

7/12/22

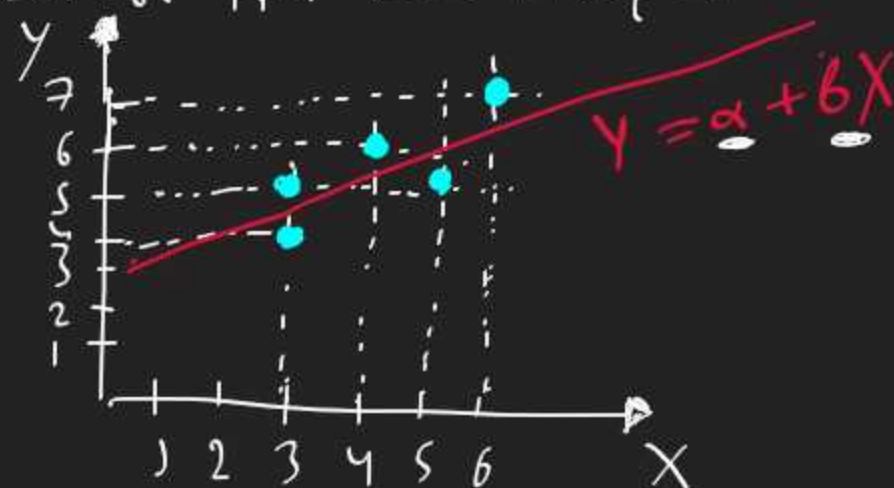
Γραφική Παρίσχυση

1

Έστω X η ανεξάρτητη μεταβλητή και
 Y η εξαρτημένη μεταβλητή

X	Y
3	4
4	6
3	5
6	7
5	5

Διάγραμμα Διασποράς (Scatter plot)



$$Y = aX + b$$

$$Y = a + bX$$

$$Y = b_0 + b_1X$$

$$Y = b_1 + b_2X$$

Υποθέτουμε πως η σχέση μεταξύ X και Y είναι γραμμική.
 και έστω $Y = \hat{a} + \hat{b}X$ η εξίσωση της ευθείας.

\hat{b} : συντελεστής ή κλίση της ευθείας

Η φυσική ερμηνεία είναι:

Για κάθε αύξηση της τιμής του X κατά μία μονάδα,
 πόσο αυξάνεται η τιμή του Y (αν $\hat{b} > 0$) ή
 πόσο μειώνεται η τιμή του Y (αν $\hat{b} < 0$)

\hat{a} : σταθερός όρος της ευθείας

Η φυσική ερμηνεία είναι:

Ποια είναι η τιμή του Y για $X=0$. (όταν δεν έχω X)

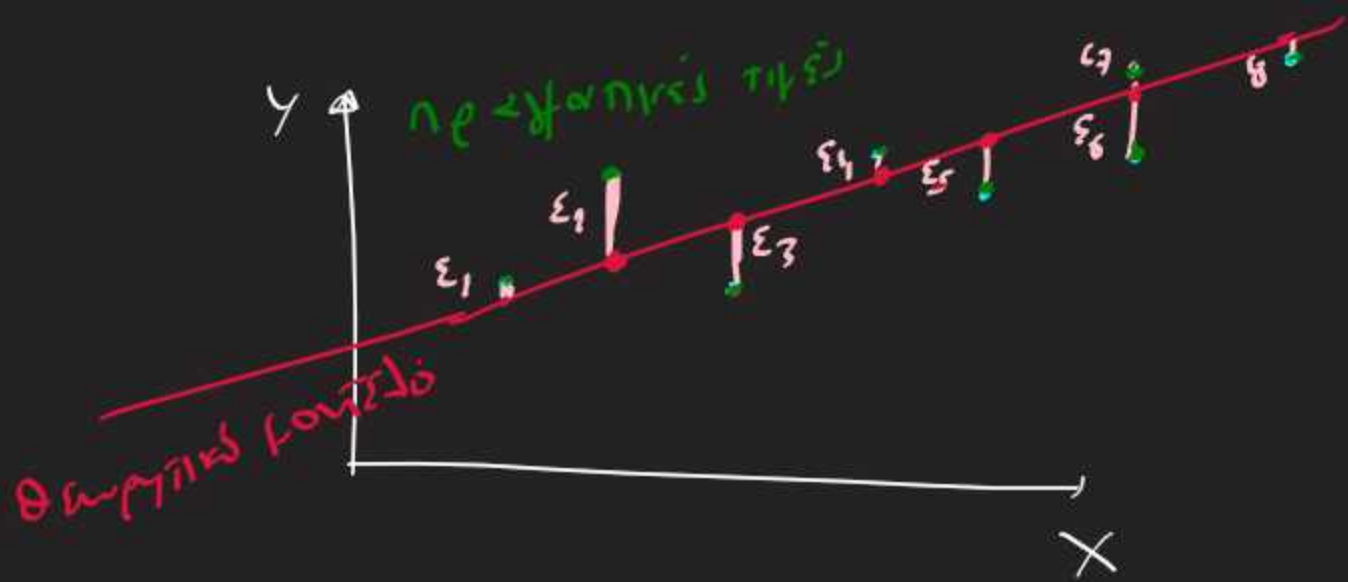
η X - kg λίπασματος που εγχύουμε σε ένα χωράφι και
 Y - η απόδοση του χωραφιού σε kg (ή σε €)

Δίνεται πως η σχέση X, Y είναι: $Y = \hat{a} + \hat{b}X$

Ερμηνεύστε το μοντέλο:

$\hat{b} = 10$ άρα για κάθε κιλό λίπασματος η απόδοση του χωραφιού αυξάνεται κατά 10 kg

$\hat{a} = 200$ άρα έχω 0 kg λίπασμα, δηλαδή δεν έχω λιπάσει τότε η απόδοση του χωραφιού θα είναι 200 kg.



σφάλμα/υπόλοιπο (error/residual) = πραγματική τιμή - θεωρητική τιμή

Ευθεία ελαχίστων τετραγώνων:

Υπολογίζουμε τους συντελεστές $\hat{\alpha}, \hat{\beta}$ προσαθώντας να ελαχιστοποιήσουμε το άθροισμα των τετραγώνων των σφαλμάτων.

$$\hat{\beta} = \frac{n \cdot \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \cdot \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$$

n: αριθμός παρατηρήσεων

Παράδειγμα: βρείτε εδώα ελαχίστων τετραγώνων για X, Y

n=8

X	Y	X·Y	X ²
15	5	75	225
13	6	78	169
11	8	88	121
9	10	90	81
9	9	81	81
6	12	72	36
5	15	75	25
4	11	44	16
72	76	603	754

Αθροιστά

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{72}{8} = 9$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{76}{8} = 9,5$$

$$\hat{\beta} = \frac{n \cdot \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \cdot \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{8 \cdot 603 - 72 \cdot 76}{8 \cdot 754 - 72^2} = \frac{4824 - 5472}{6032 - 5184} = \frac{-648}{848} = -0,76$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = \frac{76}{8} - (-0,76) \cdot \frac{72}{8} = 9,5 + 0,76 \cdot 9 = 9,5 + 6,84 = 16,34$$

$$Y = \alpha + \beta X \Leftrightarrow Y = 16,34 - 0,76 X$$

β) Ερμηνεύστε το μοντέλο: $Y = 16,34 - 0,76 \cdot X$

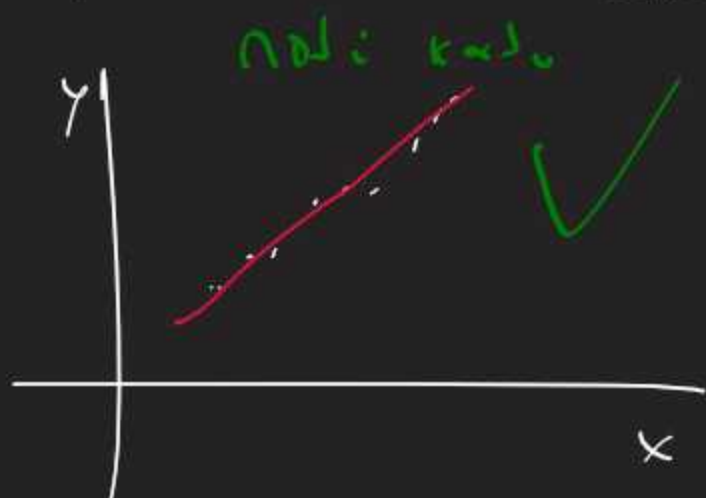
$\hat{\beta} = -0,76$ άρα όταν το X αυξάνεται κατά μία μονάδα, τότε Y μειώνεται κατά 0,76 μονάδες.

$\hat{\alpha} = 16,34$ άρα όταν $X = 0$ μονάδες, τότε $Y = 16,34$ μονάδες

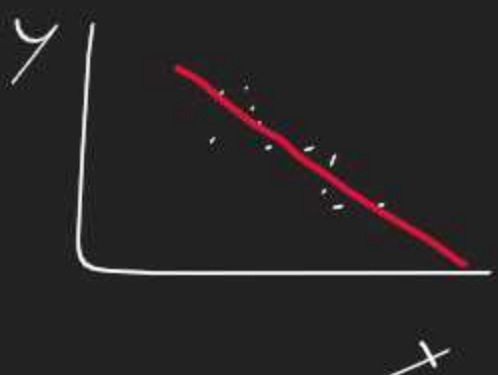
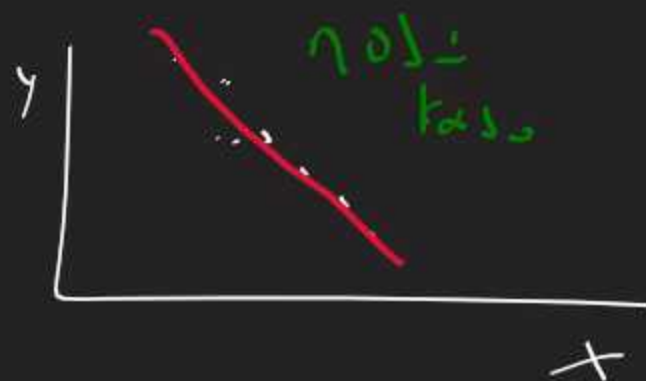
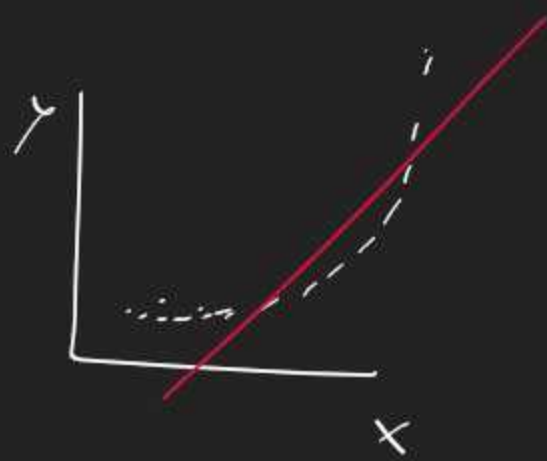
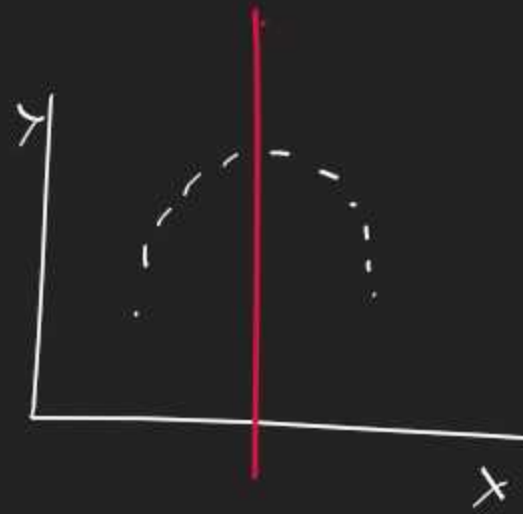
γ) Ποια είναι η αναμενόμενη τιμή του Y όταν $X = 5$ μονάδες (πρόβλεψη)

$$\begin{aligned} Y &= 16,34 - 0,76 \cdot 5 \\ &= 16,34 - 3,8 \\ &= 12,54 \end{aligned}$$

Συντελεστής Γραμμικής Συσχέτισης (Pearson)



Γραμμικό
Μοντέλο
Δω
ταπεινά



Συντελεστής συσχέτισης, Pearson που δείχνει κατά πόσο υπάρχει γραμμική συσχέτιση μεταξύ X και Y

$$r = \frac{n \cdot \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \cdot \left(\sum_{i=1}^n y_i \right)}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$