2º μάθημα 11/10/21

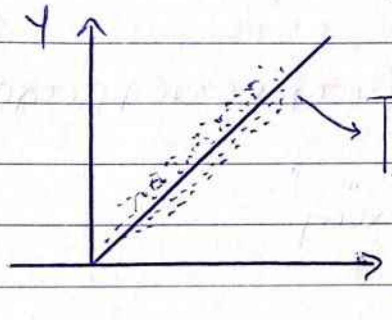
Α.Γ.Π.: $Y_i = b_0 + b_1 \cdot X_i + \varepsilon_i$, $i=1,\dots,n$

Ε.Ε.Τ. (εκτιμητές ελαχίστων τετραγώνων).

$\hat{b_1} = \dots$ → η μεταβολή της Y σε μοναδιαία μεταβολή της X.

$\hat{b_0} = \bar{Y} - \hat{b_1} \cdot \bar{X}$, για $x=0$.

Εκτιμώμενο Μοντέλο ← Χρησιμοποιείται για να προβλεφθεί η Y για συγκεκριμένη τιμή της X.



$\boxed{\hat{Y} = \hat{b_0} + \hat{b_1} \cdot x}$

(X_i, Y_i), $i=1,\dots,n$

$\hat{Y} = \hat{b_0} + \hat{b_1} \cdot x$, $\hat{Y_i} = \hat{b_0} + \hat{b_1} X_i$, $i=1,\dots,n$

Υπόλοιπα

Ορισμός: $e_i = Y_i - \hat{Y_i}$, $i=1,\dots,n$

Ερμηνεία: Εκφράζουν την απόκλιση του μοντέλου $(\hat{Y_i})$ από την πραγματικότητα (Y_i)

Κριτήριο: Αν e_i, $i=1,\dots,n$ είναι μικρά, τότε → → το μοντέλο είναι ικανοποιητικό.

<u>Ιδιότητα Υπολοίπων</u>: Στο μοντέλο της α.γ.π. $\sum_{i=1}^{n} e_i = 0$

<u>Απόδειξη</u>: $\sum_{i=1}^{n} e_i = \Sigma(Y_i - \hat{Y}_i) = \Sigma(Y_i - \hat{b_0} - \hat{b_1}X_i) =$
$= \Sigma(Y_i - \bar{Y} + \hat{b_1}\cdot\bar{X} - \hat{b_1}\cdot X_i) =$
$= \Sigma Y_i - n\bar{Y} + \hat{b_1}(\Sigma\bar{X} - \Sigma X_i) =$
$= n\bar{Y} - n\bar{Y} + \hat{b_1}\cdot(n\bar{X} - n\bar{X}) = 0$

<u>Ανάλυση της Διακύμανσης στο μοντέλο της α.γ.π.</u>:

(<u>Διακύμανσης - Μεταβλητότητας</u>: για τ.δ. $W_1, \dots, W_n$
$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (W_i - \bar{W})^2$)

Μεταβλητότητα → Δειγματική $Y_1, \dots, Y_n$ → $\cancel{\frac{1}{n-1}} \cdot \Sigma(Y_i - \bar{Y})^2$ →

→ $\boxed{\Sigma(Y_i - \bar{Y})^2}$ εκφράζει τη μεταβλητότητα των $Y_1, \dots, Y_n$

των δεδομένων Διακύμανση

• Πως διαφοροποιείται η $\Sigma(Y_i - \bar{Y})^2$ αν λάβω υπόψιν το μοντέλο της α.γ.π. ;

<u>Ολική Μεταβλ.</u> $= \underbrace{\sum_{1}^{n} (Y_i - \bar{Y})^2}_{SS_{tot}} \overset{\pm \hat{Y}_i}{=} \underbrace{\sum_{1}^{n} (\hat{Y}_i - \bar{Y})^2}_{SS_{reg}} + \underbrace{\sum_{1}^{n} (Y_i - \hat{Y}_i)^2}_{SS_{res}}$

$SS_{tot} = SS_{reg} + SS_{res}$

[Αποδ: Άσκηση]

Πίνακας Ανάλυσης για το μοντέλο της α.γ.π.:

Πηγή Μεταβλητότητας	SS	β.ε.	MS	F-πηλίκο
Παλινδρόμηση	SS_{reg}	1	$MS_{reg} = \frac{SS_{reg}}{1 = β.ε}$	$F = \frac{MS_{reg}}{MS_{res}}$
Υπόλοιπα	SS_{res}	$n-2$	$MS_{res} = \frac{SS_{res}}{2 = β.ε}$	
Ολική	SS_{tot}	$n-1$		

Γενικά: β.ε. ενός αθροίσματος τετραγώνων είναι το πλήθος των ανεξάρτητων πηγών πληροφορίας που απαιτούνται προκειμένου να υπολογιστεί το αντίστοιχο άθροισμα τετραγώνων.

• β.ε. του SS_{tot}:

$$SS_{tot} = \Sigma(Y_i - \bar{Y})^2$$

$$\bar{Y} = \frac{1}{n}\cdot\Sigma Y_i \Rightarrow Y_n + \sum_{1}^{n-1} = n\cdot\bar{Y}$$

$$Y_n - \bar{Y} = n\cdot\bar{Y} - \sum_{2}^{n-1} Y_i - \bar{Y} = (n-1)\cdot\bar{Y} - \sum_{2}^{n-1} Y_i = \sum^{n-1}\bar{Y} - \sum^{n-1} Y_i =$$

$$= -\Sigma(Y_i - \bar{Y}) \Rightarrow (Y_n - \bar{Y})^2 = \left[\sum_{2}^{n-1}(Y_i - \bar{Y})\right]^2$$

$$SS_{tot} = \sum^{n}(Y_i - \bar{Y})^2 = \sum^{n-1}(Y_i - \bar{Y})^2 = \sum^{n-1}(Y_i - \bar{Y})^2 + (Y_n - \bar{Y})^2$$

$$= \sum^{n-1}(Y_i - \bar{Y})^2 + \left[\Sigma(Y_i - \bar{Y})\right]^2.$$

Εμπειρικά: β.ε. SStot = πλήθος παρατηρήσεων − 1 = n−1
β.ε. SSμοντέλο = πλήθος ανεξ/των μετ/των.
β.ε. SSυπολοίπων = β.ε. SStot − β.ε μοντέλου.

Συντελεστής Προσδιορισμού ή Προσαρμοστικότητας

Ορισμός: $R^2 = \frac{SSreg}{SStot} = 1 - \frac{SSres}{SStot}$

Ιδιότητες:

1) R^2: καθαρός αριθμός (απαλλαγή από μον. μέτρησης)
2) $0 \leq R^2 \leq 1$.

• Αν ο R^2 παίρνει μικρές τιμές, τότε $SSreg \ll SStot$ ή $SSreg < SSres$

άτυχο μοντέλο.

• Αν ο R^2 παίρνει μεγάλες τιμές (π.χ. $\geq 0,7$), τότε $SSreg \gg SSres$, δηλ. έχω υποσχόμενο μοντέλο

Ερμηνεία του R^2: Εκφράζει το ποσοστό της μεταβλητότητας της Y που ερμηνεύεται από την X.

Υποθέσεις για τα σφάλματα:

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

σημαντικές αποδείξεις

Υποθ.1: Τα σφάλματα θεωρούνται τ.μ. $E(\varepsilon_i) = 0$
Συνέπεια στα Y_i, $E(Y_i) = b_0 + b_1 X_i$

Υποθ.2: $Var(\varepsilon_i) = \sigma^2 > 0$. Συνέπεια στα Y_i: $Var(Y_i) = \sigma^2$
με $i = 1, \dots, n$

Υποθ.3: Τα σφάλματα είναι ασυσχέτιστα μεταξύ τους, δηλ. $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$. Συνέπεια στα Y_i: Τα Y_i ασυσχέτιστα μεταξύ τους, δηλ. $Cov(Y_i, Y_j) = 0, \ \forall i \neq j$ (ως ασυσχέτιστες συνδ.)

Υποθ.4: Τα $\boxed{\varepsilon_i \sim N(0, \sigma^2)}$ Συνέπεια στα Y_i:
$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$, με $i = 1, \dots, n$.

Θεώρημα: (Κανονικότητα των ΕΕΤ)

Αν οι υποθέσεις για τα σφάλματα ικανοποιούνται, τότε:

α) $\hat{\beta}_1 \sim N\left(\beta_1, \ Var(\hat{\beta}_1) = \dfrac{\sigma^2}{\Sigma (X_i - \bar{X})^2}\right)$

β) $\hat{\beta}_0 \sim N\left(\beta_0, \ Var(\hat{\beta}_0) = \dfrac{\Sigma X_i^2}{\Sigma (X_i - \bar{X})^2} \cdot \sigma^2 = \sigma^2 \left[\dfrac{1}{n} + \dfrac{\bar{X}}{\Sigma (X_i - \bar{X})^2}\right]\right)$

Απόδειξη: α) $\hat{\beta}_1 = \dfrac{\Sigma (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\Sigma (X_i - \bar{X})^2} = \Sigma \dfrac{X_i - \bar{X}}{\Sigma (X_i - \bar{X})} \cdot Y_i$

Το $\hat{\beta}_1$ είναι γρ. συνδυασμός των Y_i
Άρα $\hat{\beta}_1 \sim$ Normal

$$E(\hat{\beta}_1) = \Sigma \frac{(X_i - \bar{X})}{\Sigma (X_i - \bar{X})^2} \cdot E(Y_i) = \Sigma \frac{(X_i - \bar{X})}{\Sigma (X_i - \bar{X})^2} \cdot (\beta_0 + \beta_1 X_i)$$

$$\Rightarrow E(\hat{\beta}_1) = \beta_0 \cdot \Sigma \frac{X_i - \bar{X}}{\Sigma (X_i - \bar{X})^2} + \beta_1 \cdot \frac{\Sigma (X_i - \bar{X}) \cdot X_i}{\Sigma (X_i - \bar{X})^2} =$$

$$= \beta_0 \cdot \frac{\cancelto{0}{\Sigma (X_i - \bar{X})}}{\Sigma (X_i - \bar{X})^2} + \beta_1 \cdot \frac{\Sigma X_i^2 - \bar{X} \cdot \Sigma X_i}{\Sigma (X_i - \bar{X})^2} =$$

$$= \beta_1 \cdot \frac{\Sigma X_i^2 - n \cdot \bar{X}^2}{\Sigma X_i^2 - 2\bar{X} \Sigma X_i + n \cdot \bar{X}^2} = \beta_1 \cdot \cancelto{1}{\frac{\Sigma X_i^2 - n \cdot \bar{X}^2}{\Sigma X_i^2 - 2n \cdot \bar{X}^2 + n \cdot \bar{X}^2}} = \beta_1$$

Άρα ο $\hat{\beta}_1$ αμερόληπτος του β_1.

$$Var(\Sigma \alpha_i W_i) = \Sigma \alpha_i^2 \cdot Var(W_i) + \sum_i \sum_j a_i a_j Cov(W_i, W_j)$$

Αν W_i, W_j ασυσχέτιστα, τότε ισχύει $Cov(W_i, W_j) = 0$.

$$Var(\Sigma \alpha_i W_i) = \Sigma \alpha_i^2 \cdot Var(W_i)$$

$$Var(\hat{\beta}_1) = Var\left(\Sigma \frac{X_i - \bar{X}}{\Sigma (X_i - \bar{X})^2} \cdot Y_i \right) =$$

$$= \Sigma \left(\frac{X_i - \bar{X}}{\Sigma (X_i - \bar{X})^2} \right)^2 \cdot Var(Y_i) = \frac{\sigma^2}{\Sigma (X_i - \bar{X})^2}$$