①

Linear Regression

$x = (x_1, x_2, \dots, x_n)$ : ind variable

$y = (y_1, y_2, \dots, y_n)$ : Dep variable.

model: $y_i = \alpha + \beta x_i$

$E[y_i | x_i] = \alpha + \beta x_i$

$\text{Variance}(y_i) = \sigma^2$

$y_i = \alpha + \beta x_i + \epsilon_i$

random error

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

$$(x_1, x_2, \dots, x_n)$$

$$(y_1, y_2, \dots, y_n)$$

To Compute

$$f(y \mid \alpha, \beta, \sigma^2, x)$$

$$= \prod_{i=1}^{n} f(y_i \mid \alpha + \beta x_i, \sigma^2)$$

Assumption:

$y_i$'s are independent Gaussian random variables with mean $\alpha + \beta x_i$ and variance $\sigma^2$

Gaussian distribution.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\prod_{i=1}^{n} f(y_i \mid \alpha + \beta x_i, \sigma^2)$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp -\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}$$

$$= f(y \mid x, \alpha, \beta, \sigma^2)$$

$$\log f = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}$$

$$\log f = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{1}^{n}\frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2}$$

for maximal likelihood estimate

$$\frac{\partial f}{\partial \alpha} = \sum_{1}^{n}\frac{(y_i - \alpha - \beta x_i)}{\sigma^2}$$

$$\frac{\partial f}{\partial \beta} = \sum_{1}^{n}\frac{(y_i - \alpha - \beta x_i)x_i}{\sigma^2}$$

$$\frac{\partial f}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{1}^{n}\frac{(y_i - \alpha - \beta x_i)^2}{(\sigma^2)^2}$$

$$\frac{\partial f}{\partial \alpha} = \sum_{1}^{n} \frac{y_i - \alpha - \beta x_i}{\sigma^2} = 0$$

i.e. $$\sum y_i = \alpha n + \beta \sum_{1}^{n} x_i$$

$$\frac{\partial f}{\partial \beta} = \sum_{1}^{n} \frac{(y_i - \alpha - \beta x_i) x_i}{\sigma^2} = 0$$

i.e. $$\alpha \sum_{1}^{n} x_i + \beta \sum_{1}^{n} x^2 = \sum_{1}^{n} x_i y_i$$

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sum_{1}^{n} y_i \\ \sum_{1}^{n} x_i y_i \end{bmatrix}$$

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} n & \sum_1^n x_i \\ \sum_1^n x_i & \sum_1^n x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_1^n y_i \\ \sum_1^n x_i y_i \end{bmatrix}$$

$$\frac{1}{n\sum_1^n x_i^2 - \left(\sum_1^n x_i\right)^2} \begin{bmatrix} \sum x_i^2 & -\sum_1^n x_i \\ -\sum_1^n x_i & n \end{bmatrix}$$

$$\beta = \frac{n\sum_1^n x_i y - \sum_1^n x_i \sum_1^n y_i}{n\sum_1^n x_i^2 - \left(\sum_1^n x_i\right)^2}$$

$$= \frac{\sum_1^n x_i y_i - n\bar{x}\bar{y}}{\sum_1^n x_i^2 - n\bar{x}^2}$$

(7)

$$\beta = \frac{\sum_1^n x_i y_i - \frac{\left(\sum_1^n x_i\right)\left(\sum_1^n y_i\right)}{n}}{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}$$

$$= \frac{\sum_1^n x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_1^n (x_i - \bar{x})^2}$$

Ncem: $\sum (x_i - \bar{x})(y_i - \bar{y})$

$$= \sum_1^n x_i y_i - \sum x_i \bar{y} - \sum \bar{x} y_i + \sum_1^n \bar{x}\bar{y}$$

$$= \sum_1^n x_i y_i - n\bar{x}^2$$

Decom: $\sum x_i^2 + n\bar{x}^2 - 2\bar{x}\sum_1^n x_i$

$$= \sum x_i^2 + n\bar{x}^2 - 2n\bar{x}^2$$

$\hat{\alpha}$, $\hat{\beta}$ : max likelihood estimates of $\alpha$ & $\beta$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$$

Estimates of $\hat{y}_i$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}\,x_i$$

$$= \bar{y} - \hat{\beta}\,\bar{x} + \hat{\beta}\,x_i$$

$$= \bar{y} + \hat{\beta}\,(x_i - \bar{x}) \quad , \; i = 1 \cdots n$$

Maximum likelihood estimati of $\sigma^2$

$$\frac{\partial \log f}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{1}^{n}\frac{(y_i - \alpha - \beta x_i)^2}{(\sigma^2)^2} = 0$$

$$\frac{n}{\cancel{2}\,\sigma^{\cancel{2}}} = \cancel{\frac{1}{2}} \times \frac{1}{(\sigma^2)^{\cancel{2}}}\sum_{1}^{n}(y_i - \alpha - \beta x_i)^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{1}^{n}(y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

## Example

| $x_i$ | $y_i$ |
|---|---|
| 2 | 10 |
| 4 | 15 |
| 8 | 10 |
| 10 | 14 |
| 12 | 8 |
| 15 | 10 |
| 19 | 17 |

$\sum x_i = 70$ $\quad$ $\sum y_i = 84$

$\bar{x} = 10$ $\quad$ $\bar{y} = 12$.

$\bar{y} \sum x_i = 840$

$\bar{x} \sum y_i = 840$

$n \bar{x} \bar{y} = 840$.