

5-12-2025

1) Υακὸ

Eclass

{ Σημειώσεις / datasets / κώδικες  
Labs  
Βιβλιογραφία  
Σημειώσεις κ. Γιαννοῦ τσος

2) Λογισμικό

1) Stata (Labs)

2) R (μάθημα)

3) ἴα αλλο δελεε (Python, Matlab, ...)

3) Μάθημα

{ Διαλέξεις

5 labs (κ. Αναστασίου)

Εργασία (μέχρι ημερὰ τελ. εξέτασης)

Τελική εξέταση (οχι σε υπολογιστή)

Πρόγραμμα

Αν. Παγ.

2/2

Π.δ.

9/2

Εναλλακτικά → 9/2

→ 2/2

Τελ. Μάθημα

Παρ.

3/1

4) Παρονοιες

# Παλινδρόμηση κ' Ανάλυση Διασποράς

## Regression Analysis of Variance

### Γραμμικά Μοντέλα Στατιστικής

Στατιστικά Μοντέλα για Επιστημικές Υφές

Μοντέλο (Υπόθεση)

Αναπαράσταση πραγματικού/φυσικού φαινομένου

Ψηφιακό/φυσικό (π.χ. μακέτα κτιρίου)

Εικονικό (π.χ. σχέδιο, εικόνα ΗΥ)

Μαθηματικό (μέσω μεταβλητών/εξισώσεων)

π.χ. νόμοι μηχανικής για κίνηση

Στατιστικό Μοντέλο

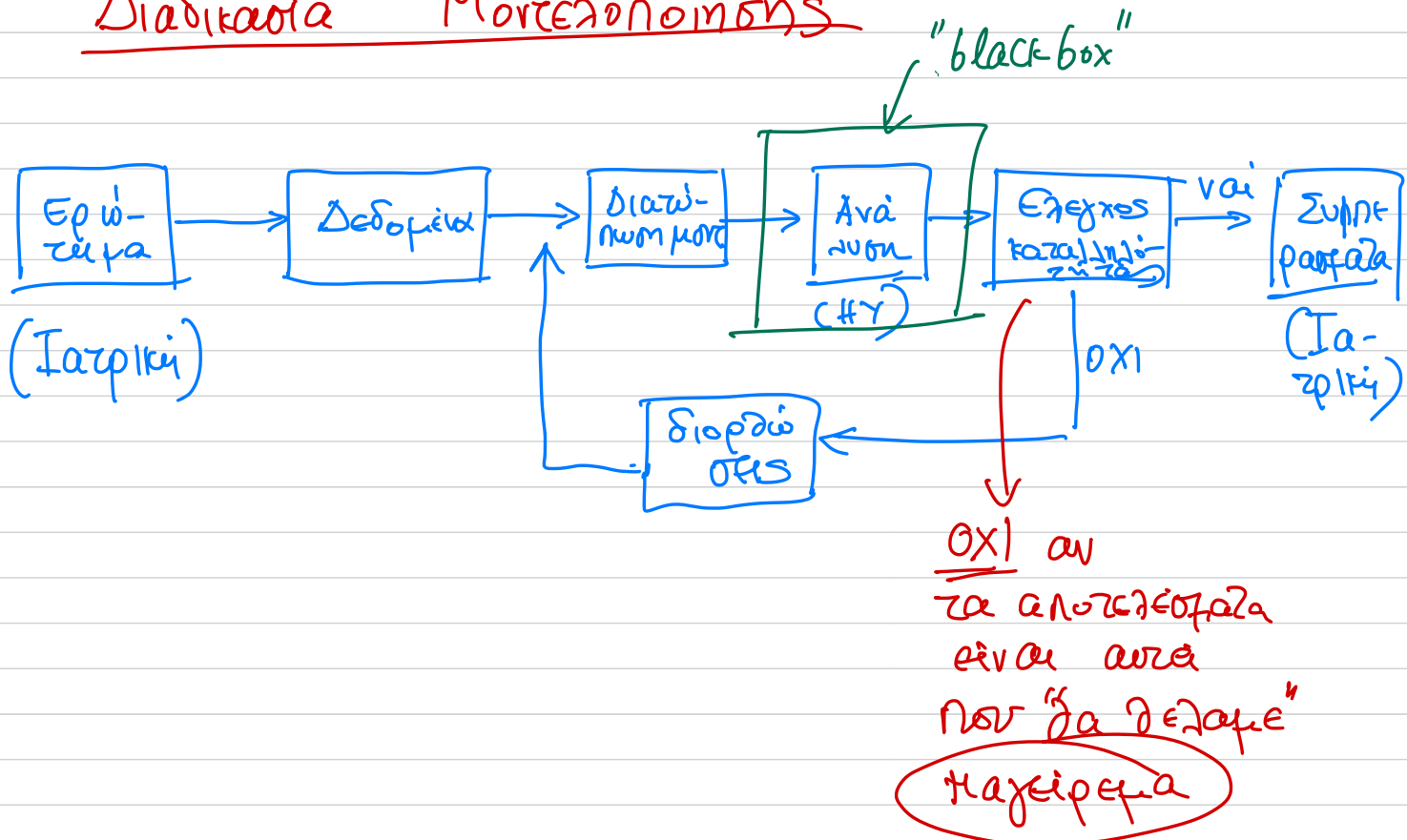
— φαινόμενα με ελλιπή πληροφόρηση  
ως προς κάποιες παραμέτρους

— συμπληρωματικότητα από δεδομένα

μέσω τυχάσις δειγματοληψίας

↳ πλαίσιο πιθανοτήτων

## Διαδικασία Μοτελοποίησης



## Στατιστικά Μοντέλα

### Γραμμικά Μοντέλα

$\left\{ \begin{array}{l} Y \text{ response-εξαρτημένη μεταβλητή} \\ \text{συνχεύσεις με} \\ X_1, X_2, \dots \text{ παράγοντες, ανεξ. μεταβλητές} \end{array} \right.$

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + \varepsilon$$

### Γενικευμένα Γραμμικά Μοντέλα

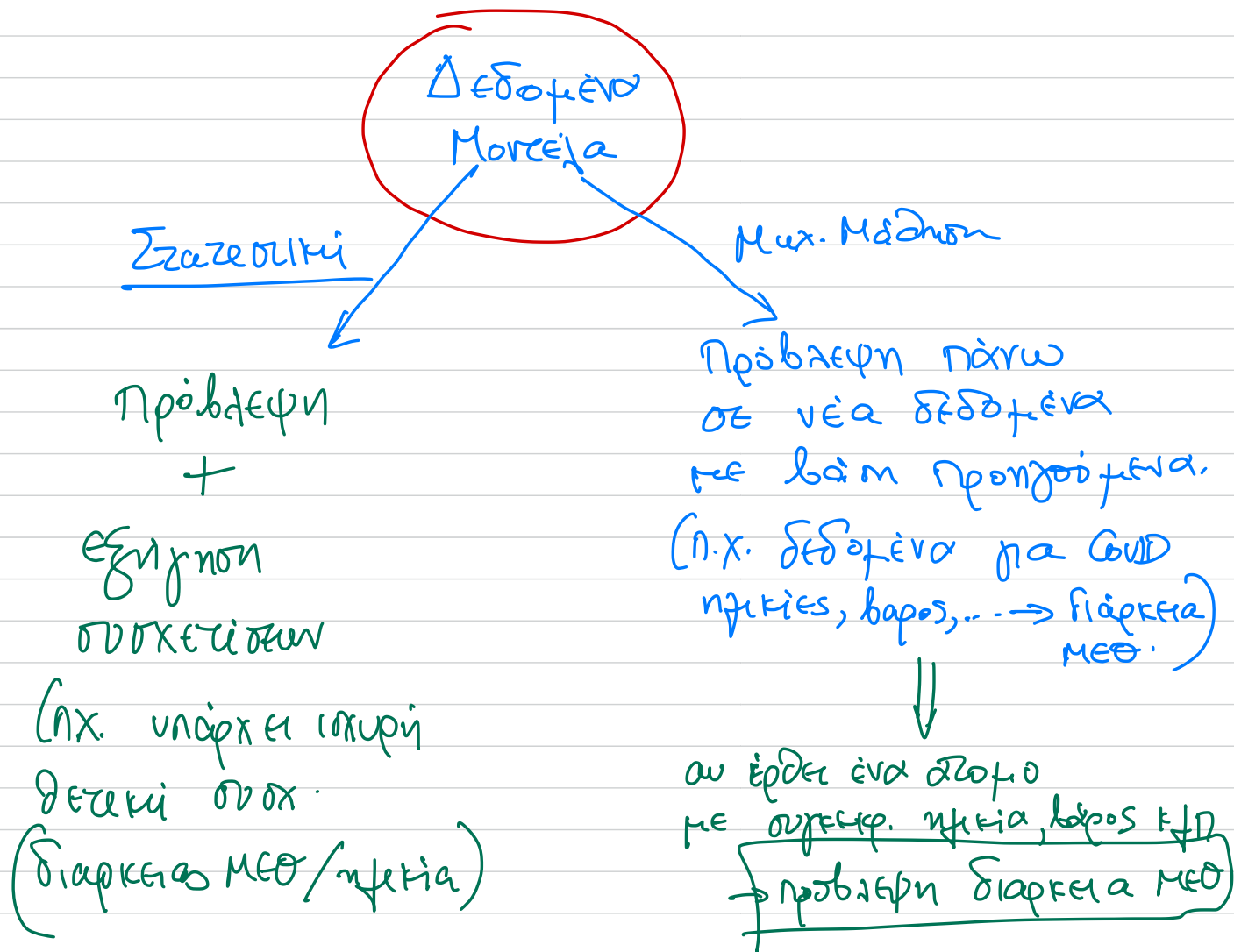
Πιο αφηρημένες σχέσεις  
 $Y = f(X_1, X_2, \dots)$

### Μοντέλα Ενδοβιωσης

$\left\{ \begin{array}{l} Y : \text{χρόνος έως γεγονός} \\ X_1, X_2, \dots \text{ παράγοντες} \end{array} \right.$

### Μοντέλα Εναρμολωμένων Μετρήσεων

# Στατιστική & Μηχανική Μάθηση





Εστω  $\mu_M = \text{Μέση τιμή } Y \text{ για male} = E(Y|\text{sex=male})$   
 $\mu_F = \text{" " " " female} = E(Y|\text{sex=female})$

t-test  $H_0: \mu_M = \mu_F$  vs.  $H_1: \mu_M \neq \mu_F$

p-value = 0.47

Δεν αποκλίνει στατ. σφ. διαφορά μέσω τιμών  $\mu_M, \mu_F$

" " " " συνθήκη score και φύλου

η μεταβλ. των score δεν μπορεί να αποδοθεί στο φύλο.

2) Εστω ότι είχαμε μια άλλη μεταβλητή:

Επάγγελμα:  $\begin{cases} \text{χειρ. εργασία} & (1) \\ \text{εργ. γραφείου} & (2) \\ \text{ενδιαμέσο} & (3) \end{cases} = \text{profession}$

$$\mu_1 = E(Y|\text{prof}=1)$$

$$\mu_2 = E(Y|\text{prof}=2)$$

$$\mu_3 = E(Y|\text{prof}=3)$$

άγνωστα  $\Rightarrow$

$$\hat{\mu}_1 = \text{εκτίμηση}$$

$$\hat{\mu}_2 = "$$

$$\hat{\mu}_3 = "$$

$\exists$  διαφορά στις μέσες αξίες στο δικτά?

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_1$ : τουλάχιστον δύο διαφέρουν

μέθοδος;

(οχι t-test)

μέθοδος ανίχνευσης διασποράς

$\Rightarrow$  γραμμικό μοντέλο

3) Ξερεφεται η μεταβλητοτητα της Y με την ηλικια?  
(παμε δρομον)

Απο διαγραμμα scatter plot  $Y \sim age$

η ηλικια σχεφεται με σκορ Y

δυσ. εζηκει μερα της μεταβλητοτητας της Y  
αλλα οχι εζ σκορηρου

Ειδη ερωτηα : Εχοντας ταβα υποψη της  
ηλικια, εχει ννημα να διερευνη-  
σουμε τον αηου παρδχοντα;

Age : ποσοτικη μεταβλητη (μεταβλητη  
κνμακα)  
(συνεχης μεταβλητη)

$H_0$  ;  $H_1$  ; ??

χρεαζεται πρωτα μωρεφο

## Μοντέλα για την $Y$

1) "Μυθετικό" Μοντέλο (περιγραφή της  $Y$ )

$$E(Y) = \mu \quad (\text{αγνωστο παράμετρος})$$

$\Downarrow$

Μοντέλο

$$Y = \mu + \varepsilon$$

$\varepsilon$  τυχαία διακύμανση

$$E(\varepsilon) = 0$$

$$(E(Y) = \mu + E(\varepsilon) = \mu \Rightarrow E(\varepsilon) = 0)$$

$$\text{Var}(Y) = \text{Var}(\varepsilon) = \sigma^2 \text{ αγνωστο.}$$

Αν υποθέσουμε  $\varepsilon \sim N(0, \sigma^2) \Rightarrow Y \sim N(\mu, \sigma^2)$   
εξέταση σε κανονική πληθυσμιακή

2)  $Y$  συσχ. με φύλο

$$\begin{cases} Y_{\text{male}} = \mu_M + \varepsilon \\ Y_{\text{fem}} = \mu_F + \varepsilon \end{cases}$$

$$E(\varepsilon) = 0$$

$$\text{Var}(\varepsilon) = \sigma^2$$

$$\text{Αν } \varepsilon \sim N(0, \sigma^2) \quad : \begin{cases} \rightarrow Y_m \sim N(\mu_M, \sigma^2) \\ \rightarrow Y_f \sim N(\mu_F, \sigma^2) \end{cases}$$

σύγκριση μέσω  
2 κανον.  
πληθυσμιακών  
(t-test)

$$\text{Έστω } S = \begin{cases} 1 & \text{αν male} \\ 0 & \text{αν female} \end{cases}$$

$$Y = \overset{\downarrow}{\mu} + S \overset{\downarrow}{\delta} + \varepsilon$$

$\mu, \delta$  άγνωστες παράμετροι

$$\text{αν } S=0 \Rightarrow Y = \overset{\mu_F}{\mu} + \varepsilon \quad (\text{female})$$

$$\text{αν } S=1 \Rightarrow Y = \underset{\mu_M}{\mu + \delta} + \varepsilon \quad (\text{male})$$

$$\mu = \mu_F$$

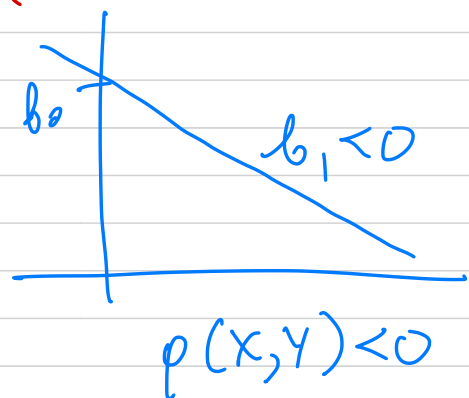
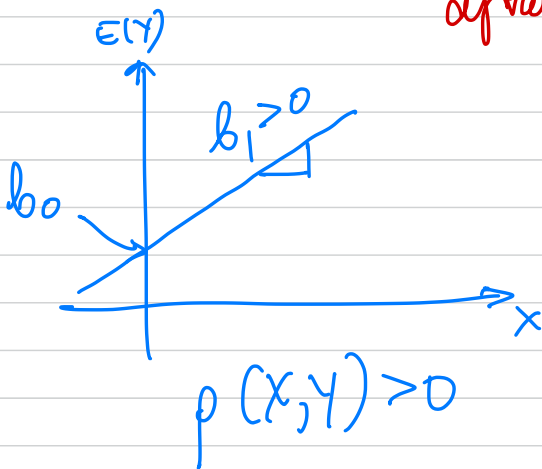
$$\delta = \mu_M - \mu_F$$

3)  $Y \sim \text{age}$

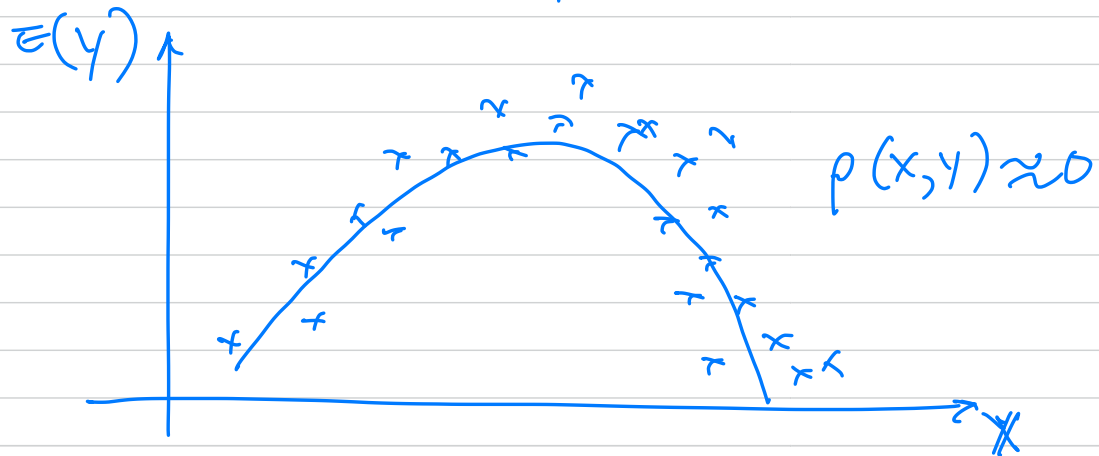
Απο διάγραμμα βλέπουμε ότι η σχέση μεταξύ  $EY$  κ'  $x = \text{age}$  είναι σχεδόν γραμμική

$$E(Y | \text{Age} = x) = b_0 + b_1 x$$

$\uparrow \quad \uparrow$   
αγνωστές παράμετροι



4) Έστω ένα διάγραμμα  $Y \sim X$

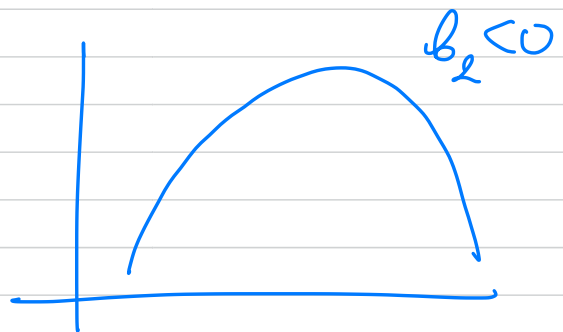
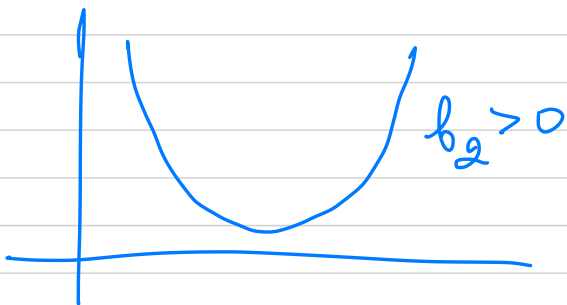


$$E(Y|X=x) = ?$$

μη γραμμική σχέση

∃ συσχέτιση οχι γραμμικά

$$Y = \underbrace{b_0 + b_1 X + b_2 X^2}_{\text{}} + \varepsilon$$



5) Συνδυασμός

$$Y = \underbrace{b_0 + b_1 S}_{\text{φύλο}} + \underbrace{b_2 X + b_3 X^2}_{\text{X}}$$

Γραμμικό μοντέλο

$N(0, \sigma^2)$

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k + \varepsilon$$

πχ.  $Y = b_0 + b_1 X + b_2 X^2$

$X_1 = X, X_2 = X^2$

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2, \quad X_3 = X_1 X_2$$

γραμμικά  
μοντέλα

→ γραμμικά ως προς τα  $b_0, b_1, \dots$  !!

Παράδειγμα με γραμμικό μοντέλου

$$E(Y) = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}}$$

μοντέλο  
λογιστικής  
παρανδρόμησης