22-1-2025

Y : price

X : city levels $\begin{cases} \text{Chi} \\ \text{Bos} \\ \text{NY} \end{cases}$

$Y = f(X) + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$

X' : κωδικοποίηση της X

πχ. $X' = \begin{matrix} \text{Chi} = 1 \\ \text{Bos} = 2 \\ \text{NY} = 3 \end{matrix}$?

$Y = b_0 + b_1 X' + \varepsilon$

$E(Y_{CHI}) = b_0 + b_1 \cdot 1 = b_0 + b_1$

$E(Y_{BOS}) = b_0 + b_1 \cdot 2 = b_0 + 2b_1$ $\Rightarrow$

$E(NY) = \quad = b_0 + 3b_1$

$E(Y_{BOS}) - E(CHI_1) = b_1$

$E(NY) - E(Y_{BOS}) = b_1$

Υπόθεση της διαφάνειας εξαρτώμενες

Αν $X'' = \begin{cases} 1, & \text{Chi} \\ 5, & \text{Bos} \\ -2, & \text{NY} \end{cases}$ $\quad Y = b_0 + b_1 X$

$$\left.\begin{aligned} E(Y_{CHI}) &= b_0 + b_1 \\ E(Y_{BOS}) &= b_0 + 5b_1 \\ E(Y_{NY}) &= b_0 - 2b_1 \end{aligned}\right\} \text{ανιστοιχο πρόβλημα}$$

Κωδικοποίηση μέσω ψευδομεταβλητών

$$X \Rightarrow \begin{cases} X_{CHI} = \begin{cases} 1 & \text{αν } X = CHI \\ 0, & \text{αν } X \neq CHI \end{cases} \\ X_{BOS} = \begin{cases} 1, & \text{αν } X = BOS \\ 0, & \text{αν } X \neq BOS \end{cases} \\ X_{NY} = \begin{cases} 1, & \text{αν } X = NY \\ 0, & \text{αν } X \neq NY \end{cases} \end{cases}$$

[Δείκτης του $X = Chi$]

$$Y = b_0 + b_1 X_{CHI} + b_2 X_{BOS} + b_3 X_{NY}$$

?

(4 αγγλικά πλάνα)

X	X_{CHI}	X_{BO}	X_{NY}
CHI	1	0	0
BOS	0	1	0
CHI	1	0	0
NY	0	0	1
BOS	0	1	0

Ανάλυση παραλλαγών 1 level (πχ. NY)

$$Y = b_0 + b_1 X_{CHI} + b_2 X_{BOS}$$

3 ανεξάρτητες παράμετροι

$\checkmark$ LSE

$$\hat{b}_0, \hat{b}_1, \hat{b}_2 \overset{?}{\Longleftrightarrow} \hat{\mu}_{CHI}, \hat{\mu}_{BOS}, \hat{\mu}_{NY}$$

Level	EY	$Y = b_0 + b_1 X_{CHI} + b_2 X_{BOS}$
CHI	$\mu_{CHI} \Longleftrightarrow$	$Y = b_0 + b_1 \cdot 1 = E(Y \mid X = CHI) = \mu_{CHI}$
BOS	$\mu_{BOS} \Longleftrightarrow$	$Y = b_0 + b_2$
NY	$\mu_{NY} \Longleftrightarrow$	$Y = b_0$

$$b_0 = \mu_{NY}$$

$$b_0 + b_1 = \mu_{CHI} \Rightarrow b_1 = \mu_{CHI} - \mu_{NY}$$

$$b_0 + b_2 = \mu_{BOS} \Rightarrow b_2 = \mu_{BOS} - \mu_{N}$$

!!

Εξαρτώνται

$\hat{b}_0 = 258.50 \Longleftrightarrow \hat{\mu}_{NY} = 258.5 = \hat{b}_0$

$\hat{b}_1 = -67,5$ $\quad \hat{\mu}_{CHI} = 191$ $\quad \hat{\mu}_{CHI} - \hat{\mu}_{NY} = -67,5 = \hat{b}_1$

$\hat{b}_2 = -56,86$ $\quad \hat{\mu}_{BOS} = 201.64$ $\quad \hat{\mu}_{BOS} - \hat{\mu}_{NY} = -56 = \hat{b}_2$

Γενικός Κανόνας Δίτιμων Μεταβλητών

Αν X : κατηγορική μεταβλητή
με k επίπεδα $X = \alpha_1, \alpha_2, \dots \alpha_k$

1) Επιλέγουμε (αυθαίρετα) ένα επίπεδο
<u>επίπεδο αναφοράς</u> (reference level)
πχ. a_k

2) Για κάθε ένα από τα υπόλοιπα επίπεδα
ορίζουμε μια δίτιμη μεταβλητή 0-1.

$$X_1 = \begin{cases} 1 & X = a_1 \\ 0 & \text{διαφ} \end{cases} \quad \dots \quad X_{k-1} = \begin{cases} 1, & X = a_{k-1} \\ 0, & \text{διαφ.} \end{cases}$$

$$Y = \underline{\underline{b_0}} + b_1 X_1 + b_2 X_2 + \dots + b_{k-1} X_{k-1}$$

<u>Ερμηνεία</u>

$$\hat{b}_0 = \hat{\mu}_{a_k} = E(Y \mid X = a_k)$$

$$\hat{b}_1 = \hat{\mu}_{a_1} - \hat{\mu}_{a_k}$$

$$\hat{b}_2 = \hat{\mu}_{a_2} - \hat{\mu}_{a_k}$$

$$\vdots$$

$$\hat{b}_{k-1} = \hat{\mu}_{a_{k-1}} - \hat{\mu}_{a_k}$$

Πίσω στο $Y = b_0 + b_1 X_{CHI} + b_2 X_{BOS}$

F-test για ολο το μοντέλο:

$H_0: b_1 = b_2 = 0$ $\qquad$ $H_1: b_1 \neq 0$ ή $b_2 \neq 0$ ή κ' τα δύο

ΟΜΩΣ:

Αν $b_1 = 0 \Rightarrow \mu_{CH} - \mu_{NY} = 0$

$b_2 = 0 \Rightarrow \mu_{BOS} - \mu_{NY} = 0$

$\Rightarrow \mu_{CHI} = \mu_{BOS} = \mu_{NY}$!

$H_0: \mu_{CHI} = \mu_{BOS} = \mu_{NY}$ $\qquad$ H_1: τουλαχ. 2 διαφορετικά.

αυτό που δείξαμε αρχικά

---

2 κατηγορικές ανεξ. μεταβλητές

$SEX = \{ M, F \}$

$SM = \{ NO, YES \}$

SEX \ SM	NO	YES
M	μ_{MN}	μ_{MY}
F	μ_{FN}	μ_{FY}

Y: response variable

$\mu_{MN} = E(Y \mid Sex = M, Sm = NO)$

Κωδικοποίηση

$X_M = 1(Sex = M)$

$X_Y = 1(Smoking = YES)$

$Y = b_0 + b_1 X_M + b_2 X_Y$ } Main effects model
Μοντέλο κύριων επιδράσεων

Sex	Smoking	X_M	X_Y	Y
F	NO	0	0	$\mu_{FN} = b_0$
F	Yes	0	1	$\mu_{FY} = b_0 + b_2$
M	NO	1	0	$\mu_{MN} = b_0 + b_1$
M	Yes	1	1	$\mu_{MY} = b_0 + b_1 + b_2$

main effects

$b_0 = \mu_{FN}$

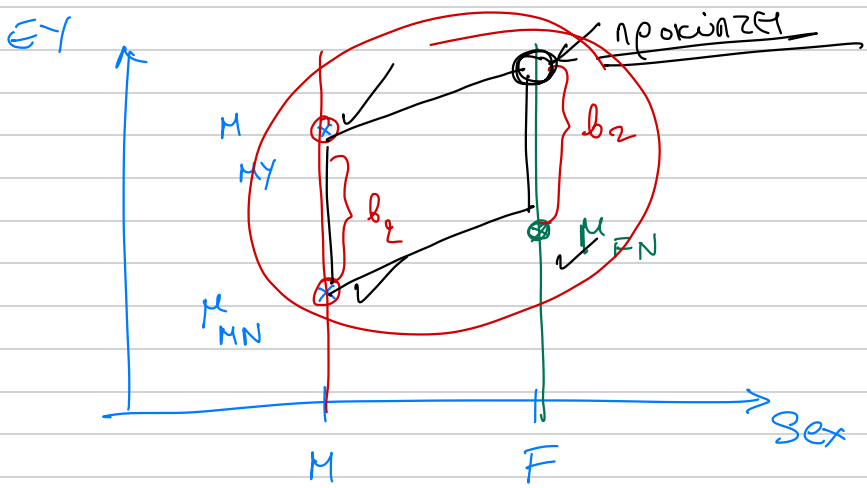
$b_1 = \mu_{MN} - \mu_{FN} = \mu_{MY} - \mu_{FY} \Rightarrow$ (1)

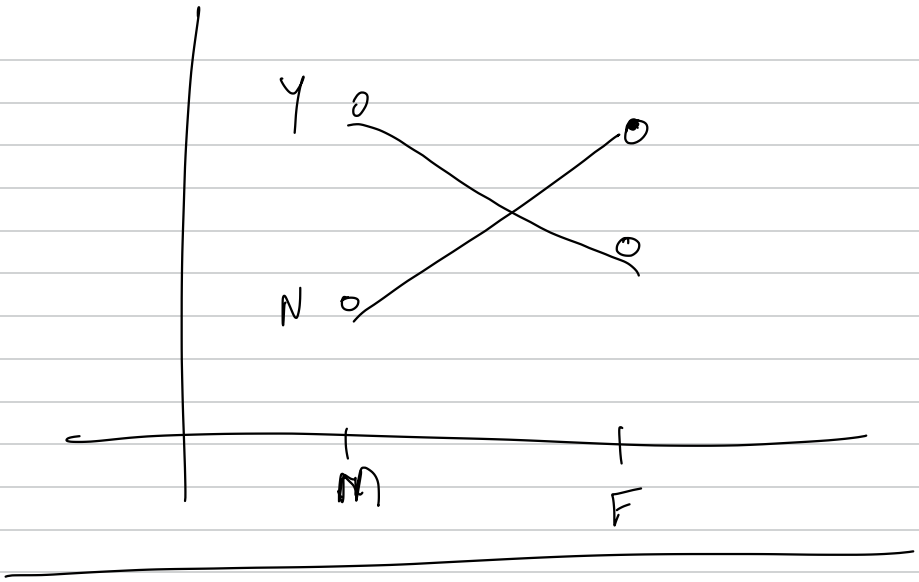
$b_2 = \mu_{FY} - \mu_{FN} = \mu_{MY} - \mu_{MN}$ (2)

$\Rightarrow \boxed{\mu_{MN} - \mu_{FN} = \mu_{MY} - \mu_{FY}.}$ περιορισμός στα μ.

(1) Effect φύλου ίδιο σε { καπνιστές / μη καπνιστές

(2) Effect καπνίσματος ίδιο σε { M / F





$$Y_0 = b_0 + b_1 X_M + b_2 X_Y + b_3 X_M \cdot X_Y$$

$$X_M = \begin{cases} 0 \\ 1 \end{cases} \qquad X_M^2 = \begin{cases} 0 \\ 1 \end{cases} \qquad X_M^2 = X_M \quad \times$$

Sex	Smok	X_M	X_Y	$X_M X_Y$	Y
F	N	0	0	0	$\mu_{FN} = b_0$
F	Y	0	1	0	$\mu_{FY} = b_0 + b_2$
M	N	1	0	0	$\mu_{MN} = b_0 + b_1$
M	Y	1	1	1	$\mu_{MY} = b_0 + b_1 + b_2 + b_3$

Effects sex

N : $\mu_{MN} - \mu_{FN} = b_1$ effect sex or Smoking = NO (En. avg)

Y : $\mu_{MY} - \mu_{FY} = b_1 + b_3$ " " " Smoki = Y

Effects of smoking

F : $\mu_{FY} - \mu_{FN} = b_2$

$M : \mu_{MY} - \mu_{MN} = b_2 + b_3$

b_3 : συντελ. αλληλεπίδρασης (interaction coefficient)

= effect (M − F) σε καν − effect (M−F) σε μη καν.

επ. ανάλ. Sun

= effect (N − Y) σε αντρες − effect (N−Y) σε γυναίκες

επ. ανάλ. φύλου

επ. ανάλ. ηλικ

π.χ. $H_0 : b_3 = 0$ $H_1 : b_3 \neq 0$

ύπαρξη αλληλεπιδράσεων

Γενικεύσεις

	Χ = χρώμα	Επαρχ	Τόπος διαμ.
reference →	R	Αντ. (επαναλ.)	Αζιάκι (επ.αναλ.)
X_G	G	$X_{\kappa\rho}$ Κέρκυρα	Λιμ. αστικό κέντρο $X_{\alpha\sigma}$
X_B	B	$X_{\Gamma\rho}$ Γραφείον	Επαρχία $X_{\epsilon\pi}$
X_Y	Y	X_M Μεικτή	

48 συνδυασμοί (μ)

Main effects

$$Y = \begin{cases} b_0 + b_1 X_G + b_2 X_B + b_3 X_Y + b_4 X_{\kappa\rho} + b_5 X_{\Gamma\rho} + b_6 X_M + \\ + b_7 X_{\alpha\sigma} + b_8 X_{\epsilon\pi} \end{cases}$$

interaction effects

$$\begin{cases} + b_9 X_G X_{\kappa\rho} + b_{10} X_G X_{\Gamma\rho} + \dots + \dots b X_M X_{\epsilon\pi} \\ + b_{\dots} X_G X_{\kappa\rho} X_{\alpha\sigma} \dots \end{cases}$$