Simple linear regression

$\rightarrow$ # of pred. var = 1

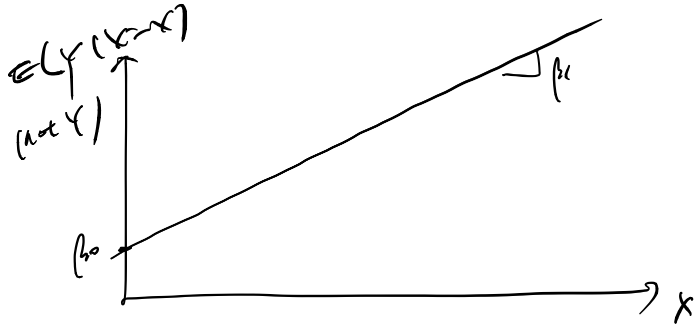
dataset

$\rightarrow \boxed{(x_1, y_1) \cdots (x_n, y_n)}$

- there is a prob. dist^n of Y for each level of X (= for given $X = x_i$)
- means of these prob dist^n vary in some fashion wrt X.

$$\Rightarrow \quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1 \ldots n, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$\Rightarrow \quad E(y_i \mid X_i = x_i) = \underline{\underline{\beta_0}} + \underline{\underline{\beta_1}} x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$



LS method. LSE of $\beta_0, \beta_1 \Rightarrow \underset{b_0}{\underset{\parallel}{\hat{\beta}_0}}, \underset{b_1}{\underset{\parallel}{\hat{\beta}_1}}$

b_0 which minimizes $\boxed{S(\beta_0, \beta_1) = \sum_{i=1}^{n} (\underbrace{y_i}_{y_i} - \underbrace{\beta_0 - \beta_1 x_i}_{E(y_i \mid x_i)})^2}$

(i.e) $b_0 (= \hat{\beta}_0) = \underset{\beta_0}{\arg\min}\ S(\beta_0, \beta_1)$

$b_1 (= \hat{\beta}_1) = \underset{\beta_1}{\arg\min}\ S(\beta_0, \beta_1)$.

$$\Rightarrow \frac{\partial S}{\partial \beta_0} = 0, \quad \frac{\partial S}{\partial \beta_1} = 0 \quad \left(\frac{\partial^2 S}{\partial \beta^2} > 0 \right)$$

① $\sum (y_i - \underbrace{\hat\beta_0 - \hat\beta_1 x_i}_{\hat y_i}) = 0$ ② $\sum x_i (y_i - \hat\beta_0 - \hat\beta_1 x_i) = 0$.

$\Downarrow$

$\sum e_i = 0$

✓ $\hat\beta_0^{\,b_0} = \bar y - \hat\beta_1 \bar x$, ✓ $\hat\beta_1^{\,b_1} = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar x)(y_i - \bar y)}{\sum (x_i - \bar x)^2}$

$S_{xy} = \sum (x_i - \bar x)(y_i - \bar y) = \sum (x_i - \bar x) y_i - \cancel{\bar y \sum (x_i - \bar x)}_{0}$

$= \sum (x_i - \bar x) y_i = \sum (y_i - \bar y) x_i$

$S_{xx} = \sum (x_i - \bar x) x_i$

$b_1 = \hat\beta_1 = \frac{\sum (x_i - \bar x) y_i}{S_{xx}} \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$ ✓

$\because$ fct of $y_i \sim N(,)$

$b_0 = \bar y - \hat\beta_1 \bar x \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar x^2}{S_{xx}}\right)\right)$

$\because cov(\bar y, \hat\beta_1) = 0$.

$V(b_0) = V(\bar y) + V(\hat\beta_1 \bar x) = \sigma^2/n + \bar x^2 Var(\hat\beta_1)$

$\sigma^2 \leftarrow \left(\hat\sigma^2 = \frac{\sum (y_i - b_0 - b_1 x_i)^2}{n-2}\right)$ $E(y_i) = \beta_0 + \beta_1 x_i$

$\uparrow b_0 \quad \uparrow b_1$

$\sigma^2 \leftarrow s^2 = \frac{\sum (x_i - \overset{\mu}{\bar x})^2}{n-1}$ y_i

$\frac{(n-2)\hat\sigma^2}{\sigma^2} \sim \chi^2_{n-2}$ · $\perp\!\!\!\perp$ b_0, b_1

by def. of t-dist^n

$z = \frac{\hat\beta_1 - \beta_1}{\sqrt{\sigma^2/ \ldots}} \sim N(0,1)$ $\perp\!\!\!\perp$ $\frac{(n-2)\hat\sigma^2}{\sigma^2} \sim \chi^2_{n-2}$.

$\parallel$

ν^2

$$\sqrt{\sigma^2 / S_{xx}}$$

$$\Rightarrow \frac{Z}{\sqrt{\chi^2 / n-2}} \sim t_{n-2}.$$

$H_0 :$ (goal goal)^c

$H_a :$ (your goal)

$\neq$

$=$

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2}.$$

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t_{n-2}.$$

#6 $\underset{\sim}{Z} = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N_2\left(\underset{\sim}{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma\right)$ $\quad |\Sigma| = \det(\Sigma)$ $\quad \Sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

dimension of Z

(a) $f_Z(z) = (2\pi)^{-\frac{2}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \underbrace{\underbrace{(\underset{\sim}{z} - \underset{\sim}{\mu})^T}_{1\times 2} \Sigma^{-1} \underbrace{(\underset{\sim}{z} - \underset{\sim}{\mu})}_{2\times 1}}\right\}$

$\Rightarrow 1\times 1 \in \mathbb{R}^1.$

$\underset{\sim}{z} = \begin{pmatrix} x \\ y \end{pmatrix}$, Σ^{-1} $\quad |\Sigma| = ad - bc.$

$f(x,y) = \boxed{}$

$v = \frac{x - \mu_x}{\sigma_x}$

$w = \frac{y - \mu_y}{\sigma_y}$

(b). Show $X \sim N(\mu_x, \sigma_x^2)$.

$$f(x) = \int f(x,y)\, dy$$

C : some constant vector

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N_2(\mu, \Sigma) \Rightarrow CZ \sim N_2\left(C\mu, C\Sigma C^T\right)$$

$$\text{(c)} \quad f_{X|Y} = \frac{f(x,y) \leftarrow (a)}{f(y) \leftarrow (b)}$$

$$= \boxed{\exp\left(\frac{X - \boxed{\alpha+\beta x}}{\boxed{\sqrt{\hat\sigma^2_{Y|X}}}}\right)^2} \times C = \boxed{}$$

#5 (b) $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

$$t^* = \boxed{\frac{\hat\beta_1 - \beta_1}{\sqrt{\hat\sigma^2 / S_{XX}}}} \sim t_{n-2}$$

rej. region $C = \left\{ X : \left| \frac{\hat\beta_1 - 0}{\sqrt{\hat\sigma^2 / S_{XX}}} \right| > t_{n-2, \alpha/2} \right\}$.

$(100)(1-\alpha)\%$

C.I for β_1 $= P_r\left(-t_{n-2,\alpha/2} < t^* < t_{n-2,\alpha/2}\right) = 1-\alpha$.

$$\Rightarrow \left(\hat\beta_1 \pm t_{\ldots} \sqrt{\hat\sigma^2 / \ldots} \right)$$

$\left(\hat\beta_1 \pm t_{n-2,\alpha/2} \sqrt{\hat\sigma^2 / S_{xx}} \right)$.

$\alpha = 0.05$

$\sigma = 2.5$

$\beta_1 = \pm 0.5$.

power

$$\gamma(\beta) = \Pr\left[\text{rej } H_0 \mid \underset{\beta_1 \neq 0}{H_0 \text{ is not true}} \right]$$

$$= \Pr\left[\left| \frac{\hat\beta_1 - 0}{\underbrace{\sqrt{\sigma^2/S_{xx}}}_{s.e(\hat\beta_1) = se}} \right| > z_{\alpha/2} \;\middle|\; \beta_1 \neq 0 \right]$$

$$= \Pr\left[\underbrace{\frac{\hat\beta_1}{se} > z_{\alpha/2}}_{} \text{ or } \underbrace{\frac{\hat\beta_1}{se} < -z_{\alpha/2}}_{} \;\middle|\; \beta_1 \neq 0 \right]$$

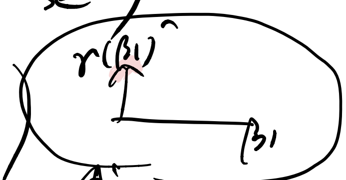
$$\frac{\hat\beta_1 - \beta_1}{se} \sim N(0,1)$$

$$= \Pr\left[\underbrace{\frac{\hat\beta_1 - \beta_1}{se}}_{"Z \sim N(0,1)} > z_{\alpha/2} - \frac{\beta_1}{se} \right] + \Pr\left[\underbrace{\frac{\hat\beta_1 - \beta_1}{se}}_{} < -z_{\alpha/2} - \frac{\beta_1}{se} \right]$$

$\Phi(x) = \Pr(Z < x)$ with $Z \sim N(0,1)$.

$$= 1 - \Pr\left(Z < z_{\alpha/2} - \frac{\beta_1}{se} \right) + \Pr\left(Z < -z_{\alpha/2} - \frac{\beta_1}{se} \right)$$

$$= 1 - \Phi\left(z_{\alpha/2} - \frac{\beta_1}{se} \right) + \Phi\left(-z_{\alpha/2} - \frac{\beta_1}{se} \right)$$



$$\gamma(\beta_1) = 1 - \Phi\left(z_{\alpha/2} - \frac{\beta_1}{se} \right) + \Phi\left(-z_{\alpha/2} - \frac{\beta_1}{se} \right)$$

$$= \gamma(-\beta_1).$$

$\Phi(x) = 1 - \Phi(-x)$.