

Μέθοδος ελαχίστων
τετραγώνων

Χρησιμοποιείται:

- Όταν έχουμε ένα σύνολο πειραματικών δεδομένων που δεν βρίσκονται ακριβώς στην ίδια ευθεία.
- Όταν γνωρίζουμε μόνο μια σειρά από πειραματικές τιμές μεγεθών που περιγράφουν ένα φαινόμενο και όχι την ακριβή μαθηματική σχέση τους.

Χρησιμοποιείται:

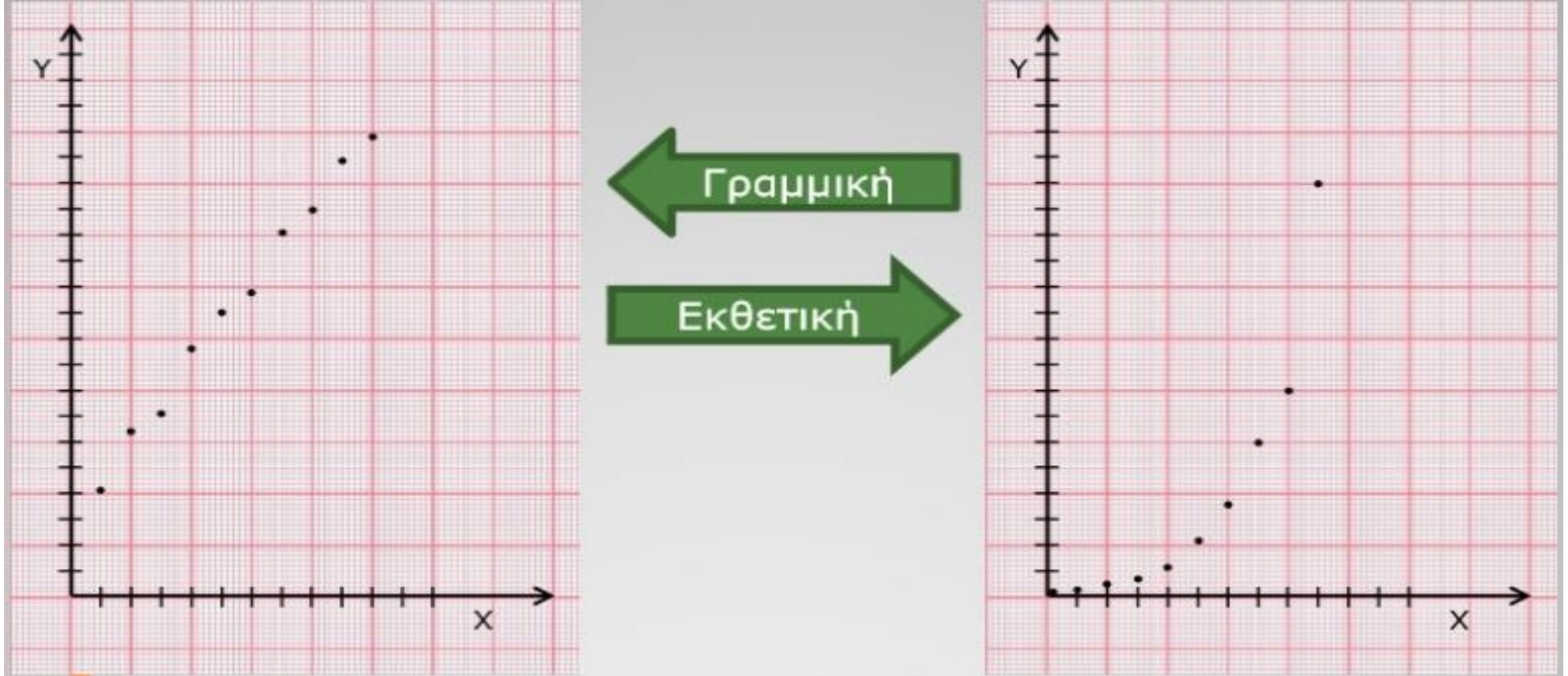
- Κατά τη μελέτη ενός φαινομένου, προσπαθούμε να προσδιορίσουμε τη μορφή της άγνωστης μαθηματικής σχέσης, στην οποία ταιριάζουν καλύτερα τα πειραματικά μας δεδομένα, ελέγχοντας μια σειρά γνωστών σχέσεων.
- Τα μεγέθη που μεταβάλλονται ανεξάρτητα από την πορεία του φαινομένου που μελετάται είναι οι *ανεξάρτητες μεταβλητές* ενώ εκείνα που μεταβάλλονται συναρτήσει αυτών και περιγράφουν το φαινόμενο είναι οι *εξαρτημένες μεταβλητές*.

Οι κυριότερες σχέσεις που εξετάζονται είναι οι:

- ◦ Γραμμική: $y = a + bx$ (ευθεία γραμμή)
- ◦ Πολυωνυμική: $y = c_0 + c_1x + c_2x^2 + \dots + c_nx^n$
- ◦ Σχέση δύναμης: $y = ax^b$
- ◦ Εκθετική: $y = ae^{bx}$
- ◦ Λογαριθμική: $y = a + b \ln x$

Στόχος της MET είναι ο προσδιορισμός των σταθερών συντελεστών a , b , c_k της σχέσης που επιλέξαμε για να περιγράψουμε το φαινόμενο.

- Αρχικά, γνωρίζοντας τις πειραματικές τιμές, σχεδιάζουμε πρόχειρα και παρατηρούμε τη μορφή του διαγράμματος, προσπαθώντας να μαντέψουμε το είδος της καμπύλης που μπορεί να ταιριάζει καλύτερα σε αυτά.



- Τότε χρειάζεται να χαράξουμε μια ευθεία γραμμή που να τα εκπροσωπεί όσο γίνεται καλύτερα. Ο υπολογισμός της ευθείας που προσεγγίζει καλύτερα τα σημεία μας γίνεται με τη μέθοδο των Ελαχίστων Τετραγώνων (Least Squares) και η διαδικασία λέγεται Γραμμική Παλινδρόμηση (Linear Regression).
- Η ευθεία αυτή συχνά δεν περιλαμβάνει κανένα από τα σημεία του διαγράμματος. Είναι όμως η μοναδική ευθεία που ικανοποιεί το κριτήριο των ελαχίστων τετραγώνων το οποίο απαιτεί:
- Το άθροισμα των τετραγώνων των αποκλίσεων όλων των σημείων από την ευθεία να είναι το ελάχιστο δυνατό.

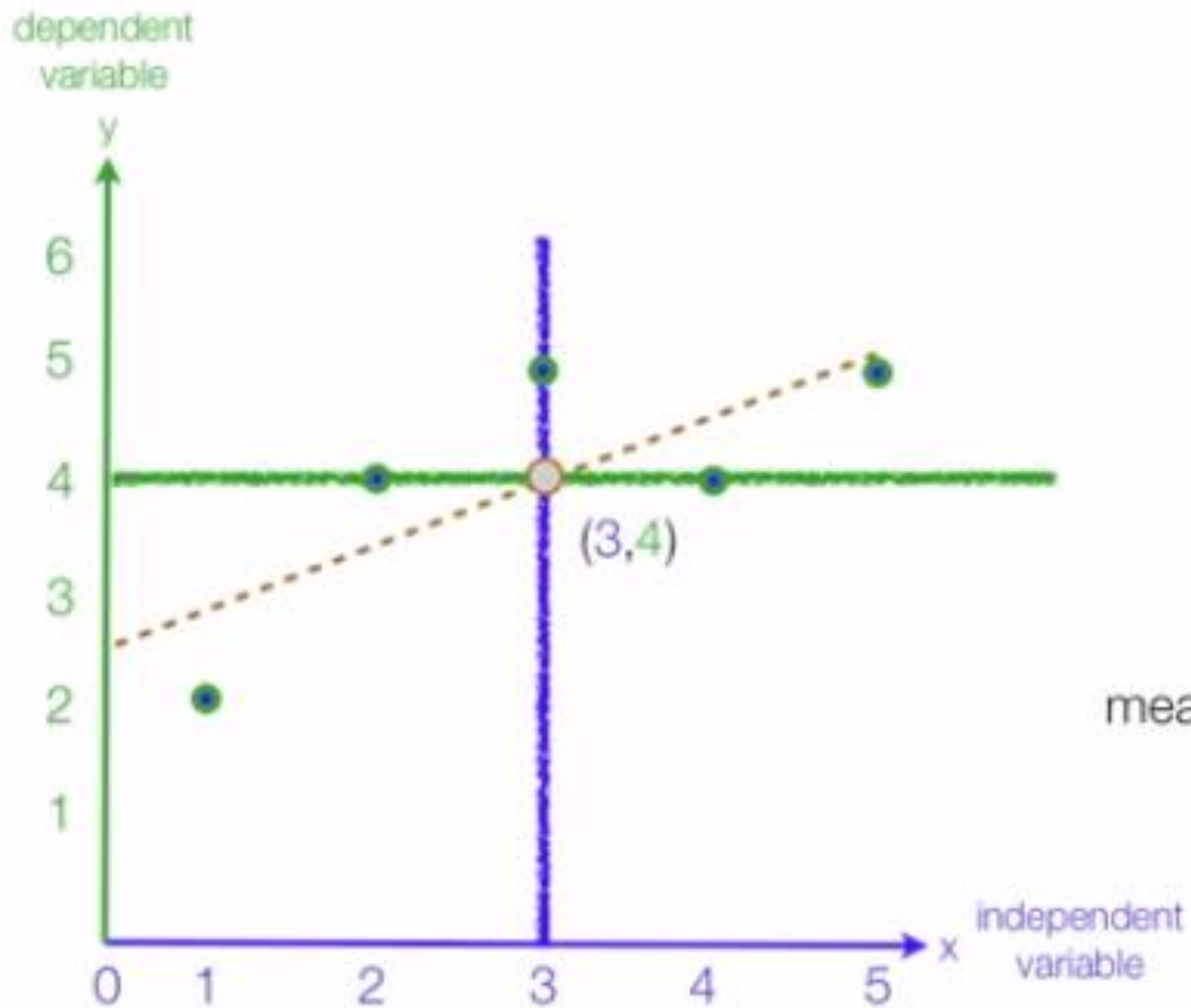
Έστω ότι πρόκειται για γραμμική σχέση της μορφής: **$y=a+bx$** .

Οι τύποι υπολογισμού των συντελεστών a και b είναι αντίστοιχα:

$$\begin{aligned} a &= \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{D}, & \delta a &= \sigma_y \sqrt{\frac{\sum_{i=1}^n x_i^2}{D}} \\ b &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{D}, & \delta b &= \sigma_y \sqrt{\frac{n}{D}} \\ D &= n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 & \text{και} & \sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_i - a - bx_i)^2}{n-2}}, \end{aligned}$$

όπου n : πλήθος των μετρήσεων και δa , δb τα σφάλματα των a και b .

Παράδειγμα



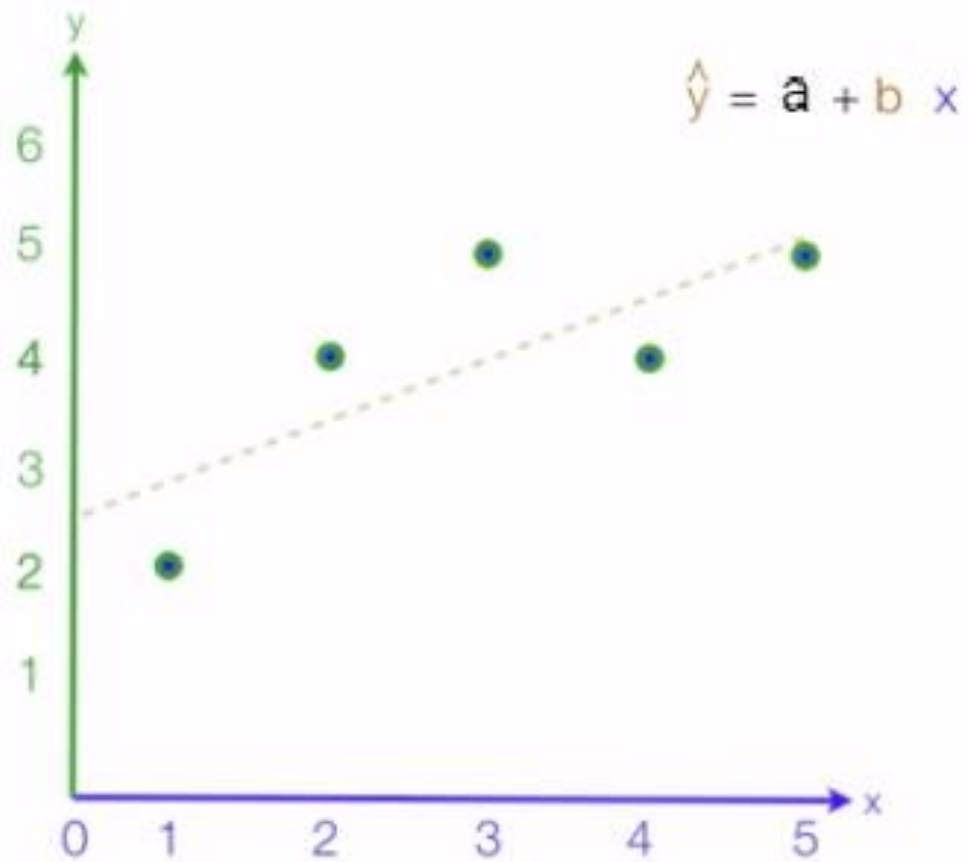
independent variable	dependent variable
x	y
1	2
2	4
3	5
4	4
5	5

mean

3

4

Παράδειγμα



mean

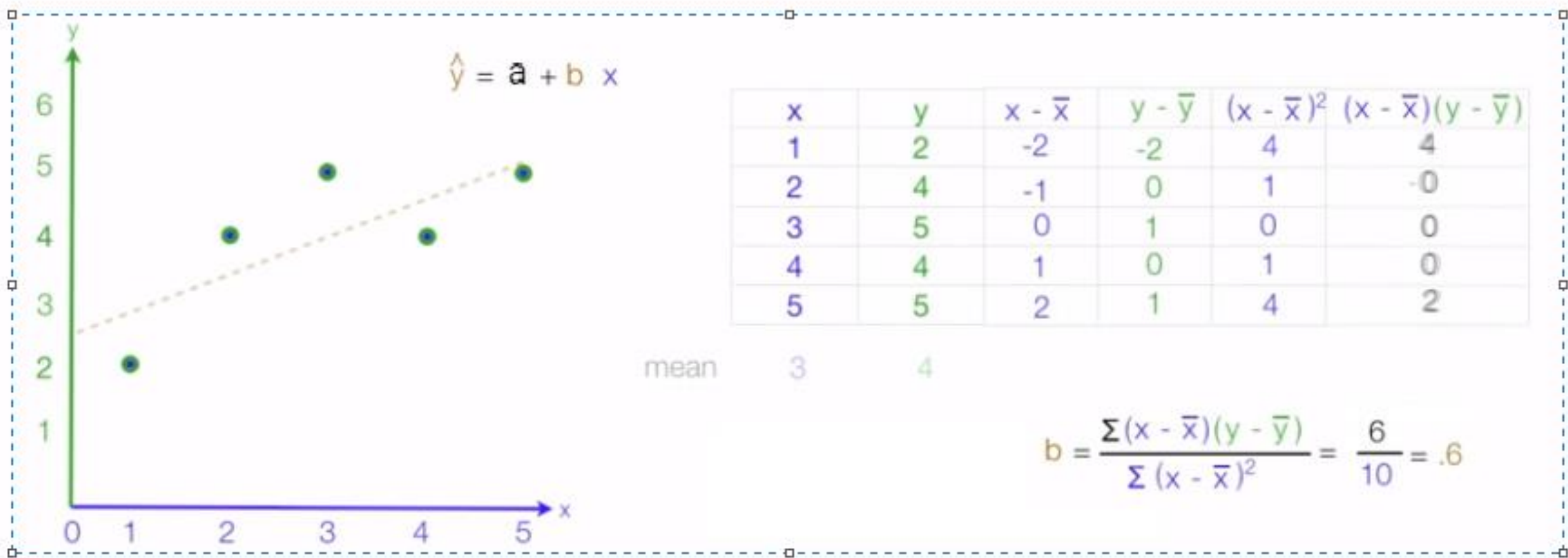
x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
1	2	-2	-2	4	4
2	4	-1	0	1	-0
3	5	0	1	0	0
4	4	1	0	1	0
5	5	2	1	4	2

3

4

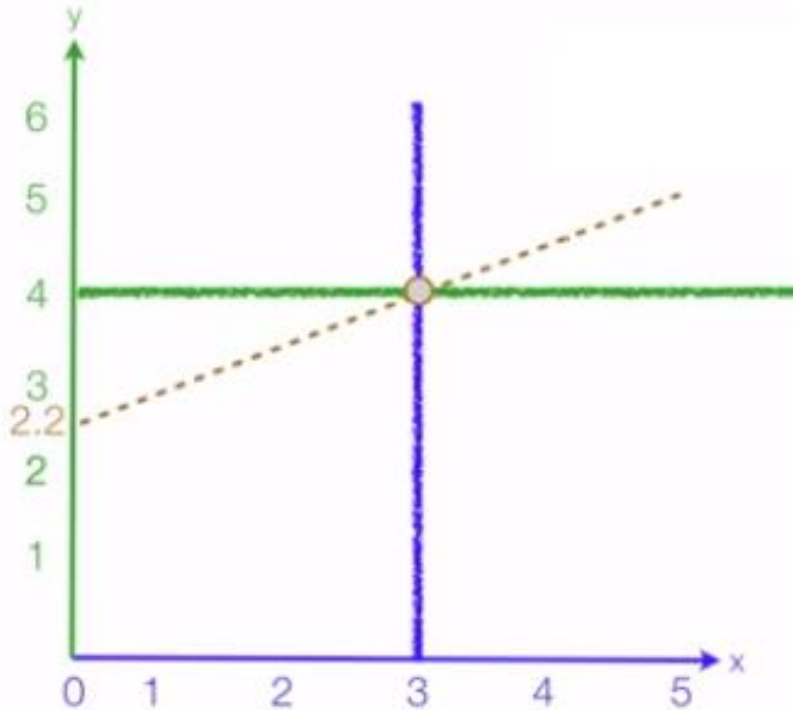
Παράδειγμα

Τώρα πρέπει να υπολογίσω την κλίση της ευθείας της γραμμικής παλινδρόμησης, δηλαδή το b :



Παράδειγμα

- Στη συνέχεια υπολογίζουμε το a , που αντιστοιχεί στο σημείο του άξονα y από το οποίο ξεκινά η ευθεία:
- Η γραμμή παλινδρόμησης πρέπει να περνά από το σημείο που τέμνονται οι μέσες τιμές, $(3,4)$
- Άρα $y=4$, $4 = a + (0.6 \cdot 3)$, **$a=2.2$**



x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
1	2	-2	-2	4	4
2	4	-1	0	1	0
3	5	0	1	0	0
4	4	1	0	1	0
5	5	2	1	4	2

mean 3 4

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = \frac{6}{10} = .6$$