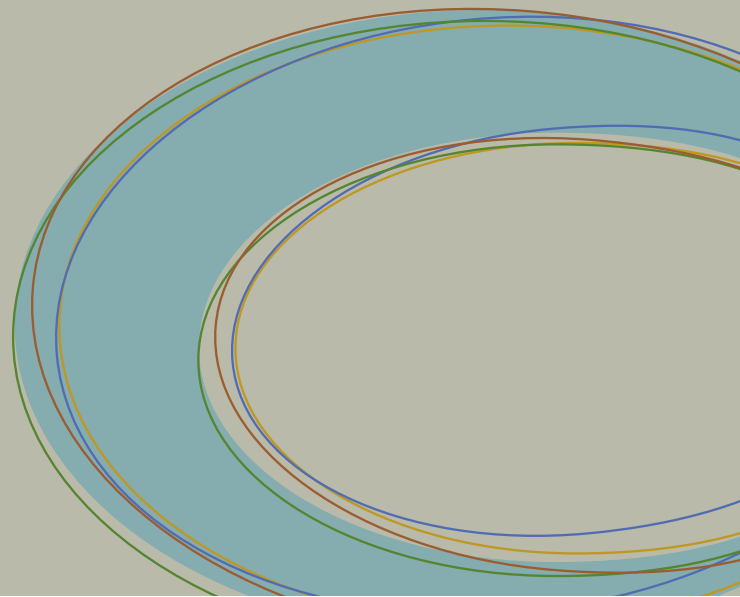


ΠΕΡΙΕΧΟΜΕΝΑ



ΜΕΡΟΣ Ι: ΒΑΣΙΚΗ ΧΡΗΣΗ 1

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή 3

1.1	Σε ποιον απευθύνεται αυτό το βιβλίο	3
1.2	Στόχοι του βιβλίου	4
1.3	Περιεχόμενα.....	4
1.3.1	Προγραμματισμός στην R	5
1.3.2	Βασικές αρχές στατιστικής.....	5
1.3.3	Μεθοδολογίες ανάλυσης δεδομένων	6
1.3.4	Εφαρμογές σε πραγματικά δεδομένα.....	6
1.4	Πλεονεκτήματα της R.....	6
1.5	Πώς είναι γραμμένο αυτό το βιβλίο.....	6
1.6	Διαβάστε περισσότερα	7
1.6.1	Βιβλία για την R.....	8
1.6.2	Διαδικτυακές πηγές	8

ΚΕΦΑΛΑΙΟ 2: Μεταβλητές, τελεστές και δεδομένα 9

2.1	Ξεκινώντας με την R	9
2.1.1	Εγκαθιστώντας την R.....	9
2.1.2	Εξοικείωση με το περιβάλλον της R.....	10
2.2	Το ολοκληρωμένο περιβάλλον R-Studio	12
2.3	Μεταβλητές στην R.....	13
2.3.1	Απλές μεταβλητές	13
2.3.2	Εκχώρηση τιμής σε μεταβλητή	14
2.3.3	Κανόνες ονομασίας μεταβλητών	14
2.3.4	Είδη μεταβλητών στην R.....	15
2.4	Τελεστές και απλές πράξεις με μεταβλητές	17
2.4.1	Αριθμητικές πράξεις.....	17
2.4.2	Λογικές πράξεις.....	18
2.5	Συναρτήσεις	19
2.5.1	Δημιουργία συναρτήσεων	19

2.6	Πολυδιάστατα δεδομένα.....	20
2.6.1	Έτοιμα (built-in) δεδομένα.....	20
2.7	Εισαγωγή δεδομένων από αρχεία.....	24
2.7.1	Ανάγνωση δεδομένων σε πίνακα	25
2.7.2	Ανάγνωση αρχείων κειμένου	25
2.7.3	Ορισμός φακέλου εργασίας και πλήρους μονοπατιού	26
2.8	Εγγραφή Αποτελεσμάτων σε αρχεία.....	27
2.9	Ανάγνωση έτοιμου κώδικα R	27

ΚΕΦΑΛΑΙΟ 3: Τύποι δεδομένων 29

3.1	Διανύσματα (vectors).....	30
3.1.1	Γενικά χαρακτηριστικά διανυσμάτων	30
3.1.2	Δημιουργία διανυσμάτων και «εξαναγκασμός» (coercion)	31
3.1.3	Εξωτερικός «εξαναγκασμός» από τον χρήστη	32
3.1.4	Συναρτήσεις διανυσμάτων	34
3.1.4.1	Συναρτήσεις αριθμητικών διανυσμάτων	34
3.1.4.2	Συναρτήσεις αλφαριθμητικών διανυσμάτων	35
3.2	Πράξεις σε διανύσματα.....	37
3.3	Πίνακες δύο διαστάσεων	39
3.3.1	Πρόσβαση σε στοιχεία πινάκων.....	41
3.3.2	Συναρτήσεις πινάκων.....	41
3.4	Πλαίσια δεδομένων.....	45
3.4.1	Εφαρμογή συναρτήσεων σε πίνακες και πλαίσια δεδομένων (apply)	45
3.5	Πίνακες περισσότερων από δύο διαστάσεων	46
3.6	Λίστες	49
3.6.1	Δημιουργία λιστών	49
3.6.2	Χειρισμός λιστών με συναρτήσεις τύπου apply()	52

ΚΕΦΑΛΑΙΟ 4: Πλαίσια δεδομένων, παράγοντες και χειρισμοί συνόλων..... 55

4.1	Χειρισμοί πλαισίων δεδομένων	55
4.1.1	Κενές τιμές (missing values).....	56
4.1.2	Συναρτήσεις ελέγχου τιμών.....	58
4.2	Παράγοντες (factors)	61
4.2.1	Δημιουργία και χειρισμός επιπέδων (levels) σε παράγοντες.....	62
4.2.2	Χρήση παραγόντων σε πλαίσια δεδομένων.....	63
4.3	Δημιουργία υποσυνόλων (subsetting).....	65
4.3.1	Δημιουργία υποσυνόλων με απλούς ελέγχους.....	66
4.3.2	Υποσύνολα με συνδυασμούς ελέγχων	66
4.3.4	Λήψη υποσυνόλων από πλαίσια δεδομένων με τη συνάρτηση subset().....	67
4.3.5	Λήψη θέσεων στοιχείων υποσυνόλων με τη συνάρτηση which()	69
4.4	Πράξεις σε σύνολα	71
4.4.1	Ύπαρξη στοιχείων συνόλου σε άλλο σύνολο.....	73
4.4.2	Συγκρίσεις μεταξύ συνόλων σε επίπεδο θέσης.....	73

ΚΕΦΑΛΑΙΟ 5: Πίνακες tibble, υποσυνολοποίηση και μετασχηματισμοί δεδομένων..... 77

5.1	Πλαίσια Δεδομένων (data frames) και πίνακες tibble.....	78
5.1.1	Δημιουργία tibble	79

5.2	Χειρισμοί δεδομένων με την <code>dplyr</code>	81
5.2.1	Προσθέτοντας νέες μεταβλητές σε σύνολα δεδομένων	81
5.2.2	Αγωγοί (<code>pipes</code>)	83
5.2.3	Υποσυνολοποίηση με τις συναρτήσεις <code>filter()</code> και <code>select()</code>	84
5.2.4	Υποσυνολοποίηση κατά γραμμή με την <code>slice()</code>	87
5.2.5	Οργάνωση με την <code>group_by()</code> και συνόψεις με τη <code>summarize()</code>	90
5.2.5.1	Συναρτήσεις <code>group()</code>	90
5.2.5.2	Δημιουργία συνόψεων	94
5.2.5.3	Συνόψεις ανά γραμμή με τη <code>rowwise()</code>	95
5.2.6	Κατάταξη με την <code>arrange()</code>	96
5.3	Τακτοποίηση δεδομένων με την <code>tidyr()</code>	98
5.3.1	Δημιουργία μακρών/σύντομων πινάκων	98
5.3.1.1	Επιμήκυνση πίνακα με την <code>pivot_longer()</code>	99
5.3.1.2	Επέκταση πίνακα με την <code>pivot_wider()</code>	100
5.3.2	Πιο πολύπλοκοι μετασχηματισμοί δεδομένων	101
5.3.3	Επέκταση και συμπλήρωση πινάκων	103
5.4	Σχεσιακά δεδομένα: Συνδυάζοντας δεδομένα από διαφορετικές πηγές	106
5.4.1	Σύνδεση δεδομένων με τις συναρτήσεις <code>join()</code>	107

ΜΕΡΟΣ II: ΠΡΟΧΩΡΗΜΕΝΗ ΧΡΗΣΗ ΚΑΙ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ 113

ΚΕΦΑΛΑΙΟ 6: Γραφικές παραστάσεις.....	115	
6.1	Ραβδογράμματα με την <code>barplot()</code>	116
6.2	Προχωρημένα ραβδογράμματα	121
6.3	Πίτες με την <code>pie()</code>	127
6.4	Διαγράμματα σκέδασης για συζευγμένα δεδομένα με την <code>plot()</code>	129
6.5	Χρήση της <code>plot()</code> για σειριακά δεδομένα	131
6.6	Πολλαπλά διαγράμματα σε μια γραφική με την <code>lines()</code>	134
6.7	Τρισδιάστατα δεδομένα με τις <code>persp()</code> και <code>filled.contour()</code>	137
6.7.1	Διαγράμματα 3D όγκου	138
6.7.1	Διαγράμματα 3D περιγραμμάτων	140
6.8	Αποθήκευση και εκτύπωση γραφικών	140
ΚΕΦΑΛΑΙΟ 7: Προχωρημένες γραφικές παραστάσεις με τη βιβλιοθήκη <code>ggplot2</code>	143	
7.1	Η βιβλιοθήκη <code>ggplot2()</code>	143
7.2	Μια πρώτη γραφική παράσταση με <code>ggplot2()</code>	144
7.2.1	Η γραμματική των γραφικών (<code>grammar of graphics</code>)	146
7.3	Μονοδιάστατα δεδομένα	147
7.4	Στατιστική αναπαράσταση μονοδιάστατων δεδομένων	149
7.5	Αριθμητικά δεδομένα συνδυασμένα με κατηγορικές μεταβλητές	152
7.6	Συνδυασμοί πολλαπλών αριθμητικών και κατηγορικών δεδομένων	155
7.7	Μορφοποίηση και στατιστικοί μετασχηματισμοί δεδομένων	157
7.7.1	Στατιστικοί μετασχηματισμοί (<code>bar</code> , <code>boxplot</code>)	158
7.7.2	Συντεταγμένες (<code>coord</code>) και όψεις (<code>facets</code>)	159
7.7.3	Μορφοποίηση αξόνων, τίτλων και κειμένου	163

ΚΕΦΑΛΑΙΟ 8: Προγραμματισμός με την R	169
8.1 Στοιχεία προγραμματισμού.....	169
8.2 Προτάσεις υπό συνθήκη – Δομές if/else και case/switch.....	170
8.2.1 Δυαδικοί If/Else έλεγχοι.....	170
8.2.2 Πολλαπλά if.....	170
8.2.3 Σύνθετες συνθήκες με σύνταξη case/switch.....	171
8.2.3.1 case_when()	171
8.2.3.2 switch.....	172
8.3 Δομές επανάληψης.....	173
8.3.1 Δομές επανάληψης for	173
8.3.2 Έλεγχος των δομών επανάληψης. while/repeat και break/next	177
8.3.3 Διανυσματική λειτουργία και συναρτήσεις apply.....	178
8.4 Συναρτήσεις.....	178
8.4.1 Δομή συναρτήσεων.....	179
8.4.2 Σταθερά και ελεύθερα ορίσματα.....	180
8.5 Εκτέλεση R κώδικα	181
8.5.1 Ανάγνωση έτοιμου κώδικα R.....	181
8.5.2 Εκτέλεση R κώδικα εκτός κονσόλας Rscript().....	182

ΚΕΦΑΛΑΙΟ 9: Χειρισμοί χαρακτήρων κειμένου στην R	187
9.1 Ανάλυση σειρών χαρακτήρων με την stringr().....	187
9.2 Μήκος και συγκόλληση σειρών χαρακτήρων.....	188
9.3 Υποσύνολα, στοίχιση και κατάταξη σειρών χαρακτήρων.....	189
9.3.1 Εξαγωγή substrings και απλές μετατροπές πεζών/κεφαλαίων.....	189
9.3.2 Κατάταξη σειρών χαρακτήρων.....	190
9.3.3 Substrings με χρήση str_split() και στοίχιση strings.....	190
9.4 Αναζήτηση, καταμέτρηση και εντοπισμός στοιχείων σε strings.....	193
9.4.1 Λογική υποσυνολοποίηση strings μέσω ταυτίσεων	193
9.5 Κανονικές εκφράσεις (regular expressions).....	195
9.5.1 Σύντομες χαρακτήρων με κανονικές εκφράσεις	197
9.5.2 Χρήση ποσοτικών τελεστών και ομαδοποίηση με παρενθέσεις	198
9.5.3 Χρήση τελεστών συγκεκριμένου (context).....	200
9.6 Ειδικές περιπτώσεις σειρών χαρακτήρων – Χρόνος και Ημερομηνίες.....	203
9.6.1 Βασικές συναρτήσεις δημιουργίας χρόνου/ημερομηνίας.....	203
9.6.2 Μετατροπές ημερομηνιών από διαφορετικές διατάξεις.....	204
9.6.3 Αποκομιδή στοιχείων από ημερομηνία/ώρα	205
9.6.4 Χρονικά διαστήματα	206
9.6.5 Υπολογισμοί διαφορών σε χρονικά διαστήματα	207

ΜΕΡΟΣ III: ΣΤΑΤΙΣΤΙΚΗ 211

ΚΕΦΑΛΑΙΟ 10: Περιγραφική στατιστική και κατανομές	213
10.1 Μέτρα κεντρικής τάσης.....	213
10.1.1 Μέση τιμή.....	214
10.1.2 Διάμεση τιμή	215
10.2 Μέτρα διασποράς	216
10.2.1 Διασπορά (variance).....	216

10.1.2	Τυπική απόκλιση.....	217
10.1.3	Τυπικό σφάλμα μέσης τιμής (standard error of the mean)	217
10.3	Ποσοστημόρια.....	218
10.4	Ιστογράμματα.....	220
10.3.1	Χρήση μεταβλητών ιστογραμμάτων.....	222
10.5	Θηκογράμματα	225
10.5.1	Παραλλαγές θηκογραμμάτων.....	227
10.6	Δειγματοληψία και τυχαία δείγματα.....	228
10.7	Συναρτήσεις προσομοίωσης κατανομών	230
10.7.1	Τυχαίοι αριθμοί από ομοιόμορφη κατανομή.....	231
10.7.2	Κανονική κατανομή.....	231
10.7.3	Προσομοίωση διωνυμικών πιθανοτήτων.....	232
10.7.4	Προσομοίωση σπάνιων γεγονότων με την κατανομή Poisson.....	233
Εφαρμογή:	Ανάλυση μαθητικών επιδόσεων	234
	Ερώτημα 1: Σχέση φύλου–επιδόσεων	235
	Ερώτημα 2: Σχέση εθνικής ομάδας προέλευσης και επιδόσεων	238
	Ερώτημα 3: σχέση επιδόσεων με μορφωτικό επίπεδο γονέων και προετοιμασία	241
ΚΕΦΑΛΑΙΟ 11: Επαγωγή και έλεγχος υποθέσεων		247
11.1	Έλεγχος κανονικότητας.....	247
11.1.1	Γραφικός έλεγχος κανονικότητας	248
11.1.2	Αριθμητικοί έλεγχοι κανονικότητας	251
11.2	Σύγκριση μέσων τιμών.....	253
11.2.1	Σε κανονικά κατανεμημένα δείγματα	254
11.2.2	Σε μη-κανονικά κατανεμημένα δείγματα.....	257
11.3	Σύγκριση λόγων και αναλογιών	259
11.3.1	Έλεγχος Fisher για πίνακες σύμπτωσης 2×2	260
11.3.2	Έλεγχος Fisher για πίνακες σύμπτωσης $> 2 \times 2$	261
11.4	Έλεγχος υπερ-εκπροσωπήσεων με την υπερ-γεωμετρική κατανομή.....	262
11.5	Στατιστικοί έλεγχοι μέσω μεταθέσεων (permutation tests).....	266
Εφαρμογή:	Ανάλυση αξίας ακινήτων στην πολιτεία της Καλιφόρνιας.....	273
	Ερώτημα 1: Σχέση ηλικίας–τιμής ακινήτου.....	274
	Ερώτημα 2: Μέγεθος νοικοκυριού και αξία ακινήτου.....	279
	Ερώτημα 3. Εξάρτηση της αξίας ακινήτων από τη γεωγραφική τους θέση	281
ΚΕΦΑΛΑΙΟ 12: Ανάλυση διακύμανσης και έλεγχοι πολλαπλών υποθέσεων		287
12.1	Πολλαπλοί ζευγαρωτοί έλεγχοι (pairwise tests).....	287
12.2	Έλεγχος πολλαπλών υποθέσεων	290
12.2.1	Διόρθωση τιμής p-value	292
12.3	Έλεγχος διακύμανσης.....	292
12.4	Ανάλυση Διακύμανσης (ANOVA)	293
12.4.1	Ανάλυση Διακύμανσης: στη θεωρία.....	294
12.4.2	Ανάλυση Διακύμανσης: στην πράξη	297
12.5	Πολύ-παραγοντική ANOVA (multi-way ANOVA).....	300
12.6	Προϋποθέσεις για τη διενέργεια ANOVA.....	306
12.6.1	Ανεξαρτησία.....	306
12.6.2	Κανονικότητα.....	306

12.6.3	Ομοσκεδαστικότητα	307
12.6.4	Ισορροπημένα και μη ισορροπημένα δείγματα.....	309
12.7	ANOVA σε μη κανονικά κατανομημένα δείγματα	310
Εφαρμογή: Ανάλυση επιπέδων γονιδιακής έκφρασης		312
Ερώτημα 1 – Διαφορικά εκφραζόμενα γονίδια μεταξύ ασθενών και υγιών δειγμάτων.....		315
Ερώτημα 2 – Αξιολόγηση θεραπειών		325
Ερώτημα 3 – Συγκριτική ανάλυση φαρμακευτικών παρεμβάσεων.....		326
ΚΕΦΑΛΑΙΟ 13: Συσχέτιση και γραμμική παλινδρόμηση		329
13.1	Συσχέτιση (Correlation).....	329
13.1.1	Γραμμική συσχέτιση Pearson (Pearson Linear Correlation).....	330
13.1.2	Συσχέτιση και ελλειπείς τιμές.....	332
13.1.2	Συσχετίσεις Κατάταξης (Rank Correlations)	333
13.2	Συνδιακύμανση και μερική συσχέτιση (covariance).....	336
13.3	Παλινδρόμηση (Regression).....	337
13.4	Απλή γραμμική παλινδρόμηση	338
13.5	Παλινδρόμηση με τη συνάρτηση $lm()$	338
13.5.1	Απλή παλινδρόμηση	338
13.5.2	Πολυωνυμική παλινδρόμηση	344
13.5.3	Πολλαπλή παλινδρόμηση	347
13.6	Στοιχεία διάγνωσης μοντέλων παλινδρόμησης.....	350
13.6.1	Προσαρμογή.....	351
13.6.2	Κανονικότητα και ακραίες τιμές.....	351
13.6.3	Ομοσκεδαστικότητα και τιμές υψηλής μόχλευσης (high leverage).....	353
13.6.4	Απόσταση Cook και επιδραστικές τιμές.....	354
13.6.5	Συνολική διάγνωση ιδιαίτερων τιμών.....	355
13.7	Σύγκριση μοντέλων και ιεράρχηση παραμέτρων.....	356
Εφαρμογή: Τιμές Lego Sets – Πόσο αξίζει αλήθεια το Millenium Falcon;.....		361
Ερώτημα 1. Σχέση τιμής και ηλικιακής ομάδας.....		362
Ερώτημα 2. Σχέση τιμής και αξιολόγησης πελατών		366
Ερώτημα 3. Ακραίες τιμές – Η περίπτωση του Millenium Falcon		368

ΜΕΡΟΣ IV: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ 377

ΚΕΦΑΛΑΙΟ 14: Μείωση διαστασιμότητας		379
14.1	Ανάλυση κύριων συνιστωσών (Principal Component Analysis, PCA)	380
14.1.1	Σχηματική ερμηνεία της PCA	380
14.1.2	Πρακτική εφαρμογή της PCA.....	382
14.1.3	Γραφική αναπαράσταση PCA.....	384
14.2	Πολυδιάστατη κλιμάκωση(Multidimensional Scaling, MDS)	387
14.2.1	Πίνακες αποστάσεων	387
14.2.2	Εφαρμογή MDS σε πίνακες αποστάσεων	388
14.3	Στοχαστική Ενσωμάτωση Γειτόνων (t-distributed Stochastic Neighbour Embedding, tSNE).....	392
14.3.1	Αρχή της Μεθοδολογίας tSNE	392
14.3.2	Αρχή της Μεθοδολογίας tSNE	393
14.4	Διερευνητική Ανάλυση Παραγόντων (Exploratory Factor Analysis, EFA).....	397
14.4.1	«Λανθάνουσες» μεταβλητές έναντι γραμμικών/μη-γραμμικών συνδυασμών.....	397

14.4.2	Εφαρμογή της EFA.....	397
14.5	Συμπεράσματα.....	403
Εφαρμογή: Κυνηγοί ταλέντων. Πώς να διαλέξουμε τον επόμενο επιθετικό μας;.....		404
	Βήμα 1. Καθαρισμός και προετοιμασία δεδομένων.....	407
	Βήμα 2. Δημιουργία και σύγκριση προφίλ Αμυντικών/Επιθετικών.....	410
	Βήμα 3. Έλεγχος προφίλ υποψηφίων και ιεράρχηση.....	414
ΚΕΦΑΛΑΙΟ 15: Ομαδοποίηση.....		421
15.1	Γενικές αρχές ομαδοποίησης.....	421
15.2	Στόχοι της ομαδοποίησης.....	422
15.3	Ιεραρχική ομαδοποίηση (Hierarchical Clustering).....	423
	15.3.1 Υπολογισμός αποστάσεων/ομοιοτήτων.....	423
	15.3.2 Εφαρμογή ιεραρχικής ομαδοποίησης.....	425
	15.3.3 Εξαγωγή ομάδων από ιεραρχικά δέντρα.....	428
	15.3.4 Ιεραρχική ομαδοποίηση σε θερμικούς χάρτες.....	431
15.4	Ομαδοποίηση διαμερισμού (Partition Clustering).....	436
	15.4.1 Υπολογισμός βέλτιστου αριθμού ομάδων.....	436
	15.4.1.1 Μέτρα συνεκτικότητας ομάδων: Within-Sum-of-Squares (WSS).....	437
	15.4.1.2 Μέτρα συνεκτικότητας ομάδων: Πολλαπλά διαγνωστικά τεστ.....	438
	15.4.2 Ομαδοποίηση κ-μέσων (k-means clustering).....	441
	15.4.3 Ομαδοποίηση διαμερισμού μεσοειδών (Partitioning Around Medoids, PAM).....	447
15.5	Ομαδοποίηση μέσω πυκνότητας.....	450
	15.5.1 Ομαδοποίηση μέσω Πυκνότητας με Εντοπισμό Θορύβου (DBSCAN).....	450
	15.5.2 Υπολογισμός ακτίνας ϵ και ελάχιστου αριθμού στοιχείων πυρήνα N_{min}	450
	15.5.2 Εφαρμογή της μεθόδου DBSCAN.....	452
15.6	Συμπεράσματα.....	454
Εφαρμογή: Διάγνωση καρκίνου του μαστού – Πότε είμαστε σίγουροι για τα άσχημα νέα;.....		455
	Βήμα 1. Ομαδοποίηση ασθενών με βάση τα μορφολογικά χαρακτηριστικά.....	458
	Βήμα 2. Ανάλυση ετερογένειας κακοήθων όγκων.....	462
	Βήμα 3. Ανάλυση παραγόντων για την ερμηνεία της ομαδοποίησης.....	468
ΚΕΦΑΛΑΙΟ 16: Ταξινόμηση.....		473
16.1	Ταξινόμηση, ομαδοποίηση και παλινδρόμηση.....	473
16.2	Εκμάθηση υπό επίβλεψη – Σύνολα εκμάθησης και Σύνολα ελέγχου.....	474
16.3	Μέθοδοι που σχετίζονται με την ομαδοποίηση – Ταξινόμηση k πλησιέστερων γειτόνων (k Nearest Neighbours, kNN).....	475
	16.3.1 Προετοιμασία δεδομένων.....	475
	16.3.2 Εφαρμογή kNN.....	476
	16.3.3 Επιλογή παραμέτρων για kNN.....	478
16.4	Διαδική ταξινόμηση με λογιστική παλινδρόμηση (Logistic Regression).....	480
	16.4.1 Εκμάθηση της λογιστικής παλινδρόμησης.....	480
	16.4.2 Εφαρμογή λογιστικής παλινδρόμησης – Η συνάρτηση predict().....	482
	16.4.3 Αξιολόγηση λογιστικής παλινδρόμησης.....	485
16.5	Γενικευμένα Γραμμικά Μοντέλα για μη-κανονικές κατανομές (Generalized Linear Models, GLM).....	488
16.6	Δέντρα Αποφάσεων (Decision Trees).....	492
	16.6.1 Θεωρητική βάση Δέντρων Αποφάσεων.....	492
	16.6.2 Εφαρμογή και αξιολόγηση δέντρων αποφάσεων.....	495

16.7	Τυχαία Δάση (Random Forests).....	501
16.7.1	Τυχαία Δάση – Εφαρμογή.....	501
16.7.2	Ιεράρχηση επεξηγηματικών μεταβλητών στα Τυχαία Δάση.....	503
16.8	Μηχανές Διανυσματικής Στήριξης (Support Vector Machines, SVM).....	504
16.8.1	Βασικές αρχές των SVM.....	504
16.8.2	Εφαρμογή των SVM.....	505
16.8.3	Ρύθμιση παραμέτρων SVM (SVM Tuning)	506
16.9	Συμπεράσματα.....	507
	Εφαρμογή: Ταξινόμηση πελατών – Σε ποιους θα πρέπει να κάνουμε έκπτωση;.....	509
	Ερώτημα 1. Ένα μοντέλο για την πιθανότητα διαφυγής.....	511
	Ερώτημα 2. Τι πρέπει να προσέξουμε στους πελάτες μας;.....	521
	Ερώτημα 3. Διαμόρφωση στρατηγικών πωλήσεων	524
	ΚΕΦΑΛΑΙΟ 17: Ανάλυση Δικτύων με την R.....	527
17.1	Εισαγωγή στη βιβλιοθήκη igraph.....	527
17.2	Δημιουργία δικτύων στην R.....	528
17.2.1	Δίκτυα μέσω απευθείας δήλωσης ακμών/κόμβων	528
17.2.2	Δίκτυα με ανάγνωση από αρχείο λίστας ακμών.....	529
17.2.3	Δίκτυα από πίνακες γειτνίασης (adjacency matrix).....	531
17.3	Αντικείμενα δικτύου igraph.....	532
17.3.1	Δομή αντικειμένων igraph	532
17.3.2	Ιδιότητες αντικειμένων igraph.....	533
17.4	Υποσυνολοποίηση δικτύων.....	537
17.5	Οπτικοποίηση δικτύων σε διαγράμματα με την plot().....	546
17.6	Τοπολογικά χαρακτηριστικά δικτύων.....	550
17.6.1	Μέγεθος και πυκνότητα δικτύου	550
17.6.2	Αποστάσεις, διάμετρος και συντομότερα μονοπάτια	551
17.7	Στατιστικές ιδιότητες δικτύων	554
17.7.1	Βαθμός κόμβων	554
17.7.2	Κεντρικότητα εγγύτητας (closeness centrality)	556
17.7.3	Διακεντρικότητα (betweenness centrality).....	558
17.7.4	Συντελεστής συσσωμάτωσης και πμοφιλία.....	559
17.8	Ομαδοποίηση δικτύων – Πυρηνικότητα, κλίκες και κοινότητες.....	560
17.8.1	Κλίκες και πυρηνικότητα.....	561
17.8.2	Κοινότητες (communities)	562
17.8.3	Αποσπασματικότητα (Modularity) δικτύων.....	566
	Εφαρμογή: Δίκτυα ανεξάρτητα κλίμακας.....	568
	A. Γενικές ιδιότητες δικτύου	570
	B. Δίκτυα ανεξάρτητα κλίμακας – Κατανομή μεγέθους κόμβων.....	570
	Γ. Σύγκριση με τυχαία δίκτυα.....	573
	ΒΙΒΛΙΟΓΡΑΦΙΑ	527