Έλεγχοι Υποθέσεων στο Μοντέλο Πολλαπλής Παλινδρ.

(1) $H_0: b_j = 0$ vs. $H_1: b_j \neq 0$

t-test.

Δίνει πληροφορία για τη σημαντικότητα της X_j όταν **η X_j μπαίνει τελευταία στο μοντέλο**

(παρουσία όλων των υπολοίπων μεταβλητών του μοντέλου).

Ελέγχω Σημαντικότητα της X_j ελέγχοντας ως προς τις άλλες μεταβλητές

Score	age	extime
Y	X_1	X_2

$Y = b_0 + b_1$ extime $\Rightarrow$ $H_0: b_1 = 0$: $H_1: b_1 \neq 0$

$p = 2 \cdot 10^{-16} \simeq 0$

extime ισχυρή συσχέτιση με το score

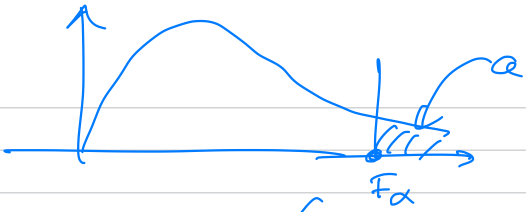
$Y = b_0 + b_1\,age + b_2\,extime \Rightarrow$ $H_0: b_2 = 0$: $H_1: b_2 \neq 0$

$p = 2 \cdot 10^{-11} \simeq 0$

ισχυρή συσχέτιση extime με score

έχοντας ελέγξει ως προς την ηλικία !!

1



$$F = \frac{MSR}{MSE} = \frac{\frac{SSR}{df_{mod}}}{\frac{SSE}{df_{er}}} \sim F(df_{mod}, df_{er})$$

F statistic συνολικού μοντέλου

Έλεγχος F για συνολικό μοντέλο

$H_0: b_1 = b_2 = \dots = b_k = 0$ $\quad H_1:$ τουλάχιστον ένα $\neq 0$

$H_0: EY = b_0$ (στο το μοντέλο μη στατιστικά σημαντικό)

$F > F_\alpha \Leftrightarrow$ reject H_0 $\quad p < \alpha$

$F \leq F_\alpha \Leftrightarrow$ accept H_0 $\quad p \geq \alpha$

Ειδική περίπτωση $(k=1)$

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

$$F = \frac{MSR}{MSE} = \frac{\frac{SSR}{1}}{\frac{SSE}{df_{er}}} \sim F(1, df_{er})$$

$H_0: b_1 = 0 \quad H_1: b_1 \neq 0$

Όμως έχουμε δει ότι

$$\left\{ \begin{array}{l} H_0 : b_1 = 0 \qquad H_1 : b_1 \neq 0 \\ \text{στο μονοπαραμετρικό} \end{array} \right\} \text{t-test για } b_1 .$$

$$t = \frac{\hat{b}_1}{s_{\hat{b}_1}}$$

Στο μοντέλο $Y = b_0 + b_1 X$

$$F = \frac{MSR}{MSE} = t^2$$

Από πιθ. $F(1, k) \overset{d}{=} t_k^2$

$$F_a = t^2_{a/2}$$

$$\underbrace{F \leq F_a}_{\text{accept } H_0} \iff t^2 \leq t^2_{a/2} \iff \underbrace{|t| \leq t_{a/2}}_{\text{accept } H_0}$$

(2) Έλεγχος F για μέρος του μοντέλου

Model 1 (full model) $Y = \boxed{b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p} + b_{p+1} X_{p+1} + \cdots + b_k X_k$

Model 2 (partial model) $Y = \boxed{b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p}$, $(p<k)$

Model 2 nested στο Model 1 (εμφωλευμένο)

Ξέρουμε $SSR(X_1, X_2, \ldots, X_p, \ldots X_k) \geq SSR(X_1, \ldots, X_p)$

$$\frac{SSE(X_1, X_2, \ldots X_p, \ldots, X_k)}{SST} \leq \frac{SSE(X_1, \ldots, X_k)}{SST}$$

Επομένως $SSR(X_1, \ldots, X_k) - SSR(X_1, \ldots, X_p) =$

$$= SSR(X_{p+1}, \ldots, X_k \mid X_1, \ldots, X_p) \geq 0$$

$$\Rightarrow R^2_{full} \geq R^2_{partial}$$

$$H_0 : \underbrace{\beta_{p+1} = \beta_{p+2} = \cdots = \beta_k = 0}_{k-p \text{ συντελεστές}} , \quad H_1 : \text{τουλάχιστον ένα} \neq 0$$

έλεγχος για μέρος του model 1

$$SSR_{full} + SSE_{full} = SSR_{par} + SSE_{par} \Rightarrow$$

$$\Rightarrow SSR_{full} - SSR_{par} = SSE_{par} - SSE_{full}$$

$$df_{er, full} = n - (k+1) = n-k-1$$

$$df_{er, par} = n - (p+1) = n-p-1 > n-k-1$$

$$df_{er,par} - df_{er,full} = k-p$$

$H_0 : b_{p+1} = b_{p+2} = \cdots = b_k = 0$, H_1 : τουλάχιστον ένα $\neq 0$

$$F = \frac{\dfrac{SSE_{par} - SSE_{full}}{\underbrace{df_{er,par} - df_{er,full}}_{=k-p}}}{MSE_{full}} \sim F(k-p,\ n-k-1)$$

accept H_0 αν $F \leq F_\alpha(k-p,\ n-k-1)$

reject H_0 αν $F > F_\alpha(k-p,\ n-k-1)$

εδ. περίπτωση 1 Ο έλεγχος F για όλο το Μοντέλο ειδική περίπτωση

full $Y = b_0 + b_1 X_1 + \cdots + b_k X_k$ full

part $Y = b_0$

$H_0 : b_1 = b_2 = \cdots = b_k = 0.$ H_1: τουλ ένα $\neq 0$

Το partial model εδώ $Y = b_0 \Rightarrow \hat{b}_0 = \bar{Y}$

$$SSE_{partial} = \sum (Y_i - \bar{Y})^2 = SST$$

$$SSE_{par} - SSE_{full} = SST - SSE_{full} = SSR_{full}$$

$$df_{er,par} - df_{er,full} = k.$$

$$F = \frac{\frac{SSR_{full}}{df_{mod,full}}}{MSE_{full}} = \frac{MSR_{full}}{MSE_{full}} = F$$

↓ για το test όλων των μοντέλων

Ειδική περίπτωση 2

full $\quad Y = b_0 + b_1 X_1 + \cdots + b_p X_p + b_{p+1} X_{p+1}$

partial $\quad Y = b_0 + b_1 X_1 + \cdots + b_p X_p$

$$\boxed{H_0 : b_{p+1} = 0 \qquad H_1 : b_{p+1} \neq 0.}$$

$\Updownarrow$

t-test b_{p+1}

Ο έλεγχος t για $b_j = 0$ ή $b_j \neq 0$ ελέγχει τη σημαντικότητα της X_j **δεδομένων** όλων των υπόλοιπων μεταβλητών (δηλ. πού η X_j μπαίνει τελευταία παρουσία όλων των άλλων).

(3) Έστω ότι στο μοντέλο

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

οι μεταβλητές μπαίνουν με τη σειρά

$$X_1, X_2, \ldots, X_k$$

Model 1 :	X_1		
" 2	X_1, X_2	F_1	$H_0: b_2 = 0 \quad H_1: b_2 \neq 0$ στο Model 2
3	X_1, X_2, X_3	F_2	$H_0: b_3 = 0 \quad H_1: b_3 \neq 0$ Model 3
⋮		⋮	
k	$X_1, X_2, X_3, \ldots, X_k$	F_k	

$F_1, F_2, \ldots, F_k$: Ftests type I (sequential Ftests)

εξαρτώνται από τη σειρά εισαγωγής

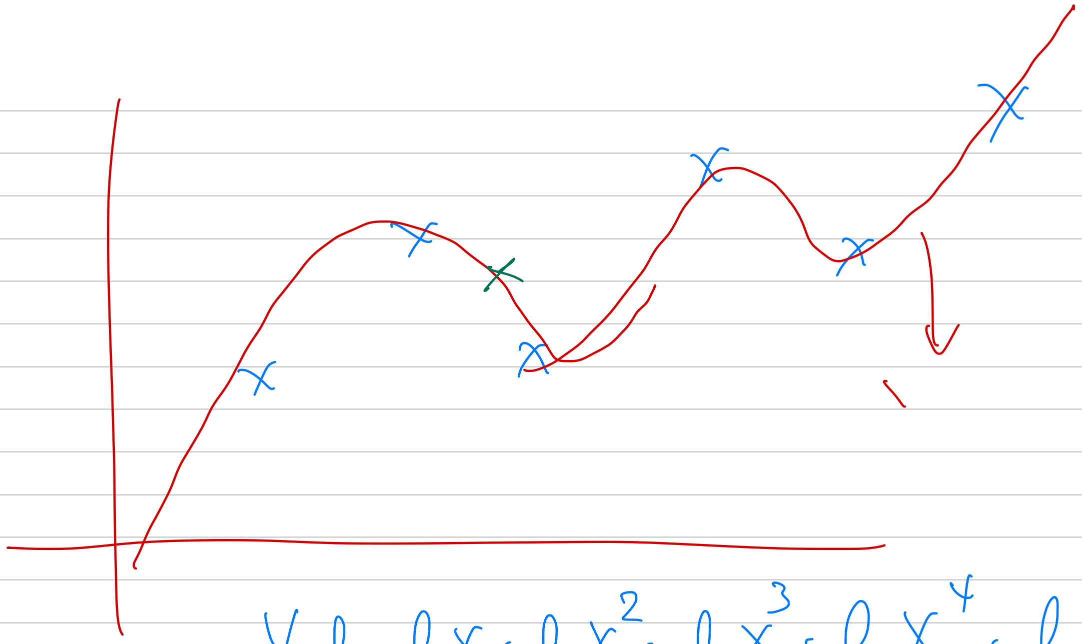
(4) Ftests type III (partial)

$H_0: b_j = 0 \qquad H_1: b_j \neq 0$ ~~στο~~ όλα στο Model K

(Η X_j τελευταία) $\Leftrightarrow$ t-tests

δεν εξαρτώνται από τη σειρά εισαγωγής

$n=6$



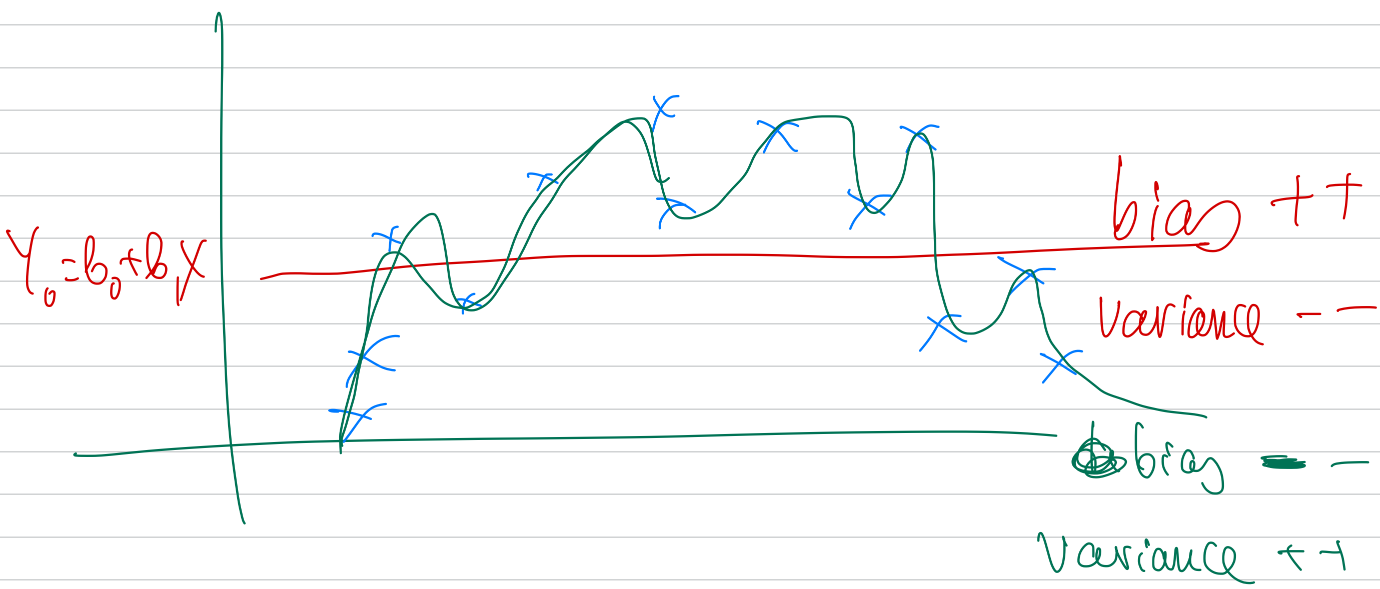
$$Y = b_0 + b_1 X + b_2 X^2 + b_3 X^3 + b_4 X^4 + b_5 X^5$$

$SSE = 0 \Rightarrow R^2 = 1$

$df_{er} = 0$

$$MSE = \frac{0}{0} \;?$$

$$F = \frac{MSR}{MSE} \;??$$



Κατηγορικές Ανεξάρτητες Μεταβλητές

1) X : κατηγορική (qualitative) τιμές : (σύνολο)
πχ. χρώμα {R, G, B}

2) X : διάταξης (ordinal)
υπάρχει διάταξη μεταξύ τιμών
πχ. Likert scale (καθόλου, λίγο, αρκετά, ...)

3) X : κλίμακας/ποσοτική (scale/quantitative)
έχει μονάδα μέτρησης

Τι σημαίνει γραμμικό μοντέλο όταν X : κατηγορική

(ANOVA Method)

Παράδειγμα

City (X)	Price (Y)
Chi	'
Chi	'
⋮	
Bos	'
⋮	'
Bos	
NY	'
NY	
⋮	'

3 παράμετροι: $\mu_{CHI}, \mu_{BOS}, \mu_{NY}$

$$Y|Chi \sim \mathcal{N}(\mu_{CHI}, \sigma^2) \qquad Y|Chi = \mu_{Chi} + \varepsilon$$

$$Y|Bos \sim \mathcal{N}(\mu_{BOS}, \sigma^2) \qquad = \mu_{BOS} + \varepsilon$$

$$Y|NY \sim \mathcal{N}(\mu_{NY}, \sigma^2) \qquad = \mu_{NY} + \varepsilon$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

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

$b_0 + b_i \cdots ?$

$$f(x) = \begin{cases} \mu_{CHI}, & X = Chi \\ \mu_{BOS}, & X = Bos \\ \mu_{NY}, & X = NY \end{cases}$$

$\hat{\mu}_{CHI} = ?$ $\quad \hat{\mu}_{BOS} = ?$, $\quad \hat{\mu}_{NY} = ?$

$\hat{\mu}_{CHI} = \overline{X}_{CHI}$, --- --- ---

$\hat{\mu}_{CHI} = 191$, $\quad \hat{\mu}_{BOS} = 201.6$, $\quad \hat{\mu}_{NY} = 258.5$

$H_0: \mu_{CHI} = \mu_{BOS} = \mu_{NY}$ $\qquad H_1$: τουλάχιστον δύο διαφέρουν

Επέκταση του t-test

Μέθοδος ANOVA (ειδική περίπτωση Γραμμ. μοντέλου

$Y = b_0 + b_1 X_1 + \cdots$

Δείγμα City

3 επίπεδα (levels): Chi, Bos, NY $\Rightarrow$

$X_1, X_2, \ldots$

ποσοτικές

$Y = b_0 + b_1 X_1$

ισοδύναμα αποτελ.

με $\hat{\mu}_{Chi}$ $\hat{\mu}_{Bos}$ ...