

10-12-2025

Απλό Γραμμικό Μοντέλο (Μονοπαραγοντικό)

Y = εξαρτημένη μεταβλητή (Ποσοτική/Κλίμακας)

X = ανεξάρτητη " (Ποσοτική/θα γράφεται)

$E(Y|X=x) = f(x)$: συνάρτηση παλινδρόμησης

Ισοδύναμα $Y|X=x = f(x) + \varepsilon$

$\varepsilon : E(\varepsilon)=0, \text{Var}(\varepsilon)=\sigma^2$

Απλό γραμμικό

τυχαία διακύμανση

π.χ.

$$Y = \underbrace{b_0 + b_1 X}_{f(x)} + \varepsilon$$

γραμμικό μοντέλο (γραμ.-β)
γραμμική παλινδρ. (γραμ.-X)

$$Y = b_0 + b_1 X + b_2 X^2$$

γραμ. μοντέλο
πολυωνυμική παλινδρόμηση

$$Y = b_0 + b_1 \sin X$$

γραμ. μοντέλο

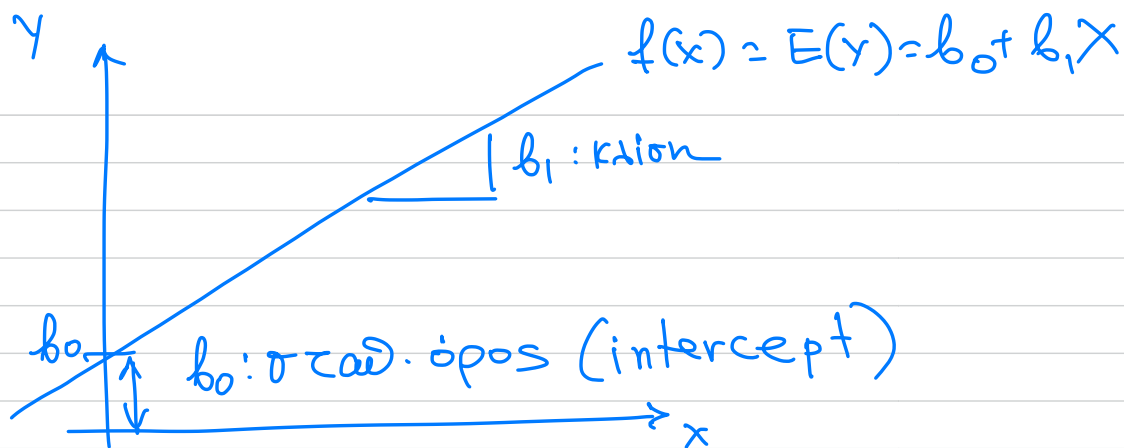
$$Y = b_0 + e^{b_1 X}$$

μη γραμμικό μοντέλο X

"Μηδενικό Μοντέλο"

$$Y = b_0 + \varepsilon$$

$E(Y) = b_0$ άγνωστη



αν $b_1 = 0 \Rightarrow E(Y) = b_0 \rightarrow$ $\nexists$ συσχέτιση μεταξύ X, Y

$$H_0 : b_1 = 0 \quad \text{vs} \quad H_1 : b_1 \neq 0$$

Δεδομένα

X	Y
x_1	y_1
x_2	y_2
$\vdots$	
x_n	y_n

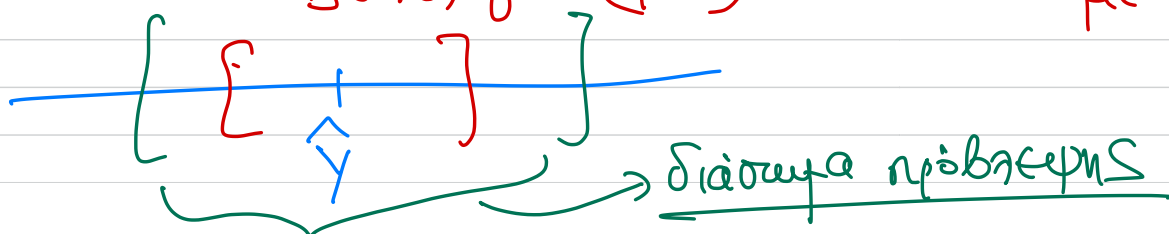
$\Rightarrow$ Εκτιμήσεις
 $\hat{b}_0, \hat{b}_1, \hat{\sigma}^2$

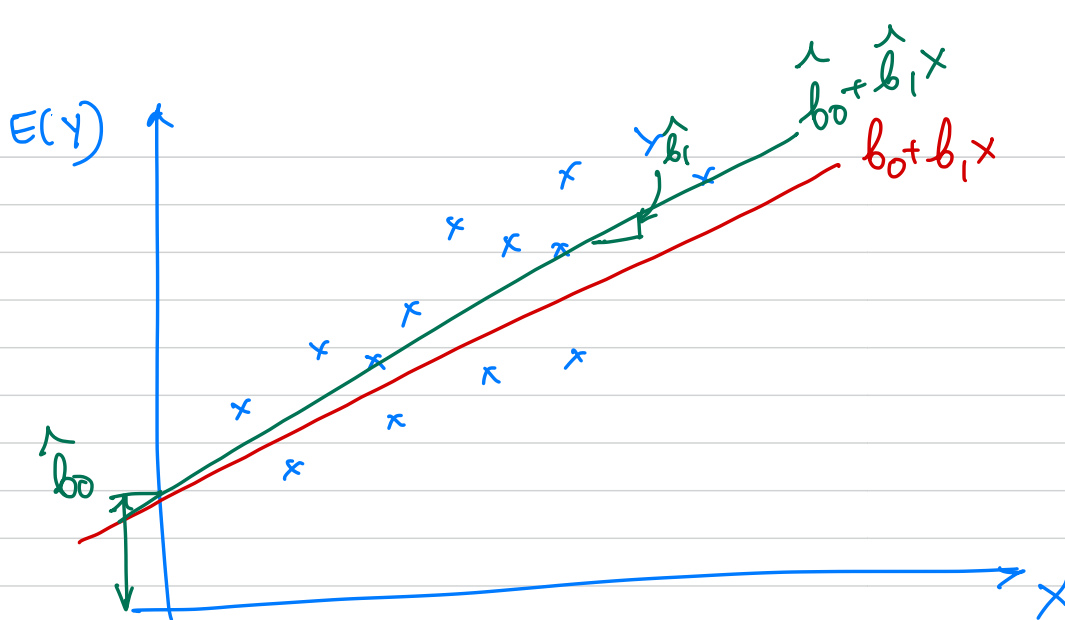
$$\hat{Y} = \hat{b}_0 + \hat{b}_1 \cdot X = \hat{f}(x)$$

(Εκτιμημένη)
συνάρτηση παλινδρόμησης

$\hat{Y}$: πρόβλεψη ως Y για $X=x$ $\leftarrow$ μεμονωμένα άτομα με $X=x$
εκτίμηση ως $E[Y|X=x]$ $\rightarrow$ εκτίμηση για μέση τιμή όλων των ατόμων με $X=x$

$\Delta \pm 95\%$ για $E(Y|X=x)$

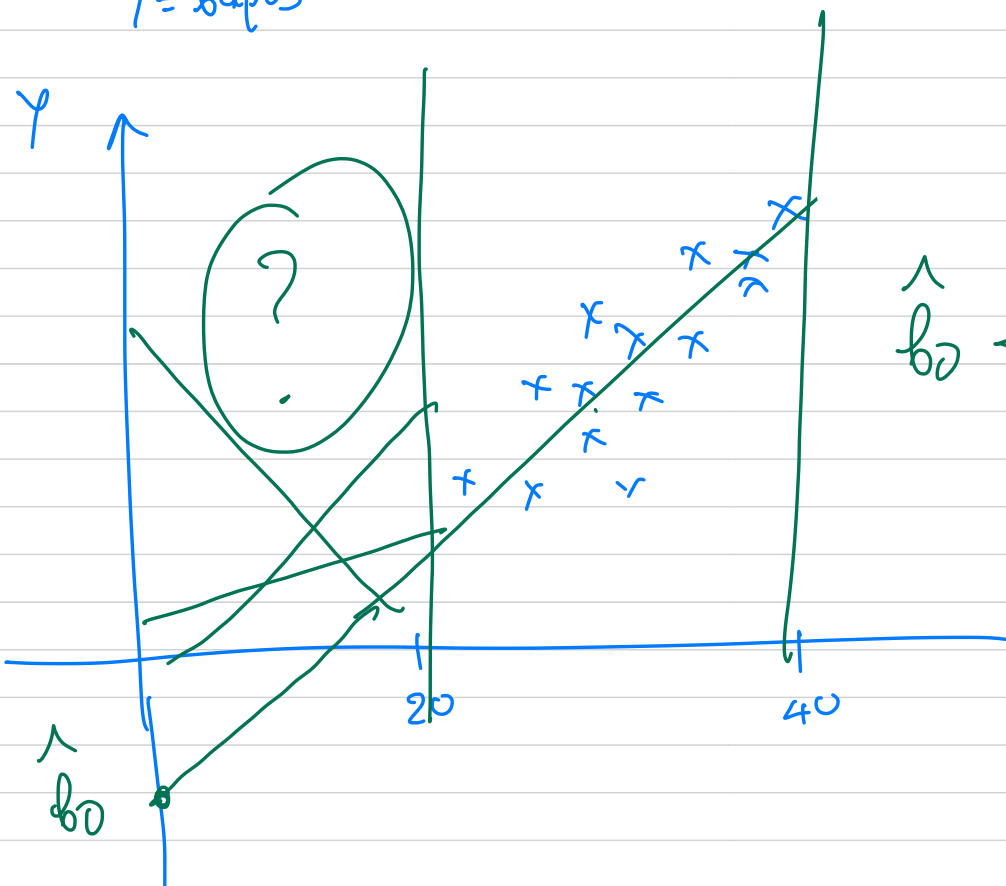




① Ερμηνεία $\hat{b}_1$: κλίση : Μεταβολή της $E(Y)$ ανά μονάδα αλλαγής της X

$$\hat{b}_0 = E(Y|X=0) \quad !! \text{ προσοχή!}$$

παράβ. $X = \text{age}$ στο δείγμα άτομα ηλικίας 20 ~ 40
 $Y = \text{βάρος}$



Γενικός κανόνας

Ερμηνείες/εκτιμήσεις/προβλέψεις για το Y
είναι ασφαλές μόνο για τιμές της X μέσα
στα όρια του δείγματος

Εκτός ορίων : extrapolation (αναπορεύεται)

② Ερμηνεία σχέσης μεταξύ X, Y

$$\hat{E}(Y) = \hat{b}_0 + \hat{b}_1 \cdot X$$

Αύξηση της X κατά 1 μονάδα προκαλεί
μεταβολή της $E(Y)$ κατά $\hat{b}_1$

ή η μεταβολή της $E(Y)$ οφείλεται στη μεταβολή X

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X \quad \Leftrightarrow \quad \hat{X} = \hat{\beta}_0 + \hat{\beta}_1 Y$$

π.χ. X = χρόνος άσκησης

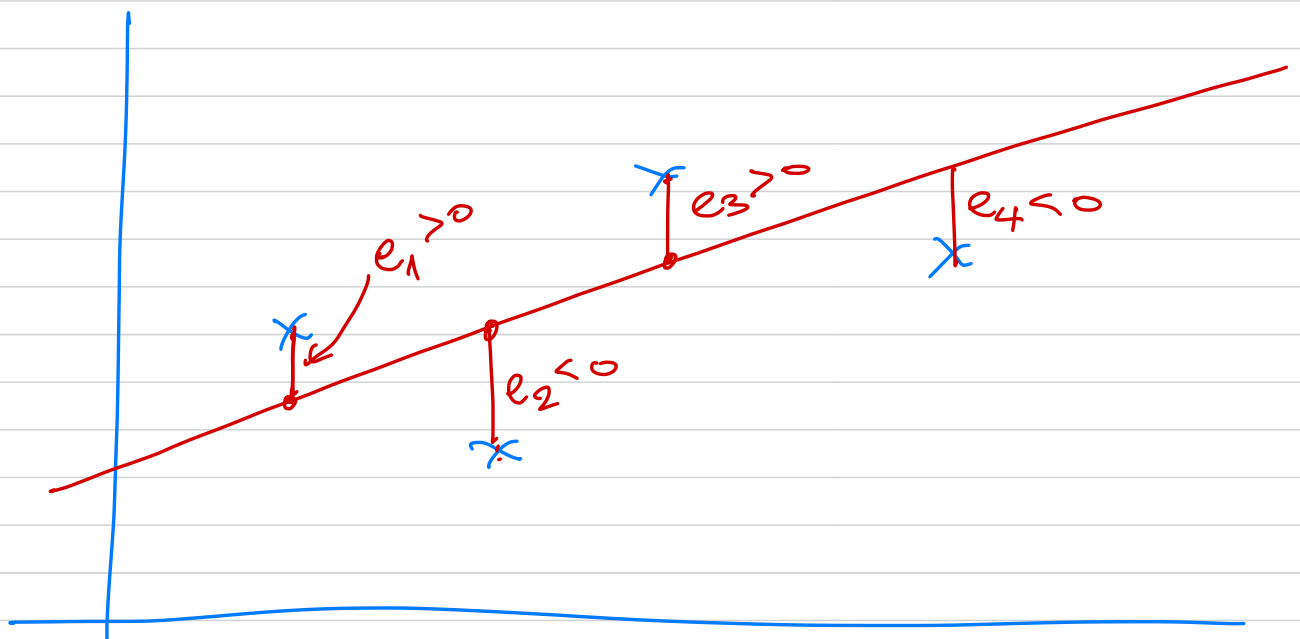
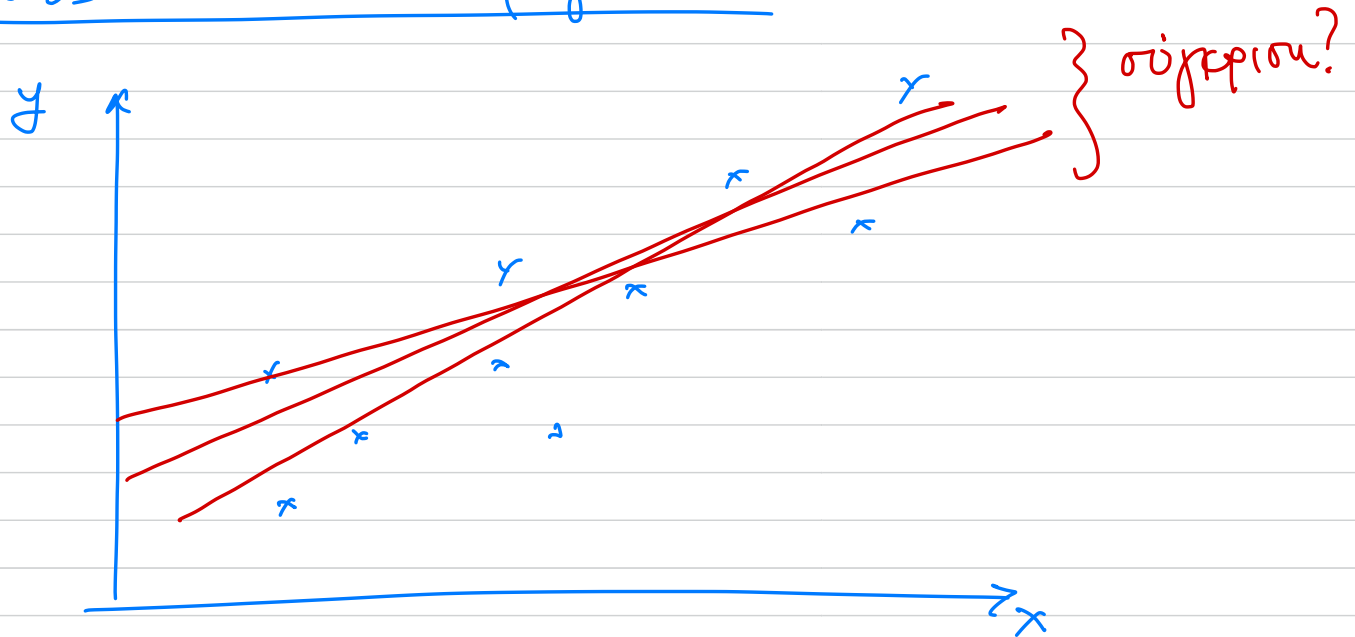
Y = σκόρ ποδοσφαιρικής επίδοσης

Συσχέτιση $\Rightarrow$ αιτιολογική σχέση
 $\Downarrow$
(causal models)



Εκτίμηση παραμέτρων $b_0, b_1, (\sigma^2)$

Μέθοδος ελαχίστων τετραγώνων



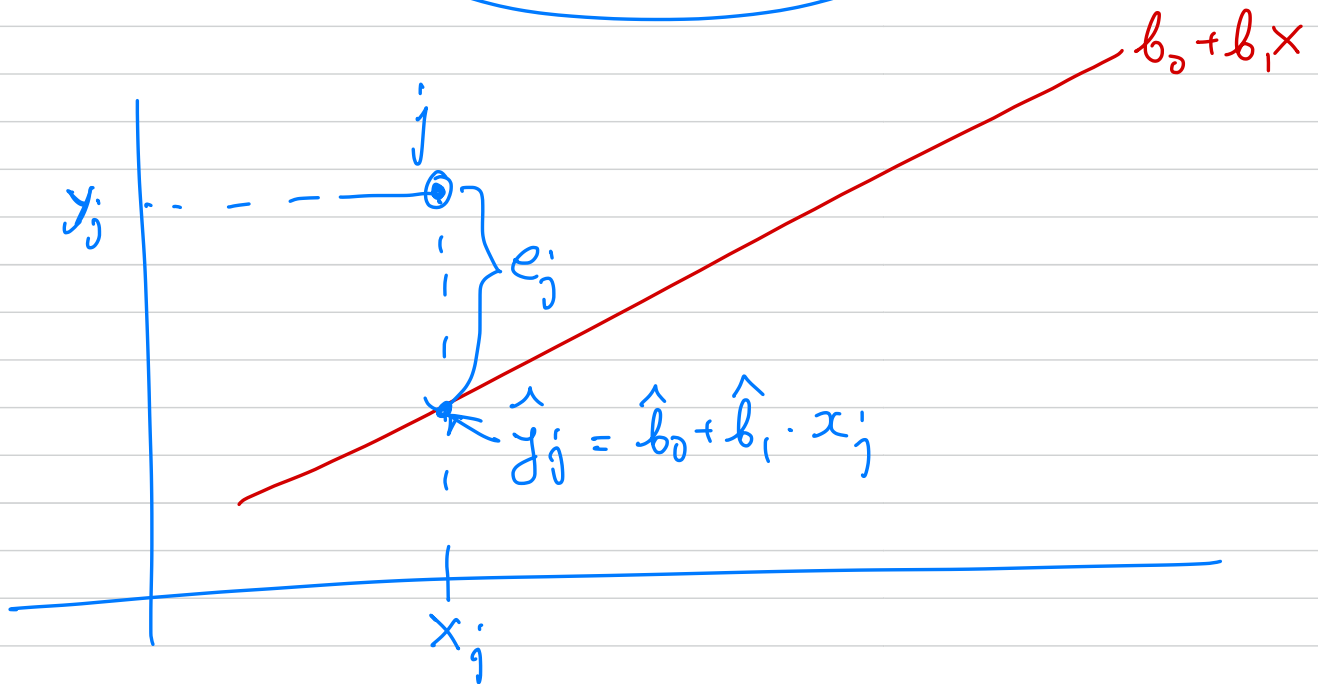
$$\sum e_j \quad \times$$

$$\sum |e_j| \quad \text{δυσκολο μαθηματικά}$$

$$\underline{\sum e_j^2} \quad \text{αριθμιο προσαρμογής}$$

Θέλουμε να βρούμε $\hat{b}_0, \hat{b}_1$ τέτοια ώστε

$$\sum e_j^2 \leftarrow \text{minimum}$$



$$e_j = y_j - \hat{y}_j = \text{κατάλοιπο της παραρ. } j$$

(residual)

$$e_j = y_j - \hat{b}_0 - \hat{b}_1 x_j$$

SSE = sum of squared errors

$$SSE = \sum_{j=1}^n (y_j - \hat{y}_j)^2 = \sum_{j=1}^n (y_j - \hat{b}_0 - \hat{b}_1 x_j)^2$$

Εκτιμήσεις ελαχ τετραγώνων των b_0, b_1 :

$$\hat{b}_0, \hat{b}_1 : SSE \leftarrow \text{minimum}$$

Τύποι :

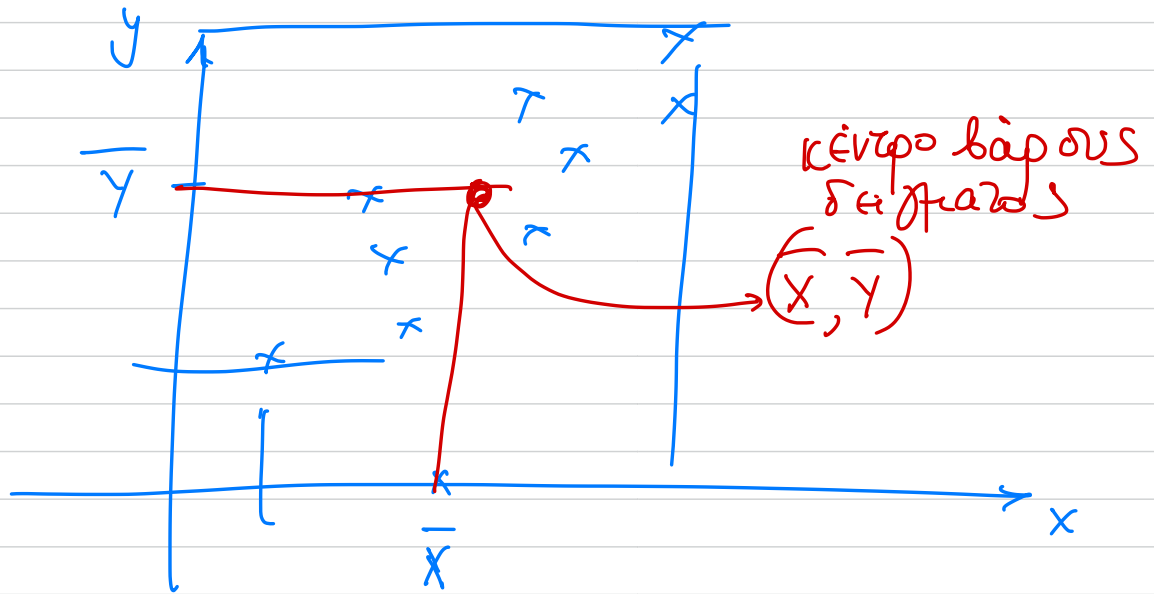
$$\begin{aligned} \hat{\beta}_1 &= \dots \\ \hat{\beta}_0 &= \dots \end{aligned} \quad \left\{ \begin{aligned} \frac{\partial SSE}{\partial b_0} &= 0 \\ \frac{\partial SSE}{\partial b_1} &= 0 \end{aligned} \right.$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}$$

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

$$\bar{y} = \frac{y_1 + \dots + y_n}{n}$$

$$\bar{y} = \hat{b}_0 + \hat{b}_1 \cdot \bar{x}$$



$\hat{b}_0$: κατανομή ?

$\hat{b}_1$: κατανομή ? ←

Γενικά αποτελέσματα

Αν $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$:

$$E(\varepsilon_i) = 0$$

$$\text{Var}(\varepsilon_i) = \sigma^2$$

$\varepsilon_1, \dots, \varepsilon_n$: αυτοχέυτα !

Τότε

$$E(\hat{b}_0) = b_0 \quad E(\hat{b}_1) = b_1$$

$\Rightarrow \hat{b}_0, \hat{b}_1$: αμερόμητες

$$Y = b_0 + b_1 X + \varepsilon$$

$$y_1 = b_0 + b_1 x_1 + \varepsilon_1$$

$$y_2 = b_0 + b_1 x_2 + \varepsilon_2$$

ε αυτοχέυτα $\Leftrightarrow$ οι παρατηρήσεις είναι
αυτοχέυτες

Ανάλυση Διασποράς

$$SSE = \sum_{j=1}^n (y_j - \hat{y}_j)^2 \leftarrow$$

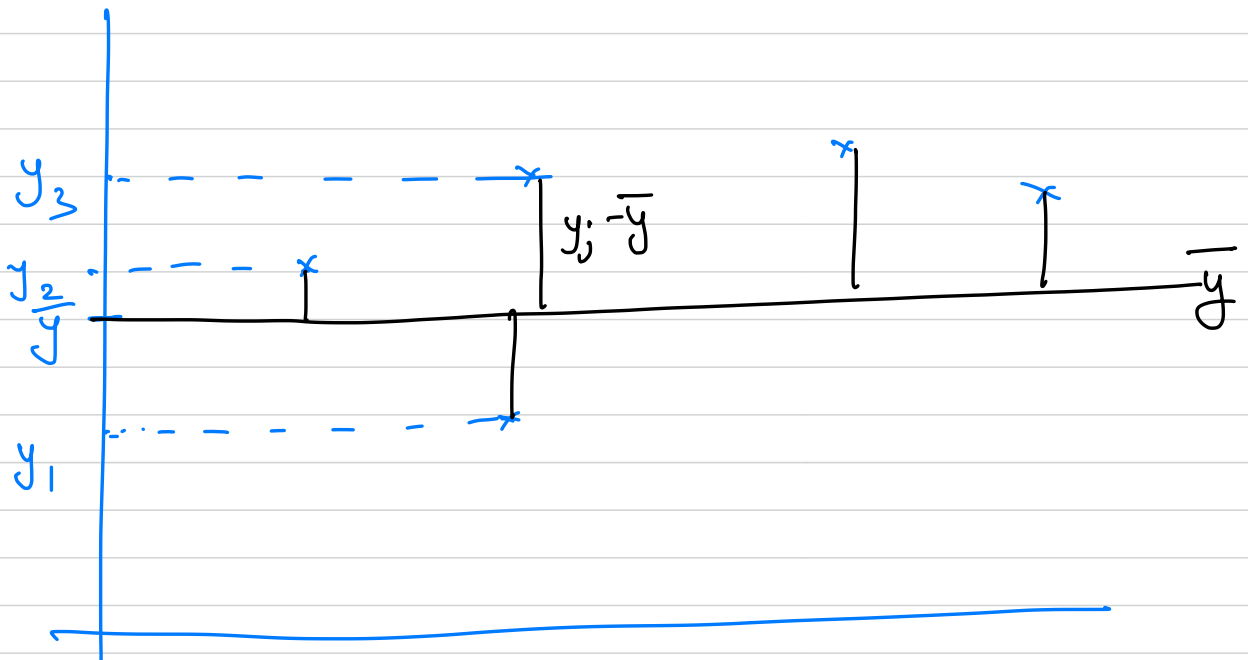
μεταβλητότητα της Y
που παραμένει ανεξήγητη
από το μοντέλο

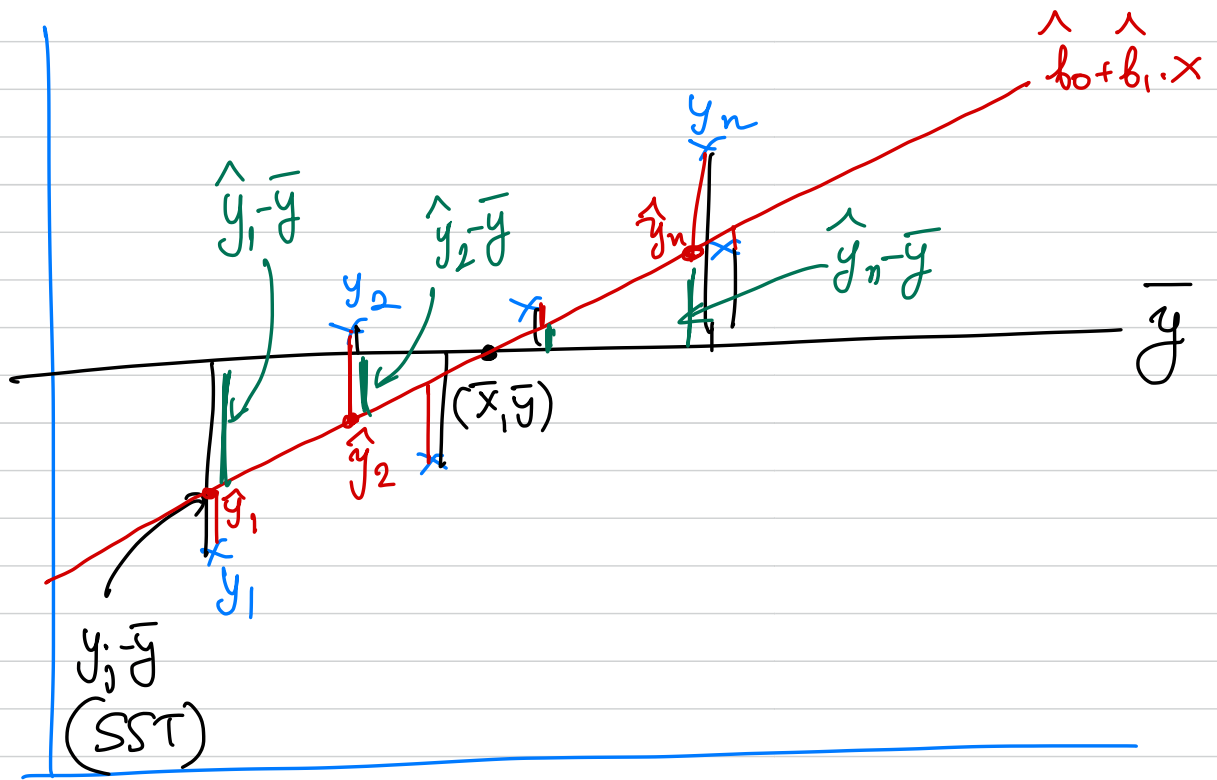
$$SST = \sum_{j=1}^n (y_j - \bar{y})^2$$

ανεξάρτητη των
μοντέλων

συνολική μεταβλητότητα
της Y στο δείγμα

$$\left(\frac{SST}{n-1} = \widehat{\text{Var}(Y)} \right)$$





$$SSE = \sum (y_i - \hat{y}_i)^2$$

$SSE < SST$

①

$$SSR = \text{Sum Squares Regression} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

μεταβλητότητα
της Y που
εξηγείται
από το μοντέλο
 $Y = b_0 + b_1 X$

Θεώρημα ανάλυσης διασποράς

Όταν τα $\hat{b}_0, \hat{b}_1$ εκτιμώνται μέσω ελαχ. τετραγώνων

τότε

$$SST = SSR + SSE$$



ανεξάρτητο
μοτέλο

$$R^2 = \frac{SSR}{SST} = \% \text{ μεταβλητότητας της } Y \text{ που εξηγείται από το } \text{μοτέλο.}$$

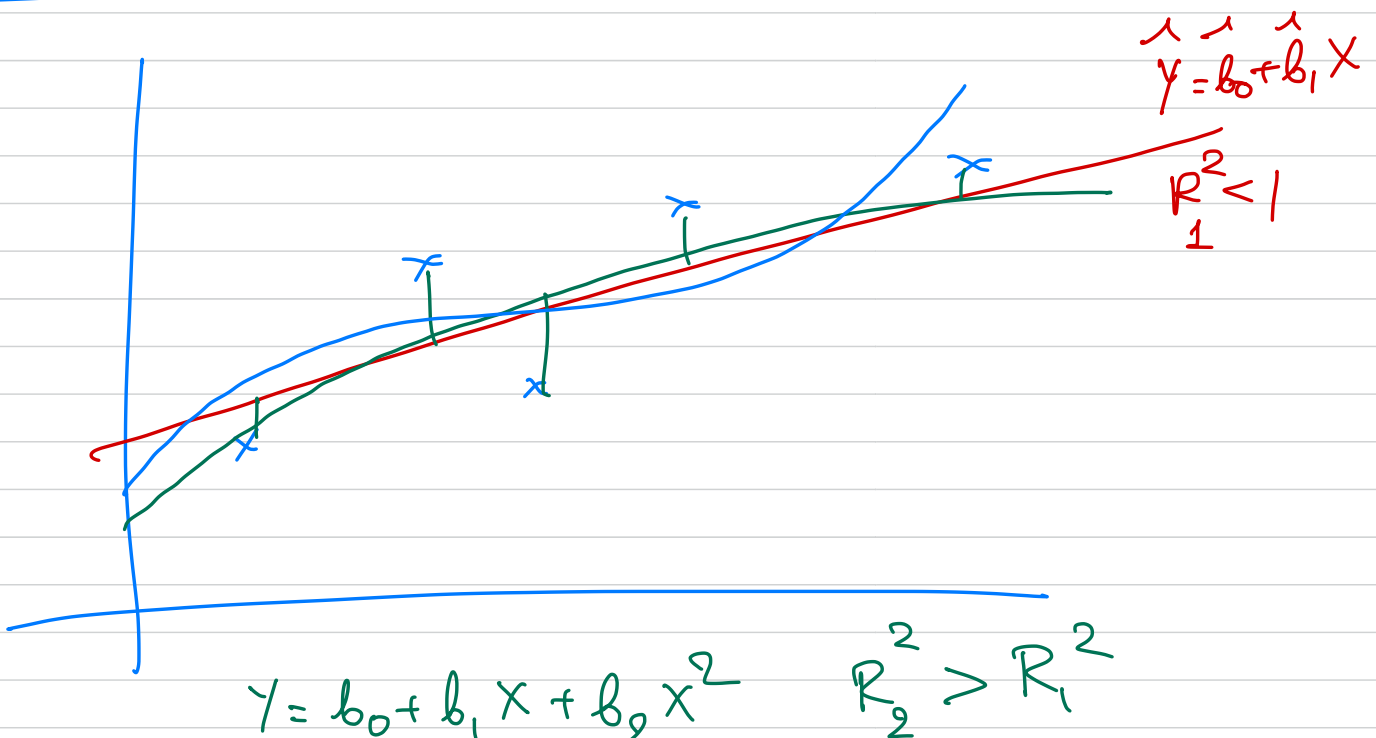
$$0 < R^2 < 1$$

$$R^2 = 1 \Rightarrow SSR = SST \Rightarrow$$

$$\Rightarrow SSE = 0$$

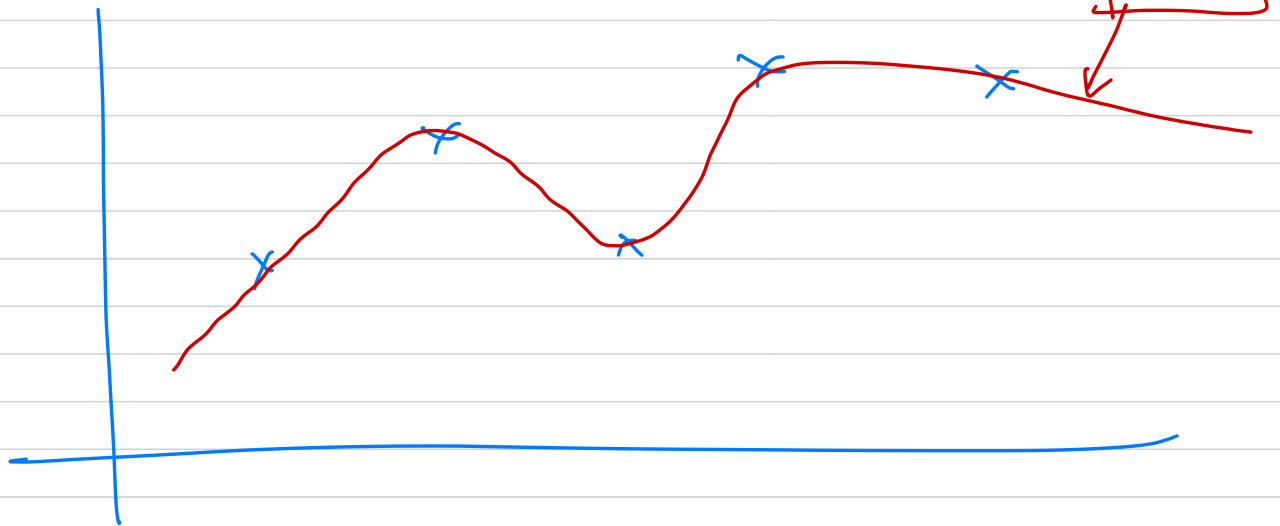
$$R^2 = \left(\hat{\rho}(X, Y) \right)^2$$

δηλ τα οφέλη
πάρω συν εικόνα
(τέλειο fit)



$$Y = b_0 + b_1X + b_2X^2 + b_3X^3$$

$$Y = b_0 + b_1X + b_2X^2 + b_3X^3 + b_4X^4$$



R^2 : επεξηγηματική ικανότητα μοντέλου

? στατιστική σημαντικότητα μοντέλου
 έλγχοι υποθέσεων
 χρειάζεσθε υπόθεση για
 την κατανομή των ϵ_i

Γραμμικά Μοντέλα $\rightarrow$ Υποθέτουμε $\epsilon_j \sim N(0, \sigma^2)$ $\left. \begin{array}{l} \epsilon_j \text{ αυτοσχευία} \end{array} \right\} \Rightarrow$

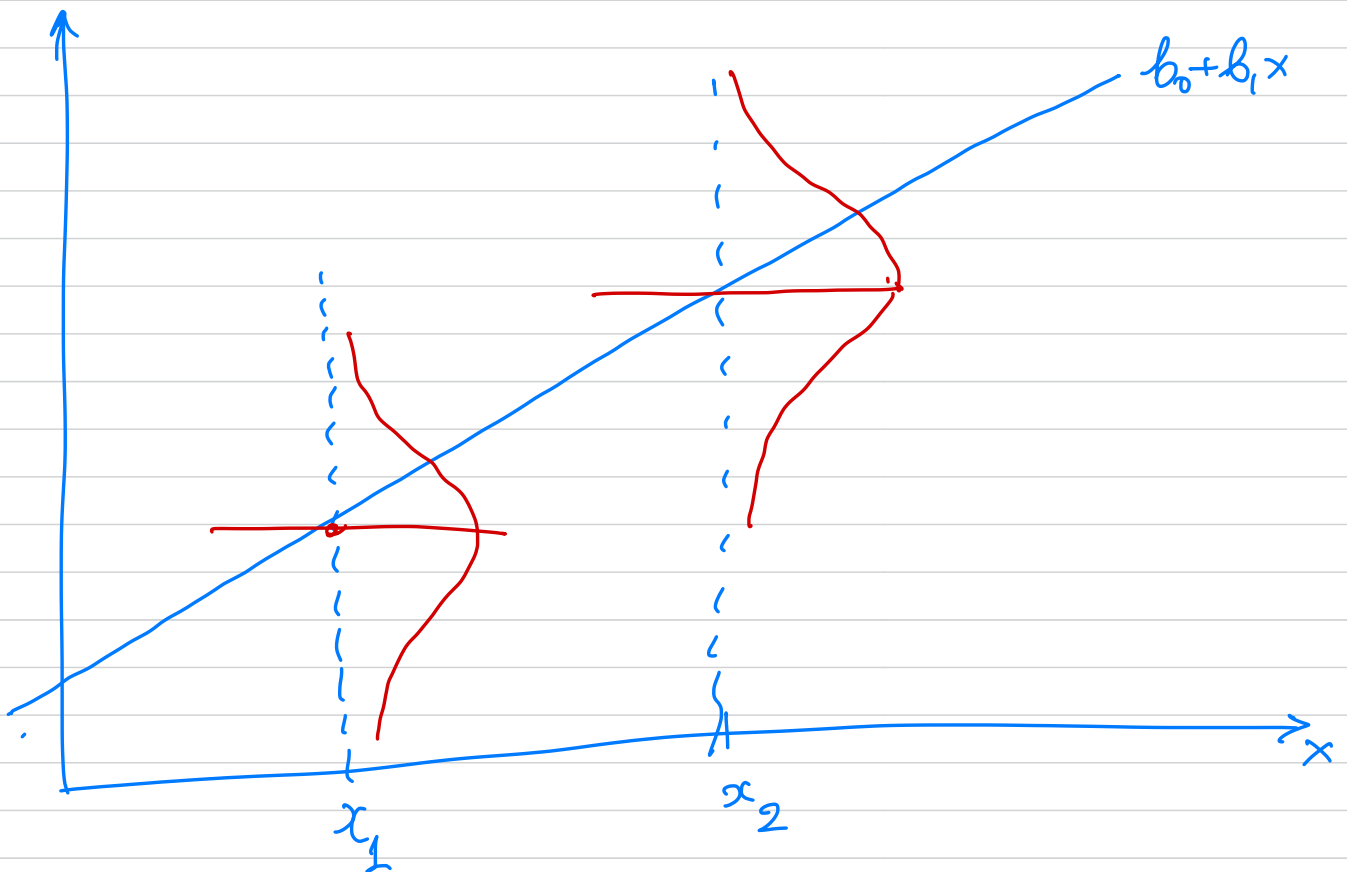
$\Rightarrow$ ϵ_j ανεξάρτητα

Υποθέσεις Παλινδρόμησης

$$Y_j = b_0 + b_1 x_j + \varepsilon_j \quad j=1, 2, \dots, n$$

- ε_j ανεξάρτητα
- $E(\varepsilon_j) = 0$
- $\text{Var}(\varepsilon_j) = \sigma^2$ ← ομοσκεδαστικότητα
- $\varepsilon_j \sim N(0, \sigma^2)$

$$Y_j | X = x_j \sim N(b_0 + b_1 x_j, \sigma^2)$$



Ελεγχος ομιαμερότητας ποτεκlor

$$H_0: b_1 = 0$$

$$H_1: b_1 \neq 0$$

t-test

$$t = \frac{\hat{b}_1 - 0}{\hat{S}_{\hat{b}_1}} = \frac{\hat{b}_1}{\hat{S}_{\hat{b}_1}}$$
