

ΑΣΚΗΣΕΙΣ ΣΤΗΝ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

ΑΣΚΗΣΗ 1. Εστω το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i=1, \dots, n$, με τα σφάλματα $\varepsilon_i \sim N(0, \sigma^2)$ και ασυσχέτιστα ανά δύο. Ναδειχθεί ότι οι εκτιμητές ελαχίστων τετραγώνων των β_0 και β_1 του μοντέλου αυτού, συμπίπτουν με τους εκτιμητές μέγιστης πιθανοφάνειάς τους. Να προσδιορισθεί ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου σ^2 .

ΑΣΚΗΣΗ 2. Εστω το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i=1, \dots, n$, με $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$ και τα $\varepsilon_i, i=1, \dots, n$, είναι ασυσχέτιστα. Αν $\hat{\beta}_0$ και $\hat{\beta}_1$ οι εκτιμητές ελαχίστων τετραγώνων των β_0 και β_1 , να δειχθούν τα ακόλουθα:

α) $E(\hat{\beta}_0) = \beta_0, E(\hat{\beta}_1) = \beta_1, Var(\hat{\beta}_0) = \sigma^2 \left\{ \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\} = \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2}, Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$.

Λέχει γίνει σε ένα από τα προηγούμενα μαθήματα.

β) $\sum_{i=1}^n X_i \varepsilon_i = 0$, με $\varepsilon_i = Y_i - \hat{Y}_i, i=1, \dots, n$, τα υπόλοιπα.

γ) $Cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2}$.

ΑΣΚΗΣΗ 3. Τα παρακάτω δεδομένα αναφέρονται στην επίδραση της θερμοκρασίας στην απόδοση μιας χημικής αντίδρασης.

Θερμοκρασία	-5	-4	-3	-2	-1	0	1	2	3	4	5
Απόδοση	1	5	4	7	10	8	9	13	14	13	18

α) Αφού ορισθεί η εξαρτημένη και η ανεξάρτητη μεταβλητή να βρεθούν οι εκτιμητές ελαχίστων τετραγώνων των παραμέτρων του μοντέλου.

β) Να κατασκευαστεί ο πίνακας ANADIA και να γίνει ένα αρχικό τεστ για τον έλεγχο της υπόθεσης $H_0: \beta_1 = 0$.

γ) Κάνοντας τις κατάλληλες υποθέσεις

ι) Να ελεγχθεί η υπόθεση $H_0: \beta_1 = 0$ σε επίπεδο σημαντικότητας $\alpha = 0.05$.

ii) Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για το β_1 .

ΑΣΚΗΣΗ 4. Για το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i=1, \dots, n$, να δειχθεί ότι:

α) $r_{\hat{X}, Y}^2 = R^2$, με $r_{X, Y}$ το δειγματικό συντελεστή συσχέτισης του Pearson και R^2 το συντελεστή προσδιορισμού.

β) Αν $r_{Y, \hat{Y}}$ είναι ο δειγματικός συντελεστής συσχέτισης του Pearson μεταξύ της παρατηρούμενης και της εκτιμώμενης τιμής του Y , να δειχθεί ότι $r_{Y, \hat{Y}}^2 = R^2$.

ΑΣΚΗΣΗ 5. Τα παρακάτω δεδομένα συγκεντρώθηκαν σε μια μελέτη που σκοπό είχε να εξετασθεί ποιοι παράγοντες επιδρούν στο βάρος ενός νεογέννητου μωρού.

Βάρος κατά τη γέννηση (gr)	1361	1588	1815	2087	2268	2404	3402	3629	3765	4083
Κοινωνικο-οικονομική κατάσταση	8	7	4	5	5	4	3	3	2	1
Σειρά γεννήσεως	4	3	4	3	2	2	2	1	1	1

- ι) Να ορισθεί η εξαρτημένη και οι ανεξάρτητες μεταβλητές.
 ιι) Να βρεθούν οι πίνακες $X'X$, $(X'X)^{-1}$, $X'Y$.
 ιιι) Να βρεθούν οι εκτιμητές ελαχίστων τετραγώνων του μοντέλου $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$, και να δοθεί ο πίνακας διακυμάνσεων-συνδιακυμάνσεών τους. Να υπολογισθεί ο δειγματικός συντελεστής συσχέτισης μεταξύ των $\hat{\beta}_1$ και $\hat{\beta}_2$.
 ιιiv) Να δοθεί ο πίνακας ANADIA και να βρεθεί ένας ημερόληπτος εκτιμητής για τη διακύμανση σ^2 των σφαλμάτων του μοντέλου.
 v) Υποθέτοντας ότι τα σφάλματα ακολουθούν κανονική κατανομή να ελεγχθούν οι υποθέσεις,
 α) $\beta_1 = \beta_2 = 0$, β) $\beta_0 = 0$, γ) $\beta_1 = 0$, δ) $\beta_2 = 0$.
 vi) Να δοθεί η φυσική ερμηνεία του εκτιμώμενου μοντέλου.

ΑΣΚΗΣΗ 6. Το γραφείο υποδοχής ασθενών ενός νοσοκομείου, θέλοντας να εξετάσει ποιοι παράγοντες μπορεί να επιδρούν στη διάρκεια παραμονής των ασθενών, που υποβάλλονται σε κάποιο είδος εγχείρησης στο νοσοκομείο, συγκέντρωσε τα παρακάτω δεδομένα για 20 τέτοιους ασθενείς.

A/A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ημέρες παραμονής Μετά την εγχείρηση	6	6	11	9	16	16	4	8	11	13	13	9	17	17	12
Αριθ. Ιατρικών Προβλημάτων	1	2	2	1	3	1	1	3	2	3	1	1	3	2	4
Ημέρες Παραμονής Πρίν την εγχείρηση	1	1	2	3	3	5	1	1	2	2	4	2	3	4	1
	16	17	18	19	20										
	6	5	12	8	9										
	1	1	3	1	2										
	1	1	2	2	2										

Κάνοντας τις κατάλληλες υποθέσεις να αναλυθούν τα δεδομένα και να διατυπωθούν τα συμπεράσματα.

ΑΣΚΗΣΗ 7. Εστω Y_1 και Y_2 δύο ανεξάρτητες τυχαίες μεταβλητές με μέσες τιμές θ και 2θ αντίστοιχα, με θ πραγματικό αριθμό. Χρησιμοποιώντας το μοντέλο της γραμμικής παλινδρόμησης να προσδιορισθεί ο εκτιμητής ελαχίστων τετραγώνων της παραμέτρου θ και να υπολογισθεί το άθροισμα τετραγώνων των υπολοίπων.

ΑΣΚΗΣΗ 8. Για το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης να δειχθεί ότι

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta).$$

ΑΣΚΗΣΗ 9. Αν σε ένα γραμμικό μοντέλο υπάρχει σταθερός όρος, τότε το άθροισμα των υπολοίπων ισούται με το μηδέν.

ΑΣΚΗΣΗ 10. Θεωρούμε το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης $Y = X\beta + \varepsilon$, με $Rank(X) = p+1$. Να δειχθεί ότι,

$$\alpha) \sum_{i=1}^n \hat{Y}_i(Y_i - \hat{Y}_i) = 0, \text{ και } \beta) \sum_{i=1}^n Var(\hat{Y}_i) = \sigma^2(p+1).$$

ΑΣΚΗΣΗ 11. Για το μοντέλο $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$, $i=1, \dots, n$, με ε_i ανεξάρτητες $N(0, \sigma^2)$, δίνονται οι κανονικές εξισώσεις:

$$10\hat{\beta}_0 + 2\hat{\beta}_1 - 6\hat{\beta}_2 = 4$$

$$2\hat{\beta}_0 + 2\hat{\beta}_1 = 6$$

$$-6\hat{\beta}_0 + 5\hat{\beta}_2 = 7.$$

α) Αν $n=10$ και $\sum_{i=1}^n Y_i^2 = 107$, να υπολογισθούν ο εκτιμητής $\hat{\beta}$ και το MS_{res} .

β) Να ελεγχθεί με τη βοήθεια ενός F -τεστ η μηδενική υπόθεση $H_0: \beta_1 = 2\beta_2$, σε επίπεδο $\alpha=5\%$.

γ) Να κατασκευαστεί ένα t -τεστ για τον έλεγχο της υπόθεσης $H_0: \beta_1 = 2\beta_2$, σε επίπεδο $\alpha=5\%$.

ΑΣΚΗΣΗ 12. Έστω το μοντέλο της απλής γραμμικής παλινδρόμησης $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, με $\sigma^2 = Var(\varepsilon_i)$, $\nu_i = E(Y_i)$, και $q_i = (Y_i - \hat{Y}_i) - (\nu_i - E(\hat{Y}_i))$, $i=1, \dots, n$. Να δειχθεί ότι

$$E\left(\sum_{i=1}^n q_i^2\right) = (n-2)\sigma^2.$$

ΑΣΚΗΣΗ 13. Έστω το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης $Y = X\beta + \varepsilon$, με $Rank(X) = p+1$.

α) Να δειχθεί ότι, $SS_{reg} = \hat{\beta}' X' Y - n\bar{Y}^2$.

β) Υπό την υπόθεση ότι το διάνυσμα των σφαλμάτων $\varepsilon \sim N_n(0, \sigma^2 I_n)$, να δειχθεί ότι οι εκτιμητές ελαχίστων τετραγώνων $\hat{\beta}$ των παραμέτρων β του μοντέλου, ταυτίζονται με τους εκτιμητές μέγιστης πιθανοφάνειας.

γ) Υπό την υπόθεση $\varepsilon \sim N_n(0, \sigma^2 I_n)$ να δειχθεί ότι $\frac{SS_{res}}{\sigma^2} \sim \chi_{n-p-1}^2$.

ΑΣΚΗΣΗ 14. Έστω το μοντέλο παλινδρόμησης

$$Y_i = \beta X_i + \varepsilon_i, \quad i=1, \dots, n,$$

για το οποίο ισχύει ότι τα σφάλματα ε_i ακολουθούν κανονική κατανομή με $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2 X_i$ και $Cov(\varepsilon_i, \varepsilon_j) = 0$, για $i, j=1, \dots, n$, $i \neq j$.

α) Να βρεθεί ο εκτιμητής μέγιστης πιθανοφάνειας (ή ο εκτιμητής ελαχίστων τετραγώνων) και να δειχθεί ότι είναι αμερόληπτος εκτιμητής του β .

β) Να δείξετε ότι οι $\beta^* = \frac{Y_1 + Y_2}{X_1 + X_2}$, $\beta^{**} = \frac{Y_n - Y_1}{X_n - X_1}$ και $\beta^{***} = \frac{X_1 Y_1 + X_2 Y_2}{X_1^2 + X_2^2}$ είναι επίσης αμερόληπτοι εκτιμητές του β .

γ) Γιατί οι εκτιμητές του προηγούμενου ερωτήματος δεν είναι αποτελεσματικοί;

ΑΣΚΗΣΗ 15. Για το μοντέλο της απλής γραμμικής παλινδρόμησης

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

να δείξετε ότι

- i) $\sum_{i=1}^n \hat{Y}_i = \sum_{i=1}^n Y_i$, και $\sum_{i=1}^n e_i = 0$, ii) $\bar{\hat{Y}} = \bar{Y}$, iii) $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$, iv) $\sum_{i=1}^n \hat{Y}_i X_i = \sum_{i=1}^n Y_i X_i$, και $\sum_{i=1}^n e_i \hat{Y}_i = 0$,
v) $\sum_{i=1}^n e_i X_i = 0$.

ΑΣΚΗΣΗ 16. Έστω ότι $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ είναι οι εκτιμητές ελαχίστων τετραγώνων του μοντέλου πολλαπλής παλινδρόμησης

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \dots, n.$$

Να βρεθούν συναρτήσεις των $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ οι εκτιμητές ελαχίστων τετραγώνων του μοντέλου

$$Y_i = \beta_0 + \beta_1 Z_{i1} + \dots + \beta_p Z_{ip} + \varepsilon_i, \quad i = 1, \dots, n,$$

όπου $Z_{ij} = c_j X_{ij}$, $j = 1, \dots, p$, με c_j , $j = 1, \dots, p$, σταθερές και $i = 1, \dots, n$.

ΑΣΚΗΣΗ 17. Για το μοντέλο της απλής γραμμικής παλινδρόμησης

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

να δείξετε ότι οι εκτιμητές ελαχίστων τετραγώνων των β_0, β_1 είναι ασυσχέτιστοι αν $\sum_{i=1}^n X_i = 0$.

ΑΣΚΗΣΗ 18. Έστω το μοντέλο παλινδρόμησης

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

για το οποίο ισχύει ότι τα σφάλματα ε_i ακολουθούν κανονική κατανομή με

$$E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \frac{\sigma^2}{X_i} \quad \text{και} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad \text{για} \quad i, j = 1, \dots, n, \quad i \neq j.$$

Να βρεθεί ο γραμμικός αμερόληπτος εκτιμητής του β με την μικρότερη διακύμανση.

ΑΣΚΗΣΗ 19. Για το κλασικό μοντέλο απλής γραμμικής παλινδρόμησης

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

να βρείτε τις συνθήκες που πρέπει να ικανοποιούνται έτσι ώστε ένας γραμμικός εκτιμητής του β_1 να είναι αμερόληπτος. Πότε αυτός είναι ο BLUE;

(Υπενθύμιση: Ένας εκτιμητής είναι γραμμικός όταν είναι της μορφής $\sum_{i=1}^n d_i Y_i$).

ΑΣΚΗΣΗ 20. Στο μοντέλο της πολλαπλής γραμμικής παλινδρόμησης $Y = X\beta + \varepsilon$, με $\text{Rank}(X) = p + 1$, τα μαθηματικοποιημένα υπόλοιπα ορίζονται

$$t_i = \frac{e_i}{\sqrt{MS_{res} \sqrt{1 - \rho_{ii}}}}, \quad i = 1, \dots, n,$$

με ρ_{ij} να συμβολίζει το στοιχείο της (i, j) -θέσης του πίνακα $P = X(X'X)^{-1}X'$. Να δειχθεί ότι

$$t_i \sim t_{n-p-1}.$$

ΑΣΚΗΣΗ 21. Έστω οι παρατηρήσεις $Y_1 = \alpha_1 + \varepsilon_1$, $Y_2 = 2\alpha_1 - \alpha_2 + \varepsilon_2$ και $Y_3 = \alpha_1 + 2\alpha_2 + \varepsilon_3$, με $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N_3(\mathbf{0}, \sigma^2 \mathbf{I}_3)$. Με τη θεωρία των γραμμικών μοντέλων να ελεγχθεί η υπόθεση $H_0: \alpha_1 = \alpha_2$.