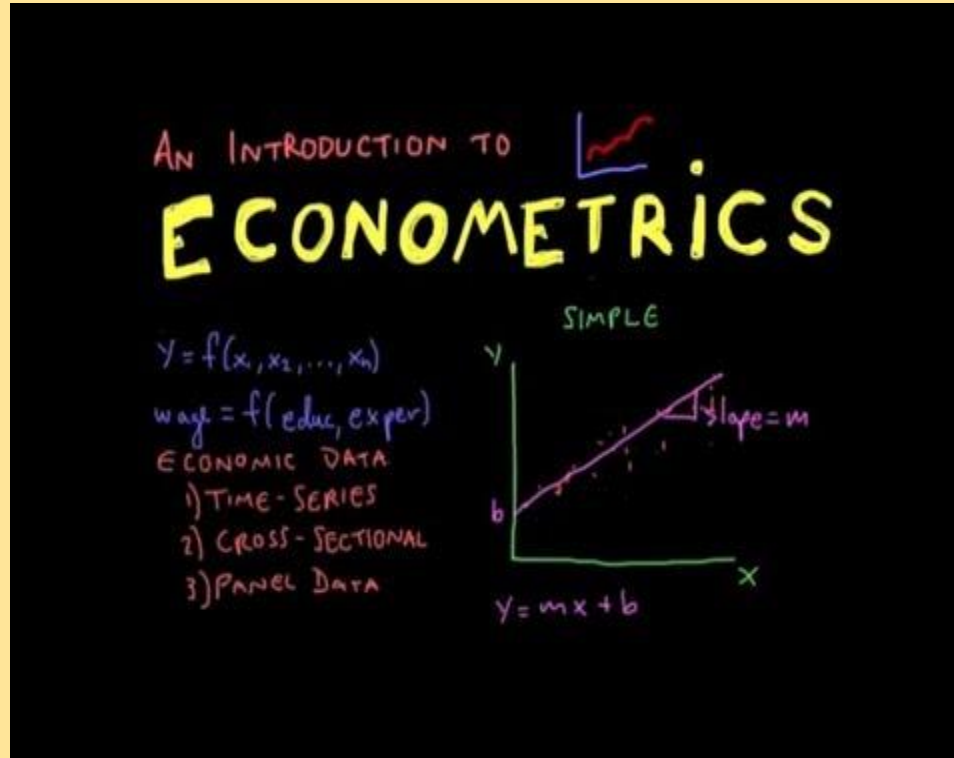


Οικονομικό Πανεπιστήμιο Αθηνών

Μάθημα: "Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρησιακά Προβλήματα"

Συνάντηση: 3η



Τι κάναμε την πρώτη εβδομάδα ; ; ;

Μάθαμε να θέτουμε σε εφαρμογή το **ΠΡΩΤΟ ΒΗΜΑ** της πολυπαραγοντικής ανάλυσης, χτίζοντας τον **συντελεστή γραμμικής συσχέτισης** για δύο μεταβλητές Ψ (εξαρτημένη) και X (ανεξάρτητη) του ενδιαφέροντός μας ! ! !

Η τιμή του συντελεστή γραμμικής συσχέτισης απλώνεται μέσα σε ένα διάστημα τιμών της μορφής $\rightarrow [-1, +1]$

Ερμηνεία: Όσο **πιο κοντά στο -1** κινείται η τιμή του συντελεστή συσχέτισης τόσο **πιο έντονα αρνητική** είναι η γραμμική σχέση που συνδέει τις μεταβλητές Ψ και X . Αντίθετα, όσο **πιο κοντά στην τιμή +1** κινείται ο συντελεστής συσχέτισης τόσο **πιο έντονα θετική** είναι η γραμμική σχέση που συνδέει τις μεταβλητές Ψ και X .

Τι κάναμε την πρώτη εβδομάδα ; ; ;

Τιμές του συντελεστή συσχέτισης **κοντά στο μηδέν** ($\rho=0$) σηματοδοτούν την **μη ύπαρξη γραμμικής σχέσης** ανάμεσα στις μεταβλητές Ψ και X .

ΠΡΟΣΟΧΗ: Πριν να προχωρήσετε στην ερμηνεία της τιμής του συντελεστή γραμμικής συσχέτισης πρέπει να επιβεβαιώσετε (μέσα σε ένα διάστημα βεβαιότητας) πως ο συντελεστής συσχέτισης **ΥΠΑΡΧΕΙ ! ! !**

Με πιο απλά λόγια, δείτε το αποτέλεσμα του ελέγχου της στατιστικής σημαντικότητας για το συντελεστή συσχέτισης.

Τι κάναμε την πρώτη εβδομάδα ; ; ;

Έλεγχος Σημαντικότητας του Συντελεστή Συσχέτισης ! ! !

H₀ (μηδενική υπόθεση): $\rho = 0$ → ΔΕΝ υπάρχει γραμμική συσχέτιση ανάμεσα στις μεταβλητές Ψ και X

H₁ (εναλλακτική υπόθεση): $\rho > 0$ ή $\rho < 0$ → υπάρχει κάποια συσχέτιση (θετική ή αρνητική) ανάμεσα στις μεταβλητές Ψ και X)

Είναι προφανές που **θέλουμε να ΑΠΟΡΡΙΨΟΥΜΕ την μηδενική υπόθεση** !!!

Τι κάναμε την πρώτη εβδομάδα ; ; ;

Έλεγχος Σημαντικότητας του Συντελεστή Συσχέτισης ! ! !

Το αποτέλεσμα του ελέγχου της στατιστικής σημαντικότητας του συντελεστή συσχέτισης δίνεται αυτόματα στο ψηφιακό περιβάλλον της εφαρμογής:

<https://www.statskingdom.com/index.html>

Πήγαινε "**βασικά στατιστικά**" → Επέλεξε "**συσχέτιση**" → Βάλε τις τιμές των μεταβλητών σου στις στήλες Ψ και X → Πάτησε "**υπολογισμός**"

Η εφαρμογή θα σου δώσει:

- 1) την τιμή του συντελεστή συσχέτισης
- 2) το διάγραμμα διασποράς
- 3) το αποτέλεσμα του ελέγχου σημαντικότητας

Τι κάναμε την πρώτη εβδομάδα ; ; ;

Έλεγχος Σημαντικότητας του Συντελεστή Συσχέτισης ! ! !

Για τιμές του **p-value < 0,05 (5%)** μπορείς να απορρίψεις την μηδενική υπόθεση της ΜΗ ύπαρξης γραμμικής σχέσης ανάμεσα στις μεταβλητές Ψ και Χ

<https://www.statskingdom.com/index.html>

Όταν το **p-value < 0,01 (1%)** τότε απορρίπτεις την μηδενική υπόθεση με μεγαλύτερη βεβαιότητα (βλ. 99%).

Αν το **p-value < 0,1 (10%)** τότε μπορείς να απορρίψεις τη μηδενική υπόθεση με πιο χαλαρό το διάστημα βεβαιότητας (βλ. 90%).

Τι πρέπει να ξέρεις μετά την 1η εβδομάδα ; ; ;

Να μπορείς να υπολογίζεις την τιμή του συντελεστή γραμμικής συσχέτισης

Να μπορείς να προχωράς στην ερμηνεία της τιμής του συντελεστή γραμμικής συσχέτισης, χαρακτηρίζοντας τη σχέση ανάμεσα στις μεταβλητές Ψ και X

Να μπορείς να διαβάζεις το αποτέλεσμα του ελέγχου της στατιστικής σημαντικότητας του συντελεστή γραμμικής συσχέτισης

Γραμμή Παλινδρόμησης (η μέθοδος των ελαχίστων τετραγώνων)

OLS

Αφού τελικά οι σχέσεις που σχηματίζονται στην πραγματικότητα ανάμεσα σε δύο παράγοντες **ΔΕΝ** είναι γραμμικές τότε γιατί προχωράμε στην κατασκευή γραμμικών μοντέλων ;

Ανάμεσα στα πραγματικά σημεία μας (X, Ψ) περνούν **άπειρες ευθείες**.

Πως επιλέγουμε την "**καλύτερη**" από αυτές ;

Η ΜΕΘΟΔΟΣ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ !!!

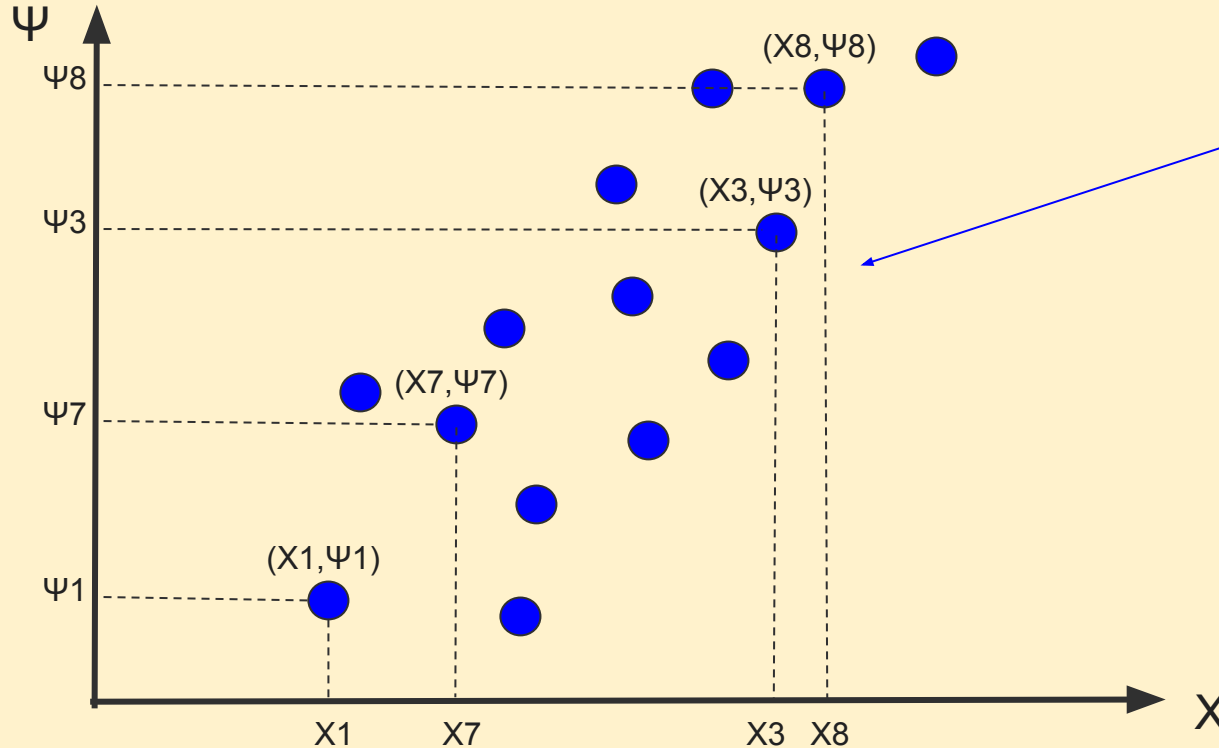
OLS

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 ($k=1$)

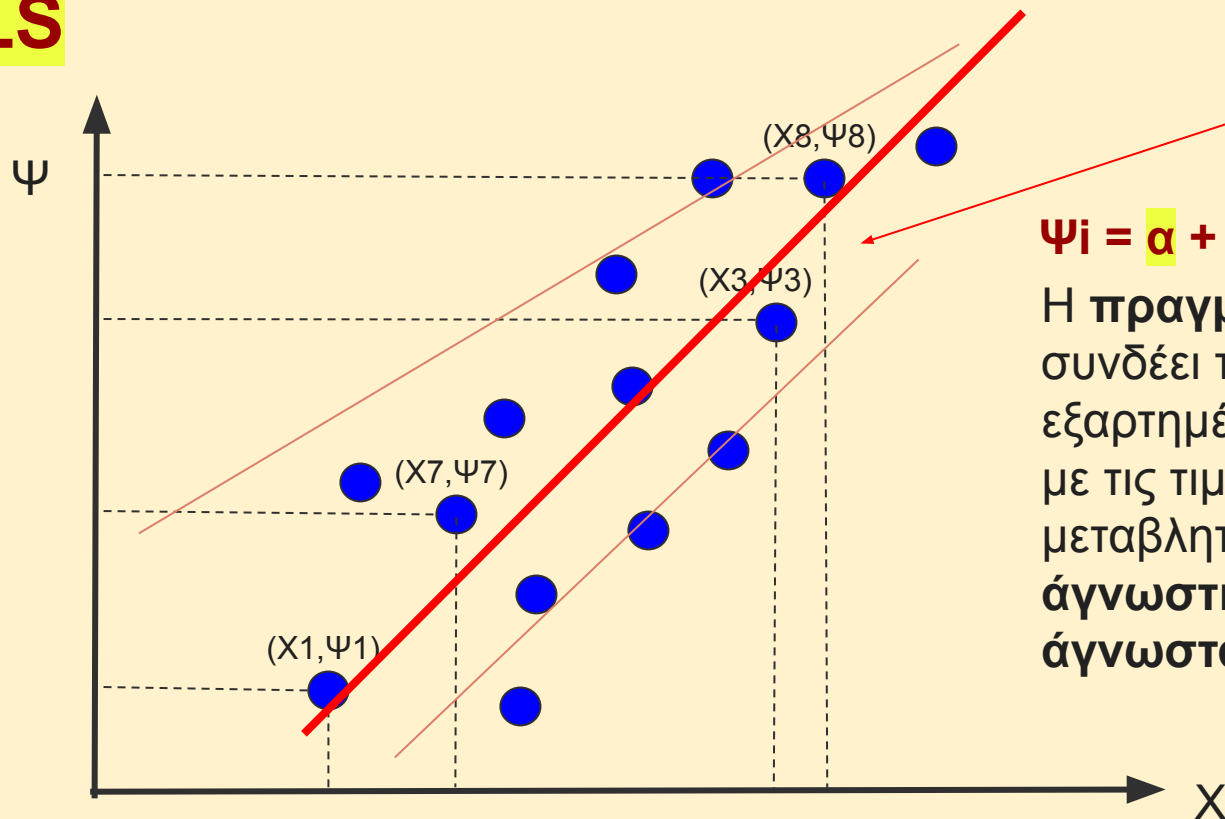
OLS



πραγματικές
τιμές (X, Ψ)

$n=14$
αριθμός
παρατηρήσεων

OLS



γραμμή
παλινδρόμησης

$$\Psi_i = \alpha + \beta * \chi_i + \epsilon_i$$

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

OLS



$(\Psi_i - \Psi_{i_εκτ.})$

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

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

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

όπου

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

X_i : οι τιμές της ανεξάρτητης μεταβλητής

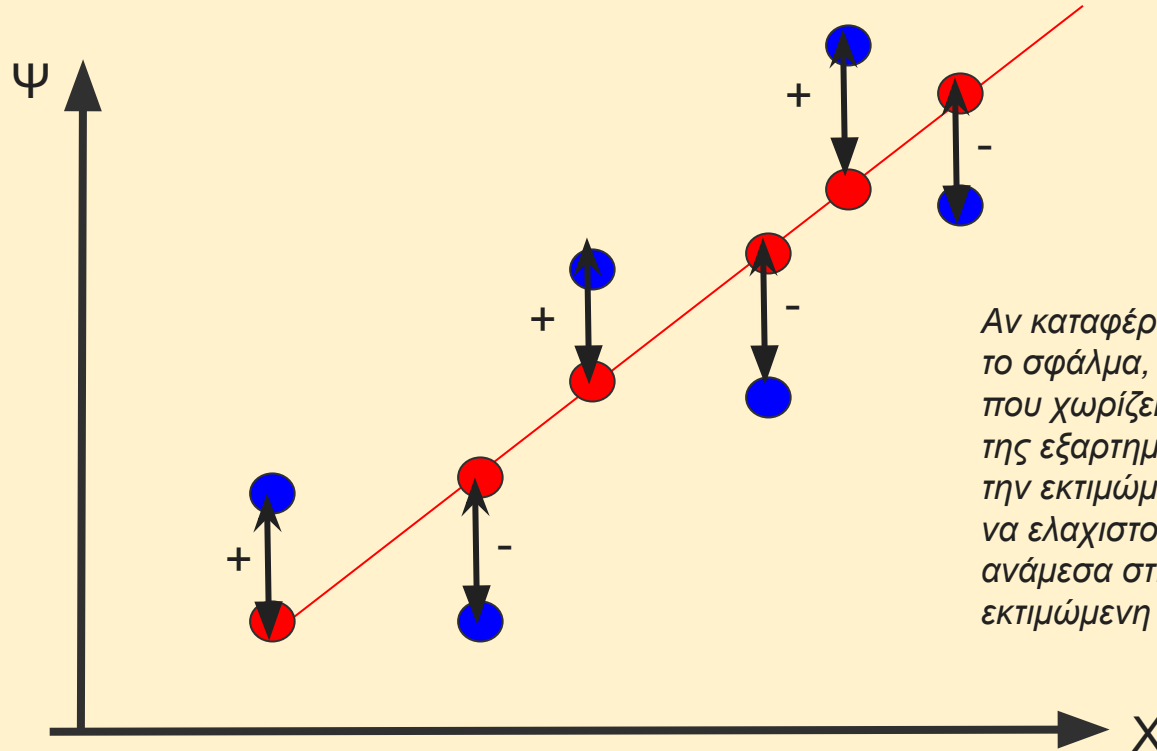
$\hat{a}$ εκτίμηση του σταθερού όρου a

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

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

Οι **εκτιμητές** των συντελεστών α 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

OLS Ελαχιστοποιώντας τα σφάλματα ...

$$\text{σφάλμα} = (\Psi_i - \hat{\Psi}_i)$$

$$\text{άθροισμα τετραγώνων των σφαλμάτων} = \sum (\Psi_i - \hat{\Psi}_i)^2$$

$$= \sum (\Psi_i - \hat{\alpha} - \hat{\beta}X_i)^2$$

OLS Ελαχιστοποιώντας τα σφάλματα ...

- Πάρτε την παράγωγο ίση με το μηδέν (ως προς $\hat{\alpha}$):

$$2 \sum (\Psi_i - \hat{\alpha} - \hat{\beta} X_i) * (-n) = 0$$

$$2 \sum (\Psi_i - \hat{\alpha} - \hat{\beta} X_i) = 0$$

$$\sum \Psi_i - n\hat{\alpha} - \hat{\beta} \sum X_i = 0$$

$$\overline{\Psi_i} - \hat{\alpha} - \hat{\beta} \overline{X_i} = 0$$

$$\hat{\alpha} = \overline{\Psi_i} - \hat{\beta} \overline{X_i}$$

OLS Ελαχιστοποιώντας τα σφάλματα ...

- Πάρτε την παράγωγο ίση με το μηδέν (ως προς $\hat{\beta}$):

$$2 \sum (\Psi_i - \hat{\alpha} - \hat{\beta} X_i) * \Sigma X_i = 0$$

$$2 \sum (\Psi_i - \bar{\Psi}_i + \hat{\beta} \bar{X}_i - \hat{\beta} X_i) * \Sigma X_i = 0$$

$$n \Sigma \Psi_i X_i - \Sigma X_i \Sigma \Psi_i - n \hat{\beta} \Sigma X_i^2 + \hat{\beta} (\Sigma X_i)^2 = 0$$

$$\hat{\beta} = \frac{n \Sigma \Psi_i X_i - \Sigma X_i \Sigma \Psi_i}{n \Sigma X_i^2 - (\Sigma X_i)^2} \quad \text{ή} \quad \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

OLS

Οι τύποι που πρέπει να ξέρεις ! ! !

$$\hat{\beta} = \frac{n \sum \psi_i X_i - \sum X_i \sum \psi_i}{n \sum X_i^2 - (\sum X_i)^2}$$

ή

$$\frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

$$\hat{\alpha} = \bar{\psi}_i - \hat{\beta} \bar{X}_i$$