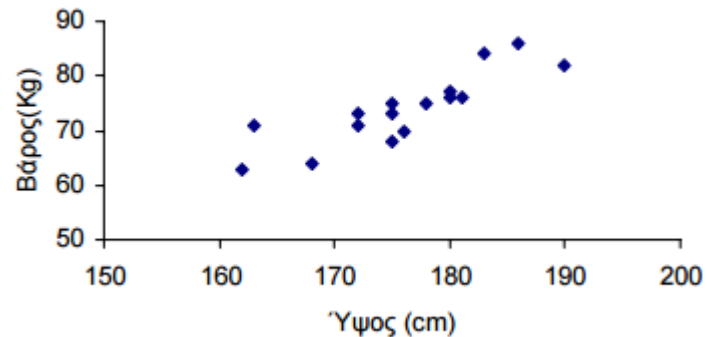


ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (LINEAR REGRESSION)

Πρόβλημα: η σχέση δύο ή περισσότερων τυχαίων μεταβλητών.

Ας πούμε η σχέση μεταξύ ύψους και βάρους των φοιτητών



1. Ερευνάμε αν υπάρχει μια σχέση μεταξύ δύο μεταβλητών
2. Αν **ναι**, καθορίζουμε αυτή τη σχέση.

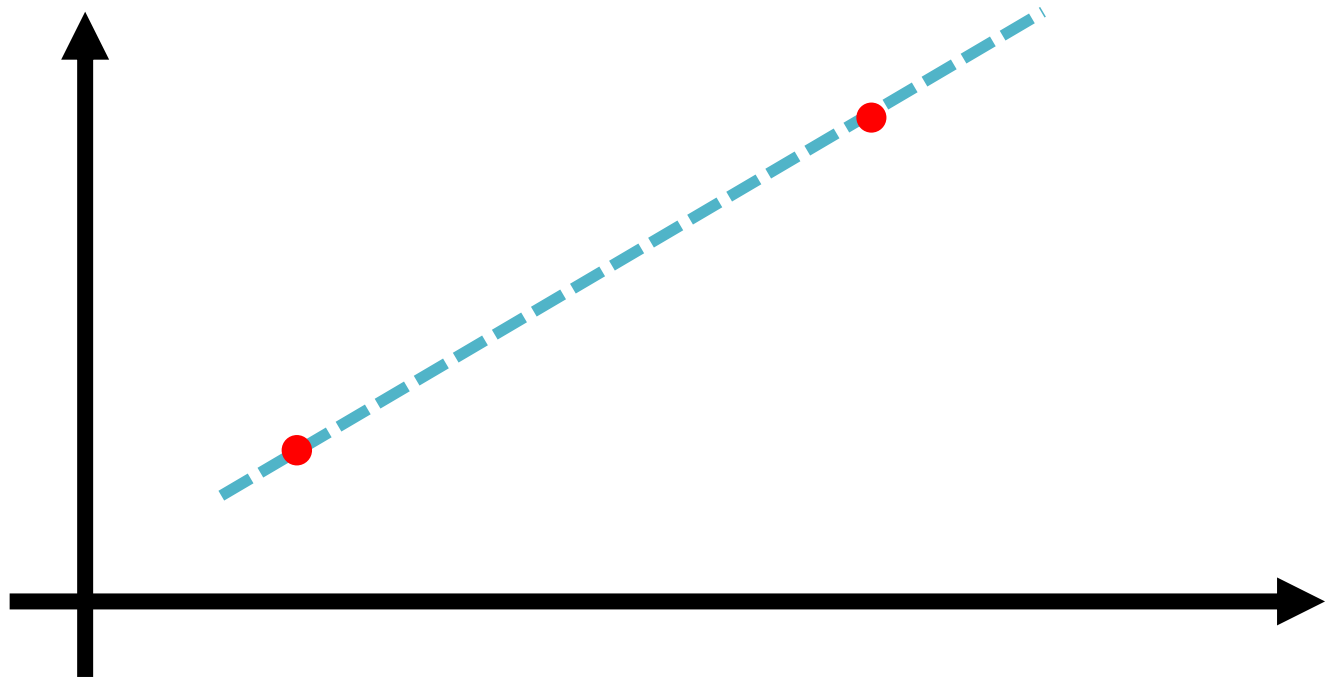
Κατασκευή της μαθηματικής
εξίσωσης που περιγράφει τη σχέση

Η απλούστερη μορφή τέτοιας σχέσης
είναι η ευθεία

$$Y = aX + b$$

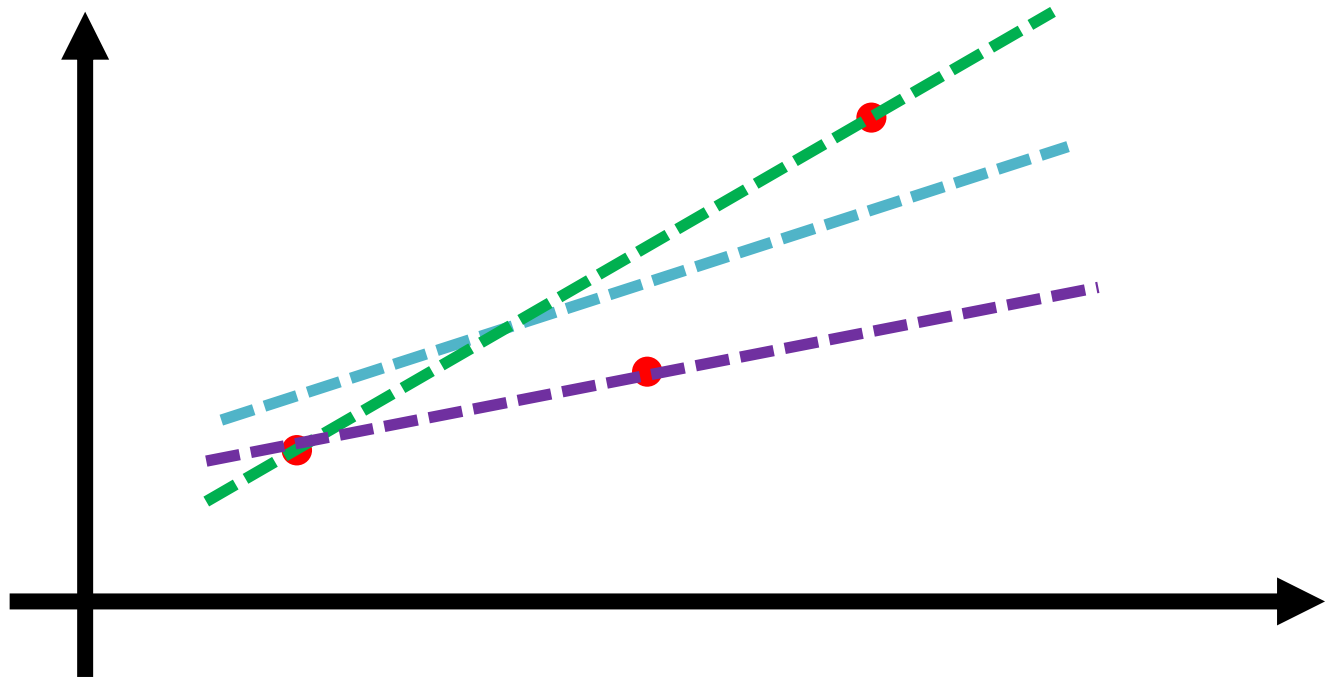
Πρέπει να βρούμε τα a και b

Για να σκεφτούμε ...



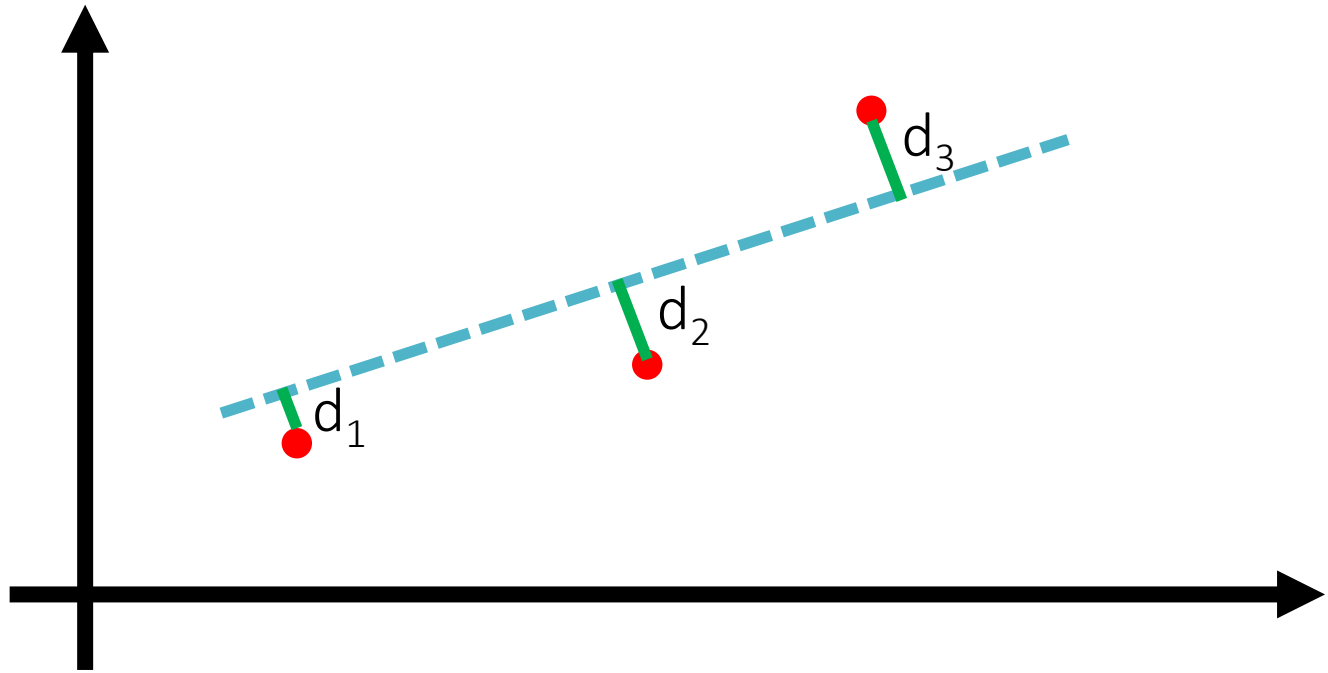
Δυο σημεία = μια και μοναδική ευθεία

Για να σκεφτούμε ...



Τρία σημεία = ποιά ευθεία???

Για να σκεφτούμε ...



Θέλω $d_1 + d_2 + d_3$ να είναι **ΕΛΑΧΙΣΤΟ**

Οπότε γενικά ...

$$a = \frac{N \sum xy - (\sum x)(\sum y)}{N \sum x^2 - (\sum x)^2}$$

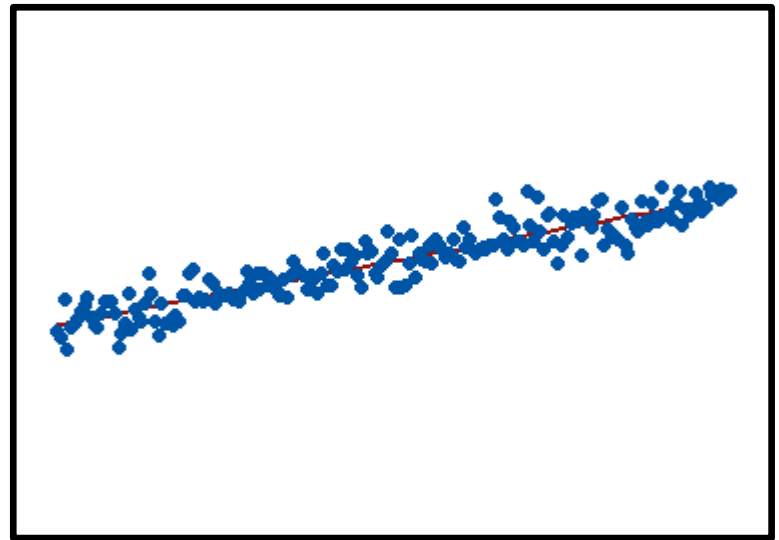
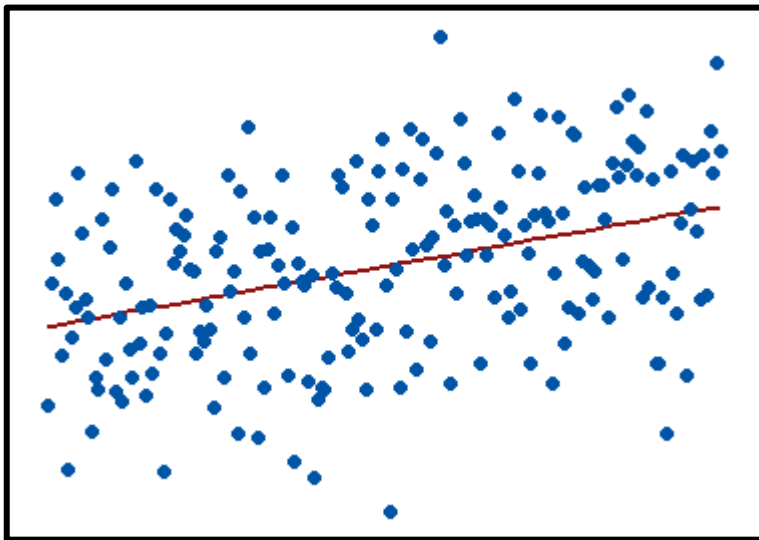
$$b = \frac{\sum y}{N} - a \frac{\sum x}{N}$$

Είναι «καλή» αυτή η ευθεία;;;

Συντελεστής προσδιορισμού $r^2 =$
το ποσοστό της συνολικής
μεταβολής του Y που οφείλεται
στη μεταβολή του X .

$$r^2 = \left[\frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}} \right]^2$$

Είναι «καλή» αυτή η ευθεία;;;



Όσο πλησιέστερα είναι στο **1**,
τόσο **ισχυρότερη** είναι η γραμμική
συσχέτιση των Y και X

Για τιμές κοντά στο **0** έπεται ότι
δεν υπάρχει καλή γραμμική
συσχέτιση.