20-12-2024 noa.webex.com/meet/aburnetas

Απλό Γραμμικό Μοντέλο

Y εξαρτ. μεταβλητή (πάντα ποσοτική)

X ανεξάρτητη μεταβλητή (ποσοτική θα γενικευτεί αργότερα

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

Εδώ $f(x) = b_0 + b_1 x$

$E(Y|X=x) = b_0 + b_1 \cdot x$, b_0, b_1 συντελεστές (άγνωστες παράμετροι)

ισοδύναμα $Y = \underbrace{b_0 + b_1 \cdot X} + \underbrace{\varepsilon}_{\text{τυχαία μεταβλητή}}$ (τυχαία απόκλιση)

$E(\varepsilon) = 0$ $\quad Var(\varepsilon) = \sigma^2$ (άγνωστη) (ανεξάρτητη του X)

Στο "μηδενικό μοντέλο" μόνο η $Y \sim N(\mu, \sigma^2)$

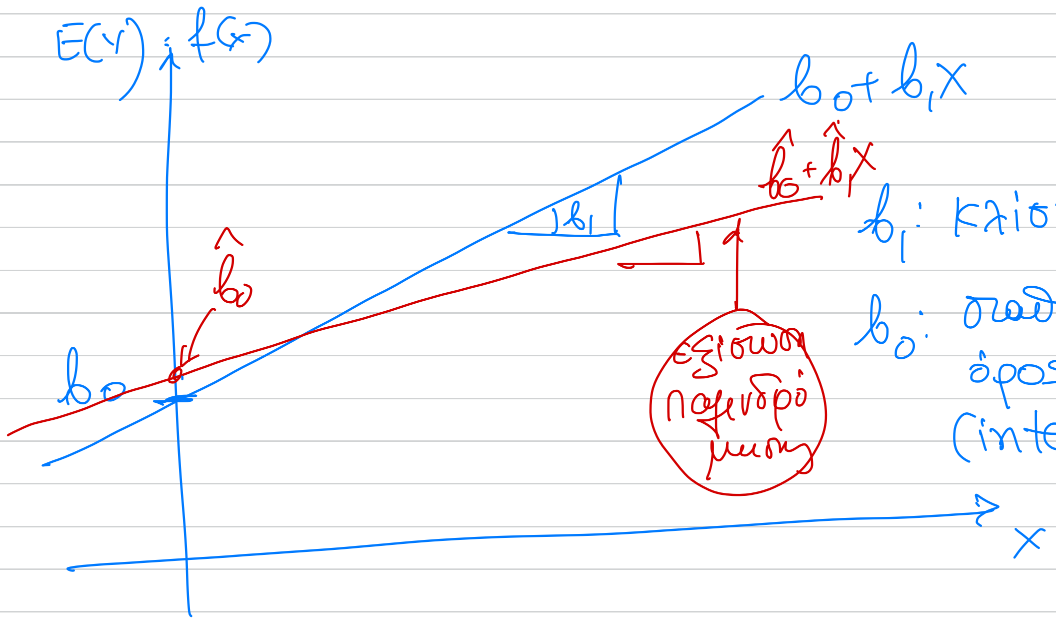
$E(Y) = \mu$

$\Leftrightarrow \quad Y = \mu + \varepsilon \, , \quad \varepsilon \sim N(0, \sigma^2)$

Δείγμα

X	Y
x_1	y_1
x_2	y_2
⋮	
x_n	y_n

$\Rightarrow$ εκτίμηση $b_0, b_1, \sigma^2, \ldots$



b_1: κλίση (slope)

b_0: σταθερός όρος (intercept)

Έστω ότι έχουμε βρει εκτιμητές $\hat{b}_0$, $\hat{b}_1$

$\hat{b}_1$

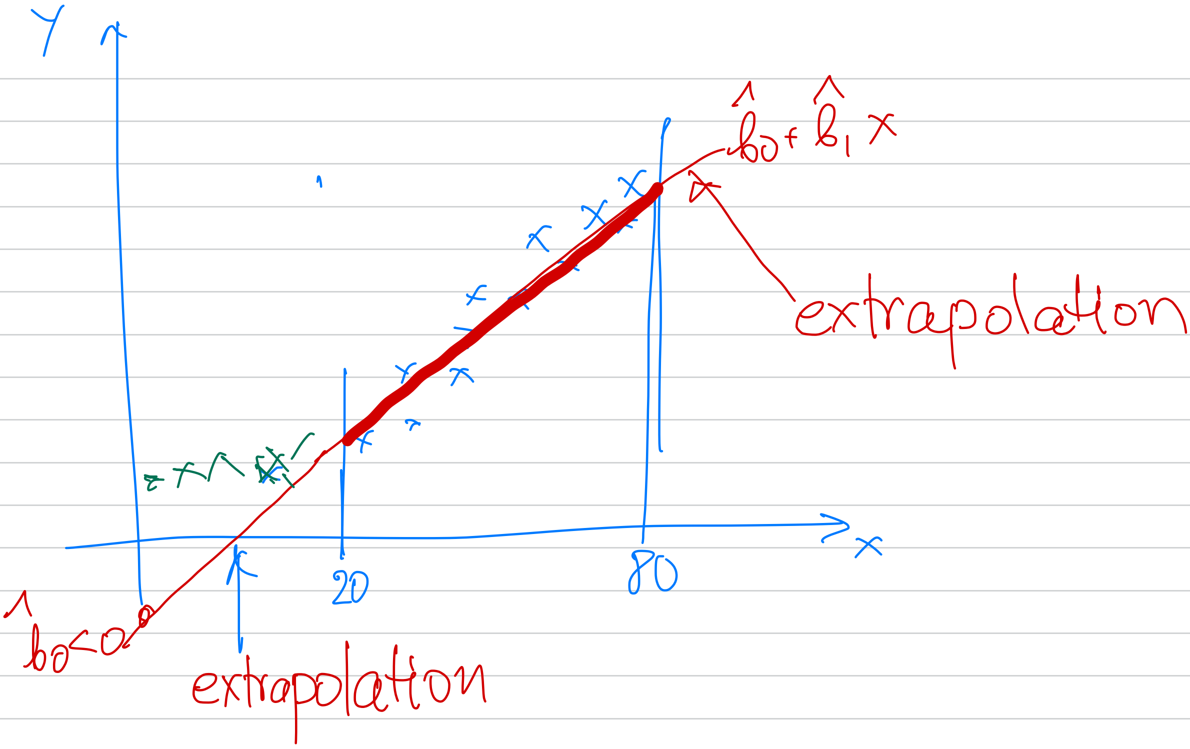
X = αριθ (έτη)

Y = αρτ. πίεση σε (mmHg)

$\hat{b}_1 = 2$

$\hat{b}_0 = 11$

$$\hat{b}_0 = \widehat{E(Y \mid X=0)}$$



2) <u>Αιτιοκρατική Ερμηνεία</u> (δεν επιτρέπεται)

Αν βρώ ότι $\boxed{X, Y \text{ συσχετίζονται}}$

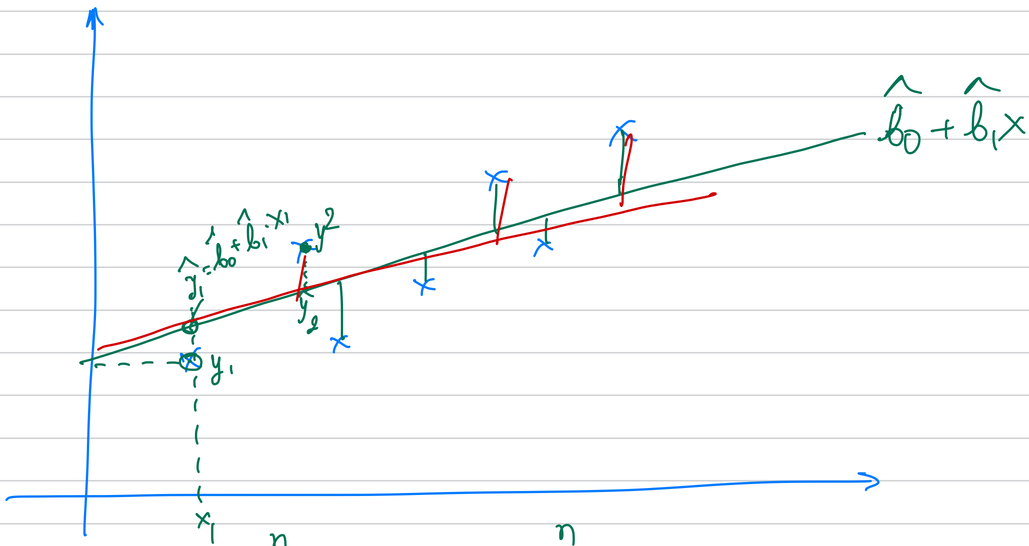
δηλ $\hat{b}_1 \neq 0$

δεν σημαίνει ότι η μια μεταβλητή <u>προκαλεί</u> τη μεταβολή της άλλης

<u>(causal models)</u>

Εκτίμηση των b_0, b_1

Μέθοδος Ελαχίστων Τετραγώνων



$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x)^2$$

sum of square erros

$$SSE(b_0, b_1) \leftarrow \min_{b_0, b_1}$$

[μέθοδος ελαχίστων τετραγώνων]

$$\left.\begin{aligned} \frac{\partial SSE}{\partial b_0} &= 0 \\ \frac{\partial SSE}{\partial \beta_1} &= 0 \end{aligned}\right\} \Rightarrow \hat{b}_0, \hat{b}_1$$

εκτιμήσεις ελαχ. τετραγώνων

(least squares estimates)

(LSE)

① **Θεώρημα** $\hat{b}_0, \hat{b}_1$ αμερόληπτες!!

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

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

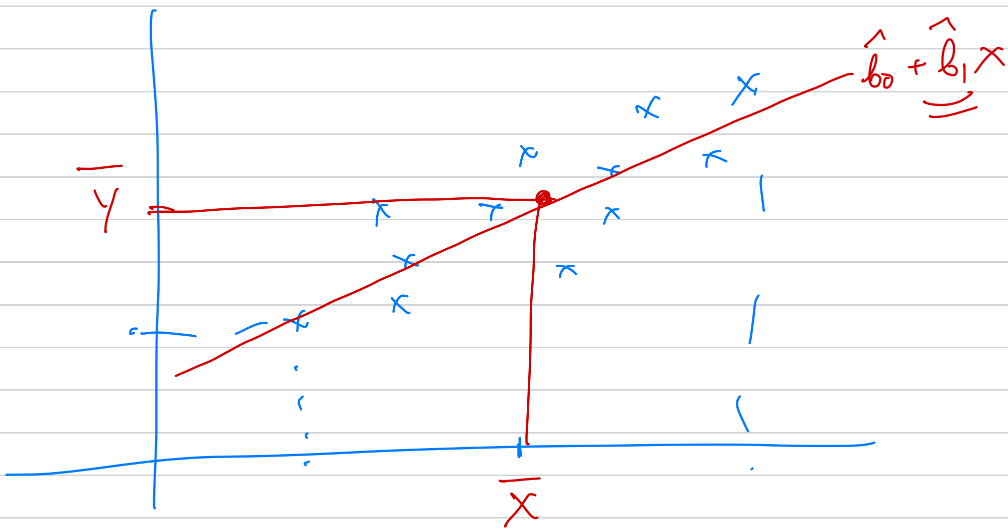
② Ιδιότητα των $\hat{b}_0$, $\hat{b}_1$

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \cdot \overline{X}$$

$$\overline{Y} = \frac{\Sigma y_i}{n} \qquad \overline{X} = \frac{\Sigma x_i}{n}$$

$$\overline{Y} = \hat{b}_0 + \hat{b}_1 \cdot \overline{X}$$

Το σημείο $(\overline{X}, \overline{Y})$ ανήκει στην ευθεία ελαχ. τετραγώνων



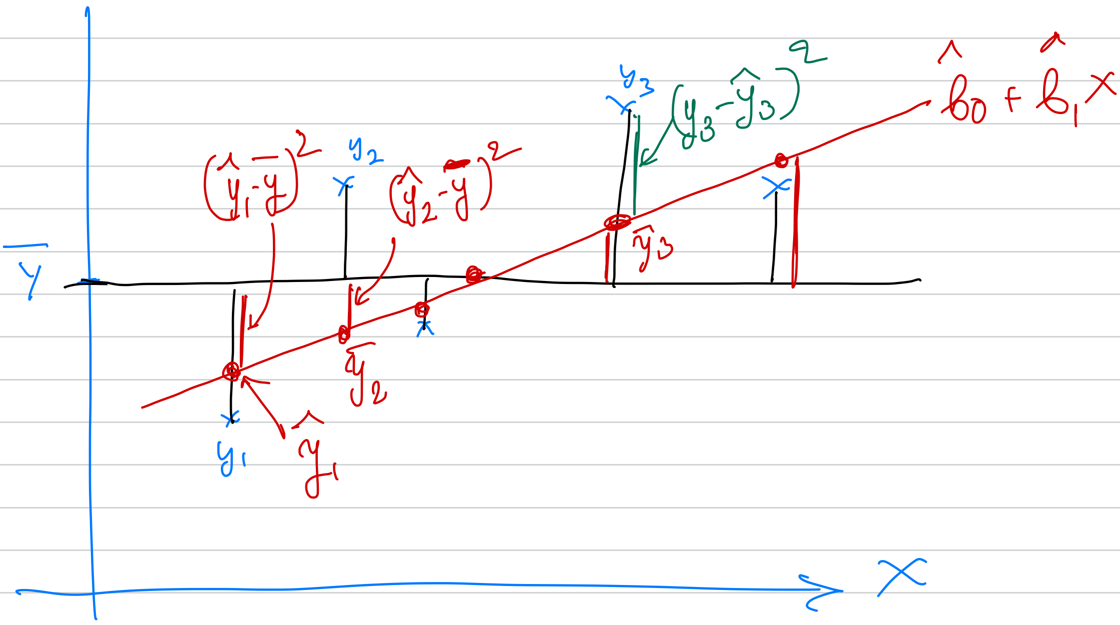
Εκτιμητική (Συμπερασματολογία)

Υπόθεση

$\boxed{\begin{array}{l}\varepsilon \sim N(0, \sigma^2) \\ \sigma^2 \text{ ανεξάρτητο του } X\end{array}}$

ομοσκεδαστικότητα

Ανάλυση Διασποράς



$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$ (εξαρτάται μόνο από τα y_i)

συνολική μεταβλητότητα του Y στο δείγμα.

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$ = άθροισμα τετραγώνων μοντέλου ή παλινδρόμησης

μεταβλητότητα του Y που εξηγείται από το X

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$ = άθροισμα τετρ. σφάλματος

μεταβλητότητα του Y που παραμένει ανεξήγητη από το X.

Θεώρημα Ανάλυσης Διασποράς

Κάτω από τη μέθοδο Ελαχ τετραγώνων

$$\boxed{SST = SSR + SSE}$$

$$0 < R^2 = \frac{SSR}{SST} < 1$$

% συνολικής μεταβλητότητας του Y που εξηγείται από τη X.

Βαθμοί Ελευθερίας

Με την υπόθεση εν $N(0,\sigma^2)$

$$\hat{b}_0 \sim N(b_0, \sigma^2_{\hat{b}_0})$$

$$\hat{b}_1 \sim N(b_1, \sigma^2_{\hat{b}_1})$$

$$SST \sim \chi^2_{n-1}$$

$$SSE \sim \chi_{n-2}$$

$$SSR \sim \chi^2_1$$

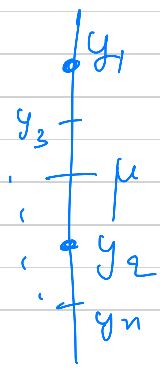
αμερόληπτη!

$$\hat{\sigma}^2 = \frac{SSE}{(n-2)}$$

$$E(\hat{\sigma}^2) = \sigma^2$$

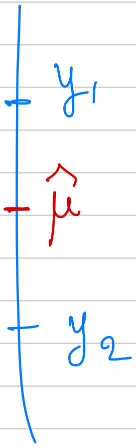
$(y_1, \dots, y_n) \sim N(\mu, \sigma^2)$ $\quad$ σ^2 άγνωστη

① μ γνωστό



$$\hat{\sigma}^2 = \frac{\sum (y_i - \mu)^2}{n}$$

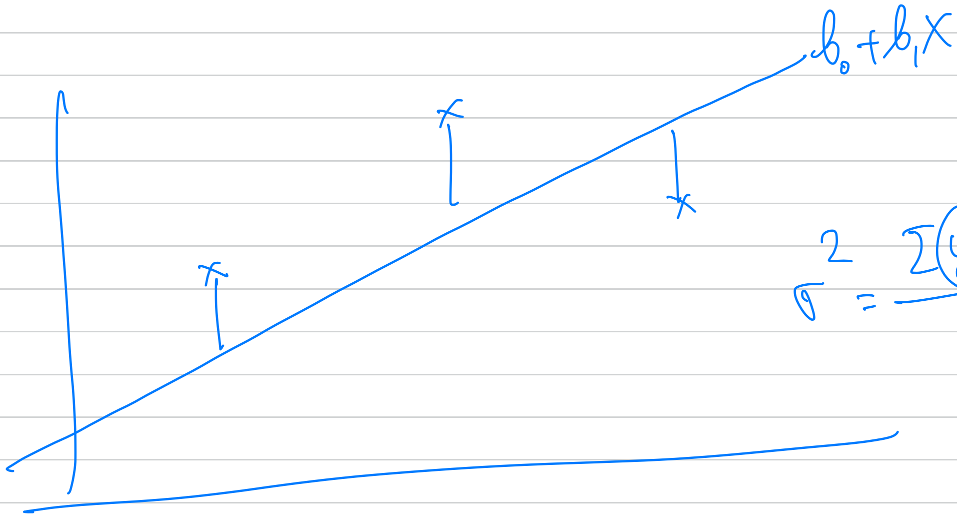
② μ άγνωστο



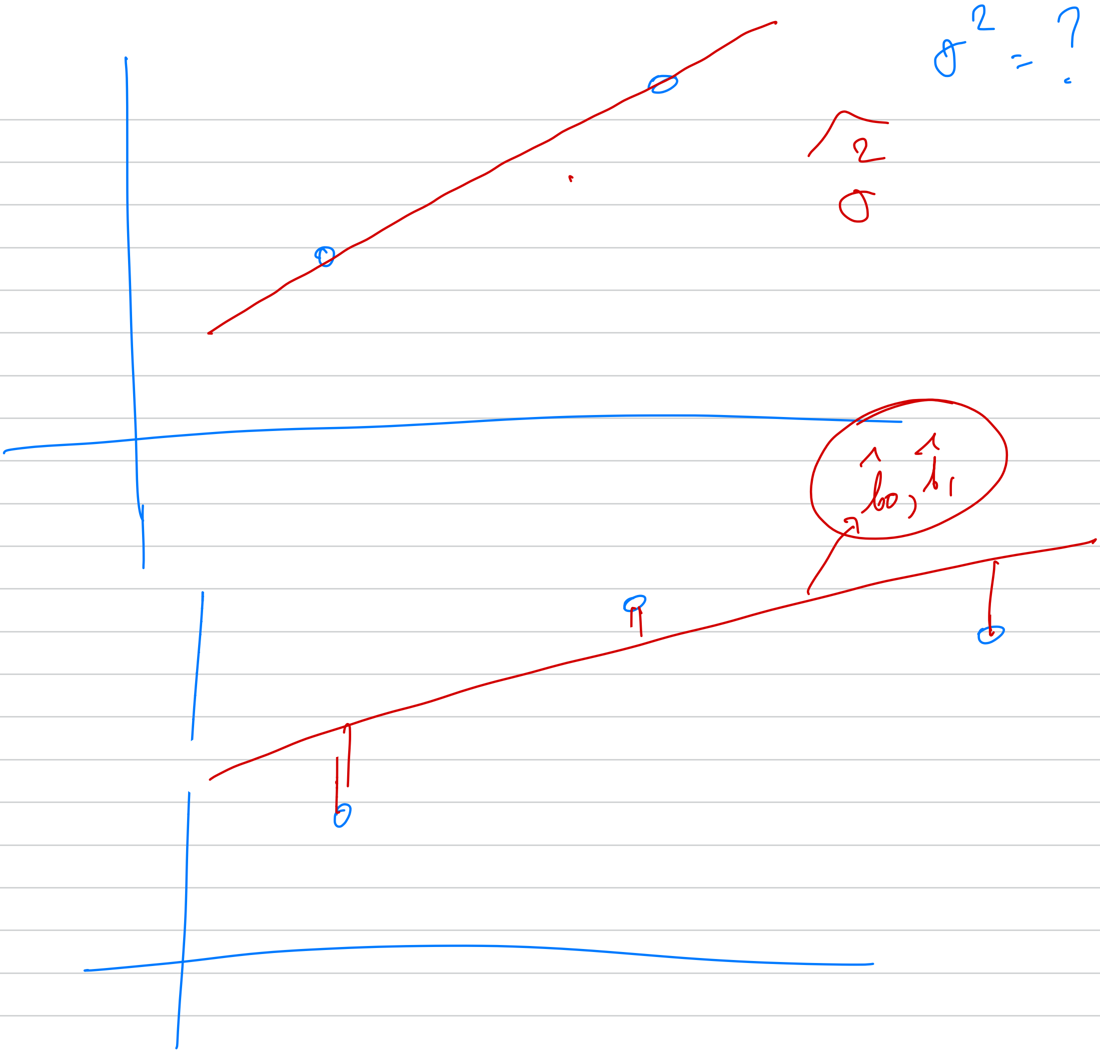
$$\hat{\sigma}^2 = \frac{\sum (x_i - \hat{\mu})^2}{n-1}$$

$Y = b_0 + b_1 x + \varepsilon$ $\quad\quad$ $\varepsilon \sim N(0, \sigma^2)$ $\quad$ σ^2 άγνωστο

① b_0, b_1 γνωστά



$$\sigma^2 = \frac{\sum (y_i - b_0 - b_1 x_i)^2}{3}$$



Κανόνας: βαθμοί ελευθερίας SST $= n-1 = df_{total}$

βαθμοί ελευθ. SSE $= n -$ αρ. b που εκτιμώνται $= df_{er}$

($= n-2$ στο $y = b_0 + b_1 \cdot x + \varepsilon$)

βαθμοί ελευθ. SSR $= \#b - 1 = df_{reg} = df_{mod.}$

$$\left.\begin{aligned} df_{tot} &= n-1 \\ df_{er} &= n - \#b \\ df_{mod} &= \#b - 1 \end{aligned}\right\} \quad \begin{aligned} df_{tot} &= df_{mod} + df_{er} \\ SST &= SSR + SSE \end{aligned}$$

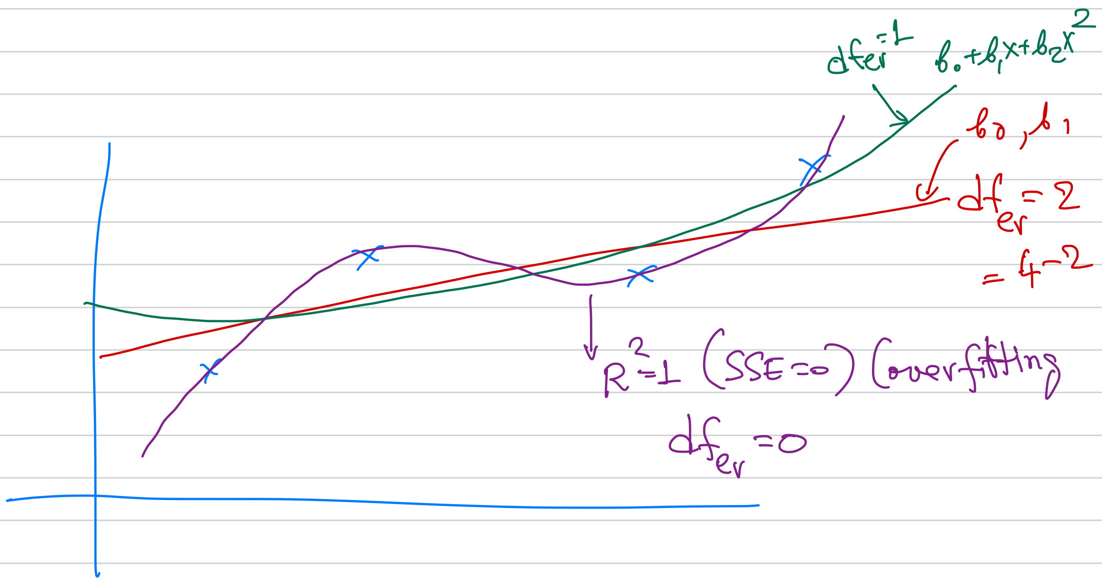
Πίνακας ANOVA

Source	SS	df	MS	F
Mod.	SSR	#b-1	MSR	
Error	SSE	n-#b	MSE	
Total	SST	n-1		

$R^2 = \frac{SSR}{SST}$

$MSR = \frac{SSR}{df_{mod}}$, $\boxed{MSE = \frac{SSE}{df_{er}}}$

$MSE = \hat{\sigma}^2$ αμερόληπτη $E(MSE) = \sigma^2$



Θα δούμε : Δ.Ε b_0, b_1

$\hat{b}_0, \hat{b}_1 \sim N$