Simple Linear Regression

① theory

$Y = \alpha + \beta x + \varepsilon$, $\mu_{Y/x} = \alpha + \beta x$ ("true" regression line)

$\hat{Y} = A + Bx$, $\hat{y} = a + bx$ (fitted regression line; 拟合回归线)

process

$Y_i = \alpha + \beta x_i + \varepsilon_i$ (random variable of y_i for a given x_i)

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ (sample mean)

$\hat{Y}_i = A + Bx_i$, $\hat{y}_i = a + bx_i$

result

$Y_0 = \alpha + \beta x_0 + \varepsilon$

$\hat{Y}_0 = A + Bx_0$, $\hat{y}_0 = a + bx_0$ (point predictor of y_0)

② (assume: $E\varepsilon = 0$, $Var(\varepsilon) = \sigma^2$; $E\varepsilon_i = 0$, $Var(\varepsilon_i) = \sigma^2$)

$EY = \alpha + \beta x$, $Var(Y) = \sigma^2$

$EY_i = \alpha + \beta x_i$, $Var(Y_i) = \sigma^2$

$EB = \beta$, $Var(B) = \frac{\sigma^2}{S_{xx}}$

$EA = \alpha$, $Var(A) = \frac{\sum_{i=1}^{n} x_i^2}{nS_{xx}}\sigma^2$

$ES^2 = \sigma^2$

$E(Y_0 - \hat{Y}_0) = 0$, $Var(Y_0 - \hat{Y}_0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]$

③ $e_i \equiv y_i - \hat{y}_i$; $SSE \equiv \sum_{i=1}^{n} e_i^2$

$b = \frac{S_{xy}}{S_{xx}}$, $a = \bar{y} - b\bar{x}$

$S_{xx} \equiv \sum_{i=1}^{n}(x_i - \bar{x})^2$, $S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$, $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$.

$B = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $A = \bar{Y} - B\bar{x}$

$S^2 \equiv \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, $s^2 = \frac{S_{yy} - bS_{xy}}{n-2}$

④ parameter : α , β , σ^2.

point estimator : a , b , s^2.

estimator : A , B , S^2

⑤ $$\frac{B-\beta}{\frac{s}{\sqrt{S_{xx}}}} \sim t\text{-dist} \quad (\nu = n-2)$$

$$\frac{A-\alpha}{s\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}}}} \sim t\text{-dist} \quad (\nu = n-2)$$

$$\frac{\hat{Y}_0 - Y_0}{s\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}} \sim t\text{-dist} \quad (\nu = n-2)$$

⑥ $(1-\alpha)\times 100\%$ C.I : β : $U\left(b,\ t_{\frac{\alpha}{2}}\frac{s}{\sqrt{S_{xx}}}\right)$.

α : $U\left(a,\ t_{\frac{\alpha}{2}}\cdot s\cdot\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}}}\right)$.

$(1-\alpha)\times 100\%$ prediction interval for y_0 : $U\left(\hat{y}_0,\ t_{\frac{\alpha}{2}}\cdot s\cdot\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$.

testing for β : $a = \beta_0 + t_\alpha\frac{s}{\sqrt{S_{xx}}}$; $a, b = \beta_0 \pm t_{\frac{\alpha}{2}}\frac{s}{\sqrt{S_{xx}}}$

for α : $a = \alpha_0 + t_\alpha \cdot s\cdot\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}}}$; $a, b = \alpha_0 \pm t_{\frac{\alpha}{2}}\cdot s\cdot\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}}}$

⑦ Sample Correlation :

$$r \equiv \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = b\sqrt{\frac{S_{xx}}{S_{yy}}} \qquad (-1 \le r \le 1)$$

$$\tilde{\chi}^2 = \sum_{i=1}^{k}\frac{(O_i - e_i)^2}{e_i} \sim \chi^2 (\nu = k-1)$$

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim Z \qquad \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t\text{-dist} \quad (\nu = n-1)$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2\text{-dist} \quad (\nu = n-1.$$