Διάλεξη 1η 8/10/2019

Παλινδρόμηση και Ανάλυση Διακύμανσης

· Μαθηματικά

↓

Μαθηματικά μοντέλα

πχ Ευθύγραμμη ομαλή κίνηση (κίνηση με σταθερή ταχύτητα)

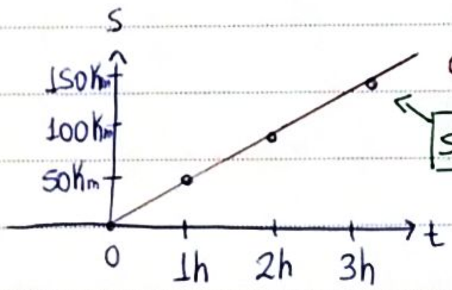
↓

Εμφανίζονται τρεις παράγοντες ή τρεις μεταβλητές

- Ταχύτητα U
- το διάνυσμα S
- ο χρόνος t.

Παράδειγμα 1

Ξέρω $U = 50\ km/h$



$as + \beta t = \gamma$

$\boxed{s = Ut}$ ← Στην ευθ. ομαλ. κιν. το διάστημα είναι ανάλογο του χρόνου t.

Για $t=0$ το $s=0$ άρα $\gamma=0$

Για $t=1$ το $s=50$ άρα $\gamma=0$ όμως $\gamma = 50a+\beta \Rightarrow 50a+\beta=0$

Για $t=2$ το $s=100$ -//- $100a+2\beta=0$

$\Rightarrow a=1,\ \beta=-50$

Παράδειγμα 2

Βάρος, Ύψος

Ενδιαφέρει η μελέτη του βάρους και του ύψους ανθρώπων

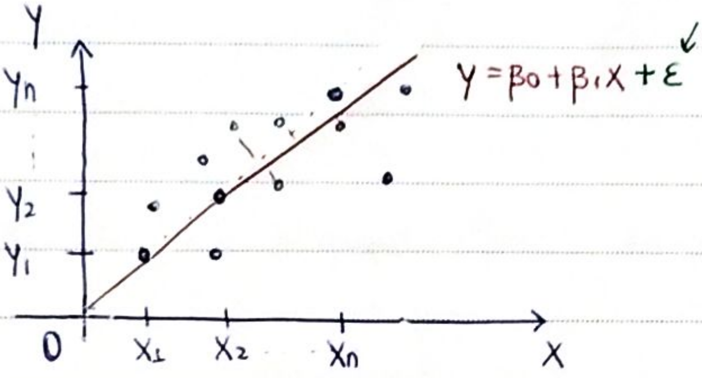
Έστω X = βάρος, Y = ύψος

▸ Ερώτημα: Υπάρχει σχέση μεταξύ X και Y;

Για να απαντήσω το ερώτημα, θα πρέπει να έχω δεδομένα.

Έστω παρατηρήσεις

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

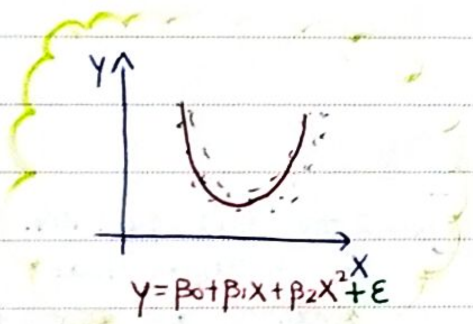


- Δεν βρίσκονται όλα απαραίτητα στην ίδια ευθεία όπως πχ 1.

Υπάρχει μεταβλητότητα

Σκόρπια σημεία που αποκλίνουν από την ευθεία

Θα υπάρχει μια γραμμική σχέση όμως υπάρχουν και σφάλματα.



▸ Μοντέλο Απλής Γραμμικής Παλινδρόμησης (αγπ)

Έστω ποσοτικές τ.μ X και Y και έστω τ.δ (τυχαίο δείγμα) $X_1, \dots, X_n$ από την X και $Y_1, \dots, Y_n$ από την Y

X	$X_1, X_2, \dots, X_n$
Y	$Y_1, Y_2, \dots, Y_n$

Το ερώτημα στο οποίο απαντά η αγπ είναι:

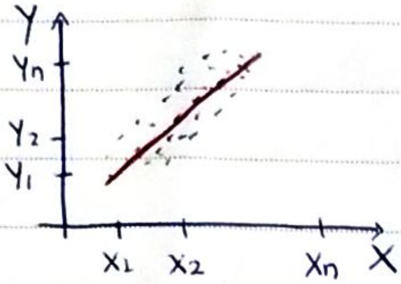
Υπάρχει γραμμική σχέση μεταξύ των Y και X;

Ποια είναι αυτή η σχέση;

Στην προσπάθεια κατασκευής γραμμικού μοντέλου που συνδέει τις X και Y

→

<u>1ο ΒΗΜΑ</u> Διάγραμμα Διασποράς



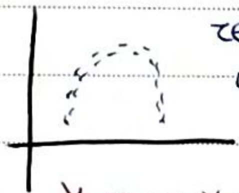
(x_i, y_i) $i=1,2,\dots,n$

Αν από το διάγραμμα διασποράς φαίνεται ότι τα σημεία (x_i,y_i) $i=1,\dots,n$ προσαρμόζονται σε μια υποτιθέμενη ευθεία, τότε αισιοδοξώ ότι το μοντέλο της αγπ μπορεί να είναι υποσχόμενο.

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

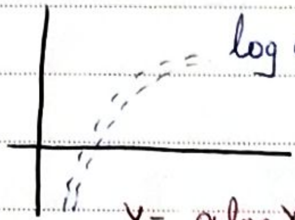
i) Για τη μορφή της σχέσης των X, Y

$Y=\beta_0+\beta_1 X$



τετραγωνική σχέση

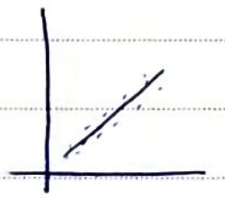
$Y=\beta_0+\beta_1 X+\beta_2 X^2$



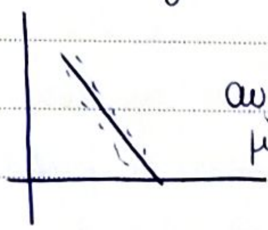
log σχέση

$Y=\alpha \log X+\beta$

ii) Αν είναι γραμμική δίνει πληροφορίες για την κατεύθυνση της σχέσης.

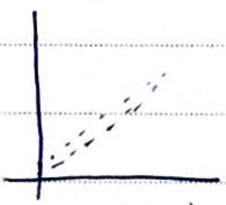


αυξάνει η μια μεταβλητή αυξάνει και η άλλη

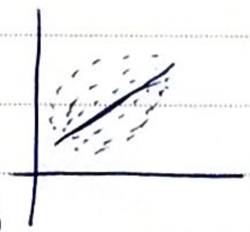


αυξάνει η μια μειώνεται η άλλη

iii) Πόσο έντονη είναι η σχέση, δηλ πόσο κοντά είναι τα (x_i,y_i) $i=1,\dots,n$ στο μοντέλο (γραμμικό ή άλλο) που τείνει να τα περιγράψει.



πολύ έντονη γραμμική σχέση



Αν το διάγραμμα διασποράς υπόσχεται γραμμικότητα τότε:

2ο ΒΗΜΑ Κατασκευή και μελέτη μοντέλου αγπ

$Y = \beta_0 + \beta_1 X + \varepsilon$ ← αγπ

Εξαρτημένη μεταβλητή ή απόκριση

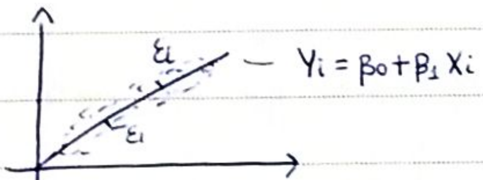
ανεξάρτητη μεταβλητή ή Επεξηγηματική μεταβλητή.

σφάλματα του μοντέλου δηλ τις αποκλίσεις των σημείων (x_i, y_i) από την ευθεία της αγπ.

Παράμετροι του μοντέλου

Στην πράξη: X είναι η μεταβλητή που είναι υπό τον έλεγχο του ερευνητή.
Y είναι η μεταβλητή που επιθυμούμε να δούμε πως μεταβάλλεται όταν μεταβάλλεται η X.

Ισοδύναμο μοντέλο αγπ $\quad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$

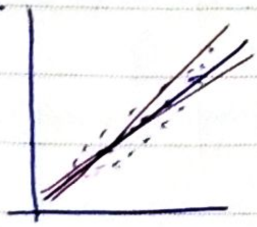


Αλλά β_0, β_1 είναι άγνωστες

Άρα για να αξιοποιηθεί το μοντέλο τα β_0 και β_1 πρέπει να εκτιμηθούν.

- Εκτιμητές Ελαχίστων Τετραγώνων (ΕΕΤ) των β_0, β_1 (Pearson ~1900)

Ιδέα:



Βρες τα β_0, β_1 έτσι ώστε τα σφάλματα ε_i $i=1,\dots,n$ να είναι τα μικρότερα δυνατά.

ή

η $S = \sum_{i=1}^{n} \varepsilon_i^2$ είναι ελάχιστη ως προς β_0 και β_1.

Άρα οι ΕΕΤ των β_0 και β_1 είναι εκείνοι που ελαχιστοποιούν ως προς β_0 και β_1 την $S = \sum_{i=1}^{n} \varepsilon_i^2$

ή

$$\hat{\beta}_0, \hat{\beta}_1 \equiv \arg\min_{\beta_0, \beta_1} S$$

$$= \arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \varepsilon_i^2$$

Εύρεση των ΕΕΤ των β_0, β_1

$$S = \sum_{i=1}^{n} \varepsilon_i^2 \underset{\text{μοντέλο}}{\overset{\text{αχη}}{=}} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\left.\begin{array}{l} \dfrac{\partial S}{\partial \beta_0} = 0 \\ \\ \dfrac{\partial S}{\partial \beta_1} = 0 \end{array}\right\} \text{λύνω το σύστημα} \ \dots \text{(πράξεις)} \Rightarrow \begin{array}{l} n\beta_0 + \beta_1 \sum X_i = \sum Y_i \\ \\ \beta_0 \sum X_i + \beta_1 \sum X_i^2 = \sum X_i Y_i \end{array}$$

Κανονικές Εξισώσεις.

Οι ΕΕΤ των β_0 και β_1 είναι:

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

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

Εκτιμώμενο Μοντέλο αγπ.

$$\hat{Y} = \hat{\beta_0} + \hat{\beta_1} X$$