

ΠΡΟΛΟΓΟΣ

Η γραμμική παλινδρόμηση χρησιμοποιείται για την μελέτη των σχέσεων μεταξύ μετρήσιμων μεταβλητών. Γενικότερα, η γραμμική στατιστική συμπερασματολογία αποτελεί ένα ευρύ πεδίο της στατιστικής ανάλυσης που περιλαμβάνει την προσαρμογή γραμμών και επιπέδων με την μέθοδο των ελαχίστων τετραγώνων, την ανάλυση διακύμανσης, την πειραματική ανάλυση, τη συσχέτιση, την πολυμεταβλητή ανάλυση και, κάποιο τρόπο, την ανάλυση χρονοσειρών.

Στο βιβλίο αυτό παρουσιάζονται μερικές από τις τεχνικές των γραμμικών στατιστικών μοντέλων που αναφέρονται κυρίως στην ανάλυση παλινδρόμησης (απλή και πολλαπλή παλινδρόμηση) και στις βασικές αρχές της ανάλυσης διακύμανσης. Οι τεχνικές αυτές εφαρμόζονται σε όλες, σχεδόν τις περιοχές επιστημονικής μελέτης όπως, π.χ. τις κοινωνικές επιστήμες, τις φυσικές και βιολογικές επιστήμες, τις επιχειρήσεις, την τεχνολογία και τις ανθρωπιστικές επιστήμες.

Για τον λόγο αυτό, η έμφαση που δίνεται στην παρουσίαση των θεμάτων συγκεντρώνεται στις εφαρμογές. Οι μαθηματικές απαιτήσεις έχουν κρατηθεί σε χαμηλό επίπεδο, ώστε ο αναγνώστης να έχει την ευχέρεια να συγκεντρώνει την προσοχή του στο στατιστικό τρόπο σκέψης κυρίως και όχι στη μαθηματική ανάλυση μεθόδων. Έμφαση όμως, έχει δοθεί και στις υποθέσεις που θα πρέπει να πληρούνται ώστε να εφαρμόζονται οι τεχνικές αυτές, όπως επίσης και στους δυνατούς τρόπους που οι υποθέσεις αυτές ελέγχονται. Αυτό γιατί όπως είναι γνωστό, οι μέθοδοι των γραμμικών μοντέλων χρησιμοποιούνται σήμερα ευρύτατα χωρίς ιδιαίτερη προσοχή στις υποθέσεις που τις συνοδεύουν, με αποτέλεσμα να εξάγονται συμπεράσματα που δεν ανταποκρίνονται στα δεδομένα από τα οποία έχουν προέλθει.

Παρότι η παρουσίαση των θεμάτων των γραμμικών μοντέλων γίνεται εξαιρετικά απλούστερη με τη χρησιμοποίηση της θεωρία πινάκων, η παρουσίαση στο βιβλίο αυτό έχει γίνει με τον παραδοσιακό τρόπο. Αυτό έγινε για τους ίδιους λόγους που αναφέρθηκαν παραπάνω. Για την διευκόλυνση του αναγνώστη, όμως, που έχει περισσότερο ενδιαφέρον και μεγαλύτερη ευχέρεια με τις τεχνικές της Γραμμικής Άλγεβρας, παρουσιάζεται η βασική θεματολογία της πολλαπλής παλινδρόμησης σε ειδική ενότητα, με τη χρησιμοποίηση της Θεωρίας Πινάκων. Έτσι, ο αναγνώστης έχει την ευχέρεια αφενός μεν να κατανοήσει τις τεχνικές αφετέρου δε αν τον ενδιαφέρει, να αντιληφθεί πώς η χρήση των μεθόδων της Γραμμικής Άλγεβρας απλουστεύει την ανάπτυξη των τεχνικών.

Ένας άλλος τομέας στον οποίο έχει δοθεί έμφαση είναι η παρουσίαση παραδειγμάτων. Ο συγγραφέας θεωρεί ότι τα παραδείγματα είναι ο καλύτερος

τρόπος για να εμπεδωθεί η θεωρία. Για την επίλυση των παραδειγμάτων χρησιμοποιούνται στατιστικά πακέτα, ώστε να είναι δυνατή η ανάλυση μεγάλου όγκου στοιχείων.

Οι μέθοδοι της πολλαπλής παλινδρόμησης παρουσιάζονται συνήθως στην Ελληνική βιβλιογραφία με τον κλασσικό θεωρητικό τρόπο. Είναι, όμως, γνωστό ότι η ανάπτυξη των στατιστικών πακέτων έχει κάνει ιδιαίτερα εύκολη τη χρησιμοποίηση των μεθόδων της πολλαπλής παλινδρόμησης και την ανάλυση στοιχείων από προβλήματα που μπορούν να αντιμετωπισθούν με τον τρόπο αυτό. Επειδή ακριβώς τα περισσότερα πρακτικά προβλήματα χρειάζονται μοντέλα στα οποία να χρησιμοποιούνται περισσότερες από μια μεταβλητές για να τα περιγράψουν, το βιβλίο δίνει έμφαση και στις διάφορες μεθόδους, εφαρμογές και τεχνικές της πολλαπλής παλινδρόμησης.

Οι στατιστικές γνώσεις που απαιτούνται από τον αναγνώστη είναι εκείνες που αναφέρονται σε ελέγχους υποθέσεως, σε διαστήματα εμπιστοσύνης και σε μερικές γνωστές κατανομές, όπως η κανονική, η t και η F .

Ιδιαίτερη προσοχή έχει δοθεί στον συμβολισμό ώστε να ανταποκρίνεται ακριβώς στις έννοιες που έχουν χρησιμοποιηθεί. Αυτό έχει γίνει παρά τον ορατό κίνδυνο να θεωρηθεί ο συμβολισμός πολύπλοκος. Άποψη του συγγραφέα, όμως, είναι ότι ο αυστηρός προσδιορισμός και χαρακτηρισμός των εννοιών που χρησιμοποιούνται βοηθά τελικά τον αναγνώστη να κατανοήσει σωστότερα και πληρέστερα τις έννοιες που αναπτύσσονται, έστω και αν αντιμετωπίζει μια προσωρινή δυσκολία με τους πολύπλοκους συμβολισμούς.

Το βιβλίο αυτό είναι το καταστάλαγμα της επεξεργασίας των παραδόσεων του συγγραφέα από σχετικά μαθήματα που έχει διδάξει σε μια σειρά από φοιτητές, Τμήματα και Πανεπιστήμια της Ελλάδας και του εξωτερικού. Η δεύτερη αυτή έκδοση έχει ορισμένες προσθήκες και βελτιώσεις σε σχέση με την προηγούμενη που κυκλοφόρησε το 1993. Μερικά από τα θέματα που έχουν προστεθεί είναι ο έλεγχος Durbin-Watson, η παρουσίαση του προβλήματος της πολυσυγγραμμικότητας και οι βασικές αρχές της μεθόδου αμφικλινούς παλινδρόμησης (ridge regression) για την αντιμετώπιση του προβλήματος αυτού.

Στο πρώτο κεφάλαιο παρουσιάζεται η βασική λογική που διέπει τα γραμμικά μοντέλα, και κυρίως τα μοντέλα απλής γραμμικής παλινδρόμησης. Επίσης αναπτύσσεται η στατιστική συμπεραματολογία για τις παραμέτρους των μοντέλων αυτών, όπως επίσης και η εξέταση των καταλοίπων για τον έλεγχο των υποθέσεων που απαιτούνται. Επίσης παρουσιάζονται και αναλύονται οι έννοιες του συντελεστή προσδιορισμού και του συντελεστή συσχέτισης. Γίνεται μια λεπτομερής παρουσίαση

δύο βασικών στατιστικών πακέτων που χρησιμοποιούνται στην ανάλυση γραμμικής παλινδρόμησης, δηλαδή του MINITAB και του SAS. Στη συνέχεια, αναπτύσσονται οι έννοιες της σειριακής συσχέτισης των καταλοίπων και ο έλεγχος Durbin-Watson για την ύπαρξη αυτοπαλινδρόμησης πρώτης τάξης. Για τις περιπτώσεις όπου οι υποθέσεις του γραμμικού μοντέλου δεν πληρούνται ή δεν είναι δυνατόν να ελεγχθούν, δίνεται η τεχνική της ,η παραμετρικής μονότονης παλινδρόμησης όπως και οι προβλέψεις που προκύπτουν από αυτήν. Στην συνέχεια παρατίθενται διάφορες εναλλακτικές μορφές του γενικού γραμμικού μοντέλου. Το κεφάλαιο αυτό συμπληρώνεται με την ανάπτυξη της θεωρίας για τη διμεταβλητή κανονική κατανομή ως εργαλείο για στατιστική συμπερασματολογία για το συντελεστή συσχέτισης ρ .

Το δεύτερο κεφάλαιο ασχολείται με την Ανάλυση Διακύμανσης κατά ένα και κατά δύο κριτήρια.

Το τρίτο και τελευταίο κεφάλαιο αναφέρεται στην πολλαπλή παλινδρόμηση, στην συμπερασματολογία για αυτήν, και στις τρεις κυριότερες μεθόδους επιλογής μεταβλητών για ένα μοντέλο πολλαπλής παλινδρόμησης. Του αποκλεισμού μεταβλητών, της προοδευτικής προσθήκης μεταβλητών και την ευρύτερα χρησιμοποιούμενη μέθοδο της βηματικής παλινδρόμησης. Τέλος εξετάζονται το κριτήριο C_p του Mallows, δίνεται συνοπτικά η παρουσίαση της γραμμικής παλινδρόμησης με τη χρήση πινάκων και αναπτύσσονται τα προβλήματα που αναφύονται από την πολλυσυγγραμμικότητα, ενώ γίνεται μια συνοπτική αναφορά στη μέθοδο της αμφικλινούς παλινδρόμησης (ridge regression) για το ξεπέραςμα των προβλημάτων αυτών.

Ο συγγραφέας θα ήθελε να ευχαριστήσει όλους εκείνους που με τις υποδείξεις τους βοήθησαν στη βελτίωση της παρουσίασης των θεμάτων από το βιβλίο αυτό πραγματεύεται.

Αθήνα 1994

Ι. Πανάρετος

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ Ι

ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

1. ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ ΚΑΙ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
 - 1.1 Εκτίμηση της Ευθείας Παλινδρόμησης με τη Μέθοδο των Ελαχίστων Τετραγώνων
 - 1.2 Συμπερασματολογία στη Γραμμική Παλινδρόμηση
 - 1.2.1 Διαστήματα Εμπιστοσύνης για το β
 - 1.2.2 Διαστήματα Εμπιστοσύνης για το α
 - 1.2.3 Διαστήματα Εμπιστοσύνης για την $\mu_{Y|x}$
 - 1.2.4 Διαστήματα Πρόβλεψης
 - 1.3 Έλεγχοι Υποθέσεων
 - 1.3.1 Έλεγχος Υποθέσεων για το β
 - 1.3.2 Μερικά Γενικά Παραδείγματα
 - 1.4 Χρησιμοποίηση των καταλοίπων για την Εξέταση των Υποθέσεων του Γραμμικού Μοντέλου
 - 1.5 Συντελεστής Προσδιορισμού
 - 1.6 Συντελεστής Συσχέτισης
2. ΧΡΗΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΠΑΚΕΤΩΝ ΣΤΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
 - 2.1 Γραμμική Παλινδρόμηση με το MINITAB
 - 2.2 Γραμμική Παλινδρόμηση με το SAS
3. ΣΕΙΡΙΑΚΟΣ ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ
 - 3.1 Σειριακή Συσχέτιση των Καταλοίπων
 - 3.2 Μοντέλο Αυτοπαλινδρόμησης Πρώτης Τάξης
 - 3.3 Έλεγχος Durbin-Watson
 - 3.4 Έλεγχος Durbin-Watson με το MINITAB και το SAS

4. ΠΑΛΙΝΔΡΟΜΗΣΗ ΤΑΞΗΣ ΜΕΓΕΘΟΥΣ
 - 4.1 Μονότονη Παλινδρόμηση
 - 4.2 Προβλέψεις στη Μονότονη Παλινδρόμηση
5. ΤΟ ΓΕΝΙΚΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ
 - 5.1 Εναλλακτικά Μοντέλα του Απλού Γραμμικού Μοντέλου:
Το Εκθετικό Μοντέλο
 - 5.2 Άλλα Μοντέλα για Διμεταβλητά Δεδομένα
 - 5.2.1 Προσαρμογή Μοντέλων με Τρεις ή Περισσότερες Παραμέτρους
 - 5.2.2 Ένα Άλλο Εναλλακτικό Μοντέλο
 - 5.2.3 Παρατηρήσεις και Συγκρίσεις
 - 5.2.4 Τελικά Συμπεράσματα
6. ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΣΤΗ ΓΡΑΜΜΙΚΗ ΣΥΣΧΕΤΙΣΗ
 - 6.1 Η Διμεταβλητή Κανονική Κατανομή
 - 6.2 Συμπερασματολογία για το ρ
 - 6.3 Έλεγχοι Υποθέσεων για $\rho=0$
 - 6.4 Διαστήματα Εμπιστοσύνης για $\rho \neq 0$
 - 6.5 Έλεγχοι Υποθέσεων για $\rho \neq 0$

ΑΣΚΗΣΕΙΣ

ΚΕΦΑΛΑΙΟ II

ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

1. ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ ΚΑΤΑ ΕΝΑ ΚΡΙΤΗΡΙΟ
 - 1.1 Πλήρως Τυχαιοποιημένος Σχεδιασμός
2. ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ ΚΑΤΑ ΔΥΟ ΚΡΙΤΗΡΙΑ
 - 2.1 Σχεδιασμός Τυχαιοποιημένων Πλήρων Σχημάτων

- 2.2 Ανάλυση Διακύμανσης για το Σχεδιασμό Τυχαιοποιημένων Πλήρων Στρωμάτων
- 2.3 Αλληλεπιδράσεις σε Πειράματα Δύο Παραγόντων
- 2.4 Σχεδιασμός Τυχαιοποιημένων Πλήρων Σχημάτων με Αλληλεπιδράσεις

ΑΣΚΗΣΕΙΣ

ΚΕΦΑΛΑΙΟ ΙΙΙ ΠΟΛΛΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

- 1. ΓΕΝΙΚΗ ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΠΟΛΛΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
 - 1.1 Μοντέλο με Δύο Ανεξάρτητες Μεταβλητές
 - 1.2 Μοντέλο με k Ανεξάρτητες Μεταβλητές
 - 1.3 Μέθοδος των Ελαχίστων Τετραγώνων
 - 1.4 Ερμηνεία του Πίνακα Πολλαπλής Παλινδρόμησης
 - 1.5 Ένας Μερικός Έλεγχος του Μοντέλου
 - 1.6 Διαστήματα Εμπιστοσύνης για τα β
 - 1.7 Συντελεστής Μερικής Συσχέτισης
 - 1.8 Ορθογωνιότητα
- 2. ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗ ΜΕΘΟΔΟ ΤΟΥ ΑΠΟΚΛΕΙΣΜΟΥ ΜΕΤΑΒΛΗΤΩΝ
 - 2.1 Η Πιθανότητα Λάθους Πρώτου Είδους
 - 2.2 Διάστημα Εμπιστοσύνης για τον Μέσο $\mu_{Y|x}$
 - 2.3 Διάστημα Πρόβλεψης για το Y
- 3. ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗ ΜΕΘΟΔΟ ΤΗΣ ΠΡΟΟΔΕΥΤΙΚΗΣ ΠΡΟΣΘΗΚΗΣ ΜΕΤΑΒΛΗΤΩΝ

4. ΜΕΘΟΔΟΣ ΤΗΣ ΒΗΜΑΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

- 4.1 Βήμα 1ο στη Μέθοδο της Βηματικής Παλινδρόμησης
- 4.2 Βήμα 2ο στη Μέθοδο της Βηματικής Παλινδρόμησης
- 4.3 Βήμα 3ο στη Μέθοδο της Βηματικής Παλινδρόμησης
- 4.4 Σύγκριση της Μεθόδου της Βηματικής Παλινδρόμησης με τη Μέθοδο Αποκλεισμού Μεταβλητών
- 4.5 Μέθοδοι Όλων των Πιθανών Παλινδρομήσεων
- 4.6 Διαστήματα Εμπιστοσύνης και Πρόβλεψης με τις Μεθόδους του Αποκλεισμού Μεταβλητών και της Βηματικής Παλινδρόμησης

5. ΤΟ ΚΡΙΤΗΡΙΟ C_p ΤΟΥ MALLOWS

6. ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ ΜΕ ΤΗ ΧΡΗΣΗ ΠΙΝΑΚΩΝ

- 6.1 Πίνακας Διακυμάνσεων-Συνδιακυμάνσεων του ε
- 6.2 Εκτιμήτριες Ελαχίστων Τετραγώνων

7. ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΠΟΛΥΣΥΓΓΡΑΜΙΚΟΤΗΤΑΣ

- 7.1 Αμφικλινής Παλινδρόμηση

ΑΣΚΗΣΕΙΣ

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΠΑΡΑΡΤΗΜΑ

Πίνακες