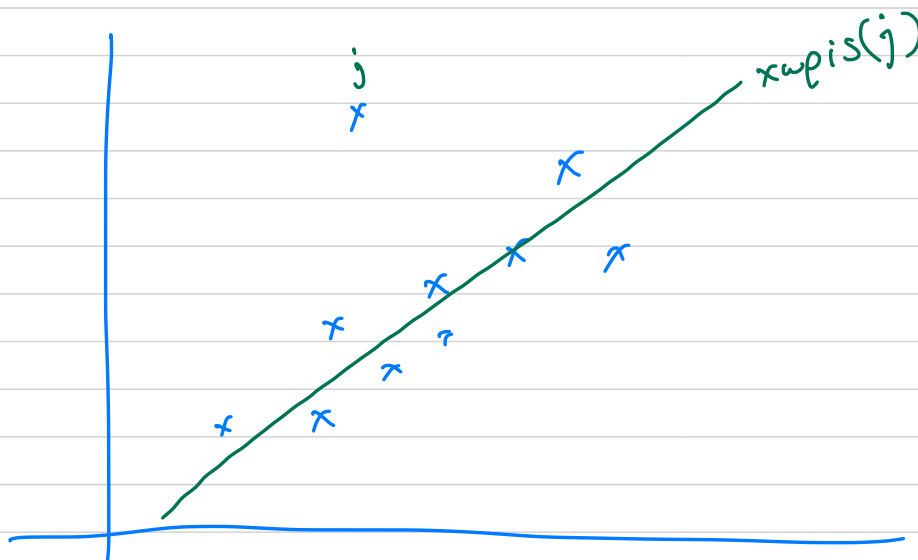
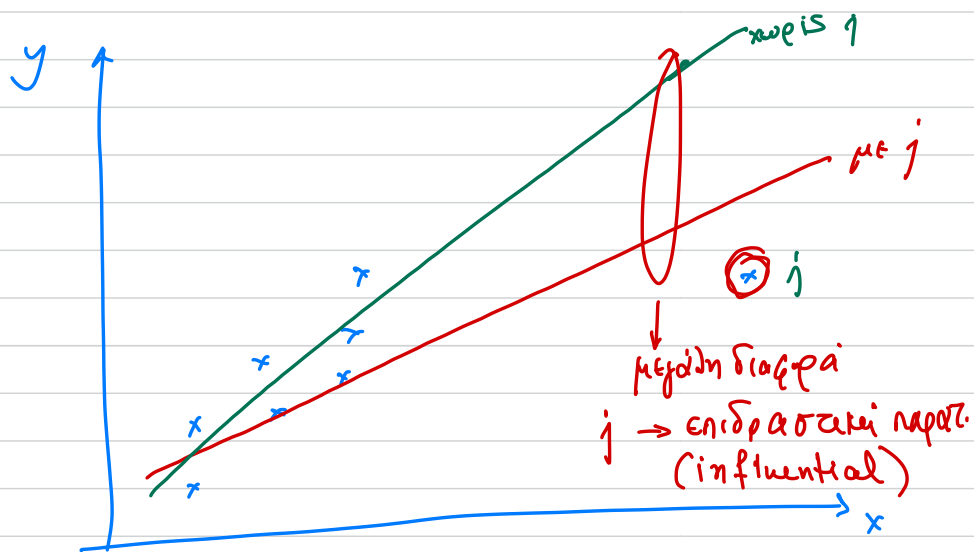


23-1-2026

Επιδραστική παρατηρήσεις (Influential Observation)



Κριτήρια

① Leverage : h_j $0 \leq h_j \leq 1$

Πινακας αραίων τιμών $h_j >$ κριτήριο
αμφι

② Απόσταση Cook (Cook's distance)

Εκτίμηση μεταβολής των $\hat{\beta}$
που προκαλεί η παρατήρηση j

π.χ. $\alpha = 0.05$ $n = 50$ $k = 5$ 16.41

$n - k - 1 = 44$

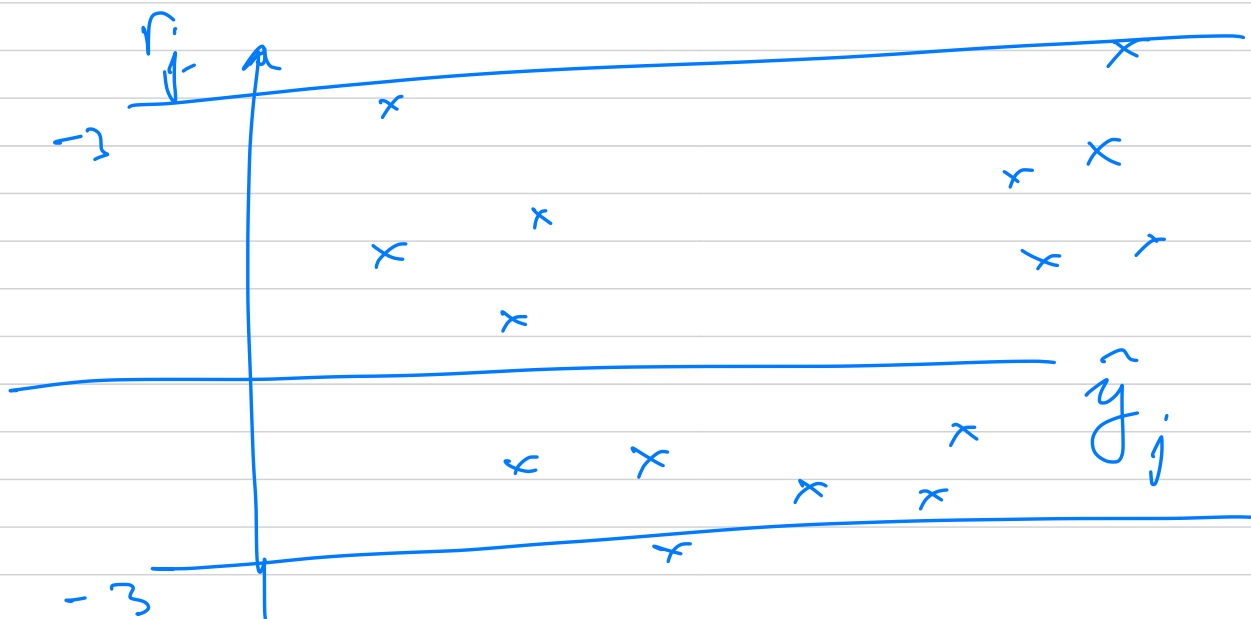
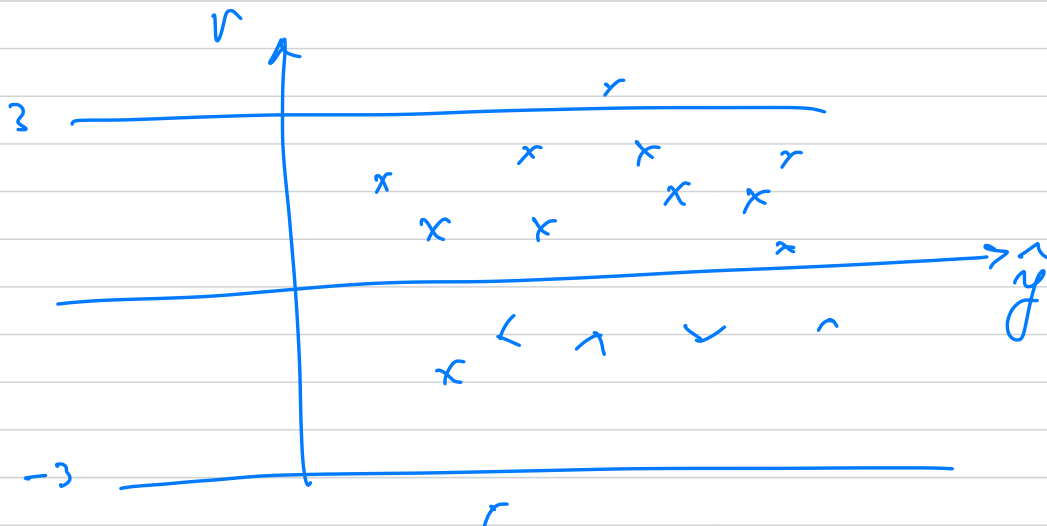
Από πίνακα: μια παρατήρηση j είναι επιδραστική

αν $44 \times d(j) > 16.41$

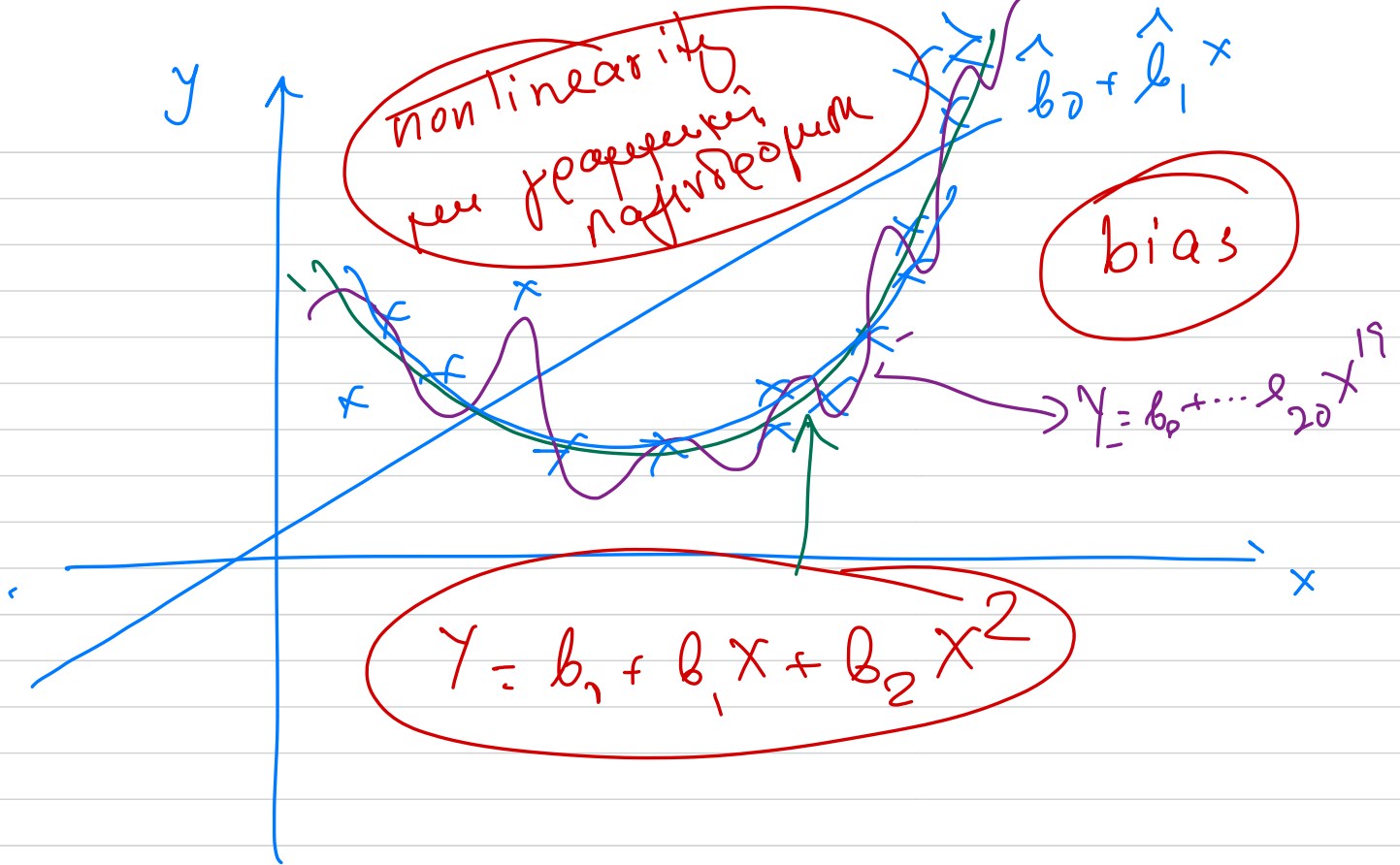
$\Leftrightarrow d(j) > \frac{16.41}{44}$

κρίσιμη τιμή
για Cook.

Διαγνωστικά Ελέγχοι



$\hat{y} = b_0 + b_1 X$



Κριτήρια Σύγκρισης Μοντέλων

Μέθοδοι Επιλογής Μοντέλων (model selection)

Full Model : $Y = b_0 + b_1 X_1 + \dots + b_k X_k$

$$k \ll n$$

$$df_{er} = n - k - 1$$

$$\left(\begin{array}{l} \text{Εμπειρικά: } df_{er} \approx 10k \\ n \approx 11k \end{array} \right)$$

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

Κριτήρια

① Ftest for part of model

$$H_0: \text{Partial} \approx \text{Full} \quad H_1: \text{Full} \text{ καλύτερο}$$

② Adjusted R^2

Τροποποιημένο R^2 ως προς R^2

ναίμεν υπόψη k, n

$$adj-R^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} \quad \begin{array}{l} \leftarrow df_{er} \text{ (0-100\%)} \\ \leftarrow df_{er} \end{array}$$

$$\text{An } p \ll n \Rightarrow n-1 \approx n-p-1$$

$$\Rightarrow \text{adj-}R^2 \approx R^2$$

$$\text{Doo to } p \rightarrow n \quad \frac{n-1}{n-p-1} \gg 1$$

$$\Rightarrow \underline{\text{adj-}R^2 < R^2}$$

1) An $\text{adj-}R^2 < R^2 \Rightarrow$ ειδική όα
 το ποσοστό
 είναι πολύ
 μεγάλο
 (overfitting)

$\text{adj-}R^2$: μέτρο σύγκρισης full-partial

3) Στατιστικό C_p - Mallows

Full (k)

Partial (p)

$$\frac{\text{SSE}(p)}{\text{MSE}(k)} \approx \frac{\text{SSE}(p)}{\text{SSE}(k)} \cdot df_{er}(\text{full})$$

Partial $\longleftrightarrow$ Full
 λίγο $\longleftrightarrow$ λίγο

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

Av Full Model $\approx$ Partial Model

$$MSE(p) \approx MSE(k)$$

$$\Rightarrow MSE(p) = \frac{SSE(p)}{n-(p+1)} \approx MSE(k) \Rightarrow$$

$$\Rightarrow \frac{SSE(p)}{MSE(k)} \approx n - (p+1)$$

$$\Rightarrow C_p \approx \cancel{n - (p+1)} - [\cancel{n - 2(p+1)}] = p+1$$

$$\text{Ozar Partial} \approx \text{Full} \Rightarrow C_p \approx p+1$$

Ozar partial unodtetoza zar full

$$\Rightarrow C_p > p+1$$

Β Γενικά κριτήρια σύγκρισης
(nested ή όχι)

Κριτήρια Πληροφορίας	AIC	Akaike Inform Criterion
	BIC	Bayesian Information Criterion

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

L : η πιθανοφάνεια του δείγματος
κύρω από us εφαρμογής LSE του b .

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

$$Y \sim N(b_0 + b_1 x, \sigma^2)$$

$$\text{Πιθανοφάνεια: } \prod_{j=1}^n f(y_j; b) = L(b)$$

$$f(y_j) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y_j - b_0 - b_1 x_j)^2}{2\sigma^2}}$$

$$L(b) = \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{SSE(b_0, b_1)}{2\sigma^2}}$$

$$\max L \iff \min SSE$$

$$\hat{b}_{MLE} = \hat{b}_{LSE}$$

$L(\hat{b}) \leftarrow$ μέγιστη πιθανοφάνεια
του δείγματος

"Καλό" Μοντέλο $\iff$ AIC μικρό

Μέθοδοι Επιλογής Μεταβλητών

Full Model $Y = b_0 + b_1 X_1 + \dots + b_k X_k$
 $\{X_1, \dots, X_k\}$

Πόσα nested models υπάρχουν;

Πόσα υποσύνολα του $\{X_1, \dots, X_k\}$ υπάρχουν;

$$\left. \begin{array}{l} X_1 : \text{vai / oxi} \\ X_2 : \text{vai / oxi} \\ \vdots \\ X_k : \text{vai / oxi} \end{array} \right\} 2 \times 2 \times \dots \times 2 = 2^k$$

$$Y = b_0 \quad (\text{μυθικός})$$

$$Y = b_0 + b_1 X_1 + \dots + b_k X_k \quad (\text{full})$$

Μέθοδος Stepwise για επιλογή μεταβλητών

① Forward method ($X_1, \dots, X_k$) υποψιαστές
μεταβλητές

α) Αν είναι μόνο ένα μονοπαριαστικό ($p=1$)
τότε θα είναι το καλύτερο

	Κριτήριο	
	State p-value	R AIC
$Y = b_0 + b_1 X_1$	P_1	
$Y = b_0 + b_1 X_2$	P_2	
$\vdots$		
$Y = b_0 + b_1 X_k$	P_k	
	t-test for b_1	

Επιλέγω το μικρότερο ή το μικρότερο p-value.

Έχω επιλέξει μια παράμετρο Penter

Αν το μικρότερο p-value ^(nx.1) $< P_{enter}$

$\Rightarrow$ η X_1 μπαίνει στο μοντέλο

$$Y = b_0 + b_1 X_1$$

β) Αν προσέθετα μια επιπλέον ποικιλία ήρω;

$Y = b_0 + b_1 X_1$		p-value t-test νέας μεταβλητής (b_2)
X_2	$Y = b_0 + b_1 X_1 + b_2 X_2$	↑ βρίσκω το min P
X_3	$Y = b_0 + b_1 X_1 + b_2 X_3$	
$\vdots$	$\vdots$	
X_k	$Y = b_0 + b_1 X_1 + b_k X_k$	

Αν το min p-value < $P_{enter} \Rightarrow$ " X_3 μπαίνει.
(εδώ X_3)

νέο $Y = b_0 + b_1 X_1 + b_2 X_3 :$

συνεχίζουμε μέχρι

σε ένα βήμα $\min P > P_{enter}$

② Backward

Ξεκινάμε από full model

Παράμετρος P_{remove}

Αν αφαιρούσαμε μια ποια θα ήταν;

Αυτή που έχει το max p στο t -test

αν αυτό είναι $\max p > P_{remove}$

η αντιστοίχη μεταβλητή φύγει

Στο ποια με $k-1$ μεταβλ. το ζωντανό

επαναλαμβάνουμε

μέχρι όφες οι μεταβλές : $p < P_{remove}$.

③ Stepwise

P_{enter} , P_{remove}

Σε κάθε βήμα $\rightarrow$ forward (επιλέγει μια να μπει)

$\rightarrow$ k' backward

Προσοχή

Αν $P_{enter} = 0.07$
 $P_{remove} = 0.05$ }

Τότε αν έχεις σε κάποιο βήμα

μια X να έχει $p = 0.06$

Καύρας

Penter $\leq$ Premove