5ο Φροντιστήριο

- Για το απλό γραμμικό μοντέλο $y_i = b_0 + b_1 x_i + u_i$, $i=1,2,\dots,n$ να υπολογιστεί η διακύμανση για τον εκτιμητή Ελαχίστων Τετραγώνων $\hat{b}_0$ ($Var(\hat{b}_0)$).

- Γνωρίζουμε ότι με τη μέθοδο Ελαχίστων Τετραγώνων προκύπτουν οι εκτιμητές για τα b_0 και b_1 που είναι:

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

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

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

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

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

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

2)

$$Var(\hat{b}_0) = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \right)^2 \cdot \sigma^2 =$$

$\left[Var(Y_i) = \sigma^2 \text{ και} \quad \text{Αν } w = \sum a_i w_i \Rightarrow Var W = \sum a_i^2 Var(w_i) \right]$

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

$$= \sigma^2 \left(\frac{\cancel{n} \cdot 1}{n^{\cancel{2}}} + \frac{\bar{x}^2 \cancel{\sum_{i=1}^{n}(x_i-\bar{x})^2}}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{\cancel{2}}} - \frac{2\cancel{n} \cdot \bar{x}}{\cancel{n}} \cdot \frac{\cancel{\sum_{i=1}^{n}(x_i-\bar{x})}^{\,0}}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \right) =$$

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

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

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

$$= \sigma^2 \frac{\sum_{i=1}^{n} x_i^2 + \cancel{2n\bar{x}^2} - \cancel{2n\bar{x}^2}}{n \cdot \sum_{i=1}^{n}(x_i-\bar{x})^2} = \boxed{\sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

✓

Ν.δ.ο. $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})\hat{u}_i = 0$, για το απλό γραμμικό μοντέλο $Y_i = b_0 + b_1 X_i + u_i$

- Είναι $\hat{u}_i = Y_i - \hat{Y}_i$, άρα έχουμε:

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

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

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

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

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

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

Ισχύει γενικά: $\bar{\hat{Y}} = \bar{Y}$

Όμως: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (*)$

$$(1) \overset{*}{\Rightarrow} = \sum_{i=1}^{n}(\hat{b}_0 + \hat{b}_1 X_i)Y_i - \sum_{i=1}^{n}(\hat{b}_0 + \hat{b}_1 X_i)^2 =$$

4)

$$= \sum_{i=1}^{n} \hat{b}_0 y_i + \sum_{i=1}^{n} \hat{b}_1 x_i y_i - \sum_{i=1}^{n} \hat{b}_0^2 - \sum_{i=1}^{n} \hat{b}_1^2 x_i^2 - \sum_{i=1}^{n} 2\hat{b}_0 \hat{b}_1 x_i$$

$$= \hat{b}_0 \sum_{i=1}^{n} y_i + \hat{b}_1 \sum_{i=1}^{n} x_i y_i - n\hat{b}_0^2 - \hat{b}_1^2 \sum_{i=1}^{n} x_i^2 - \hat{b}_0 \hat{b}_1 \sum_{i=1}^{n} x_i - \hat{b}_0 \hat{b}_1 \sum_{i=1}^{n} x_i$$

$$= \hat{b}_0 \left(\sum_{i=1}^{n} y_i - n\hat{b}_0 - \hat{b}_1 \sum_{i=1}^{n} x_i \right) +$$

1η κανονική εξίσ.

2η κανον. εξίσ.

$$+ \hat{b}_1 \left(\sum_{i=1}^{n} x_i y_i - \hat{b}_0 \sum_{i=1}^{n} x_i - \hat{b}_1 \sum_{i=1}^{n} x_i^2 \right)$$

$$= 0 + 0 = 0 \checkmark$$

Κανονικές Εξισώσεις

$$\begin{cases} \sum_{i=1}^{n} y_i = n\hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i = \hat{b}_0 \sum_{i=1}^{n} x_i + \hat{b}_1 \sum_{i=1}^{n} x_i^2 \end{cases}$$

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

i	Ποσ.	Τιμή
1	89	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	186	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 Εκτιμητές Ελαχίστων Τετραγώνων $\hat{\beta_0}$ και $\hat{\beta_1}$ και να γραφεί το εκτιμώμενο 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 $X_1, X_2, \ldots, X_{12}$ καθώς και τα $Y_1, Y_2, \ldots, Y_{12}$

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:

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

6)

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 $(X_i-\bar{X})\cdot(Y_i-\bar{Y})$ 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 εκάστοτε στήλης.

Με τα στοιχεία που έχουμε έως τώρα, μπορούμε να υπολογίσουμε τους εκτιμητές ΕΤ $\hat{b}_0, \hat{b}_1$ και να γράψουμε το εκτιμώμενο 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 , $\hat{b}_1 = \frac{-31.2}{0.045} = -693.33$

και $\hat{b}_0 = \bar{Y} - \hat{b}_1 \cdot \bar{X} = 116.25 - (-693.33)\cdot 0.95 = 794.91$

$\Rightarrow$ Εκτιμημένο υπόδειγμα : $\boxed{\hat{Y}_i = 794.91 - 693.33 X_i}$

(ii)

Άρα, τώρα μπορούμε να δημιουργήσουμε (για κάθε i υπολογίζουμε το εκτιμημένο $\hat{Y}_i$ 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 :

8)

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

n: μέγεθος δείγματος

$n-K$: βαθμοί ελευθερίας // $(n-K)$

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

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

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

RSS (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: $\underline{TSS = ESS + RSS}$

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

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

Για το συντελεστή προσδιορισμού R^2:

Από τη σχέση $TSS = ESS + RSS$ και διαιρώντας και τα δύο μέλη αυτής με TSS, θα έχουμε:

$$\underbrace{\frac{TSS}{TSS}}_{1} = \underbrace{\frac{ESS}{TSS}}_{R^2} + \frac{RSS}{TSS} \Rightarrow$$

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

Εδώ, $R^2 = 1 - \dfrac{1128.25}{22760.25} = 0.95$ ✓

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

10)

- 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: $Y_i = b_0 + b_1 X_i + u_i$

(i) Να υπολογισθεί η εκτίμηση της διακύμανσης του διαταρακτικού όρου.

(ii) Βρείτε τη διακύμανση των $\hat{b_0}, \hat{b_1}$ (που έχουν προκύψει με τη μέθοδο εκτίμησης Ελαχ-Τετρ.)

(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

- Εκτιμημένο 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 ασφαλώς τα αθροίσματα σε κάθε στήλη. Άρα, θα είναι:

$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-K} = \frac{9.69}{13-2} = 0.8813$. Οπότε: $Var(\hat{b_0}) = \sigma^2 \cdot \frac{\sum_{i=1}^{n} X_i^2}{n \cdot \sum_{i=1}^{n}(X_i-\bar{X})^2}$

$= 0.8813 \cdot \frac{2054}{13 \cdot 182} = 0.7649$

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

Για το συντελεστή προσδιορισμού R^2:

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

$$= 90.78\%$$

Σημαντικός βαθμός ερμηνευτικότητας του μοντέλου.