Επαναληπτικό Μάθημα 30-1-2024

Πρόβλημα 1

X = distance (km)

Y = density (gr/κ.εκ.)

$Y = b_0 + b_1 X + \varepsilon$ $\quad \varepsilon \sim N(0, \sigma^2)$

1-2 $\hat{b}_0 = 1.212$ (δεν έχει φυσική ερμηνεία)

$\hat{b}_1 = 0.0038$

$p = 7 \cdot 10^{-5} \simeq 0$

$R^2 = 0.47$

3 Stata

(δεν υπάρχει παρατήρηση με x=18)

1) Θα βάζαμε νέα παρατήρηση με x=18 και y κενό

2) Πάλι reg y ~ x

3) predict (sf, stdf) $\quad \hat{y} \pm t_{\alpha/2, n-2} \, stdf_{(x=18)}$

option (Std. error of forecast)

(R) για προβλέψεις σε νέες παρατηρήσεις

χρησιμοποιούμε διαφορετικό dataframe

Εδώ δημιουργώ data frame με

μία μόνο γραμμή, distance=18

	reef	dist	dens
newdata :	[NA	18	NA]

Εδώ το διάστημα πρόβλεψης

για $X=18$ $\quad 1{,}072 \le Y \le 1.487$

---

(4) jacknife, leverage, Cook's distance.

<u>Κρίσιμες τιμές</u> $\quad n=27, \; k=1$

$\underline{\text{jacknife} = 3.50}$

$\underline{\text{leverage} = 0.35}$

Cook. $n-k-1 = 27-2 = 25 \Rightarrow$ κρίσιμη τιμή ≈ 17.5 (για $25d$)

influential αν $25d > 17.5 \Rightarrow d > \frac{17.5}{25} = \underline{0.7}$

$\nexists$ outliers or influential observations

(5) shapiro.wilk. $\quad$ p-v $= 0.27$, δεν απορρ. η $N(0,1)$

Πρόβλημα 2

$P_{perc\_min} = 0.0005 < 5\%$

① $\hat{b}_1 = 0.084 \Leftarrow$ Για αύξηση του %μειον κατά 1% το %υποκατ. αυξάνεται κατά 0.084%

Stepwise-backward { Stata με βάση p-value
R με βάση AIC

② Καταλήγουμε

model2 : $undcount = b_0 + b_1\, perc\_min + b_2\, crimrate + b_3\, diffeng + b_4\, hsgrad$

③ model3 : (percmin, poverty, hsgrad). (partial)

$p = 3$

model1 : (all) (full)

υπόλοιπες

$undcount = b_0 + b_1\, percmin + b_2\, crimrate + b_3\, poverty + b_4\, diffeng + b_5\, hsgrad + b_6\, housing.$

partial F-test :

$H_0 : b_2 = b_4 = b_6 = 0$ H_1 : τουλάχ ένα $\neq 0$

Στο R τρέχουμε το full κ' το partial model

full → model1, partial model → model3

anova(model3, model1) ← F-test

pvalue ≈ 0.07 στο $\alpha = 5\%$ όχι σημαντική διαφορά

Mallow's $C_p = \dfrac{SSE(p)}{MSE(k)} - (n - 2(p+1))$

$SSE(p)$ ← model3

$MSE(k)$ ← full (model1)

$n = 27$

$p = 3$

$SSE(p) = 170.784$

$MSE(k) = 2.579$

$\Rightarrow$

$\Rightarrow C_p = 8.45$

$p+1 = 4$

$\Big\}\ C_p > p+1$

model3 υποδεέστερο του model1

(4)

housing: p-value $= 0.0087 < 1\%$

housing+poverty: p-value(housing) $= 0.085 \simeq 9\%$

Πρόβλημα 3

$$damage = b_0 + b_1\, elevc + b_2\, X_{north} + b_3\, X_{north} \times elevc$$

$$\hat{b}_0 = 37{,}87$$
$$\hat{b}_1 = -0{,}017$$
$$\hat{b}_2 = 5.39$$
$$\hat{b}_3 = 0{,}108$$

(1) South $X_{north} = 0$

$$\hat{y} = 37{,}87 - 0{,}017\, elevc$$

$\hat{b}_0 = 37{,}87$: $elevc = 0 \Rightarrow elev =$ μέσο υψόμετρο

$X_{north} = 0$

Για περιοχές South με μέσο υψόμετρο, το μέσο ποσοστό βλάβης είναι 37,87% των δέντρων

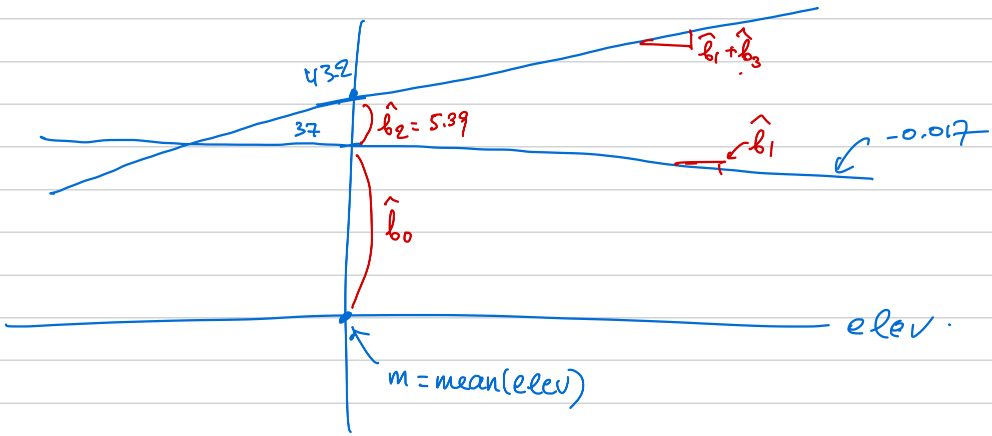
$\hat{b}_1 = -0{,}017$ επίδραση του υψομέτρου σε South

② North $\quad X_{north} = 1$

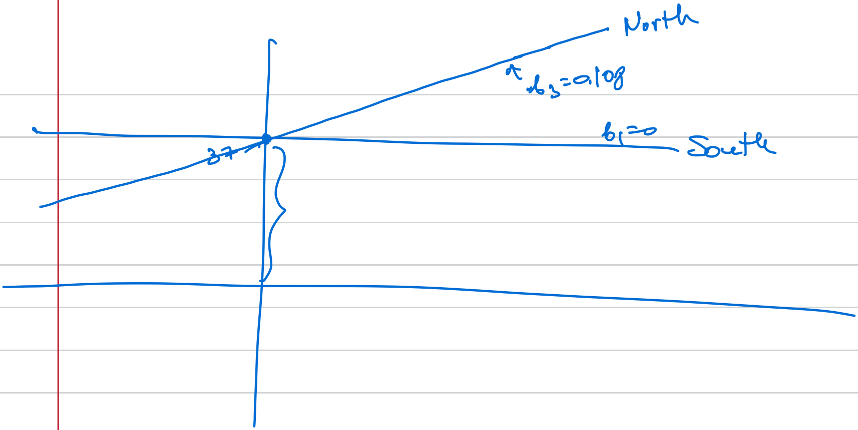
$$\hat{y} = (\underbrace{37.87}_{\hat{b}_0} + \underbrace{5.39}_{\hat{b}_2}) + (\underbrace{-0.017}_{\hat{b}_1} + \underbrace{0.108}_{\hat{b}_3})\, elevc$$

$$= 43.2 + 0.091\ elevc$$

③a <u>Ενδ. ερώτηση</u> Σχεδιάστε (προσεγγιστικά) τις δύο γραμμές λαμνδε.



③b b_1, b_2 δεν είναι στατ. σημ. $\neq 0$.


North
b3=0.108
b1=0
South
37