29-1-2025 Διαγνωστικοί Έλεγχοι [Ανάλυση Καταλοίπων (Residual Analysis)]

Υποθέσεις γραμμικού μοντέλου

$$Y = b_0 + b_1 X_1 + \dots + b_k X_k + \varepsilon$$

$\varepsilon \sim N(0, \sigma^2)$

ε_j, $j = 1, 2, \dots$, ανεξάρτητα

1. Κανονικότητα
2. Ανεξαρτησία
3. Ομοσκεδαστικότητα : $Var(\varepsilon_j) = \sigma^2$ = σταθερή

---

(1) Δείγμα

j	x_1	x_2	$\dots$	x_k	y
1	x_{11}	x_{21}		$\vdots$	y_1
2				$\vdots$	$\vdots$
$\vdots$				$\vdots$	$\vdots$
n	x_{1n}	x_{2n}		$\vdots$	y_n

(2) Αν έχω δείγμα $(e_1, \dots, e_n)$ για να ελέγξω αν ακολουθούν N

: έλεγχος Kolmogorov-Smirnov

: Shapiro-Wilk.

$\varepsilon_1, \dots, \varepsilon_n$?

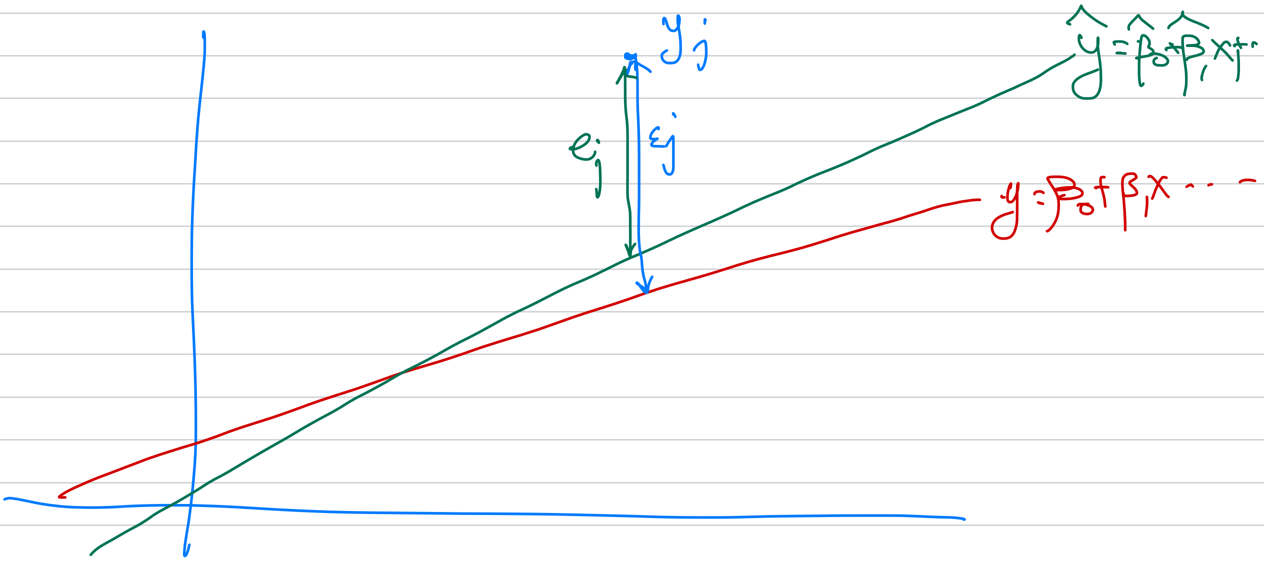
$$\varepsilon_j = y_j - (\beta_0 + \beta_1 x_{1j} + \dots + \beta_k x_{kj}) = \text{? (άγνωστα)}$$

γνωστά

άγνωστα

(3) Μπορούμε να "ανακτήσουμε" να εκτιμήσουμε τα ε_j?

$$e_j = \hat{\varepsilon}_j = y_j - (\hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \dots + \hat{\beta}_k x_{kj}) = y_j - \hat{y}_j$$



e_j : κατάλοιπο (residual) της παρατήρησης j

ΟΜΩΣ: (1) $e_j \sim \mathcal{N}(0, \sigma^2_{\varepsilon_j})$ όχι ομοσκεδαστικά;

(2) $\sum e_j = 0$ πάντα όχι ανεξάρτητα

Μετασχηματισμοί Καταλοίπων

(1) Standardized residuals

$$\tilde{e}_j = \frac{e_j}{\hat{\sigma}_{\varepsilon_j}}$$

$Var(\tilde{e}_j) = 1$

όμως όχι ανεξάρτητα

(2) Studentized residuals

$$e'_j = \frac{\tilde{e}_j}{\sqrt{1-h_j}} = \frac{e_j}{s_j\sqrt{1-h_j}}$$

h_j = leverage

③ Jackknife residuals

$$r_j = \frac{e_j}{S_{-j}\sqrt{1-h_j}}$$

ανεξάρτητα $\sim t \Longrightarrow N$ για μεγάλο n

S_{-j} : τυπ. απόκλιση του $y_j - \hat{y}_j$ σε ένα δείγμα από το οποίο έχει αφαιρεθεί η παρατήρηση j

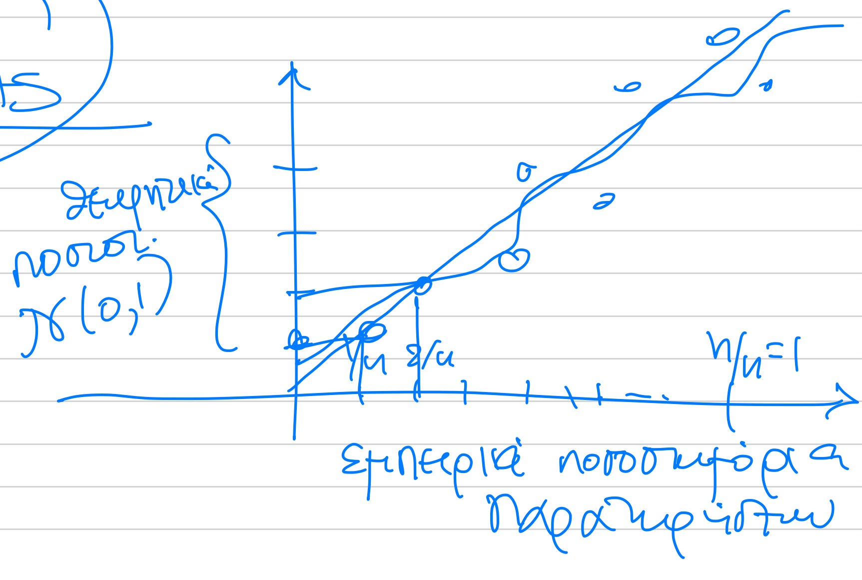
(R, Stata) → studentized ⟹ "standard"
→ jackknife → "student"

Διαγνωστικοί έλεγχοι

① Κανονικότητα

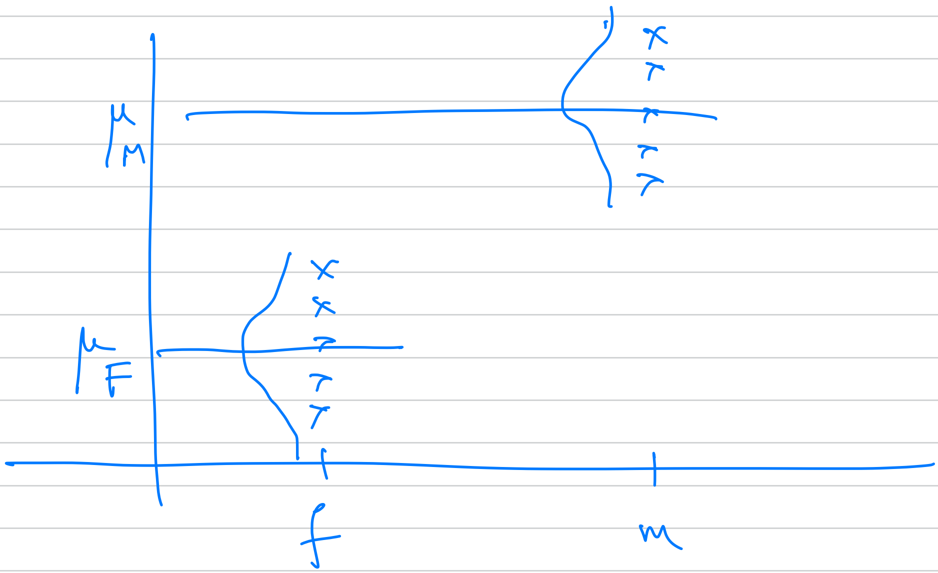
Kolmogorov-Smirnov ή Shapiro Wilk } $\boxed{\text{στα } r_j \text{ !!}}$

qq plots

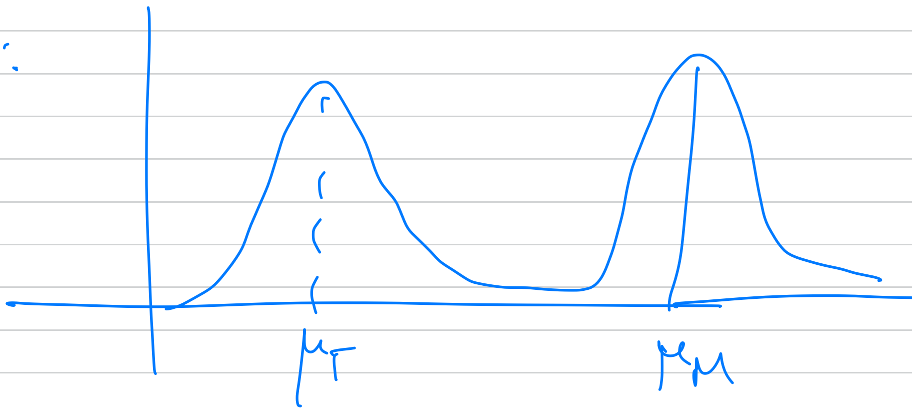


$y_i \sim N(b_0 + b_1 x_1 - \dots, \sigma^2)$ όχι με την ίδια μέση τιμή !?

πχ. $y = b_0 + b_1 \cdot X_{male}$



Ιστόγραμμα y :

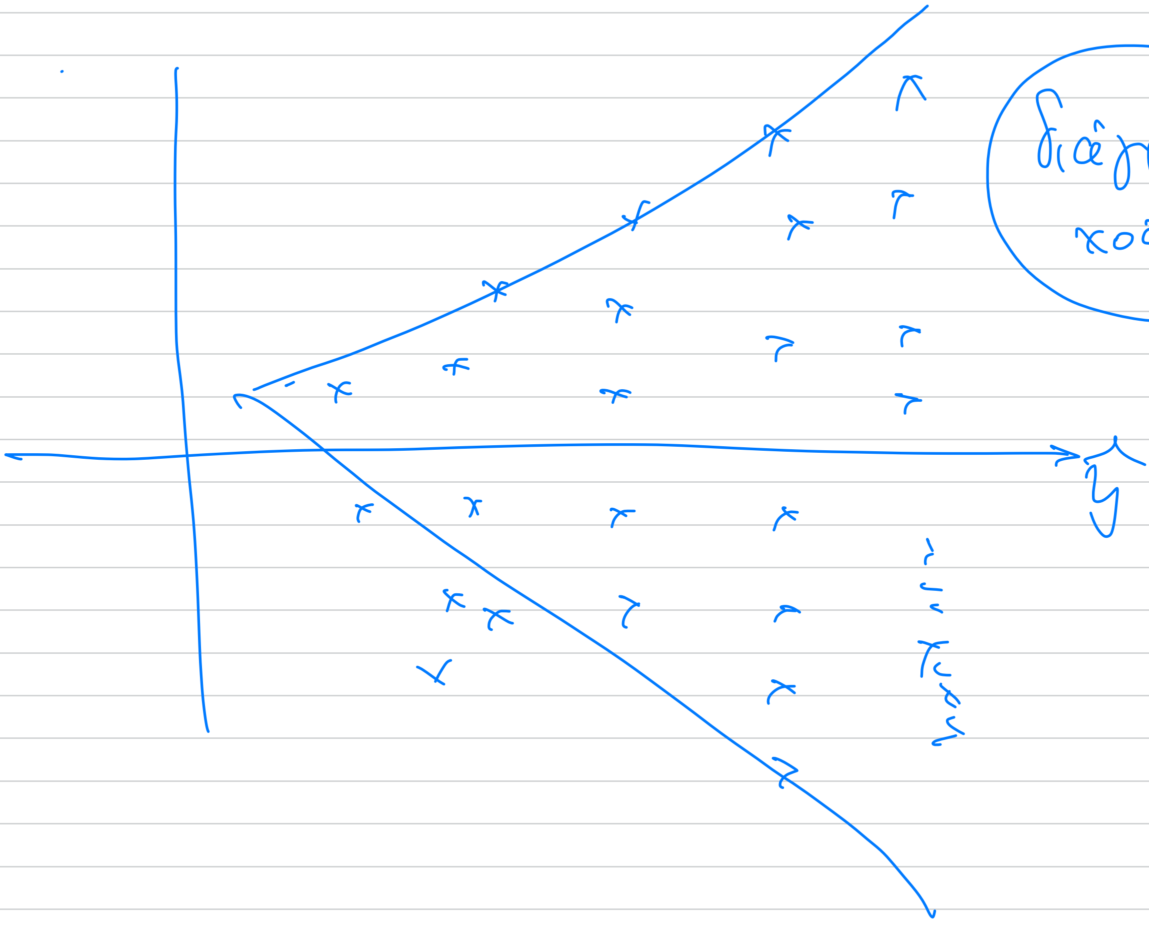


② Ομοσκεδαστικότητα

Διάγραμμα καταλοίπων

$Cov(e_j, \hat{y}_j) = 0$

$Cov(r_j, \hat{y}_j) = 0.$



διάγραμμα χοάνης

Bartlett test για ετεροσκεδαστικότητα

Ανεξαρτησία

1. Παραλλαγή μεθόδου LSE
weighted LSE

2. Μετασχηματισμός Y

$Y_1 \sim \sqrt{Y}$ π.χ.

$$\sqrt{Y} = b_0 + b_1 X + \ldots$$

Προσοχή οι ερμηνείες των b χάνονται:

$$E(\sqrt{Y}) = b_0 + b_1 X.$$

b_1 : ρυθμός μεταβολής της $E(\sqrt{Y})$ ως προς X.

δε λέει τίποτα για τη μεταβολή της $E(Y)$ ⇐ !

$$E\left(\sqrt{Y}\right) \neq \sqrt{E(Y)}$$