

3-2-2025

Πρόβλημα 1

① $R^2 = 0.4721$

$$p\text{-value}(F\text{test}) = p\text{-value}(b_1) \approx 10^{-4}$$

συνέχουν μερικοί αιώσαντες & αυτιόαυτα
σταυρικά σιμαυτά

Η αιώσαν αυτιόαυτα $\approx 50\%$ αυτιόαυτα
αυτιόαυτα

② $\text{Density} = b_0 + b_1 \text{ distance} + \varepsilon$

$\hat{b}_1 = 0.0038 \rightarrow$ για αυτιόαυτα αυτιόαυτα αυτιόαυτα
αυτιόαυτα 1 αυτιόαυτα, η αυτιόαυτα αυτιόαυτα
αυτιόαυτα αυτιόαυτα αυτιόαυτα 0.0038 αυτιόαυτα

$\hat{b}_0 = 1.212$ (δεν αυτιόαυτα αυτιόαυτα)

Αυτιόαυτα αυτιόαυτα αυτιόαυτα αυτιόαυτα
αυτιόαυτα $\hat{b}_0$ αυτιόαυτα αυτιόαυτα αυτιόαυτα αυτιόαυτα
για αυτιόαυτα = αυτιόαυτα αυτιόαυτα αυτιόαυτα.

③ $\text{stdf} = 0.1007$ (για αυτιόαυτα = 18)

$$\hat{y} = \hat{b}_0 + \hat{b}_1 18 = 1.2804$$

$$n = 27$$

$$k = 1$$

$$t_{\alpha/2, n-k-1} = t_{0.025, 25} = \dots$$

$$df_{\text{er}} = 25$$

αυτιόαυτα αυτιόαυτα αυτιόαυτα: $[1.029 \leq y \leq 1.487]$ (95%)

④ (predict ...)

κρίσιμες τιμές ($n=27, k=1$)

$$\text{Jackknife} \approx 3,50$$

$$\text{leverage} \approx 0,35$$

$$\text{Cook's distance } n-k-1=25$$

$$\text{κρίσιμη τιμή} = 17,18 \text{ για } (25 \cdot \text{Cook})$$

$$\text{influential αν } 25 \text{Cook's} > 17,18 \Rightarrow \text{Cook} > \frac{17,18}{25} \approx \underline{\underline{0,7}}$$

Outlier αν $|r_{\text{jack}}| > 3,50 \Leftrightarrow r_{\text{jack}} > 3,5 \text{ ή } r_{\text{jack}} < -3,5$
for undetected outliers

Influential

(leverage $> 0,35$) : for undetected εισπρακτικές

(Cook's distance $> 0,7$) " " "

⑤ Shapiro-Wilk

$$H_0: r \sim N$$

$$H_1: 0 < x_1 \sim N$$

$$p=0,28$$

for μπορεί να απορριφθεί
η κανονικότητα

Πρόβλημα 2

① p-value (ttest) < 0.05 για perc-min

Αν το % περριώνων αυξάνεται κατά 1% (ofes or affs
της παραμείνουν σταθερές) το ποσοστό undercount
αυξάνεται κατά μέσο όρο κατά 0.08(%)

② Μετα το backward step :

$$\text{undercount} = -1,36 + 0.063 \text{ perc-min} + 0.026 \text{ crimrate} \\ + 0.221 \cdot \text{diffeng}$$

③ Model 2: perc-min, poverty, hsgrad.

α) Πως συγκρίνουμε το model 2 με το full model?

Full: $\text{undercount} = b_0 + b_1 \text{ perc-min} + b_2 \text{ crimrate} \\ + b_3 \text{ poverty} + b_4 \text{ diffeng} \\ + b_5 \text{ hsgrad} + b_6 \text{ housing}$

← *υπόλοιπες βάρες model 2*

$$\text{Αν model 2} \equiv \text{full} \Leftrightarrow b_2 = b_4 = b_6 = 0$$

$$H_0 : b_2 = b_4 = b_6 = 0 \quad H_1 : \text{τουλάχιστον ένα} \neq 0.$$

Stata: test crimrate diffeng housing

$$P\text{-value} = 0.0706$$

Για $\alpha = 5\%$ δεν απορ. η H_0

δεν αποκλείεται στατιστικά σημαντική επίδραση

οτι το model 2 είναι υποβέβαιο.

(Για το διαγνώσκειν θα πρέπει να ξέρεις
τους τύπους για το F or partial F-test
κ' για το C_p Mallows)

⑥ Σύγκριση model με full model βάσει του C_p

$$C_p = \frac{SSE(\text{partial})}{MSE(\text{full})} - [n - 2(p+1)] \quad P = \# \text{ μεταβλητών στο partial model}$$

Εδώ $p=3$, $n=66$

Divorcan output για full model $\rightarrow MSE(\text{full}) = 2.57$
partial " $\rightarrow SSE(\text{partial}) = 226$

$$C_p = \frac{171}{2.57} - [66 - 8] = \frac{171}{2.57} - 58 \approx 8.54$$

Model 2 $\equiv$ full model αν $C_p \approx p+1 = 4$

Αν $C_p > p+1$ Model 2 υποβέβαιο

Εδώ $C_p \approx 8.5 > 4 \Rightarrow$ υποβέβαιο

④ ① undercount = $b_0 + b_1 \cdot \text{housing}$

$$p\text{-value}(b_1) = 0.009 < 0.05 \quad *$$

② undercount = $b_0 + b_1 \cdot \text{housing} + b_2 \cdot \text{poverty}$

$$p\text{-value}(b_1) = 0.085 > 0.05 \quad (\text{οχι σημαντική})$$

(a) : housing σταθεστά εμφανίζει (μείνει ως θα μείνει)

(b) : Παρνολα της poverty πάλι να είναι εμφανίζει

Ο κόσμος είναι καλύτερος ότι poverty, housing υποχρεωτικές
στο δίπλα.

↑
Αντί θα μπορούσε να σχετίζει κάποιος μια λαμβάνεται
μεταξύ housing & poverty

Πρόβλημα 3

region = 1 $\Rightarrow$ North

(ενίοτε αναφέρεται = South)

$$\text{mean}(\text{elev}) = 927.82$$

$$\text{model} \hat{=} \text{damage} = b_0 + b_1 \cdot \text{elev} + b_2 \cdot \text{region} + b_3 \cdot \text{elev} \cdot \text{region}$$

$$\hat{b}_0 = 37.87$$

$$\hat{b}_1 = -0.017$$

$$\hat{b}_2 = 5.389$$

$$\hat{b}_3 = 0.108$$

$$\text{South (region}=0) : \text{damage} = 37.87 - 0.017 \cdot \text{elev}$$

$$\text{North (region}=1) : \text{damage} = (37.87 + 5.389) + (0.108 - 0.017) \cdot \text{elev}$$

$$\text{damage} = 43.2 + 0.091 \cdot \text{elev}$$

South : elevation ποταμιώδεις (στην ανατολική ή σε ορεινά)

North : " ενδοχώρας.

② $\hat{b}_0$: region=0 $\Rightarrow$ south
elev=0 $\Rightarrow$ μέσο υψόμετρο.

Για τις πιο μέσες υψόμετρον σε νότια τοιχίες

" μέση πρόβλεψη είναι 37,87%.

$\hat{b}_1$ (region $\Rightarrow$ -main effect)

Αν το υψόμετρο αυξηθεί κατά 1 m σε ορισμένες ντολές η προσβολή μειώνεται κατά μέσο όρο κατά 0,017%

$\hat{b}_2$ main effect region (για elev = 0)

Για δύο ορισμένες στο μέσο υψόμετρο, η προσβολή στο north έχει κατά μέσο όρο 5,39% μεγαλύτερη προσβολή από την προσβολή στο south.

$\hat{b}_3$: interaction

Για μια προσβολή στο north η επίδραση της αύξησης του υψομέτρου κατά 1 m είναι 0,108% μεγαλύτερη από την αντίστοιχη επίδραση σε μια προσβολή στο south.

③ $\hat{b}_1, \hat{b}_2$ δεν είναι στατιστικά σημαντικά $\neq 0$

Αν θεσπίσει $\hat{b}_1 = 0, \hat{b}_2 = 0$ και διαφοροποιήσει οι εξισώσεις;

South damage = 37,87

North damage = 37,87 + 0.108 · elev.

