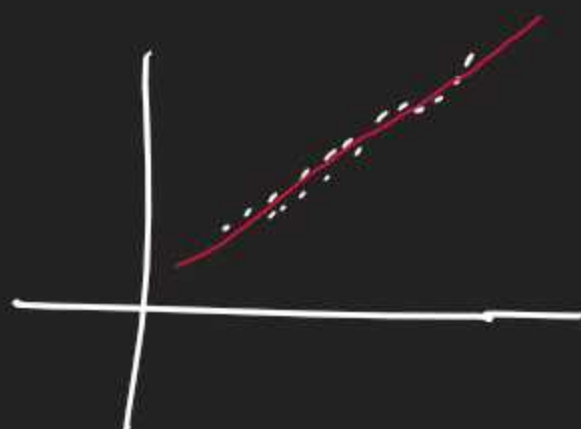


8/11/22

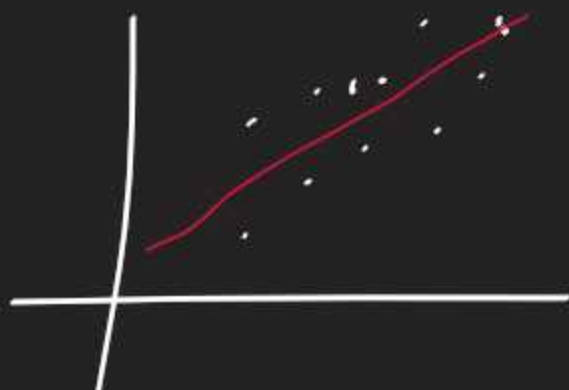
(1)

## Συντελεστής Προσδιορισμού $R^2$

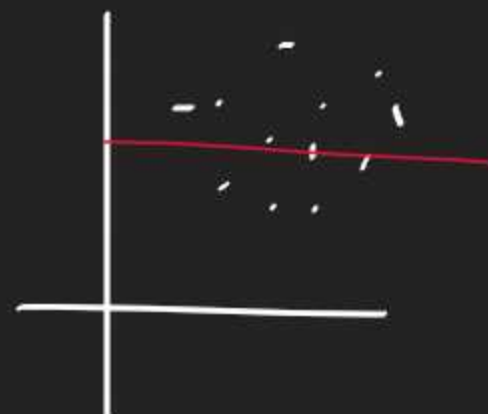
Εστω ότι έχουμε από γραμμικό πρότυπο  $\boxed{y = b_0 + b_1 x + \varepsilon}$   
και θέλουμε να ελέγξουμε πόσο καλό είναι το μοντέλο μας.



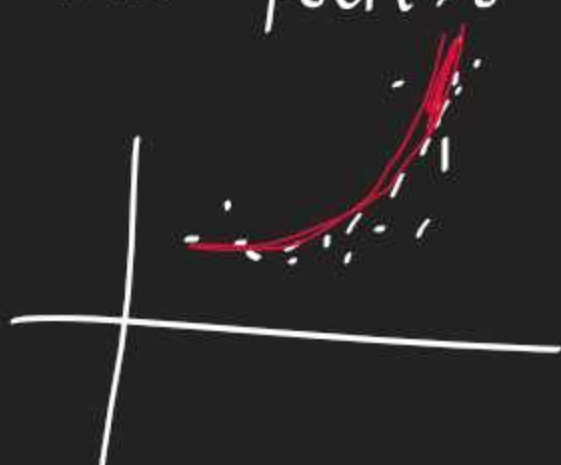
"καλό" μοντέλο



ΔΕΝ είναι καλό



κακό μοντέλο



$R^2 = 1$

$\varepsilon_i = 0$   
για κάθε  $i$

για  $0 \leq R^2 \leq 1$

$$R^2 = \frac{[\text{cov}(x, y)]^2}{s_x^2 \cdot s_y^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2} =$$

$$= \frac{\left( \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right)^2}{\left[ \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}$$

Ερμηνεία:  $R^2$  μετράει το ποσοστό μεταβλητότητας του  $Y$  που οφείλεται στη μεταβλητότητα της ανεξάρτητης μεταβλητής  $X$ .  
 $R^2$  δίνει ένα ποσοστό (καθώς και αριθμός)

Αν χρησιμοποιήσω διανύσματα  $R^2 = \frac{b_1 \cdot x^T \cdot y}{y^T \cdot y}$

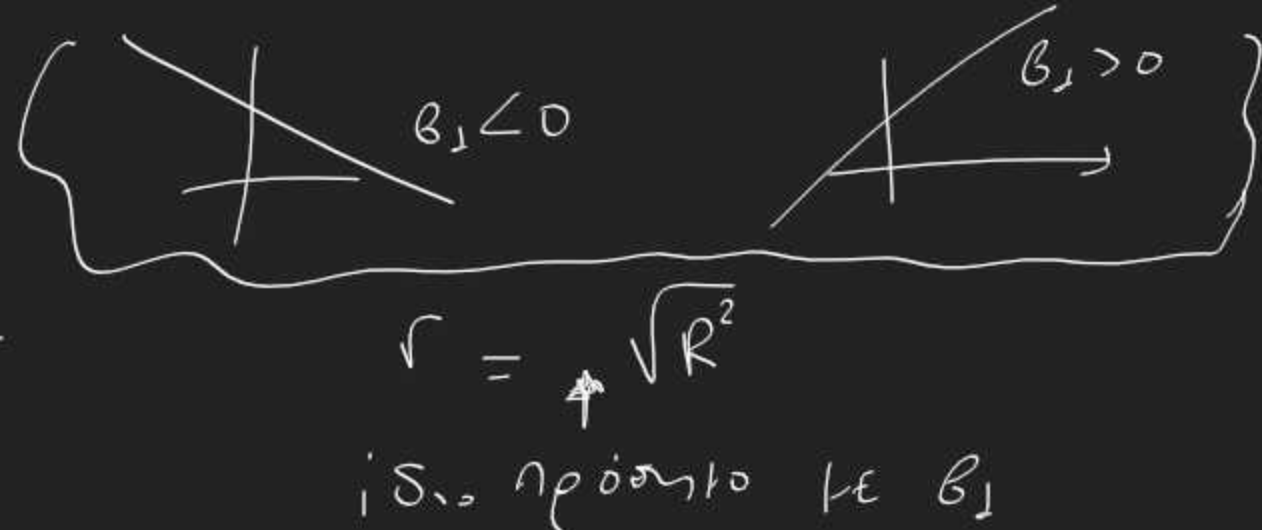
$$\boxed{y = b_0 + b_1 x + \dots + b_n x^n}$$

η  $X$  2 διαστά,  $Y$  από 5007 χωράφια

$R^2 = 0,4$ : Το 40% της μεταβλητότητας της απόδοσης οφείλεται στη μεταβλητότητα των 2 διαστάσεων

② Συντελεστής Συσχέτισης:

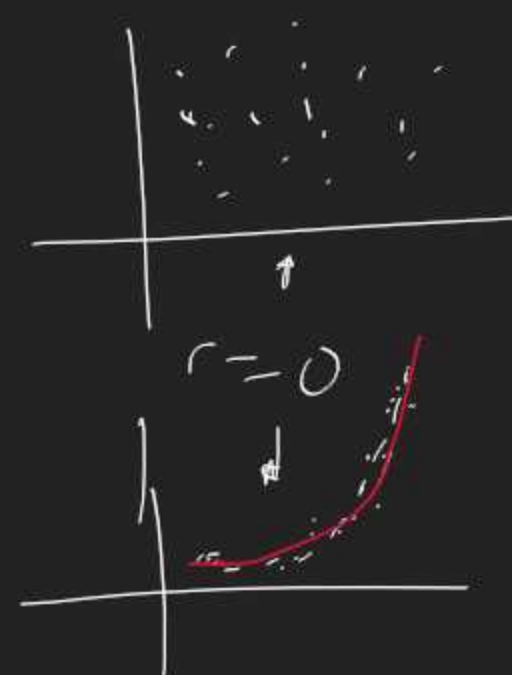
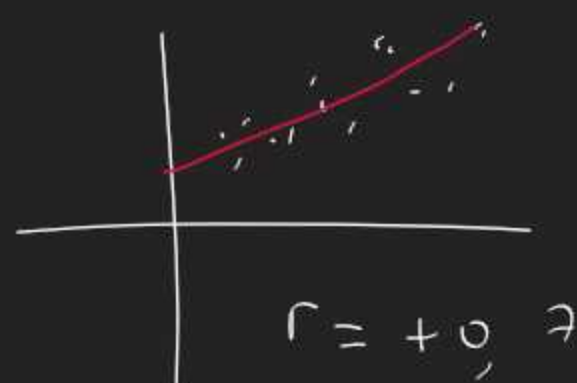
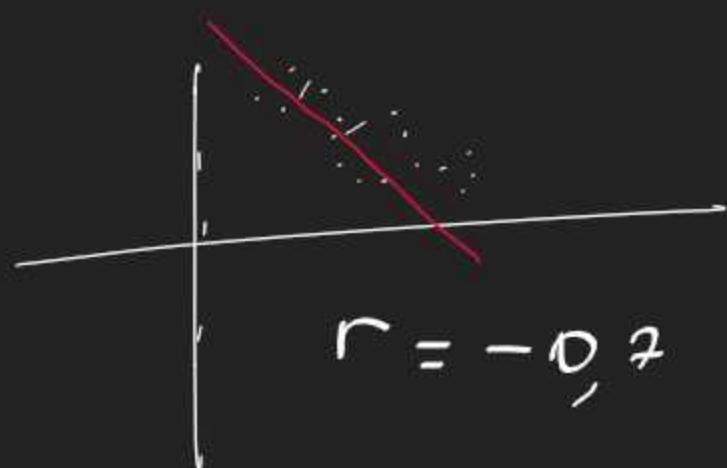
$$R; r = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}}$$



σε μορφή πινάκα:  $r = \underset{\substack{\uparrow \\ \text{κλίση ευθείας}}}{b_1} \cdot \frac{s_x}{s_y}$

σημειώ:  $-1 \leq r \leq 1$

$r$  μετράει το βαθμό γραμμικής εξάρτησης των μεταβλητών  $x, y$



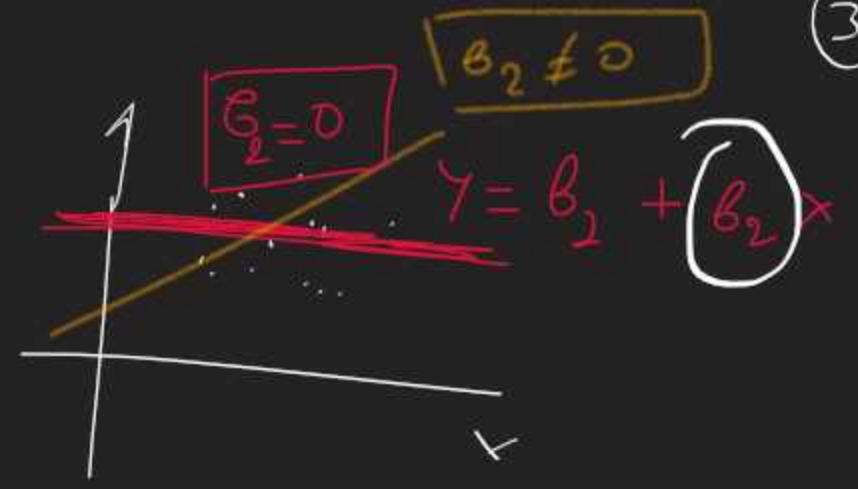
Στατιστική έρευνα του υποστίγματος

υπόθεση

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

αφίξη να έχω  
και αν ανυψώ,  
όποτε X έχει επίδραση στο Y?



Στατιστική:  $t = \frac{|\beta_2|}{s_b}$

όπου  $s_b = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{s}{s_x \cdot \sqrt{n-1}}$

Αν  $t \geq t_{n-2; \alpha}$  απορρίπτουμε  $H_0$  και

δεχόμαστε  $H_1$  άρα X έχει επίδραση στο Y  $\Rightarrow$   
η τιμή  $\beta_2$  είναι στατιστικά σημαντική.

Αν  $t < t_{n-2; \alpha}$  δεν απορρίπτουμε  $H_0$

η τιμή  
απορρίπτω

$\alpha$ : επίπεδο  
σημαντικότητας

συνήθως  $\alpha = 0,05 = 5\%$



$n_X$

X	Y
2	20
4	15
5	14
7	10
10	7

i) Γραφική πρότυπο

ii) Διασπορά

iii)  $R^2, r$

iv) Στατιστική έλεγχο υποστροφών

4

	X	Y	X.Y	X <sup>2</sup>	Y <sup>2</sup>
$x_1$	2	20			
$x_2$	4	15			
$x_3$	5	14			
$x_4$	7	10			
$x_5$	10	7			
Σολο	28	66	310	194	970

$$\bar{X} = \frac{\sum X}{n} = \frac{28}{5}$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{66}{5}$$

$$V=5$$

i)  $\beta_2 = \frac{n \sum XY - \sum X \cdot \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{5 \cdot 310 - 28 \cdot 66}{5 \cdot 194 - (28)^2} = -1,602$

ii)  $\beta_1 = \bar{Y} - \beta_2 \bar{X} = \frac{66}{5} - (-1,602) \cdot \frac{28}{5} = 22,1712$

iii)  $Y = \beta_1 + \beta_2 X + \varepsilon = 22,1712 - 1,602 X + \varepsilon$

$$Y = \beta_1 + \beta_2 X$$

$$Y = 0 + \beta X$$

iv)  $S^2 = \frac{\sum Y^2 - \beta_1 \sum Y - \beta_2 \sum X \cdot Y}{n-2} = \frac{970 - 22,1712 \cdot 66 - (-1,602) \cdot 310}{5-2}$

$$= 1,10693$$

v)  $S = \sqrt{S^2} = \sqrt{1,10693} = 1,0521$

vi)  $R^2 = \frac{(n \sum XY - \sum X \cdot \sum Y)^2}{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]} = \frac{(5 \cdot 310 - 28 \cdot 66)^2}{[5 \cdot 194 - 28^2][5 \cdot 970 - 66^2]}$

$$= 0,9664$$

vii)  $r = R = -\sqrt{R^2} = -\sqrt{0,9664} = -0,9830$

iv)  $H_0: \hat{\beta}_2 = 0$   
 $H_1: \hat{\beta}_2 \neq 0$

$$t = \frac{|\hat{\beta}_2|}{s_b} = \frac{|-1,0521|}{0,172} = 9,313$$

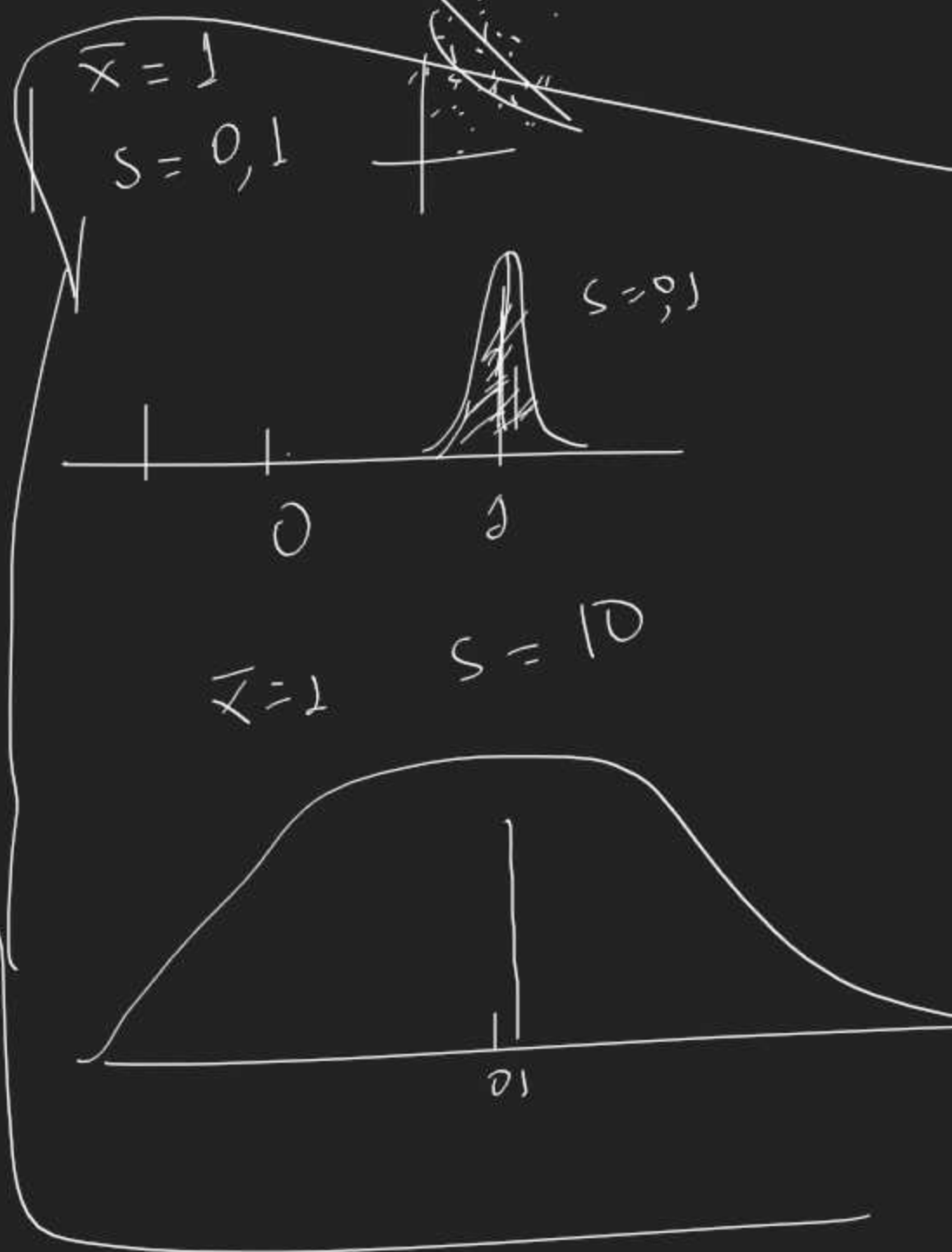
$$s_b = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{s}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}} = \frac{1,0521}{\sqrt{194 - \frac{28^2}{5}}} = \frac{1,0521}{\sqrt{37,2}} = \frac{1,0521}{6,099} = 0,172$$

critical value  
 $t_{crit} = t_{n-2; \alpha} = t_{5-2; 0,05} = t_{3; 0,05} = 2,712$

$$t = 9,313 > 2,712 = t_{crit}$$

αφα απορρίπτουμε  $H_0$

αφα  $\beta_2$  είναι στατιστικά σημαντική

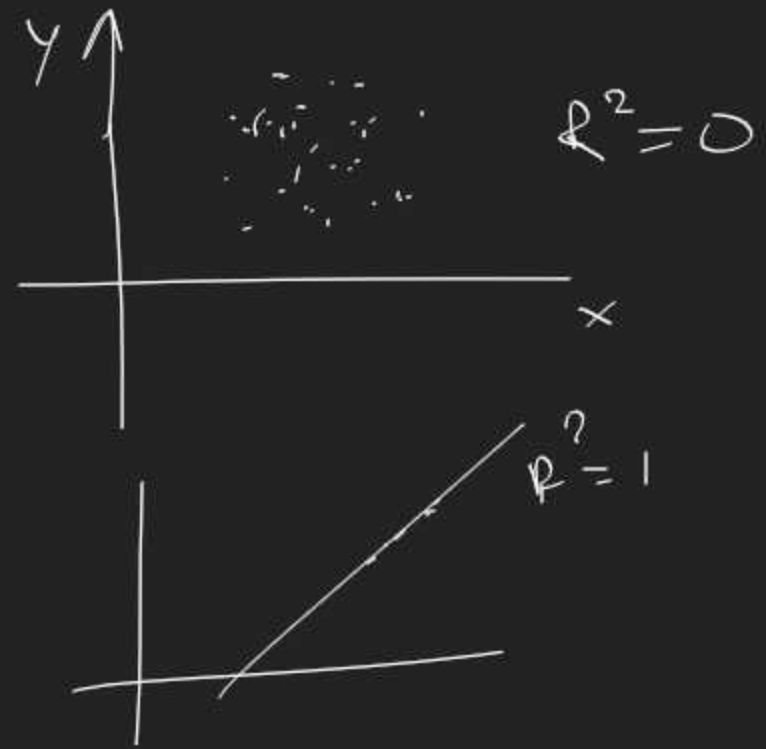


Ελέγχος υποθέσεων  $R^2$

$$H_0: \hat{R}^2 = 0$$

$$H_1: \hat{R}^2 \neq 0$$

στατιστική  $F = \left( \frac{\frac{R^2}{1}}{\frac{1-R^2}{n-2}} \right) = \frac{(n-2)R^2}{1-R^2}$



$$F_{crit} = F_{1, n-2; \alpha}$$

Αν  $F \geq F_{crit}$  τότε απορρίπτω  $H_0$

Αν  $F < F_{crit}$  τότε δεν απορρίπτω  $H_0$

Προσδιορισμός παραμέτρων:

$$F = \frac{(n-2)R^2}{1-R^2} = \frac{(5-2) \cdot 0,9664}{1-0,9664} = 87,85$$

$$F_{crit} = F_{1, 5-2; 0,05} = F_{1, 3; 0,05} = 4,24$$

πίνακας

$$F = 87,85 > 4,24 = F_{crit} \Rightarrow \text{απορρίπτω } H_0$$