09/10/17

Μάθημα 1ο

Γενικότητες

Μαθηματικά

↓

Μαθηματικά Μοντέλα

↑

Μεταβλητές (X, Y, Z, ...)

Παράδειγμα:

Ευθύγραμμης ομαλής κίνησης

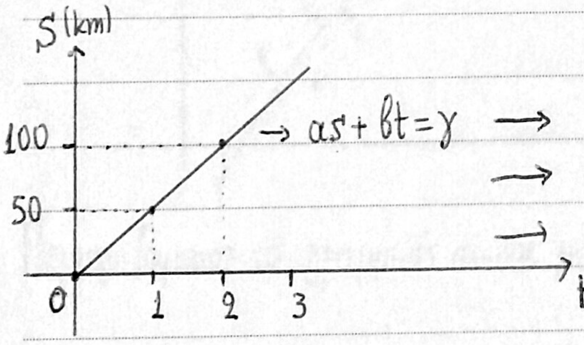
(Κίνηση σταθερή ταχύτητα)

Μεταβλητές:
- Ταχύτητα ($v = 50$ km/h)
- Χρόνος (t)
- Απόσταση (s)

} Χαρακτηριστικά γνωρίσματα

(Ψάχνω σχέση που να συνδέει τις μεταβλητές)

t	0	1	2	3	...	hours
s	0	50	100	150	...	km



$\rightarrow as + bt = \gamma$ $\rightarrow$ για $s = t = 0 \Rightarrow \gamma = 0$

$\rightarrow$ για $s = 50 \rightarrow t = 1$ άρα $50a + b = 0$

$\rightarrow$ για $s = 100 \rightarrow t = 2$ άρα $100a + 2b = 0$

$\Rightarrow s = vt$

Πειραματικά δεδομένα υπακούουν **απόλυτα το μοντέλο** $s = vt$ (αιτιοκρατικό ή ντετερμινιστικό μοντέλο)

Παράδειγμα:

Διερεύνηση σχέσης μεταξύ ύψους - βάρους

X = βάρος, Y = ύψος

Πειραματικά δεδομένα, δηλαδή ζεύγη τιμών των μεταβλητών X, Y

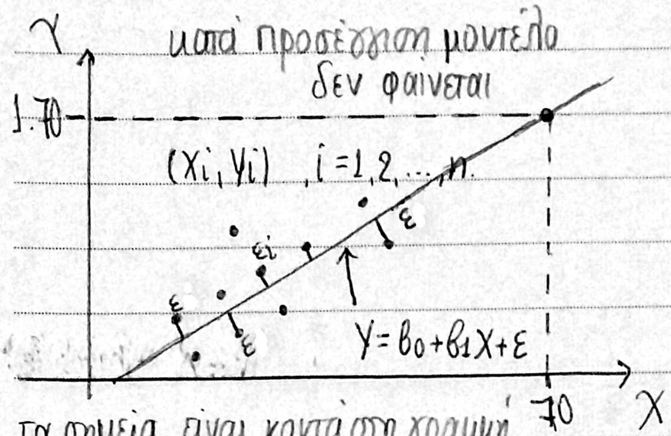
ΕΡΩΤΗΜΑ: Σχετίζεται το X με το Y? Αν σχετίζεται με ποιά σχέση?

Επιλέγω n ανθρώπους, μετρώ την X και την Y σε κάθε ένα από αυτούς.

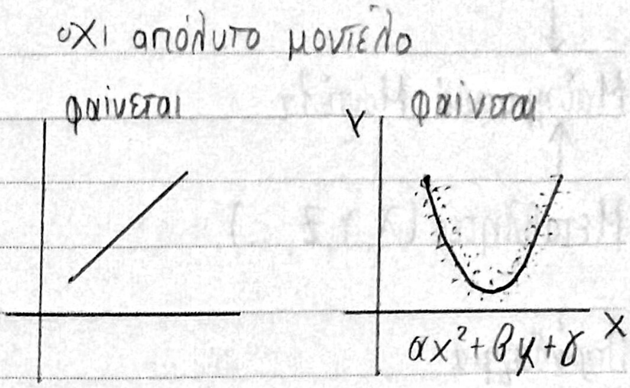
→ τυχαίο δείγμα μεγέθους n πάνω στη X.

X	x_1	x_2	...	x_n
Y	y_1	y_2	...	y_n

X, Y τυχαίες μεταβλητές

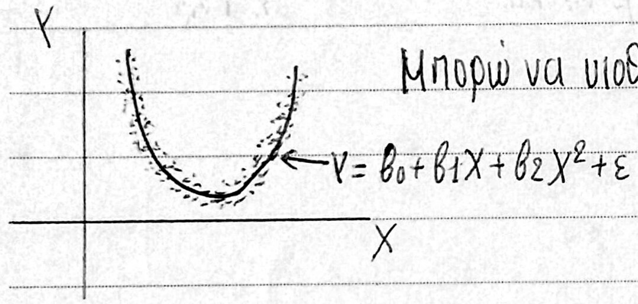


τα σημεία είναι κοντά στη γραμμή αλλά όχι απολύτως πάνω της.

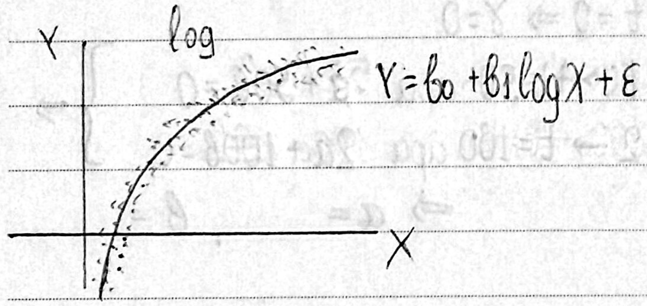


ΕΡΩΤΗΜΑ: Μπορώ να κατασκευάσω ένα μαθηματικό μοντέλο (μαθηματικό τύπο) που να συνδέει τη X με τη Y;

Δηλαδή ένα τύπο που να περικλείει τα σημεία (x_i, y_i), $i=1,\dots,n$;



Μπορώ να υιοθετήσω το μοντέλο της γραμμής? όχι



Σκοπός του μαθήματος είναι η κατασκευή γραμμικού μοντέλου που θα περιγράφει τη σχέση μεταξύ δύο (ή περισσοτέρων) τυχαίων μεταβλητών.

$Y = b_0 + b_1 X + \varepsilon$: μοντέλο απλής γραμμικής παλινδρόμησης.

$$Y = b_0 + b_1 X + \varepsilon \xleftrightarrow[\text{μοντέλο}]{\text{ισοδύναμο}} Y_i = b_0 + b_1 X_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

Απλή γραμμική παλινδρόμηση

Δύο ή περισσότερες τ.μ. π.χ. X, Y, και έστω τ.δ. $x_1, \dots, x_n$ από την X και τ.δ. $y_1, \dots, y_n$ από την Y.

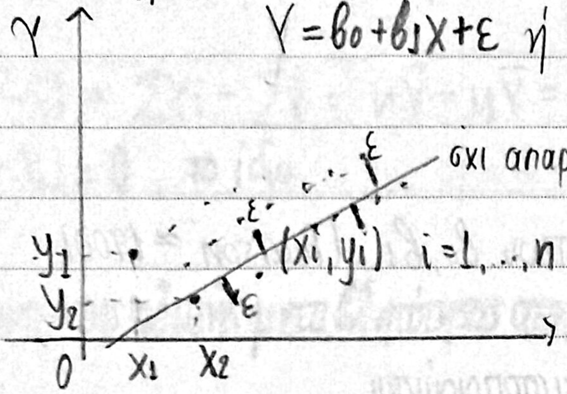
Δεδομένα:

X	$x_1 \dots\dots x_n$
Y	$y_1 \dots\dots y_n$

Ερώτημα: Υπάρχει κάποια σχέση (γραμμική?) που να συνδέει τις X, Y?

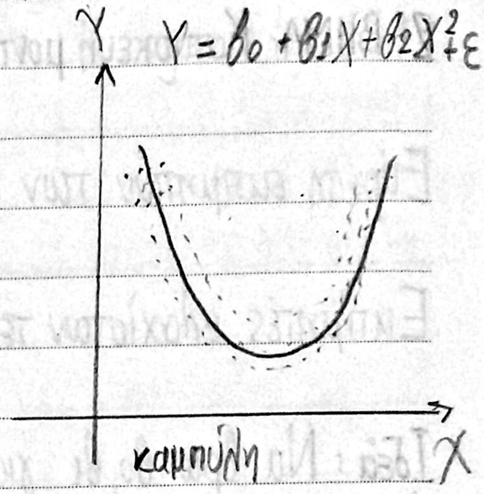
1ο ΒΗΜΑ: Διάγραμμα διασποράς $\begin{cases}\text{απεικόνιση σε σύστημα ορθογωνίων} \\ \text{αξόνων των } (x_i, y_i),\ i=1,\dots,n\end{cases}$

$Y = b_0 + b_1X + \varepsilon$ ή $y_i = b_0 + b_1x_i + \varepsilon_i$



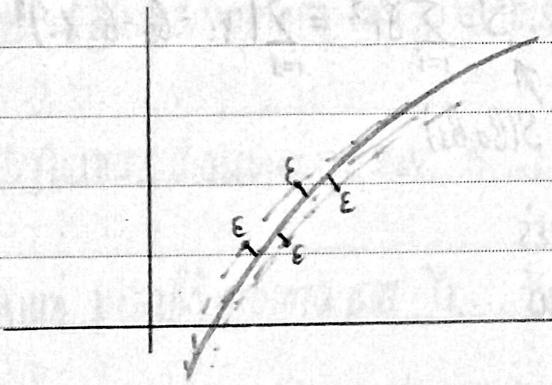
όχι απαραίτητα όλα τα σημεία πάνω στην ευθεία αλλά γύρω της θα παλινδρομούν τα δεδομένα

$Y = b_0 + b_1X + b_2X^2 + \varepsilon$



καμπύλη

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

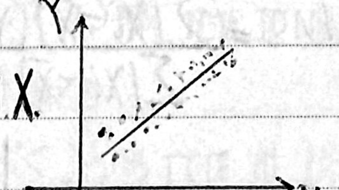


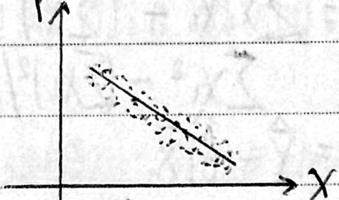
Περιοριζόμαστε σε γραμμικές σχέσεις μεταξύ X, Y.

<u>Το διάγραμμα διασποράς εκφράζει:</u>

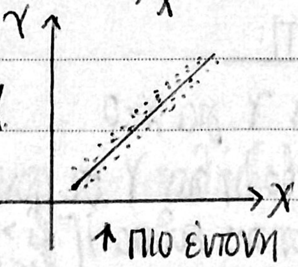
1) Τη μορφή της σχέσης μεταξύ Y και X π.χ. γραμμική σχέση $y = b_0 + b_1X + \varepsilon$

2) Κατεύθυνση της γραμμικής σχέσης π.χ.

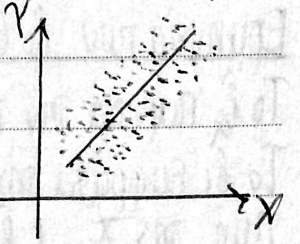




3) Πόσο έντονη είναι η γραμμική σχέση π.χ.
Ο Βαθμός της σχέσης.



↑ πιο έντονη



$Y = b_0 + b_1 X + \varepsilon$ → Μοντέλο απλής γραμμικής παλινδρόμησης (μόνο 2 μεταβλητές)

↕

$Y_i = b_0 + b_1 X_i + \varepsilon_i \quad , i = 1, 2, \dots, n$

ε_i : σφάλματα (αποστάσεις από την ευθεία της απλής γραμμ. παλινδρ.)

X_i : ανεξάρτητη ή επεξηγηματική

Y_i : εξαρτημένη ή απόκριση

b_0, b_1 : παράμετροι μοντέλου

ε : εκφράζει πιθανούς άλλους παράγοντες που επηρεάζουν την Y και τους οποίους δεν λαμβάνω υπόψη.

2ο ΒΗΜΑ: Κατασκευή μοντέλου απλής γραμμικής παλινδρόμησης

Εύρεση εκτιμητών των παραμέτρων $b_0 b_1$

Εκτιμητές ελαχίστων τετραγώνων (ΕΕΤ) των b_0, b_1 (Pearson ≈ 1900)

Ιδέα: Να βρω b_0, b_1 για τα οποία τα ε_i συνολικά ελαχιστοποιούνται.

Αρκεί να βρω τα b_0 και b_1 που ελαχιστοποιούν την $\underset{S(b_0,b_1)}{S} = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$

$$\left.\begin{aligned} \frac{\partial S}{\partial b_0} &= 0 \\ \frac{\partial S}{\partial b_1} &= 0 \end{aligned}\right\} \quad \left.\begin{aligned} n b_0 + b_1 \sum X_i &= \sum Y_i \\ b_0 \sum X_i + b_1 \sum X_i^2 &= \sum X_i Y_i \end{aligned}\right\} \text{κανονικές εξισώσεις}$$

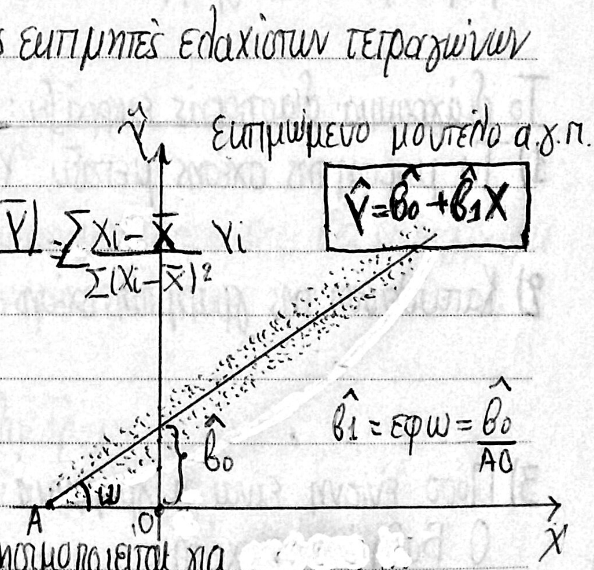
Η επίλυση των κανονικών εξισώσεων οδηγεί στους εκτιμητές ελαχίστων τετραγώνων των b_0 και b_1 που είναι:

$$\hat{b}_1 = \frac{\sum X_i Y_i - (\sum X_i)(\sum Y_i)/n}{\sum X_i^2 - (\sum X_i)^2/n} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \sum \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} Y_i$$

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

Εκτιμώμενο μοντέλο α.γ.π.

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



$\hat{b}_1 = \varepsilon\varphi\omega = \frac{\hat{b}_0}{AO}$

χρησιμοποιείται για προβλέψεις της Y για συγκεκριμένες τιμές X_i

<u>Ερμηνεία των b_0, b_1 (ΕΕΤ)</u>:

- Το $\hat{b}_0$ παριστά την τιμή της Y για X=0
- Το $\hat{b}_1$ εκφράζει την μεταβολή της Y σε μοναδιαία τιμή της X. ($\hat{b}_1$: κλίση της ευθείας) [$\hat{b}_1 > 0$ αύξηση, $\hat{b}_1 < 0$ μείωση]

3º ΒΗΜΑ: Διάγνωση Ορθότητας Μοντέλου

Υπόλοιπα: Ορίζονται ως οι αποκλίσεις του μοντέλου από την πραγματικότητα που το μοντέλο ελπίζει να περιγράψει.

ΟΡΙΣΜΟΣ: $e_i = Y_i - \hat{Y}_i \quad , i = 1, \ldots, n \quad (e_i \rightsquigarrow \varepsilon_i)$

Ιδιότητα των υπολοίπων: $\sum_{i=1}^{n} e_i = 0$ αν το μοντέλο της αχπ. έχει σταθερό όρο $\hat{\beta}_0 \neq 0$.

ΑΠΟΔΕΙΞΗ: $\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (Y_i - \hat{Y}_i) = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = \sum_{i=1}^{n} (Y_i - (\bar{Y} - \hat{\beta}_1 \bar{X}) - \hat{\beta}_1 X_i) =$
$= \sum_{i=1}^{n} (Y_i - \bar{Y}) + \hat{\beta}_1 \sum_{i=1}^{n} (\bar{X} - X_i) = 0$

- $\sum (Y_i - \bar{Y}) = \sum Y_i - \sum \bar{Y} = n\bar{Y} - n\bar{Y} = 0$
- $\sum (\bar{X} - X_i) = 0$ το ίδιο.