

ΣΤΑΤΙΣΤΙΚΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗ

ΣΕΙΡΑ 3

I) Αρχείο asfalies.txt

Με βάση ένα μοντέλο της **παλινδρόμησης Poisson** να εξεταστεί η εξάρτηση του αριθμού Y αποζημιώσεων λόγω τροχαίων ατυχημάτων ανά n συμβόλαιο, από την ηλικία του ασφαλισμένου (agecat - $X_1=0$ -νέος, 1-μεγάλος), την κατηγορία ασφαλιστρών (cartype - $X_2= 1,2,3,4$) και την περιοχή διαμονής του ασφαλισμένου (district - $X_3=1$, αν Αθήνα, $X_3=0$, αν σε άλλη πόλη).

(i) Στο πρόγραμμα να δηλωθεί η μεταβλητή X_2 ως κατηγορική π.χ. μέσω της εντολής **factor(cartype)**. Να γίνουν οι στατιστικοί έλεγχοι (Wald και Deviance), χρήση του κριτηρίου AIC.

(ii) Να κατασκευαστούν διαστήματα εμπιστοσύνης για τους εκτιμημένους συντελεστές $\hat{\beta}$ του τελικού μοντέλου και να γίνουν ερμηνείες.

(iii) Να γίνουν γραφικές παραστάσεις για τα υπόλοιπα Pearson και Deviance, index plots για τα h_{ij} , τις αποστάσεις Cook, καθώς και για τα υπόλοιπα πιθανοφάνειας.

(iv) Μπορείτε να εισάγετε και μία αλληλεπίδραση στο μοντέλο π.χ. (agecat με cartype) ή (district με cartype), ή ακόμα (agecat με district), όχι όλες μαζί. Αν κάποιες βγουν στατιστικά σημαντικές, τότε μπορείτε να κάνετε και μια γραφική παράσταση π.χ. γ/n vs agecat για τις 4 κατηγορίες cartype, όπως στο «example economy data» και στο παράδειγμα με τους γιατρούς «doctors analysis», σελ. 9. Να δοθούν ερμηνείες.

II) Αρχείο leukaemia.txt

Εξαρτημένη μεταβλητή:

Ανταπόκριση στη θεραπεία (ναι=1, όχι=0)

Συμμεταβλητές:

- 1) age – ηλικία του ασθενή
- 2) smear - ποσοστό επίστρωσης βλαστοκυττάρων
- 3) infiltrate - ποσοστό κυττάρων στο μυελό των οστών
- 4) index – δείκτης κυττάρων λευχαιμίας
- 5) blasts – βλαστοκύτταρα
- 6) temperature – υψηλότερη θερμοκρασία πριν τη θεραπεία ($\times 10^\circ \text{F}$)

(Σημειώνεται ότι για κάθε περίπτωση « $n_i=1$ », αλλά αυτή η τιμή δε φαίνεται στο αρχείο.)

Ανάλυση : **Λογιστική παλινδρόμηση**

(i) Να εξεταστεί η εξάρτηση της πιθανότητας ανταπόκρισης της θεραπείας από τις συμμεταβλητές age, smear, infiltrate, index, blasts και temperature κάνοντας χρήση των στατιστικών ελέγχων Wald και Deviance καθώς και του κριτηρίου AIC.

(ii) Να γίνουν γραφικές παραστάσεις των μερικών υπολοίπων, των υπολοίπων Deviance (με την ημι-κανονική κατανομή), index plots των h_{ii} , των αποστάσεων Cook, καθώς και των υπολοίπων πιθανοφάνειας.

(iii) Να κατασκευαστούν διαστήματα εμπιστοσύνης για τους εκτιμημένους συντελεστές $\hat{\beta}$ του τελικού μοντέλου και να γίνουν ερμηνείες.

(iv) Να εξεταστεί η προβλεπτική ικανότητα του τελικού μοντέλου μέσω μιας καμπύλης ROC.

X. ΚΑΡΩΝΗ