

Περιεχόμενα

Πρόλογος	13
1. Εισαγωγή στην R	15
1.1 Το ιστορικό και η κοινότητα της R	15
1.2 Μεταφόρτωση και εγκατάσταση R και RStudio	15
1.2.1 Εγκατάσταση της R	15
1.2.2 Εγκατάσταση του Rstudio	17
1.2.3 Εγκατάσταση πακέτων και βιβλιοθηκών	19
1.3 Περιβάλλοντα εργασίας	19
1.3.1 Περιβάλλον της R	19
1.3.2 Περιβάλλον του Rstudio	26
1.3.3 Αποθήκευση script	27
2. Ανακαλύπτοντας την R	28
2.1 Σύνταξη εντολών	28
2.2 Αριθμητικοί τελεστές	29
2.3 Λογικοί τελεστές και τελεστές σύγκρισης	29
2.4 Βασικές αριθμητικές συναρτήσεις	30
2.5 Αντικείμενα και κλάσεις αντικειμένων	31
2.6 Διανύσματα	34
2.7 Πίνακες	36
2.8 Πολλαπλοί πίνακες	42
2.9 Λίστες	42
2.10 Πλαίσια δεδομένων	45
2.11 Παράγοντες	51
2.12 Ελλιπείς τιμές	52
3. Εισαγωγή και αποθήκευση δεδομένων	53
3.1 Εισαγωγή δεδομένων από αρχείο	53
3.1.1 Εισαγωγή δεδομένων από αρχείο Excel	53
3.1.2 Εισαγωγή δεδομένων από αρχείο SPSS	55
3.2 Επιλογή περιπτώσεων από το αρχείο δεδομένων	57
3.2.1 Επιλογή μεταβλητών από σύνολο δεδομένων	57

3.2.2	Επιλογή περιπτώσεων από σύνολο δεδομένων	58
3.2.3	Επιλογή περιπτώσεων από σύνολο μεταβλητών	59
3.3	Αποθήκευση δεδομένων σε αρχείο	60
3.4	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	61
4.	Επανακωδικοποιήσεις και υπολογισμοί	63
4.1	Αντικατάσταση τιμών μεταβλητών με ελλείπουσες τιμές	63
4.2	Επανακωδικοποίηση μεταβλητών	63
4.3	Εφαρμογή στην R	63
4.3.1	Αντικατάσταση τιμών μεταβλητών με ελλείπουσες τιμές	63
4.3.2	Επανακωδικοποίηση μεταβλητών	65
4.4	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	68
5.	Περιγραφική στατιστική	72
5.1	Μέτρα θέσης (μέτρα κεντρικής τάσης)	72
5.1.1	Μέση τιμή (mean)	72
5.1.2	Διάμεσος (median)	73
5.1.3	Επικρατούσα τιμή (mode)	73
5.1.4	Τεταρτημόρια (quantiles)	74
5.1.5	Ποσοστιαία σημεία (percentiles)	74
5.2	Μέτρα διασποράς	74
5.2.1	Διακύμανση ή διασπορά (variance)	74
5.2.2	Τυπική απόκλιση (standard deviation)	75
5.2.3	Εύρος (range)	75
5.2.4	Ενδοτεταρτημοριακό εύρος (interquartile range)	75
5.2.5	Τυπικό σφάλμα (standard error)	76
5.3	Άλλα μέτρα περιγραφικής στατιστικής	76
5.3.1	Συντελεστής ασυμμετρίας (skewness coefficient)	76
5.3.2	Συντελεστής μεταβλητότητας (CV-coefficient of variation)	76
5.3.3	Συντελεστής κύρτωσης (kurtosis coefficient)	77
5.4	Εφαρμογή μέτρων περιγραφικής στατιστικής στην R	77
5.4.1	Μέση τιμή (mean)	78
5.4.2	Διάμεσος (median)	79
5.4.3	Τεταρτημόρια (quantiles)	79
5.4.4	Ποσοστιαία σημεία (percentiles)	80
5.4.5	Διακύμανση ή διασπορά (variance)	81
5.4.6	Τυπική απόκλιση (standard deviation)	81
5.4.7	Εύρος (range)	82
5.4.8	Ενδοτεταρτημοριακό εύρος (interquartile range)	83
5.4.9	Τυπικό σφάλμα (standard error)	84
5.4.10	Συντελεστής ασυμμετρίας (skewness coefficient)	85

5.4.11	Συντελεστής μεταβλητότητας (CV-coefficient of variation)	86
5.4.12	Συντελεστής κύρτωσης (kurtosis coefficient)	87
5.5	Εντολή summary()	88
5.6	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	89
6.	Πίνακες και γραφήματα περιγραφικής στατιστικής	94
6.1	Βασικοί πίνακες περιγραφικής στατιστικής	94
6.1.1	Πίνακες συχνοτήτων	94
6.1.2	Πίνακες συνάφειας	95
6.2	Βασικά γραφήματα περιγραφικής στατιστικής	96
6.2.1	Ιστόγραμμα (Histogram)	96
6.2.2	Θηκόγραμμα (box-plot)	97
6.2.3	Κυκλικό διάγραμμα (pie chart)	98
6.2.4	Διάγραμμα σκεδασμού (scatter plot)	99
6.2.5	Ραβδόγραμμα (bar chart)	99
6.3	Εφαρμογή στην R	100
6.3.1	Πίνακες συχνοτήτων	100
6.3.2	Πίνακες συνάφειας	101
6.3.3	Ιστόγραμμα (Histogram)	103
6.3.4	Θηκόγραμμα (box-plot)	106
6.3.5	Κυκλικό διάγραμμα (pie chart)	108
6.3.6	Διάγραμμα σκεδασμού (scatter plot)	110
6.3.7	Ραβδόγραμμα (bar chart)	114
6.4	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	124
7.	Στατιστική συμπερασματολογία	143
7.1	Έλεγχοι υποθέσεων	143
7.2	Σφάλματα	144
8.	Έλεγχος μέσων όρων	146
8.1	Εισαγωγή	146
8.1.1	Έλεγχος μέσου όρου ενός πληθυσμού (One Sample t-test)	146
8.1.2	Σύγκριση μέσων όρων δύο πληθυσμών – ανεξάρτητων δειγμάτων (Independent Samples t-test)	148
8.1.3	Σύγκριση μέσων όρων δύο πληθυσμών – εξαρτημένα/ζευγαρωτά δείγματα (Paired Samples t-test)	150
8.2	Προϋποθέσεις εφαρμογής των ελέγχων για τους μέσους όρους	152
8.3	Μη παραμετρικοί έλεγχοι για τους μέσους όρους	153
8.4	Παρουσίαση των αποτελεσμάτων	153
8.5	Εφαρμογή ελέγχων για τους μέσους όρους στην R	154
8.5.1	Έλεγχος μέσου όρου ενός πληθυσμού (One Sample t-test)	154

8.5.2	Σύγκριση μέσων όρων δύο πληθυσμών – ανεξάρτητων δειγμάτων (Independent Samples t-test)	158
8.5.3	Σύγκριση μέσων όρων δύο πληθυσμών – εξαρτημένα/ζευγαρωτά δείγματα (Paired Samples t-test)	164
8.5.4	Μη παραμετρικός έλεγχος μέσου όρου ενός πληθυσμού (One Sample Wilcoxon Test)	167
8.5.5	Μη παραμετρικός έλεγχος σύγκρισης μέσων όρων δύο πληθυσμών – ανεξάρτητων δειγμάτων	169
8.5.6	Μη παραμετρικός έλεγχος σύγκρισης μέσων όρων δύο πληθυσμών – εξαρτημένα/ζευγαρωτά δείγματα (Wilcoxon Signed Rank Test)	170
8.5.7	Ενδεικτικές προτάσεις παρουσίασης των αποτελεσμάτων	172
8.6	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	173
9.	Ανάλυση διακύμανσης	190
9.1	Εισαγωγή	190
9.2	Ανάλυση διακύμανσης μονής κατεύθυνσης (One-way ANOVA)	190
9.3	Προϋποθέσεις εφαρμογής One-way ANOVA	190
9.4	Εφαρμογή One-way ANOVA	191
9.5	Post hoc ανάλυση	192
9.6	Παραβίαση προϋποθέσεων εφαρμογής του ελέγχου	192
9.6.1	Παραβίαση κανονικής κατανομής – Μη παραμετρικός έλεγχος	192
9.6.2	Παραβίαση ισότητας διακυμάνσεων	193
9.6.3	Παραβίαση τυχαίων δειγμάτων	193
9.7	Παρουσίαση των αποτελεσμάτων	193
9.8	Εφαρμογή One-way ANOVA στην R	193
9.8.1	Έλεγχος Levene για τον έλεγχο ομοιογένειας των διακυμάνσεων	194
9.8.2	Ανάλυση διακύμανσης μονής κατεύθυνσης (One-way ANOVA)	196
9.8.3	Post hoc αναλύσεις	196
9.8.4	Μη παραμετρικός έλεγχος Kruskal-Wallis	199
9.8.5	Ενδεικτικές προτάσεις παρουσίασης των αποτελεσμάτων	200
9.9	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	201
10.	Έλεγχος απλής γραμμικής συσχέτισης	212
10.1	Εισαγωγή	212
10.2	Συντελεστής συσχέτισης Pearson r	212
10.2.1	Προϋποθέσεις εφαρμογής του συντελεστή συσχέτισης Pearson r	212
10.2.2	Υπολογισμός του συντελεστή συσχέτισης Pearson r	213
10.3	Εφαρμογή απλής γραμμικής συσχέτισης	214
10.4	Μη παραμετρικοί έλεγχοι συσχέτισης	215
10.4.1	Συντελεστής συσχέτισης Spearman rho	215
10.4.2	Συντελεστής συσχέτισης Kendall's tau	215

10.5	Παρουσίαση των αποτελεσμάτων	216
10.6	Εφαρμογή ελέγχου απλής γραμμικής συσχέτισης στην R	216
10.6.1	Συντελεστής συσχέτισης Pearson r	216
10.6.2	Συντελεστής συσχέτισης Spearman ρ	218
10.6.3	Συντελεστής συσχέτισης Kendall's τ	219
10.6.4	Ενδεικτικές προτάσεις παρουσίασης των αποτελεσμάτων	220
10.7	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	221
11.	Ανάλυση παλινδρόμησης	227
11.1	Εισαγωγή	227
11.2	Απλή γραμμική παλινδρόμηση	227
11.2.1	Απλή γραμμική παλινδρόμηση με εξαρτημένη συνεχή μεταβλητή	227
11.3	Εφαρμογή της απλής γραμμικής παλινδρόμησης	228
11.4	Πολλαπλή γραμμική παλινδρόμηση	229
11.5	Εφαρμογή της πολλαπλής γραμμικής παλινδρόμησης	229
11.6	Προϋποθέσεις εφαρμογής γραμμικής παλινδρόμησης	230
11.7	Παρουσίαση των αποτελεσμάτων	231
11.8	Εφαρμογή γραμμικής παλινδρόμησης στην R	232
11.8.1	Απλή γραμμική παλινδρόμηση με συνεχή εξαρτημένη μεταβλητή.....	232
11.8.2	Γραμμή πρόβλεψης στην απλή γραμμική παλινδρόμηση	234
11.8.3	Πολλαπλή γραμμική παλινδρόμηση	234
11.8.4	Προϋποθέσεις εφαρμογής γραμμικής παλινδρόμησης	236
11.8.5	Ενδεικτικές προτάσεις παρουσίασης των αποτελεσμάτων	239
11.9	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	240
12.	Έλεγχος χ^2	245
12.1	Εισαγωγή	245
12.2	Προϋποθέσεις εφαρμογής του ελέγχου χ^2	247
12.3	Εφαρμογή του ελέγχου χ^2	247
12.4	Παρουσίαση των αποτελεσμάτων	248
12.5	Εφαρμογή του ελέγχου χ^2 στην R	248
12.5.1	Εφαρμογή 1	248
12.5.2	Εφαρμογή 2	253
12.5.3	Ενδεικτικές προτάσεις παρουσίασης των αποτελεσμάτων	257
12.6	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	259
13.	Διερευνητική παραγοντική ανάλυση	268
13.1	Εισαγωγή	268
13.2	Διαδικασία εφαρμογής διερευνητικής παραγοντικής ανάλυσης	268
13.2.1	Μέθοδοι εξαγωγής παραγόντων	268

13.2.2	Περιστροφή παραγόντων	268
13.2.3	Εξαγωγή παραγόντων	269
13.3	Προϋποθέσεις εφαρμογής διερευνητικής παραγοντικής ανάλυσης	269
13.4	Αξιοπιστία και εσωτερική συνοχή παραγόντων	269
13.4.1	Cronbach's alpha	270
13.4.2	McDonald's omega	270
13.5	Εφαρμογή διερευνητικής παραγοντικής ανάλυσης	271
13.6	Παρουσίαση των αποτελεσμάτων	271
13.7	Εφαρμογή διερευνητικής παραγοντικής ανάλυσης στην R	272
13.7.1	KMO και Bartlett's Test	272
13.7.2	Ανάλυση κύριων συνιστωσών (PCA)	275
13.7.3	Cronbach's alpha	278
13.7.4	Ενδεικτικές προτάσεις παρουσίασης των αποτελεσμάτων	281
13.8	Ασκήσεις αυτοαξιολόγησης και ενδεικτικές απαντήσεις	282
14.	Συμπληρωματικοί έλεγχοι	287
14.1	Έλεγχος κανονικής κατανομής των δεδομένων	287
14.1.1	Εφαρμογή ελέγχου κανονικής κατανομής στην R	287
14.2	Έλεγχος γραμμικής σχέσης μεταξύ των μεταβλητών	290
14.2.1	Έλεγχος γραμμικής σχέσης των μεταβλητών με το διάγραμμα σκεδασμού	290
14.2.2	Εφαρμογή ελέγχου γραμμικής σχέσης των μεταβλητών στην R	291
14.3	Μέγεθος επίδρασης	293
14.3.1	Μέγεθος επίδρασης συσχέτισης	293
14.3.2	Μέγεθος επίδρασης μέσω των όρων	293
14.3.3	Μέγεθος επίδρασης ελέγχου διακύμανσης	295
14.3.4	Μέγεθος επίδρασης ελέγχου χ^2	296
14.3.5	Τιμές μεγέθους επίδρασης	297
15.	Βιβλιογραφία	298
Παράρτημα: Η διεθνής αξιολόγηση PISA		299
1.	Περιγραφή του προγράμματος	299
2.	Δημιουργία βάσης δεδομένων του PISA	299
3.	Περιγραφή της βάσης δεδομένων που αξιοποιείται	300

Πρόλογος

Η ιδέα για το συγκεκριμένο βιβλίο γεννήθηκε το 2016, όταν σε ένα επιμορφωτικό σεμινάριο του ETS στο Princeton διαπίστωσα ότι ο αριθμός των συναδέλφων που αξιοποιούσαν τις δυνατότητες της R, όχι μόνο για την ανάλυση ερευνητικών δεδομένων αλλά και για την εκπαίδευση φοιτητών στο πεδίο των Επιστημών της Αγωγής, ήταν πολύ μεγαλύτερος από τον αριθμό που είχα εκτιμήσει. Οι συζητήσεις με έπεισαν ότι έχει νόημα να επιχειρήσω (και όχι μόνο για λόγους βελτίωσης της υπολογιστικής και κριτικής σκέψης) την επαφή (και) των φοιτητών με τους οποίους συνεργάζομαι με την R. Οι πρώτες προσπάθειες εξοικείωσης έγιναν στο πλαίσιο επιλεγόμενων μαθημάτων σε Προγράμματα Μεταπτυχιακών Σπουδών Τμημάτων Επιστημών Αγωγής το 2017, ενώ το 2018 επιχειρήθηκε για πρώτη φορά η επεξεργασία ερευνητικών δεδομένων σε επιλεγόμενο μάθημα του Προπτυχιακού Προγράμματος Σπουδών. Η υποδοχή της πρότασης από το (ολιγάριθμο οφείλω να ομολογήσω) ακροατήριο των φοιτητών ήταν ενθουσιώδης, αντικαθιστώντας τις αρχικές επιφυλάξεις μου με φιλόδοξους σχεδιασμούς. Η Κατερίνα Σαργιώτη, μεταπτυχιακή φοιτήτρια τότε, ήταν από τους ανθρώπους που προκάλεσαν το επόμενο βήμα: έχοντας έρθει ήδη σε επαφή με την R στο Πρόγραμμα Μεταπτυχιακών Σπουδών, αντιμετώπισε με ενθουσιασμό την ιδέα για ένα βιβλίο αφιερωμένο στους φοιτητές των Σχολών Αγωγής που ενδιαφέρονται να «απαγκιστρωθούν» από τον (εργαλειακό) χειρισμό των εφαρμογών (και ιδιαίτερα των εμπορικών εφαρμογών). Η πρόδρομη μορφή υλικού που συμπεριλαμβάνεται στο βιβλίο δοκιμάστηκε στα πρώτα μαθήματα αξιοποίησης της R και του RStudio στην επεξεργασία και ανάλυση εμπειρικού υλικού στο Πρόγραμμα Προπτυχιακών Σπουδών του Παιδαγωγικού Τμήματος Δημοτικής Εκπαίδευσης στο Πανεπιστήμιο Ιωαννίνων και διορθώθηκε, προκειμένου να ανταποκριθεί στην καλύτερη δυνατή αξιοποίηση από τους φοιτητές. Ήταν επιλογή των συγγραφέων η έμφαση να δοθεί στην αξιοποίηση των δυνατοτήτων που παρέχουν η R και το RStudio στην αρχική ανάλυση των δεδομένων, χωρίς να απαιτείται μια (πλήρης) μαθηματική υποδομή. Τα παραδείγματα που παρουσιάζονται, όπως και οι ασκήσεις που προτείνονται, βασίζονται σε προσαρμοσμένο –προκειμένου να ανταποκριθεί στην εκπαιδευτική του στόχευση– ερευνητικό υλικό του Programme for International Student Assessment (κύμα 2015). Επισημαίνεται ότι το υλικό δεν αξιοποιείται λαμβάνοντας υπόψη το μεθοδολογικό πλαίσιο, τις απαιτήσεις, τις παραδοχές και τους περιορισμούς χρήσης του. Για τους συγγραφείς τα συγκεκριμένα δεδομένα αποτελούν απλώς μια βάση αριθμών.

Και μια δεύτερη επισήμανση: η R εξελίσσεται διαρκώς, καθώς μια πολυάριθμη κοινότητα ερευνητών την υποστηρίζει. Οι διορθώσεις αποκαθιστούν δυσλειτουργίες, αλλά και προκαλούν νέες, καθώς απαιτείται η επικαιροποίηση επιμέρους στοιχείων της.

Στη θέση <http://195.251.193.213/mod/forum/view.php?id=122> υποδεχόμαστε σχόλια, παρατηρήσεις και διορθώσεις, με την προοπτική η επόμενη έκδοση να είναι βελτιωμένη, αλλά αναρτούμε και επικαιροποιημένες οδηγίες για τις ενότητες που παρουσιάζουμε στο βιβλίο.

Είναι σημαντικό για εμάς να οικοδομήσουμε μια σχέση συνεργασίας...

Οι συγγραφείς

Αναστάσιος Εμβαλωτής, Αναπληρωτής Καθηγητής

Πανεπιστημίου Ιωαννίνων [aemvalot@uoi.gr]

Αικατερίνη Σαργιώτη, ερευνήτρια στο National Anti-Bullying Research and Resource Centre του Dublin City University [katsargioti@gmail.com]