

Απλή Γραμμική Παλινδρόμηση στην R

Εργασία 2, 18/19 Απριλίου 2019

Ο σύλλογος [Air Breizh](#) στην περιοχή της Βρεττανίας της Βορειοδυτικής Γαλλίας είναι ένας από τους 36 εγκεκριμένους (από το Υπουργείο Περιβάλλοντος της Γαλλίας) οργανισμούς επιθεώρησης της ποιότητας του αέρα, ο οποίος είναι υπεύθυνος για την περιοχή της Βρεττανίας. Ο σύλλογος αυτός μετρά από το 1994 τη μέγιστη ημερήσια συγκέντρωση σε O_3 ($\mu\text{g}/\text{ml}$), όπως επίσης συλλέγει και κάποια άλλα μετεωρολογικά δεδομένα. Θα δούμε πώς με τη βοήθεια κάποιων δεδομένων από αυτά, είναι δυνατόν να φτιάξουμε ένα μοντέλο πρόβλεψης της συγκέντρωσης του όζοντος ως συνάρτηση μιας μεταβλητής. Σε πρώτη φάση επιλέγουμε τη θερμοκρασία στις 12.00. Στην επόμενη εργασία θα φτιάξουμε ένα πιο πλήρες μοντέλο.

Προετοιμασία

- (1) Τοποθετήστε το αρχείο `ozone.txt` που βρίσκεται στην `e-class`, στον χώρο εργασίας σας στην R.
- (2) Φορτώστε τα δεδομένα, με τις επιλογές `header=T` και `row.names=1`. Κάντε `head(ozone)` για να δείτε τις πρώτες γραμμές των δεδομένων. Η εξαρτημένη μεταβλητή συμβολίζεται με `maxO3` και η ανεξάρτητη μεταβλητή `T12`.

Ερωτήματα

- (1) Κάντε αναπαράσταση των δεδομένων σε ένα γράφημα. Από το γράφημα αυτό, σας φαίνεται λογική η επιλογή μίας απλής γραμμικής παλινδρόμησης; Αιτιολογείστε.
- (2) Κάντε την παλινδρόμηση και αποθηκεύστε τα αποτελέσματα σε ένα αντικείμενο με το όνομα `reg`.
- (3) Ποιές πληροφορίες εμφανίζονται όταν καλέσουμε `reg`; Η κλάση αυτών των αντικειμένων περιέχει πολλές παραπάνω πληροφορίες: δείτε τη λίστα των χαρακτηριστικών που εμφανίζονται όταν καλέσουμε `attributes(reg)`. Εξηγήστε σύντομα 5 από αυτές.
- (4) Τι εμφανίζεται αν πληκτρολογήσουμε `summary(reg)`; Εξηγήστε αυτά που εμφανίζονται και σχολιάστε σύντομα τα αποτελέσματα με τη βοήθεια της θεωρίας και αυτών που έχουν ειπωθεί στα εργαστήρια. Επαληθεύστε την εκτίμηση της τυπικής απόκλισης των σφαλμάτων από τα δεδομένα των εκτιμώμενων σφαλμάτων. Είναι η θερμοκρασία `T12` στατιστικά σημαντική στα συνήθη επίπεδα σημαντικότητας;
- (5) Δώστε 95% διαστήματα εμπιστοσύνης των συντελεστών της παλινδρόμησης και σχολιάστε τα αποτελέσματα.
- (6) Κάντε το διάγραμμα διασποράς των δεδομένων, προσθέστε την ευθεία των ελαχίστων τετραγώνων και σχεδιάστε 95% (κατα σημείο) διαστήματα εμπιστοσύνης των μέσων αποκρίσεων της `maxO3` για τις διάφορες τιμές της `T12`. Η εντολή `matlines` μπορεί να σας είναι χρήσιμη. Δώστε μία ερμηνεία του σχήματος των καμπυλών που χαρακτηρίζουν τα όρια των διαστημάτων της εμπιστοσύνης. Τα παραπάνω διαστήματα καθορίζουν μία περιοχή που θα μπορούσαμε να πούμε ότι διαισθητικά μας δίνουν μία περιοχή εμπιστοσύνης για την ίδια την ευθεία των ελαχίστων τετραγώνων. Με ποιά έννοια πιστεύετε ότι συμβαίνει αυτό;
- (7) Κάντε ένα γράφημα, σχεδιάζοντας μία 95% περιοχή εμπιστοσύνης των συντελεστών της παλινδρόμησης. Για το σκοπό αυτό χρησιμοποιείστε τη συνάρτηση `confidenceEllipse(reg)` που ορίζεται στο πακέτο `car`. Προσθέστε στο γράφημα αυτό το ορθογώνιο που προκύπτει από τα 95% διαστήματα εμπιστοσύνης των συντελεστών της παλινδρόμησης και σχολιάστε τα αποτελέσματα.
- (8) Με τη βοήθεια της συνάρτησης `confidence.band()` που ορίζεται στο αρχείο `confidenceBand.R` κάντε το γράφημα μιας 95% περιοχής εμπιστοσύνης της ευθείας ελαχίστων τετραγώνων. Με ποιό τρόπο πιστεύετε ότι συνδέεται με το Ερώτημα 7; Σχεδιάστε

στο ίδιο γράφημα την περιοχή που καθορίσατε στο Ερώτημα 6 και συγκρίνετε τα αποτελέσματα.

- (9) Εδώ ενδιαφερόμαστε για την ποιότητα πρόβλεψης του μοντέλου. Ποια είναι η προβλεπόμενη μέγιστη συγκέντρωση του όζοντος για μία θερμοκρασία 31°C; Δώστε ένα 95% διάστημα εμπιστοσύνης της προβλεπόμενης τιμής και σχεδιάστε 95% διαστήματα εμπιστοσύνης των προβλέψεων της $maxO_3$ για τις διάφορες τιμές της T_{12} (95% διαστήματα πρόβλεψης). Σχεδιάστε στο ίδιο γράφημα τα διαστήματα εμπιστοσύνης που καθορίσατε στο Ερώτημα 6 και συγκρίνετε τα αποτελέσματα.
- (10) Κάποιος στατιστικός ισχυρίζεται ότι δεν παίζει ιδιαίτερο ρόλο αν θα επιλέξουμε ως ανεξάρτητη μεταβλητή τη θερμοκρασία στις 15.00, αντί για τις 12.00 για να εξηγήσουμε την μεταβλητή ενδιαφέροντος $maxO_3$. Εξετάστε αυτόν τον ισχυρισμό και δώστε επιχειρήματα υπέρ ή κατά.