

## ΕΦΑΡΜΟΣΜΕΝΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

### ΕΡΓΑΣΙΑ 2

**Οδηγίες για την παρουσίαση των αποτελεσμάτων:** Γράψτε ένα σύντομο κείμενο εξηγώντας τα ευρήματα και τα συμπεράσματά σας, σε γλώσσα που θα καταλάβαινε κάποιος που δεν έχει απαραίτητα εξειδικευμένες γνώσεις στατιστικής πέρα από τις κλασσικές έννοιες. Μπορείτε αν θέλετε εδώ να περιλάβετε και κάποια διαγράμματα που θεωρείτε σημαντικά. Στην εργασία που θα παραδώσετε παρουσιάστε αυτό το κομμάτι ως περίληψη στην αρχή κάθε ενότητας/προβλήματος, και τις υπόλοιπες απαντήσεις ανά ερώτημα μαζί με τις στατιστικές λεπτομέρειες και τα διαγράμματα σε παράρτημα.

Το αρχείο apt.Rdata περιέχει δεδομένα από πωλήσεις 120 διαμερισμάτων σχετικά με την τιμή πώλησης (σε 100.000 ευρώ), το μέγεθος (τ.μ.) και την πόλη.

Να τρέξετε το μοντέλο πολλαπλής παλινδρόμησης με εξαρτημένη μεταβλητή την price και ανεξάρτητες τις city και size, μαζί με όρους αλληλεπίδρασης. Για το μοντέλο αυτό απαντήστε στα παρακάτω ερωτήματα.

1. Ποια είναι η εξίσωση παλινδρόμησης για κάθε πόλη;
2. Ερμηνεύστε τους συντελεστές για τις κύριες επιδράσεις του μεγέθους.
3. Ερμηνεύστε τους συντελεστές για τις κύριες επιδράσεις της πόλης.
4. Ερμηνεύστε τους συντελεστές για τις αλληλεπιδράσεις.
5. Ποιοι συντελεστές δεν είναι στατιστικά σημαντικοί στο επίπεδο 5%; Με βάση αυτό, ποια είναι η μορφή των εξισώσεων παλινδρόμησης για κάθε τύπο εργασίας; Παραστήστε τις γραφικά.
6. Κάντε ένα έλεγχο για την υπόθεση ότι τα διαμερίσματα στην Πάτρα και τη Θεσσαλονίκη έχουν την ίδια εξίσωση παλινδρόμησης τιμής προς μέγεθος.
7. Να τρέξετε ένα νέο μοντέλο παλινδρόμησης που περιλαμβάνει και όρους δεύτερης τάξης ως προς το μέγεθος (και τους αντίστοιχους όρους παλινδρόμησης). Να παραστήσετε γραφικά τις 4 δευτεροβάθμιες εξισώσεις παλινδρόμησης που προκύπτουν (τιμή προς μέγεθος για κάθε πόλη) στο ίδιο διάγραμμα με διαφορετικό χρώμα την καθεμία.
8. Για το μοντέλο των ερωτήσεων 1-6 να γίνει ανάλυση καταλοίπων
  - a. Διαγράμματα studentized και jackknife residuals ως προς τις προβλεπόμενες τιμές και έλεγχος outliers
  - b. Διαγράμματα leverage και εντοπισμός πιθανών influential points
  - c. Διάγραμμα cook distance
  - d. Influential analysis
  - e. Διάγραμμα κανονικότητας καταλοίπων.