

ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ ΓΙΑ ΗΛΕΚΤΡΟΛΟΓΟΥΣ ΜΗΧΑΝΙΚΟΥΣ

Θέμα εργασίας στη Στατιστική

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.
- Να γραφεί στο Word ή κάποιο άλλο πρόγραμμα επεξεργασίας κειμένου.
- Να περιέχει τις απαντήσεις και λύσεις με σχόλια όπου ζητούνται, καθώς και τα αποτελέσματα από το SPSS (γραφήματα, πίνακες) με τη σειρά που ζητούνται στα ερωτήματα της εργασίας.
- Το κάθε γράφημα ή πίνακας από το SPSS θα πρέπει να έχει στον τίτλο ή επικεφαλίδα μεταξύ άλλων το ΑΕΜ του φοιτητή / της φοιτήτριας.

Η εργασία θα πρέπει να παραδοθεί ηλεκτρονικά μέσω της ιστοσελίδας του μαθήματος (<http://alexander.ee.auth.gr:8083/eTHMMY/>) το αργότερο ως τις **3/7/2010**. Θα πρέπει να παραδοθεί ηλεκτρονικά μόνο ένα αρχείο που περιέχει ολόκληρη την εργασία. Το όνομα του αρχείου θα πρέπει να είναι Stat<ΑΕΜ>.doc όπου <ΑΕΜ> είναι ο Αριθμός Ειδικού Μητρώου του φοιτητή που παραδίδει την εργασία (χωρίς τις ανισότητες), π.χ. αν το ΑΕΜ είναι 10000 το αρχείο Word θα πρέπει να έχει όνομα Stat10000.doc.

Περιγραφή προβλήματος και δεδομένα

Μας ενδιαφέρει να αναλύσουμε στοιχεία για την ετήσια εθνική παραγωγή και κατανάλωση ηλεκτρικής ενέργειας σε διάφορες χώρες του κόσμου. Τα κύρια ερωτήματα που θα μας απασχολήσουν είναι αν η παραγωγή και κατανάλωση ηλεκτρικής ενέργειας διεθνώς διαφέρει σε διαφορετικές χρονικές περιόδους, πόσο στενά συσχετίζονται (παραγωγή και κατανάλωση), καθώς και αν η κατανάλωση ηλεκτρικής ενέργειας παρουσιάζει συγκεκριμένη αυξητική τάση και ίδια για όλες τις χώρες. Για να απαντήσουμε σε αυτά τα ερωτήματα θα αναλύσουμε δείγματα από ένα σύνολο χωρών για τις οποίες έχουμε ετήσιες μετρήσεις για τον πληθυσμό, παραγωγή και κατανάλωση ηλεκτρικής ενέργειας τους. Όλα τα δεδομένα δίνονται στο αρχείο **WorldElectricity.sav** (μορφή δεδομένων SPSS) που βρίσκεται στην ιστοσελίδα του μαθήματος <http://alexander.ee.auth.gr:8083/eTHMMY/> ή εναλλακτικά στην ιστοσελίδα <http://users.auth.gr/~dkugiu/Teach/ElectricEngineer/index.html>. Τα δεδομένα στο αρχείο είναι για ετήσιες τιμές του πληθυσμού, παραγωγής και κατανάλωσης ενέργειας σε 180 χώρες και για τα έτη 1980 – 2007 [πηγή: US Energy Information Administration, <http://tonto.eia.doe.gov/cfapps/ipdbproject/IEDIndex3.cfm>]. Η οργάνωση των δεδομένων δίνεται παρακάτω, όπου η κάθε γραμμή στον πίνακα δεδομένων του αρχείου αφορά μια χώρα (180 γραμμές):

Στήλη 1: Όνομα της χώρας [Country].

Στήλη 2 – 29: Πληθυσμός της χώρας (σε εκατομμύρια) [Pop1980 – Pop2007]

Στήλη 30 – 57: Παραγωγή ηλεκτρικής ενέργειας της χώρας (σε δισεκατομμύρια κιλοβατώρες) [Gen1980 – Gen2007]

Στήλη 58 – 85: Κατανάλωση ηλεκτρικής ενέργειας της χώρας (σε δισεκατομμύρια κιλοβατώρες) [Con1980 – Con2007]

Συνοπτικά η εργασία αναφέρεται στις δύο παρακάτω μελέτες.

Μελέτη Α

Με βάση τα δείγματα από κάποιες χώρες σε δύο διαφορετικά έτη θέλουμε να διερευνήσουμε αν υπάρχει αύξηση της κατανάλωσης ηλεκτρικής ενέργειας, όπου έχουμε κανονικοποιήσει με τον πληθυσμό της κάθε χώρας στο αντίστοιχο έτος.

Μελέτη Β

Με βάση τα δείγματα από κάποιες χώρες θέλουμε να διερευνήσουμε αν συσχετίζεται η παραγωγή και η κατανάλωση ηλεκτρικής ενέργειας για κάποια έτη, καθώς επίσης και αν η κατανάλωση ηλεκτρικής ενέργειας (κανονικοποιημένη με τον πληθυσμό) παρουσιάζει αυξητική τάση με τα έτη.

Ζητήματα της εργασίας:

Μελέτη Α:

Θεωρούμε ότι ο πληθυσμός για κάθε περίοδο είναι όλες οι χώρες για τις οποίες έχουμε μετρήσεις, δηλαδή οι 180 χώρες.

Ανάλογα με το τελευταίο ψηφίο του ΑΕΜ θα χρησιμοποιήσετε το έτος στη δεκαετία από 1980 ως 1989 (για ψηφίο από 0 ως 9 αντίστοιχα) καθώς και το έτος που είναι 18 έτη μετά, δηλαδή 1998 ως 2007 (για ψηφίο από 0 ως 9 αντίστοιχα), π.χ. για ΑΕΜ που λήγει σε 0 θα χρησιμοποιηθούν τα δεδομένα για το έτος 1980 και 1998, σε 5 τα δεδομένα για το έτος 1985 και 2003 και σε 9 τα δεδομένα για το έτος 1989 και 2007.

Επιλέξτε με τυχαίο και αντιπροσωπευτικό τρόπο 60 χώρες που θα σχηματίζουν το δείγμα σας για το κάθε έτος.

Επειδή ο πληθυσμός διαφέρει σημαντικά στις διάφορες χώρες, θα πρέπει να κανονικοποιήσετε τις ετήσιες τιμές της παραγωγής και κατανάλωσης ηλεκτρικής ενέργειας, ώστε να μην επηρεάζονται άμεσα από τον πληθυσμό της κάθε χώρας. Για να το πετύχετε αυτό θα διαιρέσετε την κάθε ετήσια τιμή παραγωγής και κατανάλωσης ηλεκτρικής ενέργειας της χώρας με τον αντίστοιχο πληθυσμό της χώρας στο ίδιο έτος και το αποτέλεσμα θα το πολλαπλασιάσετε με 1000. Αυτό αντιστοιχεί στην παραγωγή ή κατανάλωση ηλεκτρικής ενέργειας σε εκατομμύρια κιλοβατώρες ανά κάτοικο της χώρας.

Τα παρακάτω ερωτήματα αφορούν τα δείγματα των κανονικοποιημένων ετήσιων τιμών παραγωγής και κατανάλωσης ηλεκτρικής ενέργειας.

1. Θεωρήστε το δείγμα των 60 χωρών στα δύο έτη που σας αντιστοιχούν σύμφωνα με το τελευταίο ψηφίο του ΑΕΜ σας.
 - Για κάθε περίπτωση (παραγωγή και κατανάλωση ηλεκτρικής ενέργειας), θα πρέπει να κάνετε **έναν** πίνακα με συνοπτικά μέτρα κεντρικής τάσης (μέση τιμή και διάμεσο) και μεταβλητότητας (διασπορά, τυπική απόκλιση, εύρος δεδομένων, πρώτο και τρίτο τεταρτομόριο). Ο πίνακας θα πρέπει να περιέχει τα μέτρα και για τα δύο δείγματα, δηλαδή για τα δύο έτη.
 - Για κάθε περίπτωση (παραγωγή και κατανάλωση ηλεκτρικής ενέργειας), θα πρέπει να κάνετε **ένα** σχήμα που να περιέχει το θηκόγραμμα για το κάθε ένα από τα δύο δείγματα (ένα για κάθε έτος).
 - Θα πρέπει να κάνετε ένα ιστόγραμμα για το κάθε ένα από τα 4 συνολικά δείγματα.

Σχολιάστε για τους πίνακες και τα σχήματα συγκρίνοντας τα αποτελέσματα για την παραγωγή και κατανάλωση ηλεκτρικής ενέργειας. Είναι η κατανομή κανονική? Υπάρχει διαφορά στην κατανομή της παραγωγής και κατανάλωσης? Υπάρχει διαφορά στα δύο έτη?

2. Εκτιμήστε τη μέση παραγωγή και κατανάλωση ηλεκτρικής ενέργειας για κάθε ένα από τα δύο έτη. Για αυτό θα χρησιμοποιήσετε διαστήματα εμπιστοσύνης σε επίπεδο 95% με βάση το κάθε δείγμα. Στην απάντησή σας συμπεριλάβετε **έναν** πίνακα από το SPSS με αποτελέσματα και για τα δύο έτη. Μπορούμε να προσδιορίσουμε με ακρίβεια τη μέση παραγωγή ή κατανάλωση ηλεκτρικής ενέργειας παγκοσμίως (σε εκατομμύρια κιλοβατώρες ανά κάτοικο της χώρας)?
3. Θέλουμε να συγκρίνουμε αν η μέση παραγωγή και κατανάλωση ηλεκτρικής ενέργειας διαφέρει στα δύο έτη. Χρησιμοποιώντας το SPSS, εκτιμήστε αν υπάρχει διαφορά, και πόση, στη μέση παραγωγή ηλεκτρικής ενέργειας στα δύο έτη. Κάνετε το ίδιο για την κατανάλωση ηλεκτρικής ενέργειας. Για να δώσετε την απάντησή σας υπολογίστε διάστημα εμπιστοσύνης για τη διαφορά μέσων τιμών σε επίπεδο εμπιστοσύνης 95% και συμπεριλάβετε τον αντίστοιχο πίνακα από το SPSS. Κάνετε το ίδιο και για την κατανάλωση ηλεκτρικής ενέργειας. Διαφέρει η μέση παραγωγή ή/και κατανάλωση ηλεκτρικής ενέργειας στα δύο έτη? Βρίσκετε να υπάρχει στατιστικά σημαντική αύξηση μεταξύ των 18 ετών?
4. Επαναλάβετε τα ερωτήματα 1 και 3 για το σύνολο των 180 χωρών. Σχολιάστε αν υπάρχει σημαντική αύξηση της παραγωγής ή/και κατανάλωσης ηλεκτρικής ενέργειας με βάση τις μετρήσεις από όλες τις χώρες. Αλλάζει η κατανομή της παραγωγής και κατανάλωσης ηλεκτρικής ενέργειας από το δείγμα των 60 χωρών στο σύνολο των 120 χωρών (με βάση τα θηκογράμματα, ιστογράμματα και συνοπτικά μέτρα)?

Μελέτη Β:

5. Για το δείγμα των 60 χωρών που σχηματίσατε στη Μελέτη Α, εξετάστε αν υπάρχει συσχέτιση της παραγωγής και κατανάλωσης ηλεκτρικής ενέργειας και αν αυτή διαφέρει στα δύο έτη, όπως αυτά ορίστηκαν στη Μελέτη Α. Για αυτό θα πρέπει να κάνετε το διάγραμμα διασποράς παραγωγής και κατανάλωσης ηλεκτρικής ενέργειας και να υπολογίσετε τον αντίστοιχο συντελεστή συσχέτισης για το κάθε έτος ξεχωριστά. Θα πρέπει και πάλι να συμπεριλάβετε τα κατάλληλα σχήματα και τους κατάλληλους πίνακες του SPSS. Διαφέρουν τα αποτελέσματα σας όταν χρησιμοποιείτε το σύνολο των 180 χωρών? Για να απαντήσετε θα πρέπει να επαναλάβετε την παραπάνω ανάλυση στο σύνολο των 180 χωρών.
6. Επαναλάβετε τη διαδικασία στο ερώτημα 5 για τον πληθυσμό και την (κανονικοποιημένη) παραγωγή ηλεκτρικής ενέργειας. Σχολιάστε αν το αποτέλεσμα από το διάγραμμα διασποράς και το συντελεστή συσχέτισης είναι αναμενόμενο.
7. Διαφοροποιούμε τώρα τη μελέτη και θέλουμε να διερευνήσουμε αν υπάρχει αυξητική τάση της ετήσιας (κανονικοποιημένης) κατανάλωσης ενέργειας, δηλαδή αν υπάρχει αύξηση της κατανάλωσης με το έτος. Επίσης θέλουμε να διερευνήσουμε αν αυτή η τάση διαφέρει με τη χώρα. Για αυτό επιλέξετε την Ελλάδα, μια χώρα της Ευρώπης και μια τρίτη χώρα εκτός Ευρώπης. Για κάθε μια από τις τρεις χώρες κάνετε τα παρακάτω:
 - Σχηματίστε το διάγραμμα διασποράς κατανάλωσης ηλεκτρικής ενέργειας προς έτος.
 - Υπολογίστε το αντίστοιχο μοντέλο παλινδρόμησης (με τη μέθοδο ελαχίστων τετραγώνων).Συγκρίνετε την καταλληλότητα του μοντέλου (δηλαδή το σφάλμα παλινδρόμησης) για κάθε μια από τις τρεις χώρες. Υπάρχει αυξητική τάση και για ποια χώρα μπορεί αυτή να προβλεφθεί καλύτερα?
8. Σχολιάστε τυχόν αμφιβολίες που μπορεί να έχετε για την εγκυρότητα των αποτελεσμάτων της Μελέτης Α και Β.