



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ & ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

Τομέας Μαθηματικών

Πολυτεχνειούπολη – Ζωγράφου ΑΘΗΝΑ - 157 80

ΤΗΛ. : 772 1774

FAX : 772 1775

ΜΑΘΗΜΑ: *Ανάλυση Δεδομένων με Η/Υ (5^ο εξάμηνο)*

ΔΙΔΑΣΚΩΝ: *Μιχάλης Λουλάκης*

ΕΡΓΑΣΙΑ 1^η

Θέμα Εργασίας: Εισαγωγή στην R και Περιγραφική Στατιστική

Άσκηση 1

Θεωρήστε τα δεδομένα του ακόλουθου συνδέσμου:

http://www.math.ntua.gr/~loulakis/info/datafiles/semfe_files/ozone1.txt.

Τα δεδομένα αφορούν σε μετρήσεις της ατμοσφαιρικής συγκέντρωσης όζοντος (O₃) στο λεκανοπέδιο του Λος Άντζελες των ΗΠΑ. Συγκεκριμένα, το δείγμα αποτελείται από 90 μετρήσεις της συγκέντρωσης του όζοντος και τεσσάρων μετεωρολογικών δεικτών. Οι μετρήσεις έλαβαν χώρα σε 90 διακριτές μέρες το έτος 1976. Συλλέχθηκαν οι κάτωθι μεταβλητές:

| Όνομα μεταβλητής | Χαρακτηρισμός | Κωδικοποίηση |
|------------------|-----------------------------------|---|
| Ozone | Ατμοσφαιρική συγκέντρωση όζοντος. | σε parts per million (ppm). |
| Wind_speed | Ταχύτητα ανέμου. | σε μίλια/ώρα. |
| Humidity | Σχετική ατμοσφαιρική υγρασία. | ως ποσοστό. |
| Visibility | Ορατότητα. | σε μίλια. |
| Season | Εποχή του χρόνου. | 1 = χειμώνας. 2 = άνοιξη. 3 =καλοκαίρι. 4 = φθινόπωρο. |

Σημείωση 1: ppm= parts per million. Αέριο όγκου ενός μικρο-λίτρου (1μL) μέσα σε όγκο ενός λίτρου αέρα (1L) θεωρείται ότι αντιστοιχεί σε συγκέντρωση ίση με 1 ppm.

Το όζον της ατμόσφαιρας έχει εν γένει μικρή συγκέντρωση σε σχέση με άλλα χημικά στοιχεία που συναντώνται στην ατμόσφαιρα. Σε αντίθεση με το όζον που βρίσκεται στην στρατόσφαιρα και προστατεύει τη γη από επιβλαβείς ακτινοβολίες, το ατμοσφαιρικό όζον πρέπει να έχει εν γένει μικρή συγκέντρωση. Αυξημένη συγκέντρωση του ατμοσφαιρικού όζοντος μπορεί να οδηγήσει, μεταξύ άλλων, σε μειωμένη πνευμονική λειτουργία και αναπνευστικά προβλήματα, π.χ. άσθμα.

- i) Στα δεδομένα μας, με τον χαρακτήρα “*” συμβολίζουμε τις αγνοούμενες τιμές. Εισάγετε τα δεδομένα στην R, αλλάζοντας κατάλληλα το σύμβολο για τις αγνοούμενες τιμές, και δημιουργήστε ένα πλαίσιο δεδομένων δίνοντας ονόματα στις 5 μεταβλητές.
- ii) Δώστε μια περιγραφική ανάλυση, για κάθε μεταβλητή ξεχωριστά, η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους και σχολιάστε τα ευρήματά σας.
- iii) Με τη βοήθεια κατάλληλων γραφημάτων δείτε αν στο δείγμα οι τιμές της συγκέντρωσης όζοντος διαφοροποιούνται ανάλογα με την εποχή κατά την οποία πραγματοποιήθηκε η μέτρηση.
- iv) Να κατασκευαστεί ο πίνακας συχνοτήτων και σχετικών συχνοτήτων για τα δεδομένα που αφορούν την σχετική ατμοσφαιρική υγρασία (Humidity) με τη χρήση 4 κλάσεων, όπου τιμές σχετικής υγρασίας έως και 29% θεωρούνται «χαμηλές (low)», τιμές σχετικής υγρασίας από 30% έως και 54% θεωρούνται «κανονικές (normal)», τιμές σχετικής υγρασίας από 55% έως και 79% θεωρούνται «υψηλές (high)» και τιμές σχετικής υγρασίας από 80% και άνω θεωρούνται «πολύ υψηλές (very high)». Δώστε αντίστοιχα (αγγλικά) ονόματα στις κατηγορίες της νέας αυτής κατηγορικής μεταβλητής. Εν συνεχεία, περιγράψτε κατάλληλα τη μεταβλητή Ozone ξεχωριστά για κάθε κατηγορία της σχετικής υγρασίας (συμπεριλάβετε γραφικές και αριθμητικές μεθόδους). Συγκρίνετε και σχολιάστε τα αποτελέσματά σας.
- v) Κατασκευάστε ένα πίνακα συνάφειας μεταξύ της μεταβλητής Season και της νέας κατηγορικής μεταβλητής για την σχετική υγρασία. Δώστε τις σχετικές συχνότητες κελιών, γραμμών και στηλών και σχολιάστε τα αποτελέσματά σας.

Άσκηση 2

Ο δειγματικός μέσος \bar{x} χρησιμοποιείται ευρέως σε πρακτικές εφαρμογές ως μέτρο θέσης. Ταυτόχρονα όμως, υπάρχει μια τεράστια ποικιλία από εναλλακτικές μορφές μέτρων θέσης που ονομάζονται σταθμικοί μέσοι με γενική μορφή:

$$\frac{\sum_{i=1}^n w(x_i)x_i}{\sum_{i=1}^n w(x_i)}$$

όπου $w(x_i)$ τα βάρη των παρατηρήσεων x_i . Ο δειγματικός μέσος \bar{x} προκύπτει από τον παραπάνω τύπο αν θέσουμε $w(x_i)=1$, $i=1,\dots,n$. Έστω διάνυσμα παρατηρήσεων $\mathbf{x}=(x_1,x_2,\dots,x_n)^T$. Να γράψετε μια δική σας συνάρτηση στην R η οποία θα παίρνει ως όρισμα το εν λόγω διάνυσμα παρατηρήσεων και θα υπολογίζει και θα επιστρέφει 3 σταθμικούς μέσους με τα εξής βάρη:

$$w_1(x_i)=1 \quad w_2(x_i)=\begin{cases} 1 & \text{if } \left|\frac{x_i-\bar{x}}{s}\right| < 1 \\ 0.5 & \text{if } 1 \leq \left|\frac{x_i-\bar{x}}{s}\right| < 2 \\ 0 & \text{if } 2 \leq \left|\frac{x_i-\bar{x}}{s}\right| \end{cases} \quad w_3(x_i)=\frac{1}{s\sqrt{2\pi}}\exp\left(-\frac{1}{2s^2}(x_i-\bar{x})^2\right), \quad i=1,\dots,n,$$

όπου s είναι η δειγματική τυπική απόκλιση.

Οδηγίες

- Η εργασία θα πρέπει να παραδοθεί στο γραφείο μου ή στην θυρίδα μου (δίπλα από την γραμματεία του Τομέα Μαθηματικών) ή στο γραφείο E2.04 της Ε.Χαριτίδου εκτυπωμένη και όχι σε ηλεκτρονική μορφή **μέχρι την Παρασκευή, 23 Δεκεμβρίου 2011 στις 2μμ. Καμιά εργασία δεν θα γίνει δεκτή μετά την ώρα αυτή.**
- Η εργασία θα πρέπει να είναι σε μορφή επίσημης αναφοράς και να περιλαμβάνει τους κώδικες της R με πλήρη επεξήγηση, γραφήματα και πίνακες με κατάλληλους τίτλους και πλήρη επεξήγηση των αποτελεσμάτων.
- Θα δοθεί ιδιαίτερη σημασία στην παρουσίαση της εργασίας. Η εργασία πρέπει να είναι κατανοητή και να περιγράφει ό,τιδήποτε χρησιμοποιήσατε πειστικά για κάποιον που δεν γνωρίζει πολλά για το αντικείμενο.

Καλή Επιτυχία