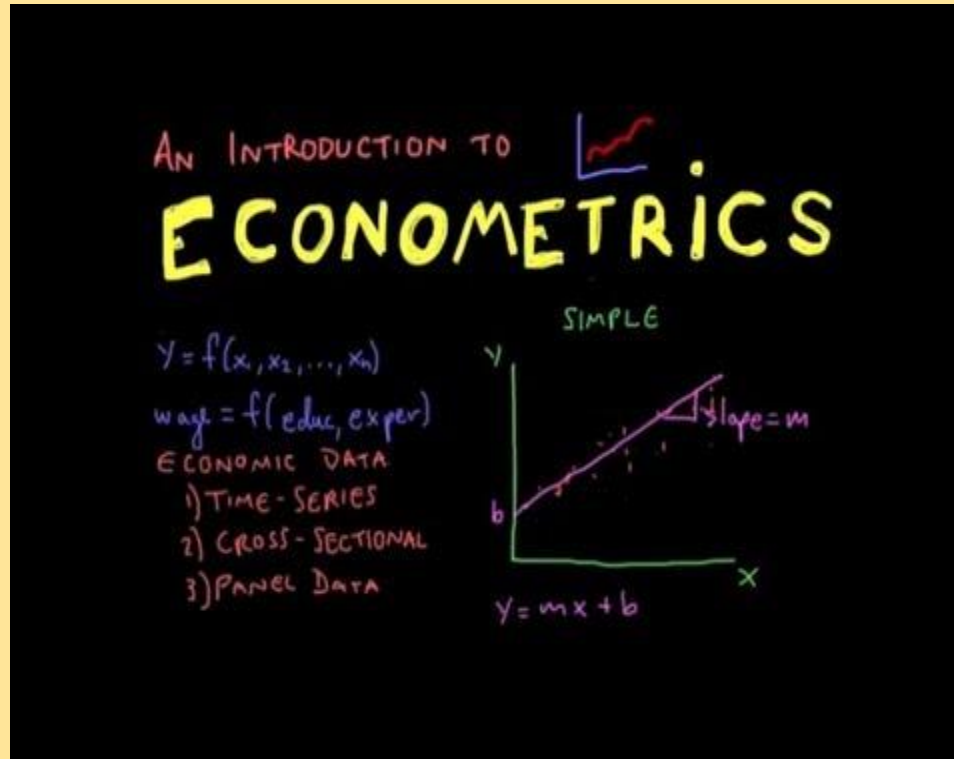


Οικονομικό Πανεπιστήμιο Αθηνών

Μάθημα: "Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρησιακά Προβλήματα"

Συνάντηση: 1η - 2η



Δρ. Τρύφωνας Λεμοντζόγλου

Οικονομικό Πανεπιστήμιο Αθηνών

Μάθημα: “Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρησιακά Προβλήματα”

Συνάντηση: 1η - 2η

ΤΟ ΥΛΙΚΟ ΜΑΣ ΣΤΟ E-CLASS:

<https://eclass.aueb.gr/courses/LOXR637/>

ΩΡΕΣ ΓΡΑΦΕΙΟΥ:

Τρίτη και Τετάρτη (17.00-18.00)
Πατησίων 95 (3ος όροφος)
κατόπιν συνεννόησης

ΕΠΙΚΟΙΝΩΝΙΑ:

trifonaslemon@aueb.gr

trifonaslemon@gmail.com

ΠΡΟΣΩΠΙΚΟΙ ΙΣΤΟΤΟΠΟΙ:

<https://economichistorydotblog.wordpress.com>

<http://trifonaslemon.blogspot.com>

Φυσικός και Κοινωνικός Πραγματικός Κόσμος

Η στατικότητα του φυσικού κόσμου

Κινηματική Εξίσωση (Φυσική):

$$v = u + a \cdot t$$



όπου

v: η ταχύτητα ενός σώματος κατά τον χρόνο **t**

u: η αρχική ταχύτητα του σώματος

a: η επιτάχυνση

t: ο χρόνος

Συγκεκριμένες συνθήκες

→ οδηγούν σε συγκεκριμένα αποτελέσματα

π.χ για **u=10**, **a=2** και **t=5** θα λαμβάνουμε πάντα το ίδιο **v**.

Η ρευστότητα του κοινωνικού κόσμου

Παράδειγμα

Εξίσωση Κατανάλωσης (Οικονομικά):

$$\Psi = \alpha + \beta * X$$

όπου

Ψ : η μηνιαία κατανάλωση ενός προϊόντος (σε ευρώ)

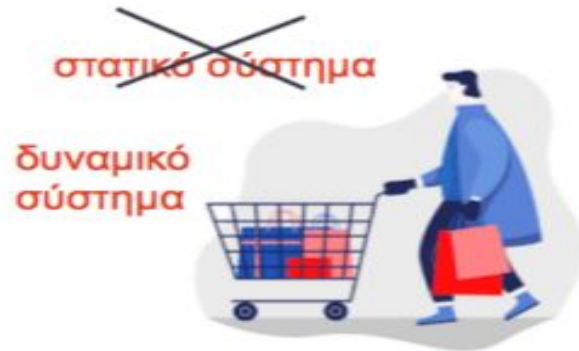
α : το ελάχιστο επίπεδο κατανάλωσης (κρατική μέριμνα)

β : ο συντελεστής που συνδέει κατανάλωση & εισόδημα

X : το μηνιαίο εισόδημα (σε ευρώ)

Συγκεκριμένες συνθήκες (βλ. ύψος εισοδήματος)

→ **ΔΕΝ** οδηγούν κατ' ανάγκη
σε συγκεκριμένα αποτελέσματα



Ο ρόλος του **εἰ** (διαταρακτικός όρος)

Παράδειγμα

Εξίσωση Κατανάλωσης (Οικονομικά):

$$\Psi = \alpha + \beta * X + \epsilon$$

όπου

Ψ: η μηνιαία κατανάλωση ενός προϊόντος (σε ευρώ)

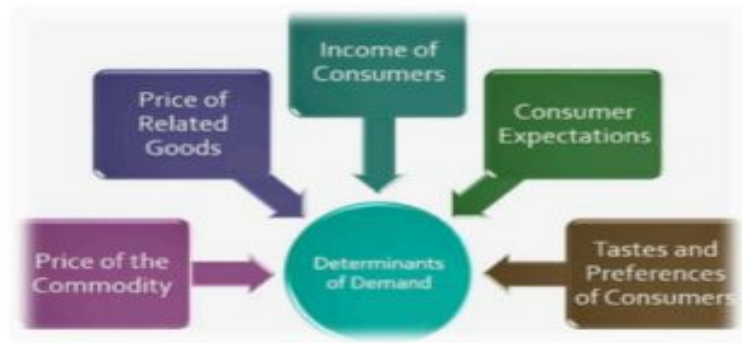
α: το ελάχιστο επίπεδο κατανάλωσης (κρατική μέριμνα)

β: ο συντελεστής που συνδέει κατανάλωση & εισόδημα

X: το μηνιαίο εισόδημα (σε ευρώ)

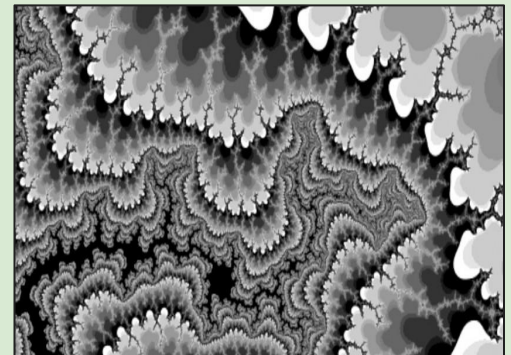
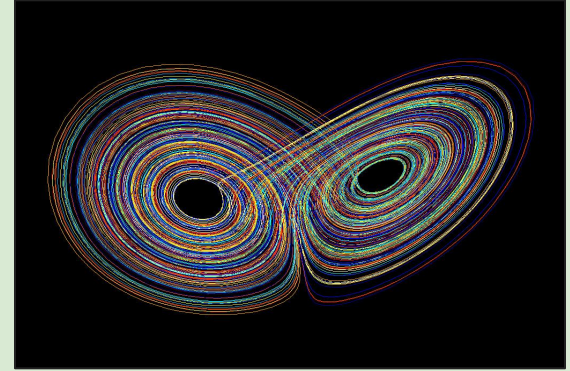
ε: τυχαίοι παράγοντες (διαταρακτικός όρος)

π.χ όλα τα άτομα με εισόδημα **X=900** ευρώ **ΔΕΝ** αγοράζουν τις ίδιες ποσότητες ενός προϊόντος



Στη φύση (πραγματικότητα) σπάνια συναντάμε γραμμικά μοτίβα!!!

Μερικά Χαοτικά Μοτίβα Συμπεριφοράς στη Φύση



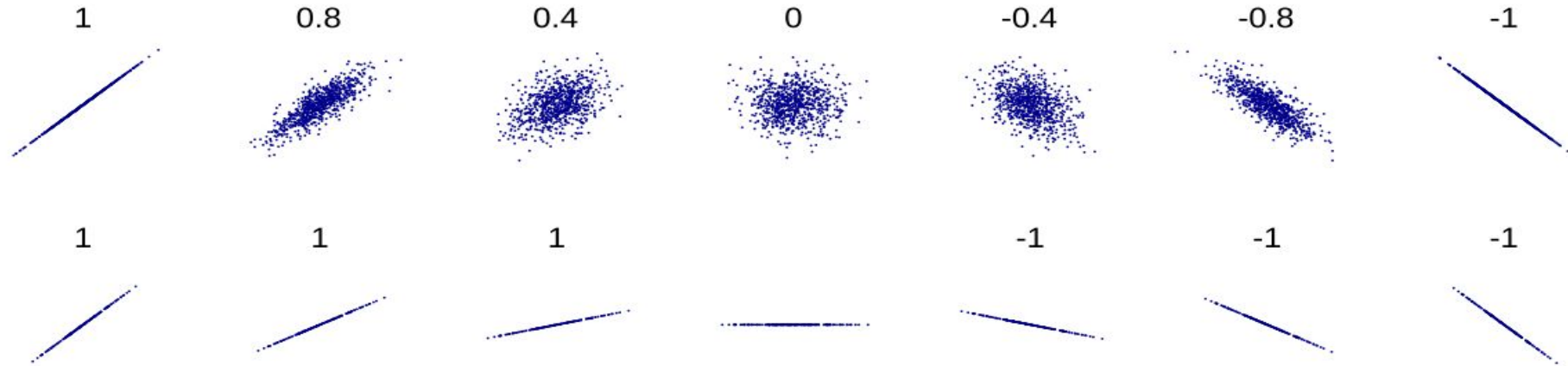
Γραμμικότητα (μια μάλλον φανταστική υπόθεση)

Ρίξτε μια ματιά στην επόμενη διαφάνεια !!!

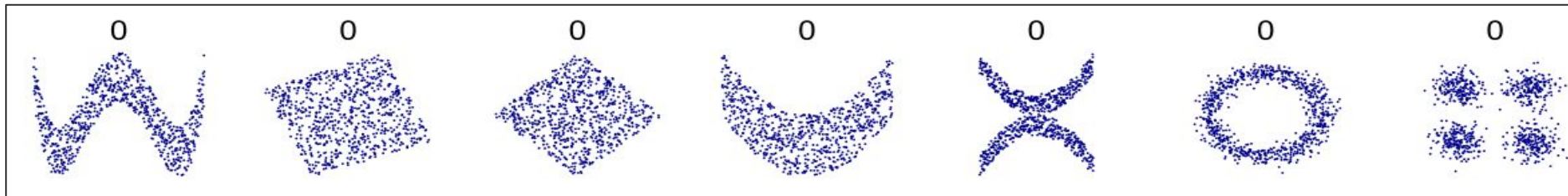
Στην πρώτη γραμμή ξεκινάμε στα αριστερά με μια **ΤΕΛΕΙΑ ΘΕΤΙΚΗ** σχέση → ο συντελεστής γραμμικής συσχέτισης είναι ο **$\rho=+1$**

Προχωρώντας προς τα δεξιά η σχέση αυτή γίνεται **ολοένα και λιγότερο έντονα θετική**. Στη μέση της πρώτης γραμμής βλέπουμε μια σχέση με **$\rho=0$** (δεν υπάρχει κάποια γραμμική σχέση). Προχωρώντας προς τα δεξιά η σχέση γίνεται **ολοένα και πιο έντονα αρνητική ($\rho<0$)** εως ότου να πάρει την τιμή **$\rho=-1$** (**ΤΕΛΕΙΑ ΑΡΝΗΤΙΚΗ** σχέση)

Γραμμικότητα (μια μάλλον φανταστική υπόθεση)



Τέλειες (αλλά μη γραμμικές) σχέσεις



Γραμμικότητα (μια μάλλον φανταστική υπόθεση)

Τι συμβαίνει στη δεύτερη γραμμή της προηγούμενης διαφάνειας;

Βλέπουμε **ΤΕΛΕΙΑ ΘΕΤΙΚΕΣ** σχέσεις με **$\rho=+1$**

και **ΤΕΛΕΙΑ ΑΡΝΗΤΙΚΕΣ** σχέσεις με **$\rho=-1$**

Στη μέση της δεύτερης γραμμής συντάμε μια **μη γραμμική σχέση** με **$\rho=0$**

Στην τρίτη γραμμή συναντάμε **ΤΕΛΕΙΕΣ** αλλά **μη γραμμικές** σχέσεις με **$\rho=0$** (δες πίσω...)

Πραγματικά διαγράμματα διασποράς με τις τιμές (Ψ_i , χ_i)

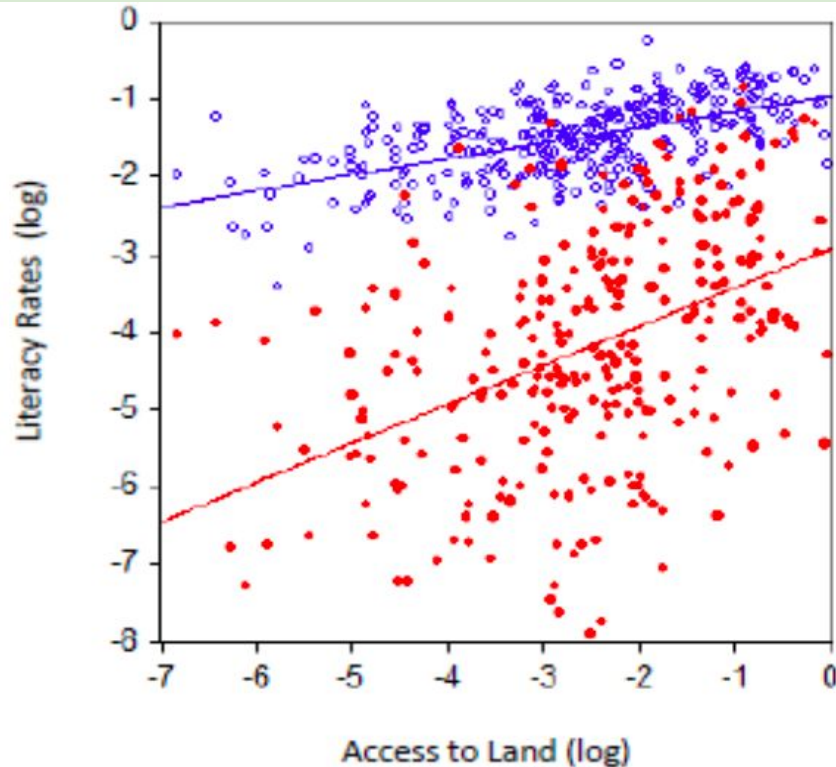
Γραμμικότητα (μια μάλλον φανταστική υπόθεση)

Ρίξτε μια ματιά στην διαφάνεια που ακολουθεί !!!

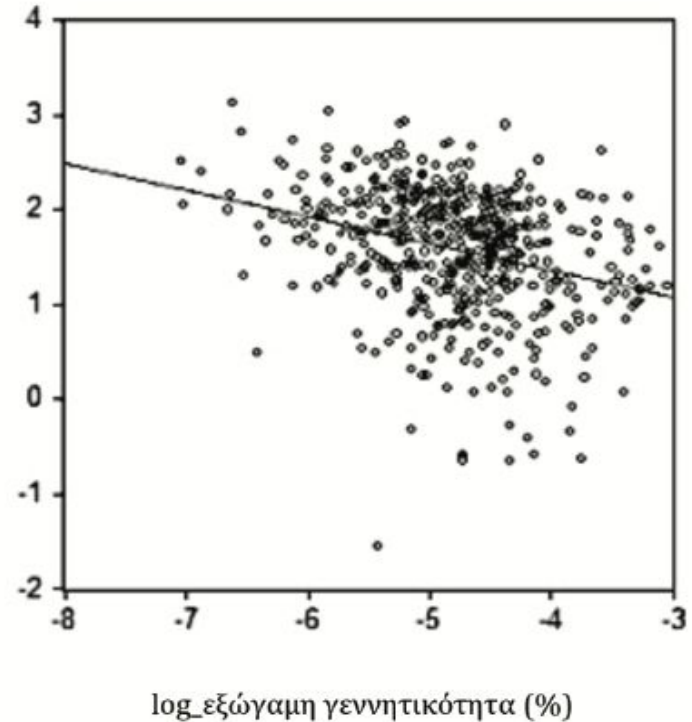
Στα αριστερά φαίνεται η σχέση που συνδέει τα **ποσοστά του εγγραμματος** για τους άνδρες και τις γυναίκες (**Ψ : εξαρτημένη μεταβλητή**) με την **πρόσβαση στην ατομική ιδιοκτησία της αγροτικής γης** (**X : ανεξάρτητη μεταβλητή**) σε **η** περιοχές της Ελλάδας στα τέλη του 1800. Η σχέση αυτή φαίνεται να είναι **ΘΕΤΙΚΗ** (πιο έντονα θετική για τους άνδρες)

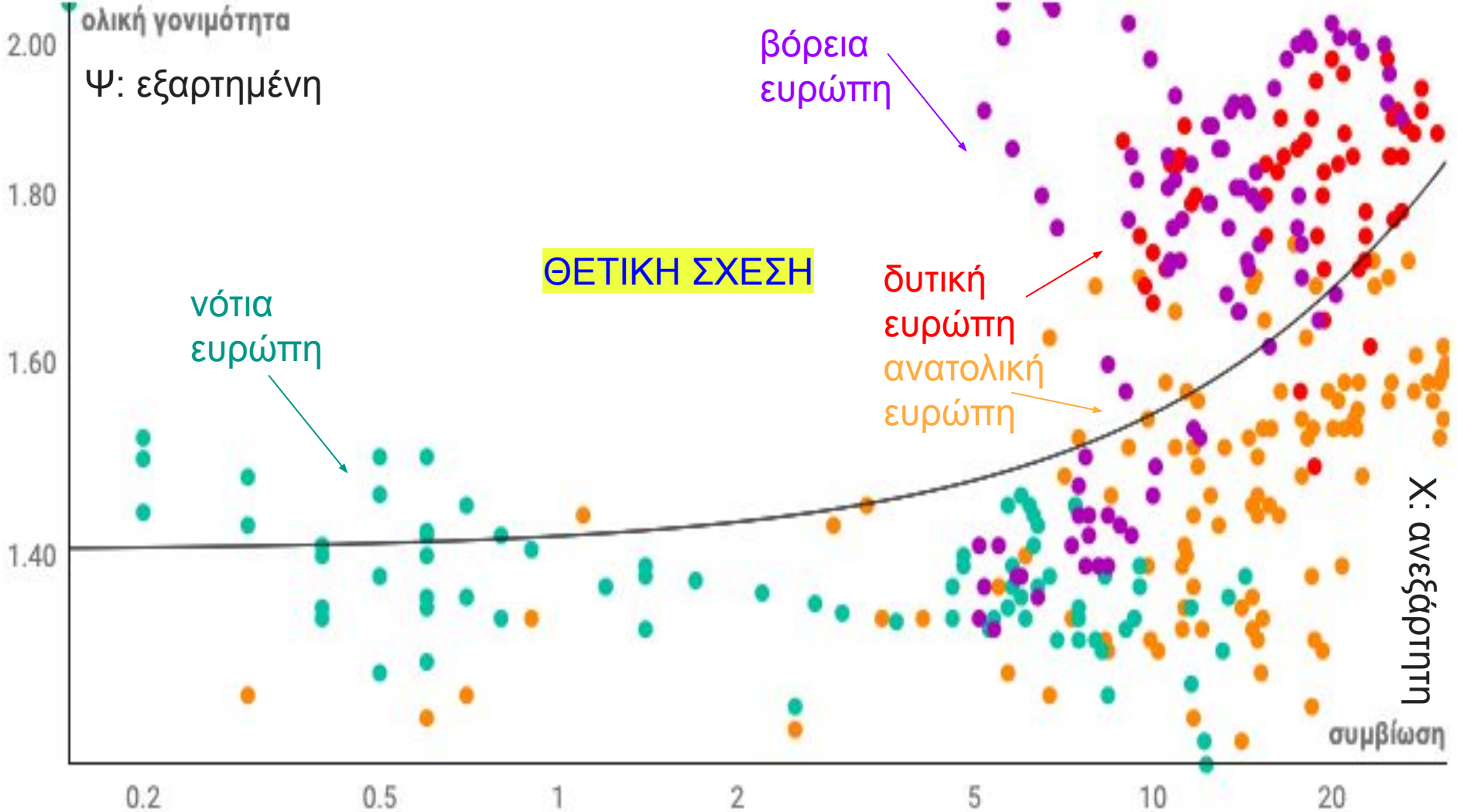
Στα δεξιά φαίνεται η σχέση που συνδέει τις τιμές της **μητρικής θνητότητας στον τοκετό** (**Ψ**) με τις τιμές της εκτός γάμου γεννητικότητας (**X**). Η σχέση αυτή φαίνεται πως ήταν **ΑΡΝΗΤΙΚΗ**

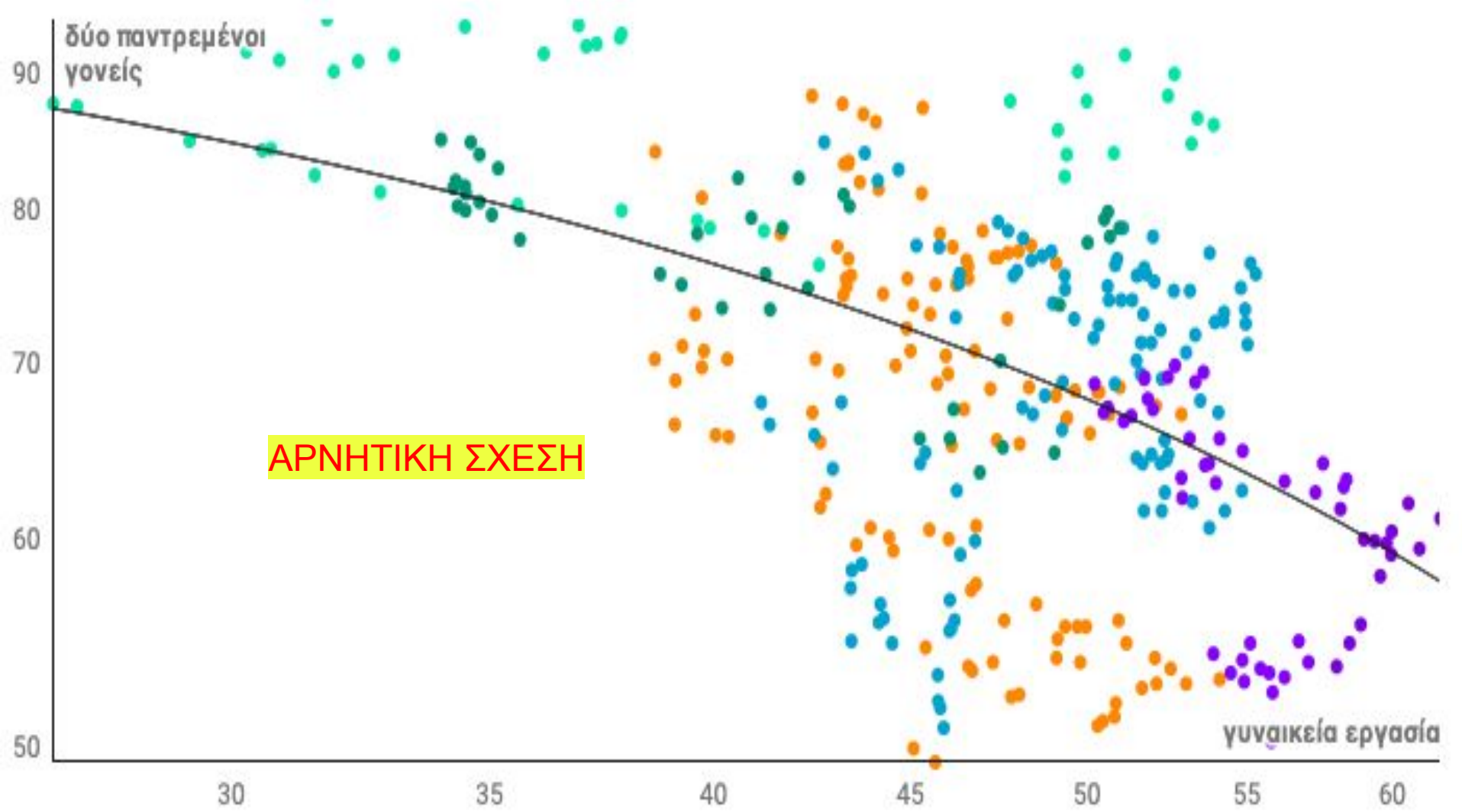
Γραμμικότητα (μια μάλλον φανταστική υπόθεση)



log_μνητρική θνησιμότητα
κατά τον τοκετό







Η ύπαρξη ΣΥΣΧΕΤΙΣΗΣ ΔΕΝ συνεπάγεται απαραίτητα και την ύπαρξη ΑΙΤΙΟΤΗΤΑΣ

Συσχέτιση και Αιτιότητα

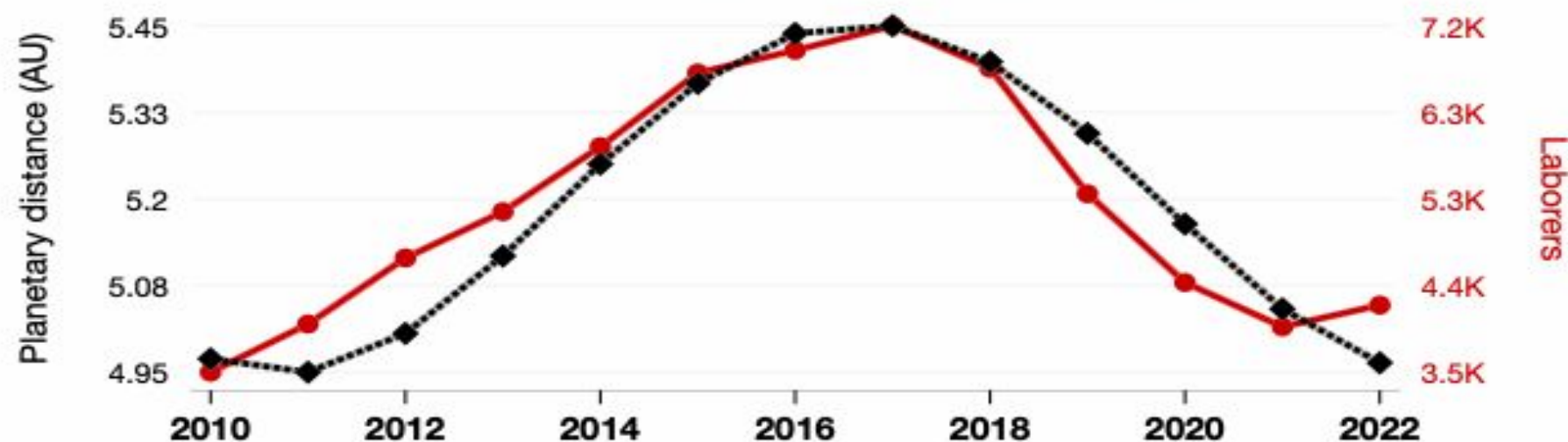


**ΣΤΙΣ διαφάνειες που
ακολουθούν μπορείς να δεις
σχεδόν ΤΕΛΕΙΕΣ
ΣΥΣΧΕΤΙΣΕΙΣ που δε
συνεπάγονται ΑΙΤΙΟΤΗΤΑ**

The distance between Jupiter and the Sun

correlates with

The number of secretaries in Alaska



◆ The average distance between Jupiter and the Sun as measured on the first day of each month · Source: Calculated using Astropy

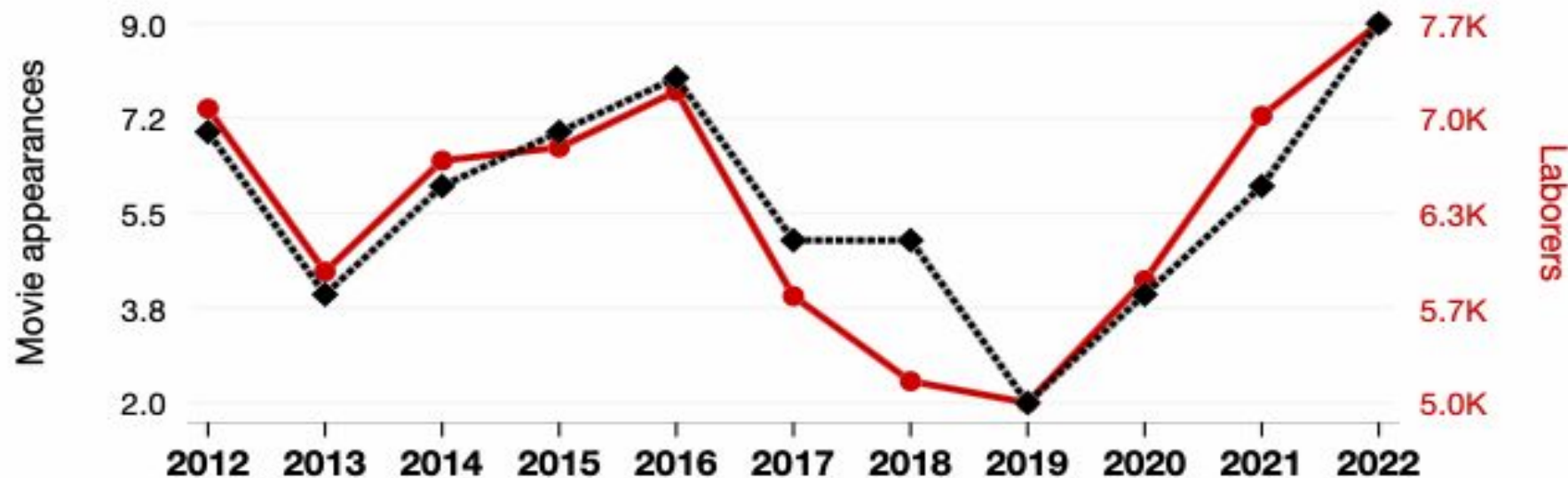
● BLS estimate of secretaries and administrative assistants, except legal, medical, and executive in Alaska · Source: Bureau of Labor Statistics

2010-2022, $r=0.950$, $r^2=0.902$, $p<0.01$ · tylervigen.com/spurious/correlation/2733

The number of movies Tom Hanks appeared in

correlates with

The number of special education teachers in Georgia



◆ The number of movies Tom Hanks appeared in · Source: The Movie DB

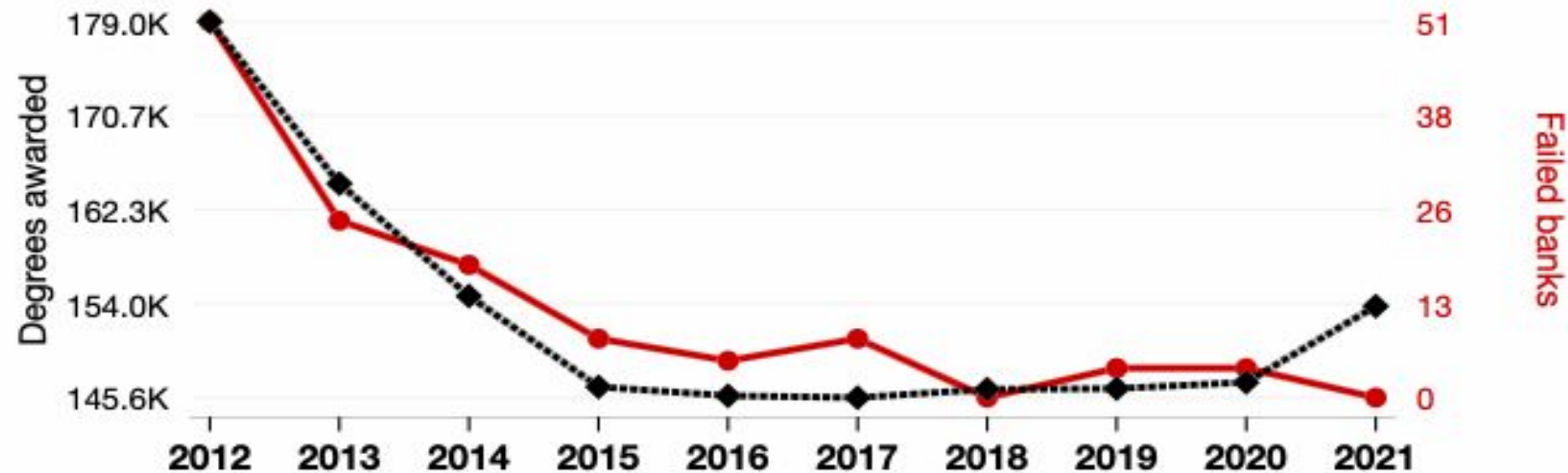
● BLS estimate of special education teachers, kindergarten and elementary school in Georgia · Source: Bureau of Labor Statistics

2012-2022, $r=0.901$, $r^2=0.811$, $p<0.01$ · tylervigen.com/spurious/correlation/5857

Master's degrees awarded in Education

correlates with

US bank failures



◆ Master's degrees conferred by postsecondary institutions in Education · Source: National Center for Education Statistics

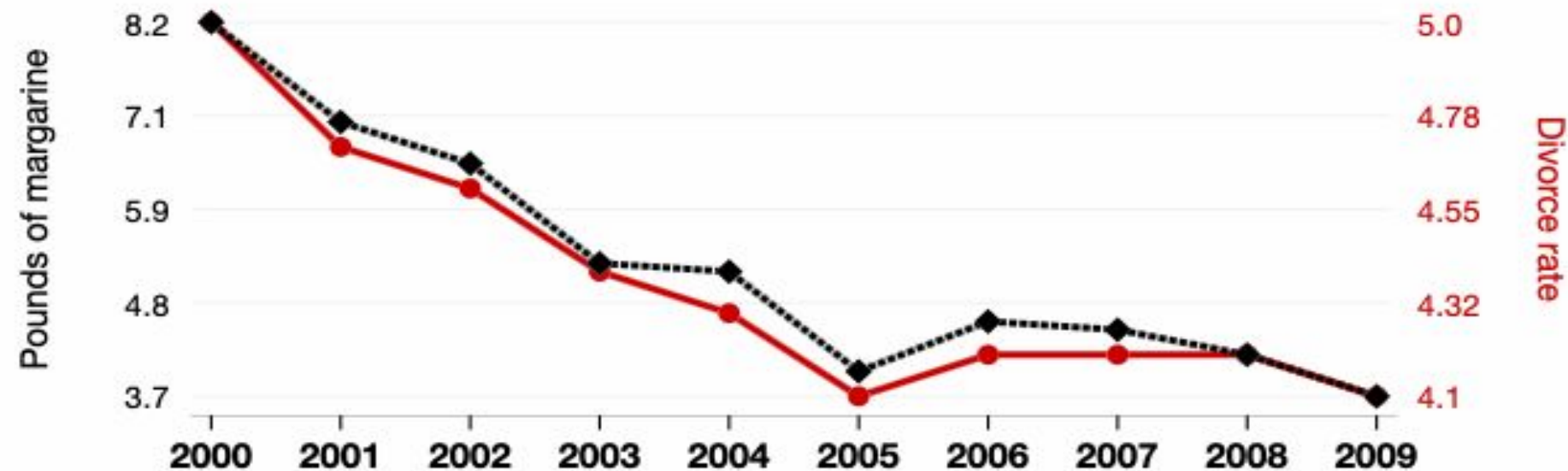
● Bank failures in the US · Source: FDIC

2012-2021, $r=0.933$, $r^2=0.870$, $p<0.01$ · tylervigen.com/spurious/correlation/2718

Per capita consumption of margarine

correlates with

The divorce rate in Maine



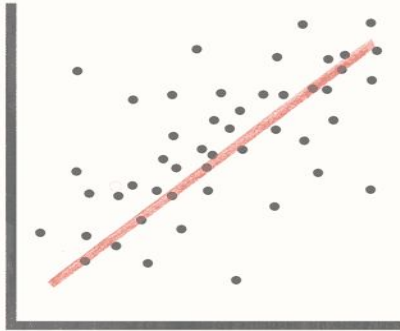
◆ Per capita consumption of margarine in the United States · Source: US Department of Agriculture

● The divorce rate in Maine · Source: CDC National Vital Statistics

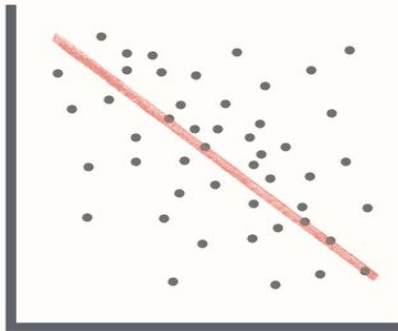
2000-2009, $r=0.993$, $r^2=0.985$, $p<0.01$ · tylervigen.com/spurious/correlation/5920

Συσχέτιση (Γραμμική) Δύο Μεταβλητών

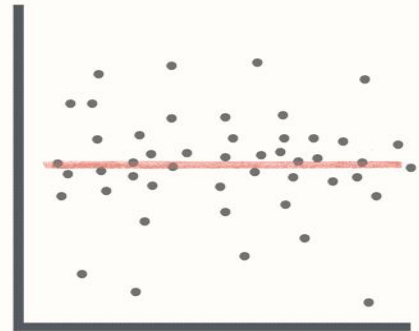
Correlation Coefficient



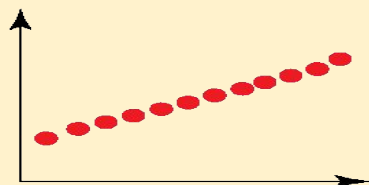
Positive Correlation



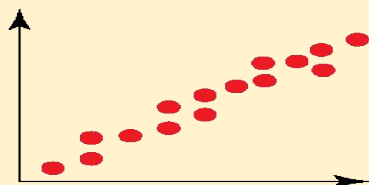
Negative Correlation



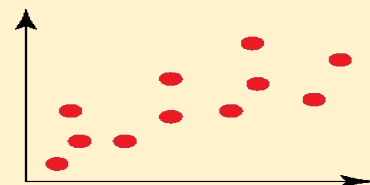
No Correlation



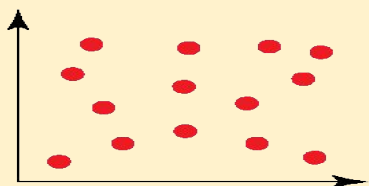
Perfect
Positive
Correlation



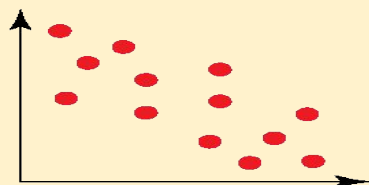
Strong
Positive
Correlation



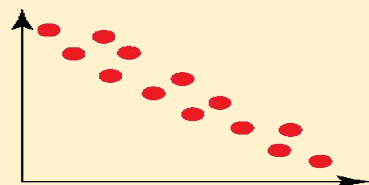
Weak
Positive
Correlation



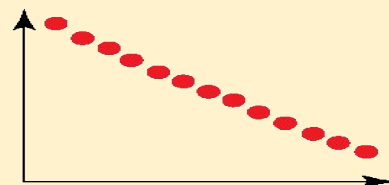
No
Correlation



Weak
Negative
Correlation



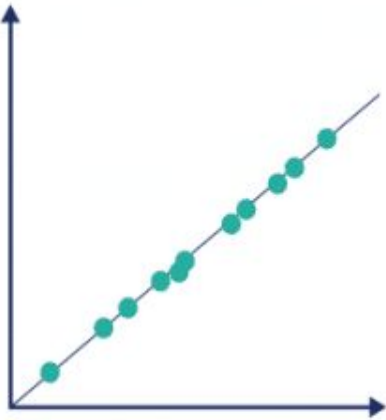
Strong
Negative
Correlation



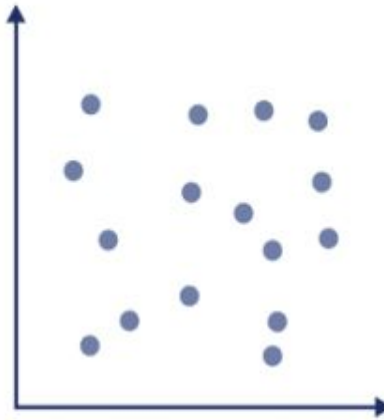
Perfect
Negative
Correlation

Συσχέτιση (Γραμμική) Δύο Μεταβλητών

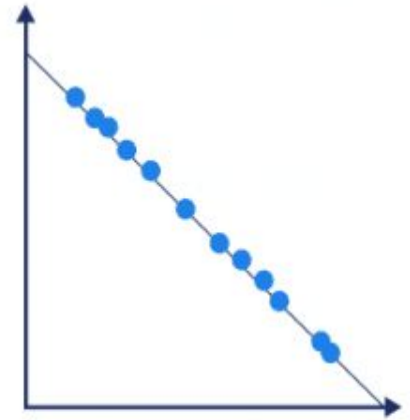
Perfect positive
correlation



Zero
correlation



Perfect negative
correlation



**ΣΤΙΣ διαφάνειες που ακολουθούν
μπορείς να δεις τα πινακάκια με
τις τιμές του συντελεστή
συσχέτισης. Παρατήρησε τις
ΘΕΤΙΚΕΣ και ΑΡΝΗΤΙΚΕΣ σχέσεις
που εντοπίζονται**

	1870		1879	
	Male Literacy %	Female Literacy %	Male Literacy %	Female Literacy %
Access to Land %	+0.47	+0.38	+0.37	+0.37
Stuck in Agriculture	-0.28	-0.27	-0.19	-0.20
Workers %	+0.34	+0.42	+0.22	+0.35
Merchants %	+0.52	+0.67	+0.54	+0.71
Urbanization %	+0.29	+0.45	+0.25	+0.41
Student Attainment %	+0.67	+0.59	+0.69	+0.68
Non-Greeks %	+0.10	+0.23	+0.09	+0.19
Widows %	+0.31	+0.19	+0.32	+0.24
Agricultural Labourers %	-0.56	-0.72	-0.46	-0.68
Farm Size	-0.30	-0.15	-0.16	-0.17
Married People %	-0.15	-0.11	-0.09	-0.05
Family Size	-0.04	-0.01	-0.05	-0.27
Sex Ratio	-0.25	-0.18	-0.14	-0.23
Priests %	+0.05	+0.11	+0.01	+0.11

ανεξάρτητες

εξαρτημένη

εξαρτημένη

?

?

	Ρυθμός οικονομικής ανάπτυξης (%)
Συνεισφορά της βιομηχανίας στο ΑΕΠ (%)	+ 0.55 ***
Συνεισφορά της πληροφόρησης και των επικοινωνιών στο ΑΕΠ (%)	+ 0.43 ***
Συνεισφορά των υπηρεσιών στο ΑΕΠ (%)	- 0.59 ***
Συνεισφορά των ακινήτων στο ΑΕΠ (%)	- 0.48 ***
Συνεισφορά του εμπορίου, των μεταφορών, της διαμονής και της εστίασης στο ΑΕΠ (%)	- 0.35 ***

Πίνακας 1: Συντελεστές γραμμικής συσχέτισης (PIIGS, 2000-2021)

*** στατιστική σημαντικότητα με βεβαιότητα 99%

	καπνός	βαμβάκι	σιτάρι	ντομάτες	πατάτες	πορτοκάλια	λεμόνια
Συγκροτήματα							
τεχν. βροχής	+0,46***	+0,80***	+0,56***	+0,32***	+0,20***	-0,03	-0,02
Αντλίες πετρ.	+0,41***	+0,83***	+0,53***	+0,36***	+0,21***	-0,01	-0,06
Αντλίες βενζ.	+0,00	-0,10**	-0,06	+0,33***	+0,33***	-0,14***	-0,08
Αντλίες							
ηλεκτρ.	-0,00	+0,21***	+0,18***	+0,32***	+0,36***	+0,34***	+0,30***
Αντλίες							
(λοιποί τύποι)	-0,07	-0,05	-0,10**	-0,01	+0,08*	-0,07	-0,07
Γεωργικοί							
ελκυστήρες	+0,27***	+0,27***	+0,35***	+0,44***	+0,37***	-0,03	+0,30***
(σύνολο)							
Γεωργικοί							
ελκυστήρες	-0,15***	-0,12***	-0,16***	+0,13***	+0,24***	-0,04	+0,38***
(μικροί)							
Μηχαν.							
ψεκαστήρες	+0,59***	+0,16***	+0,25***	+0,48***	+0,24***	-0,00	+0,38***
Θειωτήρες	+0,25***	-0,03	+0,02	+0,20***	+0,24***	-0,00	+0,53***

Πίνακας 9. Συντελεστές συσχέτισης ανάμεσα στην πρόσβαση στον τεχνολογικό εξοπλισμό και τον όγκο παραγωγής των γεωργικών προϊόντων (επαρχιακό επίπεδο)

*** στατιστική σημαντικότητα στο επίπεδο 99%

** στατιστική σημαντικότητα στο επίπεδο 95%

* στατιστική σημαντικότητα στο επίπεδο 90%

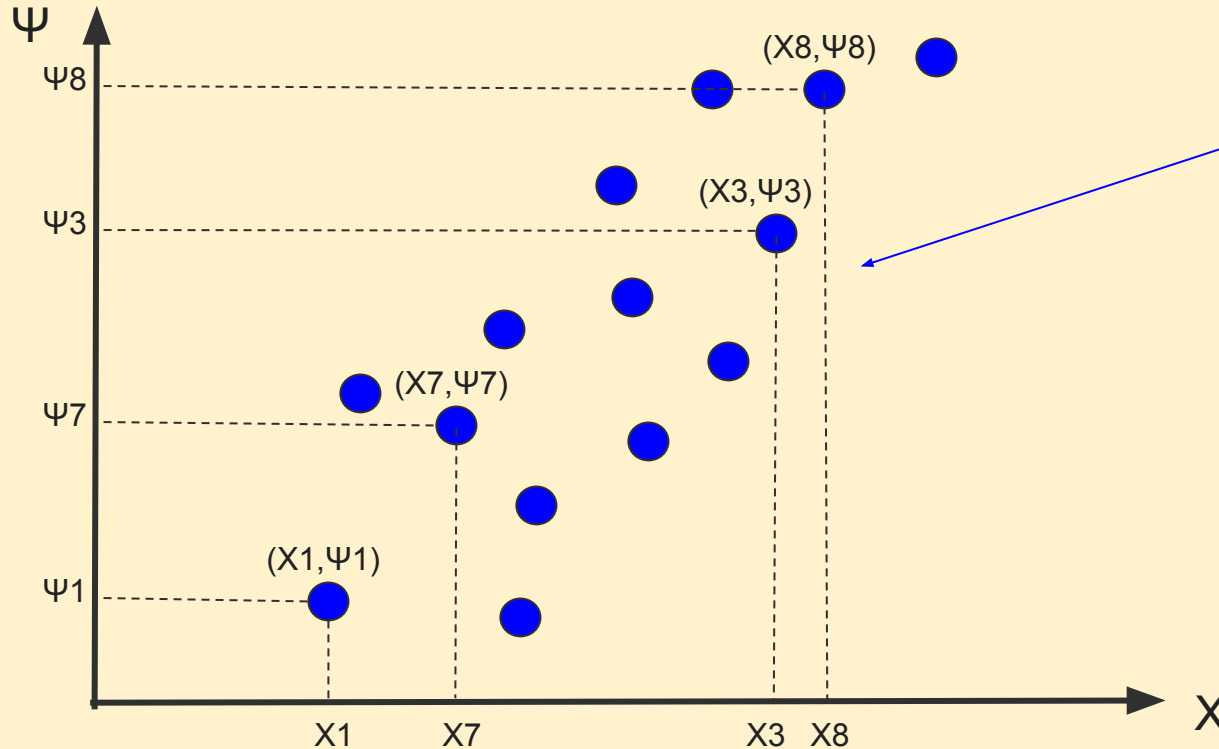
Συσχέτιση (Γραμμική) Δύο Μεταβλητών

Συντελεστής Συσχέτισης (φόρμουλα):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \longrightarrow r [-1, +1]$$

ΣΤΙΣ ΕΠΟΜΕΝΕΣ ΔΙΑΦΑΝΕΙΕΣ ΔΕΣ ΤΟΝ ΤΡΟΠΟ ΛΕΙΤΟΥΡΓΙΑΣ ΤΗΣ ΜΕΘΟΔΟΥ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

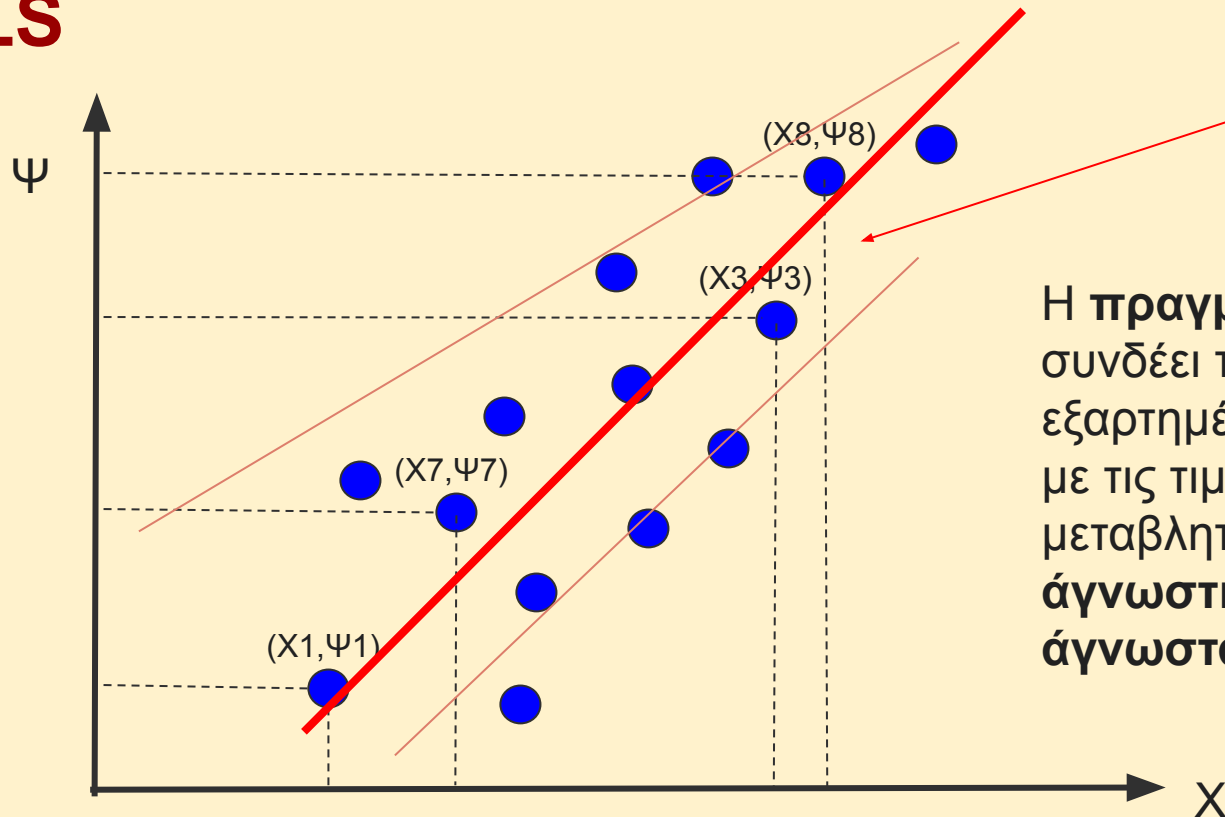
OLS



πραγματικές
τιμές (X, Ψ)

$n=14$
αριθμός
παρατηρήσεων

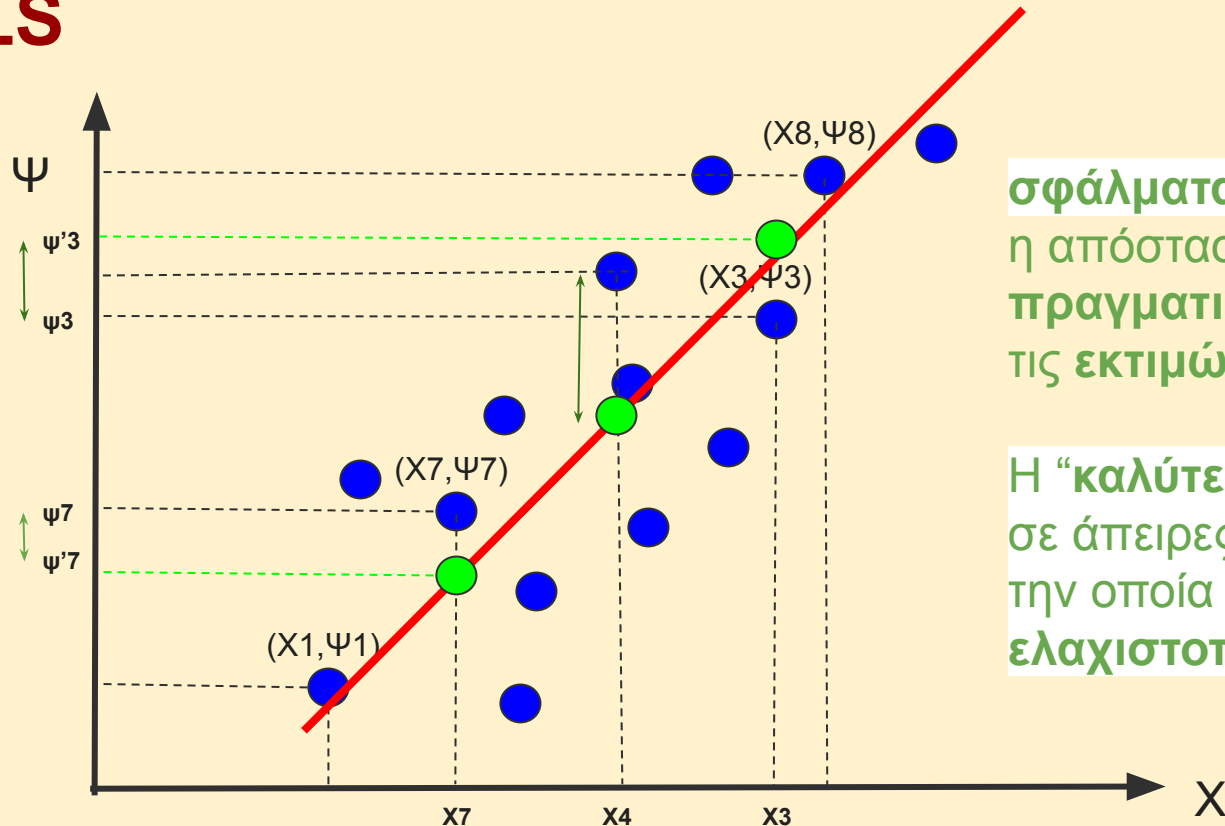
OLS



γραμμή
παλινδρόμησης

Η πραγματική σχέση που συνδέει τις τιμές της εξαρτημένης μεταβλητής (Ψ) με τις τιμές της ανεξάρτητης μεταβλητής (X) παραμένει άγνωστη (α , β και ϵ_i άγνωστα).

OLS



σφάλματα:

η απόσταση που χωρίζει τις **πραγματικές** τιμές (Ψ) από τις **εκτιμώμενες** τιμές (Ψ')

Η “**καλύτερη**” ευθεία ανάμεσα σε άπειρες θα είναι εκείνη για την οποία τα **σφάλματα ελαχιστοποιούνται !!!**

OLS

Ψ: εξαρτημένη μεταβλητή (ο παράγοντας την κίνηση του οποίου προσπαθούμε να εξηγήσουμε)

Χ: ανεξάρτητη μεταβλητή (ο προσδιοριστικός παράγοντας που υποθέτουμε πως εξηγεί ένα σημαντικό κομμάτι της κίνησης της εξαρτημένη μεταβλητής)

απλό υπόδειγμα: περιλαμβάνει μόνο 1 ανεξάρτητη μεταβλητή ($k=1$)

OLS

$$\Psi_i = \alpha + \beta \cdot X_i + \varepsilon_i$$

→ η **πραγματική σχέση** που συνδέει τα X με τα Ψ

όπου

Ψ_i : οι πραγματικές τιμές της **εξαρτημένης** μεταβλητής ($\Psi_1, \Psi_2, \Psi_3, \dots, \Psi_n$)

X_i : οι πραγματικές τιμές της **ανεξάρτητης** μεταβλητής ($X_1, X_2, X_3, \dots, X_n$)

α : ο άγνωστος **σταθερός όρος** της εξίσωσης

β : ο άγνωστος **συντελεστής διεύθυνσης** της εξίσωσης

ε_i : διαταρακτικός όρος - τυχαίοι παράγοντες

OLS

$$\widehat{\Psi}_i = \widehat{a} + \widehat{\beta} * X_i \longrightarrow \text{η **εκτιμώμενη σχέση** (γραμμή παλινδρόμησης)}$$

όπου

$\widehat{\Psi}_i$: οι εκτιμώμενες τιμές της μεταβλητής Ψ

X_i : οι τιμές της ανεξάρτητης μεταβλητής

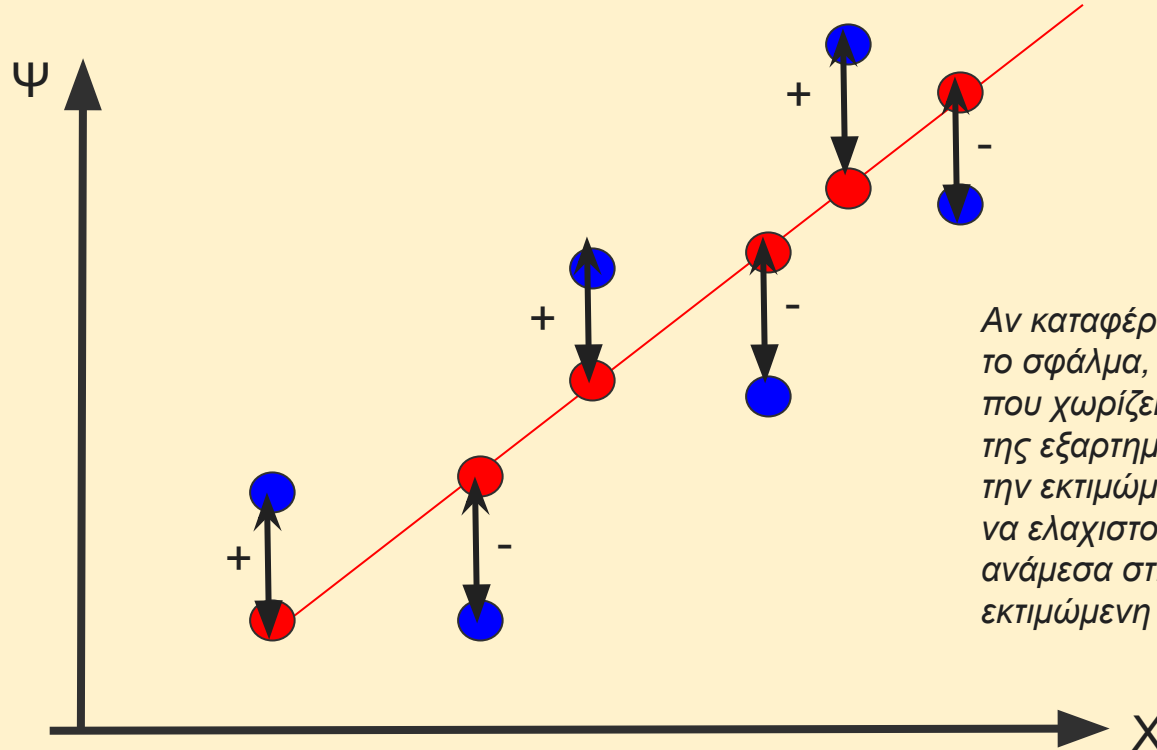
$\widehat{a}$ εκτίμηση του σταθερού όρου a

$\widehat{\beta}$ εκτίμηση του συντελεστή β

Οι πραγματικοί συντελεστές a και β είναι **άγνωστοι**.

Οι **εκτιμητές** των συντελεστών a και β μπορούν να υπολογιστούν.

OLS Ελαχιστοποιώντας τα σφάλματα ...



Αν καταφέρεις να ελαχιστοποιήσεις το σφάλμα, δηλαδή την απόσταση που χωρίζει την πραγματική τιμή της εξαρτημένης μεταβλητής από την εκτιμώμενη, τότε καταφέρνεις να ελαχιστοποιήσεις την απόσταση ανάμεσα στην πραγματική και εκτιμώμενη σχέση