6ο μάθημα 8/11/21

$\underset{\sim}{y} = X \cdot \underset{\sim}{b} + \underset{\sim}{\varepsilon}$

$\hat{b} = (X' \cdot X)^{-1} \cdot X' \cdot \underset{\sim}{y}$

$\underset{\sim}{\hat{y}} = X \cdot \underset{\sim}{\hat{b}}$

$\underset{\sim}{e} = \underset{\sim}{y} - \underset{\sim}{\hat{y}}$

$SStot = SSreg + SSres$

Υποθέσεις Σφάλματα

$\underset{\sim}{e} \sim N_n(\underset{\sim}{0}, \sigma^2 \cdot I_n)$

$\underset{\sim}{y} \sim N_n(X\underset{\sim}{b}, \sigma^2 \cdot I_n)$

Θεώρημα: Υπό τις υποθέσεις για τα σφάλματα ισχύει ότι $E(MSres) = \sigma^2$

Απόδειξη: Αλγεβρική

Συντελεστής προσδιορισμού ή Προσαρμοστικότητας

$$R^2 = \frac{SSreg}{SS_{tot}} = 1 - \frac{SSres}{SStot}$$

Όπως στην α.γ.π. R^2 = καθαρός αριθμός, $0 \leq R^2 \leq 1$

Τύπος που χρησιμοποιείται στο Spss:

R^2 - βελτιωμένος, $R^2 = \dfrac{MSreg}{SStot(\tau\epsilon\lambda)} = \dfrac{SSreg/p}{SStot/(n-1)}$

▶ Περαιτέρω ιδιότητες με την υπόθεση της κανονικότητας των σφαλμάτων, δηλ. $\underset{\sim}{e} \sim Normal_n$.

Ιδιότητα 1: Υπό την κανονικότητα των σφαλμάτων.

→ $\hat{\underset{\sim}{b}} \sim N_{p+1}\left(\underset{\sim}{b}, \sigma^2 \cdot (X' \cdot X)^{-1}\right)$

Αποδείξαμε στο πρ. μάθημα: $E(\hat{\underset{\sim}{b}}) = \underset{\sim}{b}$, $Var(\hat{\underset{\sim}{b}}) = \sigma^2 \cdot (X' \cdot X)^{-1}$.

Ισχύει ότι: Αν $\underset{\sim}{W} \sim N_n(\mu, \Sigma)$ και αν A είναι πίνακας κατάλληλης διάστασης τότε: $\underset{p \times n}{A} \cdot \underset{n \times 1}{\underset{\sim}{W}} \sim N(A \cdot \underset{\sim}{\mu}, A \cdot \Sigma \cdot A')$

δηλ. ισχύουν οι ιδιότητες πινάκων

στήλες = διάσταση

Είναι $\hat{\underset{\sim}{b}} = (X' \cdot X)^{-1} \cdot X' \cdot \underset{\sim}{Y}$

Αν εφαρμόσω για $A = (X' \cdot X)^{-1} \cdot X' \cdot \underset{\sim}{Y}$, τότε $\hat{\underset{\sim}{b}} \sim Normal$ ως γρ. μετασχ. της $\underset{\sim}{Y}$.

Ιδιότητα 2: Οι Ε.Ε.Τ. $\hat{\underset{\sim}{b}}$ είναι ΑΟΕΔ εκτιμητές.

Πιο φορμαλιστικά: Θεώρημα Gauss-Markov (*)

- Οι Ε.Ε.Τ. έχουν τη μικρότερη διακύμανση μεταξύ όλων των αμερόληπτων εκτιμητών των παραμέτρων $\underset{\sim}{b}$ που είναι γραμμικές συναρτήσεις του $\underset{\sim}{Y}$.

Ιδιότητα 3: Αν $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 \cdot I_n)$, τότε:

a) Οι ΕΕΤ $\hat{\underset{\sim}{b}}$ συμπίπτουν με τους ΕΜΠ

b) Ο ΕΜΠ της σ^2 είναι ο $\hat{\sigma}^2 = \dfrac{SS_{res}}{n}$

<u>m-διάστατη κανονική</u> $N_m(\underset{\sim}{\mu}, \Sigma)$

Έστω τ.δ. $\underset{\sim}{w} \sim \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}$. Το $\underset{\sim}{w} \sim N_{\underset{\sim}{w}}(\underset{\sim}{\mu}, \Sigma)$ αν η σ.π.π. είναι:

$$f_{\underset{\sim}{w}}(\underset{\sim}{w}) = \frac{1}{(2\pi)^{m/2} \cdot |\Sigma|^{1/2}} \cdot e^{-\frac{1}{2}(\underset{\sim}{w} - \underset{\sim}{\mu})' \Sigma^{-1} \cdot (\underset{\sim}{w} - \underset{\sim}{\mu})}$$

με $\underset{\sim}{w} \in \mathbb{R}^m$, $\Sigma > 0$ (θετικά ορισμένος πίνακας).

α) Οι Ε.Ε.Τ. προκύπτουν από ελαχιστοποίηση ως προς $\underset{\sim}{\beta}$ της $S(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \underset{\sim}{\varepsilon}' \cdot \underset{\sim}{\varepsilon} = (\underset{\sim}{y} - X\underset{\sim}{\beta})' \cdot (\underset{\sim}{y} - X\underset{\sim}{\beta})$

<u>Πιθανοφάνεια</u>: Η από κοινού κατανομή των δεδομένων στην y, δηλ. η από κοινού κατανομή του τ.δ. $\underset{\sim}{y}$. Αφού $\underset{\sim}{y} \sim N_n(X\underset{\sim}{\beta}, \sigma^2 \cdot I_n)$, η πιθανοφάνεια είναι:

$$L(\underset{\sim}{\beta}) = \frac{1}{(2\pi)^{n/2} |\sigma^2 \cdot I_n|^{1/2}} \cdot e^{-\frac{1}{2}(\underset{\sim}{y} - X\underset{\sim}{\beta})' \cdot (\sigma^2 I_n)^{-1} \cdot (\underset{\sim}{y} - X \cdot \underset{\sim}{\beta})}$$

Αλλά $(\sigma^2 \cdot I_n)^{-1} = \frac{1}{\sigma^2} \cdot I_n$

ορίζουσα $\rightarrow |aA| = a^n \cdot |A|$ $\Big\} \Rightarrow$

$$\boxed{L(\underset{\sim}{\beta}) = \frac{1}{(2\pi)^{n/2} \cdot (\sigma^2)^{n/2} \cdot \underbrace{\cancel{|I_n|}}_{1}} \cdot e^{-\frac{1}{2\sigma^2} \cdot (\underset{\sim}{y} - X\underset{\sim}{\beta})' \cdot (\underset{\sim}{y} - X\underset{\sim}{\beta})}}$$

$$\log L(\underset{\sim}{\beta}) = -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log \sigma^2 - \frac{1}{2\sigma^2} (\underset{\sim}{y} - X\underset{\sim}{\beta})' \cdot (\underset{\sim}{y} - X\underset{\sim}{\beta})$$

Οι ΕΜΠ της $\underset{\sim}{\beta}$ προκύπτουν από μεγιστοποίηση ως προς

$\underset{\sim}{b}$ της $\log L(\underset{\sim}{b})$ ή μεγιστοποίηση ως προς $\underset{\sim}{b}$ της $-(\underset{\sim}{y}-x\underset{\sim}{b})'.(\underset{\sim}{y}-x.\underset{\sim}{b})$

ή ελαχιστοποίηση ως προς b της $(\underset{\sim}{y}-x\underset{\sim}{b})'.(\underset{\sim}{y}-x\underset{\sim}{b})$

Επομένως οι ΕΕΤ και οι ΕΜΠ προκύπτουν από την ελαχιστοποίηση της ίδιας ποσότητας, άρα συμπίπτουν.

β) Ο ΕΜΠ της σ^2 θα προκύψει από μεγιστοποίηση της πιθανοφάνειας:

$$\log L(\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}.\log\sigma^2 - \frac{1}{2\sigma^2}(\underset{\sim}{y}-x\hat{\underset{\sim}{b}})'.(\underset{\sim}{y}-x\hat{\underset{\sim}{b}})$$

$$\frac{d}{d\sigma^2}\log L(\sigma^2) = 0 \Rightarrow \dots \Rightarrow \frac{1}{2\sigma^2}(\underset{\sim}{y}-x\hat{\underset{\sim}{b}})'.(\underset{\sim}{y}-x\hat{\underset{\sim}{b}}) = n$$

$$\Rightarrow \sigma^2 = \frac{(\underset{\sim}{y}-x\hat{\underset{\sim}{b}})'.(\underset{\sim}{y}-x.\hat{\underset{\sim}{b}})}{n}$$

Άρα ο ΕΜΠ της σ^2 είναι: $\hat{\sigma}^2 = \dfrac{(\underset{\sim}{y}-x\hat{\underset{\sim}{b}})'.(\underset{\sim}{y}-x\underset{\sim}{b})}{n} \Rightarrow$

$$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{SSres}{n}}$$

Είναι αμερόληπτος?

Παρατήρηση:

$$E(\hat{\sigma}^2) = E\left(\frac{SSres}{n}\right) = \frac{1}{n}E(SSres)$$

Αλλά $E(MSres) = \sigma^2 \Rightarrow E\left(\frac{SSres}{n-p-1}\right) = \sigma^2 \Rightarrow$

$E(SSres) = (n-p-1)\cdot\sigma^2$ $\Bigg\} \Rightarrow$

$$\boxed{E(\hat{\sigma}^2) = \frac{n-p-1}{n}\cdot\sigma^2}$$

Άρα, δεν είναι αμερόληπτος της σ^2.

<u>Ιδιότητα 4:</u> Υπό τις υποθέσεις για τα σφάλματα,
$\frac{SSres}{\sigma^2} \sim \chi^2_{n-p-1}$ (δεν αποδεικνύεται προς το παρόν) λόγω ελλειπών γνώσεων

<u>Ιδιότητα 5:</u> $\hat{\beta}$ και MSres είναι ανεξάρτητα.

<u>Ιδιότητα 6:</u> Τεστ Ελέγχου Ύπαρξης Παλινδρόμησης.

Για τον έλεγχο της μηδενικής υπόθεσης ότι $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$, η στατιστική συνάρτηση του τεστ είναι:

τελευταία στήλη πίνακα ΑναΔια

$$F = \frac{MSreg}{MSres} \sim F_{p,n-p-1}$$, υπό την H_0 και κ.π. $F \geq F_{p,n-p-1,\alpha}$

<u>(αν δεν μπορώ να την απορρίψω, τότε $\exists$ τουλ. ένα $\beta_i \neq 0$)</u>.

Αν η $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ απορριφθεί, τότε $\exists$ 1 τουλ. $\beta_i \neq 0$. Επομένως, έχει αξία να ενδιαφερθούμε για ελέγχους υποθέσεων $H_0: \beta_i = \beta_i^*$, με β_i^* γνωστό για $i = 1, 2, \dots, p$. Το γιατί έχει να κάνει με την ερμηνεία των $\hat{\beta}_i$, $i = 1, 2, \dots, p$.

<u>Ερμηνεία του b_i</u>: εκφράζει τη μεταβολή της Y σε μοναδιαία μεταβολή της X_i θεωρώντας των υπολοίπων X_i ως σταθερές.

<u>Ιδιότητα 7</u>: Για τον έλεγχο της $H_0: b_i = b_i^*$, $i=1,\dots,p$ η ΣΣΤ είναι η $t = \dfrac{b_i - b_i^*}{\sqrt{MS_{res}}\sqrt{C_{i+1,i+1}}}$ με κατανομή t_{n-p-1} υπό την H_0 και κ.π. $|t| \geq t_{n-p-1,\alpha/2}$ όπου $C_{i+1,i+1}$ το $(i+1,i+1)$-στοιχείο του πίνακα $(X'.X)^{-1}$.

<u>Απόδειξη</u>: Πρέπει να στηριχτώ σε εκτιμητή του b_i

δηλ. του ΕΕΤ $\hat{b}_i$, $i=1,\dots,p$

Γνωρίζω: $\hat{\underset{\sim}{b}} = (X'.X)^{-1}.X'.\underset{\sim}{Y} \sim N_{p+1}(\underset{\sim}{b}, \sigma^2.(X'.X)^{-1})$

Επίσης, γνωρίζω ότι αν $\underset{\sim}{w} \sim \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \sim N_n(\mu, \Sigma)$, τότε

η τ.μ. $W_i \sim N(\mu_i, \sigma_i^2)$, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_n \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \dots & C_{ov} \\ \vdots & \sigma_i^2 & \vdots \\ C_{ov} & \dots & \sigma_n \end{pmatrix}$ $\Rightarrow$

$\Rightarrow$ Άρα $\hat{b}_i \sim N(b_i, \sigma^2.C_{i+1,i+1})$, $i=1,\dots,p$ $\Rightarrow$

$\dfrac{\hat{b}_i - b_i}{\sigma\sqrt{C_{i+1,i+1}}} \sim N(0,1) \Rightarrow \boxed{\dfrac{\hat{b}_i - b_i^*}{\sigma\sqrt{C_{i+1,i+1}}} \sim N(0,1) \text{ υπό } H_0: b_i = b_i^*}$

Αλλά $\dfrac{SS_{res}}{\sigma^2} \sim X^2_{n-p-1}$ και είναι ανεξάρτητο του $\hat{b}_i$, επειδή $\hat{\underset{\sim}{b}}$ και MS_{res} ανεξ.

Θεωρώ $t = \dfrac{\hat{b}_i - b_i^*}{\sqrt{MS_{res}}\sqrt{C_{i+1,i+1}}} = \dfrac{(\hat{b}_i - b_i^*)/\sigma\sqrt{C_{i+1,i+1}}}{\sqrt{\dfrac{MS_{res}}{\sigma^2}}} =$

$$= \frac{(\hat{b}_i - b_i^*)/\sigma\sqrt{C_{i+1,i+1}}}{\sqrt{\dfrac{SS_{res}}{(n-p-1)\sigma^2}}} \overset{d}{=} \frac{N(0,1)}{\sqrt{X^2_{n-p-1}/(n-p-1)}} \sim t_{n-p-1} \text{ υπό } H_0$$

$\boxed{\text{Άσκηση 3}}$ φυλλάδιο

θερμ.	-5	-4	-3 ...	4	5
αποδοση	1	5	4 ...	13	18

X = θερμοκρασία (αυτήν που ελέγχει)
Y = απόδοση (εξαρτημένη μεταβλητή)

α) $\hat{b}_1 = \dfrac{\Sigma(X_i - \bar{X})\cdot(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = 1{,}44$

$\hat{b}_0 = \bar{Y} - \hat{b}_1 \cdot \bar{X} = 9{,}27$

Ερμηνεία $\hat{b}_1$
Σε μοναδιαία μεταβολή της θερ. (αύξηση) περιμένω αύξηση της απόδοσης κατά 1,44 μονάδες

Εκτιμώμενο Μοντέλο : $\boxed{\hat{Y} = 9{,}27 + 1{,}44X}$

Προσοχή! Για θερμοκρασίες από -5 έως 5

β) $H_0: b_1 = 0$

ANADIA

	SS	β.ε	MS	F-πηλίκο
Παλινδ.	226,95	1	226,95	96,17
Υπόλοιπα	21,23	9	2,36	
Ολική	248.18	10		

γ) $H_0: \beta_1 = 0$, με ε.σ. 5%.

i) $F = \dfrac{MS_{reg}}{MS_{res}} = 96.17 >> 5.12 = F_{1,9,0.05}$

Αρα, απορρίπτω την H_0.

ii) δ.ε. $\hat{\beta}_1 \pm t_{9,0.025} \sqrt{\hat{Var}(\hat{\beta}_1)} \rightsquigarrow (1.11, 1.77)$