Μάθημα 6ο

Γενική γραμμική υπόθεση

ΜΟΝΤΕΛΟ Π.Γ.Π: $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$

Γενική γραμμική υπόθεση: $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$ όπου A είναι $q \times (p+1)$ με $rank(A) = q$

$\underset{\sim}{c}$ είναι $q \times 1$ διάνυσμα

Πρακτικό ενδιαφέρον που έχει η $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$

ΠΑΡΑΔΕΙΓΜΑΤΑ:

α) Έστω $A = I_{p+1}$ και $\underset{\sim}{c} = \underset{\sim}{0}_{p+1}$. Τότε $H_0: A \cdot \underset{\sim}{\beta} = \underset{\sim}{c} \Leftrightarrow H_0: \underset{\sim}{\beta} = \underset{\sim}{0}$

β) Έστω $A = I_{p+1}$ και $\underset{\sim}{c} = \underset{\sim}{\beta}^*$ ($\underset{\sim}{\beta}^*$ γνωστό) Τότε $H_0: A \cdot \underset{\sim}{\beta} = \underset{\sim}{c} \Leftrightarrow H_0: \underset{\sim}{\beta} = \underset{\sim}{\beta}^*$

γ) Έστω $A = \begin{pmatrix} 1 & 0 & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & 0 & 1 \end{pmatrix}$

Έστω $\underset{\sim}{c} = 0$ ~~από τις συνιστώσες του $\underset{\sim}{\beta}$~~

Τότε $H_0: A \cdot \underset{\sim}{\beta} = \underset{\sim}{c} \Leftrightarrow H_0: \beta_{(1)} = 0$ όπου $\underset{\sim}{\beta}$ περιέχει κάποιες από τις συνιστώσες του

Βήματα για κατασκευή τεστ για έλεγχο H_0:

Ονομάζουμε **Πλήρες Μοντέλο** το $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$. ή

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_p X_{ip} + \varepsilon_i$, $i = 1, \dots, n$

Ονομάζουμε **H_0-Μοντέλο (H_0M)** το μοντέλο που προκύπτει από το ΠΜ αν λάβω υπόψη μου τους περιορισμούς που θέτει στα β_i η ΓΓΥ, $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$

ΠΑΡΑΔΕΙΓΜΑ: Έστω $A = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ 0 & & & & \ddots \; 0 \end{pmatrix}$, $\underset{\sim}{c} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

Τότε $A \cdot \underset{\sim}{\beta} = \underset{\sim}{c} \Leftrightarrow \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ 0 & & & & \ddots \; 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \Leftrightarrow \begin{cases} \beta_0 = 0 \\ \beta_1 = 0 \\ \beta_2 = 0 \end{cases}$, Άρα ($H_0M$) θα είναι:

(H_0M) $Y_i = \beta_3 X_{i3} + \dots + \beta_p X_{ip}$

Έστω $\hat{\underset{\sim}{\beta}}$ οι ΕΕΤ της $\underset{\sim}{\beta}$ στο ΠΜ.

Έστω $\hat{\hat{\underset{\sim}{\beta}}}_{H_0}$ οι ΕΕΤ της $\underset{\sim}{\beta}$ στο H_0Μ.

Τα εκτιμώμενα μοντέλα θα είναι $\hat{\underset{\sim}{Y}} = X\hat{\underset{\sim}{\beta}}$ και $\hat{\underset{\sim}{Y}}_{H_0} = X\hat{\underset{\sim}{\beta}}_{H_0}$ ενώ τα αντίστοιχα αθροίσματα τετραγώνων που οφείλονται στα υπόλοιπα.

$$SSres(\Pi.M) \overset{\text{ορ.}}{=} (\underset{\sim}{Y} - \hat{\underset{\sim}{Y}})'(\underset{\sim}{Y} - \hat{\underset{\sim}{Y}})$$

$$SSres(H_0M) \overset{\text{ορ.}}{=} (\underset{\sim}{Y} - \hat{\underset{\sim}{Y}}_{H_0})'(\underset{\sim}{Y} - \hat{\underset{\sim}{Y}}_{H_0})$$

Αν $SSres(\Pi M) \simeq SSres(H_0M)$ τότε η $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$ δεν μπορεί να απορριφθεί.
Άρα ένα τεστ για τον έλεγχο της H_0, $A\underset{\sim}{\beta} = \underset{\sim}{c}$ πρέπει να στηριχτεί στη διαφορά $SSres(\Pi M) - SSres(H_0M)$, και αν η διαφορά έχει μεγάλη τιμή, απορρίπτουμε $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$

Τεχνικά: Αποδεικνύεται:

(i) $SSres(H_0M) - SSres(\Pi M) = (A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})'[A(X'X)^{-1}A']^{-1}(A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})$

(ii) $\dfrac{SSres(H_0M) - SSres(\Pi M)}{\sigma^2} \sim \chi^2_q$

Επιπλέον, γνωρίζουμε ότι $\dfrac{SSres(\Pi M)}{\sigma^2} \sim \chi^2_{n-p-1}$

Επίσης, $\hat{\underset{\sim}{\beta}}$, $SSres(\Pi M)$ ανεξάρτητα.
Α… $SSres(H_0M) - SSres(\Pi M)$ είναι συνάρτηση του $\hat{\underset{\sim}{\beta}}$
$\Rightarrow SSres(H_0M) - SSres(\Pi M)$ ανεξ. $SSres(\Pi M)$
Θεωρώ τη στατιστική συνάρτηση: $F = \dfrac{[SSres(H_0M) - SSres(\Pi M)]/q}{SSres(\Pi M)/(n-p-1)}$

$$\Rightarrow F = \frac{[SSres(H_0M) - SSres(\Pi M)]/\sigma^2 q}{SSres(\Pi M)/\sigma^2(n-p-1)}$$

$$\Rightarrow F \overset{d}{=} \frac{\chi^2_q/q}{\chi^2_{n-p-1}/(n-p-1)} \underset{\text{ανεξ.}}{\equiv} F_{q,n-p-1} \text{ υπό } H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$$

Μεγάλες τιμές του F συνηγορούν σε απόρριψη της $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$

Συγκεντρωτικά: Για τον έλεγχο της ΠΥ $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$ Αλλά η στατιστική συνάρτηση τεστ είναι $F = \dfrac{(n-p-1)(A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})'[A(X'X)^{-1}A']^{-1}(A\underset{\sim}{\beta} - \underset{\sim}{c})}{q\,SS_{res}(\Pi.M)}$

με κατανομή $F_{q,\,n-p-1}$ υπό την $H_0: A\underset{\sim}{\beta} = \underset{\sim}{c}$ και κρίσιμη περιοχή $F \geq F_{q,\,n-p-1,\,a}$

Ανάλυση υπολοίπων

Υπόλοιπα: Βοηθούν στο να ελεγχθεί η ικανοποίηση των υποθέσεων για τα σφάλματα μοντέλου.

Θεωρητικές ιδιότητες υπολοίπων

Μοντέλο: $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \varepsilon \Rightarrow \underset{\sim}{\varepsilon} = \underset{\sim}{Y} - X\underset{\sim}{\beta}$

Υπόλοιπα: $\underset{\sim}{e} = Y - \hat{Y} \xrightarrow{\hat{Y} = X\hat{\beta}} \underset{\sim}{e} = \underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}$ $\Bigg\} \Rightarrow$

Τα υπόλοιπα $\underset{\sim}{e}$ είναι εκτιμητές των σφαλμάτων $\underset{\sim}{\varepsilon}$

ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΥΠΟΛΟΙΠΩΝ

ΙΔΙΟΤΗΤΑ 1: Τα υπόλοιπα είναι γραμμικές συναρτήσεις των παρατηρήσεων $\underset{\sim}{Y}$ και των σφαλμάτων $\underset{\sim}{\varepsilon}$.

ΑΠΟΔΕΙΞΗ: Πράγματι: $\underset{\sim}{e} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} = \underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}} = \underset{\sim}{Y} - X(X'X)^{-1}X'\underset{\sim}{Y} \Rightarrow$

$\Rightarrow \underset{\sim}{e} = (I_n - \underbrace{X(X'X)^{-1}X'}_{P})\underset{\sim}{Y} \Rightarrow$

$\Rightarrow \underset{\sim}{e} = (I_n - P)\underset{\sim}{Y}$ όπου $P = X(X'X)^{-1}X'$

ΠΑΡΑΤΗΡΗΣΗ: Ο P είναι
- (i) Συμμετρικός $\to P' = P$
- (ii) Ταυτοδύναμος $\to P^2 = P\cdot P = P$

(i) $P' = (X(X'X)^{-1}X')' = (X')'((X'X)^{-1})'X' = X(X'X)^{-1} = P$.

(ii) $P^2 = (X(X'X)^{-1}X')(X(X'X)^{-1}X') = X\cancel{(X'X)^{-1}}\cancel{X'X}(X'X)^{-1}X' = X(X'X)^{-1}$

$\underset{\sim}{e} = (I_n - P)\underset{\sim}{Y} \Rightarrow \underset{\sim}{e} = (I_n - P)(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}) = (I_n - P)\underset{\sim}{\varepsilon} + (I_n - P)X\underset{\sim}{\beta} \Rightarrow$

$$\underset{\sim}{e} = (I_n - P)\underset{\sim}{\varepsilon} + X\underset{\sim}{\beta} - PX\underset{\sim}{\beta} = (I_n - P)\underset{\sim}{\varepsilon} + X\underset{\sim}{\beta} - X(X'X)^{-1}\cancel{X'X}\underset{\sim}{\beta}$$

$$\underset{\sim}{e} = (I_n - P)\underset{\sim}{\varepsilon} + X\underset{\sim}{\beta} - X\underset{\sim}{\beta} \Rightarrow \underset{\sim}{e} = (I_n - P)\underset{\sim}{\varepsilon}$$

ΙΔΙΟΤΗΤΑ 2: Αν οι υποθέσεις για τα σφάλματα ικανοποιούνται ($\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$) τότε $\underset{\sim}{e} \sim N_n(\underset{\sim}{0}, (I_n - P)\sigma^2)$, όπου $P = X(X'X)^{-1}X'$

Επιπλέον $e_i \sim N(0, \sigma^2(1 - p_{ii}))$ όπου p_{ij} το (i,j)-στοιχείο του πίνακα P

ΑΠΟΔΕΙΞΗ:

Γνωρίζουμε ότι αν $\underset{\sim}{W} \sim N_n(\mu, \Sigma)$ και αν A ένας $m \times n$ πίνακας τότε:

$$\underset{\sim}{b} + \underbrace{\underset{m\times n}{A}\,\underset{n\times 1}{\underset{\sim}{W}}}_{m\times 1} \sim N_m(\underset{\sim}{b} + A\mu, A\Sigma A')$$

Επειδή τα $\underset{\sim}{e} = (I_n - P)\underset{\sim}{\varepsilon}$ είναι γραμμική συνάρτηση των $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$, τα $\underset{\sim}{e}$ έχουν n-διάστατη κανονική

Άρα, αρκεί να βρεθεί $E(\underset{\sim}{e})$ και ο $Var(\underset{\sim}{e})$.

$$E(\underset{\sim}{e}) = E[(I_n - P)\underset{\sim}{\varepsilon}] = (I_n - P)E(\underset{\sim}{\varepsilon}) = (I_n - P)\underset{\sim}{0} = \underset{\sim}{0}$$

$$Var(\underset{\sim}{e}) = Var((I_n - P)\underset{\sim}{\varepsilon}) = (I_n - P)Var(\underset{\sim}{\varepsilon})\cdot(I_n - P)' = (I_n - P)(\sigma^2 I_n)(I_n' - P') =$$
$$= \sigma^2(I_n - P)(I_n - \underset{\overset{\|}{P}}{P'}) = \sigma^2(I_n - P - P + \underset{\overset{\|}{P}}{P^2}) = \sigma^2\cdot(I_n - P).$$

Επιπλέον αφού $\underset{\sim}{e} \sim N_n(\underset{\sim}{0}, \sigma^2(I_n - P))$, κάθε $e_i \sim N(0,$ το i-διαγώνιο στοιχείο του $\sigma^2(I_n - P))$

$$\equiv N(0, \sigma^2(1 - p_{ii}))\ \forall i = 1, \dots, n$$

οπότε: $$Var(\underset{\sim}{e}) = \sigma^2 \begin{pmatrix} 1-p_{11} & -p_{12} & \cdots & -p_{1n} \\ -p_{21} & 1-p_{22} & & -p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{n1} & -p_{n2} & & 1-p_{nn} \end{pmatrix}$$

Επίσης, $Cov(e_i, e_j) = -\sigma^2 p_{ij}$ όπου p_{ij} το (i,j) στοιχείο του P.

ΙΔΙΟΤΗΤΑ 3 Τα $e_i, \hat{Y}_i$ ασυσχέτιστα ή γενικότερα $Cov(\underset{\sim}{e}, \hat{Y}) = 0$

ΑΠΟΔΕΙΞΗ: $Cov(\underset{\sim}{e}, \hat{\underset{\sim}{Y}}) = Cov((I_n - P)\underset{\sim}{Y}, X\hat{\underset{\sim}{\beta}}) = Cov((I_n-P)\underset{\sim}{Y}, X(X'X)^{-1}X'\underset{\sim}{Y}) =$

$= Cov((I_n-P)\underset{\sim}{Y}, P\underset{\sim}{Y}) \underset{=AVar(W)B'}{\overset{Cov(AW, BW)=}{=}} (I_n-P)Var(\underset{\sim}{Y})P' =$

$= (I_n-P)(\sigma^2 \cdot I_n)P = \sigma^2 (I_n - P)P = \sigma^2 (P-P^2) = \sigma^2(P-P) =$

$= \sigma^2 \cdot 0 = 0$ (μηδενικός πίνακας)

ΙΔΙΟΤΗΤΑ 4: Μαθηματικοποιημένα υπόλοιπα ή Studentized υπόλοιπα.

$t_i = \dfrac{e_i}{\hat{\sigma}\sqrt{1-p_{ii}}} = \dfrac{e_i}{\sqrt{MS_{res}}\sqrt{1-p_{ii}}}$, $i=1,\dots,n$ και p_{ii} το (i,i)-στοιχείο του P

Ιδιότητα των:

$t_i \sim t_{n-p-1}, \ \forall i=1,\dots,n$

είναι τα υπόλοιπα απαλλαγμένα από τη διακύμανση σ^2

ΑΠΟΔΕΙΞΗ:

$t_i = \dfrac{e_i}{\hat{\sigma}\sqrt{1-p_{ii}}} = \dfrac{e_i}{\sqrt{MS_{res}}\sqrt{1-p_{ii}}}$, $i=1,\dots,n$

Αποδείξαμε ότι $e_i \sim N(0, \sigma^2(1-p_{ii})) \Rightarrow$

$\Rightarrow \dfrac{e_i}{\sigma\sqrt{1-p_{ii}}} \sim N(0,1)$, $\forall i=1,\dots,n$

Γνωρίζω $\dfrac{SS_{res}}{\sigma^2} \sim \chi^2_{n-p-1}$. $\left.\begin{array}{l}\text{Επιπλέον, } \hat{\underset{\sim}{\beta}}, SS_{res} \text{ ανεξάρτητα} \\ \text{ενώ, } \underset{\sim}{e} = \underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}\end{array}\right\} \Rightarrow SS_{res}, \underset{\sim}{e}$ ανεξάρτητα

Θεωρώ: $\dfrac{\dfrac{e_i}{\sigma\sqrt{1-p_{ii}}}}{\sqrt{\left(\dfrac{SS_{res}}{\sigma^2}\right)\Big/(n-p-1)}} = \dfrac{\dfrac{e_i}{\sigma\sqrt{1-p_{ii}}}}{\sqrt{\dfrac{MS_{res}}{\sigma^2}}} = \dfrac{\dfrac{e_i}{\cancel{\sigma}\sqrt{1-p_{ii}}}}{\dfrac{\sqrt{MS_{res}}}{\cancel{\sigma}}} = \dfrac{e_i}{\sqrt{MS_{res}}\sqrt{1-p_{ii}}} \overset{op.}{=} t_i$

Άρα, $\forall i=1,\dots,n$: $t_i = \dfrac{\dfrac{e_i}{\sigma\sqrt{1-p_{ii}}}}{\sqrt{\left(\dfrac{SS_{res}}{\sigma^2}\right)\Big/(n-p-1)}} = \dfrac{N(0,1)}{\sqrt{\dfrac{\chi^2_{n-p-1}}{n-p-1}}} \sim t_{n-p-1}$

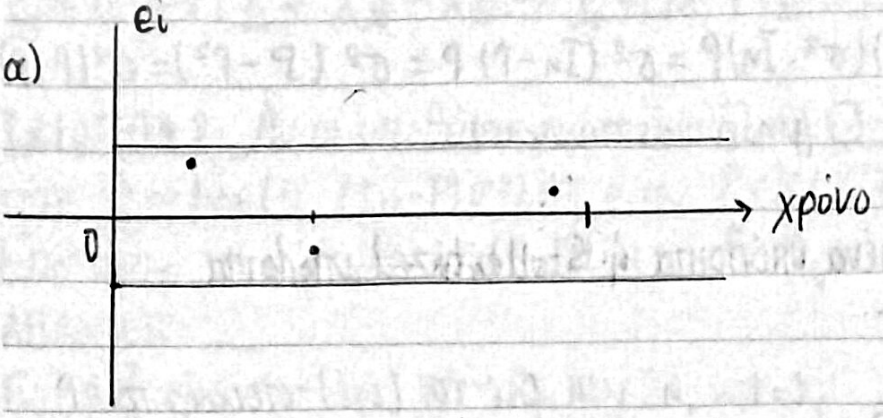
e_i ανεξ. SS_{res} γιατί τα $\hat{\underset{\sim}{\beta}}$ ανεξ. SS_{res} και

$\underset{\sim}{e} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} = \underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}$

(I) Έλεγχος Ασυσχέτιστου των σφαλμάτων

$(Cov(\varepsilon_i, \varepsilon_j) = 0 \ , i,j = 1, \dots, n \ , i \neq j)$

α)



Αν τα σημεία είναι τυχαία κατανεμημένα σε μια ζώνη γύρω από το 0 (αφού $E(e_i)=0$)
τότε έχουμε ισχυρή ένδειξη ότι τα σφάλματα είναι ασυσχέτιστα

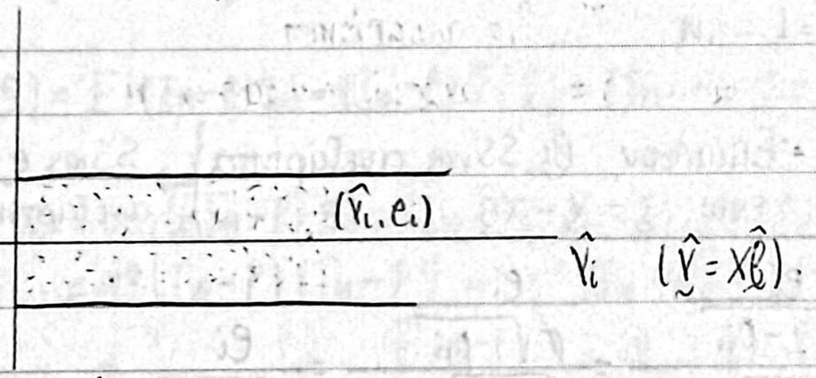
β) Μη παραμετρικό τεστ των ροών

γ) Τεστ Darbin-Watson

(II) Έλεγχος σταθερής διακύμανσης

$(Var(\varepsilon_i) = \sigma^2, i = 1, \dots, n)$

α)



Αν τα $(\hat{Y}_i, e_i)$, $i = 1, \dots, n$, κατανέμονται τυχαία σε μια ζώνη γύρω από το 0 τότε υπάρχει ισχυρή ένδειξη ότι η υπόθεση της σταθερής διακύμανσης ικανοποιείται.

Γιατί; Επειδή αποδείξαμε ότι $Cov(\underset{\sim}{e}, \underset{\sim}{\hat{Y}}) = 0$

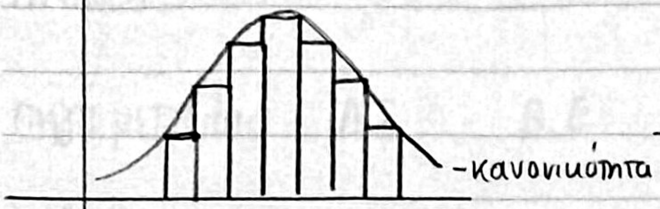
β) Τεστ Levene

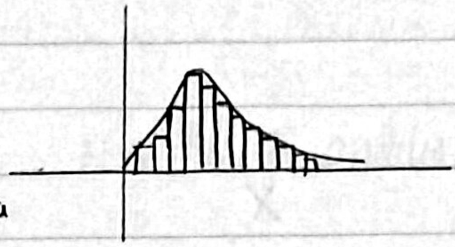
(III) Έλεγχος Κανονικότητας των σφαλμάτων

$(e_i \sim N(0,\sigma^2),\ i=1,\dots,n)$

Αρκεί να ελέγξω την κανονικότητα των $e_1,\dots,e_n$

α) Ιστόγραμμα συχνοτήτων





β) Kolmogorov-Smirnov

γ) Shapiro-Wilk.

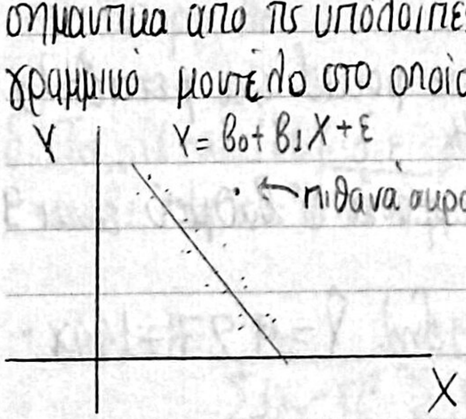
Ακραίες ή Επηρεάζουσες Παρατηρήσεις

Γενικά στη στατιστική: Ακραίες παρατηρήσεις είναι εκείνες που διαφοροποιούνται σημαντικά από τις υπόλοιπες.

Π.Χ. 2, 3, 5, 7, 4, 0, -1, 2, ..., 20 → πιθανή ακραία παρατήρηση

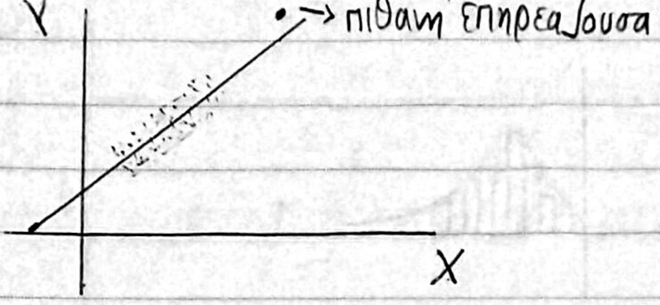
Ο μέσος όρος $\bar{X}$ επηρεάζεται σημαντικά από μια ακραία παρατήρηση, επιλέγουμε να δουλέψουμε με διάμεσο M

Ειδικότερα στα γραμμικά μοντέλα: Ακραίες παρατηρήσεις είναι εκείνες που απέχουν σημαντικά από τις υπόλοιπες. Είναι εκείνες που δεν προσαρμόζονται σε κάποιο γραμμικό μοντέλο στο οποίο προσαρμόζονται οι υπόλοιπες



Έχουν αναπτυχθεί τεχνικές για να ελέγχεται αν μια παρατήρηση είναι ή όχι ακραία. Στα γραμμικά μοντέλα, αν $|e_i|$ είναι μεγάλη, τότε η αντίστοιχη παρατήρηση είναι πιθανά ακραία

Επηρεάζουσες παρατηρήσεις: είναι γενικά παρατηρήσεις που απέχουν από τις άλλες, που μπορεί να προσαρμόζονται στο μοντέλο που προσαρμόζονται οι υπόλοιπες αλλά επηρεάζουν σημαντικά τους εκτιμητές.



Για την ανίχνευση επηρεαζουσών (έλεγχο): Τεστ που βασίζεται στην απόσταση Cook

ΑΣΚΗΣΗ 3:

Θερμοκρασία	-5	-4	1	2	5
Απόδοση	1	5	9	13	18

α.

X = Θερμοκρασία ← Ανεξάρτητη

Y = Απόδοση ← Εξαρτημένη

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

ή $Y_i = b_0 + b_1 X_i + \varepsilon_i$, $i = 1, \dots, n$

E.E.T

Μετά από πράξεις

$$\hat{b}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = 1.44$$

→ ερμηνεία ΕΕΤ.

↳ η αναμενόμενη απόδοση σε μοναδιαία μεταβολή της θερμοκρασίας είναι 1.44.

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} = 9.27$$

↳ η αναμενόμενη απόδοση για θερμοκρασία 0° βαθμοί είναι 9.27

και άρα το εκτιμώμενο μοντέλο είναι: $\hat{Y} = \hat{b}_0 + \hat{b}_1 X$ δηλ. $\hat{Y} = 9.27 + 1.44X$

εκτιμώμενου μοντέλου: Μπορώ να δίνω τιμές στο X και να προβλέπω το Y. Χρησιμοποιώ το εκτιμώμενο μοντέλο για προβλέψεις για τιμές του X κοντά στο $\bar{X}$.

Θεωρώ ότι οι υποθέσεις για τα σφάλματα ικανοποιούνται.

β. Ανάδια, αρχικό

ΑΝΑΔΙΑ

πηγή μεταβλη	Α.Τ.	Β.Ε.	M.S	F-πηλίκο
παλινδρόμηση (μοντέλο)	226.95	1	226.95	96.17
Υπόλοιπα	21.23		2.36.	
Ολική	248.18	10		

Εφόσον $SSreg >> SSres$ ή $\underbrace{MSreg >> MSres}$, Απορρ. Ho

δηλ. το μοντέλο δείχνει να 'στέκει' άρα δεν μπορώ να δεχτώ ότι το $\beta_1 = 0$

γ. Στατιστικός Test έλεγχο $H_0 : \beta_1 = 0$

t-test

F-test · το οποίο και έχω από τον πίνακα ανάδια αφού F-πηλίκο $= 96.17 > F_1, \quad = 5,12$

$\Rightarrow$ Απορρ. Ho

δ. Το $(1-a)\,100\%$ δ.ε. για την β_1

- $(\hat{\beta}_1 - t_{a,0,25}\sqrt{\hat{Var}(\hat{\beta}_1)}, \hat{\beta}_1 + t_{,0.0}\sqrt{\hat{Var}(\hat{\beta}_1)}) = (1{,}11, 1.77)$

↳ αναμενόμενη απόδοση μοναδιαίος μεταβολ. θερμοκρ.

- $\hat{Var}(\beta_1) = \dfrac{\quad}{\sum(X_i - \bar{X})^2} = \dfrac{MSres}{\sum(X_i-\bar{X})^2}$

ΑΣΚΗΣΗ 4: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , i = 1, \dots, n$

α) $r^2_{X,Y} = R^2$

β) $r^2_{X,\hat{Y}} = R^2 = r^2_{X,Y}$

ΛΥΣΗ:

α) $$R^2 = \frac{SSres}{SStot} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\sum(\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\sum(\bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} =$$

$$= \frac{\hat{\beta}_1^2 \sum(X_i - \bar{X})^2}{\sum(Y_i - \bar{Y})^2} \quad (1)$$

Αλλά: $\hat{\beta}_1 = \dfrac{\sum(X_i - \bar{X})(Y_i - \bar{X})}{\sum(X_i - \bar{X})^2}$ και $r_{X,Y} = \dfrac{\sum(X_i - \bar{Y})(Y_i - \bar{X})}{\sqrt{\sum(X_i - \bar{X})}\sqrt{\sum(Y_i - \bar{Y})}}$

δηλ. $$\hat{\beta}_1 = \left[\frac{\sum(Y_i - \bar{Y})^2}{\sum(X_i - \bar{X})^2}\right]^{1/2} \cdot r_{X,Y} \quad (2)$$

οπότε άμεσα πλέον βγαίνει το συμπέρασμα

β) • $$r_{Y,\hat{Y}} \overset{\text{op.}}{=} \frac{\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\sqrt{\sum(Y_i - \bar{Y})^2}\sqrt{\sum(\hat{Y}_i - \bar{\hat{Y}})^2}} \quad (1)$$

• $$\bar{\hat{Y}} = \frac{1}{n}\sum \hat{Y}_i = \frac{1}{n}\sum(\hat{\beta}_0 + \hat{\beta}_1 X_i) = \frac{1}{n}\hat{\beta}_0 + \hat{\beta}_1 \bar{X} = \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 \bar{X} = \bar{Y} \quad (2)$$

$$r^2_{X,\hat{Y}} = \frac{[\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})]^2}{\sum(Y_i - \bar{Y})^2 \sum(\hat{Y}_i - \bar{Y})^2} = \frac{[\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})]^2}{\sum(Y_i - \bar{Y})^2 \hat{\beta}_1^2 \sum(X_i - \bar{X})^2} \quad (3)$$

$$[\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})]^2 = [\sum(Y_i - \bar{Y})(\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{Y})]^2 = [\sum(Y_i - \bar{Y})(\quad)]^2$$
$$\overset{(4)}{=} \hat{\beta}_1^2 [\sum(Y_i - \bar{Y})(X_i - \bar{X})]^2$$

Τελικά $$r^2_{X,\hat{Y}} = \frac{[\hat{\beta}_1 [\sum(Y_i - \bar{Y})(X_i - \bar{X})]]^2}{\sum(Y_i - \bar{Y})^2 \hat{\beta}_1^2 \sum(X_i - \bar{X})^2} = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum(Y_i - \bar{Y})^2}\sqrt{\sum(X_i - \bar{X})^2}} \overset{(a)}{=} R^2$$