27/4/17

Απλή Γραμμική Παλινδρόμηση

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1, \dots, n$ | X: ελεγχόμενη
Y: εξαρτημένη τμ

Σφάλματα: $\varepsilon_i \sim (0, \sigma^2)$, $\varepsilon_i = Y_i - E(Y_i)$

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

$Y_i \sim (E(Y_i), \sigma^2)$

→ $\hat\beta_0 = \bar Y - \hat\beta_1 \bar x$, $\hat\beta_1 = \dfrac{\sum_1^n (X_i - \bar X)(Y_i - \bar Y)}{\sum (X_i - \bar X)^2}$, $\hat\sigma^2 = S^2 = \dfrac{1}{n-2} \sum e_i^2$

$\hat Y_0 = \hat\beta_0 + \hat\beta_1 X_0$ → εκτιμώμενη συνάρτηση παλινδρ.

υπόλοιπα $e_i = Y_i - \hat Y_i$, $\sum e_i = 0$, $\sum e_i x_i = 0$

Ιδιότητες των εκτιμητών των ελ. τετρ. [illegible]

$W = \sum_1^n a_i w_i$, ανεξ. τμ, $E(W) = \sum_1^n a_i E(w_i)$, $Var(W) = \sum_1^n a_i^2 Var(w_i)$

ισότητα $\sum (x_i - \bar x) = 0$

$$\hat\beta_1 = \frac{\sum_1^n (X_i - \bar X) Y_i}{\sum_1^n (X_i - \bar X)^2} - \frac{\sum_1^n (X_i - \bar X) \bar Y}{\sum_1^n (X_i - \bar X)^2} = \frac{\sum_1^n (X_i - \bar X) Y_i}{\sum_1^n (X_i - \bar X)^2}$$

$$= \sum_1^n \frac{(X_i - \bar X)}{\sum_1^n (X_i - \bar X)^2} Y_i$$

$$E(\hat\beta_1) = \sum_1^n \frac{(x_i - \bar x)}{\sum_1^n (x_i - \bar x)^2} E(Y_i) = \sum_1^n \frac{(x_i - \bar x)}{\sum_1^n (\ )^2} (\beta_0 + \beta_1 x_i) = \dots$$

$$= \frac{\sum (x_i - \bar x) \beta_0}{\sum (\ )^2} + \frac{\sum (x_i - \bar x) \beta_1 x_i}{\sum (x_i - \bar x)^2} =$$

$$= \frac{\sum (x_i-\bar{x})^{\nearrow 0}}{\sum (\ )^2} + \frac{\sum (x_i-\bar{x})\beta_1 x_i}{\sum (x_i-\bar{x})^2}$$

$$= \beta_1 \frac{\sum (x_i-\bar{x}) x_i}{\sum (x_i-\bar{x})^2}$$

$$= \beta_1 \frac{\sum^n (x_i-\bar{x}) x_i}{\sum (x_i-\bar{x})^2} - \beta_1 \frac{\sum (x_i-\bar{x})}{\sum (x_i-\bar{x})^2}$$

$$= \beta_1 \frac{\sum (x_i-\bar{x})(x_i-\bar{x})}{\sum (x_i-\bar{x})^2}$$

$$Var(\hat{\beta_1}) = \sum_i^n \left[\frac{(x_i-\bar{x})}{\sum (x_i-\bar{x})^2}\right]^2 \underbrace{Var(y_i)}_{\sigma^2}$$

$$= \sigma^2 \frac{\sum (x_i-\bar{x})^2}{\left[\sum (x_i-\bar{x})^2\right]^2}$$

$$= \sigma^2 / \sum (x_i-\bar{x})^2$$

$$\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x} = \bar{y} - \frac{\sum (x_i-\bar{x})}{\sum (x_i-\bar{x})^2}\bar{x} = \sum_{i=1}^n \left[\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_i (x_i-\bar{x})^2}\right] y_i$$

$\frac{1}{n}\sum y_i$

$$E(\hat{\beta_0}) = \sum \left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum (\ )^2}\right] (\beta_0 + \beta_1 x_i)$$

$$= \sum \left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum (\ )^2}\right] \beta_0 + \beta_1 \sum \left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum (x_i-\bar{x})^2}\right) x_i$$

(=)

$$E(\hat{\beta_0}) = \beta_0\left(\frac{n}{n} - \frac{\overbrace{\Sigma\,\bar{x}(x_i-\bar{x})}^{0}}{\Sigma(\ )^2}\right) + \beta_1\left(\underbrace{\frac{\Sigma x_i}{n}}_{\bar{x}} - \frac{\Sigma\,\bar{x}(x_i-\bar{x})\,x_i}{\underbrace{\Sigma(x_i-\bar{x})^2}_{\bar{x}}}\right)$$

$$= \beta_0$$

$$\cancel{\bar{x}\left[\frac{\Sigma(x_i-\bar{x})\,x_i}{\Sigma(x_i-\bar{x})^2} - \frac{\overbrace{\Sigma(x_i-\bar{x})\,\bar{x}}^{0}}{\Sigma(x_i-\bar{x})^2}\right]}$$

$$Var(\hat{\beta_0}) = \sum_1^n\left[\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\Sigma(x_i-\bar{x})^2}\right]^2\sigma^2$$

$$= \sigma^2\sum_1^n\left[\frac{1}{n^2} + \frac{\bar{x}^2(x_i-\bar{x})^2}{\Sigma(x_i-\bar{x})^2} - 2\cdot\frac{1}{n}\,\frac{\bar{x}(x_i-\bar{x})}{\Sigma(x_i-\bar{x})^2}\right]$$

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

$$\hat{\beta_0} = \Sigma\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\Sigma(x_i-\bar{x})^2}\right)y_i \quad , \hat{\beta_1} = \frac{\sum_1^n (x_i-\bar{x})\,y_i}{\sum_1^n (x_i-\bar{x})^2}$$

$E(\hat{\beta_0}) = \beta_0$, $E(\hat{\beta_1}) = \beta_1$,

$Var(\hat{\beta_0}) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(\ )^2}\right)$ $\quad$ $Var(\hat{\beta_1}) = \sigma^2/\Sigma(\ )^2$

Στατιστική Συμπερασματολογία

$H_0: \beta_0=0$, $H_0: \beta_1=0$, $(1-\alpha)100\%$ Δ.Ε για β_0, $(1-\alpha)100\%$ Δ.Ε για β_1

Πρόσθετη υπόθεση: $\varepsilon_i \sim N(0,\sigma^2)$ δ.δ.
$y_i \sim N(E(y_i), \sigma^2)$

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

- $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}\right)$

- $\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$ κ' S^2 ανεξ. των $\hat{\beta}_0, \hat{\beta}_1$

- $\frac{\hat{\beta}_1-\beta_1}{S/\sqrt{\Sigma(x_i-\bar{x})}} \sim t_{n-2}$

$$\frac{\dfrac{\hat{\beta}_1-\beta_1}{\sigma/\sqrt{\Sigma(x_i-\bar{x})^2}} \sim N(0,1)}{\sqrt{\dfrac{(n-2)S^2}{\sigma^2}/(n-2)} \sim \sqrt{\dfrac{\chi^2_{n-2}}{n-2}}} \sim t_{n-2}$$

$\downarrow$

$(1-\alpha)\cdot 100\%$ Δ.Ε. για β_1:

$\beta_1: \hat{\beta}_1 \pm t_{\alpha/2,\, n-2} \frac{S}{\sqrt{\Sigma(x_i-\bar{x})^2}}$

$H_0: \beta_1=0$ v $H_\alpha: \beta_1 \neq 0$

$B_1 = \frac{\hat{\beta}_1-\beta_1}{S/\sqrt{\Sigma(x_i-\bar{x})^2}} \overset{\beta_1=0}{\sim} t_{n-2}$ όταν H_0 αληθ.

κ' κρ. περ. μεγέθ. α $|B_1| \geq t_{\alpha/2,\, n-2}$ (με $\beta_1=0$)

$$\frac{\hat{\beta}_0-\beta_0}{S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}} \sim t_{n-2} \to (1-\alpha)100\%$$ Δ.Ε. για β_0:

$$\beta_0: \hat{\beta}_0 \pm t_{\alpha/2,n-2} \cdot S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}$$

$H_0: \beta_0=0$ v H_α: $\beta_0 \neq 0$

$$\beta_0 = \frac{\hat{\beta}_0-\beta_0}{S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}} \sim t_{n-2}$$ όταν H_0 αληθής

και κρίσιμη περιοχή μεγέθους α $|\beta_0| \geq t_{\alpha/2,n-2}$ (με $\beta_0=0$)

$S^2 = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}$: υπολογίζεται διαιρ...

→ Παράδειγμα ΕΒΟ

$\hat{\beta}_1 = 2.0$, $\sum(x_i-\bar{x})^2 = 3.400$, $S^2 = \frac{60}{8} = 7.5$

8 ← υπολογίζεται με t-test

$$\beta_1 = \frac{2.0-0}{\sqrt{7.5}/\sqrt{3.400}} = 42.6$$

$42.6 > 2.306 = t_{0.025,8}$ απορ. $H_0: \beta_1=0$

95% Δ.Ε. για το β_1: $2.0 \pm 2.306 \frac{\sqrt{7.5}}{\sqrt{3.400}}$

$t_{\alpha/2,n-2} = t_{0.025} = 2.306$ $\quad$ 95% $[1.89, 2.11]$

Δ.Ε. για τη συνάρτηση Παλινδρόμησης

$E(Y) = \beta_0 + \beta_1 X$

για $X = x_0$: $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ με $E(\hat{Y}_0) = \beta_0 + \beta_1 x_0 = E(Y_0)$

$$\hat{Y}_0 \sim N(E(Y_0), Var(\hat{Y}_0))$$

με $Var(\hat{Y}_0) = \sigma^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}\right)$ + δεν χρειάζεται

$$\frac{\hat{Y}_0 - E(Y_0)}{S\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma(x_0-\bar{x})^2}}} \sim t_{n-2}$$ κ' $(1-\alpha)\cdot 100\%$ Δ.Ε. για

$E(Y_0 \mid X = x_0)$: $\hat{Y} \pm t_{\alpha/2, n-2} \cdot S\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}}$

Π.χ. για $x_0 = 55$, $\hat{Y}_0 = 180$

$[118,3, 121.7]$

90% Δ.Ε.

$\alpha = 0,05$