ΕΡΩΤΗΣΕΙΣ ΠΟΛΛΑΠΛΗΣ ΕΠΙΛΟΓΗΣ

Θεωρήστε το απλό γραμμικό μοντέλο :

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, \ldots, n, \tag{1}$$

Για τα σφάλματα ϵ_i υποθέστε πως έχουν μέση τιμή 0 και είναι ομοσκεδαστικά ($var(\epsilon_i) = \sigma^2$) και ασυσχέτιστα ($cov(\epsilon_i, \epsilon_j) = 0$ για $i \neq j$). Οι ερωτήσεις 1-5 αφορούν αυτό το μοντέλο.

1. Οι παράμετροι που πρέπει να εκτιμηθούν είναι:
   - a. $(x_i, y_i),\ i = 1, \ldots, n.$
   - b. $\epsilon_i,\ i = 1, \ldots, n.$
   - c. $\beta_0, \beta_1, E(\epsilon_i^2).$
   - d. $\beta_0, \beta_1, E(\epsilon_i).$

2. Η ευθεία ελαχίστων τετραγώνων 'περνάει' πάντα από το σημείο
   - a. 0
   - b. $(0, \bar{x})$
   - c. $(0, \bar{y})$
   - d. $(\bar{x}, \bar{y})$

3. Για τα υπόλοιπα της παλινδρόμησης, $\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, ισχύει
   - a. $var(\hat{\epsilon}_i) = \sigma^2$
   - b. $var(\hat{\epsilon}_i) < var(\epsilon_i)$
   - c. $var(\hat{\epsilon}_i) > var(\epsilon_i)$
   - d. $var(\hat{\epsilon}_i) = var(\epsilon_i)$

4. Τι από τα παρακάτω ΔΕΝ ΙΣΧΥΕΙ απαραίτητα
   - a. $\sum x_i \hat{\epsilon}_i = 0$
   - b. $\sum x_i \hat{\epsilon}_i = E(\sum x_i \epsilon_i)$
   - c. $\sum x_i \hat{\epsilon}_i = \sum \hat{\epsilon}_i$
   - d. $\sum x_i \hat{\epsilon}_i > \sum \hat{\epsilon}_i$

5. Τι από από τα παρακάτω ΔΕΝ ΙΣΧΥΕΙ απαραίτητα
   - a. Το 10% διάστημα πρόβλεψης σε τιμή $x = x_0$ είναι πλατύτερο από το αντίστοιχο 99,99% διάστημα εμιστοσύνης
   - b. Το 95% διάστημα πρόβλεψης σε τιμή $x = x_0$ είναι στενότερο από το 99% διάστημα πρόβλεψης
   - c. Το 95% διάστημα πρόβλεψης σε τιμή $x = x_0$ είναι πλατύτερο από το αντίστοιχο 95% διάστημα εμιστοσύνης

d. Το 95% διάστημα πρόβλεψης σε τιμή $x = x_0$ είναι πλατύτερο από το αντίστοιχο 90% διάστημα εμιστοσύνης

Θεωρήστε το κανονικό γραμμικό μοντέλο :

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n), \end{aligned} \tag{2}$$

όπου για τον $n \times p$ πίνακα συμμεταβλητών $\boldsymbol{X}$ ισχύει $rank(X) = p$, $\mathbf{0}_n$ είναι ένα $n \times 1$ διάνυσμα με μηδενικά, $\boldsymbol{I}_n$ είναι ένας $n \times n$ ταυτοτικός πίνακας με μοναδες στην διαγώνιο και μηδενικά εκτος διαγωνίου. Οι ερωτήσεις 6-12 αφορούν αυτό το μοντέλο

6. Τι από τα παρακάτω <u>ΔΕΝ ΙΣΧΥΕΙ</u>
   a. $\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{X}^T\hat{Y}$
   b. $\boldsymbol{X}\boldsymbol{Y} = \boldsymbol{X}\hat{Y}$
   c. $\sum \hat{\epsilon}_i = 0$
   d. $\sum_{i=1}^n x_{ij}\hat{\epsilon}_i = 0,\ \forall j$

7. Ισχύει πως
   a. $\frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \sim \chi^2_n$
   b. $\frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \sim \chi^2_n$
   c. $\frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \sim F_{p,n-p}$
   d. Τίποτα από τα παραπάνω

8. Μια αμερόληπτη εκτιμήτρια της σ^2 είναι
   a. $\frac{1}{n-p}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$
   b. $\frac{1}{n-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$
   c. $\frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$
   d. $\frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$

9. Έστω πως $n = 100, p = 5$. Θέλουμε να ελέγξουμε ταυτόχρονα τις εξής 3 μηδενικές υποθέσεις: A) $\beta_0 = 2$, B) $\beta_1 = \beta_2$, Γ) $\beta_4 = 0$. Αυτό είναι ισοδύναμο με έλεγχο της μηδενικής υπόθεσης $H_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$, όπου
   a.
$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$
   b.
$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0, \end{bmatrix} \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

c.

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0, \end{bmatrix} \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

d.

$$\boldsymbol{R} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

10. (*Συνέχεια της προηγούμενης.*) Η κλασσική σ.σ.ε., έστω U, της μηδενικής υπόθεσης παραπάνω θα έχει ποιά κατανομή, κάτω από την μηδενική υπόθεση
    a. χ^2_{99}
    b. $N(0,1)$
    c. $F_{3,95}$
    d. t_{95}

11. (*Συνέχεια της προηγούμενης.*) Αν η τιμή της σ.σ.ε. παραπάνω, u, που παρατηρήσαμε ισούται με 4, τότε το $p - value$ του παραπάνω ελέγχου είναι
    a. $P(U > 4|H_0)$
    b. $2 \times P(U > 4|H_0)$
    c. $P(U > 4|H_0) + P(U < -4|H_0)$
    d. Τίποτα από τα παραπάνω

Θεωρήστε το παρακάτω output από τη προσαρμογή μοντέλου παλινδρόμησης σε μη-καπνιστές. Η απόκριση είναι ο λογάριθμος της τιμής της σπιρομέτρησης και οι συμμεταβλητές είναι η ηλικία (age, σε έτη) το ύψος (ht, σε ίντσες) και το φύλο (sex: 0 για τα κορίτσια, 1 για τα αγόρια). Τα Adj SS αναφέρονται στο άθροισμα τεραγώνων που εξηγείται από έναν παράγοντα υπό την παρουσία όλων των υπολοίπων συμμεταβλητών, ενώ τα Seq SS είναι το άθροισμα τεραγώνων που εξηγείται από έναν παράγοντα υπό την παρουσία των προηγούμενων συμμεταβλητών που εισήχθησαν στο μοντέλο . Όπως αναφέρεται και στο output οι έλεγχοι στον πίνακα Anova βασίζονται στα Seq SS. Οι ερωτήσεις 12-18 αναφέρονται σε αυτό το output.

NON-SMOKERS

Regression Analysis: ln(FEV) versus age; ht; sex

Regression Equation

ln(FEV) = -1,9059 + 0,02587 age + 0,04178 ht + 0,0290 sex

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	-1,9059	0,0820	(-2,0670; -1,7448)	-23,23	0,000	
age	0,02587	0,00366	(0,01867; 0,03306)	7,06	0,000	2,92
ht	0,04178	0,00179	(0,03826; 0,04530)	23,30	0,000	2,99
sex	0,0290	0,0120	(0,0053; 0,0526)	2,41	0,016	1,05

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC
0,142516	81,63%	81,54%	12,0567	81,36%	-617,50	-595,71

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Seq MS	F-Value	P-Value
Regression	3	52,8000	81,63%	52,8000	17,6000	866,53	0,000
age	1	40,6713	62,88%	1,0122	40,6713	2002,44	0,000
ht	1	12,0111	18,57%	11,0290	12,0111	591,37	0,000
sex	1	0,1175	0,18%	0,1175	0,1175	5,79	0,016
Error	585	11,8819	18,37%	11,8819	0,0203		
Lack-of-Fit	309	6,3964	9,89%	6,3964	0,0207	1,04	0,365
Pure Error	276	5,4854	8,48%	5,4854	0,0199		
Total	588	64,6818	100,00%				

Tests use the sequential sums of squares

12. Στην παραπάνω έρευνα έλαβαν μέρος
    a. 584 άτομα
    b. 585 άτομα
    c. 589 άτομα
    d. 588 άτομα

13. Από το output μπορούμε να πουμε πως προσεγγιστικά ισχύει
    a. $P(F > 2,41) = 0,016$, όπου $F \sim F_{1,585}$.
    b. $P(T > 2,41) = 0,008$, όπου $T \sim t_{585}$.
    c. $P(T > 2,41) = 0,016$, όπου $T \sim t_{585}$.
    d. $P(F > 5,79) = 0,008$, όπου $F \sim F_{1,585}$.

14. Βάσει του μοντέλου ισχύει πως
    a. Με 95% εμπιστοσύνη τα αγόρια είναι κατά μέσο όρο από 0,03826 ως 0,04530 ίντσες ψηλότερα από συνομίλικα κορίτσια.
    b. Με 95% εμπιστοσύνη τα αγόρια έχουν κατά μέσο όρο από 0,0053 ως 0,0526 ψηλότερη τιμή ln(FEV) από συνομίλικα κορίτσια.
    c. Με 95% εμπιστοσύνη τα αγόρια έχουν κατά μέσο όρο από 0,0053 ως 0,0526 ψηλότερη τιμή ln(FEV) από κορίτσια του ίδιου ύψους.
    d. Τίποτα από τα παραπάνω

15. Η τιμή της αμερόληπτης εκτιμήτριας της διασποράς των σφαλμάτων είναι
    a. 0,142516
    b. 0,0203
    c. 0,1175
    d. -1,9059

16. Θεωρήστε το μοντέλο $E(ln(FEV)) = \beta_0 + \beta_1 \times age + \beta_2 \times ht$, και την μηδενική υπόθεση $H_0 : \beta_2 = 0$. Η τιμή του $p-value$ του δίπλευρου ελέγχου της H_0 ισούται με
    a. $P(|T| > 23,30)$, όπου $T \sim t_{585}$.
    b. $P(T > 23,30)$, όπου $T \sim t_{585}$.
    c. $P(F > 591,37)$, όπου $F \sim F_{1,586}$.
    d. $P(F > 11,0290)$, όπου $F \sim F_{1,585}$.

17. Θεωρήσετε το μοντέλο $E(ln(FEV)) = \beta_0 + \beta_1 \times age + \beta_2 \times ht$. Μια αμερόληπτη εκτιμήτρια του β_1 ισούται με
    a. 0,029
    b. 0,04178
    c. 0,02587

d. Τίποτα από τα παραπάνω

18. Θεωρήστε το πλήρες μοντέλο $E(ln(FEV)) = \beta_0 + \beta_1 \times age + \beta_2 \times ht + \beta_3 \times sex$, και την μηδενική υπόθεση $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Η τιμή του $p - value$ του ελέγχου της H_0 ισούται με
    a. $P(|T| > 23,23)$, όπου $T \sim t_{585}$.
    b. $P(F > 866,53)$, όπου $F \sim F_{1,585}$.
    c. $P(F > 866,53)$, όπου $F \sim F_{3,585}$.
    d. $P(F > 866,53)$, όπου $F \sim F_{2,586}$.

19. Σε ποιό από τα παρακάτω μοντέλα <u>ΔΕΝ</u> μπορείτε να εφαρμόσετε τις μεθόδους γραμμικής παλινδρόμησης που διδαχτήκατε στο μάθημα για την εκτίμηση των παραμέτρων
    a. $Y_i = \beta_0 + \beta_1 x_i^2 + \epsilon_i,\ \epsilon_1, \ldots, \epsilon_n \sim^{iid} N(0, \sigma^2)$.
    b. $Y_i = e^{\beta_0} e^{\beta_1 x_i + \epsilon_i},\ \epsilon_1, \ldots, \epsilon_n \sim^{iid} N(0, \sigma^2)$.
    c. $Y_i = \frac{1}{e^{\beta_0 + \beta_1 x_i}} + \epsilon_i,\ \epsilon_1, \ldots, \epsilon_n \sim^{iid} N(0, \sigma^2)$.
    d. $Y_i = e^{\beta_0} e^{\beta_1 x_i^2 + \epsilon_i},\ \epsilon_1, \ldots, \epsilon_n \sim^{iid} N(0, \sigma^2)$.

20. Αν Z_α, $\chi^2_{n;\alpha}$, $t_{n;\alpha}$ και $F_{n_1,n_2;\alpha}$ είναι τα α-άνω ποσοστιαία σημεία των Z, χ^2_n, t_n και F_{n_1,n_2} κατανομών αντίστοιχα ($\alpha < 0.5$), τότε τι από τα παρακάτω <u>ΔΕΝ</u> ισχύει
    a. $F_{n_1,n_2;\alpha} \to \chi^2_{n_2;\alpha}$, όταν $n_1 \to \infty$.
    b. $F_{n_1,n_2;\alpha} \to \frac{1}{n_1}\chi^2_{n_1;\alpha}$, όταν $n_2 \to \infty$.
    c. $t_{n;\alpha} \to Z_\alpha$ όταν $n \to \infty$.
    d. $t_{n;\alpha} > t_{n+1;\alpha}\ \forall n$.

Σωστές Απαντήσεις:

1. c
2. d
3. b
4. d
5. a

6. b
7. a
8. a
9. c
10. c
11. a

12. c
13. b
14. d
15. b
16. c
17. d
18. c

19. c
20. a