13° online μάθημα

7/5/2020

ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ - ΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Πόσο καλή είναι η ευθεία ελαχίστων τετραγώνων $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$
ως εκτίμηση της ευθείας

$$E(Y|X) = \beta_0 + \beta_1 X$$

Από την προφανή σχέση

$$\underbrace{Y_i - \bar{Y}}_{\text{απόκλιση των } Y_i \text{ από το } \bar{Y}} = Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y}$$

μπορούμε να δείξουμε ότι

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

Όντως ισχύει ότι:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)\hat{Y}_i - \bar{Y}\sum_{i=1}^{n} (Y_i - \hat{Y}_i)$$

Όμως $\sum_{i=1}^{n} (Y_i - \hat{Y}_i) = \sum_{i=1}^{n} e_i = 0$ (Βλέπε ιδιότητες)

Επίσης $\sum_{i=1}^{n} (Y_i - \hat{Y}_i)(\hat{\beta}_0 + \hat{\beta}_1 X_i) = \hat{\beta}_0 \sum_{i=1}^{n} (Y_i - \hat{Y}_i) + \hat{\beta}_1 \sum_{i=1}^{n} x_i (Y_i - \hat{Y}_i) =$

$= 0$ λόγω κανονικών εξισώσεων.

Επομένως το $\sum_{i=1}^{n} (Y_i - \bar{Y})^2$ που είναι το άθροισμα τετραγώνων των παρατηρήσεων Y_i από τη μέση τιμή και λέγεται ΟΛΙΚΟ ΑΘΡΟΙΣΜΑ ΤΕΤΡΑΓΩΝΩΝ SS_{tot}
ισούται με το άθροισμα

$$\underbrace{\sum_{i=1}^{n} (Y_i + \hat{Y}_i)^2}_{\text{άθροισμα τετραγώνων υπολοίπων}} + \underbrace{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}_{\text{άθροισμα τετραγώνων προβλεπόμενων τιμών από τη μέση τιμή τους}}$$

Υπενθύμιση:
$\sum Y_i = \sum \hat{Y}_i$

Άρα $SS_{tot} = SS_{res} + SS_{reg}$

$\underset{\text{ΟΛΙΚΗ ΜΕΤ/ΤΑ}}{SS_{tot}} = \underset{\text{ΥΠΟΛΟΙΠΑ}}{SS_{res}} + \underset{\text{ΠΑΛ/ΣΗ}}{SS_{reg}}$

Όσο μεγαλύτερο το SS_{reg} τόσο μεγαλύτερη η σχέση εξάρτησης.

Ένας τρόπος αξιολόγησης της παλινδρόμησης αποτελεί ο συντελεστής προσδιορισμού

$$R^2 = \frac{SS_{reg}}{SS_{tot}}$$

(Προφανώς $0 \leq R^2 \leq 1$)

Δίνει το ποσοστό της μεταβλητότητας της Y που εξηγείται από την παλινδρόμηση. Ισχύει ότι

$$r_{xy}^2 = R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Θυμηθείτε ότι

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Επομένως

$$r^2 = \frac{\left(\sum(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}$$

Είναι $\sum(y_i - \bar{y})^2 = SS_{tot}$

$$\hat{b}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \Rightarrow \sum(x_i - \bar{x})(y_i - \bar{y}) = \hat{b}_1 \sum(x_i - \bar{x})^2$$

Άρα $r^2 = \dfrac{\hat{b}_1^2 \sum(x_i - \bar{x})^2}{SS_{tot}} = \dfrac{SS_{reg}}{SS_{tot}}$

Όμως $SS_{reg} = \sum_{i=1}^{n} (\hat{b}_0 + \hat{b}_1 x_i - \bar{y})^2 \overset{\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x}}{=\!=\!=} \sum_{i=1}^{n} (\bar{y} - \hat{b}_1\bar{x} + \hat{b}_1 x_i - \bar{y})^2$

$$= \hat{b}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Σε κάθε άθροισμα τετραγώνων αντιστοιχούν ορισμένοι βαθμοί ελευθερίας που ισοδυναμούν με το πλήθος των ανεξάρτητων συναρτήσεων των Y_i που απαιτούνται για τον υπολογισμό του εν λόγω αθροίσματος.

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

ΕΔΩ ΑΠΑΙΤΟΥΝΤΑΙ $n-1$ το πλήθος γιατί ισχύει ότι $\sum (Y_i - \bar{Y}) = 0$
Άρα ο όρος $Y_n - \bar{Y} = - \sum_{i=1}^{n} (Y_i - \bar{Y})$ άρα στο SS_{tot} χρειάζονται

$n-1$ συναρτήσεις Y_i.

$$SS_{reg} = \sum (\hat{Y}_i - \bar{Y})^2 = \hat{b}_1^2 \sum (x_i - \bar{x})^2$$

άρα έχει 1 β.ε. αφού μπορεί να υπολογιστεί από μια συνάρτηση των Y_i, την $\hat{b}_1$

SS_{reg} προφανώς έχει $n-2$ β.ε. καθώς

$$SS_{reg} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

και πληροί τις κανονικές εξισώσεις

$$\sum (Y_i - \hat{Y}_i) = 0$$

$$\sum x_i (Y_i - \hat{Y}_i) = 0$$

Τα παραπάνω ακολουθούν στον ακόλουθο πίνακα ΑΝΑΔΙΑ

Προέλευση ΜΕΤΑΒΛ/ΤΑΣ	Άθροισμα τετρ.	β.ε.	Μέσα τετράγωνα
ΠΑΛΙΝΔΡΟΜΗΣΗ	SS_{reg}	1	$MS_{reg} = SS_{reg} / 1$
ΥΠΟΛΟΙΠΑ	SS_{rer}	$n-2$	$MS_{rer} = SS_{rer} / n-2$
ΟΛΙΚΗ	SS_{tot}	$n-1$	

Αποδεικνύεται (Μάθημα 7ου εξαμήνου) SSreg και SSrer ανεξάρτητες

όταν $H_0: b_1 = 0$ αληθής

τότε $SSreg/\sigma^2 \sim \chi^2_1$, $SSrer/\sigma^2 \sim \chi^2_{n-2}$

δηλαδή $F = \dfrac{MSreg}{MSrer} \overset{H_0}{\sim} F_{1,n-2}$

Άρα ελέγχω την $H_0: b_1 = 0$ με περιοχή απόρριψης

$F \geq F_{a,1,n-2}$

Αποδεικνύεται ότι $E(MSrer) = \sigma^2$

$E(MSreg) = \sigma^2 + b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$

- ΣΧΟΛΙΑΣΜΟΣ ΓΙΑ ΚΑΤΑΛΛΗΛΟΤΗΤΑ ΜΟΝΤΕΛΟΥ
- ΑΝΑΛΥΣΗ ΥΠΟΛΟΙΠΩΝ
- ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ "ΓΡΑΜΜΙΚΟΠΟΙΗΣΗΣ"
- ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ ΚΑΝΟΝΙΚΟΤΗΤΑΣ
- ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ
- ΠΟΛΛΑΠΛΗ ΓΡ. ΠΑΛΙΝΔΡΟΜΗΣΗ