x_i^2}{\sum x_i} - \frac{\sum x_i}{n} \right) - \frac{\sum x_i y_i}{\sum x_i} + \frac{\sum y_i}{n} = 0$

$③ \times \sum x_i \Rightarrow \beta_1 \left(\underbrace{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \right) = \underbrace{\sum x_i y_i - \frac{\sum x_i y_i}{n}}$

$= \sum (x_i - \bar{x})^2$ $\quad = \sum \left(x - \frac{\sum x_i}{n} \right)$

① $\sum x_i^2 - 2\bar{x} \sum x_i + \sum (\bar{x})^2$ $\quad = \sum (x - \bar{x}) y_i$

$= \sum x_i^2 - 2 \frac{(\sum x_i)^2}{n} + \frac{(\sum x_i)^2}{n}$

$= \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

③

$$\Rightarrow \hat{\beta}_1 = \frac{\Sigma (x_i - \bar{x}) \cdot y_i}{\Sigma (x - \bar{x}_i)^2}$$

① 정리 $\Rightarrow \hat{\beta}_1 = \bar{y} - \hat{\beta}_1 \bar{x}$.

회귀 계수의 추정량을 구할 수 있음

이 추정량은 통계적으로 좋은 성질을 가진 추정량이 된다.

(BLUE, Best Linear Unbiased Estimator

선형이고 불편성과 최소 분산을 만족시키는 가장 효율적인 추정량)

↳ 추정량의 기댓값이 모수의 값과 일치함.

독립변수 2개 이상) $Y = X\beta + \epsilon$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & x_{21} & \cdots & x_{2p} \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}_{n\times(p+1)} \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$S(\beta) = \epsilon'\epsilon = (Y - X\beta)'(Y - X\beta)$$

$$= (Y' - \beta' X')(Y - X\beta)$$

$(1\times\cancel{(p+1)})\ (\cancel{(p+1)}\times\cancel{n})\ (\cancel{n}\times 1)$

(1×1)

$$= Y'Y - Y'X\beta - \beta' X' Y + \beta' X' X\beta$$

$(1\times\cancel{n})(\cancel{n}\times\cancel{(p+1)})(\cancel{(p+1)}\times 1)$

(1×1)

$(p+1)\times 1 \ (p$

$\Rightarrow Y'Y - 2\beta' X' Y + \beta' X' X\beta$

$Y'Y - 2\beta' X' Y + \beta^2 X' X$

$$\frac{S(\beta)}{\partial \beta} = 0 - 2X'Y + 2X'X\beta = 0$$

$(p+1)\times n \quad n\times(p+1)$

$$\Rightarrow X'X\beta = X'Y$$

$1\times\cancel{(p+1)}(\cancel{(p+1)}\times\cancel{(p+1)})$

$$\therefore \hat{\beta} = (X'X)^{-1}X'Y$$

$1\times\cancel{(p+1)}$