02/05/18 ΣΤΑΤΙΣΤΙΚΗ ①

Διάλεξη 16η

Γενικό επαναληπτικό τυπολόγιο

$Y_i = B_0 + B_1 x_i + \varepsilon_i \quad , i = 1, \ldots, n$

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

$E(Y_i) = B_0 + B_1 x_i$: άγνωστη συνάρτηση παλινδρ.

$\varepsilon_i = Y_i - E(Y_i)$

$Y_i \sim N(E(Y_i), \sigma^2)$: ανεξ τ.μ., σ^2 ανεξάρτητο του X.

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{x} \;, \quad \hat{B}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \;, \quad \hat{B}_1 \sim N\left(B_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$\hat{Y} = \hat{B}_0 + \hat{B}_1 x$: εκτιμ. συνάρτηση παλινδρόμησης.

$e_i = Y_i - \hat{Y}_i$: υπόλοιπα, $\sum_{i=1}^{n} e_i = 0$, $\sum_{i=1}^{n} e_i x_i = 0$

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

$$\frac{(n-2)S^2}{\sigma^2} \sim X^2_{n-2} \;, \quad H_0 : B_1 = 0 \;\; v \;\; H_0 : B_1 \neq 0$$

$$\text{και} \quad B_1 = \frac{\hat{B}_1 - B_1}{\frac{S}{\sqrt{\sum(x_i - \bar{x})^2}}} \sim t_{n-2}$$

$|B_1| \geq t_{\frac{\alpha}{2}, n-2}$ (ΕΒΟ, απορ H_0) ⊛ ΕΒΟ

49.6 $2306 \leftarrow t_{0.025, 8}$

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

Η ανάλυση της διακύμανσης στην παλινδρόμηση

SOS: για διακύμανση σε θεώρημα και υποθ. θέματα στις εξετάσεις.

$$y_i-\bar{y}=\underbrace{(y_i-\hat{y}_i)}_{e_i}+\underbrace{(\hat{y}_i-\bar{y})}_{\text{παλινδρόμηση}}$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(y_i-\bar{y})+2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

$$2\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)(\hat{\beta}_0+\hat{\beta}_1x_i-\bar{y})$$

$$2\sum_{i=1}^{n}(y_i-\bar{y}+\hat{\beta}_1\bar{x}-\hat{\beta}_1x_i)(\bar{y}-\hat{\beta}_1\bar{x}+\hat{\beta}_1x_i-\bar{y})$$

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

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

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

$$\rightsquigarrow \sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

$$\boxed{SS_{tot}=SS_{reg}+SS_{reg}}$$

$n-1=1+(n-2)$ Β.Ε.

$$\sum_{i=1}^{n}(y_i-\bar{y})=0$$

Συντελ. προσδιορισμού: $R^2=\frac{SS_{reg}}{SS_{tot}}100\%$.

$$SS_{reg} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{Y})^2 = \sum_{i=1}^{n}$$

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

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

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

~o~o~o~o~o

Προέλευση μεταβλητότητας	Αθροίσματα τετραγώνων (SS)	Βαθμοί ελευθερίας (Β.Ε)
Παλινδρόμηση	$SS_{reg} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$ 13600	1 1
Υπόλοιπα ή Απόκλιση από παλινδρόμηση	$SS_{reg} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$ 60	$n-2$ 8
Ολική μεταβλητότητα	$SS_{tot} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ 13.660	$n-1$ 9

(* αποτελέσματα με βάση το παράδειγμα 8.0 στο βιβλίο)

Προέλευση μεταβλητότητας	Μέσα τετραγώνων	F-test
Παλινδρόμηση	$MSreg = \frac{SSreg}{1}$ 13.600	$F = \frac{MSreg}{MSreg}$
Υπόλοιπα ή Απόκλιση από το μοντέλο	$MSreg = \frac{SSreg}{n-2} = S^2$ 7.5	
Ολική μεταβλητότητα		

$$\boxed{E(MSreg) = \frac{E(SSreg)}{n-2} = \sigma^2 = E(S^2)}$$

$$\boxed{E(MSreg) = \sigma^2 + B_1^2 \sum (x_i - \bar{x})^2}$$

$$\boxed{\frac{SSreg}{\sigma^2} \sim X_n^2 \quad \text{όταν } H_0 : \text{αληθής } (B_1 = 0)}$$

$E(MSreg) = (SSreg)$ *

$$\hat{B}_1 \sim N\left(B_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \rightarrow \frac{\hat{B}_1 - B_1}{\frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}\right) \sim N(0,1)$$

$$\frac{\hat{B}_1 - B_1}{\frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \overset{H_0}{\sim} N(0,1) \rightarrow \frac{\hat{B}_1^2 \sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2} \underset{B_1=0}{\overset{H_0}{\sim}} X_1^2$$

$$\rightarrow \frac{SSreg}{\sigma^2} \overset{H_0}{\sim} X_1^2.$$

⊛ $E(MS_{reg}) = E(SS_{reg}) = E\left(\hat{\beta}_1^2 \cdot \sum_{i=1}^{n}(x_i-\bar{x})^2\right) =$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 \left[Var(\hat{\beta}_1) + \left(E(\hat{\beta}_1)\right)^2 \right]$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 \left[\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} + \beta_1^2 \right] = \sigma^2 + \beta_1^2 \sum_{i=1}^{n}(x_i-\bar{x})^2$$

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

$$\boxed{\frac{SS_{res}}{\sigma^2} \sim \chi^2_{n-2}} \quad , \quad \boxed{\frac{SS_{reg}}{\sigma^2} \overset{H_0}{\sim} \chi^2_1}$$

$$\frac{\dfrac{SS_{reg}/1}{\sigma^2}}{(SS_{res}/\sigma^2)/(n-2)} \left(= \frac{MS_{reg}}{MS_{res}} \right) \sim F_{1,n-2}$$

F - test

$F = \dfrac{MS_{reg}}{MS_{res}} \overset{H_0}{\sim} F_{1,n-2}$, Κρ. περιοχή: $F \geq F_{\alpha,1,n-2}$

$5.32 = F_{0.05,1,8}$ απορρ. H_0

$(42.6)^2 = 1813.3$

$(2.306)^2 = 5.32$

Συντ. προσδ.: $R^2 = \dfrac{13600}{13660}\, 100\% = 99.56\%$

$e_i = Y_i - \hat{Y}_i$: υπόλοιπα

$\sum_{i=1}^{n} e_i x_i = 0$, $\sum_{i=1}^{n} e_i = 0$, $\frac{\varepsilon_i}{\sigma} \sim N(0,1)$ ή $\frac{e_i}{\sqrt{MS_{reg}}} \xrightarrow{\text{προσ}} N(0,1)$

95% των $\frac{e_i}{\sqrt{MS_{reg}}}$ στο [-2,2]

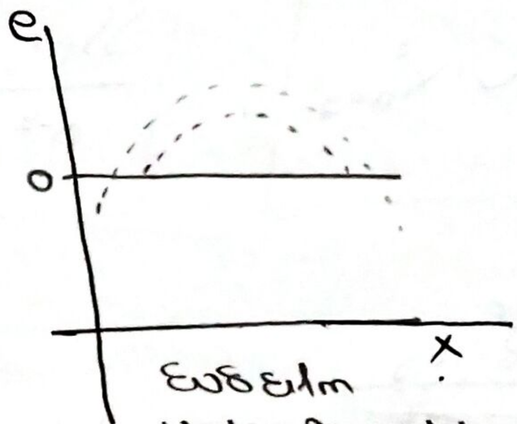
~o~o~o~o~o

▶ Έλεγχος για την καταλληλότητα του μοντέλου

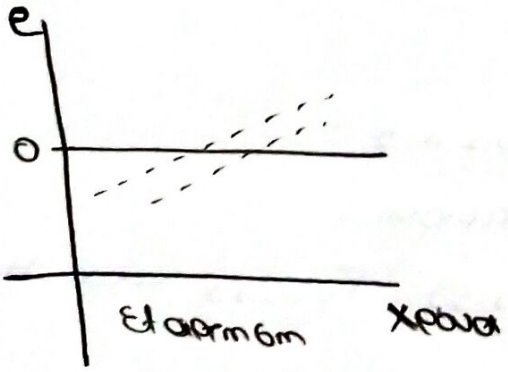
Ανάλυση Υπολοίπων (Residual Analysis)

Γραφικές παραστάσεις

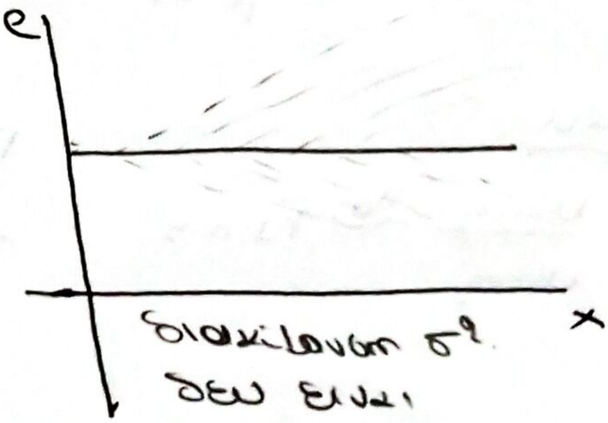
δεν υπάρχουν αποκλ. από την υπόθεση



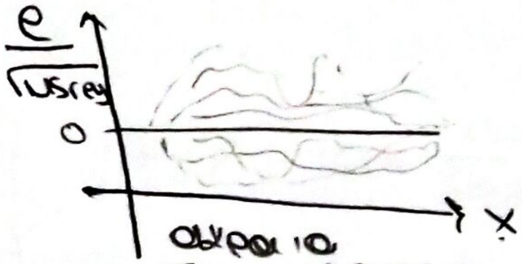
ενδείξη καμπυλόγραμμη σχέση



εξάρτηση από τον χρόνο χρόνος



διακύμανση σ^2 δεν είναι σταθερή



ακραία παρατήρηση

Παραδ. 1)

Λογαριθμικός μετασχηματισμός:

$$Y = \gamma_0 \gamma_1^{x} \cdot \varepsilon \quad , \; E(\varepsilon) = 1, \; Var(\varepsilon) = \sigma^2$$

$$\log_{10} Y = \log_{10} \gamma_0 + x \log_{10} \gamma_1 + \log \varepsilon$$

$$Y' = \beta_0 + x\beta_1 + \varepsilon'$$

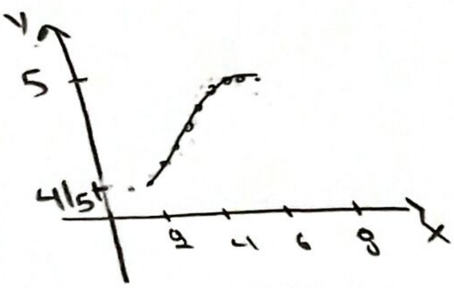
Παραδ. 2)

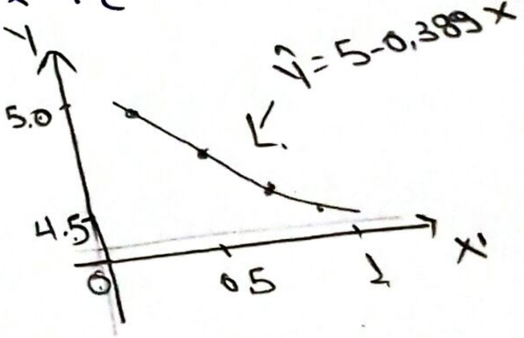
Αντίστροφος μετασχηματισμός:

$$Y = \beta_0 + \frac{\beta_1}{x} + \varepsilon$$

$$(x, y) = \left(\frac{1}{x'}, y\right) \rightsquigarrow Y = \beta_0 + \beta_1 x' + \varepsilon$$

Σχηματικά:





$$\hat{Y} = 5.0 - \frac{0.389}{x} \qquad x' = \frac{1}{x}$$