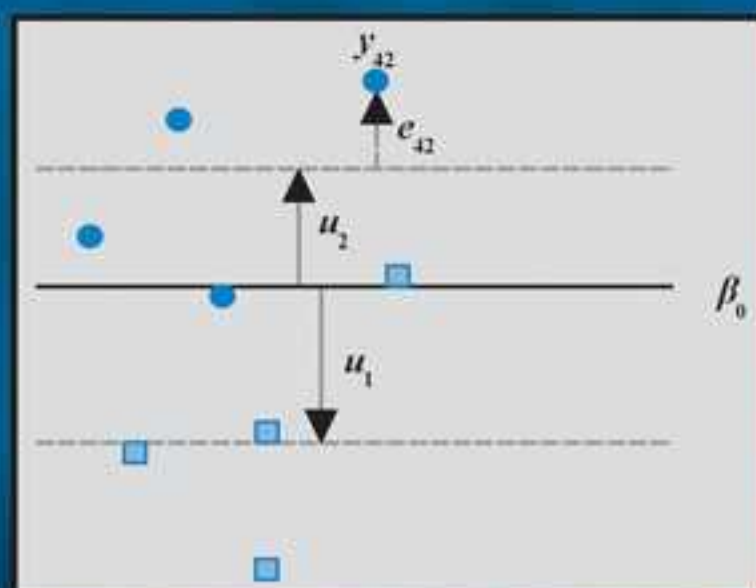


ΑΝΑΛΥΣΗ ΠΟΛΥΜΕΤΑΒΛΗΤΩΝ ΤΕΧΝΙΚΩΝ ΣΤΙΣ ΚΟΙΝΩΝΙΚΕΣ ΕΠΙΣΤΗΜΕΣ

Δεύτερη αμερικανική έκδοση



DAVID J. BARTHOLOMEW
FIONA STEELE
IRINI MOUSTAKI
JANE I. GALBRAITH



Επιστημονική επιμέλεια ελληνικής έκδοσης:
Ειρήνη Μουστάκη, *London School of Economics*

Περιεχόμενα

Πρόλογος	11
1 Προετοιμασία	17
1.1 Η δομή του βιβλίου.....	17
1.2 Η περιορισμένη χρήση των μαθηματικών	21
1.3 Μεταβλητές	28
1.4 Η γεωμετρία της πολυμεταβλητής ανάλυσης.....	31
1.5 Χρήση παραδειγμάτων	32
1.6 Εξέταση δεδομένων, μετασχηματισμοί και ελλείποντα δεδομένα	35
1.7 Βιβλιογραφία	37
2 Ανάλυση συστοιχιών	39
2.1 Ταξινόμηση στις κοινωνικές επιστήμες	39
2.2 Μερικές μέθοδοι ανάλυσης συστοιχιών	44
2.3 Γραφική παρουσίαση των αποτελεσμάτων	51
2.4 Παραγωγή της μήτρας απόστασης	55
2.5 Παράδειγμα με διαλέκτους της αγγλικής γλώσσας.....	61
2.6 Συγκρίσεις.....	69
2.7 Ομαδοποίηση μεταβλητών	71
2.8 Πρόσθετα παραδείγματα και περαιτέρω εργασία.....	72
2.9 Πρόσθετη βιβλιογραφία	87
3 Πολυδιάστατη προσαρμογή κλίμακας	89
3.1 Εισαγωγή	89
3.2 Παραδείγματα.....	92
3.3 Κλασική, διατακτική και μετρική πολυδιάστατη προσαρμογή κλίμακας	96

3.4	Σχόλια για τις υπολογιστικές διαδικασίες	101
3.5	Αξιολόγηση προσαρμογής και επιλογή πλήθους διαστάσεων.....	102
3.6	Ένα πρακτικό παράδειγμα: διαστάσεις της χρωματικής όρασης	104
3.7	Πρόσθετα παραδείγματα και περαιτέρω εργασία.....	109
3.8	Πρόσθετη βιβλιογραφία	122
4	Ανάλυση αντιστοιχίας	123
4.1	Στόχοι της ανάλυσης αντιστοιχίας.....	123
4.2	Εκτέλεση ανάλυσης αντιστοιχίας	126
4.3	Εκτέλεση ανάλυσης αντιστοιχίας: γενική μέθοδος	133
4.4	Το γράφημα διπλής προβολής	137
4.5	Ερμηνεία των διαστάσεων.....	144
4.6	Επιλογή πλήθους διαστάσεων	146
4.7	Παράδειγμα: εμπιστοσύνη στις αγορές από χώρες της Ευρωπαϊκής Ένωσης.....	149
4.8	Ανάλυση αντιστοιχίας για πίνακες πολλαπλής εισόδου	157
4.9	Πρόσθετα παραδείγματα και περαιτέρω εργασία.....	161
4.10	Πρόσθετη βιβλιογραφία	169
5	Ανάλυση κύριων συνιστωσών	171
5.1	Εισαγωγή	171
5.2	Μερικές πιθανές εφαρμογές	173
5.3	Παράδειγμα της PCA για δύο μεταβλητές	174
5.4	Γενική περιγραφή της PCA	178
5.5	Παραδείγματα.....	181
5.6	Βαθμολογίες συνιστωσών	190
5.7	Ο σύνδεσμος μεταξύ PCA και MDS, και μεταξύ PCA και CORA.....	193
5.8	Αντικατάσταση των αρχικών μεταβλητών με τις βαθμολογίες των κύριων συνιστωσών	196
5.9	Πρόσθετα παραδείγματα και περαιτέρω εργασία.....	198
5.10	Πρόσθετη βιβλιογραφία	205

6	Ανάλυση παλινδρόμησης	207
6.1	Βασικές ιδέες	207
6.2	Απλή γραμμική παλινδρόμηση	209
6.3	Ένα μοντέλο πιθανότητας για την απλή γραμμική παλινδρόμηση	214
6.4	Συμπερασματολογία για το μοντέλο της απλής γραμμικής παλινδρόμησης	217
6.5	Έλεγχος των υποθέσεων	220
6.6	Πολλαπλή παλινδρόμηση	222
6.7	Παραδείγματα πολλαπλής παλινδρόμησης	223
6.8	Εκτίμηση και συμπερασματολογία για τις παραμέτρους	225
6.9	Ερμηνεία των συντελεστών παλινδρόμησης	228
6.10	Επιλογή παλινδρομουσών μεταβλητών	232
6.11	Μετασχηματισμοί και αλληλεπιδράσεις	234
6.12	Λογιστική παλινδρόμηση	237
6.13	Ανάλυση διαδρομών (path analysis)	241
6.14	Πρόσθετα παραδείγματα και περαιτέρω εργασία	245
6.15	Πρόσθετη βιβλιογραφία	250
 7	 Παραγοντική ανάλυση	 253
7.1	Εισαγωγή στα μοντέλα λανθανουσών μεταβλητών	253
7.2	Το γραμμικό μοντέλο ενός παράγοντα	258
7.3	Το γενικό γραμμικό μοντέλο παραγόντων	262
7.4	Ερμηνεία	266
7.5	Επάρκεια μοντέλου και επιλογή πλήθους παραγόντων	269
7.6	Περιστροφή	273
7.7	Βαθμολογίες παραγόντων	278
7.8	Ένα πρακτικό παράδειγμα: ο κατάλογος απογραφής του άγχους εξετάσεων	280
7.9	Η βοήθεια της περιστροφής στην ερμηνεία	286
7.10	Σύγκριση της παραγοντικής ανάλυσης και της ανάλυσης κύριων συνιστωσών	288
7.11	Πρόσθετα παραδείγματα και περαιτέρω εργασία	291
7.12	Λογισμικό	298
7.13	Πρόσθετη βιβλιογραφία	298

8	Παραγοντική ανάλυση για δυαδικά δεδομένα	299
8.1	Μοντέλα λανθανόντων χαρακτηριστικών	299
8.2	Γιατί είναι το μοντέλο παραγοντικής ανάλυσης για μετρικές μεταβλητές ακατάλληλο για δυαδικές απαντήσεις; ..	303
8.3	Μοντέλο παραγόντων για δυαδικά δεδομένα με τη μέθοδο της θεωρίας της απάντησης ερωτήσεων	306
8.4	Καλή προσαρμογή	312
8.5	Βαθμολογίες παραγόντων	318
8.6	Περιστροφή	320
8.7	Μέθοδος των υποκείμενων μεταβλητών	321
8.8	Παράδειγμα: στάσεις απέναντι σε σεξουαλικά ζητήματα	323
8.9	Πρόσθετα παραδείγματα και περαιτέρω εργασία	330
8.10	Λογισμικό	341
8.11	Πρόσθετη βιβλιογραφία	342
9	Παραγοντική ανάλυση για διατεταγμένες κατηγορικές μεταβλητές	343
9.1	Το πρακτικό υπόβαθρο	343
9.2	Δύο μέθοδοι για τη μοντελοποίηση διατεταγμένων κατηγορικών δεδομένων	345
9.3	Η μέθοδος της συνάρτησης απάντησης ερωτήσεων	347
9.4	Παραδείγματα	355
9.5	Η μέθοδος των υποκείμενων μεταβλητών	360
9.6	Μη και μερικώς διατεταγμένες παρατηρούμενες μεταβλητές	367
9.7	Πρόσθετα παραδείγματα και περαιτέρω εργασία	372
9.8	Λογισμικό	380
9.9	Πρόσθετη βιβλιογραφία	381
10	Ανάλυση λανθανουσών κλάσεων για δυαδικά δεδομένα	383
10.1	Εισαγωγή	383
10.2	Το μοντέλο λανθανουσών κλάσεων για δυαδικά δεδομένα ...	386
10.3	Παράδειγμα: δεδομένα της στάσης απέναντι στην επιστήμη και την τεχνολογία	392

10.4	Πώς μπορούμε να ξεχωρίσουμε το μοντέλο λανθανουσών κλάσεων από το μοντέλο λανθανόντων χαρακτηριστικών; .	398
10.5	Ανάλυση λανθανουσών κλάσεων, ανάλυση συστοιχιών, και ανάλυση λανθανόντων προφίλ	400
10.6	Πρόσθετα παραδείγματα και περαιτέρω εργασία.....	402
10.7	Λογισμικό	407
10.8	Πρόσθετη βιβλιογραφία	407
11	Επιβεβαιωτική παραγοντική ανάλυση και μοντέλα δομικών εξισώσεων	409
11.1	Εισαγωγή	409
11.2	Διάγραμμα διαδρομών.....	412
11.3	Μοντέλα μέτρησης	414
11.4	Επάρκεια του μοντέλου	421
11.5	Εισαγωγή στα μοντέλα δομικών εξισώσεων με λανθάνουσες μεταβλητές	426
11.6	Το γραμμικό μοντέλο δομικών εξισώσεων	427
11.7	Ένα πρακτικό παράδειγμα	441
11.8	Επεκτάσεις	447
11.9	Πρόσθετα παραδείγματα και περαιτέρω εργασία.....	447
11.10	Λογισμικό	456
11.11	Πρόσθετη βιβλιογραφία	456
12	Πολυεπίπεδη μοντελοποίηση	459
12.1	Εισαγωγή	459
12.2	Πιθανές εφαρμογές.....	461
12.3	Σύγκριση ομάδων με χρήση πολυεπίπεδης μοντελοποίησης	462
12.4	Μοντέλο τυχαίου σταθερού όρου.....	472
12.5	Μοντέλο τυχαίας κλίσης.....	475
12.6	Θεματικές επιδράσεις	480
12.7	Πολυεπίπεδη πολυμεταβλητή παλινδρόμηση.....	484
12.8	Πολυεπίπεδη παραγοντική ανάλυση	493
12.9	Πρόσθετα παραδείγματα και περαιτέρω εργασία.....	498
12.10	Επιπλέον θέματα.....	500

12.11	Διαδικασίες και λογισμικό εκτίμησης	502
12.12	Πρόσθετη βιβλιογραφία	503
Βιβλιογραφία		505
Ευρετήριο		515

Πρόλογος

Η πολυμεταβλητή ανάλυση είναι ένα σημαντικό εργαλείο για τον κοινωνικό ερευνητή. Ο στόχος του βιβλίου αυτού είναι να βοηθήσει τους φοιτητές και τους κοινωνικούς ερευνητές με περιορισμένες γνώσεις μαθηματικών και στατιστικής να κατανοήσουν μερικές από τις κύριες πολυμεταβλητές μεθόδους, μαζί με τον τρόπο εφαρμογής τους.

Ο τίτλος της πρώτης έκδοσης ήταν μεγαλύτερος: *Η ανάλυση και ερμηνεία των πολυμεταβλητών δεδομένων στις κοινωνικές επιστήμες*. Η αλλαγή έγινε απλώς για λόγους συντομίας και δεν αντανakλά κάποια αλλαγή στον προσανατολισμό. Η ερμηνεία συνεχίζει να αποτελεί ένα στοιχείο-κλειδί.

Το βιβλίο έχει τις ρίζες του σε ένα μάθημα το οποίο δίδασκε ο πρώτος συγγραφέας για περισσότερα από 20 χρόνια σε μεταπτυχιακούς φοιτητές κοινωνικών επιστημών στο London School of Economics. Το περιεχόμενο άλλαξε με την πάροδο των χρόνων, αλλά ο κύριος στόχος παρέμεινε ο ίδιος. Οι υπόλοιποι συγγραφείς συνεισέφεραν στο μάθημα τα επόμενα χρόνια με διάφορες ιδιότητες, αλλά από τότε διασκορπίστηκαν σε διάφορα μέρη. Στη νέα του μορφή, το βιβλίο μπορεί ακόμη να αποτελέσει το βασικό σύγγραμμα κάποιου μαθήματος, αλλά κατά την άποψή μας, ο σημαντικότερος ρόλος του είναι να χρησιμεύσει ως εγχειρίδιο για τους κοινωνικούς ερευνητές.

Στην πρώτη έκδοση, εστιάσαμε αποκλειστικά στις πολυμεταβλητές μεθόδους για να αποσαφηνίσουμε τις αλληλοσυσχετίσεις μεταξύ μεταβλητών που έχουν όλες την ίδια κατάσταση. Δεν καλύψαμε, για παράδειγμα, τα αιτιώδη μοντέλα, όπως η ανάλυση παλινδρόμησης, η ανάλυση διαδρομών, τα γραμμικά μοντέλα δομικών σχέσεων, τα πολυεπίπεδα μοντέλα, κ.ο.κ. Σε αυτή τη νέα έκδοση, η θεματολογία έχει διευρυνθεί ώστε να περιλαμβάνει τέτοια θέματα. Τα νέα κεφάλαια είναι: το Κεφάλαιο 6 για την ανάλυση παλινδρόμησης, το Κεφάλαιο 11 για την επιβεβαιωτική παραγοντική ανάλυση και τα μοντέλα δομικών εξισώσεων, και το Κεφάλαιο 12 για τα πολυεπίπεδα μοντέλα. Το κεφάλαιο για την ανάλυση παλινδρόμησης αποτελεί ένα κομβικό σημείο, το οποίο όχι μόνο συνδέει τα δύο μισά του βιβλίου αλλά και αναφέρεται σε ένα θέμα που είναι

από μόνο του σημαντικό. Η ανάλυση παλινδρόμησης χαρακτηρίζει επίσης το σημείο της μετάβασης από τις περιγραφικές μεθόδους στις μεθόδους που βασίζονται σε μοντέλα, όπου η κύρια έμφαση μετατοπίζεται από την αλληλεξάρτηση στην εξάρτηση.

Είναι δύσκολο να μιλήσουμε με ακρίβεια για τα προαπαιτούμενα ενός τέτοιου μαθήματος. Ίσως ο καλύτερος τρόπος είναι να παραθέσουμε μερικά από τα θέματα που πρέπει κάποιος να έχει διδαχθεί σε εισαγωγικά μαθήματα. Αυτά πρέπει οπωσδήποτε να περιλαμβάνουν τα εξής: τη βασική περιγραφική στατιστική, στοιχεία δειγματοληψίας και συμπερασματολογίας (συμπεριλαμβανομένου του βασικού ελέγχου υποθέσεων και της εκτίμηση διαστημάτων), την αναπαράσταση κατηγορικών μεταβλητών σε πίνακες συνάφειας διπλής εισόδου, τα μέτρα συσχέτισης, τις δοκιμές ανεξαρτησίας, τη συσχέτιση και, ίσως, τη βασική ανάλυση διακύμανσης και παλινδρόμησης. Ο κατάλογος δεν είναι σε καμία περίπτωση πλήρης, αλλά τα θέματα που αναφέραμε μπορούν ενδεικτικά να προσδιορίσουν το εύρος της ύλης που θα πρέπει να έχουν καλύψει οι φοιτητές. Πέρα από τα παραπάνω, οι φοιτητές θα πρέπει να είναι εξοικειωμένοι με απλούς μαθηματικούς τύπους, καθώς και με τη χρήση και την ερμηνεία τους.

Η υλοποίηση όλων των πολυμεταβλητών μεθόδων απαιτεί τη χρήση υπολογιστή. Για όλα τα θέματα που καλύπτονται στο βιβλίο, το μόνο που χρειάζεται είναι ένας επιτραπέζιος υπολογιστής με τυπικά εμπορικά πακέτα λογισμικού, ή με το ειδικό λογισμικό που παρέχεται στον ιστότοπο του βιβλίου. Η ικανότητα χειρισμού τέτοιων προγραμμάτων σήμερα είναι κοινός τόπος μεταξύ φοιτητών και ερευνητών και, σύμφωνα με την εμπειρία μας, δεν θα αποτελέσει σημαντικό πρόβλημα.

Επειδή θέλαμε οι μαθηματικές μας απαιτήσεις να είναι οι ελάχιστες δυνατές, η παρουσίασή μας βασίζεται περισσότερο σε αριθμητικά παραδείγματα, σε λεκτικές περιγραφές και περιγραφές με εικόνες, παρά σε μαθηματικές αναπαραστάσεις. Ο στόχος μας είναι να παρέχουμε μια κατανόηση σε βάθος για τον σκοπό και τον τρόπο λειτουργίας των μεθόδων. Το συγκεκριμένο βιβλίο δεν είναι σε καμία περίπτωση ένας «οδηγός μαγειρικής». Η πολυμεταβλητή ανάλυση είναι ένα πολύπλοκο και ευαίσθητο εργαλείο, το οποίο θα μπορέσει κάποιος να μάθει καλύτερα δουλεύοντας με λεπτομερή παραδείγματα και προσπαθώντας να ερμηνεύσει τα αποτελέσματα της ανάλυσης με την καθοδήγηση ενός έμπειρου επαγγελματία. Αυτή είναι η μέθοδος που ακολουθούμε όταν διδάσκουμε, την οποία υποστηρίζουμε πάντοτε με πρακτική εργασία στον υπολογιστή.

Το βιβλίο έχει πολλά ξεχωριστά χαρακτηριστικά. Ένα από αυτά, το οποίο έχουμε ήδη αναφέρει, είναι ότι αντιμετωπίζει με σοβαρότητα και σε βάθος τα μεμονωμένα θέματα με έναν μη μαθηματικό τρόπο. Ένα δεύτερο χαρακτηριστικό είναι ότι έχουμε δώσει έμφαση στην ενότητα των μεθόδων, τις οποίες παρουσιάζουμε με τέτοια σειρά ώστε η εξάσκηση στα αρχικά στάδια να σας βοηθήσει να κατανοήσετε τις πιο δύσκολες μεθόδους που ακολουθούν αργότερα. Τα διάφορα μέρη του βιβλίου διασυνδέονται σε μεγάλο βαθμό μεταξύ τους, και θα έχετε το μέγιστο δυνατό όφελος αν ακολουθήσετε την παρουσίαση από την αρχή μέχρι το τέλος.

Υπάρχουν πολλά παραδείγματα, και ενδέχεται να παρατηρήσετε ότι τα περισσότερα προέρχονται από εφαρμογές στο Ηνωμένο Βασίλειο ή στην Ευρώπη. Αυτό είναι μια συνέπεια της απόφασής μας να επικεντρωθούμε σε υλικό με το οποίο έχουμε εργαστεί οι ίδιοι ή με το οποίο έχουμε κάποια σχέση. Πιστεύουμε ότι αυτό είναι σημαντικό, για να μπορέσουν οι αναγνώστες να έχουν την αίσθηση ότι βλέπουν την εφαρμογή και εσωτερικά και εξωτερικά.

Ένα ειδικό χαρακτηριστικό του βιβλίου είναι η λεπτομερής κάλυψη όσον αφορά τις μεθόδους λανθανουσών μεταβλητών. Οι διάφοροι κλάδοι του συγκεκριμένου θέματος εξακολουθούν να αντιμετωπίζονται συνήθως ως «στεγανά». Έτσι, για παράδειγμα, η θεωρία της απάντησης ερωτήσεων (Item Response Theory, IRT) έχει εξελιχθεί γύρω από τις ψυχομετρικές και τις εκπαιδευτικές εξετάσεις και διαθέτει τα δικά της περιοδικά και βιβλία. Υπάρχει ελάχιστη επαφή μεταξύ αυτού του πεδίου και της παραγοντικής ανάλυσης (factor analysis), για παράδειγμα. Αυτό είναι λυπηρό, επειδή τα δύο πεδία είναι εννοιολογικά ίδια — η μόνη διαφορά βρίσκεται στα επίπεδα μέτρησης των μεταβλητών. Στο παρόν βιβλίο, τέτοιες μέθοδοι συνενώνονται κάτω από μια κοινή ομπρέλα. Πιστεύουμε ότι τα μακροπρόθεσμα οφέλη αυτής της οπτικής είναι τεράστια, αλλά προκαλούνται βραχυπρόθεσμα προβλήματα. Δεν θα είναι τόσο εύκολο για εσάς να βρείτε επιπλέον παραδείγματα στη δημοσιευμένη βιβλιογραφία, τα οποία να μπορούν να συσχετιστούν άμεσα με το στυλ και τη σημειογραφία της δικής μας αντιμετώπισης. Ελπίζουμε ότι το βιβλίο θα αποτελέσει το έναυσμα μιας πιο ενοποιημένης θεώρησης για τη μοντελοποίηση λανθανουσών μεταβλητών.

Τα τελευταία χρόνια υπάρχει μια σημαντική αύξηση του ενδιαφέροντος για τα μοντέλα λανθανουσών μεταβλητών και, ειδικά τα Κεφάλαια 7-11, έχουν σχεδιαστεί να παρέχουν γερά θεμέλια για την πλήρη γκάμα των μοντέλων που

είναι διαθέσιμα σήμερα. Για εμάς, τα μοντέλα αυτά αποτελούν μια συνεκτική ομάδα από στενά συσχετισμένες μεθόδους, οι οποίες μπορούν να γίνουν σημαντικά εργαλεία για χρήση σε πολύ περισσότερους τομείς της κοινωνικής επιστημονικής έρευνας.

Το βιβλίο αποτελεί απλώς μία εισαγωγή. Οι αναγνώστες που θα ενδιαφερθούν για ένα συγκεκριμένο θέμα θα πρέπει να το διερευνήσουν περισσότερο ανατρέχοντας στην εξειδικευμένη βιβλιογραφία. Έχουμε συμπεριλάβει έναν συνοπτικό κατάλογο πρόσθετης βιβλιογραφίας στο τέλος κάθε κεφαλαίου που θα σας οδηγήσει σε σύγχρονες και περιεκτικές θεωρήσεις. Τα συγκεκριμένα βιβλία δεν παρέχουν απλώς μια έγκυρη και ευρύτερη περιγραφή από αυτή που μπορούμε να παρέχουμε εδώ, αλλά αποτελούν και σημεία εισόδου σε μια πολύ ευρύτερη βιβλιογραφία. Στο τέλος του βιβλίου υπάρχει, επίσης, ένας κατάλογος βιβλιογραφικών αναφορών, όπου μπορείτε να βρείτε τις πηγές των διαφόρων ερευνητικών αποτελεσμάτων από τα οποία αντλήσαμε στοιχεία. Συμπεριλάβαμε τις αναφορές για όσους από εσάς θελήσετε να εντρυφήσετε σε βάθος, αλλά ο κατάλογος δεν είναι σε καμία περίπτωση μια πλήρης λίστα της σχετικής ερευνητικής βιβλιογραφίας.

Θα βρείτε τον συνοδευτικό ιστότοπο του βιβλίου στη διεύθυνση:

<http://www.bristol.ac.uk/cmm/publications/aimdss-2nd-ed/>

Ο ιστότοπος περιλαμβάνει όλα τα σύνολα δεδομένων που αναφέρονται στο βιβλίο, το λογισμικό για την εκτέλεση αναλύσεων λανθανουσών μεταβλητών, και οδηγίες (στην αγγλική γλώσσα) για την εφαρμογή τους.

Η συγγραφή του βιβλίου ήταν μια συλλογική προσπάθεια στην οποία συμβάλλαμε σε διάφορα στάδια ως άτομα, σε ζευγάρια, αλλά και ως μία ομάδα. Όλοι μας έχουμε «βάλει ένα λιθαράκι» σε κάθε τμήμα του βιβλίου και η ενότητα που το διακρίνει είναι αποτέλεσμα της συνεργασίας αυτής.

Στον Πρόλογο της πρώτης έκδοσης, ευχαριστήσαμε όσους μας βοήθησαν όλα αυτά τα χρόνια με διάφορους τρόπους, και ειδικότερα τους Susanna Brown, Panagiota Tzamourani, Amani Siyam, Anastasia Kakou, Olafur Gylfason, Nick Allum, Colin Mills, Martin Knott, Rex Galbraith, Albert Satorra, και Colin Chalmers. Για την έκδοση αυτή θέλουμε να ευχαριστήσουμε τον Anders Skrondal για τις πολύτιμες συμβουλές του, τους Geoff Laslett και Elena

Erosheva για τα σχόλια που έκαναν για την πρώτη έκδοση, και, για μία ακόμη φορά, τον Rex Galbraith για την υποστήριξή του.

David Bartholomew
Fiona Steele
Irina Moustaki
Jane Galbraith

Δεκέμβριος 2007

Ανάλυση συστοιχιών

2.1 Ταξινόμηση στις κοινωνικές επιστήμες

Η ταξινόμηση είναι μία από τις βασικότερες λειτουργίες στην επιστημονική έρευνα. Είναι ιδιαίτερα σημαντική στις κοινωνικές επιστήμες, όπου συχνά δεν υπάρχει πλήρης θεωρητική κάλυψη και το πρώτο βήμα της έρευνας είναι συνήθως ο εντοπισμός κάποιου είδους μοτίβου στα δεδομένα. Οι μέθοδοι ταξινόμησης χρησιμοποιούνται εδώ και πολύ καιρό στη βιολογία, όπου η ομαδοποίηση των ατόμων σύμφωνα με το είδος και το γένος τους έχει αποτελέσει τη βάση για περαιτέρω εργασίες. Αν και ένα μέρος της δουλειάς στην ανάλυση συστοιχιών (cluster analysis) έγινε από βιολόγους που προσπαθούσαν να ταξινομήσουν φυτά, το σημαντικότερο ερέθισμα για την ανάπτυξη του αντικειμένου προήλθε από προβλήματα των κοινωνικών επιστημών, με την ευρεία ερμηνεία τους. Στην παρακάτω πολύ επιλεκτική λίστα φαίνονται οι λόγοι που θα μπορούσαν να προκαλέσουν το ενδιαφέρον μας για την εύρεση συστοιχιών και οι πρακτικοί σκοποί που θα μπορούσαν να εξυπηρετήσουν οι συστοιχίες.

- i) *Μάρκετινγκ*. Η ταχυδρομική διαφήμιση είναι πιθανό να είναι αποτελεσματικότερη αν απευθύνεται σε άτομα με παρόμοια χαρακτηριστικά, τα οποία είναι πιθανό να ανταποκριθούν με τον ίδιο τρόπο. Η τμηματοποίηση της αγοράς, όπως αποκαλείται, επιδιώκει να διαχωρίσει τον πληθυσμό-στόχο σε συστοιχίες (τμήματα), έτσι ώστε να μπορέσει να στοχεύσει καθεμία από αυτές με τρόπο που είναι πιθανότερο να επιτύχει θετική ανταπόκριση.
- ii) *Αρχαιολογία*. Τα αντικείμενα που κατασκευάστηκαν περίπου την ίδια εποχή ή από την ίδια ομάδα ανθρώπων είναι πιθανότερο να έχουν περισσότερες ομοιότητες σε σχέση με εκείνα που προέρχονται από διαφορετικές εποχές ή λαούς. Με τον σχηματισμό συστοιχιών που έχουν παρόμοια αντικείμενα, μπορούμε να ανακατασκευάσουμε ένα κομμάτι της ιστορίας μιας περιοχής.

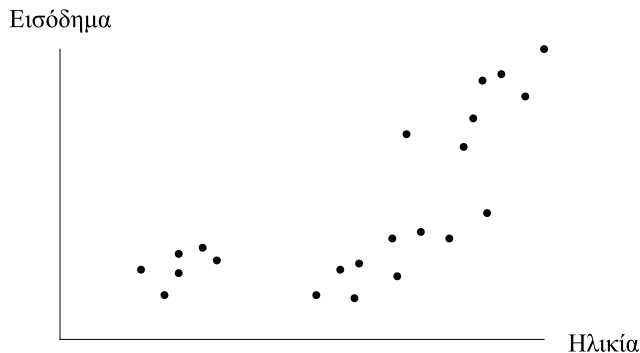
- iii) *Εκπαίδευση*. Οι επιδόσεις των σχολείων διαφέρουν μεταξύ τους, και όταν αναζητούμε τους λόγους γι' αυτή τη διαφοροποίηση, ίσως είναι χρήσιμο να ταξινομήσουμε τα σχολεία σε συστοιχίες έτσι ώστε ο ερευνητής να αναζητήσει τα κοινά στοιχεία εκείνων των σχολείων που φαίνονται να είναι γενικώς παρόμοια.

Στο κεφάλαιο αυτό θα περιγράψουμε και θα δείξουμε με παραδείγματα αρκετές μεθόδους ανάλυσης συστοιχιών, οι οποίες είναι εύκολες στην εφαρμογή και χρησιμοποιούνται ευρέως.

Το πρόβλημα που επιδιώκει να λύσει η ανάλυση συστοιχιών είναι η ομαδοποίηση ατόμων με τέτοιο τρόπο ώστε αυτά που κατανέμονται σε μια συγκεκριμένη ομάδα να είναι, κατά κάποια έννοια, κοντά το ένα στο άλλο. Αυτό μπορεί να γίνει με απλό τρόπο, αν τα αντικείμενα χαρακτηρίζονται από μία μετρήσιμη ποσότητα όπως είναι το εισόδημα. Το μόνο που πρέπει να κάνουμε είναι να ομαδοποιήσουμε εκείνα τα άτομα που έχουν παρόμοιο εισόδημα. Είναι αλήθεια ότι θα πρέπει να αποφασίσουμε τι εννοούμε με τη λέξη «παρόμοιο», αλλά αυτό θα καθοριστεί από την επιδιωκόμενη χρήση της ταξινόμησης. Το πρόβλημα είναι δυσκολότερο αν τα κριτήρια της ομοιότητας είναι υποκειμενικά ή βασίζονται σε μεγάλο πλήθος χαρακτηριστικών των αντικειμένων. Για παράδειγμα, όταν κρίνουμε την ομοιότητα δύο σχολείων, υπάρχει συνήθως ένα ολόκληρο σύνολο πιθανώς σχετικών χαρακτηριστικών, όπως είναι το μέγεθος, η τοποθεσία, τα ποσοστά των διαφόρων εθνικοτήτων, κ.ο.κ. Το ερώτημα που προκύπτει είναι πώς θα συνοψίσουμε τα διαφορετικά κομμάτια πληροφοριών έτσι ώστε να δημιουργήσουμε βάσιμα κριτήρια ομοιότητας. Αυτό το χαρακτηριστικό καθιστά την ανάλυση συστοιχιών μια πολυμεταβλητή τεχνική.

Για να δείτε τι συμβαίνει όταν τα κριτήρια των αποστάσεων βασίζονται σε περισσότερες από μία μεταβλητές, θεωρήστε την περίπτωση όπου έχουμε δύο μεταβλητές, καθεμία από τις οποίες μετριέται σε συνεχή κλίμακα. Αν τα «αντικείμενα» ήταν άνθρωποι, θα μπορούσαμε να υποθέσουμε ότι έχουμε εγγραφές που αφορούν την ηλικία και τα εισοδήματά τους, και θα θέλαμε να τους ομαδοποιήσουμε με βάση τις δύο μεταβλητές. Ας υποθέσουμε ότι σχεδιάζουμε ένα διάγραμμα στο οποίο τα άτομα είναι σημεία σε ένα επίπεδο. Τότε η θέση τους μπορεί να εμφανίζεται όπως στην Εικόνα 2.1.

Αν υπήρχε κάποια ομαδοποίηση σε συστοιχίες (clustering), θα την αναγνωρίζαμε αμέσως στην εικόνα. Στην περίπτωση αυτή, υπάρχουν τρεις τέτοιες συστοιχίες τις οποίες αναγνωρίζουμε χάρη στην ικανότητα του ματιού να εντοπί-



Εικόνα 2.1 Φανταστικά δεδομένα που δείχνουν το εισόδημα και την ηλικία αρκετών ανθρώπων

ζει μοτίβα ομαδοποίησης. Με τρεις μεταβλητές, θα μπορούσαμε να φανταστούμε σημεία σε ένα τριδιάστατο γράφημα, αλλά από εκεί και πέρα, δεν έχουμε δυνατότητα οπτικοποίησης, και χρειαζόμαστε κάποιον άλλο τρόπο αναγνώρισης των συστοιχιών. Συμπτωματικά, παρατηρήστε ότι αν είχαμε πληροφορίες μόνο για μία από τις μεταβλητές — έστω για το εισόδημα — θα βλέπαμε μόνο δύο συστοιχίες, επειδή οι δύο ηλικιακές ομάδες με χαμηλό εισόδημα θα ήταν δύσκολο να διαχωριστούν. Αυτό δείχνει πώς η πολυμεταβλητή ανάλυση δεδομένων μπορεί να αποκαλύψει περισσότερα από την ανάλυση κάθε μεταβλητής χωριστά.

Στο συγκεκριμένο παράδειγμα, η απόσταση μεταξύ δύο ατόμων ορίζεται απλώς ως η μεταξύ τους απόσταση στο γράφημα. Ωστόσο, οι πληροφορίες που έχουμε μπορεί να μην έχουν τη μορφή μετρήσεων σε συνεχή κλίμακα. Για παράδειγμα, μπορεί να υποδεικνύουν απλώς αν τα άτομα έχουν μια συγκεκριμένη ιδιότητα ή όχι. Σε τέτοιες περιπτώσεις, η γεωμετρική απεικόνιση δεν είναι διαθέσιμη, και πρέπει να γίνει κάποια προκαταρκτική εργασία για να αποφασίσουμε πώς θα μετρήσουμε την απόσταση μεταξύ των ατόμων.

Παρόλο που το παραπάνω παράδειγμα είναι στοιχειώδες, μας βοηθάει να προσδιορίσουμε τα δύο βασικά βήματα της ανάλυσης συστοιχιών:

- i) τη μέτρηση της απόστασης που χωρίζει όλα τα ζεύγη αντικειμένων
- ii) την ανάπτυξη μιας ρουτίνας, ή αλγορίθμου, για τον σχηματισμό συστοιχιών με βάση αυτές τις αποστάσεις

Μπορούμε να προσδιορίσουμε τις αποστάσεις που αναφέρονται στο (i) είτε υποκειμενικά είτε επινοώντας ένα μέτρο της απόστασης με βάση την παρατή-

ρηση μιας συλλογής μεταβλητών. Στην πρώτη περίπτωση, ο ανθρώπινος εγκέφαλος είναι εκείνος που επεξεργάζεται τις διαθέσιμες πολυμεταβλητές πληροφορίες για κάθε αντικείμενο· στη δεύτερη, η απόσταση διαμορφώνεται σύμφωνα με κάποια ορθολογική αρχή.

Πριν προχωρήσουμε, πρέπει να αποσαφηνίσουμε το νόημα της λέξης *απόσταση*, την οποία έχουμε χρησιμοποιήσει για να περιγράψουμε το πόσο μακριά βρίσκεται το ένα αντικείμενο από το άλλο. Μερικές φορές είναι πιο φυσικό να μιλάμε με όρους προσέγγισης ή *εγγύτητας* (proximity). Εναλλακτικά χρησιμοποιούνται οι όροι *ομοιότητα* και *ανομοιότητα*. Οι τελευταίοι έχουν το πλεονέκτημα ότι υποδεικνύουν μια χαλαρότερη και πιο υποκειμενική εκτίμηση της απόστασης, η οποία είναι καταλληλότερη για μερικές από τις εφαρμογές που θα συναντήσουμε. Η εγγύτητα και η ομοιότητα έχουν αντίστροφη σχέση με την απόσταση και την ανομοιότητα, και έτσι οι μετρήσεις της μίας μπορούν εύκολα να μετατραπούν σε μετρήσεις της άλλης. Θα χρησιμοποιούμε τους διάφορους όρους εναλλακτικά, αλλά θεωρήστε την «απόσταση» ως τον κύριο όρο, επειδή αποτελεί κεντρική έννοια και στην πολυδιάστατη προσαρμογή κλίμακας, την οποία θα περιγράψουμε στο Κεφάλαιο 3.

Το πρώτο στάδιο της ανάλυσης συστοιχιών είναι η κατασκευή αποστάσεων μεταξύ ζευγών αντικειμένων. Θα αναβάλλουμε την περιγραφή του τρόπου με τον οποίο γίνεται αυτό μέχρι την Ενότητα 2.4· μέχρι τότε, θα έχουμε μια σαφέστερη εικόνα του τρόπου χρήσης των αποστάσεων. Προς το παρόν, επισημαίνουμε ότι η ίδια η διαδικασία της ομαδοποίησης ξεκινάει με μια *μήτρα απόστασης*, δηλαδή έναν πίνακα στον οποίο η απόσταση μεταξύ του αντικειμένου i και του αντικειμένου j εμφανίζεται στη γραμμή i και στη στήλη j . Για παράδειγμα, αν έχουμε τέσσερα αντικείμενα, έχουμε μια *μήτρα απόστασης* 4×4 :

$$\begin{pmatrix} - & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & - & \delta_{23} & \delta_{24} \\ \delta_{31} & \delta_{32} & - & \delta_{34} \\ \delta_{41} & \delta_{42} & \delta_{43} & - \end{pmatrix}$$

όπου δ_{ij} είναι η απόσταση μεταξύ του αντικειμένου i και του αντικειμένου j . Συχνά, η μήτρα απόστασης θα είναι συμμετρική, δηλαδή θα ισχύει $\delta_{21} = \delta_{12}$, $\delta_{31} = \delta_{13}$, κ.ο.κ. Αυτό συμβαίνει επειδή οι εκτιμήσεις των αποστάσεων συνήθως δεν εξαρτώνται από τη σειρά που θα πάρουμε τα δύο αντικείμενα. Για τον λόγο αυτό, αρκεί να γράψουμε μόνο τα μισά δ — είτε αυτά που βρίσκονται στο πάνω είτε εκείνα που βρίσκονται στο κάτω τρίγωνο της μήτρας. Η διαγώνιος

μπορεί να μείνει κενή, επειδή τα συγκεκριμένα στοιχεία δεν παίζουν κανένα ρόλο στη διαδικασία της ομαδοποίησης. Τα δ_{ij} αποκαλούνται μερικές φορές παρατηρούμενες αποστάσεις ή, απλώς, *παρατηρήσεις*.

Οι μέθοδοι ανάλυσης συστοιχιών μπορούν να ταξινομηθούν γενικώς σε *ιεραρχικές* ή *μη ιεραρχικές*. Σε μια ιεραρχική μέθοδο, η διαδικασία δημιουργίας των συστοιχιών έχει παραγάγει μια ιεραρχία στην οποία υποσύνολα των συστοιχιών σε ένα επίπεδο συγκεντρώνονται για να σχηματίσουν τις συστοιχίες στο επόμενο, υψηλότερο επίπεδο. Οι ιεραρχικές μέθοδοι μπορούν, επίσης, να διαιρεθούν σε *συσσωρευτικές* (agglomerative) και σε *διαιρετικές* (divisive) μεθόδους. Σε μια συσσωρευτική μέθοδο, ξεκινάμε αντιμετωπίζοντας κάθε αντικείμενο ως συστοιχία με ένα μέλος, και στη συνέχεια συνεχίζουμε με μια σειρά βημάτων για να συγχωνεύσουμε συστοιχίες. Σε μια τέτοια μέθοδο, μόλις δύο άτομα τοποθετηθούν μαζί σε μια συστοιχία, δεν μπορούν ποτέ να διαχωριστούν σε μεταγενέστερο βήμα. Αυτό συμβαίνει επειδή οποιαδήποτε νέα συστοιχία σχηματίζεται από συστοιχίες που έχουν ήδη δημιουργηθεί σε προηγούμενα στάδια της διαδικασίας. Σε μια διαιρετική μέθοδο, ξεκινάμε από το άλλο άκρο, αντιμετωπίζοντας ολόκληρο το σύνολο των ατόμων ως μία συστοιχία και προχωράμε διαιρώντας τις υπάρχουσες συστοιχίες. Σε τέτοιες διαδικασίες, μόλις ένα ζεύγος ατόμων διαχωριστεί, δεν μπορεί να ενωθεί ξανά. Αυτό μας επιτρέπει, όπως θα δούμε παρακάτω, να αναπαραστήσουμε τα στάδια της διαδικασίας με ένα γράφημα δένδρου, όπου τα σημεία διακλάδωσης δείχνουν τα σημεία στα οποία ενώνονται ή διαχωρίζονται οι συστοιχίες.

Στις μη ιεραρχικές μεθόδους, οι συστοιχίες σχηματίζονται προσαρμόζοντας τη συμμετοχή των ατόμων (ή αντικειμένων) σε εκείνες τις συστοιχίες που υπάρχουν σε οποιοδήποτε στάδιο της διαδικασίας, με την προσθήκη ή την αφαίρεση ατόμων από αυτές. Συνήθως, οι συγκεκριμένες μέθοδοι είναι πιο δύσκολο να εφαρμοστούν και δεν χρησιμοποιούνται πολύ. Θα δείξουμε σύντομα πώς θα μπορούσαν να λειτουργήσουν τέτοιες μέθοδοι, αλλά θα επικεντρωθούμε κυρίως στις ιεραρχικές συσσωρευτικές μεθόδους.

Στο παρόν κεφάλαιο, θα θεωρήσουμε την ανάλυση συστοιχιών ως μια πλήρως περιγραφική μέθοδο, αλλά ο όρος καλύπτει επίσης μια μεγάλη γκάμα μεθόδων, κάποιες από τις οποίες βασίζονται σε μοντέλα. Θα συναντήσουμε ένα παράδειγμα στο Κεφάλαιο 10 για τις μεθόδους λανθανουσών κλάσεων (latent class methods), όπου τα άτομα πρέπει να κατανεμηθούν σε κατηγορίες που καθορίζονται μέσα στο πλαίσιο ενός μοντέλου.

2.2 Μερικές μέθοδοι ανάλυσης συστοιχιών

Θα χρησιμοποιήσουμε ένα πολύ απλό παράδειγμα που αποτελείται από πέντε μόνο άτομα, για να δείξουμε δύο από τις πιο διαδεδομένες συσσωρευτικές ιεραρχικές μεθόδους, συγκεκριμένα τη μέθοδο του *πλησιέστερου γείτονα* (nearest neighbour) και τη μέθοδο του *πιο απομακρυσμένου γείτονα* (farthest neighbour). Μερικές φορές, αναφερόμαστε σε αυτές ως μέθοδος *απλού συνδέσμου* (single linkage) και μέθοδος *πλήρους συνδέσμου* (complete linkage), αντίστοιχα. Αργότερα, θα αναφέρουμε σύντομα μερικές άλλες μεθόδους.

Για ένα παράδειγμα με πέντε μόνο άτομα, η τυπική ανάλυση δεν θα προσθέσει πολλά σε αυτά που μπορούμε να δούμε εξετάζοντας τη μήτρα με τις αποστάσεις. Άρα, δεν πρέπει να περιμένουμε να μάθουμε πολλά για την ερμηνεία των στοιχείων από το συγκεκριμένο παράδειγμα. Ο σκοπός μας είναι να ορίσουμε απλώς τα βήματα που πρέπει να ακολουθήσουμε. Φανταστείτε ότι τα δεδομένα αφορούν πέντε πελάτες ενός σουπερμάρκετ: τους Άρη, Βασίλη, Γιάννα, Δήμητρα, και Ελένη, και ότι οι πληροφορίες αποκτήθηκαν από μια έρευνα για τα γούστα και τις προτιμήσεις τους. Ας υποθέσουμε επίσης ότι οι πληροφορίες αυτές έχουν αναλυθεί και συνοψιστεί στον πίνακα των αποστάσεων, τον Πίνακα 2.1. Οι καταχωρίσεις αντιπροσωπεύουν το πόσο απέχουν τα άτομα μεταξύ τους όσον αφορά τα πιθανά αγοραστικά τους μοτίβα. Όσο μικρότερος είναι ο αριθμός, τόσο πιο κοντά βρίσκονται τα άτομα του ζεύγους. Οι επικεφαλίδες των γραμμών και των στηλών δείχνουν τον πελάτη, με το αρχικό του ονόματός του. Σε μια μελέτη μεγαλύτερης κλίμακας με πολύ περισσότερους πελάτες, ο αντικειμενικός στόχος του σούπερ μάρκετ θα μπορούσε να είναι η στόχευση της διαφήμισής του σε ομάδες πελατών με παρόμοιες απόψεις.

Ο πίνακας εμφανίζεται σε μορφή κάτω τριγώνου καθώς οι αποστάσεις θεωρούνται ότι είναι συμμετρικές, και έτσι δεν χρειάζεται να επαναλαμβάνουμε τις πληροφορίες στο πάνω μισό του πίνακα.

Πίνακας 2.1 Πίνακας απόστασης για μεθόδους ομαδοποίησης

	A	B	Γ	Δ	E
A	-				
B	3	-			
Γ	8	7	-		
Δ	11	9	6	-	
E	10	9	7	5	-

Η μέθοδος του πλησιέστερου γείτονα (ή απλού συνδέσμου)

Αυτή είναι μια συσσωρευτική μέθοδος στην οποία κάθε πελάτης θεωρείται αρχικά ως μια ανεξάρτητη συστοιχία. Πρώτα, αναζητούμε το πλησιέστερο ζεύγος ατόμων, διαδικασία που απαιτεί να σαρώσουμε όλους τους αριθμούς του πίνακα και να βρούμε τη μικρότερη καταχώριση. Αυτή είναι ο αριθμός 3 στη γραμμή Β και τη στήλη Α. Επομένως, στο πρώτο στάδιο, ο Άρης και ο Βασίλης τοποθετούνται μαζί σε μια συστοιχία. Στο επόμενο στάδιο, κατασκευάζουμε έναν νέο πίνακα απόστασης κατάλληλο για τις τέσσερις συστοιχίες που δημιουργήθηκαν στο τέλος του πρώτου σταδίου. Για να γίνει αυτό, πρέπει να καθορίσουμε πώς θα μετρηθεί η απόσταση μεταξύ ομάδων που περιέχουν περισσότερα από ένα άτομα. Στη μέθοδο του πλησιέστερου γείτονα, η απόσταση μεταξύ δύο συστοιχιών ορίζεται ως η απόσταση μεταξύ των πλησιέστερων μελών τους. Άρα, για παράδειγμα, η απόσταση μεταξύ της συστοιχίας (Άρης, Βασίλης) και του ατόμου Γιάννα είναι η μικρότερη των αποστάσεων από το Α στο Γ και από το Β στο Γ, δηλαδή ο μικρότερος από τους αριθμούς 8 και 7. Άρα, κατά την κατασκευή του νέου πίνακα απόστασης, η καταχώριση για τη γραμμή με ετικέτα Γ και τη στήλη με ετικέτα (Α,Β) είναι το 7, όπως φαίνεται στον Πίνακα 2.2.

Στο δεύτερο στάδιο, απλώς επαναλαμβάνουμε για τον Πίνακα 2.2 τη διαδικασία που εφαρμόσαμε στον Πίνακα 2.1· συγκεκριμένα, αναζητάμε τη μικρότερη τιμή. Αυτή είναι η τιμή 5, δηλαδή η απόσταση μεταξύ της Δήμητρας και της Ελένης. Άρα, στο δεύτερο στάδιο συνδυάζουμε το Δ και το Ε σε μία συστοιχία. Έχουμε πλέον τρεις συστοιχίες και χρειαζόμαστε έναν νέο πίνακα απόστασης με τις αποστάσεις μεταξύ των τριών συστοιχιών. Αυτός είναι ο Πίνακας 2.3. Η μόνη καινοτομία στο συγκεκριμένο στάδιο είναι η εύρεση της απόστασης μεταξύ των συστοιχιών (Α,Β) και (Δ,Ε), οι οποίες έχουν και οι δύο περισσότερα από ένα μέλη. Ο Άρης και η Δήμητρα απέχουν μεταξύ τους 11

Πίνακας 2.2 Πίνακας απόστασης για το πρώτο στάδιο της ομαδοποίησης πλησιέστερου γείτονα

	(Α,Β)	Γ	Δ	Ε
(Α,Β)	-			
Γ	7	-		
Δ	9	6	-	
Ε	9	7	5	-

Πίνακας 2.3 Πίνακας απόστασης για το δεύτερο στάδιο της ομαδοποίησης πλησιέστερου γείτονα

	(A,B)	Γ	(Δ,E)
(A,B)	-		
Γ	7	-	
(Δ,E)	9	6	-

μονάδες, ο Άρης και η Ελένη βρίσκονται 10 μονάδες μακριά ο ένας από τον άλλο, αλλά ο Βασίλης απέχει 9 μονάδες από τη Δήμητρα και την Ελένη. Άρα, το 9 είναι η μικρότερη απόσταση, και έτσι καταχωρίζεται στην τρίτη γραμμή και την πρώτη στήλη του Πίνακα 2.3.

Η μικρότερη καταχώριση στον Πίνακα 2.3 είναι το 6, γεγονός που δείχνει ότι στο επόμενο στάδιο πρέπει να συγχωνεύσουμε τη Γιάννα με την ομάδα (Δήμητρα, Ελένη). Τότε θα φτάσουμε σε ένα σημείο όπου υπάρχουν μόνο δύο συστοιχίες και το μοναδικό πιθανό επόμενο στάδιο είναι να τις συγχωνεύσουμε σε μία και μόνο συστοιχία. Η προηγούμενη ανάλυση μπορεί να συγκεντρωθεί σε έναν πίνακα συσσώρευσης (agglomeration table) ως εξής (Πίνακας 2.4).

Αυτό που πετύχαμε με την παραπάνω διαδικασία δεν είναι ακριβώς αυτό που ξεκινήσαμε να κάνουμε. Αντί να καταλήξουμε σε ένα και μόνο σύνολο συστοιχιών, έχουμε μια ιεραρχική ακολουθία που αρχίζει από το σύνολο των πέντε μεμονωμένων συστοιχιών με τις οποίες ξεκινήσαμε, και φτάνει στη μία συστοιχία με την οποία τελειώνει η διαδικασία. Συνεπώς, χρειαζόμαστε κάποιον τρόπο για να κρίνουμε αν ένα συγκεκριμένο σύνολο σε αυτή την ακολουθία μπορεί να κερδίσει την προσοχή μας. Θα επιστρέψουμε στο συγκεκριμένο ερώτημα αφού περιγράψουμε μια δεύτερη μέθοδο.

Πίνακας 2.4 Πίνακας συσσώρευσης για την ομαδοποίηση πλησιέστερου γείτονα

Στάδιο	Πλήθος συστοιχιών	Συστοιχίες	Επίπεδο απόστασης
Αρχικό	5	(A) (B) (Γ) (Δ) (E)	0
1	4	(A,B) (Γ) (Δ) (E)	3
2	3	(A,B) (Γ) (Δ,E)	5
3	2	(A,B) (Γ,Δ,E)	6
4	1	(A,B,Γ,Δ,E)	7

Η μέθοδος του πιο απομακρυσμένου γείτονα (ή πλήρους συνδέσμου)

Αυτή είναι ίδια με τη μέθοδο του πλησιέστερου γείτονα με την εξαίρεση ότι η απόσταση μεταξύ δύο ομάδων ορίζεται πλέον ως η απόσταση των πιο απομακρυσμένων μελών τους. Θα δείξουμε τη μέθοδο χρησιμοποιώντας το ίδιο παράδειγμα με το σούπερ μάρκετ.

Τα πρώτα δύο στάδια προχωρούν ακριβώς όπως στη μέθοδο του πλησιέστερου γείτονα επειδή, μέχρι αυτό το σημείο, ασχολούμαστε με συστοιχίες που περιέχουν μόνο ένα μέλος. Η διαφορά προκύπτει στο τρίτο στάδιο και, για τον λόγο αυτό, χρειαζόμαστε τον πίνακα απόστασης μεταξύ των τριών ομάδων που υπάρχουν στο συγκεκριμένο στάδιο. Ο πίνακας απόστασης δίνεται στον Πίνακα 2.5. Κατά τον υπολογισμό της απόστασης μεταξύ της συστοιχίας (Α,Β) και του ατόμου Γ, επιλέγουμε τη *μεγαλύτερη* από τις αποστάσεις μεταξύ του Άρη και της Γιάννας και μεταξύ του Βασίλη και της Γιάννας στον Πίνακα 2.1, δηλαδή το 8 και όχι το 7. Ομοίως, η απόσταση μεταξύ της συστοιχίας (Δ,Ε) και του ατόμου Γ είναι πλέον το 7. Τέλος, η απόσταση μεταξύ των συστοιχιών (Α,Β) και (Δ,Ε) είναι το 11, το οποίο αποτελεί τη μεγαλύτερη από τις τέσσερις αποστάσεις στην κάτω αριστερή γωνία του Πίνακα 2.1.

Η μικρότερη απόσταση στον Πίνακα 2.5 είναι το 7. Αυτό υποδεικνύει ότι στο επόμενο στάδιο θα πρέπει να συγχωνεύσουμε τη Γιάννα με τη συστοιχία (Δήμητρα, Ελένη). Ο πίνακας συσσωρεύσης που προκύπτει είναι ο ακόλουθος (Πίνακας 2.6).

Όπως βλέπετε, τα σύνολα συστοιχιών που πήραμε με τη μέθοδο του πιο απομακρυσμένου γείτονα συμπίπτουν με αυτά που προέκυψαν με τη μέθοδο του πλησιέστερου γείτονα, παρότι τα επίπεδα απόστασης στα οποία συγχωνεύονται οι συστοιχίες διαφέρουν μεταξύ τους. Γενικά, οι μέθοδοι του πλησιέστερου και του πιο απομακρυσμένου γείτονα δίνουν διαφορετικά αποτελέσματα, μερικές

Πίνακας 2.5 Πίνακας απόστασης για το δεύτερο στάδιο της ομαδοποίησης πιο απομακρυσμένου γείτονα

	(Α,Β)	Γ	(Δ,Ε)
(Α,Β)	-		
Γ	8	-	
(Δ,Ε)	11	7	-

Το βιβλίο **Ανάλυση πολυμεταβλητών τεχνικών στις Κοινωνικές Επιστήμες** είναι αποτέλεσμα της μεγάλης πείρας των συγγραφέων, οι οποίοι έχουν εργαστεί και διδάξει στο συγκεκριμένο πεδίο, και έχει στόχο να σας βοηθήσει να κατανοήσετε τον τρόπο χρήσης των κυριότερων πολυμεταβλητών μεθόδων στις Κοινωνικές Επιστήμες. Με ελάχιστες απαιτήσεις σε γνώσεις στατιστικής και μαθηματικών, αυτή η ενημερωμένη έκδοση επεκτάθηκε προκειμένου να περιλαμβάνει την ανάλυση παλινδρόμησης, την επιβεβαιωτική παραγοντική ανάλυση, τα μοντέλα δομικών εξισώσεων, και τα πολυεπίπεδα μοντέλα.

Στα πρώτα πέντε κεφάλαια οι συγγραφείς δίνουν έμφαση στη σύνοψη των δεδομένων, ενώ στο επόμενο εστιάζουν στην ανάλυση της παλινδρόμησης. Αυτό το έκτο κεφάλαιο συνδέει τα δύο μέρη του βιβλίου και σηματοδοτεί τη μετάβαση από τις περιγραφικές στις συμπερασματικές μεθόδους και από την αλληλεξάρτηση στην εξάρτηση. Στα υπόλοιπα κεφάλαια του βιβλίου περιγράφονται μέθοδοι βασισμένες σε μοντέλα που αφορούν την εξαγωγή συμπερασμάτων για τις διεργασίες οι οποίες δημιουργούν τα δεδομένα.

Οι συγγραφείς, βασιζόμενοι σε μεγάλο βαθμό σε αριθμητικά παραδείγματα, σας δίνουν την ευκαιρία να κατανοήσετε σε βάθος τον σκοπό και τον τρόπο λειτουργίας των μεθόδων, καθώς επίσης και την ερμηνεία των δεδομένων. Πολλά από τα παραδείγματα χρησιμοποιούνται σε ολόκληρο το βιβλίο για να δείξουν τις σχέσεις των διαφόρων μεθόδων. Στα περισσότερα κεφάλαια, οι συγγραφείς παρουσιάζουν προτάσεις για πρόσθετη εργασία που ξεπερνούν τις συμβατικές ασκήσεις, ενθαρρύνοντας με αυτόν τον τρόπο τους αναγνώστες να εξερευνήσουν νέους ορίζοντες στην έρευνα των Κοινωνικών Επιστημών.

Χαρακτηριστικά του βιβλίου

- Περιέχει τρία νέα κεφάλαια για την ανάλυση παλινδρόμησης, την επιβεβαιωτική παραγοντική ανάλυση και τα μοντέλα δομικών εξισώσεων, και τα πολυεπίπεδα μοντέλα
- Παρουσιάζει πολλά παραδείγματα εφαρμογής σε πραγματικά δεδομένα, όπως πολιτικές προτιμήσεις, κοινωνικές στάσεις, βαθμολογίες εξετάσεων, και δεδομένα από τον χώρο των εργασιακών σχέσεων
- Καλύπτει μεθόδους που συνοψίζουν, περιγράφουν και διερευνούν πολυμεταβλητά σύνολα δεδομένων
- Εδραιώνει μια ενοποιημένη θεώρηση για τα μοντέλα λανθανουσών μεταβλητών παρέχοντας λεπτομερή κάλυψη διαφόρων μεθόδων, όπως η θεωρία της απάντησης ερωτήσεων και η παραγοντική ανάλυση
- Στον συνοδευτικό ιστότοπο θα βρείτε όλα τα σύνολα δεδομένων του βιβλίου, λογισμικό για να εκτελείτε αναλύσεις λανθανουσών μεταβλητών, καθώς και οδηγίες (στην αγγλική γλώσσα) για την εφαρμογή των αναλύσεων



ΕΚΔΟΣΕΙΣ
ΚΛΕΙΔΑΡΙΘΜΟΣ

Δομοκού 4, Σταθμός Λαρίσης, 10440 ΑΘΗΝΑ, Τηλ. 210-5237635
www.klidarithmos.gr info@klidarithmos.gr

ISBN 978-960-461-441-7



9 789604 614417