10-1-2024

Κατηγορικές Μεταβλητές

Αλληλεπιδράσεις (Interaction effects)

Αν X : κατηγορική επίπεδα $\begin{cases} a \\ b \\ c \\ d \end{cases}$

Επιλέγουμε αυθαίρετα επίπεδο αναφοράς (π.χ. a)

Ορίζουμε μια binary μεταβλητή για καθένα από τα υπόλοιπα

$$X_b = \begin{cases} 1 & \text{αν } X = b \\ 0, & \text{αν } X \neq b \end{cases}$$

$$X_c = \begin{cases} 1 & \text{αν } X = c \\ 0 & X \neq c \end{cases}$$

$$X_d = \begin{cases} 1, & \text{αν } X = d \\ 0; & X \neq d. \end{cases}$$

$$Y = b_0 + b_1 X_b + b_2 X_c + b_3 X_d$$

$$b_0 = E(Y \mid X = a) = \mu_a \leftarrow \text{επίπεδο αναφοράς}$$

$$\left.\begin{aligned} b_1 &= \mu_b - \mu_a = E(Y \mid X = b) - E(Y \mid X = a) \\ b_2 &= \mu_c - \mu_a \\ b_3 &= \mu_d - \mu_a \end{aligned}\right\}$$

διαφορά μέσων τιμών σε κάθε επίπεδο από την μέση τιμή του Y στο επ. αναφοράς.

F-test :

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ $\qquad$ H_1 : τουλάχιστον ένα $\neq 0$

$$\Updownarrow$$

$H_0: \mu_a = \mu_b = \mu_c = \mu_d$ $\qquad$ H_1 : τουλάχιστον δύο διαφορετικά

$$Y \sim X \quad \left\{ \begin{matrix} a \\ b \\ c \\ d \end{matrix} \right.$$

Επέκταση 1 : 2 κατηγορικές μεταβλητές

(Two-way ANOVA)

Παράδειγμα

Test φυσικής κατάστασης $\left\{ \begin{array}{l} Y = \text{score} \\ G = \text{sex} \in \{M, F\} \\ S = \text{smoking} \in \{\text{No}, \text{Yes}\} \end{array} \right.$

$G: \quad X_F = \begin{cases} 1, & \text{Female} \\ 0, & \text{Male} \end{cases}$

$S: \quad X_Y = \begin{cases} 1, & \text{Smoking} = \text{Yes} \\ 0, & \text{"} \quad = \text{No} \end{cases}$

$Y = b_0 + b_1 X_F + b_2 X_Y \quad \leftarrow \quad ?$

G	S	
M	N	μ_{MN}
M	Y	μ_{MY}
F	N	μ_{FN}
F	Y	μ_{FY}

4 ελεύθερες παράμετροι

??

Όμως στο μοντέλο $Y = b_0 + b_1 X_F + b_2 X_Y$

$b_0, b_1, b_2 \rightarrow$ 3 παράμετροι

G	S	X_F	X_Y	$Y = b_0 + b_1 X_F + b_2 X_Y$
M	N	0	0	$Y = b_0 = \mu_{MN}$
M	Y	0	1	$Y = b_0 + b_2 = \mu_{MY}$
F	N	1	0	$Y = b_0 + b_1 = \mu_{FN}$
F	Y	1	1	$Y = b_0 + b_1 + b_2 = \mu_{FY}$

$b_0 = \mu_{MN}$

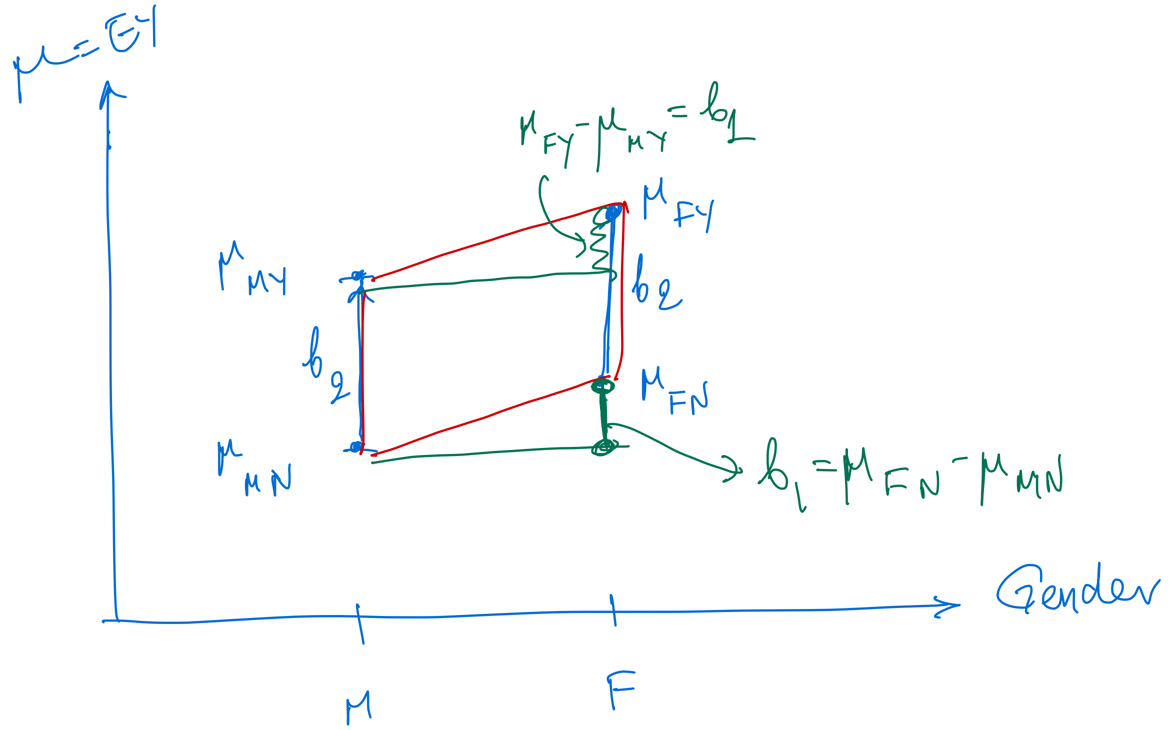
$b_1 = \mu_{FN} - \mu_{MN}$ = επίδραση του φύλου στους μη καπνιστές

$= \mu_{FY} - \mu_{MY}$ = " " " " καπνιστές

$$b_2 = \boxed{\mu_{MY} - \mu_{MN} = \mu_{FY} - \mu_{FN}}$$

$\mu_{MY} - \mu_{MN}$ = (effect) επίδραση καπνίσματος στους άνδρες

$\mu_{FY} - \mu_{FN}$ = " " στις γυναίκες



Τι γίνεται [πως μπορούμε να το διαγνώσουμε κ' να το εκτιμήσουμε]

αν η επίδραση του καπνίσματος διαφέρει μεταξύ των δύο φύλων ή

η επίδραση του φύλου διαφέρει μεταξύ καπνιστών κ' μη καπνιστών;

ΧΡΕΙΑΖΟΜΑΣΤΕ ΔΙΑΦΟΡΕΤΙΚΟ ΜΟΝΤΕΛΟ.

$$Y = b_0 + b_1 X_F + b_2 X_Y$$

π.χ. - $Y = b_0 + b_1 X_F + b_2 X_F^2 + b_3 X_Y + b_4 X_Y^2$

$X_F^2 = X_F$ (επειδή $X_F \in \{0,1\}$)

όχι

$$Y = b_0 + b_1 X_F + b_2 X_Y + b_3 X_F X_Y$$

G	S	X_F	X_Y	$Y = b_0 + b_1 X_F + b_2 X_Y + b_3 X_F X_Y$
M	N	0	0	$\mu_{MN} = b_0$
M	Y	0	1	$\mu_{MY} = b_0 + b_2$
F	N	1	0	$\mu_{FN} = b_0 + b_1$
F	Y	1	1	$\mu_{FY} = b_0 + b_1 + b_2 + b_3$

$\mu_{MY} - \mu_{MN} = b_2$ = επίδραση καπν. σε M

$\mu_{FY} - \mu_{FN} = b_2 + b_3$ = " " σε F

$$\mu_{FN} - \mu_{MN} = b_1$$

$$\mu_{FY} - \mu_{MY} = b_1 + b_3$$

$$b_3 = \underbrace{(\mu_{FY} - \mu_{FN})} - \underbrace{(\mu_{MY} - \mu_{MN})}$$

$$= (\mu_{FY} - \mu_{MY}) - (\mu_{FN} - \mu_{MN})$$

διαφορά επιδράσεων

όρος αλληλεπίδρασης

(interaction)

$$Y = b_0 + \underbrace{\overbrace{b_1 X_F}^{G} + \overbrace{b_2 X_Y}^{S}}_{\text{κύριες επιδράσεις (main effects)}} + \underbrace{b_3 X_F X_Y}_{\text{αλληλεπιδράσεις (interaction effects)}}$$

b_1 : επίδραση φύλου όταν καπνίστρια στο επίπεδο αναφοράς

b_2 " καπνίσματος " φύλο " " "

Έλεγχος $H_0: b_3 = 0$ $H_1: b_3 \neq 0$

Αν accept H_0 $\Rightarrow$ δεν υπάρχει ένδειξη αλληλεπίδρασης.

Γενίκευση

$$X_1 = \begin{cases} a & \text{αναφ} \\ b \\ c \\ d \end{cases} \qquad X_2 = \begin{cases} 1 & \text{αναφ.} \\ 2 \\ 3 \\ 4 \end{cases}$$

$\downarrow$

X_b $\quad Z_2$

X_c $\quad Z_3$

X_d $\quad Z_4$

$$Y = b_0 + \underbrace{b_1 X_b + b_2 X_c + b_3 X_d}_{\text{main effects } X_1} + \underbrace{b_4 Z_2 + b_5 Z_3 + b_6 Z_4}_{\text{main effects } X_2}$$

[main effects model]

$$\left. \begin{aligned} &+ b_7 X_b Z_2 + b_8 X_b Z_3 + b_9 X_b Z_4 \\ &+ b_{10} X_c Z_2 + b_{11} X_c Z_3 + b_{12} X_c Z_4 \\ &+ b_{13} X_d Z_2 + b_{14} X_d Z_3 + b_{15} X_d Z_4 \end{aligned} \right\} \text{interaction effects.}$$

Analysis of Covariance (ANACOVA)

Παράγοντες $\begin{cases} \text{κατηγορικές μεταβλητές} \\ + \\ \text{ποσοτικές } '' \end{cases}$

Παράδειγμα

Y = score φυσ- κατάστασης

A = age (ποσοτική)

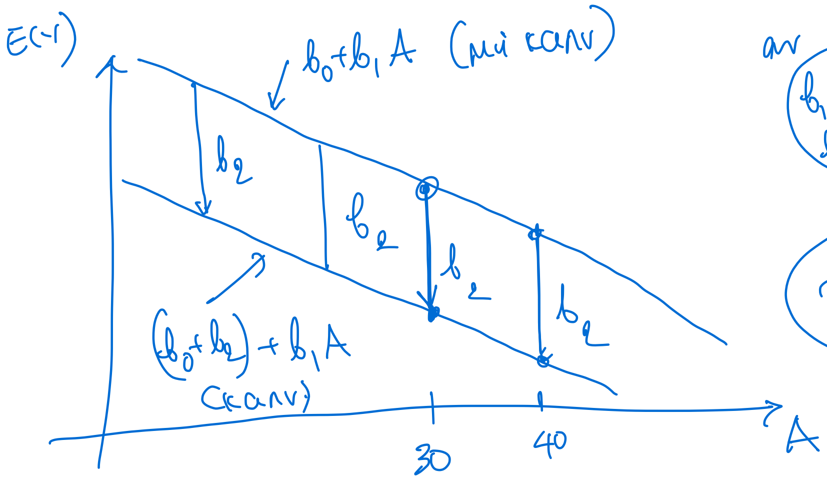
S = κάπνισμα $(N|Y)$ $\begin{cases} N : \text{επ. αναφ.} \\ X_Y = 1, \text{ Smoking=Yes} \end{cases}$

$$Y = b_0 + \underbrace{b_1 A}_{\text{effect age}} + \underbrace{b_2 X_Y}_{\text{effect smoking}}$$

Smoking : No $\Rightarrow$ $X_Y = 0$ $\Rightarrow$ $Y = b_0 + b_1 \cdot A$ ← ευθεία Y~age για μη καπνι

Yes $\Rightarrow$ $X_Y = 1$ $\Rightarrow$ $Y = b_0 + b_1 A + b_2$

$= (b_0 + b_2) + b_1 A$ ← ευθεία Y~age καπνιστές



αν $b_1 < 0$, $b_2 < 0$

Παράλληλες !!

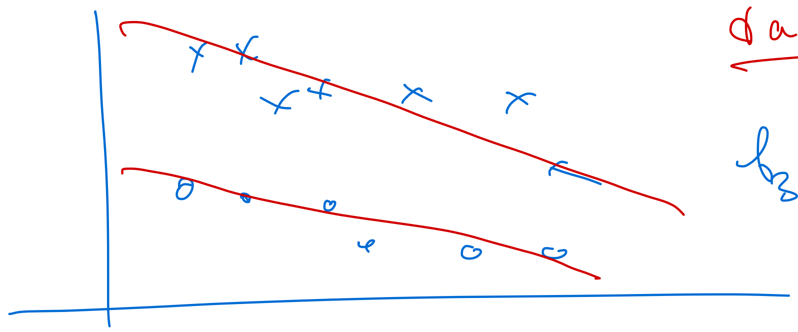
Επίδραση ηλικίας σε καπνιστές $= b_1$

'' '' σε μη καπν $= b_1$

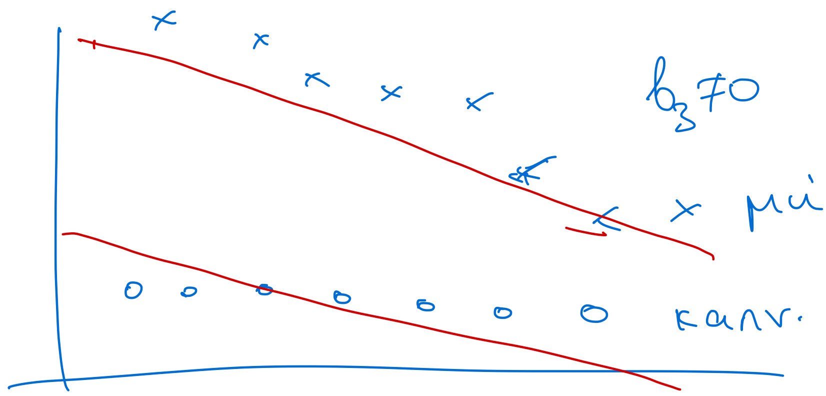
Επίδραση του καπνίσματος $= b_2$ για όλες τις τιμές του Age.

---

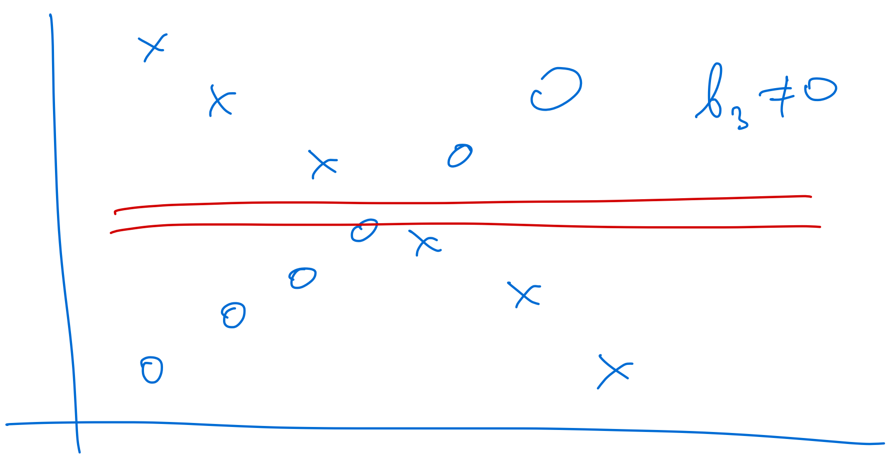
dataset



$b_3 \approx 0$.



$b_3 \neq 0$



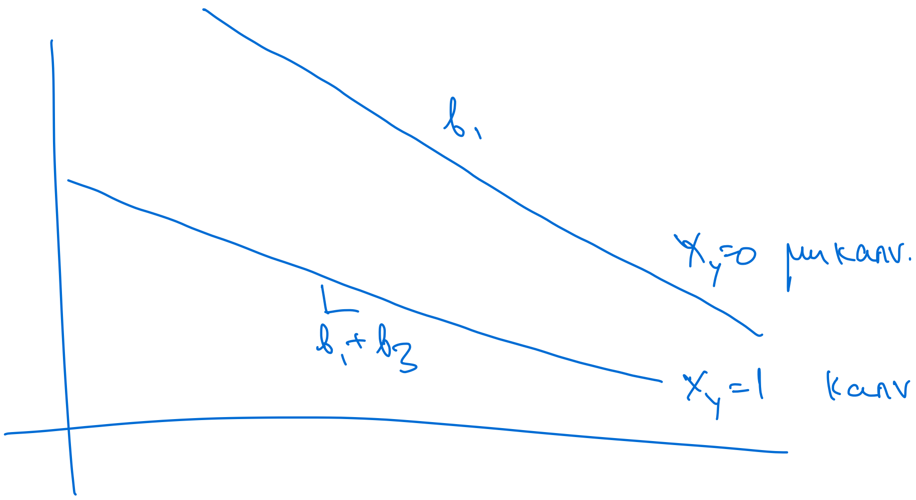
$b_3 \neq 0$

Μοντέλο αλληλεπιδράσεων

$$Y = b_0 + b_1 A + b_2 X_Y + b_3 X_Y A$$

Μη καπν. $X_Y = 0$: $Y = b_0 + b_1 A$

καπν. $X_Y = 1$: $Y = (b_0 + b_2) + (b_1 + b_3) \cdot A$



$H_0 : b_3 = 0$. $H_1 : b_3 \neq 0$ έλεγχος για ύπαρξη αλληλεπιδράσεων

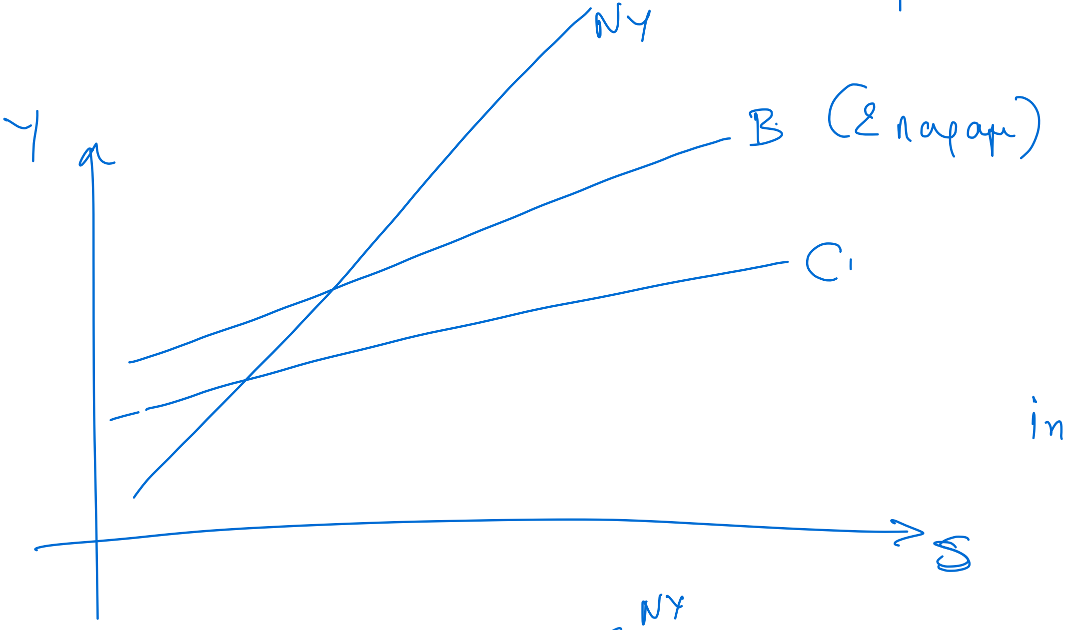
Παράδ. 2 Τιμές διαμερισμάτων

Y = τιμή

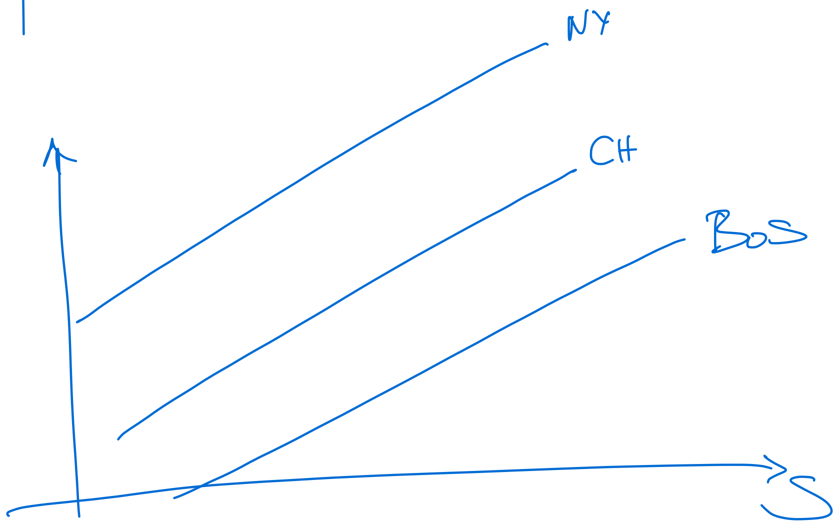
S = μέγεθος (σε τμ). (ποσοτική)

C = city $\begin{cases} NY \leftarrow \text{αναφ.} \\ B \quad X_B \\ Chi \quad X_{CH} \end{cases}$

$$Y = b_0 + b_1 S + b_2 X_B + b_3 X_{CH} + b_4 X_B S + b_5 X_{CH} \cdot S$$



interactions



main effects

$$Y = b_0 + b_1 S + b_2 X_{CH} + b_3 X_B$$

Interactions Model

$$Y = b_0 + b_1 S + b_2 X_B + b_3 X_{CH} + b_4 X_B S + b_5 X_{CH} \cdot S$$

NY $\quad Y = b_0 + b_1 S$

CH $\quad Y = (b_0 + b_3) + (b_1 + b_5) S$ $\quad$ διαφορά επίδρασης του S μεταξύ NY και CH. (NY: αναφ.)

B $\quad Y = (b_0 + b_2) + (b_1 + b_4) S$ $\quad$ διαφ. επίδρασης του S μεταξύ NY κ' Boston (NY: επ. αναφ.)