

Στατιστικά Προγράμματα, 9-11-05, Homework #1

Το Γεωδυναμικό Ινστιτούτο του Αστεροσκοπείου Αθηνών έχει καταγράψει στοιχεία που αφορούν τους σεισμούς που έγιναν στον Ελλαδικό χώρο από το έτος 1964 έως και σήμερα. Για κάθε σεισμό καταγράφονται οι μεταβλητές:

1. MAGN (magnitude): Το μέγεθος του σεισμού στην κλίμακα Richter
2. LAT (latitude), 3. LONG (Longitude) γεωγραφικό πλάτος (σε μοίρες βόρεια από τον ισημερινό) και γεωγραφικό μήκος επικέντρου του σεισμού (σε μοίρες ανατολικά του Greenwich)
4. DEPTH: το εστιακό βάθος του σεισμού σε km
5. YEAR, 6. MONTH, 7. DAY, 8. HOUR, 9. MIN, 10. SEC: χρόνος εκδήλωσης σεισμού

Τα δεδομένα βρίσκονται σε αρχεία txt στην ιστοσελίδα του γεωδυναμικού Ινστιτούτου:

<http://www.gein.noa.gr/services/cat.html>

Ο κάθε φοιτητής θα βασιστεί στα στοιχεία των ετών $1983+2x$, $1984+2x$ όπου x είναι ο τελευταίος αριθμός του μητρώου του (προσοχή να μην χαθούν δεδομένα κατά την επεξεργασία). Δεδομένου ότι οι ασθενείς σεισμοί δεν καταγράφονται πάντοτε από τα όργανα, κρατήστε μόνο τους σεισμούς με μέγεθος ≥ 3.2 (διαγράφοντας τους ασθενέστερους) ώστε να μπορεί να θεωρηθεί ότι τα στοιχεία που παραμένουν είναι πλήρη όσον αφορά το μέγεθος.

- Σε κάποιους υπολογιστές μπορεί να χρειαστεί, πριν ανοίξετε τα αρχεία txt με το SPSS, να αντικαταστήσετε τις υποδιαστολές «.» σε «,» (π.χ. με το Wordpad χρησιμοποιώντας replace).
- Να δώσετε σωστό όνομα και τύπο σε κάθε μεταβλητή
- Προσέξτε να κρατήσετε μόνο τις πλήρεις παρατηρήσεις που αφορούν τα έτη $1983+2x$, $1984+2x$ (αν υπάρχουν και άλλες στα αρχεία να διαγραφούν)
- Η μεταβλητή Month να παίρνει τις τιμές 1,2,...,12 αντί JAN,...,DEC (τα JAN,...,DEC να δηλωθούν ως value labels)

ΕΡΩΤΗΜΑΤΑ

- 1) (i) Να δοθεί ένας πίνακας και ένα ραβδόγραμμα με το πλήθος των σεισμών ανά μέγεθος. (ii) Πόσοι σεισμοί έχουν μέγεθος ≥ 4 , πόσοι ≥ 4.5 και πόσοι ≥ 5 ;
- 2) Να απαντηθούν τα ερωτήματα του (1) ανά έτος (τα ραβδογράμματα να γίνουν σε κοινό σχήμα, stacked)
- 3) (i) Να εκτιμήσετε σημειακά και με δ.ε. 95% το μέσο μέγεθος των σεισμών. (ii) Ελέγξτε αν είναι ίσο με 3.55 (αμφίπλευρο και μονόπλευρα τεστ). (iii) Να εκτιμήσετε την τυπική απόκλιση του μεγέθους. (iv) Τι μέγεθος έχουν οι τρεις μεγαλύτεροι σεισμοί, πότε και πού καταγράφηκαν και με τι εστιακό βάθος;
- 4) Να κατασκευάσετε τρία διαγράμματα διασποράς (scatterplot) που να δείχνουν τις θέσεις των σεισμικών epicenters, το πρώτο να περιλαμβάνει σεισμούς μεγέθους ≥ 3.2 , το δεύτερο σεισμούς μεγέθους ≥ 4 , και το τρίτο σεισμούς μεγέθους ≥ 5 (χρησιμοπ. διαφορετικό χρώμα για τα δύο έτη)
- 5) Να εξετάσετε αν οι σεισμοί που καταγράφονται στην περιοχή της Κεφαλονιάς, Ζακύνθου και Λευκάδας ($\text{long} < 21.1$, $37.2 < \text{lat} < 39$) έχουν διαφορετική μέση ένταση από τους σεισμούς που καταγράφονται στην υπόλοιπη Ελλάδα (βασιστείτε στους σεισμούς με ένταση ≥ 4). Υπάρχει στατιστικά σημαντική διαφορά στις αντίστοιχες διασπορές;
- 6) (i) Να εκτιμήσετε μέσω δ.ε. 95% το μέσο εστιακό βάθος των σεισμών με μέγεθος ≥ 4 . (ii) Ελέγξτε αν είναι μεγαλύτερο από 20km (μονόπλευρο τεστ).
- 7) (i) Να δοθεί ένα ιστόγραμμα με τα εστιακά βάθη των σεισμών (40 κλάσεις από 0 έως 200km). (ii) Να δοθεί ένα scatterplot που να δείχνει τα μεγέθη των σεισμών σε σχέση με τα εστιακά τους βάθη. (iii) Τι ποσοστό των σεισμών με μέγεθος ≥ 4 έχουν εστιακό βάθος ≥ 20 km.
- 8) Με βάση τους σεισμούς με ένταση ≥ 4 : (i) Να ελέγξετε αν το μέσο μέγεθος των σεισμών με «μεγάλο» εστιακό βάθος (≥ 15 km) είναι διαφορετικό από το μέσο μέγεθος των σεισμών με «μικρό» ε-

στιακό βάθος ($< 15\text{km}$). (ii) Οι σεισμοί με «μεγάλο» εστιακό βάθος ακολουθούν την ίδια κατανομή με τους σεισμούς με «μικρό» εστιακό βάθος (χρησιμ. δυο διαφορετικά απαραμετρικά τεστ);

9) (i) Να φτιάξετε ένα ραβδόγραμμα που να δείχνει το πλήθος των σεισμών ανά ώρα ημέρας (24 ράβδοι). (ii) Ένας τυχαίος σεισμός έντασης ≥ 4 εμφανίζεται με την ίδια πιθανότητα όλες τις ώρες της ημέρας (χ^2 τεστ);

10) Να φτιάξετε ένα κοινό γράφημα με τα δ.ε. 95% για το μέσο μέγεθος των σεισμών ανά ώρα (24 δ.ε.)

11) Υποστηρίζεται ότι οι περισσότεροι σεισμοί μεγέθους ≥ 4 καταγράφονται τις απογευματινές ώρες (από 17:00 έως και 22:59). Επαληθεύεται κάτι τέτοιο από τα δεδομένα; (Binomial test, χ^2 τεστ)

12) (i) Να φτιάξετε ένα ραβδόγραμμα που να δείχνει το πλήθος των σεισμών ανά μήνα και ανά έτος (12×2 ράβδοι). (ii) Να φτιάξετε όλα τα Boxplots των μεγεθών των σεισμών ανά μήνα και ανά έτος (σε κοινό γράφημα, 12×2 boxplots). (iii) Να φτιάξετε ένα κοινό γράφημα με τα δ.ε. 95% για το μέσο μέγεθος των σεισμών ανά μήνα και ανά έτος (σε κοινό γράφημα, 12×2 δ.ε.).

13) Ποια κατανομή νομίζετε ότι ακολουθούν τα δευτερόλεπτα (μεταβλητή SEC) που καταγράφονται σεισμοί; Να το επαληθεύσετε με κατάλληλο έλεγχο (Ιστόγραμμα, Q-Q plot, K-S test).

14) Αν $N(r)$ είναι το πλήθος των σεισμών μεγέθους r , να φτιάξετε ένα scatterplot της $\log_{10}N(r)$ σε σχέση με την r . Να σχολιάσετε τι παρατηρείτε. Διερευνήσετε αν μπορούμε να διατυπώσουμε κάποιο νόμο με βάση την παρατήρηση αυτή.

15) Να κατασκευάσετε έναν πίνακα που να δείχνει το πλήθος των σεισμών με ένταση στα διαστήματα $A = [3.2, 3.5]$, $B = [3.6, 3.9]$, $C = [4, 4.3]$, $D = [4.4, 4.7]$, $E = [4.8 \text{ και άνω}]$ ανά έτος. Έχει διατυπωθεί η άποψη ότι στον ελλαδικό χώρο ένας σεισμός ανήκει σε μία από τις παραπάνω πέντε κατηγορίες με αντίστοιχες πιθανότητες: 60%, 30%, 7%, 2%, 1%. Επαληθεύεται αυτό από τα δεδομένα (χ^2 τεστ);

16) Με βάση σεισμούς μεγέθους ≥ 4 : (i) Το μέσο μέγεθος μεταξύ των δύο ετών είναι το ίδιο; (ii) Η κατανομή των μεγεθών των σεισμών μεταξύ των δυο ετών είναι ίδια; (χρησιμ. δυο διαφορετικά απαραμετρικά τεστ)

17) Να ελέγξετε αν τα μεγέθη των σεισμών είναι ανεξάρτητα από τους χρονικά γειτονικούς σεισμούς (runs test με βάση το mean και το median). Τι συμπεραίνετε; Υπάρχει θετική ή αρνητική τοπική εξάρτηση των μεγεθών των σεισμών;

18) Με βάση τις παρατηρήσεις του έτους 1983+2x: Να κατασκευάσετε μία νέα μεταβλητή TIME η οποία να δείχνει το χρόνο εκδήλωσης του σεισμού σε ώρες από την αρχή του έτους (π.χ. 3523.5 σημαίνει 3523 ώρες και 30 λεπτά). *Υπόδειξη:* χρησιμοποιήστε και μια νέα μεταβλητή DAYSB η οποία για κάθε τιμή y της MONTH να δείχνει το πλήθος των ημερών πριν από τον μήνα y (JAN $\rightarrow$ 0, FEB $\rightarrow$ 31, MAR $\rightarrow$ 31+28, ...). Να κατασκευάσετε το ιστόγραμμα των χρόνων αυτών (20 κλάσεις από 0 έως 8760) για σεισμούς με μέγεθος ≥ 3.2 και για σεισμούς με μέγεθος ≥ 4 .

19) Να κατασκευάσετε μια νέα μεταβλητή η οποία θα καταγράφει τους ενδιάμεσους χρόνους (σε ώρες) μεταξύ των εκδηλώσεων σεισμών με μέγεθος ≥ 4 : (i) Να κατασκευάσετε το ιστόγραμμα των τιμών της. (ii) Ποιος είναι ο μεγαλύτερος ενδιάμεσος χρόνος και ποια η εκτίμηση του μέσου και της τυπικής απόκλισης ενός ενδιάμεσου χρόνου; Κάθε πόσες ώρες περίπου γίνεται ένας σεισμός μεγέθους ≥ 4 σε όλο τον ελλαδικό χώρο;

20) Να ελέγξετε αν οι ενδιάμεσοι χρόνοι μεταξύ των σεισμών μεγέθους ≥ 4 ακολουθούν την εκθετική κατανομή (Q-Q plot, K-S)

21) Ελέγξετε αν οι ενδιάμεσοι χρόνοι μεταξύ των σεισμών μεγέθους ≥ 4 ακολουθούν την κατανομή Γάμμα (Q-Q plot, K-S). Δώστε τις εκτιμήσεις των παραμέτρων της Γάμμα.

(Όλοι οι έλεγχοι να γίνουν σε ε.σ. $\alpha = 5\%$)