



Εφαρμοσμένη Επιχειρησιακή Στατιστική (Business Statistics & Data Mining) χρησιμοποιώντας διαδεδομένα λογισμικά ανάλυσης

Η ευρεία διάδοση των υπολογιστών μα την παράλληλη αύξηση των δυνατοτήτων τους καθώς και η μεγάλη ανάπτυξη και εξάπλωση των στατιστικών λογισμικών είχε σαν αποτέλεσμα πολλές στατιστικές μέθοδοι να γίνουν γνωστές σε ένα πλατύ κοινό.

Στο σεμινάριο αυτό εξετάζονται τέτοιες μέθοδοι και αποτελούν μέρος μιας γενικότερης μεθοδολογίας γνωστές με το όνομα <<γραμμικά μοντέλα>>.

Τα γραμμικά μοντέλα εξακολουθούν να έχουν την σπουδαιότητά τους για τρεις κυρίως λόγους. Πρώτον, διότι υπάρχουν φαινόμενα τα οποία μπορούν να περιγραφούν από γραμμικά μοντέλα. Δεύτερον, διότι υπάρχουν φαινόμενα τα οποία εκ πρώτης όψης δεν φαίνονται να είναι γραμμικά όμως μπορούν να μετατραπούν σε γραμμικά με κατάλληλους μετασχηματισμούς. Τρίτον, διότι η κατανόηση τους είναι μια απαραίτητη προϋπόθεση για την μελέτη των μη γραμμικών μοντέλων. Στενά συνδεδεμένη με την γραμμική παλινδρόμηση είναι και η <<ανάλυση διακύμανσης>> κατά ένα, κατά δύο,.....,κατά κ παράγοντες.

Το σεμινάριο αυτό απευθύνεται κυρίως σε μαθηματικούς. Όμως και ο μη μαθηματικός ο οποίος το μόνο που θέλει είναι να γνωρίζει την χρήση των μεθόδων αυτών μπορεί να επωφεληθεί από το σεμινάριο αυτό. Ειδικά, ακόμη και η μελέτη των παραδειγμάτων θα μπορούσε να φανεί χρήσιμη στον μη ειδικό.

Αναλυτικά στο σεμινάριο αυτό θα μελετηθούν οι παρακάτω ενότητες:

• ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

1. Γενικά
2. Το απλό γραμμικό μοντέλο
3. Εκτίμηση των παραμέτρων
4. Ανάλυση διακύμανσης
5. Συντελεστής Προσδιορισμού R^2
6. Ιδιότητες των εκτιμητών ελαχίστων τετραγώνων
7. Εκτίμηση του σ^2
8. Διστήματα εμπιστοσύνης και έλεγχος υποθέσεων
9. Το F-Test για την παλινδρόμηση
10. Συσχέτιση μεταξύ των X και Y
11. Ερμηνεία του προβλεπόμενου μοντέλου της παλινδρόμησης
12. Έλλειψη προσαρμογής (Lack of fit)

- **ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ**
 1. Περιγραφή των δεδομένων και του μοντέλου
 2. Εκτίμηση των παραμέτρων
 3. Έλεγχος των υποθέσεων
 4. Επρμηγεία των εκτιμητών
 5. Μερικός συντελεστής συσχέτισεως
 6. Υποθέσεις για τις ανεξάρτητες μεταβλητές

- **ΑΝΑΛΥΣΗ ΤΩΝ ΥΠΟΛΟΙΠΩΝ**
 1. Έλεγχος των υποθέσεων για τα σφάλματα
 2. Έλεγχος ύπαρξης ασυσχέτιστων σφαλμάτων
 3. Έλεγχος σταθερής διακύμανσης
 4. Έλεγχος κανονικότητας σφαλμάτων
 5. Έλεγχος ορθότητας του μοντέλου
 6. Έλεγχος γραμμικότητας του μοντέλου
 7. Εισαγωγή νέων όρων στο μοντέλο
 8. Ακραίες παρατηρήσεις
 9. Επηρεάζουσες παρατηρήσεις

- **ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ**
 1. Ανάλυση του προβλήματος
 2. Πολλαπλές συγκρίσεις
 3. Μέθοδος ελάχιστης σημαντικής διαφοράς
 4. Μέθοδος του Tukey
 5. Μέθοδος του Scheffe
 6. Διαμέριση του SS_{τ}
 7. Πειράματα με δύο παράγοντες
 8. Ανάλυση διακύμανσης κατά δύο παράγοντες με μια παρατήρηση ανά κυψελίδα
 9. Διαμέριση των SS_A και SS_B
 10. Πλήρως τυχαιοποιημένα μπλόκ
 11. Ανάλυση διακύμανσης κατά δύο παράγοντες με ίσο αριθμό παρατηρήσεων (>1) ανά κυψελίδα
 12. Αλληλεπίδραση
 13. Ανάλυση διακύμανσης κατά δύο παράγοντες με άνισο αριθμό παρατηρήσεων ανά κυψελίδα

Στις παραπάνω ενότητες θα δοθεί έμφαση τόσο στο θεωρητικό όσο και στο πρακτικό επίπεδο με την χρήση ηλεκτρονικών υπολογιστών (Laptops) και διαδεδομένων λογισμικών στατιστικής ανάλυσης. Τα δεδομένα των παραδειγμάτων που θα υλοποιηθούν θα είναι από τον τραπεζικό χώρο, τον χώρο των τηλεπικοινωνιών και τον χώρο της ιατρικής. Θα δοθεί μεγάλη έμφαση στην επεξεργασία των δεδομένων, στην κύρια ανάλυση τους και στην παρουσίαση των αποτελεσμάτων κατά τρόπο ακριβή, συνεπή και κατανοητό.