3/5/17

Απλή Γραμμική Παλινδρόμηση

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , i = 1, \dots, n$

$\varepsilon_i \sim N(0, \sigma^2)$ κ' ανεξ. τμ

$Y_i \sim N(E(Y_i), \sigma^2) \quad , \varepsilon_i = Y_i - E(Y_i)$

$E(Y_i) = \beta_0 + \beta_1 x_i$: άγνωστη συν. παλινδρ.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \quad , \quad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}$$

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \qquad (\hat{Y}_0 = 10.0 + 2.0 x_0)$$

$$e_i = Y_i - \hat{Y}_i \quad , \quad \sum e_i = 0 , \quad \sum e_i x_i = 0$$

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

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

$H_0: \beta_1 = 0 \ v \ H_1: \beta_1 \neq 0$

$\alpha = 5\%$ 2.350 4.26

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-2} \sum_1^n e_i^2 = \frac{1}{n-2} \sum_1^n (Y_i - \hat{Y}_i)^2$$

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{S / \sqrt{\sum (x_i - \bar{x})^2}} \sim t_{n-2} \quad , \quad |T_1| \geq t_{\alpha/2, n-2} \ (\text{για } \beta_1 = 0)$$

$(1-\alpha) \cdot 100\%$ Δ.Ε. για β_1: $\hat{\beta}_1 \pm t_{\alpha/2, n-2} \dfrac{S}{\sqrt{\sum (x_i - \bar{x})^2}}$

Η ανάλυση της Διακύμανσης
στην Παλινδρόμηση

$\bar{\hat{y}} = \bar{y}$

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

$$\sum_1^n (y_i - \bar{y})^2 = \sum_1^n (y_i - \hat{y}_i)^2 + \sum_1^n (\hat{y}_i - \bar{y})^2 + 2\sum_1^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

$\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$

Δεν μας χρειάζεται

$$+2\sum_1^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})$$

$$+2\sum_1^n (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i)(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y})$$

$$= \dots + 2\sum_1^n [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})][\hat{\beta}_1 (x_i - \bar{x})]$$

$$= \dots + 2\hat{\beta}_1 \sum_1^n (x_i - \bar{x})(y_i - \bar{y}) - 2\hat{\beta}_1^2 \sum_1^n (x_i - \bar{x})^2$$

$$= \dots + 2\hat{\beta}_1 \sum (x_i - \bar{x})(y_i - \bar{y}) \underbrace{- 2\hat{\beta}_1 \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \sum (x_i - \bar{x})^2}_{0}$$

Άρα $\sum_1^n (y_i - \bar{y})^2 = \sum_1^n (y_i - \hat{y}_i)^2 + \sum_1^n (\hat{y}_i - \bar{y})^2$

Ολικό άθροισμα τετραγώνων SS_{tot}	Άθροισμα Τετρ. Υπολ. SS_{res}	Άθρ. Τετραγ. Παλινδρ. SS_{reg}

Β.Ε. $n-1 = n-2 + 1$

$\sum (y_i - \bar{y}) = 0$ γιατί μας χρειάζεται το β_1

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

$R^2 = SSreg/SStot$

Βαθμοί Ελευθερίας $n-1 = n-2+1$ γιατί μας χρησιμεύει το $\hat{\beta}_1$

$\sum(y_i-\bar{y})=0$

ΑΝΑΔΙΑ

Προέλευση Μεταβλητότητας	Αθροίσματα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Τετράγωνα
Παλινδρόμηση	$SSreg = \sum^n(\hat{y}_i-\bar{y})^2 = \hat{\beta}_1^2\sum^n(x_i-\bar{x})^2 = 13600$ *	1 1	$MSreg = SSreg = 13600$
Υπόλοιπο Απόκλιση από Παλινδρ.	$SSres = \sum(y_i-\hat{y}_i)^2 = 60$	$n-2$ 8	$MSres = \frac{SSres}{n-2} = s^2 = 7,5$
Συνολική Μεταβολή	$SStot = \sum^n(y_i-\bar{y})^2 = 13660$	$n-1$ 9	$R^2 = \frac{13600}{13660} = 0,9956$ ή $99,56\%$

* Αποτελέσματα για το Παράδειγμα του ΕΒΟ

$E(s^2) = E(MSres) = \sigma^2$ ~ πάντα ✓ + δεν θα ζητηθεί ✓

$E(MSreg) = \sigma^2 + \beta_1^2\sum(x_i-\bar{x})^2$

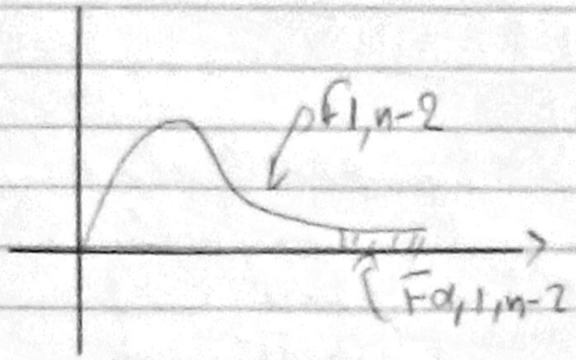
$F = \frac{MSreg}{MSres} \sim ;$

$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2_{n-2} \Rightarrow \frac{SSres}{\sigma^2} \sim \chi^2_{n-2}$

$\frac{SS_{reg}}{\sigma^2} \overset{\beta_1=0}{\sim} \chi^2_1 \leftarrow \frac{SS_{res}}{\sigma^2} \sim \chi^2_{n-2}$

άρα $F = \frac{MS_{reg}}{MS_{res}} = \frac{\frac{SS_{reg}}{\sigma^2}/1}{\frac{SS_{res}}{\sigma^2}/(n-2)} \sim F_{1,n-2}$ όταν H_0 αληθής

Κρ. περ. μεγέθους α: $F \geq F_{\alpha,1,n-2}$



$t^2_{n-2} = F_{1,n-2}$

Συνέχεια του πίνακα

Προέλευση Μεταβλητότητας	Άθροισμα Τετρ.	Β.Ε.	Μέσα Τετρ.	F-test
Παλινδρόμηση				

$4.26^2 = 18.133$
$2.306 = 5.32$

$F = \frac{MS_{reg}}{MS_{res}} = 18.13$

Υπόλοιπο

κρ. περ. β_0, β_1

Απόκλιση από Παλινδρ.

$F \geq F_{\alpha,1,n-2}$

$18.13 > 5.32$

$$E(MS_{reg}) = E(SS_{reg}) = E\left[\hat\beta_1^2 \sum_{i}^{n} (x_i-\bar{x})^2\right] = \sum_{i}^{n}(x_i-\bar{x})^2 \left(E(\hat\beta_1^2)\right)$$

$$= \sum_{i}^{n}(x_i-\bar{x})^2 \left[Var(\hat\beta_1) + \left(E(\hat\beta_1)\right)^2\right]$$

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

$\frac{SS_{reg}}{\sigma^2} \sim ;$ | $\hat\beta_1 \sum(x_i-\bar{x})^2 = SS_{reg}$ | $\hat\beta_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum(x_i-\bar{x})^2}\right) \rightarrow$

$\rightarrow \frac{\hat\beta_1-\beta_1}{\sigma/\sqrt{\sum(x_i-\bar{x})^2}} \sim N(0,1) \rightarrow \frac{(\hat\beta_1-\beta_1)\sqrt{\sum(x_i-\bar{x})^2}}{\sigma} \sim N(0,1)$

$\frac{(\hat\beta_1-\beta_1)^2 \sum(x_i-\bar{x})^2}{\sigma^2} \sim \chi^2_1$ για $\beta_1=0$,

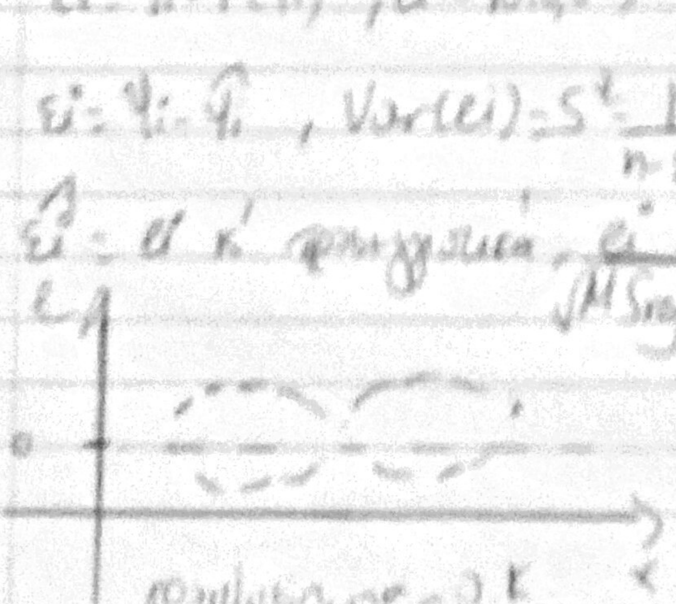
$\frac{SS_{reg}}{\sigma^2} \sim \chi^2_1$

Έλεγχος για την καταλληλότητα του Μοντέλου
Ανάλυση Υπολοίπων (Residual Analysis)

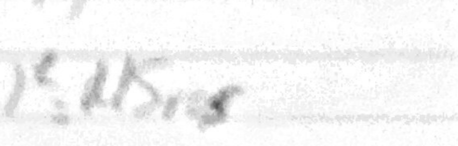
$\varepsilon_i = y_i - E(y_i)$, $\varepsilon_i \sim N(0,\sigma^2)$, $\frac{\varepsilon_i}{\sigma} \sim N(0,1)$

$e_i = y_i - \hat{y}_i$, $Var(e_i) = s^2 = \frac{1}{n-2}\sum_{i}^{n}(e_i-\bar{e})^2 = MS_{res}$

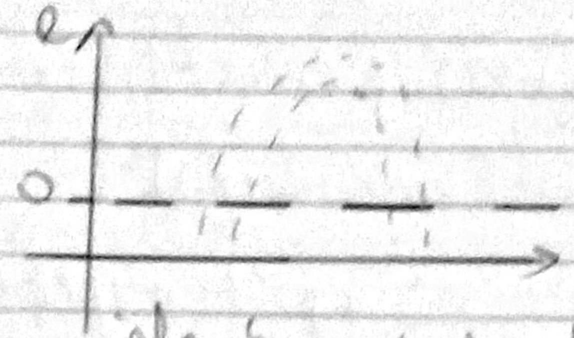
$\hat\varepsilon_i = e_i$ κι' προσεγγιστικά $\frac{e_i}{\sqrt{MS_{res}}} \sim N(0,1)$



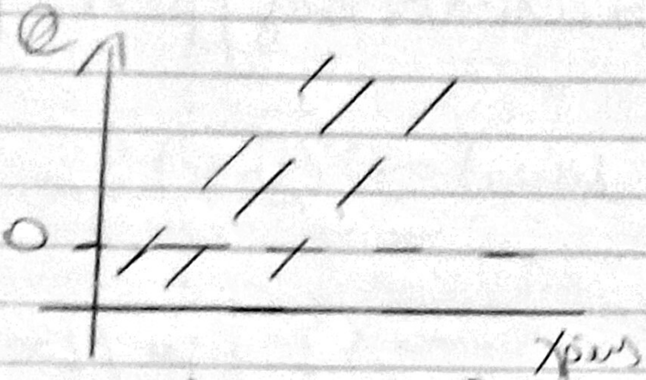
Γράφημα των e_i ως προς x (σταθερή διακύμανση)



95% των $\frac{e_i}{\sqrt{MS_{res}}}$ είναι $(-2,2)$



Ύπαρξη καμπυλόγραμμης σχέσης



μη σταθερή διακύμανση