

5.10. Γραμμικό μοντέλο πολλών ερμηνευτικών μεταβλητών

Το γενικευμένο κλασικό μοντέλο:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + e_i$$

εκτιμάται συνήθως με την μέθοδο των πινάκων.

	Y	X	
		X_1	X_2
\rightarrow			
\rightarrow			
\rightarrow			

αλλά αλλάζουν οι διαστάσεις του πίνακα X , αρα έχω περισσότερα πράγματα

$$b_{OLS} = (X'X)^{-1}X'Y$$

$$X'X = \begin{pmatrix} T & \sum x_1 & \sum x_2 & \dots & \sum x_k \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \dots & \sum x_1 x_k \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \dots & \sum x_2 x_k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_k & \sum x_1 x_k & \sum x_2 x_k & \dots & \sum x_k^2 \end{pmatrix}$$

ε γενικλοησι εα!

Y	X_1	X_2
10	1	3
12	4	5
22	6	7

$$X = \begin{pmatrix} 1 & 3 \\ 4 & 5 \\ 6 & 7 \end{pmatrix}$$

$$X' = \begin{pmatrix} 1 & 4 & 6 \\ 3 & 5 & 7 \end{pmatrix}$$

$$\sigma^2 = \frac{\sum e_i^2}{T-k-1} = \frac{Y'Y - b_{OLS}'X'Y}{T-k-1}$$

επηρεάζει ως προς κύρια διαγώνια

ο πίνακας αυτός χρησιμοποιείται για να δούμε αν οι μεταβλητές που χρησιμοποιούμε είναι απαραίτητες ή όχι για το μοντέλο και

$$\sigma^2 (X'X)^{-1} = \begin{bmatrix} \text{Var}(b_0) & \text{Cov}(b_0 b_1) & \dots & \text{Cov}(b_0 b_k) \\ \text{Cov}(b_1 b_0) & \text{Var}(b_1) & \dots & \text{Cov}(b_1 b_k) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(b_k b_0) & \text{Cov}(b_k b_1) & \dots & \text{Var}(b_k) \end{bmatrix}$$

1

Συνολικά Άθροισμα Τετραγώνων

$$ΣΑΤ = ΣΣΤ = Y' \cdot Y - T \cdot \bar{Y}^2$$

Ερμηνευτικά: Δίνει την απόκλιση του Y από \bar{Y}

Άθροισμα Τετραγώνων Παλινδρόγησης

$$ΑΤΠ = ΣΣΡ = b'_{OLS} \cdot X' \cdot Y - T \cdot \bar{Y}^2$$

Ερμηνευτικά: Δίνει την απόκλιση του \hat{Y} από \bar{Y}

Άθροισμα Τετραγώνων Σφάλματος

$$ΑΤΣ = ΣΣΕ = Y' \cdot Y - b'_{OLS} \cdot X' \cdot Y$$

Ερμηνευτικά: Δίνει την απόκλιση \hat{Y} από Y

Χρησιμοποιούν
στον
υπολογισμό
του R^2

$$R^2 = \frac{ΣΣΡ}{ΣΣΤ}$$

Αποδοχή Υποθέσεων

3

Ελέγχεται η μηδενική υπόθεση:

$$H_0: \beta_i = 0 \text{ με } H_1: \beta_i \neq 0$$

σε επίπεδο σημαντικότητας $\alpha\%$.

από πίνακα
bols

Χρησιμοποιούμε τον t -λόγο $= \frac{b_i}{\sqrt{\text{var } b_i}}$ όπου $\sqrt{\text{Var}(b_i)}$ το τυπικό σφάλμα

του b_i που δίνεται από την τετραγωνική ρίζα του αντίστοιχου διαγώνιου στοιχείου του πίνακα $\sigma^2 (X'X)^{-1}$.

από πίνακα

Μακρο-οικονομικά

Να βάλω η τεταβλητή X_i
στο μοντέλο του Y ή όχι?

αν.δ.σ. Y : $Y = 5 + 20X_1 + 0.5X_2 + 0.02X_3$
 b_0 b_1 b_2 b_3
 τυπικά σφάλματα

Πόσο στατιστικά σημαντική είναι η τεταβλητή X_i :

α : επίπεδο σημαντικότητας (συνηθως 10% ; 5% ; 1%)

Για να πραγματοποιηθεί ο έλεγχος, θα πρέπει να συγκριθεί ο t -λόγος με την τιμή της $t_{\alpha/2; T-k-1}$ από τον πίνακα της t -κατανομής. Εάν $t_{\alpha/2; T-k-1} < t\text{-ratio}(b_i)$ η μηδενική υπόθεση H_0 δεν γίνεται δεκτή που σημαίνει ότι η συγκεκριμένη ανεξάρτητη - ερμηνευτική μεταβλητή επηρεάζει σημαντικά την εξαρτημένη μεταβλητή.

$$t_{crit} = t_{\frac{\alpha}{2}; T-k-1}$$

βαθμιά
εξελιγμένη

Παράδειγμα

Έστω τα δεδομένα:

Y	8	10	13	14	15
X ₁	1	2	3	4	5
X ₂	6	5	4	5	3

όπου Y η μεταβλητή που εκφράζει το ύψος ενός φυτού, X₁: τα επίπεδα βροχής και X₂: την ξηρασία (ανομβρία). (i) Να εκτιμηθεί το μοντέλο: $\hat{y} = b_0 + b_1x_1 + b_2x_2$. (ii) Ποια από τις δύο ερμηνευτικές μεταβλητές είναι σημαντική για το ύψος του φυτού; (για $\alpha = 10\%$ ε.σ.).

Y	X ₁	X ₂
8	1	6
10	2	5
13	3	4
14	4	5
15	5	3

$$X = \begin{bmatrix} 1 & 6 \\ 2 & 5 \\ 3 & 4 \\ 4 & 5 \\ 5 & 3 \end{bmatrix} \quad X^T = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 5 & 4 & 5 & 3 \end{bmatrix}$$

$$X^T \cdot X = \begin{bmatrix} 5 & 15 & 23 \\ 15 & 55 & 63 \\ 23 & 63 & 111 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 26.7 & -2.7 & -4 \\ -2.7 & 0.325 & 0.375 \\ -4 & 0.375 & 0.625 \end{bmatrix}$$

$$X^T \cdot Y = \begin{bmatrix} 60 \\ 198 \\ 265 \end{bmatrix}$$

$$b_{OLS} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = (X^T X)^{-1} X^T Y = \begin{bmatrix} 26.7 & -2.7 & -4 \\ -2.7 & 0.325 & 0.375 \\ -4 & 0.375 & 0.625 \end{bmatrix} \begin{bmatrix} 60 \\ 198 \\ 265 \end{bmatrix} = \begin{bmatrix} 7.4 \\ 1.725 \\ -0.125 \end{bmatrix}$$

$$\begin{aligned} b_0 &= 7.4 \\ b_1 &= 1.725 \\ b_2 &= -0.125 \end{aligned}$$

Άρα

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$y = 7.4 + 1.725 x_1 - 0.125 x_2$$

$$\hat{\sigma}^2 = \frac{Y^T Y - b_{OLS}^T X^T Y}{T - k - 1} = \frac{754 - 752.425}{5 - 2 - 1} = \frac{1.575}{2} = 0.7875$$

$$\hat{\sigma}^2 (X^T X)^{-1} = 0.7875 \cdot \begin{bmatrix} 26.7 & -2.7 & -4 \\ -2.7 & 0.325 & 0.375 \\ -4 & 0.375 & 0.625 \end{bmatrix} = \begin{bmatrix} 21 & -0.26 & -0.50 \\ -0.26 & 0.26 & 0.26 \\ -0.50 & 0.26 & 0.50 \end{bmatrix}$$

$$t_{crit} = t_{\frac{\alpha}{2}, T-k-1} = t_{0.05, 5-2-1} = t_{0.05, 2} = 2.92$$

$$\left. \begin{aligned} 3.38 &> 2.92 \\ t &> t_{crit} \end{aligned} \right\} \text{Άρα } X_2 \text{ είναι}$$

$$\text{για μεταβλητή } X_1: t = \frac{b_1}{\sqrt{\text{var}(b_1)}} = \frac{1.725}{\sqrt{0.26}} = \frac{1.725}{0.509} = 3.38$$

στην κριτική τιμή $\alpha = 10\%$

$$t_{crit} = 2.92$$

Για σταθμίζοντας X_2 : $t = \frac{|b_2|}{\sqrt{\text{Var } b_2}} = \frac{|-0.125|}{\sqrt{0.5}} = \frac{0.125}{0.71} = 0.176$

$t < 2.92$ ⑤
 Άρα X_2 είναι
 στατιστικά σημαντική
 για $\alpha = 10\%$