Απλή Γραμμική Παλινδρόμηση

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , \quad i = 1, \ldots, n$

τ.μ. $\varepsilon_i \sim (0, \sigma^2)$: τυχαία σφάλματα

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

$\min \sum_{i=1}^{n} \varepsilon_i^2 \to \hat\beta_0, \hat\beta_1$ $\qquad \varepsilon_i = Y_i - E(Y_i)$

$\hat\beta_0 = \bar{Y} - \hat\beta_1 \bar{X}$ $\qquad e_i = Y_i - \hat{Y}_i$: υπόλοιπα

$\hat\beta_1 = \dfrac{\sum_{i=1}^{n}(X_i - \bar X)(Y_i - \bar Y)}{\sum_{i=1}^{n}(X_i - \bar X)^2}$ $\qquad \sum e_i = 0$

$\sum e_i X_i = 0$

$\hat Y_0 = \hat\beta_0 + \hat\beta_1 X_0$ (για $X = X_0$) $\to$ (ΕΒΟ: $\hat Y_0 = 10{,}0 + 2{,}0 \cdot X_0$)

Ιδιότητες των εκτιμητών ελαχ. τετραγ.

$W = \sum a_i W_i$, $E(W) = \sum a_i E(W_i)$, $Var(W) = \sum a_i^2 Var(W_i)$ για W_i, $i = 1, \ldots, n$ ανεξ τ.μ.

$$\hat\beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar X) Y_i}{\sum (X_i - \bar X)^2} - \cancel{\frac{\sum_{i=1}^{n}(X_i - \bar X)\bar Y}{\sum (X_i - \bar X)^2}}^{\,0}$$

$$E(\hat\beta_1) = \sum_{i=1}^{n} \frac{X_i - \bar X}{\sum_{i=1}^{n}(X_i - \bar X)^2} E(Y_i) = \sum_{i=1}^{n} \frac{X_i - \bar X}{\sum (X_i - \bar X)^2}(\beta_0 + \beta_1 X_i)$$

$$= \beta_0 \frac{\sum (X_i - \bar X)}{\sum (X_i - \bar X)^2} + \beta_1 \frac{\sum (X_i - \bar X) X_i}{\sum (X_i - \bar X)^2} - \beta_1 \frac{\cancel{\sum (X_i - \bar X)}^{\,0} \bar X}{\sum (X_i - \bar X)^2} = \beta_1 \frac{\sum (X_i - \bar X)(X_i - \bar X)}{\sum (X_i - \bar X)^2} = \beta_1$$

$E(\hat\beta_1) = \beta_1 \Rightarrow$ ο $\hat\beta_1$ είναι αμερόληπτος εκτιμητής

$$Var(\hat\beta_1) = \sum \left[\frac{(X_i - \bar X)}{\sum (X_i - \bar X)} \right]^2 \cdot Var(Y_i) = \frac{\sum (X_i - \bar X)^2}{(\sum (X_i - \bar X)^2)^2} \sigma^2 = \frac{\sigma^2}{\sum (X_i - \bar X)^2}$$

Άρα $E(\hat\beta_1) = \beta_1$ και $Var(\hat\beta_1) = \dfrac{\sigma^2}{\sum (X_i - \bar X)^2}$

$$\hat\beta_0 = \frac{1}{n} \sum_{i=1}^{n} Y_i - \bar X \frac{\sum_{i=1}^{n}(X_i - \bar X) Y_i}{\sum (X_i - \bar X)^2} = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{\bar X (X_i - \bar X)}{\sum_{i=1}^{n}(X_i - \bar X)^2} \right) Y_i$$

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

$$= \beta_0 \sum \frac{1}{n} - \beta_0 \bar X \frac{\cancel{\sum (X_i - \bar X)}^{\,0}}{\sum (X_i - \bar X)^2} + \beta_1 \frac{\sum X_i}{n} - \beta_1 \bar X \frac{\sum (X_i - \bar X) X_i}{\sum (X_i - \bar X)^2} =$$

$= \beta_0 - 0 + \beta_1 \bar X - \beta_1 \bar X = \beta_0$ Άρα $E(\hat\beta_0) = \beta_0 \Rightarrow \hat\beta_0$ αμερόληπτος

$$Var(\hat{\beta}_0) = \sum\left[\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right]^2 Var(Y_i) =$$

$$= \sigma^2 \sum\left(\frac{1}{n^2} + \frac{\bar{x}^2(x_i-\bar{x})^2}{[\sum(x_i-\bar{x})^2]^2} - 2\frac{1}{n}\frac{\bar{x}\cancel{(x_i-\bar{x})}^{0}}{\sum(x_i-\bar{x})^2}\right) =$$

$$= \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2} + 0\right]$$

Τελικά, $E(\hat{\beta}_0)=\beta_0$, $Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)$

<u>Μέσο τετραγωνικό σφάλμα</u>

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i)^2$$

$n-2$ βαθμοί ελευθερίας

Αν $\beta_1 = 0$, δεν υπάρχει γραμμική σχέση μεταξύ των X, Y

<u>Στατιστική Συμπερασματολογία</u>

$Y_i \sim N(\beta_0+\beta_1 X_i, \sigma^2)$ δηλ. $\varepsilon_i \sim N(0,\sigma^2)$, ανεξ. τ.μ.

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

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

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

Μπορεί να αποδειχθεί: $\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$

Δηλ. $\beta_1 = \frac{\hat{\beta}_1-\beta_1}{S/\sqrt{\sum(x_i-\bar{x})^2}} \sim t_{n-2}$ $\quad(1-a)100\%$ ΔΕ για β_1:

$$\hat{\beta}_1 \pm t_{a/2,n-2}\frac{S}{\sqrt{\sum(x_i-\bar{x})^2}}$$

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

$(1-a)100\%$ ΔΕ για β_0: $\hat{\beta}_0 \pm t_{a/2,n-2}\, S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}$

<u>Στο πχ. (ΕΒΟ)</u>:

95% ΔΕ για το β_1: $[1.89, 2.11]$

$t_{0.025,8} = 2.306$, $S^2 = \frac{60}{8} = 7.5$, $\sum(x_i-\bar{x})^2 = 3{,}400$

Για να ελέγξουμε την υπόθεση:

$H_0: \beta_1 = 0$ και $H_a: \beta_1 \neq 0$ χρησιμοποιούμε το στατιστικό:

β_1 (με $\beta_1 = 0$) δηλ. $\beta_1 = \frac{\hat{\beta}_1 - 0}{S/\sqrt{\sum(x_i-\bar{x})^2}} \overset{H_0}{\sim} t_{n-2}$ και κρ. περ. μεγέθους a

$|\beta_1| \geq t_{a/2,n-2}$

<u>Για το πχ. της ΕΒΟ</u>

$\beta_1 = \frac{2.0}{\sqrt{7.5}/\sqrt{3400}} = 42.6$ η κρ. περ.: $|\beta_1| \geq t_{0.025,8} (=2.306)$

Επειδή $42.6 > 2.306$ απορ. την H_0 δηλ. $\exists$ γραμμ. σχέση μεταξύ των X,Y

Για να ελέγξουμε την υπόθεση ότι $H_0: \beta_0 = 0$ και $H_a: \beta_0 \neq 0$

Χρησιμοποιούμε το στατιστικό: $\beta_0 = \dfrac{\hat{\beta}_0 - 0}{S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}} \overset{H_0}{\sim} t_{n-2}$ και

κρ. περ. μεγέθους α: $|\beta_0| \geq t_{\alpha/2, n-2}$

Δ.Ε. για τη συνάρτηση παλινδρόμησης

$E(Y_0) = \beta_0 + \beta_1 X_i$ και εκτιμ. συνάρτηση παλινδρ. $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_0$ για $X = X_0$

Χρειαζόμαστε την κατανομή του $\hat{Y}_0$:

$$\hat{Y}_0 \sim N\left[\beta_0 + \beta_1 X_0 = E(Y_0), \sigma^2\left(\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{\sum(x_i-\bar{x})^2}\right)\right] \text{ και } \frac{(n-2)S^2}{\sigma^2} \sim X^2_{n-2}$$

$$\frac{\hat{Y}_0 - E(Y_0)}{S\sqrt{\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{\sum(x_i-\bar{x})^2}}} \sim t_{n-2} \Rightarrow (1-\alpha)100\% \text{ ΔΕ για } E(Y_0): \hat{Y}_0 \pm t_{\alpha/2, n-2} S\sqrt{\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{\sum(x_i-\bar{x})^2}}$$

Για το πχ. της ΕΒΟ

ΕΒΟ: 90% ΔΕ για $E(Y|X_0 = 55)$

$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0 = 10 + 2 * 55 = 120$, $t_{0.05, 8} = 1,860$

$S^2 = 7.5$, $\sum(x_i - \bar{x})^2 = 3,400$