ΣΤΑΤΙΣΤΙΚΗ

Διάλεξη 15η

26/04/2018

Υπενθύμιση

Απλή Γραμμική Παλινδρόμηση

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \, , \, i = 1, \dots, n$ || x: ανεξάρτ. μετ. / y: εξαρτ. μετ.

$\varepsilon_i \sim (0, \sigma^2)$: τυχαία σφάλματα, ασυσχέτιστα κ.λπ.

$E(Y_i) = \beta_0 + \beta_1 x_i$: άγνωστη συνάρτηση παλινδρόμησης.

$\varepsilon_i = Y_i - E(Y_i)$

$$\boxed{\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i} \quad (1)$$

$$\boxed{\sum_{i=1}^{n} x_i Y_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2} \quad (2)$$

κανονικές εξισώσεις.

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$

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

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$

Υπόλοιπα:

$e_i = Y_i - \hat{Y}_i$

$\hat{\beta}_0 = 10.0 \, , \quad \hat{\beta}_1 = 2.0$

$\hat{Y} = 10.0 + 2.0x$.

Ιδιότητες της εκτιμώμενης ευθείας παλινδρόμησης:

(i) Το κέντρο βάρους των παρατηρήσεων $(\bar{x},\bar{y})$ είναι σημείο και ευθεία διότι:

για $x=\bar{x}$ είναι $\hat{y}=\hat{\beta}_0+\hat{\beta}_1\bar{x}=\bar{y}-\hat{\beta}_1\bar{x}+\hat{\beta}_1 x=\bar{y}$

(ii) $\sum_{i=1}^{n} e_i=0$ διότι $\sum_{i=1}^{n}(y_i-\hat{y}_i)=\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)$

$$=\sum_{i=1}^{n} y_i-n\hat{\beta}_0-\hat{\beta}_1\sum_{i=1}^{n}x_i$$

$=0$ από την 1 κανονική εξίσωση

(iii) $\sum_{i=1}^{n} e_i x_i=\sum_{i=1}^{n}(y_i-\hat{y}_i)x_i=\sum_{i=1}^{n}x_i(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)=0$

από την 2 κανονική εξίσωση

Ιδιότητες των εκτιμητών ελαχίστων τετραγώνων:

$$\omega=\sum_{i=1}^{n}a_i\omega_i$$

$$E(\omega)=\sum_{i=1}^{n}a_i E(\omega_i)$$

$Var(\omega)=\sum_{i=1}^{n}a_i^2 Var(\omega_i)$ για ω_i ανεξ. τ.μ.

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}-\cancel{\frac{\overbrace{\sum_{i=1}^{n}(x_i-\bar{x})}^{0}\bar{y}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

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

$$E(\hat{B}_1) = \sum_{i=1}^{n}\frac{(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}E(y_i) = \sum_{i=1}^{n}\frac{(x_i-\bar{x})}{(x_i-\bar{x})^2}(B_0+B_1x_i)$$

$$= \cancelto{0}{\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}B_0 + \frac{\sum_{i=1}^{n}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}B_1x_i$$

$$= B_1\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})x_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \frac{\sum_{i=1}^{n}(x_i-\bar{x})\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}B_1 = B_1$$

$$Var(\hat{B}_1) = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}Var(y_i) = \sigma^2\cdot\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

$$= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\hat{B_0} = \bar{Y} - \hat{B}_1 \bar{x} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)Y_i$$

$$E(\hat{B_0}) = \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)(B_0 + B_1 x_i)$$

$$= B_0 \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) + B_1 \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)x_i$$

$$= B_0 \cdot \left(\frac{n}{n} - \cancel{\frac{\bar{x}\sum_{i=1}^{n}(x_i-x)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}^{\,0}\right) + B_1\left(\frac{\sum_{i=1}^{n} x_i}{n} - \frac{\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})x_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$= B_0 + B_1 \cdot 0 = B_0$$

$$Var(\hat{B_0}) = \sigma^2 \cdot \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^2$$

$$= \sigma^2 \cdot \sum_{i=1}^{n}\left(\frac{1}{n^2} + \frac{\bar{x}^2(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2} - 2\frac{1}{n}\frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

Στατιστική Συμπερασματολογία:

$\varepsilon_i \sim N(0, \sigma^2)$ και ανεξ τ.μ.:

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

$$\boxed{\hat{\beta}_0 = \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)y_i}$$

ή ανεξ τ.μ. : $y_i \sim N\left(E(y_i) = \beta_0 + \beta_1 x_i, \sigma^2\right)$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)\right)$$

<u>και</u> $$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

~o~o~o~o~o~o~o~o~o~o~o~

$$B_1 = \frac{\hat{\beta}_1 - \beta_1}{\frac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{n-2}$$

$$B_0 = \frac{\hat{\beta}_0 - \beta_0}{S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{n-2}$$

$(1-\alpha)100\%$ Δ.Ε. για B_1 : $\hat{B}_1 \pm t_{\frac{\alpha}{2}, n-2} \dfrac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

$(1-\alpha)100\%$ Δ.Ε. για B_0 : $\hat{B}_0 \pm t_{\frac{\alpha}{2}, n-2} \cdot S\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

$H_0 : B_0 = 0$ v $H_a : B_0 \neq 0$, και η κρισιμη περιοχή είναι
$|B_0| \geq t_{\alpha/2, n-2}$ (για $B_0 = 0$)

$H_0 : B_1 = 0$ v $H_a : B_1 \neq 0$ και η κρισιμη περιοχη είναι
$|B_1| \geq t_{\alpha/2, n-2}$ (για $B_1 = 0$)
όταν Ho αληθές.

~o~o~o~o~o~o~o

Για το 5.2 παράδειγμα:

$S^2 = 7.5 = \dfrac{60}{8}$

$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 3400$

$t_{0.025,8} = 2.306$

$B_1 = \dfrac{2.0 - 0}{\sqrt{7.5}/\sqrt{3400}} = 42.6$

$42.6 > 2.306$ απορρ Ho,
υπάρχει γραμμ σχέση.

95% ΔΕ για το B_1 : $2.0 \pm 2.306 \cdot \dfrac{\sqrt{7.5}}{\sqrt{3400}} \Rightarrow [1.89, 2.11]$

Δ.Ε για την συνάρτηση παλινδρόμησης :

$E(Y) = B_0 + B_1 x$ και $\hat{Y}_0 = \hat{B}_0 + \hat{B}_1 x_0$ για $x = x_0$

με $\boxed{E(\hat{Y}_0) = B_0 + B_1 x_0 = E(Y_0)}$

$$\boxed{Var(\hat{Y}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)}$$

$$\hat{Y}_0 \sim N\left(B_0 + B_1 x_0 = E(Y_0) ,\ \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) \right)$$

$$\frac{\hat{Y}_0 - E(Y_0)}{S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \sim t_{n-2}$$ και έτσι:

$(1-a)\%$ Δ.Ε για $E(Y_0)$:

$$\hat{Y}_0 \pm t_{a/2, n-2} \cdot S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Για $x_0 = 55$, $\hat{Y}_0 = 10 + 2 * 55 = 120$ και Δ.Ε. 95% :

$$120 \pm 1.860 \cdot \sqrt{7.5}\sqrt{\frac{1}{10} + \frac{(55-50)^2}{\sqrt{3.400}}}$$