

Θέμα εργασίας

Οδηγίες: Σε κάθε φοιτητή/φοιτήτρια αντιστοιχεί ένας αριθμός από 1 ως 10 σύμφωνα με το τελευταίο ψηφίο του αριθμού ειδικού μητρώου (ΑΕΜ) του (π.χ. στο φοιτητή ή στη φοιτήτρια με αριθμό ειδικού μητρώου 4002 αντιστοιχεί το 2 και με 4000 το 10). Αυτός ο αριθμός καθορίζει τα δεδομένα που θα χρησιμοποιήσει. **Αν ο φοιτητής ή η φοιτήτρια χρησιμοποιήσει λάθος δεδομένα η εργασία του/της δε θα αξιολογηθεί.**

Οδηγίες:

Η εργασία θα αξιολογηθεί με άριστα τη **μία μονάδα**. Θα πρέπει:

- Να είναι ατομική. Η συνεργασία πάνω στα ερωτήματα είναι επιθυμητή αλλά η αντιγραφή απαγορεύεται.
- Στην πρώτη σελίδα να αναφέρεται το όνομα και το ΑΕΜ του φοιτητή / της φοιτήτρια.
- Να περιέχει τις απαντήσεις και λύσεις με σχόλια, όπως ζητούνται και με τη σειρά που ζητούνται.
- Τα γραφήματα και οι πίνακες αποτελεσμάτων από το SPSS θα πρέπει να παρουσιάζονται με τη σειρά που ζητούνται και σε κατάλληλα σημεία μέσα στο κείμενο της εργασίας.
- Το κάθε γράφημα από το SPSS θα πρέπει να έχει στον τίτλο το ΑΕΜ του φοιτητή / της φοιτήτριας και ο κάθε πίνακας από το SPSS θα πρέπει να έχει στην επικεφαλίδα το ΑΕΜ του φοιτητή / της φοιτήτριας.
- Η εργασία θα πρέπει να είναι γραμμένη στον υπολογιστή σε πρόγραμμα επεξεργασίας κειμένου, όπως Word.

Για να αξιολογηθεί η εργασία ο φοιτητής θα πρέπει να στείλει με ηλεκτρονικό ταχυδρομείο το αρχείο της εργασίας ως συνημμένο στη διεύθυνση **dkugiu@auth.gr** μέχρι τις **30/6/2014**. Το όνομα του αρχείου θα πρέπει να είναι **stat<ΑΕΜ>.doc**, όπου το <ΑΕΜ> θα πρέπει να αντικατασταθεί με τον αριθμό του ΑΕΜ (χωρίς τις ανισότητες), και doc δηλώνει ότι είναι γραμμένο σε Word (αλλιώς με κάποιο άλλο κωδικό μορφής αρχείου, π.χ. docx, pdf, ps). Για παράδειγμα για φοιτητή με κωδικό 4002 το αρχείο σε Word θα πρέπει να έχει το όνομα **stat4002.doc**.

Περιγραφή προβλήματος:

Τα αίτια των πυρκαγιών δεν είναι γενικά γνωστά και πιστεύεται πως σημαντικό ρόλο παίζουν οι κλιματικές και μετεωρολογικές συνθήκες. Για την αποτίμηση του κινδύνου πυρκαγιάς χρησιμοποιείται ένα σύστημα μετεωρολογικών δεικτών, γνωστό ως forest Fire Weather Index (FWI). Το σύστημα αυτό περιλαμβάνει μεταξύ άλλων τους δείκτες Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC) και το Initial Spread Index (ISI). Άλλοι παράγοντες / δείκτες μπορεί επίσης να χρησιμοποιηθούν για να εκτιμήσουν τον κίνδυνο πυρκαγιών όπως η θερμοκρασία, η σχετική υγρασία, η ταχύτητα ανέμου, η βροχόπτωση, καθώς επίσης και η γεωγραφική θέση της περιοχής.

Τα παραπάνω μεγέθη μετρήθηκαν σε μια μεγάλη χρονική περίοδο σε 517 διαφορετικές μέρες σε μια περιοχή (τα συγκεκριμένα στοιχεία παραλείπονται) και δίνονται στο αρχείο δεδομένων SPSS `ForestFires.sav` στην ιστοσελίδα του μαθήματος <http://users.auth.gr/dkugiu/Teach/ChemicalEngineer/index.html> ή εναλλακτικά στο blackboard (<https://blackboard.lib.auth.gr> με κωδικό μαθήματος 10U001). Στις μετρήσεις προστέθηκε η πληροφορία του αν έγινε πυρκαγιά ή όχι την αντίστοιχη χρονική στιγμή (μέρα) μέτρησης καθώς και ο μήνας της μέτρησης (η πληροφορία της ημέρας δεν χρησιμοποιήθηκε). Οι στήλες του αρχείου είναι ως εξής:

1. **X**: άξονας x χωρικών συντεταγμένων περιοχής: 1 to 9
2. **Y**: άξονας y χωρικών συντεταγμένων περιοχής: 2 to 9
3. **FFMC**: FFMC δείκτης συστήματος FWI: 18.7 to 96.20
4. **DMC**: DMC δείκτης συστήματος FWI: 1.1 to 291.3
5. **DC**: DC δείκτης συστήματος FWI: 7.9 to 860.6
6. **ISI**: ISI δείκτης συστήματος FWI: 0.0 to 56.10
7. **temp**: θερμοκρασία σε βαθμούς Κελσίου: 2.2 to 33.30
8. **RH**: σχετική υγρασία σε ποσοστά: 15.0 to 100
9. **wind**: ταχύτητα ανέμου σε km/h: 0.40 to 9.40
10. **rain**: βροχόπτωση σε mm/m² : 0.0 to 6.4
11. **fire**: πυρκαγιά με κάψιμο σημαντική έκτασης: 0-> δεν υπήρξε πυρκαγιά, 1-> υπήρξε πυρκαγιά
12. **month**: μήνας του έτους: από 'jan' ως 'dec'
13. **monthcode**: κωδικός μηνός του έτους: από 1 ως 12.

Το ενδιαφέρον της εργασίας εστιάζεται στα παρακάτω θέματα: Α) αν υπάρχει διαφορά στο επίπεδο των δεικτών στις στήλες 1-10 όταν εμφανίζεται ή δεν εμφανίζεται πυρκαγιά, και Β) αν η θερμοκρασία (στήλη 7) συσχετίζεται με (και επηρεάζει ή επηρεάζεται από) τους υπόλοιπους δείκτες καθώς και το μήνα του έτους.

Ζητήματα της εργασίας:Μέρος Α

Στο μέρος Α ο κάθε φοιτητής θα επιλέξει το δείκτη που είναι στη στήλη του αρχείου δεδομένων που αντιστοιχεί στο τελευταίο ψηφίο του αριθμού ΑΕΜ του (π.χ. αν λήγει σε 0 θα χρησιμοποιήσει το δείκτη βροχόπτωσης στη στήλη 10). Από ολόκληρο το δείγμα των 517 ημερών σε 247 ημέρες δεν υπήρχε πυρκαγιά (η τιμή στη στήλη fire είναι 0) και στις υπόλοιπες 270 ημέρες υπήρχε πυρκαγιά (η τιμή στη στήλη fire είναι 1). Αυτά είναι τα δύο **μεγάλα δείγματα Α1 και Β1** αντίστοιχα που θα χρησιμοποιήσετε για το δείκτη σας. Επίσης από κάθε ένα δείγμα θα επιλέξετε τυχαία 100 ημέρες και αυτά θα είναι τα **μικρότερα δείγματα Α2 και Β2** για το δείκτη σας. Η μεταβλητή ενδιαφέροντος παρακάτω αναφέρεται ως **δείκτης** και αντιστοιχεί στο δείκτη που σας αντιστοιχεί (ένας από τους 10 δείκτες στις στήλες 1 ως 10 του πίνακα δεδομένων).

1. Για τα δείγματα Α1 και Β1 των παρατηρήσεων του δείκτη με ή χωρίς πυρκαγιά, υπολογίστε τα μέτρα κεντρικής τάσης (μέση τιμή και διάμεσο) και μεταβλητότητας (διασπορά, τυπική απόκλιση, εύρος δεδομένων, πρώτο και τρίτο τεταρτομόριο). Σχηματίστε τα δύο θηκογράμματα σε ένα σχήμα. Με βάση το αντίστοιχο θηκογράμμα και τα αντίστοιχα συνοπτικά μέτρα σχολιάστε την κατανομή του δείκτη για την περίπτωση εμφάνισης πυρκαγιάς και μη εμφάνισης πυρκαγιάς. Διαφέρουν οι δύο κατανομές του δείκτη ως προς το κέντρο και τη μεταβλητότητα τους;
2. Με βάση το δείγμα Α1 του δείκτη υπολογίστε το 95% διάστημα εμπιστοσύνης για την τυπική απόκλιση του δείκτη όταν δεν υπάρχει πυρκαγιά. Κάνετε το ίδιο για το δείγμα Β1 και σχολιάστε αν η τυπική απόκλιση του δείκτη φαίνεται να είναι στα ίδια επίπεδα όταν δεν υπάρχει πυρκαγιά και όταν υπάρχει πυρκαγιά. [Εδώ δε σας βοηθά το SPSS!].
3. Υπολογίστε με βάση το δείγμα Α1 το 95% διάστημα εμπιστοσύνης (δ.ε.) για τη μέση τιμή του δείκτη όταν δεν υπάρχει πυρκαγιά. Κάνετε το ίδιο για το δείγμα Β1. Συγκρίνετε και σχολιάστε τα δύο διαστήματα εμπιστοσύνης.
4. Ένας βασικός σκοπός αυτής της μελέτης είναι η σύγκριση της μέσης τιμής του δείκτη όταν υπάρχει και όταν δεν υπάρχει πυρκαγιά. Για αυτό θεωρήστε τα δύο δείγματα Α1 και Β1 και υπολογίστε το κατάλληλο 95% δ.ε. για να εκτιμήσετε αν η μέση τιμή του δείκτη διαφέρει όταν εμφανίζεται και δεν εμφανίζεται πυρκαγιά. Αν διαφέρει εκτιμήστε επίσης τη διαφορά. Κάνετε το ίδιο σε επίπεδο εμπιστοσύνης 99%. Συγκρίνετε και σχολιάστε τα δύο διαστήματα εμπιστοσύνης.
5. Επαναλάβετε τα βήματα 1-4 για τα μικρότερα δείγματα Α2 και Β2. Διαφέρουν τα αποτελέσματα σας στο ερώτημα 4;

Μέρος Β

Στο δεύτερο μέρος θα χρησιμοποιήσετε τα δεδομένα από δύο μεταβλητές (δείκτες), τη θερμοκρασία στη στήλη 7 και τον δείκτη στην στήλη που αντιστοιχεί στο τελευταίο ψηφίο του αριθμού ΑΕΜ σας. Αν το τελευταίο ψηφίο του ΑΕΜ σας λήγει σε 7 θα χρησιμοποιήσετε ως δεύτερη μεταβλητή των κωδικό του μήνα στη στήλη 13. Για τις περιπτώσεις που ο δεύτερος δείκτης είναι ο άξονας Χ (στήλη 1), ο άξονας Υ (στήλη 2) ή ο κωδικός μήνα (στήλη 13), θέλουμε να διερευνήσουμε αν η θερμοκρασία επηρεάζεται από αυτόν το δείκτη, ενώ στις άλλες περιπτώσεις θέλουμε να διερευνήσουμε αν η θερμοκρασία επηρεάζει το δείκτη. Σε αντιστοιχία με το Μέρος Α, θα χρησιμοποιήσετε το ζευγάρι μεταβλητών όλων των 517 παρατηρήσεων (**μεγάλο δείγμα**) και αυτό των 200 παρατηρήσεων (**μικρότερο δείγμα**) (δε θα κάνετε διάκριση σε ύπαρξη ή μη πυρκαγιάς).

6. Σχεδιάστε το διάγραμμα διασποράς για το ζευγάρι των μεταβλητών του μεγάλου δείγματος που σας αντιστοιχεί και υπολογίστε τον αντίστοιχο συντελεστή συσχέτισης. Με βάση το διάγραμμα διασποράς και το συντελεστή συσχέτισης, σχολιάστε αν φαίνεται να υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών που σας αντιστοιχούν.
7. Υπολογίστε το μοντέλο γραμμικής παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων για το ζευγάρι των μεταβλητών του μεγάλου δείγματος που σας αντιστοιχεί σύμφωνα με τις παραπάνω οδηγίες.
8. Επαναλάβετε τα βήματα 6-7 για το μικρότερο δείγμα.
9. Συγκρίνετε την τυπική απόκλιση των υπολοίπων για το μοντέλο γραμμικής παλινδρόμησης που εκτιμήθηκε στο μεγάλο δείγμα και στο μικρότερο δείγμα. Σχολιάστε επίσης αν κάποιο από τα δύο μοντέλα γραμμικής παλινδρόμησης είναι κατάλληλο για πρόβλεψη.