
Περιεχόμενα

Πρόλογος	19
Σύμβολα και συντομογραφίες	21
1 Ανάλυση δεδομένων	23
1.1 Εισαγωγή	23
1.2 Δεδομένα	24
1.2.1 Παρουσίαση δεδομένων: Γραφικές παραστάσεις	26
1.2.2 Παρουσίαση δεδομένων: Πίνακες δεδομένων και συχνότητες	32
1.3 Μέτρα θέσης	37
1.3.1 Μέσος	38
1.3.2 Διάμεσοι, τεταρτημόρια, ποσοστημόρια και επικρατούσα τιμή	40
1.3.3 Παρατηρήσεις στα μέτρα θέσης	45
1.4 Μέτρα διασποράς	47
1.4.1 Διακύμανση και τυπική απόκλιση	47
1.4.2 Ροπές	50
1.4.3 Ασυμμετρία και κυρτότητα	51
1.4.4 Η διασπορά μέσω διαγραμμάτων	55
1.4.5 Μετασχηματισμός δεδομένων	59
1.5 Διμετάβλητα δεδομένα	61
1.5.1 Διμετάβλητα δεδομένα σε πίνακα διπλής εισόδου	63
1.5.2 Απλή γραμμική παλινδρόμηση	67
1.6 Ασκήσεις και λύσεις	73
2 Πιθανότητες και τυχαίες μεταβλητές	85
2.1 Εισαγωγή	85
2.2 Θεωρία πιθανοτήτων	86
2.2.1 Η έννοια της πιθανότητας	86
2.2.2 Η πιθανότητα ως σχετική συχνότητα	92
2.2.3 Πιθανότητα και αξιώματα	94

2.2.4	Υπό συνθήκη πιθανότητα και ανεξαρτησία	98
2.3	Τυχαίες μεταβλητές	106
2.3.1	Διακριτές τυχαίες μεταβλητές	106
2.3.2	Θεωρητικές κατανομές διακριτών τυχαίων μεταβλητών	112
2.3.3	Μη διακριτές τυχαίες μεταβλητές	122
2.3.4	Θεωρητικές κατανομές μη διακριτών τυχαίων μεταβλητών	128
2.3.5	Συνάρτηση πιθανότητας πολλών μεταβλητών	150
2.3.6	Συναρτήσεις τυχαίων μεταβλητών	155
2.4	Εκτιμητική και χρήση κατανομών	159
2.4.1	Μέθοδος των ροπών	159
2.4.2	Μέθοδος μέγιστης πιθανοφάνειας	160
2.4.3	Μπεϋσιανή μέθοδος εκτίμησης	164
2.4.4	Προσεγγιστικές μέθοδοι εκτίμησης: Η μέθοδος Newton	167
2.5	Ιδιότητες εκτιμητριών συναρτήσεων	168
2.5.1	Αμεροληψία, συνέπεια, αποτελεσματικότητα	168
2.5.2	Κεντρικό οριακό θεώρημα	171
2.5.3	Φράγμα Cramer-Rao	173
2.6	Ασκήσεις και λύσεις	176
3	Στατιστική επαγωγή	185
3.1	Εισαγωγή	185
3.2	Στατιστικοί έλεγχοι υποθέσεων	185
3.3	Στάδια στατιστικού ελέγχου	187
3.3.1	Σχόλια στον έλεγχο υπόθεσης	191
3.4	Έλεγχοι της διωνυμικής αναλογίας	192
3.4.1	Μονόπλευρος έλεγχος p	192
3.4.2	Αμφίπλευρος έλεγχος p	196
3.4.3	Έλεγχοι διωνυμικής αναλογίας δύο πληθυσμών	199
3.5	Έλεγχοι του μέσου	203
3.5.1	Έλεγχοι του μέσου όταν σ γνωστό	203
3.5.2	Έλεγχοι του μέσου όταν σ άγνωστο	207
3.5.3	Έλεγχοι του μέσου δύο πληθυσμών	209
3.6	Έλεγχοι της διακύμανσης	213
3.7	Έλεγχοι ενσωμάτωσης	219
3.8	Έλεγχος ανεξαρτησίας	223
3.9	Έλεγχος μεταβλητότητας	227
3.10	Έλεγχος μεταβλητότητας για συνδυασμούς	231
3.11	Έλεγχοι κανονικότητας	233
3.12	Διαστήματα εμπιστοσύνης	236
3.12.1	Διαστήματα εμπιστοσύνης του μέσου	237

3.12.2	Διαστήματα εμπιστοσύνης του μέσου δύο πληθυσμών . .	240
3.12.3	Διαστήματα εμπιστοσύνης της διωνυμικής αναλογίας . .	241
3.12.4	Διαστήματα εμπιστοσύνης της διακύμανσης	244
3.12.5	Διαστήματα εμπιστοσύνης εκτιμητριών μέγιστης πιθανο- φάνειας	245
3.13	Παλινδρόμηση και επαγωγή	247
3.13.1	Έλεγχοι υπόθεσης στην παλινδρόμηση	247
3.13.2	Διαστήματα εμπιστοσύνης στην παλινδρόμηση	250
3.14	Ασκήσεις και λύσεις	255
4	Πέραν της κανονικότητας...	263
4.1	Εισαγωγή	263
4.2	Έλεγχος τύπου Wilcoxon για έναν πληθυσμό	265
4.3	Έλεγχος τύπου Wilcoxon για ζεύγη δύο πληθυσμών	267
4.4	Έλεγχος τύπου Wilcoxon για δύο πληθυσμούς με αθροιστική ιεράρχηση	270
4.5	Έλεγχος τύπου Kruskal-Wallis για άνω των δύο πληθυσμούς . . .	272
4.6	Έλεγχος τύπου Friedman για άνω των δύο πληθυσμούς	275
4.7	Έλεγχος συσχέτισης τύπου Spearman	277
4.8	Έλεγχοι τάσεων	281
4.8.1	Ο έλεγχος των Mann-Kendall	281
4.8.2	Ο έλεγχος των Mann-Kendall με εποχικότητα	282
4.9	Προσδιορισμός μεγέθους τάσεων κατά Sen	283
4.10	Έλεγχος μεταβολής σημείου/ων (ή σημείου/ων ρήξης)	284
4.11	Ασκήσεις και λύσεις	285
5	Εισαγωγή στο R	289
5.1	Εγκατάσταση του R	289
5.1.1	Εγκατάσταση βιβλιοθηκών στο R	289
5.2	Βασικές πράξεις στο R	291
5.2.1	Αριθμητικές πράξεις	291
5.2.2	Ορισμός αντικειμένων	292
5.2.3	Διανύσματα	292
5.3	Συναρτήσεις στο R	293
5.4	Αναζήτηση βοήθειας στο R	295
6	Απεικόνιση δεδομένων στο R	297
6.1	Εισαγωγή δεδομένων	297
6.2	Απεικόνιση δεδομένων	298

7	Ανάλυση δεδομένων στο R	307
7.1	Μονομετάβλητα μέτρα στο R	307
7.2	Διμετάβλητα μέτρα στο R	311
8	Πιθανότητες στο R	313
8.1	Τυχαίες μεταβλητές στο R	313
8.2	Διακριτές κατανομές στο R	314
8.3	Μη διακριτές κατανομές στο R	316
9	Στατιστική επαγωγή στο R	323
9.1	Έλεγχοι υπόθεσης στο R	323
9.1.1	Έλεγχος ανεξαρτησίας στο R	325
9.1.2	Έλεγχος μεταβλητότητας στο R	326
9.1.3	Έλεγχοι κανονικότητας στο R	327
9.2	Διαστήματα εμπιστοσύνης στο R	329
9.3	Απλή γραμμική παλινδρόμηση στο R	330
9.4	Ειδικά θέματα παλινδρόμησης στο R	335
10	Πέραν της κανονικότητας στο R	341
10.1	Έλεγχος τύπου Wilcoxon για έναν πληθυσμό στο R	341
10.2	Έλεγχος τύπου Wilcoxon για ζεύγη δύο πληθυσμών στο R	341
10.3	Έλεγχος τύπου Wilcoxon στο R για δύο πληθυσμούς με αθροιστική ιεράρχηση	342
10.4	Έλεγχος τύπου Kruskal-Wallis στο R για άνω των δύο πληθυσμούς	343
10.5	Έλεγχος τύπου Friedman στο R για άνω των δύο πληθυσμούς	343
10.6	Έλεγχος συσχέτισης τύπου Spearman στο R	344
10.7	Έλεγχοι υπόθεσης των Mann-Kendall στο R	345
10.8	Προσδιορισμός μεγέθους τάσεων κατά Sen στο R	347
10.9	Έλεγχος μεταβολής σημείου/ων (ή σημείου/ων ρήξης) στο R	348
I	Πίνακες δεδομένων	353
II	Πίνακες κατανομών	359
	Αντί επιλόγου	373

Κατάλογος σχημάτων

1.1	Κατανομή των ελληνικών εμπορικών πλοίων 100 KOX και άνω κατά ομάδες ηλικιών την 31η Μαΐου 2011	26
1.2	Δύναμη του ελληνικού εμπορικού στόλου σε πλοία χωρητικότητας 100 KOX και άνω κατά κατηγορίες για την περίοδο από το 2002 έως και το 2011 (Μάιος). Οι κατηγορίες αφορούν το σύνολο (SYN.), τα φορτηγά (FOR.), τα δεξαμενόπλοια (DEK.) και τα επιβατηγά και λοιπά (EP. & LOIP.)	27
1.3	Πραγματική κυριότητα εμπορικού στόλου: Ελλάδα (GR), Ιαπωνία (YP), Κίνα (CHIN), Γερμανία (GER), Ν. Κορέα (S. KOR), Σιγκαπούρη (SG), ΗΠΑ (US), ΗΒ (UK), Ταϊβάν (TW), Νορβηγία (NOR), Δανία (DK), Βερμούδες (BER), Λοιπές (LOIP.)	28
1.4	Ραβδόγραμμα συχνοτήτων χωρικών δεδομένων. Ελληνικά σχολεία και εκκλησίες στη Μικρά Ασία του 1912. Εξετάζονται ο αριθμός των εκκλησιών (ΕΚΚΛ.) και των κληρικών (ΚΛΗΡ.) στο άνω διάγραμμα, και των δασκάλων (DASK.) καθώς και των μαθητών (MATH.) στο κάτω διάγραμμα για τις περιοχές-επαρχίες Σμύρνης (SM.), Τραπεζούντας (TR.), Προύσας (PR.), Σεβάστειας (SEB.), Ικονίου (IKON.), Χρυσούπολης (XR.), Νικομήδειας (ΝΙΚ.), Γαλατίας (GAL.), Αρδανελίων (DAR.) και Κασταμονής (KAST.) . . .	29
1.5	Ραβδόγραμμα συχνοτήτων και διάγραμμα τιμών χρονολογικών σειρών συστηματικών διαστημάτων: Αποτεφρωμένα στρέμματα στον νομό Ηρακλείου σε μηνιαία βάση από τον Ιανουάριο του 2008 έως και τον Δεκέμβριο του 2012	30
1.6	Διάγραμμα χρονολογικών σειρών μη συστηματικών διαστημάτων: Έλληνες (Ρωμιοί) κάτοικοι της Κωνσταντινούπολης κατά τα έτη από το 1844 έως και το 1994	31
1.7	Διάγραμμα χωρο-χρονικών σειρών: Μηνιαία βροχόπτωση (ύψος σε χιλιοστά) για την περίοδο από τον Σεπτέμβριο του 2005 έως τον Αύγουστο του 2008 σε περιοχές του νομού Ρεθύμνης, με βάση μετρήσεις από τους σταθμούς Ανωγείων (ΑΝ.), Δρυμίσκου (DR.), Ρεθύμνου (RETH.) και Λευκογείων (LEUK.)	31

1.8	Ραβδόγραμμα συχνοτήτων των ενδεικτικών ημερήσιων τιμών αμόλυδης βενζίνης για τους 51 νομούς της χώρας με ημερομηνία παρατήρησης τη 10η Ιουνίου 2014	32
1.9	Γράφημα αθροιστικών συχνοτήτων των ενδεικτικών ημερήσιων τιμών αμόλυδης βενζίνης για τους 51 νομούς της χώρας με ημερομηνία παρατήρησης τη 10η Ιουνίου 2014	33
1.10	Συχνότητες και αθροιστικές συχνότητες του αριθμού γεννήσεων ανά ώρα εντός ενός εικοσιτετραώρου	36
1.11	Ραβδόγραμμα συχνοτήτων με άνισα εύρη κλάσεων των ενδεικτικών ημερήσιων τιμών αμόλυδης βενζίνης για τους 51 νομούς της χώρας με ημερομηνία παρατήρησης τη 10η Ιουνίου 2014	37
1.12	Γραμμή συχνοτήτων από δεδομένα ετήσιας μισθοδοσίας (σε ν.μ.) 21 εργαζομένων σε μια Ανώνυμη Εταιρεία. Η κάθετη διακεκομμένη γραμμή συμβολίζει τη μέση τιμή και η κάθετη μη διακεκομμένη γραμμή τη διάμεσο	46
1.13	Αρνητική συμμετρία, συμμετρία και θετική ασυμμετρία μέσω καμπυλών κατανομής συχνοτήτων και θέσης στατιστικών μέτρων όπως ο μέσος $\bar{x}$, η διάμεσος M και η επικρατούσα τιμή m	52
1.14	Ραβδόγραμμα συχνοτήτων των ενδεικτικών ημερήσιων τιμών αμόλυδης βενζίνης για τους 51 νομούς της χώρας με ημερομηνία παρατήρησης τη 10η Ιουνίου 2014. Η κάθετη διακεκομμένη γραμμή συμβολίζει τη μέση τιμή και η κάθετη μη διακεκομμένη γραμμή τη διάμεσο. Καμπύλη κανονικής κατανομής ως πρότυπο συμμετρίας και μεσοκύρτωσης	55
1.15	Θηκόγραμμα για τα ύψη τριών ομάδων ανδρών: πρώτη ομάδα (Α), δεύτερη ομάδα (Β) και τρίτη ομάδα (C)	56
1.16	Ραβδόγραμμα ανά μήνα του αριθμού των αποτεφρωμένων στρεμμάτων στον νομό Ηρακλείου κατά την περίοδο από το 2008 έως και το 2012. Οι μήνες αποτυπώνονται ως εξής: Ιανουάριος (Ι), Φεβρουάριος (F), Μάρτιος (MAR), Απρίλιος (Α), Μάιος (Μ), Ιούνιος (ΙΟΥΝ), Ιούλιος (ΙΟΥΛ), Αύγουστος (ΑΥ), Σεπτέμβριος (S), Οκτώβριος (Ο), Νοέμβριος (Ν), Δεκέμβριος (D)	57
1.17	Ραβδογράμματα και θηκογράμματα τριών ειδών δεδομένων: συμμετρικά δεδομένα (Α), δεδομένα με αρνητική συμμετρία (Β) και δεδομένα με θετική συμμετρία (C). Η κάθετη διακεκομμένη γραμμή συμβολίζει τη μέση τιμή και η κάθετη μη διακεκομμένη γραμμή τη διάμεσο	58

1.18	Θηκογράμματα για τα δεδομένα των αριθμών γεννήσεων ενός ει- κοσιτετραώρου: δίχως μετασχηματισμό (x), με λογαριθμικό μετα- σχηματισμό $[\log(x)]$ και με μετασχηματισμό της τετραγωνικής ρί- ζας ($\sqrt{x}$)	60
1.19	Παραδείγματα συσχέτισης για x και y μεταβλητές: τέλεια θετική συσχέτιση (A), θετική συσχέτιση (B), τέλεια αρνητική συσχέτιση (C), αρνητική συσχέτιση (D), σχεδόν μηδενική συσχέτιση (E) . . .	63
1.20	Εξίσωση παλινδρόμησης βάσει των διμετάβλητων δεδομένων x και y για τα έτη 1950-1961 (αριστερά). Εξίσωση παλινδρόμησης βάσει των διμετάβλητων δεδομένων x και y για τα έτη 1950-1973 (δεξιά)	71
2.1	Ραβδόγραμμα σχετικών συχνοτήτων των ενδεικτικών ημερήσιων τιμών αμόλυδης βενζίνης για τους 51 νομούς της χώρας με ημε- ρομηνία παρατήρησης τη 10η Ιουνίου 2014	94
2.2	Δειγματικός χώρος και ενδεχόμενα (από κανόνες πιθανοτήτων)	96
2.3	Δενδρικό διάγραμμα πιθανοτήτων για τα ενδεχόμενα του παρα- δείγματος 2.17	105
2.4	Ραβδόγραμμα συναρτήσεων πιθανότητας και αθροιστικής πι- θανότητας της τ.μ. X , που αναφέρεται στον αριθμό των συμβο- λαίων ανά εβδομάδα σε ασφαλιστική εταιρεία	110
2.5	Ραβδόγραμμα συναρτήσεων πιθανότητας μιας ομοιόμορφης τ.μ. X που ορίζεται στον χώρο $\Omega_x = \{1, 2, 3, 4, 5, 6\}$	113
2.6	Ραβδόγραμμα συναρτήσεων πιθανότητας Bernoulli τ.μ. ($p = 0,7$)	114
2.7	Ραβδόγραμμα συναρτήσεων πιθανότητας διωνυμικών τ.μ. με $n = 20$ και αντίστοιχες πιθανότητες $p = 0,1$, $p = 0,25$, $p = 0,5$ και $p = 0,75$	116
2.8	Ραβδόγραμμα συναρτήσεων πιθανότητας και αθροιστικής πι- θανότητας μιας διωνυμικής τ.μ. X με $n = 10$, $p = 0,6$	117
2.9	Ραβδόγραμμα συναρτήσεων πιθανότητας Poisson τ.μ. με $\lambda = 4$.	122
2.10	Οι πιθανότητες $P(X \leq x)$, $P(X > x)$ και $P(\alpha \leq X \leq \beta)$ για μη διακριτή τ.μ. X	124
2.11	Διάγραμμα συνάρτησης αθροιστικής συνάρτησης πυκνότητας πι- θανότητας της τ.μ. X που αναφέρεται στο ύψος της καταβαλλό- μενης αποζημίωσης από ασφαλιστική εταιρεία	126
2.12	Διάγραμμα συναρτήσεων πυκνότητας πιθανότητας και αθροιστι- κής συνάρτησης πυκνότητας πιθανότητας μιας εκθετικής X με $\lambda = 2$ ή $\beta = 1/2$	127
2.13	Συνάρτηση πυκνότητας πιθανότητας ομοιόμορφης τ.μ. που ορί- ζεται στον δειγματικό χώρο $[1, 4]$	129

2.14	Συνάρτηση πυκνότητας πιθανότητας εκθετικών τ.μ. με τιμές για την παράμετρο λ ίσες με 0,2, 0,5, 1, 2 και 5	131
2.15	Συνάρτηση πυκνότητας πιθανότητας Pareto τ.μ. για $x_0 = 2$ και παράμετρο α με τιμές ίσες με 0,2, 0,5, 0,75, 1 και 2	132
2.16	Συνάρτηση πυκνότητας πιθανότητας κανονικών τ.μ. με μέση τιμή ίση με 2 και τυπικές αποκλίσεις με τιμές ίσες με 0,5, 1, 1,2, 1,5 και 2	133
2.17	Πιθανότητες και όρια αυτών για την κανονική κατανομή	134
2.18	Πιθανότητες από το παράδειγμα 2.32: (α) $P(Z > 0)$, (β) $P(Z > 1,15)$, (δ) $P(Z \leq -1,05)$ και (ε) $P(-1,49 < Z < 1,85)$	138
2.19	Συνάρτηση πυκνότητας πιθανότητας λογαριθμικοκανονικής τ.μ. με μέση τιμή ίση με 1 και τυπικές αποκλίσεις με τιμές ίσες με 0,5, 1, 1,2, 1,5 και 2	142
2.20	Συνάρτηση πυκνότητας πιθανότητας της μονοπαράμετρης Γάμια τ.μ. με τιμές k ίσες με 0,5, 1, 1,5, 2 και 3	144
2.21	Συνάρτηση πυκνότητας πιθανότητας χ^2 τ.μ. με βαθμούς ελευθερίας ίσους με 1, 2, 5, 10 και 20	146
2.22	Συνάρτηση πυκνότητας πιθανότητας Student-t τ.μ. με μέση τιμή ίση με 0, διακύμανση ίση με 1 και βαθμούς ελευθερίας ίσους με 2, 5, 10, 50 και 100	147
2.23	Εμπειρική συνάρτηση πυκνότητας πιθανότητας απόδοσης ενός χρεογράφου (με κυματοειδή γραμμή) και των Student-t τ.μ. με μέση τιμή ίση με 0, διακύμανση ίση με 1 και βαθμούς ελευθερίας ίσους με 2, 5, 10, 50 και 100	147
2.24	Συνάρτηση πυκνότητας πιθανότητας F τ.μ. με διάφορα ζεύγη βαθμών ελευθερίας	149
2.25	Ύστερη συνάρτηση πυκνότητας πιθανότητας του ενδεχομένου $\theta = a$	165
2.26	Πρότερη (αριστερά) και ύστερη (δεξιά) συνάρτηση πυκνότητας πιθανότητας της παραμέτρου λ	167
3.1	Περιοχές H_0 και H_1 από τον μονόπλευρο έλεγχο υπόθεσης (ΕΥ1), καθώς και μέτρο p (ΕΥ2) στον ίδιο έλεγχο της ενότητας 3.2	189
3.2	Περιοχές H_0 και H_1 για μονόπλευρους και αμφίπλευρους ελέγχους υπόθεσης (ΕΥ1) για αναλογία p	194
3.3	Περιοχές $H_0 : p = 0,005$ και $H_1 : p \neq 0,005$ από τον αμφίπλευρο έλεγχο υπόθεσης (ΕΥ1), καθώς και μέτρο p (ΕΥ2) στον έλεγχο του παραδείγματος 3.2	199
3.4	Περιοχές $H_0 : \mu \leq 50$ και $H_1 : \mu > 50$ από τον μονόπλευρο έλεγχο υπόθεσης (ΕΥ1), καθώς και μέτρο p (ΕΥ2) στον έλεγχο του παραδείγματος 3.6	208

3.5	Περιοχές H_0 και H_1 για μονόπλευρους και αμφίπλευρους ελέγχους υπόθεσης (ΕΥ1) για τη διακύμανση σ^2	214
3.6	Περιοχές $H_0 : \sigma_{\kappa\nu}^2 \leq 8$ και $H_1 : \sigma_{\kappa\nu}^2 > 8$ από τον μονόπλευρο έλεγχο υπόθεσης διακύμανσης (ΕΥ1) , καθώς και μέτρο p (ΕΥ2) στον έλεγχο του παραδείγματος 3.8	217
3.7	Θηκογράμματα δεδομένων κατανάλωσης καυσίμου ανά 100 χλμ. στις 20.000 στροφές το δευτερόλεπτο για δεδομένα από τα εργοστάσια 1, 2 και 3	230
3.8	Διάστημα εμπιστοσύνης του πληθυσμιακού μέσου μ με πιθανότητα 95%	236
4.1	Διάγραμμα παρατηρήσεων του δείκτη γονιμότητας y και του ποσοστού αποχώρησης από τη δευτεροβάθμια εκπαίδευση x σε δέκα επαρχίες της Ελδετίας	279
6.1	Ιστόγραμμα συχνοτήτων αφίξεων πελατών σε ταμείο εξυπηρέτησης δημόσιου οργανισμού ανά ώρα	299
6.2	Διάγραμμα μηνιαίων αεροπορικών αφίξεων (σε χιλ.) στα αεροδρόμια Ηρακλείου (IR.), Χανίων (CH.) και Ρόδου (ROD.) για το έτος 2013.	301
6.3	Πίτα μηνιαίων αεροπορικών αφίξεων στο αεροδρόμιο των Χανίων για το έτος 2013	301
6.4	Πίτα συνολικών (ετήσιων) αεροπορικών αφίξεων στα αεροδρόμια Χανίων, Ηρακλείου και Ρόδου για το έτος 2013	302
6.5	Ραβδόγραμμα συχνοτήτων σε διάρκεια ζωής ασθενών (σε ημέρες) με ίσες τάξεις. Επίσης, αποτυπώνονται η δειγματική μέση τιμή (διακεκομμένη) και η διάμεσος (μη διακεκομμένη)	304
7.1	Θηκόγραμμα αεροπορικών αφίξεων στα αεροδρόμια Ηρακλείου (IR.), Χανίων (CH.) και Ρόδου (ROD.) για το έτος 2013	311
7.2	Διμετάβλητη απεικόνιση της μεταδλητής x (taxut.) που αναφέρεται στην ταχύτητα αυτοκινήτου και της μεταδλητής y (apost.) που αναφέρεται στην απόσταση ακινητοποίησης	312
8.1	Ιστογράμματα συχνοτήτων από ρίψη ζαριού $n = 4$ φορές, $n = 40$ φορές, $n = 400$ και $n = 4.000$ φορές αντίστοιχα (παράδειγμα διακριτής ομοιόμορφης κατανομής)	315
8.2	Διαγράμματα και ιστογράμματα συχνοτήτων ομοιόμορφα κατανεμημένων ψευδοσειρών στον χώρο $[0, 1]$ (μη διακριτές) για $n = 10$ και $n = 1.000$	317

9.1	Απλή γραμμική παλινδρόμηση της μεταβλητής x (taxut.), που αναφέρεται ως ταχύτητα αυτοκινήτου, και της y (apost.), που αναφέρεται ως απόσταση ακινητοποίησης	331
9.2	Πρόβλεψη και 95% διάστημα εμπιστοσύνης στην απλή γραμμική παλινδρόμηση, όπου η μεταβλητή x (taxut.) αναφέρεται ως ταχύτητα αυτοκινήτου και η y (apost.) ως απόσταση ακινητοποίησης	333
9.3	Διαγνωστικά διαγράμματα καταλοίπων από απλή γραμμική παλινδρόμηση της μεταβλητής x (taxut.) ως ταχύτητας αυτοκινήτου και της y (apost.) ως απόστασης ακινητοποίησης	334
9.4	Μη γραμμική παλινδρόμηση με χρήση μετασχηματισμού Box-Cox για $\lambda = 0,4$, όπου η μεταβλητή x (taxut.) αναφέρεται ως ταχύτητα αυτοκινήτου και η y (apost.) ως απόσταση ακινητοποίησης	338
9.5	Διαγνωστικά διαγράμματα καταλοίπων στη μη γραμμική παλινδρόμηση με χρήση μετασχηματισμού Box-Cox για $\lambda = 0,4$, όπου η μεταβλητή x (taxut.) αναφέρεται ως ταχύτητα αυτοκινήτου και η y (apost.) ως απόσταση ακινητοποίησης	340
10.1	Διάγραμμα μηνιαίων μέσων θερμοκρασιών από τον σταθμό Ηρακλείου Κρήτης κατά την περίοδο από το 1960 έως και το 2004	345

Κατάλογος πινάκων

1.1	Συναρτήσεις συχνοτήτων των X και Y	63
2.1	Πίνακας αθροιστικής συνάρτησης πυκνότητας πιθανότητας της τυποποιημένης κανονικής κατανομής για διάφορες τιμές z . . .	135
2.2	Συναρτήσεις πιθανοτήτων των X και Y	150
II.1	Πίνακας αθροιστικής συνάρτησης πιθανότητας $P(X \leq x)$ της διωνυμικής κατανομής	360
II.2	Πίνακας αθροιστικής συνάρτησης πιθανότητας $P(X \leq x)$ της κατανομής Poisson	362
II.3	Πίνακας αθροιστικής συνάρτησης πυκνότητας πιθανότητας $P(Z \leq z)$ της τυποποιημένης κανονικής κατανομής	363
II.4	Πίνακας της Student-t κατανομής	364
II.5	Πίνακας της X^2 κατανομής	365
II.6	Πίνακας της F κατανομής για $\alpha = 5\%$	366
II.7	Πίνακας της F κατανομής για $\alpha = 1\%$	367
II.8	Πίνακας κριτικών τιμών $D_{n,\alpha}$ ελέγχου Kolmogorov-Smirnov . . .	368
II.9	Πίνακας κριτικών τιμών (άνω και κάτω όριο) ελέγχου τύπου Wilcoxon	369
II.10	Πίνακας κριτικών τιμών (άνω και κάτω όριο) ελέγχου αθροιστικής ιεράρχησης τύπου Wilcoxon	370
II.11	Πίνακας κριτικών τιμών (άνω και κάτω όριο) ελέγχου αθροιστικής ιεράρχησης τύπου Wilcoxon	370
II.12	Πίνακας κριτικών τιμών ελέγχου τύπου Spearman	371

Πρόλογος

[...] εάν κατάσχω τον λόγον και
μη μεταδώ, τότε πένης εμί, όταν
δε καταβάλω, τότε ευπορώτερος
γίνομαι, εάν μη καταβάλω τον
λόγον, τότε μόνος πλουτώ, εάν
δε καταβάλω, μετά πάντων υμών
τον καρπόν αποφέρομαι.

Ιωάννης ο Χρυσόστομος,
«Περι ελεημοσύνης
και εις τα Δέκα Παρθένους»

Λίγα φυσικά φαινόμενα ή ανθρώπινες συμπεριφορές χαρακτηρίζονται από βεβαιότητα. Αν και υπάρχουν φυσικοί νόμοι αλλά και επιστημονικές θεωρίες που τα ορίζουν, κανείς δεν μπορεί με βεβαιότητα να μιλήσει για την ακριδή τους έκβαση. Στατιστική είναι η επιστήμη που διερευνά την αβεβαιότητα μέσω της συλλογής δείγματος τέτοιων φαινομένων, της οργάνωσης, αλλά και της μελέτης τους. Οι εφαρμογές της στατιστικής καλύπτουν όλο σχεδόν το φάσμα των επιστημών, καθιστώντας την ένα πολύτιμο εργαλείο. Έτσι, βλέπουμε εφαρμογές της στατιστικής σε κλάδους ή επιστήμες όπως η αρχαιολογία, η βιολογία, η δημογραφία, οι οικονομικές επιστήμες, η διοίκηση επιχειρήσεων, η γεωφυσική, η σεισμολογία, η ιατρική, η ψυχολογία και πολλές άλλες. Στην πράξη, η στατιστική ανάλυση είναι χρήσιμη, για παράδειγμα, στη χρονολόγηση αντικειμένων από ανασκαφές, στην επιδημιολογική ανάλυση, στην πρόβλεψη πληθυσμιακών μεταβολών, στην πρόβλεψη οικονομικών κύκλων, στην οργάνωση και πρόβλεψη πωλήσεων, στην πρόγνωση επικίνδυνης σεισμικής δραστηριότητας, στην ανάλυση της αποτελεσματικότητας μιας φαρμακευτικής αγωγής, στην πρόβλεψη μιας ανθρώπινης συμπεριφοράς ύστερα από κάποιο ερέθισμα κ.ά.

Στο παρόν βιβλίο θα καλύψουμε τρεις βασικές ενότητες. Στο πρώτο κεφάλαιο εξετάζουμε την ανάλυση δεδομένων, αναπτύσσοντας θέματα που έχουν να κάνουν με την ταξινόμηση, την απλή περιγραφική στατιστική ανάλυση και την παρουσίαση δεδομένων τυχαίων φαινομένων. Πρόκειται για μια αρχική προσέγγιση των δεδομένων για το πώς πρέπει να μεταφέρουμε πληροφορία σε σχέση

με το δείγμα ενός τυχαίου φαινομένου. Αυτή η πληροφόρηση στηρίζεται σε βασικά στατιστικά μέτρα και γραφικές παραστάσεις. Επίσης, γίνεται μια πρώτη αναφορά στη χρήση και την ανάλυση δύο τυχαίων φαινομένων ταυτόχρονα (διμετάβλητη ανάλυση) και στην απλή παλινδρόμηση. Στο δεύτερο κεφάλαιο θα καλύψουμε τη θεωρία πιθανοτήτων, που αποτελεί το βασικό εργαλείο της στατιστικής ανάλυσης δεδομένων. Αρχικά, αναπτύσσουμε τη στατιστική θεωρία μέσω νόμων και αξιωμάτων και μελετάμε τον τρόπο διαμόρφωσης στατιστικών υποδειγμάτων. Στη συνέχεια, ορίζουμε την έννοια της τυχαίας μεταβλητής και τη συνδέουμε με τη θεωρία πιθανοτήτων, για να φτάσουμε στην έννοια της κατανομής πιθανοτήτων. Αντικείμενο του τρίτου κεφαλαίου αποτελεί η επαγωγική στατιστική. Εδώ γίνεται σύνδεση μεταξύ δειγματικών και πληθυσμιακών (γενικών) συμπερασμάτων για το προς εξέταση τυχαίο φαινόμενο μέσω των κατανομών πιθανοτήτων. Στόχος μας είναι ο προσδιορισμός διαστημάτων εμπιστοσύνης και η πραγματοποίηση ελέγχων στατιστικής υπόθεσης για γνωστά στατιστικά μέτρα. Το κεφάλαιο ολοκληρώνεται με τη χρήση επαγωγικής στατιστικής στην ανάλυση παλινδρόμησης. Τέλος, στα υπολοιπα κεφάλαια μελετάμε την εφαρμογή των πιο πάνω εννοιών χρησιμοποιώντας το στατιστικό λογισμικό R.

Η παρούσα έκδοση καλύπτει κυρίως τις ανάγκες των φοιτητών των πρώτων ετών φοίτησης στο Τμήμα Οικονομικών Επιστημών του Πανεπιστημίου Κρήτης. Επίσης, ενδείκνυται και για φοιτητές ή αποφοίτους λοιπών πανεπιστημιακών σχολών, αλλά και για πρακτικούς της στατιστικής ανάλυσης, αφού περιλαμβάνει πολλά θέματα στατιστικής θεωρίας, καθώς και εφαρμογές. Έχει δοθεί ιδιαίτερη προσοχή, έτσι ώστε να υπάρχουν αρκετά πρακτικά παραδείγματα που αφορούν την καθημερινή λήψη αποφάσεων, στηριζόμενα πάντα στην απαραίτητη για την κατανόησή τους στατιστική θεωρία. Ευελπιστώ ότι αυτό το βιβλίο, στο πλαίσιο της κατανόησης των πραγμάτων που μας περιβάλλουν, θα μεταφέρει στον αναγνώστη το αίσθημα της αβεβαιότητας το οποίο διέπει κάθε λήψη απόφασης αναφορικά με τις ανθρώπινες αλλά και τις φυσικές συμπεριφορές.

Τέλος, θα ήθελα να ευχαριστήσω τα στελέχη έκδοσης του εκδοτικού οίκου Πεδίο, και πιο συγκεκριμένα τον κ. Άγη Πετρόπουλο και την κ. Έφη Ρουμπούλα, για την αρωγή τους στην επιτυχή έκδοση του παρόντος βιβλίου.

Ρέθυμνο, Αύγουστος 2022
ΓΕΩΡΓΙΟΣ ΚΩΝ. ΤΣΙΩΤΑΣ

Σύμβολα και συντομογραφίες

ν.μ.	Νομισματικές μονάδες
τ.μ.	Τυχαία(ες) μεταβλητή(ές)
σ.(π.)π.	Συνάρτηση (πυκνότητας) πιθανότητας
α.σ.(π.)π.	Αθροιστική συνάρτηση (πυκνότητας) πιθανότητας
$\forall$	Για κάθε
ω	Δειγματικό σημείο
Ω	Δειγματικός χώρος
$\mathcal{R}$	Σύνολο πραγματικών αριθμών
$\sum$	Άθροισμα
$\prod$	Γινόμενο
$\propto$	Αναλογικό
$\rightarrow$	Τείνει προς
$\sim$	Κατανέμεται ως
$\perp$	Ανεξάρτητη
sup	Ανώτερη
f_i, F_i	Συχνότητα, αθροιστική συχνότητα
$f(\cdot, \cdot), P(\cdot, \cdot)$	Από κοινού συχνότητα, από κοινού πιθανότητα
$F(x), P(X \leq x)$	Αθροιστική συνάρτηση (πυκνότητας) πιθανότητας της x
$f_X(x), P(X = x)$	Συνάρτηση (πυκνότητας) πιθανότητας της x
$f_{X,Y}(x, y), f_{X Y}(x y)$	Από κοινού και δεσμευμένη συνάρτηση (πυκνότητας) πιθανότητας για x και y
$P(\cdot \cdot)$	Δεσμευμένη πιθανότητα
$\phi(x), \Phi(x)$	Συνάρτηση πυκνότητας πιθανότητας και αθροιστική σ.π.π.
$E(\cdot)$	Κανονικής τ.μ. x
$\mu, \bar{x}$	Μαθηματική ελπίδα ή μέσος
M	Πληθυσμιακός και δειγματικός μέσος
m	Διάμεσος
$V(\cdot)$	Επικρατούσα τιμή
σ^2, s^2	Διακύμανση
Q_i	Πληθυσμιακή και δειγματική διακύμανση
P_i	Τεταρτημόριο i
$Cov(\cdot, \cdot)$	Ποσοστημόριο i
ρ	Συνδιακύμανση
$\mathcal{U}(k)$	Συντελεστής συσχέτισης
$\mathcal{B}(p)$	Ομοιόμορφη (διακριτή) κατανομή με παράμετρο k
$\mathcal{B}(n, p)$	Bernoulli κατανομή με παράμετρο p
$\mathcal{P}(\lambda)$	Διωνυμική κατανομή με παραμέτρους n και p
$\mathcal{U}(\alpha, \beta)$	Poisson κατανομή με παράμετρο λ
$\mathcal{E}(\lambda)$	Ομοιόμορφη (μη διακριτή) κατανομή με παραμέτρους α και β
$\mathcal{PA}(\alpha, x_0)$	Εκθετική κατανομή με παράμετρο λ
$\mathcal{N}(\mu, \sigma^2)$	Pareto κατανομή με παραμέτρους α και x_0
$\mathcal{LG}(\mu, \sigma^2)$	Κανονική (ή κατά Gauss) κατανομή με μέσο μ και διακύμανση σ^2
$\mathcal{GA}(k)$	Λογαριθμικοκανονική κατανομή με μέσο μ και διακύμανση σ^2
$\mathcal{GA}(k, \beta)$	Γάμμα κατανομή (μονοπαράμετρο) με παράμετρο k
χ_n^2	Γάμμα κατανομή (διπαράμετρο) με παραμέτρους k και β
T_n	χ^2 κατανομή με παράμετρο n
$\mathcal{F}_{n_1, n_2}$	Student-t κατανομή με παράμετρο n
$L(\cdot)$	F κατανομή με παραμέτρους n_1 και n_2
$l(\cdot)$	Συνάρτηση πιθανοφάνειας
	Λογαριθμική συνάρτηση πιθανοφάνειας