

## Θέμα εργασίας

### 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 αρχή (πρώτη σελίδα) να αναφέρεται το όνομα και το ΑΕΜ του φοιτητή / της φοιτήτριας.
- 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.

Η εργασία θα πρέπει να παραδοθεί ηλεκτρονικά μέσω της ιστοσελίδας του μαθήματος στο elearning (<https://elearning.auth.gr/course/view.php?id=6010>) το αργότερο ως τις **30/6/2019**. Θα πρέπει να παραδοθεί ηλεκτρονικά μόνο ένα αρχείο που περιέχει ολόκληρη την εργασία. Το όνομα του αρχείου θα πρέπει να είναι Stat<ΑΕΜ>.docx όπου <ΑΕΜ> είναι ο Αριθμός Ειδικού Μητρώου του φοιτητή που παραδίδει την εργασία (χωρίς τις ανισότητες) και docx δηλώνει ότι είναι γραμμένο σε Word (αλλιώς με κάποιο άλλο κωδικό μορφής αρχείου, π.χ. doc, pdf, ps), π.χ. αν το ΑΕΜ είναι 9102 το αρχείο Word θα πρέπει να έχει όνομα Stat9102.doc. **Παρακαλώ να εξετάσετε αν το όνομα του αρχείου είναι σωστό πριν το υποβάλλετε. Αρχείο με όνομα που δεν είναι στην προβλεπόμενη μορφή μπορεί να αγνοηθεί και η εργασία να μην αξιολογηθεί!**

### Περιγραφή προβλήματος και δεδομένα

Ένας βασικός δείκτης της ποιότητας (μόλυνσης) του αέρα είναι το PM<sub>25</sub> (αιωρούμενα σωματίδια με διάμετρο μικρότερη από 2.5 μικρόμετρα). Για την καλή ποιότητα του αέρα η Ευρωπαϊκή Ένωση έχει θέσει ως όριο η συγκέντρωση των PM<sub>25</sub> να μην ξεπερνά τα 25 μg/m<sup>3</sup>. Θέλουμε να διερευνήσουμε τη στάθμη του PM<sub>25</sub> σε διάφορες χώρες της Ευρώπης καθώς και την εξάρτηση της από το υψόμετρο (altitude). Η μελέτη βασίζεται σε στοιχεία από σταθμούς 11 χωρών της Ευρώπης που έγιναν το έτος 2015 (πηγή: <http://dataservice.eea.europa.eu/>). Τα δεδομένα βρίσκονται στο αρχείο δεδομένων SPSS στην ιστοσελίδα του μαθήματος AirQuality2015.sav. Οι μεταβλητές (στήλες) και για τα δύο αρχεία είναι:

country → Τα ονόματα των 11 χωρών της Ευρώπης

countryNo → Οι αύξοντες αριθμοί των 11 χωρών της Ευρώπης

PM25Level → η μέση μέτρηση του δείκτη PM<sub>25</sub> σε μg/m<sup>3</sup> για το έτος 2015.

Altitude → το υψόμετρο (σε μέτρα)

### Ζητήματα της εργασίας

Η χώρα που σας αντιστοιχεί έχει αύξοντα αριθμό ίδιο με το τελευταίο ψηφίο του ΑΕΜ σας (το 0 αντιστοιχεί στη χώρα Νο 10, Μεγάλη Βρετανία, United Kingdom). Από τη χώρα αυτήν επιλέξετε με τυχαίο και αντιπροσωπευτικό τρόπο ένα δείγμα 20 καταγραφών (μέσο PM<sub>25</sub> το έτος 2015 σε 20 διαφορετικούς σταθμούς). Αυτό είναι το δείγμα Α. Από τη χώρα με αύξοντα αριθμό 11, Ιταλία, επιλέξετε με τυχαίο και αντιπροσωπευτικό τρόπο ένα μικρό δείγμα 20 καταγραφών. Αυτό είναι το δείγμα Β.

### Μελέτη Α

Θα μελετήσετε την κατανομή του δείκτη PM<sub>25</sub>, και κυρίως τη μέση τιμή του, στις δύο χώρες που σας αντιστοιχούν.

1. Σχολιάστε την *κατανομή* του δείκτη PM<sub>25</sub> στις δύο χώρες Α και Β με βάση τα αντίστοιχα δείγματα των μετρήσεων από 20 σταθμούς. Θα πρέπει να συμπεριλάβετε τα παρακάτω, χρησιμοποιώντας το SPSS:
  - Έναν πίνακα με συνοπτικά μέτρα κεντρικής τάσης (μέση τιμή και διάμεσο), μεταβλητότητας (διασπορά, τυπική απόκλιση, εύρος δεδομένων, πρώτο και τρίτο τεταρτομόριο). Ο πίνακας θα πρέπει να περιέχει τα μέτρα και για τα δύο δείγματα.
  - Ιστόγραμμα για το κάθε δείγμα.
  - Ένα σχήμα που να περιέχει τα δύο θηκογράμματα, ένα για κάθε δείγμα.

Να γίνουν αναλυτικά σχόλια για τον πίνακα και τα σχήματα συγκρίνοντας τα αποτελέσματα για τα δύο δείγματα. Υπάρχουν διαφορές στην κατανομή για τη στάθμη του δείκτη PM<sub>25</sub> στις δύο χώρες με βάση τα δύο δείγματα;

2. Κάνετε το ίδιο με το ερώτημα 1 αλλά για το δείγμα όλων των διαθέσιμων μετρήσεων PM<sub>25</sub> για τις δύο χώρες που σας αντιστοιχούν. Έχουν αλλάξει οι κατανομές του PM<sub>25</sub> στις δύο χώρες και οι διαφορές μεταξύ τους με βάση τα μεγαλύτερα δείγματα;

3. Για κάθε ένα από τα δείγματα Α και Β των 20 παρατηρήσεων, ελέγξτε με κατάλληλο 95% διάστημα εμπιστοσύνης αν η μέση τιμή του δείκτη PM25 μπορεί να είναι  $25 \mu\text{g}/\text{m}^3$  (το όριο για την καλή ποιότητα του αέρα της ΕΕ). Στην απάντησή σας συμπεριλάβετε κατάλληλο πίνακα αποτελεσμάτων στο SPSS και για τα δύο δείγματα και σχολιάστε τα αποτελέσματα. Συγκρίνετε επίσης την ακρίβεια της εκτίμησης της μέσης τιμής του δείκτη PM25 στις δύο χώρες παρουσιάζοντας κατάλληλο πίνακα από το SPSS για το διάστημα εμπιστοσύνης σε επίπεδο εμπιστοσύνης 95%. Σημειώστε τυχόν επιφύλαξη που έχετε για την εγκυρότητα των αποτελεσμάτων.
4. Με βάση τα δείγματα Α και Β, ελέγξτε αν η μέση τιμή του δείκτη PM25 μπορεί να είναι ίδια στις δύο χώρες χρησιμοποιώντας κατάλληλο 95% διάστημα εμπιστοσύνης. Στην απάντησή σας συμπεριλάβετε κατάλληλο πίνακα αποτελεσμάτων στο SPSS και σχολιάστε τα αποτελέσματα. Σημειώστε τυχόν επιφύλαξη που έχετε για την εγκυρότητα των αποτελεσμάτων.
5. Επαναλάβετε την ίδια διαδικασία στο βήμα 4, αλλά ως δείγματα Α και Β θεωρείστε όλες τις διαθέσιμες παρατηρήσεις του PM25 για την κάθε χώρα που σας αντιστοιχεί. Διαφέρουν οι απαντήσεις σας για τη σύγκριση του μέσου δείκτη PM25 στις δύο χώρες από αυτές για τα μικρά δείγματα στο βήμα 4;

### **Μελέτη Β:**

Θέλουμε να εξετάσουμε αν εξαρτάται η στάθμη του δείκτη PM25 από το υψόμετρο του σταθμού στον οποίο γίνεται η μέτρηση. Για αυτό θεωρούμε τα ζευγαρωτά μικρά δείγματα των 20 καταγραφών του δείκτη PM25 (μέση τιμή στο έτος 2015) σε 20 διαφορετικούς σταθμούς της χώρας και τα αντίστοιχα υψόμετρα που βρίσκονται οι σταθμοί. Η ανάλυση θα γίνει και για τα δύο δείγματα Α και Β των χωρών που σας αντιστοιχούν (τα ίδια που χρησιμοποιήσατε στη Μελέτη Α).

6. Χρησιμοποιώντας το SPSS, κάνετε κατάλληλο διάγραμμα διασποράς και υπολογίστε τον αντίστοιχο συντελεστή συσχέτισης για το δείκτη PM25 και το υψόμετρο για κάθε μια από τις δύο χώρες ξεχωριστά χρησιμοποιώντας το αντίστοιχο δείγμα. Σχολιάστε αν εξαρτάται η στάθμη του δείκτη PM25 από το υψόμετρο για κάθε μια από τις δύο χώρες.
7. Για κάθε ένα από τα δύο ζευγάρια δειγμάτων (για την περιοχή Α και Β) για το δείκτη PM25 και το υψόμετρο και χρησιμοποιώντας το SPSS, εκτιμήστε το μοντέλο γραμμικής παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων. Σχολιάστε το κάθε ένα από τα δύο μοντέλα που εκτιμήσατε με έμφαση στην καταλληλότητα του μοντέλου για προβλέψεις. Στην απάντησή σας θα συμπεριλάβετε τους πίνακες του SPSS που δίνουν την εκτίμηση των παραμέτρων του μοντέλου (σταθερός όρος, κλίση και τυπική απόκλιση σφαλμάτων παλινδρόμησης). Φαίνεται τα δύο μοντέλα παλινδρόμησης να είναι τα ίδια;
8. Προβλέψτε, αν γίνεται, με το κάθε ένα από τα δύο μοντέλα που εκτιμήσατε στο ερώτημα 7 (για το δείγμα Α και Β) τη στάθμη του δείκτη PM25 αν γνωρίζουμε ότι το υψόμετρο είναι 25 μ και αν είναι 250 μ. Προσεγγιστικά ποια είναι η ακρίβεια της πρόβλεψης για τη χώρα Α και τη χώρα Β;
9. Όπως στο ερώτημα 7, υπολογίστε την ευθεία ελαχίστων τετραγώνων της στάθμης του δείκτη PM25 από το υψόμετρο για την κάθε χώρα χρησιμοποιώντας όλο το δείγμα των διαθέσιμων ζευγαρωτών παρατηρήσεων. Συγκρίνετε για κάθε χώρα ξεχωριστά το μοντέλο παλινδρόμησης με βάση το δείγμα των 20 καταγραφών και με βάση το δείγμα όλων των διαθέσιμων καταγραφών.