

ΕΞΕΤΑΣΗ ΕΡΓΑΣΤΗΡΙΟΥ ΓΡΑΜΜΙΚΩΝ ΜΟΝΤΕΛΩΝ 2025-2026

ΔΙΑΡΚΕΙΑ ΕΞΕΤΑΣΗΣ: 3 ΩΡΕΣ

Τα δεδομένα που περιέχονται στο αρχείο `project.sav` προέρχονται από μία έρευνα που πραγματοποιήθηκε στις περιοχές του Λαυρίου και της Ελευσίνας σχετικά με τις επιπτώσεις της έκθεσης στο μόλυβδο στη σωματική ανάπτυξη παιδιών ηλικίας 5 έως 10 ετών. Η εξαρτημένη μεταβλητή είναι η περίμετρος της κεφαλής των παιδιών μετρημένη σε cm (head). Η ανεξάρτητη μεταβλητή lead δίνει τη συγκέντρωση του μολύβδου στο αίμα των παιδιών σε $\mu\text{g/dL}$. Άλλες επεξηγηματικές μεταβλητές είναι οι:

Age: ηλικία παιδιού

Sex: φύλο παιδιού (0: κορίτσι, 1: αγόρι)

Town: περιοχή κατοικίας (0: Ελευσίνα, 1: Λαύριο)

Hemoglobin: επίπεδα αιμοσφαιρίνης στο αίμα του παιδιού σε $\mu\text{g/dL}$

Iron: επίπεδα σιδήρου στο αίμα του παιδιού σε $\mu\text{g/dL}$

Protein: ποσοστό συγκέντρωσης πρωτεϊνών στον οργανισμό του παιδιού

Motherheight: ύψος μητέρας σε m

Motherweight: βάρος μητέρας σε kg

Motherage: ηλικία μητέρας

Fatherheight: ύψος πατέρα σε m

Fatherweight: βάρος πατέρα σε kg

Fatherage: ηλικία πατέρα

Απαντήστε τα παρακάτω ερωτήματα χρησιμοποιώντας το στατιστικό πρόγραμμα R για τις απαιτούμενες αναλύσεις. Παρουσιάστε συνοπτικές απαντήσεις και σχόλια σε ένα αρχείο Word, χρησιμοποιώντας όποια αποτελέσματα από το output της R κρίνετε σκόπιμο, συμπεριλαμβανομένων πινάκων και γραφημάτων. Το αρχείο Word με τις απαντήσεις σας, καθώς και το αρχείο της R με τον κώδικα που χρησιμοποιήθηκε, θα παραδοθούν στο τέλος της εξέτασης μέσω της eclass.

Ερωτήματα

Ερώτημα 1

Υπολογίστε μονοδιάστατα περιγραφικά στατιστικά μέτρα για όλες τις συνεχείς μεταβλητές που θα χρησιμοποιήσετε, καθώς και τις συσχετίσεις μεταξύ τους. Φτιάξτε ιστογράμματα που

να δείχνουν την κανονικότητα ή όχι αυτών των μεταβλητών. Σχολιάστε.

Ερώτημα 2

Φτιάξτε σημειογράμματα που να δείχνουν την ύπαρξη ή μη γραμμικής σχέσης ανάμεσα στην εξαρτημένη μεταβλητή και στις συνεχείς ανεξάρτητες. Χρησιμοποιήσετε ως μεταβλητή ομαδοποίησης την town. Ποια τα συμπεράσματά σας;

Ερώτημα 3

Να διερευνηθεί γραφικά και περιγραφικά εάν διαφοροποιείται η περίμετρος της κεφαλής του παιδιού ανάλογα με την πόλη. Να γίνει ο κατάλληλος έλεγχος υποθέσεων και να υπολογιστεί το αντίστοιχο p-value, αφού πρώτα ελέγξετε αν ισχύουν οι απαιτούμενες υποθέσεις για τον έλεγχο. Τι συμπεράσματα βγάζετε;

Ερώτημα 4

Να προσαρμοστούν τα παρακάτω μοντέλα πολλαπλής γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή τη head:

$$M_1 : \quad \text{head} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{town} + \varepsilon$$

$$M_2 : \quad \text{head} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{town} + \beta_4 \text{lead} + \varepsilon$$

$$M_3 : \quad \text{head} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{town} + \beta_4 \text{lead} + \beta_5 \text{motherheight} + \beta_6 \text{fatherheight} + \varepsilon$$

Για κάθε μοντέλο να παρουσιαστούν:

- (α) Το R^2 .
- (β) Το προσαρμοσμένο R^2 .
- (γ) Ο συνολικός έλεγχος σημαντικότητας του μοντέλου.
- (δ) Οι στατιστικά σημαντικές μεταβλητές σε επίπεδο σημαντικότητας 5%.
- (ε) Σύντομη σύγκριση των τριών μοντέλων με βάση το R^2 και το προσαρμοσμένο R^2 .
- (στ) Να επιλεγεί το καταλληλότερο μοντέλο και να αιτιολογηθεί η επιλογή.

Ερώτημα 5

Με βάση το μοντέλο που επιλέχθηκε στο προηγούμενο ερώτημα:

- (α) Να ερμηνευθούν οι συντελεστές των επεξηγηματικών μεταβλητών του επιλεγμένου μοντέλου.
- (β) Να προβλεφθεί η περίμετρος κεφαλής για παιδί με τα εξής χαρακτηριστικά:

sex = 1, age = 8, town = 1, lead = 6.87, motherheight = 1.61, fatherheight = 1.78.

- (γ) Να δοθεί 95% διάστημα εμπιστοσύνης για τη μέση προβλεπόμενη τιμή.

Ερώτημα 6

Για το μοντέλο που επιλέχθηκε στο ερώτημα 4:

- (α) Να κατασκευαστεί ιστόγραμμα των τυποποιημένων καταλοίπων.
- (β) Να κατασκευαστεί Q-Q plot των τυποποιημένων καταλοίπων.
- (γ) Να κατασκευαστεί διάγραμμα καταλοίπων ως προς τις προβλεπόμενες τιμές.
- (δ) Να σχολιαστεί αν φαίνεται να ικανοποιούνται οι βασικές υποθέσεις της γραμμικής παλινδρόμησης: κανονικότητα, γραμμικότητα, ομοσκεδαστικότητα και ανεξαρτησία.

Ερώτημα 7 (Bonus)

Προσαρμόστε το μοντέλο πολλαπλής παλινδρόμησης που να περιλαμβάνει τις μεταβλητές age, sex, town, lead, motherheight, fatherheight, καθώς και το γινόμενο της ψευδομεταβλητής town με τη μεταβλητή lead. Παρουσιάστε αποτελέσματα για το R^2 και το προσαρμοσμένο R^2 , την ανάλυση διασποράς και τον συνολικό έλεγχο για την ύπαρξη γραμμικής σχέσης, καθώς και για τους επιμέρους ελέγχους στατιστικής σημαντικότητας των συντελεστών αυτού του μοντέλου. Τι εκφράζει το γινόμενο town*lead στο μοντέλο που προσαρμόσατε? Σχολιάστε.

Ερώτημα 8

Προσαρμόστε ένα μοντέλο ANOVA με δύο παράγοντες, τις μεταβλητές sex και town, και απαντητική μεταβλητή την head. Ποιο είναι το πιο πλούσιο μοντέλο ANOVA που μπορεί να προσαρμοστεί σε αυτά τα δεδομένα; Ποιο είναι το καταλληλότερο μοντέλο ANOVA για αυτά τα δεδομένα; Σε ποιο μοντέλο παλινδρόμησης αντιστοιχεί το μοντέλο ANOVA στο οποίο καταλήξατε και πώς συγκρίνεται με τα μοντέλα που προσαρμόσατε στο ερώτημα 4 ως προς το προσαρμοσμένο R^2 ;