

31-1-2025

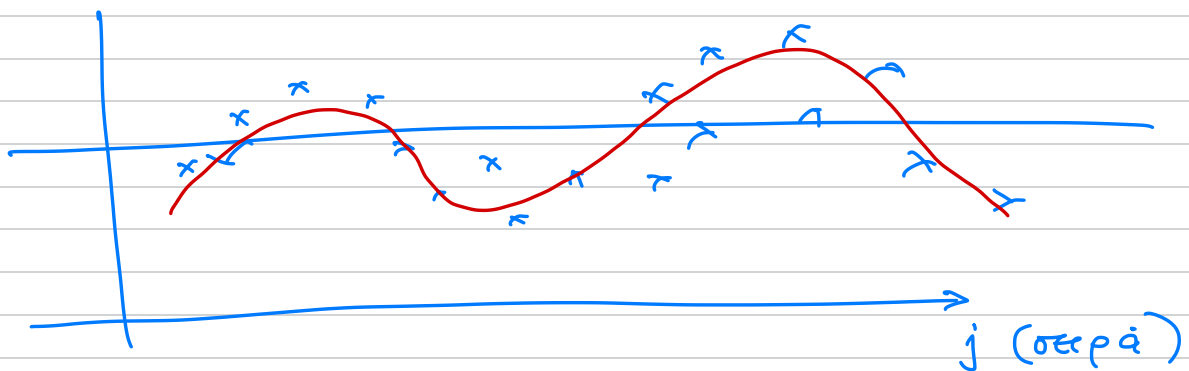
Έλεγχος κανονικότητας καταλοίπων ✓
" ομοσκεδαστικότητα ✓

Έλεγχος αυτοσυσχέτισης καταλοίπων

$\varepsilon_1, \varepsilon_2, \dots$ ανεξάρτητες $\Rightarrow \hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots$ ανεξάρτητες

Παραβιάζεται συχνά σε περίπτωση χρονοσειρών
(οι παρατηρήσεις είναι σε μια χρονική σειρά
που ενέχει συσχέτιση)

Αυτοσυσχέτιση: συσχέτιση μεταξύ παρατηρήσεων
σε μια χρονοσειρά.



Έλεγχος Durbin-Watson για αυτοσυσχέτιση
καταλοίπων

d : Durbin-Watson statistic

$$0 \leq d \leq 4$$

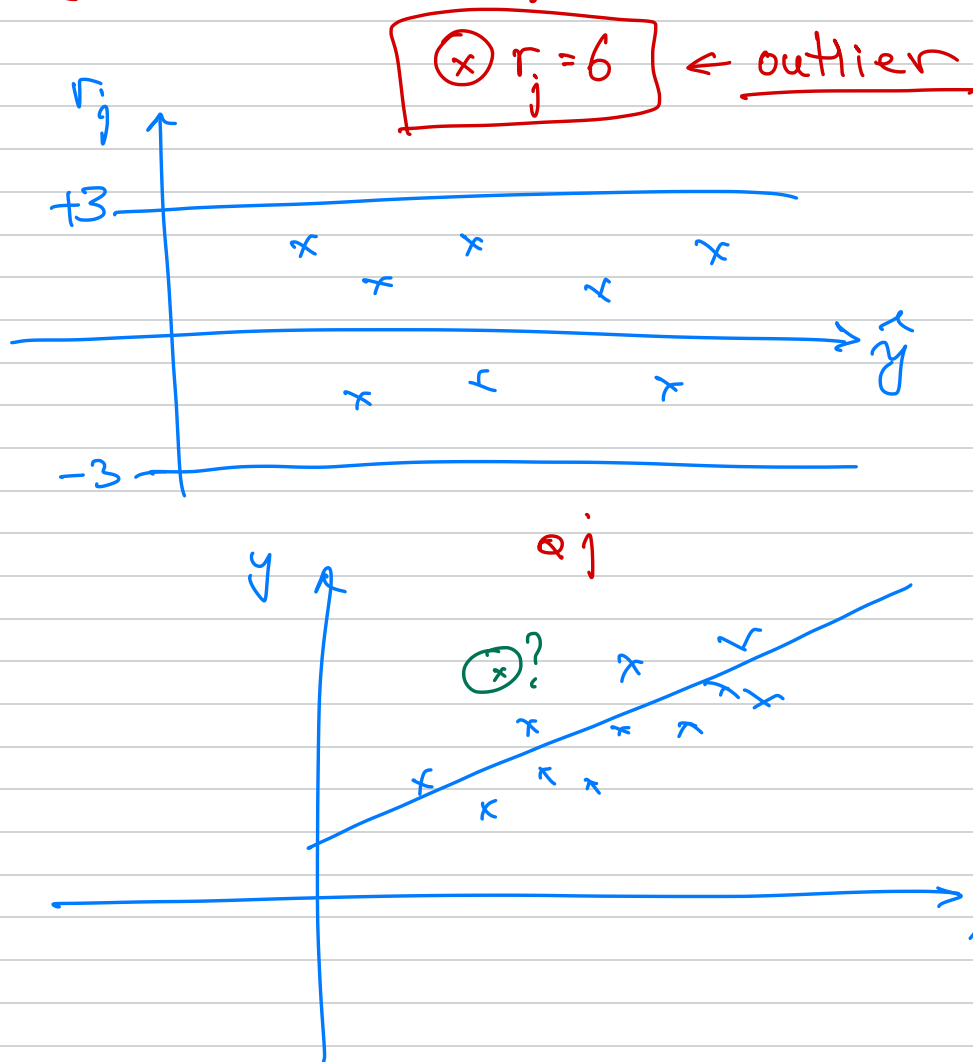
$d \approx 2$: δεν υπάρχει αυτοσυσχέτιση

$d < 2$: " αρνητική αυτοσυσχ.

$d > 2$: " θετική αυτοσυσχ.

(p-value : προσεγγιστικά)

Εκτός κ' Ενδιαστικές Παρατηρήσεις (Outliers, Influential Observations).



① κριτήριο outlier

② Αιτιολόγηση

↓ α) Ελεγχος σωστής καταγραφής

β) Ελεγχος κριτηρίων αποκλεισμού

γ) Επανεπαξιολογείται η ανάλυση χωρίς outliers

Παρουσιάζονται και οι δύο αναφορές

Κριτήριο χαρακτηρισμού ως outlier

(στατιστικός έλεγχος βασίζεται στο $|r_i|$)

Πινάκες αραιών τιμών (εξαρ: n, k)

μεγ. δείγματος
μεγέθος ποσότητας

Bonferroni correction

Πολλά έλεγχα πάνω στο ίδιο μόνιμο/δεδομένα

Έλεγχος 1, α : $P(\sigmaφαλμ) = \alpha$

" 2 : $P(\sigmaφαλμ) = \alpha$

$P(\text{κάνα σφάλμα στους δύο ελέγχους}) \approx$ (ανεξάρτητοι έλεγχοι)

$$P(\text{κ' τα δύο σωστά}) \approx (1-\alpha)(1-\alpha) = (1-\alpha)^2$$

$$P(\text{συνάξιμων εα σφαλμ}) = 1 - (1-\alpha)^2$$

Αν κάνουμε m ελέγχους

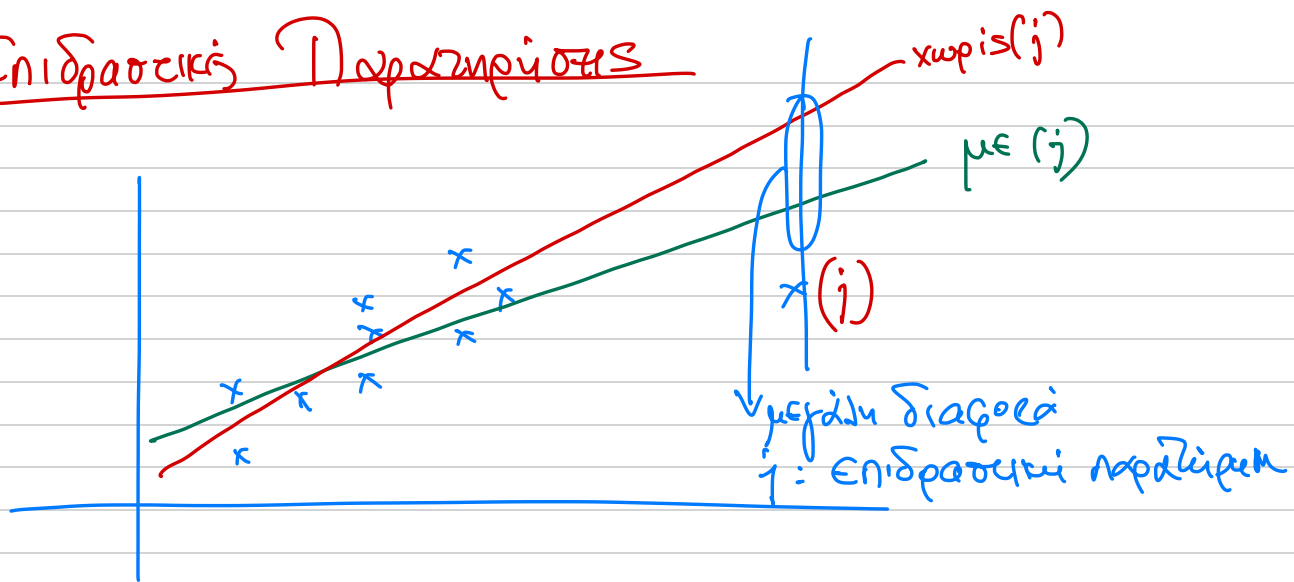
$$P(\text{συνά. ένα σφαλμ}) \approx 1 - (1-\alpha)^m \approx m \cdot \alpha$$

$$\text{Όταν } \alpha \approx 0, m \text{ μεγάλο } (1-\alpha)^m \approx 1 - m \cdot \alpha$$

$$\text{Αν θέλουμε } m\alpha = 0.05 \Rightarrow \alpha = \frac{0.05}{m}$$

↓
Διόρθωση Bonferroni

Επιδραστικές Παρατηρήσεις



Κριτήρια

- ① Leverage (Μόχλευση) : h_j
 $0 \leq h_j \leq 1$

Πινακας αερίων εφών

- ② Απόσταση Cook (Cook's distance)
Εξήγηση τετατοποίηση $\hat{b}$

Κριτήρια Σύγκρισης - Επιλογής Μοντέλου

Full model : $Y = b_0 + b_1 X_1 + \dots + b_k X_k$
 $k \ll n$ $\left(\begin{array}{l} d_{\text{fer}} \approx 10 \cdot k \\ (n \approx 11 \cdot k) \end{array} \right)$

Partial Model $Y = b_0 + b_1 X_1 + \dots + b_p X_p$ ($p < k$)

① H_0 : Partial $\approx$ Full H_1 : όχι

Ελεγχος F για μέτρο του μεγέθους

② Adjusted- R^2

Τροποποιημένος τύπος για R^2 (adj- $R^2 < R^2$)
Παίρνει υπόψη k, n

③ C_p -Mallows statistic

$$C_p = \frac{SSE(p)}{MSE(k)} - [n - 2(p+1)]$$

Αν τα δύο μοντέλα είναι σχετικά
ισοδύναμα τότε $MSE(p) \approx MSE(k)$

Αν όχι είναι $MSE(p) > MSE(k)$

$$\text{Ομω } MSE(p) = \frac{SSE(p)}{n-p-1}$$

$$SSE(p) = MSE(p) \cdot (n-p-1)$$

$$C_p = \frac{MSE(p)}{MSE(k)} (n-p+1) - [n - 2(p+1)]$$

Αν ισοδύναμα $C_p \approx n - (p+1) - [n - 2(p+1)]$
 $= p+1$

Αν partial υποδείκτης $C_p > (p+1)$

Σύγκριση γενικών (οχι nested) μοντέλων

Κριτήρια Πληροφωρίας

AIC : Akaike Information Criterion

BIC : Bayesian " "

$$\underline{AIC} = 2(k - \log L)$$

L : πιθανοφάνεια των b στο μοντέλο

Μικρές τιμές των $\underline{AIC}$ $\rightarrow$ αξιολογώ
μοντέλο

Σύγκριση επιρροής

k μεταξύ οχι nested μοντέλων.

Stepwise Regression

Full Model $Y \sim (X_1, X_2, X_3, \dots, X_k)$

Πόσα διαφορετικά μοντέλα υπάρχουν;

$$2 \times 2 \times \dots \times 2 = 2^k$$

Karvonen : Διεφομετρικός μεταβλητός
λεξιδιαφαινόμενος (για να έχουμε
εξήγηση)

Interaction μαζί με main effects.

① Forward

①
$$\begin{aligned} Y &= b_0 + b_1 X_1 \\ Y &= b_0 + b_1 X_2 \\ &\vdots \\ Y &= b_0 + b_1 X_k \end{aligned}$$
 } επιλέγω αυτό με
το μικρότερο p-value
έως το X_2

②
$$\begin{aligned} Y &= b_0 + b_1 X_2 + b_2 X_1 \\ &\quad + b_2 X_3 \\ &\quad + b_2 X_4 \leftarrow \text{μικρότερο } \underline{p=0.20} \\ &\quad \vdots \\ &\quad + b_2 X_k \end{aligned}$$
 μικρότερο p
Αν $p < p_{\text{enter}}$
πρόσθετο

② Backward

③ Stepwise