

6° Φροντιστήριο

► Για το ανθό γραμμικό μοντέλο $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i=1,2,\dots,n$ να υπολογιστεί η διακύμανση για τον ευτιμητή ελαχίστων τετραγώνων $\hat{\beta}_0 (\text{Var}(\hat{\beta}_0))$.

► Γνωρίζουμε ότι με τη μέθοδο ελαχίστων τετραγώνων προκύπτουν οι ευτιμητές για τα β_0 και β_1 που είναι:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1) \text{ και } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2)$$

$$\text{όπου } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{Από (2)} \Rightarrow \hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} - \bar{x} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \hat{\beta}_1$$

$$\Rightarrow \hat{\beta}_0 = \sum_{i=1}^n \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \cdot y_i$$

Άρα, για τη διακύμανση του $\hat{\beta}_0$, θα έχουμε:

9/

$$\text{Var}(\hat{\beta}_0) = \sum_{i=1}^n \left(\frac{1}{n} - \frac{\bar{X}(x_i - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2} \right)^2 \sigma^2 \rightarrow \left[\begin{array}{l} \text{Var}(Y_i) = \sigma^2 \text{ var} \\ \text{Av } W = \sum a_i W_i \Rightarrow \\ \text{Var } W = \sum a_i^2 \text{Var}(W_i) \end{array} \right]$$

$$= \sigma^2 \sum_{i=1}^n \left(\frac{1}{n^2} + \frac{\bar{X}^2 (x_i - \bar{X})^2}{\left(\sum_{i=1}^n (x_i - \bar{X})^2 \right)^2} - \frac{2}{n} \cdot \frac{\bar{X} (x_i - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2} \right) =$$

$$= \sigma^2 \left(\frac{n-1}{n^2} + \frac{\bar{X}^2 \sum_{i=1}^n (x_i - \bar{X})^2}{\left(\sum_{i=1}^n (x_i - \bar{X})^2 \right)^2} - \frac{2n\bar{X} \cdot \sum_{i=1}^n (x_i - \bar{X})}{n \sum_{i=1}^n (x_i - \bar{X})^2} \right) =$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \right) = \sigma^2 \left(\frac{\sum_{i=1}^n (x_i - \bar{X})^2 + n \cdot \bar{X}^2}{n \sum_{i=1}^n (x_i - \bar{X})^2} \right) =$$

$$= \sigma^2 \left(\frac{\sum_{i=1}^n (x_i^2 + \bar{X}^2 - 2\bar{X}x_i) + n\bar{X}^2}{n \sum_{i=1}^n (x_i - \bar{X})^2} \right) = \sigma^2 \left(\frac{\sum_{i=1}^n x_i^2 + n\bar{X}^2 - 2\bar{X} \sum_{i=1}^n x_i + n\bar{X}^2}{n \sum_{i=1}^n (x_i - \bar{X})^2} \right) =$$

$$= \sigma^2 \left(\frac{\sum_{i=1}^n x_i^2 + n\bar{X}^2 - 2\bar{X} \cdot n\bar{X} + n\bar{X}^2}{n \sum_{i=1}^n (x_i - \bar{X})^2} \right) =$$

$$= \sigma^2 \frac{\sum_{i=1}^n x_i^2 + 2n\bar{X}^2 - 2n\bar{X}^2}{n \cdot \sum_{i=1}^n (x_i - \bar{X})^2} = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{X})^2} \quad \checkmark$$

- Δίνεται το ακόλουθο δείγμα παρατηρήσεων για την τιμή (X_i) και την ποσότητα (Y_i) μιας μάκρας υαφές.

i	Ποσ	Τιμή
1	87	1
2	86	1
3	74	1
4	79	1
5	68	1
6	84	1
7	139	0.95
8	122	0.95
9	102	0.95
10	180	0.85
11	179	0.85
12	187	0.85

Η γραμμική σχέση που συνδέει τις μεταβλητές δίνεται από το υπόδειγμα:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- (i) Να υπολογιστούν οι Ειτιμητές Ελαχίστων Τετραγώνων $\hat{\beta}_0$ και $\hat{\beta}_1$ και να γραφεί το ειτιμώμενο υπόδειγμα.
 (ii) Να βρεθεί η διακύμανση του διαταρακτικού όρου.

- (iii) Να υπολογιστούν τα αθροίσματα τετραγώνων TSS, ESS και RSS, καθώς και ο συντελεστής προσδιορισμού R^2 .

- Προφανώς, είναι $n=12$ ✓

Μας δίνεται τα $X_1, X_2, \dots, X_{12}$ καθώς και τα $Y_1, Y_2, \dots, Y_{12}$

Υπολογίζουμε τα $\bar{X} = \frac{1}{12} \sum_{i=1}^{12} X_i$ και $\bar{Y} = \frac{1}{12} \sum_{i=1}^{12} Y_i$

Στη συνέχεια βρίσκουμε τις διαφορές:

$$X_i - \bar{X} \quad \text{και} \quad Y_i - \bar{Y}$$

Δηλ. δημιουργούμε τη στήλη

$$\begin{matrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_{12} - \bar{x} \end{matrix}$$

και

$$\begin{matrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_{12} - \bar{y} \end{matrix}$$

Στη συνέχεια κατασκευάζουμε το γινόμενο των 2 προηγούμενων στηλών, δηλ τη στήλη $(x_i - \bar{x})(y_i - \bar{y})$ και ύστερα τις στήλες $(x_i - \bar{x})^2$ και $(y_i - \bar{y})^2$

ΠΡΟΣΟΧΗ

Σημειώνουμε ότι, στο τέλος κάθε στήλης υπολογίζουμε το άθροισμα των όρων της! Δηλ. στον πίνακα που έχουμε δημιουργήσει, η τελευταία γραμμή (η 13η) θα έχει ως στοιχεία τα αθροίσματα από τα δεδομένα των 12 γραμμών της ελάχιστης στήλης.

Με τα στοιχεία που έχουμε έως τώρα, μπορούμε να υπολογίσουμε τους ευριττές $\hat{\beta}_0, \hat{\beta}_1$ και να γράψουμε το ευτιμώμενο μοντέλο.

i	y_i	x_i	$y_i - \bar{y}$	$x_i - \bar{x}$	$(y_i - \bar{y})(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$\hat{y}_i$	$\hat{u}_i$	$\hat{u}_i^2$
1	89	1	-27.25	0.05	-1.3625	0.0025	742.5625	81.5833	7.4167	55.0069
⋮										
12	187	0.85	70.75	-0.1	-7.075	0.01	5005.5625	185.5833	1.4167	2.0069
Sum	1395	11.40	0.00	0.00	-31.200	0.045	22760.25	1395.0	0.00	1128.25

Έτσι έχουμε : $\bar{X} = \frac{1}{12} \sum_{i=1}^{12} X_i = \frac{11.40}{12} = 0.95$

$$\bar{Y} = \frac{1}{12} \sum_{i=1}^{12} Y_i = \frac{1395}{12} = 116.25$$

και $\sum_{i=1}^{12} (X_i - \bar{X})(Y_i - \bar{Y}) = -31.2$, ενώ $\sum_{i=1}^{12} (X_i - \bar{X})^2 = 0.045$

Άρα, $b_1^{\wedge} = \frac{-31.2}{0.045} = -693.33$

και $b_0^{\wedge} = \bar{Y} - b_1^{\wedge} \cdot \bar{X} = 116.25 - (-693.33) \cdot 0.95 = 794.91$

$\Rightarrow$ Επιτηρημένο υπόδειγμα : $\hat{Y}_i = 794.91 - 693.33 \cdot X_i$

(ii)

Άρα, τώρα μπορούμε να δημιουργήσουμε (για κάθε i υπολογίζουμε το επιτηρημένο $\hat{Y}_i$ που αντιστοιχεί στο X_i) τη στήλη με τα $\hat{Y}_i$ και στη συνέχεια για κάθε i βρίσκουμε το $\hat{u}_i = Y_i - \hat{Y}_i$.

Ύστερα, δημιουργούμε μια στήλη με τα $\hat{u}_i^2$.

Επομένως, μπορούμε ν' απαντήσουμε στο ερώτημα αναφορικά με τη διακύμανση του διαταρακτικού όρου :

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-k} = \frac{\sum_{i=1}^{12} \hat{u}_i^2}{12-(2)} = \frac{1128.25}{10} = 112.825 \quad \checkmark$$

μέγεθος
δείγματος

βαθμὶ
ἐλευθερίας

//
(n-k)

$K=2$, δηλαδή όσοι είναι
οι εὑριστῆς τοῦ μοντέλου, δηλ. 2
($\hat{b}_0, \hat{b}_1$)

(iii) Είναι:

$$TSS \text{ (Total Sum of Squares)} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$ESS \text{ (Explained Sum of Squares)} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$RSS \text{ (Residual Sum of Squares)} = \sum_{i=1}^n \hat{u}_i^2$$

Άρα, θα έχουμε:

$$TSS = \sum_{i=1}^{12} (Y_i - \bar{Y})^2 = 22760.25$$

$$RSS = \sum_{i=1}^{12} \hat{u}_i^2 = 1128.25$$

Γενικά ισχύει: $TSS = ESS + RSS$

$$\Rightarrow ESS = TSS - RSS \Rightarrow$$

$$\Rightarrow ESS = 22760.25 - 1128.25 = 21632 \quad \checkmark$$

Για το συντελεστή προσδιορισμού R^2 :

Από τη σχέση $TSS = ESS + RSS$ και

διαίρωντας και τα δύο μέλη αυτής με TSS , θα έχουμε:

$$\frac{TSS}{TSS} = \left(\frac{ESS}{TSS} \right) + \frac{RSS}{TSS} \Rightarrow$$

$\downarrow$ $\downarrow$
 1 R^2

$$\Rightarrow R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad 0 < R^2 < 1$$

Εδώ, $R^2 = 1 - \frac{1128.25}{22760.25} = 0.95 \quad \checkmark$

✓ Μεγάλο R^2 , σημαίνει υψηλός (σημαντικός) βαθμός
ερμηνευσιμότητας του μοντέλου.

8)

- Δίνεται δείγμα 13 παρατηρήσεων για τις μεταβλητές Y_i (μέσο ωρομίσθιο) και X_i (έτη εκπαίδευσης). Η γραμμική σχέση που τις συνδέει με βάση την οικονομική θεωρία είναι: $Y_i = \beta_0 + \beta_1 X_i + u_i$

- (i) Να υπολογισθεί η επιρροή της διακύμανσης του διαταρακτικού όρου.
 (ii) Βρείτε τη διακύμανση των $\hat{\beta}_0, \hat{\beta}_1$ (που έχουν προκύψει) με τη μέθοδο επιρροής ελαχ-τεταρ.
 (iii) Υπολογίστε το συντελεστή προσδιορισμού R^2 .

n	Y_i	X_i
1	4.4567	6
2	5.77	7
3	5.9787	8
4	7.3317	9
5	7.3182	10
6	6.5844	11
7	7.8182	12
8	7.8351	13
9	11.0223	14
10	10.6738	15
11	10.8361	16
12	13.615	17
13	13.531	18

► Επιτηρημένο υπόδειγμα: $\hat{Y}_i = -0.01445 + 0.724097 X_i$

Με βάση τις $n=13$ παρατηρήσεις για τα X_i και Y_i που έχουμε, υπολογίζουμε τα $\bar{X}$ και $\bar{Y}$ και δημιουργούμε πίνακα με στήλες: $X_i, Y_i, X_i - \bar{X}, Y_i - \bar{Y}, (X_i - \bar{X})^2, (Y_i - \bar{Y})^2, X_i^2, \hat{Y}_i, \hat{u}_i, \hat{u}_i^2$ και αργότερα τα αθροίσματά σε κάθε στήλη. Άρα, θα είναι:

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-k} = \frac{9.69}{13-2} = 0.8813 \quad \text{Οπότε: } \text{Var}(\hat{\beta}_0) = \frac{1}{n} \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$= 0.8813 \cdot \frac{2054}{13 \cdot 122} = 0.7649$$

(9)

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{0.8813}{182} = 0.004842$$

Για το συντελεστή προσδιορισμού R^2 :

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum_{i=1}^{13} \hat{u}_i^2}{\sum_{i=1}^{13} (y_i - \bar{y})^2} = 1 - \frac{9.69}{105.1183} = 0.9078$$

= 90.78 %

Σημαντικός βαθμός ερμηνευστικότητας
του μοντέλου