2024-01-17

Διαγνωστικοί Έλεγχοι (Έλεγχοι Καταλοίπων)

Γραμμικό Μοντέλο

$$Y = \underline{b_0 + b_1 X} + \varepsilon$$

ε τυχαία απόκλιση

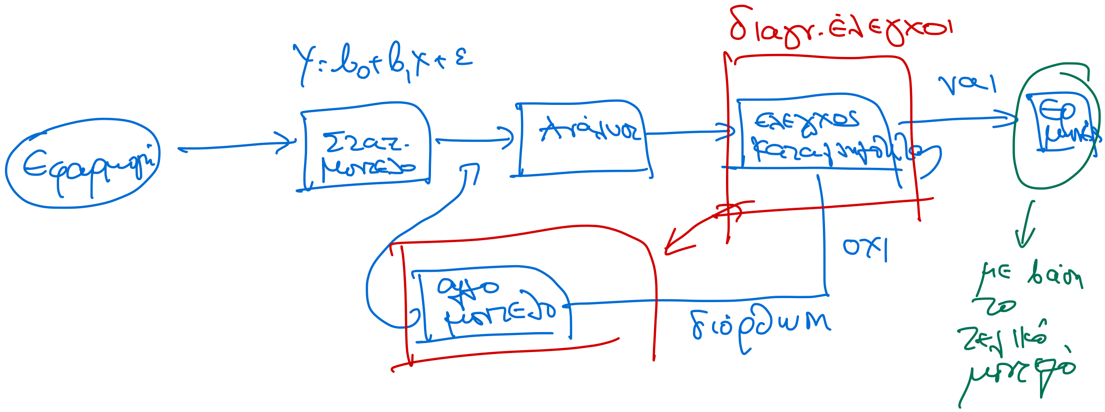
Δείγμα $\quad Y_j = b_0 + b_1 x_j + \varepsilon_j \qquad j = 1, \dots, n$

Υποθέσεις

1. $\varepsilon_j \sim N(0, \sigma^2)$ — Κανονικότητα
2. $Var(\varepsilon_j) = \sigma^2$ — Ομοσκεδαστικότητα
3. ε_j ανεξάρτητα μεταξύ των παρατηρήσεων — Ανεξαρτησία

b_0, b_1, σ^2 άγνωστες

(1)



(2) Έλεγχος για έκτροπες κ' επιδραστικές παρατηρήσεις

Μέρος 1 : <u>Έλεγχος Υποθέσεων του Μοντέλου</u>

Οι υποθέσεις αναφέρονται στα $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$

Αν είχα δείγμα τιμών των $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ θα μπορούσα να κάνω κάποιους στατιστικούς ελέγχους

Το δείγμα μας όμως είναι

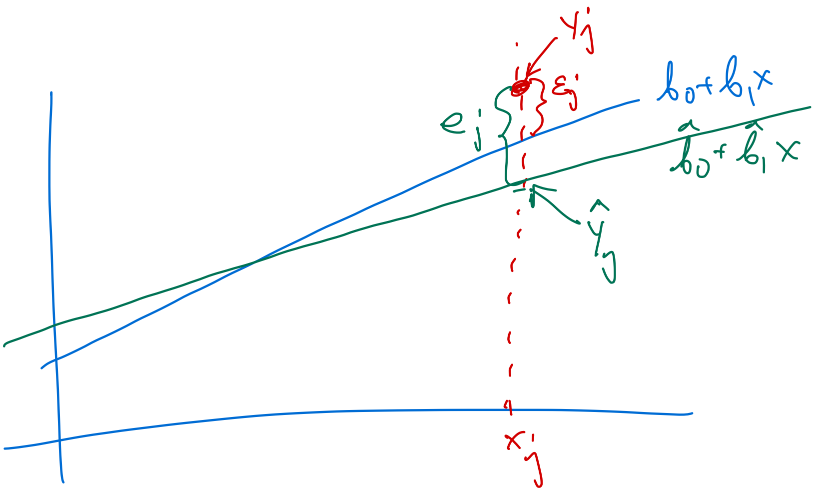
$$\begin{pmatrix} X_1 & Y_1 \\ \vdots & \vdots \\ X_n & Y_n \end{pmatrix} \longrightarrow \hat{b}_0, \hat{b}_1, \hat{\sigma}^2$$

$$Y_j = b_0 + b_1 X_j + \varepsilon_j \Rightarrow \varepsilon_j = Y_j - b_0 - b_1 X_j$$

$\boxed{\text{άγνωστες τιμές}}$!!

"Εκτιμήσεις" των πραγματικών ε_j :

$$e_j = Y_j - \hat{b}_0 - \hat{b}_1 x_j = Y_j - \hat{Y}_j = \text{κατάλοιπο (residual) της παρατήρησης } j$$



Προβλήματα

1. $e_j = Y_j - \hat{b}_0 - \hat{b}_1 x_j \quad , j=1,\dots,n$

$\hat{b}_0 , \hat{b}_1$ εξαρτώνται από <u>όλες</u> τις παρατηρήσεις

Επίσης ισχύει ότι για τις εκτιμήτριες LSE

τότε $\frac{1}{n}\sum e_j = 0 \Rightarrow e_1, \dots, e_n$ δεν είναι ανεξάρτητα

stata (stdp)

② $Var(e_j) = s_j^2$ εξαρτάται από το j (από το x_j)

δεν είναι η ίδια για όλες τις παρατηρήσεις

Χρειαζόμαστε $\boxed{\text{μετασχηματισμούς των καταλοίπων}}$

Μετασχηματισμοί Καταλοίπων

① Τυποποιημένα Κατάλοιπα (Standardized)

$$\tilde{e}_j = \frac{e_j - 0}{s_j} \Rightarrow Var(\tilde{e}_j) = \frac{Var(e_j)}{s_j^2} = 1$$

② Studentized Κατάλοιπα.

$$r_j = \frac{\tilde{e}_j}{\sqrt{1-h_j}} = \frac{e_j}{s_j\sqrt{1-h_j}}$$

h_j = leverage της παρατήρησης j

H : hat matrix $\quad h_j$: διαγώνια στοιχεία του H.

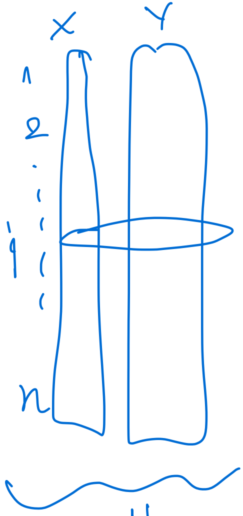
$$0 < h_j < 1$$

Αποδεικνύεται $\quad r_j$ προσεγγιστικά ανεξάρτητα $\sim t$

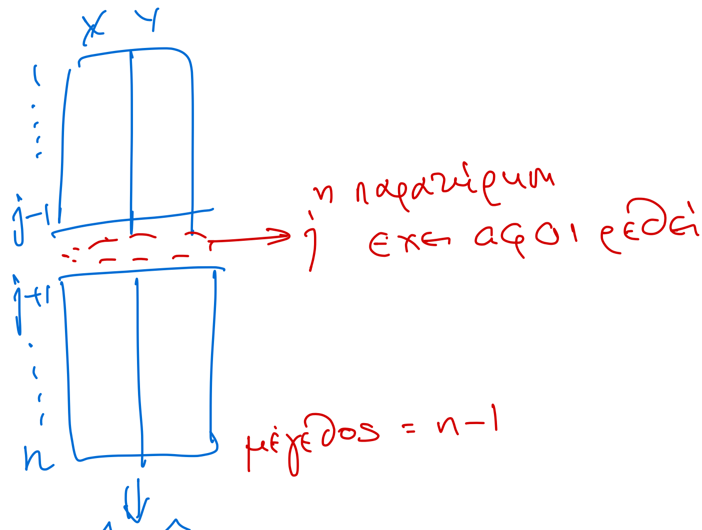
③ Κατάλοιπα Jackknife

$$\tilde{r}_j = \frac{e_j}{S_{(-j)}\sqrt{1-h_j}}$$

$S_{(-j)}$ η τυπική απόκλιση της παρατήρησης j που προέρχεται από εκτιμήσεις του αρχικού δείγματος χωρίς την παρατήρηση j



$\hat{b}_0, \hat{b}_1, s_j$



$\hat{b}_0, \hat{b}_1, s_{(-j)}$

$$\boxed{\text{Resampling methods} \begin{cases} \text{Bootstrap} \\ \text{Jacknife} \end{cases}}$$

Ισχύει ότι $\tilde{r}_j = \dfrac{e_j}{s_{(-j)}\sqrt{1-h_j}}$

Είναι <u>ασυσχέτιστα</u>

$$\tilde{r}_j \sim t_{n-k-2} =$$

k αρ. μεταβλητών στο μοντέλο

$n-k-2 = (n-1) - (k+1) = \text{dfer}$

στο δείγμα $(n-1)$

Επίσης για ~~n-k-2~~ μεγάλο $\underline{t_{n-k-2} \approx N}$

Προσοχή

Στό Stata και στο R: τα studentized res → standard; τα jackknife → student.

Διαγνωστικοί Έλεγχοι

- Διαγράμματα
- Έλεγχοι Υποθέσεων

① **Διαγράμματα Καταλοίπων**

$$\boxed{\tilde{r}_j \sim \hat{Y}_j}$$

(ισχύει στα $\tilde{r}_j, \hat{Y}_j$ ασυσχέτιστα)