

Ανάλυση Παλινδρόμησης Εργασία 2025-26

1. Στο αρχείο `drugtrial.dta` έχει καταγραφεί το σκορ της επίδρασης διαφορετικής δόσης ενός φαρμάκου σε 24 άνδρες και γυναίκες.

α) Δώστε περιγραφικά ανά φύλο και δόση φαρμάκου. Επηρεάζει η δόση το σκορ; Το σκορ διαφέρει ανά φύλο; Δώστε τα σχετικά μοντέλα γραμμικής παλινδρόμησης και ερμηνεύστε τους συντελεστές.

β) Εφαρμόστε κατάλληλο μοντέλο γραμμικής παλινδρόμησης. Δώστε την εξίσωση του μοντέλου με τις κύριες επιδράσεις και την αλληλεπίδραση των δυο ανεξάρτητων μεταβλητών. Πώς ερμηνεύονται οι συντελεστές;

γ) Υπολογίστε το αναμενόμενο σκορ ανά δόση σε άνδρες και γυναίκες.

2. Στο αρχείο `goats.dta` περιλαμβάνονται δεδομένα από ένα τυχαίο δείγμα από 40 κατσίκες, σε 20 από τις οποίες ακολουθήθηκε μια πιο εντατική διαίτα (treatment) την οποία θέλουμε να αξιολογήσουμε ως προς την αύξηση βάρους (weightgain).

α) Δώστε περιγραφικά στατιστικά ανά διαίτα. Η διαφορά που παρατηρείται στην αύξηση βάρους είναι στατιστικά σημαντική; Δώστε το σχετικό μοντέλο γραμμικής παλινδρόμησης και ερμηνεύστε τους συντελεστές.

β) Το αρχικό βάρος του ζώου (`initialwt`) σχετίζεται με το βάρος που πήρε το κάθε ζώο; Περιγράψτε τη σχέση και δώστε το σχετικό μοντέλο γραμμικής παλινδρόμησης (ερμηνεύστε τους συντελεστές).

γ) Αξιολογήστε την επίδραση της διαίτας λαμβάνοντας υπόψη πιθανή συσχέτιση της πρόσληψης βάρους με το αρχικό βάρος. Δώστε το σχετικό μερικό F-test.

δ) Ελέγξτε γραφικά για πιθανή αλληλεπίδραση. Δώστε την εξίσωση του μοντέλου με τις κύριες επιδράσεις και την αλληλεπίδραση των δυο ανεξάρτητων μεταβλητών. Πώς ερμηνεύονται οι συντελεστές; Αν είναι εφικτό, δώστε τη σταθμισμένη (για αρχικό βάρος) επίδραση της διαίτας στην πρόσληψη βάρους.

3. Θέλουμε να μελετήσουμε ποιοι από τους καταγεγραμμένους παράγοντες επηρεάζουν το σκορ ικανοποίησης για τη ζωή (**lifesat**) των 20 συμμετεχόντων, επτά χρόνια μετά το κολέγιο (αρχείο `lifesat.dta`).

- ι. Ελέγξτε όλα τα μονοπαραγοντικά μοντέλα. Δώστε το στικτόγραμμα με την ευθεία ελαχίστων τετραγώνων για το πιο στατ. σημαντικό μονοπαραγοντικό μοντέλο. Δώστε το μοντέλο που περιλαμβάνει όλες τις μεταβλητές που ήταν στατιστικά σημαντικές στα μονοπαραγοντικά σε επίπεδο $\alpha=10\%$. Σχολιάστε το πολυπαραγοντικό μοντέλο.

- ii. Με βάση τη διαδικασία «forward» ποια μεταβλητή επιλέγεται στο πρώτο βήμα; Ποια στο δεύτερο; Δώστε το κατάλληλο F-test σε κάθε βήμα και συγκρίνετε το με το 95^ο εκατοστημόριο της F-κατανομής.
- iii. Εφαρμόστε την «forward» διαδικασία στο Stata/R. Εξηγήστε τι γίνεται σε κάθε βήμα και περιγράψτε το μοντέλο που επελέγη.
- iv. Εφαρμόστε την «backward» διαδικασία στο Stata/R. Εξηγήστε τι γίνεται σε κάθε βήμα, περιγράψτε το τελικό μοντέλο και συγκρίνετε τα αποτελέσματα με αυτά της διαδικασίας «forward».
- v. Εφαρμόστε την «stepwise backward» διαδικασία στο Stata/R (με p-value εξόδου 10% and p-value εισόδου 5% για το Stata). Εξηγήστε το επιπλέον βήμα της διαδικασίας αυτής και σχολιάστε τα αποτελέσματα.
- vi. Για τα μοντέλα από τα ερωτήματα i, iii, iv και v υπολογίστε το R^2_{adj} , το μέσο τετραγωνικό σφάλμα, την Mallows' Cp statistic και το συνολικό p-value του μοντέλου. Ποιο μοντέλο επιλέγετε;
- vii. Να ελεγχθεί αν πληρούνται οι προϋποθέσεις του μοντέλου που επιλέξατε καθώς και η εφαρμογή του.
- viii. Να δοθεί η εξίσωση του τελικού μοντέλου. Να ερμηνευθούν οι συντελεστές του μοντέλου. Ποιο είναι το αναμενόμενο σκορ ικανοποίησης για τη ζωή του και το σχετικό 95% διάστημα πρόβλεψης για το άτομο 21;