



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 3^η

Θέμα Εργασίας: Ανάλυση Παλινδρόμησης

Σημείωση: Η εργασία που θα παραδώσετε δεν πρέπει να ξεπερνά τις **25** σελίδες και πρέπει να είναι γραμμένη με μέγεθος γραμματοσειράς **12**.

Άσκηση 1

Θεωρήστε τα δεδομένα του ακόλουθου συνδέσμου:

http://www.math.ntua.gr/~loulakis/info/datafl1semfe_files/ozone330.txt.

Τα δεδομένα αφορούν σε μετρήσεις της ατμοσφαιρικής συγκέντρωσης όζοντος (O₃) στο λεκανοπέδιο του Λος Άντζελες των ΗΠΑ. Συγκεκριμένα, το δείγμα αποτελείται από **330** μετρήσεις της συγκέντρωσης του όζοντος και τεσσάρων μετεωρολογικών δεικτών. Οι μετρήσεις έλαβαν χώρα σε **330** 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) Εισάγετε τα δεδομένα στην R και δημιουργήστε ένα πλαίσιο δεδομένων δίνοντας ονόματα στις 5 μεταβλητές. Περιγράψτε κατάλληλα τις μεταβλητές σας.
- ii) Με τη χρήση της εντολής *sample*, διαλέξτε τυχαία 300 παρατηρήσεις. Δουλέψτε από εδώ και στο εξής με το εν λόγω σετ δεδομένων.
- iii) Προσαρμόστε το απλό γραμμικό μοντέλο με μεταβλητή απόκρισης Y την μεταβλητή “Ozone” και επεξηγηματική μεταβλητή X την μεταβλητή “Humidity”. Δώστε το γράφημα της ευθείας ελαχίστων τετραγώνων. Ερμηνεύστε τους εκτιμητές των συντελεστών του παραπάνω γραμμικού μοντέλου. Δώστε 95% διαστήματα εμπιστοσύνης για τους συντελεστές του εν λόγω μοντέλου.
- iv) Εξηγήστε πλήρως τα αποτελέσματα που παίρνετε από την R με τη χρήση της εντολής *summary* στο παραπάνω απλό γραμμικό μοντέλο.
- v) Ελέγξτε τις προϋποθέσεις του απλού γραμμικού μοντέλου που προσαρμόσατε παραπάνω. Σχολιάστε πλήρως και λεπτομερώς τα ευρήματά σας.
- vi) Προσαρμόστε το γενικό γραμμικό μοντέλο με μεταβλητή απόκρισης Y την μεταβλητή “Ozone” και επεξηγηματικές μεταβλητές όλες τις υπόλοιπες μεταβλητές. Ερμηνεύστε τους εκτιμητές των συντελεστών του παραπάνω γραμμικού μοντέλου. Δώστε 95% διαστήματα εμπιστοσύνης για τους συντελεστές του εν λόγω μοντέλου.
- vii) Εξηγήστε πλήρως τα αποτελέσματα που παίρνετε από την R με τη χρήση της εντολής *summary* στο παραπάνω γενικό γραμμικό μοντέλο. Ποια είναι η κατηγορία αναφοράς της κατηγορικής μεταβλητής “Season”;
- viii) Ελέγξτε τις προϋποθέσεις του γενικού γραμμικού μοντέλου που προσαρμόσατε παραπάνω. Σχολιάστε πλήρως και λεπτομερώς τα ευρήματά σας.
- ix) Εκτιμήστε, σημειακά και με τη βοήθεια ενός 95% διαστήματος εμπιστοσύνης, με βάση το παραπάνω γενικό γραμμικό μοντέλο, την αναμενόμενη τιμή Όζοντος (μεταβλητή “Ozone”) μιας φθινοπωρινής μέρας κατά την οποία καταγράφεται σχετική ατμοσφαιρική υγρασία 57%, ορατότητα ίση με 110 μίλια και ταχύτητα ανέμου ίση με 2 μίλια/ώρα.
- x) Θεωρήστε ως κατηγορία αναφοράς της κατηγορικής μεταβλητής “Season” το καλοκαίρι και προσαρμόστε ξανά το γενικό γραμμικό μοντέλο σας. Ερμηνεύστε και πάλι τους εκτιμητές των συντελεστών του γραμμικού σας μοντέλου.

- Η εργασία θα πρέπει να παραδοθεί στο γραφείο μου ή στην θυρίδα μου (δίπλα από την γραμματεία του Τομέα Μαθηματικών) ή στο γραφείο Ε2.05 της Ε.Χαριτίδου εκτυπωμένη και όχι σε ηλεκτρονική μορφή **μέχρι την Δευτέρα, 5 Μαρτίου 2012 στις 14:00**. 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 οτιδήποτε χρησιμοποιήσατε πειστικά για κάποιον που δεν γνωρίζει πολλά για το αντικείμενο.

Καλή Επιτυχία