

Μεθοδολογία Έρευνας και Εισαγωγή στη Στατιστική Ανάλυση Δεδομένων με το IBM SPSS STATISTICS

ΜΙΛΤΙΑΔΗΣ ΧΑΛΙΚΙΑΣ
ΑΛΕΞΑΝΔΡΑ ΜΑΝΩΛΕΣΣΟΥ
ΠΑΝΑΓΙΩΤΑ ΛΑΛΟΥ



Ελληνικά Ακαδημαϊκά Ηλεκτρονικά
Συγγράμματα και Βοηθήματα
www.kallipos.gr

HEALLINK
Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ



ΕΣΠΑ
2007-2013
Ευρωπαϊκό Κοινωνικό Ταμείο

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΜΙΛΤΙΑΔΗΣ ΧΑΛΙΚΙΑΣ
Αναπληρωτής Καθηγητής
ΑΕΙ Πειραιά ΤΤ

ΑΛΕΞΑΝΔΡΑ ΜΑΝΩΛΕΣΟΥ
Msc Biostatistics

ΠΑΝΑΓΙΩΤΑ ΛΑΛΟΥ
Phd Mathematics

***Μεθοδολογία Έρευνας και Εισαγωγή
στη Στατιστική Ανάλυση Δεδομένων
με το IBM SPSS STATISTICS***



Ελληνικά Ακαδημαϊκά Ηλεκτρονικά
Συγγράμματα και Βοηθήματα
www.kallipos.gr

Μεθοδολογία Έρευνας και Εισαγωγή στη Στατιστική Ανάλυση Δεδομένων με το IBM SPSS STATISTICS

Συγγραφή

Μιλτιάδης Χαλικιάς

Αλεξάνδρα Μανωλέσου

Παναγιώτα Λάλου

Κριτικός αναγνώστης

Χρήστος Κίτσος

Ομότιμος καθηγητής Τεχνολογικού Εκπαιδευτικού Ιδρύματος Αθηνών

Συντελεστές έκδοσης

Γλωσσική Επιμέλεια: Μέμα Χαραλάμπους

Τεχνική Επεξεργασία: Ιωάννης Λιόμας

ISBN: 978-960-603-123-6

Copyright © ΣΕΑΒ, 2015



Το παρόν έργο αδειοδοτείται υπό τους όρους της άδειας Creative Commons Αναφορά Δημιουργού - Μη Εμπορική Χρήση - Όχι Παράγωγα Έργα 3.0. Για να δείτε ένα αντίγραφο της άδειας αυτής επισκεφτείτε τον ιστότοπο <https://creativecommons.org/licenses/by-nc-nd/3.0/gr/>

ΣΥΝΔΕΣΜΟΣ ΕΛΛΗΝΙΚΩΝ ΑΚΑΔΗΜΑΪΚΩΝ ΒΙΒΛΙΟΘΗΚΩΝ

Εθνικό Μετσόβιο Πολυτεχνείο

Ηρώων Πολυτεχνείου 9, 15780 Ζωγράφου

www.kallipos.gr

Πίνακας περιεχομένων

Πίνακας περιεχομένων.....	5
Πίνακας συντομεύσεων-ακρωνύμια	14
Πρόλογος.....	15
Εισαγωγή	16
Κεφάλαιο 1.....	18
Σύνοψη.....	18
Προσπαιτούμενη γνώση	18
1.1 Εισαγωγή	18
1.2 Επιστημολογία και φιλοσοφικά ρεύματα	19
1.2.1 Εισαγωγή	19
1.2.2. Βασικές έννοιες.....	19
1.2.3. Σκοπός και γένεση των επιστημών.....	20
1.2.4. Τα πρώτα βήματα της φιλοσοφίας και των επιστημών.....	21
1.2.5. Ερωτήματα της γνώσης-φιλοσοφικά ρεύματα και μεθοδολογικές προτάσεις.....	23
1.2.6. Επιστημάνσεις - συμπεράσματα	25
1.3. Σκοποί επιστημονικής έρευνας.....	25
1.3.1 Διερεύνηση	25
1.3.2 Περιγραφή	26
1.3.3 Ερμηνεία και αναζήτηση αιτιότητας	26
Αιτιότητα.....	26
Συνθήκες εξαγωγής αποτελέσματος.....	27
1.4 Η Λογική της έρευνας (Παραγωγική-Επαγωγική)	27
1.5 Μονάδες Ανάλυσης	27
1.6 Η χρονική διάσταση των ερευνών	28
1.7 Δειγματοληψία	29
1.7.1 Βασικές έννοιες.....	29
Στατιστικός πληθυσμός.....	29
Μεταβλητές	30
1.7.2 Καθορισμός μεγέθους δείγματος	31
Μέτρα θέσεως (περίπτωση πρωτογενών δεδομένων).....	31
Μέτρα Διασποράς (πρωτογενών δεδομένων)	32
Κριτήριο καθορισμού μεγέθους δείγματος με βάση το τυπικό σφάλμα.....	33
Κριτήριο καθορισμού μεγέθους δείγματος με βάση τον συντελεστή παραλλακτικότητας	34
Κριτήριο καθορισμού μεγέθους δείγματος με βάση το διάστημα εμπιστοσύνης.....	34
Παράδειγμα 1	36

Παράδειγμα 2	37
1.7.3 Τα είδη δειγματοληψίας.....	39
1.7.3.1 Μη πιθανοτική δειγματοληψία.....	39
1.7.3.2 Πιθανοτική δειγματοληψία.....	42
1.8 Τα είδη των ερευνών.....	43
1.8.1 Έρευνα πεδίου.....	43
1.8.1.1 Παραδείγματα κατάλληλα για έρευνα πεδίου	43
1.8.1.2 Οι διάφοροι ρόλοι του ερευνητή.....	44
1.8.1.3 Συνεντεύξεις	44
1.8.1.4 Ομάδες εστίασης.....	45
1.8.1.5 Δεοντολογία ποιοτικής έρευνας πεδίου	46
1.8.2 Δειγματοληπτική έρευνα	46
1.8.2.1 Ερωτηματολόγια.....	46
1.8.2.2. Πλεονεκτήματα δειγματοληπτικών ερευνών.....	47
1.8.2.3 Μειονεκτήματα δειγματοληπτικών ερευνών.....	47
1.8.2.4 Δευτερογενής έρευνα.....	47
1.8.2.5 Δεοντολογία της δειγματοληπτικής έρευνας.....	48
1.8.3 Πειραματικοί σχεδιασμοί.....	48
1.8.3.1 Θέματα κατάλληλα για πειράματα.....	48
1.8.3.2 Η διαδικασία του πειράματος.....	48
1.8.3.3 Προ-έλεγχος και μετα-έλεγχος	49
1.8.3.4 Το κλασικό πείραμα.....	49
1.8.3.5 «Ταίριασμα» και «Τυχαιοποίηση»	50
1.8.3.6 Εγκυρότητα	50
1.8.3.7 Πηγές έλλειψης εσωτερικής εγκυρότητας.....	51
1.8.3.8 Πλεονεκτήματα και μειονεκτήματα	51
1.8.3.9 Δεοντολογία και πειράματα.....	51
1.8.4 Μη αντιδραστικές μέθοδοι ανάλυσης	52
1.8.4.1 Πλεονεκτήματα και μειονεκτήματα	52
1.8.4.2 Δεοντολογία μη αντιδραστικών ερευνών	53
Βιβλιογραφικές Αναφορές Κεφαλαίου.....	54
Ελληνόγλωσσες	54
Ξενόγλωσσες.....	54
Κεφάλαιο 2.....	56
Σύνοψη.....	56
Προσπαιτούμενη γνώση	56
2.1. Εισαγωγή.....	56
2.2.1 Ερωτηματολόγιο	56

2.2.2 Δείγματα ερωτηματολογίων.....	58
Ερωτηματολόγιο 1.....	59
Ερωτηματολόγιο 2.....	65
Ερωτηματολόγιο Υφισταμένων.....	65
Στοιχεία Ερωτηθέντων.....	69
Ερωτηματολόγιο Προϊσταμένων.....	71
Στοιχεία Ερωτηθέντων.....	73
2.2.3 Ερωτηματολόγια ατομικής συμπλήρωσης.....	75
2.2.4 Ταχυδρόμηση και επιστροφή.....	76
2.2.5 Παρακολούθηση των επιστροφών.....	76
2.2.6 Επαναληπτικές αποστολές.....	76
2.2.7 Ποσοστό απόκρισης.....	76
2.2.8 Δειγματοληπτικές έρευνες με χρήση συνεντεύξεων.....	76
2.2.9 Συντονισμός και έλεγχος.....	77
2.2.10 Τηλεφωνικές έρευνες.....	77
2.2.11 Ποσοστά απόκρισης σε έρευνες μέσω συνεντεύξεων.....	78
2.2.12 Διαδικτυακές έρευνες.....	78
2.2.13 Δομή δειγματοληπτικής έρευνας.....	78
2.3 Μελέτη περίπτωσης: Κατασκευή ερωτηματολογίου δειγματοληπτικής έρευνας.....	79
2.3.1 Σκοπός και στόχοι μελέτης.....	79
Θεωρητικοί στόχοι.....	79
Ερευνητικοί στόχοι.....	80
2.3.2 Διατύπωση ερευνητικών ερωτημάτων.....	80
1ος Στόχος (Θεωρητικός).....	80
2ος Στόχος (Θεωρητικός).....	80
1ος Στόχος (Ερευνητικός).....	80
2ος Στόχος (Ερευνητικός).....	80
3ος Στόχος (Ερευνητικός).....	81
2.3.3 Διατύπωση ερευνητικών υποθέσεων.....	81
Συμπεράσματα.....	82
Ερωτηματολόγιο 3.....	82
Πρώτη Θεματική Ενότητα.....	82
Δεύτερη Θεματική Ενότητα.....	83
Τρίτη Θεματική Ενότητα.....	86
Προαιρετικές Πληροφορίες.....	88
Παρατήρηση.....	89
2.4 Δημοσκοπήσεις.....	89
2.4.1 Εισαγωγικά.....	89

2.4.2 Οι δημοσκοπήσεις σήμερα	90
2.4.3 Σχετικές έρευνες.....	90
2.4.4 Τρόποι διεξαγωγής των δημοσκοπήσεων	91
2.4.4.1 Exit-Polls.....	91
2.4.4.2 Δημοσκοπήσεις μέσω SMS.....	92
2.4.5 Αξιοπιστία εταιρειών δημοσκοπήσεων.....	92
Βιβλιογραφικές Αναφορές Κεφαλαίου.....	93
Ελληνόγλωσσες	93
Ξενόγλωσσες.....	93
Κεφάλαιο 3.....	95
Σύνοψη.....	95
Προσπαιτούμενη γνώση	95
3.1 Το Περιβάλλον του SPSS.....	95
Data View	96
Variable View.....	97
3.2 Μενού File.....	99
3.2.1 Καταχώριση δεδομένων στο IBM SPSS	99
Χειρόγραφο Καταγραφή Δεδομένων	99
Μεταφορά Δεδομένων από Αρχείο Excel.....	100
File > Open > Data.....	100
File > Save As	102
3.2.2 Κωδικοποίηση Δεδομένων	102
Ερωτηματολόγια.....	102
Μεταβλητές	102
Παρατηρήσεις	102
3.3 Μενού Edit.....	103
3.4 Μενού View	103
3.5 Μενού Data.....	103
3.5.1 Sort Cases	103
3.5.2 Sort Variables.....	104
3.5.3 Transpose.....	105
3.5.4 Select Cases.....	105
3.5.5 Merge Files	108
Merge > Add Cases	108
Merge > Add Variables	109
3.5.6 Split File.....	112
3.5.7 Weight Cases	112
3.5.8 Restructure.....	113

3.6 Μενού Transform	121
3.6.1 Recode	121
Transform > Recode > Into Same Variable	121
Old Values	122
New Values	122
Transform > Recode > Into Different Variable	123
3.6.2 Compute	124
3.6.3 Replace Missing Values	125
Μέθοδοι Διαχείρισης Ελλειπουσών Τιμών	126
3.6.4 Random Number Generator	127
3.7 Μενού Analyze	127
Βιβλιογραφικές Αναφορές Κεφαλαίου	128
Ελληνόγλωσσες	128
Ξενόγλωσσες	128
Κεφάλαιο 4	129
Σύνοψη	129
Προσπαιτούμενη γνώση	129
4.1 Βασικές Έννοιες και Ορισμοί	129
Περιγραφική στατιστική (Descriptive Statistics)	129
Βασικές Στατιστικές Έννοιες	130
Στατιστικός πληθυσμός (population)	130
Στατιστική μεταβλητή (Variable)	130
Είδη Μεταβλητών	130
4.2 Frequencies	131
Ποιοτικές μεταβλητές	131
Συνεχείς μεταβλητές	133
Percentile Values	134
Quartiles - Τεταρτημόρια	134
Percentiles - Ποσοστημόρια	135
Central Tendency	135
Mean - Αριθμητικός μέσος ή Μέση τιμή	135
Median - Διάμεσος	135
Mode - Επικρατούσα τιμή	135
Sum - Άθροισμα	135
Dispersion	136
Std. Deviation - Τυπική απόκλιση	136
Variance - Διακύμανση	136
Range - Έυρος	136

S.E. Mean - Τυπικό σφάλμα.....	136
Distribution	136
Skewness - Συντελεστής Ασυμμετρίας.....	137
Kurtosis - Συντελεστής Κύρτωσης.....	137
4.3 Descriptives	139
4.4 Explore.....	140
Trimmed Mean - Ξακρισμένη μέση τιμή.....	142
Confidence Interval for Mean - Διάστημα Εμπιστοσύνης για τον μέσο	142
4.5 Crosstabs	144
Count.....	145
Percentages.....	146
Residuals.....	146
4.6 Correlate.....	147
Correlation Coefficients	148
Test of Significance	149
Παραμετρικός έλεγχος.....	149
Μη Παραμετρικός έλεγχος.....	150
4.7 Multiple Response.....	152
4.7.1 Analyze > Multiple Response > Define Multiple Response Sets	152
4.7.2 Analyze > Multiple Response > Frequencies	153
4.7.3 Analyze > Multiple Response > CrossTabs	154
4.8 Graphs	155
4.9 Ραβδογράμματα - Bar Charts.....	156
Simple	157
Clustered.....	159
4.10 Γραμμικά - Line	159
Simple	160
Multiple.....	162
4.11 Γράφημα Περιοχής - Area.....	163
Simple	163
4. 12 Κυκλικά (πιτογράμματα) - Pie.....	165
4.13 High - Low	167
Simple	168
Clustered.....	169
4.14 Θηκογράμματα ή Πλαισίου Απολήξεων BoxPlots	170
Simple	171
Clustered.....	172
4.15 Διασποράς - Scatter/ Dot	173

Simple	174
Matrix Scatter	175
4.16 Ιστογράμματα - Histogram	176
Βιβλιογραφικές Αναφορές Κεφαλαίου	178
Ελληνόγλωσσες	Error! Bookmark not defined.
Ξενόγλωσσες.....	Error! Bookmark not defined.
Κεφάλαιο 5.....	179
Σύνοψη.....	179
Προσπαιτούμενη γνώση	179
5.1 Εισαγωγή	179
5.2 Βασικές Έννοιες και Ορισμοί	179
5.3 Γραφικοί Έλεγχοι Κανονικότητας	181
5.3.1 Ιστογράμματα.....	181
5.3.2 P-P Plots.....	181
5.3.3 Q-Q Plots	183
5.4 Στατιστικοί Έλεγχοι Κανονικότητας.....	184
5.4.1 One-Sample Kolmogorov-Smirnov (K-S) Test.....	184
Υποθέσεις του κριτηρίου:	184
5.4.2 Shapiro-Wilk Test.....	186
Βιβλιογραφικές Αναφορές Κεφαλαίου	Error! Bookmark not defined.
Ελληνόγλωσσες	Error! Bookmark not defined.
Ξενόγλωσσες.....	Error! Bookmark not defined.
Κεφάλαιο 6.....	189
Σύνοψη.....	189
Προσπαιτούμενη γνώση	189
6.1 T-Test για τον έλεγχο Μέσης Τιμής Πληθυσμού.....	189
Μορφές της ελεγχουσυνάρτησης	190
6.2 T-test για τον έλεγχο της διαφοράς δύο μέσων από ανεξάρτητα δείγματα	191
6.3 Mann-Whitney U	193
6.4 T-test για τον έλεγχο της διαφοράς δύο μέσων από συσχετιζόμενα δείγματα.....	196
6.5 Wilcoxon	198
6.6 Means.....	200
6.7 One-Way ANOVA για ανεξάρτητα δείγματα.....	201
Υποθέσεις.....	201
Ανάλυση της μεθόδου	202
Tukey Test	206
Σημείωση	206
6.8 Kruskal-Wallis H.....	209

6.9 Two-Way ANOVA for Independent Samples	212
Ανάλυση της μεθόδου	212
Υποθέσεις.....	212
Παράδειγμα	213
6.10 Friedman	219
6.11 X²-test.....	222
Λόγος Συμπληρωματικών Πιθανοτήτων	225
Ερμηνεία.....	226
6.12 Mc Nemar test	227
6.13 Παράρτημα: Crosstabs > Cell.....	229
Συχνότητες (Counts).....	229
Ποσοστά (Percentages).....	229
Υπόλοιπα (Residuals).....	230
6.14 Παράρτημα: Crosstabs > Statistics	231
Βιβλιογραφικές Αναφορές Κεφαλαίου.....	234
Ελληνόγλωσσες	234
Ξενόγλωσσες.....	234
Κεφάλαιο 7.....	235
Σύνοψη.....	235
Προσπαιτούμενη γνώση	235
7.1 Απλή και Πολλαπλή Γραμμική Παλινδρόμηση (Linear Regression).....	235
Πολλαπλή γραμμική παλινδρόμηση.....	240
F-έλεγχος για τη σημαντικότητα όλου του μοντέλου (overall F-test).....	242
Partial F-tests (variables-added-in-order ή Type I F-tests)	243
Variables-added-last ή Type III F-tests	244
Υποθέσεις του πολλαπλού γραμμικού μοντέλου.....	244
Βήματα για την επιλογή του «βέλτιστου» μοντέλου παλινδρόμησης	246
Σημείωση.....	248
Εφαρμογή παραδείγματος στο SPSS.....	257
Regression Equation.....	258
Έλεγχος Υπολοίπων.....	259
Γραφικές Παραστάσεις Πρόσθετων Μεταβλητών και Μερικών Υπολοίπων	262
Διάγραμμα Πρόσθετων Μεταβλητών	262
Διάγραμμα Μερικών Υπολοίπων.....	262
7.2 Λογιστική Παλινδρόμηση (Logistic Regression).....	263
Σημεία προσοχής.....	267
ROC Curve.....	272
Βιβλιογραφικές Αναφορές Κεφαλαίου.....	275

Ελληνόγλωσσες	275
Ξενόγλωσσες.....	275
ΓΛΩΣΣΑΡΙ.....	276

Πίνακας συντομεύσεων-ακρωνύμια

ANOVA	Analysis of Variance
APA	American Psychological Association
FPR	False Positive Rate
IBM	International Business Machines Corporation
IEEE	Institute of Electrical and Electronics Engineers
ISBN	International Standard Book Number
MLA	Modern Language Association
NPV	Negative Predictive Value
OR	Odds Ratio
PDF	Portable Document Format
PPV	Positive Predictive Value
ROC	Receiver Operating Characteristic
SPSS	Statistical Package for the Social Sciences
TPR	True Positive Rate
VIF	Variance Inflation Factor
XHTML	eXtensible HyperText Markup Language
XML	EXtensible Markup Language

Πρόλογος

Το βιβλίο αυτό χρηματοδοτήθηκε από το πρόγραμμα Κάλλιπος και προορίζεται για φοιτητές ελληνικών ΑΕΙ. Μπορεί να φανεί χρήσιμο, όμως, σε όποιον επιθυμεί να πραγματοποιήσει σωστά μια έρευνα και σε όποιον επιθυμεί να αναλύσει δεδομένα έρευνας με το IBM SPSS STATISTICS.

Τα πρώτα δύο κεφάλαια προέρχονται από υλικό που είχα χρησιμοποιήσει κατά τη διδασκαλία του μαθήματος «Μεθοδολογία Έρευνας για Διοικητικά Στελέχη» για τα έτη 2013-2014 και 2014-2015 στο τμήμα Διοίκησης Επιχειρήσεων του ΑΕΙ Πειραιά ΤΤ. Τα υπόλοιπα κεφάλαια (3-7) αποτελούν μία εισαγωγή στο στατιστικό πακέτο SPSS που είναι το πιο διαδεδομένο στατιστικό πακέτο στην Ελλάδα.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω τους αποφοίτους του τμήματος Διοίκησης Επιχειρήσεων Γιάννη Πέρο και Γιάννη Φλουρή για τη συμβολή τους στην παράγραφο των πειραματικών σχεδιασμών και στην παράγραφο των δημοσκοπήσεων αντίστοιχα. Ακόμα θα ήθελα να ευχαριστήσω τις αποφοίτους του μεταπτυχιακού προγράμματος «Διεθνής Διοικητική των Επιχειρήσεων», Ευαγγελία Τριανταφύλλου, Ιωάννα Νταβή και Γεωργία Τσιτμηδέλη για την επισύναψη των ερωτηματολογίων των διπλωματικών τους εργασιών.

Οποιαδήποτε παρατήρηση για ενδεχόμενα λάθη και παραλείψεις είναι καλοδεχούμενη και θα ληφθεί υπόψη σε μελλοντική επανέκδοση.

M. Χαλικιάς

Εισαγωγή

Στόχος του βιβλίου είναι να γίνει χρήσιμο σε αναγνώστες που δεν έχουν ιδιαίτερες γνώσεις στατιστικής ή φιλοσοφίας της επιστήμης. Γι' αυτό και όπου υπάρχει ανάγκη γίνεται μια θεωρητική αναφορά σε στατιστικούς όρους.

Στο πρώτο κεφάλαιο επιχειρείται μια σύντομη αναφορά στα είδη της επιστημονικής έρευνας. Για την πληρέστερη κατανόηση του θέματος κρίθηκε απαραίτητο πρώτα να δοθούν κάποια στοιχεία για την επιστημολογία (τη γένεση και την εξέλιξή της). Ακολούθως καθορίζονται οι σκοποί της επιστημονικής έρευνας, διευκρινίζονται οι απλές στατιστικές έννοιες (πληθυσμός, δείγμα, μεταβλητή, κ.ά.), καθώς και τα τεχνικά θέματα της δειγματοληψίας και της μεθοδολογίας καθορισμού μεγέθους δείγματος. Διεξοδικά εξετάζεται τόσο ο τρόπος διεξαγωγής της έρευνας, όσο και τα πλεονεκτήματα και μειονεκτήματα της έρευνας πεδίου, των πειραματικών σχεδιασμών και των μη αντιδραστικών ερευνών.

Στο δεύτερο κεφάλαιο εξετάζονται τεχνικά ζητήματα της δειγματοληπτικής έρευνας, η οποία θεωρείται ως η πιο διαδεδομένη ερευνητική μέθοδος. Συγκεκριμένα ελέγχεται η εγκυρότητα και η αξιοπιστία ερωτηματολογίου, η παρουσίαση και ανάλυση αποτελεσμάτων, καθώς και η σύγκριση των ευρημάτων με άλλες μεθόδους έρευνας. Τέλος καταγράφονται τα πλεονεκτήματα και μειονεκτήματα της δειγματοληπτικής έρευνας.

Στο τρίτο κεφάλαιο γίνεται λόγος για την καταχώρηση δεδομένων στο IBM SPSS STATISTICS καθώς και για τις βασικές εντολές διαχείρισης των δεδομένων από το SPSS. Αρχικά γίνεται αναφορά στο περιβάλλον του SPSS και συγκεκριμένα στις επιλογές Run the Tutorial, Type in data, Run an existing query, Create new query using Database Wizard και στα φύλλα *Data View* και *Variable View*. Ακόμα γίνεται αναφορά στην καταχώρηση δεδομένων στο IBM SPSS (χειρόγραφη καταγραφή δεδομένων και μεταφορά δεδομένων από αρχείο Excel), και στην κωδικοποίηση δεδομένων ερωτηματολογίων. Από το μενού *Edit View Data* αναλύονται οι εντολές *Sort Variables*, *Transpose*, *Select Cases*, *Merge Files* (ενοποίηση δεδομένων από διαφορετικά αρχεία), *Split File* (διαχωρισμό των παρατηρήσεων του αρχείου σε μικρότερες εικονικές διαμερίσεις), *Weight Cases* (στάθμιση σύμφωνα με τις τιμές μιας άλλης μεταβλητής), *Restructure* (αναδιοργάνωση του αρχείου μετατρέποντας τις μεταβλητές σε παρατηρήσεις και αντίστροφα) *Recode* (επανακωδικοποίηση εφαρμόζεται για τη δημιουργία νέας μεταβλητής), *Compute* (δημιουργία νέων μεταβλητών υπό την εφαρμογή μιας μαθηματικής έκφρασης), *Replace Missing Values* (αντικατάσταση των ελλειπουσών τιμών).

Στο τέταρτο κεφάλαιο γίνεται αναφορά στον υπολογισμό μέτρων και γραφημάτων με χρήση του στατιστικού πακέτου. Αρχικά περιέχονται βασικές έννοιες και ορισμοί για την περιγραφική στατιστική, τις έννοιες του πληθυσμού και δείγματος και τα είδη των μεταβλητών. Γίνεται αναφορά στον υπολογισμό περιγραφικών στατιστικών στο SPSS μέσω των διαδρομών *Descriptive Statistics > Frequencies*, *Descriptive Statistics > Descriptives*, *Descriptive Statistics > Explore*, *Descriptive Statistics > Crosstabs*. Παράλληλα εξηγούνται οι στατιστικές έννοιες όπου χρειάζεται.

Γίνεται αναφορά στην έννοια της συσχέτισης, στους συντελεστές συσχέτισης και στον τρόπο υπολογισμού τους (μέσω της επιλογής *Correlate*). Γίνεται αναλυτική περιγραφή της εντολής *Multiple Response (Analyze > Multiple Response > Define Multiple Response Sets)*. Τέλος γίνεται αναφορά στα διαγράμματα μέσω του μενού *Graphs*. Συγκεκριμένα περιγράφονται τα Ραβδογράμματα, τα Γραμμικά, τα Γράφηματα Περιοχής, τα Κυκλικά, Τα Υψηλών-Χαμηλών τιμών (High-Low), τα Θηκογράμματα ή Πλαισίου Απολήξεων, τα Διασποράς και τα Ιστογράμματα.

Στο πέμπτο κεφάλαιο περιέχονται βασικές έννοιες και ορισμοί του ελέγχου υποθέσεων, γραφικοί έλεγχοι κανονικότητας μέσω των ιστογραμμάτων και των P-P plots. Στατιστικοί έλεγχοι κανονικότητας One-Sample Kolmogorov-Smirnov (K-S) Test και Shapiro-Wilk Test.

Στο έκτο κεφάλαιο παρουσιάζονται οι γνωστότεροι παραμετρικοί και μη παραμετρικοί έλεγχοι. Συγκεκριμένα οι έλεγχοι: T-Test για τον έλεγχο Μέσης Τιμής Πληθυσμού, T-test για τον έλεγχο της διαφοράς δύο μέσων από ανεξάρτητα δείγματα, Mann-Whitney U, T-test για τον έλεγχο της διαφοράς δύο μέσων από συσχετιζόμενα δείγματα, Wilcoxon, Means, One-Way ANOVA για ανεξάρτητα δείγματα, Kruskal-Wallis, Two-Way ANOVA για ανεξάρτητα δείγματα, Friedman, X²-test, Mc Nemar test.

Τέλος, στο τελευταίο κεφάλαιο γίνεται εκτενής αναφορά στην απλή και πολλαπλή γραμμική παλινδρόμηση, και περιληπτική αναφορά στη λογιστική παλινδρόμηση. Σε κάθε περίπτωση παρέχονται απλές εφαρμογές και παραδείγματα.

Συνοψίζοντας θα λέγαμε ότι το βιβλίο συνδυάζει την παρουσίαση μεθόδων έρευνας, με έμφαση στις δειγματοληπτικές έρευνες, με την παρουσίαση σχετικά απλής στατιστικής ανάλυσης δεδομένων στο SPSS. Λόγω του σκοπού του βιβλίου καθώς και της περιορισμένης έκτασής του δεν περιέχονται άλλες μορφές γενικευμένων γραμμικών μοντέλων καθώς και μέθοδοι πολυμεταβλητής στατιστικής ανάλυσης (π.χ. factor, cluster, discriminant) που μπορούν να είναι χρήσιμα εργαλεία σε όλα τα είδη ερευνών. Το κενό αυτό ευελπιστούμε ότι θα καλυφθεί σε μελλοντικό σύγγραμμα.

Οι συγγραφείς

Μ. Χαλικιάς, Α. Μανωλέσου, Π. Λάλου

Κεφάλαιο 1

Τα είδη έρευνας

Σύνοψη

Στο κεφάλαιο αυτό επιχειρείται μια σύντομη αναφορά στα είδη της επιστημονικής έρευνας. Για την πληρέστερη κατανόηση του θέματος κρίθηκε απαραίτητο πρώτα να δοθούν κάποια στοιχεία για την επιστημολογία (τη γένεση και την εξέλιξή της). Ακολούθως καθορίζονται οι σκοποί της επιστημονικής έρευνας, διευκρινίζονται οι απλές στατιστικές έννοιες (πληθυσμός, δείγμα, μεταβλητή κ.ά.), τα τεχνικά θέματα της δειγματοληψίας και η μεθοδολογία καθορισμού μεγέθους δείγματος. Διεξοδικά εξετάζεται τόσο ο τρόπος διεξαγωγής της έρευνας, όσο και τα πλεονεκτήματα και μειονεκτήματα της έρευνας πεδίου, των πειραματικών σχεδιασμών και των μη αντιδραστικών ερευνών. Στις δειγματοληπτικές έρευνες θα προσπαθήσουμε να εμβαθύνουμε περισσότερο στο δεύτερο κεφάλαιο.

Προαπαιτούμενη γνώση

Ως προαπαιτούμενη γνώση θεωρείται μόνο η δυνατότητα χρήσης των στατιστικών πινάκων της κανονικής κατανομής και της κατανομής Student για τον καθορισμό μεγέθους δείγματος. Καλύτερη κατανόηση του κεφαλαίου όμως είναι δυνατόν να επιτευχθεί εάν ο αναγνώστης έχει γνωστικό υπόβαθρο επιστημολογίας, προκειμένου να κατανοήσει καλά τα φιλοσοφικά ρεύματα έρευνας ώστε να αντιληφθεί ουσιαστικά και όχι μηχανιστικά τις επιστημονικές μεθόδους που θα κληθεί να εφαρμόσει στη μελλοντική επαγγελματική ενασχόλησή του. Η ικανή εξοικείωσή του επίσης με απλές στατιστικές έννοιες θα του επιτρέψουν να γνωρίσει καλύτερα τον τρόπο υπολογισμού του μεγέθους ενός δείγματος. Σημαντικά και ενδιαφέροντα στοιχεία στα θέματα αυτά μπορεί κάποιος να αναζητήσει στα βιβλία των Στ. Θεοφανίδη, Σ. Θεοφανίδη & Γ. Μπένου, Χ. Φράγγου, Ε. Babbie, J. Mason, M. Grawitz, J. W. Creswell, Y. C. Gagnon. (βλ. βιβλιογραφικές αναφορές).

1.1 Εισαγωγή

Έχοντας ήδη ορίσει ως επιθυμητή την επαφή με κάποια επιστημολογικά προβλήματα, θα επιχειρήσουμε μια προσέγγιση της γένεσης και της εξέλιξης του γνωστικού φαινομένου, θα αναφερθούμε στον τρόπο με τον οποίο αναπτύχθηκε ιστορικά ο έλεγχος της ορθότητας της γνώσης και ποια είναι τα όρια του κύρους της. Στο επόμενο βήμα, θα εξετάσουμε τις μεθόδους που χρησιμοποιήθηκαν για την επιστημονική τεκμηρίωση της γνώσης στις διάφορες επιστήμες και θα έρθουμε σε μια πρώτη γνωριμία με τη μεθοδολογία και τα προβλήματά της.

Ο καθορισμός των σκοπών της έρευνας θα είναι το άλλο θέμα που θα μας απασχολήσει. Η χρήση της επιστημονικής έρευνας γενικά, αλλά και ειδικότερα στη στατιστική γίνεται πιο εύκολη και αποτελεσματική, εάν κάποιος ξέρει τα αίτια δημιουργίας του φαινομένου που εξετάζει, τί ανάγκες καλύπτει και πως αυτό λειτουργεί σε βάθος. Οι γνώσεις αυτές θα του επιτρέψουν να προσδιορίσει ακριβέστερα τί θέλει να ανακαλύψει, αλλά και να σχεδιάσει ορθότερα την έρευνα που καλείται να διεξάγει, αφού η επιστημονική έρευνα δεν είναι τίποτα παραπάνω από παρατήρηση και ερμηνεία των ορθών στοιχείων που απαιτούνται για να υλοποιηθεί ένα οποιοδήποτε σχέδιο. Για αυτό πρέπει κάθε φορά να συγκεκριμενοποιείται επακριβώς, τί πρόκειται να παρατηρηθεί και να αναλυθεί, το γιατί και το πώς.

Ο σχεδιασμός έρευνας περιλαμβάνει τις εξής βασικές παραμέτρους.

- Βασικοί σκοποί της έρευνας,
- Λογική της έρευνας,
- Μονάδες ανάλυσης (τί ή ποιος) θα μελετηθεί,
- Εναλλακτικοί τρόποι διαχείρισης του χρόνου στην έρευνα (συγχρονικές και διαχρονικές μελέτες).

1.2 Επιστημολογία και φιλοσοφικά ρεύματα

1.2.1 Εισαγωγή

Σήμερα είναι γενικά αποδεκτό, ότι η επιστημονική έρευνα στον κάθε τομέα είναι αποτέλεσμα παρατηρήσεων.

Ποιο όμως είναι το αντικείμενο μιας επιστήμης, πότε πρωτοεμφανίστηκαν, τί ανάγκες κάλυπταν, γιατί λέμε ότι οι επιστήμες γεννήθηκαν στην Ελλάδα, αφού αξιόλογα τεχνικά επιτεύγματα κατά πολύ παλαιότερα υπάρχουν σε πολλές χώρες (Αίγυπτο, Βαβυλωνία κ.λπ.);

Για τη διαλεύκανση αυτών των ερωτημάτων κρίνεται απαραίτητη μια σύντομη ιστορική αναδρομή στο επιστημονικό φαινόμενο καθώς και στη σχέση του με τη φιλοσοφία. Αφού οριοθετηθεί το ζήτημα σαφώς, θα επιχειρήσουμε μία σύντομη αναφορά των μεθόδων που η επιστημολογία χρησιμοποιεί με έμφαση τη δειγματοληπτική έρευνα, η οποία έχει πολλές εφαρμογές στις επιστήμες διοίκησης και οικονομίας καθώς και στις κοινωνικές επιστήμες.

Αναμφίβολα θεωρείται πλέον κοινός τόπος ότι ο κόσμος και η κοινωνία είναι ένα γοητευτικό μεν σύστημα, ιδιαίτερα όμως περίπλοκο, αντιφατικό και δυσκολονόητο, στο οποίο συντελούνται άπειρες διεργασίες. Η κατάκτηση της γνώσης, όπως είναι φυσικό, έγινε σταδιακά, μέσα από πολλές πλάνες και λαθεμένες αντιλήψεις. Από όλους αναγνωρίζεται πλέον ότι υπάρχουν πάρα πολλά ζητήματα που τα αγνοούμε ακόμη, και ότι σ' αυτό το γεγονός οφείλεται και η ποικιλία των φιλοσοφικών και ερευνητικών ρευμάτων στα οποία θα αναφερθούμε πιο κάτω.

Η επιστήμη είναι το εργαλείο που δημιούργησε ο άνθρωπος στην προσπάθειά του να κατανοήσει συγκεκριμένους τομείς της πραγματικότητας γύρω του, ώστε να προβλέπει γεγονότα που μπορεί να συμβούν, με σκοπό να τα ελέγχει, ή και να τα αλλάζει προς όφελός του, αν είναι εφικτό σε συγκεκριμένους υλικούς τομείς (Θεοδωρίδης 1955). Εύκολα γίνεται αντιληπτό ότι ο σκοπός της είναι συγγενικός με αυτόν της φιλοσοφίας, που όμως έχει στόχο να γνωρίσει και να οργανώσει τον κόσμο στο σύνολό του και όχι μόνο σε επί μέρους ζητήματα. Άμεσα προκύπτουν λοιπόν τα επόμενα ερωτήματα: πότε πρωτοεμφανίζονται η φιλοσοφία και η επιστήμη; ποια η σχέση των δύο αυτών γνωστικών αντικειμένων και ποια τα όριά τους;

Η συστηματική εξέταση φιλοσοφικών θεμάτων, δηλαδή η τάση του ανθρώπου να κατακτήσει την «καθολική γνώση» στηριζόμενος στη λογική εμφανίζεται πρώτα ιστορικά. Λίγο αργότερα, όμως, αρχίζει να γίνεται αντιληπτή η δυσκολία του όλου εγχειρήματος, οπότε σταδιακά αρχίζει να αυτονομείται από τη φιλοσοφία η κάθε επιστήμη, ακολουθώντας θα έλεγε κανείς μία παράλληλη πορεία με αυτήν του γνωσιολογικού προβλήματος (που είναι ένα από τα ζητήματα που εξετάζει η φιλοσοφία). Η γνωσιολογία αναπτύχθηκε αργότερα από τα άλλα δύο φιλοσοφικά πεδία (οντολογικό και ηθικό), όπως αποδεικνύεται ιστορικά, αν και είναι, φυσικά, το πιο καθοριστικό, αφού στην περίπτωση που ο άνθρωπος δεν μπορεί να αποκτήσει έγκυρη γνώση, δύσκολα μπορεί να αντιληφθεί κάποιος πώς νομιμοποιείται να συζητάει άλλα φιλοσοφικά θέματα. Από το σημαντικό αυτό πεδίο της γνωσιολογίας αυτονομήθηκε βαθμιαία η επιστημολογία, δηλ. η επιστήμη που μελετά την εξέλιξη των επιστημών, τον τρόπο που αυτές λειτουργούν και τα θεμέλια στα οποία στηρίζουν την ανάπτυξή τους, δηλ. τη μεθοδολογία που ακολουθούν για την τεκμηρίωση των εκάστοτε συγκεκριμένων καταστάσεων που εξετάζουν.

1.2.2. Βασικές έννοιες

Η επιστημολογία πραγματεύεται προβλήματα της επιστημονικής γνώσης. Ο όρος «επιστημολογία» δεν είναι αρχαιοελληνικός, όπως θα μπορούσε κάποιος να υποθέσει, αλλά ένας νεολογισμός Ευρωπαίων φιλοσόφων. Ο όρος αποδίδεται ως επιστημολογία ελληνικά, epistemology αγγλικά, epistemologie γαλλικά κ.λπ., αυτολεξεί σημαίνει λόγος περί της επιστήμης, από το αρχαιοελληνικό ρήμα επίσταμαι δηλαδή γνωρίζω και το ουσιαστικό λόγος (Suchting 2006). Η χρήση της λέξης, αν και για ορισμένους ταυτίζεται με αυτήν της γνωσιολογίας, σήμερα γενικά θεωρείται ότι πρέπει να είναι πιο περιορισμένη (Μπιτσάκης, 1987), αφού η γνωσιολογία ασχολείται όχι μόνο με την εξέλιξη των επιστημών, αλλά και με τη φύση της γνώσης αυτής καθαυτής. Απαντά σε ερωτήματα σχετικά με το πώς απέκτησε ο άνθρωπος τη δυνατότητα να σκέπτεται, να μαθαίνει, να γνωρίζει και γενικότερα εξετάζει το θέμα φιλοσοφικά, ενώ η επιστημολογία ερευνά τη γένεση των επιστημών και την εξέλιξή τους, αναδεικνύοντας τα αίτια που καθόρισαν την όποια εξέλιξη (θεωρία - πείραμα - χαρακτήρα των επιστημονικών κρίσεων κ.λπ.) και τη λογική που διέπει κάθε επιστήμη, ανάλογα με

τη φύση του αντικειμένου και την κατάταξή τους (φυσικές επιστήμες, μαθηματικά, κοινωνικές επιστήμες κ.λπ.).

Σε μια προσπάθεια να διακρίνουμε καλύτερα τα όρια φιλοσοφίας και επιστήμης θα μπορούσαμε να πούμε ότι η φιλοσοφία παράγει καθολική γνώση, γενικεύει τα αποτελέσματα των επιστημών καλύπτοντας την ανάγκη του ανθρώπου για οργάνωση, τάξη και ενότητα και μας δίνει μια συνολική εικόνα του κόσμου. Εξετάζει τις δυνατότητες της γνώσης, δημιουργώντας και βαθαινοντας το γνωστικό επίπεδο, δημιουργεί συστήματα φιλοσοφικά με τα οποία προσπαθεί να ερμηνεύσει τον κόσμο (παράγει δηλαδή ιδεολογία), θέτει ηθικά ζητήματα, που είναι απαραίτητα στην πρακτική ζωή, μας βοηθάει στην ανάπτυξη της κριτικής σκέψης και απαντά στα μεταφυσικά προβλήματα. Βασικό της εργαλείο είναι οι κατηγορίες (η ομαδοποίηση των εννοιών σε κατά το δυνατό μεγαλύτερα σύνολα και η ταξινόμησή τους). Το αντικείμενό της δεν είναι αυστηρά προσδιορισμένο και η μεθοδολογία της ποικίλλει. Ο κάθε καινοτόμος φιλόσοφος κρίνει τα υπάρχοντα φιλοσοφικά συστήματα έχοντας ως στόχο τη δημιουργία μιας νέας κοσμοαντίληψης για την ορθότερη οργάνωση του κόσμου ώστε να ορίσει με περισσότερη επιτυχία τη θέση του ανθρώπου στην κοινωνία.

Η επιστήμη αντίθετα παράγει γνώσεις σε συγκεκριμένους τομείς, που έχουν αποκτηθεί βάσει συγκεκριμένων μεθόδων και τεχνικών, ανάλογα με το είδος της επιστήμης και είναι γενικά αποδεκτές από τους ειδικούς. Ο κάθε επιστήμονας στηρίζεται στις γνώσεις των προκατόχων του, τις οποίες έχει ως βάση για να εμβαθύνει περισσότερο στα όποια προβλήματα είναι ακόμα αναπάντητα. Στόχος της είναι να αναζητήσει μεθόδους που θα της επιτρέψουν να κατανοήσει τη λειτουργία του συγκεκριμένου τομέα πληρέστερα, ώστε η όποια παρέμβασή της να είναι πιο αποτελεσματική, αφού θα επιτρέπει την πρόβλεψη και τον έλεγχο των μελλοντικών γεγονότων, κατά τη διάρκεια της εφαρμογής της.

1.2.3. Σκοπός και γένεση των επιστημών

Ένα πρώτο ζήτημα που τίθεται προς απάντηση είναι ποιες είναι οι απαρχές της επιστήμης και η σχέση της με τη φιλοσοφία, όπως ήδη αναφέρθηκε.

Η γένεση της προεπιστημονικής γνώσης πραγματοποιήθηκε ιστορικά σε πολύ μακρινούς χρόνους, και κάποια γεγονότα σχετικά με αυτή θα μείνουν για πάντα άγνωστα. Όταν οι ειδικοί π.χ. προσπαθούν να απαντήσουν στο ερώτημα αν τα μαθηματικά ανακαλύφθηκαν ή εφευρέθηκαν, σίγουρα η όποια απάντηση εξαρτάται από την ιδιοσυγκρασία του κάθε επιστήμονα, αφού πλέον ο πρώτος άνθρωπος που σκέφτηκε κάποια θεωρήματα κ.λπ. έχει χαθεί για πάντα. Ακόμα αναμφίβολα όποια αρχαιολογική ανακάλυψη κι αν πραγματοποιηθεί σχετικά με το θέμα, ποτέ δε θα μπορεί κανείς να ισχυριστεί ότι τα ευρήματα αυτά είναι σίγουρα τα παλαιότερα και πολύ περισσότερο δεν θα μπορεί να πετύχει την πλήρη ανασύσταση του όποιου πνευματικού υπόβαθρου των δημιουργών αυτού του επιτεύγματος. Είναι δυνατόν μόνο να γίνουν ορισμένες εικασίες, που ασφαλώς θα στηρίζονται και αυτές με τη σειρά τους σε κάποια κοσμοαντίληψη. Παρόλα αυτά τα δεδομένα που διασώθηκαν, μας επιτρέπουν να σχηματίσουμε μια, έστω μερική, εικόνα για την εμφάνιση του φαινομένου της γνώσης στον κόσμο.

Αναμφίβολα οι πρώτοι άνθρωποι δεν ήταν σε θέση να αντιλαμβάνονται ορθά τις αιτίες των σεισμών ή άλλων φυσικών φαινομένων, ή δεν ήξεραν τις πραγματικές αιτίες που άλλοτε πετύχαιναν και άλλοτε αποτύγγαναν στο κυνήγι τους, ή στο ψήσιμο των αγγείων τους κ.λπ. Ήταν όμως επιτακτική ανάγκη για αυτούς να καταλάβουν τη φύση και τις δυνάμεις της για να μπορέσουν να επιζήσουν πρωταρχικά οι ίδιοι, και αργότερα να βελτιώσουν τη ζωή τους όσο μπορούσαν. Η ανάγκη αυτή γέννησε τα πρώτα σπέρματα της φιλοσοφίας και της επιστήμης. Για να πετύχουν τους σκοπούς αυτούς αναγκάζονταν οι άνθρωποι να παρατηρούν, να σκέφτονται και να κάνουν κάποιες γενικεύσεις που συνήθως κατέληγαν σε κοσμοαντιλήψεις που ενείχαν πολλά στοιχεία μυθοπλασίας, αλλά και κάποια σπέρματα αλήθειας που τους βοηθούσαν να συμβιώνουν και να θέτουν τα προβλήματά τους σε κάποια τάξη. Ταυτόχρονα με τις ίδιες διαδικασίες περίπου δημιουργούσαν τα πρώτα εργαλεία και τις τεχνικές τους για την άμεση ικανοποίηση των πρακτικών αναγκών τους. Η αναζήτηση λοιπόν των πηγών της φιλοσοφίας και των προεπιστημονικών γνώσεων οδηγεί αναμφίβολα στις θρησκευτικές και τις μυθολογικές δοξασίες των διάφορων λαών και στις πρώτες τεχνολογικές τους κατακτήσεις.

Επόμενο ήταν ότι οι άνθρωποι θα αντιλαμβάνονταν κάποια στιγμή ότι οι απαντήσεις τους ήταν ανεπαρκείς και χρειάζονταν είτε συμπλήρωση, είτε αναθεώρηση κ.λπ., έτσι βαθμιαία τα προβλήματα και οι

απαντήσεις έπαψαν να είναι απλοϊκά και άρχισε να γίνεται σταδιακά αντιληπτή η πολυπλοκότητα και η πολυμορφία τους. Η στιγμή που κάποιος θα αισθανόταν την ανάγκη τακτοποίησης αυτού του γνωστικού χάους όλο και πλησίαζε. Και αυτό συνέβη όταν ένας σημαντικός θησαυρός γνώσεων είχε πλέον κατακτηθεί και επομένως το βήμα αυτό ήταν δυνατό να πραγματοποιηθεί. Αν ο βαθμός συσσώρευσης κάποιων γνώσεων ήταν απαραίτητος, άλλο τόσο χρειαζόνταν και οικονομικές και κοινωνικές συνθήκες που θα ευνοούσαν την ανάπτυξη νέων αντιλήψεων. Ο συνδυασμός αυτών των παραγόντων επιτεύχθηκε στις πόλεις της Ιωνίας στη Μ. Ασία (ανάπτυξη του εμπορίου, ναυτιλίας, άμεση γνωριμία άλλων πολιτισμών και σκέψεων κ.λπ. και ταυτόχρονα ανάπτυξη της πολιτικής σημασίας των δήμων) και κάπως έτσι δημιουργήθηκαν οι συνθήκες που επέτρεψαν στον άνθρωπο να περάσει από τις διάφορες θρησκευτικές αντιλήψεις και τη μυθολογία στη φιλοσοφία σε πρώτο βήμα και λίγο αργότερα από την προεπιστήμη στην επιστήμη. Αυτό που διαφοροποιεί τους Ίωνες φιλοσόφους του 6ου αι. π.Χ. από τη μέχρι τότε σκέψη είναι ότι οι κοσμοερμηνείες τους στηρίχτηκαν σε φυσικές εξηγήσεις των φαινομένων και ότι βασικό εργαλείο σε αυτή τη διάσταση είχαν τη λογική τους. Ο κόσμος δε δημιουργήθηκε πλέον γι' αυτούς, όπως τον εξιστορούσε ο Ησίοδος στη Θεογονία, από τους έρωτες της Γης και του Ουρανού, αλλά από το νερό (για τον Θαλή), ή από τον αέρα (για τον Αναξίμενη) κ.λπ.

Οι πρώτες θεωρίες των Ιώνων σοφών, αν και απλοϊκές συχνά, ήταν τολμηρές και επιχειρήσαν να δώσουν καθολική ερμηνεία του κόσμου, δηλαδή έθεσαν τα θεμέλια της φιλοσοφίας. Οι δυσκολίες που συναντούσαν, όμως, καθώς και τα λογικά κενά που κάποιος μπορούσε να εντοπίσει σ' αυτές, παρακίνησαν κάποιους άλλους ανθρώπους, για να καλύψουν πληρέστερα κάποιες ανάγκες τους, να στραφούν όχι τόσο στη γενική ερμηνεία του κόσμου, αλλά στη μελέτη συγκεκριμένων φαινομένων, έστω και αν δεν γνώριζαν τη σχέση που αυτά μπορούσαν να έχουν με τον υπόλοιπο κόσμο, έτσι γεννήθηκαν οι επιστήμες. Η προεπιστημονική εφεύρεση ήταν συμπτωματική και όχι συστηματική, περιοριζόταν σε καθαρά ωφελμιστικό στόχο, χωρίς να ενδιαφέρεται να αναζητήσει ριζικές λύσεις που θα βελτίωναν τα μέχρι στιγμής δεδομένα (Hawkes & Wooley, 1963). Ένα παράδειγμα προεπιστημονικής γνώσης μας δίνει μία επιγραφή που βρέθηκε στη Νινευί, τον 7ο αι. π.Χ. και η οποία δίνει συστηματικές οδηγίες για το πώς κάποιος μπορεί να κατασκευάσει φαγεντιανή¹. Στο κείμενο καταγράφεται λεπτομερώς τί πρέπει να γίνει για να επανέλθει με επιτυχία η όποια επιχειρούμενη κατασκευή φαγεντιανής. Από τη μελέτη της επιγραφής προκύπτει ότι οι τεχνικοί της εποχής είχαν κάνει κάποια πειράματα, γεγονός που θεωρείται το πρώτο βήμα προς την επιστήμη. Στο σχετικό κείμενο, όμως, δεν υπάρχει ούτε αιτιολόγηση γιατί πρέπει να ακολουθηθεί αυτή η μέθοδος και όχι άλλη ούτε αναφέρονται ποιες οι όποιες αλλαγές υπέστησαν τα υλικά, στοιχεία που σίγουρα είχαν τη δυνατότητα να παρατηρήσουν μέχρι την παραγωγή του τελικού προϊόντος οι τεχνικοί, οι οποίοι όμως δεν επιχειρήσαν τη σχετική καταγραφή. Οι επιστήμες που αναπτύχθηκαν, όμως, στον «αρχαιοελληνικό κόσμο» στηρίζονταν πια σε αυστηρά δομημένο και συνεκτικό σύστημα προτάσεων (θεωρημάτων και πορισμάτων) με βάση ένα σύνολο ορισμών, αποδείξεων, κ.λπ. και θα παρήγαγαν γνώση με γενική αποδοχή π.χ. την ευκλείδεια γεωμετρία.

1.2.4. Τα πρώτα βήματα της φιλοσοφίας και των επιστημών

Από όλα τα ζητήματα που τίθενται επί τάπητος από τους φιλοσόφους: μεταφυσικά, ηθικά γνωσιολογικά, το τελευταίο αναδύεται ως το θεμελιώδες, όπως ήδη αναφέρθηκε.

Ερευνώντας τη φύση της γνώσης, οι πρώτοι φιλόσοφοι εστιάζουν το ενδιαφέρον τους κυρίως στον ορισμό της και όχι τόσο στη μεθοδολογία της. Ο Λεύκιππος και ο Δημόκριτος επιχειρούν να ερμηνεύσουν τα διάφορα φυσικά φαινόμενα που προσεγγίζουν στηριγμένοι σε παρατηρήσεις, δηλαδή στον εμπειρισμό, δυστυχώς όμως το «Περί νου», όπου ο Δημόκριτος αναπτύσσει τις απόψεις του συστηματικά, δεν έχει διασωθεί. Ο Πλάτωνας αναζητά τις απαντήσεις του για τον κόσμο στη νόηση. Κατά τη γνώμη του, οι ιδέες-έννοιες δε βρίσκονται στον αισθητό κόσμο, αλλά είναι αιώνιες, αμετάβλητες και συλλαμβάνονται με το νου. Ο Αριστοτέλης, λίγο αργότερα, δεν διστάζει να έλθει σε αντιπαράθεση με τον Πλάτωνα και να θέσει τα θεμέλια της λογικής (των μεθόδων της νόησης και της εξαγωγής συμπερασμάτων), τα οποία πιστεύει ότι πρέπει να στηρίζονται στη μελέτη του υλικού κόσμου. Λίγο αργότερα οι στωικοί εντοπίζουν τις αδυναμίες

¹Φαγεντιανή ονομάζεται μια σύνθετη ύλη που κατασκευάζεται από κονιορτοποιημένο χαλαζία, αμμόλιθο ή πυριτόλιθο και από ένα διάλυμα ανθρακικού νατρίου. Το υγρό μίγμα αυτών των υλικών τοποθετούνταν σε μήτρες και ψηνόταν σε υψηλή θερμοκρασία, περίπου στους 870ο C

των θεωριών των δύο μεγάλων φιλόσοφων της αρχαιότητας και θέτουν τα θεμέλια της σχετικότητας στη γνώση.

Εντωμεταξύ συντελείται και η μερική αυτονόμηση των επιστημών. Στο βαθμό που η κάθε επιστήμη αποκτά τη δική της μέθοδο και τα δικά της κριτήρια για ανάπτυξη είναι πλέον δυνατό να προχωρήσει μόνη της. Οι πρώτες επιστήμες που θα αυτονομηθούν από τη φιλοσοφία θα είναι τα μαθηματικά, η αστρονομία και η ιατρική (ήδη από τα τέλη του 5αι. π.Χ.).

Τα μαθηματικά επωφελούνται πρώτα, γιατί μελετούν αρχικά τουλάχιστον γενικές ιδιότητες των σωμάτων, σχήματα, ποσότητες κ.λπ., ιδιότητες που ο άνθρωπος μπορεί να συλλάβει ευκολότερα εξετάζοντας τα αντικείμενα και να εμβαθύνει σε αυτές. Με την πρόοδο π.χ. της χωρομετρίας οι άνθρωποι φθάνουν στη γεωμετρία, κατά την παρατήρηση των άστρων σε συνδυασμό με την παράλληλη μελέτη των μαθηματικών αρχίζει η ανάπτυξη της τριγωνομετρίας κ.λπ. Με τη σειρά της η αστρονομία, μια επιστήμη που δε στηρίζεται στο πείραμα, αλλά στην παρατήρηση των ουράνιων σωμάτων, επωφελούμενη και από την πρόοδο στα μαθηματικά παίρνει ιδιαίτερη ώθηση, αφού είναι απαραίτητη τόσο στη γεωργία, όσο και στη ναυσιπλοΐα. Στην ιατρική, παρά τις ελλειπείς γνώσεις ανατομίας και φυσιολογίας, προφανώς λόγω της αμεσότητας των ζητημάτων, προωθείται επίσης η προσεχτική κλινική ανάλυση σε συνδυασμό με την ιατρική δεοντολογία, που βασίζεται στον ορθολογισμό της εποχής. Τα θεμέλια της ιατρικής επιστήμης δημιουργούνται αυτή την εποχή από τον Ιπποκράτη. Οι άλλες επιστήμες παραμένουν σε όλη την αρχαιότητα συνδεδεμένες με τη φιλοσοφία. Η πειραματική διαδικασία που θα τους έδινε τεράστια ώθηση επρόκειτο να εδραιωθεί βραδύτερα με αργά και κοπιώδη βήματα για να μπορέσει με τον Γαλιλαίο και τον Νεύτωνα να φθάσει στα πρώτα ικανοποιητικά αποτελέσματα.

Με την καθιέρωση του χριστιανισμού οι γνωσιολογικές θεωρίες επανεξετάζονται, φυσικά, με αποτέλεσμα οι πατέρες της εκκλησίας να υιοθετήσουν όσες απόψεις των Ελλήνων φιλοσόφων δεν έρχονται σε αντίθεση με τον χριστιανισμό. Ούτε ο φιλόσοφος, ούτε ο επιστήμονας, μπορούν να διερευνούν θέματα που έρχονται σε αντίθεση με τις γραφές (έτσι ανοίγει ο δρόμος για διάφορες αλχημικές τάσεις και απόκρυφες φιλοσοφίες που θα αναπτυχθούν στα περιθώρια της χριστιανικής σκέψης).

Κατά την περίοδο του Μεσαίωνα κυριαρχεί ο σχολαστικισμός που βασίζεται κύρια στις αριστοτελικές απόψεις και έχει ως βασικό χαρακτηριστικό την ταύτιση της φιλοσοφίας με τον χριστιανισμό. Παράλληλα φυσικά συνυπάρχουν διάφορες νεοπλατωνικές αντιλήψεις σε συνδυασμό με τον πυθαγόρειο επιστημονικό προσανατολισμό.

Κάποια στιγμή όμως οι εμπειρικές διαπιστώσεις επιστημόνων, που στρέφονται στην παρατήρηση και σε στοιχειώδη πειράματα για την απόκτηση γνώσεων φθάνουν στη διατύπωση του πρώτου κριτηρίου οριοθέτησης των σύγχρονων επιστημών: ο υλικός κόσμος είναι προσιτός μέσω των αισθήσεων (Losee, 1993). Είναι γνωστή η θεωρία του Occam [το ξυράφι του Όκαμ - (Occam's Razor)], σύμφωνα με την οποία η «οικονομία» στην επιστήμη, δηλαδή η γρηγορότερη και αποτελεσματικότερη μέθοδος για την απόκτηση γνώσης των φυσικών φαινομένων είναι η άμεση παρατήρηση. Ο ίδιος, φυσικά πρεσβεύει ότι η έννοια του Θεού δεν μπορεί να γίνει αντικείμενο των φυσικών επιστημών αλλά είναι και πρέπει να μείνει το αναμφισβήτητο έργο της θεολογίας και ότι τις «καθολικές έννοιες» μπορεί να τις αντιλαμβάνεται μόνο ο Θεός και να τις αποκαλύπτει στους πιστούς. Οι επιστήμες οφείλουν να εγκαταλείψουν την αναζήτηση της βαθύτερης ουσίας των πραγμάτων και να προσανατολιστούν στη μελέτη των φυσικών φαινομένων. Με αυτό τον τρόπο, όμως, καθιερώνονται δύο αλήθειες: η επιστημονική και η θεολογική, σε περίπτωση αντιφατικών απόψεων, η εξ αποκαλύψεως άποψη θεωρείται ορθότερη. Καρπός, όμως, της θεωρίας αυτής είναι η οριοθέτηση της ανθρώπινης γνώσης, αναγνωρίζεται η αξιοπιστία των αισθήσεων και είναι δυνατόν πια να θεμελιωθούν βαθμιαία οι φυσικές επιστήμες. Ο επιστήμονας επιτρέπεται τώρα να πιστεύει στις αλήθειες της φύσης και να τις διερευνά ως κάτι υπαρκτό, και όχι ως μια απλή υπόθεση χωρίς φυσική αντιστοιχία.

Για να υπάρξουν βέβαια ουσιαστικά αποτελέσματα στην έρευνα θα χρειαστεί χρόνος ακόμη. Η πραγματική πρόοδος θα επέλθει από ανακαλύψεις που δημιουργούν διαδοχικά πλήγματα στις αντιλήψεις του Αριστοτέλη, που κυριαρχούν μέχρι την εποχή εκείνη. Από τις πρώτες απόψεις που καταρρίπτονται είναι αυτή σύμφωνα με την οποία ένα σώμα διατηρείται σε κίνηση μόνο με τη δράση μιας συνεχούς εξωτερικής δύναμης. Κατά την αριστοτελική άποψη, ένα βλήμα που κινείται μέσω του αέρα οφείλει τη συνεχή κίνησή του σε δίνες ή δονήσεις που υπάρχουν στο περιβάλλον. Σε περίπτωση απουσίας αυτών δυνάμεων, το σώμα μένει ακίνητο σχεδόν αμέσως. Ο Jean Buridan, ένας Γάλλος ιερέας, όμως, παρατηρεί ότι η κίνηση διατηρείται από μια ιδιότητα του σώματος-βλήματος: την ώθηση. Το βλήμα αποκτά αυτή την ιδιότητα όταν τίθεται σε

κίνηση, ο ίδιος θα υποστηρίξει ότι η ώθηση μειώνεται λόγω της αντίστασης του αέρα και της βαρύτητας στην πορεία. Στη συνέχεια ο Κοπέρνικος αποκαλύπτει νέους φυσικούς νόμους και η θεωρία ότι η θέση της γης είναι στο κέντρο του σύμπαντος εγκαταλείπεται οριστικά. Ο Γαλιλαίος συμβάλλει ουσιαστικά στην πρόοδο της μεθοδολογίας με την καθιέρωση της ευρετικής διαδικασίας (ανάλυση για εντοπισμό αιτιών και σύνθεση για την επιβεβαίωση των προτάσεων) στα μαθηματικά. Η φύση αντιμετωπίζεται πλέον ως ένα μηχανικό σύστημα, η φυσική μαθηματικοποιείται και είναι ένα βήμα πια από την καθιέρωση του πειράματος. Η αρτιότητα της επιστημονικής διαδικασίας θα συμπληρωθεί λίγο αργότερα από τον Νεύτωνα με τις παρατηρήσεις του για τη φύση του φωτός. Η αναζήτηση της επιστημονικής αλήθειας έχει βρει το δρόμο της και είναι έτοιμη να αυτονομηθεί πλέον οριστικά από τη φιλοσοφία. Όσο αργά ήταν τα πρώτα βήματα των επιστημών μέχρι τότε, τόσο αυτά επιταχύνονται ευνοούμενα από τις κοινωνικοοικονομικές συνθήκες. Το γεγονός θα αναδείξει φυσικά την αναγκαιότητα απόκτησης και νέων επιστημονικών μεθόδων, που όμως θα βρίσκονται σε στενή συνάφεια πάντοτε με τις γνωσιολογικές και φιλοσοφικές εξελίξεις.

1.2.5. Ερωτήματα της γνώσης - φιλοσοφικά ρεύματα και μεθοδολογικές προτάσεις

Απαραίτητο νομίζουμε είναι να δούμε κάπως πιο συγκεκριμένα το γνωσιολογικό ζήτημα, αφού από αυτό όπως ήδη αναφέραμε δημιουργήθηκε η επιστημολογία.

Αν και το αντικείμενο της γνώσης είχε απασχολήσει ήδη τους αρχαίους Έλληνες και ο Πλάτων την όριζε ως την αληθή δόξα μετά λόγου, δηλαδή ως την πεποίθηση που προέρχεται μεν από τις αισθήσεις, αλλά έχει υποστεί τον απαραίτητο κριτικό έλεγχο της λογικής και έχει βρεθεί αληθής, θα χρειαστούν αρκετοί αιώνες για να θεμελιωθεί η γνωσιολογία ως αυτοτελής και συστηματικός κλάδος της φιλοσοφίας, από τους R. Descartes (1596-1650), J. Locke (1632-1704) και Em. Kant (1724-1804). Ο Descartes είναι ο πρώτος που αντιλαμβάνεται την ανάγκη εξέτασης της μεθόδου που χρησιμοποιούσε μέχρι τότε η γνωσιολογία. Δέχεται το λόγο ως το όργανο για την κατάκτηση της γνώσης. Ακόμα αναγνωρίζει την ύπαρξη αξιωμάτων από τα οποία με αυστηρούς απαγωγικούς συλλογισμούς μπορεί κάποιος να κατακτήσει τη γνώση περαιτέρω. Ο Locke μελετά το ίδιο πρόβλημα, θεωρώντας όμως ως όργανο της γνώσης τις αισθήσεις και μεταθέτοντας το θέμα της έρευνας που μέχρι τότε ήταν η μεταφυσική (δηλαδή η προέλευση του κόσμου και η βαθύτερη ουσία του), στη μελέτη της ανθρώπινης νόησης, δηλαδή στην πραγματικότητα. Λίγο αργότερα ο Kant επιχειρεί να συγκεράσει τις 2 αυτές απόψεις υποστηρίζοντας ότι είναι απαραίτητη η συνεργασία του νου και των αισθήσεων για τη δημιουργία των γνώσεων.

Εξετάζοντας τα βασικά ερωτήματα που θέτει η μεθοδολογικά οργανωμένη πλέον γνωσιολογία διαπιστώνουμε ότι την απασχολούν:

- Το αντικείμενο της γνώσης,
- Η πηγή της γνώσης,
- Το εάν είναι δυνατή η απόκτηση της απόλυτης γνώσης.

Η εφαρμογή της επιστημονικής έρευνας στο σύντομο, σχετικά, χρονικό διάστημα ανάπτυξής της, όπως ήταν φυσικό, στηρίζεται στο πείραμα, ή τον έλεγχο των όποιων παρατηρήσεων των ερευνητών δημιουργώντας μια σειρά κριτήρια για την επαλήθευση των υποθέσεών της.

Η πιο παλιά χρησιμοποιούμενη μεθοδολογία, στις νεότερες επιστήμες, είναι αυτή του εμπειρισμού που στηρίζεται σε πειράματα που ακολουθούν είτε την παραγωγική είτε την επαγωγική μέθοδο και στην εξ αυτών εξαγωγή συμπερασμάτων. Η σαφήνεια της πειραματικής διαδικασίας βασικά προκύπτει από τη δυνατότητα που μας παρέχουν τα διαδοχικά βήματα αυτής της διαδικασίας να τα παρατηρήσουμε και να τα καταγράψουμε με τις αισθήσεις μας. Αυτή ακριβώς η δυνατότητα της αισθητηριακής παρατήρησης μας επιτρέπει να φθάσουμε στην αξιολόγηση του πειράματος και να αποκτήσουμε κάποια βεβαιότητα για το τί συμβαίνει στον κόσμο. Έχοντας συνείδηση των ορίων του πειράματος και της παρατήρησης οι ειδικοί, που την εφαρμόζουν, προτείνουν και το σχετικό περιορισμό του επιστημονικού αντικειμένου, αφού η όποια επαλήθευση δεν μπορεί να ολοκληρωθεί με λογική αυστηρότητα και ο έλεγχος μπορεί να είναι μόνο ενδεικτικός.

Ο θετικισμός θεωρεί ότι κάθε αξιόπιστη γνώση βασίζεται στην εμπειρία, πιστεύει, όμως, ότι η διατύπωση γενικών νόμων είναι δυνατή όταν επαληθεύεται με τη μαρτυρία γεγονότων, τα οποία να μεν δεν προέρχονται από άμεση παρατήρηση, αλλά αποτελούν έννοιες άλλων γενικών επιστημονικών θεωριών.

Ο Karl Raimund Popper (1902-1994) προσπάθησε να λύσει το πρόβλημα της επαλήθευσης στηριζόμενος στη θεμελιώδη ασυμμετρία μεταξύ επαλήθευσης και διάψευσης μιας θεωρίας. Ανέδειξε δηλαδή το ότι ως αλήθεια αποδεχόμαστε, γενικά, κάτι που αντιστοιχεί με την πραγματικότητα. Είναι πιο λογικό να συνδυάσουμε δύο πράγματα που μοιάζουν, παρά δύο πράγματα που δεν μοιάζουν· το τραπέζι, π.χ. μπορεί να συνδυαστεί με έννοιες όπως η συνδιάσκεψη, το φαγητό κ.λπ., αλλά όχι με αταίριαστες έννοιες, όπως το φεγγάρι, ο ρουχισμός κ.ά. Ο ίδιος επισημαίνει ότι δεν είναι λογικό να επικροτούμε συσχετισμούς αταίριαστων πραγμάτων. Κατά τη γνώμη του η ορθή επιστημονική θεωρία ξεχωρίζει από τις άλλες ακριβώς επειδή μας προμηθεύει κριτήρια βάσει των οποίων μπορούμε να συλλάβουμε σε ποιες περιπτώσεις δεν είναι δυνατό να εφαρμοστεί μία θεωρία, τότε δηλαδή αυτή διαψεύδεται.

Ο εμπειρικός ρεαλισμός ένοιωσε επίσης την ανάγκη περιορισμού του επιστημονικού αντικειμένου, καθώς στάθηκε στη διττή φύση των εξεταζόμενων προβλημάτων. Κάποια θέματα, κατά την άποψή του, είναι ατομικά-ιδιωτικά, δηλαδή ο νους για να συλλάβει και να εξετάσει ένα ζήτημα, πρέπει να το έχει ο ίδιος νοιώσει, με την έννοια ότι αν κάποιος δεν πονά, το μυαλό του δεν έχει κάτι να παρατηρήσει και να αντιληφθεί τί είναι ο πόνος. Η επιστήμη σ' αυτή την περίπτωση δεν έχει συγκεκριμένο πεδίο (δύο άνθρωποι δεν μπορούν να έχουν την ίδια άποψη για τον πόνο, ή την επιθυμία κ.λπ.). Άλλα θέματα είναι δημόσια και ο νους έχει τη δυνατότητα να τα εξετάσει αντικειμενικά (τη φύση, τα αντικείμενα κ.λπ.) γιατί υπάρχουν ανεξάρτητα από αυτόν, και ο καθένας μπορεί να τα εξετάσει και να τα κατανοήσει σύμφωνα με την κοινή λογική.

Ο κριτικός ρεαλισμός αξιοποιώντας τα επιτεύγματα της επιστήμης επιχειρεί με την αποκάλυψη των δομών των φαινομένων και την κατανόησή τους να συλλάβει κάποιες γενικές αλήθειες. Δεν απαιτείται, κατά τη γνώμη του, άπειρη σειρά επαληθεύσεων, αφού η γνώση των εσωτερικών νομοτελειακών λειτουργιών που καθορίζουν το φαινόμενο αποτελεί και την απόδειξη της ισχύος μιας πρότασης. Γνωρίζουμε, π.χ. ότι ο ήλιος θα ανατείλει αύριο, όχι γιατί έτσι η εμπειρία μας έχει δείξει, αλλά γιατί ξέρουμε πια από τους νόμους του Νεύτωνα την κίνηση του ήλιου και που ακριβώς βρίσκεται αυτός κάθε στιγμή. Θεωρεί δε ότι με τον εντοπισμό των μηχανισμών εξέλιξης των φαινομένων, ανοίγει και η προοπτική περαιτέρω επίδρασης σε αυτά. Οι επικριτές της θεωρίας, όμως, επισημαίνουν ότι και πάλι δεν είναι δυνατή η απόλυτη βεβαιότητα, αφού η αντίληψη των δομών των φαινομένων δεν είναι άμεση, αλλά στηρίζεται σε λογικές ερμηνείες.

Η φαινομενολογία με ιδρυτή τον Edmund Husserl (1859-1938) επιχειρεί να περιορίσει τα σφάλματα του εμπειρισμού, του θετικισμού κ.λπ. Προσπαθεί να ξεκαθαρίσει το ερευνητικό πεδίο από τα συμπτωματικά και υποκειμενικά στοιχεία της παρατήρησης με στόχο η ουσία του φαινομένου να μείνει ελεύθερη και να γίνει καλύτερα αντιληπτή. Ένα σχετικό παράδειγμα μπορούμε να δούμε στην άποψη του Σεφέρη για τον Κάλβο. Ο Σεφέρης, όπως ο ίδιος αναφέρει, είχε διαβάσει Κάλβο αρκετές φορές, η κρίση του δε γι' αυτόν διέφερε από ανάγνωση σε ανάγνωση, ο Κάλβος τον άφηγε αδιάφορο αρχικά, ενώ μετά από κάποια στιγμή κατέληξε να τον θαυμάζει. Η ταυτότητα της ποιητικής τέχνης του Κάλβου είναι ανεξάρτητη, όμως, από την ψυχική διάθεση του Σεφέρη ή άλλου ποιητή και αυτή ενδιαφέρει την έρευνα που καλείται να αποφύγει την πολυγωνμία και να εντοπίσει την καθαυτό αξία του Κάλβου.

Διάφορα σύγχρονα ρεύματα της ερμηνευτικής κοινωνιολογίας θεωρούν ότι για τη μελέτη των προβλημάτων του ανθρώπου απαιτείται πιο εξειδικευμένη μεθοδολογία και στρέφουν την προσοχή τους σε εργαλεία μέτρησης κατάλληλα για τη μελέτη της καθημερινής ανθρώπινης πραγματικότητας και πώς αυτή νοηματοδοτείται στην απλή βιωματική εμπειρία. Το γεγονός όμως ότι οι σχετικές έρευνες διενεργούνται σε μικρό αριθμό ατόμων περιορίζει την εγκυρότητα των αποτελεσμάτων και τις όποιες δυνατότητες γενίκευσης.

Ο σύγχρονος πραγματισμός, τέλος, αναζητά επίσης νέα εργαλεία μέτρησης για να κατανοήσει τις ενέργειες της ανθρώπινης γνώσης σε όλη την πολυπλοκότητά τους (αξιακή, γνωσιολογική και εμπειρική). Σύμφωνα με τους πραγματιστές η «ωφελιμότητα» και η «αλήθεια» δεν είναι πια διαχωρισμένες η μία από την άλλη, αλλά εμφανίζονται ως οι δύο όψεις της αυτής πραγματικότητας.

1.2.6. Επιστημάνσεις συμπεράσματα

Σήμερα έχοντας διανύσει ένα αρκετά μακρύ διάστημα εφαρμογής σημαντικών επιστημονικών προγραμμάτων μπορούμε να υποστηρίξουμε ότι παρά τις όποιες κριτικές δέχτηκε η φιλοσοφία κατά την ιστορική διαδρομή της είναι σαφές ότι στο σύμπαν τα πάντα βρίσκονται σε κάποια σχέση μεταξύ τους. Ο χωρισμός της φιλοσοφίας από τις επιστήμες κρίθηκε απαραίτητος λόγω των δυσκολιών στην κατάκτηση της γνώσης, ώστε να διευκολυνθεί η μελέτη των διάφορων αντικειμένων. Δεν παύει, όμως, να είναι ένας τεχνητός χωρισμός, η ανάγκη για καθολική γνώση δεν μπορεί να αγνοηθεί. Οι όποιες επιστημονικές πεποιθήσεις, είναι ειδικές και δεν απαντούν σε όλα τα θέματα που θέτει η ανθρώπινη φύση (ηθικά, μεταφυσικά), έτσι αναπόφευκτα δημιουργείται η ανάγκη εντοπισμού, αλλά και ξεπερασμού των ορίων των δύο πεδίων. Φιλοσοφία και επιστήμες είναι ουσιαστικά στενά συνδεδεμένες.

Πολλές από τις επιστημονικές διενέξεις που προκλήθηκαν η ίδια η πρόοδος της γνώσης είτε τις επιλύει (π.χ. μηχανισμός αισθητηριακών δεδομένων), είτε τις θέτει σε νέα βάση (η θεωρία εξέλιξης του Δαρβίνου κ.ά.).

Η ιστορική αναδρομή αναδεικνύει ότι η επιστήμη και η φιλοσοφία δεν εξελίσσονται ευθύγραμμα, αλλά αυτό συμβαίνει μόνον όταν οι κοινωνικές και οικονομικές συνθήκες είναι ώριμες και το επιτρέπουν. Πολλά ερωτήματα εξάλλου παραμένουν ανοιχτά γύρω από τη φύση της επιστήμης και σχετικά με το πώς αυτή λειτουργεί. Τέτοια ερωτήματα συνήθως δεν ανακύπτουν κατά τη μελέτη μιας συγκεκριμένης επιστήμης, αλλά τα αντιμετωπίζει η φιλοσοφία, ή οι ίδιοι οι επιστήμονες όταν αισθανθούν την ανάγκη να φιλοσοφήσουν.

Η απόλυτη αλήθεια δεν είναι δυνατό να επιτευχθεί ούτε από τη φιλοσοφία, ούτε από την επιστήμη, τουλάχιστον με τα σημερινά δεδομένα. Οι επιτυχίες, όμως, των επιστημών δείχνουν πολλές φορές αντιστοιχία γνώσης και αντικειμενικής πραγματικότητας.

Από πολλούς αναγνωρίζεται η ανάγκη ύπαρξης διαλόγου μεταξύ των διάφορων ερευνητικών μεθόδων ώστε να ανασυγκροτηθεί το νόημα των επιστημονικών όρων προς μία κατεύθυνση ομογενοποίησής τους και να δημιουργηθούν οι προϋποθέσεις για βαθύτερη κατανόηση της επιστημονικής προόδου.

1.3. Σκοποί επιστημονικής έρευνας

Κατά τον Earl Babbie (2012) τρεις είναι οι συνηθέστεροι και χρησιμότεροι σκοποί μιας έρευνας η διερεύνηση, η περιγραφή και η ερμηνεία.

1.3.1 Διερεύνηση

Η διερευνητική προσέγγιση προκύπτει συνήθως όταν ο ερευνητής εξετάζει ένα νέο αντικείμενο. Η σχετική έρευνα διεξάγεται μέσω συνεντεύξεων, συζητήσεων, χρήσης ομάδων εστίασης ή καθοδηγούμενων συζητήσεων σε μικρές ομάδες κ.ά. Οι παραπάνω τεχνικές χρησιμοποιούνται συχνά σε έρευνες αγοράς. Πολλές φορές ο λόγος διεξαγωγής της έρευνας δεν είναι μόνο η επιθυμία του ερευνητή για την καλύτερη κατανόηση ενός θέματος, αλλά και ο εντοπισμός της ορθής μεθόδου για τη συγκέντρωση των κατάλληλων πληροφοριών για έρευνα. Για παράδειγμα, μπορεί να πραγματοποιηθεί διερευνητική έρευνα προκειμένου να ελεγχθεί η δυνατότητα διεξαγωγής μιας πιο εκτεταμένης μελέτης ή προκειμένου να διαμορφωθούν οι μέθοδοι που θα χρησιμοποιηθούν σε επόμενη μελέτη (ένας ερευνητής, π.χ. πάει σε μία περιοχή που μόλις σταμάτησαν οι εχθροπραξίες για να εξετάσει τις συνθήκες διαβίωσης, τη σύνθεση πληθυσμού, κ.λπ.).

Συνοψίζοντας θα λέγαμε ότι οι διερευνητικές εμπειρικές μελέτες χρησιμοποιούνται κάθε φορά που ένας ερευνητής αντιμετωπίζει ένα καινούργιο θέμα. Το βασικό μειονέκτημα των διερευνητικών μελετών είναι ότι σπάνια προσφέρουν επιστημονικά τεκμηριωμένες απαντήσεις σε ερευνητικά ερωτήματα, αν και μπορούν να υποδείξουν πιθανές απαντήσεις και μπορούν να χρησιμοποιηθούν ώστε να επιλεγούν οι κατάλληλες ερευνητικές μέθοδοι, που θα ήταν σε θέση να προσφέρουν οριστικές απαντήσεις. Η κυριότερη αιτία που οι διερευνητικές μελέτες είναι σπάνια έγκυρες και αξιόπιστες είναι ότι η δειγματοληψία που χρησιμοποιείται συνήθως είναι ευκαιριακή.

1.3.2 Περιγραφή

Συχνά ο ερευνητής αρκείται στην περιγραφή των όσων παρατήρησε. Οι επιστημονικές περιγραφές βασίζονται σε ακριβή δεδομένα και είναι συνήθως πιο ακριβείς και έγκυρες από άλλα είδη ερευνών.

Παράδειγμα περιγραφικής μελέτης είναι η απογραφή των κατοίκων της Ελλάδας, που πραγματοποιήθηκε από την Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ) το 2011. Ξεκίνησε τον Φεβρουάριο και ολοκληρώθηκε τον Μάιο. Είχε ως σκοπό την απογραφή τόσο του πληθυσμού, όσο και των κάθε είδους κτηρίων της χώρας. Όπως σε όλες τις απογραφές εστίασε στην καταγραφή του «μόνιμου» πληθυσμού της Ελλάδας, και όχι του «πραγματικού». Καταχώρισε δηλαδή τους απογραφόμενους στον πίνακα της περιοχής, όπου δήλωσαν ότι ζουν μόνιμα το τελευταίο δωδεκάμηνο, όχι στο σημείο που βρίσκονταν την ημέρα της απογραφής.

Το μειονέκτημα των περιγραφικών μελετών είναι ότι καταγράφουν και περιγράφουν προβλήματα, αλλά δεν εντοπίζουν, ούτε αναλύουν αιτιακές σχέσεις.

1.3.3 Ερμηνεία και αναζήτηση αιτιότητας

Οι ερμηνευτικές μελέτες αναζητούν την αιτία των αποτελεσμάτων. Σε αντίθεση με τα προηγούμενα είδη ερευνών αντικείμενό τους δεν είναι η καταγραφή, αλλά η έγκυρη και τεκμηριωμένη ερμηνεία γεγονότων. Για παράδειγμα, το ποσοστό πωλήσεων δύο αναψυκτικών είναι περιγραφική έρευνα, το γιατί υπερτερούν οι πωλήσεις του Α σε σχέση με αυτές του Β είναι ερμηνευτική.

Θέματα ερμηνευτικών ερευνών μπορεί να είναι ενδεικτικά: 1) Η αναζήτηση των αιτίων που προκαλούν τις εκδηλώσεις βίας στα σχολεία. 2) Ποιοι είναι οι λόγοι επιτυχίας ενός ροκ συγκροτήματος; 3) Γιατί οι γυναίκες ψηφίζουν άντρες υποψηφίους κατά τις βουλευτικές εκλογές;

Αιτιότητα

Στις ερμηνευτικές μελέτες μας ενδιαφέρει να εξετάσουμε την αιτία διάφορων γεγονότων, των οποίων η αιτιακή σχέση εξηγείται είτε με τη **νομοθετική εξήγηση** (nomothetic explanation) είτε με την **ιδιογραφική** (idiographic explanation).

Με τη νομοθετική εξήγηση προσπαθούμε να καταγράψουμε κάποιους από τους παράγοντες (ανεξάρτητες μεταβλητές) που ερμηνεύουν ορισμένες αιτίες ενός φαινομένου. Αντίθετα με το ιδιογραφικό μοντέλο αναζητούμε την πλήρη ερμηνεία των αιτίων ενός φαινομένου.

Στη νομοθετική εξήγηση η αιτιότητα δεν είναι ούτε πλήρης, ούτε αποκλειστική. Για παράδειγμα, μπορεί να εξετάζεται η αιτία που προκάλεσε μια αρρώστια σε κάποιον ασθενή, π.χ. είναι γνωστό ότι το κάπνισμα συσχετίζεται με τον καρκίνο του πνεύμονα, αυτό δεν σημαίνει, ότι το κάπνισμα είναι η αποκλειστική αιτία της εμφάνισης αυτής της ασθένειας σε κάποιον άνθρωπο ή ότι όποιος καπνίζει θα παρουσιάσει καρκίνο στους πνεύμονες κάποια στιγμή.

Παράδειγμα ιδιογραφικής προσέγγισης έχουμε στην ακόλουθη περίπτωση: μια ομάδα ποδοσφαίρου, προκειμένου να αγωνιστεί σε ένα παιχνίδι στο εξωτερικό, μεταφέρεται αυθημερόν με αεροπλάνο στη χώρα διεξαγωγής του αγώνα. Κατά την πτήση, όμως, δημιουργούνται επικίνδυνα καιρικά φαινόμενα με αποτέλεσμα οι επιβάτες να φθάσουν στον προορισμό τους σε πολύ κακή ψυχολογική και σωματική κατάσταση. Η απόδοση της ομάδας στον αγώνα είναι ιδιαίτερα μειωμένη σε σχέση με άλλες φορές και η αιτία γι' αυτό είναι προφανής και πλήρης (η προηγηθείσα ταλαιπωρία), έτσι μπορούμε να πούμε ότι προσεγγίσαμε με ακρίβεια την ερμηνεία του αποτελέσματος χρησιμοποιώντας μια ιδιογραφική προσέγγιση. Παρατηρούμε ότι η αιτία προηγείται του αποτελέσματος. Αυτή η σχέση (αιτίας και αποτελέσματος) καλείται αιτιολογική. Ένα ακόμη παράδειγμα στο οποίο ακολουθείται η ίδια χρονική τάξη είναι η οπαδική προτίμηση των παιδιών. Είναι γνωστό ότι η ποδοσφαιρική ομάδα που επιλέγουν τα παιδιά συσχετίζεται με την αντίστοιχη προτίμηση του πατέρα. Προφανώς σε πολύ σπανιότερες περιπτώσεις μπορεί η επιλογή του πατέρα να επηρεαστεί από αυτήν του παιδιού!

Σε μία αιτιολογική σχέση οι μεταβλητές σχετίζονται είτε ανάλογα είτε αντιστρόφως ανάλογα. Υπάρχει όμως και η περίπτωση η συσχέτιση να είναι απλά στατιστική και όχι αιτιολογική. Παράδειγμα το κάπνισμα εκτός από τον καρκίνο του πνεύμονα συνδέεται με κίτρινο δόντι από τη νικοτίνη.

Εύκολα θα μπορεί να διαπιστώσει κάποιος στατιστική συσχέτιση του κιτρινισμού των δοντιών με τον καρκίνο του πνεύμονα. Η συσχέτιση αυτή όμως δεν είναι αιτιολογική και ως εκ τούτου δεν είναι ενδιαφέρουσα επιστημονικά και συνήθως παραλείπεται.

Συχνά προκύπτουν σφάλματα λόγω παρανόησης της αιτιότητας. Σύμφωνα με τη νομοθετική προσέγγιση «οι εξαιρέσεις δεν επιβεβαιώνουν τον κανόνα», για παράδειγμα, είναι γνωστό ότι οι μεγαλύτεροι ηλικίας άνθρωποι είναι πιο συχνά θρησκευόμενοι από τους νεότερους. Αυτό, όμως, δεν σημαίνει ότι δεν υπάρχουν άθεοι μεγάλης ηλικίας ή θρησκευόμενοι νέοι κ.λπ. Εάν χρησιμοποιηθεί ιδιογραφική εξήγηση σε αυτή την περίπτωση, η έρευνα δεν θα δώσει σωστά αποτελέσματα, αφού η αιτιότητα δεν είναι πλήρης. Ένα αντιπαράδειγμα πλήρους αιτιότητας είναι και το κάπνισμα, το οποίο, ως γνωστό, είναι αιτία καρκίνου του πνεύμονα, προφανώς όμως δεν είναι η μοναδική αιτία (πολλοί ασθενείς μπορεί να έχουν αυτή την πάθηση χωρίς να είναι καπνιστές!). Η συσχέτιση επίσης κάποιες φορές ισχύει στη μειοψηφία των περιπτώσεων! Κάποιος νέος οδηγός είναι πιθανότερο, π.χ. μέσα στο χρονικό διάστημα ενός μηνός να προκαλέσει ατύχημα παραβιάζοντας τον ΚΟΚ, από ένα πιο έμπειρο οδηγό, αυτό δεν σημαίνει όμως ότι η πλειοψηφία των νέων οδηγών υποπίπτει σε αντίστοιχες παραβάσεις.

Συνθήκες εξαγωγής αποτελέσματος

Αναγκαία καλούμε τη συνθήκη που πρέπει να υπάρχει απαραίτητα για να προκύψει ένα συγκεκριμένο αποτέλεσμα. Είναι, για παράδειγμα, αναγκαίο να παρακολουθήσει κάποιος μαθήματα αγγλικής γλώσσας για να μπορέσει να αποκτήσει έστω το πρώτο δίπλωμα. Η παρακολούθηση, όμως, απλά σχετικών μαθημάτων δεν αρκεί για την επίτευξη του σκοπού αυτού (χρειάζεται η επιτυχία σε κάποιες εξετάσεις κ.λπ). Ικανή αποκαλούμε μια συνθήκη όταν αυτή υποχρεωτικά μας οδηγεί στο επιθυμητό αποτέλεσμα, π.χ. για να στεφθεί πρωταθλήτρια μία ομάδα στο ποδόσφαιρο πρέπει να συμμετέχει στο σχετικό πρωτάθλημα (αναγκαία συνθήκη) αλλά και να συγκεντρώσει στο τέλος της χρονιάς τη μεγαλύτερη βαθμολογία από τις υπόλοιπες ομάδες (ικανή συνθήκη).

Μια συνθήκη που θεωρείται ταυτόχρονα αναγκαία και ικανή αποτελεί φυσικά πολύ σημαντικό εύρημα σε μια έρευνα, όμως, σπάνια αυτό εντοπίζεται σε κοινωνικές έρευνες (αντίθετα συχνά απαντάται στα μαθηματικά και γενικότερα στις θετικές επιστήμες).

1.4 Η Λογική της έρευνας (Παραγωγική - Επαγωγική)

Οι δύο ερευνητικές προσεγγίσεις που αναδεικνύουν τη λογική μιας μελέτης και από τις οποίες ο ειδικός καλείται να επιλέξει είναι η παραγωγική και η επαγωγική (Κυριαζόπουλος & Σαμαντά, 2011).

Ειδικότερα, η παραγωγική ερευνητική προσέγγιση δίνει τη δυνατότητα για μετάβαση από το «γενικό» στο «ειδικό», ενώ η επαγωγική ακολουθεί την αντίθετη κατεύθυνση, από το «ειδικό» προς το «γενικό». Στην πράξη αυτό σημαίνει ότι στις έρευνες που εφαρμόζεται η παραγωγική λογική αρχικά εξετάζεται η υπάρχουσα θεωρία και με βάση αυτήν αναπτύσσονται συγκεκριμένες ερευνητικές υποθέσεις προς έλεγχο. Ενώ στις επαγωγικές έρευνες γίνεται προσπάθεια τα εμπειρικά στοιχεία να ενταχθούν στα πλαίσια κάποιας θεωρίας.

Παράδειγμα: έστω ότι θέλουμε να ερευνήσουμε αν ένα φροντιστήριο είναι καλό για την προετοιμασία μαθητών στις πανελλήνιες εξετάσεις. Χρησιμοποιούμε παραγωγική προσέγγιση, αν επιχειρήσουμε να διαπιστώσουμε ότι λειτουργεί με ολιγομελή τμήματα, με έμπειρους καθηγητές, αν υπάρχει κεντρικός συντονισμός, αν διεξάγονται διαγωνίσματα κ.λπ., και ανάλογα να προβλέψουμε το επίπεδο εκπαίδευσης που προσφέρεται και κατά συνέπεια την ποιότητά του. Επαγωγική προσέγγιση εφαρμόζουμε εάν εξετάσουμε τις επιδόσεις μαθητών που πηγαίνουν σε διάφορα φροντιστήρια, και στηριζόμενοι σε αυτό το στοιχείο καταλήξουμε στο συμπέρασμα ότι κάποιο φροντιστήριο είναι καλύτερο από άλλο αφού οι μαθητές που φοίτησαν σε αυτό έβγαλαν καλύτερες επιδόσεις στις εξετάσεις κ.ο.κ.

1.5 Μονάδες Ανάλυσης

Παρά το γεγονός ότι συνήθως είναι εύκολα αναγνωρίσιμη η μονάδα ανάλυσης, δηλαδή η μονάδα (συνήθως κάποιο άτομο) που δίνει την πληροφορία, υπάρχουν περιπτώσεις που αυτή δεν είναι εύκολα αναγνωρίσιμη και καμιά φορά προκύπτουν λάθη λόγω επιλογής εσφαλμένης μονάδας ανάλυσης. Εκτός από άτομα σε μια

εμπειρική έρευνα μονάδες ανάλυσης μπορεί να είναι ομάδες (νοικοκυριά, κόμματα), οργανώσεις (επιχειρήσεις, νοσοκομεία), ανθρώπινα δημιουργήματα (βιβλία, πίνακες ζωγραφικής, τραγούδια, κ.λπ.). Παράδειγμα στην έρευνα για το επίπεδο φτώχειας ενός πληθυσμού έχει αποδειχτεί ότι «ένα νοικοκυριό» είναι καλύτερη μονάδα ανάλυσης από τυχόν εξέταση μεμονωμένων ατόμων.

Η ανεπιτυχής επιλογή μονάδων ανάλυσης καλείται οικολογικό σφάλμα (ecological fallacy) και συνεπάγεται φυσικά και λανθασμένα συμπεράσματα. Για παράδειγμα, ας εξετάσουμε την περίπτωση συμμετοχής της Ελλάδας στο διαγωνισμό της Eurovision με ένα ελληνόφωνο τραγούδι!. Ας θεωρήσουμε ότι θέλοντας να τεκμηριώσουμε την άποψη ότι η ελληνόφωνη μουσική έχει μεγαλύτερη απήχηση στους δυτικοευρωπαίους σε σχέση με τους ανατολικοευρωπαίους βλέπουμε τις βαθμολογίες που έδωσαν οι χώρες της ανατολικής Ευρώπης (Ρωσία, Πολωνία, κ.λπ) και οι χώρες της δυτικής (Αγγλία, Γερμανία, κ.λπ). Στην Αγγλία, Γαλλία, Γερμανία ζει όμως μεγάλος αριθμός Ελλήνων μεταναστών και φοιτητών οι οποίοι υποστήριζαν το τραγούδι. Έτσι το καλό αποτέλεσμα σε αυτές τις χώρες δεν οφείλεται στις προτιμήσεις του ιθαγενούς πληθυσμού. Το λάθος σ' αυτή την περίπτωση προήλθε γιατί μονάδα ανάλυσης έπρεπε να είναι ο ψηφοφόρος και όχι οι χώρες που συμμετείχαν (έτσι ώστε να λαμβάνεται υπόψη η ιθαγένεια του ψηφοφόρου και όχι ο τόπος διαμονής του).

Ο αναγωγισμός (reductionism) είναι μία μέθοδος που βοήθησε ιδιαίτερα την ανάπτυξη των επιστημών (η προσέγγιση ενός θέματος μέσω της κατανόησης των μερών που το απαρτίζουν), αλλά όπως και άλλες μέθοδοι έχει συγκεκριμένα όρια απόδοσης, τα οποία όταν τα ξεπερνάμε, λέμε ότι διαπράττουμε αναγωγικό σφάλμα, το οποίο είναι σύνθηρες όταν χρησιμοποιούνται ακατάλληλες μονάδες ανάλυσης. Σ' αυτή την περίπτωση επιχειρούμε να εξαγάγουμε κάποιο αποτέλεσμα ερμηνεύοντάς το βάσει ενός παράγοντα, που τον θεωρούμε ως πιο σημαντικό, χωρίς να υπολογίζουμε ότι η πολυπλοκότητα του φαινομένου απαιτεί άλλη προσέγγιση.

Αναγωγικό σφάλμα, π.χ. διαπράττουμε όταν αναλύοντας τη μακροβιότητα των κατοίκων μιας περιοχής της Κρήτης θεωρούμε ότι αυτή οφείλεται στη μεσογειακή διατροφή. Στην περίπτωση αυτή το συμπέρασμα μπορεί να είναι σωστό, εν μέρει, αλλά η ερμηνεία είναι ελλιπής: έχουμε απομονώσει έναν μόνο παράγοντα, γιατί η μακροβιότητα μπορεί να οφείλεται και σε πάρα πολλούς άλλους λόγους, όπως άθληση, καλή ιατρική περίθαλψη, έλλειψη άγχους, περιβαλλοντολογικές συνθήκες κ.ά.

Άλλο παράδειγμα έχουμε αν προβλέψουμε ότι η Ρεάλ Μαδρίτης θα γίνει πρωταθλήτρια Ευρώπης την επόμενη χρονιά, απλά επειδή είναι η ομάδα με το μεγαλύτερο προϋπολογισμό στην Ευρώπη.

1.6 Η χρονική διάσταση των ερευνών

Σημαντικός παράγοντας στη διεξαγωγή μιας έρευνας είναι ο χρόνος. Όπως είδαμε και στη νομοθετική εξήγηση η χρονική αλληλουχία των γεγονότων και των καταστάσεων καθορίζει αιτίες και αποτελέσματα. Οι ερευνητές διαχωρίζουν το σχεδιασμό της έρευνάς τους σε συγχρονικές και διαχρονικές μελέτες.

Συγχρονική είναι μια μελέτη που βασίζεται σε παρατηρήσεις που αντιπροσωπεύουν ένα συγκεκριμένο σημείο στο χρόνο και αφορούν σε μια συγκεκριμένη κατηγορία πληθυσμού. Παραδείγματα είναι όλες οι δημοσκοπήσεις που αφορούν στην πρόθεση ψήφου των κατοίκων μιας χώρας. Ακόμα συγχρονική μελέτη έχουμε όταν ένας ερευνητής θέλει να καταγράψει την αποδοχή ενός προϊόντος σε ένα πληθυσμό σε μια δεδομένη χρονική στιγμή κ.ο.κ.

Οι διαχρονικές μελέτες περιλαμβάνουν δεδομένα που αντλήθηκαν σε διαφορετικά χρονικά διαστήματα. Τις διαχρονικές έρευνες τις χωρίζουμε σε τρεις κατηγορίες.

- Η μελέτη τάσης (trend study): καταγράφει ένα δεδομένο χαρακτηριστικό κάποιου πληθυσμού σε διάφορες χρονικές περιόδους. Παράδειγμα τέτοιας μελέτης είναι η μελέτη του ποσοστού των αναλφάβητων στην Ελλάδα το 1920, 1940, 1960 και 1980.
- Η μελέτη κοόρτης (cohort study) παρατηρεί τις μεταβολές αντιλήψεων, συνηθειών, κ.λπ σε ένα συγκεκριμένο υποπληθυσμό, σε διάφορες χρονικές στιγμές. Παράδειγμα είναι η μελέτη της γνώμης των γεννηθέντων το 1950 ως προς το πολιτικό σύστημα της χώρας τους (συνήθως σε τέτοιες έρευνες μελετάται αν αλλάζει ο προσανατολισμός των ψηφοφόρων ανάλογα με την ηλικία, αν είναι αριστερός, δεξιός, κ.λπ). Έτσι μπορούμε να μετράμε τον πολιτικό προσανατολισμό των γεννηθέντων το 1950 ανά δεκαετία μετά το 1970.

- Η μελέτη πάνελ (panel study) αντλεί τα δεδομένα της από το ίδιο σύνολο ατόμων ανά τακτά χρονικά διαστήματα. Για παράδειγμα, μελετάμε τη γνώμη για τη θρησκεία των ίδιων ατόμων ανά δεκαπέντε χρόνια. Είναι σαφές ότι τα πιο τεκμηριωμένα αποτελέσματα θα τα εξάγουμε από μελέτες πάνελ. Όμως το σύνολο των ερωτηθέντων (πάνελ) είναι δύσκολο να διατηρηθεί το ίδιο με την πάροδο του χρόνου. Συγκεκριμένα οι συνηθέστερες απώλειες προέρχονται από την άρνηση των συμμετεχόντων να συνεχίσουν την έρευνα, καμιά φορά όμως, οφείλονται στη φυσική απώλεια των μονάδων ανάλυσης, ή αλλαγή τόπου κατοικίας τους, κ.λπ.

Οι ερευνητές μπορούν να εξαγάγουν κατά προσέγγιση συμπεράσματα για διαχρονικές διαδικασίες, ακόμα και όταν έχουν στη διάθεσή τους μόνο συγχρονικά δεδομένα. Πιο συγκεκριμένα υπονοούν διαδικασίες που εξελίσσονται με τη πάροδο του χρόνου, εξάγουν λογικά συμπεράσματα και ζητούν από τα άτομα να ανακαλέσουν παλαιότερες συμπεριφορές. Αυτές οι μελέτες μπορεί να χαρακτηριστούν ως «κατά προσέγγιση» διαχρονικές.

1.7 Δειγματοληψία

1.7.1 Βασικές έννοιες

Σε σύγχρονες έρευνες συχνά χρησιμοποιούνται μετρήσεις αντιπροσωπευτικές του πληθυσμού.

Ειδικά στις περιπτώσεις ερευνών όπου οι μονάδες ανάλυσης είναι άνθρωποι η απογραφή ή η μέτρηση των σχετικών παραμέτρων σε όλα τα άτομα του υπό μελέτη συνόλου (πληθυσμού), εφόσον αυτός είναι μεγάλος είναι κατά κανόνα αδύνατη ή πολύ δαπανηρή.

Καταλήγουμε λοιπόν στη λήψη αντιπροσωπευτικού δείγματος, το οποίο αν επιλεγεί σωστά, σύμφωνα με τους κανόνες της δειγματοληψίας, μπορεί να χρησιμοποιηθεί για να γίνουν αναφορές στον πληθυσμό. Το θέμα γενικά είναι αρκετά σημαντικό. Οι μέθοδοι που ακολουθούνται ποικίλλουν ανάλογα με τα χαρακτηριστικά, την κατανομή του πληθυσμού και το μέγεθος του δείγματος.

Μία από τις πρώτες εφαρμογές της δειγματοληψίας έγινε το 1786 από τον Pierre Simon Laplace, ο οποίος υπολόγισε τον πληθυσμό της Γαλλίας βάσει ενός δείγματος, καθώς και τις πιθανολογικές εκτιμήσεις τυχόν λάθους. Αξίζει να αναφερθεί μία από τις πρώτες δημοσκοπήσεις για την πρόθεση ψήφου σε εκλογές, η οποία πραγματοποιήθηκε στις ΗΠΑ το 1936, και κατέληξε σε λάθος συμπέρασμα (2.400.000 ψηφοφόροι συμμετείχαν σε αυτή). Τα αποτελέσματα της έρευνας έδειξαν τότε μεγάλο προβάδισμα του ρεπουμπλικάνου υποψηφίου Alf Landon, αλλά μετά τη διεξαγωγή των εκλογών πρόεδρος των ΗΠΑ διαπιστώθηκε ότι εξελέγη με πολύ άνετη νίκη ο δημοκρατικός υποψήφιος Franklin Roosevelt (τα ποσοστά ήταν 60,7% έναντι 39,3%). Το λάθος αυτό οφείλεται στο ότι η έρευνα πραγματοποιήθηκε τηλεφωνικά, αν και η χρήση του συγκεκριμένου μέσου επικοινωνίας δεν ήταν ευρέως διαδεδομένη την εποχή αυτή. Κατά συνέπεια έλαβαν μέρος στη δημοσκόπηση σχετικά ευκατάστατοι ψηφοφόροι, οι οποίοι προτιμούσαν τον ρεπουμπλικάνο υποψήφιο, η μεγάλη πλειοψηφία των πολιτών, όμως, που είχε χαμηλά εισοδήματα ψήφισε τον Roosevelt.

Από τη εποχή που στην στατιστική ανάλυση δεδομένων άρχισαν να χρησιμοποιούνται ηλεκτρονικοί υπολογιστές η δειγματοληψία αποτελεί απαραίτητο εργαλείο για ερευνητές όλων των επιστημών.

Διάσημα εγχειρίδια δειγματοληψίας έχουν γράψει οι Snedecor και Cochran (1956), Seber (1973), Begon (1979), Krebs (1999), Govindarajulu (1999) κ.ά. (βλ. βιβλιογραφικές αναφορές).

Ακολούθως θα ασχοληθούμε με τον καθορισμό του μεγέθους του δείγματος και με τα κυριότερα είδη δειγματοληψίας.

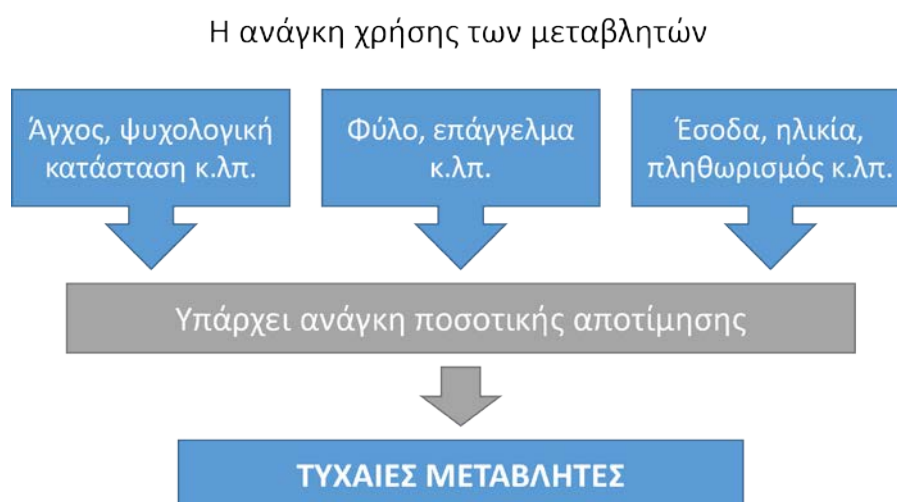
Στατιστικός πληθυσμός

Στατιστικός πληθυσμός ονομάζεται το σύνολο των μονάδων ανάλυσης, π.χ. άνθρωποι, ζώα, επιχειρήσεις κ.λπ., οι οποίες μελετώνται κατά την πραγματοποίηση της έρευνας. Το πλήθος των μονάδων ανάλυσης του δείγματος συμβολίζεται με n , ενώ του πληθυσμού με N . Ο πληθυσμός διακρίνεται σε άπειρο και πεπερασμένο ανάλογα με τη φύση της έρευνας.

Ο καθορισμός του πληθυσμού δεν είναι πάντοτε σαφής. Αν θέλουμε, π.χ. να συλλέξουμε δεδομένα εισοδήματος για την οικονομική κατάσταση φοιτητών ΑΕΙ και η έρευνα διεξάγεται στον χώρο του πανεπιστημίου επιδιώκουμε σ' αυτήν τη συμμετοχή ενεργών φοιτητών. Αν θέλουμε να μελετήσουμε δημογραφικά χαρακτηριστικά μακροχρόνια ανέργων, τότε πρέπει πρώτα από όλα να εξειδικεύσουμε τον όρο «μακροχρόνια άνεργος», εφόσον η χρήση του όρου σε αυτήν την περίπτωση ποικίλλει ανάλογα με τον πληθυσμό στον οποίο αναφέρεται. Ανάλογα, λοιπόν, με το σκοπό της έρευνας από τις μονάδες ανάλυσης του πληθυσμού, μέσω ενός δείγματος, επιλέγονται οι κατάλληλες προς μελέτη. Η όλη διαδικασία ονομάζεται δειγματοληψία. Για να επιλεγούν ορθά οι μονάδες ανάλυσης του δείγματος απαραίτητο είναι να κατανοηθεί η έννοια της μεταβλητής.

Μεταβλητές

Ο όρος μεταβλητή χρησιμοποιείται όταν θέλουμε να αποδώσουμε κάποιο χαρακτηριστικό ή ιδιότητα σε ένα πρόσωπο, αντικείμενο, κατάσταση, κ.λπ., το οποίο σκοπεύουμε να καταμετρήσουμε για την έρευνά μας. Για να μελετήσουμε καλύτερα τα αποτελέσματα ενός πειράματος τύχης, π.χ. αντιστοιχούμε σε κάθε απλό ενδεχόμενο έναν πραγματικό αριθμό. Για παράδειγμα, αν μελετούμε τα ενήλικα άτομα μιας πόλης ως προς το ύψος τους, τότε σε κάθε άτομο αντιστοιχούμε κάποιον αριθμό που δηλώνει το ύψος του. Με αυτό τον τρόπο έχουμε ορίσει τη μεταβλητή $X = \text{«ύψος ατόμου»}$. Αν ερευνούμε 50 οικογένειες με τρία παιδιά μιας πόλης, ορίζουμε τη μεταβλητή $Y = \text{«οικογένεια με τρία παιδιά»}$. Είναι εύκολα αντιληπτό ότι οι τιμές που παίρνουν οι μεταβλητές X, Y είναι διαφορετικού είδους. Πράγματι η μεταβλητή X παίρνει τιμές μεταξύ του συνόλου $=\{1.50, \dots, 2.12\}$ ενώ η Y παίρνει τιμές μεταξύ του συνόλου $=\{0, 1, 2, 3, \dots, 50\}$.



Εικόνα 1.1: Διαφορετικά είδη εννοιών που ποσοτικοποιούνται μέσω τυχαίων μεταβλητών

Από τις παραπάνω μεταβλητές γίνεται φανερό ότι υπάρχουν διαφορετικά είδη μεταβλητών ανάλογα με τον δειγματοχώρο και έτσι έχουμε τις εξής κατηγορίες:

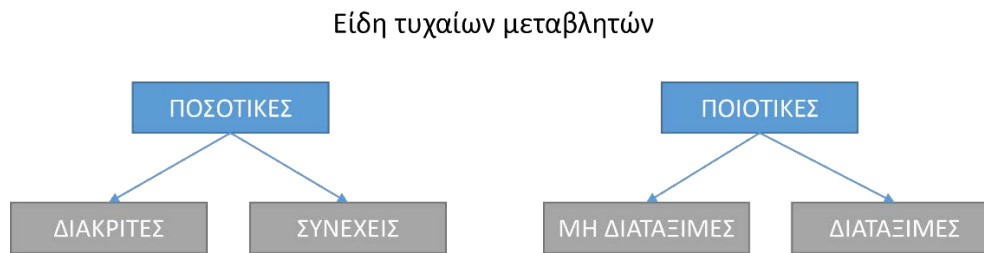
α) Ποσοτικές (quantitative)

Είναι οι μεταβλητές που μπορούν να μετρηθούν. Σ' αυτή την περίπτωση η συνάρτηση τυχαίας μεταβλητής δίνει ως αποτέλεσμα τη μέτρηση χωρίς τη μονάδα π.χ. 70 αντί για 70kg κ.ά., οπότε έχουμε δύο είδη μεταβλητών: αυτές που παίρνουν οποιαδήποτε τιμή σε ένα διάστημα πραγματικών αριθμών και ονομάζονται συνεχείς (ηλικία, βάρος, εισόδημα) και αυτές που παίρνουν συγκεκριμένες τιμές, συνήθως ακέραιες, και λέγονται διακριτές (αριθμός παιδιών οικογένειας κ.ά.).

β) Ποιοτικές (qualitative, categorical)

Είναι οι μεταβλητές που δεν μπορούν να μετρηθούν. Σε αυτή την περίπτωση η αντιστοίχιση τιμών του δειγματοχώρου με τους πραγματικούς αριθμούς είναι θέμα ορισμού και δεν έχει αριθμητική υπόσταση. Παράδειγμα αν εξετάζουμε το φύλο ενός ασθενή, μπορεί να γίνει η αντιστοίχιση «1» στον άντρα και «0» στη

γυναίκα ή το ανάποδο. Σε αυτή την περίπτωση διακρίνουμε πάλι δύο περιπτώσεις: τις μεταβλητές που εμπεριέχουν την έννοια της διάταξης στις τιμές που παίρνουν και ονομάζονται διατάξιμες (εξέλιξη νόσου, κάπνισμα) και στις μη διατάξιμες (φύλο, πάσχοντες ή μη από μια νόσο, κ.λπ.).



Εικόνα 1.2: Τα είδη των μεταβλητών

1.7.2 Καθορισμός μεγέθους δείγματος

Το πρώτο που ενδιαφέρει σε μία δειγματοληπτική έρευνα είναι το πλήθος των παρατηρήσεων, που πρέπει να έχει το δείγμα, ώστε τα αποτελέσματα να έχουν ένα συγκεκριμένο βαθμό αξιοπιστίας

Είναι αυτονόητο ότι ο βαθμός εγκυρότητας του δείγματος μεγαλώνει με την αύξηση του δείγματος. Αν και το κυριότερο μέτρο εγκυρότητας δεν είναι το μέγεθος του δείγματος, αλλά η αντιπροσωπευτικότητά του που προκύπτει από τη σωστή επιλογή της μεθόδου.

Για τον καθορισμό μεγέθους του δείγματος πρωταρχικό ρόλο παίζει η διασπορά των παρατηρήσεων του πληθυσμού. Αν θελήσουμε, π.χ. να εκτιμήσουμε το μέσο μισθό των δημοσίων υπαλλήλων της δευτεροβάθμιας εκπαίδευσης στην Ελλάδα, θα θεωρήσουμε ως πληθυσμό όλους τους καθηγητές δευτεροβάθμιας στα δημόσια σχολεία των οποίων οι μισθοί έχουν μια διακύμανση σ (όχι πολύ μεγάλη). Αν όμως το ζητούμενο είναι η εκτίμηση του μέσου μισθού των πτυχιούχων λογιστικής στην Ελλάδα, η διασπορά είναι πολύ μεγαλύτερη από πριν και κατά συνέπεια και το μέγεθος του δείγματος είναι πολύ μεγαλύτερο. Στην ακραία περίπτωση που η διασπορά είναι μηδέν αρκεί μία παρατήρηση, προκειμένου να εκτιμήσουμε το μέσο μισθό (π.χ. αν θέλουμε να εκτιμήσουμε το μέσο μισθό πρωτοετών στρατιωτικής σχολής).

Παρακάτω παρατίθενται βασικοί τύποι περιγραφικών στατιστικών:

Μέτρα θέσεως (περίπτωση πρωτογενών δεδομένων)

α) Μέσος αριθμητικός ή μέσος όρος ή μέση τιμή

Αν οι τιμές της μεταβλητής X είναι x_1, x_2, \dots, x_n

Ο μέσος αριθμητικός υπολογίζεται από τη σχέση:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Εξίσωση 1.1

β) Διάμεσος

Η διάμεσος δ_n παρατηρήσεων, x_1, x_2, \dots, x_n , της μεταβλητής X είναι η τιμή της παρατήρησης που βρίσκεται στο μέσο, όταν οι παρατηρήσεις διαταχθούν σε αύξουσα ή φθίνουσα σειρά. Διακρίνουμε δυο περιπτώσεις:

1. Το πλήθος n των παρατηρήσεων είναι άρτιος αριθμός. Στην περίπτωση αυτή η διάμεσος είναι το ημίαθροισμα των δύο μεσαίων παρατηρήσεων.
2. Το πλήθος n των παρατηρήσεων είναι περιττός αριθμός. Στην περίπτωση αυτή η διάμεσος είναι η μεσαία παρατήρηση.

Παρατήρηση: Η διάμεσος δεν επηρεάζεται από παρατηρήσεις οι οποίες βρίσκονται πολύ μακριά από τον κύριο όγκο των δεδομένων (outliers). Το αντίθετο συμβαίνει με τον μέσο αριθμητικό του οποίου η τιμή είναι ευαίσθητη σε τέτοιες παρατηρήσεις.

γ) Επικρατούσα τιμή

Όπως προκύπτει και από την ονομασία είναι η παρατήρηση με τη μεγαλύτερη συχνότητα.

δ) Πρώτο τεταρτημόριο Q_1

Το πρώτο τεταρτημόριο Q_1 (Q_1 quartile) διαιρεί τα δεδομένα σε δύο μέρη, έτσι ώστε, όταν τα δεδομένα είναι διατεταγμένα κατ' αύξουσα σειρά μεγέθους, το μέρος με τις μικρότερες παρατηρήσεις να αντιστοιχεί στο 25% των δεδομένων.

ε) Τρίτο τεταρτημόριο Q_3

Το τρίτο τεταρτημόριο Q_3 (Q_3 quartile) διαιρεί τα δεδομένα σε δύο μέρη, έτσι ώστε, όταν τα δεδομένα είναι διατεταγμένα κατ' αύξουσα σειρά μεγέθους, το μέρος με τις μεγαλύτερες παρατηρήσεις να αντιστοιχεί στο 25% των δεδομένων.

Μέτρα Διασποράς (πρωτογενών δεδομένων)

α) Εύρος

Το εύρος R (Range) είναι το απλούστερο από όλα τα μέτρα διασποράς και ορίζεται ως η διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής ενός συνόλου στατιστικών παρατηρήσεων:

$$R = x_{\max} - x_{\min}$$

Εξίσωση 1.2

β) Ενδοτεταρτημοριακό εύρος

Μέτρο διασποράς είναι και το ενδοτεταρτημοριακό εύρος IR που ορίζεται από τον τύπο:

$$IR = Q_3 - Q_1$$

γ) Διακύμανση και Τυπική απόκλιση

Αν οι τιμές της μεταβλητής X είναι x_1, x_2, \dots, x_n , η διακύμανση (Variance) υπολογίζεται από τη σχέση:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Εξίσωση 1.3

Η διακύμανση εκφράζεται σε μονάδες που αντιστοιχούν στο τετράγωνο των μονάδων της μεταβλητής. Η ανάγκη για ένα μέτρο διασποράς που να εκφράζεται στις ίδιες μονάδες με τις αρχικές (ούτως ώστε να μπορεί να συνεκτιμάται σε συνδυασμό και με τη μέση τιμή) οδήγησε στη χρησιμοποίηση της τετραγωνικής ρίζας της διακύμανσης, η οποία ονομάζεται τυπική απόκλιση:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Εξίσωση 1.4

δ) Συντελεστής μεταβλητότητας (coefficient of variation)

Ας θεωρήσουμε τους μηνιαίους μισθούς, σε ευρώ, πέντε υπαλλήλων δύο εταιρειών Α και Β. Για την Α έχουμε τους μισθούς 10.100, 10.050, 10.020, 10.000, 10.010 και για τη Β έχουμε τους μισθούς 1.100, 1.050, 1.020, 1.000, 1.010. Παρατηρούμε ότι και οι δύο εταιρείες έχουν ίδια μέτρα διασποράς (τυπική απόκλιση, εύρος, κ.λπ.). Παρόλα αυτά αν κάποιος υπάλληλος της πρώτης εταιρείας υποστεί μείωση μισθού 1.000 ευρώ, τότε αυτό θα έχει γι' αυτόν πολύ μικρότερες συνέπειες από ό,τι μια αντίστοιχη μείωση στο μισθό ενός υπαλλήλου της εταιρείας Β.

Από το παραπάνω παράδειγμα φαίνεται η ανάγκη να οριστεί ένα καινούργιο μέτρο, το οποίο δεν θα αντικατοπτρίζει μόνο τη διασπορά των δεδομένων, αλλά και τις επιπτώσεις που έχει αυτή η διασπορά στην πειραματική μονάδα. Το μέτρο αυτό συμβολίζεται με CV, ονομάζεται συντελεστής μεταβλητότητας και δίνεται από το λόγο της τυπικής απόκλισης (s) και του αριθμητικού μέσου:

$$CV = \frac{s}{\bar{x}}$$

Εξίσωση 1.5

Εάν ο μέσος όρος είναι κοντά στο μηδέν ή πολύ μεγάλος, τότε ο συντελεστής μεταβλητότητας καθίσταται αναξιόπιστος. Επίσης, θα πρέπει να ελεγχθεί προσεκτικά η αξιοπιστία του συντελεστή μεταβλητότητας, όταν μία κατανομή παρουσιάζει ασυμμετρία.

Ο συντελεστής μεταβλητότητας μπορεί επίσης να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης μίας εμπειρικής κατανομής συχνοτήτων, αν η μέση τιμή και η τυπική απόκλιση s μεταβάλλονται με τέτοιο τρόπο, ώστε ο CV να παραμένει σταθερός. Η αξία αυτής της μεθόδου είναι μεγαλύτερη, αν τα δεδομένα είναι τέτοια που δεν διευκολύνουν τον υπολογισμό των \bar{x} και s.

Κριτήριο καθορισμού μεγέθους δείγματος με βάση το τυπικό σφάλμα

Έστω τυχαία μεταβλητή X (π.χ. τα εισοδήματα καθηγητών δευτεροβάθμιας) προερχόμενη από πληθυσμό με μέση τιμή μ και διασπορά σ^2 . Αν πάρουμε ένα δείγμα μεγέθους n θα εκτιμήσουμε το μέσο εισόδημα του πληθυσμού από το δειγματικό μέσο:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Εξίσωση 1.6

Η διακύμανση του δειγματικού μέσου ονομάζεται τυπικό σφάλμα και δίνεται από τον τύπο:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Εξίσωση 1.7

Αρα για δεδομένο τυπικό σφάλμα και γνωστή διασπορά του πληθυσμού μπορούμε να εκτιμήσουμε το μέγεθος του δείγματος από τον τύπο:

$$n = \frac{\sigma^2}{\sigma_{\bar{x}}^2}$$

Εξίσωση 1.8

Εφόσον το τυπικό σφάλμα είναι ένα μέτρο διασποράς, είναι προφανές ότι όσο μικρότερες τιμές έχει, τόσο και πιο αντιπροσωπευτική είναι η τιμή της δειγματικής εκτίμησης.

Το πρόβλημα της εύρεσης της διασποράς σ^2 του πληθυσμού συνήθως αντιμετωπίζεται με εκτίμηση της σ^2 από μικρό στοιχειώδες δείγμα. Οπότε ο τύπος γίνεται:

$$n = \frac{s^2}{s_{\bar{x}}^2}$$

Εξίσωση 1.9

Ο τύπος αυτός δίνει το μέγεθος για προκαθορισμένο δείγμα και για προκαθορισμένο τυπικό σφάλμα, μειονεκτεί όμως στο ότι δε λαμβάνει υπόψη τη μέση τιμή της μεταβλητής, που θέλουμε να εκτιμήσουμε. Έτσι μπορεί να έχουμε ίδια προκαθορισμένη τιμή τυπικού σφάλματος, π.χ. $\sigma_{\bar{x}} = 1$ για μία μεταβλητή με μέση τιμή $\mu = 1.000.000$, και για μία με μέση τιμή $\mu = 1$ κάτι που προφανώς δεν είναι σωστό.

Κριτήριο καθορισμού μεγέθους δείγματος με βάση τον συντελεστή παραλλακτικότητας

Με τον συντελεστή παραλλακτικότητας:

$$C = \frac{\sigma_{\bar{x}}}{\mu}$$

Εξίσωση 1.10

Υπάρχει η δυνατότητα να αντιμετωπιστεί το μειονέκτημα του τυπικού σφάλματος, που προαναφέρθηκε, γιατί λαμβάνεται υπόψη στην τιμή του τυπικού σφάλματος και η μέση τιμή του πληθυσμού.

Οπότε για συγκεκριμένη τιμή του συντελεστή παραλλακτικότητας C έχουμε:

$$C = \frac{\sigma_{\bar{x}}}{\mu} \Leftrightarrow C = \frac{\sigma}{\mu \cdot \sqrt{n}} \Leftrightarrow n = \frac{\sigma^2}{\mu^2 \cdot C^2}$$

Εξίσωση 1.11

Η μέση τιμή αλλά και η διακύμανση του πληθυσμού συνήθως δεν είναι γνωστές και εκτιμώνται από στοιχειώδες δείγμα. Σ' αυτή την περίπτωση έχουμε τον τύπο:

$$n = \frac{s^2}{\bar{x}^2 \cdot C^2}$$

Εξίσωση 1.12

Κριτήριο καθορισμού μεγέθους δείγματος με βάση το διάστημα εμπιστοσύνης

Στην περίπτωση που θέλουμε να καθορίσουμε μέγεθος δείγματος για συγκεκριμένο διάστημα εμπιστοσύνης, π.χ. θέλουμε να εκτιμήσουμε το ποσοστό ενός κόμματος στις εκλογές, ενώ το εύρος του διαστήματος δεν θέλουμε να ξεπεράσει το 3%. Το διάστημα εμπιστοσύνης όταν είναι γνωστή η διασπορά του πληθυσμού είναι γνωστό ότι δίνεται από τον τύπο:

$$\left[\bar{x} - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

Εξίσωση 1.13

Άρα το μήκος του διαστήματος είναι (κατά το ήμισυ):

$$d = Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Εξίσωση 1.14

Από τη σχέση αυτή προκύπτει ότι ο τύπος για το μέγεθος του δείγματος είναι:

$$n = Z_{1-\alpha/2}^2 \cdot \frac{\sigma^2}{d^2}$$

Εξίσωση 1.15

Το πρόβλημα της εύρεσης της διακύμανσης σ^2 του πληθυσμού συνήθως αντιμετωπίζεται με εκτίμηση της s^2 από μικρό στοιχειώδες δείγμα. Οπότε ο τύπος γίνεται:

$$d = Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Εξίσωση 1.16

Για μικρά δείγματα θα χρησιμοποιείται η κατανομή Student αντί για την κανονική άρα το μέγεθος του δείγματος είναι:

$$n = \left(t_{n-1, 1-\alpha/2} \cdot \frac{s}{d} \right)^2$$

Εξίσωση 1.17

Στην περίπτωση αυτή ο υπολογισμός είναι πιο σύνθετος, γιατί το μέγεθος του δείγματος περιέχεται και στην κατανομή. Αντί για σταθερό μήκος d μπορεί να δίνεται το διάστημα ως ποσοστό της παραμέτρου της μέσης τιμής, π.χ. 0,1μ ή γενικά $k\mu$ οπότε το μέγεθος του δείγματος είναι:

$$n = \left(Z_{1-\alpha/2} \cdot \frac{\sigma}{k \cdot \mu} \right)^2$$

Εξίσωση 1.18

Για άγνωστη διακύμανση πληθυσμού και μεγάλο δείγμα έχουμε:

$$n = \left(Z_{1-\alpha/2} \cdot \frac{s}{k \cdot \bar{x}} \right)^2,$$

Εξίσωση 1.19

Και για μικρό δείγμα έχουμε:

$$n = \left(t_{n-1, 1-\alpha/2} \cdot \frac{s}{k \cdot \bar{x}} \right)^2,$$

Εξίσωση 1.20

Παράδειγμα 1

Έστω ότι θέλουμε να υπολογίσουμε το μέγεθος του δείγματος με πιθανότητα 95% ($\alpha = 0,05$) προκειμένου το διάστημα εμπιστοσύνης να περιέχει με ακρίβεια 10% το μέσο ετήσιο εισόδημα των εκπαιδευτικών πρωτοβάθμιας και δευτεροβάθμιας εκπαίδευσης στην Ελλάδα (δηλαδή με εύρος διαστήματος $\pm 0,10\mu$ και $k = 0,10$). Από μικρό δείγμα υπολογίστηκαν:

$$\bar{x} = 16.500$$

και

$$s^2 = 40.000$$

Να επαναληφθεί η διαδικασία προκειμένου να εκτιμηθεί το μέγεθος, αν θέλουμε το εύρος της εκτίμησης να μην ξεπερνά τα 200 ευρώ με πιθανότητα 95%.

Μέγεθος Δείγματος	Βαθμοί Ελευθερίας	$t_{n-1,0.975}$	$\sqrt{n_1} = t_{n-1,1-\alpha/2} \cdot \frac{s}{k \cdot \mu}$	n_1	$\sqrt{n_2} = t_{n-1,1-\alpha/2} \cdot \frac{s}{d}$	n_2
2	1	12,706	30,80242	948,7893	25,412	645,7697
3	2	4,303	10,43152	108,8165	8,606	74,06324
4	3	3,182	7,713939	59,50486	6,364	40,5005
5	4	2,776	6,729697	45,28882	5,552	30,8247
6	5	2,571	6,232727	38,84689	5,142	26,44016
7	6	2,447	5,932121	35,19006	4,894	23,95124
8	7	2,365	5,733333	32,87111	4,73	22,3729
9	8	2,306	5,590303	31,25149	4,612	21,27054
10	9	2,262	5,483636	30,07027	4,524	20,46658
11	10	2,228	5,401212	29,17309	4,456	19,85594
12	11	2,201	5,335758	28,47031	4,402	19,3776
13	12	2,179	5,282424	27,90401	4,358	18,99216
14	13	2,160	5,236364	27,4195	4,32	18,6624
15	14	2,145	5,2	27,04	4,29	18,4041
16	15	2,131	5,166061	26,68818	4,262	18,16464
17	16	2,12	5,139394	26,41337	4,24	17,9776
18	17	2,11	5,115152	26,16478	4,22	17,8084
19	18	2,101	5,093333	25,94204	4,202	17,6568
20	19	2,093	5,073939	25,74486	4,186	17,5226
21	20	2,086	5,05697	25,57294	4,172	17,40558
22	21	2,08	5,042424	25,42604	4,16	17,3056
23	22	2,074	5,027879	25,27957	4,148	17,2059
24	23	2,069	5,015758	25,15782	4,138	17,12304
25	24	2,064	5,003636	25,03638	4,128	17,04038
26	25	2,06	4,993939	24,93943	4,12	16,9744
27	26	2,056	4,984242	24,84267	4,112	16,90854
28	27	2,052	4,974545	24,7461	4,104	16,84282
29	28	2,048	4,964848	24,64972	4,096	16,77722
30	29	2,045	4,957576	24,57756	4,09	16,7281

Πίνακας 1.1: Τιμές του n για τις παραμέτρους του παραδείγματος

Παρατηρούμε ότι με το κριτήριο του ποσοστού επί της μέσης τιμής το ζητούμενο δείγμα είναι 6 (όπου υπάρχει η μικρότερη απόσταση του αριθμού της πρώτης και της τέταρτης στήλης (6-6,232727), ενώ με το δεύτερο κριτήριο 5 (όπου η αντίστοιχη απόσταση είναι 5-5, 552). Παρατηρούμε ότι και στις δύο περιπτώσεις το μέγεθος δείγματος είναι μικρό γιατί η τυπική απόκλιση $s=200$ θεωρείται σχετικά μικρή σε σχέση με τη μέση τιμή, αλλά και τα κριτήρια δεν ήταν αυστηρά (ο συντελεστής $k = 0,1$ είναι μεγάλος, όπως και το εύρος $d = 200$).

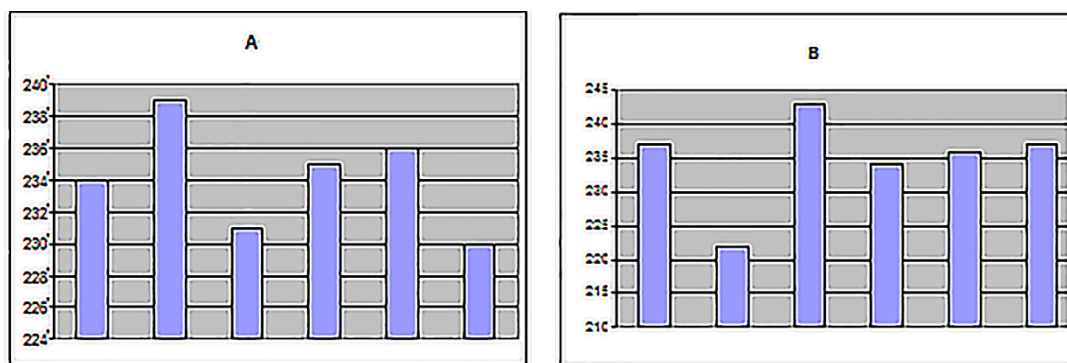
Παράδειγμα 2

Καταγράφεται ο χρόνος επίλυσης ενός προβλήματος από ηλεκτρονικούς υπολογιστές με τα ίδια χαρακτηριστικά προκειμένου να βρεθεί ο χρόνος επίλυσης του προβλήματος. Εφαρμόζοντας το ίδιο πρόβλημα παίρνουμε πέντε προκαταρκτικά δείγματα από έξι υπολογιστές. Να υπολογιστεί το μέγεθος των δειγμάτων που προκύπτει από τις παραπάνω μεθόδους και να σχολιαστούν τα αποτελέσματα.

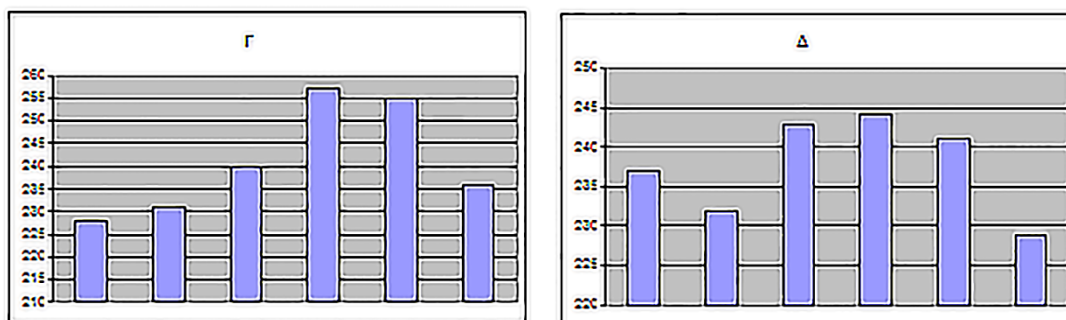
A/A	X ₁	X ₂	X ₃	X ₄	X ₅
1	234	237	228	237	222
2	239	222	231	232	234
3	231	243	240	243	218
4	235	234	257	244	245
5	236	236	255	241	221
6	230	237	236	229	222

Πίνακας 1.2: Δεδομένα 2^ο παραδείγματος

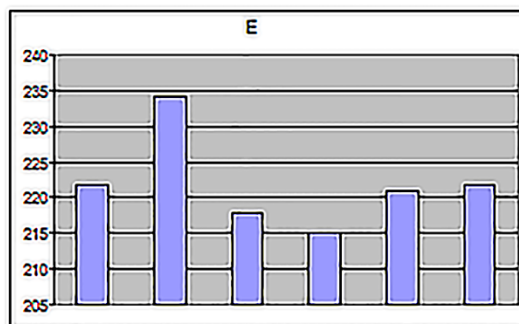
Για τις έρευνες αυτές προκύπτουν τα παρακάτω διαγράμματα:



Εικόνα 1.3: Διαγράμματα με δεδομένα το χρόνο επίλυσης (A και B)



Εικόνα 1.4: Διαγράμματα με δεδομένα το χρόνο επίλυσης (Gamma και Delta)



Εικόνα 1.5: Διάγραμμα με δεδομένα το χρόνο επίλυσης (E)

Καθώς και τα εξής:

	A	B	Γ	Δ	E
\bar{x}	234,1667	234,8333	241,1667	237,6667	222,0000
σ	3,311596	6,968979	12,22157	6,121002	6,480741

Πίνακας 1.3: Μέσες τιμές και τυπικές αποκλίσεις των δεδομένων

Θα θεωρήσουμε αρχικά την τιμή του τυπικού σφάλματος ως:

$$s_{\bar{x}} = 2$$

Εξίσωση 1.21

Οπότε τα συγκεκριμένα πέντε ζητούμενα μεγέθη δείγματος βρίσκονται από τον τύπο:

$$n = \frac{s^2}{s_{\bar{x}}^2}$$

Εξίσωση 1.22

n_A	n_B	n_Γ	n_Δ	n_E
2,7416667	12,14167	37,34167	9,366667	10,5000

Πίνακας 1.4: Πίνακας που έχουν καταγραφεί τα ζητούμενα μεγέθη του δείγματος

Στη συνέχεια θα υπολογίσουμε την τιμή του συντελεστή παραλλακτικότητας $C = 0,01$. Οπότε τα πέντε ζητούμενα μεγέθη δείγματος βρίσκονται από τον τύπο:

$$n = \frac{s^2}{\bar{x}^2 \cdot C^2}$$

Εξίσωση 1.23

και καταλήγουμε στον Πίνακα 5:

n_A	n_B	n_Γ	n_Δ	n_E
1,9999747	8,806814	25,68143	6,632977	8,522036

Πίνακας 1.5: Πίνακας που έχουν καταγραφεί τα ζητούμενα μεγέθη του δείγματος

Ακολούθως θα δεχτούμε ως δεδομένο το μήκος του διαστήματος εμπιστοσύνης ($d = 5$), οπότε τα ζητούμενα μεγέθη δείγματος βρίσκονται από τον τύπο:

$$n = Z_{1-\alpha/2}^2 \cdot \frac{\sigma^2}{d^2}$$

Εξίσωση 1.24

και καταλήγουμε στον Πίνακα 6:

n_A	n_B	n_Γ	n_Δ	n_E
1,6851819	7,462948	22,95228	5,757278	6,453888

Πίνακας 1.6: Πίνακας που έχουν καταγραφεί τα ζητούμενα μεγέθη του δείγματος

Τέλος θα πάρουμε ως δεδομένο και το μήκος του διαστήματος εμπιστοσύνης (2% της μέσης τιμής, δηλαδή $k = 0,02$), οπότε τα ζητούμενα μεγέθη δείγματος βρίσκονται από τον τύπο:

$$n = \left(Z_{1-\alpha/2} \cdot \frac{s}{k \cdot \bar{x}} \right)^2$$

Εξίσωση 1.25

και καταλήγουμε στον Πίνακα 7:

n_A	n_B	n_Γ	n_Δ	n_E
1,9207757	8,458064	24,66444	6,370311	8,184563

Πίνακας 1.7: Πίνακας που έχουν καταγραφεί τα ζητούμενα μεγέθη του δείγματος

Συνοψίζοντας καταλήγουμε στον Πίνακα 8:

	Πίνακας 4	Πίνακας 5	Πίνακας 6	Πίνακας 7
n_A	2,7416667	1,9999747	1,6851819	1,9207757
n_B	12,14167	8,806814	7,462948	8,458064
n_Γ	37,34167	25,68143	22,95228	24,66444
n_Δ	9,366667	6,632977	5,757278	6,370311
n_E	10,5000	8,522036	6,453888	8,184563

Πίνακας 1.8: Πίνακας που έχουν καταγραφεί τα ζητούμενα μεγέθη του δείγματος

1.7.3 Τα είδη δειγματοληψίας

1.7.3.1 Μη πιθανοτική δειγματοληψία

Όταν η εξαγωγή δείγματος βασίζεται σε τεχνικές κατά τις οποίες δε χρησιμοποιούνται οι νόμοι των πιθανοτήτων, τότε η όλη διαδικασία καλείται μη πιθανοτική δειγματοληψία. Συνήθως είναι ευκαιριακές διαδικασίες και εφαρμόζονται για πιλοτικές έρευνες και όχι για έρευνες επιστημονικού κύρους. Σ' αυτή την περίπτωση συνήθως εφαρμόζονται τρία δειγματοληπτικά σχέδια:

- **Δειγματοληψία ευκαιρίας (convenience sampling):** Είναι η συλλογή όσο το δυνατό μεγαλύτερου δείγματος, κατά τη διάρκειά της συγκεντρώνουμε όλες τις παρατηρήσεις στις

οποιές έχουμε εύκολη πρόσβαση. Για παράδειγμα, έστω ότι ένας ερευνητής ενδιαφέρεται να συλλέξει πληροφορίες για την ικανοποίηση των πελατών από ένα κατάστημα. Εφαρμογή ευκαιριακής δειγματοληψίας θα έχουμε αν ο ερευνητής πάει μια μέρα στο κατάστημα και δίνει ερωτηματολόγια στους πελάτες που βρίσκονται στο κατάστημα. Άλλο παράδειγμα είναι τα ρεπορτάζ στην τηλεόραση που ασχολούνται με την άποψη του κόσμου για τις τιμές των προϊόντων στην οδό Ερμού, κ.λπ. Προφανώς ο συγκεκριμένος τρόπος συλλογής του δείγματος δεν αντιπροσωπεύει τον πληθυσμό και δεν έχει επιστημονική εγκυρότητα. Η χρήση του συγκεκριμένου σχεδίου γίνεται κυρίως για πιλοτικές έρευνες και όχι για εξαγωγή συμπερασμάτων. Πρέπει να τονιστεί ότι συχνά παρατηρείται λανθασμένη σύγχυση της ευκαιριακής δειγματοληψίας με την απλή τυχαία δειγματοληψία που θα εξετάσουμε στην παράγραφο 1.7.3.2.

- **Δειγματοληψία κρίσεως ή σκόπιμη δειγματοληψία (judgement sampling):** Η μέθοδος αυτή θυμίζει σε μεγάλο βαθμό την ευκαιριακή, με τη διαφορά ότι η συλλογή των μονάδων γίνεται με κάποια επιλογή του ερευνητή, έτσι ώστε το δείγμα να είναι κατά τη γνώμη του πιο αντιπροσωπευτικό. Αν κατά τη δειγματοληψία ευκαιρίας για την ικανοποίηση πελατών, που αναφέραμε στο προηγούμενο παράδειγμα, ο ερευνητής κάνει μία πρόχειρη επιλογή σε ποιους δίνει το ερωτηματολόγιο, τότε η μέθοδος πλέον καλείται δειγματοληψία κρίσεως. Ο συγκεκριμένος ερευνητής έχοντας δώσει τα τρία πρώτα ερωτηματολόγια σε γυναίκες, το τέταρτο το δίνει σε άντρα, γιατί πιστεύει έτσι ότι το δείγμα θα είναι πιο αντιπροσωπευτικό. Όπως σε όλα τα σχέδια δειγματοληψίας και το συγκεκριμένο στερείται επιστημονικής εγκυρότητας και δεν συναντάται σε επιστημονικές δημοσιεύσεις, αλλά σε πιλοτικές έρευνες.
- **Δειγματοληψία της χιονοστιβάδας (snowball sampling):** Είναι μία μέθοδος κατά την οποία ο ερωτώμενος καλείται να βρει και να υποδείξει άλλους συμμετέχοντες στην έρευνα. Παρά το ότι η διαδικασία αυτή φαίνεται αρκετά παράδοξη, εντούτοις χρησιμοποιείται και σε επιστημονικές έρευνες. Ο λόγος χρήσης της μεθόδου είναι ότι σε πολλές περιπτώσεις ο πληθυσμός είναι αδύνατο να καταγραφεί, έτσι ώστε να υπάρχει δειγματοληπτικό πλαίσιο και να πραγματοποιηθεί η επιθυμητή δειγματοληψία, επίσης είναι δύσκολη η ανεύρεση των μονάδων ανάλυσης. Παράδειγμα έστω ότι θέλουμε την πραγματοποίηση έρευνας για ζητήματα που αφορούν στους λαθρομετανάστες. Προφανώς η πρόσβαση σε αυτούς τους ανθρώπους είναι πολύ δύσκολη και τη διευκολύνει σε πολύ μεγάλο βαθμό η ύπαρξη ενός λαθρομετανάστη, ο οποίος θα θέλει να βοηθήσει στην έρευνα μιλώντας σε αυτούς που γνωρίζει και προτρέποντάς τους να συμμετάσχουν στην έρευνα.
- **Ποσοτική δειγματοληψία (quota sampling):** Γίνεται συχνά σύγχυση του σχεδίου έρευνας που εφαρμόζεται σ' αυτή τη μέθοδο με τα αντίστοιχα της πιθανοτικής δειγματοληψίας. Στο ποσοτικό δειγματοληπτικό σχέδιο η συλλογή δείγματος πραγματοποιείται εξασφαλίζοντας συγκεκριμένα ποσοστά σε κάποιες παραμέτρους που κατά την κρίση του ερευνητή είναι οι σημαντικότερες. Προφανώς τα ποσοστά αυτά αντιπροσωπεύουν τον πληθυσμό. Έστω ότι το ζητούμενο είναι η εύρεση ενός δείγματος από τους φοιτητές του τμήματος Διοίκησης Επιχειρήσεων του ΑΕΙ Πειραιά ΤΤ. Από τα στοιχεία της γραμματείας βρέθηκε ότι το ποσοστό των φοιτητριών είναι 65% και των φοιτητών 35%. Ακόμα από τα στοιχεία της γραμματείας προκύπτει ότι το 18% είναι πρωτοετείς, το 16% δευτεροετείς, το 15% τριτοετείς, 15% τεταρτοετείς, και το υπόλοιπο 36% είναι φοιτητές επί πτυχίω.

Έτος	Φύλο		Σύνολο
	Άνδρες	Γυναίκες	
Πρωτοετείς			18%
Δευτεροετείς			16%
Τριτοετείς			15%
Τεταρτοετείς			15%
Επι πτυχίω			36%
Σύνολο	65%	35%	100%

Πίνακας 1.9: Καταγραφή μεταβλητών με τα αντίστοιχα ποσοστά

Έτος	Φύλο		Σύνολο
	Άνδρες	Γυναίκες	
Πρωτοετείς	$65\% \cdot 18\% = 11,7\%$	6,3%	18%
Δευτεροετείς	10,40%	5,6%	16%
Τριτοετείς	9,75%	5,25%	15%
Τεταρτοετείς	9,75%	5,25%	15%
Επι πτυχίω	23,4%	12,6%	36%
Σύνολο	65%	35%	100%

Πίνακας 1.10: Ποσοστά ανά κατηγορία

Έστω ότι το δείγμα αποτελείται από 1.000 φοιτητές-φοιτήτριες τότε θα πάρουμε τις παρακάτω παρατηρήσεις από κάθε κατηγορία:

Έτος	Φύλο		Σύνολο
	Άνδρες	Γυναίκες	
Πρωτοετείς	117	63	180
Δευτεροετείς	104	56	160
Τριτοετείς	97,5	52,5	150
Τεταρτοετείς	97,5	52,5	150
Επι πτυχίω	234	126	360
Σύνολο	650	350	1000

Πίνακας 1.11: Συχνότητες ανα κατηγορία

Η μέθοδος αυτή, αν και μπορεί σε πολλές περιπτώσεις να δώσει αντιπροσωπευτικά δείγματα, στερείται εγκυρότητας για το λόγο ότι η στάθμιση στηρίζεται σε περιορισμένο αριθμό παραγόντων (ενός, δύο ή τριών) που ο ερευνητής αυθαίρετα θεώρησε ότι είναι οι πιο σημαντικοί. Στην πράξη, όμως, ο ερευνητής δεν είναι σε θέση να προεξοφλήσει, αν κάποιος από τους παράγοντες είναι σημαντικός. Στο προηγούμενο παράδειγμα αν η έρευνα αφορά στο εισόδημα των φοιτητών, μπορεί οι παράγοντες φύλο και έτος να μην είναι σημαντικοί, αλλά να χρειάζεται να εξεταστούν άλλοι παράγοντες, όπως εάν οι φοιτητές προέρχονται από αστικό κέντρο, ή από επαρχία, αν παρακολουθούν τα μαθήματα ή όχι, κ.λπ.

1.7.3.2 Πιθανοτική δειγματοληψία

- **Απλή τυχαία δειγματοληψία (simple random sampling):** Από ένα πληθυσμό πεπερασμένου πλήθους μονάδων ανάλυσης εκλέγεται τυχαία δείγμα χωρίς επανάθεση. Ο όρος τυχαία δεν σημαίνει ότι εκλέγουμε στην τύχη όποιους θέλουμε από τον πληθυσμό κάποιων ατόμων. Η τυχαιότητα εξασφαλίζεται με χρήση τυχαίων αριθμών κατά την επιλογή των ατόμων ή με χρήση κάλπης.

Η επανάθεση εξασφαλίζει ότι οι πειραματικές μονάδες του δείγματος είναι διαφορετικές μεταξύ τους. Έτσι αν έχουμε ένα πληθυσμό N μελών και θέλουμε να υπολογίσουμε το πλήθος των δειγμάτων (n στοιχείων το καθένα), αυτό δίνεται από τον υπολογισμό N μελών ανά n στοιχείων το καθένα:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Εξίσωση 1.26

- **Συστηματική δειγματοληψία (systematic sampling):** Κατά τη συστηματική δειγματοληψία πρώτα από όλα αριθμούνται όλα τα μέλη του πληθυσμού. Αν ο πληθυσμός είναι N μονάδων ανάλυσης και θέλουμε να εξάγουμε δείγμα n από αυτές, πρώτα από όλα υπολογίζουμε το ηλίκο:

$$k = \frac{N}{n}$$

Εξίσωση 1.27

Παίρνουμε το ακέραιο μέρος του αριθμού αυτού: $\lambda = [k]$.

Μετά τον υπολογισμό του λ επιλέγεται τυχαία αριθμός ρ μεταξύ του 1 και του λ . Τα n μέλη του δείγματος θα έχουν τους εξής αύξοντες αριθμούς στο πλαίσιο τους $\rho, \rho+\lambda, \rho+2\lambda, \dots, \rho+(n-1)\lambda$.

Η διαδικασία αυτή είναι πολύ εύκολα εφαρμόσιμη στην περίπτωση που οι πειραματικές μονάδες είναι είδη αριθμημένες, π.χ. φάκελοι νοσοκομείου. Λάθος, που μπορεί να προκύψει από τη διαδικασία αυτή, είναι αν η αρίθμηση κρύβει και μία περιοδικότητα στα δεδομένα, π.χ. αν αναφερόμαστε σε εφημερίδες και το βήμα $\lambda=30$ θα προκύπτουν φύλλα εφημερίδων με ίδιες ημερομηνίες κάθε μήνα (π.χ. 20/6-20/7-19/8, κ.λπ).

- **Στρωματοποιημένη δειγματοληψία (stratified sampling):** Κατά τη στρωματοποιημένη δειγματοληψία ο πληθυσμός διαιρείται σε στρώματα (strata), από τα οποία, αφού καθοριστούν, εξάγονται δείγματα από το καθένα με τη μέθοδο της απλής τυχαίας δειγματοληψίας. Η μέθοδος έχει καλή εφαρμογή σε περιπτώσεις που ο πληθυσμός είναι ανομοιόμορφος. Ο καθορισμός των στρωμάτων γίνεται με κριτήριο τη διασπορά εντός των στρωμάτων. Συγκεκριμένα το επιδιωκόμενο είναι μέσα στα στρώματα να υπάρχει όσο το δυνατό μικρότερη διασπορά και ανάμεσα στα στρώματα όσο το δυνατό μεγαλύτερη. Ένα δείγμα που προκύπτει από στρωματοποιημένη δειγματοληψία μπορεί να είναι: είτε αναλογικό (ο αριθμός των μονάδων ανάλυσης που επιλέγονται να είναι ανάλογος του μεγέθους του δείγματος) είτε μη αναλογικό.
- **Δειγματοληψία κατά συστάδες (cluster sampling):** Στη δειγματοληψία κατά συστάδες ο πληθυσμός διαιρείται σε συστάδες (clusters), κάθε μία από αυτές θα αντιπροσωπεύει ένα νέο πληθυσμό. Οπότε η δειγματοληψία πραγματοποιείται σε δύο φάσεις: στην πρώτη με απλή τυχαία δειγματοληψία χωρίς επανάθεση επιλέγεται δείγμα από τις συστάδες και στη δεύτερη γίνεται απογραφή των συστάδων. Παρατηρούμε ότι η συγκεκριμένη μέθοδος μοιάζει αρκετά με την στρωματοποιημένη, παρόλα αυτά οι διαφορές στην εφαρμογή είναι αρκετά διακριτές. Συγκεκριμένα για τη μεγαλύτερη δυνατή αποτελεσματικότητα στη δειγματοληψία συστάδων πρέπει, σε αντίθεση με αυτό που συμβαίνει στη στρωματοποιημένη δειγματοληψία, μεταξύ των συστάδων να υπάρχει όσο το δυνατό μικρότερη διασπορά, ενώ μέσα στις συστάδες όσο το δυνατό μεγαλύτερη.

- **Διπλή δειγματοληψία (double sampling):** Κατά τη διαδικασία στατιστικού ελέγχου ποιότητας και αξιοπιστίας προϊόντων (quality control), κυρίως στον τομέα της βιομηχανικής παραγωγής, πολλές φορές έλεγχος ενός δείγματος σημαίνει και καταστροφή του προϊόντος. Παράδειγμα αν θέλουμε να ελέγξουμε την παραγωγή σοκολάτας, ως προς τη γεύση. Στην περίπτωση αυτή το δείγμα που εξάγεται είναι η πρώτη παρτίδα παραγόμενων προϊόντων. Οπότε με βάση τα αποτελέσματά της, βγάζουμε συμπεράσματα και για τα υπόλοιπα παραγόμενα προϊόντα.

1.8 Τα είδη των ερευνών

1.8.1 Έρευνα πεδίου

Η έρευνα πεδίου είναι ένα είδος έρευνας που δίνει τη δυνατότητα στον ερευνητή να παρατηρεί το αντικείμενο έρευνας στο περιβάλλον που πραγματοποιείται. Συνήθως δεν περιλαμβάνει ποσοτική ανάλυση. Απλά καταγράφονται παρατηρήσεις και μια ολοκληρωμένη οπτική για το αντικείμενο που εξετάζεται. Ο σκοπός ανάλογων ερευνών είναι συχνά διερευνητικός, χωρίς να αποκλείεται να είναι περιγραφικός ή ερμηνευτικός.

1.8.1.1 Παραδείγματα κατάλληλα για έρευνα πεδίου

Η έρευνα πεδίου είναι κατάλληλη, όπως ήδη αναφέρθηκε, για ζητήματα που η μελέτη τους είναι επιτυχής όταν διεξάγεται στις συνθήκες που αυτά γεννιούνται και αναπτύσσονται, ώστε να παρακολουθεί κάποιος εύκολα τις διαφοροποιήσεις που παρατηρούνται σε αυτά με την πάροδο του χρόνου.

Τέτοια ζητήματα είναι:

- Διερεύνηση των συνθηκών εργασίας που επικρατούν σε μια βιομηχανία, εργοστάσιο, κ.λπ., έστω για παράδειγμα η λειτουργία ενός ορυχείου χρυσού στη Ν. Αφρική. Παρά το γεγονός ότι πληροφορίες μπορεί να εξαχθούν με προσέγγιση των εργαζομένων με κάποια ερωτηματολόγια, είναι πολύ προτιμότερο ο ερευνητής να βρεθεί ο ίδιος επί τόπου και να παρακολουθήσει τις συνθήκες εργασίας, αλλά και τις πρακτικές της εταιρείας.
- Μελέτη της απήχησης στον κόσμο ενός νέου πολιτικού προσώπου. Ο ερευνητής παρακολουθεί τις κεντρικές ομιλίες του πολιτικού, καταγράφοντας τον παλμό, την απήχηση, τα συνθήματα και τα χαρακτηριστικά του κόσμου που συμμετέχουν.
- Δημοσιογραφική κάλυψη των επιπτώσεων ενός συμβάντος, π.χ. του τυφώνα Κατρίνα στις ΗΠΑ.
- Η παρατήρηση και καταγραφή των αντιδράσεων των πολιτών που στέκονται στην ουρά για κάποια δουλειά, π.χ. των ανέργων που περιμένουν για το σχετικό επίδομα του ΟΑΕΔ, κ.λπ.
- Η μελέτη της ζωής ανθρώπων πετυχημένων στη δουλειά τους, οι οποίοι έχουν μια πολύ καλή κοινωνική θέση και έχουν ξεκινήσει από χαμηλά. Κάποιοι από αυτούς ανέβηκαν με πολύ κόπο και κάποιοι απλά ήταν πολύ τυχεροί. Το θέμα μελέτης θα αφορούσε στις τυχόν αλλαγές στον χαρακτήρα τους και στη συμπεριφορά τους. Ο ερευνητής πρέπει να έρθει κοντά με τους ανθρώπους αυτούς και με ένα μικρό κύκλο ανθρώπων που τους γνωρίζουν από παλιά για να πάρει πληροφορίες και να βγάλει τα συμπεράσματά του.
- Μελέτη σχέσεων συνεργαζόμενων ατόμων, π.χ. των παικτών μιας αθλητικής ομάδας. Για να πραγματοποιηθεί τέτοιου είδους έρευνα πρέπει να υπάρχει παρατηρητής από μέσα, δηλαδή ο ερευνητής να μπει στην ομάδα, να ζήσει την ατμόσφαιρα στις προπονήσεις και το κλίμα στα αποδυτήρια.

1.8.1.2 Οι διάφοροι ρόλοι του ερευνητή

Οι ρόλοι είναι τρεις:

- **Συμμετέχων:** Σε αυτή την περίπτωση ο ερευνητής δρα ο ίδιος στο πεδίο που μελετά, με σκοπό να μην γίνει αντιληπτός από τα υποκείμενα μελέτης του. Φυσικά θα πρέπει να έχει ένα επίπεδο γνώσεων σχετικά με το θέμα που μελετά, γιατί αλλιώς κινδυνεύει να τον καταλάβουν. Βασικότερο μειονέκτημα της μεθόδου αυτής είναι ότι ο ερευνητής από τη στιγμή που συμμετέχει μπορεί να επηρεάσει το πεδίο μελέτης του εκφράζοντας κάποια άποψη. Είναι χαρακτηριστικό ότι εκτός από απόφαση που μπορεί να κληθεί να πάρει ο ερευνητής από το υπόλοιπο περιβάλλον για τις επόμενες δράσεις των ατόμων που συμμετέχουν, μπορεί και η απλή παρουσία του, η γνώμη του για διάφορα θέματα να επηρεάσει τη δράση της ομάδας. Άλλο μειονέκτημα είναι ότι μπορεί να επηρεαστεί και ο ίδιος από τα γεγονότα που παρακολουθεί χάνοντας την αντικειμενικότητά του. Σε κάθε περίπτωση γεννάται και σημαντικό θέμα δεοντολογίας, δεδομένου ότι για να υπάρχει εγκυρότητα και ο ερευνητής να έχει μία σωστή εικόνα δε θα πρέπει να γνωστοποιεί στο περιβάλλον που δρα την ιδιότητά του.
- **Παρατηρητής:** Είναι αυτός που απλά παρακολουθεί χωρίς να λαμβάνει μέρος σε καμία δραστηριότητα. Και εδώ υπάρχουν κίνδυνοι. Όπως στην περίπτωση διαδήλωσης με έντονα επεισόδια. Ο ερευνητής δεν μπορεί απλά να κάθεται σε μία άκρη με το μπλοκάκι του και να βλέπει τους διαδηλωτές (τα υποκείμενα της έρευνας) να πετάνε πέτρες στους αστυνομικούς, γιατί τότε θα γίνει αμέσως αντιληπτός και οι διαδηλωτές πιθανότατα να αλλάξουν συμπεριφορά απέναντί του και στη συγκεκριμένη περίπτωση μπορεί να διατρέξει και κάποιο κίνδυνο. Επίσης αν περιορίζεται μόνο στην απλή παρατήρηση, δεν θα είναι εύκολο ο ίδιος να αποκτήσει ολοκληρωμένη εικόνα για το αντικείμενο της μελέτης του.
- **Συμμετέχων-Παρατηρητής:** Αυτός ο ρόλος είναι μάλλον και ο πιο ορθός, γιατί επιτρέπει στον ερευνητή και να συμμετέχει και να παρατηρεί. Μπορεί να κατευθύνει τη μελέτη του, κατά την κρίση του, αλλά και να λαμβάνει υπόψη του γεγονότα, που τυχόν υποδεικνύουν κάποιες απρόβλεπτες διαστάσεις, που με τη σειρά τους τον οδηγούν σε διαφοροποιήσεις του αρχικού σχεδιασμού.

Σχέσεις με το αντικείμενο μελέτης

Υπάρχει κίνδυνος κατά τη συμμετοχή του σε μία έρευνα πεδίου ο ερευνητής να χάσει την αντικειμενικότητά του. Να βλέπει με συμπάθεια είτε τα φυσικά πρόσωπα μελέτης (π.χ. τους συμμετέχοντες σ' ένα κόμμα), είτε το αντικείμενο μελέτης (π.χ. την ιδεολογία του κόμματος). Η οπτική γωνία του ερευνητή, που βλέπει τα γεγονότα αποστασιοποιημένα και αντικειμενικά, καλείται ημική.

Πρέπει να τονιστεί ότι η αντικειμενική (ημική) οπτική δεν είναι πάντα το ζητούμενο! Σ' ένα θέμα, π.χ. όπως τα αισθήματα των κατοίκων ενός απολυταρχικού καθεστώτος σε χώρα της Αφρικής, που ο ηγέτης τους θεωρείται ότι έχει θρησκευτικές υπερεξουσίες, ίσως η επιλογή Ευρωπαίου αντικειμενικού ερευνητή δεν είναι καλή. Αντίθετα κάποιος θρησκευόμενος πολίτης αυτής της χώρας μπορεί να δώσει καλύτερη εικόνα. Η οπτική γωνία που έχει ο ερευνητής που υιοθετεί τις απόψεις του αντικειμένου μελέτης του λέγεται ητική.

1.8.1.3 Συνεντεύξεις

Όταν για τη βαθύτερη κατανόηση και ανάλυση κάποιου θέματος απαιτείται η γνώση της σκέψης, των κινήτρων, των συναισθημάτων, κ.λπ. των ερωτωμένων, τότε οι συνεντεύξεις είναι η πιο πρόσφορη μέθοδος για να αντληθούν οι απαραίτητες πληροφορίες.

Παρά το ότι κάποιος σχεδιασμός έχει προηγηθεί, στόχος είναι η συζήτηση να διεξαχθεί κατά το δυνατόν αβίαστα. Η συνέντευξη για να είναι επιτυχής στηρίζεται στην αλληλεπίδραση συνεντευκτική και συνεντευξιζόμενου. Σε πολλές περιπτώσεις μπορεί ένα μέρος της συνέντευξης να αναλωθεί σε κάτι που είναι εκτός σχεδιασμού, αν κριθεί ότι μπορεί να εξηγήσει στάσεις συμπεριφοράς, κ.λπ.

Ένα λογικό μέτρο για τον χρόνο ομιλίας συνεντευκτή και συνεντευξιαζόμενου είναι οι απαντήσεις από τον ερωτώμενο να καλύπτουν τουλάχιστο το 95% του συνολικού χρόνου της συνέντευξης και οι ερωτήσεις-τοποθετήσεις του συνεντευκτή το πολύ το 5% του χρόνου.

Συνοψίζοντας ο ερευνητής κατά τη διάρκεια της συνέντευξης θα πρέπει να προσέξει ότι:

- Ο τρόπος διατύπωσης μιας ερώτησης καθορίζει την απάντηση που θα αποσπάσει ο ερευνητής, ενώ από τις απαντήσεις σ' αυτές διαμορφώνονται οι επόμενες ερωτήσεις.
- Ο ερευνητής κατά τη διάρκεια μιας συνέντευξης θα πρέπει να μπορεί να ακούει, να σκέφτεται και να μιλά σχεδόν ταυτόχρονα.
- Πρέπει να είναι καλός ακροατής, δηλαδή πρέπει να κοιτάει και να ακούει με ενδιαφέρον τον ερωτώμενο και να του αφήνει την πρωτοβουλία σε περιπτώσεις σύντομων παύσεων. Δεν πρέπει να είναι βέβαια παθητικός και να μην μιλά, αλλά, όπως ήδη αναφέρθηκε, η ομιλία του να μην ξεπερνά συνολικά το 5% της διάρκειας της συνέντευξης.
- Θα πρέπει να καθοδηγεί τη συζήτηση προς την κατεύθυνση που επιθυμεί έμμεσα.
- Αν διακόπτει τον ερωτώμενο, μεταφράζεται ως έλλειψη ενδιαφέροντος. Η ενθάρρυνση του ερευνητή ως προς τον ερωτώμενο, για περαιτέρω ανάπτυξη του θέματος, αντιθέτως, δηλώνει το γνήσιο ενδιαφέρον του ερευνητή.
- Θα πρέπει να έχει προκαθορίσει συγκεκριμένα τα βασικά ζητήματα για να πετύχει ομαλές και λογικές μεταβάσεις από το ένα ζήτημα στο άλλο (Rubin, H. & Rubin, R., 1995).
- Ο ερευνητής δεν πραγματοποιεί αυθόρμητη συζήτηση, αλλά επίσημη. Αυτό σημαίνει ότι η επιθυμία να εμφανιστεί ο ίδιος ως ενδιαφέρον άτομο είναι αντιπαραγωγική. Πρέπει να κάνει τον συνεντευξιαζόμενο να νιώθει αυτός ως ενδιαφέρον άτομο, ακούγοντάς τον περισσότερο παρά μιλώντας ο ίδιος.

Τα στάδια της διαδικασίας συνέντευξης είναι:

- Θεματοποίηση: Καθορισμός σκοπού και στόχων της συνέντευξης.
- Σχεδιασμός: Διατύπωση διαδικασίας συνέντευξης.
- Συνέντευξη.
- Απομαγνητοφώνηση, δηλ. μεταγραφή σε κείμενο της συνέντευξης.
- Ανάλυση: Μελέτη των συλλεχθέντων πληροφοριών σύμφωνα με το σκοπό της έρευνας.
- Επαλήθευση: Έλεγχος αξιοπιστίας και της εγκυρότητας των συμπερασμάτων.
- Έκθεση: Δημοσίευση συμπερασμάτων έρευνας.

1.8.1.4 Ομάδες εστίασης

Οι έρευνες με ομάδες εστίασης είναι έρευνες με συχνή εφαρμογή στις επιστήμες Διοίκησης και Οικονομίας. Στις έρευνες αυτές ο ερευνητής είναι συντονιστής μιας συζήτησης που διεξάγεται με μία ομάδα εστίασης (focus group). Η τεχνική αυτή είναι ίδια μ' αυτή που χρησιμοποιεί ένας δημοσιογράφος που συντονίζει μια πολιτική εκπομπή με πολλούς καλεσμένους στην τηλεόραση.

Χρησιμοποιείται τακτικά σε έρευνες αγοράς για την αξιολόγηση προϊόντων και ειδών εμπορεύματος. Βασίζεται σε δομημένες, ημιδομημένες ή μη δομημένες συνεντεύξεις.

Συγκεντρώνονται 5-12 άτομα σε ιδιωτικό και άνετο περιβάλλον για να συζητήσουν ένα συγκεκριμένο ζήτημα. Για την επιλογή των ατόμων προηγείται δειγματοληψία, η οποία σπάνια είναι πιθανοτική (άλλωστε ο περιορισμένος αριθμός του δείγματος δεν διευκολύνει ώστε αυτό να είναι αντιπροσωπευτικό). Συχνά το δείγμα αποτελείται από άτομα που θεωρούνται κατάλληλα για την έρευνα, π.χ. είναι εν δυνάμει πελάτες ενός κινητού που θα βγει σε λίγο στην αγορά.

Ο ερευνητής θα πρέπει να δίνει το λόγο σε όλη την ομάδα για το ίδιο ερώτημα. Ο χρόνος θα πρέπει να μοιράζεται εξίσου, ώστε να μην μονοπωλούν κάποιοι τη συζήτηση. Ο ίδιος έχει επίσης την ευθύνη για μια εποικοδομητική και πολιτισμένη συζήτηση.

Παρακάτω παρατίθενται τα πλεονεκτήματα και τα μειονεκτήματα των ομάδων εστίασης σύμφωνα με τον Krueger (1988).

Πλεονεκτήματα

- Είναι μια κοινωνικά προσανατολισμένη ερευνητική μέθοδος που συλλαμβάνει δεδομένα της πραγματικής ζωής σε ένα κοινωνικό περιβάλλον.
- Είναι ευέλικτη.
- Έχει υψηλή εγκυρότητα.
- Προσφέρει άμεσα αποτελέσματα.
- Έχει μικρό κόστος.

Μειονεκτήματα

- Ο ερευνητής έχει μειωμένη ικανότητα ελέγχου σε σχέση με τις ατομικές συνεντεύξεις.
- Δυσκολότερη ανάλυση δεδομένων.
- Οι συντονιστές πρέπει να έχουν ειδικές δεξιότητες.
- Η αντιπαράθεση ερωτώμενων στην ομάδα εστίασης μπορεί να προκαλέσει προβλήματα.
- Δύσκολη συγκέντρωση ομάδων.
- Η συζήτηση πρέπει να διεξαχθεί σε πρόσφορο περιβάλλον.

Ενδιαφέρουσα είναι η άποψη του που υποστηρίζει ότι η έρευνα σε ομάδες εστίασης μπορεί να είναι ο κατάλληλος τρόπος για να γίνει μία σωστή και πλήρης σύνταξη ερωτηματολογίου δειγματοληπτικής έρευνας.

1.8.1.5 Δεοντολογία ποιοτικής έρευνας πεδίου

- Μία έρευνα πεδίου, για να έχει έγκυρο αποτέλεσμα, πρέπει να μην γίνεται αντιληπτή η παρουσία του ερευνητή, αυτό έχει, όμως, ως αποτέλεσμα ανυποψίαστοι άνθρωποι να γίνονται αντικείμενα μελέτης ερευνητικών σκοπών, ή ακόμα χειρότερα να καλλιεργούνται ανθρώπινες σχέσεις στρατηγικά σχεδιασμένες.
- Ελλοχεύει ο κίνδυνος, όπως και σε άλλες έρευνες, προσωπικά δεδομένα που εκμυστηρεύονται οι πληροφοριοδότες στον ερευνητή να βγουν στη δημοσιότητα.

1.8.2 Δειγματοληπτική έρευνα

Οι δειγματοληπτικές έρευνες διεξάγονται από ερευνητές για τη συλλογή δεδομένων με σκοπό είτε τη διερεύνηση, είτε την περιγραφή, είτε την ερμηνεία διαφόρων ζητημάτων. Πρέπει να σημειωθεί, όμως, ότι η δομή αυτών των ερευνών και τα τεχνικά στοιχεία θα αναφερθούν στο επόμενο κεφάλαιο. Το εργαλείο διεξαγωγής μιας έρευνας είναι το ερωτηματολόγιο. Τρόποι εξαγωγής των δεδομένων είναι: οι προσωπικές συνεντεύξεις μέσω ερωτηματολογίων και τηλεφώνου (με τη βοήθεια του υπολογιστή), και η ατομική συμπλήρωση ερωτηματολογίου που αποστέλλεται ταχυδρομικά ή διαδικτυακά. Θα γίνει αναλυτική αναφορά στη σύνταξη του ερωτηματολογίου με παραδείγματα λανθασμένης σύνταξης ερωτηματολογίου, αλλά και λανθασμένων ερωτήσεων στο δεύτερο κεφάλαιο. Ακόμα, σε δειγματοληπτικές έρευνες εφαρμόζεται δευτερογενής ανάλυση των δεδομένων, κατά την οποία κάποια δεδομένα που έχουν συλλεχθεί για μία έρευνα ξαναχρησιμοποιούνται σε κάποια άλλη μεταγενέστερη.

1.8.2.1 Ερωτηματολόγια

Αρχικά, τα ζητήματα που απασχολούν τους ερευνητές για τη διεξαγωγή δειγματοληπτικών ερευνών είναι πολλά και στη μεγάλη πλειοψηφία τους έχουν ως επίκεντρο τους ανθρώπους και τα ζητήματα που τους απασχολούν, επομένως, μονάδες ανάλυσης μιας δειγματοληπτικής έρευνας είναι συνήθως μεμονωμένα άτομα ή διάφορες ομάδες ατόμων.

Μια δειγματοληπτική έρευνα είναι η πιο κατάλληλη μέθοδος για τη συγκέντρωση στοιχείων από ένα μεγάλο πληθυσμό. Ακόμα, οι έρευνες αυτές παρέχουν τη μέτρηση της αντίληψης των απόψεων και κατευθύνσεων ενός πληθυσμού. Χαρακτηριστικό παράδειγμα τέτοιων ερευνών αποτελούν οι δημοσκοπήσεις.

1.8.2.2. Πλεονεκτήματα δειγματοληπτικών ερευνών

Οι δειγματοληπτικές έρευνες έχουν τα εξής πλεονεκτήματα:

1. Αφορούν σε ζητήματα μεγάλων πληθυσμών, τα οποία δύσκολα διερευνώνται με άλλες μεθόδους.
2. Λόγω της καταχώρισης των απαντήσεων στις κλειστές ερωτήσεις η επεξεργασία και η συμπερασματολογία είναι διαδικασίες σχετικά απλές.

1.8.2.3 Μειονεκτήματα δειγματοληπτικών ερευνών

Οι δειγματοληπτικές έρευνες έχουν, όμως, και μειονεκτήματα, τα οποία είναι:

1. Απώλεια πληροφορίας που οφείλεται στο «στρίμωγμα» δεδομένων σε γενικές κατηγορίες, για αυτό χαρακτηρίζονται και «άκαμπτες».
2. Το θέμα που μελετάται μπορεί να μην είναι μετρήσιμο μέσω ερωτηματολογίων.
3. Η ορθή διεξαγωγή τους απαιτεί πολύ χρόνο και χρήμα.
4. Αν γίνει λάθος στον αρχικό σχεδιασμό, π.χ. σύνταξη ερωτηματολογίου δεν μπορεί να διορθωθεί.

1.8.2.4 Δευτερογενής έρευνα

Η συσσώρευση καταγεγραμμένων γνώσεων επιτρέπει συχνά σε πολλούς επιστήμονες να αντλήσουν από αυτές κατάλληλα στοιχεία για τα δικά τους ενδιαφέροντα. Υπάρχουν περιπτώσεις που τα δεδομένα μιας έρευνας επαναχρησιμοποιούνται σε μία άλλη με εντελώς διαφορετικό προσανατολισμό, σκοπό και στόχους. Για παράδειγμα, επαναχρησιμοποιούνται απαντήσεις σε ερωτηματολόγια, απογραφικά στοιχεία και μετρήσεις πειραμάτων, κ.λπ. Τα πρώτα χρόνια συλλογής δευτερογενών δεδομένων τα δεδομένα συλλέγονταν σε έντυπη μορφή σε βιβλιοθήκες. Τώρα πλέον προφανώς φυλάσσονται σε ηλεκτρονικές διευθύνσεις.

Ακόμα υπάρχει πληθώρα ερευνητικών κέντρων που υπάρχει η δυνατότητα συλλογής στοιχείων όπως τα εξής:

1. Η Ελληνική Τράπεζα Κοινωνικών Δεδομένων (ΕΤΚΔ),
2. Η Ειδική Τράπεζα Πληροφοριών (ΕΤΠ),
3. Το Αρχείο Κοινωνικών Δεδομένων και Δεικτών (ΑΚΔΔ),
4. Το Περιβάλλον Διαχείρισης Κοινωνικών Δεδομένων (ΠΔΚΔ),
5. Ο Κόμβος Δευτερογενούς Επεξεργασίας (ΚΔΕ).

Ενδεικτικά τέτοια κέντρα στις ΗΠΑ είναι:

1. Το κέντρο Roper για την Έρευνα της κοινής γνώμης του Πανεπιστημίου του Κονέκτικατ,
2. Το Εθνικό Κέντρο Έρευνας της κοινής γνώμης του Πανεπιστημίου του Σικάγο.

Προφανές πλεονέκτημα της δευτερογενούς έρευνας είναι η οικονομία. Επίσης υπάρχει και η δυνατότητα της μετα-ανάλυσης όπου ο ερευνητής συγκεντρώνει πληροφορίες από παλαιότερες έρευνες για ένα συγκεκριμένο ζήτημα.

Βασικό μειονέκτημα είναι η εγκυρότητα, εφόσον οι ερευνητές συλλέγουν τα δεδομένα για ένα συγκεκριμένο σκοπό, και δεν είναι εύκολο τα δεδομένα αυτά να είναι κατάλληλα για άλλα ερευνητικά ενδιαφέροντα.

1.8.2.5 Δεοντολογία της δειγματοληπτικής έρευνας

Η μέθοδος αυτή δεν παρουσιάζει πολλά δεοντολογικά ζητήματα, γιατί ο ρόλος του ερευνητή είναι ξεκάθαρος και δεν αποκρύβεται από τον ερωτώμενο. Δεοντολογικά ζητήματα μπορούν να προκύψουν με τη διαχείριση των προσωπικών δεδομένων. Είναι χαρακτηριστικό ότι συχνά πραγματοποιούνται δημοσκοπήσεις, κατά τις οποίες χρησιμοποιείται κάλλη, ώστε ο ερωτώμενος να αποφασίζει ανεπηρέαστος και μετά ο ερευνητής συνδέει την «ψήφο» με τις υπόλοιπες ερωτήσεις εν αγνοία του ερωτώμενου.

1.8.3 Πειραματικοί σχεδιασμοί

Ο όρος πείραμα για τους περισσότερους ανθρώπους είναι ταυτισμένος με τις επιστήμες της φυσικής, της χημείας, της ιατρικής, κ.λπ. (Κίτσος, 1994). Οι πειραματιστές, όμως, πέρα από τις ελεγχόμενες συνθήκες εργαστηρίου έχουν τη δυνατότητα να πραγματοποιήσουν πειράματα στην κοινωνία.

1.8.3.1 Θέματα κατάλληλα για πειράματα

Στους πειραματικούς σχεδιασμούς ο σκοπός της έρευνας είναι συνήθως ερμηνευτικός και όχι περιγραφικός, αφού ο στόχος της έρευνας είναι η άσκηση κάποιας επίδρασης σε μία υπάρχουσα κατάσταση και η εξέταση των αποτελεσμάτων αυτής, εξετάζεται δηλ. αν σημειώθηκαν κάποιες αλλαγές στη συγκεκριμένη κατάσταση μετά την εφαρμογή της όποιας επίδρασης. Τα πειράματα ταιριάζουν περισσότερο σε ερευνητικά προγράμματα που αφορούν σε περιορισμένες και σαφείς έννοιες και προτάσεις. Παράδειγμα: Ας υποθέσουμε ότι θέλουμε να μειώσουμε τον βαθμό προκατάληψης έναντι των γυναικών με διευθυντική θέση μέσα στις επιχειρήσεις. Πολλοί θεωρούν ότι οι γυναίκες δεν μπορούν να υποστηρίξουν μια θέση με τόσες ευθύνες και απαιτήσεις, επειδή δεν έχουν τα απαραίτητα προσόντα και (για πολλούς) την απαραίτητη νοημοσύνη! Εάν κάνουμε γνωστή τη συμβολή τους σε βάθος χρόνου στην ανάπτυξη πολλών επιχειρήσεων, υποθέτουμε ότι το στοιχείο αυτό θα επιτρέψει τη μείωση της σχετικής προκατάληψης. Επομένως αποφασίζουμε να ελέγξουμε τη συγκεκριμένη υπόθεση σε πειραματικό επίπεδο.

Βήματα:

- 1) Ελέγχουμε μια ομάδα υποκειμένων προκειμένου να καθορίσουμε το επίπεδο προκατάληψης (η μέτρηση της προκατάληψης μπορεί να γίνει π.χ. με συμπλήρωση ερωτηματολογίου).
- 2) Με διάφορα μέσα (π.χ. ντοκιμαντέρ) δίνουμε πληροφορίες σχετικά με το θέμα που εξετάζουμε, παρέχονται δηλαδή στοιχεία που αποδεικνύουν την επιτυχημένη επαγγελματική δράση γνωστών γυναικών.
- 3) Μετά το ντοκιμαντέρ μετράμε εκ νέου το επίπεδο της σχετικής προκατάληψης ώστε να διαπιστώσουμε εάν το ντοκιμαντέρ συνέβαλε στη μείωση της προκατάληψης ή όχι.

1.8.3.2 Η διαδικασία του πειράματος

Η πειραματική διαδικασία έχει δύο στάδια τον προ-έλεγχο και το μετα-έλεγχο (μέτρηση πριν και μετά τη δράση των αποτελεσμάτων). Σ' αυτήν παίρνουν μέρος οι πειραματικές ομάδες και οι ομάδες ελέγχου. Πειραματική ομάδα είναι η ομάδα μονάδων ανάλυσης που υφίσταται το πείραμα-τη δράση. Ομάδα ελέγχου είναι η ομάδα μονάδων ανάλυσης που χωρίς να υφίσταται κάποια δράση καταμετρώνται οι μονάδες ανάλυσης προκειμένου να πραγματοποιείται σύγκριση με την πειραματική ομάδα.

Στο πείραμα θα χρησιμοποιηθούν δύο ειδών μεταβλητές: η εξαρτημένη, που μπορεί να είναι σύνολο ποσοτικών μεταβλητών και είναι το αποτέλεσμα του πειράματος και η ανεξάρτητη που είναι μία ποιοτική μεταβλητή δύο τιμών: α) «η μονάδα ανάλυσης έχει δεχθεί την επίδραση», β) «η μονάδα ανάλυσης δεν έχει δεχθεί την επίδραση». Στο παράδειγμα που προαναφέρθηκε η εξαρτημένη μεταβλητή είναι «η προκατάληψη κατά τις αποτελεσματικότητας των γυναικών σε διοικητικές θέσεις» (η οποία μπορεί να έχει μετρηθεί είτε με

συνέντευξη είτε με τη συμπλήρωση ερωτηματολογίου). Ενώ η ανεξάρτητη είναι η μεταβλητή με τις τιμές: «η μονάδα ανάλυσης έχει δει το ντοκιμαντέρ», «η μονάδα ανάλυσης δεν έχει δει το ντοκιμαντέρ».

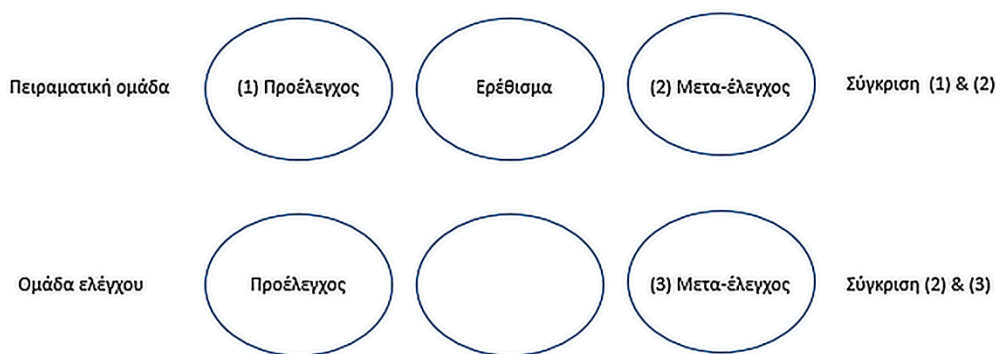
Πρέπει όμως να πούμε πως η ανεξάρτητη μεταβλητή σε ένα πείραμα μπορεί να γίνει εξαρτημένη μεταβλητή σε ένα άλλο πείραμα. Για παράδειγμα, η προκατάληψη στο προηγούμενο παράδειγμα είναι η εξαρτημένη μεταβλητή, αλλά σ' ένα άλλο πείραμα, που εξετάζει το αποτέλεσμα της προκατάληψης στην εκλογική συμπεριφορά, θα μπορούσε να είναι η ανεξάρτητη μεταβλητή.

1.8.3.3 Προ-έλεγχος και μετα-έλεγχος

Σε κάθε έρευνα το βασικό θέμα που ανακύπτει είναι η εγκυρότητάς της: το πρόβλημα αυτό επισημαίνεται κυρίως στο στάδιο του μετα-ελέγχου, γιατί στην πρώτη διεξαγωγή του ερωτηματολογίου οι ερωτώμενοι μπορεί (και πρέπει!) να αγνοούν το σκοπό του, ενώ στη δεύτερη διεξαγωγή είναι σχεδόν βέβαιο ότι θα αντιληφθούν το λόγο για τον οποίο διενεργείται η έρευνα και έτσι υπάρχει μεγάλη πιθανότητα να επηρεαστούν και να δώσουν άλλες απαντήσεις από αυτές που θα έδιναν αν ήταν ανεπηρέαστοι. Για παράδειγμα, κάποιος που συνειδητοποιεί ότι καταμετράται η προκατάληψή του κατά των γυναικών, το πιθανότερο είναι να δώσει απαντήσεις που να μην προδίδουν τέτοια προκατάληψη. Η μεροληψία στις απαντήσεις είναι το μεγάλο πρόβλημα που έχουν να αντιμετωπίσουν οι ερευνητές σε πειραματικούς σχεδιασμούς οικονομικών και κοινωνικών επιστημών. Αξιοσημείωτο είναι ότι μεροληψία προκύπτει ακόμα και σε θετικές επιστήμες (ιατρική, κ.λπ.) όπως θα αναφέρουμε πιο κάτω.

1.8.3.4 Το κλασσικό πείραμα

Το πρόβλημα της μεροληψίας των απαντήσεων αντιμετωπίστηκε από τους ερευνητές με τη χρησιμοποίηση δύο ομάδων. Το αποτέλεσμα της δράσης δεν θα μετράται από τα αποτελέσματα της πειραματικής ομάδας πριν και μετά τη επίδραση, όπως θα περίμενε κανείς, αλλά από τη σύγκριση μονάδων ανάλυσης που έχουν δεχτεί τη δράση (πειραματική ομάδα) και μονάδων ανάλυσης που δεν την έχουν δεχτεί (ομάδα ελέγχου). Το κλασσικό πείραμα παρουσιάζεται στο σχήμα που ακολουθεί.



Εικόνα 1.6: Το κλασσικό πείραμα

Στην αρχή έχουμε δύο ομάδες ελέγχου, οι οποίες με κριτήρια πιθανοτικής δειγματοληψίας θα πρέπει να μοιάζουν μεταξύ τους. Πραγματοποιείται ο προ-έλεγχος, δηλ. καταμετρώνται ως προς την εξαρτημένη μεταβλητή (π.χ. την προκατάληψη κατά της «καταλληλότητας» των γυναικών για την ανάληψη διοικητικών καθηκόντων) οι μονάδες ανάλυσης της πειραματικής ομάδας και της ομάδας ελέγχου. Στην πειραματική ομάδα ασκείται η επιλεγείσα επίδραση και ξαναμετράται ως προς την εξαρτημένη μεταβλητή (μετα-έλεγχος), μετα-έλεγχος διενεργείται και στην ομάδα ελέγχου, παρά το γεγονός ότι αυτή δεν υποβάλλεται στη σχετική δοκιμασία. Το συμπέρασμα για το αν η δράση επηρεάζει το αποτέλεσμα, εξάγεται από τη σύγκριση των μετα-ελέγχων των δύο ομάδων. Με αυτό τον τρόπο δίνεται η δυνατότητα να υπολογίσουμε τη μεροληπτικότητα που προέρχεται από τις απαντήσεις που δίνει κάποιος ο οποίος συνειδητοποιεί ότι είναι «υποκείμενο πειράματος». Στο παράδειγμα της έρευνας, κάποιος που συνειδητοποιεί ότι η άποψή του για την «καταλληλότητα» των γυναικών-διευθυντών γίνεται αντικείμενο μελέτης, είναι λογικό να δώσει

μεροληπτικές απαντήσεις στο μετα-έλεγχο για να μη δείχνει προκατειλημμένος. Έχοντας την ομάδα ελέγχου μπορούμε να εκτιμήσουμε το βαθμό της προκατάληψης και να συγκρίνουμε τα αποτελέσματα αυτά με το μετα-έλεγχο της πειραματικής ομάδας.

Στο συγκεκριμένο πείραμα τα αναμενόμενα αποτελέσματα είναι: α) ότι το ντοκιμαντέρ δεν παίζει ρόλο και έτσι ο μετα-έλεγχος και για τις δύο ομάδες δίνει τον ίδιο βαθμό προκατάληψης και β) ότι το ντοκιμαντέρ παίζει ρόλο και η προκατάληψη στην πειραματική ομάδα είναι μικρότερη από αυτή της ομάδας ελέγχου.

Το πόσο μεγάλο βαθμό παίζει η προκατάληψη φαίνεται από το «φαινόμενο Χόθορν»: Σε μία έρευνα στις αρχές του αιώνα οι ερευνητές ενδιαφέρονταν να μάθουν τους τρόπους με τους οποίους μπορούσε να αυξηθεί η παραγωγικότητα των εργαζομένων. Για αυτό το λόγο μελέτησαν τις συνθήκες εργασίας στο τηλεφωνικό κέντρο της εταιρείας στην περιοχή Χόθορν του Σικάγου. Τα αποτελέσματα της έρευνας ήταν ενθαρρυντικά, όταν βελτιώνονταν οι συνθήκες φωτισμού. Έτσι βγήκε το συμπέρασμα ότι οι συνθήκες φωτισμού βελτιώνουν την παραγωγικότητα. Για να είναι ασφαλές το συμπέρασμα επανέφεραν τις συνθήκες φωτισμού σταδιακά στις αρχικές συνθήκες, τότε με έκπληξη διαπίστωσαν ότι η παραγωγικότητα συνέχισε να βελτιώνεται! Έτσι το καταληκτικό συμπέρασμα είναι ότι η παραγωγικότητα βελτιωνόταν όχι λόγω φωτισμού, αλλά επειδή οι εργαζόμενοι ένοιωθαν πως τους έδιναν προσοχή οι ερευνητές.

Ενδιαφέρον έχει το φαινόμενο της προκατάληψης του ερευνητή! Σε πολλές περιπτώσεις, κυρίως, όταν η μέτρηση είναι υποκειμενική, ο ερευνητής μπορεί να επηρεαστεί από τη γνώση ότι το υποκείμενο ανήκει στην πειραματική ομάδα. Λόγου χάρι ο γιατρός ερευνητής μπορεί να βλέπει βελτίωση στην κινητική κατάσταση ενός ασθενή, όταν πιστεύει ότι έχει πάρει ένα κατάλληλο φάρμακο. Για την περίπτωση αυτή εφαρμόζεται το «διπλά τυφλό πείραμα». Συγκεκριμένα στην περίπτωση δύο πειραματικών ομάδων, σύμφωνα με τις οποίες η μία λαμβάνει ψευδοφάρμακο, ενώ η άλλη λαμβάνει το πειραματικό φάρμακο, οι υπεύθυνοι ερευνητές δεν πρέπει να γνωρίζουν τί φάρμακο έχει χορηγηθεί στην κάθε ομάδα, ούτε όμως και οι ομάδες πρέπει να το γνωρίζουν αυτό, ούτως ώστε τα αποτελέσματα να είναι έγκυρα χωρίς μεροληψία. Ο τρόπος αυτός αποκαλείται «διπλά τυφλό πείραμα».

1.8.3.5 «Ταίριασμα» και «Τυχαιοποίηση»

Για να είναι σωστά εφαρμόσιμος ο κλασικός σχεδιασμός, όπως αναφέρθηκε και πιο πάνω, θα πρέπει η ομάδα ελέγχου και η πειραματική ομάδα να μοιάζουν όσο το δυνατό περισσότερο σε κάποια χαρακτηριστικά. Αυτό επιτυγχάνεται ή με πιθανοτική δειγματοληψία ή με «ταίριασμα». Το συνηθέστερο παράδειγμα πιθανοτικής δειγματοληψίας είναι όταν και οι δύο ομάδες συλλέγονται με απλή δειγματοληψία από τον πληθυσμό. Σε τέτοια περίπτωση λαμβάνεται βεβαίως υπόψη και το μέγεθος του δείγματος που θα πρέπει να είναι ικανοποιητικό.

Αν ο πληθυσμός είναι μικρός, συνήθως, ακολουθείται η διαδικασία «τυχαιοποίησης», δηλαδή χωρίζεται με τυχαίο τρόπο (κλήρωση) σε πειραματική ομάδα και ομάδα ελέγχου. Η «τυχαιοποίηση» είναι γενικά καλύτερος τρόπος από το «ταίριασμα», όταν δηλαδή για κάθε παρατήρηση από τη μία ομάδα, επιλέγουμε μία από την άλλη που της «ταιριάζει». Το κριτήριο για το «ταίριασμα» είναι η ομοιότητα σε περιορισμένο πλήθος παραγόντων, που ο ερευνητής κρίνει ως σημαντικότερους (π.χ. ηλικία, φύλο, οικογενειακή κατάσταση, κ.λπ.). Στην περίπτωση αυτή το μειονέκτημα της διαδικασίας είναι το ίδιο με την ποσοστική δειγματοληψία, οι παράγοντες που λαμβάνονται υπόψη είναι λίγοι και ο ερευνητής μπορεί να κάνει λάθος εκτίμηση για τους παράγοντες που επηρεάζουν την εξαρτημένη μεταβλητή που εξετάζεται. Το «ταίριασμα» πλεονεκτεί της «τυχαιοποίησης» και της πιθανοτικής δειγματοληψίας, μόνο όταν το πλήθος των μονάδων ανάλυσης είναι μικρό.

1.8.3.6 Εγκυρότητα

Υπάρχουν δύο ειδών παράγοντες οι οποίοι επηρεάζουν αρνητικά την πειραματική έρευνα, οι παράγοντες που αποτελούν πηγή έλλειψης εσωτερικής εγκυρότητας και οι παράγοντες που αποτελούν πηγή έλλειψης εξωτερικής εγκυρότητας.

1.8.3.7 Πηγές έλλειψης εσωτερικής εγκυρότητας

Οι Campbell και Stanley (1963) και οι Cook και Campbell (1979) εξέτασαν πηγές που επηρεάζουν την εσωτερική εγκυρότητα μιας έρευνας, όπως το ιστορικό της έρευνας, την ωρίμανση, την επίδραση του ελέγχου, την επίδραση των εργαλείων μέτρησης, την στατιστική παλινδρόμηση, τα σφάλματα μεροληπτικής επιλογής, την «πειραματική θνησιμότητα» και την αποθάρρυνση συμμετοχής στην έρευνα.

1. Ένα γεγονός που γίνεται γνωστό κατά τη διάρκεια ενός πειράματος μπορεί να επηρεάσει το πείραμα. Για παράδειγμα, μία έρευνα η οποία εξέταζε την εμπιστοσύνη των πολιτών στην αμερικανική κυβέρνηση θα μπορούσε να επηρεαστεί από την αποκάλυψη του Snowden ότι οι αμερικανικές και βρετανικές κυβερνήσεις εφαρμόζουν πρόγραμμα μαζικής παρακολούθησης.
2. Ανεξάρτητα με το αν το πείραμα είναι μακροχρόνιο ή όχι, το υποκείμενο επηρεάζεται από τον ψυχολογικό-ανθρώπινο παράγοντα. Αυτό σημαίνει ότι το υποκείμενο μπορεί να βαρεθεί, να πεινάσει/διψάσει, να κουραστεί, ή αν αναφερόμαστε σε μία μακροχρόνια περίοδο να έχει μία διαφορετική στάση απέναντι στο θέμα λόγω αλλαγής των αντιλήψεών του.
3. Στην περίπτωση που το υποκείμενο γνωρίζει ότι ελέγχεται ίσως προσπαθήσει να δώσει τις απαντήσεις τις οποίες νομίζει ότι θέλουμε ή εκείνες που θα το κάνουν να φαίνεται καλύτερο.
4. Στην περίπτωση που οι ομάδες που χρησιμοποιούνται στο πείραμα έχουν ακραίες παρατηρήσεις, τότε παρατηρείται στατιστική παλινδρόμηση. Ορισμένες φορές διεξάγουμε πειράματα σε υποκείμενα τα οποία ξεκινούν με ακραίες τιμές για την εξαρτημένη μεταβλητή. Σ' αυτή την περίπτωση υπάρχει ο κίνδυνος οι αλλαγές που θα παρατηρηθούν να οφείλονται στο γεγονός ότι τα υποκείμενα εκκινούν από ακραία θέση και όχι λόγω επίδρασης του πειραματικού αποτελέσματος. Αν δοκιμάζουμε, π.χ. μία νέα μέθοδο προπόνησης σε μία ομάδα, η οποία είναι τελευταία στην κατάταξη του πρωταθλήματος, το γεγονός ότι ίσως βελτιώνεται δεν μπορεί να συνδεθεί με το πείραμα, αφού ακόμα και χωρίς πειραματικό ερέθισμα η βελτίωση είναι η μόνη αλλαγή που θα μπορούσε να συμβεί.
5. Οι μονάδες ανάλυσης μπορεί να «πεθάνουν» κατά την έρευνα: Η κατηγορία αυτή δεν αναφέρεται στον βιολογικό θάνατο των υποκειμένων, αλλά στον θάνατο σε πειραματικό επίπεδο. Οποιοσδήποτε λόγος υποχρεώσει το υποκείμενο να εγκαταλείψει το πείραμα μετά τον προ-έλεγχο, ανήκει πλέον σε αυτήν την κατηγορία, αφού επηρεάζει τα τελικά αποτελέσματα.

1.8.3.8 Πλεονεκτήματα και μειονεκτήματα

Το βασικό πλεονέκτημα ενός ελεγχόμενου πειράματος έγκειται στην απομόνωση της επίδρασης της ανεξάρτητης μεταβλητής. Αυτό σημαίνει ότι η μεταβολή των στοιχείων από τον προ-έλεγχο στον μετα-έλεγχο οφείλεται στην επίδραση του πειράματος, αν οι μονάδες ανάλυσης δεν έχουν υποστεί μεταβολή με κάποιο άλλο τρόπο. Ακόμα το πείραμα δίνει τη δυνατότητα επανάληψης, γι' αυτό είναι συνήθως πιο εύκολο από την επανάληψη δειγματοληπτικών ερευνών.

Μία δυσκολία του πειράματος (μειονέκτημα), που παρουσιάζεται, οφείλεται στο γεγονός ότι στην πραγματικότητα είναι δύσκολο να απομονώσουμε πλήρως τις συνθήκες πειράματος, όπως και να εξαλείψουμε την προκατάληψη των υποκειμένων του πειράματος, π.χ. στις απαντήσεις που δίνουν.

1.8.3.9 Δεοντολογία και πειράματα

Ζητήματα που προκύπτουν από τη διεξαγωγή ενός πειράματος είναι τα εξής: τα υποκείμενα, όπως και στην έρευνα πεδίου, δεν πρέπει να γνωρίζουν ότι μελετώνται, γιατί έτσι δημιουργούνται συνθήκες μεροληπτικότητας στη συμπεριφορά και στις απαντήσεις τους. Ακόμα ένα πείραμα μπορεί να δημιουργήσει τραύματα ψυχικά και σωματικά ή και προβλήματα υγείας (αν πρόκειται για ιατρικό πείραμα).

1.8.4 Μη αντιδραστικές μέθοδοι ανάλυσης

Μέχρι τώρα εξετάστηκαν μέθοδοι που, λίγο πολύ, προϋποθέτουν αλληλεπίδραση με τις μονάδες ανάλυσης, αφού κατά κανόνα η μονάδα ανάλυσης είναι άνθρωπος και για να δώσει την αναγκαία πληροφορία, χρειάζεται να υπάρξει αλληλεπίδραση μεταξύ αυτού και του ερευνητή. Ακολούθως θα γίνει αναφορά σε μεθόδους που δεν προϋποθέτουν αλληλεπίδραση με τις μονάδες ανάλυσης, γι' αυτό λέγονται και μη αντιδραστικές μέθοδοι. Πολύ ωραίο παράδειγμα έχει ο Eugene Webb στο βιβλίο του Unobtrusive Research (1966) όπου παρουσιάζει παραδείγματα μη αντιδραστικών ερευνών. Ο Webb παραλληλίζει τη δουλειά που πρέπει να κάνει ο ερευνητής μ' αυτή που κάνει ένας ντεντέκτιβ για να εξιχνιάσει μια υπόθεση. Ένα από τα πιο εντυπωσιακά παραδείγματα, που παρουσιάζει, είναι αυτό, κατά το οποίο ο ερευνητής προκειμένου να καταλήξει στο πιο ενδιαφέρον εύρημα ενός μουσείου, παρατηρεί τη φθορά του ξύλινου πατώματος μπροστά από κάθε εύρημα!

Διακρίνουμε δύο τύπους μη αντιδραστικών μεθόδων: α) τις έρευνες περιεχομένου και β) τις έρευνες καταγεγραμμένων στατιστικών.

Έρευνες περιεχομένου είναι αυτές, που οι μονάδες ανάλυσης είναι καταγεγραμμένα αποτελέσματα ανθρώπινων δραστηριοτήτων ή επικοινωνιών (ιστοσελίδες, βιβλία κ.λπ.).

Για μια έρευνα που αφορά στις συνθήκες εργαζομένων στο χρηματιστήριο της Νέας Υόρκης τη δεκαετία του 1970, π.χ. μπορεί να χρησιμοποιηθεί υλικό όπως φωτογραφίες, συλλογή άρθρων από τις εφημερίδες της εποχής, αγγελίες εφημερίδων για εργασία, δεδομένα με τιμές μετοχών της εποχής, κ.λπ.

Άλλο παράδειγμα είναι η σύνδεση τηλεοπτικών διαφημίσεων με το περιεχόμενο των ταινιών που προβάλλονται. Παράδειγμα η σύνδεση των πολιτικών δημοσιογραφικών εκπομπών (talk show) με προϊόντα αντρικά ή γυναικεία προϊόντα. Το αποτέλεσμα αυτό θα επιβεβαιωθεί ή θα απορριφθεί με απλή καταγραφή των διαφημίσεων που παίχτηκαν τον τελευταίο μήνα σε όλα τα talk show της τηλεόρασης. Ακόμα πολύ ενδιαφέρον έχει, αν αλλάξει ο προσανατολισμός των διαφημίσεων σε μία συγκεκριμένη πολιτική εκπομπή, όπου γνωστός δημοσιογράφος καλεί κάθε φορά έναν πολιτικό (αν υπάρχει δηλαδή, διαφορά ανάλογα με τον πολιτικό προσανατολισμό του υποψηφίου (π.χ. αν είναι δεξιός, αριστερός κ.ο.κ.)

Οι έρευνες καταγεγραμμένων στατιστικών είναι αυτές στις οποίες χρησιμοποιούνται καταγεγραμμένα στατιστικά είτε πρωτογενών δεδομένων είτε επεξεργασμένων, τα οποία μπορεί να παρέχονται από την Ελληνική Στατιστική Αρχή, τη Eurostat, διάφορους επιστημονικούς οργανισμούς, όπως το ΕΚΚΕ, τον Παγκόσμιο Οργανισμό Υγείας, συνδικαλιστικούς φορείς, όπως την ΑΔΕΔΥ, τη ΓΕΣΕΕ, κ.λπ. Ας δούμε μερικά παραδείγματα:

1. Έστω ότι θέλουμε να συνδυάσουμε την εκλογική απήχηση των δύο κορυφαίων κομμάτων σε μία χώρα με την οικονομική κατάσταση των ψηφοφόρων. Μία έρευνα που θα είχε διαθέσιμα ποσοστά των κομμάτων και κατά κεφαλήν εισοδήματα, θα μπορούσε να δώσει τεκμηριωμένες απαντήσεις σε αυτό το ζήτημα.
2. Έστω ότι θέλουμε να διερευνήσουμε την επίδραση της οικονομικής κατάστασης κρατών σε σχέση με το επίπεδο υγειονομικής περίθαλψης των κατοίκων. Θα χρησιμοποιηθούν υφιστάμενα στατιστικά που αφορούν στο ΑΕΠ των χωρών, στο προσδόκιμο ζωής των κατοίκων, στον αριθμό των γιατρών ανά 10.000 κατοίκους, στον αριθμό νοσοκομειακών κλινών ανά 1.000 κατοίκους κ.λπ.

1.8.4.1 Πλεονεκτήματα και μειονεκτήματα

Το μεγάλο πλεονέκτημα των μη αντιδραστικών μεθόδων είναι ότι είναι οικονομικές, ως επί το πλείστον πραγματοποιούνται από ένα ερευνητή, ο οποίος έχει να συλλέξει δεδομένα συνήθως με μικρό ή μηδενικό κόστος. Επίσης επιτρέπει τη διόρθωση των σφαλμάτων στο σχεδιασμό ή στην καταγραφή. Αντίθετα π.χ. με τις δειγματοληπτικές έρευνες, αν προκύψει κάποιο πρόβλημα, μπορεί να συλλεχθούν επιπλέον μονάδες ανάλυσης ή να μελετηθούν άλλα άτομα. Ακόμα σε αντίθεση με πειραματικούς σχεδιασμούς και δειγματοληπτικές μεθόδους, επειδή οι μονάδες ανάλυσης δεν είναι άνθρωποι, η πραγματοποίηση μιας έρευνας δεν δημιουργεί ψυχικές και σωματικές βλάβες σε ανθρώπους (ή πειραματόζωα). Ακόμα για την άρτια πραγματοποίησή τους δεν χρειάζεται να ειπωθούν ψέματα ώστε να αποκρυφτεί η ιδιότητα του ερευνητή (όπως στην έρευνα πεδίου).

1.8.4.2 Δεοντολογία μη αντιδραστικών ερευνών

Επειδή δεν γίνεται απόκρυψη του ρόλου του ερευνητή σ' αυτού του είδους τις έρευνες, τα δεοντολογικά θέματα είναι από μικρά έως ανύπαρκτα. Το μόνο πρόβλημα που μπορεί να υπάρχει είναι ως προς τη δημοσιοποίηση προσωπικών δεδομένων σε μία έρευνα, π.χ. που κάνει χρήση αλληλογραφίας μέσω e-mails κ.λπ.

Βιβλιογραφικές Αναφορές Κεφαλαίου

Ελληνόγλωσσες

Η τεχνική της Φαγεντιανής. Ανακτήθηκε 15 Σεπτεμβρη, 2015 από

<http://www.fhw.gr/chronos/02/crete/gr/artsandtechnology/fayence.html>

Θεοδωρίδης, Χ. (1955). *Εισαγωγή στη Φιλοσοφία*. Αθήνα: Εκδόσεις του κήπου.

Θεοφανίδης, Σ. & Μπένου, Γ. (2002). *Μεθοδολογία της επιστημονικής σκέψης και έρευνας*. Αθήνα: Εκδόσεις ΕΚΕΒΙ

Κίτσος, Χ. (1994). *Στατιστική Ανάλυση πειραματικών σχεδιασμών*. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.

Κυριαζόπουλος, Π. Γ. & Σαμαντά, Ε. (2011). *Μεθοδολογία έρευνας εκπόνησης διπλωματικών εργασιών*. Αθήνα: Σύγχρονη εκδοτική.

Μπιτσάκης, Ε. (1987). Επιστημολογία: Ορισμός, Ρεύματα και Λειτουργία, *Ελληνική Φιλοσοφική Επιθεώρηση (ΕΦΕ)* 4, σσ. 127-145.

Φράγγος, Χ. (2004). *Μεθοδολογία έρευνας και ανάλυση δεδομένων*. Αθήνα: Εκδόσεις Interbooks.

Babbie, E. (2011). *Εισαγωγή στην κοινωνική έρευνα*. Αθήνα: Εκδόσεις Κριτική Α.Ε.

Creswell, J. W. (2012). *Εκπαιδευτική έρευνα. Σχεδιασμός, διεξαγωγή και αξιολόγηση της ποσοτικής και ποιοτικής έρευνας*. Αθήνα: Εκδόσεις ΕΚΕΒΙ

Grawitz, M. (2006). *Μέθοδοι των Κοινωνικών Επιστημών, Τόμοι Α' και Β'*, Αθήνα: Βιβλιόπολις ΑΕΒΕ. Προϊόντα Πνευματικής Δημιουργίας.

Losee, J. (1993). *Φιλοσοφία της Επιστήμης. Μια ιστορική εισαγωγή*. Θεσσαλονίκη: Εκδόσεις Βάνιας.

Mason, J. (2003). *Η διεξαγωγή της ποιοτικής έρευνας*. Αθήνα: Εκδόσεις: ΕΚΕΒΙ.

Hawkes, J. & Wooley L. (1963) *Ιστορία της ανθρωπότητας*, Αθήνα: Εκδόσεις Ελληνική Παιδεία.

Ξενόγλωσσες

Babbie, E. (2007). *The practice of social research*. Thomson Wadsworth. Belmont, CA.

Begon, M. (1979). *Investigating animal abundance: capture-recapture for biologists*. Edward Arnold (Publishers) Ltd.

Campbell, D. T., Stanley, J. C. & Gage, N. L. (1963). *Experimental and quasi-experimental designs for research* (No. 04; Q175, C3.). Boston: Houghton Mifflin.

Cook, T. D., Campbell, D.T. & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings* (Vol. 351). Boston: Houghton Mifflin.

Gagnon, Y. C. (2010). *The case study as research method*. PUQ.

Govindarajulu, Z. (1999). *Elements of sampling theory and methods*. Prentice Hall.

Krebs, C. J. (1999). *Ecological methodology* (Vol. 620). Menlo Park, California: Benjamin/Cummings.

Krueger, R. A. (1997). *Analyzing and reporting focus group results* (Vol. 6). Sage publications.

Morgan, D. L. (1993). *Successful focus groups: Advancing the state of the art* (Vol. 156). Sage Publications.

Rubin, H. J. & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data*. Sage.

Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters*. London: Griffin.

Snedecor, G. W. & Cochran, W. G. (1956). *Statistical Methods*. Iowa State College Press.

Suchting, W. (2006). Epistemology. *Historical Materialism*, 14(3), pp. 331-345.

Webb, E. J., Campbell, D. T., Schwartz, R. D. & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences* (Vol. 111). Chicago: Rand McNally.

Κεφάλαιο 2

Δειγματοληπτικές Έρευνες

Σύνοψη

Στο κεφάλαιο αυτό εξετάζονται τεχνικά ζητήματα δειγματοληπτικών ερευνών οι οποίες θεωρούνται ως η πιο διαδεδομένη ερευνητική μέθοδος στις κοινωνικές επιστήμες και στις επιστήμες διοίκησης και οικονομίας. Ελέγχεται η εγκυρότητα και η αξιοπιστία ερωτηματολογίου, η παρουσίαση και η ανάλυση αποτελεσμάτων και η σύγκριση των ευρημάτων με άλλες μεθόδους έρευνας και τέλος καταγράφονται τα πλεονεκτήματα και μειονεκτήματα δειγματοληπτικών ερευνών. Πιο εκτεταμένη αναφορά γίνεται στις δημοσκοπήσεις που αποτελούν ειδική κατηγορία δειγματοληπτικών ερευνών.

Προσπαιτούμενη γνώση

Προσπαιτούμενη γνώση ουσιαστικά δεν υπάρχει. Καλύτερη κατανόηση του κεφαλαίου όμως θα είναι δυνατή αν ο αναγνώστης έχει γνώσεις δειγματοληψίας (που μπορούν να προκύψουν από τη σωστή μελέτη του προηγούμενου κεφαλαίου). Τα εγχειρίδια που παρουσιάζουν με πληρότητα τις δειγματοληπτικές μεθόδους έρευνας είναι πολλά, ενδεικτικά αναφέρονται τα βιβλία των E. Babbie, D. Morgan, και M. Grawitz (βλ. βιβλιογραφικές αναφορές) ακόμα αναφέρονται ενδεικτικά τα βιβλία των R. Alvarez et.al και G. Michelat και J.P. Thomas για τον χώρο των δημοσκοπήσεων. Όπως έχω αναφέρει και στον πρόλογο σημαντική είναι η συμβολή των φοιτητών μου Γ. Φλουρή, Ι. Νταβή, Γ. Τσιμηδέλη και Ε. Τριανταφύλλου.

2.1. Εισαγωγή

Το πιο διαδεδομένο είδος έρευνας στις επιστήμες διοίκησης και οικονομίας και στις κοινωνικές επιστήμες είναι οι δειγματοληπτικές έρευνες, οι οποίες εφαρμόζονται σε όλους τους σχετικούς ερευνητικούς τομείς και διακρίνονται τόσο για την ευκολία της ανάλυσης των δεδομένων που προκύπτουν, όσο και για την ευκολία εξαγωγής συμπερασμάτων.

2.2.1 Ερωτηματολόγιο

Το ερωτηματολόγιο αποτελεί το θεμελιώδες στοιχείο σε κάθε δειγματοληπτική έρευνα, αλλά και στα πειράματα, στις έρευνες πεδίου και σε άλλες δραστηριότητες που χρειάζονται για τη συγκέντρωση πληροφοριών, στοιχείων και δεδομένων. Έτσι η κατασκευή ενός ερωτηματολογίου είναι πολύ σημαντική, επειδή αυτό παρέχει ουσιαστικά τα δεδομένα της έρευνας, οπότε θα πρέπει να τηρούνται κάποιες αρχές για να είναι αξιοποιήσιμη η πληροφορία που προκύπτει.

Η μορφή ενός ερωτηματολογίου είναι επίσης πολύ σημαντική για το είδος και τη διατύπωση των ερωτήσεων που αυτό περιλαμβάνει. Ένα ερωτηματολόγιο θα πρέπει να είναι οργανωμένο, σαφές, σύντομο, και να εμπεριέχει τις αναγκαίες οδηγίες και υποδείξεις.

Για να επιτυγχάνονται τα προαναφερθέντα θα πρέπει να προηγείται ο κατάλληλος σχεδιασμός και έλεγχος. Αρχικά επιλέγεται η κατάλληλη μορφή ερωτήσεων. Οι ερωτήσεις μπορούν να είναι δύο ειδών: είτε κλειστού τύπου όπου ο ερωτώμενος καλείται να επιλέξει μεταξύ συγκεκριμένων απαντήσεων είτε ανοιχτού τύπου όπου ο ερωτώμενος απαντά στην ερώτηση συμπληρώνοντας το κενό περιθώριο που προβλέπεται για να καταχωρίσει την απάντησή του. Έτσι η επιλογή της μορφής των ερωτήσεων γίνεται με κριτήριο την αποτελεσματικότητα, την ευελιξία, το ενδιαφέρον, την ομοιογένεια και κυρίως την καταλληλότητα στην επεξεργασία των δεδομένων.

Προφανώς οι κλειστές ερωτήσεις υπερτερούν στην ευκολία ανάλυσης των δεδομένων, ενώ οι ανοιχτές δίνουν επιπλέον δυνατότητα στον ερωτώμενο να ξεδιπλώσει τη σκέψη του.

Σε γενικές γραμμές θα πρέπει να προηγούνται ανοιχτές ερωτήσεις με ανάλογο περιεχόμενο και κλειστές, που θα έπονται. Προϋποθέσεις για ένα σωστό ερωτηματολόγιο θεωρούνται:

- **Η διατύπωση σαφών και μονοσήμαντων ερωτήσεων:** Συχνά ο ερευνητής παρασυρόμενος από τη γνώση που έχει σε κάποιο θέμα θεωρεί αυτονόητα πράγματα που δεν είναι. Παράδειγμα: λάθος μπορεί να είναι η ερώτηση «Ποια είναι η γνώμη σας για το νέο νομοσχέδιο της παιδείας που αφορά στη μεταγραφή των φοιτητών;». Στο ερώτημα αυτό σίγουρα θα υπάρχουν και ερωτώμενοι που δεν γνωρίζουν σε τι αφορά το συγκεκριμένο νομοσχέδιο.
- **Η αποφυγή διφορούμενων ερωτήσεων:** Οι συντάκτες συνήθως επιθυμούν μια και μοναδική απάντηση σε ένα ερώτημα που μπορεί να περιέχει ποικίλες απαντήσεις, γι' αυτό πρέπει οι δυνατές απαντήσεις να μην επικαλύπτονται. Παράδειγμα: στην ερώτηση «Η Ελλάδα πρέπει να σκληρύνει τη στάση της απέναντι στην Τουρκία στο θέμα των γκρίζων ζωνών του Αιγαίου»; Παρά το ότι κάποιιοι θα απαντήσουν ξεκάθαρα ναι ή ξεκάθαρα όχι, σημαντικό ποσοστό των ερωτώμενων θα θέλει διευκρίνιση στο τί είναι σκλήρυνση της στάσης καθώς και το ποιες θεωρείται ότι θα είναι οι επιπτώσεις.
- **Η αντίληψη της ικανότητας απάντησης των ερωτώμενων:** Πολλές φορές οι ερωτώμενοι νομίζουν ότι ξέρουν να απαντήσουν, αλλά οι απαντήσεις που δίνουν δεν είναι σωστές. Παράδειγμα: στην ερώτηση «Το κυκλοφοριακό συγκοινωνιακό σύστημα στην Αθήνα έχει βελτιωθεί σε σχέση με τη δεκαετία του ενενήντα»; Είναι μία ερώτηση που ο ερωτώμενος θα πρέπει να ανακαλέσει στη μνήμη του την κατάσταση της συγκοινωνίας 25 χρόνια πριν. Σε τέτοια περίπτωση είναι δύσκολο να περιμένουμε αξιόπιστες απαντήσεις. Επειδή επίσης μόνιμα οι συντάκτες ερωτηματολογίου έχουν την αμφιβολία αν είναι αξιόπιστες οι απαντήσεις των ερωτώμενων, σε περιπτώσεις που οι απαντήσεις αφορούν σε προσωπικά βιώματα, καλό είναι να υπάρχουν ερωτήσεις ελέγχου αξιοπιστίας, π.χ. με μια ερώτηση διαπίστωσης ενός ιστορικού γεγονότος της εποχής που καλούμε τους ερωτώμενους να μας δώσουν πληροφορίες.
- **Η απροθυμία των ερωτώμενων να απαντήσουν:** Πολλές φορές οι ερωτώμενοι είναι απρόθυμοι και διστακτικοί να βοηθήσουν δίνοντας απάντηση σε ερωτήσεις που θίγουν πολιτικά, ηθικά ζητήματα κ.ά. και τίθενται από κάποιον που δεν γνωρίζουν. Χαρακτηριστικό είναι ότι σε απολυταρχικά καθεστώτα οι ερωτώμενοι διστάζουν να απαντήσουν ακόμα και σε ερωτήματα που αφορούν στη γνώση ενός προϊόντος κ.ο.κ. Γεγονός που πρέπει να λαμβάνεται υπόψη και κατά τη δημιουργία σχετικών ερωτήσεων. Οι ερωτήσεις που εμπεριέχονται σε ένα ερωτηματολόγιο θα πρέπει να είναι σχετικές με τους ερωτώμενους και κατανοητές από αυτούς, καθώς δεν πρέπει να τους ζητούν να σκεφθούν για θέματα που δεν τους έχουν απασχολήσει. Ωραίο παράδειγμα ελέγχου αξιοπιστίας σε ένα ερωτηματολόγιο αναγνωσιμότητας πολιτικών είναι αυτό που εμπεριέχει στο ερωτηματολόγιο ερωτήματα σχετικά με την ύπαρξη ενός πολιτικού που δεν υπάρχει κ.ο.κ.
- **Η αποτύπωση σύντομων ερωτήσεων:** Οι ερωτώμενοι κουράζονται από πολύπλοκα και μεγάλης έκτασης ερωτηματολόγια, οι σύντομες και σαφείς ερωτήσεις είναι ιδανικές.
- **Η αποφυγή ερωτήσεων αρνητικού περιεχομένου:** Πάντα ελλοχεύει ο κίνδυνος να παρανοήσουν οι ερωτώμενοι κάτι και λέξεις όπως το «δεν», «ναι ή όχι» συμβαίνει να δημιουργούν κάποια προκατάληψη στους ερωτώμενους. Οι περισσότεροι ερωτώμενοι συμπληρώνουν γρήγορα το ερωτηματολόγιο και είναι πιθανό να μην προσέξουν ότι η ερώτηση είναι αρνητική. Παράδειγμα λάθους ερώτησης: «Η Ελλάδα δεν πρέπει να είναι υπέρ της ένταξης της Τουρκίας στην Ευρωπαϊκή Ένωση;».
- **Η παράλειψη προκατειλημμένων και μεροληπτικών ερωτήσεων και όρων:** Επειδή τέτοιες λέξεις μπορεί να επηρεάσουν τον ερωτώμενο να απαντήσει σε μια συγκεκριμένη κατεύθυνση ή μπορεί να υπαρξει η ταύτιση κάποιου ατόμου με μια συγκεκριμένη κατεύθυνση (κάτι που οδηγεί σε έλλειψη εγκυρότητας). Παράδειγμα λάθους διατύπωσης: «Συμφωνείτε με την άποψη των δανειστών της Ελλάδας ότι οι Έλληνες δεν εργάζονται όσο οι άλλοι Ευρωπαίοι (!);».
- **Η μορφή των ερωτήσεων:** Η μορφή των ερωτήσεων ενός ερωτηματολογίου μπορεί να δίνει τη δυνατότητα στους ερωτώμενους να επιλέξουν την απάντηση από μια λίστα πιθανών

απαντήσεων (μέσα σε κουτάκια ή με κωδικό αριθμό δίπλα σε αυτές από ένα πρόγραμμα επεξεργασίας του υπολογιστή). Επίσης, μπορεί να εμπεριέχονται ερωτήσεις συνάφειας, η σωστή χρήση τους μπορεί να κάνει πιο εύκολο το έργο των ερωτώμενων και τις απαντήσεις πιο σαφείς και αποτελεσματικές. Σε αυτή την περίπτωση θα πρέπει να περιλαμβάνονται οδηγίες στην επικεφαλίδα κάθε σελίδας όπου αναγράφεται η ερώτηση συνάφειας. Μια άλλη μορφή είναι οι ερωτήσεις πίνακα ή κλίμακας Likert που είναι ένας πίνακας που περιέχει ερωτήσεις με πέντε ή επτά πιθανές απαντήσεις (π.χ. πολύ δυσαρεστημένος/η, δυσαρεστημένος/η, ούτε δυσαρεστημένος/η, ούτε ικανοποιημένος/η, ικανοποιημένος/η, πολύ ικανοποιημένος/η). Τα πλεονεκτήματα αυτής της μορφής είναι ότι χρησιμοποιείται εποικοδομητικά και αποτελεσματικά ο χώρος, είναι πιο γρήγορη η συμπλήρωση του ερωτηματολογίου και έχουν τη δυνατότητα οι ερωτώμενοι να συγκρίνουν τις απαντήσεις τους. Υπάρχουν, όμως, και κάποια μειονεκτήματα σε αυτή τη μορφή ερωτηματολογίου: δεν αποτυπώνεται από τη διάταξη των απαντήσεων η ιδιοσυγκρασία του κάθε ερωτώμενου, ακόμα λόγω της δομής των απαντήσεων πολλές φορές ο ερωτώμενος μπαίνει στον πειρασμό να απαντά με ταχύτητα, κάνοντας πρόχειρη ανάγνωση, επειδή θεωρεί ότι λίγο πολύ όλες οι απαντήσεις κινούνται στην ίδια κατεύθυνση.

- **Η διάταξη των ερωτήσεων:** Η σειρά των ερωτήσεων ενός ερωτηματολογίου είναι σημαντική καθώς μπορεί να επηρεάσει τις απαντήσεις των ερωτώμενων. Παράδειγμα: έστω ότι υπάρχει μια ανοιχτή ερώτηση «Ποιο είναι το σημαντικότερο πρόβλημα της Ελλάδας σήμερα;» και μετά από λίγες ερωτήσεις ακολουθεί η ερώτηση κλειστού τύπου «Συμφωνείτε με την άποψη ότι η ανεργία των νέων είναι σε δυσθεώρητα όρια;» με απαντήσεις τύπου Likert καθόλου, λίγο, αρκετά, πολύ, πάρα πολύ. Προφανώς αν η δεύτερη ερώτηση γίνει πριν από την πρώτη, τότε οι απαντήσεις στην πρώτη ερώτηση θα είναι τελείως διαφορετικές. Πολλές φορές οι ερευνητές-συντάκτες προβαίνουν σε μια τυχαία σειρά των στοιχείων του ερωτηματολογίου. Το καλύτερο είναι να γίνεται εκτίμηση των συνεπειών της σειράς των ερωτήσεων. Είναι διαφορετική η σειρά στα ατομικής συμπλήρωσης ερωτηματολόγια και στις συνεντεύξεις. Ακόμα, τα δημογραφικά στοιχεία θα πρέπει να τοποθετούνται στο τέλος ενός ερωτηματολογίου και όχι στην αρχή.

Όπως αναφέρθηκε παραπάνω σε ένα ερωτηματολόγιο είναι αναγκαίο να περιέχονται ακριβείς οδηγίες και εισαγωγικά σχόλια για να διευκολύνεται ο ερωτώμενος ως προς την συμπλήρωσή του. Στις ερωτήσεις ανοικτού τύπου θα πρέπει να προσδιορίζεται το μέγεθος των απαντήσεων των ερωτώμενων σε κάθε ερώτηση από τους συντάκτες. Τα ερωτηματολόγια που περιέχουν ενότητες ή υποενότητες θα πρέπει να περιλαμβάνουν και μια εξήγηση από κάτω και κάποιες ερωτήσεις μπορεί να χρειάζονται ειδικές οδηγίες με σκοπό να απαντηθούν σωστά από τους ερωτώμενους.

Τέλος, θα πρέπει οι συντάκτες να προβαίνουν σε ένα προέλεγχο του ερωτηματολογίου πριν την δημοσίευσή του. Ένας τρόπος είναι η ανάγνωσή του, δηλαδή να ζητηθεί από τους συντάκτες να διαβάσουν το ερωτηματολόγιο άλλοι άνθρωποι και να διατυπώσουν τις απόψεις τους για αυτό ή να πραγματοποιήσουν μιας μικρής έκτασης πιλοτική έρευνα.

Συνοψίζοντας, θα πρέπει οι ερευνητές να διατυπώσουν το σκοπό της έρευνάς τους και να συντάξουν το ερωτηματολόγιο που θα εξυπηρετεί τον σκοπό αυτό λαμβάνοντας υπόψη όλα τα παραπάνω, στη συνέχεια θα δούμε παραδείγματα μελετών περίπτωσης σχεδιασμού ερωτηματολογίων.

2.2.2 Δείγματα ερωτηματολογίων

Κατωτέρω παρατίθενται κάποια ερωτηματολόγια που περιλαμβάνουν όλα τα στοιχεία που αναπτύξαμε παραπάνω λεπτομερώς. Τα συγκεκριμένα ερωτηματολόγια συντάχθηκαν από τις απόφοιτες του μεταπτυχιακού προγράμματος «Διεθνής Διοικητική των Επιχειρήσεων» Ιωάννα Νταβή και Γεωργία Τσιμηδέλη στα πλαίσια των διπλωματικών τους εργασιών, επιβλέπων καθηγητής ήταν ο Μ. Χαλικιάς.

Ερωτηματολόγιο 1

Θέμα: «Η επίδραση της οικονομικής κρίσης στην ισορροπία εργασιακής και προσωπικής ζωής (work life balance). Μελέτη περίπτωσης: Η Εθνική Τράπεζα».

Συντάκτης ερωτηματολογίου: Ιωάννα Νταβή

1^η Ενότητα: Ισορροπία Εργασίας-Οικογένειας

(Κλίμακα Likert)

Πόσο συχνά έχουν συμβεί σε σας τα ακόλουθα κατά τη διάρκεια του τελευταίου χρόνου;	Σχεδόν ποτέ	Μερικές φορές	Συχνά	Πολύ συχνά	Πάντα
1. Έχω έρθει στο σπίτι από τη δουλειά πολύ κουρασμένος/η για να κάνω τις απαραίτητες δουλειές του σπιτιού					
2. Είναι δύσκολο για μένα να εκπληρώσω τις υποχρεώσεις μου εξαιτίας του χρόνου που δαπανώ στη δουλειά					
3. Το άγχος και η κούραση σχετικά με τις επαγγελματικές ευθύνες πρέπει να μεταφέρονται στο σπίτι;					
4. Επιστρέφω εκνευρισμένος από τη δουλειά με συνέπεια να συγκρούομαι με την οικογένεια μου					
5. Το άγχος και η κούραση σχετικά με τις οικογενειακές ευθύνες πρέπει να μεταφέρονται στην εργασία;					
6. Μου είναι δύσκολο να συγκεντρωθώ στη δουλειά εξαιτίας των υποχρεώσεων του σπιτιού					

Πίνακας 2.1: Πίνακας ερωτηματολογίου 1 (πρώτη ενότητα)

2^η Ενότητα: Εργασιακή Ικανοποίηση

(Κλίμακα Likert)

Παρακαλώ δηλώστε τον βαθμό που συμφωνείτε με τις παρακάτω προτάσεις	Συμφωνώ έντονα	Συμφωνώ	Διαφωνώ	Διαφωνώ έντονα	Δεν μπορώ να διαλέξω
1. Είμαι ικανοποιημένος από την εργασία μου					
2. Η δουλειά μου προσφέρει προοπτικές για εξέλιξη					
3. Στη δουλειά μου έχω την ευκαιρία να αποκτήσω νέες γνώσεις και να εξελιχθώ επαγγελματικά					
4. Πληρώνομαι καλά για τη δουλειά μου					
5. Είμαι ικανοποιημένος από το σχεδιασμό των αδειών / διακοπών					
6. Επικοινωνώ καλά με τους συναδέλφους μου στη δουλειά					
7. Επικοινωνώ καλά με τους προϊσταμένους μου στη δουλειά					
8. Οι ανώτεροί μου, μου παρέχουν αρκετή υποστήριξη όταν χρειάζεται					
9. Υπάρχει ευελιξία όταν προκύπτουν					

προσωπικά προβλήματα					
10. Πληρώνομαι ικανοποιητικά για τη δουλειά που προσφέρω					
11. Στην εργασία μου, εγώ έχω τον έλεγχο της κατάστασης					
12. Υπάρχουν ευκαιρίες για περεταίρω εκπαίδευση και εκμάθηση νέων δεξιοτήτων					
13. Είναι σαφής ο ρόλος μου και οι ευθύνες μου στην εργασία μου					
14. Αξιοποιούνται οι ικανότητες μου στη δουλειά μου					
15. Οι σημερινές ικανότητές μου ανταποκρίνονται καλά στα καθήκοντά μου					
16. Νιώθω περήφανος/η για τη δουλειά που κάνω					

Πίνακας 2.2: Πίνακας ερωτηματολογίου 1 (δεύτερη ενότητα)

3^η Ενότητα: Εργασιακή Πίεση

(Κλίμακα Likert)

Παρακαλώ δηλώστε τον βαθμό που συμφωνείτε με τις παρακάτω προτάσεις	Συμφωνώ έντονα	Συμφωνώ	Διαφωνώ	Διαφωνώ έντονα	Δεν μπορώ να διαλέξω
1. Υπάρχει εξουθενωτικός ρυθμός εργασίας και πιεστικές προθεσμίες					
2. Το περιβάλλον στην εργασία μου αλλάζει συνεχώς και γρήγορα					
3. Πρέπει να επιτυγχάνω διαρκώς υψηλότερους στόχους για να παραμείνω στη δουλειά μου					
4. Υπάρχει μεγάλος φόρτος εργασίας, πίεση στη δουλειά πέρα από τα όρια αντοχής μου					
5. Υπάρχει ανταγωνισμός στη δουλειά μου					
6. Συχνά εργάζομαι πέραν του κανονικού ωραρίου					
7. Δεν έχω αρκετό χρόνο για να διεκπεραιώνω τη δουλειά μου					
8. Η δουλειά μου είναι πνευματικά απαιτητική					

Πίνακας 2.3: Πίνακας ερωτηματολογίου 1 (τρίτη ενότητα)

4^η Ενότητα: Πρακτικές για τη συμφιλίωση εργασιακής και προσωπικής ζωής

(Κλίμακα Likert)

Παρακαλώ δηλώστε τον βαθμό που νομίζετε ότι οι παρακάτω επιλογές θα βοηθούσαν στην ισορροπία της εργασιακής και προσωπικής ζωής σας λαμβάνοντας υπόψη τις εκάστοτε καθημερινές ανάγκες σας εκτός δουλειάς	Συμφωνώ έντονα	Συμφωνώ	Διαφωνώ	Διαφωνώ έντονα	Δεν μπορώ να διαλέξω
1. Η δυνατότητα να προσέρχεστε και να αποχωρείτε στην εργασία σας αργότερα από το καθιερωμένο ωράριο					
2. Η δυνατότητα επιλογής ενός ωραρίου που θα ταιρίαζε καλύτερα στις οικογενειακές και κοινωνικές δεσμεύσεις σας					
3. Η δυνατότητα επιλογής μειωμένου ωραρίου					
4. Η δυνατότητα επιλογής να εργάζεστε ορισμένες μέρες την εβδομάδα					
5. Η δυνατότητα να εργάζεστε από το σπίτι teleworking					

Πίνακας 2.4: Πίνακας ερωτηματολογίου 1 (τέταρτη ενότητα)

5^η Ενότητα: Δημογραφικά Στοιχεία

Παρακαλώ δηλώστε αν συμφωνείτε ή διαφωνείτε με την παρακάτω πρόταση

1. Στις σημερινές οικονομικές συνθήκες, ο μισθός μου δεν είναι επαρκής για να τα βγάλω πέρα:

Συμφωνώ	
Διαφωνώ	

Πίνακας 2.5: Πίνακας ερωτηματολογίου 1 (ικανοποίηση από το μισθό)

2. Φύλο:

Ανδρας	
Γυναίκα	

Πίνακας 2.6: Πίνακας ερωτηματολογίου 1 (φύλο)

3. Ηλικία:

Κάτω των 25 ετών	26-35 ετών	36-45 ετών	46-55 ετών	Άνω των 55 ετών
---------------------	---------------	---------------	---------------	--------------------

Πίνακας 2.7: Πίνακας ερωτηματολογίου 1 (ηλικία)

4. Οικογενειακή κατάσταση:

Παντρεμένος/η	Διαζευγμένος/η	Ανύπαντρος/η	Χήρος/α
---------------	----------------	--------------	---------

Πίνακας 2.8: Πίνακας ερωτηματολογίου 1 (οικογενειακή κατάσταση)

5. Πόσα παιδιά και εξαρτώμενα άτομα έχετε στις πιο κάτω κατηγορίες; Παρακαλώ σημειώστε τον ακριβή αριθμό:

	Αριθμός
Προσχολική ηλικία	
Δημοτικό	
Γυμνάσιο/ Λύκειο	
Τριτοβάθμια Εκπαίδευση	
Στρατιωτική θητεία	
Ενήλικες	

Πίνακας 2.9: Πίνακας ερωτηματολογίου 1 (αριθμός παιδιών)

6. Επίπεδο εκπαίδευσης:

Γυμνάσιο/ Λύκειο	
Απόφοιτος μεταλκειακής εκπαίδευσης (ΓΕΚ, δημόσια - ιδιωτική Μέση σχολή)	
Πανεπιστήμιο	
Μεταπτυχιακό	

Πίνακας 2.10: Πίνακας ερωτηματολογίου 1 (επίπεδο εκπαίδευσης)

7. Συμβόλαιο εργασίας:

Μόνιμος	
Συμβασιούχος	

Πίνακας 2.11: Πίνακας ερωτηματολογίου 1 (σχέση εργασίας)

8. Πόσο χρονικό διάστημα εργάζεστε σε αυτή τη θέση;

		Έτη
--	--	-----

Πίνακας 2.12: Πίνακας ερωτηματολογίου 1 (αριθμός ετών προϋπηρεσίας)

9. Απασχόληση:

Πλήρης	
Μερική	

Πίνακας 2.13: Πίνακας ερωτηματολογίου 1 (είδος απασχόλησης)

10. Η εργασία σας είναι σχετική με τις σπουδές σας;

ΝΑΙ	
-----	--

OXI	
-----	--

Πίνακας 2.14: Πίνακας ερωτηματολογίου 1 (συνάφεια με σπουδές)

11. Μηνιαίο εισόδημα σε ευρώ:

500-1.000	
1.001-1.500	
1.501-2.000	
Άνω των 2.000	

Πίνακας 2.15: Πίνακας ερωτηματολογίου 1 (μηνιαίο εισόδημα)

12. Ο/η σύζυγος/σύντροφός σας εργάζεται;

Ναι	
Όχι	

Πίνακας 2.16: Πίνακας ερωτηματολογίου 1 (απασχόληση συζύγου)

13. Υπό κανονικές συνθήκες, καθημερινά ποιος είναι ο συνολικός χρόνος για τη μετάβασή σας στην εργασία και την επιστροφή στο σπίτι:

	Λεπτά
--	-------

Πίνακας 2.17: Πίνακας ερωτηματολογίου 1 (χρόνος μετάβασης στην εργασία)

Ερωτηματολόγιο 2

Θέμα: «Σχέση υφιστάμενου-προϊστάμενου και η επίδραση στην εργασιακή ικανοποίηση και απόδοση του εργαζομένου».

Συντάκτης ερωτηματολογίου: Γεωργία Τσιτμηδέλη

Ερωτηματολόγιο Υφισταμένων

Αρχή Φόρμας

1. Σε ποιο βαθμό επηρεάζεται η εργασιακή σας ικανοποίηση και απόδοση από τους παρακάτω παράγοντες;

	Καθόλου	Λίγο	Σχετικά	Αρκετά	Πολύ
Το επίπεδο του στρες και η καθημερινή πίεση	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η αίσθηση ότι αμείβεστε δίκαια	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ευκαιρία εξέλιξης	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η σιγουριά της εργασίας	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Οι πολιτικές της επιχείρησης	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Το περιεχόμενο της εργασίας	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η σχέση σας με τον προϊστάμενο	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.18: Πίνακας ερωτηματολογίου 2 (επιρροή εργασιακής ικανοποίησης και απόδοσης)

2. Σε ποιον βαθμό τα παρακάτω στοιχεία της σχέσης μεταξύ προϊστάμενου και υφιστάμενου επιδρούν στην εργασιακή ικανοποίηση και απόδοσή σας;

	Καθόλου	Λίγο	Σχετικά	Αρκετά	Πολύ
Η αναγνώριση της δουλειά σας από τον προϊστάμενο	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ανοιχτή επικοινωνία με τον προϊστάμενο σας	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ομαδικότητα και η κοινή προσπάθεια για επίτευξη στόχων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Η συνεργασία με τον προϊστάμενο	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η κατανομή ευθυνών	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η αναλυτική περιγραφή των καθηκόντων και των υποχρεώσεών σας	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ενθάρρυνση ανάληψης πρωτοβουλιών	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η δυνατότητα συμμετοχής στη λήψη αποφάσεων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ενθάρρυνση υλοποίησης καινοτόμων προγραμμάτων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Οι ευκαιρίες για συνεχιζόμενη επαγγελματική κατάρτιση	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ικανότητα του προϊστάμενου να παρακινεί	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η δυνατότητα ανάπτυξης ικανοτήτων μέσα από τη συμμετοχή σε εκπαιδευτικά προγράμματα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η υποστήριξη σε ζητήματα που ενέχουν δυσκολίες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η δυνατότητα ανάληψης περισσότερων αρμοδιοτήτων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η εργασιακή ευελιξία στο τμήμα που εργάζομαι	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.19: Πίνακας ερωτηματολογίου 2 (επιρροή σχέση προϊστάμενου-υφιστάμενου στην εργασιακή ικανοποίηση και απόδοση)

3. Στο τμήμα που εργάζομαι ο προϊστάμενός μου για να είναι αποτελεσματικός θα πρέπει να:

	Διαφωνώ απόλυτα	Διαφωνώ	Ούτε διαφωνώ ούτε συμφωνώ	Συμφωνώ	Συμφωνώ απόλυτα
Ενθαρρύνει την ενεργό συμμετοχή στη λήψη αποφάσεων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ενδυναμώνει τη συναδελφικότητα και την ομαδικότητα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Διαμορφώνει από κοινού στόχους και να τους υλοποιεί	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ενδιαφέρεται για τα συναισθήματα των υφισταμένων του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ενδιαφέρεται για τις προσωπικές ανάγκες των υφισταμένων του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Επιβραβεύει τη καλή απόδοση	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αντιλαμβάνεται τις αλλαγές ως ευκαιρίες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Χρησιμοποιεί αποτελεσματικά την αναπτυξιακή μέθοδο επιθεώρησης του προσωπικού μέσω της παρακολούθησης της εργασίας του εκάστοτε υπαλλήλου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Τροφοδοτεί με νέες ιδέες τους υφισταμένους του για την αντιμετώπιση των κοινών προβλημάτων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Προωθεί και να μεταδίδει τις προσδοκίες του για υψηλή απόδοση	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Εντοπίζει τις σημαντικές διαφορές μεταξύ των υφισταμένων του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ηγείται περισσότερο με τις πράξεις του παρά με τα λόγια του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Παρακολουθεί συστηματικά τις ενέργειες του κάθε υπαλλήλου για να είναι σίγουρος ότι συνδέονται άμεσα με τα επιδιωκόμενα αποτελέσματα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αντιμετωπίζει τους κινδύνους, να αξιοποιεί τις ευκαιρίες και να «παρακολουθεί το περιβάλλον» υπέρ της κλινικής	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αναγνωρίζει τις σπουδαίες επιδόσεις	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αναπτύσσει πρωτοβουλίες και να υποστηρίζει προγράμματα που διευκολύνουν τη δημιουργία ενός θετικού κλίματος	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.20: Πίνακας ερωτηματολογίου 2 (αποτελεσματικότητα προϊστάμενου)

4. Σε ποιον βαθμό επηρεάζεται η απόδοσή σας στην εργασία από τα παρακάτω;

	Καθόλου	Λίγο	Σχετικά	Αρκετά	Πολύ
Την ανάθεση παραπάνω αρμοδιοτήτων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Τη δυνατότητα ανάληψης πρωτοβουλιών	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Την επιβράβευση από τους προϊστάμενους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Τη δυνατότητα εξέλιξης	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Τη κατανόηση των προσωπικών σας ιδιαιτεροτήτων από τον προϊστάμενο	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Την ίση μεταχείριση	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Το ομαδικό πνεύμα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.21: Πίνακας ερωτηματολογίου 2 (επιρροή στην απόδοση)

5. Πόσο σας εκφράζουν τα παρακάτω σχετικά με την «ασορροπημένη» σχέση μεταξύ του προϊστάμενου και του υφιστάμενου:

	Διαφωνώ απόλυτα	Διαφωνώ	Ούτε διαφωνώ ούτε συμφωνώ	Συμφωνώ	Συμφωνώ απόλυτα
Ο υφιστάμενος πρέπει να γνωρίζει τις εξωτερικές πιέσεις και τους στόχους του προϊστάμενου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Κάθε υφιστάμενος πρέπει να γνωρίζει τις δικές του ανάγκες και ικανότητες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο υφιστάμενος εξαρτάται από τον ανώτερό του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο προϊστάμενος πρέπει να εξωτερικεύει τις επιθυμίες και τις ιδιαιτερότητές του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο προϊστάμενος είναι αναγκαίο να ενημερώνεται και για τις επιτυχίες και τις αποτυχίες της ομάδας του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Για το πόσα πράγματα πρέπει να ενημερωθεί εξαρτάται από τη διάθεση και την εμπιστοσύνη του ως προς τους υφισταμένους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Πρέπει να υπάρχει εμπιστοσύνη και εντιμότητα μεταξύ των δύο μελών	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Πρέπει ο υφιστάμενος να «βλέπει» τη θετική πλευρά στη δουλειά ακόμα και αν δέχεται επικριτικά σχόλια από τον προϊστάμενό του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο υφιστάμενος πρέπει να ελέγχει τις αντιδράσεις του απέναντι στον προϊστάμενό του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.22: Πίνακας ερωτηματολογίου 2 (στοιχεία ισορροπίας σχέσης προϊστάμενου-υφισταμένου)

Στοιχεία Ερωτηθέντων

1. Φύλο:

Αντρας	<input type="radio"/>
Γυναίκα	<input type="radio"/>

Πίνακας 2.23: Πίνακας ερωτηματολογίου 2 (φύλο)

2. Σε ποιο ηλικιακό γκρουπ ανήκετε;

Κάτω από 25	<input type="radio"/>
26-30	<input type="radio"/>
31-45	<input type="radio"/>
46-60	<input type="radio"/>
61 και άνω	<input type="radio"/>

Πίνακας 2.24: Πίνακας ερωτηματολογίου 2 (ηλικία)

3. Ποιο είναι το ανώτερο επίπεδο σπουδών που έχετε συμπληρώσει;

Απόφοιτος Λυκείου	<input type="radio"/>
Απόφοιτος ΑΕΙ/ ΤΕΙ	<input type="radio"/>
Κάτοχος Μεταπτυχιακού Διπλώματος	<input type="radio"/>
Άλλο:	

Πίνακας 2.25: Πίνακας ερωτηματολογίου 2 (εκπαιδευτικό επίπεδο)

Σε ποιο μαιευτήριο εργάζεστε;

4. Ποιος είναι ο τίτλος της θέσης εργασίας σας;

5. Πόσο χρονικό διάστημα εργάζεστε στο μαιευτήριο;

>1 χρόνο	<input type="radio"/>
1-5 χρόνια	<input type="radio"/>
6-10 χρόνια	<input type="radio"/>
<10 χρόνια	<input type="radio"/>

Πίνακας 2.26: Πίνακας ερωτηματολογίου 2 (έτη προϋπηρεσίας)

6. Πόσο χρονικό διάστημα συνεργάζεστε με τον προϊστάμενό σας;

>1 χρόνο	<input type="radio"/>
1-5 χρόνια	<input type="radio"/>
6-10 χρόνια	<input type="radio"/>
<10 χρόνια	<input type="radio"/>

Πίνακας 2.27: Πίνακας ερωτηματολογίου 2 (έτη συνεργασίας με τον συγκεκριμένο προϊστάμενο)

7. Σχόλια

Ερωτηματολόγιο Προϊσταμένων

Αρχή Φόρμας:

1. Ποια είναι η άποψή σας για τα παρακάτω;

	Διαφωνώ απόλυτα	Διαφωνώ	Ούτε διαφωνώ ούτε συμφωνώ	Συμφωνώ	Συμφωνώ απόλυτα
Υπάρχει συσχέτιση ανάμεσα στην εργασιακή ικανοποίηση και την αποδοτικότητα των εργαζομένων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο ρόλος του προϊστάμενου επιδρά στην εργασιακή ικανοποίηση και απόδοση του υφιστάμενου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η εργασιακή ικανοποίηση είναι αναγκαία για την ευημερία της κλινικής	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η δραστηριοποίηση των υφισταμένων αποτελεί μέλημα των ανώτερων στελεχών	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η επίτευξη των στόχων εξαρτάται από τη συνεργασία του ανθρώπινου δυναμικού	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.28: Πίνακας ερωτηματολογίου 2 (επιρροή της σχέσης προϊσταμένου-υφισταμένου στην εργασιακή απόδοση)

2. Προσδιορίστε το βαθμό που σας αντιπροσωπεύουν τα παρακάτω:

Στο τμήμα που ηγούμαι θα πρέπει να:

	Διαφωνώ απόλυτα	Διαφωνώ	Ούτε διαφωνώ ούτε συμφωνώ	Συμφωνώ	Συμφωνώ απόλυτα
Βοηθάω σε διάφορα ζητήματα που ενέχουν δυσκολίες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μεταδίδω γνώσεις	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αναλαμβάνουν οι υφιστάμενοι περισσότερες ευθύνες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Υπάρχει περισσότερη εργασιακή ευελιξία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Υπάρχει αμοιβαία εμπιστοσύνη	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Υπάρχει ανεξαρτησία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Υπάρχει θετική υποστήριξη	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Υπάρχει ανοιχτή επικοινωνία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Υπάρχει αυτονομία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Φέρουν οι υφιστάμενοι εις πέρας γρήγορα μια εργασία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Είναι πιο αφοσιωμένοι οι υφιστάμενοι	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Υπάρχει αυστηρότητα στους τύπους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Εκτελούν οι υφιστάμενοι μόνο ό,τι περιλαμβάνει η θέση εργασία τους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μην υπάρχει διάθεση για εργασία στον ίδιο βαθμό κάθε μέρα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Χαίρουν όλοι ανεξαιρέτως την εγνόιά μου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Επικρατεί πολλές φορές άγχος και stress	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.29: Πίνακας ερωτηματολογίου 2 (ρόλος του προϊσταμένου)

3. Προσδιορίστε σε ποιο βαθμό τα παρακάτω συντελούν στην προσωπική σας επιτυχία:

	Καθόλου	Λίγο	Σχετικά	Αρκετά	Πολύ
Η επιβράβευση των υφισταμένων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ανάθεση αρμοδιοτήτων στους υφιστάμενους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η δυνατότητα ανάληψης πρωτοβουλιών	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ανάπτυξη - εξέλιξη των υφιστάμενων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ικανότητα προσαρμογής στις ιδιαιτερότητες της προσωπικότητας των υφισταμένων	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η ίση μεταχείριση του προσωπικού	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Το ομαδικό πνεύμα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.30: Πίνακας ερωτηματολογίου 2 (παράγοντες επιτυχίας προϊσταμένου)

4. Πόσο σας εκφράζουν τα παρακάτω σχετικά με την «ισορροπημένη» σχέση μεταξύ του προϊστάμενου και του υφιστάμενου.

	Διαφωνώ απόλυτα	Διαφωνώ	Ούτε διαφωνώ ούτε συμφωνώ	Συμφωνώ	Συμφωνώ απόλυτα
Ο υφιστάμενος πρέπει να γνωρίζει τις εξωτερικές πιέσεις και τους στόχους του προϊστάμενου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Κάθε υφιστάμενος πρέπει να γνωρίζει τις δικές του ανάγκες και ικανότητες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο υφιστάμενος εξαρτάται από τον ανώτερό του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο προϊστάμενος πρέπει να εξωτερικεύει τις επιθυμίες και τις ιδιαιτερότητές του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο προϊστάμενος είναι αναγκαίο να ενημερώνεται και για τις επιτυχίες και τις αποτυχίες της ομάδας του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Για το πόσα πράγματα πρέπει να ενημερωθεί εξαρτάται από τη διάθεση και την εμπιστοσύνη του ως προς τους υφισταμένους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Πρέπει να υπάρχει εμπιστοσύνη και εντιμότητα μεταξύ των δύο μελών	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Πρέπει ο υφιστάμενος να «βλέπει» τη θετική πλευρά στη δουλειά ακόμα και αν δέχεται επικριτικά σχόλια από τον προϊστάμενο του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ο υφιστάμενος πρέπει να ελέγχει τις αντιδράσεις του απέναντι στον προϊστάμενό του	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Πίνακας 2.31: Πίνακας ερωτηματολογίου 2 (ισορροπία σχέσης προϊσταμένου-υφισταμένου)

Στοιχεία Ερωτηθέντων

1. Φύλο:

Αντρας	<input type="radio"/>
Γυναίκα	<input type="radio"/>

Πίνακας 2.32: Πίνακας ερωτηματολογίου 2 (φύλλο)

2. Σε ποιο ηλικιακό γκρουπ ανήκετε;

Κάτω από 25	<input type="radio"/>
26-30	<input type="radio"/>
31-45	<input type="radio"/>
46-60	<input type="radio"/>
61 και άνω	<input type="radio"/>

Πίνακας 2.33: Πίνακας ερωτηματολογίου 2 (ηλικία)

3. Ποιο είναι το ανώτερο επίπεδο σπουδών που έχετε συμπληρώσει;

Απόφοιτος Λυκείου	<input type="radio"/>
Απόφοιτος ΑΕΙ/ ΤΕΙ	<input type="radio"/>
Κάτοχος Μεταπτυχιακού Διπλώματος	<input type="radio"/>
Άλλο:	

Πίνακας 2.34: Πίνακας ερωτηματολογίου 2 (εκπαιδευτικό επίπεδο)

Σε ποιο μαιευτήριο εργάζεστε;

4. Ποιος είναι ο τίτλος της θέσης εργασίας σας;

5. Πόσο χρονικό διάστημα εργάζεστε στο μαιευτήριο;

>1 χρόνο	<input type="radio"/>
1-5 χρόνια	<input type="radio"/>
6-10 χρόνια	<input type="radio"/>
<10 χρόνια	<input type="radio"/>

Πίνακας 2.35: Πίνακας ερωτηματολογίου 2 (έτη προϋπηρεσίας)

6. Πόσο χρονικό διάστημα είστε προϊστάμενος;

>1 χρόνο	<input type="radio"/>
1-5 χρόνια	<input type="radio"/>
6-10 χρόνια	<input type="radio"/>
<10 χρόνια	<input type="radio"/>

Πίνακας 2.36: Πίνακας ερωτηματολογίου 2 (έτη προϋπηρεσία σε θέση προϊσταμένου)

Στα παραπάνω ερωτηματολόγια παρατηρούμε ότι τα θέματα που αναλύονται (π.χ. επιρροή εργασίας στην οικογένεια, ικανοποίηση από τη δουλειά κ.ο.κ.) δεν καλύπτονται από μία ερώτηση, αλλά από σειρά ερωτήσεων που καλύπτουν το θέμα σφαιρικά. Συνήθως σε τέτοιες περιπτώσεις κατά τη στατιστική ανάλυση χρησιμοποιείται factor analysis, όπου από ένα σετ πολλών συσχετισμένων μεταβλητών προκύπτει ένα σετ λίγων και ασυσχέτιστων μεταβλητών που περιέχει σημαντικό μέρος της πληροφορίας του αρχικού σετ δεδομένων.

2.2.3 Ερωτηματολόγια ατομικής συμπλήρωσης

Οι βασικές μέθοδοι υποβολής ερωτηματολογίου σε ένα δείγμα είναι:

- Τα παραδοσιακά τυπωμένα ερωτηματολόγια ατομικής συμπλήρωσης που συμπληρώνονται είτε ταχυδρομικά είτε μέσω συνεντεύξεων,
- Οι έρευνες μέσω τηλεφώνου,
- Οι έρευνες μέσω διαδικτύου.

Η πιο συχνή μορφή ερωτηματολογίου ατομικής συμπλήρωσης είναι η ταχυδρομική έρευνα. Η διαδικασία έχει ως εξής: ένας εργαζόμενος στην έρευνα παραδίδει το ερωτηματολόγιο στο σπίτι του ερωτώμενου και του εξηγεί σχετικά τα θέματα που θέτει η έρευνα. Έπειτα αφήνει το ερωτηματολόγιο προκειμένου να συμπληρωθεί και αργότερα περνά να το παραλάβει.

Μπορεί, επίσης, να υπάρξει συνδυασμός τεχνικών, δηλ. της παράδοσης στο σπίτι και της ταχυδρόμησης του ερωτηματολογίου. Γενικά φαίνεται να είναι υψηλότερο το ποσοστό απαντήσεων σ' αυτή την περίπτωση, απ' όταν η έρευνα διεξάγεται απλά ταχυδρομικά.

2.2.4 Ταχυδρόμηση και επιστροφή

Στην έρευνα μέσω ταχυδρομείου, αποστέλλονται τα ερωτηματολόγια με σκοπό να συμπληρωθούν από τους ερωτώμενους και αναμένεται να ταχυδρομηθούν συμπληρωμένα. Στη μέθοδο αυτή υπάρχει ένα πρόβλημα: η μη επιστροφή του ερωτηματολογίου.

Για να μειωθούν στο ελάχιστο οι περιπτώσεις αυτές καλό είναι τεχνικά ζητήματα όπως το μέγεθος του φακέλου, η ευκολία αποστολής του φακέλου (όπου πρέπει σίγουρα να αναγράφεται η διεύθυνση του ερευνητή και να προσφέρεται το γραμματόσημο), ο καθορισμός του τρόπου αποστολής (συστημένο ή απλό, αν το επιτρέπει ο προϋπολογισμός των οικονομικών της έρευνας προφανώς θα πρέπει να είναι συστημένο) να αντιμετωπίζονται έγκαιρα και να επιλύονται με τον καλύτερο τρόπο.

2.2.5 Παρακολούθηση των επιστροφών

Αποτελεί πάντα σημαντική πληροφορία πόσοι και **ποιοι** απαντούν στα ερωτηματολόγια, γιατί έτσι μπορούμε να αποφύγουμε κάποια μεροληπτικά αποτελέσματα. Παράδειγμα: αν σε μία ιατρική έρευνα δούμε ότι το μεγάλο ποσοστό των ερωτηματολογίων που δεν απαντήθηκαν ήταν τα ερωτηματολόγια που είχαν δοθεί σε βαριά ασθενείς, το γεγονός θα αξιολογηθεί διαφορετικά από την περίπτωση που οι ερωτώμενοι, που δεν απάντησαν, ανήκαν σε άλλες κοινωνικές ομάδες.

2.2.6 Επαναληπτικές αποστολές

Καλό είναι σε ερωτώμενους που δεν απάντησαν να γίνουν επαναληπτικές αποστολές. Ο απλούστερος τρόπος, είναι τους σταλεί το ερωτηματολόγιο ξανά. Άλλος τρόπος είναι η αποστολή επιστολής σε όσους δεν επέστρεψαν το ερωτηματολόγιο, σε αυτήν την περίπτωση καλύτερα είναι να επισυνάπτεται στην επιστολή αντίγραφο του ερωτηματολογίου. Ο δεύτερος τρόπος πρέπει να προσεχθεί ώστε να μην φέρει εκνευρισμό στον ερωτώμενο. Οι επαναληπτικές είναι αποτελεσματική μέθοδος, καθώς ανεβάζει το ποσοστό απόκρισης μιας ταχυδρομικής έρευνας.

2.2.7 Ποσοστό απόκρισης

Σημαντικό είναι να εξετάσουμε το ποσοστό των ατόμων του δείγματος που απάντησαν στο ερωτηματολόγιο. Προφανώς όσο μεγαλύτερο είναι το δείγμα και όσο μεγαλύτερο είναι το ποσοστό απόκρισης τόσο το καλύτερο. Αυτό όμως δεν σημαίνει ότι υπάρχει κάποιο ποσοστό το οποίο εξασφαλίζει αντιπροσωπευτικό δείγμα. Όπως είπαμε και πριν έχει σημασία ποιοι δεν απαντούν (2.2.5).

Υψηλό ποσοστό απόκρισης σημαίνει μικρό ενδεχόμενο εμφάνισης μεροληψίας, ενώ υψηλό ποσοστό «μη απόκρισης», είναι ένα αρνητικό σημάδι. Συνήθως ποσοστό μεγαλύτερο του 70% εξασφαλίζει αντιπροσωπευτικά αποτελέσματα. Ενδιαφέρον έχει η παρουσία κινήτρων (π.χ. οικονομικών) που αποδεδειγμένα αυξάνουν το ποσοστό απόκρισης. Ακόμα όπως θα ειπωθεί και παρακάτω οι πρόσωπο με πρόσωπο συνεντεύξεις αυξάνουν το ποσοστό αυτό.

2.2.8 Δειγματοληπτικές έρευνες με χρήση συνεντεύξεων

Η συνέντευξη είναι μια εναλλακτική μέθοδος συλλογής δεδομένων δειγματοληπτικών ερευνών. Οι ερωτήσεις γίνονται προφορικά και οι απαντήσεις καταγράφονται από τους ερευνητές. Διεξάγονται συνήθως πρόσωπο με πρόσωπο ή μέσω τηλεφώνου. Σε έρευνες μεγάλου δείγματος τις περισσότερες φορές χρειάζεται πάνω από ένας συνεντευκτής.

Πλεονεκτήματα: υψηλό ποσοστό απόκρισης, οι ερωτώμενοι είναι περισσότερο απρόθυμοι να απορρίψουν μια συνέντευξη, ενώ με ευκολία μπορούν να πετάξουν το ερωτηματολόγιο που τους ταχυδρομήθηκε στα σκουπίδια. Επίσης οι συνεντευκτές μπορούν να διαφυλάξουν την ομαλή ροή των απαντήσεων, όταν υπάρχουν ερωτήσεις που προκαλούν σύγχυση, αφού ο ερευνητής μπορεί να διευκρινίσει ό,τι θέλει ή και να διορθώσει τον ερωτώμενο όταν υπάρχει παρανόηση της ερώτησης.

Στην περίπτωση πολλών συνεντευκτών θα πρέπει να διασφαλίζεται η ομοιομορφία των αποτελεσμάτων γι' αυτό θα πρέπει να τους δίνονται οδηγίες και να υπάρχει συντονισμός.

Η εμφάνιση του συνεντευκτή, ο τρόπος που απευθύνει τις ερωτήσεις και γενικά η συμπεριφορά του είναι απαραίτητο να μην επηρεάζουν την αντίληψη του ερωτώμενου για μια ερώτηση ή μια απάντηση. Για εξοικονόμηση χρόνου και χρήματος ο συνεντευκτής κάνει την έρευνα σε συγκεκριμένη γεωγραφική περιοχή. Η ουδετερότητα είναι πολύ σημαντική για τέτοια δείγματα, καθώς αν ο συνεντευκτής επηρεάσει με κάποιο τρόπο το δείγμα, τότε θα επηρεαστούν οι απαντήσεις.

Καλό θα ήταν οι ερευνητές-συνεντευκτές να προσέχουν:

- την εμφάνισή τους, να είναι προσεκτική και όχι εξεζητημένη,
- να έχουν καλή γνώση του ερωτηματολογίου πριν ξεκινήσει η διαδικασία της συνέντευξης,
- την πιστή αναπαραγωγή των ερωτήσεων,
- την καταγραφή των απαντήσεων με ηλεκτρονικά μέσα,
- να δίνονται διευκρινίσεις όταν οι απαντήσεις δείχνουν ότι δεν υπάρχει κατανόηση της ερώτησης.

2.2.9 Συντονισμός και έλεγχος

Όπως αναφέρθηκε για τον σωστό συντονισμό χρειάζεται εκπαίδευση των συνεντευκτών για τη διεξαγωγή της έρευνας. Αυτή ξεκινά βέβαια από τη συζήτηση περί γενικών κανόνων και μελέτη του ερωτηματολογίου, αλλά δεν αρκεί. Είναι πάντα καλό να υπάρχει μια προετοιμασία προδιαγραφών που θα συνδέονται με το κάθε ερωτηματολόγιο. Οι προδιαγραφές είναι επεξηγηματικά και διευκρινιστικά σχόλια για το πώς πρέπει να αντιμετωπίζονται οι δυσκολίες ή οι ασάφειες. Θα πρέπει να περιέχουν οδηγίες για τον χειρισμό τέτοιων καταστάσεων.

Μετά, τη μελέτη του ερωτηματολογίου πρέπει να κατανοηθεί πώς θα λειτουργήσει η συνέντευξη, για αυτό το καλύτερο είναι να γίνει μια επίδειξη συνέντευξης χωρίς διακοπές, με διευκρινίσεις στο τέλος. Μετά τη συνέντευξη επίδειξης οι συνεντευκτές μπορούν να κάνουν εξάσκηση παίρνοντας συνέντευξη ο ένας στον άλλο.

Χαρακτηριστικό είναι ότι στην απογραφή πληθυσμού οι συνεντευκτές παρακολουθούν σεμινάρια και ανά είκοσι άτομα είναι υπό την παρακολούθηση ενός υπευθύνου.

Πολλές φορές πριν οι συνεντευκτές πάρουν τις πρώτες συνεντεύξεις συμμετέχουν σε πιλοτική έρευνα προκειμένου να αντιμετωπίσουν τα όποια προβλήματα σε κανονικές συνθήκες.

2.2.10 Τηλεφωνικές έρευνες

Πλεονεκτήματα

- Είναι πολύ πιο οικονομικές από τις ταχυδρομικές.
- Κερδίζεται πολύ χρόνος σε σχέση με τις ταχυδρομικές.
- Η διαδικασία απλής τυχαίας δειγματοληψίας γίνεται σχετικά εύκολα εφόσον είναι γενικά εύκολη η πρόσβαση σε καταλόγους σταθερών τηλεφώνων.
- Η επιρροή του συνεντευκτή στον ερωτώμενο είναι πολύ περιορισμένη.

- Ο συνεντευκτής δεν διατρέχει κανένα κίνδυνο εφόσον δεν έρχεται σε επαφή με κόσμο και δεν χρειάζεται να επισκεφθεί ενδεχομένως επικίνδυνες περιοχές.

Μειονεκτήματα

- Συνήθως προκύπτουν από καταλόγους σταθερών τηλεφώνων και έτσι αποκλείονται αυτοί που δεν έχουν τηλέφωνο ή χρησιμοποιούν μόνο κινητό.
- Στους καταλόγους σταθερών τηλεφώνων δεν περιλαμβάνονται συνήθως απόρρητοι αριθμοί.
- Υπάρχει μεγάλη πιθανότητα να κλείσει ο ερωτώμενος το τηλέφωνο πριν το τέλος της συνέντευξης.
- Υπάρχει μεγάλη προκατάληψη σε αυτούς που δέχονται τηλέφωνα λόγο των διαδεδομένων πωλήσεων μέσω τηλεφώνων.

2.2.11 Ποσοστά απόκρισης σε έρευνες μέσω συνεντεύξεων

Το ποσοστό απόκρισης δεν είναι σταθερό στην πορεία του χρόνου. Μέχρι το τέλος της δεκαετίας του ενενήντα που πρωτοεμφανίστηκαν μαζικά οι δημοσκοπήσεις στην Ελλάδα και πραγματοποιούνταν κυρίως με συνεντεύξεις πρόσωπο με πρόσωπο το ποσοστό αυτό ήταν γενικά πολύ υψηλό. Μετά όμως η παρουσία πολλών εταιριών έφερε έναν κορεσμό στον κόσμο που υπό διάφορες συνθήκες είχε πρόθεση να συμμετάσχει. Τα τελευταία χρόνια για λόγους οικονομίας στη μεγάλη πλειοψηφία πραγματοποιούνται τηλεφωνικές έρευνες. Το ποσοστό απόκρισης είναι πάρα πολύ μικρό, λόγω των τηλεπωλήσεων που έχουν παίξει έναν καθοριστικό ρόλο στο σοβαρό πρόβλημα που αντιμετωπίζουν οι παραδοσιακές τηλεφωνικές έρευνες. Επιπλέον οι τηλεφωνητές και τα φωνητικά μηνύματα έχουν συμβάλει στο σοβαρό πρόβλημα που έχει προκύψει. Έτσι το ποσοστό απόκρισης θα είναι ένα σημαντικό ζήτημα που θα εξακολουθεί να πλήττει τις δειγματοληπτικές έρευνες γνώμης.

2.2.12 Διαδικτυακές έρευνες

Η διάδοση του διαδικτύου είχε άμεσο αντίκτυπο στη δημιουργία των διαδικτυακών ερευνών. Είναι ένα είδος έρευνας το οποίο παρά τα προφανή προβλήματα κερδίζει έδαφος συνεχώς, ειδικά στις έρευνες αγοράς. Η ιδανική τους εφαρμογή γίνεται όταν ο πληθυσμός είναι οι επισκέπτες ενός ιστοτόπου (website).

Σ' αυτή την περίπτωση ο ερευνητής και συντάκτης των ερωτήσεων καλό είναι να προσέξει τα ακόλουθα:

- Να περιορίσει κατά το δυνατό τον χρόνο της έρευνας.
- Αν χρειάζεται η χρήση κωδικού να αποφευχθεί η χρήση ενός «αριθμού» κ.λπ., καλύτερα να καταχωρείται το όνομα ή το e-mail σε κάποιο πεδίο.
- Να γνωστοποιηθεί ότι τα αποτελέσματα θα κοινοποιηθούν στους συμμετέχοντες στην έρευνα. Πολλές φορές αυτό το κίνητρο λειτουργεί ευεργετικά.
- Καλό είναι να υπάρχει σχεδιασμός στην ημερομηνία αποστολής των e-mail στους συμμετέχοντες, ανάλογα με το θέμα της έρευνας.
- Το link της έρευνας πρέπει να είναι σε εμφανές σημείο και άμεσα προσβάσιμο.
- Πολύ ευεργετικά λειτουργεί ένα σύστημα ανταμοιβής των συμμετεχόντων με υλική επιβράβευση σε περίπτωση ανταπόκρισης.

2.2.13 Δομή δειγματοληπτικής έρευνας

1. Πρώτα από όλα καταγράφεται ο σκοπός της έρευνας.
2. Με βάση τον σκοπό καταγράφονται οι επιμέρους στόχοι που μπορεί να αφορούν στη θεωρητική τεκμηρίωση του θέματος (θεωρητικοί στόχοι), ή στη μελέτη και τεκμηρίωση συμπερασμάτων και υποθέσεων από τα δεδομένα της δειγματοληπτικής μας έρευνας. Προφανώς κατά την εκπλήρωση των θεωρητικών στόχων θα πρέπει να γίνεται εννοιολόγηση

των θεμάτων που θα απασχολήσουν τον ερευνητή και να έχουν συλλεχθεί παραπομπές της βιβλιογραφίας από σχετικές έρευνες. Η επισκόπηση της βιβλιογραφίας μπορεί να εμπλουτίσει τους ερευνητικούς στόχους που έχουμε θέσει εξ αρχής, γι' αυτό καλό είναι να προηγείται των ερευνητικών στόχων και υποθέσεων.

3. Αφού τεθούν οι θεωρητικοί στόχοι θα πρέπει να αντιστοιχήσουμε ερωτήματα που πρέπει να απαντηθούν είτε δίνοντας ορισμούς (εννοιολόγηση) είτε από τη μελέτη άλλων εργασιών. Αντίστοιχα στους ερευνητικούς στόχους θα πρέπει να αντιστοιχήσουμε ερευνητικά ερωτήματα στη μελέτη καθώς και ερευνητικές προτάσεις (ερευνητικές υποθέσεις) που θα θέλουμε να δούμε αν ισχύουν ή όχι. Είναι χαρακτηριστικό ότι ερευνητικές υποθέσεις που έχουν εξεταστεί σε άλλες εργασίες καλό είναι να επανεξετάζονται. Πρέπει να σημειωθεί ότι η αιτιακή ανάλυση είναι σημαντική για τον έλεγχο των υποθέσεων. Θα πρέπει να οριστούν οι μεταβλητές που συνδέονται αιτιακά, καθώς και ο τρόπος μέτρησής τους. Επίσης, μπορεί να οριστεί ο βαθμός της σχέσης μεταξύ των δύο μεταβλητών. Είναι σημαντικός ο καθορισμός του ελέγχου, έτσι ώστε να αποδειχτεί ότι οι σχέσεις δεν είναι πλαστές.
4. Έχοντας όλα τα παραπάνω κατά νου κατασκευάζουμε το ερωτηματολόγιο το οποίο εκτός των άλλων θα πρέπει να μας δώσει δεδομένα κατάλληλα για επεξεργασία. Σ' αυτή τη φάση θα αποφασίσουμε για τα υποκείμενα της μελέτης και τη στατιστική ανάλυση.
5. Σχεδιασμός πιθανοτικής δειγματοληψίας ώστε να διατεθεί το ερωτηματολόγιο. Θα πρέπει να ληφθούν όλες οι παράμετροι δειγματοληψίας που αναφέρθηκαν, να γίνει καθορισμός μεγέθους δείγματος, αλλά και να εκτιμηθεί σωστά η διαδικασία σε χρόνο και κόστος.
6. Στατιστική Ανάλυση δεδομένων. Τα ερωτήματα και οι υποθέσεις θα πρέπει να απαντώνται με στατιστικούς ελέγχους και όχι περιγραφικά.
7. Εξαγωγή συμπερασμάτων λαμβάνοντας υπόψη τις αδυναμίες του δείγματος (σχεδόν πάντα θα υπάρχουν!).

Η παρακάτω μελέτη περίπτωσης προέρχεται από την απόφοιτη του μεταπτυχιακού προγράμματος «Διεθνής Διοικητική των Επιχειρήσεων» Ευαγγελία Τριανταφύλλου στα πλαίσια της διπλωματικής της εργασίας, επιβλέπων καθηγητής ήταν ο Μ. Χαλικιάς.

2.3 Μελέτη περίπτωσης: Κατασκευή ερωτηματολογίου δειγματοληπτικής έρευνας

Η Επίδραση της Οικονομικής Ύφεσης στα Συστήματα Αμοιβών και Ανταμοιβών σε Μεσαίες και Μεγάλες Επιχειρήσεις στην περιοχή Αττικής. Συντάκτης ερευνητικής πρότασης και ερωτηματολογίου: Τριανταφύλλου Ευαγγελία.

2.3.1 Σκοπός και στόχοι μελέτης

Η παρούσα ερευνητική εργασία έχει ως σκοπό τη μελέτη και διερεύνηση της δομής και της φιλοσοφίας των συστημάτων αμοιβών και ανταμοιβών εργαζομένων, όπως εφαρμόζονται από τις ελληνικές μεσαίες και μεγάλες επιχειρήσεις (με επικέντρωση στη γεωγραφική περιοχή της Αττικής) και την καταγραφή και διακρίβωση των συνεπειών της οικονομικής ύφεσης στις πολιτικές απασχόλησης και αμοιβών των παραπάνω επιχειρήσεων. Παράλληλα θα διερευνηθεί τόσο σε θεωρητικό, όσο και σε εμπειρικό επίπεδο, το φαινόμενο της ακαμψίας των μισθών και οι παράγοντες που οδηγούν σε αυτό, στο πλαίσιο των μεσαίων και μεγάλων επιχειρήσεων.

Για την επίτευξη του παραπάνω σκοπού, η μελέτη θα εστιάσει στους ακόλουθους στόχους:

Θεωρητικοί στόχοι

Θ.1. Θεωρητική μελέτη και ανάλυση της έννοιας της αμοιβής και των συστημάτων αμοιβών και ανταμοιβών, ως βασική λειτουργία της Διαχείρισης Ανθρώπινων Πόρων.

Θ.2. Μελέτη των βασικών θεωριών, που ερμηνεύουν το φαινόμενο και τα αίτια της ακαμψίας των αμοιβών, σε περίοδο οικονομικής ύφεσης.

Ερευνητικοί στόχοι

E.1. Να διερευνηθούν τα δομικά στοιχεία των συστημάτων αμοιβών, που εφαρμόζονται από τις μεσαίες και μεγάλες επιχειρήσεις και να εξεταστεί η συσχέτισή τους με τα «δημογραφικά» τους χαρακτηριστικά (έτη δραστηριότητας, μέγεθος, σύνθεση ανθρώπινου δυναμικού, κ.λπ.).

E.2. Να διερευνηθούν οι επιπτώσεις της υφιστάμενης οικονομικής ύφεσης στα οικονομικά μεγέθη των προς εξέταση εταιρειών και να διακριβωθούν οι μεταβολές στις ακολουθούμενες από τις επιχειρήσεις πολιτικές εργασίας και αμοιβών.

E.3. Να διερευνηθεί το φαινόμενο της ακαμψίας μισθών στις μεσαίες και μεγάλες επιχειρήσεις και να διακριβωθούν τα αίτια που οδηγούν σε αυτό.

2.3.2 Διατύπωση ερευνητικών ερωτημάτων

Μετά την παρουσίαση του σκοπού και των θεωρητικών και ερευνητικών στόχων της παρούσας μελέτης, ακολουθεί η διατύπωση των επιμέρους ερευνητικών ερωτημάτων, που ανακύπτουν από το θέμα.

1ος Στόχος (Θεωρητικός)

Θεωρητική μελέτη και ανάλυση της έννοιας της αμοιβής και των συστημάτων αμοιβών και ανταμοιβών, ως βασική λειτουργία της Διαχείρισης Ανθρώπινων Πόρων.

Θ.1.1. Πώς ορίζεται η έννοια της αμοιβής;

Θ.1.2. Πώς ορίζεται η έννοια του συστήματος αμοιβής και ανταμοιβής;

Θ.1.3. Ποιος είναι ο στρατηγικός ρόλος του συστήματος αμοιβής και ανταμοιβής στα πλαίσια των οικονομικών οργανισμών;

Θ.1.4. Ποια είναι τα συστατικά στοιχεία ενός συστήματος αμοιβής και ανταμοιβής;

Θ.1.5. Ποιοι παράγοντες επηρεάζουν τον σχεδιασμό ενός συστήματος αμοιβής και ανταμοιβής;

Θ.1.6. Σε ποια θεωρητικά μοντέλα βασίζεται ο σχεδιασμός ενός συστήματος αμοιβής και ανταμοιβής;

2ος Στόχος (Θεωρητικός)

Μελέτη των βασικών θεωριών, που ερμηνεύουν το φαινόμενο και τα αίτια της ακαμψίας των αμοιβών, σε περίοδο οικονομικής ύφεσης.

Θ.2.1. Πώς ορίζεται η έννοια της ακαμψίας αμοιβών και γιατί είναι σημαντική για την επιχείρηση;

Θ.2.2. Ποιες είναι οι θεωρητικές προσεγγίσεις που μελετούν το φαινόμενο της ακαμψίας;

1ος Στόχος (Ερευνητικός)

Διερεύνηση των δομικών στοιχείων των συστημάτων αμοιβών, που εφαρμόζονται από τις μεσαίες και μεγάλες επιχειρήσεις και της ενδεχόμενης συσχέτισής τους με τα «δημογραφικά» τους χαρακτηριστικά.

E.1. Τα «δημογραφικά» χαρακτηριστικά μιας επιχείρησης διαφοροποιούν τη δομή των εφαρμοζόμενων συστημάτων αμοιβής;

2ος Στόχος (Ερευνητικός)

Διερεύνηση των επιπτώσεων της υφιστάμενης οικονομικής ύφεσης στις ακολουθούμενες από τις επιχειρήσεις πολιτικές εργασίας και αμοιβών.

E.2.1. Έχουν μεταβάλλει οι επιχειρήσεις τις πολιτικές εργασίας και αμοιβών, κατά την περίοδο της οικονομικής κρίσης;

E.2.2. Ποιοι παράγοντες επηρεάζουν τις πολιτικές εργασίας και αμοιβών;

3ος Στόχος (Ερευνητικός)

Διερεύνηση του φαινομένου της ακαμψίας αμοιβών στις μεσαίες και μεγάλες επιχειρήσεις και διακρίβωση των αιτίων που οδηγούν σε αυτό.

Ε.3.1. Υπάρχει ακαμψία στους μισθούς των εργαζομένων στις ελληνικές μεσαίες και μεγάλες επιχειρήσεις, σε περίοδο οικονομικής ύφεσης;

Ε.3.2. Ποιοι μηχανισμοί-αίτια οδηγούν στην εμφάνιση φαινομένων ακαμψίας αμοιβών στις επιχειρήσεις;

Ε.3.3. Υπάρχει συσχέτιση των αιτίων που οδηγούν σε φαινόμενα ακαμψίας με τα χαρακτηριστικά της επιχείρησης (μέγεθος, σύνθεση ανθρώπινου δυναμικού);

2.3.3 Διατύπωση ερευνητικών υποθέσεων

Στην παρούσα ενότητα συνοψίζονται οι ερευνητικές υποθέσεις, όπως προέκυψαν από τη θεωρητική προσέγγιση, του υπό διερεύνηση προβλήματος, και αντιστοιχίζονται με τους ερευνητικούς στόχους της μελέτης (Πίνακας 37).

Ερευνητικός Στόχος	Υποθέσεις
1ος Ερευνητικός Στόχος Διερεύνηση των δομικών στοιχείων των συστημάτων αμοιβών σε μεσαίες και μεγάλες επιχειρήσεις και της συσχέτισή τους με τα «δημογραφικά» χαρακτηριστικά των επιχειρήσεων	H1: Υπάρχει θετική συσχέτιση μεταξύ του μεγέθους της επιχείρησης και της χρήσης ευέλικτων μορφών αμοιβής
2ος Ερευνητικός Στόχος Διερεύνηση των επιπτώσεων της υφιστάμενης οικονομικής ύφεσης στις ακολουθούμενες από τις επιχειρήσεις πολιτικές εργασίας και αμοιβών	H2: Υπάρχει θετική συσχέτιση μεταξύ του μεγέθους της επιχείρησης και της χρήσης μη-μισθολογικών μεθόδων, για τον περιορισμό του εργατικού τους κόστους
	H3: Η ένταση των επιπτώσεων της οικονομικής κρίσης στην επιχείρηση, σχετίζεται με την εφαρμογή των στρατηγικών περιορισμού του εργατικού τους κόστους
	H4: Υπάρχει συσχέτιση μεταξύ της συνδικαλιστικής οργάνωσης των εργαζομένων και της εφαρμογής έμμεσων στρατηγικών περιορισμού του μη-μισθολογικού κόστους
3ος Ερευνητικός Στόχος Διερεύνηση του φαινομένου της ακαμψίας μισθών στις υπό εξέταση επιχειρήσεις και διακρίβωση των αιτίων που οδηγούν σε αυτό	H5: Οι επιχειρήσεις αποφεύγουν να μειώσουν τους μισθούς τους ακόμη και όταν αντιμετωπίζουν πτώση της ζήτησης των προϊόντων τους
	H6: Υπάρχει θετική συσχέτιση μεταξύ της ύπαρξης ακαμψίας των μισθών και της εφαρμογής αυστηρού θεσμικού πλαισίου λειτουργίας της αγοράς εργασίας και συλλογικών συμβάσεων καθορισμού των κατώτατων ορίων αμοιβών
	H7: Υπάρχει συσχέτιση μεταξύ των μηχανισμών ακαμψίας και των δημογραφικών χαρακτηριστικών των επιχειρήσεων

Πίνακας 2.37: Αντιστοίχιση Ερευνητικών Υποθέσεων και Ερευνητικών Στόχων

Συμπεράσματα

- Πρακτική εφαρμογή αποτελεσμάτων (managerial implications),
- Συστάσεις για μελλοντική έρευνα,
- Περιορισμοί έρευνας.

Ερωτηματολόγιο 3

Θέμα: «Η επίδραση της οικονομικής ύφεσης στα συστήματα αμοιβών και ανταμοιβών σε μεσαίες και μεγάλες επιχειρήσεις, στην περιοχή της Αττικής».

Πρώτη Θεματική Ενότητα

1.1 Εκτός του μισθού, ποια από τα ακόλουθα στοιχεία περιλαμβάνει η επιχείρησή σας, στο σύστημα αμοιβής που εφαρμόζει;

	Υψηλόβαθμα διοικητικά στελέχη	Ειδικευμένο τεχνικό προσωπικό	Διοικητικοί υπάλληλοι	Ανειδίκευτοι εργάτες
Πρόγραμμα διανομής ή προνομιακής αγοράς μετοχών				
Συμμετοχή στα κέρδη				
Bonus βασισμένο στην ατομική απόδοση / Στόχους				
Bonus βασισμένο στην ομαδική απόδοση / Στόχους				

Πίνακας 2.38: Πίνακας ερωτηματολογίου 3 (συστήματα αμοιβών)

1.2 Η επιχείρησή σας παρέχει στους εργαζόμενους, κάποια από τα ακόλουθα προνόμια, πέραν των νόμιμων υποχρεώσεων της;

Πρόγραμμα διακοπής εργασίας	
Άδεια μητρότητας/πατρότητας/γονική άδεια	
Συνταξιοδοτικό πρόγραμμα	
Εκπαιδευτική άδεια	
Πρόγραμμα ιδιωτικής ασφάλειας υγείας	
Κανένα από τα παραπάνω	

Πίνακας 2.39: Πίνακας ερωτηματολογίου 3 (προνόμια εργαζομένων)

Δεύτερη Θεματική Ενότητα

2.1 Την τελευταία πενταετία έχετε μειώσει τους μισθούς των εργαζομένων της επιχείρησης και αν ναι σε τι ποσοστό του προσωπικού;

Όχι	<input type="radio"/>
Ναι	Σε <input type="text"/> % του προσωπικού

Πίνακας 2.40: Πίνακας ερωτηματολογίου 3 (μειώσεις μισθών)

2.2 Παρακαλώ δηλώστε κατά πόσο είναι αποτρεπτική, καθεμιά από τις παρακάτω προτάσεις στην απόφαση του οργανισμού σας, να μην μειώσει τους μισθούς του προσωπικού της.

	Καθόλου	Λίγο	Αρκετά	Πάρα πολύ
Δεν επιτρέπεται από την ισχύουσα νομοθεσία (π.χ. συλλογικές συμβάσεις)				
Θα είχε αρνητικό αντίκτυπο στο ηθικό των εργαζομένων				
Θα οδηγούσε σε μείωση της έντασης και της ποιότητας της εργασίας των εργαζομένων				
Θα είχε αρνητικό αντίκτυπο στη φήμη της εταιρείας, ως εργοδότη				
Θα οδηγούσε σε ενδεχόμενη αποχώρηση των πιο παραγωγικών εργαζομένων				
Η αντικατάσταση παλαιών εργαζομένων, με νεότερους και χαμηλότερα αμειβόμενους εργαζόμενους, θα οδηγούσε τους εναπομείναντες παλαιούς σε άρνηση συνεργασίας με το νέο προσωπικό				
Λόγω της αποχώρησης εργαζομένων, θα δημιουργούνταν υψηλά κόστη πρόσληψης και εκπαίδευσης των νέων εργαζομένων				
Η αποχώρηση παλαιών εργαζομένων, θα σήμαινε και απώλεια τμήματος της τεχνογνωσίας της επιχείρησης				
Θα δημιουργούσε δυσκολίες στην προσέλκυση νέων εργαζομένων				
Στους εργαζόμενους δεν αρέσουν οι απρόβλεπτες μειώσεις αποδοχών. Προτιμούν να έχουν ένα σταθερό σύστημα αμοιβής				

Οι εργαζόμενοι συγκρίνουν τις αποδοχές που λαμβάνουν εργαζόμενοι των ίδιων προσόντων στον ίδιο κλάδο εργασίας

Πίνακας 2.41: Πίνακας ερωτηματολογίου 3 (παράγοντες συγκράτησης των μισθών)

2.3 Παρακαλώ δηλώστε, κατά πόσο είναι σχετική καθεμιά από τις παρακάτω προτάσεις ως στρατηγική αντιμετώπισης, μιας απρόσμενης μείωσης της ζήτησης των προϊόντων της επιχείρησής σας.

	Καθόλου	Λίγο	Αρκετά	Πάρα πολύ
Μείωση τιμών πώλησης προϊόντων και υπηρεσιών				
Μείωση περιθωρίου κέρδους				
Μείωση παραγωγής				
Μείωση κόστους				

Πίνακας 2.42: Πίνακας ερωτηματολογίου 3 (στρατηγικές απρόσμενης ζήτησης)

2.3.1 Παρακαλώ αξιολογήστε τη σημαντικότητα καθενός από τα ακόλουθα μέτρα για τη μείωση του κόστους της επιχείρησης.

	Καθόλου	Λίγο	Αρκετά	Πάρα πολύ
Μείωση μισθών				
Απολύσεις μόνιμου προσωπικού				
Γενική μείωση του προσωπικού της επιχείρησης (συμπεριλαμβανομένων και των εργαζομένων με συμβάσεις ορισμένου χρόνου)				
Χρήση υπηρεσιών εταιρειών «προσωρινής απασχόλησης»				
Μετατροπή συμβάσεων πλήρους απασχόλησης σε μερικής (ευέλικτες μορφές απασχόλησης)				
Ενθάρρυνση αποχώρησης εργαζομένων, μέσω προγραμμάτων πρόωρης συνταξιοδότησης				

Πρόσληψη νέων εργαζομένων σε χαμηλότερα επίπεδα αμοιβής από αυτούς που αποχώρησαν οικειοθελώς				
Μείωση ή κατάργηση των ατομικών και ομαδικών βραβείων απόδοσης (bonus)				
Μείωση ή κατάργηση πρόσθετων παροχών (π.χ. πακέτα υγείας, διατακτικές κ.ά.)				
Αλλαγή προγραμμάτων εργασίας / εργασίας σε βάρδιες				
Πάγωμα ή επιβράδυνση προαγωγών				
Δεν θα προσπαθούσαμε να μειώσουμε το κόστος εργασίας				

Πίνακας 2.43: Πίνακας ερωτηματολογίου 3(παράγοντες μείωσης κόστους)

2.3.2 Έχετε χρησιμοποιήσει κάποιο/κάποια από τα παραπάνω μέτρα, την τελευταία πενταετία;

1. Μείωση μισθών	<input type="radio"/>
2. Απολύσεις μόνιμου προσωπικού	<input type="radio"/>
3. Γενική μείωση του προσωπικού της επιχείρησης (συμπεριλαμβανομένων και των εργαζομένων με συμβάσεις ορισμένου χρόνου)	<input type="radio"/>
4. Χρήση υπηρεσιών εταιρειών «προσωρινής απασχόλησης»	<input type="radio"/>
5. Μετατροπή συμβάσεων πλήρους απασχόλησης σε μερικής (ευέλικτες μορφές απασχόλησης)	<input type="radio"/>
6. Ενθάρρυνση αποχώρησης εργαζομένων, μέσω προγραμμάτων πρόωρης συνταξιοδότησης	<input type="radio"/>
7. Πρόσληψη νέων εργαζομένων σε χαμηλά επίπεδα αμοιβής, σε θέσεις ατόμων που αποχώρησαν οικειοθελώς	<input type="radio"/>
8. Μείωση ή κατάργηση των ατομικών και ομαδικών βραβείων απόδοσης (bonus)	<input type="radio"/>
9. Μείωση ή κατάργηση πρόσθετων παροχών (π.χ. πακέτα υγείας, διατακτικές κ.ά.)	<input type="radio"/>
10. Αλλαγή προγραμμάτων εργασίας / εργασίας σε βάρδιες	<input type="radio"/>
11. Πάγωμα ή επιβράδυνση προαγωγών	<input type="radio"/>

12. Δεν προσπαθήσαμε να μειώσουμε το κόστος εργασίας



Πίνακας 2.44: Πίνακας ερωτηματολογίου 3 (μέτρα περικοπών που έχουν ληφθεί)

2.4 Λαμβάνοντας υπόψη τις αλλαγές στο ευρύτερο οικονομικό περιβάλλον και ειδικότερα στην αγορά εργασίας, τα τελευταία πέντε έτη, πιστεύετε ότι έχει γίνει ευκολότερη η προσαρμογή των μισθών, προκειμένου να μειωθεί το εργατικό κόστος;

Ναι	<input type="radio"/>
Όχι	<input type="radio"/>

Πίνακας 2.45: Πίνακας ερωτηματολογίου 3 (επιρροή οικονομικής κρίσης στην προσαρμογή μισθών)

2.4.1 Αν απαντήσατε **Ναι**, παρακαλώ υποδείξτε το λόγο που θεωρείτε σημαντικότερο (επιλέξτε μόνο μια απάντηση)

Εξαιτίας της αυξημένης ανεργίας, υπάρχουν περισσότεροι διαθέσιμοι υποψήφιοι, που είναι διαθέσιμοι να εργαστούν για λιγότερα χρήματα, από το μέσο μισθό στον οποίο αποτιμάται η αντίστοιχη θέση εργασίας	<input type="radio"/>
Τα συνδικαλιστικά σωματεία είναι λιγότερο ισχυρά σε σχέση με το παρελθόν	<input type="radio"/>
Οι διαρθρωτικές αλλαγές στην αγορά εργασίας, αφήνουν λιγότερο προστατευμένους τους εργαζόμενους από όσο στο παρελθόν	<input type="radio"/>

Πίνακας 2.46: Πίνακας ερωτηματολογίου 3 (αιτίες προσαρμογής μισθών)

Τρίτη Θεματική Ενότητα

3.1 Ποιο είναι το έτος ίδρυσης της επιχείρησής σας;

3.2 Ποιος είναι ο κύριος τομέας δραστηριότητας της επιχείρησής σας;

Βιομηχανία - Μεταποίηση	<input type="radio"/>	Χονδρικό - Λιανικό Εμπόριο	<input type="radio"/>
Τομέας Ενέργειας	<input type="radio"/>	Μεταφορές - Αποθήκευση	<input type="radio"/>
Κατασκευή	<input type="radio"/>	Παροχή υπηρεσιών	<input type="radio"/>
Ενημέρωση επικοινωνία	<input type="radio"/>		
Άλλο (παρακαλώ προσδιορίστε)	<input type="text"/>		

Πίνακας 2.47: Πίνακας ερωτηματολογίου 3(τομείς της επιχείρησης)

3.3 Η επιχείρησή σας είναι μέλος κάποιας εργοδοτικής οργάνωσης;

Ναι	<input type="radio"/>
Όχι	<input type="radio"/>

Πίνακας 2.48: Πίνακας ερωτηματολογίου 3 (εργοδοτική οργάνωση)

3.4 Πόσα άτομα απασχολεί συνολικά η επιχείρησή σας, κατά την περίοδο διενέργειας της έρευνας;

(εξαιρούνται οι εργαζόμενοι με συμβάσεις ανεξαρτήτων υπηρεσιών και γενικότερα οι εξωτερικοί συνεργάτες της επιχείρησης)

Αριθμός εργαζομένων

3.5 Πώς κατανέμεται, κατά προσέγγιση, το προσωπικό της επιχείρησής σας, στις ακόλουθες ομάδες απασχόλησης;

Χαμηλής ειδίκευσης προσωπικό (εργάτες παραγωγής, βοηθητικό προσωπικό κ.λ.π.)	<input type="text"/>
Διοικητικοί υπάλληλοι	<input type="text"/>
Ειδικευμένο τεχνικό προσωπικό και εργοδηγοί	<input type="text"/>
Υψηλά ειδικευμένοι υπάλληλοι και στελέχη διοίκησης	<input type="text"/>

Πίνακας 2.49: Πίνακας ερωτηματολογίου 3(κατανομή προσωπικού)

3.6 Σε τί ποσοστό του συνολικού προσωπικού της επιχείρησής, ο μισθός/ημερομίσθιο (ονομαστικός) είναι στα όρια των κατώτατων νομοθετημένων ορίων (Ε.Γ.Σ.Σ.Ε. ή άλλης κλαδικής ή ομοιοεπαγγελματικής σύμβασης εργασίας) ή πλησίον αυτών (+/- 10%);

3.7 Υπάρχει σωματείο εργαζομένων στην επιχείρησή σας;

Ναι	<input type="radio"/>
Όχι	<input type="radio"/>

Πίνακας 2.50: Πίνακας ερωτηματολογίου 3 (εργατικό σωματείο)

3.7.1 Αν απαντήσατε ΝΑΙ, τί ποσοστό των εργαζομένων συμμετέχουν σε αυτό;

Ποσοστό εργαζομένων στο σωματείο %

3.8 Συγκρίνοντας τον κύκλο εργασιών της επιχείρησής σας κατά το προηγούμενο οικονομικό έτος, θα λέγατε ότι σε σχέση με εκείνον του 2009 (αρχή κρίσης) είναι:

Πολύ χαμηλότερος	<input type="radio"/>
Χαμηλότερος	<input type="radio"/>
Περίπου ο ίδιος	<input type="radio"/>
Υψηλότερος	<input type="radio"/>
Πολύ Υψηλότερος	<input type="radio"/>

Πίνακας 2.51: Πίνακας ερωτηματολογίου 3 (κύκλος εργασιών)

Προαιρετικές Πληροφορίες

Όνομα του ατόμου που απαντά στην έρευνα:

Θέση στην εταιρεία:

Ηλεκτρονική διεύθυνση επικοινωνίας:

Παρατήρηση

Στο ερωτηματολόγιο θα πρέπει να δίνεται η δυνατότητα να απαντηθούν τα ερευνητικά ερωτήματα (πολλές φορές ερευνητικά ερωτήματα είναι και ερωτήσεις του ερωτηματολογίου). Ακόμα θα πρέπει να δίνεται η δυνατότητα απάντησης στις ερευνητικές υποθέσεις μέσω κατάλληλων στατιστικών ελέγχων. Ως εκ τούτου θα πρέπει να υπάρχουν οι ανάλογες ερωτήσεις που θα καταχωρηθούν ως μεταβλητές για την πραγματοποίηση του ελέγχου.

2.4 Δημοσκοπήσεις

2.4.1 Εισαγωγικά

Το πιο διαδεδομένο είδος δειγματοληπτικών ερευνών στην Ελλάδα είναι οι δημοσκοπήσεις. Οι δημοσκοπήσεις ουσιαστικά αποτελούν μέρος του σημερινού πολιτικού συστήματος παγκοσμίως και η επίδρασή τους δεν περιορίζεται στην περιγραφή και στην καταγραφή του πολιτικού σκηνικού, αλλά σε μεγάλο βαθμό επιδρά στη διαμόρφωσή του.

Είναι χαρακτηριστικό ότι επιλογή αποφάσεων υποψηφίων κ.ο.κ. εξαρτάται από αποτελέσματα δημοσκοπήσεων.

Συχνά οι δημοσκοπήσεις έχουν πέσει έξω ειδικά τα τελευταία χρόνια όπου η ψήφος στις ελληνικές βουλευτικές εκλογές έχει γίνει ιδιαίτερα ρευστή. Παρόλα αυτά η ισχύς της επιρροής τους παραμένει και αποτελέσματά τους χρησιμοποιούνται ως επιχειρήματα στον πολιτικό διάλογο. Οι δημοσκοπήσεις γίνονται εργαλεία χάραξης στρατηγικής είτε σε έρευνες αγοράς είτε σε στρατηγική πολιτικών στελεχών και πολιτικών κομμάτων.

Υπάρχουν δύο μεγάλες κατηγορίες δημοσκοπήσεων: α) Οι δημοσκοπήσεις που παρουσιάζονται στα ΜΜΕ και β) Οι δημοσκοπήσεις που αφορούν ιδιώτες, κόμματα, επιχειρήσεις, διαφημιστικές εταιρίες κ.ο.κ. Η πρώτη κατηγορία δημοσκοπήσεων διαθέτει και το μεγαλύτερο τζίρο και εξασφαλίζει και δημοσιοποίηση των αποτελεσμάτων.

Για τις δημοσκοπήσεις στην Ελλάδα έχει γίνει πολύ καλή αναφορά στο διαδικτυακό άρθρο του Χρ. Βερναρδάκη στο συλλογικό τόμο αφιέρωμα στη Λαμπίρη Δημάκη (Αφουξενίδης & Αλεξάκης 2006). Στην Ελλάδα οι δημοσκοπήσεις πραγματοποιούνται από το 1989 και μετά. Πριν βρουν έδαφος και εφαρμοστούν οι δημοσκοπήσεις στην Ευρώπη και στις ΗΠΑ είχε καλλιεργηθεί το κατάλληλο υπόβαθρο δεοντολογίας και αξιοπιστίας για τη διενέργειά τους. Η μεγάλη αυτή καθυστέρηση σε σχέση με τις υπόλοιπες ευρωπαϊκές χώρες, οφείλεται πρώτα στην αργή εξέλιξη των επιστημονικών εξελίξεων σε σχέση με τις ευρωπαϊκές χώρες αλλά και στην καχυποψία με την οποία αντιμετώπιζε ο κόσμος τις δημοσκοπήσεις λόγω της δικτατορίας και του εμφυλίου. Τεράστια ώθηση και προβολή έδωσαν τα exit polls που εμφανίστηκαν στις εκλογές του 1996. Η οικονομική ανάπτυξη τη δεκαετία του 90 στην Ελλάδα σε συνδυασμό με την άνθιση των ιδιωτικών καναλιών που ήταν οι βασικοί πυλώνες προβολής τους και χρηματοδότησής τους δημιούργησαν την ανάπτυξη των δημοσκοπήσεων και την εγκαθίδρυσή τους στα ελληνικά δρώμενα.

Στην Ελλάδα οι δημοσκοπήσεις επηρεάζουν ποικιλοτρόπως τη διαμόρφωση της κοινής γνώμης, γιατί και οι δημοσιογράφοι προσαρμόζουν την άποψή τους και την τοποθέτηση που ακολουθούν με βάση τα αποτελέσματα δημοσκοπήσεων. Ακόμα η βαρύτητα των θεμάτων και η ατζέντα των συζητήσεων υιοθετούνται από δημοσκοπήσεις με στόχο πάντα το μεγαλύτερο ενδιαφέρον και τη μεγαλύτερη τηλεθέαση.

Η βαρύτητα αυτή έχει ως αποτέλεσμα πολλές φορές την κατάχρηση, τη δόλια χρήση τους και την σκόπιμη παρερμηνεία των αποτελεσμάτων τους.

Οι συγκεκριμένες έρευνες είναι συγχρονικές και αφορούν στη στιγμή που πραγματοποιούνται όπως όλες οι δειγματοληπτικές έρευνες. Όπως παρατηρεί και ο Grémy (1993), οι απαντήσεις, σε μεγάλο βαθμό, αποτελούν απόρροια της διατύπωσης της ερώτησης που πραγματοποιείται.

Όπως αναφέρει ο Χρ. Βερναρδάκης υπάρχουν δύο βασικές φιλοσοφίες έρευνας για την κοινωνική έρευνα και κατ' επέκταση, για τις δημοσκοπήσεις και τις δειγματοληπτικές έρευνες. Η πρώτη στηρίζεται στα ποσοστά των απαντήσεων και την επεξεργασία με ποσοτική μέθοδο, ενώ η δεύτερη αφορά την εκτίμηση της στάσης των ερωτώμενων με ποιοτική επεξεργασία αποτελεσμάτων, ώστε οι απαντήσεις να προσλαμβάνονται

ως μια ολότητα (π.χ ο χαρακτηρισμός των απαντήσεων του ερωτώμενου ότι αποτελεί αριστερή στροφή σε σχέση με τις απόψεις που είχε πριν από ένα γεγονός κ.ο.κ.).

Η πρώτη προσέγγιση είναι πιο διαδεδομένη και εύκολη στην επεξεργασία, τη δεύτερη την ανέδειξαν οι μελετητές κοινωνικής συμπεριφοράς (Michelat & Thomas, 1966· Michelat & Simon, 1977). Πρακτικά τα δύο όρια των προσεγγίσεων αυτών δεν είναι πάντοτε ξεκάθαρα και ο συνδυασμός τους είναι η ενδεδειγμένη διαδικασία (Rivière & Martelli-Banégas, 2002· Grunberg, 2002).

2.4.2 Οι δημοσκοπήσεις σήμερα

Ως δημοσκόπηση ορίζεται το αποτέλεσμα της δειγματοληπτικής έρευνας, που γίνεται σε ένα σύνολο ανθρώπων, με σκοπό να ακουστεί η γνώμη του πληθυσμού για ένα συγκεκριμένο ζήτημα.

Εκτός της πρόθεσης ψήφου και των πολιτικών θεμάτων οι δημοσκοπήσεις, ασχολούνται κυρίως με τους τομείς των πωλήσεων, το μάρκετινγκ, έρευνες αγοράς, κ.λπ.

Στην Ελλάδα υπάρχουν περίπου 30 ιδιωτικές εταιρείες δημοσκοπήσεων. Αξιοσημείωτο είναι ότι σε πρόσφατες αμερικανικές εκλογές, το κόστος ήταν έξι εκατομμύρια δολάρια (\$6.000.000) για περίπου διακόσιες δημοσκοπήσεις που διενεργήθηκαν.

Λόγω της σημασίας των δημοσκοπήσεων πολλές φορές έχει αμφισβητηθεί αν λένε την αλήθεια. Σε περίπτωση που τα λάθη δεν είναι επιτηδευμένα διακρίνουμε τρεις κατηγορίες αποτελεσμάτων που προϋποθέτουν και άλλη μορφή ανάλυσης.

Η πρώτη είναι οι συνηθισμένες δημοσκοπήσεις όπου καταγράφεται προεκλογικά η πρόθεση ψήφου, η δεύτερη είναι τα exit-polls δηλαδή οι δημοσκοπήσεις που πραγματοποιούνται την ημέρα των εκλογών, έξω από εκλογικά τμήματα και η τρίτη οι εκτιμήσεις και οι προβλέψεις αποτελεσμάτων με βάση τα πρώτα δεδομένα (π.χ. τα πρώτα καταγεγραμμένα ψηφοδέλτια). Η εκτίμηση γίνεται με εφαρμογή στατιστικών μοντέλων. Η δεύτερη και η τρίτη, αφορούν αποκλειστικά την ημέρα των εκλογών, έτσι το αποτέλεσμα τους κρίνεται άμεσα από το κοινό και από το πολιτικό σύστημα. Είναι σχεδιασμένες έτσι ώστε να προσφέρουν πολύ μεγαλύτερη ακρίβεια μέτρησης σε σχέση με τις απλές δημοσκοπήσεις. Για τη δεύτερη κατηγορία καθοριστική είναι η σωστή δειγματοληψία, από εκεί και πέρα τα περιθώρια λάθους είναι ελάχιστα και έχουν να κάνουν με τη διατύπωση των ερωτήσεων. Στην τρίτη κατηγορία η εκτίμηση της πρόθεσης ψήφου αυτών που δεν συνεργάστηκαν με την έρευνα είναι ιδιαίτερα δύσκολη και απαιτεί χρήση στατιστικών μοντέλων.

Η πρώτη περίπτωση δημοσκοπήσεων είναι δύσκολο να δώσει ακριβή αποτελέσματα. Ακόμα τα αποτελέσματα των δημοσκοπήσεων αυτών είναι πρακτικά αδύνατο να ελεγχθούν. Τα λάθη προέρχονται κυρίως από άρνηση συμμετοχής, κακή δειγματοληψία αλλά και λανθασμένη πρόβλεψη για το ποσοστό αναποφάσιστων.

Είναι γεγονός ότι έχει δημιουργηθεί η αίσθηση ότι οι δημοσκοπήσεις, αντί να καταγράφουν τη στάση της κοινής γνώμης, στην πραγματικότητα τη διαμορφώνουν. Το φαινόμενο αυτό οφείλεται κυρίως στην αυξημένη προβλεπτική ικανότητα που αποδίδεται στις δημοσκοπήσεις σε σχέση με αυτήν που πραγματικά έχουν. Το προβάδισμα κάποιου κόμματος στις δημοσκοπήσεις μπορεί να δημιουργήσει το σύνδρομο του νικητή (bandwagon effect), το οποίο με τη σειρά του είναι ικανό να παρασύρει τους ταλαντευόμενους ψηφοφόρους και έτσι ο νικητής με βάση τις δημοσκοπήσεις ενισχύεται στην εκλογική διαδικασία. Ακόμα η επιρροή των δημοσκοπήσεων έχει να κάνει και με την επιλογή κομμάτων ή ατόμων που χρειάζεται να «ενισχυθούν», εφόσον σύμφωνα με τις δημοσκοπήσεις η πραγματική πρώτη προτίμηση του ψηφοφόρου έχει εξασφαλισμένο καλό ή κακό αποτέλεσμα.

Αξιοσημείωτο είναι ότι στην Ελλάδα για κάποια χρόνια απαγορεύτηκε η δημοσιοποίηση δημοσκοπήσεων για ορισμένο χρονικό διάστημα πριν από τις εκλογές. Αυτό όμως δεν κρίθηκε χρήσιμο ή αποτελεσματικό, εφόσον δεν ήταν δυνατό να απαγορευτεί η δημοσιοποίηση σε άλλες χώρες ή μέσω internet κ.λπ.

2.4.3 Σχετικές έρευνες

Οι σύγχρονες έρευνες στοχεύουν κυρίως στην άμεση και γρήγορη πρόβλεψη μέσω internet. Κατά την έρευνα των Metaxas et al. (2011), εφάρμοσαν μοντέλα πρόβλεψης των μέσων κοινωνικής δικτύωσης για πολιτικές δημοσκοπήσεις κατά τις δύο τελευταίες εκλογές μελών του Κογκρέσου στις ΗΠΑ.

Το μεγάλο πρόβλημα που εντοπίστηκε είναι ότι είναι εύκολο να δημιουργηθούν πολλοί ψεύτικοι λογαριασμοί και προφίλ χρηστών, ώστε να ενισχυθεί το μήνυμα υπέρ κάποιου υποψηφίου όπως έχουν επισημάνει οι (Mustafaraj et al., 2011).

Είναι βέβαιο ότι δεδομένα μέσων κοινωνικής δικτύωσης αυξάνουν την προβλεψιμότητα των αποτελεσμάτων των εκλογών. Αλλά είναι δύσκολο να δημιουργηθεί μια μεθοδολογία λόγω της αλματώδους εξέλιξής τους (Livne et al., 2011).

Οι O'Connor, B. et.al (2010) προσπάθησαν να χρησιμοποιήσουν δεδομένα ερωτηματολογίων που εξήχθησαν από μέσα κοινωνικής δικτύωσης, τα λεγόμενα στοιχεία της «αίσθησης» και του γενικότερου κλίματος που μετρώνται από τα κείμενα που περιέχονται στα tweets του Twitter.

Οι Visser, P. S., Krosnick, J. A., Marquette, J., και Curtin, M. (1996), τονίζουν τη διαφορά στο βαθμό εγκυρότητας των δημοσκοπήσεων μέσω ταχυδρομείου που εφαρμόζονταν τη δεκαετία του 80 με τις σημερινές και με αυτό τον τρόπο θέλουν να τονίσουν ότι το μέλλον των δημοσκοπήσεων είναι μέσω internet.

2.4.4 Τρόποι διεξαγωγής των δημοσκοπήσεων

Οι τεχνικές πραγματοποίησης των δημοσκοπήσεων είναι αυτές που χρησιμοποιούνται στις δειγματοληπτικές έρευνες (μέσω τηλεφώνου, μέσω email, μέσω ταχυδρομείου). Όμως λόγω της φύσης των ερευνών μπορούν να πραγματοποιηθούν και με sms και βέβαια ξεχωριστή σημασία έχουν τα exit-polls.

2.4.4.1 Exit-Polls

Οι χρηματοδότες των exit-polls είναι κατά βάση τα ΜΜΕ. Η ακρίβειά τους είναι συνήθως πολύ μεγάλη και είναι χαρακτηριστικό ότι για αμφιλεγόμενου κύρους εκλογικές αναμετρήσεις έχουν χρησιμοποιηθεί προκειμένου να εντοπίσουν πιθανή νοθεία αποτελεσμάτων. Η κυριότερη διαφορά εκτός από τον χώρο και τον χρόνο που γίνονται είναι ότι ο ερωτώμενος καλείται να απαντήσει στο τί έχει πραγματικά ψηφίσει και όχι τι προτίθεται να κάνει.

Ακόμα χρησιμοποιούνται για να συνδέσουν την ψήφο με κοινωνικά χαρακτηριστικά στα οποία καλείται να απαντήσει ο ερωτώμενος. Αυτή η πληροφορία μπορεί να προέλθει μόνο από τη διαδικασία της δημοσκόπησης των exit-polls, καθώς είναι γνωστό ότι η ψηφοφορία είναι μυστική και ανυπόγραφη. Αξιοσημείωτη αποτυχία των exit-polls σημειώθηκε το 1992 στη Μεγάλη Βρετανία και τις χρονιές 2012 και 2015 στις Ελληνικές εκλογές. Το λάθος στις Ελληνικές εκλογές αποδόθηκε, στο γεγονός ότι ψηφοφόροι συγκεκριμένων κομμάτων αρνήθηκαν να συμμετάσχουν στη διαδικασία. Τέτοιες περιπτώσεις, τα τελευταία χρόνια αποτελούν εξαίρεση, αφού τις πιο πολλές φορές τα exit-polls είναι πετυχημένα.

Αξιοσημείωτο είναι το αποτέλεσμα του exit-poll στις προεδρικές εκλογές της Ουκρανίας το 2004 όπου αποκαλύφθηκε ότι υπήρχε νοθεία.

Τα σφάλματα στα exit-polls μπορεί να προέλθουν από:

1. Λάθη κατά τη δειγματοληψία. Μην ξεχνάμε ότι οι πληθυσμοί στις εκλογικές αναμετρήσεις είναι πολύ μεγάλοι, ενώ το δείγμα πρακτικά δύσκολα ξεπερνά τους 5.000 συμμετέχοντες. Το ιδανικό είναι να υπάρχει εκπροσώπηση από όλα τα εκλογικά κέντρα, αλλά πολλές φορές αυτό είναι αδύνατο.
2. Μερικοί άνθρωποι μπορεί να έχουν πραγματικά συμμετάσχει στη διαδικασία για το τί ψηφισαν, για ένα συγκεκριμένο υποψήφιο ή κόμμα, αλλά μπορεί εκούσια να δήλωσαν κάτι άλλο.
3. Συνήθως η ανταπόκριση στα exit-polls εξαρτάται από το κόμμα που ψηφίστηκε. Οι υποστηρικτές κάποιων κομμάτων τείνουν να συμμετέχουν στη διαδικασία, ενώ κάποιων άλλων όχι.

2.4.4.2 Δημοσκοπήσεις μέσω SMS

Στην περίπτωση που οι ερωτήσεις είναι περιορισμένες (π.χ. η πρόθεση ψήφου) χρησιμοποιείται όλο και πιο συχνά η διαδικασία των sms. Με αποστολή και των ερωτήσεων και των απαντήσεων με sms.

Τα πλεονεκτήματα είναι αρκετά καθώς σχεδόν όλοι είναι πλέον χρήστες κινητών τηλεφώνων, ενώ η μεγάλη πλειοψηφία έχει εξοικείωση με τα sms. Τα κυριότερα πλεονεκτήματα είναι:

- Ταχύτατη απάντηση και γρήγορη πρόσβαση στα αποτελέσματα, που συγκεντρώνονται πολύ εύκολα. Ο αναμενόμενος χρόνος απόκρισης δεν ξεπερνά τα δέκα λεπτά.
- Επιπλέον πληροφορίες. Μέσω των νέων τεχνολογιών, είναι πιο εύκολη η συλλογή περισσότερων πληροφοριών, όπως είναι η γεωγραφική θέση και η δραστηριότητα του κινητού, πάντα με τη συγκατάθεση του χρήστη, αλλιώς υφίσταται θέμα δεοντολογίας.

Η παρουσία των sms είναι συνηθισμένη σε τηλεοπτικές εκπομπές, όπου ο θεατής καλείται να απαντήσει σε κάποιο ερώτημα στέλνοντας sms. Αυτού του είδους οι «δημοσκοπήσεις» στερούνται εντελώς εγκυρότητας.

2.4.5 Αξιοπιστία εταιρειών δημοσκοπήσεων

Λόγω του ρόλου των δημοσκοπήσεων στη διαμόρφωση της κοινής γνώμης και της επιρροής στο πολιτικό σκηνικό οι εταιρείες πρέπει να υπόκεινται σε ελέγχους αξιοπιστίας και εγκυρότητας των αποτελεσμάτων. Στην Ελλάδα οι εταιρείες έχουν συγκροτήσει σύλλογο τον ΣΕΔΕΑ (Σύλλογος Εταιρειών Δημοσκοπήσεων και Ερευνών Αγοράς) που αριθμεί αυτή τη στιγμή 23 εταιρείες. Ανάλογος διεθνής οργανισμός είναι η ESOMAR (European company of SOcial and MArket Research) που όμως δεν περιλαμβάνει εταιρείες, αλλά επιστήμονες που δραστηριοποιούνται στον χώρο των δημοσκοπήσεων και των ερευνών αγοράς. Οι εταιρείες μέλη του ΣΕΔΕΑ οφείλουν να τηρούν τους αντίστοιχους κώδικες δεοντολογίας και επαγγελματικής πρακτικής που η ESOMAR εκδίδει. Ανάλογος οργανισμός της ESOMAR είναι ο WAPOR (World Association for Public Opinion Research) Παγκόσμιος Οργανισμός Κοινής Γνώμης, του οποίου μέλη είναι επιστήμονες που δραστηριοποιούνται ερευνητικά και επαγγελματικά σε όλο τον κόσμο.

Βιβλιογραφικές Αναφορές Κεφαλαίου

Ελληνόγλωσσες

- Αφουξενίδης, Α. - Αλεξάκης, Μ. επιμ. (2006). *Πολιτική Κοινωνιολογία: Έξι κείμενα αφιερωμένα στην Ιωάννα Λαμπίρη- Δημάκη*. Αθήνα: Παπαζήσης.
- Νταβή, Ι. (2013). *Οι επιπτώσεις της οικονομικής κρίσης στην ισορροπία εργασιακής και προσωπικής ζωής των εργαζομένων στις ελληνικές τράπεζες*. Μεταπτυχιακή εργασία, Διοίκησης Επιχειρήσεων ΑΕΙ Πειραιά ΤΤ.
- Τριανταφύλλου, Ε. (2013). *Η επίδραση της οικονομικής ύφεσης στα συστήματα αμοιβών και ανταμοιβών σε μεσαίες και μεγάλες επιχειρήσεις, στην περιοχή Αττικής*. Μεταπτυχιακή εργασία, Διοίκησης Επιχειρήσεων ΑΕΙ Πειραιά ΤΤ.
- Τσιμηδέλη, Γ. (2012). *Σχέση υφιστάμενου-προϊστάμενου και η επίδραση στην εργασιακή ικανοποίηση και απόδοση του εργαζομένου που αναπτύσσεται στον Ιδιωτικό και Δημόσιο Μαιευτικό Κλάδο*. Μεταπτυχιακή εργασία, Διοίκησης Επιχειρήσεων ΑΕΙ Πειραιά ΤΤ.
- Φλουρής, Ι. (2015). *Δημοσκοπήσεις: Μεθοδολογία και εφαρμογές τους σε εκλογικό αποτέλεσμα*. Πτυχιακή εργασία, Διοίκησης Επιχειρήσεων ΑΕΙ Πειραιά ΤΤ.
- Babbie, E. (2011). *Εισαγωγή στην κοινωνική έρευνα*. Αθήνα: Εκδόσεις Κριτική Α.Ε.
- Grawitz, M. (2006). *Μέθοδοι των Κοινωνικών Επιστημών, Τόμοι Α' και Β'*, Αθήνα: Βιβλιόπολις ΑΕΒΕ. Προϊόντα Πνευματικής Δημιουργίας.

Ξενόγλωσσες

- Babbie, E. (2007). *The practice of social research* Thomson Wadsworth. Belmont, CA.
- Grémy, J. P. (1993). Questions et réponses: quelques résultats sur les effets de la formulation des questions dans les sondages. *Sociétés contemporaines*, 16(1), 165-176.
- Grunberg, G., Mayer, N., & Sniderman, P. M. (2002). *La démocratie à l'épreuve: une nouvelle approche de l'opinion des Français*. Les Presses de Sciences Po.
- Livne, A., Simmons, M. P., Adar, E. & Adamic, L. A. (2011). The Party Is Over Here: Structure and Content in the 2010 Election. *ICWSM*, 11, pp. 17-21.
- Metaxas, P. T., Mustafaraj, E. & Gayo-Avello, D. (2011). How (not) to predict elections. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 2011 IEEE Third International Conference on (pp. 165-171). IEEE.
- Michelat, G. & Thomas, J. P. H. (1966). *Dimensions du nationalisme: enquête par questionnaire (1962)* (Vol. 143). Presses de la Fondation nationale des sciences politiques.
- Michelat, G. & Simon, M. (1977). *Classe, religion et comportement politique*. Presses de Sciences Po.
- Morgan, D. L. (1993). *Successful focus groups: Advancing the state of the art* (Vol. 156). Sage Publications.
- Mustafaraj, E. & Metaxas, P. T. (2010). From obscurity to prominence in minutes: Political speech and real-time search.
- Mustafaraj, E., Finn, S., Whitlock, C. & Metaxas, P. T. (2011). Vocal minority versus silent majority: Discovering the opinions of the long tail. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 2011 IEEE Third International Conference on (pp. 103-110). IEEE.
- O'Connor, B., Balasubramanyan, R., Routledge, B. R. & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *ICWSM*, 11(pp. 122-129), 1-2.
- Rivière, E. & Martelli-Banégas, D. (2002). Chapitre 10. Pour un usage apaisé des sondages. *Académique*, pp. 293-309.

Visser, P. S., Krosnick, J. A., Marquette, J. & Curtin, M. (1996). Mail surveys for election forecasting? An evaluation of the Columbus Dispatch poll. *Public Opinion Quarterly*, 60(2), pp. 181-227.

Κεφάλαιο 3

Εισαγωγική Διαχείριση του IBM SPSS

Σύνοψη

Στο κεφάλαιο αυτό γίνεται αρχικά μία αναφορά στο περιβάλλον του SPSS και συγκεκριμένα στις επιλογές *Run the Tutorial*, *Type in data*, *Run an existing query*, *Create new query using Database Wizard* και αναλυτικά στα φύλλα *Data View* και *Variable View*. Γίνεται αναφορά στο Μενού *File* και συγκεκριμένα στις επιλογές *Data*, *Syntax*, *Output* και *Script*. Περιγράφεται η καταχώρηση δεδομένων στο IBM SPSS (χειρόγραφη καταγραφή δεδομένων και μεταφορά Δεδομένων από Αρχείο Excel) και η κωδικοποίηση δεδομένων ερωτηματολογίων. Από το μενού *Edit View Data* αναλύονται οι εντολές *Sort Variables*, *Transpose*, *Select Cases*, *Merge Files* (ενοποίηση δεδομένων από διαφορετικά αρχεία), *Split File* (διαχωρισμός των παρατηρήσεων του αρχείου σε μικρότερες εικονικές διαμερίσεις), *Weight Cases* (στάθμιση σύμφωνα με τις τιμές μιας άλλης μεταβλητής), *Restructure* (αναδιοργάνωση του αρχείου μετατρέποντας τις μεταβλητές σε παρατηρήσεις και αντίστροφα), *Recode* (επανακωδικοποίηση εφαρμόζεται για τη δημιουργία νέας μεταβλητής), *Compute* (δημιουργία νέων μεταβλητών υπό την εφαρμογή μιας μαθηματικής (αριθμητικής) έκφρασης), *Replace Missing Values* (αντικατάσταση των ελλειπουσών τιμών).

Προαπαιτούμενη γνώση

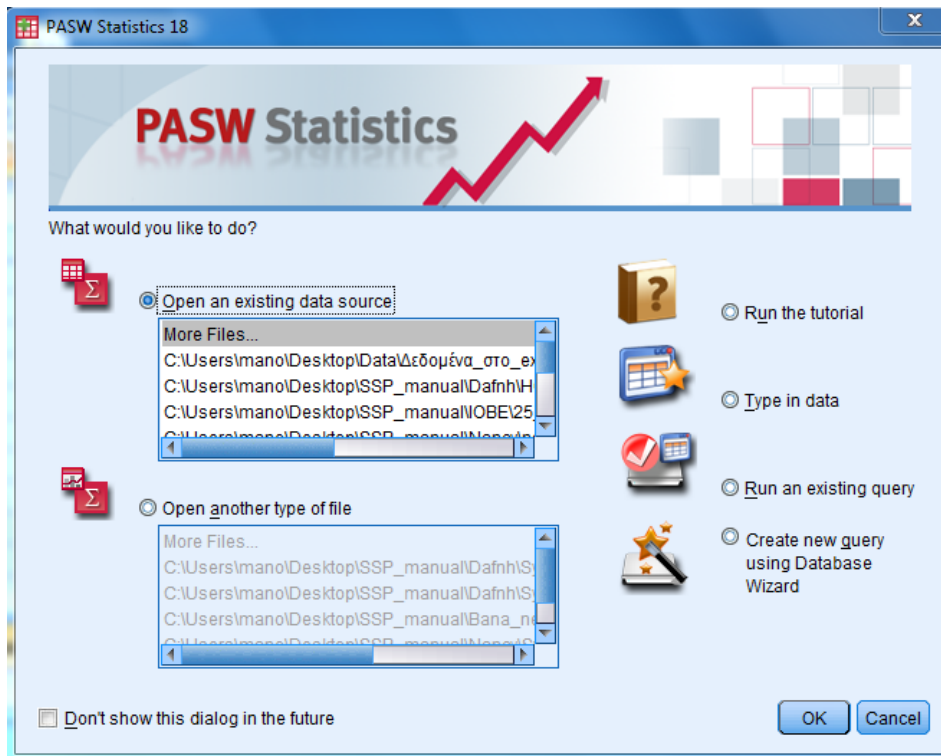
Προαπαιτούμενη γνώση δεν υπάρχει. Καλύτερη κατανόηση του κεφαλαίου είναι δυνατή σε αναγνώστες που έχουν εξοικείωση με χρήση υπολογιστών. Τα εγχειρίδια που παρουσιάζουν με πληρότητα την ύλη του συγκεκριμένου κεφαλαίου καθώς και των υπόλοιπων που αφορούν στη χρήση SPSS αναφέρονται ενδεικτικά στα βιβλία των D. Howitt, και D. Cramer, Carver και Nash, Coakes και Steed, Field, Huber και Norousis.

3.1 Το Περιβάλλον του SPSS

Ξεκινώντας την εφαρμογή εμφανίζεται το παράθυρο διαλόγου βλ. Εικόνα 3.1. Εάν επιθυμείτε να ανοίξετε ένα αρχείο δεδομένων «.sav» που υπάρχει ήδη διαθέσιμο τότε αφήνετε την αρχική επιλογή *Open an existing data source* και πατάμε το πλήκτρο αποδοχής *OK*.

Εάν επιθυμείτε να ανοίξετε ένα αρχείο δεδομένων διαφορετικού τύπου που υπάρχει ήδη διαθέσιμο τότε επιλέγετε *Open another type of file* και πατάμε το πλήκτρο αποδοχής *OK*.

Εάν επιθυμείτε να δημιουργήσετε ένα νέο αρχείο δεδομένων, δηλαδή θέλετε ένα νέο λογιστικό φύλλο εργασίας, τότε ακυρώνετε το παράθυρο διαλόγου από το πλήκτρο ακύρωσης *Cancel*.



Εικόνα 3.1: Εισαγωγική Ενότητα του SPSS

Η ενότητα διαθέτει ακόμα τις επιλογές:

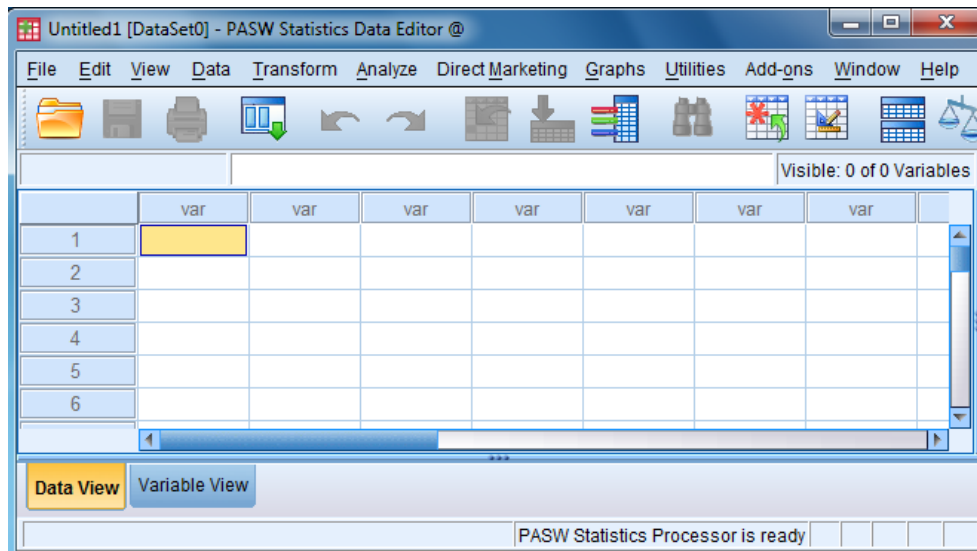
- *Run the Tutorial* - Για να παρακολουθήσετε το πρόγραμμα εκμάθησης.
- *Type in data* - Για να εισάγετε δεδομένα πληκτρολογώντας.
- *Run an existing query* - Για να εκτελέσετε μια εντολή.
- *Create new query using Database Wizard* - Για να δημιουργήσετε μια εντολή χρησιμοποιώντας το πρόγραμμα Database Wizard.

Επιλέγοντας *Cancel* εμφανίζεται η Εικόνα 3.2 δηλαδή ένας editor παρόμοιος με λογιστικού φύλλου Excel προκειμένου να πραγματοποιηθεί νέα καταγραφή δεδομένων.

Η ενότητα αυτή περιλαμβάνει τη γραμμή (menu) επιλογών, τη γραμμή συντομευμένων ενεργειών (εικονίδια) και δύο υποενότητες τη Data View και τη Variable View που θα αναλυθούν στη συνέχεια.

Data View

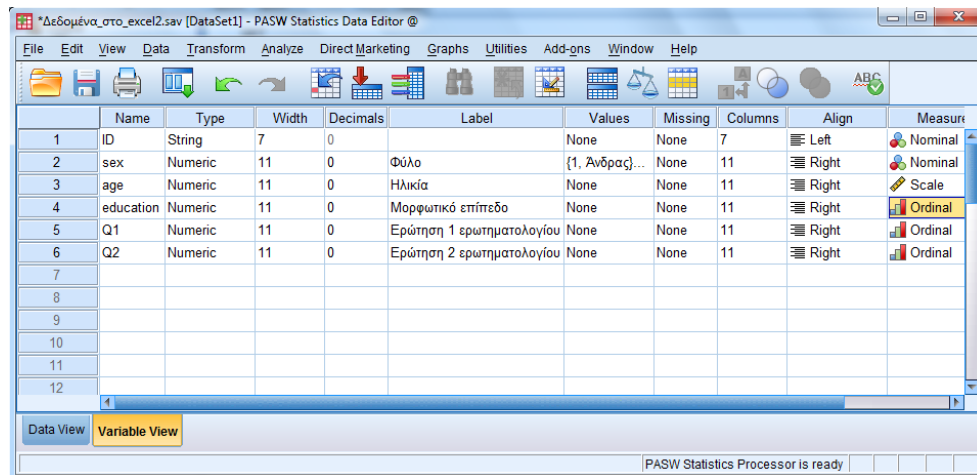
Η συγκεκριμένη υποενότητα αποτελείται από το λογιστικό φύλλο των δεδομένων, στο οποίο κατά την χειρόγραφη καταγραφή δεδομένων ο χρήστης εισάγει (πληκτρολογεί) απευθείας σε αυτό τις τιμές των παρατηρήσεων (περισσότερα βλ. στην παράγραφο «Καταχώρηση δεδομένων στο IBM SPSS»).



Εικόνα 3.1: Η ενότητα διαχείρισης του SPSS

Variable View

Η συγκεκριμένη υποενότητα περιλαμβάνει το σύνολο των μεταβλητών που περιέχονται στο *Data View*, και τα χαρακτηριστικά αυτών. Αναλυτικότερα, εάν η καταγραφή των δεδομένων γίνεται χειρόγραφα, τότε η εφαρμογή δίνει μια αρχική ονομασία στις μεταβλητές (π.χ. VAR00001), όπως φαίνεται και από την Εικόνα 3.4. Καθώς οι ονομασίες αυτές δεν είναι βοηθητικές για τον αναλυτή, παρέχεται η δυνατότητα μετονομασίας στην υποενότητα *Variable View*.



Εικόνα.3.3: Ενότητα Variable View

Τα χαρακτηριστικά που καταγράφονται για κάθε μεταβλητή είναι τα εξής:

Στήλη	Επεξήγηση
Name	Σ' αυτό το κελί καταγράφουμε μια σύντομη ονομασία της μεταβλητής, αρχίζοντας με γραμματικό χαρακτήρα. Αν και η εφαρμογή υποστηρίζει ελληνικούς και λατινικούς χαρακτήρες, προτείνεται η χρήση λατινικών χαρακτήρων για να υπάρχει μελλοντική συμβατότητα του αρχείου δεδομένων με παλαιότερες εκδόσεις. Το πεδίο είναι αλφαριθμητικό. Η αναλυτική ονομασία της μεταβλητής καταγράφεται στη στήλη Label.
Type	Σ' αυτή τη στήλη καθορίζεται ο τύπος δεδομένων της μεταβλητής. Η αρχική επιλογή είναι «Numeric», ενώ όταν η μεταβλητή δέχεται αλφαριθμητικά δεδομένα τότε καθορίζεται ως «String». Αν η μεταβλητή αφορά ημερομηνία τότε επιλέγεται το «Date».
Width	Σ' αυτή τη στήλη καθορίζεται ο μέγιστος αριθμός χαρακτήρων που θα εισαχθούν για την τιμή της μεταβλητής.
Decimals	Σ' αυτή τη στήλη καθορίζεται το πλήθος των δεκαδικών ψηφίων των τιμών της μεταβλητής.
Label	Σ' αυτή τη στήλη καταγράφεται η ονομασία (πλήρης) της μεταβλητής.
Values	Σ' αυτή τη στήλη καθορίζονται οι τιμές της μεταβλητής, δηλ. κωδικές ονομασίες και ερμηνεία αυτών. Για παράδειγμα, για τη μεταβλητή φύλο (sex) εάν αυτή έχει κωδικοποιηθεί ως «1-Ανδρας» και «2- Γυναίκα», τότε οι τιμές 1 & 2 καταγράφονται στην θέση Value ενώ οι περιγραφές «Ανδρας» και «Γυναίκα» στο πεδίο Value Label του κάθε κωδικού.
Missing	Σ' αυτή τη στήλη καθορίζονται οι τιμές οι οποίες θα ληφθούν από την εφαρμογή ως ελλείπουσες παρατηρήσεις. Συνήθως οι καταχωρητές των δεδομένων στις περιπτώσεις απουσίας απαντήσεων καταγράφουν τιμές «99» ή «999» ή κάποια άλλη τιμή. Επίσης εάν το κελί παραμείνει κενό (χωρίς τιμή), τότε και αυτή η περίπτωση θεωρείται από την εφαρμογή ως ελλείπουσα παρατήρηση. Ο διαχωρισμός είναι ότι όταν ο αναλυτής εξάγει μια ανάλυση, π.χ. έναν περιγραφικό πίνακα συχνότητας, τότε οι πραγματικές περιπτώσεις ελλειπουσών παρατηρήσεων (κενά κελιά) ξεχωρίζουν από την ένδειξη <i>Missing System</i> .
Columns	Σ' αυτή τη στήλη καθορίζεται το πλάτος της στήλης της μεταβλητής που θα εμφανίζεται στο περιβάλλον <i>Data View</i> .
Align	Σ' αυτή τη στήλη καθορίζεται ο τρόπος στοίχισης των δεδομένων της μεταβλητής.
Measure	Σ' αυτή τη στήλη καθορίζεται η κλίμακα μέτρησης της μεταβλητής. Οι επιλογές που προσφέρονται είναι: Scale - εάν η μεταβλητή είναι συνεχής, Nominal - εάν η μεταβλητή είναι ποιοτική μη διατάξιμη, Ordinal - εάν η μεταβλητή είναι ποιοτική διατάξιμη.

Πίνακας 3.1: Τα χαρακτηριστικά που καταγράφονται για κάθε μεταβλητή

Οποιοδήποτε από τα παραπάνω στοιχεία είναι δυνατό να τροποποιηθούν από τον αναλυτή. Με δεξί κλικ στην στήλη αρίθμησης των μεταβλητών παρέχονται οι δυνατότητες Αντιγραφής, *Copy*, Επικόλλησης, *Paste*, Διαγραφής, *Clear* και Εισόδου Νέας Μεταβλητής, *Insert New Variable*. Ας δούμε όμως τώρα πιο διεξοδικά ορισμένες κεντρικές ομάδες (μενού), με τις σημαντικότερες δυνατότητές τους, που διαθέτει το περιβάλλον του SPSS.

3.2 Μενού File

Περιέχει εντολές για την διαχείριση του αρχείου, όπως δημιουργία νέου αρχείου, *New*, άνοιγμα υπάρχοντος αρχείου, *Open* και *Open Database*, αποθήκευση, *Save* και *Save As*, εκτύπωση, *Print*, εμφάνιση πληροφοριών για τους τύπους και τα ονόματα των μεταβλητών *Display Data File Information* κ.ά.

Από τις επιλογές αυτές αξίζει να επικεντρωθούμε στις επιλογές της δημιουργίας νέου αρχείου, *New*, μέσω του οποίου καθορίζεται το είδος του παραθύρου που θα ανοίξει, όπου:

- *Data*: Για τη δημιουργία ενός νέου λογιστικού φύλλου καταγραφής δεδομένων.
- *Syntax*: Για τη δημιουργία ενός νέου περιβάλλοντος καταγραφής εντολών. Ο τρόπος γραφής των εντολών προϋποθέτει βασικές προγραμματιστικές γνώσεις και δεδομένου ότι η παρούσα εργασία απευθύνεται σε πιο αρχάριους αναλυτές, δεν θα γίνει αναφορά σε αυτές. Να αναφέρουμε όμως ότι εάν κάποιος αναλυτής επιθυμεί να αρχίσει να μαθαίνει τον τρόπο που δέχεται το SPSS τις εντολές, μπορεί αντί να εκτελεί κάθε πλαίσιο διαλόγου από το πλήκτρο *OK*, να χρησιμοποιεί το πλήκτρο *PASTE* το οποίο ανοίγει αυτόματα το περιβάλλον *Syntax*, έχοντας ήδη καταγράψει τον αντίστοιχο κώδικα.
- *Output*: Για την εμφάνιση του περιβάλλοντος των αποτελεσμάτων της κάθε ενέργειάς μας. Οποιαδήποτε ενέργεια εκτελεί ένας αναλυτής στο αρχείο δεδομένων του (από τη μετατροπή μιας μεταβλητής, μέχρι την οποιαδήποτε ανάλυση), καταγράφονται στο συγκεκριμένο περιβάλλον. Επίσης, το αρχείο αυτό αποθηκεύεται και μπορεί να ανοιχτεί από τον αναλυτή χωρίς να είναι απαραίτητη η ύπαρξη αρχείου δεδομένων γι' αυτό.
- *Script*: Από το συγκεκριμένο περιβάλλον ο αναλυτής έχει τη δυνατότητα να συντάξει πλήρη προγράμματα σε γλώσσες προγραμματισμού, όπως η Basic ή η Python, και να έχει πρόσβαση στις λειτουργίες του SPSS.

Στην επόμενη παράγραφο όμως θα ασχοληθούμε πιο συγκεκριμένα με την καταχώρηση δεδομένων στην εφαρμογή που είναι το προαπαιτούμενο κομμάτι για να ξεκινήσουμε την οποιαδήποτε ανάλυση.

3.2.1 Καταχώριση δεδομένων στο IBM SPSS

Για να προβούμε σε οποιαδήποτε στατιστική επεξεργασία στο SPSS, απαιτείται αρχικά η καταγραφή των δεδομένων στο πρόγραμμα.

Η εισαγωγή των δεδομένων δύναται να πραγματοποιηθεί:

1. είτε με χειρόγραφη καταγραφή αυτών,
2. είτε με τη μεταφορά αυτών από κάποιο άλλο λογισμικό. Το πιο ευρέως διαδεδομένο λογισμικό στο οποίο πραγματοποιείται η αρχική καταγραφή των δεδομένων είναι το Excel.

Χειρόγραφη Καταγραφή Δεδομένων

Η χειρόγραφη καταγραφή των δεδομένων πραγματοποιείται γράφοντας τα δεδομένα απευθείας πάνω στο λογιστικό φύλλο του SPSS. Τονίζουμε ότι μια ορθή καταγραφή είναι κάθε στήλη να αντιστοιχεί σε μια μεταβλητή και κάθε γραμμή να αντιστοιχεί στα στοιχεία του κάθε ατόμου / αντικειμένου. Σημειώνουμε ότι σε περίπτωση που κάποιος καταχωρητής τα έχει καταγράψει ανάποδα, δηλ. τις μεταβλητές στις γραμμές και τα χαρακτηριστικά στις στήλες, τότε ο αναλυτής θα πρέπει μέσω της επιλογής *Data > Transpose*, να αντιστρέψει τον πίνακα των δεδομένων του αρχείου.

	VAR00001	VAR00002	VAR00003	VAR00004	var
1	ABC0001	1,00	10,00	11,00	
2					
3					
4					

Εικόνα 3.2: Εισαγωγή Δεδομένων στο λογιστικό φύλλο του SPSS

Επίσης, για την ορθή καταγραφή των δεδομένων μας θα πρέπει αρχικά να πραγματοποιηθεί μια προεργασία κωδικοποίησής τους.

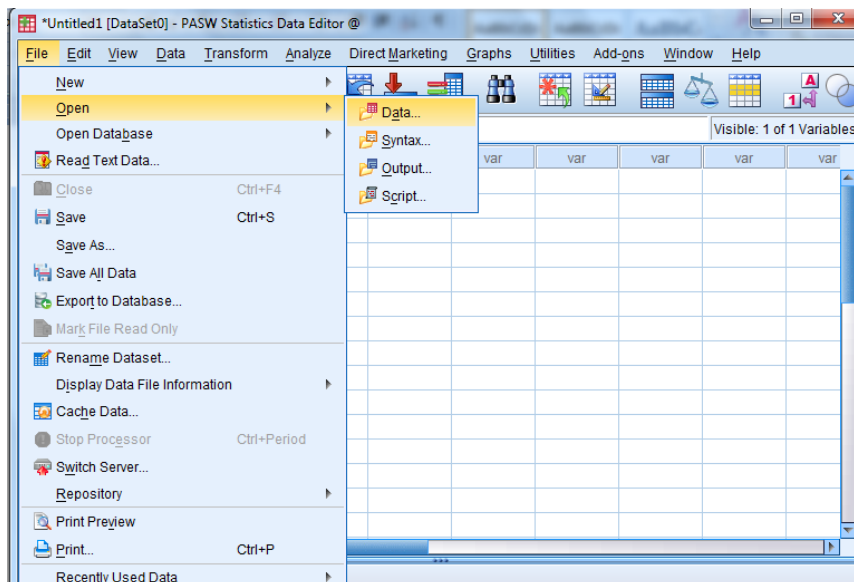
Για παράδειγμα, εάν στην κατοχή μας διαθέτουμε ένα πλήθος απαντημένων ερωτηματολογίων, για την επεξεργασία της στατιστικής ανάλυσης θα πρέπει οι απαντήσεις των ερωτώμενων να καταγραφούν στο SPSS. Έτσι, πριν την καταγραφή τους είναι σημαντικό να ορίσουμε εξαρχής μια κωδικοποίηση στα:

1. Ερωτηματολόγια.
2. Στις μεταβλητές.
3. Στις απαντήσεις των ερωτήσεων πολλαπλής επιλογής, ή σε οποιαδήποτε μεταβλητή περιέχεται στην ανάλυση.

Μεταφορά Δεδομένων από Αρχείο Excel

Για να εισάγουμε τα δεδομένα επιλέγουμε από την γραμμή εντολών:

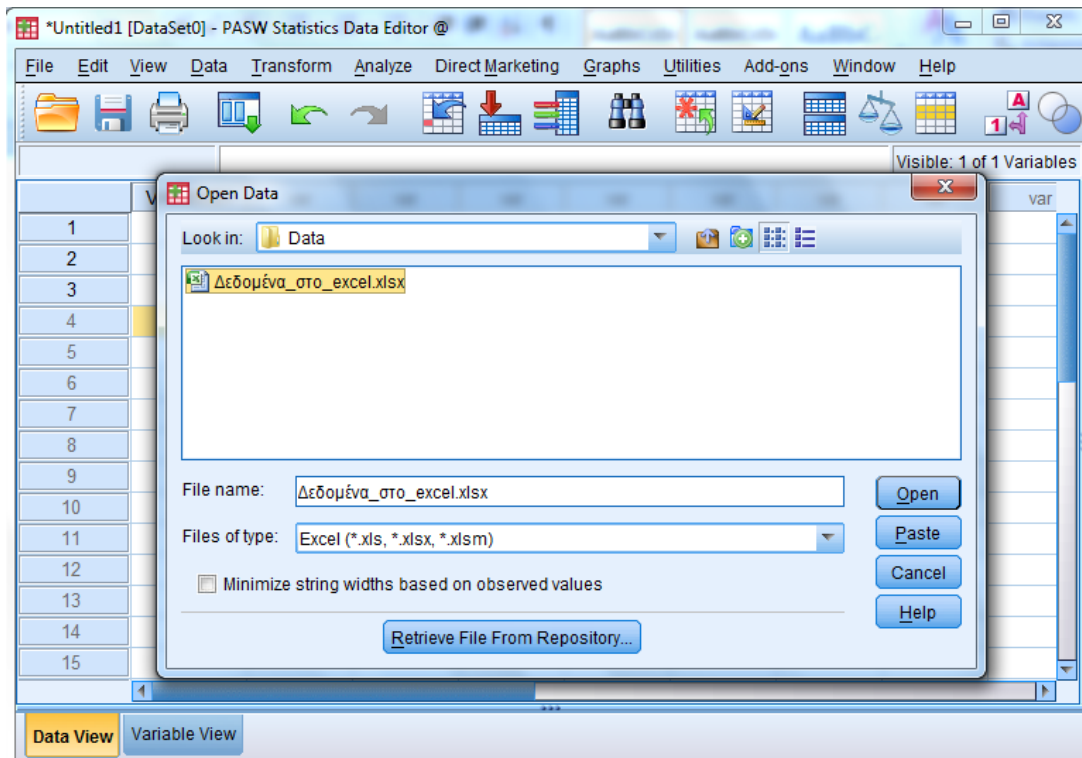
File > Open > Data



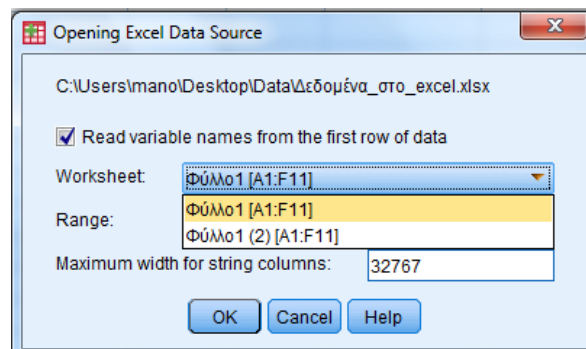
Εικόνα 3.3: Άνοιγμα αρχείου δεδομένων

Στη συνέχεια αναζητάμε τη θέση αποθήκευσης του αρχείου δεδομένων, φροντίζοντας την επιλογή του κατάλληλου τύπου αρχείου, π.χ. Excel για αρχεία με κατάληξη.xls, .xlsx.

Η διαδικασία συνεχίζεται με την επιλογή *Open*, όπου στη συνέχεια επιλέγουμε το φύλλο *Worksheet* από το οποίο θα αντλήσουμε τα δεδομένα μας, καθώς στο αρχείο *Excel* δύναται να υπάρχουν δεδομένα σε περισσότερα από ένα λογιστικά φύλλα.



Εικόνα 3.4: Αναζήτηση του αρχείου δεδομένων



Εικόνα 3.5: Επιλογή του λογιστικού φύλλου με τα δεδομένα

Με την ολοκλήρωση, *OK*, πραγματοποιείται η εισαγωγή των δεδομένων στο λογιστικό φύλλο του *SPSS*.

	country	pop	fert	lifeexp	lifeexpb	hlthgdp	students	terteduc	unemmw	unempop
1	1	10192	1,53	80,5	74,3	8,6	2137	357	9,5	17,7
2	2	5295	1,72	78,8	73,9	8,3	973	183	5,2	7,3
3	3	82057	1,36	80,6	74,5	10,3	14497	2098	9,3	9,1
4	4	10511	1,29	80,6	75,5	8,4	1904	374	10,9	29,6
5	5	39348	1,15	82,4	75,1	7,0	8087	1746	18,8	26,2
6	6	58728	1,75	82,4	74,8	9,4	12008	2027	11,8	20,0
7	7	3694	1,93	79,1	73,5	6,8	1000	143	7,5	6,5
8	8	57563	1,20	81,8	75,5	8,2	9203	1869	11,8	30,7
9	9	424	1,68	80,5	73,7	6,0	62	2	2,7	7,3
10	10	15654	1,63	80,6	75,2	8,7	3136	461	4,0	5,6
11	11	8075	1,34	80,9	74,7	8,0	1426	247	4,5	5,3
12	12	957	1,46	78,9	71,7	7,7	2076	352	5,2	8,9
13	13	5147	1,70	80,8	73,5	6,9	1101	250	11,4	21,3
14	14	8848	1,50	81,9	76,9	7,9	1962	281	8,3	11,3
15	15	50000	1,72	79,7	74,8	6,8	13238	1938	6,3	12,7

Εικόνα 3.6: Η υποενότητα Data View

File > Save As

Είτε με την ολοκλήρωση της καταγραφής των δεδομένων είτε σε ενδιάμεσο στάδιο, αποθηκεύουμε το αρχείο μας από την επιλογή *File > Save As*, ορίζοντας την τοποθεσία αποθήκευσης. Ενδιάμεσες καταχωρήσεις των τροποποιήσεων του αρχείου θα πρέπει να πραγματοποιούνται για την αποφυγή της απώλειάς τους, από την επιλογή *File > Save*.

3.2.2 Κωδικοποίηση Δεδομένων

Ερωτηματολόγια

Στην περίπτωση που τα δεδομένα μας προέρχονται από ερωτηματολόγια θα πρέπει να γίνεται μια κωδικοποίηση των ερωτηματολογίων, δηλώνοντας στο κάθε ερωτηματολόγιο έναν μοναδικό κωδικό, ώστε στο μέλλον να είμαστε σε θέση να γνωρίζουμε από ποιο ερωτηματολόγιο προέρχονται οι παρατηρήσεις των μεταβλητών μας. Η συγκεκριμένη κωδικοποίηση καταγράφεται σχεδόν πάντα στην πρώτη στήλη του λογιστικού φύλλου.

Μεταβλητές

Στις περιπτώσεις των ερωτηματολογίων, συναντάμε συνήθως πολυπληθείς ομάδες ερωτήσεων, για τις οποίες για την πιο γρήγορη και άμεση καταγραφή τους ορίζουμε μια σύντομη περιγραφή, όπως «Q001», «Q002» κ.ο.κ. Ωστόσο, η πλήρης περιγραφή καταγράφεται, όπως είδαμε, στο πεδίο *Label* της ενότητας *Variable View*.

Παρατηρήσεις

Όταν τα δεδομένα, οι παρατηρήσεις, οι τιμές των μεταβλητών μας, ή αλλιώς οι απαντήσεις των ερωτήσεων είναι περιγραφικές, όταν δηλαδή με άλλα λόγια διαχειριζόμαστε ποιοτικές μεταβλητές, τότε κωδικοποιούμε τις απαντήσεις μας, και στο λογιστικό φύλλο καταγράφουμε απευθείας τις κωδικοποιήσεις αυτών. Για παράδειγμα, για τη μεταβλητή «Φύλλο» μπορούμε να ορίσουμε «1 - Άνδρας» και «2 - Γυναίκα». Έτσι στην καταγραφή των δεδομένων θα γραφτούν απευθείας οι τιμές «1» και «2», αντίστοιχα. Η συγκεκριμένη προσέγγιση εκτός από το γεγονός ότι εξυπηρετεί την πιο γρήγορη καταγραφή των δεδομένων, ενισχύει τη βέλτιστη ποιότητα δεδομένων, καθώς εάν χρησιμοποιούσαμε String / Text τιμές, υπάρχει κίνδυνος

διαφορετικότητας μεταξύ ιδίων τιμών, αφού μια τιμή «άνδρας» (με μικρό άλφα) είναι διαφορετική από την τιμή «Άνδρας» (με κεφαλαίο άλφα).

Καλό θα είναι πριν οποιαδήποτε καταγραφή των δεδομένων, να υπάρχει μια αρχική συνεννόηση του αναλυτή με τους καταχωρητές, προκειμένου να αποφεύγονται μελλοντικά προβλήματα, δυσκολίες και τλαιπωρία του αναλυτή κατά τον «καθαρισμό» των δεδομένων.

3.3 Μενού Edit

Περιέχει εντολές διαδικασιών ανάκλησης, *Undo*, αντιγραφής, *Copy*, αποκοπής, *Cut*, επικόλλησης δεδομένων, *Paste*, καθώς και εισαγωγής μεταβλητών, *Insert Variable*. Οι εντολές αυτές δεν θα αναλυθούν περισσότερο καθώς η χρήση τους είναι γνωστή και κοινή στα περισσότερα προγράμματα, και θεωρούμε ότι ο χρήστης είναι εξοικειωμένος στη χρήση υπολογιστών γενικότερα.

3.4 Μενού View

Περιέχει εντολές ρύθμισης του πλαισίου *Data Editor*, δηλ. του πλαισίου εμφάνισης των δεδομένων των μεταβλητών που ανοίγει αυτόματα με την εκκίνηση του προγράμματος.

Έτσι μέσω των επιλογών παρέχονται οι δυνατότητες απόκρυψης ή διαμόρφωσης της γραμμής εργαλείων *Status Bar*, απόκρυψης ή εμφάνισης του *Data Editor*, διαχείρισης της γραμματοσειράς της οθόνης *Fonts*, καθώς και εμφάνισης των τιμών ή της περιγραφής των τιμών των μεταβλητών *Values Labels*.

Ομοίως, οι εντολές αυτές δεν θα αναλυθούν περισσότερο καθώς η χρήση τους είναι απλή και οι χρήστες θα μπορούσαν να πειραματιστούν σε αυτές. Επίσης αφορούν σε μορφολογικά τμήματα της εφαρμογής οπότε δεν κρίνεται αναγκαία η εκμάθησή τους από την παρούσα εργασία.

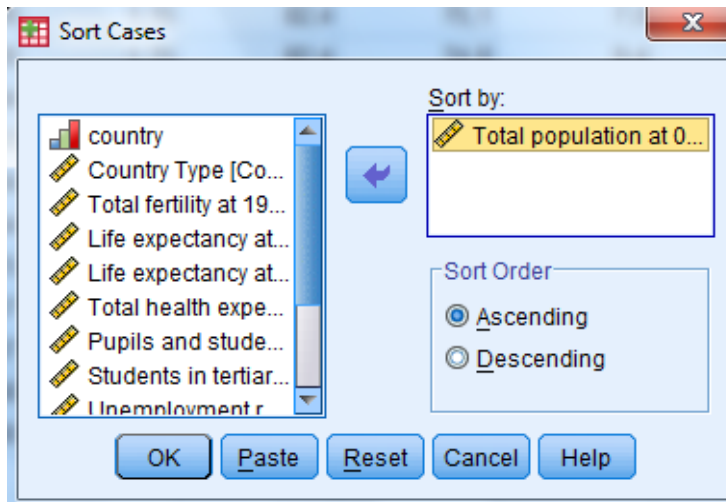
3.5 Μενού Data

Περιέχει εντολές διαχείρισης και τροποποίησης των δεδομένων, όπως επιλογή των δεδομένων που θα αναλυθούν *Select Cases*, ταξινόμηση δεδομένων *Sort Cases*, διαχωρισμού αρχείου *Split File* ή ενοποίησης αρχείων *Merge Files*. Στο συγκεκριμένο μενού επιλογών περιέχονται αρκετά σημαντικές δυνατότητες χειρισμού των μεταβλητών, εκ των οποίων αξίζει να παρουσιάσουμε τα εξής:

3.5.1 Sort Cases

Η δυνατότητα προσφέρεται από την επιλογή *Data > Sort Cases*, με σκοπό την ταξινόμηση των δεδομένων (γραμμών) με βάση μιας ή περισσότερων μεταβλητών.

Στο παράθυρο διαλόγου που προκύπτει ο χρήστης καλείται να προσθέσει τη μεταβλητή ή τις μεταβλητές βάσει των οποίων θα πραγματοποιηθεί η ταξινόμηση, και στη συνέχεια να ορίσει εάν η ταξινόμηση θα είναι αύξουσα *Ascending* ή φθίνουσα *Descending*. Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*.

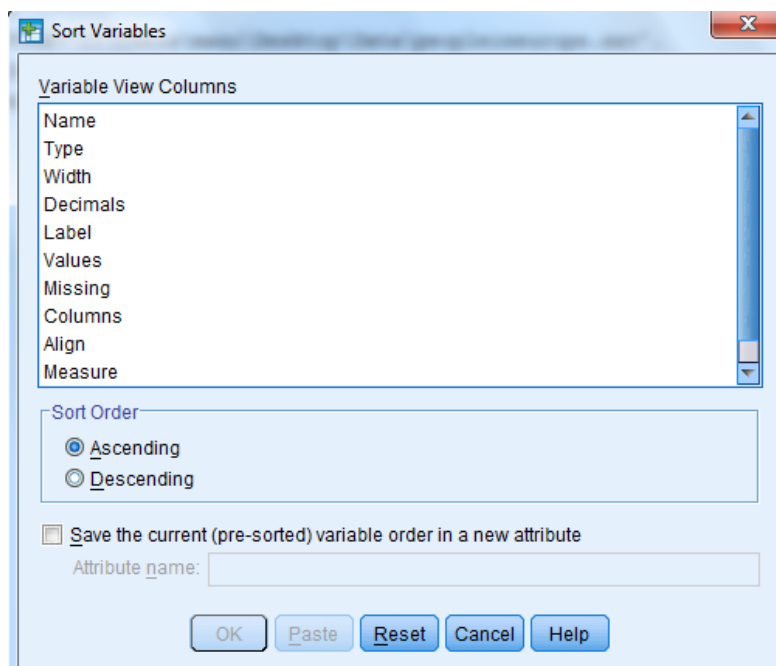


Εικόνα 3.7: Η υποενοότητα Sort Cases

3.5.2 Sort Variables

Η δυνατότητα προσφέρεται από την επιλογή *Data > Sort Variables*, με σκοπό την ταξινόμηση των δεδομένων με βάση κάποιο χαρακτηριστικό μιας ή περισσότερων μεταβλητών του αρχείου.

Ορισμένες επιλογές ταξινόμησης που προσφέρονται από το παράθυρο διαλόγου είναι η ονομασία της μεταβλητής, ο τύπος, ο αριθμός των ψηφίων της κ.ά.

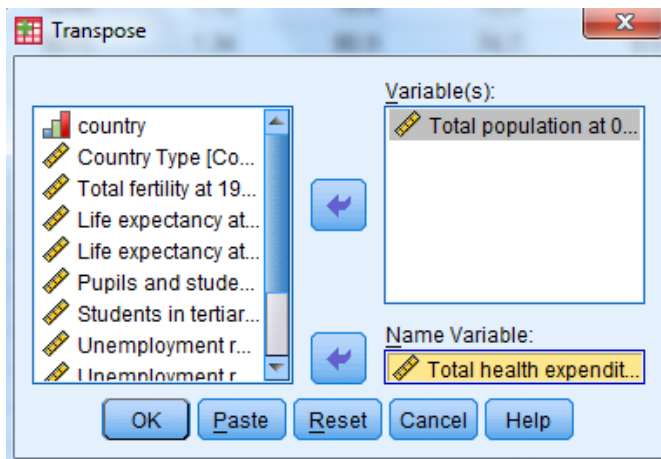


Εικόνα 3.8: Η υποενοότητα Sort Variables

Ομοίως ο αναλυτής έχει τη δυνατότητα να ορίσει εάν η ταξινόμηση θα είναι αύξουσα *Ascending* ή φθίνουσα *Descending*. Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*.

3.5.3 Transpose

Η δυνατότητα προσφέρεται από την επιλογή *Data > Transpose*, με σκοπό την αντιμετάθεση των γραμμών (χαρακτηριστικά ατόμων ή αντικειμένων) και των στηλών (μεταβλητών) του αρχείου. Έτσι δημιουργείται ένα νέο αρχείο το οποίο έχει τις στήλες ως γραμμές και τις γραμμές ως στήλες. Να αναφέρουμε ότι η συγκεκριμένη δυνατότητα εξυπηρετεί περιπτώσεις όπου οι καταχωρητές έχουν καταγράψει με «λανθασμένο» τρόπο τα δεδομένα, π.χ. σε κάποιο ενδιάμεσο τύπου αρχείο, όπως Excel. Και μιλάμε για λανθασμένο τρόπο καταχώρησης, διότι τα στατιστικά προγράμματα χειρίζονται τα δεδομένα έχοντας ως κοινό παρονομαστή ότι κάθε στήλη αντιστοιχεί σε κάποια μεταβλητή, ενώ κάθε γραμμή σε κάποια οντότητα (άτομο ή αντικείμενο).



Εικόνα 3.9: Η υποενότητα *Transpose*

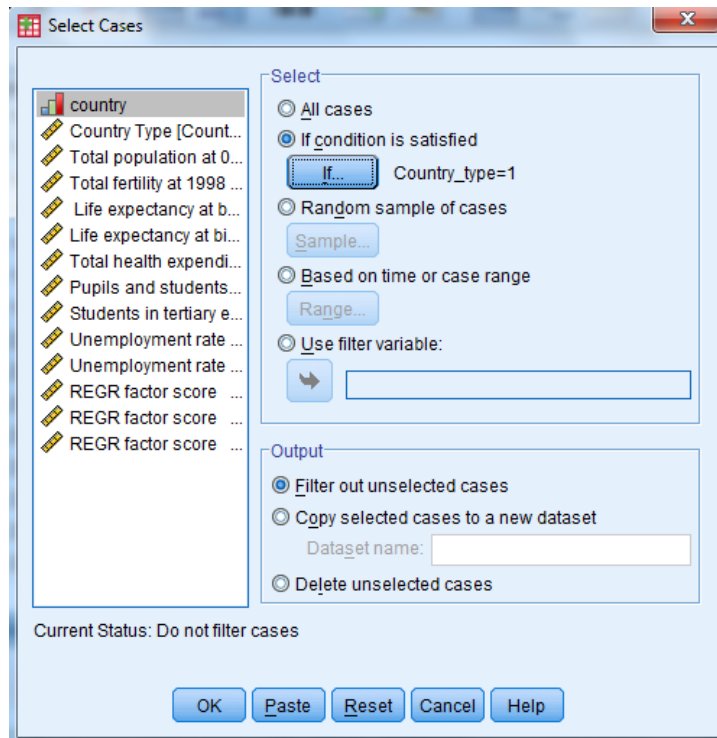
Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*.

3.5.4 Select Cases

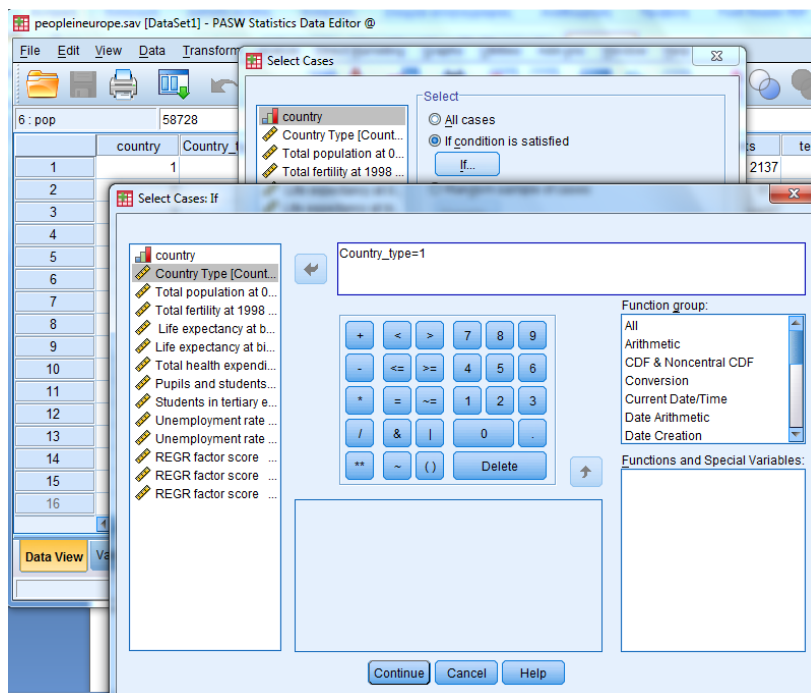
Υπάρχουν περιπτώσεις που επιθυμούμε να πραγματοποιήσουμε ανάλυση δεδομένων σε συγκεκριμένες παρατηρήσεις, οι οποίες ικανοποιούν ορισμένα κριτήρια. Σε αυτή την περίπτωση η προσθήκη φίλτρου στην επιλογή των παρατηρήσεων πραγματοποιείται από την ενότητα : *Data > Select Cases*.

Από αυτή την ενότητα παρέχονται οι επιλογές:

- *All Cases*: Για την επιλογή όλων των παρατηρήσεων. Αυτή χρησιμοποιείται κυρίως για την παύση ενός φίλτρου που έχει ήδη εφαρμοστεί, ώστε η μελλοντική ανάλυση να γίνει και πάλι στο σύνολο των δεδομένων του αρχείου.
- *If condition is satisfied*: Για να ορίσουμε τις συνθήκες επιλογής των παρατηρήσεων. Η υποενότητα που εμφανίζεται είναι σχεδόν ίδια με αυτή της εντολής *Transform > Compute* όπου στο πλαίσιο πάνω δεξιά της ενότητας καταγράφεται η «μαθηματική» έκφραση, ενώ επίσης από το πλαίσιο *Function Group* παρέχονται διάφορες μαθηματικές, σχεσιακές ή λογικές συναρτήσεις. Εφαρμόζοντας το φίλτρο, εάν η συνθήκη που έχει καταγραφεί ισχύει, τότε η παρατήρηση επιλέγεται, αλλιώς όχι.



• Εικόνα 3.10: Η υποενότητα Select Cases

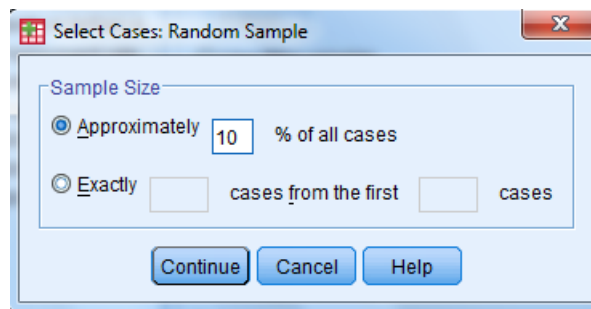


• Εικόνα 3.11: Menu Data > Select Cases > If condition is satisfied

- *Random samples of cases*: Για να γίνει μέσω τυχαίας δειγματοληψίας επιλογή τυχαίου υποσυνόλου παρατηρήσεων. Ο αναλυτής μπορεί να επιλέξει δείγμα κατά προσέγγιση (*Approximately*) μεγέθους ίσου με οριζόμενο ποσοστό, π.χ. 10% των συνολικών παρατηρήσεων, ή ακριβώς (*Exactly*) με το οριζόμενο πλήθος παρατηρήσεων.

Σε περίπτωση που επιλεγεί η δεύτερη μέθοδος, ο αναλυτής μπορεί να ορίσει και το πλήθος των παρατηρήσεων του αρχικού αρχείου από το οποίο θα επιλεγεί το δείγμα, συνεπώς σ' αυτή την περίπτωση το

οριζόμενο πλήθος θα πρέπει να είναι μικρότερο ή ίσο από το πλήθος των παρατηρήσεων του αρχικού αρχείου.

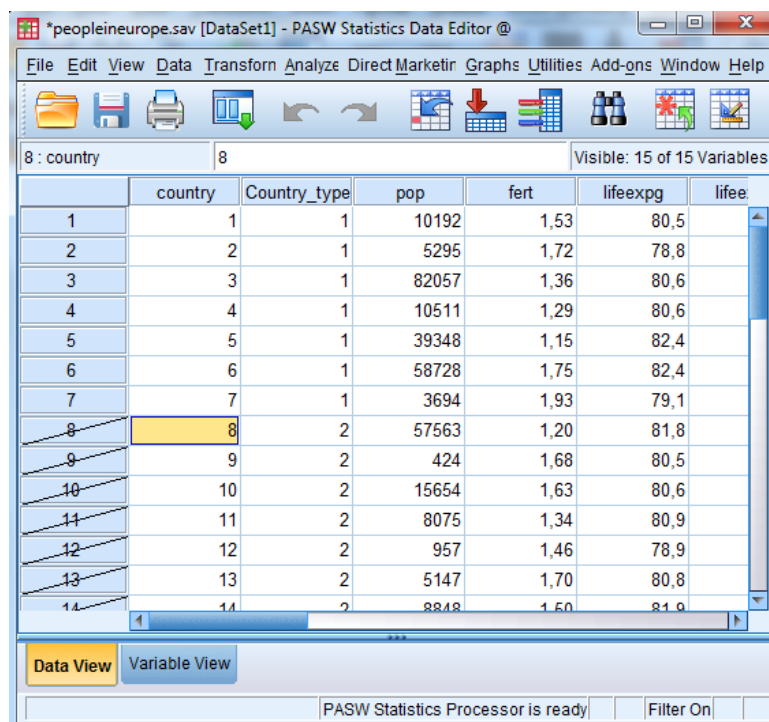


Εικόνα 3.12: Η υποενοότητα *Random samples of cases*

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*.

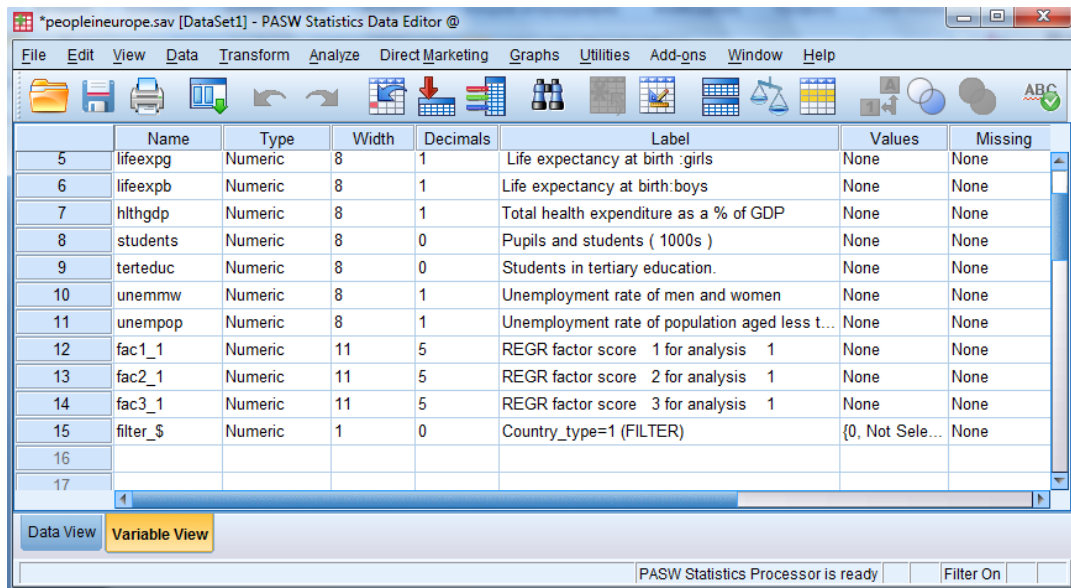
Με την ολοκλήρωση των ενεργειών μας επιβεβαιώνουμε ότι έχουμε ολοκληρώσει ορθά τη διαδικασία, ελέγχοντας τα αποτελέσματα των ενεργειών μας:

1. Στην υποενοότητα *Data View* εμφανίζεται η διαγραφή των μη επιλεγμένων παρατηρήσεων. Οι παρατηρήσεις που έχουν την ένδειξη της διαγραφής, δεν θα ληφθούν υπόψη στην οποιαδήποτε ανάλυση πραγματοποιήσει ο αναλυτής.



Εικόνα 3.13: Εμφάνιση των διαγραμμένων παρατηρήσεων

2. Στην υποενοότητα *Variable View* έχει δημιουργηθεί η μεταβλητή φίλτρου η οποία φέρει την ονομασία *filter_\$*, και στην περιγραφή της οποίας μπορούμε να δούμε λεπτομέρειες για το φίλτρο που εφαρμόστηκε.



Εικόνα 3.14: Δημιουργία της μεταβλητής φίλτρου

Η πάυση του φίλτρου πραγματοποιείται είτε από την επιλογή *Data > Select Cases > All Cases*, είτε απλά διαγράφοντας τη μεταβλητή *filter_\$* από την επιλογή *Clear* στο περιβάλλον *Variable View*.

3.5.5 Merge Files

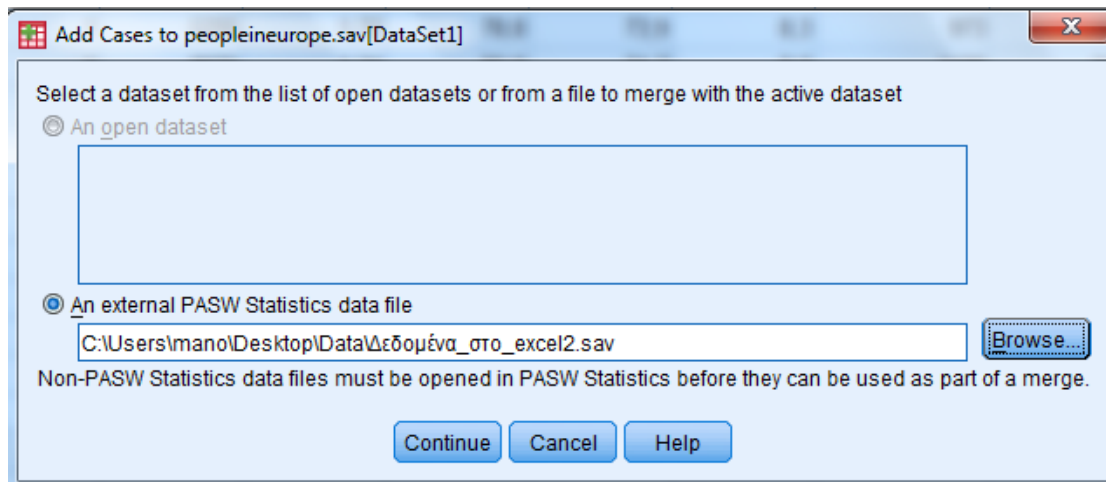
Η δυνατότητα προσφέρεται από την επιλογή *Data > Merge* με σκοπό την ενοποίηση δεδομένων από διαφορετικά αρχεία, και αυτό γιατί συμβαίνει συχνά η καταχώρηση των δεδομένων να γίνεται από περισσότερους του ενός καταχωρητές για εξοικονόμηση χρόνου. Έτσι στο τέλος προκύπτουν διαφορετικά αρχεία δεδομένων τα οποία θα πρέπει να ενοποιηθούν σε ένα.

Από το σημείο αυτό παρέχονται δύο επιλογές:

- *Merge > Add Cases* - Με σκοπό τη δημιουργία ενός νέου αρχείου δεδομένων, όπου ο αναλυτής θα έχει απομονώσει ορισμένες μεταβλητές από ένα μεγαλύτερο αρχείο, και την επιλογή.
- *Merge > Add Variables* - Με σκοπό την δημιουργία ενός νέου συγκεντρωτικού αρχείου δεδομένων, όπου ο αναλυτής θα συγκεντρώσει σε ένα αρχείο μεταβλητές από διαφορετικά αρχεία.

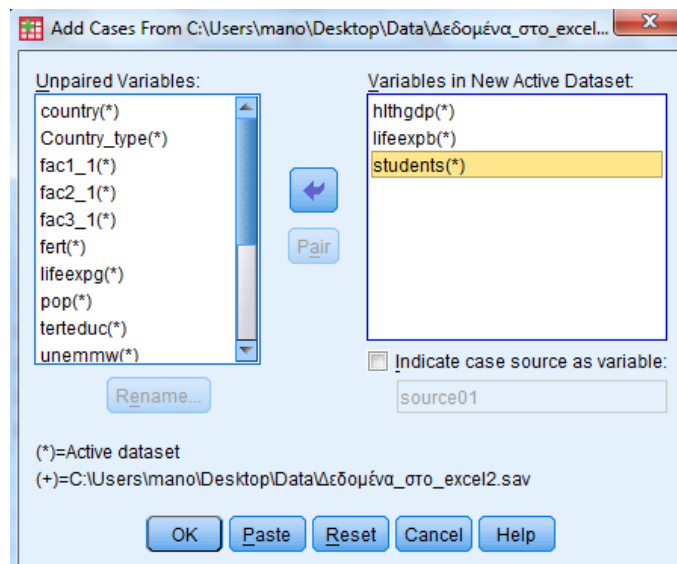
Merge > Add Cases

Στο πρώτο πλαίσιο διαλόγου ο αναλυτής ορίζει το αρχείο από το οποίο θα αντλήσει δεδομένα. Αυτό μπορεί να είναι είτε κάποιο αρχείο που είναι ήδη ανοιγμένο *An open dataset*, είτε να ορίσει το μονοπάτι στο οποίο βρίσκεται αποθηκευμένο το αρχείο *An external PASW Statistics data file*.



Εικόνα 3.15: Η υποενότητα Merge > Add Cases > Step 1

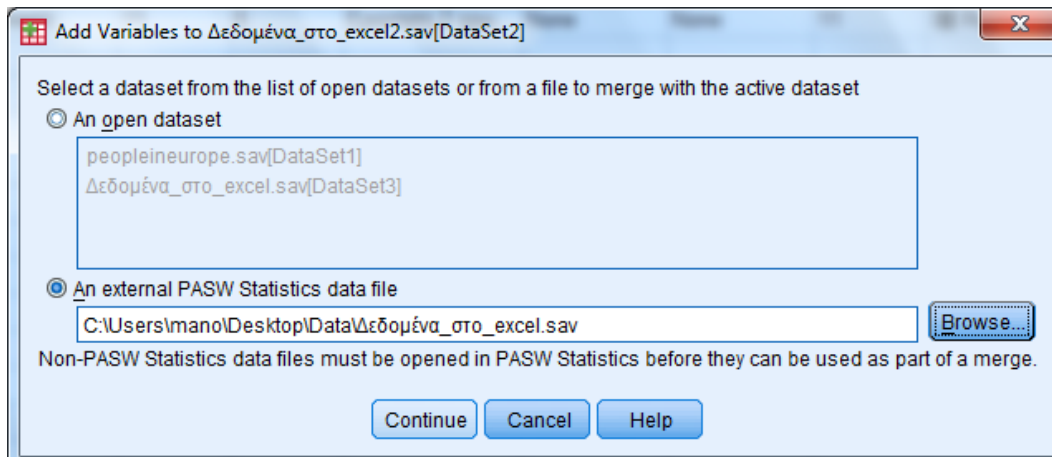
Στη συνέχεια στο δεύτερο πλαίσιο διαλόγου θα οριστούν στο σημείο *Variables in New Active Dataset* οι μεταβλητές που θα δημιουργήσουν το νέο αρχείο.



Εικόνα 3.16: Η υποενότητα Merge > Add Cases > Step 2

Merge > Add Variables

Ομοίως, στο πρώτο πλαίσιο διαλόγου ο αναλυτής ορίζει το αρχείο από το οποίο θα λάβει τις επιπρόσθετες μεταβλητές. Αυτό μπορεί να είναι είτε κάποιο αρχείο που είναι ήδη ανοιγμένο *An open dataset* είτε να ορίσει το μονοπάτι στο οποίο βρίσκεται αποθηκευμένο το αρχείο *An external PASW Statistics data file*.



Εικόνα 3.17: Η υποενότητα Merge > Add Variables > Step 1

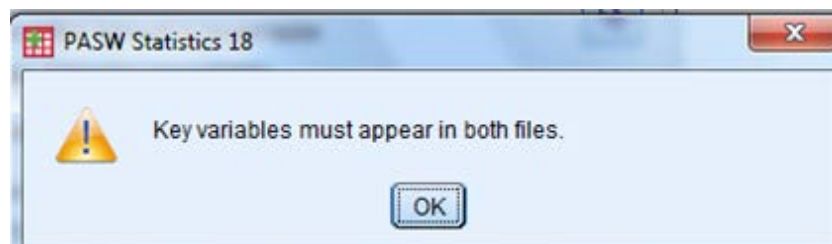
Στη συνέχεια στο δεύτερο πλαίσιο διαλόγου έχουμε:

- *Excluded Variables*: Σ' αυτό το παράθυρο δηλώνονται οι μεταβλητές που δεν θα προστεθούν στο αρχείο.
- *New Active Dataset*: Εμφανίζονται οι μεταβλητές που περιέχονται στο επιλεγμένο αρχείο.
- *Key Variables*: Το συγκεκριμένο πλαίσιο ενεργοποιείται όταν επιλεγεί *Match cases on key variables in sorted files* με σκοπό να δηλωθεί η μεταβλητή - κλειδί (ID) σύμφωνα με την οποία θα γίνει η αναγνώριση των δεδομένων σε κάθε οντότητα (γραμμή του αρχείου).

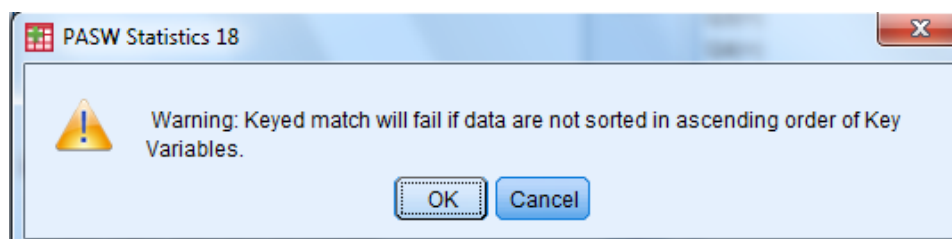
Για να επιτευχθεί μια σωστή αντιστοίχιση των παρατηρήσεων κάθε οντότητας (ατόμου ή αντικειμένου αναλόγως τί μελετάμε) θα πρέπει:

1. η μεταβλητή κλειδί να περιέχεται και στα δύο αρχεία, και
2. οι εγγραφές να έχουν ήδη ταξινομηθεί με βάση το κλειδί (βλ. εντολή *Sort Cases*).

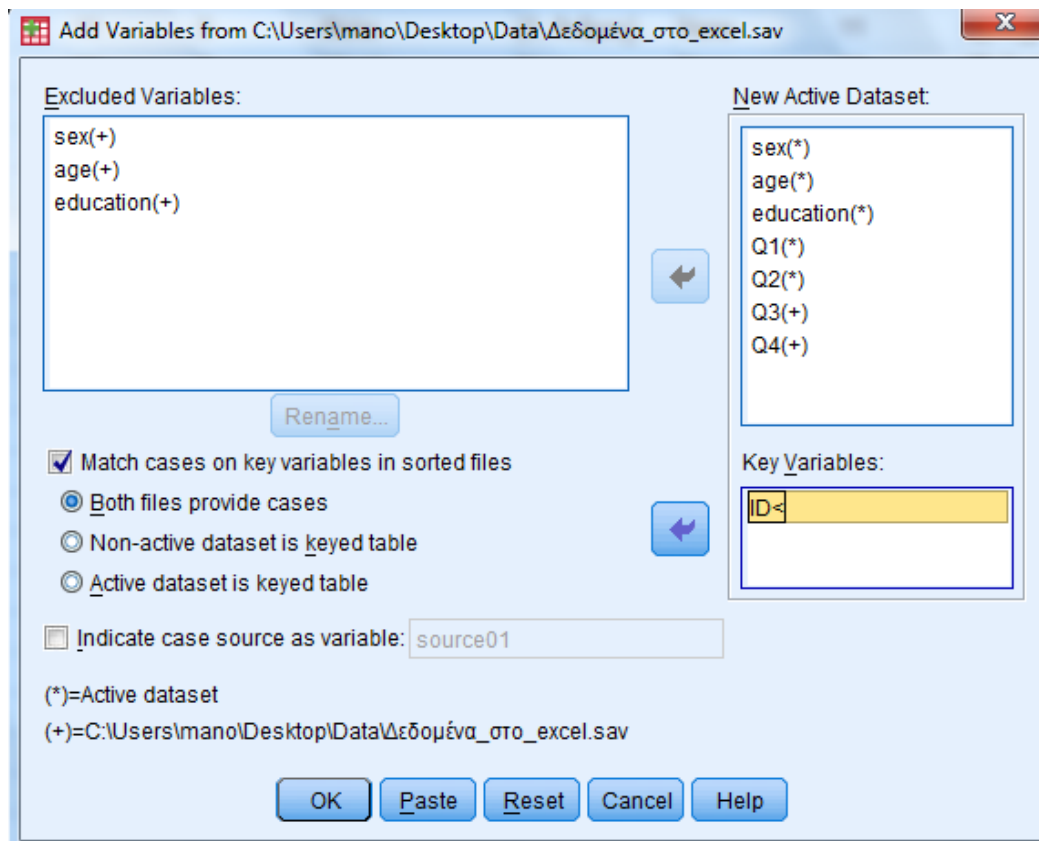
Η ίδια η εφαρμογή εμφανίζει μήνυμα διακοπής για την πρώτη περίπτωση και προειδοποιητικό μήνυμα για τη δεύτερη περίπτωση.



Εικόνα 3.18: Μήνυμα διακοπής ώστε το κλειδί να δηλωθεί και στα δύο προς ενοποίηση αρχεία



Εικόνα 3.19: Προειδοποιητικό μήνυμα ταξινόμησης των δεδομένων με βάση το κλειδί αντιστοίχισης



Εικόνα 3.20: Η υποενότητα Merge > Add Variables > Step 2

Στο παραπάνω παράδειγμα έχουμε δύο αρχεία με κοινές τις μεταβλητές *sex*, *age*, *education*, οι οποίες επειδή υπάρχουν και στα δύο αρχεία, εξαιρούνται από την ενοποίηση για να μην έχουμε διπλοκαταχωρήσεις. Και στα δύο αρχεία περιέχεται η μεταβλητή ταυτοποίησης της κάθε γραμμής *ID*, η οποία και ορίζεται ως η μεταβλητή κλειδί.

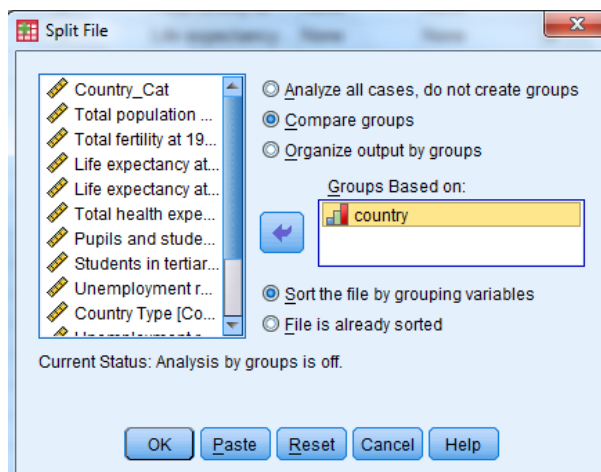
Με την ολοκλήρωση της διαδικασίας προκύπτει ένα αρχείο που περιέχει το σύνολο των μεταβλητών.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ID	String	7	0		None	None	7	Left	Nominal
2	sex	Numeric	11	0	Φύλο	{1, Άνδρας}...	None	11	Right	Nominal
3	age	Numeric	11	0	Ηλικία	None	None	11	Right	Scale
4	education	Numeric	11	0	Μορφωτικό επίπεδο	None	None	11	Right	Ordinal
5	Q1	Numeric	11	0	Ερώτηση 1 ερωτηματολογίου	None	None	11	Right	Ordinal
6	Q2	Numeric	11	0	Ερώτηση 2 ερωτηματολογίου	None	None	11	Right	Ordinal
7	Q3	Numeric	11	0	Ερώτηση 3 ερωτηματολογίου	None	None	11	Right	Nominal
8	Q4	Numeric	11	0	Ερώτηση 5 ερωτηματολογίου	None	None	11	Right	Nominal

Εικόνα 3.21: Αποτέλεσμα μετά την ενοποίηση Merge

3.5.6 Split File

Η δυνατότητα προσφέρεται από την επιλογή *Data > Split File*, με σκοπό την αντίθετη διαδικασία από την επιλογή *Merge*, δηλ. τον διαχωρισμό των παρατηρήσεων του αρχείου σε μικρότερες εικονικές διαμερίσεις. Ο διαχωρισμός επιτυγχάνεται με βάση τις τιμές μιας ή περισσότερων μεταβλητών *Compare Groups* με σκοπό την εφαρμογή της επόμενης ανάλυσης στο επιλεγμένο υποσύνολο. Η διαμέριση πραγματοποιείται ιεραρχικά αρχίζοντας από την πρώτη μεταβλητή στο πλαίσιο *Groups Based on*, στη συνέχεια εσωτερικά στην κάθε κατηγορία αυτής με τις τιμές της δεύτερης κ.ο.κ.



Εικόνα 3.22: Η υποενότητα *Split File*

Διαπιστώνουμε ότι η συγκεκριμένη δυνατότητα διαφέρει κατά πολύ από την επιλογή *Select Cases*, και επιπρόσθετα θα πρέπει να λάβουμε υπόψη ότι για την εφαρμογή της διαμέρισης θα πρέπει οι παρατηρήσεις να έχουν ταξινομηθεί από την επιλογή *Sort Cases*. Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*.

3.5.7 Weight Cases

Η δυνατότητα προσφέρεται από την επιλογή *Data > Weight Cases*, με σκοπό οι παρατηρήσεις να σταθμιστούν σύμφωνα με τις τιμές μιας άλλης μεταβλητής. Οι τιμές της μεταβλητής που χρησιμοποιούνται για τη στάθμιση μιας άλλης μεταβλητής θεωρούνται ως το πλήθος των επαναλήψεων (*replications*) της κάθε παρατήρησης από την αστάθμητη μεταβλητή.

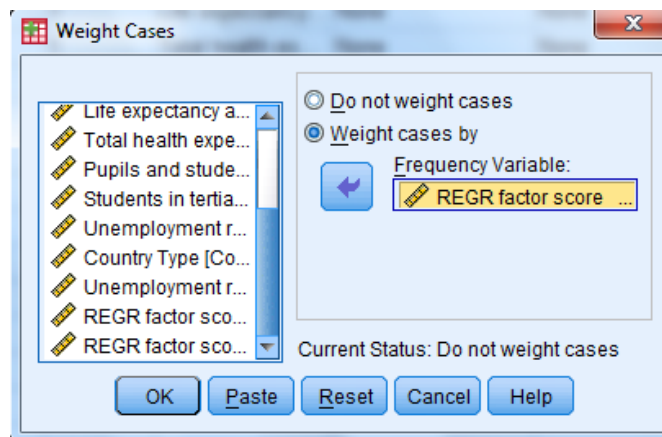
Η μεταβλητή που χρησιμοποιείται για να σταθμίσει μια άλλη, ενδέχεται είτε να είναι μια αρχική μεταβλητή του αρχείου είτε μια μεταβλητή που έχει προκύψει από κάποια ανάλυση (π.χ. παλινδρόμηση).

Επίσης, η χρήση αυτής της επιλογής γίνεται όταν η καταχώρηση δεδομένων έχει γίνει με ομαδοποίηση δεδομένων. Για παράδειγμα, έστω ότι έχουμε ένα δείγμα 100 παρατηρήσεων όπου κάποιος έχει μετρήσει το πλήθος των τιμών μιας μεταβλητής (έστω το πλήθος παιδιών σε μια οικογένεια), και αντί να λάβουμε αναλυτική γνώση όλων των παρατηρήσεων, λαμβάνουμε το πλήθος της κάθε παρατήρησης, δηλ. το πλήθος της τιμής «0 παιδιά», της τιμής «1 παιδί» κ.ο.κ. Για να μπορέσει ο ερευνητής να εξάγει περιγραφικούς δείκτες γι' αυτά τα δεδομένα θα πρέπει να χρησιμοποιήσει αυτή τη δυνατότητα καθώς στην πραγματικότητα η τιμή «1» εμφανίστηκε 22 φορές στο δείγμα.

Πλήθος παιδιών	Συχνότητα n_i
0	8
1	22
2	7
3	3
Σύνολο	40

Εικόνα 3.23: Πίνακας συχνοτήτων της μεταβλητής πλήθος παιδιών σε μια οικογένεια

Στο πεδίο *Frequency Variable* καταγράφεται η μεταβλητή στάθμησης των δεδομένων.



Εικόνα 3.24: Η υποεπίπτωση *Weight Cases*

Η διαδικασία ολοκληρώνεται με το πλήκτρο OK.

3.5.8 Restructure

Η δυνατότητα προσφέρεται από την επιλογή *Data > Restructure*, με σκοπό την αναδιοργάνωση του αρχείου μετατρέποντας τις μεταβλητές σε παρατηρήσεις, και αντίστροφα. Η εντολή αυτή διαφέρει από την εντολή *Transpose*, καθώς δεν αλλάζει απλά τη διάσταση του πίνακα εναλλάσσοντας γραμμές με στήλες, αλλά αναδιοργανώνει τη δομή των μεταβλητών και των παρατηρήσεων σύμφωνα με κανόνες.

Η αναδιοργάνωση συνήθως χρησιμοποιείται για την εξαγωγή στατιστικών αναλύσεων σε περιπτώσεις επαναλαμβανόμενων μετρήσεων, δηλ. μετρήσεων του ίδιου χαρακτηριστικού ενός άτομου που έχουν καταγραφεί σε διαφορετικές χρονικές στιγμές, και στο αρχείο έχουν καταγραφεί σε διαφορετικές στήλες.

Έστω για παράδειγμα ότι έχουμε τις ακόλουθες τιμές, βλ. Εικόνα 3.27:

	id	exertype	diet	time1	time2	time3	var
1	2	1	1	90	92	93	
2	3	1	1	97	97	94	
3	4	1	1	80	82	83	
4	5	1	1	91	92	91	
5	6	1	2	83	83	84	
6	7	1	2	87	88	90	
7	8	1	2	92	94	95	
8	9	1	2	97	99	96	
9	10	1	2	100	97	100	
10	11	2	1	86	86	84	
11	12	2	1	93	103	104	
12	13	2	1	90	92	93	
13	14	2	1	95	96	100	

Εικόνα 3.25: Δεδομένα σε Wide Format

Στόχος μας είναι το αρχείο να λάβει μια μορφή όπου τα δεδομένα των τριών μεταβλητών *time1*, *2*, *3* θα περιέχονται σε μία στήλη (μια μεταβλητή). Δύο είναι οι περιπτώσεις / μεθοδολογίες που θα συναντήσουμε:

- *Data > Restructure > Restructure selected variables into cases.*

Με σκοπό την αναδιοργάνωση από Wide σε Long Format.

- *Data > Restructure > Restructure selected cases into variables.*

Με σκοπό την αναδιοργάνωση από Long σε Wide Format. Η μέθοδος της αναδιοργάνωσης του αρχείου εξαρτάται από την ανάλυση που θα πραγματοποιηθεί, καθώς και από την πρωταρχική μορφή του αρχείου.

Στο παράδειγμα, υπό την υπάρχουσα μορφή των δεδομένων ο αναλυτής μπορεί μόνο να διεξάγει:

- *GLM Repeated Measures*
- *Paired Sample T- test*
- *2 Repeated Samples*
- *k - Repeated Samples*

Σε αυτές τις περιπτώσεις δεν υπάρχει η ανάγκη αναδιοργάνωσης.

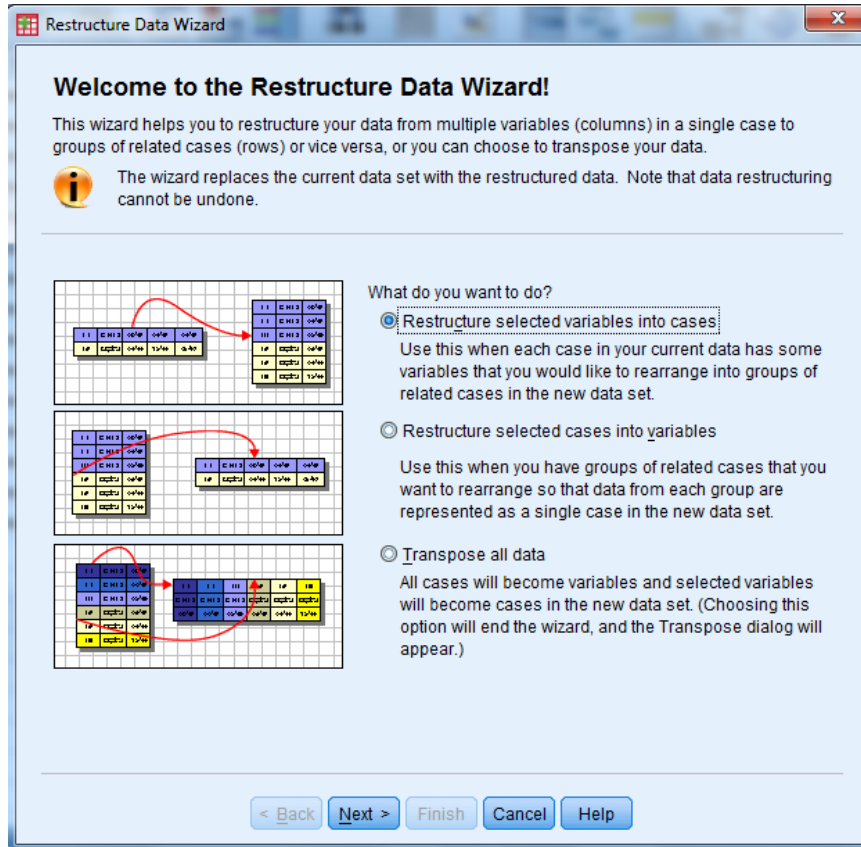
Όμως, αυτές δεν είναι παρά ορισμένες περιπτώσεις, και καθώς η πλειοψηφία των στατιστικών αναλύσεων απαιτούν μια μορφή όπου τα δεδομένα του αρχείου θα είναι σε εκτενή μορφή (long) με ομαδοποίηση παρατηρήσεων και όχι μεταβλητών, οπότε το αρχείο πρέπει να αναδιοργανωθεί.

Στο σημείο αυτό να αναφέρουμε ξανά ότι είναι σημαντικό να γίνεται μια αρχική συνεννόηση του αναλυτή με τους καταχωρητές των δεδομένων για να αποφεύγονται μελλοντικές καθυστερήσεις επεξεργασίας και αναδιοργάνωσης του αρχείου δεδομένων. Βέβαια καθώς τα δεδομένα ενδέχεται να προέρχονται από οποιαδήποτε πηγή και σε οποιαδήποτε μορφή, πρέπει να γνωρίζουμε το είδος της ανάλυσης που θα πραγματοποιηθεί έτσι ώστε να διαμορφώσουμε κατάλληλα το αρχείο στην επιθυμητή μορφή. Αναλυτικότερα για την ολοκλήρωση της διαδικασίας ακολουθούμε 7 βήματα.

Βήμα 1 από 7

Επιλογή της μεθόδου ανάλογα με την αρχική μορφή των δεδομένων μας, και το είδος της ανάλυσης που θα πραγματοποιήσουμε. Στο παράδειγμα μας (βλ. παραπάνω) τα δεδομένα είναι σε Wide μορφή οπότε επιλέγουμε:

Data > Restructure > Restructure selected variables into cases και στη συνέχεια το πλήκτρο *Next*.

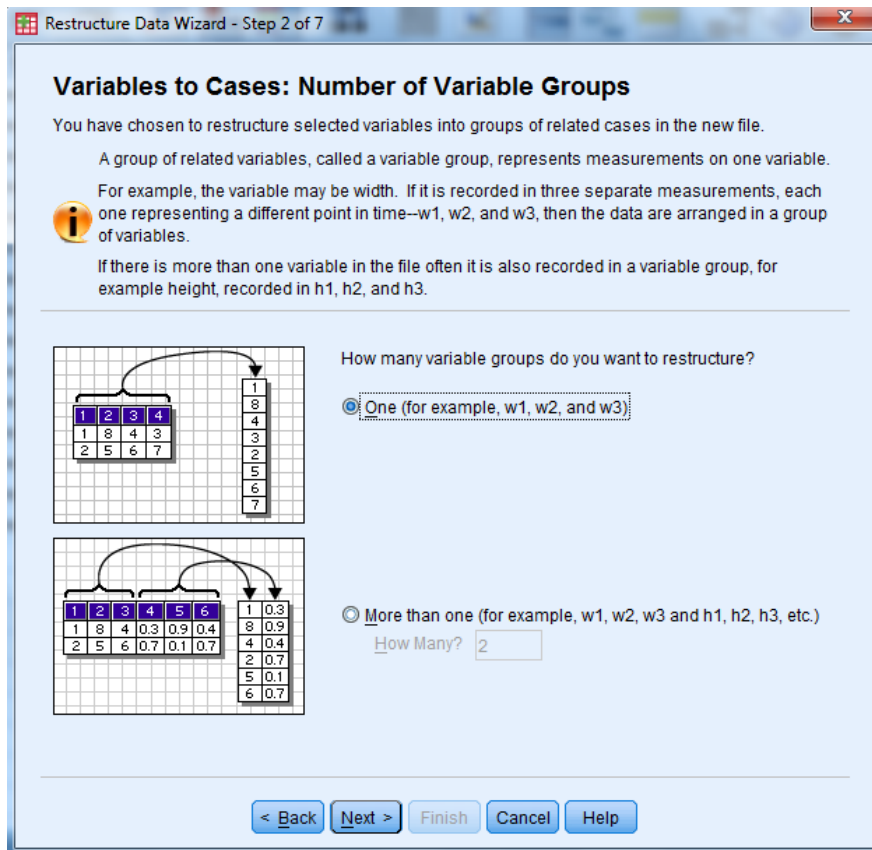


Εικόνα 3.26: Επιλογή μεθόδου αναδιοργάνωσης. Βήμα 1 από 7

Βήμα 2 από 7

Επιλέγουμε το πλήθος των ομαδοποιημένων μεταβλητών που υπάρχουν στο αρχείο. Στο παράδειγμα (βλ. παραπάνω) οι τρεις μεταβλητές *time1*, *time 2*, *time 3* αποτελούν μια ομάδα μεταβλητών καθώς αναφέρονται σε μετρήσεις του ίδιου χαρακτηριστικού που πραγματοποιήθηκαν σε διαφορετικές χρονικές στιγμές.

Έτσι, επιλέγουμε την επιλογή *One*. Σε αντίθετη περίπτωση θα επιλέγαμε *More than one* και θα ορίζαμε το πλήθος των ομάδων μεταβλητών.



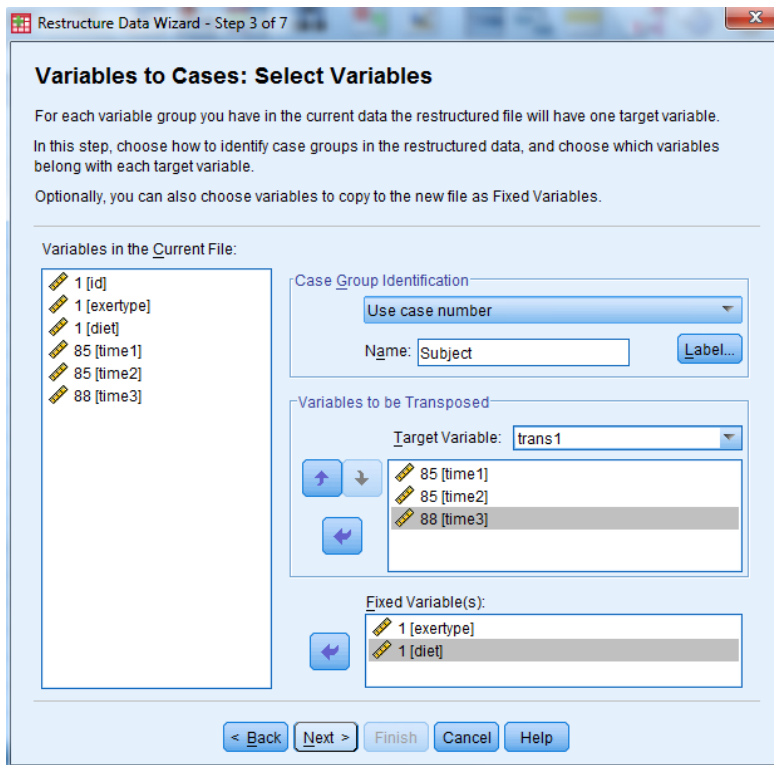
Εικόνα 3.27: Επιλογή πλήθους ομάδων μεταβλητών. Βήμα 2 από 7

Βήμα 3 από 7

Στο σημείο αυτό ορίζουμε τις μεταβλητές. Στο πλαίσιο *Case Group Identification* επιλέγουμε τη μεταβλητή εκείνη που ταυτοποιεί τις παρατηρήσεις. Στην προκειμένη περίπτωση θα δημιουργήσουμε τη μεταβλητή *Subject*, για την οποία καταγράφουμε το όνομά της στο πεδίο *Label*.

Στο πεδίο *Target Variable* δηλώνουμε το όνομα της νέας μεταβλητής, και στη συνέχεια εντός του πλαισίου προσθέτουμε τις επαναλαμβανόμενες μεταβλητές, *Time1*, *Time2*, *Time3*.

Τέλος στο πλαίσιο *Fixed Variable(s)* θα καταγραφούν εκείνες οι μεταβλητές που είναι σταθερές σε κάθε γραμμή (όπως *sex*, *age*), δηλ. τα αμετάβλητα χαρακτηριστικά του ατόμου.

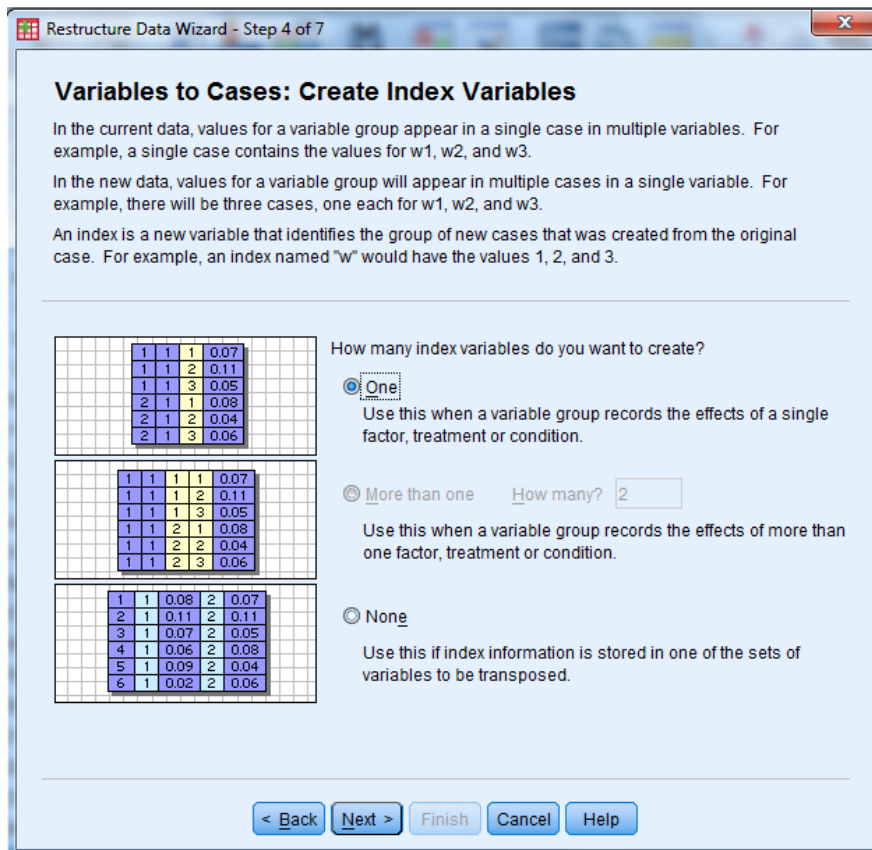


Εικόνα 3.28: Ορισμός των μεταβλητών. Βήμα 3 από 7

Βήμα 4 από 7

Σ' αυτό το βήμα δηλώνουμε τη δημιουργία μεταβλητών-δεικτών (Indexes), οι οποίες μας βοηθούν να αναγνωρίσουμε της επαναλήψεις. Οι μεταβλητές αυτές απαιτούνται σε ορισμένες στατιστικές αναλύσεις (π.χ. Repeated Measurement Analysis) για την αναγνώριση των υποκειμένων. Η εφαρμογή παρουσιάζει ένδειξη για τη μορφή που θα λάβουν οι δείκτες, και ο αναλυτής στο σημείο αυτό καλείται να ορίσει το πλήθος τους.

Εάν στο παράδειγμά μας είχαμε περισσότερες από μια ομάδες μεταβλητών, τότε θα ήταν αναγκαία η δημιουργία περισσότερων από μια μεταβλητών-δεικτών.



Εικόνα 3.29: Επιλογή πλήθους μεταβλητών-δεικτών. Βήμα 4 από 7

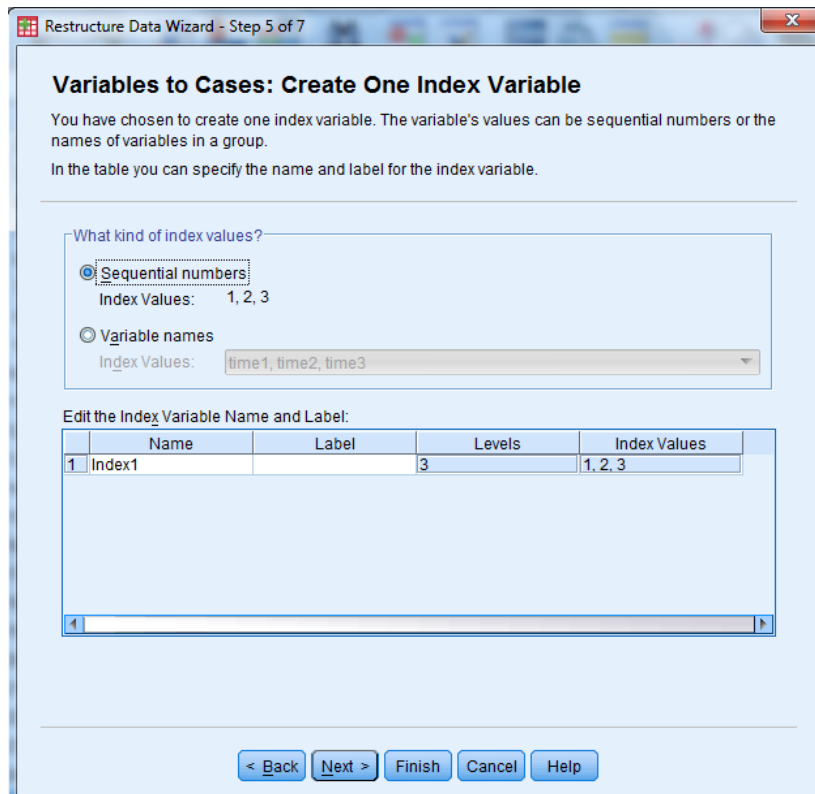
Βήμα 5 από 7

Σ' αυτό το βήμα ορίζουμε τις τιμές, την ονομασία και την περιγραφή των μεταβλητών-δεικτών. Οι τιμές που δύναται να λάβουν είναι είτε διαδοχικοί αριθμοί *sequential numbers* είτε το όνομα της μεταβλητής από την οποία προέκυψαν *variable name*.

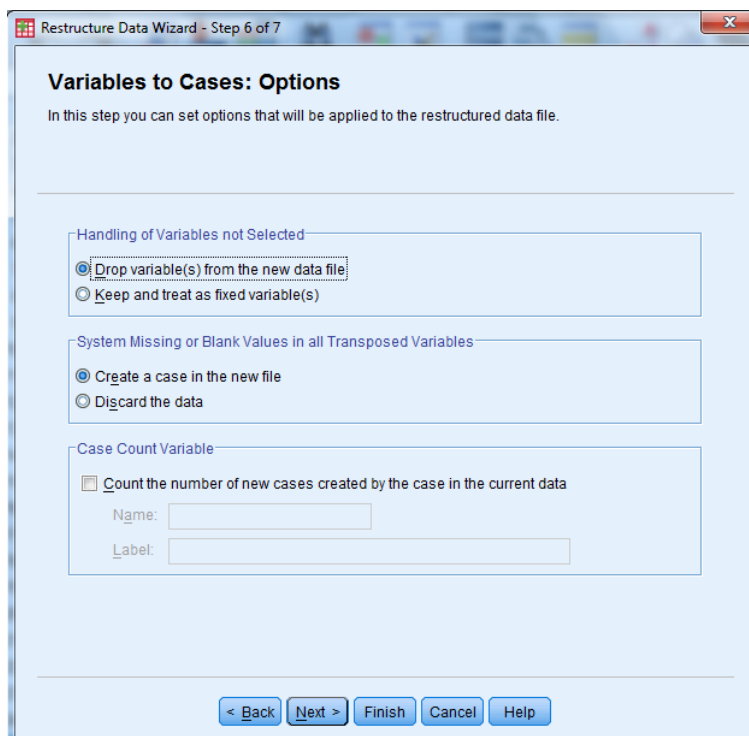
Στο πλαίσιο *Edit the Index Variable Name and Label* καταγράφονται η ονομασία και η περιγραφή των μεταβλητών-δεικτών.

Βήμα 6 από 7

Σε αυτό δηλώνουμε ορισμένες παραμέτρους (Options) που αφορούν στη μορφή του αναδιοργανωμένου αρχείου ως προς τις μεταβλητές που δεν θα αναδιοργανωθούν *Handling of Variables not Selected*, τις ελλείπουσες παρατηρήσεις *System Missing*, και τη δημιουργία μιας νέας μεταβλητής *Create Count Variable* που θα μετρά το πλήθος των παρατηρήσεων που προκύπτουν για κάθε παρατήρηση του αρχικού αρχείου.



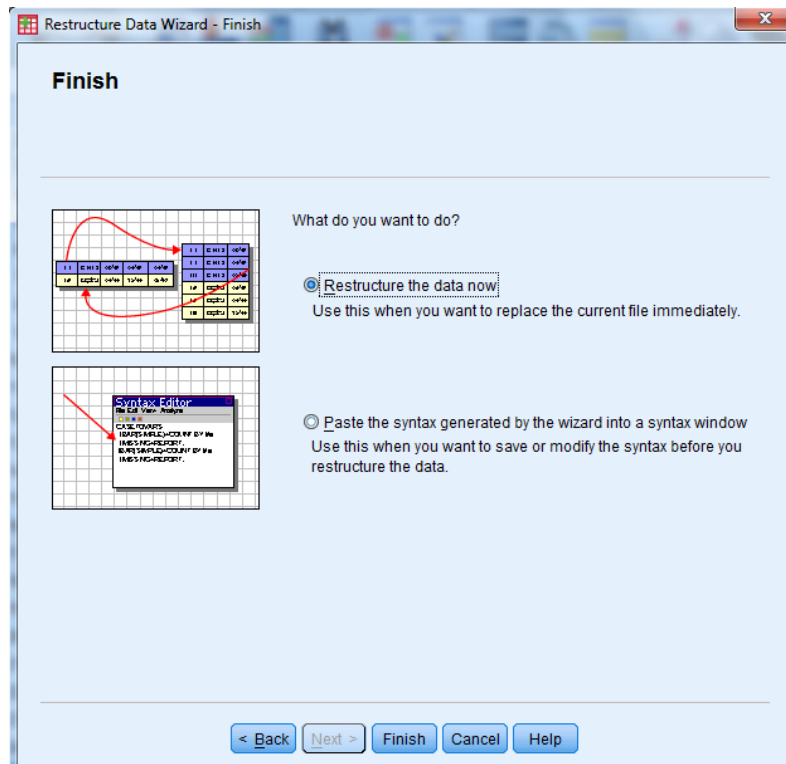
Εικόνα 3.30: Ορισμός στοιχείων της μεταβλητής-δείκτη. Βήμα 5 από 7



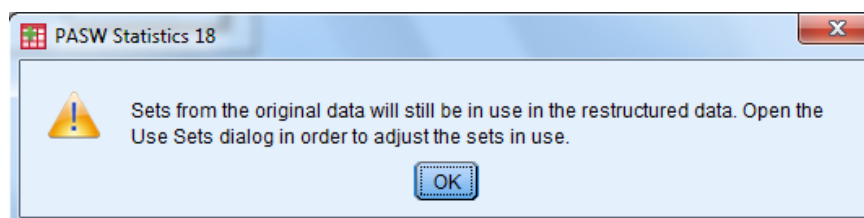
Εικόνα 3.31: Ορισμός λοιπών στοιχείων διαχείρισης. Βήμα 6 από 7

Βήμα 7 από 7

Στο τελευταίο βήμα ολοκληρώνεται η διαδικασία της αναδιοργάνωσης *Finish*. Ωστόσο, παρέχεται η δυνατότητα επανάκτησης της αρχικής μορφής των δεδομένων.



Εικόνα 3.32: Ολοκλήρωση της διαδικασίας. Βήμα 7 από 7



Εικόνα 3.33: Μήνυμα δυνατότητας επανάκτησης του αρχικού αρχείου

	Subject	exertype	diet	Index1	trans1	var	var	v
1	1	1	1	1	90			
2	1	1	1	2	92			
3	1	1	1	3	93			
4	2	1	1	1	97			
5	2	1	1	2	97			
6	2	1	1	3	94			
7	3	1	1	1	80			
8	3	1	1	2	82			
9	3	1	1	3	83			
10	4	1	1	1	91			
11	4	1	1	2	92			
12	4	1	1	3	91			
13	5	1	2	1	83			

Εικόνα 3.34: Το αποτέλεσμα της αναδιοργάνωσης. Δεδομένα σε long format

3.6 Μενού Transform

Περιέχει εντολές για τη δημιουργία νέων μεταβλητών ή τον μετασχηματισμό υπαρχόντων, βασιζόμενες σε ορισμένες συνεχείς κατανομές ή σε μαθηματικές και λογικές σχέσεις υπαρχόντων δεδομένων. Από τις εντολές αυτές θα ασχοληθούμε εκτενέστερα με τις εξής:

3.6.1 Recode

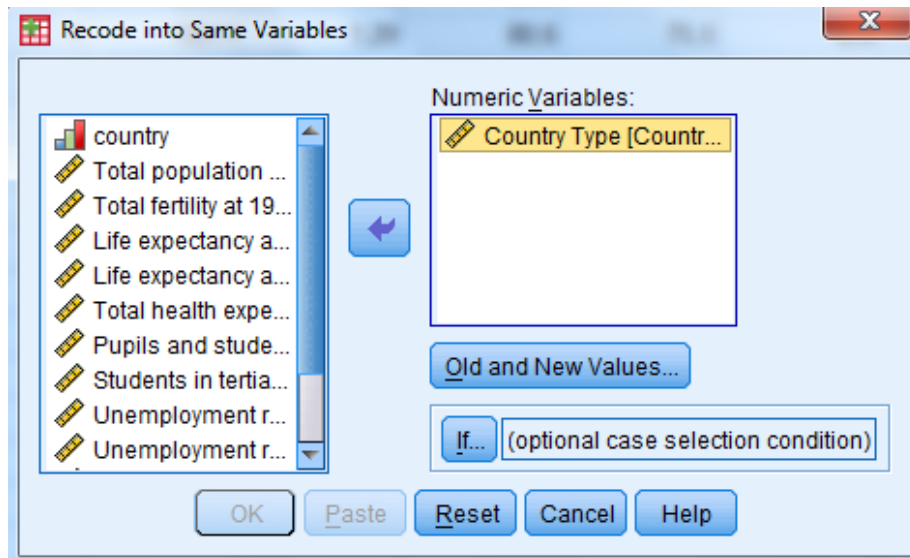
Η δυνατότητα προσφέρεται από την επιλογή *Transform > Recode*, με σκοπό την επανακωδικοποίηση των τιμών μιας μεταβλητής.

Η εντολή χωρίζεται σε δύο κατηγορίες:

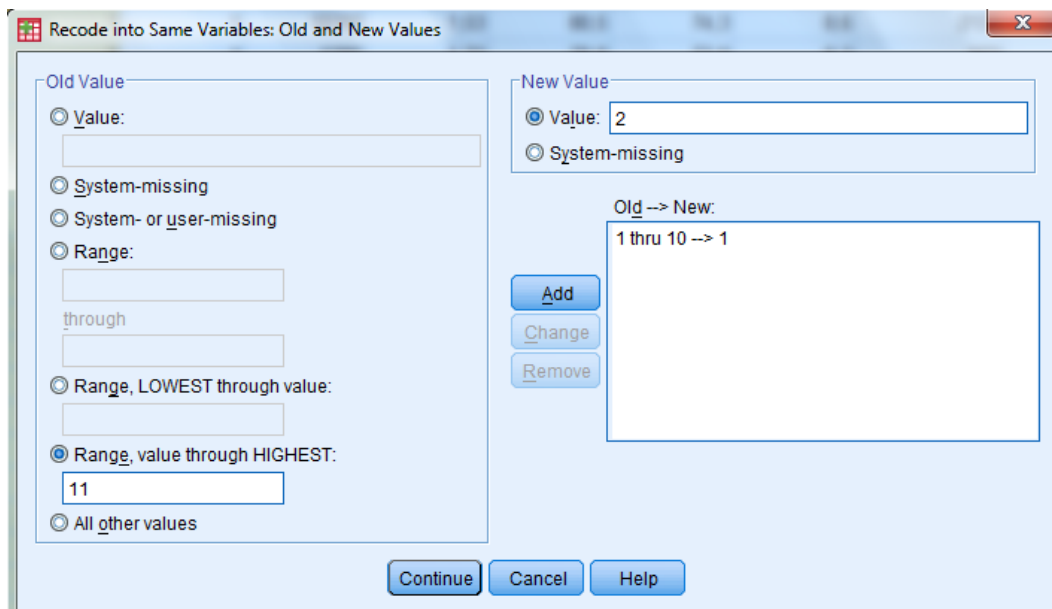
- *Transform > Recode > Into Same Variable* - Για την εφαρμογή της επανακωδικοποίησης των τιμών μιας μεταβλητής εντός της ίδιας μεταβλητής.
- *Transform > Recode > Into Different Variable* - Για την εφαρμογή της επανακωδικοποίησης των τιμών μιας μεταβλητής σε μια νέα μεταβλητή, διατηρώντας την αρχική μεταβλητή αμετάβλητη.

Transform > Recode > Into Same Variable

Στο πρώτο πλαίσιο διαλόγου ο αναλυτής ορίζει στο πλαίσιο *Numeric Variables* τη μεταβλητή ή τις μεταβλητές που θα μετασχηματίσει, και στη συνέχεια στο πλαίσιο διαλόγου *Old and New Values* τις παλαιές και τις νέες τιμές.



Εικόνα 3.35: Η υποενότητα Recode Into Same Variables



Εικόνα 3.36: Η υποενότητα Old and New Values

Έστω ότι ο ερευνητής θέλει να μετασχηματίσει μια ποσοτική μεταβλητή σε μια ποιοτική μεταβλητή, ομαδοποιώντας τα δεδομένα της συνεχούς μεταβλητής σε κλάσεις. Ο ορισμός των νέων τιμών πραγματοποιείται στο υποπλαίσιο διαλόγου *Old & New Values*, όπου αναλυτικά έχουμε:

Old Values

Στο σημείο αυτό δηλώνονται οι αρχικές τιμές, ή το εύρος τιμών, ή ακόμα και οι ελλείπουσες τιμές της υπάρχουσας μεταβλητής που θα επανακωδικοποιηθούν. Ο χειρισμός του εύρους μπορεί να οριστεί είτε για τα ανώτατα, είτε για κατώτατα άκρα.

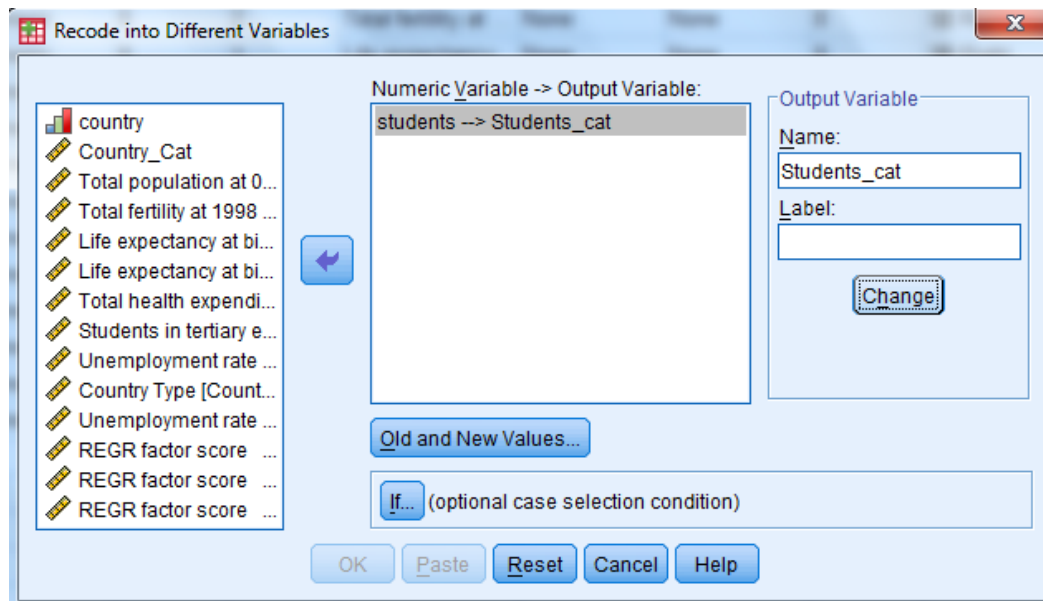
New Values

Στο σημείο αυτό ορίζονται οι νέες τιμές της μεταβλητής που θα αντικαταστήσουν τις αρχικές για κάθε επιλογή.

Με το πλήκτρο *Add* κάθε αντιστοίχιση μεταφέρεται στο υποπλαίσιο *Old > New*, ενώ από τα πλήκτρα *Change* και *Remove* παρέχεται η δυνατότητα τροποποίησης ή διαγραφής μιας αντιστοίχισης. Οι αντιστοιχίσεις εμφανίζονται ταξινομημένες σύμφωνα πάντα με τις αρχικές τιμές της μεταβλητής.

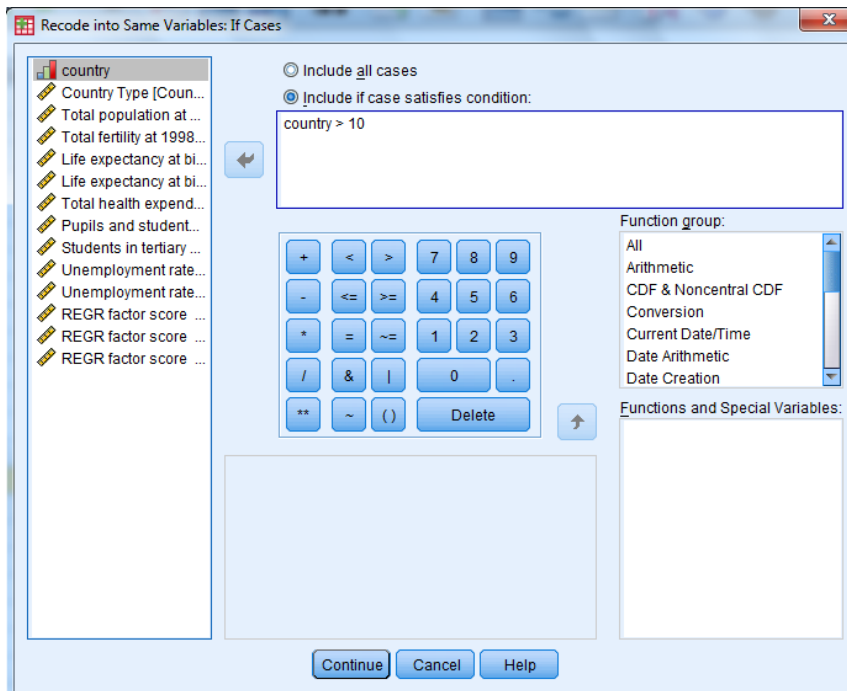
Transform > Recode > Into Different Variable

Με αυτή την επιλογή, η αρχική μεταβλητή παραμένει αναλλοίωτη και η επανακωδικοποίηση εφαρμόζεται για τη δημιουργία νέας μεταβλητής. Η διαδικασία είναι όμοια με την προηγούμενη επιλογή, με τη διαφορά ότι πλέον ο χρήστης θα πρέπει να ορίσει τη σύντομη και την πλήρη ονομασία της νέας μεταβλητής στο πλαίσιο *Output Variable* στα πεδία *Name* και *Label*, και επιπρόσθετα από το πλήκτρο *Change* να δηλώσει ποια μεταβλητή αντικαθιστά την επιλεγμένη υπάρχουσα.



Εικόνα 3.37: Η υποενότητα *Recode into Different Variables*

Σε περιπτώσεις που η επανακωδικοποίηση είναι περισσότερο πολύπλοκη, ή για παράδειγμα πρέπει να εφαρμοστεί σε ένα υποσύνολο των αρχικών τιμών, τότε αυτό επιτυγχάνεται από το πλήκτρο *If (optional case selection condition)* υποπλαίσιο διαλόγου *If cases*. Το παράθυρο που εμφανίζεται είναι σχεδόν ίδιο με αυτό της επιλογής *Compute* παρέχοντας τη δυνατότητα μετασχηματισμού εφαρμόζοντας μαθηματικές πράξεις ή συναρτήσεις. Προκαθορισμένες συναρτήσεις περιέχονται στη λίστα *Function Group*.



Εικόνα 3.38: Η υποενοότητα *If cases*

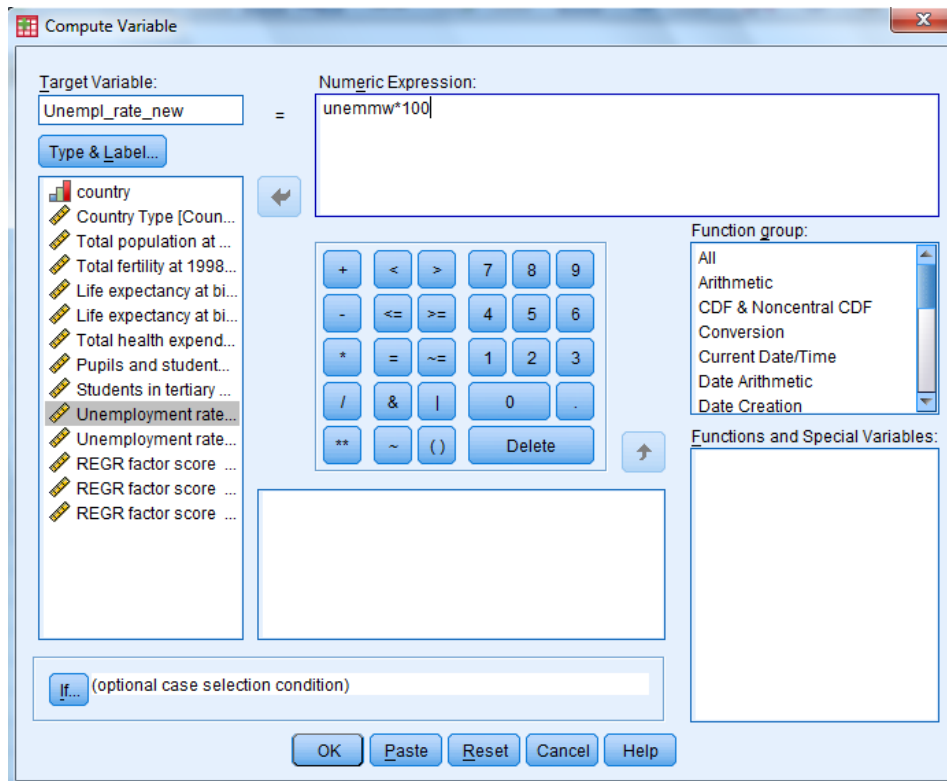
Μια συνήθης περίπτωση επανακωδικοποίησης των τιμών μιας μεταβλητής είναι όταν ο αναλυτής έχει να διαχειριστεί διατάξιμες μεταβλητές από ερωτηματολόγιο κοινωνικοψυχολογικών μετρήσεων, και θέλει να κάνει αλλαγή στην κλίμακα μιας διατάξιμης μεταβλητής. Για παράδειγμα, οι τιμές 1 - 5 (κλίμακα Likert) να τροποποιηθούν σε «-1», «-0,5», «0», «+0,5», «+1», ώστε στη συνέχεια ο χρήστης να προβεί σε δημιουργία δείκτη που άμεσα θα λαμβάνει την ένδειξη εάν οι προτιμήσεις των ερωτώμενων είναι «θετικές» ή «αρνητικές».

3.6.2 Compute

Η δυνατότητα προσφέρεται από την επιλογή *Transform > Compute*, με σκοπό τη δημιουργία νέων μεταβλητών υπό την εφαρμογή μιας μαθηματικής (αριθμητικής) έκφρασης.

Στο πεδίο *Target Variable* καταγράφεται η σύντομη ονομασία της νέας μεταβλητής, ενώ στο δεξιό μέρος *Numeric Expression* καταγράφεται η μαθηματική έκφραση (ο τύπος μετασχηματισμού) της μεταβλητής ή η μαθηματική έκφραση των μεταβλητών που θα οδηγήσουν στη δημιουργία της νέας μεταβλητής.

Η χρήση της συγκεκριμένης επιλογής συνηθίζεται σε μετασχηματισμούς των μονάδων μέτρησης, π.χ. από GHz σε MHz, ή στη δημιουργία μεταβλητών υψωμένων σε δυνάμεις για τη μετέπειτα χρήση τους σε μοντέλα παλινδρόμησης. Αξίζει να σημειωθεί ότι κατά τη διαδικασία υπολογισμού της νέας μεταβλητής παρέχεται πληθώρα μαθηματικών και στατιστικών συναρτήσεων στο πλαίσιο *Function Group*.



Εικόνα 3.39: Η υποενοότητα Compute

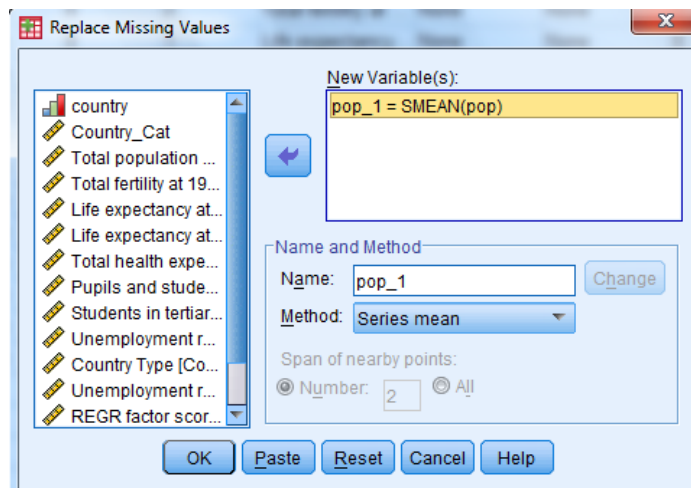
Παράδειγμα

Έστω ότι ο ερευνητής έχει καταγράψει μόνο το έτος γέννησης των ερωτώμενων και θέλει να υπολογίσει την ηλικία τους. Σ' αυτή την περίπτωση η μαθηματική έκφραση που θα καταγραφεί στο πλαίσιο *Numeric Expression* είναι : «2015-brthyear», όπου brthyear η μεταβλητή «Έτος Γέννησης».

3.6.3 Replace Missing Values

Η δυνατότητα προσφέρεται από την επιλογή *Transform > Replace Missing Values*, με σκοπό την αντικατάσταση των ελλειπουσών τιμών, όταν αυτές υπάρχουν, καθώς δημιουργούν προβλήματα στις στατιστικές αναλύσεις. Το κυριότερο πρόβλημα είναι το ότι μειώνουν το πλήθος των προς επεξεργασία δεδομένων, με αποτέλεσμα να μην ισχύουν βασικές υποθέσεις της μεθοδολογίας. Στις περιπτώσεις που η εφαρμογή διαπιστώνει ότι οι υποθέσεις της κάθε μεθοδολογίας καταστρατηγούνται, τότε εμφανίζει προειδοποιητικό μήνυμα (Warning) στην αρχή των αποτελεσμάτων της ανάλυσης, για ενημέρωση του αναλυτή, ώστε αυτός στη συνέχεια να προβεί σε κάποια μέθοδο διαχείρισής τους.

Σ' αυτές τις περιπτώσεις η αντικατάσταση των ελλειπουσών τιμών κρίνεται αναγκαία και μπορεί να γίνει με την αντικατάσταση αυτών των μεταβλητών (των μεταβλητών που έχουν ελλείπουσες τιμές) από νέες μεταβλητές οι οποίες προκύπτουν από τις αρχικές έχοντας αντικαταστήσει της ελλείπουσες τιμές με κάποια τιμή. Η τιμή που θα καταγραφεί στην θέση της ελλείπουσας τιμής προκύπτει μέσω της επιλεγμένης μεθόδου.



Εικόνα 3.40: Η υποενότητα Replace Missing Values

Οι νέες μεταβλητές που θα προκύψουν έχουν ως ονομασία την ονομασία της αρχικής μεταβλητής (τους έξι πρώτους χαρακτήρες) και αριθμημένη κατάληξη, διατηρώντας τα value labels της αρχικής (εφόσον υπάρχουν).

Στο πλαίσιο *New Variable* ορίζονται οι μεταβλητές οι οποίες έχουν πρόβλημα ελλειπουσών τιμών και στο πεδίο *Method* επιλέγεται η μέθοδος διαχείρισής τους.

Μέθοδοι Διαχείρισης Ελλειπουσών Τιμών

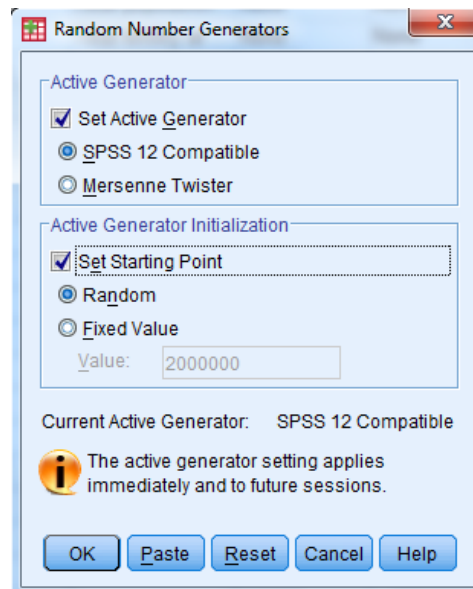
Η εφαρμογή διαθέτει τις ακόλουθες μεθόδους:

- *Mean*: Ωστε οι ελλείπουσες τιμές να αντικατασταθούν από τον μέσο όρο των υπαρχόντων παρατηρήσεων της μεταβλητής.
- *Mean or nearby points*: Ωστε οι ελλείπουσες τιμές να αντικατασταθούν από τον μέσο όρο των παρατηρήσεων που βρίσκονται γύρω από την ελλείπουσα τιμή. Το πλήθος των παρατηρήσεων που θα ληφθούν υπόψη (πάνω και κάτω από την ελλείπουσα παρατήρηση) ορίζεται στο πεδίο *span of nearby points*. Η συγκεκριμένη δυνατότητα είναι σύνηθες να χρησιμοποιείται όταν η μεταβλητή είναι διατάξιμη και η τιμή της ελλείπουσας τιμής εξαρτάται από τη θέση της.
- *Median or nearby points*: Ωστε οι ελλείπουσες τιμές να αντικατασταθούν από τη διάμεσο των παρατηρήσεων που βρίσκονται γύρω από την ελλείπουσα τιμή. Το πλήθος των παρατηρήσεων που θα ληφθούν υπόψη (πάνω και κάτω από την ελλείπουσα παρατήρηση) ορίζεται στο πεδίο *span of nearby points*. Η συγκεκριμένη δυνατότητα είναι σύνηθες να χρησιμοποιείται όταν η μεταβλητή είναι διατάξιμη και η τιμή της ελλείπουσας τιμής εξαρτάται από τη θέση της.
- *Linear Interpolation*: Ωστε η αντικατάσταση να πραγματοποιηθεί με τη μέθοδο της γραμμικής παρεμβολής, χρησιμοποιώντας τις παρατηρήσεις πριν και μετά τη θέση της ελλείπουσας τιμής. Εάν οι παρατηρήσεις πριν και μετά είναι επίσης ελλείπουσες τιμές, τότε η μέθοδος δεν μπορεί να εφαρμοστεί.
- *Linear Trend at point*: Ωστε οι ελλείπουσες τιμές να αντικατασταθούν με τη χρήση υποδείγματος γραμμικής παλινδρόμησης. Δηλαδή έχουμε ως εξαρτημένη μεταβλητή την ίδια τη μεταβλητή και ως ανεξάρτητη τη θέση της κάθε τιμής της μεταβλητής στο αρχείο. Ομοίως η μέθοδος αυτή πρέπει να επιλέγεται εάν υπάρχει εκ των προτέρων γνώση ότι η τιμή της ελλείπουσας σχετίζεται με τη θέση στην οποία βρέθηκε.

3.6.4 Random Number Generator

Η δυνατότητα προσφέρεται από την επιλογή *Transform > Random Number Generators*, με σκοπό την παραγωγή τυχαίων αριθμών από προκαθορισμένους αλγορίθμους (γεννήτριες) τυχαίων αριθμών.

Στην ενότητα *Active Generator*, επιλέγεται το είδος του αλγορίθμου και στην ενότητα *Active Generator Initialization* στο πεδίο *Set Starting Point* οι τιμές εκκίνησης του αλγορίθμου. Ο αναλυτής μπορεί να ορίσει μια σταθερή τιμή στο πεδίο *Value* της επιλογής *Fixed Values* έτσι ώστε σε κάθε στάδιο του αλγορίθμου να λαμβάνεται η συγκεκριμένη τιμή.



Εικόνα 3.41: Η υποενότητα *Random Number Generators*

3.7 Μενού *Analyze*

Αποτελεί τον βασικό κορμό και τον λόγο ύπαρξης αυτού του προγράμματος, καθώς περιέχει όλες τις εντολές στατιστικής ανάλυσης που προσφέρονται από το πρόγραμμα. Οι εντολές αυτές θα περιγραφούν αναλυτικότερα στα επόμενα κεφάλαια του κειμένου αυτού.

Βιβλιογραφικές Αναφορές Κεφαλαίου

Ελληνόγλωσσες

Howitt, D., Cramer, D. (Επιμέλεια: Σ. Κοντάκος) (2011). *Στατιστική με το SPSS 16.0*. Αθήνα: Εκδόσεις Κλειδάριθμος.

Ξενόγλωσσες

Carver, R. & Nash, J. (2011). *Doing Data Analysis with SPSS: Version 18.0*. 5th Edition Easton Cengage Learning.

Coakes, S. J. & Steed, L. G. (1999). *SPSS: Analysis without anguish: Versions 7.0, 7.5, 8.0 for Windows*. Brisbane: Jacaranda Wiley.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London:Sage.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, pp. 799-821.

Marques de Sa, J. P. (2007). *Applied Statistics Using SPSS. STATISTICA, MATLAB and R*, 2nd Edition, Porto: Universidade de Porto

Norusis Marija, J. (2002). *SPSS 11.0 Guide to data analysis*. Upper Saddle River New Jersey: Prentice Hall.

Κεφάλαιο 4

Περιγραφική Στατιστική - Γραφήματα

Σύνοψη

Αρχικά περιέχονται βασικές έννοιες και ορισμοί για την περιγραφική στατιστική, τις έννοιες του πληθυσμού και του δείγματος και τα είδη των μεταβλητών. Γίνεται αναφορά στον υπολογισμό περιγραφικών στατιστικών στο SPSS μέσω των διαδρομών *Descriptive Statistics > Frequencies*, *Descriptive Statistics > Descriptives*, *Descriptive Statistics > Explore*, *Descriptive Statistics > Crosstabs*.

Παράλληλα εξηγούνται οι στατιστικές έννοιες που χρειάζονται, ενδεικτικά αναφέρουμε *Trimmed Mean* - Ξακρισμένη μέση τιμή και *Confidence Interval for Mean* - Διάστημα Εμπιστοσύνης για τον μέσο.

Γίνεται αναφορά στην έννοια της συσχέτισης, στους συντελεστές συσχέτισης και στον τρόπο υπολογισμού τους (μέσω της επιλογής *Correlate*). Γίνεται αναλυτική περιγραφή της εντολής *Multiple Response* (*Analyze > Multiple Response > Define Multiple Response Sets*). Και τέλος γίνεται αναφορά στα διαγράμματα μέσω του μενού *Graphs*. Συγκεκριμένα περιγράφονται τα Ραβδογράμματα *Bar Charts* (*Simple* και *Clustered*), τα Γραμμικά *Line* (επιλογές *Simple* και *Multiple*) τα Γραφήματα Περιοχής *Area*, τα Κυκλικά *Pie*, Τα Υψηλών - Χαμηλών τιμών *High - Low* (επιλογές *Simple* και *Clustered*) τα Θηκογράμματα ή Πλαισίου Απολήξεων *BoxPlots* (επιλογές *Simple* και *Clustered*), τα Διασποράς *Scatter/ Dot* (επιλογές *Simple* και *Matrix* και *Scatter*) και τα Ιστογράμματα, *Histogram*.

Προσπαιτούμενη γνώση

Καλύτερη κατανόηση του κεφαλαίου προκύπτει αν ο αναγνώστης έχει γνώσεις περιγραφικής στατιστικής (ύλη που κατά βάση καλύπτεται στο πρώτο κεφάλαιο). Τα εγχειρίδια που παρουσιάζουν με πληρότητα την ύλη του συγκεκριμένου κεφαλαίου είναι αυτά που αναφέρθηκαν στο Κεφάλαιο 3.

4.1 Βασικές Έννοιες και Ορισμοί

Ως **Στατιστική (Statistic)** ορίζεται η επιστήμη που απασχολείται με τη συγκέντρωση (συλλογή), οργάνωση, παρουσίαση και ανάλυση δεδομένων.

Δηλαδή ασχολείται με την:

- Συλλογή και Επιλογή ενός δείγματος από τον πληθυσμό και λήψη μετρήσεων από αυτό.
- Οργάνωση των δεδομένων με την χρήση των Η/Υ μέσω διαφόρων λογιστικών φύλλων (π.χ. Excel) ή στατιστικών πακέτων (π.χ. SPSS, STATA, SAS, R, SPlus, Minitab κ.ά).
- Παρουσίαση περιγραφικών χαρακτηριστικών των δεδομένων με χρήση πινάκων συχνοτήτων, γραφημάτων και περιγραφικών μέτρων.
- Στατιστική συμπερασματολογία με χρήση μοντέλων [π.χ. γραμμική παλινδρόμηση (*linear regression*)].

Η επιστήμη της Στατιστικής χωρίζεται σε δύο μεγάλους κλάδους, την Περιγραφική Στατιστική (*Descriptive Statistics*), η οποία θα αναλυθεί στο παρόν κεφάλαιο, και την Στατιστική Συμπερασματολογία (*Inferential Statistics*), η οποία θα αναλυθεί στα επόμενα κεφάλαια.

Περιγραφική στατιστική (Descriptive Statistics)

Σκοπός της Περιγραφικής Στατιστικής είναι η συνοπτική παρουσίαση των δεδομένων. Αποτελεί ένα στατιστικό εργαλείο με σκοπό τη συγκέντρωση ταξινόμηση και παρουσίαση πρωτογενών δεδομένων σε κατανοητή μορφή. Γίνεται με τη χρήση πινάκων (συχνοτήτων, διπλής εισόδου), γραφημάτων

(ραβδογράμματα, θηκογράμματα, διασποράς κ.ά.), και στατιστικών μέτρων (μέτρα κεντρικής τάσης, μέτρα μεταβλητότητας και μέτρα αναφορικά με τη μορφή της κατανομής συχνότητας).

Βασικές Στατιστικές Έννοιες

Στατιστικός πληθυσμός (population)

Με τον όρο αυτό καλούμε το σύνολο των ατόμων ή αντικειμένων στα οποία αναφέρονται οι παρατηρήσεις μας. Τα στοιχεία του συνόλου αυτού καλούνται στατιστικές μονάδες ή άτομα. Ο πληθυσμός μας μπορεί να είναι πεπερασμένος ή άπειρος, υπαρκτός ή ιδεατός. Τα άτομα αυτά διακρίνονται για κάποιο χαρακτηριστικό ή ιδιότητα.

Στατιστική μεταβλητή (Variable)

Η ιδιότητα ή το χαρακτηριστικό, ως προς το οποίο ερευνάται ένας πληθυσμός καλείται **στατιστική μεταβλητή (variable)**. Κάθε πληθυσμός έχει διαφορετικά χαρακτηριστικά (ιδιότητες), όπως για παράδειγμα μια ομάδα ατόμων μπορούν να διαφέρουν ως προς την ηλικία, το φύλο, το βάρος κ.ά. Για να μετρήσουμε τα μεταβαλλόμενα χαρακτηριστικά χρειάζεται να χρησιμοποιήσουμε είτε κατάλληλα εργαλεία μέτρησης είτε κατάλληλες κωδικοποιήσεις με σκοπό τη διεξαγωγή μετρήσεων για κάθε χαρακτηριστικό (μεταβλητή). Τα αριθμητικά δεδομένα (δηλαδή το πεδίο τιμών των μεταβλητών) λέγονται παρατηρήσεις (observations) ή μετρήσεις. Οι μετρήσεις αυτές ονομάζονται τιμές των μεταβλητών. Είναι προφανές ότι οι τιμές διαφοροποιούνται μεταξύ των ατόμων ή μεταξύ των στατιστικών χαρακτηριστικών. Οι τιμές των μεταβλητών μπορεί να είναι πραγματικοί αριθμοί (π.χ. βάρος = 75kg) ή φράσεις (π.χ. φύλο = άνδρας).

Είδη Μεταβλητών

Η φύση των μεταβλητών καθορίζει το είδος των δεδομένων. Οι μεταβλητές διακρίνονται σε **ποσοτικές** (quantitative) και **ποιοτικές** (qualitative) ανάλογα με το εάν οι τιμές εκφράζουν αριθμητικά ή ονομαστικά δεδομένα.

Οι ποσοτικές μεταβλητές διακρίνονται με τη σειρά τους σε **συνεχείς** (continues) και **διακριτές** (discrete) ανάλογα με το αν είναι συνεχείς ή διακριτές. Συνεχείς μπορούν να ονομαστούν οι μεταβλητές που δύναται να λάβουν μια οποιαδήποτε τιμή, έστω και θεωρητική, εντός ενός διαστήματος (π.χ. ύψος, βάρος). Από την άλλη μεριά, διακριτές μπορούν να οριστούν οι μεταβλητές οι οποίες λαμβάνουν ακέραιο αριθμό (π.χ. αριθμός παιδιών σε μία οικογένεια).

Οι ποιοτικές μεταβλητές διακρίνονται σε μη διατάξιμες (nominal) και **διατάξιμες** (ordinal). Οι μεταβλητές, οι οποίες δίνουν τη δυνατότητα στον ερευνητή να διατάξει και να διαβαθμίσει τις κατηγορίες που προκύπτουν από τις τιμές ονομάζονται διατάξιμες [π.χ. επίπεδα εκπαίδευσης, γνώμη καταναλωτή (Καθόλου, Λίγο, Αρκετά, Πολύ, Πάρα Πολύ)]. Οι υπόλοιπες που δεν παρέχουν τη δυνατότητα διάταξης αλλά με βάση τα χαρακτηριστικά που εκφράζουν οι τιμές τους επιτρέπουν απλά και μόνο τη διάκριση ορισμένων κατηγοριών ονομάζονται μη διατάξιμες (π.χ. φύλο, οικογενειακή κατάσταση).

Είναι σημαντικό η αναγνώριση του είδους της μεταβλητής να γίνει ορθά, καθώς σε αυτό βασίζεται η επιλογή των στατιστικών αναλύσεων που θα εφαρμοστούν.

Πιο συγκεκριμένα:

Είδος Μεταβλητής	Περιγραφική Στατιστική Ανάλυση
Ποιοτική - Μη Διατάξιμη	Κατανομές Συχνοτήτων και κατάλληλα Διαγράμματα (π.χ. ραβδογράμματα, κυκλικά).
Ποιοτική - Διατάξιμη ή Ποσοτική - Διακριτή	Δείκτες τάσης, θέσης, διασποράς (εύρος) και κατάλληλα Διαγράμματα (π.χ. ραβδογράμματα ή ιστογράμματα).
Ποσοτική - Συνεχής	Δείκτες τάσης, θέσης, διασποράς, λοξότητας και κυρτότητας και κατάλληλα Διαγράμματα (π.χ. ιστογράμματα, γραμμής, διασποράς).

Πίνακας 4.1: Αναγνώριση του είδους μιας μεταβλητής

Οι ενότητες της εφαρμογής SPSS που θα αναλυθούν στο παρόν κεφάλαιο, και που αφορούν στον κλάδο της περιγραφικής στατιστικής, είναι οι:

- *Descriptive Statistics* > Frequencies,
- *Descriptive Statistics* > Descriptives,
- *Descriptive Statistics* > Explore,
- *Descriptive Statistics* > Crosstabs,
- *Tables*,
- *Correlate*,
- *Multiple Response*,
- *Graphs*.

4.2 Frequencies

Η δυνατότητα προσφέρεται από την επιλογή *Analyze* > *Descriptive Statistics* > *Frequencies*, με σκοπό την εξαγωγή πινάκων συχνοτήτων Ποιοτικών ή Ποσοτικών Διακριτών μεταβλητών. Η εντολή χρησιμοποιείται ως επί το πλείστον για ποιοτικές μεταβλητές, ωστόσο δύναται να χρησιμοποιηθεί και για την ανάλυση ποσοτικών μεταβλητών, όπως θα δούμε παρακάτω.

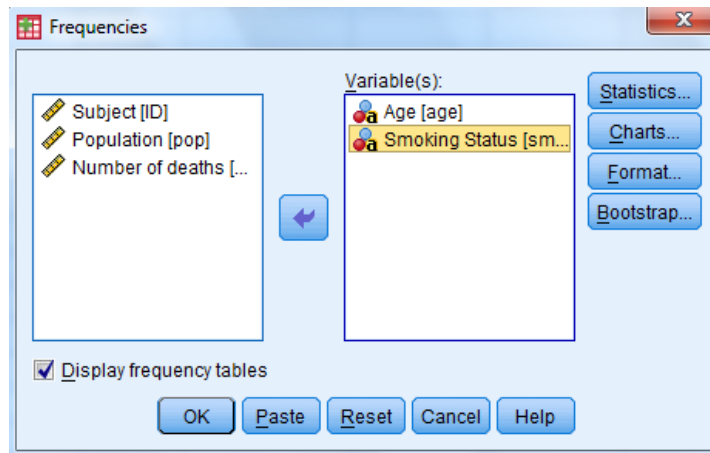
Ποιοτικές μεταβλητές

Η εντολή χρησιμοποιείται για την εξαγωγή πινάκων συχνοτήτων. Οι κατανομές συχνότητας (frequency distributions) αποτελούν μία από τις πιο γνωστές μεθόδους παρουσίασης ενός συνόλου δεδομένων. Στόχος είναι να δοθεί μία πλήρης εικόνα για την κατανομή των δεδομένων στην κλίμακα μέτρησής της.

Τα ποιοτικά δεδομένα, είτε αυτά είναι τύπου κατηγορίας (nominal) ή τύπου διάταξης (ordinal), δεν επιδέχονται μαθηματικές πράξεις. Η μόνη δυνατότητα που έχουμε είναι να καταμετρήσουμε τις διάφορες τιμές της μεταβλητής – δηλ. να υπολογίσουμε τη **συχνότητα** εμφάνισης κάθε τιμής και να παρουσιάσουμε τα αποτελέσματά μας σε ένα **πίνακα απλής εισόδου** (one-way table). Για την εξαγωγή πινάκων διπλής εισόδου χρησιμοποιούμε την εντολή *Analyze* > *Descriptive Statistics* > *Crosstabs*, η οποία θα αναλυθεί στην ομώνυμη παράγραφο.

Για την εξαγωγή των πινάκων συχνοτήτων θα πρέπει να είναι επιλεγμένη η παράμετρος εμφάνισης των πινάκων συχνοτήτων *Display frequency tables*, αλλιώς στην ανάλυση επιτρέπεται η χρήση ποσοτικών μεταβλητών και η εξαγωγή περιγραφικών μέτρων σε αυτές.

Η πιο διαδεδομένη χρήση της συγκεκριμένης εντολής είναι για την περιγραφική ανάλυση ποιοτικών μεταβλητών, καθώς για τις ποσοτικές μεταβλητές μπορούμε να χρησιμοποιήσουμε και άλλες εντολές, όπως είναι η εντολή *Descriptives*, και η εντολή *Explore*.



Εικόνα 4.42: Το menu Frequencies

Στο πλαίσιο *Variable(s)* τοποθετούμε τις μεταβλητές για τις οποίες θα εξάγουμε πίνακες (κατανομές) συχνοτήτων. Επίσης από την υποενότητα *Charts* μπορούμε να δημιουργήσουμε κυκλικά διαγράμματα *Pie Chart* ή ραβδογράμματα *Bar Chart*. Περισσότερα για τα χαρακτηριστικά των διαγραμμάτων βλ. στην ομώνυμη παράγραφο. Στο σημείο *Chart Values* ορίζουμε εάν στο διάγραμμα θα εμφανίζονται οι απόλυτες *Frequencies* ή οι σχετικές συχνότητες *Percentages* της μεταβλητής.

Για παράδειγμα, στους πίνακες συχνοτήτων που προκύπτουν για τις δύο ποιοτικές μη διατάξιμες μεταβλητές «Ηλικιακή Ομάδα» και «Καπνιστική Συνήθεια» λαμβάνουμε:

Frequency Table

Age

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 40-44	4	11,1	11,1	11,1
45-59	4	11,1	11,1	22,2
50-54	4	11,1	11,1	33,3
55-59	4	11,1	11,1	44,4
60-64	4	11,1	11,1	55,6
65-69	4	11,1	11,1	66,7
70-74	4	11,1	11,1	77,8
75-79	4	11,1	11,1	88,9
80+	4	11,1	11,1	100,0
Total	36	100,0	100,0	

Smoking Status

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid cigarPipeOnly	9	25,0	25,0	25,0
cigaretteOnly	9	25,0	25,0	50,0
cigarettePlus	9	25,0	25,0	75,0
no	9	25,0	25,0	100,0
Total	36	100,0	100,0	

Εικόνα 4.43: Πίνακες συχνοτήτων για την ηλικιακή ομάδα και την καπνιστική συνήθεια

Στην πρώτη στήλη εμφανίζονται οι περιγραφές (ονομασίες) των τιμών της ποιοτικής μεταβλητής.

Στη δεύτερη στήλη εμφανίζεται η (απόλυτη) **Συχνότητα (Frequency)** των τιμών της κάθε τιμής, δηλ. το πλήθος εμφάνισης της κάθε παρατήρησης x_i .

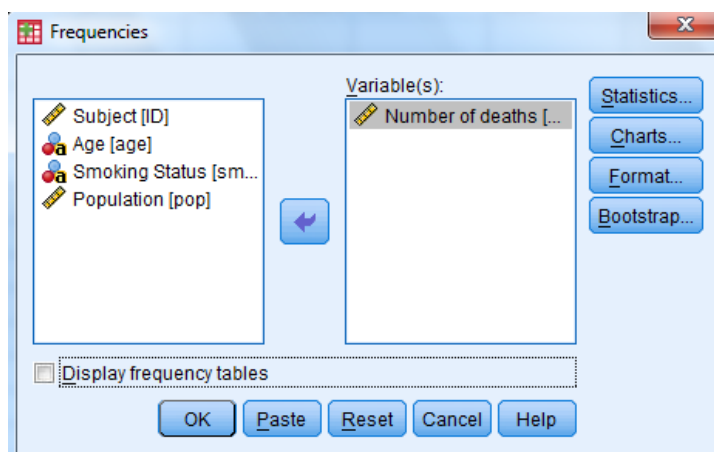
Στην τρίτη στήλη εμφανίζεται η **Σχετική Συχνότητα (Relative Frequency)** f_i της τιμής x_i , η οποία προκύπτει αν από τη συχνότητα διαιρέσουμε το μέγεθος n του δείγματος. Η σχετική συχνότητα πολλές φορές καταγράφεται ως ποσοστό **Percent** επί τοις εκατό ($f_i\%$).

Στην τέταρτη στήλη εμφανίζεται η σχετική συχνότητα αφού έχουν αφαιρεθεί από το σύνολο των παρατηρήσεων οι ελλείπουσες τιμές. Συνεπώς πλέον το μέγεθος n του δείγματος είναι μικρότερο και το εμφανιζόμενο ποσοστό αντιστοιχεί στην πραγματική σχετική συχνότητα με βάση το έγκυρο (valid) πλήθος των παρατηρήσεων. Εύλογα διαπιστώνει κανείς ότι όταν δεν υπάρχουν ελλείπουσες τιμές οι δύο στήλες μας παρέχουν την ίδια πληροφορία.

Στην πέμπτη στήλη εμφανίζεται η **Αθροιστική Συχνότητα (cumulative percent)**, δηλ. το πλήθος των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i εκφρασμένο ως ποσοστό. Για παράδειγμα, η αθροιστική συχνότητα στη μεταβλητή «Ηλικιακή Ομάδα» συνεπάγεται ότι το 11,1% των ατόμων είναι ηλικίας έως 44 ετών, το 22,2% ηλικίας έως 59 ετών, ..., το 77,8% ηλικίας έως 74 ετών, κ.ο.κ.

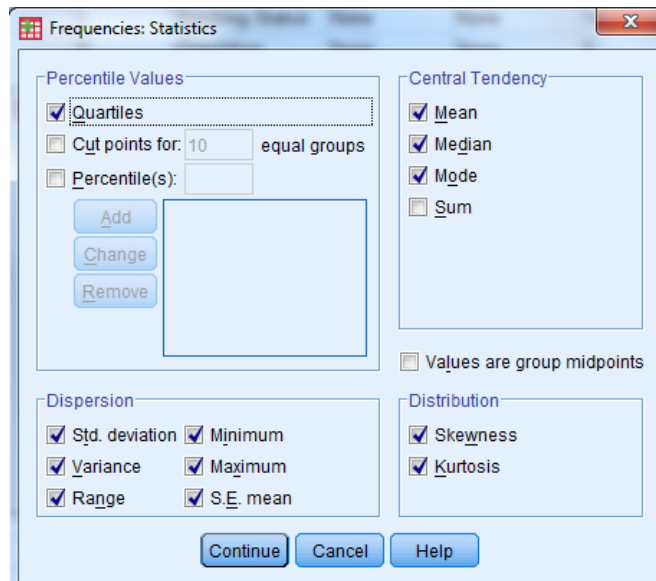
Συνεχείς μεταβλητές

Προαναφέραμε ότι μπορούμε να χρησιμοποιήσουμε την ίδια εντολή *Analyze > Descriptive Statistics > Frequencies*, με σκοπό την ανάλυση ποσοτικών μεταβλητών. Αυτό επιτυγχάνεται μόνο εάν ΔΕΝ είναι επιλεγμένη η παράμετρος εμφάνισης των πινάκων συχνοτήτων *Display frequency tables*.



Εικόνα 4.44: Το menu *Frequencies* με συνεχή μεταβλητή

Σ' αυτή την περίπτωση μέσω της υποενότητας *Statistics* επιλέγουμε τους περιγραφικούς δείκτες, ενώ από την υποενότητα *Charts* επιλέγουμε τη δημιουργία Ιστογράμματος *Histograms* με ή χωρίς την εμφάνιση της καμπύλης της κανονικής κατανομής *With normal curve* για οπτικό έλεγχο της κανονικότητας των δεδομένων της μεταβλητής.



Εικόνα 4.45: Το menu Frequencies - Statistics

Αναλυτικά από την υποενότητα *Statistics* παρέχονται τα ακόλουθα περιγραφικά μέτρα:

Percentile Values

Αποτελείται από μέτρα με σκοπό τον εντοπισμό της θέσης της κατανομής κατά μήκος του άξονα των τιμών της μεταβλητής. Αυτά ονομάζονται μέτρα θέσης (**location measures**) και τα πιο γνωστά είναι τα τεταρτημόρια και τα εκατοστημόρια.

Quartiles - Τεταρτημόρια

- Τα Τεταρτημόρια υποδιαιρούν το σύνολο των διατεταγμένων κατά αύξουσα σειρά παρατηρήσεων σε τέσσερις ισοπληθείς ομάδες, κάθε μια από τις οποίες περιέχει το 25% των παρατηρήσεων. Συγκεκριμένα:
- Το πρώτο τεταρτημόριο (first quartile), που συμβολίζεται με Q_1 , είναι η τιμή της παρατήρησης μέχρι την οποία περιλαμβάνεται το 25% των διατεταγμένων παρατηρήσεων του δείγματος.
- Το δεύτερο τεταρτημόριο (second quartile), που συμβολίζεται με Q_2 , είναι η τιμή της παρατήρησης μέχρι την οποία περιλαμβάνεται το 50% των διατεταγμένων παρατηρήσεων του δείγματος. Δηλαδή, το δεύτερο τεταρτημόριο των παρατηρήσεων είναι η Διάμεσος.
- Το τρίτο τεταρτημόριο (third quartile), που συμβολίζεται με Q_3 , είναι η τιμή της παρατήρησης μέχρι την οποία περιλαμβάνεται το 75% των διατεταγμένων παρατηρήσεων του δείγματος.

Η **θέση** των τεταρτημορίων υπολογίζεται από τους ακόλουθους μαθηματικούς τύπους, όπου n το μέγεθος του δείγματος.

$$Q_1 = \frac{n+1}{4}$$

Εξίσωση 4.1

$$Q_2 = \frac{2(n+1)}{4}$$

Εξίσωση 4.2

$$Q_3 = \frac{3(n+1)}{4}$$

Εξίσωση 4.3

Percentiles - Ποσοστημόρια

Τα εκατοστημόρια είναι τα σημεία που υποδιαιρούν το σύνολο των διατεταγμένων κατά αύξουσα σειρά παρατηρήσεων σε 100 ισοπληθείς ομάδες. Έτσι έχουμε ότι το 25^ο εκατοστημόριο θα έχει από τα αριστερά του το 25% των παρατηρήσεων και από δεξιά του το υπόλοιπο 75%, το 50^ο εκατοστημόριο θα διαιρέσει τις παρατηρήσεις στη μέση άρα ισούται με τη Διάμεσο, κ.ο.κ. Εάν τώρα γενικεύσουμε αυτό το ποσοστό τότε το k- οστό εκατοστημόριο θα έχει από τα αριστερά του το k% των παρατηρήσεων και από τα δεξιά του το (100 - k)%. Η εφαρμογή, εκτός από τα τεταρτημόρια, παρέχει στον αναλυτή τη δυνατότητα να ορίσει εκείνος το σημείο τομής των παρατηρήσεων σε οποιονδήποτε αριθμό ομάδων ίδιου μεγέθους μέσω του πεδίου *Cut points for n equal groups*. Προφανώς για n = 4 λαμβάνουμε τα τεταρτημόρια.

Τέλος, από την επιλογή *Percentile(s)* μπορούν να υπολογιστούν συγκεκριμένα εκατοστημόρια (π.χ το 5^ο, 10^ο κ.ο.κ.) επιλέγοντας στο αντίστοιχο πεδίο το ποσοστό που χαρακτηρίζει το εκατοστημόριο και καταχωρώντας το με το πλήκτρο *Add*.

Central Tendency

Τα δεδομένα εμφανίζουν μια τάση να περιστρέφονται γύρω από μια κεντρική τιμή που εκφράζει την τυπική ή μέση τιμή τους. Αυτά ονομάζονται μέτρα κεντρικής τάσης (*central tendency measures*) και τα πιο γνωστά είναι ο αριθμητικός, ο αρμονικός και ο γεωμετρικός μέσος όρος. Από την εφαρμογή SPSS παρέχεται ο αριθμητικός μέσος.

Mean - Αριθμητικός μέσος ή Μέση τιμή

Η μέση τιμή ενός συνόλου n παρατηρήσεων ορίζεται ως το άθροισμα των παρατηρήσεων διά του πλήθους των παρατηρήσεων:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Εξίσωση 4.4

Στο σημείο αυτό παρέχονται και μέτρα θέσης όπως:

Median - Διάμεσος

Διάμεσος ενός δείγματος n παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται ως η μεσαία παρατήρηση, όταν το n είναι περιττός αριθμός, ή ο μέσος όρος (ημιάθροισμα) των δύο μεσαίων παρατηρήσεων όταν το n είναι άρτιος αριθμός.

Mode - Επικρατούσα τιμή

Είναι η παρατήρηση με τη μεγαλύτερη συχνότητα.

Sum - Άθροισμα

Το άθροισμα όλων των παρατηρήσεων.

Dispersion

Αποτελείται από μέτρα ποσοτικής αξιολόγησης της συγκέντρωσης των τιμών μιας μεταβλητής γύρω από την κεντρική τιμή τους. Αυτά ονομάζονται μέτρα διασποράς (**measures of variability**) και τα πιο γνωστά είναι η διακύμανση, η τυπική απόκλιση, και το εύρος).

Std. Deviation - Τυπική απόκλιση

Τυπική απόκλιση ορίζεται ως η τετραγωνική ρίζα της διακύμανσης.

$$s = \sqrt{S^2}$$

Εξίσωση 4.5

Variance - Διακύμανση

Διακύμανση ορίζεται ως ο αριθμητικός μέσος των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής από τον αριθμητικό μέσο

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Εξίσωση 4.6

Range - Έυρος

Εύρος ορίζεται ως η διαφορά της ελάχιστης *Minimum* παρατήρησης από τη μέγιστη *Maximum* παρατήρηση, δηλαδή:

$$R = X_{\max} - X_{\min}$$

Εξίσωση 4.7

S.E. Mean - Τυπικό σφάλμα

Ορίζεται ως η τυπική απόκλιση του μέσου όρου ενός δείγματος.

$$s.e = \frac{s}{\sqrt{n}}$$

Εξίσωση 4.8

Distribution

Αποτελείται από μέτρα που περιγράφουν την μορφή της κατανομής συχνότητας, δηλ. μέτρα λοξότητας (measures of skewness) και μέτρα κύρτωσης (kurtosis measures). Η ασυμμετρία αναφέρεται στην απόκλιση του διαγράμματος της κατανομής συχνοτήτων από το διάγραμμα της κανονικής κατανομής, ενώ η κυρτότητα αναφέρεται στο πόσο πεπλατυσμένο είναι το διάγραμμα την κατανομής συχνοτήτων.

Skewness - Συντελεστής Ασυμμετρίας

Δίνεται από τη σχέση:

$$\beta_1 = \frac{\sum_{i=1}^n z_i^3}{n}$$

Εξίσωση 4.9

Όπου:

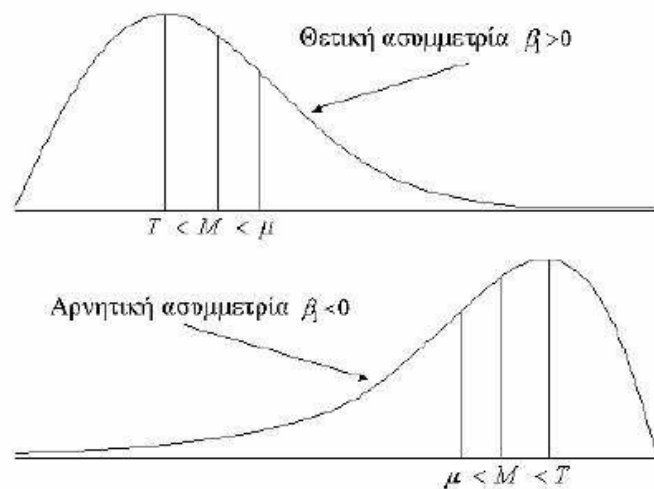
$$z_i = \frac{x_i - \bar{x}}{s}$$

Εξίσωση 4.10

Η οποία καλείται τυποποιημένη μορφή των παρατηρήσεων.

Για:

- $\beta_1 = 0$, έχουμε συμμετρία,
- $\beta_1 < 0$, έχουμε αρνητική ή αριστερή ασυμμετρία,
- $\beta_1 > 0$, έχουμε θετική ή δεξιά ασυμμετρία.



Εικόνα 4.46: Μορφή καμπύλης συχνότητας με Θετική και Αρνητική ασυμμετρία

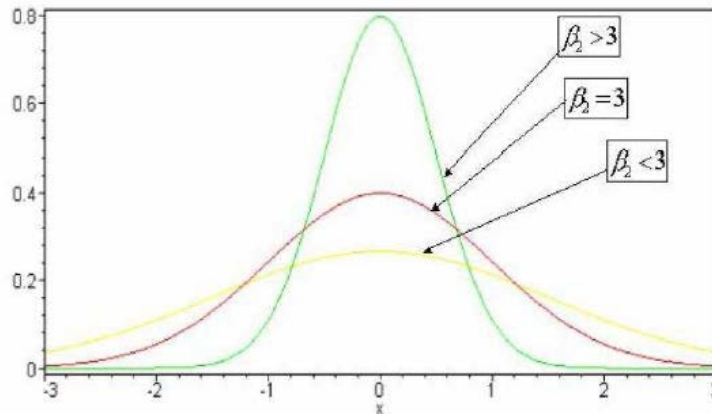
Kurtosis - Συντελεστής Κύρτωσης

Χαρακτηρίζει το ύψος της κορυφής της κατανομής και δίνει πληροφορίες για την αιχμηρότητα της καμπύλης. Οι κατανομές διακρίνονται σε λεπτόκυρτες, μεσόκυρτες και πλατύκυρτες.

$$\beta_2 = \frac{\sum_{i=1}^n z_i^4}{n}$$

Εξίσωση 4.11

- $\beta_2 = 3$, έχουμε μεσόκυρτη κατανομή,
- $\beta_2 < 3$, έχουμε πλατύκυρτη κατανομή,
- $\beta_2 > 3$, έχουμε λεπτόκυρτη κατανομή.



Εικόνα 4.47: Μορφή καμπύλης συχνότητας για διάφορες τιμές του συντελεστή κύρτωσης

Στα αποτελέσματα μαζί με τις τιμές των συντελεστών ασυμμετρίας και κύρτωσης, εμφανίζονται και τα τυπικά τους σφάλματα.

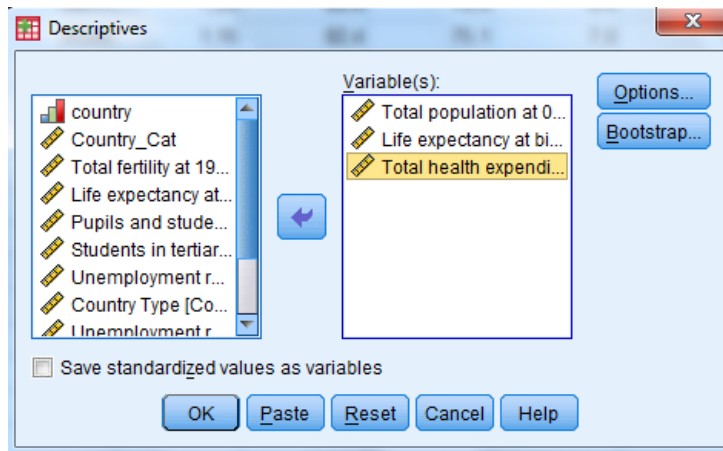
Η παράμετρος *Values are group midpoints* επιλέγεται όταν οι τιμές της μεταβλητής είναι τα κεντρικά σημεία ομαδοποιημένων δεδομένων (για παράδειγμα στην μεταβλητή Age στον πίνακα συχνοτήτων στην Εικόνα 4.2, για την ηλικιακή ομάδα 40-44 στο αρχείο των δεδομένων θα είχαμε καταγράψει μόνο την τιμή 42,5 (κεντρική τιμή του διαστήματος), κ.ο.κ.]. Τότε, ενεργοποιώντας την παράμετρο η διάμεσος και τα εκατοστημόρια εκτιμώνται με τη μέθοδο της παρεμβολής (interpolation).

Statistics		
Number of deaths		
N	Valid	36
	Missing	0
Mean		253,61
Median		171,50
Mode		120
Std. Deviation		262,597
Variance		68957,444
Skewness		1,489
Std. Error of Skewness		,393
Kurtosis		1,435
Std. Error of Kurtosis		,768
Range		999
Minimum		2
Maximum		1001
Percentiles	25	75,50
	50	171,50
	75	316,00

Εικόνα 4.48: Αποτελέσματα από την εντολή *Frequencies*

4.3 Descriptives

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Descriptive Statistics > Descriptives*, με σκοπό την εξαγωγή περιγραφικών μέτρων σε ποσοτικές μεταβλητές. Η εντολή είναι ένα υποσύνολο των δυνατών αποτελεσμάτων της εντολής *Explore* η οποία αναλύεται στην επόμενη ενότητα.



Εικόνα 4.49: Το menu Descriptives

Οι ποσοτικές μεταβλητές για τις οποίες θα λάβουμε τα περιγραφικά μέτρα εισάγονται στο πλαίσιο *Variable(s)*. Με την επιλογή της παραμέτρου *Save standardized values as variables* δημιουργούνται και αποθηκεύονται ως μεταβλητές στο αρχείο δεδομένων, οι τυποποιημένες τιμές των παρατηρήσεων των μεταβλητών, οι οποίες προκύπτουν από τη σχέση:

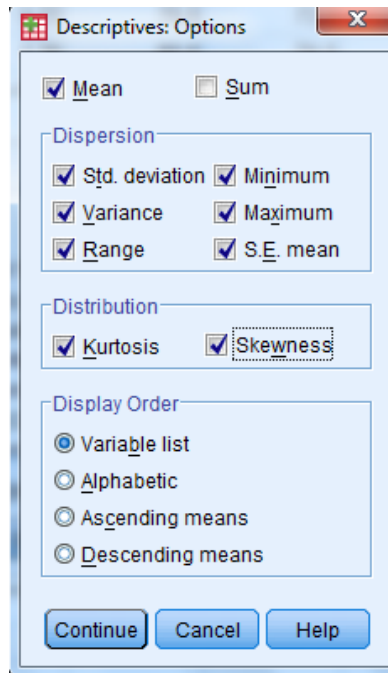
$$z_i = \frac{x_i - \bar{x}}{s}$$

Εξίσωση 4.12

Οι τυποποιημένες τιμές μιας μεταβλητής που ακολουθεί κανονική κατανομή, έχουν την ιδιότητα να ακολουθούν την κανονική κατανομή με μηδενική μέση τιμή και μοναδιαία διασπορά. Η κατανομή αυτή συμβολίζεται ως $N(0,1)$ και ονομάζεται τυπική κανονική κατανομή. Επίσης, οι τιμές αυτές είναι καθαροί αριθμοί, καθώς έχει γίνει εξάλειψη των μονάδων μέτρησης των παρατηρήσεων, και μπορούν να χρησιμοποιηθούν για τη σύγκριση μεταβλητών διαφορετικών μονάδων μέτρησης.

Από το πλήκτρο *Options* προσφέρεται η επιλογή των περιγραφικών μέτρων τα οποία έχουν ήδη παρουσιαστεί στην προηγούμενη ενότητα, όπως ο μέσος, η διασπορά, η τυπική απόκλιση, το εύρος και οι συντελεστές ασυμμετρίας και κύρτωσης.

Επιπρόσθετα, ο αναλυτής από την παρούσα ενότητα έχει τη δυνατότητα να διαμορφώσει την ταξινόμηση των μεταβλητών *Display Order* στην εμφάνιση των αποτελεσμάτων είτε με βάση τη λίστα των μεταβλητών σύμφωνα με τη σειρά που έχουν προστεθεί στο πλαίσιο *Variable(s)*, είτε αλφαβητικά *Alphabetic*, είτε κατά αύξοντα αριθμό μέσης τιμής *Ascending Means*, είτε κατά φθίνοντα αριθμό μέσης τιμής *Descending Means*.



Εικόνα 4.50: Το menu Descriptives - Options

Η διαφορά από την εντολή *Frequencies* είναι ότι η συγκεκριμένη εντολή λειτουργεί αποκλειστικά για ποσοτικές μεταβλητές, και τα αποτελέσματά της εμφανίζονται σε συγκεντρωτικό πίνακα που επιτρέπει πιο εύκολα τη σύγκριση των αποτελεσμάτων μεταξύ των ποσοτικών μεταβλητών.

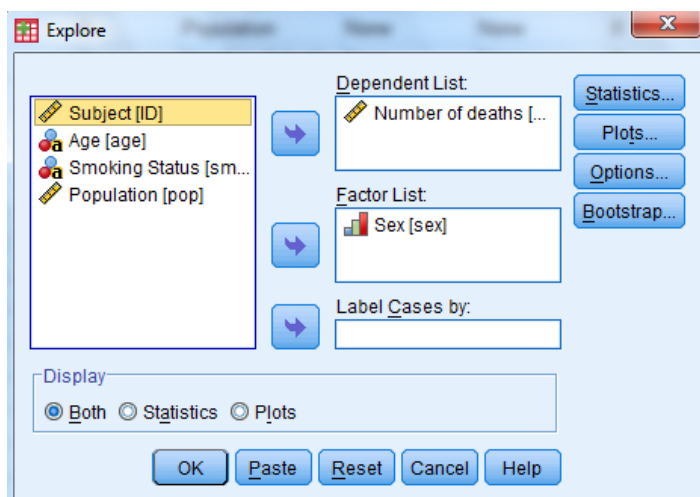
	Descriptive Statistics										
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Total population at 01/01/1998:1000s	15	81633	424	82057	24372,20	27121,193	7,356E8	1,035	,580	-,423	1,121
Life expectancy at birth : girls	15	3,6	78,8	82,4	80,633	1,1629	1,352	-,059	,580	-,733	1,121
Total health expenditure as a % of GDP	15	4,3	6,0	10,3	7,933	1,1185	1,251	,310	,580	,148	1,121
Valid N (listwise)	15										

Εικόνα 4.51: Αποτελέσματα από την εντολή Descriptives

4.4 Explore

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Descriptive Statistics > Explore*, με σκοπό την εξαγωγή περιγραφικών μέτρων σε ποσοτικές μεταβλητές. Η συγκεκριμένη εντολή διαφέρει κατά πολύ από την εντολή *Descriptives*, διότι παρέχει πολλές επιπρόσθετες δυνατότητες, όπως η εξαγωγή διαστημάτων εμπιστοσύνης, η ανίχνευση ακραίων τιμών, ο έλεγχος κανονικότητας των τιμών της μεταβλητής μέσω των στατιστικών Shapiro-Wilks και Kolmogorov-Smirnov, η εξαγωγή περιγραφικών μέτρων σε υποομάδες ποιοτικών μεταβλητών (Factors), και η δημιουργία διαγραμμάτων Box - Plot, Ιστογραμμάτων και Stem & Leaf plots.

Αναλυτικότερα:

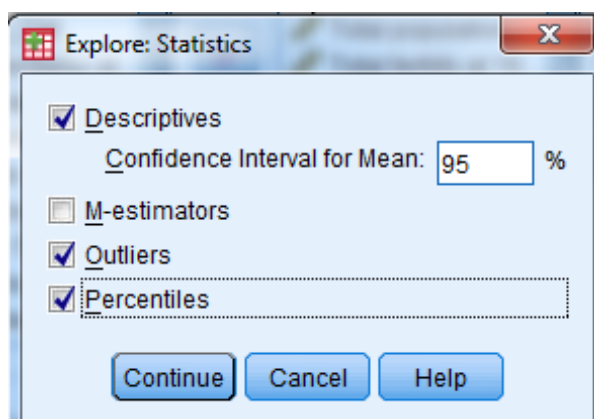


Εικόνα 4.52: Το menu Explore

Στο πλαίσιο *Dependent List* ορίζονται οι ποσοτικές μεταβλητές, ενώ στο πλαίσιο *Factor List* δηλώνουμε την (ή τις) ποιοτική μεταβλητή που θα χωρίσει το δείγμα σε υποομάδες με σκοπό την εξαγωγή των περιγραφικών μέτρων σ' αυτή. Δηλαδή, για κάθε υποομάδα θα λάβουμε ξεχωριστά περιγραφικά μέτρα.

Στο σημείο *Display* επιλέγουμε εάν στο Output θα εμφανιστούν μόνο τα περιγραφικά μέτρα *Statistics*, μόνο τα διαγράμματα *Plots* ή και τα δύο *Both*.

Από το πλήκτρο *Statistics* επιλέγουμε την εξαγωγή των βασικών περιγραφικών μέτρων *Descriptives* με διαστήματα εμπιστοσύνης *Confidence Interval for Mean* ορίζοντας το επιθυμητό ποσοστό.



Εικόνα 4.53: Το menu Explore - Statistics

Τα περιγραφικά μέτρα που εξάγονται από την επιλογή αυτή είναι:

- η μέση τιμή και το 95% δ.ε. της μέσης τιμής,
- η «Ξακρισμένη» μέση τιμή,
- η διάμεσος,
- η τυπική απόκλιση,
- η μέγιστη και η ελάχιστη παρατήρηση,
- το ενδοτεταρτημοριακό εύρος που ορίζεται ως η διαφορά του πρώτου τεταρτημορίου από το τρίτο τεταρτημόριο,
- οι συντελεστές κύρτωσης και ασυμμετρίας.

Trimmed Mean - Ξακρισμένη μέση τιμή

Ορίζεται ως η μέση τιμή που προκύπτει εάν από το σύνολο των δεδομένων αφαιρεθούν το 5% των ελάχιστων και το 5% των μέγιστων παρατηρήσεων. Η μέση τιμή που προκύπτει από το υπόλοιπο 90% του δείγματος καλείται «ξακρισμένη» (trimmed) μέση τιμή, και εξυπηρετεί την ανάγκη εξάλειψης πιθανόν αλλοιώσεων από την ύπαρξη ακραίων τιμών (outliers). Ως ακραία τιμή ορίζεται μια παρατήρηση πολύ μικρότερη ή πολύ μεγαλύτερη από τις υπόλοιπες.

Confidence Interval for Mean - Διάστημα Εμπιστοσύνης για τον μέσο

Το $(1-a)\%$ δ.ε. της μέσης τιμής προκύπτει από τη σχέση:

$$\left(\bar{x} - t_{n-1,a} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,a} \frac{s}{\sqrt{n}}\right)$$

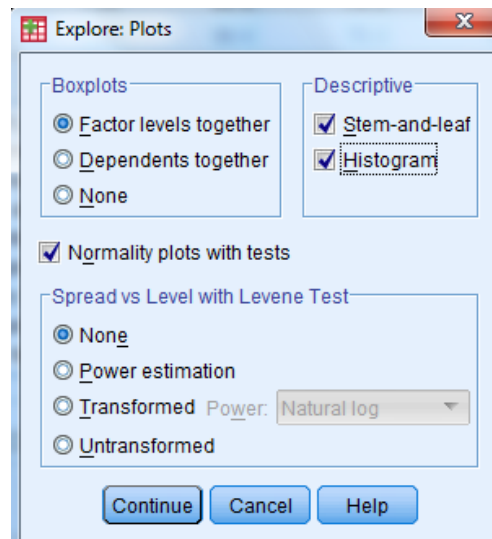
Εξίσωση 4.13

Για $a = 5\%$ λαμβάνουμε το 95% διάστημα εμπιστοσύνης της μέσης τιμής, το οποίο εκφράζει την πιθανότητα η εκτιμώμενη παράμετρος (μέση τιμή) να περιέχεται εντός του διαστήματος.

Επιπρόσθετα, από αυτή την ενότητα, μπορούμε να λάβουμε τις 5 μικρότερες και 5 μεγαλύτερες (ακραίες) τιμές *Outliers*, καθώς και τα ποσοστημόρια *Percentiles*, τα οποία έχουμε ήδη αναλύσει στην προηγούμενη παράγραφο.

Από το πλήκτρο *Plots* μπορούμε να επιλέξουμε τη δημιουργία θηκογραμμάτων για έλεγχο ακραίων τιμών *Box - Plot*, περιγραφικών όπως τα *Stem - and - leaf* και τα Ιστογράμματα *Histogram* και *Q - Q Plots* ελέγχου κανονικότητας *Normality plots with tests* (Περισσότερα για τα χαρακτηριστικά των διαγραμμάτων, βλ. στην ομώνυμη παράγραφο).

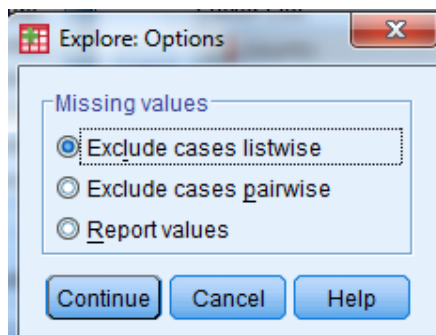
Επίσης με την επιλογή της παραμέτρου *Normality plots with tests* λαμβάνουμε και τα αποτελέσματα των ελέγχων κανονικότητας *Shapiro-Wilks* και *Kolmogorov-Smirnov*, οι οποίοι θα αναλυθούν σε επόμενο κεφάλαιο.



Εικόνα 4.54: Το menu Explore - Plots

Από το πλήκτρο *Options* γίνεται η διαχείριση των ελλειπουσών τιμών, όπου:

- *Exclude cases listwise*, οι παρατηρήσεις που έχουν την ελλείπουσα τιμή είτε σε κάποια ποσοτική μεταβλητή (πλαίσιο *dependents*), είτε σε κάποιον παράγοντα (πλαίσιο *Factors*) και εξαιρούνται από τις αναλύσεις. Η συγκεκριμένη επιλογή είναι προκαθορισμένη και συχνά δεν γίνεται αντιληπτό ότι αν π.χ. έχουμε δηλώσει τρεις ποσοτικές μεταβλητές και κάποια εξ αυτών έχει ελλείπουσες τιμές, τότε ολόκληρη η σειρά θα εξαιρεθεί από την ανάλυση και τα περιγραφικά μέτρα θα προκύψουν στο ίδιο μέγεθος δείγματος και για τις τρεις μεταβλητές.
- *Exclude cases pairwise*, αν μια ποσοτική μεταβλητή έχει ελλείπουσα τιμή σε όλες τις ομάδες της ποιοτικής σπότε εξαιρείται από την ανάλυση, αλλιώς όχι.
- *Report values*, οι παρατηρήσεις με ελλείπουσες τιμές αναλύονται ως ξεχωριστή κατηγορία.



Εικόνα 4.55: Το menu *Explore - Options*

Αποτελέσματα από την εντολή *Explore*.

			Percentiles					
Sex			Percentiles					
			5	10	25	50	75	90
Weighted Average (Definition 1)	Number of deaths	Male	4,00	18,00	55,00	149,00	432,00	901,00
		Female	2,00	2,80	120,00	189,00	295,00	616,40
Tukey's Hinges	Number of deaths	Male			59,00	149,00	312,50	
		Female			120,00	189,00	253,00	

Εικόνα 4.56: Αποτελέσματα - Ποσοστημόρια - από την εντολή *Explore*.

			Tests of Normality					
Sex			Kolmogorov-Smirnov ^a			Shapiro-Wilk		
			Statistic	df	Sig.	Statistic	df	Sig.
Number of deaths	Male		,331	19	,000	,767	19	,000
	Female		,239	17	,011	,864	17	,018

a. Lilliefors Significance Correction

Εικόνα 4.57: Αποτελέσματα - Έλεγχοι Κανονικότητας - από την εντολή *Explore*

Descriptives				Statistic	Std. Error
Sex					
Number of deaths	Male	Mean		266,05	70,687
		95% Confidence Interval for Mean	Lower Bound	117,54	
			Upper Bound	414,56	
		5% Trimmed Mean		239,78	
		Median		149,00	
		Variance		94937,386	
		Std. Deviation		308,119	
		Minimum		4	
		Maximum		1001	
		Range		997	
	Interquartile Range		377		
	Skewness		1,450	,524	
	Kurtosis		,922	1,014	
	Female	Mean		239,71	50,672
		95% Confidence Interval for Mean	Lower Bound	132,29	
			Upper Bound	347,13	
5% Trimmed Mean			223,01		
Median			189,00		
Variance			43650,596		
Std. Deviation			208,927		
Minimum			2		
Maximum			778		
Range			776		

Εικόνα 4.58: Αποτελέσματα - Περιγραφικά Μέτρα - από την εντολή Explore

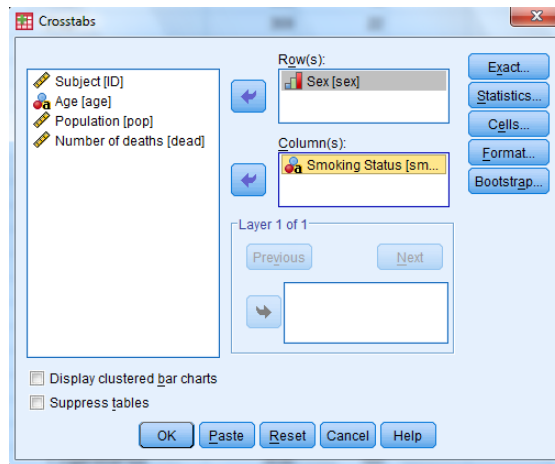
Extreme Values					
Sex				Case Number	Value
Number of deaths	Male	Highest	1	23	1001
			2	24	901
			3	33	689
			4	25	613
			5	34	432
		Lowest	1	11	4
			2	1	18
			3	3	19
			4	13	38
			5	4	55
	Female	Highest	1	32	778
			2	22	576
			3	31	514
			4	26	337
			5	18	253
Lowest		1	10	2	
		2	12	3	
		3	2	22	
		4	9	120	
		5	8	120	

Εικόνα 4.59: Αποτελέσματα - Ακραίες Παρατηρήσεις - από την εντολή Explore

4.5 Crosstabs

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Descriptive Statistics > Crosstabs*, με σκοπό τη δημιουργία δισδιάστατων περιγραφικών πινάκων συνάφειας μεταξύ ποιοτικών ή ποσοτικών διακριτών μεταβλητών.

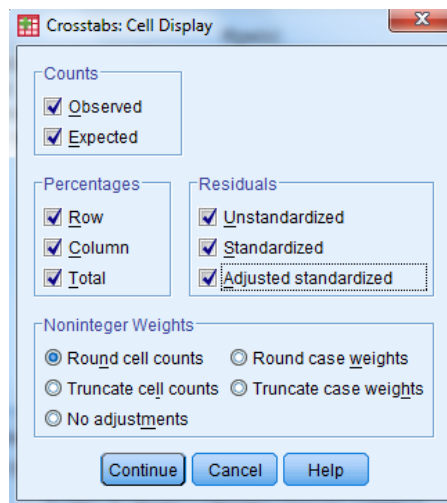
Επιπρόσθετα, η συγκεκριμένη εντολή χρησιμοποιείται για τη σύγκριση και τον έλεγχο συσχέτισης μεταξύ των ποιοτικών ή ποσοτικών διακριτών μεταβλητών, σημεία τα οποία θα αναλυθούν σε επόμενο κεφάλαιο στα πλαίσια των ελέγχων υποθέσεων. Στο παρόν κεφάλαιο θα αναλυθούν τα σημεία εκείνα τα οποία αφορούν στην εξαγωγή περιγραφικών στοιχείων.



Εικόνα 4.60: To menu Crosstabs

Επιλέγοντας την παράμετρο *Display clustered bar charts* εμφανίζονται ομαδοποιημένα ραβδογράμματα για τη γραφική απεικόνιση των συχνοτήτων των τιμών των μεταβλητών, ενώ με την παράμετρο *Suppress tables* δεν θα εμφανιστούν οι πίνακες συχνοτήτων. Προφανώς για τους σκοπούς του παρόντος κεφαλαίου δεν θα επιλέξουμε ποτέ αυτή την παράμετρο!

Στα πλαίσια *Row(s)* και *Column(s)* ορίζουμε τις μεταβλητές, οι τιμές των οποίων θα εμφανίζονται στις γραμμές και στις στήλες του πίνακα, αντίστοιχα. Για να είναι ο πίνακας πιο εμφανίσιμος συνηθίζουμε να ορίζουμε τη μεταβλητή με τις περισσότερες τιμές στη θέση των γραμμών.



Εικόνα 4.61: To menu Crosstabs - Cell Display

Από το πλήκτρο *Cells* επιλέγουμε την προβολή πρόσθετων στοιχείων σε κάθε κελί του πίνακα π.χ. σχετικών συχνοτήτων, ή την εμφάνιση υπολοίπων. Συγκεκριμένα:

Count

- *Observed*: Επιλέγεται ώστε σε κάθε κελί του πίνακα να εμφανίζονται οι παρατηρούμενες συχνότητες.
- *Expected*: Επιλέγεται ώστε σε κάθε κελί του πίνακα να εμφανίζονται οι αναμενόμενες συχνότητες.

Percentages

- *Row*: Επιλέγεται ώστε να εμφανίζονται οι σχετικές συχνότητες % επί των γραμμών.
- *Column*: Επιλέγεται ώστε να εμφανίζονται οι σχετικές συχνότητες % επί των στηλών.
- *Total*: Επιλέγεται ώστε να εμφανίζονται οι σχετικές συχνότητες % επί του συνόλου των παρατηρήσεων.

Residuals

- *Unstandardized*: Επιλέγεται για να εμφανιστούν τα μη τυποποιημένα υπόλοιπα από τη διαφορά παρατηρούμενων και αναμενόμενων συχνοτήτων.
- *Standardized*: Επιλέγεται για να εμφανιστούν τα τυποποιημένα υπόλοιπα από τη διαφορά παρατηρούμενων και αναμενόμενων συχνοτήτων.
- *Adjusted standardized*: Επιλέγεται για να εμφανιστούν τα αναπροσαρμοσμένα υπόλοιπα. Υπό τη μηδενική υπόθεση ότι οι 2 μεταβλητές είναι ανεξάρτητες, τα αναπροσαρμοσμένα υπόλοιπα θα ακολουθούν την τυπική κανονική κατανομή, δηλαδή έχουν μέση τιμή 0 και τυπική απόκλιση 1. Έτσι ένα προσαρμοσμένο υπόλειμμα που είναι πάνω από 1,96 (κατά συνθήκη χρησιμοποιείται 2,0) δείχνει ότι ο αριθμός των παρατηρήσεων στο εν λόγω κελί είναι σημαντικά μεγαλύτερος από ό,τι θα αναμενόταν εάν η μηδενική υπόθεση ήταν αληθής, με επίπεδο σημαντικότητας 0,05. Ένα αναπροσαρμοσμένο υπόλοιπο μικρότερο από -2,0 υποδεικνύει ότι ο αριθμός των παρατηρήσεων στο εν λόγω κελί είναι σημαντικά μικρότερος απ' ό,τι θα αναμενόταν εάν η μηδενική υπόθεση ήταν αληθής. Έτσι ανάλογα με το πρόσημο του αναπροσαρμοσμένου υπολοίπου, διαπιστώνουμε εάν σε κάθε κελί υπάρχουν λιγότερες ή περισσότερες παρατηρήσεις απ' ό,τι θα περίμενε κανείς αν οι 2 μεταβλητές ήταν ανεξάρτητες. Εάν ο πίνακας συνάφειας είναι 2x2, τότε όλα τα προσαρμοσμένα κατάλοιπα θα έχουν την ίδια απόλυτη τιμή, με ακριβώς 2 από αυτά να είναι αρνητικά.
- *Noninteger Weights*: Χρησιμοποιείται για να περικόψουμε ή να στρογγυλοποιήσουμε τις δεκαδικές τιμές από τις τιμές των παρατηρήσεων σε κάθε κελί. Γενικά, οι τιμές σε κάθε κελί είναι ακέραιες τιμές, δεδομένου ότι αντιπροσωπεύουν το πλήθος των περιπτώσεων σε κάθε συνδυασμό τιμών. Αν όμως στο αρχείο μας οι τιμές αυτές είναι σταθμισμένες σε σχέση με μια άλλη μεταβλητή βαρύτητας με δεκαδικές τιμές (για παράδειγμα 1,25), τότε οι παρατηρήσεις σε κάθε κελί μπορεί επίσης να είναι δεκαδικές.

Από το πλήκτρο *Statistics* γίνεται η επιλογή των ελέγχων υποθέσεων και δεικτών συνάφειας μεταξύ των ποιοτικών ή ποσοτικών διακριτών μεταβλητών. Οι συγκεκριμένες δυνατότητες θα αναλυθούν στο επόμενο κεφάλαιο.

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

Sex * Smoking Status Crosstabulation

			Smoking Status				Total
			cigarPipeOnly	cigaretteOnly	cigarettePlus	no	
Sex	Male	Count	3	5	6	5	19
		Expected Count	4,8	4,8	4,8	4,8	19,0
		% within Sex	15,8%	26,3%	31,6%	26,3%	100,0%
		% within Smoking Status	33,3%	55,6%	66,7%	55,6%	52,8%
		% of Total	8,3%	13,9%	16,7%	13,9%	52,8%
		Residual	-1,8	,3	1,3	,3	
		Std. Residual	-,8	,1	,6	,1	
	Adjusted Residual	-1,3	,2	1,0	,2		
Sex	Female	Count	6	4	3	4	17
		Expected Count	4,3	4,3	4,3	4,3	17,0
		% within Sex	35,3%	23,5%	17,6%	23,5%	100,0%
		% within Smoking Status	66,7%	44,4%	33,3%	44,4%	47,2%
		% of Total	16,7%	11,1%	8,3%	11,1%	47,2%
		Residual	1,8	-,3	-1,3	-,3	
		Std. Residual	,8	-,1	-,6	-,1	
	Adjusted Residual	1,3	-,2	-1,0	-,2		
Total	Count	9	9	9	9	36	
	Expected Count	9,0	9,0	9,0	9,0	36,0	
	% within Sex	25,0%	25,0%	25,0%	25,0%	100,0%	
	% within Smoking Status	100,0%	100,0%	100,0%	100,0%	100,0%	
	% of Total	25,0%	25,0%	25,0%	25,0%	100,0%	

Εικόνα 4.62: Αποτελέσματα από την επιλογή Crosstabs

Από τον πίνακα ελέγχουμε οπτικά και συγκρίνουμε τις σχετικές συχνότητες επί των γραμμών ή των στηλών, ή για το σύνολο του δείγματος. Είμαστε επίσης ιδιαίτερα προσεκτικοί ως προς τα ποσοστά που συγκρίνουμε καθώς οι σχετικές συχνότητες επί των γραμμών αθροίζουν στο 100% οριζοντίως, οι σχετικές συχνότητες επί των στηλών αθροίζουν στο 100% καθέτως, ενώ οι σχετικές συχνότητες επί του συνολικού δείγματος αθροίζουν στο 100% μόνο στο τελευταίο κελί.

4.6 Correlate

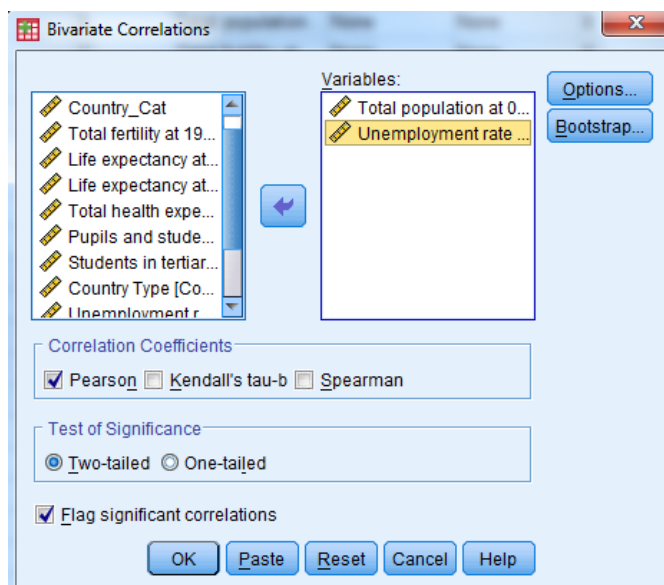
Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Correlate > Bivariate*, με σκοπό τον έλεγχο ύπαρξης γραμμικής σχέσης μεταξύ συνεχών μεταβλητών. Ο έλεγχος γίνεται με τον υπολογισμό των συντελεστών συνάφειας (r) των Pearson (παραμετρικό) και Spearman (μη παραμετρικό).

Οι συντελεστές συσχέτισης είναι καθαροί αριθμοί (χωρίς μονάδες), που μας δείχνουν το μέγεθος και την κατεύθυνση της γραμμικής συσχέτισης μεταξύ δύο ποσοτικών μεταβλητών X και Y , και λαμβάνουν τιμές στο διάστημα $-1, 1$, όπου:

- $p = 1 \Rightarrow$ Ισχυρή θετική συσχέτιση,
- $p = -1 \Rightarrow$ Ισχυρή αρνητική συσχέτιση,
- $p = 0 \Rightarrow$ Απουσία συσχέτισης.

Γραφικά η διερεύνηση της κατεύθυνσης της συσχέτισης (θετική, αρνητική ή απουσία συσχέτισης) πραγματοποιείται από τα διαγράμματα *Scatter plots*, βλ. ομώνυμη παράγραφο.

Στο πλαίσιο *Variables* ορίζουμε τις μεταβλητές για τις οποίες θα υπολογιστεί ο πίνακας των (ανά δύο) συντελεστών συσχέτισης.



Εικόνα 4.63: Το menu Bivariate Correlations

Correlation Coefficients

- *r* - *Pearson* - Προϋπόθεση για την εφαρμογή του είναι οι μεταβλητές να κατανέμονται κανονικά. Ο συντελεστής δίνεται από τη σχέση:

$$r = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}}$$

Εξίσωση 4.14

Όπου:

$$Cov(x, y) = E(x - \bar{x})(y - \bar{y})$$

Εξίσωση 4.15

Για την εκτίμηση του συντελεστή συσχέτισης χρησιμοποιούμε το δειγματικό συντελεστή συσχέτισης, ο οποίος δίνεται από τη σχέση:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Εξίσωση 4.16

- *r* - *Spearman* - Χρησιμοποιείται όταν οι μεταβλητές δεν κατανέμονται κανονικά. Είναι ίδιος με τον παραμετρικό συντελεστή του Pearson, με τη διαφορά ότι αντί να χρησιμοποιηθούν οι δειγματικές παρατηρήσεις x_i και y_i των μεταβλητών X και Y αντίστοιχα, χρησιμοποιούνται οι θέσεις (ranks - βαθμοί) αυτών, όπως έχουν καταταξιωθεί στο δείγμα. Ο συντελεστής δίνεται από τη σχέση:

$$r = \frac{\sum_{i=1}^n (x_{ri} - \bar{x})(y_{ri} - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{ri} - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_{ri} - \bar{y})^2}}$$

Εξίσωση 4.17

Η ισοδύναμα:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Εξίσωση 4.18

Όπου, d_i είναι η διαφορά των σχετικών θέσεων (ranks - βαθμών) των τιμών των x_i και y_i .

Test of Significance

Παραμετρικός έλεγχος

Εφόσον ο πληθυσμιακός συντελεστής συσχέτισης ρ εκτιμάται από τον δειγματικό συντελεστή r , πρέπει να αξιολογηθεί, και καθώς έχουμε υποθέσει ότι οι μεταβλητές X και Y κατανέμονται κανονικά, η υπόθεση που ελέγχεται είναι η:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Εξίσωση 4.19

Υπό το κριτήριο:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Εξίσωση 4.20

Όπου:

$$\text{Var}(r) = \frac{1-r^2}{n-2}$$

Εξίσωση 4.21

Η υπόθεση απορρίπτεται για τιμές:

$$|t| > t_{\alpha/2;n}$$

Εξίσωση 4.22

Μη Παραμετρικός έλεγχος

Ο έλεγχος για τον μη παραμετρικό συντελεστή συσχέτισης, είναι παρόμοιος. Η υπόθεση που ελέγχεται είναι η:

$$H_0 : \rho_s = 0$$

$$H_1 : \rho_s \neq 0$$

Εξίσωση 4.23

Υπό το κριτήριο:

$$t = \frac{r_s}{\sqrt{\frac{1-r_s^2}{n-2}}}$$

Εξίσωση 4.24

Όπου:

$$Var(r) = \frac{1-r_s^2}{n-2}$$

Εξίσωση 4.25

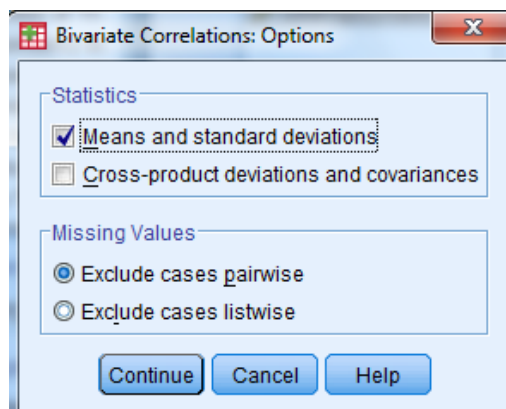
Η υπόθεση απορρίπτεται για τιμές:

$$|t| > t_{\alpha/2;n}$$

Εξίσωση 4.26

Στο συγκεκριμένο σημείο επιλέγουμε εάν ο έλεγχος που θα πραγματοποιηθεί είναι δίπλευρος *Two-tailed* ή μονόπλευρος *One-tailed*. Περισσότερα αναφορικά με τους ελέγχους υποθέσεων θα παρουσιαστούν σε επόμενο κεφάλαιο, παρόλα αυτά σε αυτό το σημείο παραθέτουμε τα παραπάνω στοιχεία καθώς αφορούν στον συντελεστή συσχέτισης, και δεν θα επανέλθουμε σ' αυτό σε άλλο σημείο του παρόντος συγγράμματος.

Flag significant correlations, με την επιλογή της παραμέτρου, όσοι συντελεστές συσχέτισης είναι στατιστικά σημαντικοί σε επίπεδο σ.σ. $\alpha = 5\%$, απεικονίζονται με έναν αστερίσκο, ενώ όσοι είναι στατιστικά σημαντικοί σε επίπεδο σ.σ. $\alpha = 1\%$ απεικονίζονται με δύο αστερίσκους, προκειμένου να είναι άμεσα αντιληπτοί από τον αναλυτή.



Εικόνα 4.64: Το menu *Bivariate Correlations - Options*

Από το πλήκτρο *Options* επιλέγουμε στο πλαίσιο *Statistics* τη μια ή και τις δύο από τις ακόλουθες παραμέτρους για τον συντελεστή Pearson:

- *Means and standard deviations* - Για να εμφανιστούν ο μέσος και η τυπική απόκλιση για κάθε μεταβλητή καθώς και το πλήθος των έγκυρων (μη ελλειπουσών τιμών) παρατηρήσεων. Η διαχείριση των ελλειπουσών τιμών πραγματοποιείται μέσω των κάτωθι παραμέτρων.
- *Cross-product deviations and covariances* - Ωστε να εμφανιστούν για κάθε ζεύγος μεταβλητών. Η «cross-product» τυπική απόκλιση ισούται με το άθροισμα των γινομένων των διαφορών των παρατηρήσεων από τη μέση τιμή τους (διορθωμένες παρατηρήσεις - mean corrected), δηλ. ο αριθμητής του συντελεστή συσχέτισης r . Η συνδιακύμανση είναι ένα μη τυποποιημένο μέτρο της σχέσης μεταξύ δύο μεταβλητών και ισούται με το «cross-product» διαιρούμενο με $N-1$.

Από το πλαίσιο *Missing Values* γίνεται η διαχείριση των ελλειπουσών τιμών επιλέγοντας μιας εκ των δύο παραμέτρων.

- *Exclude cases pairwise* - Μ' αυτή την επιλογή εξαιρούνται από τον υπολογισμό του συντελεστή συσχέτισης οι περιπτώσεις με ελλείπουσες τιμές για ένα ή και για τα δύο ζεύγη μεταβλητών.
- *Exclude cases listwise* - Μ' αυτή την επιλογή από τον υπολογισμό των συντελεστών συσχέτισης εξαιρούνται όλες οι περιπτώσεις που υπάρχει μια ελλείπουσα τιμή σε οποιαδήποτε από τις μεταβλητές.

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*, όπου και λαμβάνουμε τον περιγραφικό πίνακα με τον μέσο, την τυπική απόκλιση και το πλήθος των μεταβλητών, και τον πίνακα με τον συντελεστή συσχέτισης. Αναλυτικά στον πίνακα (Εικόνα 24) λαμβάνουμε τους συντελεστές συσχέτισης για όλους τους ανά δύο συνδυασμούς των μεταβλητών που έχουν δηλωθεί στο πλαίσιο *Variables*. Η διαγώνιος του πίνακα φέρει πάντα την τιμή «1», καθώς η συσχέτιση της μεταβλητής με τον εαυτό της είναι πάντα «1».

Σε κάθε κελί εμφανίζονται τρεις τιμές, όπου η πρώτη είναι ο συντελεστής συσχέτισης ($r = 0,413$), στη δεύτερη το p -value από τον έλεγχο (μηδενική υπόθεση) ισότητας του συντελεστή με το μηδέν (p -value = $0,126$ - άρα η υπόθεση δεν απορρίπτεται), και στην τρίτη το πλήθος των κοινών παρατηρήσεων μεταξύ των μεταβλητών. Εάν μια από τις δύο μεταβλητές έχει ελλείπουσα τιμή, τότε δεν θα χρησιμοποιηθεί ούτε η αντίστοιχη τιμή της άλλης μεταβλητής. Η ερμηνεία είναι κοινή και για τους δύο τύπους συντελεστών συσχέτισης.

Descriptive Statistics

	Mean	Std. Deviation	N
Total population at 01/01/1998.1000s	24372,20	27121,193	15
Unemployment rate of men and women	8,480	4,1509	15

Correlations

		Total population at 01/01/1998. 1000s	Unemployment rate of men and women
Total population at 01/01/1998.1000s	Pearson Correlation	1	,413
	Sig. (2-tailed)		,126
	N	15	15
Unemployment rate of men and women	Pearson Correlation	,413	1
	Sig. (2-tailed)	,126	
	N	15	15

Εικόνα 4.65: Αποτελέσματα από την επιλογή *Correlate*

4.7 Multiple Response

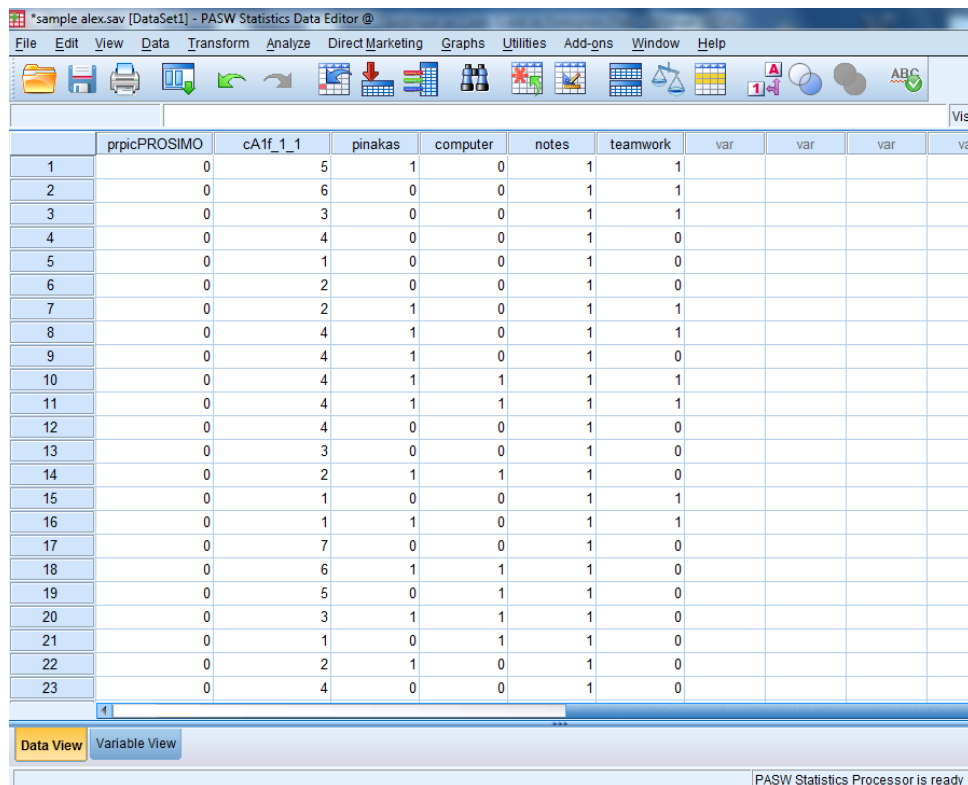
Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Multiple Response*, με σκοπό τη διαχείριση ερωτήσεων πολλαπλής επιλογής στις οποίες ο ερωτώμενος έχει τη δυνατότητα να επιλέξει περισσότερες από μία απαντήσεις.

Σε αυτή την περίπτωση κάθε δυνατή απάντηση θα οριστεί ως μια ξεχωριστή μεταβλητή την οποία ο ερωτώμενος είτε επέλεξε είτε όχι. Συνεπώς θα δημιουργηθούν μεταβλητές-δείκτες για την κάθε δυνατή απάντηση, με τιμές «1» εάν ο ερωτώμενος επέλεξε την απάντηση και «0» εάν όχι.

Έχουμε για παράδειγμα την ερώτηση «Τί μέσα χρησιμοποιείτε για την παράδοση του μαθήματος σας;», με πιθανές απαντήσεις:

- Μαυροπίνακας,
- Η/Υ και projector,
- Σημειώσεις,
- Ομαδικές εργασίες.

Η καταχώρηση των απαντήσεων, όπως φαίνεται και από την εικόνα, πραγματοποιείται με το να διαχειριζόμαστε την κάθε απάντηση ως ξεχωριστή μεταβλητή, που είτε επιλέχτηκε είτε όχι.



	prpicPROSIMO	cA1f_1_1	pinakας	computer	notes	teamwork	var	var	var	var
1	0	5	1	0	1	1				
2	0	6	0	0	1	1				
3	0	3	0	0	1	1				
4	0	4	0	0	1	0				
5	0	1	0	0	1	0				
6	0	2	0	0	1	0				
7	0	2	1	0	1	1				
8	0	4	1	0	1	1				
9	0	4	1	0	1	0				
10	0	4	1	1	1	1				
11	0	4	1	1	1	1				
12	0	4	0	0	1	0				
13	0	3	0	0	1	0				
14	0	2	1	1	1	0				
15	0	1	0	0	1	1				
16	0	1	1	0	1	1				
17	0	7	0	0	1	0				
18	0	6	1	1	1	0				
19	0	5	0	1	1	0				
20	0	3	1	1	1	0				
21	0	1	0	1	1	0				
22	0	2	1	0	1	0				
23	0	4	0	0	1	0				

Εικόνα 4.66: Δεδομένα της ερώτησης πολλαπλών απαντήσεων «Τί μέσα χρησιμοποιείτε για την παράδοση του μαθήματος σας;»

Οπότε, για την εξαγωγή συχνοτήτων σε αυτή την περίπτωση επιλέγουμε την εντολή *Multiple Responses*.

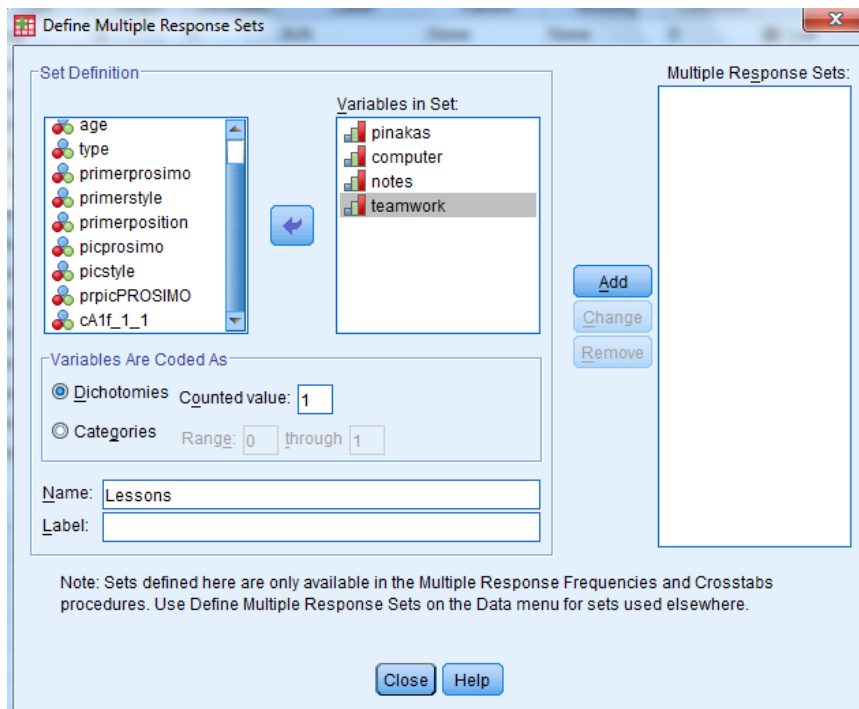
4.7.1 Analyze > Multiple Response > Define Multiple Response Sets

Αρχικά ορίζουμε το σύνολο των μεταβλητών που συνθέτουν την ερώτηση. Στο πλαίσιο *Variables in Set* δηλώνουμε όλες τις (δείκτριες) μεταβλητές που αποτελούν τις απαντήσεις της ερώτησης.

Στο πλαίσιο *Variables are coded* επιλέγουμε εάν οι μεταβλητές είναι δίτιμες *Dichotomies* και ορίζουμε την τιμή που θα λαμβάνεται ως επιλογή (1 - Ναι). Εάν οι τιμές που έχουν δοθεί σε κάθε μεταβλητή

είναι περισσότερες από δύο (δεν είναι δίτιμη) τότε θα πρέπει να ορίσουμε ποιες τιμές θα ανήκουν σε κάθε κατηγορία. Αυτό γίνεται στα πεδία *Categories* και στο *Range*.

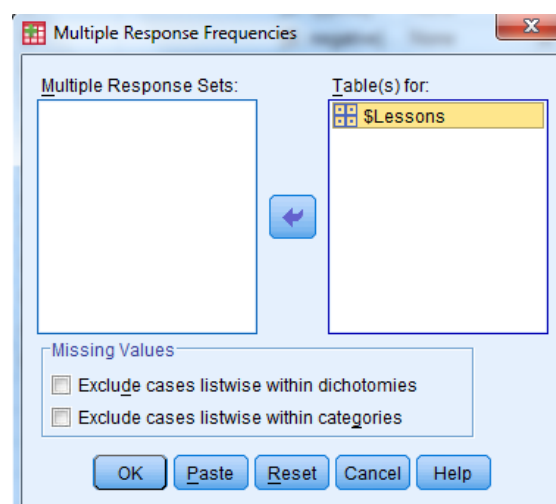
Στη συνέχεια μεταφέρουμε το σετ των μεταβλητών στο πλαίσιο *Multiple Response Sets*, το οποίο όπως διαπιστώνουμε λαμβάνει την ονομασία που έχουμε καταγράψει στο πεδίο *Name*.



Εικόνα 4.67: To menu *Multiple Response*> *Define Multiple Response Sets*

4.7.2 Analyze > Multiple Response > Frequencies

Έχοντας ήδη δημιουργήσει την ομάδα των μεταβλητών που συνθέτουν την ερώτηση, μπορούμε να δημιουργήσουμε την κατανομή πολλαπλών απαντήσεων για την ερώτηση. Αυτό επιτυγχάνεται από την εντολή *Analyze > Multiple Response > Frequencies*, όπου ορίζουμε το σετ που δημιουργήσαμε στο πλαίσιο *Table(s) for*.



Εικόνα 4.68: To menu *Multiple Response > Frequencies*

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*, και λαμβάνουμε τις απόλυτες και σχετικές συχνότητες της κάθε μεταβλητής-απάντησης. Όπως βλέπουμε συνολικά δόθηκαν 67 θετικές απαντήσεις. Η τρίτη στήλη μας παρέχει τα ποσοστά (σχετική συχνότητα) της κάθε μεταβλητής ξεχωριστά, έτσι για την απάντηση «Σημειώσεις» το 100% του δείγματος δήλωσε ότι χορηγεί σημειώσεις. Για την απάντηση «Μαυροπίνακας» το 50% του δείγματος (15/30) δήλωσε ότι κάνει χρήση αυτού, ενώ επί του συνόλου των θετικών απαντήσεων (15/67) αυτό το ποσοστό ανέρχεται σε 22,4%.

\$Lessons Frequencies

	Responses		Percent of Cases
	N	Percent	
\$Lessons ^a pinakas	15	22,4%	50,0%
computer	9	13,4%	30,0%
notes	30	44,8%	100,0%
teamwork	13	19,4%	43,3%
Total	67	100,0%	223,3%

a. Dichotomy group tabulated at value 1.

Εικόνα 4.69: Αποτελέσματα από την εντολή *Multiple Response > Frequencies*

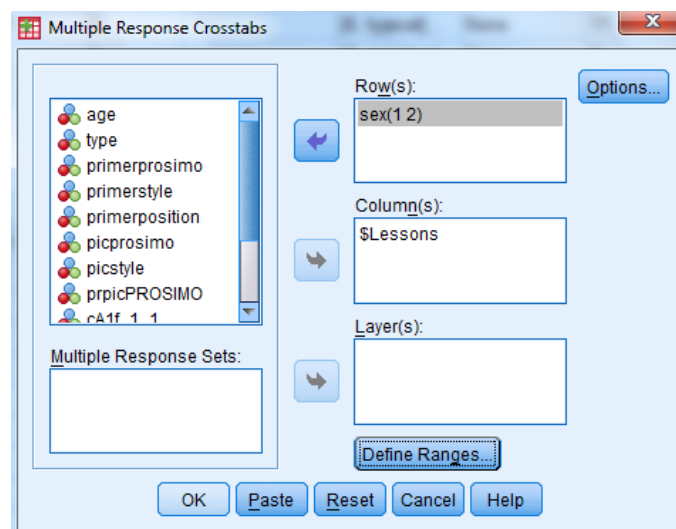
4.7.3 Analyze > Multiple Response > Crosstabs

Επιπλέον, από την εντολή *Analyze > Multiple Response > Crosstabs* μπορούμε να λάβουμε πίνακες συνάφειας με τις κατανομές συχνοτήτων σε συνδυασμό και με άλλες ποιοτικές ή ποσοτικές διακριτές μεταβλητές.

Στο πλαίσιο *Row(s)* τοποθετούμε τη μεταβλητή που θα εμφανίζεται στις γραμμές του πίνακα συνάφειας, ενώ στο πλαίσιο *Column(s)* τοποθετούμε τη μεταβλητή που θα εμφανίζεται στις στήλες του πίνακα.

Για την επιλεγμένη ποιοτική ή ποσοτική διακριτή μεταβλητή εμφανίζονται δύο λατινικά ερωτηματικά, τα οποία είναι οι θέσεις που θα πρέπει να δηλώσουμε το εύρος των τιμών της μεταβλητής. Αυτό γίνεται από το πλήκτρο *Define Ranges*, όπου καταγράφουμε την ελάχιστη *Minimum* και την μέγιστη τιμή *Maximum* που λαμβάνει η μεταβλητή.

Από την υποενότητα *Options* έχουμε τη δυνατότητα να δηλώσουμε εάν στον πίνακα συνάφειας θα εμφανίζονται οι σχετικές συχνότητες *Cell Percentages*, ανά γραμμή *Row*, ανά στήλη *Column*, ή σύνολο *Total*. Προσοχή θα πρέπει να δοθεί στην επιλογή *Percentages Based On*, όπου δηλώνουμε αν τα ποσοστά θα υπολογιστούν επί των ερωτώμενων *Cases*, ή επί των απαντήσεών τους *Responses*.



Εικόνα 4.70: Το menu *Multiple Response Crosstabs*

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*. Η ερμηνεία του πίνακα είναι όπως και προηγουμένως.

sex*\$Lessons Crosstabulation

			\$Lessons ^a				Total
			pinakas	computer	notes	teamwork	
sex	Boy	Count	10	4	15	6	15
	Girl	Count	5	5	15	7	15
Total		Count	15	9	30	13	30

Percentages and totals are based on respondents.
a. Dichotomy group tabulated at value 1.

Εικόνα 4.71: Αποτελέσματα από την εντολή *Multiple Response > Crosstabs*

4.8 Graphs

Περιέχει τη δυνατότητα γραφικής απεικόνισης των δεδομένων μέσω προκαθορισμένων τύπων διαγραμμάτων. Γράφημα (plot) ονομάζεται μια γραφική αναπαράσταση μιας ή περισσότερων μεταβλητών. Τα γραφήματα είναι χρήσιμα διότι μας παρέχουν άμεση πληροφόρηση για το σχήμα και τη μορφή της κατανομής των μεταβλητών, ή την από κοινού σχέση τους.

Βασικά χαρακτηριστικά για τη δημιουργία ενός γραφήματος είναι:

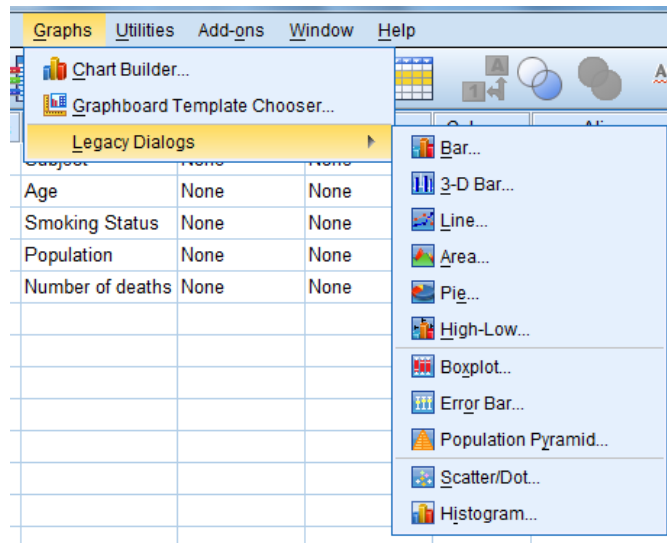
- Να ξεκαθαρίσουμε τους στόχους και τη ζητούμενη πληροφορία που επιθυμούμε να αντλήσουμε.
- Να επιλέξουμε το κατάλληλο είδος γραφήματος.
- Να δημιουργήσουμε ένα ξεκάθαρο και κατανοητό γράφημα σύμφωνο με το αρχικό ερώτημα.

Η επιλογή *Legacy Dialogs* διαφέρει από την εντολή *Chart Builder* στο ότι μας επιτρέπει να δημιουργήσουμε βασικά γραφήματα επιλέγοντας εκ των προτέρων το είδος του γραφήματος που θα δημιουργήσουμε.

Το *Chart Builder*, ωστόσο, είναι μια πραγματική μηχανή γραφικών που δημιουργήθηκε με την προσαρμογή του προγράμματος στην εξέλιξη της τεχνολογίας για να προσφέρει στους χρήστες μεγαλύτερη ευελιξία στη γραφική απεικόνιση, οδηγώντας τον χρήστη καθώς του παρέχει τη δυνατότητα να βλέπει πώς θα διαμορφωθεί το γράφημα με κάθε επιλογή του.

Καθώς αυτή η επιλογή μοιάζει πολύ με τον τρόπο δημιουργίας του Excel, θα παρουσιάσουμε τις δυνατές επιλογές του *Legacy Dialogs* κάνοντας μια αναφορά στο είδος του κάθε γραφήματος. Όλα τα διαγράμματα μπορούν να μεταφερθούν σε άλλο πρόγραμμα π.χ. Word, Powerpoint με τη διαδικασία αντιγραφής *Copy*, επικόλλησης *Paste*.

Οι τύποι των διαγραμμάτων που προσφέρονται είναι:

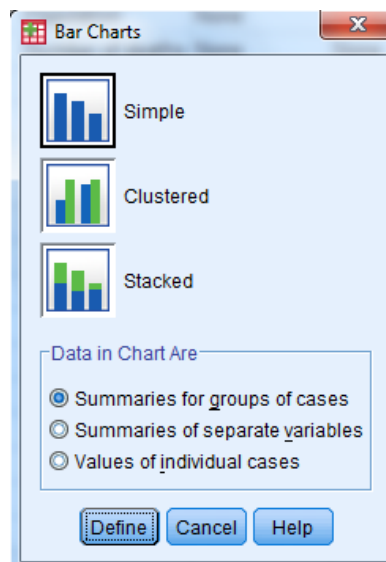


Εικόνα 4.72: Τύποι γραφημάτων που παρέχονται από το menu *Graphs > Legacy Dialogs*

4.9 Ραβδογράμματα - Bar Charts

Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > Bar*, για τη δημιουργία ραβδογράμματος. Τα ραβδογράμματα χρησιμοποιούνται για τη γραφική απεικόνιση των συχνοτήτων (απόλυτων και σχετικών) ποιοτικών μεταβλητών. Αποτελούνται από οριζόντιες στήλες που οι βάσεις τους βρίσκονται στον οριζόντιο (πιο σπάνια στον κάθετο) άξονα. Σε κάθε τιμή της μεταβλητής X αντιστοιχεί μια ορθογώνια στήλη της οποίας το ύψος είναι ίσο με την αντίστοιχη συχνότητα ή σχετική συχνότητα (ή ακόμα και με τις τιμές κάποιας άλλης μεταβλητής, π.χ. μέσος). Έτσι έχουμε αντίστοιχα ραβδογράμματα συχνοτήτων και ραβδογράμματα σχετικών συχνοτήτων. Τόσο η απόσταση μεταξύ των στηλών, όσο και το πλάτος αυτών καθορίζονται αυθαίρετα από τον δημιουργό.

Τα ραβδογράμματα μπορούν να είναι απλά *Simple*, όταν αφορούν μια μεταβλητή, ομαδοποιημένα *Clustered* ή *Stacked*, όταν αφορούν σε περισσότερες από μια μεταβλητές.



Εικόνα 4.73: Το αρχικό menu Bar Chart

- *Data in Chart Are* - Ορίζουμε το είδος των δεδομένων που είναι στη λίστα μεταβλητών.
- *Summaries for groups of cases* - Είναι η επιλογή που χρησιμοποιούμε καθώς μας δείχνει το πλήθος των περιπτώσεων που περιέχονται σε κάθε τιμή της μεταβλητής.
- *Summaries of separate variables* – Μ' αυτή την επιλογή μπορούμε να λάβουμε αποτελέσματα έχοντας σε κάθε ράβδο π.χ. το μέσο άλλων μεταβλητών.
- *Values of individual cases*.

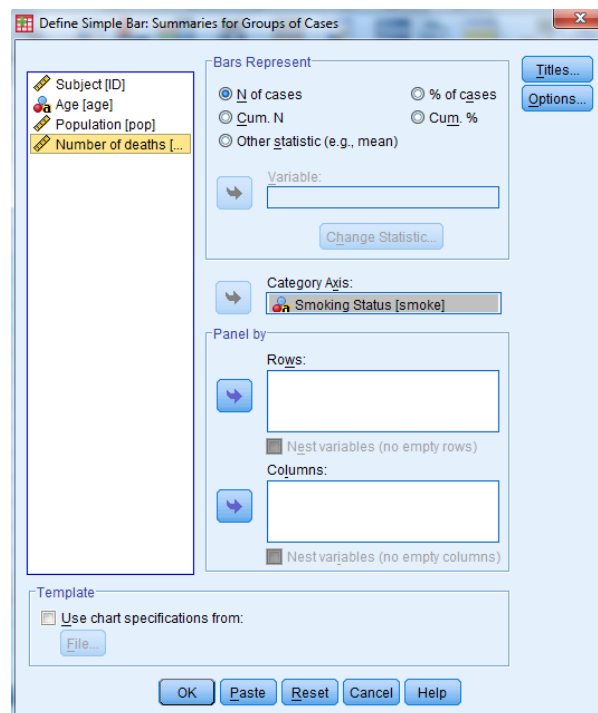
Συνεχίζουμε στην επόμενη ενότητα με το πλήκτρο *Define*. Θα παρουσιάσουμε και τις δύο δυνατές επιλογές *Simple* και *Clustered* καθώς εξυπηρετούν τις πιο συνήθεις ανάγκες απεικόνισης των μεταβλητών μας.

Simple

Στο *Bars Represent* επιλέγουμε σε τί θα αναφέρονται οι ράβδοι του γραφήματος. Οι δυνατές επιλογές είναι :

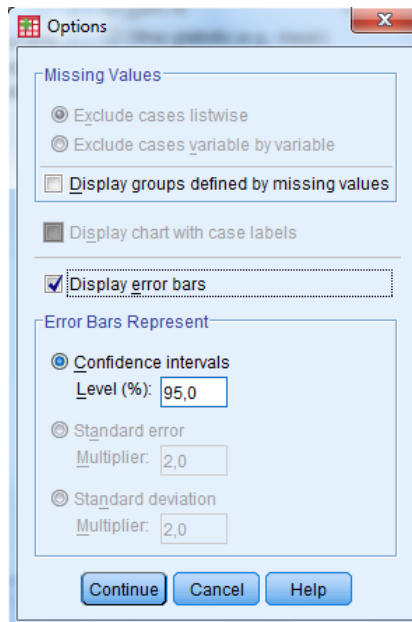
- *N of cases*: Το πλήθος των παρατηρήσεων (συχνότητα).
- *% of cases*: Το ποσοστό των παρατηρήσεων (σχετική συχνότητα).
- *Cum N*: Η αθροιστική συχνότητα.
- *Cum %*: Η αθροιστική σχετική συχνότητα.
- *Other statistic* (π.χ. *mean*): Κάποιο άλλο στατιστικό όπως η μέση τιμή μιας μεταβλητής που θα οριστεί στο πεδίο *Variable*.

Στο πεδίο *Category Axis* δηλώνουμε την ποιοτική μεταβλητή για την οποία θα δημιουργήσουμε το γράφημα.



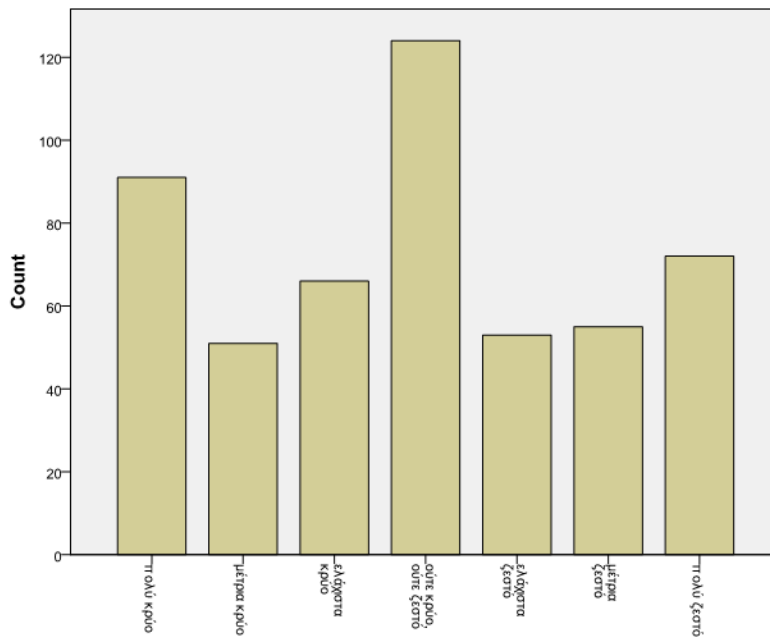
Εικόνα 4.74: Η κεντρική ενότητα της επιλογής *Bar Chart > Simple*

Από το πλήκτρο *Titles* δηλώνουμε τους τίτλους του γραφήματος, ενώ από το πλήκτρο *Options* μπορούμε να επιλέξουμε την εμφάνιση διαστημάτων σφάλματος (π.χ. συν ή πλην μια τυπική απόκλιση) για τη μέση τιμή σε κάθε ράβδο.



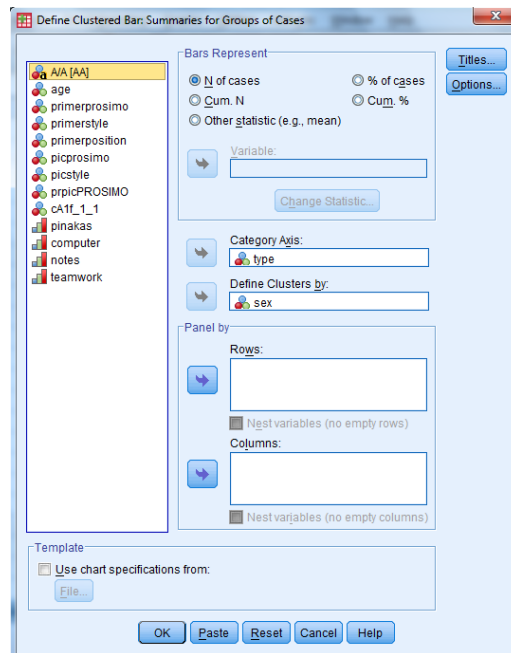
Εικόνα 4.75: Το menu Bar Chart > Simple >Options

Η διαδικασία ολοκληρώνεται με το πλήκτρο OK όπου και λαμβάνουμε το γράφημα:



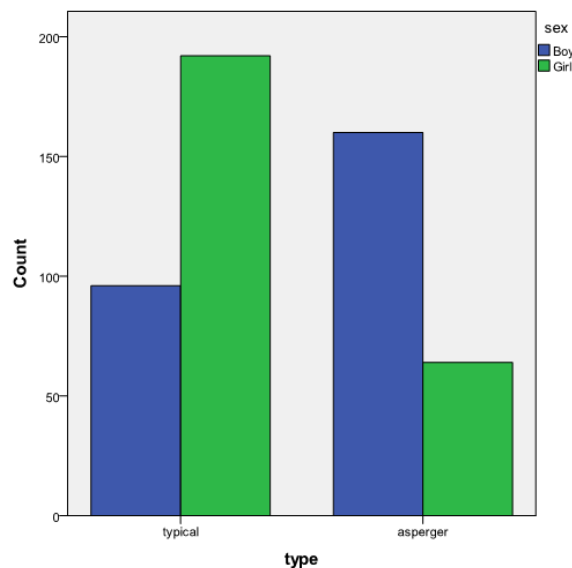
Εικόνα 4.76: Αποτελέσματα από την εντολή Bar Chart > Simple

Clustered



Εικόνα 4.77: Η κεντρική ενότητα της επιλογής *Bar Chart > Clustered*

Για τα ομαδοποιημένα ραβδογράμματα ορίζουμε επιπρόσθετα στο πεδίο *Define Clusters by* τη δεύτερη ποιοτική ή ποσοτική διακριτή μεταβλητή που θα διαμορφώσει τις ομάδες. Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα:



Εικόνα 4.78: Αποτελέσματα από την εντολή *Bar Chart > Clustered*

4.10 Γραμμικά - Line

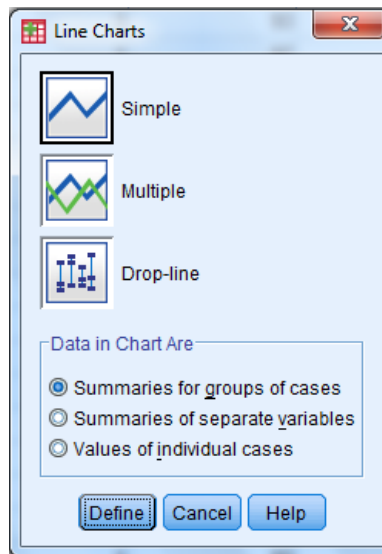
Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > Line*, για τη δημιουργία διαγράμματος γραμμής, το οποίο χρησιμοποιείται για την απεικόνιση των τιμών μιας μεταβλητής συνήθως σε σχέση με τον χρόνο. Τα διαγράμματα αυτά χρησιμοποιούνται για την απεικόνιση ποσοτικών μεταβλητών, διαφορών στους μέσους διαχρονικά ή υπό συνθήκες, ή μεταξύ ομάδων. Έτσι στον άξονα x τοποθετούμε τη χρονική (διατάξιμη) μεταβλητή (time) και στον y άξονα τοποθετούμε την ποσοτική μεταβλητή (ή τις τιμές των μέσων κ.λπ).

Τα γραμμικά διαγράμματα μπορούν να είναι απλά *Simple*, όταν αφορούν μια μεταβλητή, ή ομαδοποιημένα *Multiple*, όταν αφορούν σε περισσότερες από μια μεταβλητές.

Data in Chart Are, ορίζουμε το είδος των δεδομένων που είναι στη λίστα μεταβλητών.

- *Summaries for groups of cases* - Είναι η επιλογή που χρησιμοποιούμε καθώς μας δείχνει το πλήθος των περιπτώσεων που περιέχονται σε κάθε κατηγορία.
- *Summaries of separate variables* – Μ' αυτή την επιλογή μπορούμε να λάβουμε αποτελέσματα έχοντας σε κάθε ράβδο π.χ. τον μέσο άλλων μεταβλητών.
- *Values of individual cases*.

Συνεχίζουμε στην επόμενη ενότητα με το πλήκτρο *Define*.



Εικόνα 4.79: Το αρχικό menu Line Charts

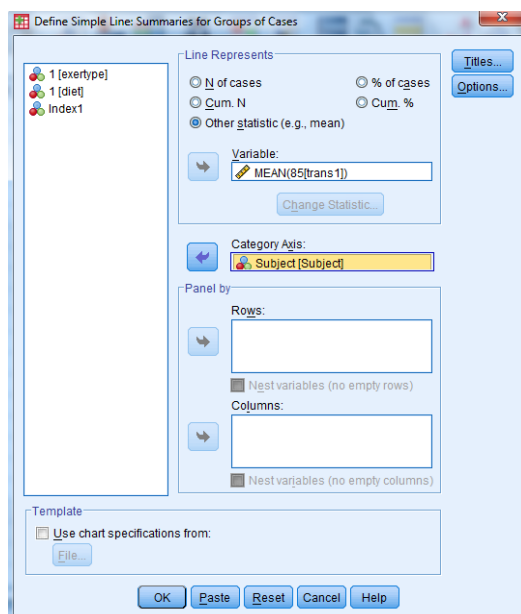
Θα παρουσιάσουμε και τις δύο δυνατές επιλογές *Simple* και *Multiple* καθώς εξυπηρετούν τις πιο σύνθητες ανάγκες απεικόνισης των μεταβλητών μας.

Simple

Στο *Line Represents* επιλέγουμε σε τί θα αναφέρονται οι γραμμές του γραφήματος. Οι δυνατές επιλογές είναι:

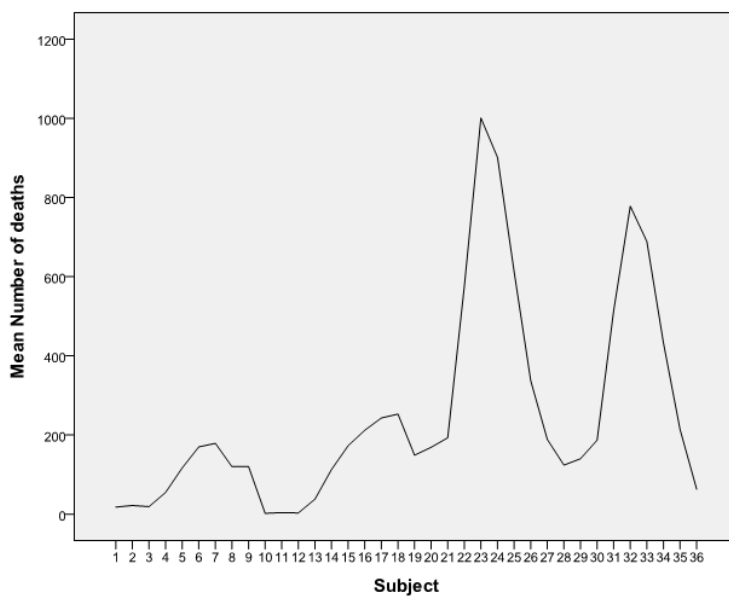
- *N of cases*: Το πλήθος των παρατηρήσεων (συχνότητα).
- *% of cases*: Το ποσοστό των παρατηρήσεων (σχετική συχνότητα).
- *Cum N*: Η αθροιστική συχνότητα.
- *Cum %*: Η αθροιστική σχετική συχνότητα.
- *Other statistic* (π.χ. *mean*): Κάποιο άλλο στατιστικό όπως η μέση τιμή μιας μεταβλητής που θα οριστεί στο πεδίο *Variable*.

Στο πεδίο *Category Axis* δηλώνουμε τη χρονική (διατάξιμη) μεταβλητή. Από το πλήκτρο *Titles* δηλώνουμε τους τίτλους του γραφήματος, από το πλήκτρο *Options* μπορούμε να επιλέξουμε την εμφάνιση διαστημάτων σφάλματος (π.χ. συν ή πλην μια τυπική απόκλιση).



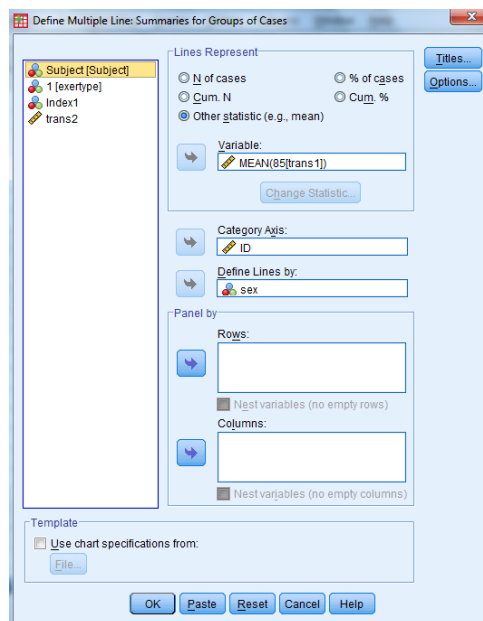
Εικόνα 4.80: Η κεντρική ενότητα της επιλογής *Line Charts > Simple*

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα της Εικόνας 4.40.



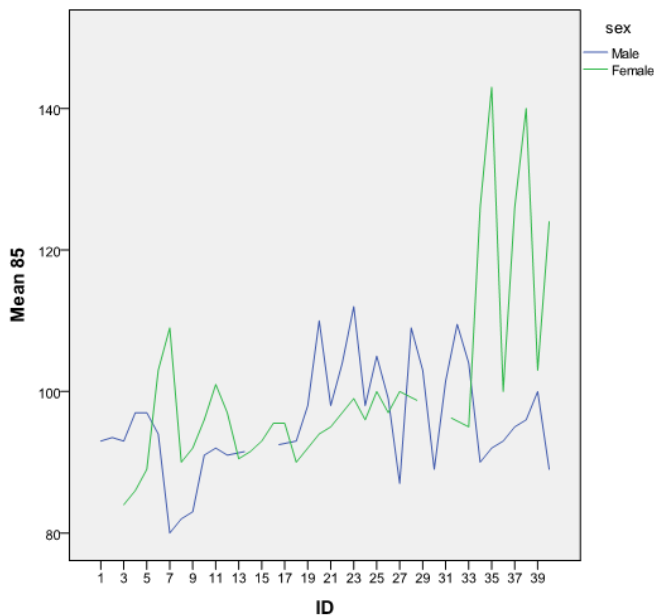
Εικόνα 4.81: Αποτελέσματα από την εντολή *Line Chart > Simple*

Multiple



Εικόνα 4.82: Η κεντρική ενότητα της επιλογής *Line Charts > Multiple*

Για τις πολλαπλές γραμμές ορίζουμε επιπρόσθετα στο πεδίο *Define Lines by* την ποιοτική (ή ποσοτική διακριτή) μεταβλητή που θα διαμορφώσει την κάθε γραμμή. Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα της Εικόνας 4.42.

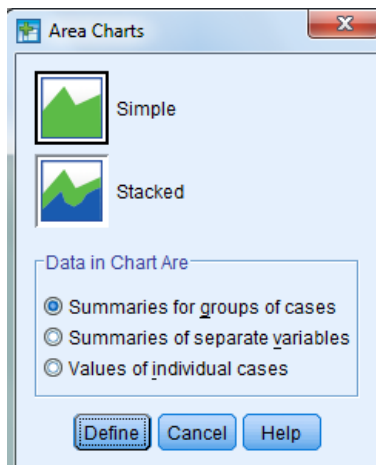


Εικόνα 4.83: Αποτελέσματα από την εντολή *Line Chart > Multiple*

4.11 Γράφημα Περιοχής - Area

Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > Area*, για τη δημιουργία γραφήματος περιοχής. Ένα γράφημα περιοχής αφορά ποσοτικά δεδομένα και η δημιουργία του βασίζεται στο γράφημα γραμμής που είδαμε παραπάνω, όπου η περιοχή μεταξύ του άξονα και της δημιουργημένης γραμμής (line chart) γεμίζει με κάποιο χρώμα. Το γράφημα αυτό χρησιμοποιείται κυρίως για να συγκρίνει κανείς με μια περιοχή γραφήματος δύο ή περισσότερες ποσότητες.

Τα διαγράμματα περιοχής μπορούν να είναι απλά *Simple*, όταν αφορούν μια μεταβλητή, ή στοιβαγμένα *Stacked*, όταν αφορούν περισσότερες από μια μεταβλητές.



Εικόνα 4.84: Το αρχικό menu Area Charts

- *Data in Chart Are* - Ορίζουμε το είδος των δεδομένων που είναι στη λίστα μεταβλητών.
- *Summaries for groups of cases* - Είναι η επιλογή που χρησιμοποιούμε καθώς μας δείχνει το πλήθος των περιπτώσεων που περιέχονται σε κάθε κατηγορία.
- *Summaries of separate variables*- Με αυτή την επιλογή μπορούμε να λάβουμε αποτελέσματα έχοντας σε κάθε ράβδο π.χ. τον μέσο άλλων μεταβλητών.
- *Values of individual cases*.

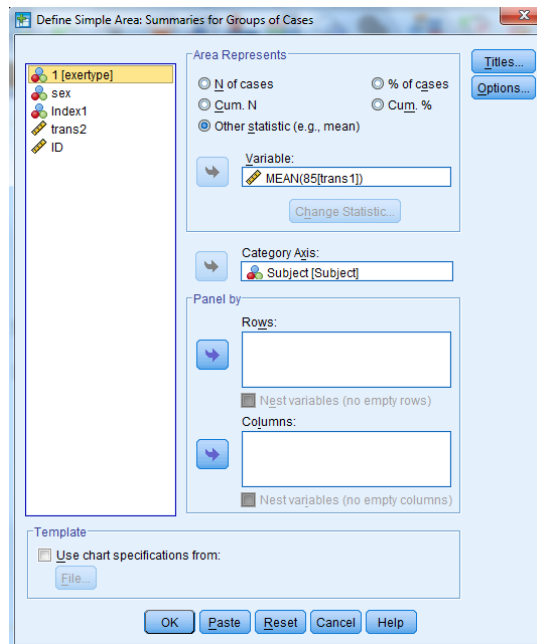
Συνεχίζουμε στην επόμενη ενότητα με το πλήκτρο *Define*.

Simple

Στο *Area Represents* επιλέγουμε σε τί θα αναφέρονται οι περιοχές του γραφήματος. Οι δυνατές επιλογές είναι:

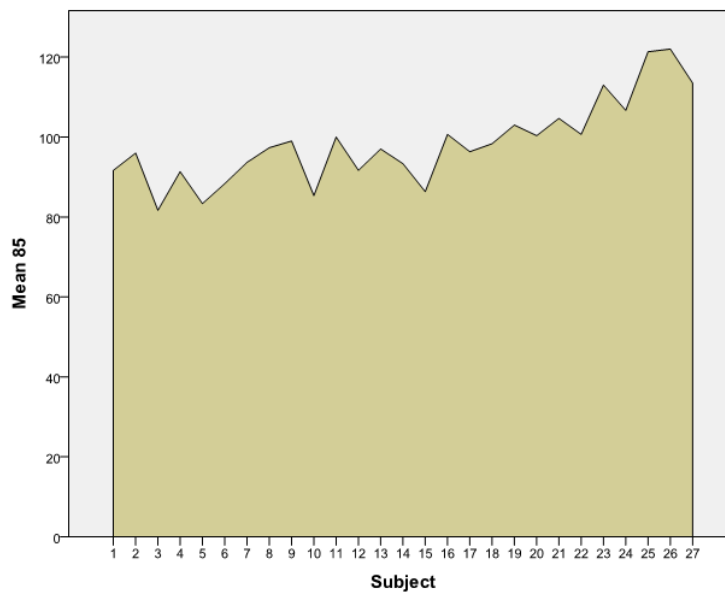
- *N of cases* - Το πλήθος των παρατηρήσεων (συχνότητα).
- *% of cases* - Το ποσοστό των παρατηρήσεων (σχετική συχνότητα).
- *Cum N* - Η αθροιστική συχνότητα.
- *Cum %* - Η αθροιστική σχετική συχνότητα.
- *Other statistic (e.g. mean)* - Κάποιο άλλο στατιστικό όπως η μέση τιμή μιας μεταβλητής που θα οριστεί στο πεδίο *Variable*.

Στο πεδίο *Category Axis* δηλώνουμε τη χρονική (διατάξιμη) μεταβλητή. Από το πλήκτρο *Titles* δηλώνουμε τους τίτλους του γραφήματος, από το πλήκτρο *Options* μπορούμε να επιλέξουμε την εμφάνιση διαστημάτων σφάλματος (π.χ. συν ή πλην μια τυπική απόκλιση).



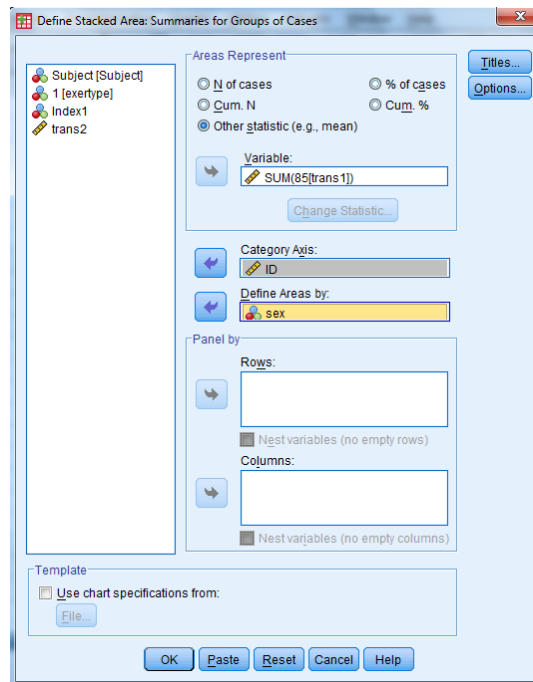
Εικόνα 4.85: Η κεντρική ενότητα της επιλογής *Area Charts > Simple*

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα της Εικόνας 4.45.



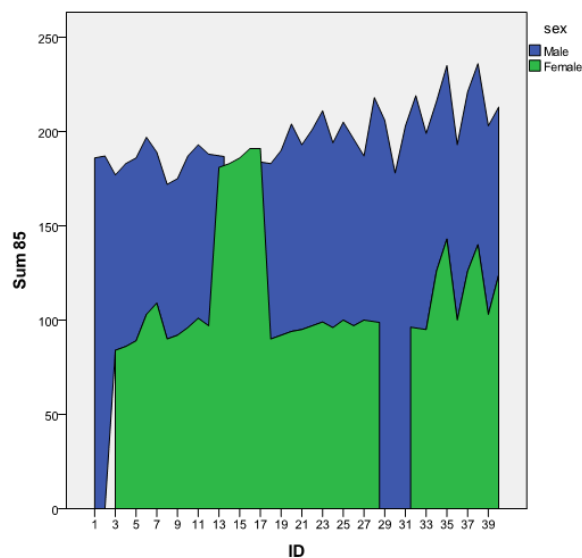
Εικόνα 4.86: Αποτελέσματα από την εντολή *Area Chart > Simple*

Για τη σύγκριση περισσότερων από μιας ποσότητας ορίζουμε επιπρόσθετα στο πεδίο *Define Areas by* την ποιοτική (ή ποσοτική διακριτή) μεταβλητή που θα διαμορφώσει την κάθε περιοχή.



Εικόνα 4.87: Η κεντρική ενότητα της επιλογής Area Charts > Stacked

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα της Εικόνας 4.47.

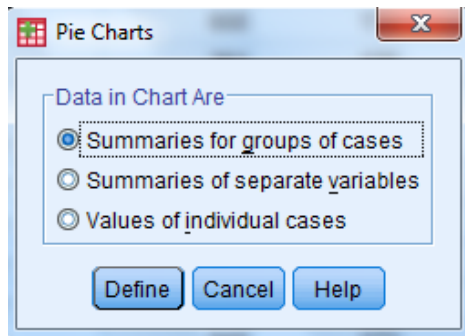


Εικόνα 4.88: Αποτελέσματα από την εντολή Area Chart > Stacked

4.12 Κυκλικά (πιτογράμματα) - Pie

Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > Pie*, για τη δημιουργία κυκλικού διαγράμματος.

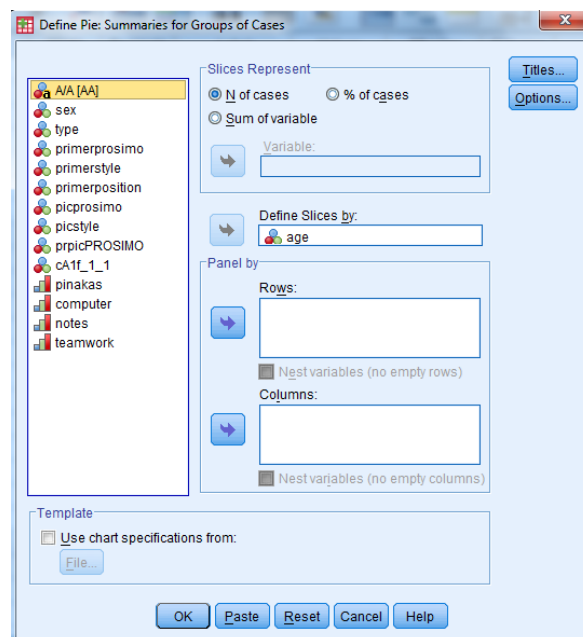
Το κυκλικό διάγραμμα χρησιμοποιείται για τη γραφική απεικόνιση των ποιοτικών μεταβλητών. Είναι ένας κυκλικός δίσκος χωρισμένος σε κυκλικούς τομείς, τα εμβαδά των οποίων (ή ισοδύναμα τα τόξα) είναι ανάλογα προς τις αντίστοιχες απόλυτες ή σχετικές συχνότητες των τιμών της μεταβλητής.



Εικόνα 4.89: Το αρχικό menu Pie Charts

- *Data in Chart Are* - Ορίζουμε το είδος των δεδομένων που είναι στη λίστα μεταβλητών.
- *Summaries for groups of cases* - Είναι η επιλογή που χρησιμοποιούμε καθώς μας δείχνει το πλήθος των περιπτώσεων που περιέχονται σε κάθε κατηγορία.
- *Summaries of separate variables* - Με αυτή την επιλογή μπορούμε να λάβουμε αποτελέσματα έχοντας σε κάθε ράβδο π.χ. τον μέσο άλλων μεταβλητών.
- *Values of individual cases.*

Συνεχίζουμε στην επόμενη ενότητα με το πλήκτρο *Define*.

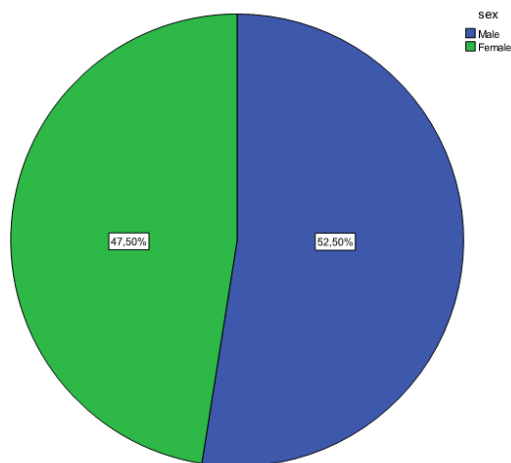


Εικόνα 4.90: Η κεντρική ενότητα της επιλογής Pie Chart

Στο *Slices Represent* επιλέγουμε σε τί θα αναφέρονται οι κυκλικοί δίσκοι του γραφήματος. Οι δυνατές επιλογές είναι:

- *N of cases* - Το πλήθος των παρατηρήσεων (συχνότητα).
- *% of cases* - Το ποσοστό των παρατηρήσεων (σχετική συχνότητα).
- *Cum N* - Η αθροιστική συχνότητα.
- *Cum %* - Η αθροιστική σχετική συχνότητα.
- *Other statistic (e.g. mean)* - Κάποιο άλλο στατιστικό όπως η μέση τιμή μιας μεταβλητής που θα οριστεί στο πεδίο *Variable*.

Στο πεδίο *Defined Slices by* δηλώνουμε την ποιοτική μεταβλητή για την οποία θα δημιουργήσουμε το γράφημα. Από το πλήκτρο *Titles* δηλώνουμε τους τίτλους του γραφήματος. Μπορούμε να επιλέξουμε και άλλες προσαρμοσμένες επιλογές, όπως να διαχωρίσουμε τα δεδομένα σε γραμμές και στήλες, τα οποία όμως δεν είναι υποχρεωτικά για τη δημιουργία ενός βασικού κυκλικού γραφήματος. Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*, όπου και λαμβάνουμε το γράφημα. Με διπλό κλικ στο γράφημα μπορούμε να προβούμε σε μορφολογική επεξεργασία του.



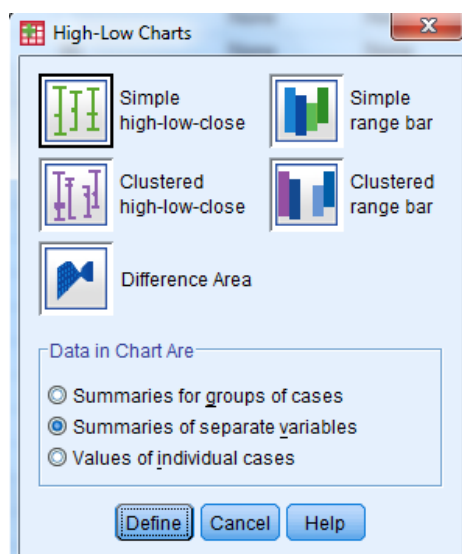
Εικόνα 4.91: Αποτελέσματα από την εντολή *Pie Chart*

4.13 High - Low

Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > High - Low*, για τη δημιουργία διαγράμματος Υψηλών - Χαμηλών τιμών. Το συγκεκριμένο είναι ένας τύπος γραφήματος που χρησιμοποιείται για να δείξει την εξέλιξη της τιμής του χρηματοπιστωτικού μέσου με την πάροδο του χρόνου.

Κάθε κάθετη γραμμή στο διάγραμμα δείχνει το εύρος τιμών (υψηλότερες, χαμηλότερες και τιμές κλεισίματος μιας μετοχής) σε σχέση με τη μονάδα του χρόνου. Θα μπορούσαν να δοθούν διάφορες αποχρώσεις ανάλογα με το αν οι τιμές αυξήθηκαν ή μειώθηκαν κατά την περίοδο που εξετάζονται.

Τα διαγράμματα Υψηλών - Χαμηλών τιμών μπορούν να είναι απλά *Simple*, όταν αφορούν σε μια μεταβλητή, ή ομαδοποιημένα *Clustered* όταν αφορούν σε περισσότερες από μια μεταβλητές. Θα παρουσιάσουμε τις επιλογές *Simple* και *Clustered* που είναι οι πιο συνήθεις.



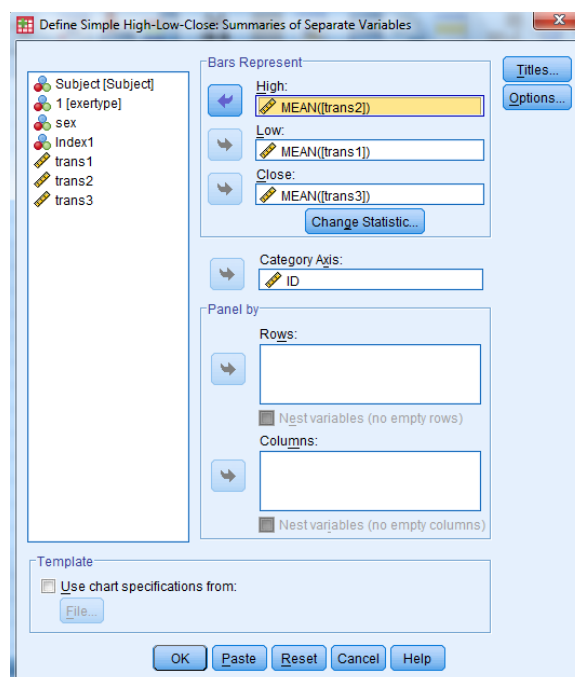
Εικόνα 4.92: Το αρχικό menu *High - Low Charts*

- *Data in Chart Are* - Ορίζουμε το είδος των δεδομένων που είναι στη λίστα μεταβλητών.
- *Summaries for groups of cases* - Είναι η επιλογή που χρησιμοποιούμε καθώς μας δείχνει το πλήθος των περιπτώσεων που περιέχονται σε κάθε κατηγορία.
- *Summaries of separate variables* - Με αυτή την επιλογή μπορούμε να λάβουμε αποτελέσματα έχοντας σε κάθε ράβδο π.χ. τον μέσο άλλων μεταβλητών.
- *Values of individual cases.*

Συνεχίζουμε στην επόμενη ενότητα με το πλήκτρο *Define*.

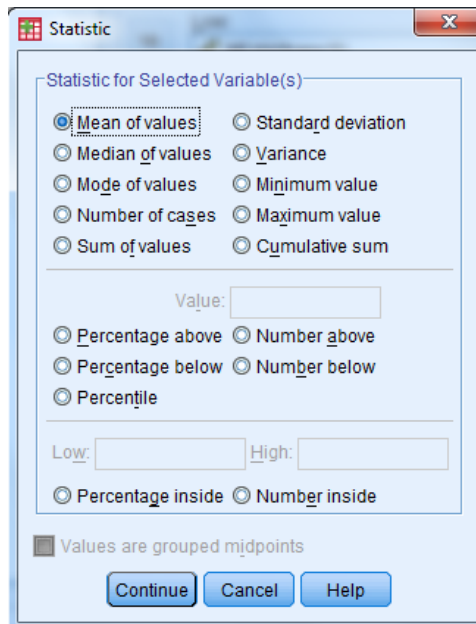
Simple

Στα πεδία *High-Low-Close* ορίζουμε την μεταβλητή που μας δίνει την υψηλότερη, την χαμηλότερη και την τιμή κλεισίματος, π.χ. της μετοχής. Στο πεδίο *Category Axis* δηλώνουμε τη χρονική (διατάξιμη) μεταβλητή. Από το πλήκτρο *Titles* δηλώνουμε τους τίτλους του γραφήματος.



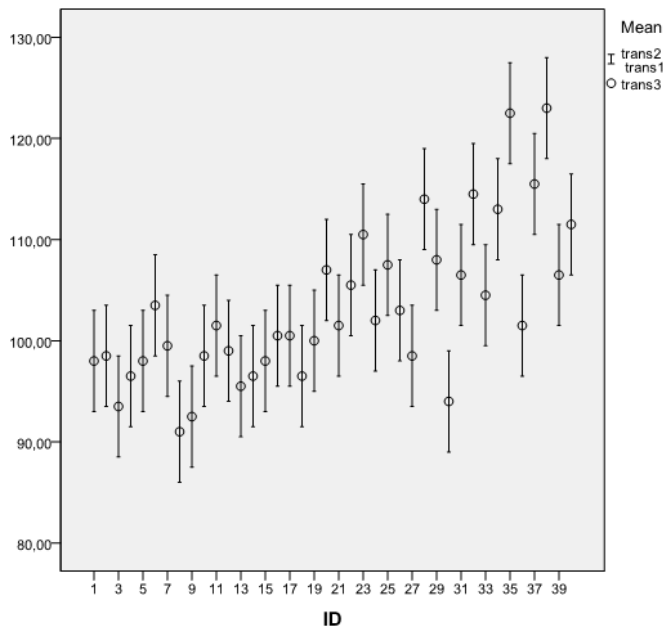
Εικόνα 4.93: Η κεντρική ενότητα της επιλογής *High - Low Charts > Simple*

Από το πλήκτρο *Statistic* επιλέγουμε το περιγραφικό μέτρο που θα απεικονίζεται στο γράφημα. Η πιο συνηθής τιμή που λαμβάνεται και προκαθορισμένη από την εφαρμογή είναι ο μέσος.



Εικόνα 4.94: Το menu από την εντολή High - Low Chart > Simple > Statistic

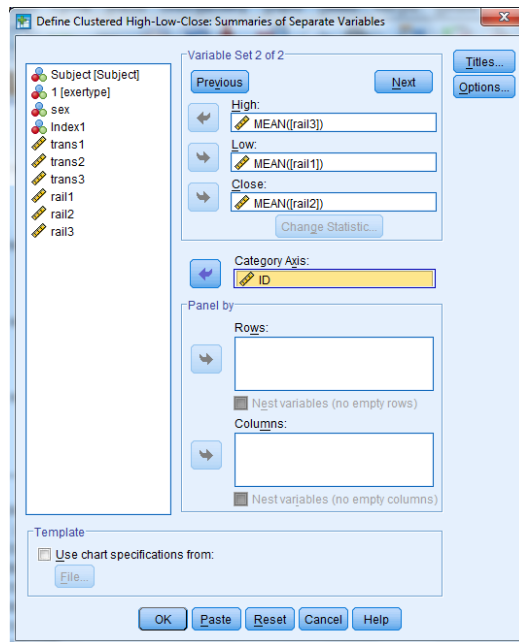
Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK*, όπου και λαμβάνουμε το γράφημα της Εικόνας 4.54.



Εικόνα 4.95: Αποτελέσματα από την εντολή High - Low Chart > Simple

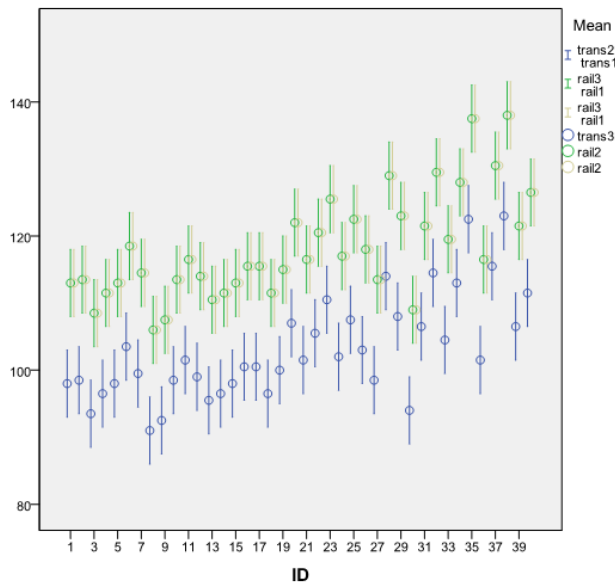
Clustered

Για την απεικόνιση των τιμών περισσότερων από μιας «μετοχής» ορίζουμε τις τιμές τους στα αντίστοιχα πεδία με την αρχική «μετοχή». Τα πεδία εμφανίζονται με το πλήκτρο *Next*.



Εικόνα 4.96: Η κεντρική ενότητα της επιλογής High - Low Charts > Clustered

Η διαδικασία ολοκληρώνεται με το πλήκτρο OK, όπου και λαμβάνουμε το γράφημα της Εικόνας 4.56.



Εικόνα 4.97: Αποτελέσματα από την εντολή High - Low Chart > Clustered

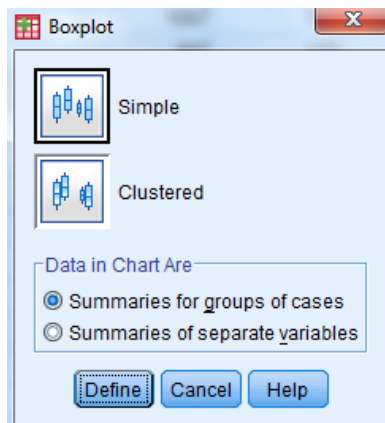
4.14 Θηκογράμματα ή Πλαισίου-Απολήξεων BoxPlots

Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > Box Plot*, για τη δημιουργία διαγράμματος πλαισίου-απολήξεων. Τα διαγράμματα πλαισίου-απολήξεων (ή αλλιώς θηκογράμματα) απεικονίζουν την κατανομή των ποσοτικών μεταβλητών και παρέχουν τη δυνατότητα διερεύνησης της ύπαρξης ακραίων τιμών σε μια μεταβλητή. Ο κορμός του διαγράμματος αποτελείται από ένα ορθογώνιο παραλληλόγραμμο του οποίου η κάτω πλευρά (βάση) απεικονίζει το πρώτο τεταρτημόριο Q_1 , ενώ η πάνω πλευρά (κορυφή) το τρίτο τεταρτημόριο Q_3 . Η γραμμή εντός του παραλληλογράμμου απεικονίζει τη διάμεσο των παρατηρήσεων. Η μέγιστη τιμή του διαγράμματος (άνωτερο φράγμα) δίνεται από τη σχέση $Q_3 + 1.5(Q_3 - Q_1)$ και η ελάχιστη (κατώτερο φράγμα) από την σχέση $Q_1 - 1.5(Q_3 - Q_1)$.

Τα θηκογράμματα μπορούν να είναι απλά *Simple*, όταν αφορούν σε μια μεταβλητή είτε ομαδοποιημένα *Clustered*, όταν αφορούν σε περισσότερες από μια μεταβλητές.

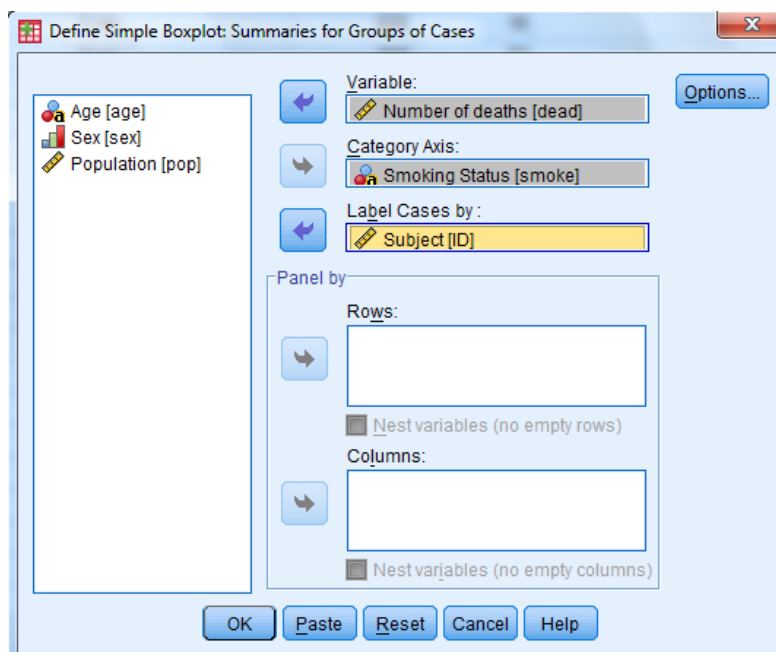
- *Data in Chart Are*: Ορίζουμε το είδος των δεδομένων που είναι στη λίστα μεταβλητών.
- *Summaries for groups of cases*: Είναι η επιλογή που χρησιμοποιούμε καθώς μας δείχνει το πλήθος των περιπτώσεων που περιέχονται σε κάθε κατηγορία.
- *Summaries of separate variables*: Μ' αυτή την επιλογή μπορούμε να λάβουμε αποτελέσματα έχοντας σε κάθε πλαίσιο π.χ. τον μέσο άλλων μεταβλητών.

Συνεχίζουμε στην επόμενη ενότητα με το πλήκτρο *Define*.



Εικόνα 4.98: Το αρχικό menu *Boxplot*

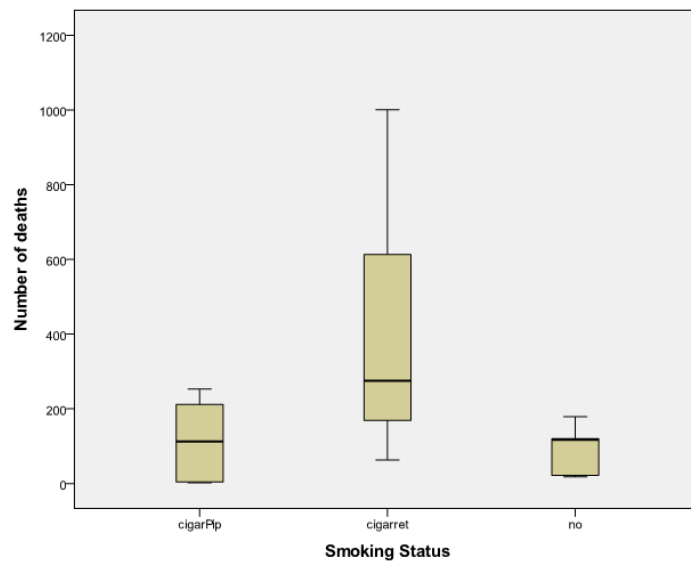
Simple



Εικόνα 4.99: Η κεντρική ενότητα της επιλογής *Boxplot > Simple*

Στο πεδίο *Variable* επιλέγουμε την ποσοτική μεταβλητή για την οποία θα δημιουργηθεί το γράφημα. Στο πεδίο *Category Axis* μπορούμε να δηλώσουμε μια ποιοτική μεταβλητή για να δημιουργήσουμε ξεχωριστό θηκόγραμμα για την κάθε κατηγορία της. Στο πεδίο *Label Cases by* δηλώνουμε τη μεταβλητή η οποία ονοματίζει τις παρατηρήσεις μας, συνήθως είναι η μεταβλητή *Subject* ή *ID*. Από το πλήκτρο *Titles* δηλώνουμε τους τίτλους του γραφήματος.

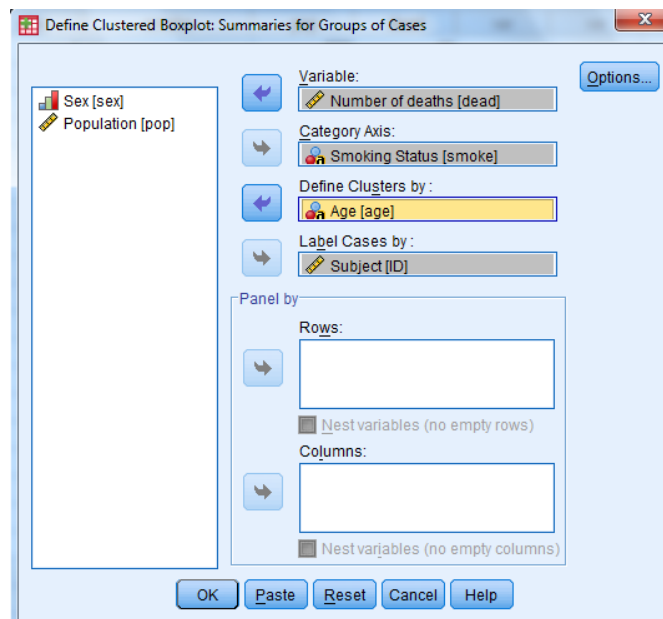
Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα.



Εικόνα 4.100: Αποτελέσματα από την εντολή *Boxplot > Simple*

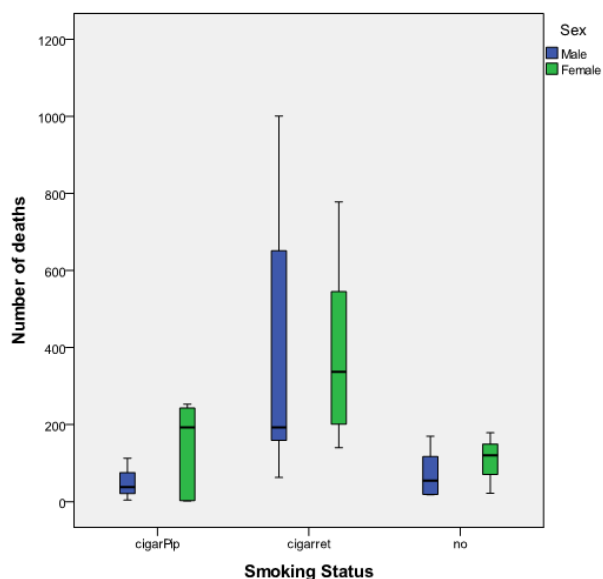
Clustered

Για τη δημιουργία γραφήματος με την απεικόνιση και δεύτερης ποιοτικής (ή ποσοτικής διακριτής) μεταβλητής, ορίζουμε στο πεδίο *Define Clusters by* τη δεύτερη ποιοτική (ή ποσοτική διακριτή) μεταβλητή.



Εικόνα 4.101: Η κεντρική ενότητα της επιλογής *Boxplot > Clustered*

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα:

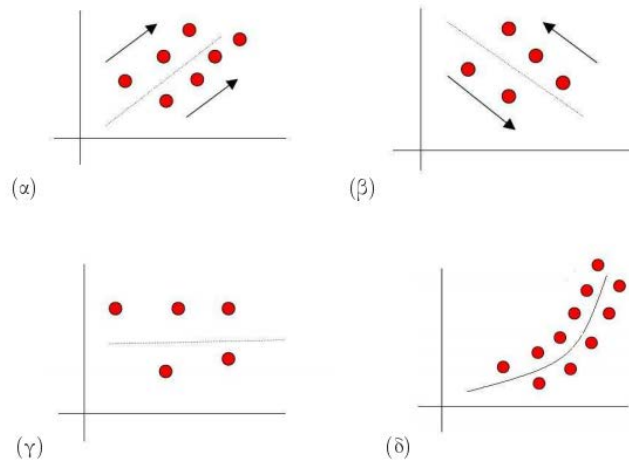


Εικόνα 4.102: Αποτελέσματα από την εντολή *Boxplot > Clustered*

4.15 Διασποράς - Scatter/ Dot

Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > Scatter / Dot*, για την δημιουργία διαγράμματος σημείων ή διασποράς.

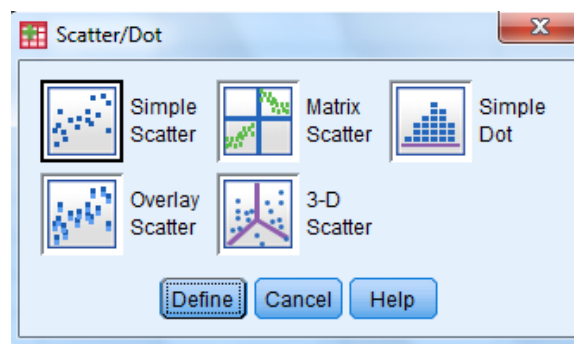
Το διάγραμμα σημείων ονομάζεται και διάγραμμα διασποράς και αποτελεί μια γραφική απεικόνιση του τρόπου συσχέτισης μεταξύ ποσοτικών μεταβλητών, περιγράφοντας τη δυσδιάστατη κατανομή τους. Κάθε σημείο απεικονίζει ένα ζεύγος παρατηρήσεων των δύο μεταβλητών. Με το διάγραμμα διασποράς εντοπίζονται εύκολα συσχετίσεις και ακραίες τιμές. Η απλή συσχέτιση ασχολείται με τον βαθμό με τον οποίο τα σημεία συγκεντρώνονται γύρω από μια ευθεία χωρίς να προσδιορίζεται ποια είναι ακριβώς αυτή η γραμμή που διέρχεται μέσα από το νέφος των σημείων. Γραμμική συσχέτιση εμφανίζεται όταν τα σημεία όλων των παρατηρήσεων τείνουν να συγκεντρωθούν γύρω από μια ευθεία και μη γραμμική όταν τα σημεία τείνουν να συγκεντρωθούν γύρω από μια καμπύλη. Θετική συσχέτιση όταν δύο μεταβλητές τείνουν να μεταβάλλονται προς την ίδια κατεύθυνση (σχήμα α, Εικόνα 4.62), σε αυτή την περίπτωση οι τιμές αυξομειώνονται ταυτόχρονα. Αρνητική συσχέτιση όταν οι δύο μεταβλητές τείνουν να μεταβάλλονται προς την αντίθετη κατεύθυνση (σχήμα β, Εικόνα 4.62), και μηδενική συσχέτιση όταν οι μεταβολές των τιμών της μιας ποσοτικής μεταβλητής δεν σχετίζονται με τις τιμές της άλλης ποσοτικής μεταβλητής. Σε αυτή την περίπτωση τα σημεία του διαγράμματος είναι διασκορπισμένα σε όλο το μήκος του διαγράμματος (σχήμα γ, Εικόνα 4.62). Μη γραμμική συσχέτιση έχουμε όταν οι μεταβολές των τιμών της μιας μεταβλητής συνδέονται με τις μεταβολές της άλλης με μη γραμμική μορφή. Τα σημεία του νέφους τείνουν να συγκεντρωθούν γύρω από μια καμπύλη (σχήμα δ - Εικόνα 4.62).



Εικόνα 4.103: Γενικές μορφές διαγραμμάτων διασποράς

Οι μορφές διαγραμμάτων διασποράς που προσφέρονται από την εφαρμογή είναι:

- *Simple Scatter*: Για τη δημιουργία απλού διαγράμματος διασποράς μεταξύ δύο ποσοτικών μεταβλητών.
- *Matrix Scatter*: Όταν έχουμε περισσότερες από δύο μεταβλητές, οπότε δημιουργούμε τα ανά δύο διαγράμματα διασποράς τους.
- *Simple Dot*: Μοιάζει με το ιστόγραμμα και χρησιμοποιείται για την απεικόνιση της κατανομής μιας ποσοτικής μεταβλητής. Τα δεδομένα είναι ομαδοποιημένα, αλλά αντί για μια τιμή σε κάθε κλάση όλες οι παρατηρήσεις της κάθε κλάσης εμφανίζονται στοιβαγμένες. Τα διαγράμματα αυτά ονομάζονται και διαγράμματα πυκνότητας.
- *Overlay Scatter*.
- *3-D Scatter*: Για τη δημιουργία τρισδιάστατου γραφήματος.

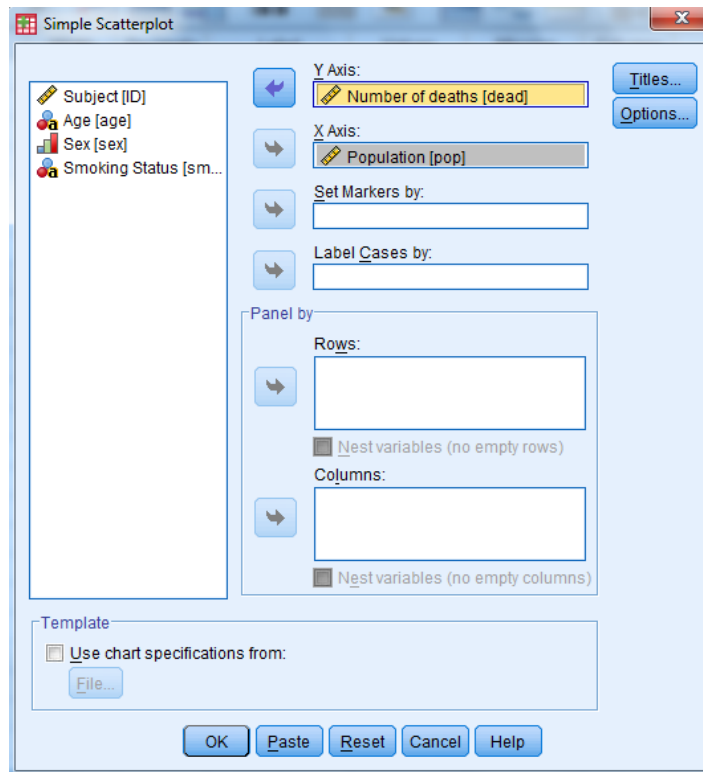


Εικόνα 4.104: Το αρχικό menu Scatter / Dot

Συνεχίζουμε στην επόμενη ενότητα με το πλήκτρο *Define*.

Simple

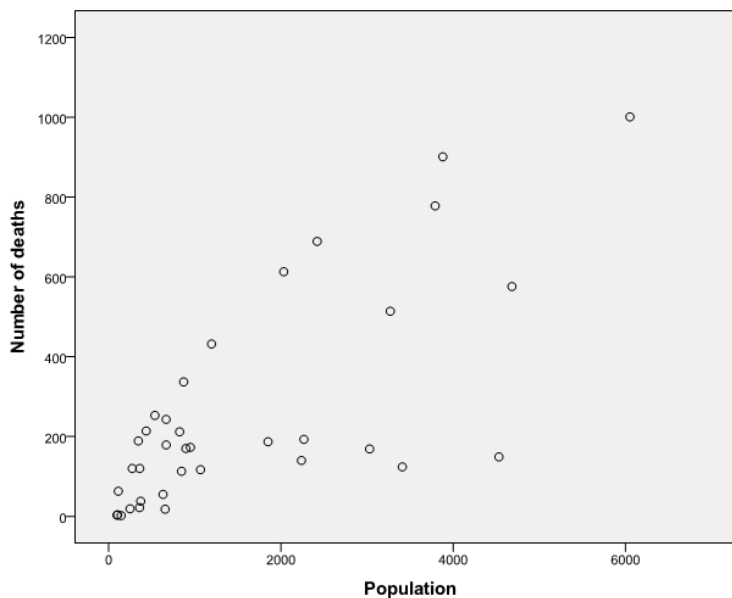
Στα πεδία *Y Axis* και *X Axis* ορίζουμε τις δύο ποσοτικές μεταβλητές.



Εικόνα 4.105: Η κεντρική ενότητα της επιλογής Scatter / Dot > Simple

Δεν είναι υποχρεωτικό, αλλά στο πεδίο *Label Cases by* μπορούμε να δηλώσουμε τη μεταβλητή που ονοματίζει τις παρατηρήσεις μας. Αν δεν δηλώσουμε κάποια μεταβλητή τότε οι ακραίες παρατηρήσεις μπορούν να εμφανιστούν με την αριθμητική τιμή τους.

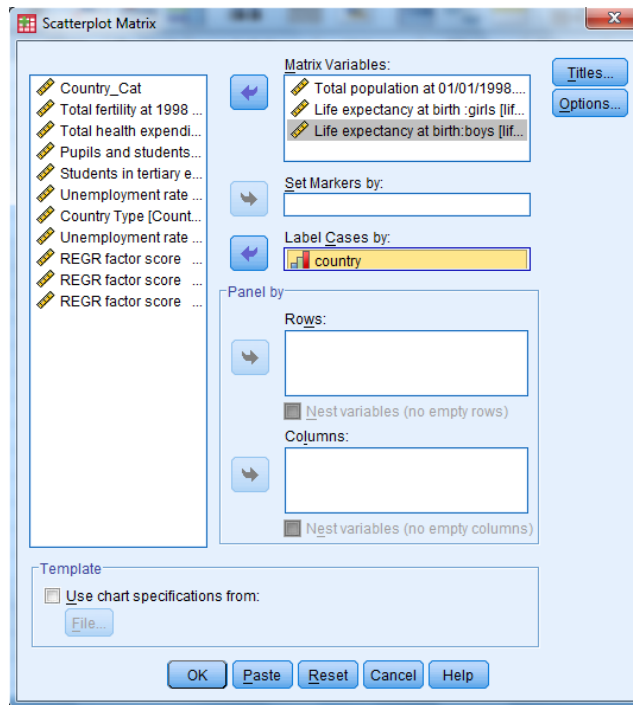
Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα.



Εικόνα 4.106: Αποτελέσματα από την εντολή Scatterplot > Simple

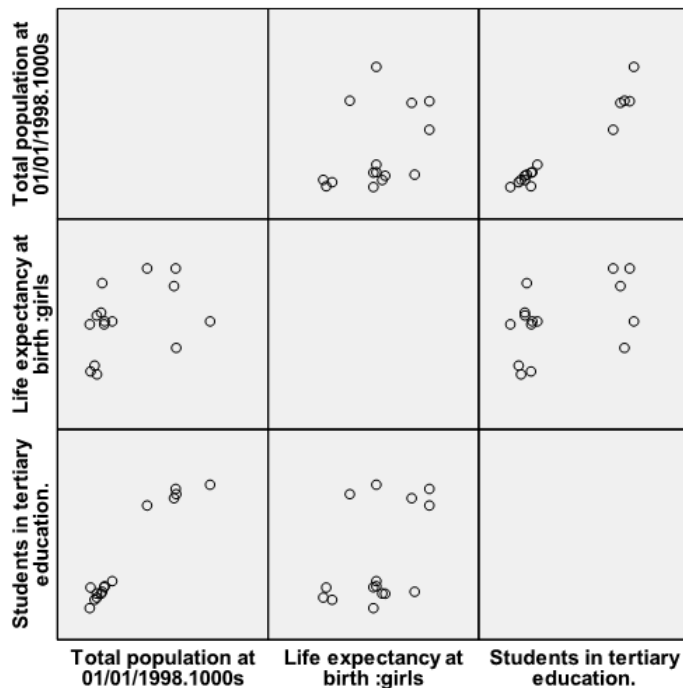
Matrix Scatter

Στο πλαίσιο *Matrix Variables* ορίζουμε τις ποσοτικές μεταβλητές για τις οποίες θα δημιουργηθούν ανά δύο τα διαγράμματα διασποράς.



Εικόνα 4.107: Η κεντρική ενότητα της επιλογής Scatterplot > Matrix Scatter

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα.



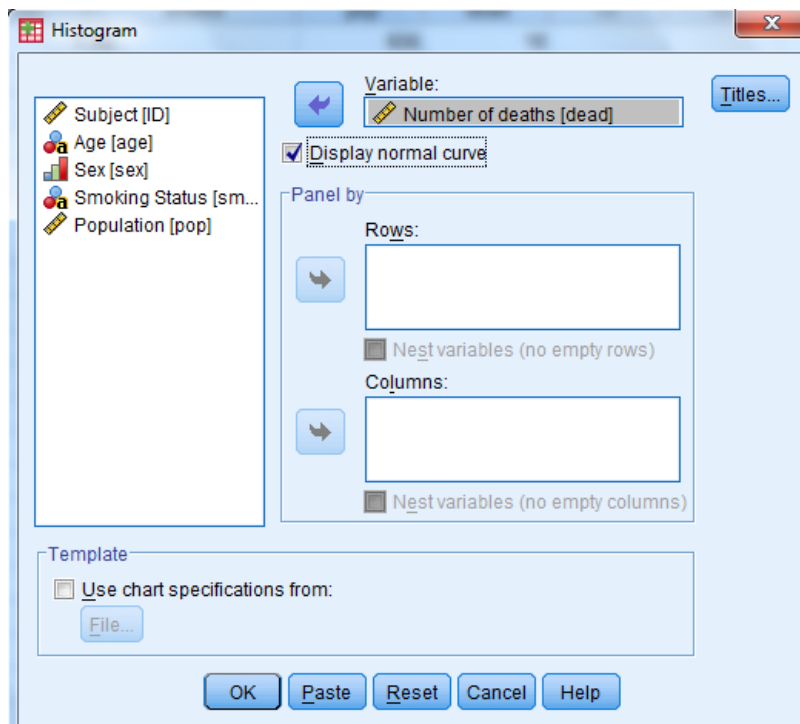
Εικόνα 4.108: Αποτελέσματα από την εντολή Scatterplot > Matrix Scatter

4.16 Ιστογράμματα - Histogram

Η δυνατότητα προσφέρεται από την επιλογή *Graphs > Legacy Dialogs > Histogram*, για τη δημιουργία ιστογράμματος. Το ιστόγραμμα απεικονίζει την κατανομή μιας ποσοτικής μεταβλητής με τη βοήθεια ράβδων. Κάθε ράβδος αντιστοιχεί σε ένα διάστημα τιμών και το ύψος είναι ανάλογο της συχνότητας των παρατηρήσεων που περιέχονται σ' αυτό. Συχνά απεικονίζουμε και τη γραμμή της κανονικής κατανομής για να

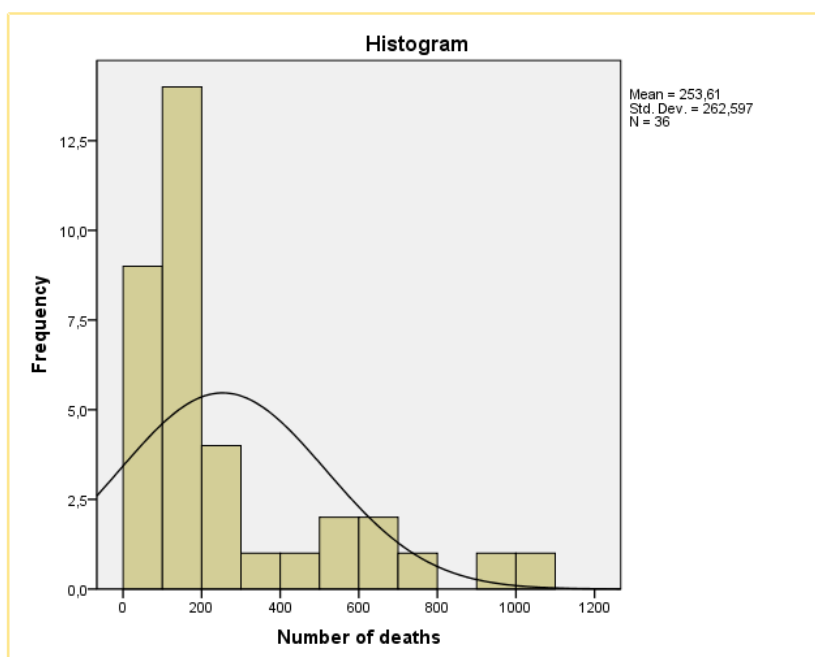
συγκρίνουμε εάν η κατανομή της ποσοτικής μας μεταβλητής είναι κανονική ή όχι. Αντίθετα με το ραβδόγραμμα στο οποίο ο x άξονας παριστάνει κατηγορίες, ο οριζόντιος άξονας ενός ιστογράμματος είναι μια αριθμητική κλίμακα. Κατασκευάζεται βασισμένο στον πίνακα συχνοτήτων ομαδοποιημένων παρατηρήσεων.

Στο πεδίο *Variable* ορίζουμε την ποσοτική μεταβλητή, και επιλέγοντας την παράμετρο *Display normal curve* δηλώνουμε την εμφάνιση της καμπύλης της κανονικής κατανομής.



Εικόνα 4.109: Η κεντρική ενότητα της επιλογής Histogram

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε το γράφημα.



Εικόνα 4.110: Αποτελέσματα από την εντολή Histogram

Βιβλιογραφικές Αναφορές Κεφαλαίου

Ελληνόγλωσσες

Howitt, D., Cramer, D. (Επιμέλεια: Σ. Κοντάκος) (2011). *Στατιστική με το SPSS 16.0*. Αθήνα: Εκδόσεις Κλειδάριθμος.

Ξενόγλωσσες

Carver, R. & Nash, J. (2011). *Doing Data Analysis with SPSS: Version 18.0*. 5th Edition Easton Cengage Learning.

Coakes, S. J. & Steed, L. G. (1999). *SPSS: Analysis without anguish: Versions 7.0, 7.5, 8.0 for Windows*. Brisbane: Jacaranda Wiley.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London:Sage.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, pp. 799-821.

Marques de Sa, J. P. (2007). *Applied Statistics Using SPSS. STATISTICA, MATLAB and R*, 2nd Edition, Porto: Universidade de Porto

Norusis Marija, J. (2002). *SPSS 11.0 Guide to data analysis*. Upper Saddle River New Jersey: Prentice Hall.

Κεφάλαιο 5

Βασικές έννοιες ελέγχων υποθέσεων και έλεγχοι κανονικότητας

Σύνοψη

Βασικές έννοιες και ορισμοί του ελέγχου υποθέσεων, γραφικοί έλεγχοι κανονικότητας μέσω των ιστογραμμάτων (διαδρομές *Analyze > Descriptive Statistics > Frequencies > Charts*, επιλέγοντας την παράμετρο Show normal curve on Histogram, *Analyze Descriptive Statistics > Explore > Plots*, επιλέγοντας την παράμετρο Histogram, *Graphs > Legacy Dialogs > Histogram*), των P-P plots (διαδρομές *Analyze > Descriptive Statistics > P-P Plots*, *Analyze > Descriptive Statistics > Explore > Plots*, επιλέγοντας την παράμετρο Normality plots with tests), των Q-Q Plots (διαδρομές *Analyze > Descriptive Statistics > Q-Q Plots*, *Analyze > Descriptive Statistics > Explore > Plots*). Στατιστικοί έλεγχοι κανονικότητας α) One-Sample Kolmogorov-Smirnov (K-S) Test (μέσω των διαδρομών *Analyze > Descriptive Statistics > Explore > Plots* επιλέγοντας την παράμετρο Normality plots with tests και *Analyze > Nonparametric tests > Legacy Dialogs > 1 sample K-S*), β) Shapiro-Wilk Test μέσω της διαδρομής *Analyze > Descriptive Statistics > Explore > Plots* επιλέγοντας την παράμετρο Normality plots with tests).

Προαπαιτούμενη γνώση

Καλύτερη κατανόηση του κεφαλαίου προκύπτει αν ο αναγνώστης έχει καλή θεωρητική γνώση της κανονικής κατανομής. Τα εγχειρίδια που παρουσιάζουν με πληρότητα την ύλη του συγκεκριμένου κεφαλαίου είναι αυτά που αναφέρθηκαν στο κεφάλαιο 3.

5.1 Εισαγωγή

Στατιστική Συμπερασματολογία είναι ο κλάδος της Στατιστικής που έχει ως σκοπό τη εξαγωγή νόμων, κανόνων και συμπερασμάτων τα οποία ξεπερνούν το επίπεδο των παρατηρήσεων, και γενικεύονται στο σύνολο του πληθυσμού.

Γενίκευση κάνουμε όταν ξεκινώντας από την πληροφορία που δίνει ένα ή περισσότερα αντιπροσωπευτικά δείγματα από κάποιον πληθυσμό, διατυπώνουμε μια πρόταση-υπόθεση (στατιστική υπόθεση) για τον πληθυσμό. Έτσι στο παρόν κεφάλαιο, αφήνουμε κατά μέρος την περιγραφική στατιστική ανάλυση και μπαίνουμε στον χώρο της συμπερασματολογίας (επαγωγής).

Προσοχή: Για να ισχύει η γενίκευση θα πρέπει το δείγμα στο οποίο θα βασιστούμε για την εξαγωγή συμπερασμάτων, να είναι όσο το δυνατόν πιο αντιπροσωπευτικό του πληθυσμού που μελετάμε και φυσικά να ισχύουν οι προϋποθέσεις εφαρμογής της εκάστοτε μεθοδολογίας.

5.2 Βασικές Έννοιες και Ορισμοί

Η στατιστική υπόθεση (statistical hypothesis) μπορεί να είναι μια οποιαδήποτε «στατιστική» πρόταση (για περιγραφικά μέτρα, κατανομές πληθυσμών, στοχαστικές διαδικασίες, κ.λπ.) την ορθότητα ή μη της οποίας θέλουμε να εξετάσουμε με βάση τις παρατηρήσεις που διαθέτουμε.

Έστω ότι μελετάμε μια παράμετρο θ για την οποία υποθέτουμε ότι μπορεί να λάβει μια συγκεκριμένη τιμή. Αυτή την υπόθεση την ονομάζουμε μηδενική υπόθεση (null hypothesis) και τη συμβολίζουμε με H_0 .

Έτσι στο παράδειγμά μας η μηδενική υπόθεση διαμορφώνεται ως $H_0: \theta = \theta_0$. Σε περίπτωση που η H_0 απορριφθεί τότε θεωρούμε ότι ισχύει η εναλλακτική υπόθεση (alternative hypothesis), η οποία συμβολίζεται ως H_1 και η οποία δύναται να λάβει μια από τις ακόλουθες μορφές: $H_1: \theta \neq \theta_0$, η παράμετρος να είναι διάφορη από την τιμή θ_0 (δίπλευρος έλεγχος).

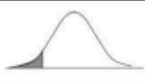


$H_1: \theta < \theta_0$, η παράμετρος να είναι μικρότερη από την τιμή θ_0 (μονόπλευρος έλεγχος).

$H_1: \theta > \theta_0$, η παράμετρος να είναι μεγαλύτερη από την τιμή θ_0 (μονόπλευρος έλεγχος).

Η επιλογή μονόπλευρου ή δίπλευρου ελέγχου εξαρτάται από το αντικείμενο που μελετάμε.

Τιμές της παραμέτρου θ «κοντά» στη θ_0 υποστηρίζουν την ορθότητα της H_0 , ενώ τιμές της παραμέτρου θ «μακριά» από τη θ_0 δεν την υποστηρίζουν. Έτσι, χωρίζουμε τις δυνατές τιμές της παραμέτρου θ σε αυτές για τις οποίες η H_0 απορρίπτεται, και αυτές για τις οποίες δεν μπορούμε να απορρίψουμε την H_0 . Η περιοχή όπου η H_0 απορρίπτεται ονομάζεται περιοχή απόρριψης (rejection region) και συμβολίζεται με R .

Η απόφαση για την απόρριψη ή μη της H_0 γίνεται βάση πιθανοτήτων γι' αυτόν τον λόγο ορίζουμε επίπεδο εμπιστοσύνης $(1-\alpha)$ για την απόφαση ελέγχου. Το α λέγεται επίπεδο σημαντικότητας (significance level) και καθορίζει τη διαχωριστική γραμμή μεταξύ των περιοχών απόρριψης ή μη. Όταν το α μικραίνει, μικραίνει και η περιοχή απόρριψης R . Η πιο συνήθης τιμή για το α είναι $\alpha = 5\%$. Το επίπεδο σημαντικότητας που υπολογίζεται από τα δεδομένα μας και το οποίο συγκρίνουμε με το οριζόμενο επίπεδο σημαντικότητας α ονομάζεται p -value.

One-Tail Test (left tail)	Two-Tail Test	One-Tail Test (right tail)
$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$
$H_1 : \mu < \mu_0$	$H_1 : \mu \neq \mu_0$	$H_1 : \mu > \mu_0$
		

Εικόνα 5.111: Έλεγχοι Υποθέσεων για τη μέση τιμή ενός πληθυσμού. Δεξιά και αριστερά είναι οι μονόπλευροι έλεγχοι

Με βάση την H_0 προσδιορίζουμε τη «στατιστική έλεγχο» (test statistic) της εκάστοτε υπόθεσης μας. Οι στατιστικοί έλεγχοι διακρίνονται σε δύο κατηγορίες: τους παραμετρικούς, που βασίζονται σε ελεγχοσυναρτήσεις με γνωστή κατανομή, και τους μη παραμετρικούς που βασίζονται σε άλλες ιδιότητες της παραμέτρου.

Προσοχή: Οι παραμετρικοί έλεγχοι υποθέσεων βασίζονται σε γνωστές κατανομές, οι οποίες φέρουν τις ιδιότητες της κανονικής κατανομής, συνεπώς για να αποφανθούμε εάν θα χρησιμοποιηθεί παραμετρικός ή μη παραμετρικός έλεγχος, αρχικά εξετάζουμε αν οι παρατηρήσεις μας ακολουθούν κανονική κατανομή.

Συνοπτικά τα στάδια ενός στατιστικού ελέγχου είναι:

- Ορίζεται η μηδενική υπόθεση H_0 .
- Ορίζεται η εναλλακτική υπόθεση H_1 .
- Ορίζεται το επίπεδο σημαντικότητας α .
- Επιλέγεται η κατάλληλη ελεγχοσυνάρτηση.
- Ορίζεται η περιοχή απόρριψης R της H_0 .

Σε έναν έλεγχο υπόθεσης υπάρχει περίπτωση να γίνουν δύο ειδών σφάλματα.

1. Ως Σφάλμα Τύπου I ορίζεται η πιθανότητα απόρριψης της μηδενικής υπόθεσης H_0 , ενώ αυτή είναι σωστή. Η πιθανότητα του σφάλματος Τύπου I συμβολίζεται με α .

$$\alpha = P [\text{απόρριψη της } H_0 \mid H_0 \text{ είναι σωστή}]$$

Εξίσωση 5.1

2. Ως Σφάλμα Τύπου II ορίζεται η πιθανότητα αποδοχής της μηδενικής υπόθεσης H_0 , ενώ αυτή είναι λανθασμένη. Η πιθανότητα του σφάλματος Τύπου II συμβολίζεται με β .

$$\beta = P[\text{αποδοχή της } H_0 \mid H_0 \text{ είναι λάθος}]$$

Εξίσωση 5.2

Ισχύς (power) ενός στατιστικού ελέγχου ονομάζεται η πιθανότητα απόρριψης της H_0 , ενώ είναι πράγματι ψευδής, δηλ. είναι η πιθανότητα να αποφύγουμε ένα σφάλμα Τύπου II.

$$\gamma = 1 - \beta = P[\text{απόρριψη της } H_0 \mid H_0 \text{ είναι λάθος}]$$

Εξίσωση 5.3

Επίπεδο σημαντικότητας ενός ελέγχου ονομάζεται η πιθανότητα να παρατηρηθεί μια τιμή του στατιστικού μεγαλύτερη από αυτήν που έδωσε το δείγμα των παρατηρήσεων.

$$p\text{-value} = P[T(Y) \text{ πιο ακραία από την } T(Y_{obs}) \mid \theta, H_0]$$

Εξίσωση 5.4

5.3 Γραφικοί Έλεγχοι Κανονικότητας

Για να αποφασιστεί εάν θα χρησιμοποιηθεί παραμετρικός ή μη παραμετρικός έλεγχος, αρχικά εξετάζουμε αν οι παρατηρήσεις μας ακολουθούν κανονική κατανομή. Ο έλεγχος κανονικότητας δύναται να πραγματοποιηθεί είτε γραφικά είτε μέσω στατιστικών ελέγχων.

Η υπόθεση που εξετάζεται είναι εάν η κατανομή των δεδομένων είναι η κανονική κατανομή, ή όχι:

$$H_0 : f(x) = N(\mu, \sigma^2)$$

$$H_1 : f(x) \neq N(\mu, \sigma^2)$$

Εξισώσεις 5.5

Γραφικά, ο έλεγχος κανονικότητας πραγματοποιείται με τη χρήση ιστογράμματος, ή διαγραμμάτων P-P και Q-Q plots. Αναλυτικότερα:

5.3.1 Ιστογράμματα

Η παραγωγή ιστογραμμάτων έχει ήδη αναλυθεί σε προηγούμενες παραγράφους, μέσω των επιλογών:

- *Analyze > Descriptive Statistics > Frequencies > Charts*, επιλέγοντας την παράμετρο Show normal curve on Histogram.
- *Analyze > Descriptive Statistics > Explore > Plots*, επιλέγοντας την παράμετρο Histogram.
- *Graphs > Legacy Dialogs > Histogram*.

5.3.2 P-P Plots

Η δυνατότητα προσφέρεται από τις επιλογές:

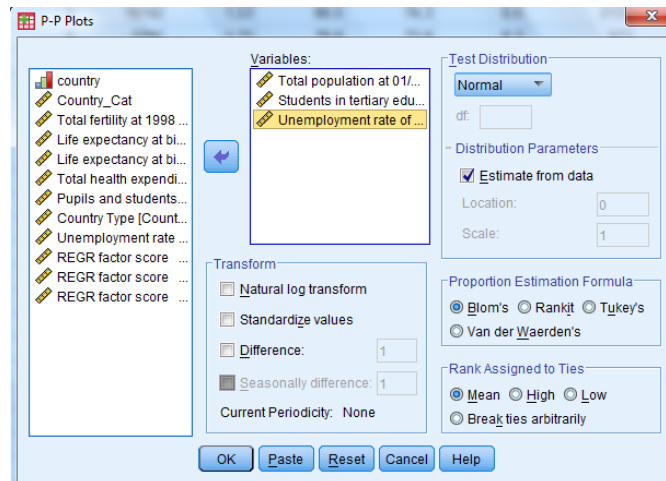
- *Analyze > Descriptive Statistics > P-P Plots*.
- *Analyze > Descriptive Statistics > Explore > Plots*, επιλέγοντας την παράμετρο Normality plots with tests με σκοπό τον έλεγχο κανονικότητας μιας ποσοτικής μεταβλητής.

Το P-P plot (probability-probability plot or percent-percent plot) είναι ένα γράφημα πιθανότητας για την αξιολόγηση του πόσο στενά συμφωνούν δύο σύνολα δεδομένων, βασιζόμενες στις αθροιστικές τους συναρτήσεις κατανομής.

Από το γράφημα εξετάζουμε εάν τα δεδομένα μας συμπίπτουν ή τείνουν στην ευθεία γραμμή του γραφήματος. Εάν τα σημεία τείνουν προς την ευθεία γραμμή συνεπάγεται ότι τα δεδομένα ακολουθούν την κατανομή που έχει οριστεί στο Test Distribution.

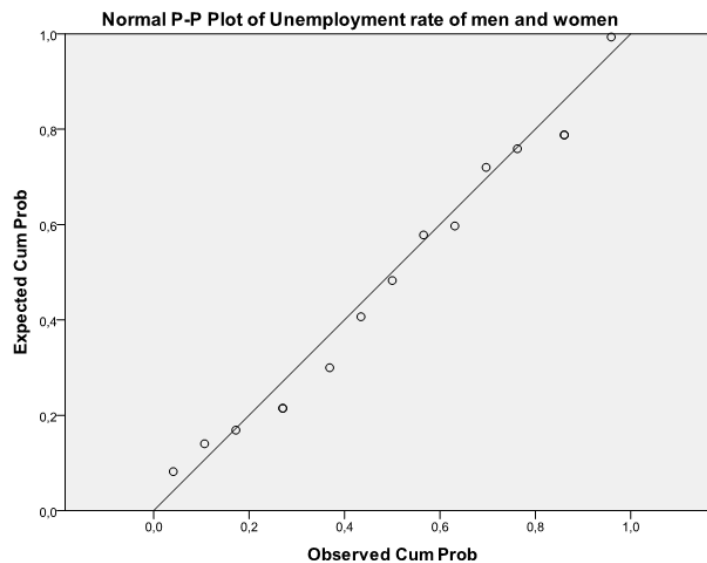
Στο πλαίσιο Variables επιλέγουμε τις μεταβλητές που θα εξεταστούν ως προς την κανονικότητά τους, επιλέγοντας στο πεδίο Test Distribution την κανονική (Normal) κατανομή.

Η επιλογή της κανονικής κατανομής είναι η προκαθορισμένη επιλογή της εφαρμογής, ωστόσο παρέχονται και πολλές άλλες γνωστές κατανομές όπως: οι Student, Pareto, Weibull, Uniform κ.ά.



Εικόνα 5.112: Το menu P-P plot

Η διαδικασία ολοκληρώνεται με το πλήκτρο OK όπου και λαμβάνουμε ένα γράφημα για κάθε μεταβλητή.



Εικόνα 5.113: Αποτελέσματα από την εντολή P-P plot

5.3.3 Q-Q Plots

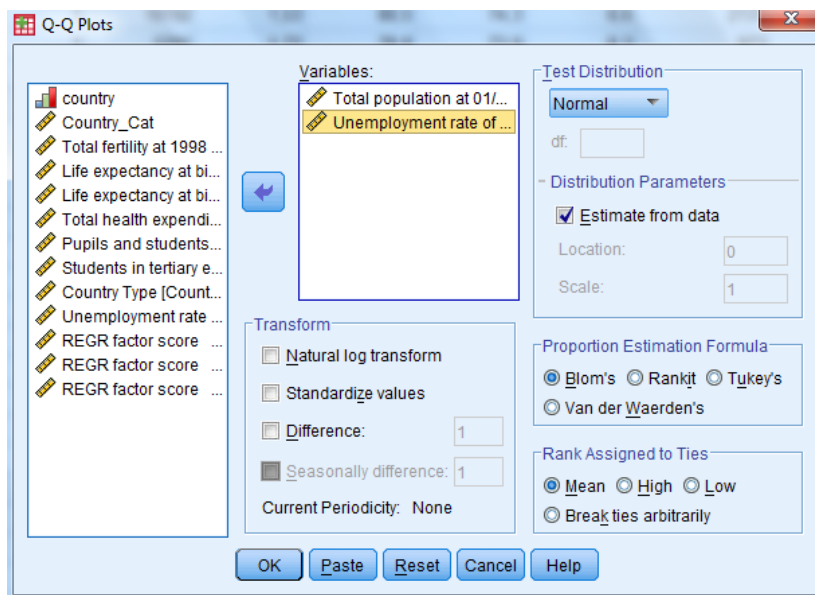
Η δυνατότητα προσφέρεται από τις επιλογές:

- *Analyze > Descriptive Statistics > Q-Q Plots.*
- *Analyze > Descriptive Statistics > Explore > Plots*, επιλέγοντας την παράμετρο Normality plots with tests με σκοπό τον έλεγχο κανονικότητας μιας ποσοτικής μεταβλητής.

Ένα Q-Q plot (το «Q» προέρχεται από την λέξη quantile) είναι ένα γράφημα πιθανότητας για τη γραφική σύγκριση δύο κατανομών πιθανότητας, απεικονίζοντας τα ποσοστημόρια της μιας σε σχέση με την άλλη.

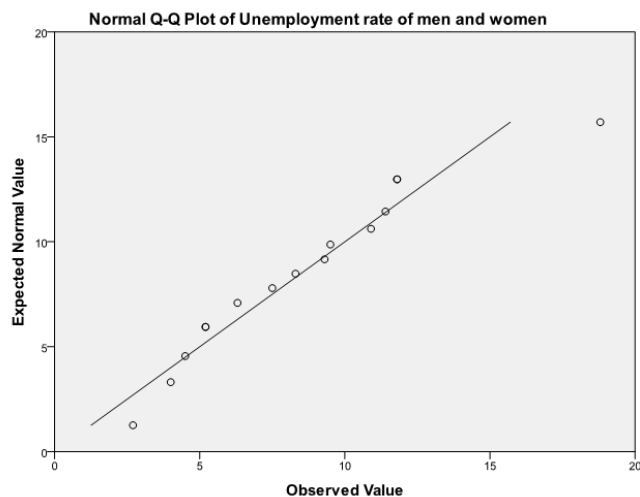
Το γράφημα Q-Q χρησιμοποιείται για να συγκρίνουμε τα σχήματα των συναρτήσεων κατανομής, παρέχοντας μια γραφική άποψη για το πώς ιδιότητες, όπως η θέση, η κλίμακα και η ασυμμετρία είναι παρόμοιες ή διαφορετικές στις δύο κατανομές. Επίσης, χρησιμοποιείται για να συγκρίνει τις συλλογές δεδομένων ή θεωρητικές κατανομές. Η χρήση του γραφήματος για τη σύγκριση δύο δειγμάτων δεδομένων μπορεί να θεωρηθεί ως μια μη-παραμετρική προσέγγιση για τη σύγκριση των κατανομών τους. Γενικότερα η χρήση γραφημάτων Q-Q είναι μια πιο ισχυρή προσέγγιση από την κοινή τεχνική της σύγκρισης των ιστογραμμάτων των δύο δειγμάτων.

Εάν οι δύο κατανομές είναι ίσες, τότε τα σημεία του Q-Q plot θα βρίσκονται στην ευθεία $y = x$. Εάν οι κατανομές έχουν γραμμική σχέση, τα σημεία του γραφήματος θα βρίσκονται περίπου σε μια γραμμή, αλλά όχι κατ' ανάγκη στην ευθεία $y = x$. Γενικότερα τα γραφήματα Q-Q είναι συχνά τοξοειδούς ή «S» μορφής, υποδεικνύοντας ότι η μία από τις κατανομές είναι πιο ασύμμετρη από την άλλη.



Εικόνα 5.114: Το menu Q-Q plot

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε ένα γράφημα για κάθε μεταβλητή.



Εικόνα 5.115: Αποτελέσματα από την εντολή Q-Q plot

5.4 Στατιστικοί Έλεγχοι Κανονικότητας

Για να αποφασιστεί αν θα χρησιμοποιηθεί παραμετρικός ή μη παραμετρικός έλεγχος, αρχικά εξετάζουμε την κανονικότητα των παρατηρήσεών μας.

Οι στατιστικοί έλεγχοι που παρέχονται για τον έλεγχο της κανονικότητας είναι αυτοί των Kolmogorov-Smirnov και των Shapiro-Wilk.

Η υπόθεση που εξετάζεται είναι αν η κατανομή των δεδομένων είναι η κανονική κατανομή, ή όχι.

5.4.1 One-Sample Kolmogorov-Smirnov (K-S) Test

Η δυνατότητα προσφέρεται από τις επιλογές :

- *Analyze > Descriptive Statistics > Explore > Plots*, επιλέγοντας την παράμετρο Normality plots with tests.
- *Analyze > Nonparametric tests > Legacy Dialogs > 1 sample K-S*, με σκοπό τον έλεγχο κανονικότητας μιας ποσοτικής μεταβλητής.

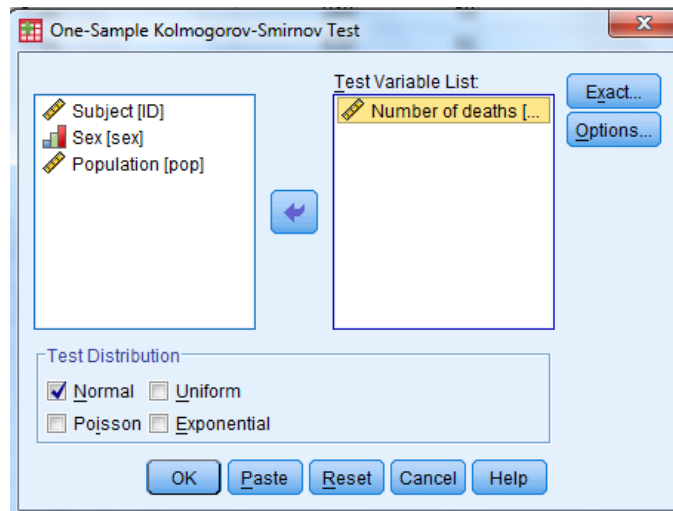
Το κριτήριο K-S είναι ένας μη παραμετρικός έλεγχος που χρησιμοποιείται για να εξετάσει την καλή προσαρμογή ενός τυχαίου δείγματος σε μία δεδομένη κατανομή, και βασίζεται στη διαφορά της εμπειρικής συνάρτησης κατανομής που προέρχεται από το δείγμα, και της αναμενόμενης συνάρτησης κατανομής υπό την υπόθεση της κανονικότητας, ή της οποιαδήποτε μηδενικής υπόθεσης H_0 η οποία μπορεί να είναι Κανονική, Ομοιόμορφη, Poisson ή Exponential.

Υποθέσεις του κριτηρίου:

Το κριτήριο Kolmogorov-Smirnov υποθέτει ότι οι παράμετροι της υπό έλεγχο κατανομής προσδιορίζονται εκ των προτέρων. Αυτή η προσέγγιση υπολογίζει τις παραμέτρους από το δείγμα. Για κάθε κατανομή οι παράμετροι που εκτιμώνται είναι:

Κατανομή	Παράμετρος
Κανονική κατανομή	Η δειγματική μέση τιμή και η δειγματική
Ομοιόμορφη κατανομή	τυπική απόκλιση
Poisson και Exponential κατανομή	Το εύρος (ελάχιστη και μέγιστη τιμή)
	Η δειγματική μέση τιμή

Πίνακας 5.1: Πίνακας που δείχνει τη σχέση κατανομής-παραμέτρου



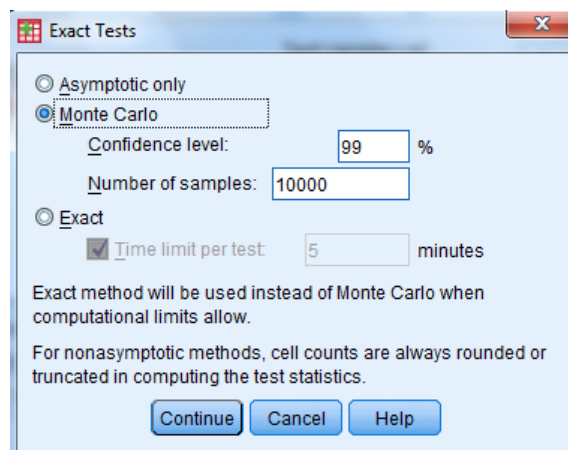
Εικόνα 5.116: Το menu One-Sample Kolmogorov-Smirnov Test

Στο πλαίσιο Test Variables List ορίζονται οι μεταβλητές για τις οποίες θα υπολογιστεί το κριτήριο κανονικότητας. Το κριτήριο υπολογίζεται ξεχωριστά για κάθε μεταβλητή.

Στο πλαίσιο Test Distribution επιλέγεται η κατανομή σύγκρισης. Οι δυνατές επιλογές είναι:

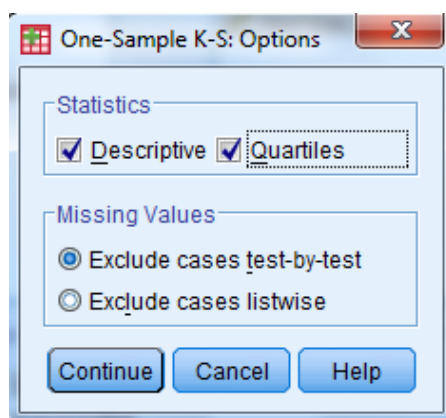
- Normal,
- Poisson,
- Uniform,
- Exponential.

Από το πλήκτρο Exact επιλέγοντας την παράμετρο Monte Carlo ζητάμε από την εφαρμογή να χρησιμοποιήσει την τεχνική της προσομοίωσης για τον έλεγχο της κανονικότητας, όπου διεξάγει 10.000 (προκαθορισμένη τιμή) ελέγχους κανονικότητας και για κάθε έναν υπολογίζει το p -value. Στο τέλος εμφανίζει τον μέσο όρο αυτών των 10.000 p -values και ένα 99% δ.ε. γι' αυτόν. Ωστόσο, η συνήθης επιλογή είναι η προκαθορισμένη Asymptotic only.



Εικόνα 5.117: Το menu One-Sample Kolmogorov-Smirnov Test -Options

Από το πλήκτρο Options παρέχεται η δυνατότητα εξαγωγής περιγραφικών μέτρων, τα ποσοστημόρια της κάθε μεταβλητής, και η διαχείριση των ελλειπουσών τιμών. Τα περιγραφικά μέτρα που παράγονται είναι η μέση τιμή, η τυπική απόκλιση, η ελάχιστη και η μέγιστη παρατήρηση και το πλήθος των έγκυρων παρατηρήσεων. Από τα ποσοστημόρια απεικονίζονται το 25^ο, το 50^ο, και το 75^ο ποσοστημόριο, δηλ. το 1^ο, 2^ο (διάμεσος) και 3^ο τεταρτημόριο.



Εικόνα 5.118: Το menu One-Sample Kolmogorov-Smirnov Test-Options

Η διαδικασία ολοκληρώνεται με το πλήκτρο OK όπου και λαμβάνουμε τους ακόλουθους πίνακες.

Descriptive Statistics								
	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Number of deaths	36	253,61	262,597	2	1001	75,50	171,50	316,00

Εικόνα 5.119: Περιγραφικά μέτρα από την εντολή One-Sample Kolmogorov-Smirnov Test

One-Sample Kolmogorov-Smirnov Test			Number of deaths
N			36
Normal Parameters ^{a, b}	Mean		253,61
	Std. Deviation		262,597
Most Extreme Differences	Absolute		,254
	Positive		,254
	Negative		-,169
Kolmogorov-Smirnov Z			1,526
Asymp. Sig. (2-tailed)			,019

a. Test distribution is Normal.

b. Calculated from data.

Εικόνα 5.120: Αποτελέσματα από την εντολή One-Sample Kolmogorov-Smirnov Test

Από τα αποτελέσματά μας διαπιστώνουμε ότι η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας 5%, συνεπώς η μεταβλητή δεν ακολουθεί κανονική κατανομή.

5.4.2 Shapiro-Wilk Test

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Descriptive Statistics > Explore > Plots* επιλέγοντας την παράμετρο Normality plots with tests, με σκοπό τον έλεγχο κανονικότητας μιας ποσοτικής μεταβλητής.

Το κριτήριο Shapiro-Wilk είναι ένας ακόμα πολύ γνωστός μη παραμετρικός έλεγχος για το αν οι παρατηρήσεις μίας μεταβλητής προέρχεται από κανονική κατανομή.

Επιλέγοντας την παράμετρο Normality plots with tests, όταν οι τιμές της μεταβλητής είναι σταθμισμένες και οι τιμές των σταθμίσεων είναι μη αέριαιες, τότε το κριτήριο των Shapiro-Wilk

υπολογίζεται όταν το σταθμισμένο (weighted) μέγεθος του δείγματος είναι μεταξύ 3-50. Εάν οι τιμές της μεταβλητής δεν είναι σταθμισμένες ή οι σταθμίσεις είναι ακέραιες τιμές, τότε το κριτήριο υπολογίζεται για μέγεθος δείγματος μεταξύ 3 και 5.000.

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε τον ακόλουθο πίνακα:

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Number of deaths	,254	36	,000	,807	36	,000

a. Lilliefors Significance Correction

Εικόνα 5.121: Αποτελέσματα από την εντολή Explore-Normality plots with tests

Από τα αποτελέσματα διαπιστώνουμε, ομοίως με τον προηγούμενο έλεγχο, ότι η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας 5%, συνεπώς η μεταβλητή δεν ακολουθεί κανονική κατανομή.

Βιβλιογραφικές Αναφορές Κεφαλαίου

Ελληνόγλωσσες

Howitt, D., Cramer, D. (Επιμέλεια: Σ. Κοντάκος) (2011). *Στατιστική με το SPSS 16.0*. Αθήνα: Εκδόσεις Κλειδάριθμος.

Ξενόγλωσσες

Carver, R. & Nash, J. (2011). *Doing Data Analysis with SPSS: Version 18.0*. 5th Edition Easton Cengage Learning.

Coakes, S. J. & Steed, L. G. (1999). *SPSS: Analysis without anguish: Versions 7.0, 7.5, 8.0 for Windows*. Brisbane: Jacaranda Wiley.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London:Sage.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, pp. 799-821.

Marques de Sa, J. P. (2007). *Applied Statistics Using SPSS. STATISTICA, MATLAB and R*, 2nd Edition, Porto: Universidade de Porto

Norusis Marija, J. (2002). *SPSS 11.0 Guide to data analysis*. Upper Saddle River New Jersey: Prentice Hall.

Κεφάλαιο 6

Παραμετρικοί και μη παραμετρικοί έλεγχοι υποθέσεων

Σύνοψη

Το κεφάλαιο αυτό παρουσιάζει τους γνωστότερους παραμετρικούς και μη παραμετρικούς ελέγχους θεωρητικά και στο πρόγραμμα. Συγκεκριμένα του ελέγχου: T-Test για τον έλεγχο Μέσης Τιμής Πληθυσμού, T-test για τον έλεγχο της διαφοράς δύο μέσων από ανεξάρτητα δείγματα, Mann-Whitney U, T-test για τον έλεγχο της διαφοράς δύο μέσων από συσχετιζόμενα δείγματα, Wilcoxon, Means, One-Way ANOVA για ανεξάρτητα δείγματα, Kruskal-Wallis, Two-Way ANOVA for Independent samples Friedman, χ^2 test, Mc Nemar test.

Σημειώνεται ότι τα δεδομένα των εφαρμογών είναι κατά πλειοψηφία δεδομένα που χρησιμοποιήθηκαν κατά τις παραδόσεις των αντίστοιχων μαθημάτων του ΠΜΣ «Βιοστατιστικής» των πανεπιστημίων Αθηνών και Ιωαννίνων. Για τις εφαρμογές των 6.2, 6.6. και 6.11 χρησιμοποιήθηκαν δεδομένα από την αντίστοιχη σελίδα του πανεπιστημίου του Princeton (βλ. βιβλιογραφικές αναφορές). Πολύτιμες πληροφορίες για τη συγκεκριμένη ύλη υπάρχουν στον αντίστοιχο ιστότοπο του προγράμματος (βλ. βιβλιογραφικές αναφορές). Τέλος για την παράγραφο 6.14 αξιόλογη πηγή ήταν η διπλωματική εργασία του Γ. Κουτσοχέρα (βλ. βιβλιογραφικές αναφορές).

Προαπαιτούμενη γνώση

Απαιτείται πολύ καλή γνώση του προηγούμενου κεφαλαίου καθώς και θεωρητική γνώση των κατανομών κατανομής και χ^2 , και εξοικείωση με τη διαδικασία ελέγχου υποθέσεων (έννοιες της ελεγχουσυνάρτησης και της περιοχής απόρριψης της μηδενικής υπόθεσης).

6.1 T-Test για τον έλεγχο Μέσης Τιμής Πληθυσμού

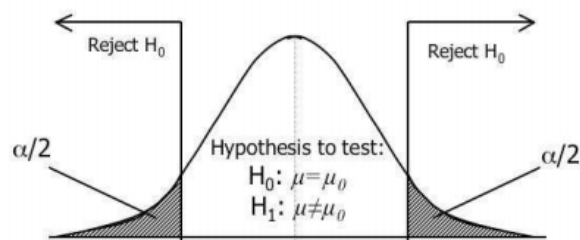
Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Compare Means > One-Sample T-test*, με σκοπό τον έλεγχο της ισότητας της μέσης τιμής ενός πληθυσμού με μια συγκεκριμένη τιμή μ_0 . Δηλαδή οι υποθέσεις που εξετάζονται είναι:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Εξίσωση 6.1

Η στατιστική συνάρτηση του ελέγχου διαμορφώνεται ανάλογα με το εάν το δείγμα μας προέρχεται από κανονικό πληθυσμό ή όχι, με γνωστή ή άγνωστη διασπορά. Η υπόθεση απορρίπτεται όταν η τιμή του κριτηρίου είναι κατά απόλυτη τιμή μεγαλύτερη από την τιμή που θα λάμβανε εάν ίσχυε η μηδενική υπόθεση και δίνεται από τους αντίστοιχους πίνακες της κανονικής, ή της t-student κατανομής.



Εικόνα 6.122: Περιοχές απόρριψης σε δίπλευρο έλεγχο

Μορφές της ελεγχοσυνάρτησης

1^η Περίπτωση. Διασπορά γνωστή:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Εξίσωση 6.2

Απορρίπτουμε την H_0 εάν η περιοχή απόρριψης είναι:

$$R = \{|z| > z_{\alpha/2}\}$$

Εξίσωση 6.3

2^η Περίπτωση. Διασπορά άγνωστη, $n \geq 30$:

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Εξίσωση 6.4

Απορρίπτουμε την H_0 εάν η περιοχή απόρριψης είναι:

$$R = \{|z| > z_{\alpha/2}\}$$

Εξίσωση 6.5

3^η Περίπτωση. Διασπορά άγνωστη, $n < 30$

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

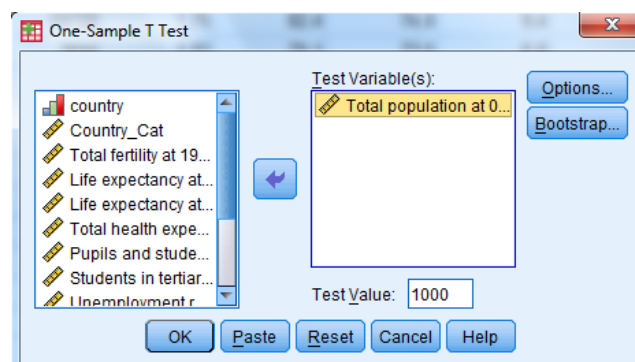
Εξίσωση 6.6

Απορρίπτουμε την H_0 εάν η περιοχή απόρριψης είναι:

$$R = \{|t| > t_{n-1; \alpha/2}\}$$

Εξίσωση 6.7

Στο περιβάλλον του SPSS, στο πεδίο *Test Variable(s)* εισάγουμε την ή τις συνεχείς μεταβλητές που θα εξεταστούν και στο πεδίο *Test Value* εισάγουμε την υπό εξέταση τιμή μ_0 .



Εικόνα 6.123: Το menu One-Sample T-test

Από το πλήκτρο *Options* ορίζουμε το επίπεδο σημαντικότητας για τη δημιουργία των διαστημάτων εμπιστοσύνης της διαφοράς των δύο μέσων (της μέσης τιμής και της υπό εξέτασης τιμής).

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Total population at 01/01/1998.1000s	15	24372,20	27121,193	7002,662

One-Sample Test						
	Test Value = 1000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Total population at 01/01/1998.1000s	3,338	14	,005	23372,200	8352,98	38391,42

Εικόνα 6.124: Περιγραφικά μέτρα της μεταβλητής και αποτελέσματα ελέγχου της ισότητας της μέσης τιμής με την τιμή 1000

Στον πρώτο πίνακα παρουσιάζονται το πλήθος παρατηρήσεων, η μέση τιμή, η τυπική απόκλιση και το τυπικό σφάλμα μέσης τιμής. Διαπιστώνεται ότι η μέση τιμή 24.372,2 διαφέρει από την υπό εξέταση τιμή $\mu_0 = 1.000$. Η διαφορά αυτή είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%, όπως επιβεβαιώνεται από τα αποτελέσματα του ελέγχου ($p\text{-value} = 0,005 < 0,05$).

6.2 T-test για τον έλεγχο της διαφοράς δύο μέσων από ανεξάρτητα δείγματα

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Compare Means > Independent-Samples T-test*, με σκοπό τον έλεγχο της συσχέτισης μιας συνεχούς μεταβλητής με μια δίτιμη κατηγορική μεταβλητή, στην οποία οι κατηγορίες αφορούν μετρήσεις που γίνονται σε διαφορετικά άτομα. Για παράδειγμα, εξετάζουμε το πλήθος θανάτων σε σχέση με το φύλο. Η μεταβλητή «Φύλο» είναι δίτιμη και οι μετρήσεις διενεργούνται σε διαφορετικά άτομα: άνδρες / γυναίκες.

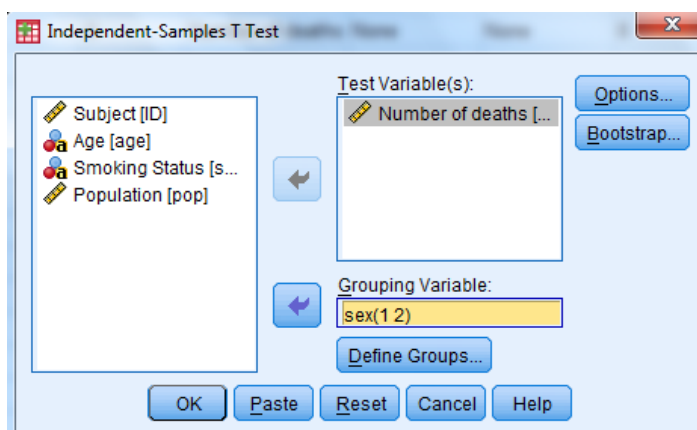
Στην πραγματικότητα η υπόθεση που εξετάζεται είναι η ισότητα των μέσων των δύο ανεξάρτητων ομάδων.

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Εξίσωση 6.8

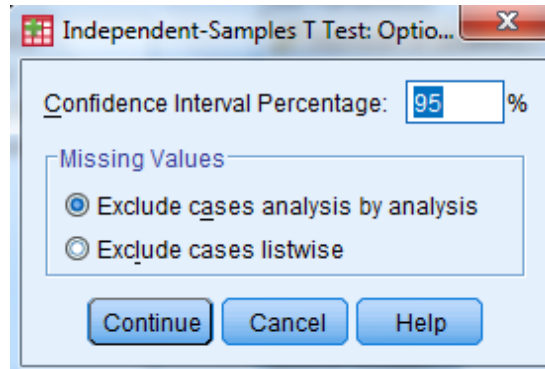
Εάν η υπόθεση της ισότητας δεν απορριφθεί τότε θα συνεπάγεται ότι οι δύο μέσοι είναι ίδιοι στις δύο ομάδες, συνεπώς η κατηγοριοποίηση δεν οδηγεί σε διαφορετικά αποτελέσματα, ή με άλλα λόγια η κατηγορική μεταβλητή δεν σχετίζεται με τη συνεχή μεταβλητή.



Εικόνα 6.125: Το menu Independent-Samples T-test

Στο πεδίο *Test Variable(s)* εισάγουμε την ή τις συνεχείς μεταβλητές που θα εξεταστούν, ενώ στο πεδίο *Grouping Variable* εισάγουμε την κατηγορική μεταβλητή. Καθώς ενδέχεται η κατηγορική που θα επιλεγεί να έχει περισσότερες από δύο κατηγορίες. Μέσω του πλήκτρου *Define Groups* ορίζουμε τους κωδικούς των δύο κατηγοριών που θα χρησιμοποιηθούν για την εφαρμογή του T-test.

Από το πλήκτρο *Options* ορίζουμε το επίπεδο σημαντικότητας για τη δημιουργία των διαστημάτων εμπιστοσύνης της διαφοράς των δύο μέσων $\mu_1-\mu_2$.



Εικόνα 6.126: Το menu Independent-Samples T-test > Options

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε δύο πίνακες.

Στον πρώτο πίνακα παρουσιάζονται τα περιγραφικά μέτρα (στο συγκεκριμένο παράδειγμα του πλήθους θανάτων) στις δύο ομάδες. Απ' αυτά φαίνεται ότι οι άνδρες έχουν κατά μέσο όρο μεγαλύτερο πλήθος θανάτων σε σχέση με τις γυναίκες.

Sex	N	Mean	Std. Deviation	Std. Error Mean
Number of deaths Male	19	266,05	308,119	70,687
Female	17	239,71	208,927	50,672

Εικόνα 6.127: Περιγραφικά μέτρα του πλήθους θανάτων στις δύο ομάδες

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Number of deaths	Equal variances assumed	2,962	,094	,297	34	,769	26,347	88,833	-154,184	206,877
	Equal variances not assumed			,303	31,804	,764	26,347	86,973	-150,855	203,549

Εικόνα 6.128: Αποτελέσματα του T-test για ανεξάρτητα δείγματα

Στον δεύτερο πίνακα λαμβάνουμε τα αποτελέσματα του ελέγχου. Στον πίνακα αυτόν εμφανίζονται δύο εγγραφές για τον έλεγχο T-test της ισότητας των δύο μέσων. Αυτό οφείλεται στο ότι για τον υπολογισμό χρησιμοποιούνται διαφορετικοί τύποι, αναλόγως εάν ισχύει ή όχι η ισότητα των διασπορών στις δύο ομάδες. Συνεπώς για να αποφανθούμε ποια αποτελέσματα θα χρησιμοποιήσουμε, ελέγχουμε αρχικά τον έλεγχο Levene για την ισότητα των διασπορών. Ο έλεγχος υποδεικνύει ότι η υπόθεση της ισότητας δεν μπορεί να απορριφθεί σε επίπεδο 5% ($p\text{-value} = 0,094 > 0,05$), συνεπώς θα χρησιμοποιηθούν τα αποτελέσματα του ελέγχου T-test από την πρώτη γραμμή, από όπου και διαπιστώνουμε ότι η υπόθεση της ισότητας δεν

απορρίπτεται ($p\text{-value} = 0,769 > 0,05$), άρα το μέσο πλήθος θανάτων μεταξύ ανδρών και γυναικών είναι σχεδόν το ίδιο.

6.3 Mann-Whitney U

Πρόκειται για τον αντίστοιχο μη παραμετρικό έλεγχο της ισότητας δύο μέσων μεταξύ ανεξάρτητων δειγμάτων, τα οποία όμως δεν ακολουθούν κανονική κατανομή. Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Non Parametric Tests > Independent-Samples*.

Το κριτήριο Mann-Whitney είναι ένας βαθμολογικός έλεγχος που εξετάζει την υπόθεση ότι το άθροισμα των βαθμών των θέσεων των παρατηρήσεων του ενός δείγματος είναι ίσο με το άθροισμα των βαθμών των θέσεων των παρατηρήσεων του άλλου δείγματος.

$$H_0 : W_A = W_B$$

$$H_1 : W_A \neq W_B$$

Εξίσωση 6.9

Εάν η υπόθεση ότι τα αθροίσματα είναι ίσα δεν απορριφθεί, τότε συνεπάγεται ότι η κατανομή των παρατηρήσεων της μιας ομάδας είναι σχεδόν ίδια με την κατανομή της άλλης και ότι οι παρατηρήσεις εμφανίζονται με τυχαίο τρόπο τόσο στη μια όσο και στην άλλη ομάδα. Συνεπώς, η διαφοροποίηση λόγω της κατηγορικής μεταβλητής δεν σχετίζεται με τη συνεχή μεταβλητή.

Η ελεγχουσυνάρτηση βασίζεται στη σχέση:

$$U = \min(U_A, U_B)$$

Εξίσωση 6.10

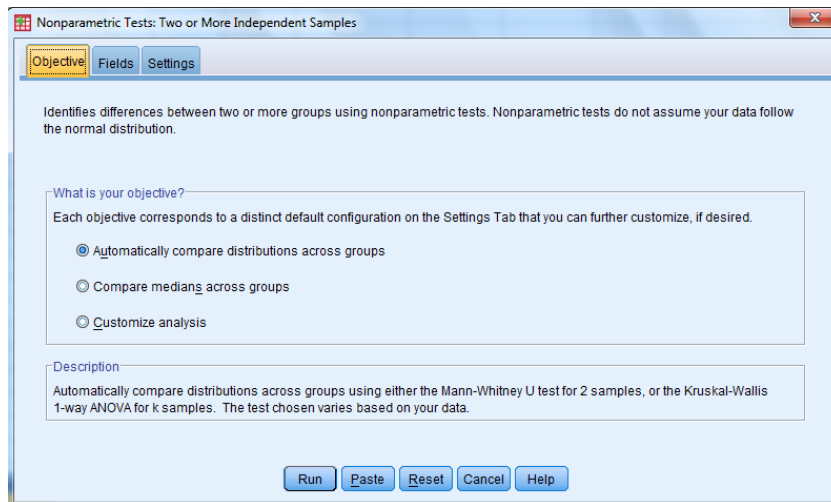
Όπου:

$$U_A = n_A n_B + \frac{n_A(n_A + 1)}{2} - W_A$$

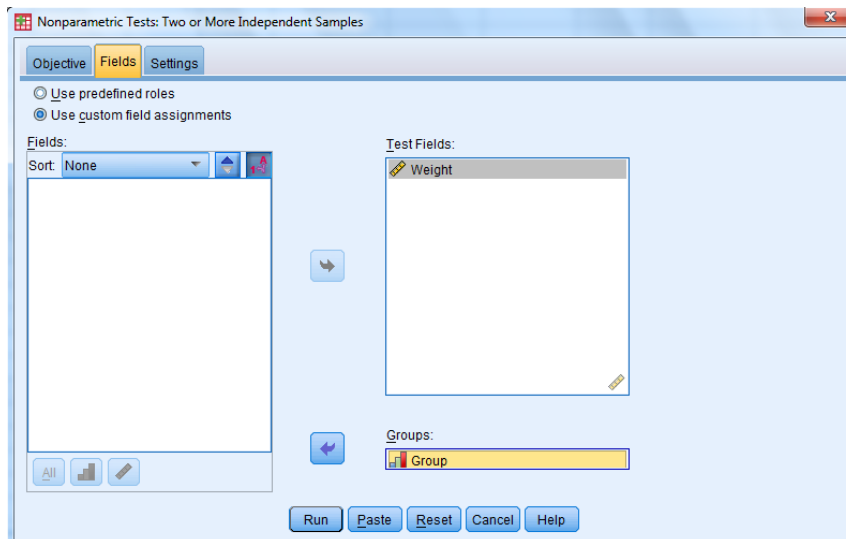
$$U_B = n_A n_B + \frac{n_B(n_B + 1)}{2} - W_B$$

Εξίσωση 6.11

Έστω ότι έχουμε δύο ομάδες ποντικών (ίδιας ηλικίας και φύλου) στα οποία μετά από τυχαιοποίηση χορηγήθηκε ένα πρόγραμμα διατροφής υψηλής (ομάδα Α) και χαμηλής (ομάδα Β) περιεκτικότητας πρωτεϊνών. Στο τέλος του προγράμματος μετρήθηκε το βάρος των ποντικών ώστε να εξεταστεί ποιο είδος διατροφής επιδρά στο βάρος των ποντικών.

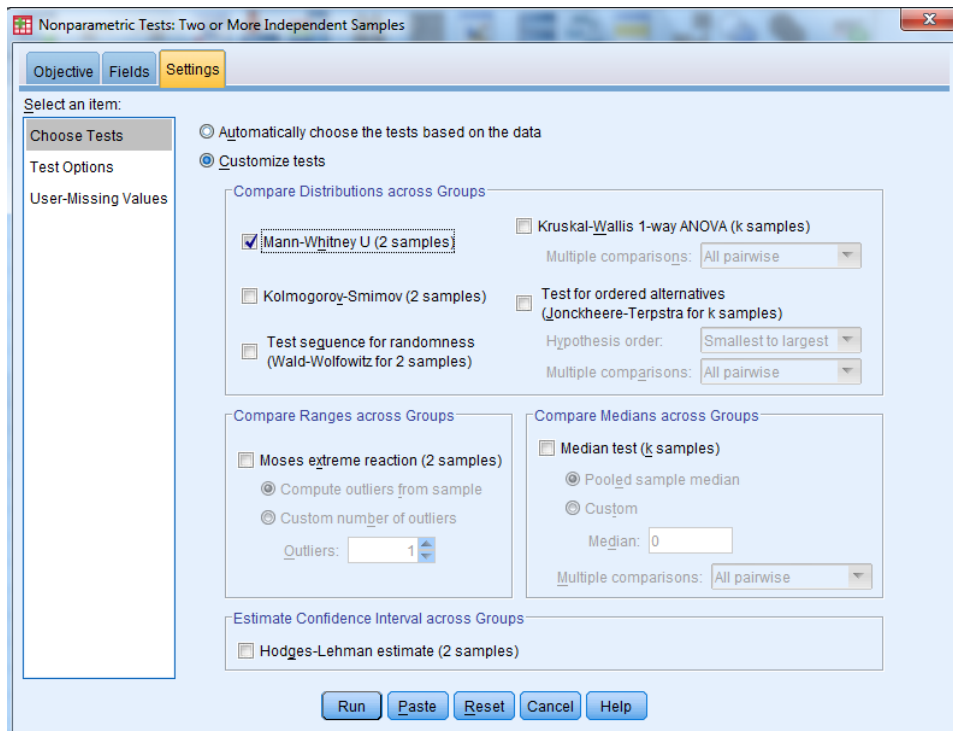


Εικόνα 6.129: Το menu *Nonparametric Tests: Two or More Independent-Samples*



Εικόνα 6.130: Το menu *Nonparametric Tests: Two or More Independent-Samples > Fields*

Στο πεδίο *Test Fields* ορίζουμε τη συνεχή μεταβλητή, και στο πεδίο *Groups* την κατηγορική μεταβλητή.



Εικόνα 6.131: Το menu *Nonparametric Tests: Two or More Independent-Samples* > *Settings*

Στη συνέχεια στην ενότητα *Settings* επιλέγουμε το κριτήριο σύγκρισης των δύο ομάδων *Mann-Whitney U (2 samples)*.

Η διαδικασία ολοκληρώνεται με το πλήκτρο *RUN* όπου και λαμβάνουμε:

Ranks				
	Group	N	Mean Rank	Sum of Ranks
Weight	Group A	12	12.13	145.50
	Group B	8	8.06	64.50
	Total	20		

Εικόνα 6.132: Περιγραφικά μέτρα των βαθμών των θέσεων των παρατηρήσεων

Το SPSS μας δίνει το μέσο της ιεράρχησης ή το μέσο της κατάταξης (Mean Rank) και όχι περιγραφικά μέτρα για τις δύο μεταβλητές ή για τις μετρήσεις της ίδιας μεταβλητής. Στις περιπτώσεις που η κανονικότητα παραβιάζεται συνηθίζουμε να δίνουμε τη διάμεσο ή το εύρος. Καθώς αυτά τα περιγραφικά δεν παρέχονται αυτόματα από το SPSS τα εξάγουμε από την εντολή Means, βλ. ομόνομη ενότητα.

Test Statistics ^b	
	Weight
Mann-Whitney U	28.500
Wilcoxon W	64.500
Z	-1.506
Asymp. Sig. (2-tailed)	.132
Exact Sig. [2*(1-tailed Sig.)]	.135 ^a

a. Not corrected for ties.

b. Grouping Variable: Group

Εικόνα 6.133: Αποτελέσματα του ελέγχου Mann-Whitney U

Ο έλεγχος υποδεικνύει ότι η υπόθεση της ισότητας δεν μπορεί να απορριφθεί σε επίπεδο 5% ($p\text{-value} = 0,132 > 0,05$), άρα δεν μπορούμε να ισχυριστούμε ότι οι δύο μέσοι διαφέρουν.

6.4 T-test για τον έλεγχο της διαφοράς δύο μέσων από συσχετιζόμενα δείγματα

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Compare Means > Paired-Samples T-test*, με σκοπό τον έλεγχο της συσχέτισης μιας συνεχούς μεταβλητής με μια δίτιμη κατηγορική μεταβλητή, στην οποία οι κατηγορίες αφορούν μετρήσεις που γίνονται στα ίδια άτομα πριν και μετά από τη μεσολάβηση ενός γεγονότος (π.χ. χορήγηση φαρμάκου). Για παράδειγμα, εξετάζουμε εάν οι άνθρωποι χάνουν βάρος, εάν υποβληθούν σε μια δίαιτα υψηλή σε υδατάνθρακες και χαμηλά λιπαρά. Τα δύο δείγματα καλούνται εξαρτημένα (correlated) και οι παρατηρήσεις ζευγαρωτές παρατηρήσεις.

Καθώς οι μετρήσεις εφαρμόζονται στο ίδιο άτομο, το δείγμα που προκύπτει προέρχεται από τη διαφορά των δύο μετρήσεων (πριν και μετά το γεγονός). Η υπόθεση που εξετάζεται είναι εάν ο μέσος όρος αυτών των διαφορών ισούται με το μηδέν ή όχι.

$$H_0 : \bar{d} = 0$$

$$H_1 : \bar{d} \neq 0$$

Εξίσωση 6.12

Η ελεγχουσυνάρτηση είναι η:

$$t = \frac{\bar{d}}{\hat{\sigma}_D} \sim t_{N-1}$$

Εξίσωση 6.13

Όπου:

$$\bar{d} = \sum_{i=1}^n \frac{x_i - y_i}{n}$$

Εξίσωση 6.14

και

$$\hat{\sigma}_d = s_d \sqrt{\frac{1}{N}}$$

Εξίσωση 6.15

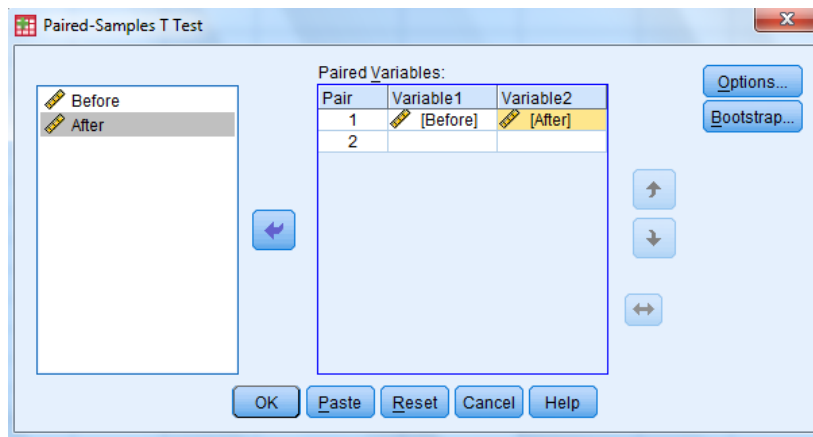
Η υπόθεση απορρίπτεται όταν $t > t_{N-1}$.

Συνεπώς, εάν η υπόθεση της ισότητας δεν απορριφθεί τότε θα συνεπάγεται ότι οι διαφορές είναι κατά μέσο όρο ίσες με το μηδέν, δηλ. ότι το γεγονός δεν επηρέασε την αρχική κατάσταση.

Στο πεδίο *Paired Variables* στις θέσεις *Variable1* & *Variable2* εισάγουμε τις δύο μεταβλητές που αντιπροσωπεύουν τις μετρήσεις πριν και μετά το γεγονός.

Προσοχή! Διαπιστώνουμε ότι η μορφή των δεδομένων για να πραγματοποιηθεί η ανάλυση είναι Wide Format. Εάν τα δεδομένα μας δεν είναι σε αυτή την μορφή, τότε βλ. ενότητα *Restructure* για τον μετασχηματισμό.

Από το πλήκτρο *Options* ορίζουμε το επίπεδο σημαντικότητας για τη δημιουργία των διαστημάτων εμπιστοσύνης.



Εικόνα 6.134: Το menu Paired-Samples T-test

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε δύο πίνακες.

Στον πρώτο πίνακα παρουσιάζονται τα περιγραφικά μέτρα στις δύο ομάδες. Απ' αυτά φαίνεται ότι υπάρχει μια μείωση της μέσης τιμής πριν και μετά.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before	65,780	15	5,1988	1,3423
	After	64,207	15	5,2380	1,3524

Εικόνα 6.135: Περιγραφικά μέτρα των δύο ομάδων

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Before - After	1,5733	,4301	,1110	1,3352	1,8115	14,169	14	,000

Εικόνα 6.136: Αποτελέσματα του T-test για εξαρτημένα δείγματα

Ο δεύτερος πίνακας παρουσιάζει τα αποτελέσματα του ελέγχου από όπου διαπιστώνουμε ότι η διαφορά είναι σε επίπεδο 5% ($p\text{-value} < 0,001$), συνεπώς η διαφορά (μείωση) είναι στατιστικά σημαντική και το γεγονός (δίαιτα) συνετέλεσε στη μείωση της μέσης τιμής (σε μείωση του βάρους).

6.5 Wilcoxon

Ο έλεγχος αυτός είναι ο μη παραμετρικός ισοδύναμος του T-test για εξαρτημένα δείγματα. Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Nonparametric Tests > Related Samples*.

Το κριτήριο Wilcoxon είναι ένας βαθμολογικός έλεγχος που εξετάζει την υπόθεση ότι το άθροισμα των βαθμών των αρνητικών διαφορών είναι ίσο με το άθροισμα των βαθμών των θετικών διαφορών που προκύπτουν από τη διαφορά των τιμών των δύο δειγμάτων.

Η ελεγχουσυνάρτηση βασίζεται στη σχέση:

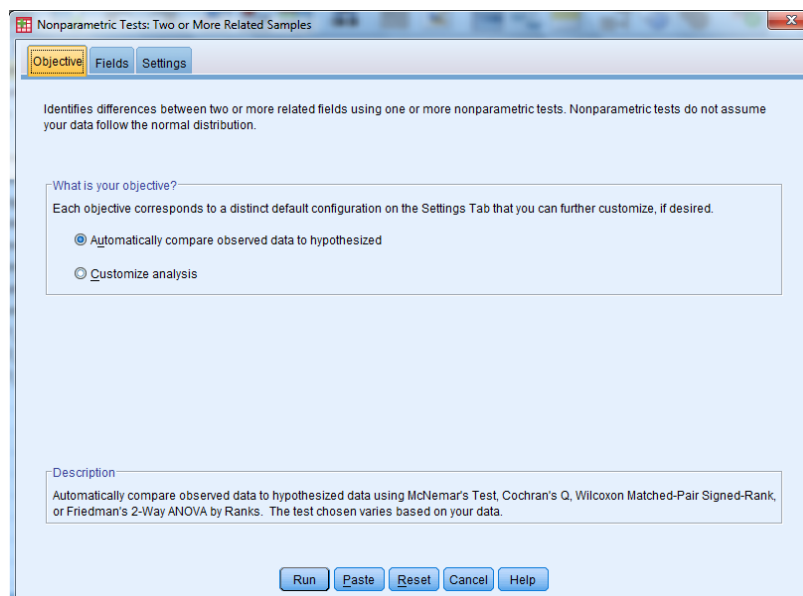
$$T = \min(T^+, T^-)$$

Εξίσωση 6.16

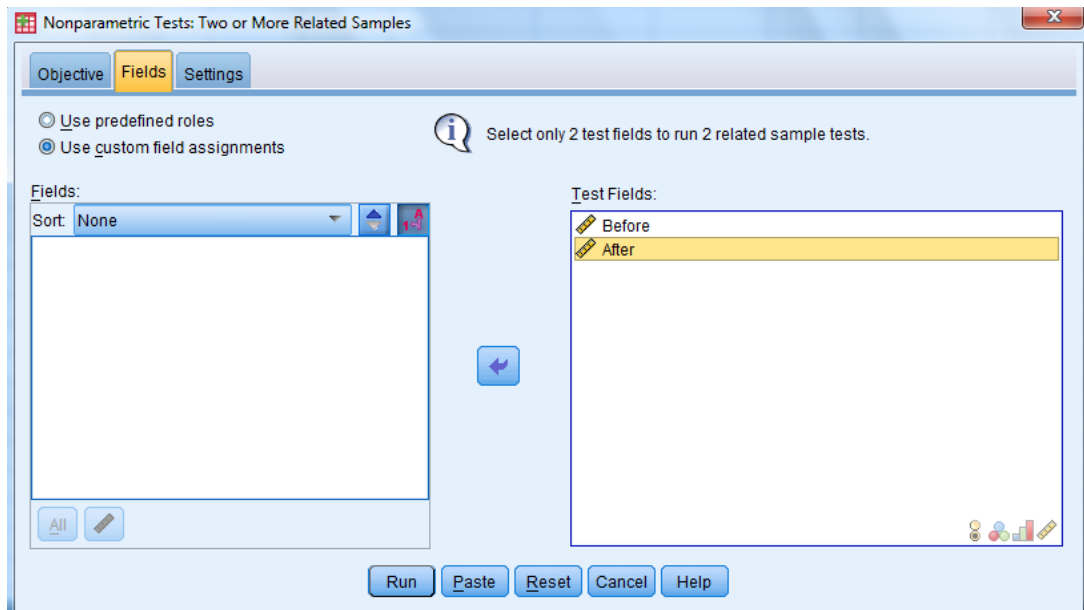
Όπου:

T^+ : το άθροισμα των βαθμών των θετικών διαφορών και

T^- : το άθροισμα των βαθμών των αρνητικών διαφορών μεταξύ των δύο ομάδων (πριν και μετά).



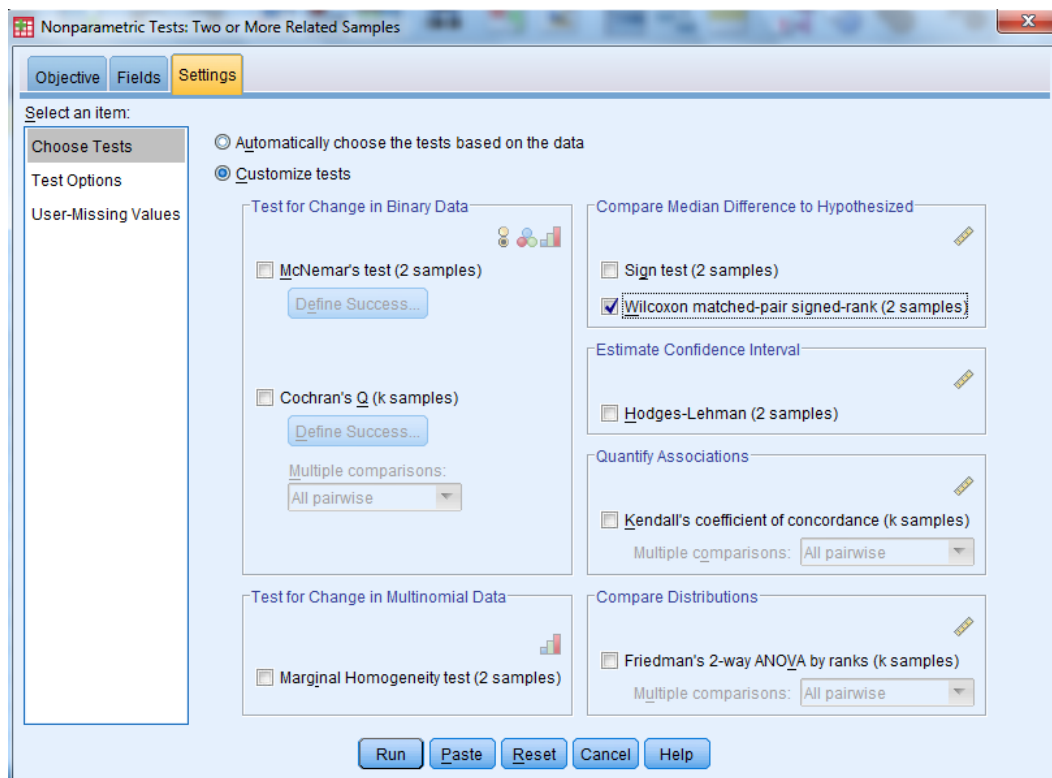
Εικόνα 6.137: Το menu Nonparametric Tests: Two or More Related-Samples



Εικόνα 6.138: Το menu *Nonparametric Tests: Two or More Independent-Samples > Fields*

Στο πεδίο *Test Fields* ορίζουμε τις δύο μεταβλητές που αντιπροσωπεύουν τις μετρήσεις πριν και μετά το γεγονός.

Στη συνέχεια στην ενότητα *Settings* επιλέγουμε το κριτήριο σύγκρισης των δύο ομάδων *Wilcoxon matched-pair signed-rank (2 samples)*.



Εικόνα 6.139: Το menu *Nonparametric Tests: Two or More Independent-Samples > Settings*

Η διαδικασία ολοκληρώνεται με το πλήκτρο *RUN* όπου και λαμβάνουμε:

		N	Mean Rank	Sum of Ranks
After - Before	Negative Ranks	15 ^a	8.00	120.00
	Positive Ranks	0 ^b	.00	.00
	Ties	0 ^c		
	Total	15		

a. After < Before

b. After > Before

c. After = Before

Εικόνα 6.140: Περιγραφικά μέτρα των βαθμών των θέσεων των παρατηρήσεων

Test Statistics ^b	
	After - Before
Z	-3.410 ^a
Asymp. Sig. (2-tailed)	.001

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Εικόνα 6.141: Αποτελέσματα του ελέγχου Wilcoxon

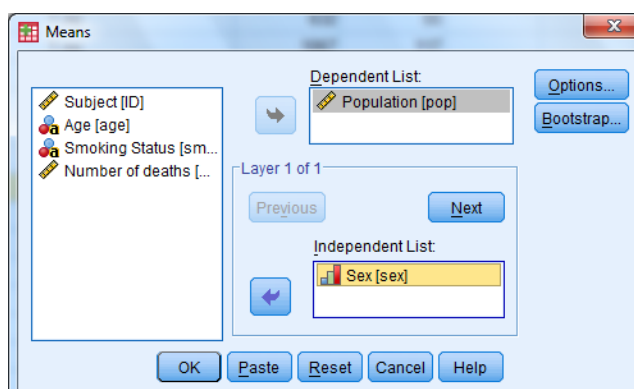
Το SPSS μας δίνει το μέσο της ιεράρχησης ή το μέσο της κατάταξης (Mean Rank) και όχι περιγραφικά μέτρα για τις δύο μεταβλητές ή για τις μετρήσεις της ίδιας μεταβλητής. Στις περιπτώσεις που η κανονικότητα παραβιάζεται συνηθίζουμε να δίνουμε τη διάμεσο ή το εύρος. Καθώς αυτά τα περιγραφικά δεν παρέχονται αυτόματα από το SPSS τα εξάγουμε από την εντολή Means, βλ. ομόνομη ενότητα.

Ο δεύτερος πίνακας παρουσιάζει τα αποτελέσματα του ελέγχου απ' όπου διαπιστώνουμε ότι σε επίπεδο 5% ($p\text{-value} < 0,001$) απορρίπτεται η H_0 , συνεπώς η διαφορά (μείωση) είναι στατιστικά σημαντική και το γεγονός (δίαιτα) συνετέλεσε στη μείωση της μέσης τιμής (σε μείωση του βάρους).

6.6 Means

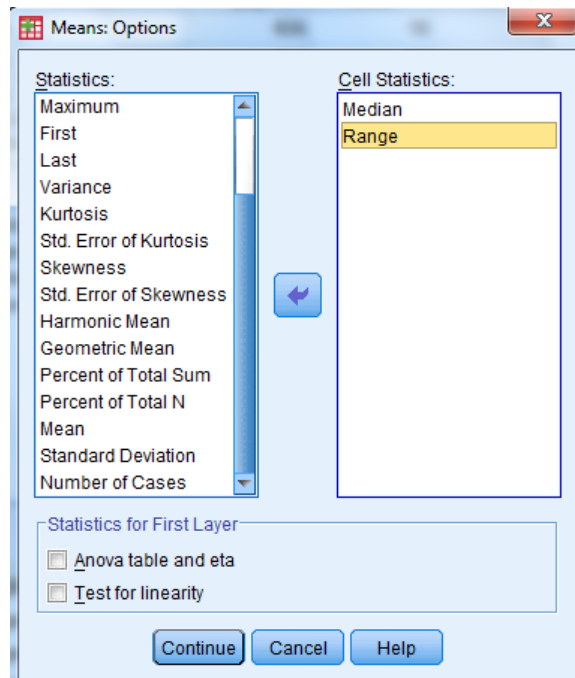
Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Compare Means > Means*, με σκοπό την εξαγωγή περιγραφικών μέτρων σε ομάδες δεδομένων.

Στο πεδίο *Dependent List* εισάγουμε τη συνεχή μεταβλητή για την οποία θα εξαχθούν τα περιγραφικά μέτρα και στο πεδίο *Independent List* την κατηγορική μεταβλητή ομαδοποίησης των δεδομένων.



Εικόνα 6.142: Το menu Means

Από το πλήκτρο *Options* επιλέγουμε τα στατιστικά που επιθυμούμε να εξαχθούν εισάγοντάς τα στο πεδίο *Cell Statistics*.



Εικόνα 6.143: Το menu Means > Options

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

Report

Population

Sex	Median	Range	Mean
Male	1195,00	5948	1874,00
Female	667,00	4584	1206,82
Total	858,50	5954	1558,94

Εικόνα 6.144: Αποτελέσματα εντολής Means

Ο πίνακας παρουσιάζει τα περιγραφικά μέτρα τόσο για την κάθε ομάδα, όσο και για το σύνολο του δείγματος.

6.7 One-Way ANOVA για ανεξάρτητα δείγματα

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Compare Means > One- Way ANOVA*, με σκοπό τον έλεγχο συσχέτισης μεταξύ μιας συνεχούς μεταβλητής με μια κατηγορική η οποία έχει περισσότερες από δύο κατηγορίες. Ανάλογα με τις κατηγορίες- «επίπεδα» της κατηγορικής μεταβλητής η μέθοδος προσπαθεί να εξηγήσει τη μεταβλητότητα που οφείλεται στα διαφορετικά επίπεδα της συνεχούς μεταβλητής. Ακόμα ερευνάται η μεταβλητότητα που υπάρχει εντός της κάθε κατηγορίας «επίπεδο» αποδίδοντας σε αυτήν το ανάλογο «ποσοστό».

Υποθέσεις

Για την εφαρμογή της μεθόδου υποθέτουμε ότι:

- Οι πληθυσμοί / επίπεδα από τους οποίους προέρχονται οι μετρήσεις έχουν την ίδια διασπορά:

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

Εξίσωση 6.17

- Οι πληθυσμοί κατανέμονται κανονικά.
- Οι μετρήσεις σε κάθε δείγμα (επίπεδο) είναι ανεξάρτητες και κατανέμονται κανονικά.

Εφόσον ισχύουν οι παραπάνω υποθέσεις, αναρωτιόμαστε γιατί να μην γίνει εφαρμογή του ελέγχου T-test λαμβάνοντας ανά δύο τα ζεύγη των επιπέδων της κατηγορικής μεταβλητής, ώστε να συγκριθούν οι ομάδες που διαφέρουν μεταξύ τους. Αυτό δεν μπορεί να εφαρμοστεί, γιατί θα έπρεπε να εφαρμοστούν:

$$\binom{k}{2}$$

Εξίσωση 6.18

Έλεγχοι για τα k ανεξάρτητα δείγματα, γιατί εάν το κάθε T-test απαιτούσε ένα επίπεδο στατιστικής σημαντικότητας α (σφάλμα Τύπου I), τότε το επίπεδο στατιστικής σημαντικότητας για την συνολική ανάλυση θα ήταν $k\alpha$. Αυτό οδηγεί στο ότι το επίπεδο σημαντικότητας της κάθε μηδενικής υπόθεσης θα έπρεπε να περιοριστεί σε: $\alpha = 5\% / k$.

Με αυτόν τον τρόπο η περιοχή απόρριψης της μηδενικής υπόθεσης περιορίζεται πολύ. Ως εκ τούτου αντί των πολλαπλών T-test εφαρμόζουμε ένα F-test.

Ανάλυση της μεθόδου

Η συνολική μεταβλητότητα των δεδομένων που εκφράζεται μέσω τετραγωνικών αποκλίσεων δίνεται από τη σχέση:

$$SST = SS_{bg} + SS_{wg}$$

Εξίσωση 6.19

Όπου:

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

Εξίσωση 6.20

Η διακύμανση όλων των παρατηρήσεων γύρω από τον κοινό δειγματικό μέσο.

$$\bar{X} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}$$

Εξίσωση 6.21

Ο ολικός δειγματικός μέσος όπου:

$$N = n_1 + \dots + n_k$$

Εξίσωση 6.22

$$SS_{wg} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

Εξίσωση 6.23

Η διακύμανση των παρατηρήσεων γύρω από τον μέσο κάθε δείγματος, δηλ. «μεταβλητότητα» μέσα στο κάθε δείγμα (within groups).

$$SS_{bg} = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2$$

Εξίσωση 6.24

Η διακύμανση των δειγματικών μέσων γύρω από τον ολικό μέσο, δηλαδή «μεταβλητότητα» μεταξύ των δειγμάτων (between groups).

Η ανάλυση διακύμανσης ερευνά αν οι διαφορές ανάμεσα στους μέσους των διαφόρων «επιπέδων» της τυχαίας μεταβλητής X οφείλονται σε «τυχαίους» παράγοντες:

$$SS_{wg} = X_{ij} - \bar{X}_i$$

Εξίσωση 6.25: Μέτρο μεταβλητότητας στο κάθε δείγμα

ή σε συστηματική διαφορά λόγω των διαφόρων επιπέδων της τυχαίας μεταβλητής X :

$$SS_{bg} = \bar{X}_i - \bar{X}$$

Εξίσωση 6.26: Μέτρο μεταβλητότητας μεταξύ των δειγμάτων

ANOVA Table

Source of variability	Sum of Squares SS	Degrees of freedom $d.f.$	Mean of SS MS	Ratio F
Between Groups	SS_{bg}	$k - 1$	$MSS_{bg} = \frac{SS_{bg}}{k-1}$	$F = \frac{MSS_{bg}}{MSS_{wg}}$
Within Groups	SS_{wg}	$N - k$	$MSS_{wg} = \frac{SS_{wg}}{N-k}$	
Total variability	SST	$N - 1$		

Εικόνα 6.145: One-way ANOVA Table

Η υπόθεση που εξετάζεται είναι εάν οι μέσοι όλων των ομάδων είναι ίσοι μεταξύ τους, ή υπάρχει τουλάχιστον ένας μέσος που διαφέρει.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k = \mu$$

$$H_1 : \exists 0 < i < j \leq k : \mu_i \neq \mu_j$$

Εξίσωση 6.27

Η ελεγχουσυνάρτηση δίνεται από τη σχέση:

$$F = \frac{MSS_{bg}}{MSS_{wg}} \sim F_{k-1, N-k}$$

Εξίσωση 6.28

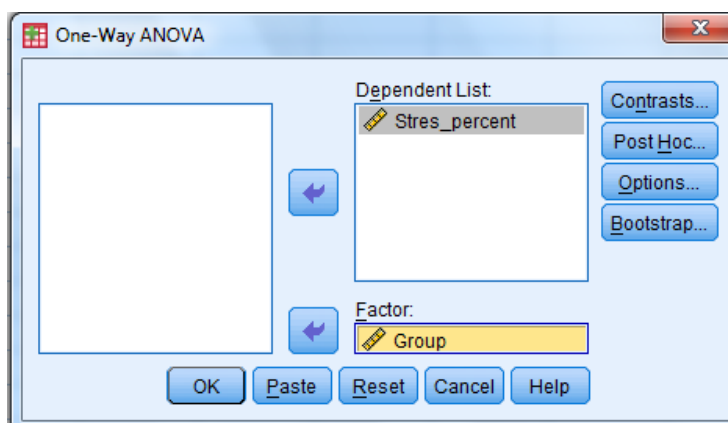
Όπου:

MSS_{bg} , η μέση μεταβλητότητα μεταξύ των ομάδων, και MSS_{wg} , η μέση μεταβλητότητα εντός των ομάδων.

Η υπόθεση απορρίπτεται όταν $F > F_{k-1, k; \alpha}$. Τα αποτελέσματα του ελέγχου παρουσιάζονται στην Εικόνα 6.24.

Έστω ότι έχουμε έναν πληθυσμό από 20 ποντίκια σε καθεστώς άγχους χωρισμένα σε 4 ισοπληθείς ομάδες, στα οποία χορηγείται ενδοφλεβίως ένα φάρμακο σε διαφορετικές δόσεις. Εξετάζουμε εάν η δοσολογία του φαρμάκου επιδρά στη μείωση των ποσοστών ταραχής.

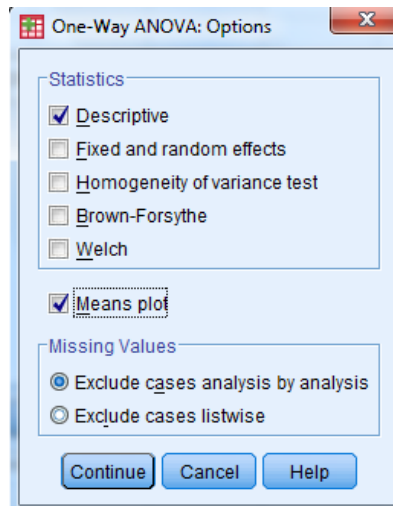
Στο πεδίο *Dependent List* εισάγουμε την/ις υπό εξέταση συνεχή/είς μεταβλητή/ές (π.χ. τη μεταβλητή «ποσοστό άγχους») και στο πεδίο *Factor* την κατηγορική μεταβλητή. Στο παράδειγμά μας είναι η μεταβλητή που καθορίζει σε ποια ομάδα δοσολογίας ανήκει το κάθε ποντίκι.



Εικόνα 6.146: Το menu One-way ANOVA

Από το πλήκτρο *Options* ορίζουμε την εμφάνιση πρόσθετων πληροφοριών και ελέγχων, όπως:

- *Descriptive*, για την εμφάνιση περιγραφικών μέτρων (μέση τιμή, τυπική απόκλιση, τυπικό σφάλμα, 95% Δ.Ε. του μέσου, ελάχιστη και μέγιστη παρατήρηση) της κάθε κατηγορίας.
- *Fixed and Random Effects*, για την εκτίμηση της επίδρασης της κατηγορικής μεταβλητής στη συνεχή μεταβλητή μέσω ενός υποδείγματος σταθερών επιδράσεων και μέσω ενός υποδείγματος τυχαίων επιδράσεων.
- *Homogeneity of variance test*, για τον έλεγχο της ισότητας των διασπορών μεταξύ των ομάδων. Ο έλεγχος που υπολογίζεται είναι αυτός του Levene. Εάν ο έλεγχος απορριφθεί τότε η χρήση της μεθόδου δεν συνιστάται δεδομένου ότι παραβιάζεται η υπόθεση εφαρμογής της μεθόδου. Σε αυτή την περίπτωση θα πρέπει να γίνει χρήση της μη παραμετρικής μεθόδου Kruskal-Wallis, ή των ελέγχων Brown-Forsythe και Welch.
- *Means Plot*, για την εμφάνιση διαγράμματος των μέσων της κάθε κατηγορίας.
- *Missing Values*, για τον χειρισμό των ελλειπουσών τιμών.

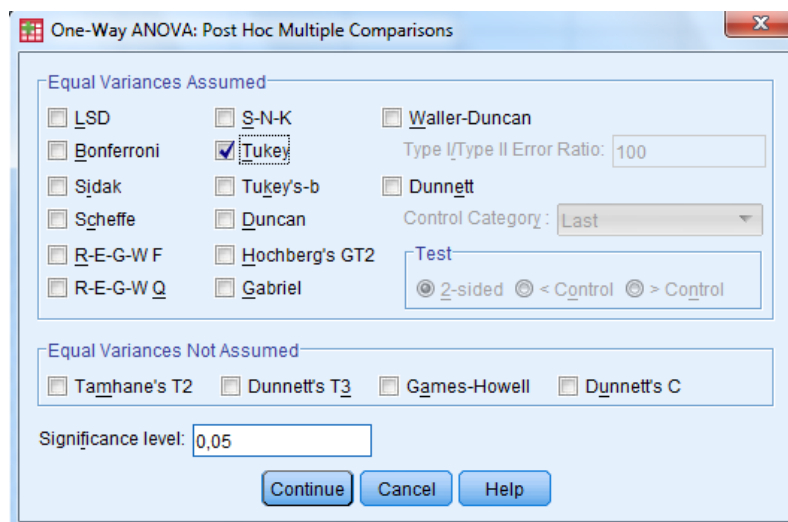


Εικόνα 6.147: Το menu One-way ANOVA > Options

Σε περίπτωση που ο έλεγχος One-way ANOVA απορριφθεί, δεν είναι σε θέση να μας δώσει ποιος ήταν ο μέσος που διέφερε σε σχέση με τους υπόλοιπους. Για να προσδιοριστούν οι μέσοι που διαφέρουν μεταξύ τους, θα πρέπει να γίνουν πολλαπλές συγκρίσεις μεταξύ των ομάδων (ανά δύο). Επειδή όπως προαναφέραμε θα πρέπει να ληφθεί υπόψη η μείωση στο επίπεδο σημαντικότητας λόγω των πολλαπλών συγκρίσεων, έχουν διαμορφωθεί ορισμένες μέθοδοι, όπως:

- Έλεγχοι που βασίζονται σε προσεγγίσεις pairwise multiple comparisons tests ή range tests (Tukey, Hochberg's GT2, Gabriel, Scheffe),
- Έλεγχοι που κάνουν μόνο πολλαπλές συγκρίσεις των μέσων τιμών ανά δύο (Bonferroni, Sidak, Dunnett C, LSD-Least Significant Difference).

Η πιο ευρέως διαδεδομένη είναι ο έλεγχος του Tukey.



Εικόνα 6.148: Οι δυνατές επιλογές για Post-Hoc Multiple Comparisons

Tukey Test

Είναι μια διαδικασία πολλαπλών συγκρίσεων που χρησιμοποιείται κυρίως όταν τα μεγέθη των k δειγμάτων είναι ίσα, και εξετάζει την υπόθεση:

$$H_0 : \mu_L = \mu_S$$

$$H_1 : \mu_L \neq \mu_S$$

Εξίσωση 6.29

Ο έλεγχος βασίζεται στον υπολογισμό της στατιστικής συνάρτησης Q η οποία ορίζεται ως:

$$Q = \frac{\bar{X}_L - \bar{X}_S}{\sqrt{\frac{MSS_{wg}}{N_{p/s}}}}$$

Εξίσωση 6.30

Όπου:

$$\bar{X}_L$$

Ο μέγιστος δειγματικός μέσος

$$\bar{X}_S$$

Ο ελάχιστος δειγματικός μέσος

$$N_{p/s}$$

Το πλήθος των μετρήσεων σε κάθε δείγμα.

Όλοι οι μέσοι συγκρίνονται ανά δύο και απορρίπτουμε τη μηδενική υπόθεση όταν $Q > Q_\alpha$, όπου η στατιστική συνάρτηση $Q \sim$ κατανομή (k, df_{wg}) .

Σημείωση

Αν τα δείγματα έχουν διαφορετικό μέγεθος, τότε το $N_{p/s}$ υπολογίζεται από τον αρμονικό μέσο των μεγεθών των δειγμάτων:

$$N_{p/s} = \frac{k}{\frac{1}{n_1} + \dots + \frac{1}{n_k}}$$

Εξίσωση 6.31

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					1	5		
2	5	25,0400	2,42343	1,08379	22,0309	28,0491	22,80	27,70
3	5	22,5000	3,36378	1,50433	18,3233	26,6767	19,30	27,80
4	5	22,3000	1,96850	,88034	19,8558	24,7442	19,60	23,90
Total	20	24,6750	3,67364	,82145	22,9557	26,3943	19,30	33,50

Εικόνα 6.149: Περιγραφικά μέτρα των ομάδων

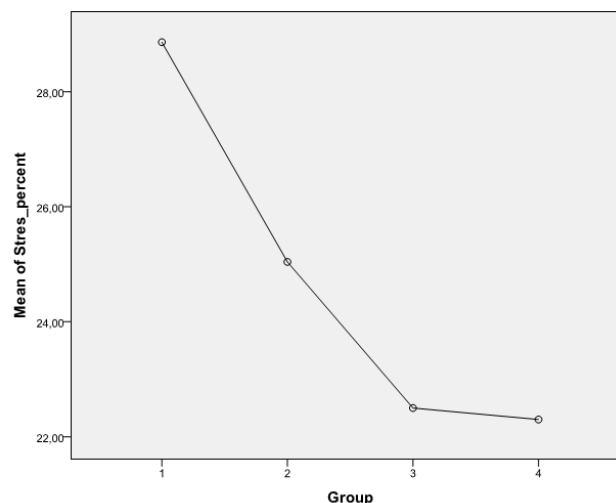
ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	140,094	3	46,698	6,423	,005
Within Groups	116,324	16	7,270		
Total	256,418	19			

Εικόνα 6.150: ANOVA Table με τα αποτελέσματα του F-test

Στα αποτελέσματα λαμβάνουμε τον πίνακα με τα περιγραφικά μέτρα των μέσων των ομάδων, και τον πίνακα ANOVA με τα αποτελέσματα του F-test, από τον οποίο διαπιστώνουμε ότι η μηδενική υπόθεση απορρίπτεται ($p\text{-value} = 0,005$), συνεπώς υπάρχει τουλάχιστον ένας μέσος που διαφέρει σημαντικά από τους υπόλοιπους. Αναφορικά με το παράδειγμά μας μπορεί να θεωρηθεί ότι υπάρχουν σημαντικές διαφορές στους μέσους των τεσσάρων ομάδων, επομένως η αύξηση της δοσολογίας που χορηγήθηκε στα ποντίκια μειώνει τα επίπεδα στρες.

Για το λόγο αυτό θα πρέπει να ληφθεί υπόψη και ο πίνακας πολλαπλών συγκρίσεων μέσω κάποιας από τις μεθόδους της ενότητας Post Hoc Multiple Comparisons.



Εικόνα 6.151: Γραφική απεικόνιση των μέσων τιμών των ομάδων

Multiple Comparisons

		Stres_percent Tukey HSD			95% Confidence Interval	
(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
1	2	3,82000	1,70532	,155	-1,0589	8,6989
	3	6,36000*	1,70532	,009	1,4811	11,2389
	4	6,56000*	1,70532	,007	1,6811	11,4389
2	1	-3,82000	1,70532	,155	-8,6989	1,0589
	3	2,54000	1,70532	,466	-2,3389	7,4189
	4	2,74000	1,70532	,403	-2,1389	7,6189
3	1	-6,36000*	1,70532	,009	-11,2389	-1,4811
	2	-2,54000	1,70532	,466	-7,4189	2,3389
	4	,20000	1,70532	,999	-4,6789	5,0789
4	1	-6,56000*	1,70532	,007	-11,4389	-1,6811
	2	-2,74000	1,70532	,403	-7,6189	2,1389
	3	-,20000	1,70532	,999	-5,0789	4,6789

*. The mean difference is significant at the 0.05 level.

Εικόνα 6.152: Πίνακας Πολλαπλών συγκρίσεων μέσω της μεθόδου Tukey

Ο πίνακας Multiple Comparisons παρουσιάζει τις διαφορές των μέσων τιμών μεταξύ των ομάδων ανά δύο, σημειώνοντας με αστερίσκο τις στατιστικά σημαντικές διαφορές.

Από την εφαρμογή του Tukey test έχουμε:

$$\sqrt{\frac{MSS_{wg}}{N_{p/s}}} = \sqrt{\frac{7,27}{5}} = 1.2058$$

Εξίσωση 6.32

$Q_{\alpha=0.05} = 4,05$ και $Q_{\alpha=0.01} = 5,2$ ποσοστιαία σημεία της Q .

Οπότε η διαφορά των μέσων των ομάδων 1 & 2:

$$\bar{X}_L - \bar{X}_S < Q_{\alpha} \sqrt{\frac{MSS_{wg}}{N_{p/s}}} \Rightarrow 3.82 < 4.88$$

Εξίσωση 6.33

Άρα η H_0 δεν απορρίπτεται.

Η διαφορά των μέσων των ομάδων 1 & 3:

$$\bar{X}_L - \bar{X}_S < Q_{\alpha} \sqrt{\frac{MSS_{wg}}{N_{p/s}}} \Rightarrow 6.36 > 4.88$$

Εξίσωση 6.34

Άρα η H_0 απορρίπτεται κ.ο.κ. Επιπρόσθετα στα αποτελέσματα εμφανίζεται ο ακόλουθος πίνακας Homogeneous Subsets στον οποίο παρουσιάζονται τα ομοιογενή υποσύνολα που δύναται να δημιουργηθούν από τις υφιστάμενες ομάδες. Στο παράδειγμά μας διαπιστώνουμε ότι μπορούν να δημιουργηθούν δύο

υποσύνολα στα οποία είτε το πρώτο θα περιέχει τις ομάδες 4,3 και 2 και το δεύτερο την ομάδα 1 μόνη της, είτε το πρώτο θα περιέχει τις ομάδες 4 και 3 και το δεύτερο τις ομάδες 2 και 1.

Homogeneous Subsets

Stres_percent

Tukey HSD^a

Group	N	Subset for alpha = 0.05	
		1	2
4	5	22,3000	
3	5	22,5000	
2	5	25,0400	25,0400
1	5		28,8600
Sig.		,403	,155

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5,000.

Εικόνα 6.153: Έλεγχος ομοιογένειας των ομάδων με σκοπό την ομαδοποίηση τους

6.8 Kruskal-Wallis H

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Non parametric Tests > Legacy Dialogs > k Independent Samples*, με σκοπό τον έλεγχο συσχέτισης μεταξύ μιας συνεχούς μεταβλητής με μια κατηγορική, η οποία έχει περισσότερες από δύο κατηγορίες με ίσο ή διαφορετικό πλήθος ατόμων ανά κατηγορία. Η μέθοδος χρησιμοποιείται όταν δεν ισχύουν οι υποθέσεις εφαρμογής της One-Way ANOVA, δηλ. όταν οι κατανομές δεν είναι κανονικές και οι διασπορές δεν είναι ίσες, ή όταν η μεταβλητή είναι διατεταγμένη.

Η μέθοδος βασίζεται στην ίδια λογική των ελέγχων εξετάζοντας αν ένα σύνολο k ανεξάρτητων ομάδων προέρχονται από τον ίδιο πληθυσμό. Σε αντίθεση με τον έλεγχο One-way ANOVA που εξετάζει την ισότητα των μέσων, εδώ χρησιμοποιείται το άθροισμα των βαθμών των θέσεων των παρατηρήσεων. Η μέθοδος είναι μια επέκταση του κριτηρίου Mann-Whitney U για k ανεξάρτητες ομάδες. Κατ' επέκταση η υπόθεση που εξετάζεται είναι αν οι ομάδες έχουν ίσο μέσο βαθμών (δηλ. τυχαία διάταξη) ή κάποια ομάδα διαφοροποιεί τη διάταξη των παρατηρήσεων.

Για την εφαρμογή της μεθόδου οι παρατηρήσεις διατάσσονται κατά αύξουσα σειρά και σε κάθε παρατήρηση προσδιορίζεται η σχετική της θέση (rank). Στη συνέχεια υπολογίζονται οι ποσότητες:

$$SS_k = \sum_{j=1}^k \frac{R_j^2}{n_j}$$

Εξίσωση 6.35

Όπου:

$$R_j = \sum_{i=1}^{n_j} r_{ij}$$

Εξίσωση 6.36

Το άθροισμα των θέσεων της κάθε j-ομάδας, το οποίο είναι το αντίστοιχο «άθροισμα τετραγώνων των θέσεων μεταξύ των ομάδων»:

$$SS_r = \sum_{j=1}^k \sum_{i=1}^{n_j} r_{ij}^2$$

Εξίσωση 6.37

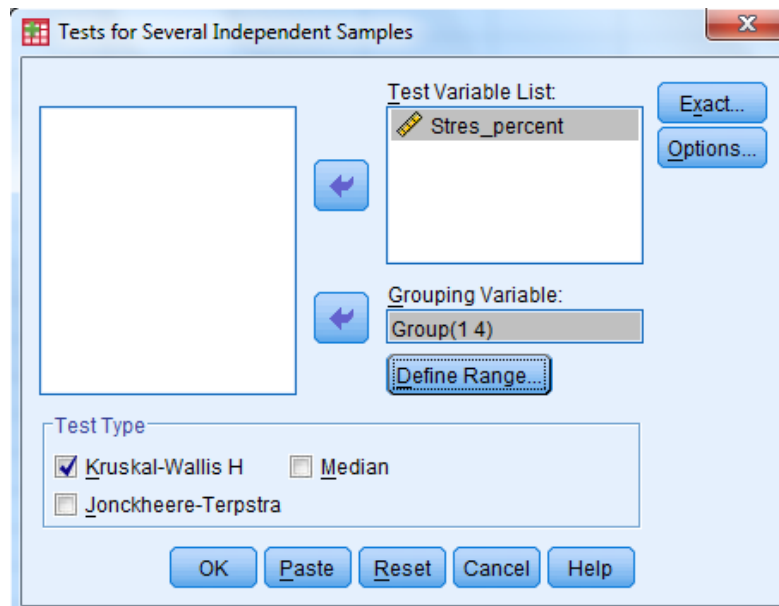
Το «συνολικό άθροισμα τετραγώνων των θέσεων των παρατηρήσεων όλων των ομάδων».

Η ελεγχουσυνάρτηση προκύπτει από τη σχέση:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1) \sim X_{k-1}^2$$

Εξίσωση 6.38

Η υπόθεση απορρίπτεται όταν: $H > X_{k-1;\alpha}^2$



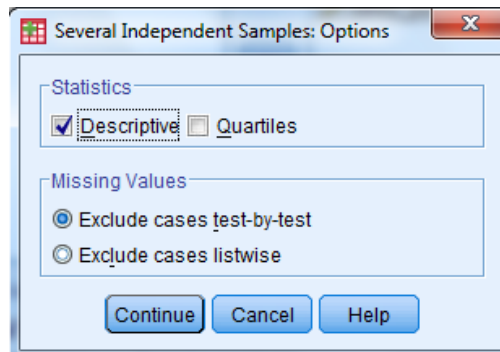
Εικόνα 6.154: Το menu εκτέλεσης Kruskal-Wallis H

Στο πεδίο *Test Variable List* εισάγουμε την/ις υπό εξέταση συνεχή/είς μεταβλητή/ές (π.χ. την μεταβλητή «ποσοστό άγχους») και στο πεδίο *Grouping Variable* την κατηγορική μεταβλητή. Στη συνέχεια από το πλήκτρο *Define Range* ορίζουμε το εύρος των κωδικών των επιπέδων της κατηγορικής μεταβλητής (ελάχιστο και μέγιστο κωδικό) που θα χρησιμοποιηθούν στην ανάλυσή μας.

Σ' αυτή την ενότητα παρέχονται επιπρόσθετα οι έλεγχοι Median και Jonckheere-Terpstra:

- *Median Test*, η οποία είναι μια μη παραμετρική διαδικασία που βασίζεται στον έλεγχο της συνάφειας μεταξύ των ομάδων και του πλήθους των παρατηρήσεων που είναι μικρότερες ή μεγαλύτερες από τη διάμεσο.
- *Jonckheere-Terpstra* που χρησιμοποιείται αντί του Kruskal-Wallis όταν μεταξύ των επιπέδων των ομάδων υπάρχει σχετική διάταξη, ώστε να εξεταστεί το ενδεχόμενο της ύπαρξης γραμμικής σχέσης, μεταξύ της διάταξης και της εξαρτημένης μεταβλητής.

Από το πλήκτρο *Options* επιλέγουμε την εμφάνιση των περιγραφικών μέτρων (Descriptive) και των τεταρτημορίων (Quartiles) για την εξαρτημένη μεταβλητή.



Εικόνα 6.155: Το menu εκτέλεσης Kruska -Wallis H > Options

Η διαδικασία ολοκληρώνεται με το πλήκτρο OK όπου και λαμβάνουμε:

Descriptive Statistics								
	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Stres_percent	20	24,6750	3,67364	19,30	33,50	22,1250	23,8000	27,6750
Group	20	2,50	1,147	1	4	1,25	2,50	3,75

Εικόνα 6.156: Περιγραφικά μέτρα των ομάδων

Ranks			
	Group	N	Mean Rank
Stres_percent	1	5	16,80
	2	5	11,20
	3	5	6,80
	4	5	7,20
	Total	20	

Εικόνα 6.157: Οι μέσοι των βαθμών των ομάδων. Ο έλεγχος εξετάζει την ισότητά τους

Test Statistics ^{a,b}	
	Stres_percent
Chi-square	9,258
df	3
Asymp. Sig.	,026

a. Kruskal Wallis Test

b. Grouping Variable:
Group

Εικόνα 6.158: Αποτελέσματα του ελέγχου Kruskal-Wallis

Από τα αποτελέσματα του ελέγχου Kruskal-Wallis διαπιστώνουμε ότι η υπόθεση απορρίπτεται ($p\text{-value} = 0,026$).

Σε περίπτωση που ο έλεγχος Kruskal-Wallis απορριφθεί, δεν είναι σε θέση να μας δώσει ποιος ήταν ο μέσος που διέφερε σε σχέση με τους υπόλοιπους. Για να προσδιοριστούν οι μέσοι που διαφέρουν μεταξύ τους, θα πρέπει να γίνουν πολλαπλές συγκρίσεις μεταξύ των ομάδων (ανά δύο).

Σε αυτή την περίπτωση ο ερευνητής μπορεί χειροκίνητα να προβεί σε επί μέρους εφαρμογή του Mann-Whitney U test για όλους τους ανά δύο πιθανούς συνδυασμούς, διορθώνοντας το επίπεδο σημαντικότητας διαιρώντας το με το πλήθος των συγκρίσεων (μέθοδος Bonferroni).

6.9 Two-Way ANOVA for Independent Samples

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > General Linear Model > Univariate*, με σκοπό τον έλεγχο συσχέτισης μεταξύ μιας συνεχούς μεταβλητής με δύο κατηγορικές μεταβλητές. Σε περίπτωση που θέλουμε να εξετάσουμε τη συσχέτιση με περισσότερες από δύο κατηγορικές, τότε εφαρμόζεται μια three way ANOVA, κ.ο.κ.

Ανάλυση της μεθόδου

Στην ανάλυση διακύμανσης άνα δύο παράγοντες έχουμε δύο κατηγορικές μεταβλητές X_1 (παράγοντας A) και X_2 (παράγοντας B), και μια εξαρτημένη ποσοτική μεταβλητή Y . Έτσι, οι υποθέσεις που ελέγχουμε είναι εάν:

- Τα διαφορετικά επίπεδα τιμών της X_1 επηρεάζουν τις τιμές της Y .
- Τα διαφορετικά επίπεδα τιμών της X_2 επηρεάζουν τις τιμές της Y .
- Υπάρχει αλληλεπίδραση (interaction) μεταξύ των μεταβλητών X_1 και X_2 .

Έλλειψη αλληλεπίδρασης σημαίνει ότι η επίδραση της μεταβλητής X_1 στην Y είναι ίδια για κάθε επίπεδο της μεταβλητής X_2 και αντίστροφα.

Για να εξεταστούν οι παραπάνω υποθέσεις, θα καταλήξουμε σε τρία F -test.

Υποθέσεις

Έστω ότι η μεταβλητή X_1 έχει r επίπεδα και η μεταβλητή X_2 έχει c επίπεδα. Οι παρατηρήσεις συμβολίζονται ως Y_{ijk} .

Όπου:

$i = 1, \dots, r$: τα επίπεδα της μεταβλητής X_1 ,

$j = 1, \dots, c$: τα επίπεδα της μεταβλητής X_2 ,

$k = 1, \dots, N_T$: το σύνολο των μετρήσεων.

Οι παρατηρήσεις εκφράζονται από τη σχέση:

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

Εξίσωση 6.39

Όπου:

$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

Εξίσωση 6.40

Για την εφαρμογή της ανάλυσης ως προς δύο παράγοντες θα πρέπει να ισχύουν:

- Οι πληθυσμοί από τους οποίους προέρχονται οι παρατηρήσεις να κατανέμονται κανονικά:

$$Y_{ijk} \sim N(\mu_{ij}, \sigma^2)$$

Εξίσωση 6.41

- Οι πληθυσμοί από τους οποίους προέρχονται οι παρατηρήσεις να έχουν ίσες διασπορές:

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

Εξίσωση 6.42

Παράδειγμα

Θέλουμε να εξετάσουμε την επίδραση και την αλληλεπίδραση δύο τύπων ναρκωτικών Α και Β, στη φυσιολογική διέγερση. Έχουμε 40 εργαστηριακά ποντίκια που χωρίζονται σε τέσσερις ομάδες των δέκα ατόμων. Κάθε ομάδα έλαβε μια συγκεκριμένη δόσοληψία από τα ναρκωτικά Α και Β (0 μονάδες και 1 μονάδα). Η τυχαία μεταβλητή Y είναι η φυσιολογική διέγερση.

Στόχος μας είναι να κατασκευάσουμε τον ANOVA-Table οπότε εργαζόμαστε ως εξής:

1^{ον}: Υπολογίζουμε τους μέσους όρους των ομάδων.

2^{ον}: Υπολογίζουμε τα τετραγωνικά αθροίσματα για την κάθε ομάδα, όπου:

$$SS = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N}$$

$$SST = SS_{bg} + SS_{wg}$$

$$SS_{wg} = SS_{g1} + SS_{g2} + SS_{g3} + SS_{g4}$$

$$SS_{bg} = SST - SS_{wg}$$

$$SS_{bg} = SS_r + SS_c + SS_{IN}$$

Εξισώσεις 6.43

Δηλαδή η μεταβλητότητα μεταξύ των ομάδων αναλύεται στη μεταβλητότητα μεταξύ των επιπέδων r , των επιπέδων c και στην αλληλεπίδραση.

3^{ον}: Υπολογίζουμε τους βαθμούς ελευθερίας:

$$df_r = r - 1$$

$$df_c = c - 1$$

$$df_{IN} = (r - 1)(c - 1)$$

$$df_T = N_T - 1$$

$$df_{error} = df_{wg} = N_T - rc$$

$$df_{bg} = rc - 1$$

$$df_T = df_{wg} + df_{bg}$$

$$df_{bg} = df_r + df_c + df_{IN}$$

Εξισώσεις 6.44

4^{ον}: Υπολογίζουμε τα:

$MS_{wg}, MS_r, MS_c, MS_{IN}$ (ή αλλιώς MS_{rc})

$$MSS_r = \frac{SS_r}{df_r}$$

$$MSS_c = \frac{SS_c}{df_c}$$

$$MSS_{IN} = \frac{SS_{IN}}{df_{IN}} \text{ και } MSS_{error} = MSS_{wg} = \frac{SS_{wg}}{df_{wg}}.$$

Εξισώσεις 6.45

5^{ον}: Υπολογίζονται τα F -tests:

- 1) F -test για τον έλεγχο της επίδρασης των επιπέδων της X_1 στις τιμές της Y . Η υπόθεση που εξετάζεται είναι:

H_0 : Δεν υπάρχει στατιστικά σημαντική διαφορά στους μέσους μεταξύ των δύο γραμμών.

H_1 : Υπάρχει στατιστικά σημαντική διαφορά στους μέσους μεταξύ των δύο γραμμών.

$$F_r = \frac{MSS_r}{MSSE}$$

Εξίσωση 6.46

Απορρίπτουμε την H_0 όταν:

$$F_r > F_{r-1, N_T - rc; \alpha}$$

Εξίσωση 6.47

- 2) F -test για τον έλεγχο της επίδρασης των επιπέδων της X_2 στις τιμές της Y . Η υπόθεση που εξετάζεται είναι:

H_0 : Δεν υπάρχει στατιστικά σημαντική διαφορά στους μέσους μεταξύ των δύο στηλών.

H_1 : Υπάρχει στατιστικά σημαντική διαφορά στους μέσους μεταξύ των δύο στηλών.

$$F_c = \frac{MSS_c}{MSSE}$$

Εξίσωση 6.48

Απορρίπτουμε την H_0 όταν:

$$F_c > F_{c-1, N_T - rc; \alpha}$$

Εξίσωση 6.49

- 3) F -test για τον έλεγχο ύπαρξης αλληλεπίδρασης των επιπέδων της X_1 στις τιμές της X_2 . Η υπόθεση που εξετάζεται είναι:

H_0 : Δεν υπάρχει αλληλεπίδραση μεταξύ γραμμών και στηλών.

H_1 : Υπάρχει αλληλεπίδραση μεταξύ γραμμών και στηλών.

$$F_{IN} = \frac{MSS_{IN}}{MSSE}$$

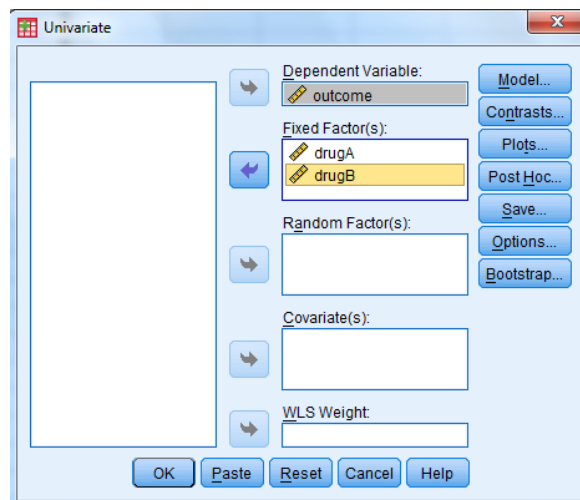
Εξίσωση 6.50

Απορρίπτουμε την H_0 όταν:

$$F_{IN} > F_{(r-1)(c-1), N_T - rc; \alpha}$$

Εξίσωση 6.51

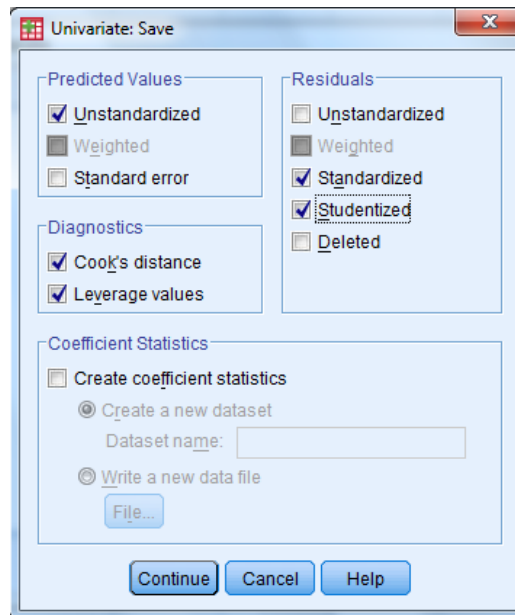
Από την επιλογή *Analyze > General Linear Model > Univariate* ορίζουμε στο πεδίο *Dependent Variable* την εξαρτημένη μεταβλητή και στο πεδίο *Fixed Factor(s)* τις κατηγορικές μεταβλητές.



Εικόνα 6.159: Το menu Univariate

Από το πλήκτρο *Save* ο χρήστης έχει τη δυνατότητα να αποθηκεύσει διάφορες τιμές που προβλέπονται από το μοντέλο, σφάλματα (residuals) ή συναφή μέτρα αξιολόγησης της ανάλυσης, τα οποία αποθηκεύονται ως νέες μεταβλητές στο φύλλο δεδομένων. Αναλυτικότερα:

- *Predicted Values*: Περιέχει επιλογές με σκοπό τον υπολογισμό τιμών που προκύπτουν από προβλέψεις του μοντέλου.
- *Unstandardized*: Η τιμή πρόβλεψης για την εξαρτημένη μεταβλητή.
- *Weighted*: Οι σταθμισμένες μη τυποποιημένες προβλεπόμενες τιμές. Η δυνατότητα προσφέρεται μόνο αν η μεταβλητή έχει προηγουμένως δηλωθεί ως σταθμισμένη.
- *Standard error*: Μια εκτίμηση της τυπικής απόκλισης του μέσου όρου της εξαρτημένης μεταβλητής για τις περιπτώσεις που έχει τις ίδιες τιμές των ανεξάρτητων μεταβλητών.
- *Diagnostics*: Μας παρέχει μέτρα για τον εντοπισμό ασυνήθιστων περιπτώσεων. Π.χ. ακραίων παρατηρήσεων ή παρατηρήσεων επίδρασης στο μοντέλο.
- *Cook's distance*: Είναι ένα μέτρο για τον εντοπισμό ακραίων παρατηρήσεων.
- *Leverage values*: Είναι ένα μέτρο για τον εντοπισμό παρατηρήσεων με μεγάλη επίδραση (μόγλευση) στο μοντέλο.



Εικόνα 6.160: Το menu Univariate > Save

Residuals: Περιέχει τις επιλογές αποθήκευσης διαφόρων μορφών σφαλμάτων του μοντέλου, όπου σφάλμα νοείται ως η διαφορά μεταξύ της προβλεπόμενης τιμής και της πραγματικής παρατήρησης.

Unstandardized: Μη τυποποιημένα κατάλοιπα.

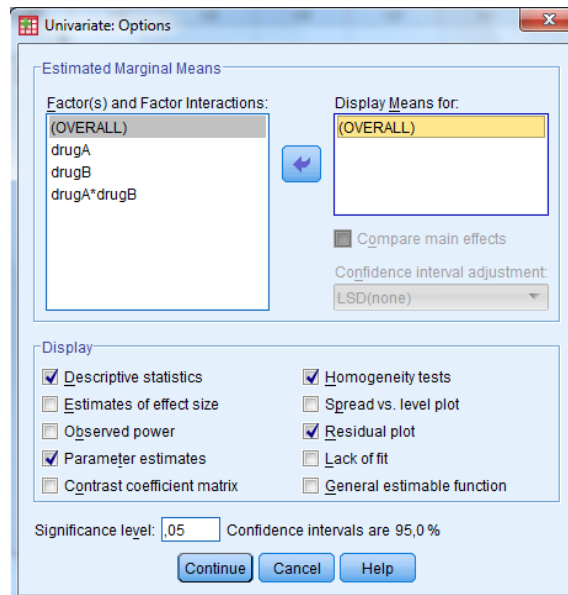
Weighted: Σταθμισμένα μη τυποποιημένα σφάλματα. Η δυνατότητα προσφέρεται μόνο αν η μεταβλητή έχει προηγουμένως δηλωθεί ως σταθμισμένη.

Standardized: Τυποποιημένα σφάλματα. Εάν ισχύει η υπόθεση της κανονικότητας, τότε τα τυποποιημένα σφάλματα έχουν μέση τιμή μηδέν και τυπική απόκλιση 1.

Studentized: Τα σφάλματα διαιρούνται με την εκτιμήτρια της τυπικής απόκλισης η οποία διαφέρει ανάλογα με την κάθε περίπτωση και εξαρτάται από την απόσταση της κάθε παρατηρούμενης τιμής των ανεξάρτητων μεταβλητών από τη μέση τιμή της ανεξάρτητης μεταβλητής.

Deleted: Τα σφάλματα που προκύπτουν όταν η συγκεκριμένη παρατήρηση εξαιρείται από τον υπολογισμό του συντελεστή παλινδρόμησης. Είναι η διαφορά μεταξύ της τιμής της εξαρτημένης μεταβλητής και της προσαρμοσμένης προβλεπόμενης τιμής.

Coefficient Statistics: Δημιουργεί έναν πίνακα διακυμάνσεων - συνδιακυμάνσεων των εκτιμώμενων παραμέτρων του μοντέλου είτε σε μια νέα βάση δεδομένων στην τρέχουσα ενότητα είτε σε νέο εξωτερικό αρχείο. Επίσης για κάθε εξαρτημένη μεταβλητή, θα υπάρξει μια σειρά εκτιμώμενων παραμέτρων, μια σειρά στατιστικά σημαντικών τιμών από ελέγχους T-test των εκτιμήσεων των συντελεστών του μοντέλου και μια σειρά από τους βαθμούς ελευθερίας των σφαλμάτων. Για ένα πολυπαραγοντικό μοντέλο υπάρχουν παρόμοιες σειρές για κάθε εξαρτημένη μεταβλητή. Αυτό το αρχείο δύναται να χρησιμοποιηθεί από άλλες διαδικασίες που έχουν τη δυνατότητα να διαβάζουν αρχεία πινάκων (μήτρας).



Εικόνα 6.161: Το menu Univariate > Options

Από το πλήκτρο *Options* λαμβάνουμε (προαιρετικά) διάφορα στατιστικά στοιχεία τα οποία υπολογίζονται βάσει ενός μοντέλου σταθερών επιδράσεων (fixed effects). Αναλυτικά:

- *Estimated Marginal Means*: Σ' αυτή την ενότητα επιλέγουμε τους παράγοντες και τις αλληλεπιδράσεις για τους οποίους θα υπολογιστούν οι εκτιμώμενοι μέσοι των πληθυσμών των κατηγοριών των παραγόντων. Οι μέσοι αυτοί είναι προσαρμοσμένοι σε σχέση με τις συμμεταβλητές, αν υπάρχουν. Οι αλληλεπιδράσεις είναι διαθέσιμες μόνο αν έχουν οριστεί αλληλεπιδράσεις στο προσαρμοσμένο (custom) μοντέλο.
- *Compare main effects*: Παρέχει μη διορθωμένες κατά ζεύγη συγκρίσεις μεταξύ των εκτιμώμενων μέσων για κάθε κύρια επίδραση του μοντέλου, τόσο για συγκρίσεις μεταξύ των ομάδων όσο και εντός των ομάδων. Η δυνατότητα προσφέρεται μόνο αν έχουν δηλωθεί κύριες επιδράσεις στην ενότητα Display Means For.
- *Confidence interval adjustment*: Επιλέγουμε τη μέθοδο προσαρμογής του επιπέδου στατιστικής σημαντικότητας για τις πολλαπλές συγκρίσεις. Μπορούμε να επιλέξουμε μεταξύ των μεθόδων LSD, Bonferroni, ή Sidak. Η δυνατότητα προσφέρεται μόνο αν έχει επιλεγεί η παράμετρος Compare main effects.

Η επιλογή Display Εμφανίζει πίνακες στατιστικών στοιχείων, όπως:

- *Descriptive statistics*: Παρέχει έναν πίνακα με περιγραφικά μέτρα όπως μέσο, τυπική απόκλιση και μέγεθος δείγματος για όλες τις εξαρτημένες μεταβλητές.
- *Estimates of effect size*: Παρέχει την εκτιμήτρια η^2 για κάθε επίδραση και κάθε εκτιμώμενη παράμετρο. Η εκτιμήτρια η^2 περιγράφει την αναλογία της συνολικής μεταβλητότητας που αναλογεί σε κάθε παράγοντα.
- *Observed power*: Παρέχει την ισχύ του ελέγχου όταν η εναλλακτική υπόθεση βασίζεται στην παρατηρούμενη τιμή.
- *Parameter estimates*: Παρέχει τον πίνακα με τους εκτιμώμενους συντελεστές του μοντέλου, τις τυπικές τους αποκλίσεις, τα t tests, τα διαστήματα εμπιστοσύνης και την ισχύ (p-values) των ελέγχων.
- *Homogeneity tests*: Παρέχει τον έλεγχο του Levene για τον έλεγχο της ομοιογένειας (ισότητας) της διασποράς της εξαρτημένης μεταβλητής σε όλα τα επίπεδα συνδυασμών των ομάδων που προκύπτουν από τις κατηγορικές μεταβλητές. Επίσης, περιλαμβάνονται οι έλεγχοι Box's M test της ομοιογένειας του πίνακα συνδιασποράς της εξαρτημένης μεταβλητής.

- *Spread-versus-level*: Χρησιμοποιείται για τον έλεγχο των υποθέσεων του μοντέλου, και εμφανίζεται όταν δεν υπάρχουν παράγοντες.
- *Residual plots*: Χρησιμοποιείται για τον έλεγχο των υποθέσεων του μοντέλου, π.χ. με τη δημιουργία ενός γραφήματος διασποράς των προβλεπόμενων παρατηρήσεων σε σχέση με τα τυποποιημένα σφάλματα της εξαρτημένης μεταβλητής, ώστε να εξεταστούν οι υποθέσεις κανονικότητας, ομοσκεδαστικότητας και ανεξαρτησίας.
- *Lack of fit*: Για τον έλεγχο της περίπτωσης που η σχέση μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών μπορεί να περιγραφεί επακριβώς από το μοντέλο.
- *General estimable function*: Μας επιτρέπει την κατασκευή ειδικών (custom) υποθέσεων βασιζόμενες σε μετασχηματισμούς γενικών συναρτήσεων.
- *Significance level*: Ορίζεται το επίπεδο στατιστικής σημαντικότητας που θα χρησιμοποιείται από τις αναλύσεις (π.χ. post hoc tests, ή διαστήματα εμπιστοσύνης). Η τιμή αυτή χρησιμοποιείται επίσης για τον υπολογισμό της παρατηρούμενης ισχύος του ελέγχου (p-value).

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

Descriptive Statistics

Dependent Variable: outcome

drugA	drugB	Mean	Std. Deviation	N
0	0	20,4300	2,41387	10
	1	23,8200	2,73813	10
	Total	22,1250	3,05543	20
1	0	24,2700	2,80913	10
	1	27,8100	3,67225	10
	Total	26,0400	3,66382	20
Total	0	22,3500	3,22156	20
	1	25,8150	3,75882	20
	Total	24,0825	3,87529	40

Εικόνα 6.162: Περιγραφικά μέτρα των ομάδων

Levene's Test of Equality of Error Variances^a

Dependent Variable: outcome

F	df1	df2	Sig.
,587	3	36	,627

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + drugA + drugB + drugA * drugB

Εικόνα 6.163: Έλεγχος της ισότητας των διασπορών

Tests of Between-Subjects Effects

Dependent Variable: outcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	273,391 ^a	3	91,130	10,505	,000
Intercept	23198,672	1	23198,672	2674,139	,000
drugA	153,272	1	153,272	17,668	,000
drugB	120,062	1	120,062	13,840	,001
drugA * drugB	,056	1	,056	,006	,936
Error	312,307	36	8,675		
Total	23784,370	40			
Corrected Total	585,698	39			

a. R Squared = ,467 (Adjusted R Squared = ,422)

Εικόνα 6.164: ANOVA Table

Parameter Estimates

Dependent Variable: outcome

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	27,810	,931	29,858	,000	25,921	29,699
[drugA=0]	-3,990	1,317	-3,029	,005	-6,661	-1,319
[drugA=1]	0 ^a
[drugB=0]	-3,540	1,317	-2,688	,011	-6,211	-,869
[drugB=1]	0 ^a
[drugA=0] * [drugB=0]	,150	1,863	,081	,936	-3,628	3,928
[drugA=0] * [drugB=1]	0 ^a
[drugA=1] * [drugB=0]	0 ^a
[drugA=1] * [drugB=1]	0 ^a

a. This parameter is set to zero because it is redundant.

Εικόνα 6.165: Ο πίνακας των συντελεστών (b) του γραμμικού μοντέλου

Από τον πίνακα διαπιστώνουμε ότι δεν υπάρχει αλληλεπίδραση ($p\text{-value} = 0,936 > 0,005$), ενώ οι δύο πρώτες υποθέσεις απορρίπτονται στα επίπεδα στατιστικής σημαντικότητας $\alpha = 0,05$ και $\alpha = 0,01$. Συνεπώς οι διαφορές μεταξύ των μέσων των δύο στηλών και των δύο γραμμών είναι στατιστικά σημαντικές που σημαίνει ότι 1 μονάδα του ναρκωτικού A παράγει μεγαλύτερη διέγερση από 0 μονάδες του ναρκωτικού και παρόμοια 1 μονάδα του ναρκωτικού B παράγει μεγαλύτερη διέγερση από 0 μονάδες του ίδιου ναρκωτικού. Οπότε 1 μονάδα από τα ναρκωτικά A και B γενικά αυξάνει τη σωματική διέγερση χωρίς ταυτόχρονα να υπάρχει αλληλεπίδραση μεταξύ των δύο ναρκωτικών.

6.10 Friedman

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Non parametric Tests > Legacy Dialogs > Related Samples*, με σκοπό τον έλεγχο συσχέτισης επαναλαμβανόμενων μετρήσεων μιας συνεχούς μεταβλητής. Η μέθοδος χρησιμοποιείται όταν δεν ισχύουν οι υποθέσεις εφαρμογής της Repeated Measures ANOVA, δηλ. όταν οι κατανομές δεν είναι κανονικές και οι διασπορές δεν είναι ίσες. Η μέθοδος είναι μια επέκταση του κριτηρίου sign test για k συσχετιζόμενα δείγματα.

Για την εφαρμογή της μεθόδου οι παρατηρήσεις διατάσσονται κατά αύξουσα σειρά και σε κάθε παρατήρηση προσδιορίζεται η σχετική της θέση (rank) και στη συνέχεια υπολογίζεται το άθροισμα των σχετικών θέσεων.

Η ελεγχουσυνάρτηση προκύπτει από τη σχέση:

$$T = \frac{12}{nk(k+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3h(k+1) \sim X_{k-1}^2$$

Εξίσωση 6.52

Όπου:

$$R_j = \sum_{i=1}^{n_j} r_{ij}$$

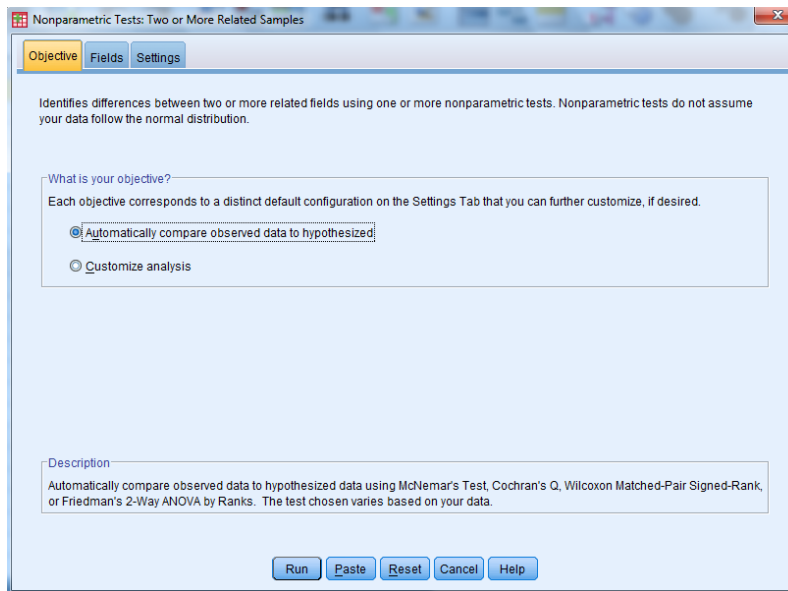
Εξίσωση 6.53

Το άθροισμα των θέσεων της κάθε j -ομάδας.

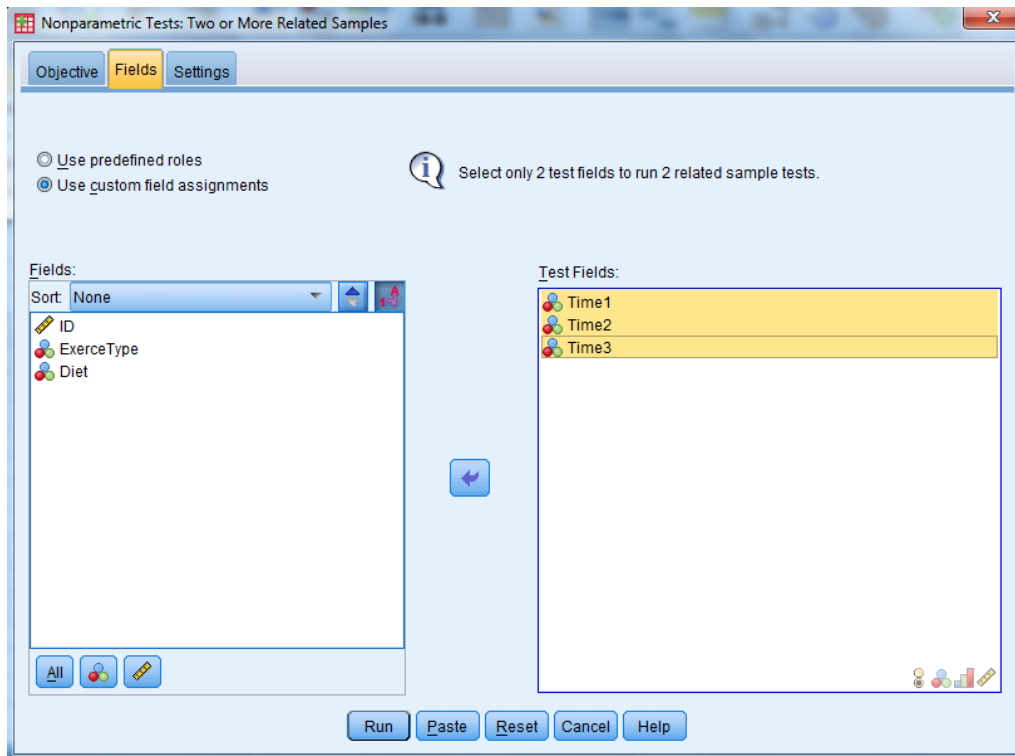
Η υπόθεση απορρίπτεται όταν:

$$T > X_{k-1; \alpha}^2$$

Εξίσωση 6.54

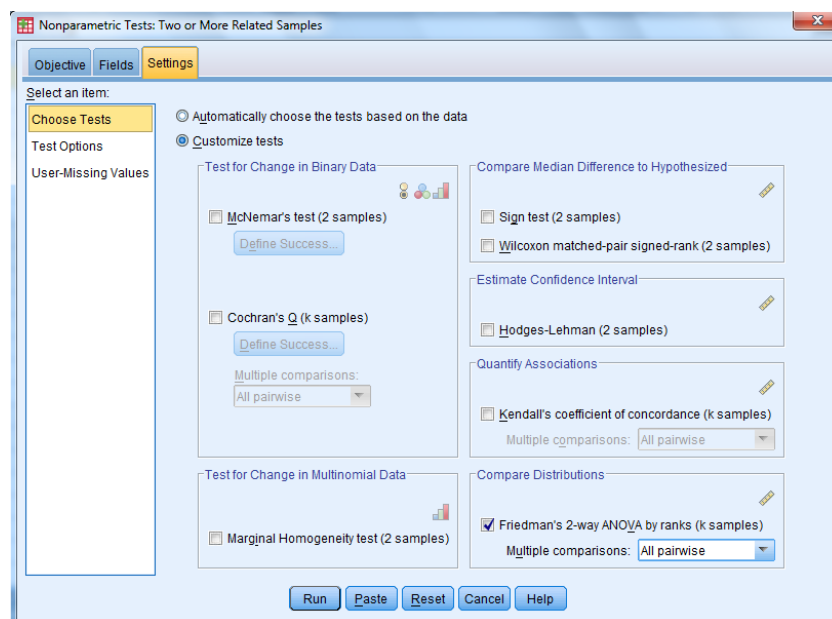


Εικόνα 6.166: Το menu Non parametric Tests: Two or More Related Samples > Step 1 Objective



Εικόνα 6.167: Το menu *Non parametric Tests: Two or More Related Samples* > Step 2 Fields

Στο πεδίο *Test Fields* εισάγουμε την ή τις επαναλαμβανόμενες μετρήσεις:



Εικόνα 6.168: Το menu *Non parametric Tests: Two or More Related Samples* > Step 3 Settings

Στην ενότητα *Settings* επιλέγουμε τον έλεγχο *Friedman's Two-way ANOVA by ranks (k samples)* για τη σύγκριση των κατανομών των επαναλαμβανόμενων μετρήσεων.

Σε αυτή την ενότητα παρέχονται επιπρόσθετα οι έλεγχοι:

- *Mc Nemar Test (2 samples)*: Βλ. επίσης ομώνυμη ενότητα.

- *Cohran's Q (k samples)*: Είναι μια μη παραμετρική διαδικασία η οποία ελέγχει κατά πόσο k δοκιμές (π.χ. θεραπείες) δίτιμων μεταβλητών έχουν ταυτόσημα αποτελέσματα. Ο έλεγχος είναι ισοδύναμος του Mc Nemar Test.

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

	Mean	Std. Deviation	N
Time1	93,41	6,062	29
Time2	102,10	14,477	29
Time3	105,00	18,949	29

Εικόνα 6.169: Περιγραφικά μέτρα των μέσων

	Mean Rank
Time1	1.26
Time2	2.26
Time3	2.48

Εικόνα 6.170: Οι μέσοι των βαθμών των ομάδων

N	29
Chi-Square	25.748
df	2
Asymp. Sig.	.000

a. Friedman Test

Εικόνα 6.171: Αποτελέσματα του ελέγχου Friedman

Από τα αποτελέσματα του ελέγχου διαπιστώνουμε ότι η υπόθεση της ισότητας των μέσων απορρίπτεται, συνεπώς υπάρχει διαφορά μεταξύ των επαναλαμβανόμενων μετρήσεων.

6.11 X²-test

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Descriptive Statistics > Crosstabs*, με σκοπό τον έλεγχο για την ύπαρξη σχέσης μεταξύ δύο ποιοτικών μεταβλητών.

Έχουμε ήδη αναφέρει ότι περιγραφικά οι από κοινού κατανομές συχνότητας δύο ποιοτικών μεταβλητών απεικονίζονται σε έναν πίνακα συνάφειας (contingency table) διπλής εισόδου.

Οι υποθέσεις που εξετάζονται είναι:

H_0 : Δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών A & B, ή αλλιώς οι παρατηρηθείσες συχνότητες είναι ίσες με τις αναμενόμενες.

H_1 : Υπάρχει συσχέτιση μεταξύ των μεταβλητών A & B, ή αλλιώς οι παρατηρηθείσες συχνότητες δεν είναι ίσες με τις αναμενόμενες.

Η στατιστική συνάρτηση δίνεται από τη σχέση:

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \sim X^2_{(r-1)(c-1);a}$$

Εξίσωση 6.55

Όπου:

O_{ij} : Οι παρατηρούμενες τιμές από κάθε κελί και e_{ij} : οι αναμενόμενες τιμές κάθε κελιού, οι οποίες υπολογίζονται από το πηλίκο του γινομένου του οριζόντιου και του κάθετου αθροίσματος του κάθε κελιού, προς το γενικό σύνολο.

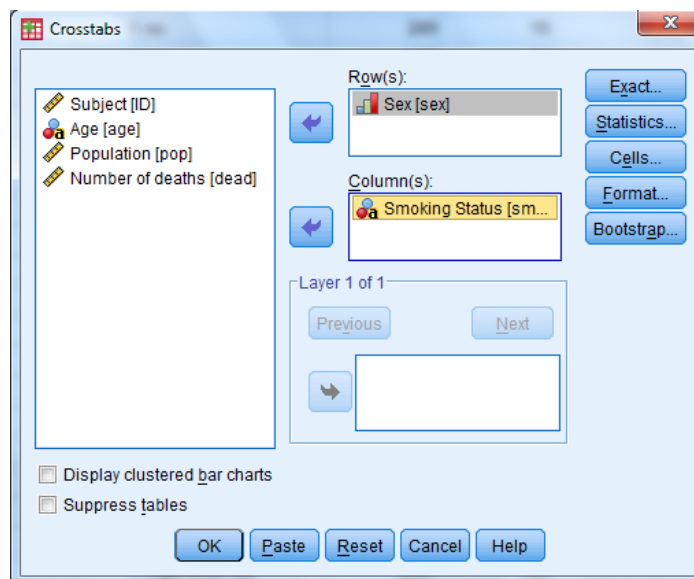
Υποθέσεις του ελέγχου:

- Όλες οι αναμενόμενες τιμές είναι μεγαλύτερες του ένα
- Περισσότερες από το 80% των αναμενόμενων τιμών ξεπερνά το 5

Όταν οι υποθέσεις εφαρμογής του X^2 τεστ δεν ισχύουν, τότε το SPSS εμφανίζει σημείωση που το αναφέρει και τότε λαμβάνουμε υπόψη το Fisher's Exact Test.

Έστω, ότι έχουμε τις μεταβλητές «Φύλο» και «Συνήθεια Καπνίσματος» και θέλουμε να εξετάσουμε την πιθανή συσχέτισή τους από τα αποτελέσματα των μετρήσεων σε μια ομάδα θεραπείας. Συνεπώς:

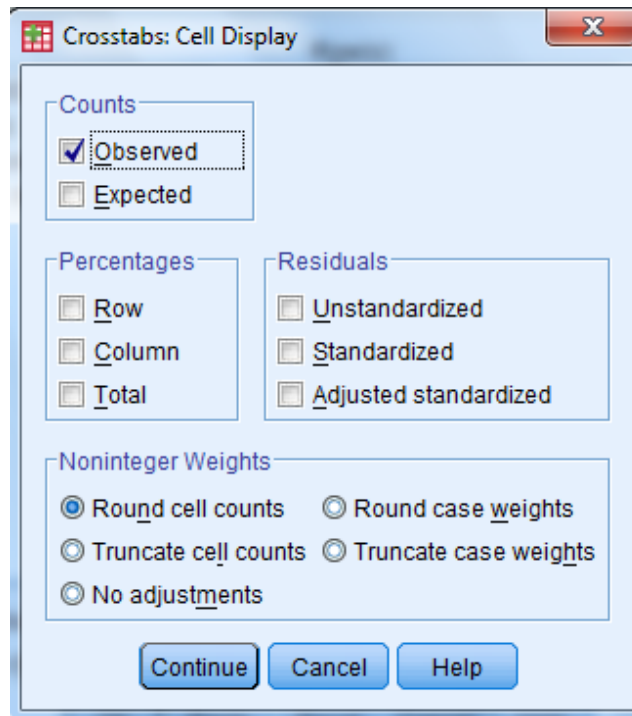
Οι δύο μεταβλητές εισάγονται στα πεδία Row(s) και Column(s). Συνηθίζουμε να εισάγουμε στις γραμμές τη μεταβλητή με τις περισσότερες κατηγορίες για λόγους ομοιομορφίας του πίνακα συνάφειας που θα δημιουργηθεί.



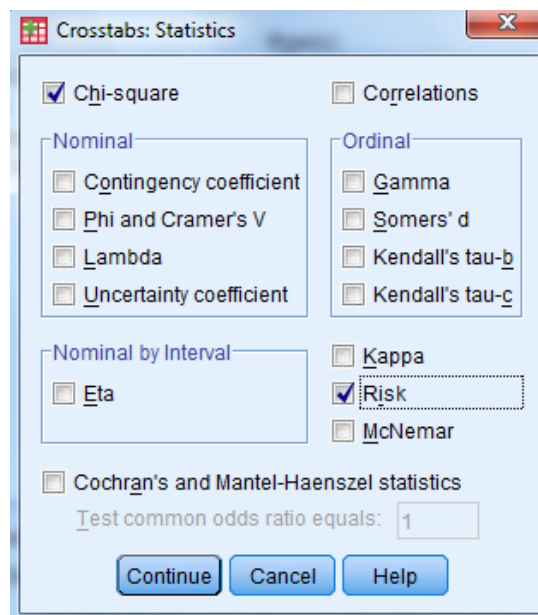
Εικόνα 6.172: Το menu Crosstabs

Από το πλήκτρο *Cell* αν επιθυμούμε επιλέγουμε επιπρόσθετα *Expected* για να εμφανιστούν οι αναμενόμενες συχνότητες. Περισσότερα για τις επιλογές του πλήκτρου *Cell* βλ. Παράρτημα «Crosstabs > Cell».

Από το πλήκτρο *Statistics* επιλέγουμε *Chi-square* για να υπολογιστεί το στατιστικό κριτήριο. Περισσότερα για τις επιλογές του πλήκτρου *Statistics* βλ. Παράρτημα «Crosstabs > Statistics».



Εικόνα 6.173: To menu Crosstabs > Cell



Εικόνα 6.174: To menu Crosstabs > Statistics

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

Sex * Smoking Status Crosstabulation

			Smoking Status				Total
			cigarPipeOnly	cigaretteOnly	cigarettePlus	no	
Sex	Male	Count	3	5	6	5	19
		Expected Count	4,8	4,8	4,8	4,8	19,0
	Female	Count	6	4	3	4	17
		Expected Count	4,3	4,3	4,3	4,3	17,0
Total		Count	9	9	9	9	36
		Expected Count	9,0	9,0	9,0	9,0	36,0

Εικόνα 6.175: Πίνακας συνάφειας του Φύλου με την Καπνιστική Συνήθεια

Ο πίνακας συνάφειας εμφανίζει τις παρατηρηθείσες και τις αναμενόμενες συχνότητες. Παράδειγμα, για το πρώτο κελί η αναμενόμενη τιμή δίνεται από τη σχέση $(19 \cdot 9) / 36$.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,118 ^a	3	,548
Likelihood Ratio	2,150	3	,542
N of Valid Cases	36		

a. 8 cells (100,0%) have expected count less than 5. The minimum expected count is 4,25.

Εικόνα 6.176: Αποτελέσματα του χ^2 τεστ

Από τον πίνακα διαπιστώνουμε ότι η μηδενική υπόθεση δεν μπορεί να απορριφθεί (p -value = 0,548 > 0,05) συνεπώς δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ του φύλου και τις καπνιστικής συνήθειας.

Παρακάτω παραθέτουμε μια περίπτωση δεδομένων στα οποία οι υποθέσεις εφαρμογής του χ^2 τεστ δεν ισχύουν, οπότε το SPSS εμφανίζει σχετική σημείωση που το αναφέρει, και σε αυτή την περίπτωση λαμβάνουμε υπόψη το Fisher's Exact Test.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,026 ^b	1	,873		
Continuity Correction ^a	,000	1	1,000		
Likelihood Ratio	,026	1	,873		
Fisher's Exact Test				1,000	,598
Linear-by-Linear Association	,025	1	,875		
N of Valid Cases	26				

a. Computed only for a 2x2 table

b. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 3,81.

Εικόνα 6.177: Αποτελέσματα του Fisher Exact τεστ στο παράδειγμα Διακοπής Σχολείου με Διακοπή Θεραπευτικής Αγωγής

Μέσω του χ^2 τεστ ελέγχουμε ανά δύο την ύπαρξη συσχετίσεων (μονοπαραγοντική ανάλυση). Ο έλεγχος αυτός είναι ένα μέτρο που μας δείχνει αν οι μεταβλητές συσχετίζονται ή όχι, δηλαδή δεν μας λέει το πόσο συσχετίζονται, παρά μόνο την ύπαρξη ή όχι κάποιας συσχέτισης.

Λόγος Συμπληρωματικών Πιθανοτήτων

Συνεπώς, για να εξεταστεί ο βαθμός συσχέτισης, στις περιπτώσεις 2x2 πινάκων συνάφειας, όπου ο έλεγχος προκύψει στατιστικά σημαντικός, υπολογίζουμε τα μέτρα κινδύνου π.χ. το OR (Odds Ratio).

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Διακοπή Σχολείου (OXI / NAI)	1,143	,224	5,841
For cohort Διακοπή Θεραπευτικής Σχέσης = OXI	1,059	,522	2,147
For cohort Διακοπή Θεραπευτικής Σχέσης = NAI	,926	,367	2,337
N of Valid Cases	26		

Εικόνα 6.178: Εκτίμηση μέτρων του βαθμού κινδύνου, στο παράδειγμα Διακοπής Σχολείου με Διακοπή Θεραπευτικής Αγωγής

Έστω ότι έχουμε τον ακόλουθο γενικό πίνακα, στον οποίο παρουσιάζονται οι συχνότητες ατόμων που εκτέθηκαν σε έναν παράγοντα και εμφάνισαν ή όχι μια ασθένεια. Εάν προκύψει ότι ο παράγοντας σχετίζεται με την ασθένεια θα πρέπει να διερευνηθεί ο βαθμός αυτής της συσχέτισης.

		Ασθένεια		Σύνολο
		Ναι (D^+)	Όχι (D^-)	
Παράγοντας	Ναι (E^+)	a	b	$a + b$
	Όχι (E^-)	c	d	$c + d$
Σύνολο		$a + c$	$b + d$	n

Εικόνα 6.179: 2x2 πίνακας συνάφειας Ασθένειας και Παράγοντα

Το OR είναι ο λόγος των συμπληρωματικών πιθανοτήτων, δηλαδή του λόγου πιθανότητας εμφάνισης της ασθένειας στα άτομα που εκτέθηκαν στον παράγοντα, προς τον λόγο πιθανότητας εμφάνισης της ασθένειας στα άτομα που δεν εκτέθηκαν στον παράγοντα, και δίνεται από τη σχέση:

$$OR = \frac{\frac{P(D^+ | E^+)}{P(D^- | E^+)}}{\frac{P(D^+ | E^-)}{P(D^- | E^-)}} = \frac{\frac{a / (a + c)}{c / (a + c)}}{\frac{b / (b + d)}{d / (b + d)}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$$

Εξίσωση 6.56

Ερμηνεία

- Όταν $OR = 1$ τότε δεν υπάρχει συσχέτιση μεταξύ της ασθένειας και της έκθεσης στον παράγοντα.
- Όταν $OR = a > 1$ τότε η έκθεση είναι επιβαρυντική για την ασθένεια, δηλαδή η πιθανότητα εμφάνισης της ασθένειας είναι a φορές μεγαλύτερη από την πιθανότητα μη εμφάνισης της ασθένειας. Δηλαδή υπάρχει μεγαλύτερη πιθανότητα, τόσες φορές όσες είναι το OR ασθένειας αν υπάρχει έκθεση στον παράγοντα, σε σχέση με την πιθανότητα ασθένειας αν δεν υπάρχει έκθεση στον παράγοντα.
- Όταν $OR = a < 1$ τότε η έκθεση στον παράγοντα είναι προστατευτική, δηλαδή εάν εκτεθείς στον παράγοντα έχεις $1-a$ φορές λιγότερη πιθανότητα να ασθηνήσεις.

Στο παράδειγμα Διακοπής Σχολείου με Διακοπή Θεραπευτικής Αγωγής, έχουμε $OR = 1,143$ δηλαδή οι ασθενείς που δεν διακόπτουν τη θεραπευτική αγωγή έχουν 1,143 φορές περισσότερη πιθανότητα να μη διακόψουν το σχολείο σε σχέση με αυτούς που διέκοψαν τη θεραπευτική αγωγή.

6.12 Mc Nemar test

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Descriptive Statistics > Crosstabs*, επιλέγοντας Mc Nemar από το πλήκτρο *Statistics*.

Σκοπός είναι να εξεταστεί η συμφωνία δύο κατηγορικών μεταβλητών, όταν οι παρατηρήσεις εμφανίζουν αντιστοιχία κατά ζεύγη. Ο έλεγχος αυτός είναι ο μη παραμετρικός ισοδύναμος του *T-test* για εξαρτημένα δείγματα. Σ' αυτή την περίπτωση τα δεδομένα παρουσιάζονται σε έναν πίνακα συνάφειας της μορφής:

		Κατηγορική B		Σύνολο
		+	-	
Κατηγορική A	+	c_1	d_1	$c_1 + d_1$
	-	d_2	c_2	$d_2 + c_2$
Σύνολο		$c_1 + d_2$	$d_1 + c_2$	n

Εικόνα 6.180: 2x2 Πίνακας συνάφειας κατά ζεύγη δεδομένων κατηγορικών μεταβλητών

Η υπόθεση που εξετάζεται είναι:

H_0 : Δεν υπάρχει διαφορά μεταξύ των απαντήσεων μεταβλητών *A* και *B*

H_1 : Υπάρχει διαφορά μεταξύ των απαντήσεων μεταβλητών *A* και *B*

Η ελεγχοσυνάρτηση δίνεται από τη σχέση:

$$X^2 = \frac{(d_1 - d_2)^2}{d_1 + d_2} \sim X^2_{1;a}$$

Εξίσωση 6.57

Ελεγχοσυνάρτηση με διόρθωση κατά Yates:

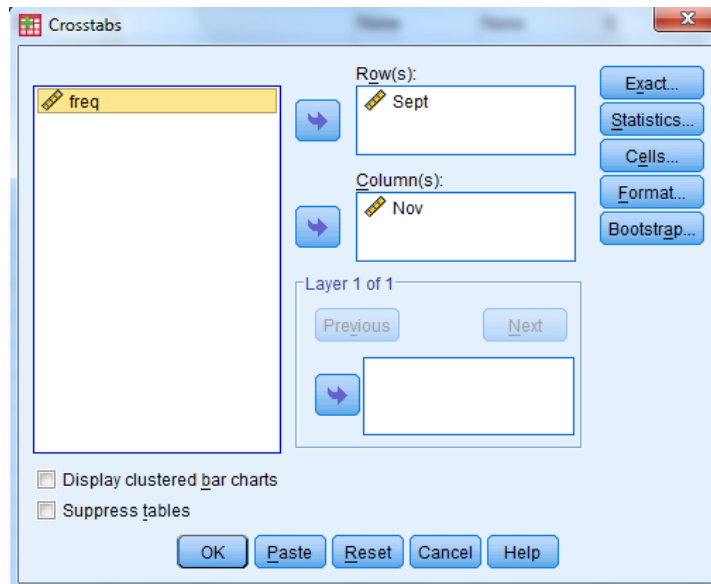
$$X^2 = \frac{(|d_1 - d_2| - 1)^2}{d_1 + d_2} \sim X^2_{1;a}$$

Εξίσωση 6.58

Η υπόθεση απορρίπτεται όταν:

$$X^2 > X^2_{1;a}$$

Εξίσωση 6.59



Εικόνα 6.181: Το menu Crosstabs

Από το πλήκτρο *Statistics* επιλέγουμε Mc Nemar για να υπολογιστεί το στατιστικό κριτήριο. Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:

Sept * Nov Crosstabulation

Count

		Nov		Total
		,00	1,00	
Sept	,00	60	15	75
	1,00	5	20	25
Total		65	35	100

Εικόνα 6.182: Πίνακας συνάφειας των ατόμων που κρυολόγησαν τους μήνες Σεπτέμβριο και Νοέμβριο

Τον πίνακα συνάφειας των συχνοτήτων των παρατηρήσεων για τους συνδυασμούς των κατηγορικών μεταβλητών.

Chi-Square Tests

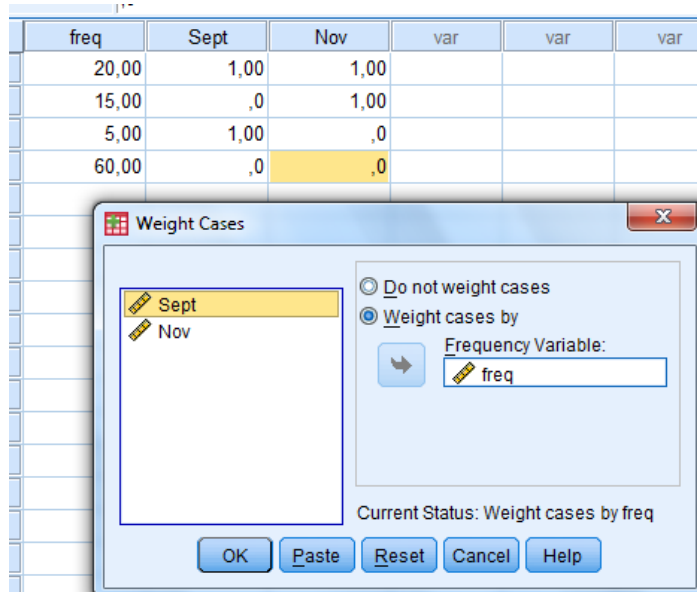
	Value	Exact Sig. (2-sided)
McNemar Test		,041 ^a
N of Valid Cases	100	

a. Binomial distribution used.

Εικόνα 6.183: Αποτελέσματα του ελέγχου

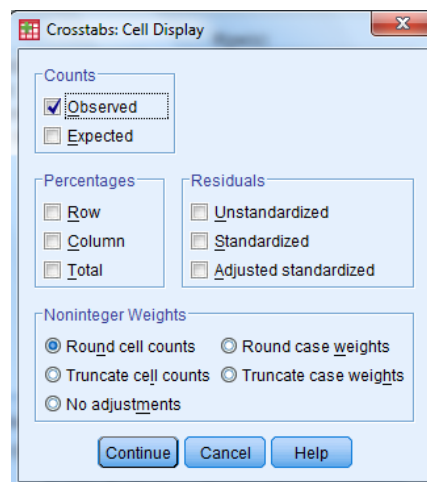
Η υπόθεση που εξετάζεται είναι εάν υπάρχει διαφορά στην πιθανότητα νόσησης μεταξύ των μηνών Σεπτεμβρίου και Νοεμβρίου. Από τα αποτελέσματα του ελέγχου διαπιστώνουμε ότι η μηδενική υπόθεση απορρίπτεται ($p\text{-value} = 0,041 < 0,05$), συνεπώς δεν υπάρχει διαφορά στην πιθανότητα νόσησης μεταξύ των δύο μηνών.

Σημείωση: Στην περίπτωση που τα δεδομένα έχουν γραφεί σε *Wide Format*, τότε θα πρέπει να σταθμιστούν μέσω της εντολής *Weight Cases* (βλ. ομώνυμη ενότητα).



Εικόνα 6.184: Στάθμιση δεδομένων που έχουν καταγραφεί σε Wide Format

6.13 Παράρτημα: Crosstabs > Cell



Εικόνα 6.185: Το menu Crosstabs > Cell

Συχνότητες (Counts)

- *Observed*: Επιλέγεται για να εμφανιστούν οι παρατηρούμενες συχνότητες των συνδυασμών των κατηγοριών της /των ποιοτικής/ών μεταβλητής/τών.
- *Expected*: Επιλέγεται για να εμφανιστούν οι αναμενόμενες συχνότητες των συνδυασμών των κατηγοριών της /των ποιοτικής/ών μεταβλητής/τών.

Ποσοστά (Percentages)

- *Row*: Επιλέγεται για να εμφανιστούν οι σχετικές συχνότητες (%) ως προς τα αθροίσματα στις γραμμές του πίνακα.
- *Column*: Επιλέγεται για να εμφανιστούν οι σχετικές συχνότητες (%) ως προς τα αθροίσματα στις στήλες του πίνακα.

- *Total*: Επιλέγεται για να εμφανιστούν οι σχετικές συχνότητες (%) ως προς το σύνολο των παρατηρήσεων.

Υπόλοιπα (Residuals)

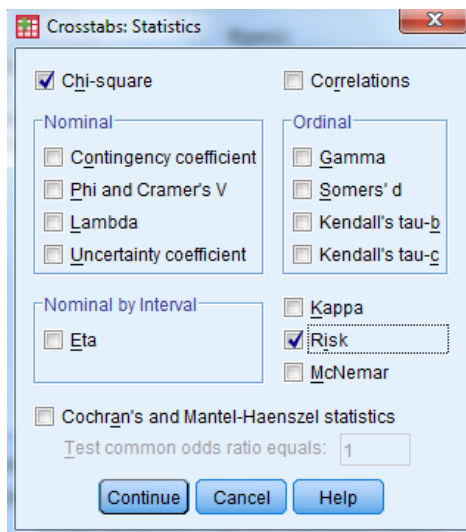
- *Unstandardized*: Επιλέγεται για να εμφανιστούν τα μη τυποποιημένα υπόλοιπα από τη διαφορά παρατηρούμενων και αναμενόμενων συχνοτήτων.
- *Standardized*: Επιλέγεται για να εμφανιστούν τα τυποποιημένα υπόλοιπα από τη διαφορά παρατηρούμενων και αναμενόμενων συχνοτήτων.
- *Adjusted standardized*: Επιλέγεται για να εμφανιστούν τα αναπροσαρμοσμένα υπόλοιπα. Ισχύει ότι με τη μηδενική υπόθεση ότι οι 2 μεταβλητές είναι ανεξάρτητες, τα αναπροσαρμοσμένα υπόλοιπα θα ακολουθούν την τυπική κανονική κατανομή, δηλαδή έχουν μέση τιμή 0 και τυπική απόκλιση 1. Έτσι, ένα προσαρμοσμένο υπόλοιπο που είναι πάνω από 1,96 (κατά συνθήκη χρησιμοποιείται 2,0) δείχνει ότι ο αριθμός των παρατηρήσεων στο εν λόγω κελί είναι σημαντικά μεγαλύτερος από ό,τι θα αναμενόταν εάν η μηδενική υπόθεση ήταν αληθής, με επίπεδο σημαντικότητας 0,05. Ένα αναπροσαρμοσμένο υπόλοιπο μικρότερο από -2,0 υποδεικνύει ότι ο αριθμός των παρατηρήσεων στο εν λόγω κελί είναι σημαντικά μικρότερος από ό,τι θα αναμενόταν εάν η μηδενική υπόθεση ήταν αληθής. Έτσι, ανάλογα με το πρόσημο του αναπροσαρμοσμένου υπολοίπου, διαπιστώνουμε αν σε κάθε κελί υπάρχουν λιγότερες ή περισσότερες παρατηρήσεις από ό,τι θα περίμενε κανείς αν οι δύο μεταβλητές ήταν ανεξάρτητες. Αν ο πίνακας συνάφειας είναι 2x2, τότε όλα τα προσαρμοσμένα σφάλματα θα έχουν την ίδια απόλυτη τιμή, με ακριβώς 2 από αυτά να είναι αρνητικά.
- *Noninteger Weights*: Συνήθως κάθε κελί περιέχει ακέραιους αριθμούς, δεδομένου ότι αποτυπώνουν τις συχνότητες εμφάνισης. Ωστόσο, στις περιπτώσεις που τα δεδομένα σταθμίζονται από μεταβλητή με βαρύτητα κλασματικών τιμών (π.χ. 1,25), τότε επίσης ενδέχεται ο αριθμός που θα προκύψει σε κάθε κελί να είναι κλασματικός. Σ' αυτές τις περιπτώσεις, υπάρχουν επιλογές με σκοπό την περικοπή ή στρογγυλοποίηση πριν ή μετά τον υπολογισμό των συχνοτήτων σε κάθε κελί.

Οι επιλογές που παρέχονται είναι:

- *Round Cell Counts*: Επιλέγεται έτσι ώστε οι σταθμισμένες συχνότητες των κελιών να χρησιμοποιηθούν όπως είναι, αλλά οι αθροιστικές σταθμισμένες τιμές στρογγυλοποιούνται πριν από τον υπολογισμό των στατιστικών μέτρων.
- *Truncate Cell Counts*: Επιλέγεται έτσι ώστε οι σταθμισμένες συχνότητες των κελιών να χρησιμοποιηθούν όπως είναι, αλλά οι αθροιστικές σταθμισμένες τιμές να περικοπούν πριν από τον υπολογισμό των στατιστικών μέτρων. Σ' αυτή την περίπτωση δεν γίνεται στρογγυλοποίηση στις τιμές αλλά αποκοπή από ένα σημείο και μετά. Για παράδειγμα, οι αριθμοί 1,4142 και 1,4987 περικόπτονται και οι δύο στον αριθμό 1,4.
- *Round Case Weights*: Επιλέγεται έτσι ώστε κάθε στάθμιση να στρογγυλοποιείται πριν τη χρήση της.
- *Truncate Case Weights*: Επιλέγεται έτσι ώστε κάθε στάθμιση να περικόπτεται πριν τη χρήση της.
- *No adjustments*: Επιλέγεται έτσι ώστε να μην απαιτούνται προσαρμογές. Οι σταθμίσεις χρησιμοποιούνται όπως είναι, το ίδιο και οι κλασματικές συχνότητες των κελιών. Ωστόσο, όταν πρέπει να υπολογιστούν ακριβή στατιστικά μέτρα (τα οποία είναι διαθέσιμα μόνο μέσω της επιλογής Exact tests), τότε οι αθροιστικές σταθμίσεις των κελιών είτε θα στρογγυλοποιηθούν είτε θα περικοπούν πριν τη χρήση τους στον υπολογισμό των στατιστικών μέτρων.

6.14 Παράρτημα: Crosstabs > Statistics

Από το πλήκτρο *Statistics* γίνεται η επιλογή των ελέγχων υποθέσεων και δεικτών συνάφειας μεταξύ των κατηγορικών μεταβλητών.



Εικόνα 6.186: Το menu *Crosstabs > Statistics*

- *Chi-square*: Επιλέγεται για τον υπολογισμό των ελέγχων Pearson X^2 τεστ, likelihood-ratio X^2 τεστ, Fisher X^2 exact test, και X^2 τεστ κατά Yates (διόρθωση συνέχειας) σε πίνακες 2x2. Το Fisher X^2 exact test υπολογίζεται για πίνακες 2x2 όταν δεν ισχύουν οι υποθέσεις του X^2 test, δηλ. όταν υπάρχει αναμενόμενη τιμή μικρότερη του ένα, ή όταν περισσότερες από το 80% των αναμενόμενων είναι μικρότερες του 5. Για πίνακες ανεξαρτήτων πλήθους γραμμών και στηλών υπολογίζονται τα Pearson X^2 τεστ και likelihood-ratio X^2 τεστ, ενώ όταν και οι δύο μεταβλητές είναι ποσοτικές τότε προσφέρεται και ο έλεγχος γραμμικής τάσης (the linear-by-linear association test).
- *Correlations*: Επιλέγεται για τον υπολογισμό του πίνακα συσχετίσεων, για πίνακες στους οποίους οι γραμμές και οι στήλες περιέχουν διατεταγμένες τιμές. Οι συσχετίσεις αποδίδονται από τον συντελεστή συσχέτισης του Spearman, r (μόνο για αριθμητικά δεδομένα). Περισσότερα για τον συντελεστή συσχέτισης r του Spearman βλ. παράγραφο *Correlations*.
- *Nominal*: Περιέχει επιλογές με σκοπό τον υπολογισμό στατιστικών μέτρων συνάφειας για ονομαστικά δεδομένα (κατηγορικά). Σε αυτή την υποενοότητα μπορούμε να επιλέξουμε:
- *Contingency coefficient*: Ο συντελεστής συνάφειας είναι ένα μέτρο της σύνδεσης συσχέτισης των μεταβλητών. Οι τιμές του κυμαίνονται μεταξύ 0 και 1, με τιμή 0 να μην υποδεικνύει συσχέτιση μεταξύ των γραμμών και των στηλών (δηλ. συσχέτισης των μεταβλητών) και τιμές κοντά στο 1 να υποδεικνύουν υψηλό βαθμό συσχέτισης μεταξύ των μεταβλητών. Η μέγιστη δυνατή τιμή εξαρτάται από τον αριθμό των σειρών και των στηλών σε έναν πίνακα.
- *Phi and Cramer's V*: Ο συντελεστής συνάφειας ϕ (Yule, 1912) είναι ένα μέτρο που βασίζεται στο X^2 του Pearson και προκύπτει από τη ρίζα του πηλίκου του X^2 με το μέγεθος δείγματος,

$$\phi = \sqrt{\frac{X^2}{n}}$$

Εξίσωση 6.60

Στη βιβλιογραφία ο συντελεστής αναφέρεται και ως συντελεστής μέσης τετραγωνικής συνάφειας του Pearson (Pearson's coefficient of mean square contingency). Για πίνακες 2x2 ο συντελεστής ϕ λαμβάνει τιμές στο διάστημα $-1, 1$ χωρίς ωστόσο αυτό να είναι απόλυτο.

Ο συντελεστής V (Crammer, 1946) είναι μια επέκταση του συντελεστή ϕ για $I \times J$ πίνακες συνάφειας με $I, J > 2$ και προκύπτει από τη σχέση:

$$V = \sqrt{\frac{\phi^2}{q-1}}$$

Εξίσωση 6.61

Όπου:

$$q = \min\{I, J\}$$

Εξίσωση 6.62

Εκφράζοντας τη συνάφεια μεταξύ δύο μεταβλητών ως ποσοστό της μέγιστης δυνατής μεταβλητότητάς τους. Ο συντελεστής V λαμβάνει τιμές στο διάστημα $[0,1]$ με $V = 1$ όταν οι δύο μεταβλητές έχουν ίσα περιθώρια αθροίσματα. Καθώς τα περιθώρια αθροίσματα διαφέρουν όλο και περισσότερο η τιμή του V πλησιάζει το 0.

- *Lambda*: Ο συντελεστής συνάφειας λ (Goodman & Kruskal, 1954) ανήκει στα μέτρα προγνωστικής συνάφειας, και αποτελεί ένα μέτρο της αναλογικής μείωσης του σφάλματος, όταν οι τιμές της ανεξάρτητης μεταβλητής χρησιμοποιούνται για να προβλέψουν τις τιμές της εξαρτημένης μεταβλητής. Ο συντελεστής λ λαμβάνει τιμές στο διάστημα $[0,1]$, με $\lambda = 0$, όταν οι μεταβλητές είναι ανεξάρτητες, και $\lambda = 1$, όταν με βάση τη μια μεταβλητή μπορεί να γίνει τέλεια πρόγνωση των τιμών της άλλης.
- *Uncertainty coefficient*: Ο συντελεστής αβεβαιότητας U (Theil's uncertainty coefficient, 1972) ανήκει και αυτός στα μέτρα προγνωστικής συνάφειας και επίσης, είναι γνωστός και ως συντελεστής εντροπίας (entropy coefficient). Διαφοροποιείται από τον συντελεστή λ με την έννοια ότι λαμβάνει υπόψη του ολόκληρη την κατανομή της εξαρτημένης μεταβλητής και όχι μόνο την πιο συχνή κατηγορία της (όπως συμβαίνει στον συντελεστή λ). Λαμβάνει τιμές στο διάστημα $[0,1]$, με $U = 0$ όταν οι μεταβλητές είναι ανεξάρτητες, και $U = 1$, όταν με βάση τη μια μεταβλητή μπορεί να γίνει τέλεια πρόγνωση των τιμών της άλλης. Για παράδειγμα, $U = 0,83$ συνεπάγεται ότι η γνώση της μιας μεταβλητής, μειώνει το σφάλμα πρόβλεψης των τιμών της άλλης κατά 83%. Το πρόγραμμα υπολογίζει και τις συμμετρικές και τις μη συμμετρικές εκδοχές του συντελεστή αβεβαιότητας.
- *Ordinal*: Περιέχει επιλογές με σκοπό τον υπολογισμό στατιστικών μέτρων συνάφειας για διατάξιμα δεδομένα. Σ' αυτή την υποενότητα μπορούμε να επιλέξουμε:
- *Gamma*: Ο συντελεστής γ (Goodman & Kruskal, 1954) είναι ένα μέτρο διατακτικής συνάφειας, που βασίζεται στη διαφορά «συμφωνιών» (concordant) και «ασυμφωνιών» (discordant). Λαμβάνει τιμές στο διάστημα $[-1, 1]$, με $\gamma = 1$ ή $\gamma = -1$ να υποδεικνύει ισχυρή θετική ή αρνητική συσχέτιση, ενώ για $\gamma = 0$ μικρή ή καμία συσχέτιση.
- *Somers' d*: Ο συντελεστής D (Somers, 1962) είναι ένα μέτρο για την πρόβλεψη της εξαρτημένης μεταβλητής γνωρίζοντας την ανεξάρτητη. Είναι παραλλαγή του μέτρου Gamma, λαμβάνοντας υπόψη τυχόν ισοβαθμίες (ties) των ζευγών των παρατηρήσεων. Λαμβάνει τιμές στο διάστημα $[-1, 1]$, με $D = 1$ όταν δεν υπάρχουν «ασύμφωνα» ζεύγη μεταξύ των μεταβλητών X και Y , και η διάταξη των X είναι σε πλήρη συμφωνία με τη διάταξη των τιμών της Y (ισχυρή συσχέτιση μεταξύ των δύο μεταβλητών), ενώ $D = -1$ όταν δεν υπάρχουν «σύμφωνα» ζεύγη μεταξύ των μεταβλητών X και Y , και η διάταξη των τιμών της X είναι σε πλήρη ασυμφωνία με τη διάταξη των τιμών της Y .
- *Kendall's tau-b*: Ο συντελεστής τ (Kendall, 1938) είναι ένα διατακτικό μέτρο συνάφειας, το οποίο έχει τρεις παραλλαγές. Ο συντελεστής τ_b λαμβάνει υπόψη τις περιπτώσεις ισοβαθμιών (ties). Λαμβάνει τιμές στο διάστημα $[-1, 1]$, με το πρόσημο να καθορίζει την κατεύθυνση της σχέσης, όπως και στα προηγούμενα μέτρα.

- *Kendall's tau-c*: Ο συντελεστής τ (Kendall, 1938) είναι ένα διατακτικό μέτρο συνάφειας, το οποίο έχει τρεις παραλλαγές. Ο συντελεστής τ_c , σε αντίθεση με τον προηγούμενο δείκτη, αγνοεί τις περιπτώσεις ισοβαθμιών (ties). Λαμβάνει και αυτός τιμές στο διάστημα $[-1, 1]$, με το πρόσημο ομοίως, να καθορίζει την κατεύθυνση της σχέσης.
- *Nominal by Interval*: Περιέχει επιλογές με σκοπό τον υπολογισμό στατιστικών μέτρων συνάφειας μεταξύ μιας (κωδικοποιημένης) κατηγορικής και μιας ποσοστικής μεταβλητής.
- *Eta*: Ο συντελεστής η είναι ένα μέτρο συσχέτισης μεταξύ δύο μεταβλητών. Λαμβάνει τιμές στο διάστημα $[0, 1]$, με $\eta = 0$ όταν δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών και $\eta = 1$ ή κοντά στο 1, όταν υπάρχει υψηλός βαθμός συσχέτισης. ο συντελεστής είναι κατάλληλος όταν η μια εξαρτημένη μεταβλητή μετράται σε κλίμακα (π.χ. το εισόδημα) και η άλλη ανεξάρτητη μεταβλητή είναι κατηγορική, με περιορισμένο αριθμό κατηγοριών (π.χ. το φύλο).
- *Kappa*: Ο συντελεστής k Cohen είναι ένα στατιστικό μέτρο της συμφωνίας μεταξύ των αξιολογήσεων δύο βαθμολογητών όταν και οι δύο βαθμολογούν το ίδιο αντικείμενο. Λαμβάνει τιμές στο διάστημα $[0, 1]$, με την τιμή 1 να δηλώνει τέλεια συμφωνία. Για να γίνει ο υπολογισμός του συντελεστή θα πρέπει και οι δύο μεταβλητές να έχουν ίδιο τύπο δεδομένων (string ή numeric) και το ίδιο καθορισμένο μήκος.
- *Risk*: Ο συντελεστής κινδύνου, υπολογίζεται για πίνακες 2×2 και είναι ένα μέτρο του βαθμού συσχέτισης δύο παραγόντων. Ορίζεται ως ο λόγος της επίπτωσης (αναλογία πιθανοτήτων) δύο ομάδων με διαφορετική έκθεση σε έναν παράγοντα. Αν το διάστημα εμπιστοσύνης του συντελεστή κινδύνου περιέχει την τιμή 1, τότε δεν είμαστε σε θέση να υποθέσουμε ότι κάποιος παράγοντας συνδέεται με το αποτέλεσμα/συμβάν. Η αναλογία πιθανοτήτων μπορεί να χρησιμοποιηθεί ως μια εκτίμηση του σχετικού κινδύνου ή όταν η εμφάνιση του παράγοντα είναι σπάνια.
- *McNemar*: Πρόκειται για μία μη παραμετρική στατιστική δοκιμασία για τον έλεγχο της συμφωνίας δύο κατηγορικών μεταβλητών, όταν οι παρατηρήσεις εμφανίζουν αντιστοιχία κατά ζεύγη. Ο έλεγχος αυτός είναι ο μη παραμετρικός ισοδύναμος του T-test για εξαρτημένα δείγματα, βλ. ομώνυμη ενότητα.
- *Cochran's and Mantel-Haenszel statistics*: Μ' αυτή την επιλογή λαμβάνονται οι στατιστικές Cochran και Mantel-Haenszel για τον έλεγχο της ανεξαρτησίας μεταξύ μιας δίτιμης μεταβλητής παράγοντα και μιας δίτιμης μεταβλητής απόκρισης, που εξαρτάται από μια μεταβλητή που ορίζει ένα ή περισσότερα στρώματα (ελέγχου). Θα πρέπει να ληφθεί υπόψη ότι ενώ άλλα στατιστικά μέτρα υπολογίζονται ξεχωριστά για κάθε στρώμα, οι στατιστικές Cochran και Mantel-Haenszel υπολογίζονται μία φορά για όλες τις στρώσεις.

Βιβλιογραφικές Αναφορές Κεφαλαίου

Ελληνόγλωσσες

Κουτσογέρας, Γ. (2010). *Μέτρα συνάφειας και μέτρα ασυμμετρίας για πίνακες συνάφειας*. Μεταπτυχιακή εργασία, τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιά.

Ξενόγλωσσες

Cramér, H. (1946). *Mathematical Methods of Statistics* (Princeton New Jersey: *Princeton Mathematical Series, NJ*, pp. 367-369.

Goodman, L. A. & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49(268), pp. 732-764.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, pp. 81-93.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review*, pp. 799-811.

Theil, H. (1972). *Statistical decomposition analysis; with applications in the social and administrative sciences* Amsterdam: North Holland Publishing Company.

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, (75-76) pp. 579-652.

IBM Knowledge Center. Retrieved August 30, 2015 from http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/idh_glm_u_sav.htm

Κεφάλαιο 7

Γραμμική και λογιστική παλινδρόμηση

Σύνοψη

Στο κεφάλαιο αυτό γίνεται ανάλυση της μεθοδολογίας της απλής και πολλαπλής γραμμικής παλινδρόμησης, και αναφορά στη μεθοδολογία της λογιστικής παλινδρόμησης. Γίνεται αρκετά λεπτομερής θεωρητική αναφορά σε κάθε μεθοδολογία καθώς και επεξήγηση των επιλογών και των παραμέτρων που παρέχονται από το πρόγραμμα. Σε κάθε περίπτωση υπάρχουν εφαρμογές σε δειγματικά αρχεία δεδομένων και επεξήγηση των αποτελεσμάτων. Σημειώνεται ότι χρησιμοποιήθηκαν δύο αρχεία δεδομένων (solar.sav και carcrash.sav) τα οποία χρησιμοποιούνται στο μάθημα Στατιστικά Πρότυπα του ΔΠΜΣ, Μαθηματική προτυποποίηση σε σύγχρονες τεχνολογίες και στην οικονομία του ΕΜΠ (υπεύθυνη μαθήματος Χ. Καρώνη).

Προσπαιτούμενη γνώση

Για την καλύτερη κατανόηση του κεφαλαίου αυτού οι αναγνώστες θα πρέπει να έχουν παρακολουθήσει το μάθημα των γενικευμένων γραμμικών μοντέλων σε προχωρημένο επίπεδο, καθώς μέσω των επιλογών του προγράμματος γίνεται εκτενής αναφορά σε πληθώρα κριτηρίων και ελέγχων. Ακόμα ως προσπαιτούμενη είναι η βασική γνώση των στατιστικών ελέγχων, της εκτιμητικής (κυρίως η έννοια των διαστημάτων εμπιστοσύνης), των κατανομών Student, X^2 και κανονικής, των πράξεων μεταξύ πινάκων καθώς και η γνώση της έννοιας των αντίστροφων πινάκων.

Από τα εγχειρίδια που παρουσιάζουν με πληρότητα την ύλη του συγκεκριμένου κεφαλαίου αναφέρονται ενδεικτικά τα βιβλία των Carver και Nash, Καρώνη, Οικονόμου και Καρώνη, Coakes και Steed, Field, Norusis, Howit Cramer και Montgomery and Peck κ.ά. (βλ. βιβλιογραφικές αναφορές). Ακόμα για το συγκεκριμένο κεφάλαιο όπως για τα Κεφάλαια 3-6 σημαντική πηγή πληροφοριών αποτελεί ο ιστότοπος του SPSS.

7.1 Απλή και Πολλαπλή Γραμμική Παλινδρόμηση (Linear Regression)

Στην απλή παλινδρόμηση ο στόχος είναι να βρεθεί η μαθηματική σχέση που συνδέει τις παρατηρήσεις δύο μεταβλητών. Έχουμε ήδη δει ότι ο συντελεστής συσχέτισης μας δείχνει το αν και κατά πόσο δύο μεταβλητές σχετίζονται, χωρίς όμως να μας παρέχει τον τρόπο με τον οποίο μεταβάλλονται οι τιμές της μιας σε σχέση με τις τιμές της άλλης.

Αρχικά, γίνεται η γραφική παράσταση των δεδομένων (scatter plot) ώστε να προσεγγιστεί η μορφή της καμπύλης που τα συνδέει. Οι παρατηρήσεις των μεταβλητών απεικονίζονται σε ένα σύστημα ορθογωνίων συντεταγμένων. Παρατηρώντας στη συνέχεια το διάγραμμα διασποράς προκύπτει το συμπέρασμα ότι δεν υπάρχει καμπύλη που να ενώνει όλα τα σημεία και για τον λόγο αυτό αναζητάμε εκείνη την καμπύλη που προσαρμόζεται όσο το δυνατό καλύτερα στα δεδομένα που υπάρχουν. Η απλούστερη μορφή συσχέτισης δύο μεταβλητών είναι η γραμμική, δηλ. θα μελετηθεί η απλή γραμμική παλινδρόμηση και στη συνέχεια θα γίνει μια γενίκευση για περισσότερες από δύο μεταβλητές (πολλαπλή γραμμική παλινδρόμηση).

Η γραμμική συσχέτιση εκφράζεται μαθηματικά από τη σχέση:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Εξίσωση 7.1

Όπου:

Y η εξαρτημένη μεταβλητή, x η ανεξάρτητη μεταβλητή, β_0 η τεταγμένη, δηλ. η τιμή της εξαρτημένης για $x = 0$, β_1 η κλίση της ευθείας, και ε_i το τυχαίο σφάλμα με $E(\varepsilon_i) = 0$, που είναι η απόκλιση της Y από την ευθεία γραμμικής παλινδρόμησης $E(Y_i) = \beta_0 + \beta_1 x$.

Η ευθεία γραμμικής παλινδρόμησης δεν περνάει απαραίτητα από όλα τα σημεία των παρατηρήσεων Y_i . Αυτό έχει ως αποτέλεσμα στις περισσότερες περιπτώσεις για κάθε x_i να παράγεται μια προβλεπόμενη τιμή για τη μεταβλητή:

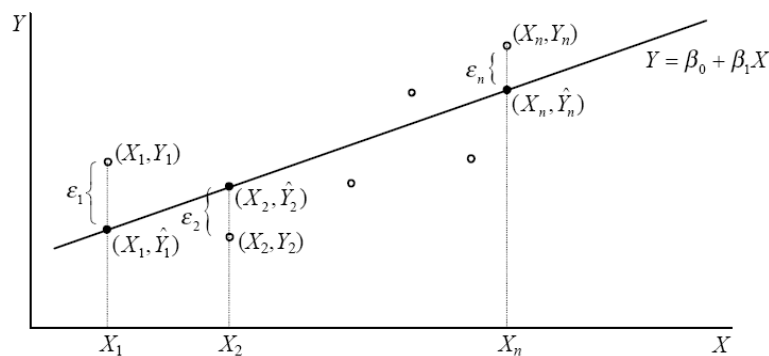
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Εξίσωση 7.2

Όπου:

$$Y_i = \hat{Y}_i + \varepsilon_i$$

Εξίσωση 7.3



Εικόνα 7.187: Γραφική απεικόνιση των δεδομένων της ευθείας παλινδρόμησης και των σφαλμάτων

Όπως βλέπουμε και από το παραπάνω σχήμα, σκοπός μας είναι να πετύχουμε μια καλή προσαρμογή της ευθείας στα δεδομένα, δηλαδή να ελαχιστοποιηθούν όσο το δυνατόν οι αποκλίσεις (σφάλματα ε_i) των δεδομένων από τις εκτιμώμενες τιμές. Συνεπώς, για την εκτίμηση των παραμέτρων β_0 και β_1 χρησιμοποιείται η μέθοδος των ελάχιστων τετραγώνων. Με τη μέθοδο αυτή η ευθεία που προσαρμόζεται καλύτερα στα δεδομένα που έχουν δοθεί είναι αυτή που ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων (καταλοίπων-residuals).

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

Εξίσωση 7.4

Από τη σχέση αυτή λαμβάνουμε τις πρώτες μερικές παραγώγους, τις οποίες θέτουμε ίσες με το μηδέν και από το 2x2 γραμμικό σύστημα εκτιμάμε τις παραμέτρους, όπου:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Εξίσωση 7.5

Και

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Εξίσωση 7.6

Αυτό που εξετάζουμε στα γραμμικά μοντέλα ($Y = B_0 + \beta_1 x_i$), είναι η ύπαρξη ή μη γραμμικής σχέσης μεταξύ της εξαρτημένης (dependent) μεταβλητής Y και της ανεξάρτητης (independent) μεταβλητής X . Οπότε ορίζουμε ως **μηδενική υπόθεση**: «Δεν υπάρχει γραμμική σχέση μεταξύ Y και X », δηλαδή:

$$H_0 : \hat{\beta}_1 = 0$$

Εξίσωση 7.7

Έναντι της **εναλλακτικής υπόθεσης**: «Υπάρχει γραμμική σχέση μεταξύ Y και X », δηλαδή:

$$H_1 : \hat{\beta}_1 \neq 0$$

Εξίσωση 7.8: Δίπλευρος έλεγχος

Ο έλεγχος γίνεται σε $(1-\alpha)\%$ επίπεδο σημαντικότητας και η στατιστική συνάρτηση υπό την H_0 είναι:

$$T = \frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)}$$

Εξίσωση 7.9

Ακολουθεί την t -κατανομή με $n-2$ βαθμούς ελευθερίας (T-test). Περιοχή απόρριψης: απορρίπτουμε την H_0 αν:

$$|T| > t_{n-2; \alpha/2}$$

Εξίσωση 7.10: Δίπλευρος έλεγχος

Αντίστοιχος έλεγχος για τη διατομή (intercept) β_0 δεν έχει νόημα. Στο σύνολο των παρατηρήσεων η ολική διακύμανση, δηλαδή η ολική μεταβλητότητα των παρατηρήσεων (SST), δίδεται από τη σχέση:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2$$

Εξίσωση 7.11

Η οποία αναλύεται σε δύο μέρη:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n [(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})]^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

Εξίσωση 7.12

Αποδεικνύεται ότι:

$$2 \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i) = 0$$

Εξίσωση 7.13

Ένα μέρος εξηγείται από την ευθεία παλινδρόμησης (regression sum of squares):

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

Εξίσωση 7.14

Ένα άλλο παραμένει ανεξηγήμενο (error sum of squares):

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Εξίσωση 7.15

Οπότε:

$$SST = SSR + SSE$$

Εξίσωση 7.16

Η ανάλυση γίνεται για να ερμηνευτεί η μεταβλητότητα των παρατηρήσεων που μελετώνται. Βέβαια ο αρχικός στόχος είναι να ελαττωθεί το μέρος της συνολικής μεταβλητότητας που παραμένει ανεξηγήμενο.

Συνεπώς, ένα μέτρο για να κρίνουμε αν η ευθεία παλινδρόμησης προσαρμόζεται ικανοποιητικά στα δεδομένα είναι ο συντελεστής προσδιορισμού (coefficient of determination) που δίνεται από τη σχέση:

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

Εξίσωση 7.17

Πρέπει να είναι όσο το δυνατό μεγαλύτερος. Αν $R^2 = 0,78$ σημαίνει ότι 78% της μεταβλητότητας των αποδόσεων της επένδυσης σχετίζεται με τις αποδόσεις δείκτη αναφοράς και το 22% δεν συνδέεται. Μ' άλλα λόγια ο συντελεστής προσδιορισμού R^2 εκφράζει το ποσοστό της συνολικής μεταβλητότητας που εξηγείται από το μοντέλο παλινδρόμησης. Μια παραλλαγή του συντελεστή προσδιορισμού είναι ο τροποποιημένος συντελεστής προσδιορισμού (adjusted coefficient of the termination) που δίνεται από τη σχέση:

$$R^2_{adj} = 1 - \frac{\frac{SSE}{n-2}}{\frac{SST}{n-1}} = 1 - (1 - R^2) \frac{n-1}{n-2}$$

Εξίσωση 7.18

Χρησιμοποιείται για τη σύγκριση nested μοντέλων. Όλα τα παραπάνω συνοψίζονται στον πίνακα Ανάλυσης Διασποράς της Παλινδρόμησης (ANOVA-Table).

ANOVA-Table στην απλή γραμμική παλινδρόμηση					
Source (Πηγή διακύμανσης)	SS (Αθροισμα τετραγώνων)	Df (β.ε.)	MS (Μέσα τετραγώνων)	F	Prob > F
Regression (Μοντέλο)	$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	1	$MSR = SSR$	$F = \frac{MSR}{MSE}$	$P(F > F_{1,n-2;a})$
Residuals (Σφάλματα)	$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	n-2	$MSE = \frac{SSE}{n-2}$		
Total (Ολικό)	$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$	n-1			

Πίνακας 7.1: F-test για το γραμμικό μοντέλο

Είδαμε ότι η ολική διακύμανση των δεδομένων χωρίζεται σε δύο μέρη, ένα εξαιτίας της παλινδρόμησης και ένα που παραμένει ανερμήνευτο.

Από τη θεωρία πιθανοτήτων ξέρουμε ότι:

$$\frac{SSR}{\sigma^2} = \frac{(n-2)MSE}{\sigma^2} \sim X_{n-2}^2$$

Εξίσωση 7.19

Από την άλλη αν $\beta_0 = 0$ έχουμε ότι:

$$\frac{SSR}{\sigma^2} \sim X_1^2$$

Εξίσωση 7.20

Ο λόγος αυτών των δύο θα μας δώσει:

$$F = \frac{MSR}{MSE} \sim F_{1,n-2;a}$$

Εξίσωση 7.21

Ο F-έλεγχος για το αν υπάρχει γραμμική σχέση μεταξύ των X και Y (δηλαδή για το αν μια ευθεία είναι χρήσιμη για την ερμηνεία μέρους της μεταβλητότητας των δεδομένων) βασίζεται στην παρατήρηση ότι:

$$E(MSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

Εξίσωση 7.22

Και:

$$E(MSE) = E(S^2) = \sigma^2$$

Εξίσωση 7.23

Όταν το μοντέλο παλινδρόμησης είναι σωστά καθορισμένο. Εάν η $\beta_1 \approx 0$ (slope), δηλαδή η παλινδρόμηση δεν προσθέτει τίποτα επιπλέον για την ερμηνεία της μεταβλητότητας των παρατηρήσεων τα δύο μέσα αθροίσματα τετραγώνων MSR και MSE και έχουν την ίδια μέση τιμή σ^2 . Επομένως έχουμε το λόγο:

$$\frac{MSR}{MSE} < 1$$

Εξίσωση 7.24

Αν η υπόθεση της γραμμικής σχέσης των X και Y δεν ευσταθεί τότε έχουμε το λόγο:

$$\frac{MSR}{MSE} > 1$$

Εξίσωση 7.25

Αν ευσταθεί η γραμμική σχέση ($\beta_1 \neq 0$). Ο F -έλεγχος για την ύπαρξη ή όχι γραμμικής σχέσης των X και Y ορίζεται ως εξής:

H_0 : δεν υπάρχει γραμμική σχέση μεταξύ των X και Y

Εξίσωση 7.26

H_1 : υπάρχει γραμμική σχέση μεταξύ των X και Y

Εξίσωση 7.27

$$F = \frac{MSR}{MSE} \sim F_{1,n-2;\alpha}$$

Εξίσωση 7.28

Απορρίπτουμε την H_0 αν:

$$F > F_{1,n-2;\alpha}$$

Εξίσωση 7.29

Στην απλή γραμμική παλινδρόμηση το F -test είναι ισοδύναμο με το T -test για μηδενική κλίση. Αυτό συμβαίνει επειδή $T^2 = F$ όπου $T \sim t_{n-2;\alpha}$ και $F \sim F_{1,n-2;\alpha}$.

Πολλαπλή γραμμική παλινδρόμηση

Η πολλαπλή γραμμική παλινδρόμηση (Multiple Linear Regression) είναι επέκταση της απλής γραμμικής παλινδρόμησης από το μοντέλο.

$$Y_i = \beta_0 + \beta_1 x_i$$

Εξίσωση 7.30

Μεταφερόμαστε στο πολλαπλό γραμμικό μοντέλο:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad i = 1, \dots, n$$

Εξίσωση 7.31

Ο όρος γραμμικό μοντέλο αναφέρεται στις παραμέτρους β_j , $j = 1, \dots, p-1$ (και όχι στις ανεξάρτητες μεταβλητές x_{ij}) και κατ' επέκταση τα μοντέλα:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i \quad i = 1, \dots, n$$

Εξίσωση 7.32: Πολυωνυμικό μοντέλο

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad i = 1, \dots, n$$

Εξίσωση 7.33: Μοντέλο με αλληλεπίδραση

Θεωρούνται ως πολλαπλά γραμμικά μοντέλα. Απεικόνιση της πολλαπλής παλινδρόμησης με χρήση πινάκων. Για n ανεξάρτητες παρατηρήσεις έχουμε:

$$\begin{aligned} Y_1 &= \beta_0 X_{10} + \beta_1 X_{11} + \dots + \beta_{p-1} X_{1,p-1} + \varepsilon_1 \\ Y_2 &= \beta_0 X_{20} + \beta_1 X_{21} + \dots + \beta_{p-1} X_{2,p-1} + \varepsilon_2 \\ &\dots\dots\dots \\ Y_n &= \beta_0 X_{n0} + \beta_1 X_{n1} + \dots + \beta_{p-1} X_{n,p-1} + \varepsilon_n \end{aligned}$$

Εξισώσεις 7.34

Ισοδύναμα:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = X \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \Rightarrow Y = X\beta + \varepsilon$$

Εξίσωση 7.35

Όπου:

$$X = \begin{bmatrix} X_{10} & X_{11} & \dots & X_{1,p-1} \\ X_{20} & X_{21} & \dots & X_{2,p-1} \\ \dots & \dots & \dots & \dots \\ X_{n0} & X_{n1} & \dots & X_{n,p-1} \end{bmatrix}$$

Εξίσωση 7.36: Πίνακας σχεδιασμού (Design Matrix)

Ομοίως, με την απλή παλινδρόμηση για την εκτίμηση των συντελεστών:

$\hat{\beta}_j, j = 1, \dots, p - 1$, εφαρμόζουμε τη μέθοδο των ελαχίστων τετραγώνων

Εξίσωση 7.37

$$SSE(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \varepsilon_i \dots + \hat{\beta}_{p-1} X_{i,p-1}) \right]^2$$

$$= (Y - X \hat{\beta})^T (Y - X \hat{\beta})$$

Εξίσωση 7.38

Παίρνοντας τις πρώτες μερικές παραγώγους προκύπτει ένα σύστημα με p εξισώσεις και p αγνώστους. Ισοδύναμα:

$$X \hat{\beta} = Y \Rightarrow X^T X \beta = X^T Y \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

Εξίσωση 7.39

Οι εκτιμήτριες β μπορούν να υπολογιστούν είτε λύνοντας το $p \times p$ σύστημα είτε πολλαπλασιάζοντας τους πίνακες:

ANOVA-Table στο πολλαπλό γραμμικό μοντέλο					
Source (Πηγή μεταβλητότητας)	SS (Άθροισμα τετραγώνων)	Df (β.ε.)	MS (Μέσα τετραγώνων)	F	Prob > F
Regression (Μοντέλο)	$SSR = \hat{\beta}^T X^T Y - n\bar{Y}^2$	p-1	$MSR = \frac{SSR}{p-1}$	$F = \frac{MSR}{MSE}$	$P(F > F_{p-1, n-p; \alpha})$
Residuals (Σφάλματα)	$SSE = Y^T Y - \hat{\beta}^T X^T Y$	n-p	$MSE = \frac{SSE}{n-p}$		
Total (Ολικό)	$SST = Y^T Y - n\bar{Y}^2$	n-1			

Πίνακας 7.2: F- έλεγχοι στην πολλαπλή γραμμική παλινδρόμηση

F-έλεγχος για τη σημαντικότητα όλου του μοντέλου (overall F-test)

Όπως και στην απλή παλινδρόμηση, έτσι και εδώ, έχουμε έναν F- έλεγχο για όλο το μοντέλο. Οπότε:

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_{p-1} = 0$$

Εξίσωση 7.40

Έναντι της:

$$H_1 : \text{τουλάχιστον ένα } \beta_j \neq 0, j = 1, \dots, p - 1$$

Εξίσωση 7.41

Συνεπώς:

$$F = \frac{MSR}{MSE} = \frac{SSR / (p-1)}{SSE / (n-p)} \sim F_{p-1, n-p; \alpha}$$

Εξίσωση 7.42

Απορρίπτεται η H_0 για:

$$F > F_{p-1, n-p; \alpha}$$

Εξίσωση 7.43

Το SSR λαμβάνει τη συνεισφορά όλων των ανεξάρτητων μεταβλητών. Ωστόσο δεν συνεισφέρουν όλες το ίδιο σημαντικά στην ερμηνεία της μεταβλητότητας των Y_i και επίσης κάποιες από αυτές σχετίζονται μεταξύ τους, άρα δεν είναι αναγκαίες μέσα στο μοντέλο.

Partial F-tests (variables-added-in-order ή Type I F-tests)

Ένας τρόπος για να ελεγχθεί η μεμονωμένη συνεισφορά κάθε ανεξάρτητης μεταβλητής στο τετραγωνικό άθροισμα της παλινδρόμησης (SSR) είναι η χρήση των partial F -tests. Το SSR μπορεί να αναλυθεί στο άθροισμα:

$$SSR = SS(\beta_1) + SS(\beta_2 | \beta_1) + \dots + SS(\beta_{p-1} | \beta_1, \dots, \beta_{p-2})$$

Εξίσωση 7.44

Όπου, $SS(\beta_2 | \beta_1)$ το επιπλέον άθροισμα τετραγώνων από την προσθήκη της X_2 στο μοντέλο παλινδρόμησης, δηλαδή η επιπλέον μεταβλητότητα της Y που ερμηνεύεται από την προσθήκη της X_2 στο μοντέλο αφού έχει ήδη συμπεριληφθεί η X_1 .

Παρατηρούμε ότι η σειρά της προσθήκης των ανεξάρτητων μεταβλητών είναι τρομερά σημαντική για τον υπολογισμό αυτών των partial F -tests. Εάν θέλουμε απλά να εξετάσουμε το κατά πόσο για παράδειγμα μια X_1 είναι σημαντική για την πρόβλεψη της Y τότε θα κάνουμε μια μονοπαραγοντική ανάλυση (απλή γραμμική παλινδρόμηση).

Για να απαντήσουμε όμως στο ερώτημα το κατά πόσο η προσθήκη της X_2 συνεισφέρει στατιστικά σημαντικά στην πρόβλεψη της Y αφού έχουμε ήδη λογαριάσει τη συνεισφορά της X_1 , χρησιμοποιούμε ένα partial F -test. Όπου:

$$F = \frac{SSR(\beta_1, \beta_2) - SSR(\beta_1) / 1}{SSR(\beta_1, \beta_2) / (n-p)}$$

Εξίσωση 7.45

Πιο γενικά:

$$F(X_1^*, X_2^*, \dots, X_p^* / X_1, X_2, \dots, X_p) = \frac{[SSR(\beta_1^*, \beta_2^*, \dots, \beta_k^*, \beta_1, \dots, \beta_p) - SSR(\beta_1, \dots, \beta_p)] / k}{SSR(\beta_1^*, \beta_2^*, \dots, \beta_k^*, \beta_1, \dots, \beta_p) / (n-p-k-1)}$$

Εξίσωση 7.46

Ένας άλλος τρόπος για να ελέγχουμε αν η προσθήκη μιας νέας μεταβλητής, αφού έχουμε ήδη p μεταβλητές στο μοντέλο, συμβάλλει στατιστικά σημαντικά στην πρόβλεψη της Y είναι η χρήση ενός T -test.

Ορίζουμε:

$$H_0 : \beta_p = 0$$

Εξίσωση 7.47

Έναντι της:

$$H_1 : \beta_p \neq 0$$

Εξίσωση 7.48

$$T = \frac{\hat{\beta}_p}{s.e.(\hat{\beta}_p)} \sim t_{n-p-2}$$

Εξίσωση 7.49

Απορρίπτουμε την H_0 για:

$$T > t_{n-p-2;\alpha}$$

Εξίσωση 7.50

Παρατήρηση:

$$T^2 = \text{partial}F(X^* | X_1, \dots, X_p)$$

Εξίσωση 7.51

Variables-added-last ή Type III F-tests

Ένας τελευταίος F-έλεγχος είναι ο Type III F-test που βασίζεται στο άθροισμα τετραγώνων κάθε μεταβλητής αν είναι γνωστές όλες οι υπόλοιπες μεταβλητές του μοντέλου. Εάν έχουμε p -μεταβλητές τότε υπολογίζουμε τα:

$$SSR(X_1 | X_2, X_3, \dots, X_p)$$

$$SSR(X_2 | X_1, X_3, \dots, X_p)$$

.....

$$SSR(X_p | X_1, X_2, \dots, X_{p-1})$$

Εξισώσεις 7.52

Όπου η υπό εξέταση μεταβλητή προστίθεται κάθε φορά τελευταία στο μοντέλο.

Υποθέσεις του πολλαπλού γραμμικού μοντέλου

1. **Ανεξαρτησία:** Οι παρατηρήσεις Y_i πρέπει να είναι στατιστικά ανεξάρτητες μεταξύ τους (όταν λαμβάνονται πολλαπλές μετρήσεις από το ίδιο υποκείμενο τότε θα πρέπει να χρησιμοποιηθούν άλλες τεχνικές αναλύσεις).
2. **Γραμμικότητα:** Ο μέσος των παρατηρήσεων Y_i για κάθε συνδυασμό των X_1, X_2, \dots, X_{p-1} είναι ένας γραμμικός συνδυασμός αυτών:

$$\mu_{Y|X_1, X_2, \dots, X_{p-1}} = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1}$$

Εξίσωση 7.53

3. **Ομοσκεδαστικότητα:** Η διασπορά των Y_i είναι ίδια για κάθε συνδυασμούς X_1, X_2, \dots, X_{p-1} με:

$$\sigma_{Y/X_1, X_2, \dots, X_{p-1}}^2 \equiv \sigma^2 \text{ ή αλλιώς } \sigma_{Y/X_1, X_2, \dots, X_{p-1}}^2 \equiv \sigma^2$$

Εξίσωση 7.54

4. **Κανονικότητα:** Για κάθε συνδυασμό X_1, X_2, \dots, X_{p-1} η μεταβλητή Y κατανέμεται κανονικά:

$$Y_i \sim N(\mu_{Y|X_1, \dots, X_{p-1}}, \sigma^2) \text{ ή αλλιώς } \varepsilon_i \sim N(0, \sigma^2)$$

Εξίσωση 7.55

Κατ' επέκταση των παραπάνω οι προϋποθέσεις για την εφαρμογή της μεθόδου είναι:

1. Τα σφάλματα:

$$\varepsilon_i = Y_i - \hat{Y}_i$$

Εξίσωση 7.56

Έχουν μηδενική μέση τιμή και άγνωστη διακύμανση σ^2 .

2. Τα $\varepsilon_1, \dots, \varepsilon_n$ είναι ασυσχέτιστα δηλαδή:

$$\forall i \neq j \text{ έχουμε: } \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

Εξίσωση 7.57

3. Άμεση συνέπεια των παραπάνω υποθέσεων είναι ότι κάθε παρατήρηση Y_i έχει:

$$E(Y_i) = \beta_0 + \beta_1 x_i = \mu_{Y/X}$$

Εξίσωση 7.58

Με κοινή διακύμανση σ^2 και ότι οι παρατηρήσεις Y_i και Y_j όπου $i \neq j$ είναι ασυσχέτιστες.

4. Τα σφάλματα ε_i ακολουθούν $N(0, \sigma^2)$, το οποίο συνεπάγεται ότι τα ε_i εκτός από ασυσχέτιστα είναι και ανεξάρτητα μεταξύ τους.

Ο έλεγχος της εγκυρότητας του μοντέλου γίνεται με τη χρήση των σφαλμάτων (καταλοίπων). Τα σφάλματα χρησιμοποιούνται επίσης, για την ανίχνευση παρατηρήσεων που συμπεριφέρονται αντιφατικά με όλο το μοντέλο. Πολύ υψηλά σφάλματα μπορεί να υποδεικνύουν μια ασυνήθιστα υψηλή παρατήρηση ή απλά μια κοινή εφαρμογή. Από την άλλη, μια παρατήρηση μπορεί να αξίζει να ληφθεί υπόψη ακόμα και αν το σφάλμα της είναι μικρό, επειδή καθορίζει σε μεγάλο βαθμό την ευθεία παλινδρόμησης.

Αναλυτικότερα τα διαγνωστικά ελέγχου των υποθέσεων του πολλαπλού μοντέλου παλινδρόμησης θα παρουσιαστούν ακολούθως μέσα από τις δυνατές επιλογές του προγράμματος.

Βήματα για την επιλογή του «βέλτιστου» μοντέλου παλινδρόμησης

Ορίζουμε το μέγιστο δυνατό μοντέλο (πλήρες) δηλαδή το μοντέλο που θα έχει:

- όλους τους predictors $X_i, i=1, \dots, k$,
- όλους τους μεγαλύτερου βαθμού όρους (που επιθυμούμε),
- τις μεταξύ τους αλληλεπιδράσεις.

Εάν το πλήθος των μεταβλητών- predictors και της διατομής είναι $k+1=n$ (όπου n το πλήθος των παρατηρήσεων), τότε οι βαθμοί ελευθερίας του σφάλματος θα είναι μηδέν. Έτσι δεν θα έχει νόημα πόσους predictors θα επιλέξουμε γιατί:

$$R^2 = 1 - \frac{SSE}{SST} = 1,$$

Εξίσωση 7.59

αδιαφορώντας αν το μοντέλο παλινδρόμησης είναι λογικό ή όχι.

Ελάχιστοι απαιτούμενοι βαθμοί ελευθερίας για το σφάλμα $n-k-1 > 10$. Μια άλλη πρόταση είναι η επιλογή $n > 5k$ ή ακόμα $n \geq 10k$. Έτσι, αν έχουμε 50 παρατηρήσεις το μεγαλύτερο μοντέλο θα έχει predictors μεταξύ $k=5$ και $k-10$.

Ορίζουμε το κριτήριο με το οποίο θα γίνει η επιλογή των predictors X_i . Υποθέτουμε ότι το πλήρες μοντέλο είναι το:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \dots + \beta_k X_k + \varepsilon$$

Εξίσωση 7.60

Το μειωμένο μοντέλο είναι το:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \dots + \varepsilon \text{ όπου: } p \leq k$$

Εξίσωση 7.61

Υποψήφια κριτήρια για την επιλογή μοντέλου είναι:

α) Ο συντελεστής προσδιορισμού R^2 :

$$R^2_{(Y|X_1, \dots, X_p)} = 1 - \frac{SSE}{SST}$$

Εξίσωση 7.62α

Μεγιστοποιεί το πλήρες μοντέλο. Η εναλλακτική του μορφή είναι:

$$R^2_{adj} = \frac{SSE / (n-p-1)}{SST / (n-1)} = 1 - (1-R^2) \left(\frac{n-1}{n-p-1} \right)$$

Εξίσωση 7.62β

Προσαρμόζει το R^2 για το μέγεθος του κάθε μοντέλου.

β) Η F -στατιστική συνάρτηση με την οποία συγκρίνουμε το πλήρες με το περιορισμένο μοντέλο:

$$F_p = \frac{[SSE_{(p)} - SSE_{(k)}] / (k-p)}{MSE_{(k)}} \text{ όπου } MSE_{(k)} = \frac{SSE_{(k)}}{n-k-1}$$

Εξίσωση 7.63

Η F_p μετράει σχετικές αλλαγές στο σφάλμα, έπειτα από την προσθήκη ενός πλήθους από $k-p$ predictors και συγκρίνεται με την $F_{(k-p),(n-k-1)}$ κατανομή.

γ) Η διασπορά στο περιορισμένο μοντέλο:

$$MSE_{(p)} = \frac{SSE_{(p)}}{n-p-1}$$

Εξίσωση 7.64

Θέλουμε να ελαχιστοποιηθεί.

δ) Η C_p Mallow's στατιστική συνάρτηση:

$$C_p = \frac{SSE_{(p)}}{\hat{\sigma}^2} - [n - 2(p+1)]$$

Εξίσωση 7.65

Όπου $\hat{\sigma}$: η καλύτερη εκτίμηση της διασποράς των Y και συνήθως:

$$\hat{\sigma}^2 = MSE_{(k)}$$

Εξίσωση 7.66

Σημειώνουμε ότι αν θεωρήσουμε ως καλύτερη εκτίμηση της διασποράς την $MSE_{(k)}$, τότε $C_p = k+1$ για το πλήρες μοντέλο. Έτσι αν ένα περιορισμένο p - μεταβλητών μοντέλο έχει διασπορά κοντά στο $MSE_{(k)}$ τότε $C_p \approx p+1$.

Όλα τα προαναφερθέντα κριτήρια εμπεριέχουν την ίδια πληροφορία. Έτσι, ισχύουν οι σχέσεις:

$$F_p = \frac{(R_k^2 - R_p^2) / (k-p)}{(1 - R_k^2) / (n-k-1)}$$

Εξίσωση 7.67

Και:

$$C_p = (k-p)F_p + (2p-k+1)$$

Εξίσωση 7.68

Ελέγχουμε την αξιοπιστία (reliability) του γραμμικού μοντέλου (split-sample προσέγγιση):

Για να ελέγξουμε την αξιοπιστία του μοντέλου, δηλαδή το κατά πόσο το μοντέλο κάνει καλή πρόβλεψη, ακολουθούμε την εξής διαδικασία:

- α) Χωρίζουμε τα δεδομένα μας σε δύο μέρη, ένα δοκιμαστικό δείγμα (training sample) και ένα κύριο δείγμα (validation sample).
- β) Προσαρμόζουμε το «βέλτιστο» μοντέλο στο δοκιμαστικό δείγμα

$$\hat{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p$$

Εξίσωση 7.69

Συγκεκριμένα θέτουμε:

$$\{Y_i\}_{i=1}^{n_i}$$

Εξίσωση 7.70

Το διάνυσμα των n_i προσαρμοσμένων τιμών Y_i από το δοκιμαστικό δείγμα με:

$$R^2_{(1)} = R^2(Y_1 / X_1, X_2, \dots, X_p) = r^2(Y_1, \hat{Y}_1)$$

Εξίσωση 7.71

γ) Χρησιμοποιούμε τα:

$$\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$$

Εξίσωση 7.72

Για να υπολογίσουμε το διάνυσμα των προσαρμοσμένων Y_i από το κύριο δείγμα βασιζόμενο στους συντελεστές παλινδρόμησης που υπολογίστηκαν για το δοκιμαστικό μοντέλο με:

$$R^2_{(2)} = r^2(Y_2, \hat{Y}_2)$$

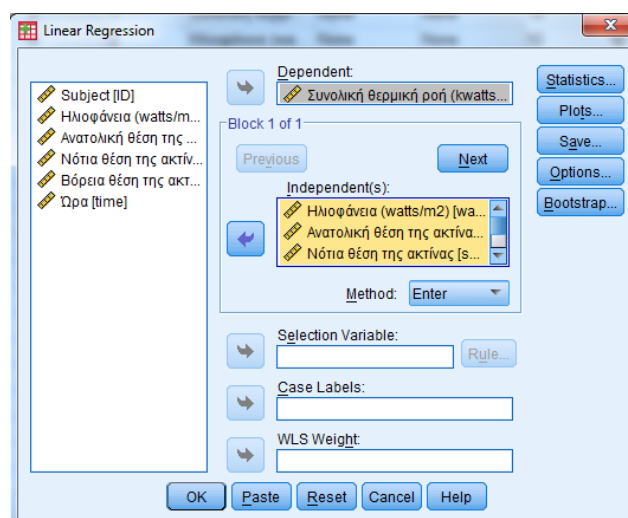
Εξίσωση 7.73

Η ποσότητα $R^2_{(1)} - R^2_{(2)}$ καλείται shrinkage on cross-validation. Τιμές της shrinkage μικρότερες από 0,10 υποδεικνύουν ένα αξιόπιστο μοντέλο και τιμές κοντά στο 0,9 χαρακτηρίζουν ένα μοντέλο ως προβληματικό.

Σημείωση

Για να πραγματοποιηθεί η split-sample προσέγγιση θα πρέπει να έχουμε ένα αρκετά μεγάλο δείγμα ώστε να είμαστε σε θέση να το χωρίσουμε. Ο διαχωρισμός των δύο ομάδων γίνεται με μεθόδους δειγματοληψίας, όπως απλή τυχαία δειγματοληψία, στρωματοποιημένη και άλλα.

Για την εκτέλεση της μεθόδου στο SPSS, η δυνατότητα προσφέρεται από την επιλογή *Analyze > Regression > Linear Regression*.



Εικόνα 7.188: Το menu Linear Regression

Στο πλαίσιο *Dependent* ορίζουμε την εξαρτημένη συνεχή μεταβλητή, και στο πλαίσιο *Independent(s)* εισάγουμε τις ανεξάρτητες προς εξέταση μεταβλητές. Στη συνέχεια στο πεδίο *Method* επιλέγουμε τη μέθοδο επιλογής του βέλτιστου μοντέλου. Οι δυνατές επιλογές είναι:

- *Enter*: Με αυτή τη μέθοδο εισάγονται στο μοντέλο όλες οι μεταβλητές ανεξάρτητα από το αν είναι στατιστικά σημαντικές σε σχέση με την εξαρτημένη μεταβλητή.
- *Stepwise*: Είναι μια παραλλαγή της διαδικασίας forward, που μας επιτρέπει να επανεξετάσουμε αν κάποια από τις μεταβλητές (εκτός από αυτήν που έχει μόλις εισαχθεί στο μοντέλο), μπορεί να αφαιρεθεί χρησιμοποιώντας τη διαδικασία backward. **Βήμα 1**, αρχικά ορίζεται ένα μέγιστο *p-value* για την αφαίρεση κάποιας μεταβλητής και ένα ελάχιστο για την είσοδο κάποιας μεταβλητής. Έτσι ώστε $p\text{-value}_{OUT} < p < p\text{-value}_{IN}$ για να είμαστε σίγουροι ότι καμιά μεταβλητή που θα αφαιρεθεί από το μοντέλο δεν θα μπορεί να ξαναμπει στο ίδιο βήμα της διαδικασίας. **Βήμα 2**, στη συνέχεια, υπολογίζονται όλα τα Τύπου III partial F-tests για κάθε υποψήφια μεταβλητή που δεν είναι ακόμα προσαρμοσμένη στο μοντέλο (adjusted) με όσες είναι ήδη μέσα στο μοντέλο. Εάν το χαμηλότερο *p-value* είναι χαμηλότερο από το $p\text{-value}_{IN}$ η μεταβλητή εισέρχεται. **Βήμα 3**, έπειτα, υπολογίζονται όλα τα partial F-tests για όλες τις μεταβλητές. Εάν το υψηλότερο *p-value* είναι υψηλότερο από το $p\text{-value}_{OUT}$ τότε η μεταβλητή απομακρύνεται. Το μοντέλο ξαναπροσαρμόζεται και το **Βήμα 3** επαναλαμβάνεται μέχρις ότου καμιά μεταβλητή να μην μπορεί να μπει στο μοντέλο. Η διαδικασία πηγαίνει στο **Βήμα 2** και συνεχίζεται μέχρις ότου καμιά μεταβλητή να μην μπορεί ούτε να μπει ούτε να αφαιρεθεί από το μοντέλο.
- *Remove*: Σ' αυτή τη μέθοδο όλες οι μεταβλητές (σε ένα μπλοκ) απομακρύνονται σε ένα μόνο βήμα.
- *Backward*: Είναι μια διαδικασία επιλογής στην οποία αρχικά εισάγονται στο μοντέλο όλες οι μεταβλητές και στη συνέχεια απομακρύνονται μία προς μία διαδοχικά. **Βήμα 1** προσαρμόζεται ένα πλήρες μοντέλο και ορίζεται ένα μέγιστο *p-value* για την απομάκρυνση κάθε παράγοντα (για παράδειγμα 10%). **Βήμα 2**, πραγματοποιούνται όλα τα partial F-tests. **Βήμα 3**, αν το υψηλότερο *p-value* είναι μεγαλύτερο από το *p-value* που ορίσαμε στο **Βήμα 1** τότε η αντίστοιχη μεταβλητή απομακρύνεται. **Βήμα 4**, αν καμιά μεταβλητή δεν απομακρυνθεί, η διαδικασία σταματάει και οι εναπομείνουσες μεταβλητές ορίζουν το «βέλτιστο» μοντέλο. Εάν όμως, κάποια μεταβλητή απομακρυνθεί, τότε το νέο μοντέλο προσαρμόζεται (θεωρείται ως πλήρες) και η διαδικασία επαναλαμβάνεται από το **Βήμα 2** μέχρις ότου καμιά μεταβλητή να μην μπορεί να απομακρυνθεί.
- *Forward*: Είναι μια διαδικασία στην οποία οι μεταβλητές εισέρχονται σταδιακά στο μοντέλο. **Βήμα 1**, προσδιορίζεται ένα μέγιστο *p-value* για την είσοδο κάθε μεταβλητής στο μοντέλο. **Βήμα 2**, εκτελούνται όλες οι μονοπαραγοντικές αναλύσεις, δηλαδή προσαρμόζονται όλες οι μεταβλητές κάθε μια μόνη της. **Βήμα 3**, υπολογίζονται όλες οι F στατιστικές για τα απλά μοντέλα παλινδρόμησης του Βήματος 2. **Βήμα 4**, αν το χαμηλότερο *p-value* του Τύπου III F-test είναι μεγαλύτερο από το *p-value* που ορίσαμε στο **Βήμα 1** τότε σταματάμε, αλλιώς ενσωματώνουμε αυτή τη μεταβλητή στο μοντέλο. **Βήμα 5**, για τις εναπομείνουσες μεταβλητές που δεν είναι ακόμα στο μοντέλο, υπολογίζουμε τα Τύπου III partial F-tests ελέγχοντάς τα (controlling) για όλες τις μεταβλητές του μοντέλου. Έπειτα, επανερχόμαστε στο **Βήμα 4** και επαναλαμβάνουμε τη διαδικασία έως ότου καμιά μεταβλητή να μην μπορεί να μπει στο μοντέλο.

Ανεξάρτητα από τη μέθοδο που θα επιλέξει ο κάθε ερευνητής για την επιλογή του «βέλτιστου» (optimum) μοντέλου παλινδρόμησης, θα πρέπει να έχει υπόψη του ότι το μοντέλο θα πρέπει να πληροί δύο βασικές αρχές.

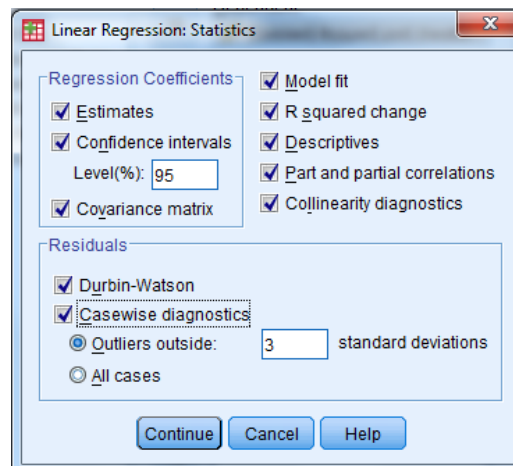
1. *Αξιοπιστία (Reliability)*: Εάν στόχος μας είναι να προβλέψουμε το *Y* τότε το μοντέλο που προβλέπει καλύτερα το *Y* λέγεται ότι είναι «ακριβές» (be reliable). Στην περίπτωση αυτή ο βασικός στόχος είναι η ακριβής (accurate) εκτίμηση της *Y* και όχι τόσο οι ιδιαιτερότητες του ίδιου του μοντέλου.

2. *Εγκυρότητα (Validity)*: Όταν δίνουμε έμφαση στη σχέση της Y με τις X_i τότε εμβαθύνουμε στην πρόβλεψη των συντελεστών της παλινδρόμησης, αναζητώντας ένα έγκυρο (valid) γραμμικό μοντέλο παλινδρόμησης.

Στο πλαίσιο *Selection Variable* μπορούμε να ορίσουμε μια μεταβλητή καθώς και τη συνθήκη Rule μέσω της οποίας θα απομονώσουμε το υποσύνολο των δεδομένων στο οποίο επιθυμούμε να εφαρμόσουμε την παλινδρόμηση.

Από το πλήκτρο *Statistics* μπορούμε να επιλέξουμε την εμφάνιση των αποτελεσμάτων της παλινδρόμησης και κυρίως διαγνωστικών μέτρων ελέγχου αυτής. Αναλυτικότερα έχουμε:

- *Estimates*: Εμφανίζει τον πίνακα με τους συντελεστές παλινδρόμησης B , τα τυπικά σφάλματα αυτών, τους τυποποιημένους συντελεστές παλινδρόμησης, την τιμή του T -test και το αποτέλεσμα του δίπλευρου ελέγχου.



Εικόνα 7.189: To menu *Linear Regression > Statistics*

- *Confidence intervals*: Εμφανίζει τα διαστήματα εμπιστοσύνης των συντελεστών παλινδρόμησης και του πίνακα συνδιακυμάνσεων, σύμφωνα με το οριζόμενο επίπεδο στο πεδίο *Level*.
- *Covariance matrix*: Εμφανίζει έναν πίνακα διακυμάνσεων-συνδιακυμάνσεων των συντελεστών παλινδρόμησης. Υπενθυμίζουμε ότι τα στοιχεία της διαγωνίου του πίνακα είναι οι διακυμάνσεις, ενώ τα στοιχεία εκτός διαγωνίου είναι οι συνδιακυμάνσεις.
- *Model fit*: Εμφανίζει κατάσταση των μεταβλητών που εισήχθησαν ή/και αφαιρέθηκαν από το μοντέλο, καθώς και τις τιμές διαγνωστικών μέτρων καταλληλότητας του μοντέλου, όπως τον συντελεστή προσδιορισμού R^2 , τον προσαρμοσμένο συντελεστή προσδιορισμού R^2 , το τυπικό σφάλμα της εκτίμησης και τον πίνακα ανάλυσης διασποράς ANOVA.
- *R squared change*: Εμφανίζει τις ενδιάμεσες μεταβολές του συντελεστή προσδιορισμού R^2 οι οποίες λαμβάνουν χώρα κατά την προσθήκη ή αφαίρεση των ανεξάρτητων μεταβλητών στο μοντέλο. Εάν η είσοδος μιας μεταβλητής οδηγεί σε σημαντική αύξηση της τιμής του συντελεστή προσδιορισμού, τότε αυτό συνεπάγεται ότι η μεταβλητή είναι σημαντική για το μοντέλο.
- *Descriptives*: Εμφανίζει πίνακα συνήθων περιγραφικών μέτρων των μεταβλητών του μοντέλου, όπως το πλήθος έγκυρων παρατηρήσεων, τη μέση τιμή και την τυπική απόκλιση. Επιπρόσθετα, εμφανίζεται και ο πίνακας των συσχετίσεων μεταξύ των μεταβλητών.
- *Part and Partial Correlations*: Εμφανίζει τις μερικές συσχετίσεις, δηλ. τη συσχέτιση η οποία απομένει μεταξύ δύο μεταβλητών μετά από την αφαίρεση της συσχέτισης που οφείλεται στην αμοιβαία ένωσή τους με τις υπόλοιπες μεταβλητές. Η συσχέτιση μεταξύ της εξαρτημένης

μεταβλητής και μιας ανεξάρτητης μεταβλητής, όταν οι γραμμικές επιδράσεις των άλλων ανεξάρτητων μεταβλητών στο μοντέλο έχουν αφαιρεθεί από αυτές.

- *Collinearity diagnostics*: Για την πραγματοποίηση του ελέγχου πολυσυγγραμμικότητας στο μοντέλο. Συγγραμμικότητα (Collinearity) ή πολυσυγγραμμικότητα (multicollinearity) εμφανίζεται όταν μια ανεξάρτητη μεταβλητή είναι μια γραμμική συνάρτηση μιας άλλης ανεξάρτητης μεταβλητής. Σε περίπτωση που υπάρχουν υψηλές συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών, αυτό αποτελεί ισχυρή ένδειξη πολυσυγγραμμικότητας. Η πολυσυγγραμμικότητα δημιουργεί προβλήματα, όχι στην προσαρμογή του υποδείγματος, αλλά στο να διαπιστωθούν πραγματικά οι στατιστικά σημαντικοί παράγοντες στην επίδραση της εξαρτημένης μεταβλητής, και αυτό γιατί η παρουσία της πολυσυγγραμμικότητας οδηγεί σε αυξημένα τυπικά σφάλματα. Στην ακραία περίπτωση που μια ανεξάρτητη μεταβλητή είναι γραμμικός συνδυασμός μερικών ή όλων των υπόλοιπων ανεξάρτητων μεταβλητών, τότε προκύπτει ότι η ορίζουσα του πίνακα $X'X$ ισούται με μηδέν, το οποίο καθιστά τον πίνακα $X'X$ η αντιστρέψιμο, και κατ' επέκταση δεν μπορεί να γίνει ο υπολογισμός των συντελεστών παλινδρόμησης:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

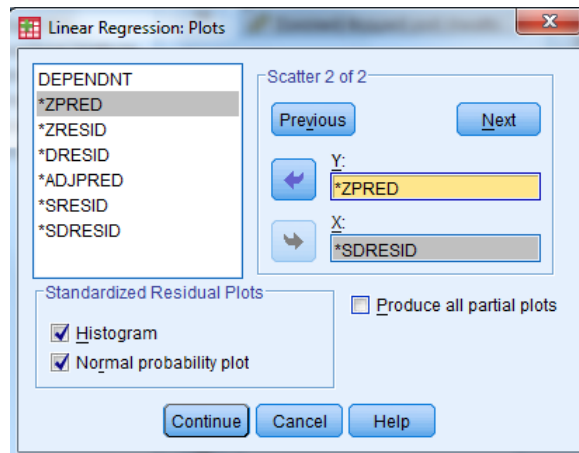
Εξίσωση 7.74: Συντελεστές παλινδρόμησης

Μέτρα για τον έλεγχο της πολυσυγγραμμικότητας είναι ο παράγοντας μεγέθυνσης διασποράς (*Variance inflation factor-VIF*), ο οποίος παρέχεται σε κάθε μεταβλητή και μας δείχνει πόσο αυξάνεται η διασπορά ενός εκτιμημένου συντελεστή παλινδρόμησης.

Όταν υπάρχουν συσχετίσεις μεταξύ των επεξηγηματικών μεταβλητών. Τιμές του $VIF > 5$ θεωρούνται ως ένδειξη πολυσυγγραμμικότητας. Για να μειωθεί το φαινόμενο της πολυσυγγραμμικότητας αφαιρούμε από το μοντέλο τους μη σημαντικούς παράγοντες.

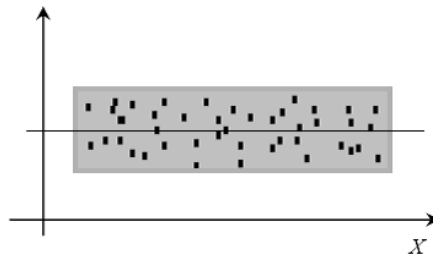
- *Durbin-Watson*: Για την πραγματοποίηση του ελέγχου Durbin-Watson προκειμένου να διαπιστωθεί η σειριακή συσχέτιση των σφαλμάτων (αυτοσυσχέτιση). Αυτοσυσχέτιση έχουμε όταν σε μια μεταβλητή οι μετρήσεις λαμβάνονται σειριακά κατά τον χρόνο και υπάρχει συσχέτιση μεταξύ των Y_t και Y_{t-1} , η μέτρηση στον χρόνο t σε σχέση με την προηγούμενη παρατήρηση. Η ύπαρξη αυτοσυσχέτισης ή serial correlation μεταξύ των παρατηρήσεων εξετάζεται με το Durbin-Watson test.
 - Αν $d < d_L$ ή $d > 4 - d_L$ υπάρχει αυτοσυσχέτιση.
 - Αν $d < d_u$ ή $d > 4 - d_u$ δεν υπάρχει αυτοσυσχέτιση.
 - Αλλιώς δεν μπορούμε να αποφανθούμε.
- *Casewise Diagnostics*: Για την πραγματοποίηση διαγνωστικών ελέγχων με βάση το κριτήριο επιλογής, όπως των ακραίων τιμών (τιμών μεγαλύτερων από προκαθορισμένο μέγεθος που προκύπτει από την προσθαφαίρεση των τυπικών αποκλίσεων). Στόχος είναι να εξεταστούν πιθανές ακραίες τιμές (outliers) και παρατηρήσεις υψηλής επίδρασης (influential-points). Μια ακραία τιμή (outliers) είναι μια ασυνήθιστη τιμή που δεν συμφωνεί με το pattern των υπόλοιπων δεδομένων. Μια παρατήρηση υψηλής επίδρασης μπορεί να είναι μια ακραία τιμή, αλλά, επίσης, μπορεί να είναι μια τιμή των δεδομένων που να έχει μεγάλη συνεισφορά στο σχηματισμό της ευθείας παλινδρόμησης. Το μέτρο που χρησιμοποιείται για την εύρεση της επιρροής μιας παρατήρησης στο μοντέλο είναι η Cook's distance. Περισσότερα για την Cook's distance βλ. στην ανάλυση των επιλογών του πλήκτρου *Save*.

Από το πλήκτρο *Plots* μπορούμε πολύ αποτελεσματικά να εξετάσουμε αν ισχύουν οι υποθέσεις του μοντέλου, καθώς και να εξεταστεί η προσαρμογή του. Έτσι χρησιμοποιούμε γραφήματα των σφαλμάτων έναντι των προσαρμοσμένων τιμών (\hat{Y}_i) και μελετάμε τα patterns που εμφανίζουν. Ο λόγος που χρησιμοποιούμε γραφήματα των σφαλμάτων έναντι των προσαρμοσμένων τιμών (\hat{Y}_i) και όχι των παρατηρούμενων (Y_i), είναι γιατί τα σφάλματα συσχετίζονται με τις παρατηρούμενες τιμές, ενώ με τις προσαρμοσμένες όχι.



Εικόνα 7.190: Το menu Linear Regression > Plots

Ο χρήστης μπορεί να δηλώσει τη δημιουργία γραφημάτων διαφόρων συνδυασμών, αλλάζοντας τα στοιχεία των τεταγμένων Y και X μέσω των πλήκτρων *Next* και *Previous*.

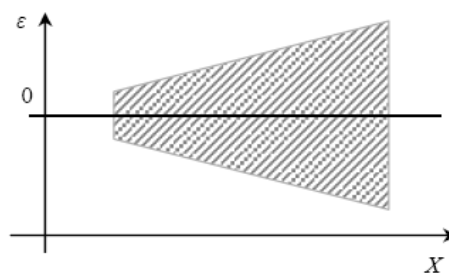


Κατάλληλο το γραμμικό μοντέλο

Εικόνα 7.191: Μοντέλα υπολοίπων (πρότυπη γραφική παράσταση)

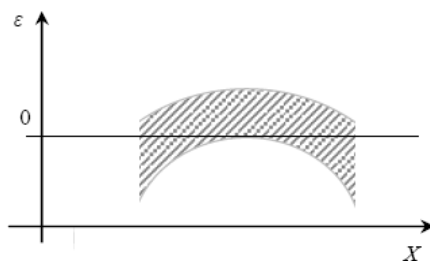
Τα προβλήματα που μπορούν να ανιχνευθούν από τα γραφήματα των σφαλμάτων είναι:

1. Η διασπορά να μην είναι σταθερή, δηλαδή να μην υπάρχει ομοσκεδαστικότητα. Στην περίπτωση αυτή το γράφημα συνήθως θα «ανοίγει σαν χωνί» (will funneling) καθώς τα Y_i θα αυξάνονται.



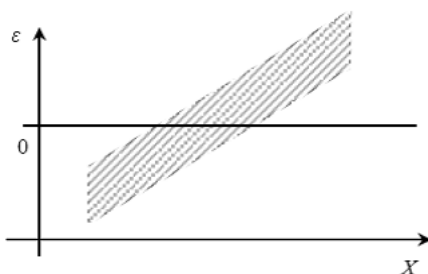
Εικόνα 7.192: Μοντέλα σφαλμάτων με μη σταθερή διακύμανση

2. Οι παρατηρήσεις να αποκλίνουν από το γραμμικό μοντέλο. Τότε θα υπάρχει ένα συστηματικό pattern των σφαλμάτων που θα υποδεικνύει τη χρήση επιπλέον όρων (τετραγωνικών ή πολυωνυμικών, curvilinear) στο μοντέλο παλινδρόμησης.



Εικόνα 7.193: Μοντέλα σφαλμάτων με μη γραμμικό μοντέλο

3. Η ύπαρξη τάσης με την πάροδο του χρόνου. Τα σφάλματα τότε θα είναι «roughly» ανεξάρτητα. Όταν υπάρχει χρονική τάση ή άλλου είδους εξάρτηση στα δεδομένα, το μοντέλο δεν είναι ανεξάρτητο.



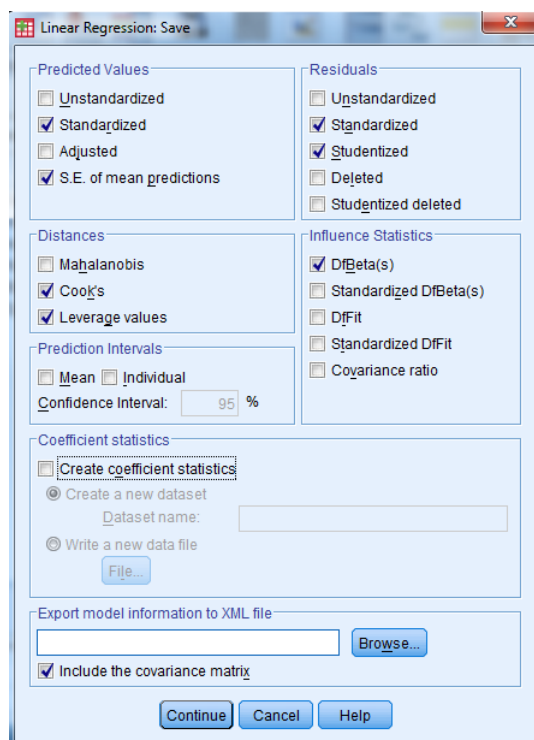
Εικόνα 7.194: Μοντέλα υπολοίπων με εξάρτηση μεταξύ των ϵ_i

Η υπόθεση της κανονικότητας εξετάζεται με έλεγχο της κανονικότητας των σφαλμάτων μέσω των ελέγχων Kolmogorov-Smirnov και Shapiro-Wilks, και γραφικά από τα διαγράμματα Q-Q και P-P plots (βλέπε ομώνυμες ενότητες). Η εφαρμογή έχει τη δυνατότητα να δημιουργήσει ιστόγραμμα, Q-Q και P-P plots των τυποποιημένων σφαλμάτων, από τις επιλογές *Histogram* και *Normal Probability Plots*, αντίστοιχα.

Από το πλήκτρο *Save* μπορούμε να αποθηκεύσουμε τις εκτιμώμενες παρατηρήσεις, τα σφάλματα, και άλλα χρήσιμα στατιστικά για τους διαγνωστικούς ελέγχους του μοντέλου. Μέσω αυτών αντλούμε τα στοιχεία εκείνα που θα μας οδηγήσουν στην επιλογή του «βέλτιστου» μοντέλου.

Σχετικά με τις εκτιμώμενες παρατηρήσεις Predicted Values έχουμε:

- *Unstandardized*: Η μη τυποποιημένη εκτιμώμενη παρατήρηση, δηλ. η τιμή που προβλέπει το μοντέλο για τις παρατηρήσεις των ανεξάρτητων μεταβλητών.
- *Standardized*: Οι τυποποιημένες εκτιμώμενες παρατηρήσεις, οι οποίες προκύπτουν μέσω μετασχηματισμού, όπου η μέση προβλεπόμενη τιμή αφαιρείται από την προβλεπόμενη τιμή, και η διαφορά διαιρείται με την τυπική απόκλιση των προβλεπόμενων τιμών. Οι τυποποιημένες προβλεπόμενες τιμές έχουν μέση τιμή μηδέν και τυπική απόκλιση ίση με τη μονάδα.
- *Adjusted*: Η προβλεπόμενη τιμή για μια παρατήρηση, κατά την οποία η παρατήρηση αυτή εξαιρείται από τον υπολογισμό των συντελεστών παλινδρόμησης.
- *S.E. of mean predictions*: Τα τυπικά σφάλματα των προβλεπόμενων τιμών (τα οποία αποτελούν και την τυπική απόκλιση της μέσης τιμής).



Εικόνα 7.195: Το menu Linear Regression > Save

Σχετικά με τα σφάλματα (Residuals) έχουμε:

- *Unstandardized*: Η διαφορά μεταξύ μιας παρατήρησης και της εκτιμώμενης τιμής που προβλέπεται από το μοντέλο.

$$\varepsilon_i = Y_i - \hat{Y}_i$$

Εξίσωση 7.75

- *Standardized*: Τα τυποποιημένα σφάλματα είναι τα σφάλματα διαιρεμένα με την τυπική του απόκλιση αφού η μέση τιμή αυτών ισούται με μηδέν.

$$z_i = \frac{\varepsilon_i}{s} \text{ με } E(\varepsilon_i) = 0 \text{ και } Var(\varepsilon_i) = 1$$

Εξίσωση 7.76

- *Studentized*: Το σφάλμα διαιρείται από μια εκτίμηση της τυπικής του απόκλισης, η οποία ποικίλλει από παρατήρηση σε παρατήρηση, αναλόγως της απόστασης μεταξύ της κάθε παρατήρησης της εξαρτημένης από τις μέσες τιμές των ανεξάρτητων μεταβλητών.

$$r_i = \frac{\varepsilon_i}{s\sqrt{1-h_i}}$$

Εξίσωση 7.77

Όπου h_i είναι το i -οστό διαγώνιο στοιχείο (leverage) του πίνακα $H = X(X^T X)^{-1} X^T$ (hat matrix) $r_i \sim t_{n-p-1}$ (για p -μεταβλητές).

- *Deleted*: Είναι μια τροποποίηση των τυποποιημένων σφαλμάτων σύμφωνα με την οποία χρησιμοποιείται μια εναλλακτική εκτιμήτρια της σ^2 , η οποία προκύπτει με την προσαρμογή του μοντέλου με όλες τις παρατηρήσεις εξαιρώντας την i -οστή παρατήρηση.

$$r_i' = \frac{\varepsilon_i}{S_{(i)}\sqrt{1-h_i}}$$

Εξίσωση 7.78

Όπου $S_{(i)}^2$ είναι η εκτιμήτρια της διασποράς σ^2 που προκύπτει από την προσαρμογή του μοντέλου στις $n-1$ παρατηρήσεις, έχοντας δηλ. εξαιρέσει την i -οστή παρατήρηση.

- *Studentized deleted*: Τα deleted σφάλματα διαιρεμένα με το τυπικό τους σφάλμα. Η διαφορά μεταξύ των Studentized deleted και των Studentized σφαλμάτων υποδεικνύει πόση διαφορά εξαλείφει η παρατήρηση στην πρόβλεψή της.

Οι επιλογές της ενότητας Distances έχουν ως σκοπό τον εντοπισμό περιπτώσεων με ασυνήθιστους συνδυασμούς παρατηρήσεων στις ανεξάρτητες μεταβλητές, όπου ενδέχεται να έχουν μεγάλο αντίκτυπο στο μοντέλο παλινδρόμησης. Αναλυτικά:

- *Mahalanobis*: Ένα μέτρο για την εξέταση του κατά πόσο οι τιμές μιας υπόθεσης σχετικά με τις ανεξάρτητες μεταβλητές διαφέρουν από το μέσο όρο του συνόλου των τιμών. Μεγάλες τιμές απόστασης Mahalanobis προσδιορίζουν ότι το αντικείμενο έχει ακραίες τιμές σε μία ή περισσότερες από τις ανεξάρτητες μεταβλητές.
- *Cook's*: Όπως προαναφέραμε είναι το μέτρο που χρησιμοποιείται για την εύρεση της επιρροής μιας παρατήρησης στο μοντέλο. Αυτή η στατιστική συνάρτηση μετράει την αλλαγή στους συντελεστές παλινδρόμησης β_i όταν η i -οστή παρατήρηση αφαιρείται από τα δεδομένα. Η στατιστική είναι ανάλογη ενός σταθμισμένου μέσου των τετραγωνικών διαφορών των β_j και $\beta_{j(i)}$ και εκφράζεται με όρους των studentized σφαλμάτων (r_i) και των leverages (h_i) από τον τύπο:

$$d_i = \left(\frac{1}{p+1} \right) r_i^2 \left(\frac{h_i}{1-h_i} \right) = \frac{\varepsilon_i^2 h_i}{(p+1)s^2(1-h_i)}$$

Εξίσωση 7.79

Με άλλα λόγια, αν μια παρατήρηση είναι σημείο επιρροής τότε θα υπάρχει μια σημαντική διαφορά μεταξύ των τιμών των εκτιμώμενων συντελεστών παλινδρόμησης, αν αυτή η παρατήρηση απουσιάζει. Πρακτικά αν $d_i > 1$ προτείνεται περαιτέρω μελέτη της συγκεκριμένης παρατήρησης.

- *Leverage values*: Είναι οι τιμές της διαγώνιου (στοιχεία h_{ii}) του πίνακα προβολής $H = X(X'X)^{-1}X'$ που ονομάζεται μόχλευση (leverage). Αν ένα σημείο έχει πολύ μεγάλη μόχλευση, τότε το σημείο αυτό επιδρά στην προσαρμογή της παλινδρόμησης ώστε η ευθεία παλινδρόμησης να περνά πολύ κοντά από το σημείο αυτό. Σημεία με μόχλευση $h_{ij} > 2p/n$ θεωρούνται υποψήφια σημεία επιρροής και χρίζουν περαιτέρω εξέτασης.

Στα διαστήματα εμπιστοσύνης Prediction Intervals παρέχονται τα άνω και κάτω όρια για τη μέση ή μεμονωμένη τιμή πρόβλεψης. Το ποσοστό του διαστήματος εμπιστοσύνης καθορίζεται στο πεδίο *Confidence Interval*.

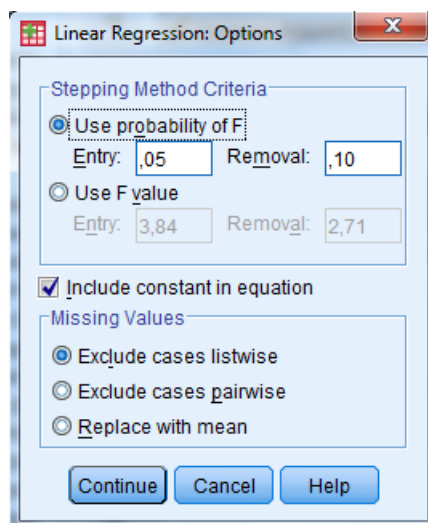
- *Mean*: Χρησιμοποιείται για την εμφάνιση διαστήματος εμπιστοσύνης για τη μέση προβλεπόμενη τιμή.
- *Individual*: Χρησιμοποιείται για την εμφάνιση διαστήματος εμπιστοσύνης για την εξαρτημένη μεταβλητή σε μια και μοναδική περίπτωση.

Influence Statistics εκφράζουν την αλλαγή των συντελεστών παλινδρόμησης $DfBeta(s)$ και των προβλεπόμενων τιμών ($DfFit$) που προκύπτουν από τον αποκλεισμό μιας συγκεκριμένης περίπτωσης. Αναλυτικά:

- $DfBeta(s)$: Είναι η διαφορά στην τιμή του συντελεστή παλινδρόμησης B που προκύπτει από τον αποκλεισμό μιας συγκεκριμένης περίπτωσης από την ανάλυση παλινδρόμησης. Οι τιμές αυτές υπολογίζονται για όλους τους συντελεστές παλινδρόμησης του μοντέλου, συμπεριλαμβανομένου και του σταθερού όρου.
- $Standardized DfBeta(s)$: Είναι η τυποποιημένη διαφορά στο συντελεστή παλινδρόμησης. Υπό εξέταση περιπτώσεις είναι αυτές με απόλυτη τιμή μεγαλύτερη του $2 / \sqrt{N}$ όπου N είναι ο αριθμός των περιπτώσεων.
- $DfFit$: Είναι η μεταβολή στην προβλεπόμενη τιμή που προκύπτει από τον αποκλεισμό μιας συγκεκριμένης περίπτωσης.
- $Standardized DfFit$: Είναι οι τυποποιημένες τιμές των παραπάνω διαφορών. Υπό εξέταση περιπτώσεις είναι αυτές με απόλυτη τιμή μεγαλύτερη του $2 \sqrt{\rho/N}$ όπου ρ είναι ο αριθμός των παραμέτρων στο μοντέλο και N είναι ο αριθμός των περιπτώσεων.
- $Covariance Ratio$: Είναι η αναλογία της ορίζουσας του πίνακα συνδιασποράς, όταν μια συγκεκριμένη περίπτωση εξαιρείται από τον υπολογισμό των συντελεστών παλινδρόμησης, με την ορίζουσα του πίνακα συνδιασποράς και όλες τις περιπτώσεις να περιλαμβάνονται στην ανάλυση. Αν η αναλογία είναι κοντά στο 1, συνεπάγεται ότι η περίπτωση αυτή δε μεταβάλλει σημαντικά τη μήτρα συνδιακύμανσης.

Από το πλήκτρο *Options* παρέχονται οι επιλογές:

- *Stepping Method Criteria*: Στο σημείο αυτό ορίζεται ο τρόπος επιλογής για το αν μια μεταβλητή θα εισαχθεί ή θα αφαιρεθεί από το μοντέλο, όταν εκτελούνται οι αυτόματες διαδικασίες επιλογής μοντέλου (forward, backward ή stepwise). Ο χρήστης μπορεί είτε να ορίσει επίπεδα στατιστικής σημαντικότητας Use Probability of F είτε να δηλώσει απευθείας την τιμή του κριτηρίου F-test Use F Value. Έτσι, μια μεταβλητή θα εισαχθεί στο μοντέλο αν το επίπεδο σημαντικότητάς της είναι μικρότερο από την τιμή εισαγωγής (Entry) και θα αφαιρεθεί αν είναι μεγαλύτερο από την τιμή αποκλεισμού (Removal). Αντίστοιχα, θα εισαχθεί αν η τιμή του F-test είναι μεγαλύτερη από την τιμή του F-test κριτηρίου εισαγωγής, και θα αφαιρεθεί αν είναι μικρότερη από την τιμή αποκλεισμού.



Εικόνα 7.196: Το menu Linear Regression > Options

- *Include constant in equation*: Η παράμετρος είναι προεπιλεγμένη, έτσι ώστε στο μοντέλο παλινδρόμησης να συμπεριλαμβάνεται ο σταθερός όρος. Επιλέγοντας την παράμετρο αυτή θα

πρέπει να έχουμε υπόψη ότι τα μοντέλα που θα προκύψουν δεν είναι συγκρίσιμα με ένα μοντέλο στο οποίο περιέχεται η σταθερά, καθώς για παράδειγμα ο συντελεστής προσδιορισμού R_2 δεν μπορεί να ερμηνευτεί με τον συνηθισμένο τρόπο.

Τέλος, σε αυτή την ενότητα γίνεται και η διαχείριση των ελλειπουσών τιμών.

Εφαρμογή παραδείγματος στο SPSS

Χρησιμοποιήθηκε το αρχείο δεδομένων solar.sav (βλ. σύνοψη) όπου περιέχονται δεδομένα ηλιακής ενέργειας $N = 29$ συλλεκτών. Στόχος είναι να μελετηθούν οι παράγοντες που επηρεάζουν το ύψος της συνολικής θερμικής ροής ανά τετραγωνικό μέτρο ($Y = \text{kwhatts/m}^2$). Οι παράγοντες που μετρήθηκαν είναι:

- X_1 - «Ηλιοφάνεια (watts/m²)»,
- X_2 - «Ανατολική θέση ακτίνας (inches)»,
- X_3 - «Νότια θέση ακτίνας (inches)»,
- X_4 - «Βόρεια θέση ακτίνας (inches)»,
- X_5 - «Η ώρα της ημέρας (σε δεκαδική κλίμακα)».

Η εισαγωγή των μεταβλητών εμφανίζεται στην Εικόνα 7.2. Στη συνέχεια γίνεται επιλογή των κατάλληλων μέτρων όπως αυτά περιγράφηκαν παραπάνω. Τέλος, η διαδικασία ολοκληρώνεται με το πλήκτρο OK όπου και λαμβάνουμε:

Descriptive Statistics			
	Mean	Std. Deviation	N
Συνολική θερμική ροή (kwhatts) ανά επιφάνεια	249,638	22,8988	29
Ηλιοφάνεια (watts/m ²)	754,4741	79,79682	29
Ανατολική θέση της ακτίνας	35,0966	1,41699	29
Νότια θέση της ακτίνας	35,5293	1,85297	29
Βόρεια θέση της ακτίνας	16,7021	,90818	29
Ώρα	13,2245	1,94602	29

Εικόνα 7.197: Περιγραφικά μέτρα των μεταβλητών του μοντέλου

Correlations							
		Συνολική θερμική ροή (kwhatts) ανά επιφάνεια	Ηλιοφάνεια (watts/m ²)	Ανατολική θέση της ακτίνας	Νότια θέση της ακτίνας	Βόρεια θέση της ακτίνας	Ώρα
Pearson Correlation	Συνολική θερμική ροή (kwhatts) ανά επιφάνεια	1,000	,628	,102	,112	-,849	-,352
	Ηλιοφάνεια (watts/m ²)	,628	1,000	-,204	-,107	-,634	-,583
	Ανατολική θέση της ακτίνας	,102	-,204	1,000	-,329	-,117	-,067
	Νότια θέση της ακτίνας	,112	-,107	-,329	1,000	,287	,696
	Βόρεια θέση της ακτίνας	-,849	-,634	-,117	,287	1,000	,685
	Ώρα	-,352	-,583	-,067	,696	,685	1,000
	Sig. (1-tailed)	Συνολική θερμική ροή (kwhatts) ανά επιφάνεια		,000	,299	,281	,000
Ηλιοφάνεια (watts/m ²)		,000		,145	,291	,000	,000
Ανατολική θέση της ακτίνας		,299	,145		,041	,272	,366
Νότια θέση της ακτίνας		,281	,291	,041		,065	,000
Βόρεια θέση της ακτίνας		,000	,000	,272	,065		,000
Ώρα		,031	,000	,366	,000	,000	
N		Συνολική θερμική ροή (kwhatts) ανά επιφάνεια	29	29	29	29	29
	Ηλιοφάνεια (watts/m ²)	29	29	29	29	29	29
	Ανατολική θέση της ακτίνας	29	29	29	29	29	29
	Νότια θέση της ακτίνας	29	29	29	29	29	29
	Βόρεια θέση της ακτίνας	29	29	29	29	29	29
	Ώρα	29	29	29	29	29	29

Εικόνα 7.198: Πίνακας Συσχετίσεων μεταξύ των παραγόντων

Κάτω από τον συντελεστή συσχέτισης εμφανίζεται το p-value του ελέγχου σημαντικότητας της τιμής του συντελεστή, δηλ. η υπόθεση που εξετάζεται είναι αν οι δύο μεταβλητές είναι ασυσχέτιστες, ή όχι.

$$H_0 : p = 0$$

$$H_1 : p \neq 0$$

Εξίσωση 7.80

Από τον πίνακα των συσχετίσεων διαπιστώνουμε στατιστικά σημαντική συσχέτιση σε επίπεδο 1% μεταξύ των παραγόντων:

- $X_1 - X_4$: «Ηλιοφάνεια (watts/m²)» και «Βόρεια θέση ακτίνας (inches)»,
- $X_1 - X_5$: «Ηλιοφάνεια (watts/m²)» και «Ωρα της ημέρας (σε δεκαδική κλίμακα)»,
- $X_3 - X_5$: «Νότια θέση ακτίνας (inches)» και «Ωρα της ημέρας (σε δεκαδική κλίμακα)»,
- $X_4 - X_5$: «Βόρεια θέση ακτίνας (inches)» και «Ωρα της ημέρας (σε δεκαδική κλίμακα)».

Ειδικότερα ο παράγοντας «Ωρα της ημέρας» πιθανόν να πρέπει να αφαιρεθεί από το μοντέλο, καθώς όχι μόνο δεν είναι στατιστικά σημαντικός στο μοντέλο (T-test p -value = 0,199), αλλά επίσης έχει στατιστικά σημαντικές συσχετίσεις με 3 από τους υπόλοιπους 4 παράγοντες.

Τα αποτελέσματα μπορούν να θεωρηθούν αναμενόμενα δεδομένου ότι ανάλογα με την πορεία του ήλιου εντός της ημέρας, άρα και μεταβολή της ώρας εντός της ημέρας, αλλάζει η θέση της ακτίνας του ήλιου και η έκταση/ένταση της ηλιοφάνειας.

Coefficients ^a										
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics		
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF	
1	(Constant)	324,675	96,105		3,377	,003	125,767	523,383		
	Ηλιοφάνεια (watts/m ²)	,067	,029	,235	2,321	,029	,007	,127	,432	2,316
	Ανατολική θέση της ακτίνας	2,562	1,248	,159	2,053	,052	-,019	5,144	,739	1,353
	Νότια θέση της ακτίνας	3,820	1,460	,309	2,618	,015	,801	6,840	,316	3,164
	Βόρεια θέση της ακτίνας	-22,931	2,706	-,909	-8,475	,000	-28,529	-17,334	,383	2,612
	Ωρα	2,390	1,808	,203	1,322	,199	-1,349	6,129	,187	5,352

a. Dependent Variable: Συνολική θερμική ροή (kwatts) ανά επιφάνεια

Εικόνα 7.199: Πίνακας συντελεστών του μοντέλου

Regression Equation

Το μοντέλο που προσαρμόσαμε είναι το εξής:

Συνολική Θερμική Ροή (kwatts/m²) = 324,6 + 0,067 Ηλιοφάνεια(watts/m²) + 2,56 Ανατολική (inches) + 3,82 Νότια (inches) - 22,93 Βόρεια (inches) + 2,39 Ωρα (σε δεκαδική κλίμακα).

Για Δυτική θέση ακτίνας, χωρίς ηλιοφάνεια στις 00:00 (τα μεσάνυχτα), η συνολική θερμική ροή ανά τετραγωνικό μέτρο εκτιμάται κατά μέσο όρο κατά 324.6 kwatts/m².

Σε περίπτωση που η ακτίνα τοποθετηθεί Ανατολικά ο μέσος όρος της συνολικής θερμικής ροής αυξάνεται κατά 2,56 kwatts/m² σε σχέση με το αν η ακτίνα παραμείνει σε Δυτική θέση. Εάν η ακτίνα τοποθετηθεί Νότια ή Βόρεια, η μεταβολή που εκτιμάται να προκύψει κατά μέσο όρο ανέρχεται σε αύξηση κατά 3,82 kwatts/m² και μείωση κατά 22,93 kwatts/m², αντίστοιχα.

Επίσης για αύξηση της ηλιοφάνειας κατά 10 watts η συνολική θερμική ροή αναμένεται να αυξηθεί κατά 6,73 kwatts/m². Τέλος, ομοίως η μεταβολή της ώρας της ημέρας κατά 1 ώρα αναμένεται να οδηγήσει κατά μέσο όρο σε αύξηση της συνολικής θερμικής ροής.

Από τους ελέγχους T-test της ισότητας των συντελεστών του υποδείγματος με το μηδέν, προκύπτει ότι ο παράγοντας «Ωρα της ημέρας» δεν είναι στατιστικά σημαντικός στην ερμηνεία της συνολικής θερμικής ροής (p -value = 0,199 > 0,05) σε επίπεδο σημαντικότητας 5%, και οριακά ο παράγοντας «Ανατολική θέση της ακτίνας» (p -value = 0,052 > 0,05) σε επίπεδο σημαντικότητας 5%.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13193,204	5	2638,641	40,766	,000 ^a
	Residual	1488,724	23	64,727		
	Total	14681,928	28			

a. Predictors: (Constant), Ωρα, Ανατολική θέση της ακτίνας, Ηλιοφάνεια (watts/m2), Βόρεια θέση της ακτίνας, Νότια θέση της ακτίνας

b. Dependent Variable: Συνολική θερμική ροή (kwatts) ανά επιφάνεια

Εικόνα 7.200: ANOVA Table

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,948 ^a	,899	,877	8,0453	,899	40,766	5	23	,000	1,597

a. Predictors: (Constant), Ωρα, Ανατολική θέση της ακτίνας, Ηλιοφάνεια (watts/m2), Βόρεια θέση της ακτίνας, Νότια θέση της ακτίνας

b. Dependent Variable: Συνολική θερμική ροή (kwatts) ανά επιφάνεια

Εικόνα 7.201: Περιγραφικά μέτρα των μεταβλητών του μοντέλου

Το συγκεκριμένο υπόδειγμα έχει καλή προσαρμογή (F -test value = 40.77, p -value < 0.000) και εξηγεί το 78,84% ($R^2_{adj} = 78,84\%$) της μεταβλητότητας της συνολικής θερμικής ροής.

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				(Constant)	Ηλιοφάνεια (watts/m2)	Ανατολική θέση της ακτίνας	Νότια θέση της ακτίνας	Βόρεια θέση της ακτίνας	Ωρα
1	1	5,967	1,000	,00	,00	,00	,00	,00	,00
	2	,025	15,312	,00	,08	,00	,00	,00	,08
	3	,005	34,890	,00	,26	,08	,01	,03	,12
	4	,002	61,190	,00	,00	,27	,40	,40	,10
	5	,001	78,055	,01	,56	,03	,57	,04	,43
	6	,000	184,597	,99	,10	,62	,42	,54	,27

a. Dependent Variable: Συνολική θερμική ροή (kwatts) ανά επιφάνεια

Εικόνα 7.202: Πίνακας διαγνωστικών πολυσυγραμμικότητας

Τα μέτρα ελέγχου της πολυσυγραμμικότητας Ανοχή (tolerance) και VIF έχουν υπολογιστεί και παρουσιάζονται στον πίνακα των συντελεστών παλινδρόμησης. Τιμές του $VIF > 5$ θεωρούνται ως ένδειξη πολυσυγραμμικότητας. Από τα αποτελέσματα επιβεβαιώνουμε την προηγούμενη διαπίστωση ότι ο παράγοντας «Ωρα της ημέρας» δύναται να εξαιρεθεί από το μοντέλο της παλινδρόμησης.

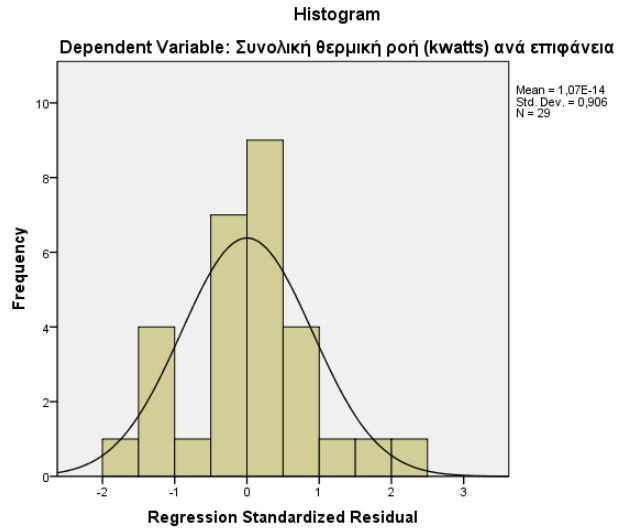
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	193,496	278,075	249,638	21,7068	29
Residual	-13,6756	17,4002	,0000	7,2917	29
Std. Predicted Value	-2,586	1,310	,000	1,000	29
Std. Residual	-1,700	2,163	,000	,906	29

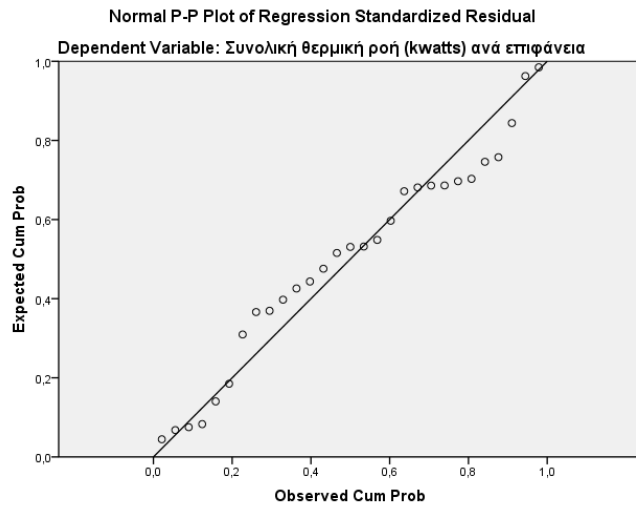
a. Dependent Variable: Συνολική θερμική ροή (kwatts) ανά επιφάνεια

Εικόνα 7.203: Πίνακας περιγραφικών μέτρων των προβλεπόμενων τιμών και των διαφορών ειδών σφαλμάτων

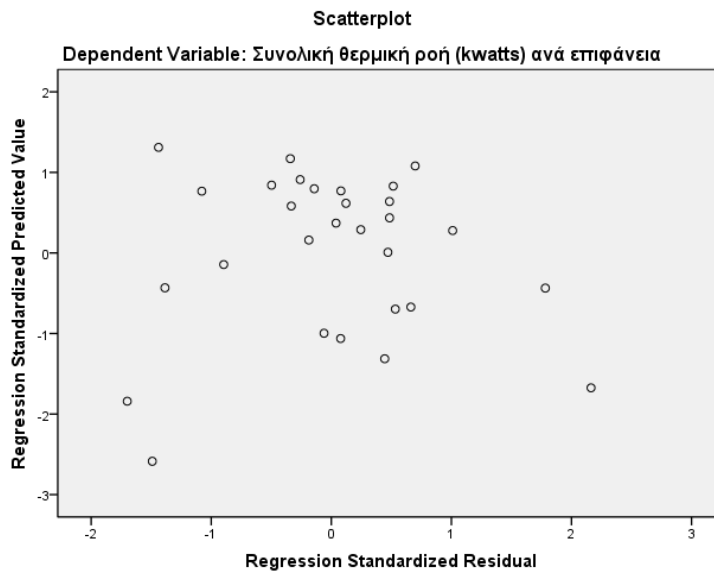
Έλεγχος Υπολοίπων



Εικόνα 7.204: Ιστόγραμμα για τον έλεγχο κανονικότητας των τυποποιημένων υπολοίπων

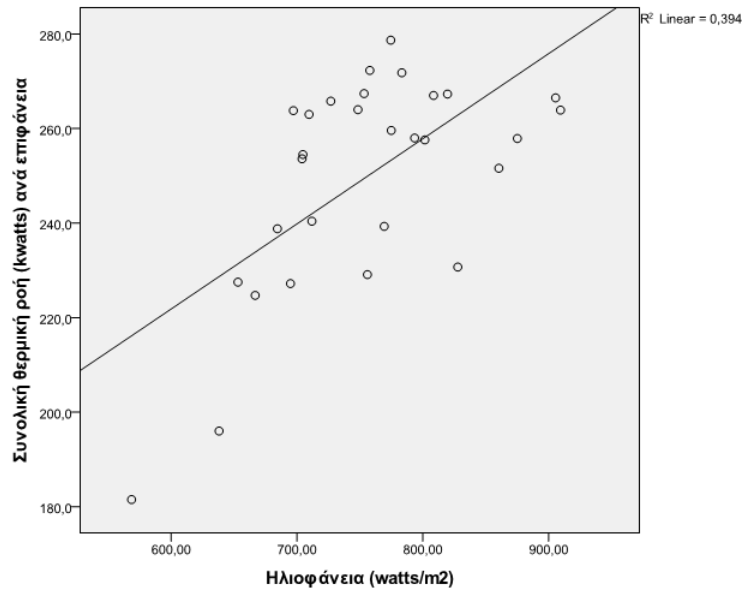


Εικόνα 7.205: P-P plot για τον έλεγχο κανονικότητας των τυποποιημένων υπολοίπων



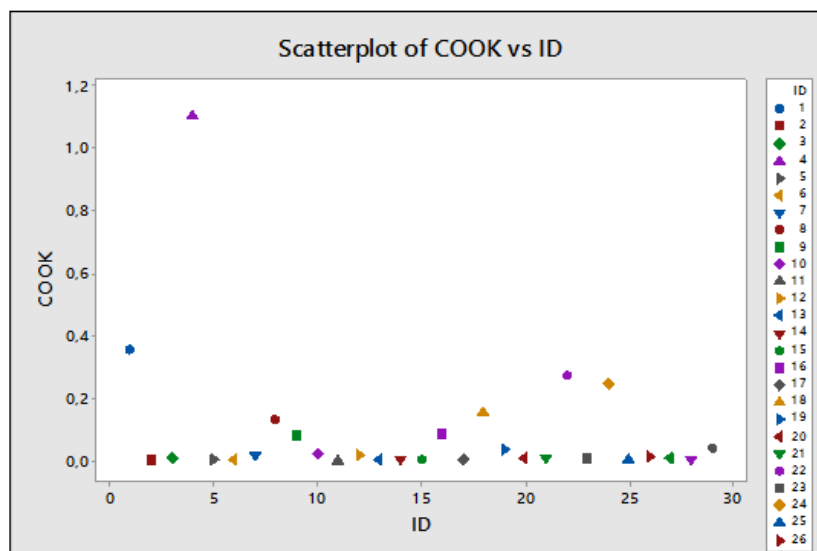
Εικόνα 7.206: Έλεγχος ομοσκεδαστικότητας των τυποποιημένων υπολοίπων σε σχέση με τις εκτιμώμενες τιμές της εξαρτημένης μεταβλητής

Διαπιστώνουμε ότι τα σφάλματα κατανέμονται με τυχαίο τρόπο, συνεπώς ισχύει και η υπόθεση της ομοσκεδαστικότητας. Για να κατασκευάσουμε την ευθεία παλινδρόμησης σε ένα διάγραμμα διασποράς μεταξύ των δύο μεταβλητών, π.χ. «Συνολική θερμική ροή» και «Ηλιοφάνεια», με διπλό κλικ πάνω στο γράφημα επιλέγουμε *Chart editor > Elements > Fit line at Total*.



Εικόνα 7.207: Προσαρμογή μοντέλου παλινδρόμησης μεταξύ της εξαρτημένης «Συνολική θερμική ροή» και της ανεξάρτητης «Ηλιοφάνεια»

Στη συνέχεια έχοντας ορίσει την αποθήκευση των αποστάσεων Cook, κατασκευάζουμε ένα διάγραμμα για τον έλεγχο ύπαρξης σημείων επιρροής. Από το διάγραμμα προκύπτει ότι η 4^η παρατήρηση αποτελεί σημείο επιρροής.



Εικόνα 7.208: Αποστάσεις Cook για τον έλεγχο ύπαρξης σημείων επιρροής

Γραφικές Παραστάσεις Πρόσθετων Μεταβλητών και Μερικών Υπολοίπων

Για να εξετάσουμε αν υπάρχει η ανάγκη στο μοντέλο μας να εισαχθεί η μεταβλητή «Ωρα της Ημέρας» θα χρησιμοποιούμε το διάγραμμα πρόσθετων μεταβλητών (added variable plot) και μερικών υπολοίπων (partial residual plot).

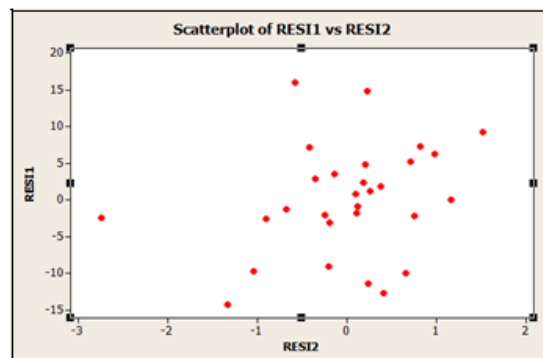
Υπενθυμίζουμε ότι οι παράγοντες που μετρήθηκαν είναι:

- X_1 - «Ηλιοφάνεια (watts/m²)»,
- X_2 - «Ανατολική θέση ακτίνας (inches)»,
- X_3 - «Νότια θέση ακτίνας (inches)»,
- X_4 - «Βόρεια θέση ακτίνας (inches)»,
- X_5 - «Η ώρα της ημέρας (σε δεκαδική κλίμακα)».

Διάγραμμα Πρόσθετων Μεταβλητών

Σ' αυτή την τεχνική, στο πρώτο βήμα εφαρμόζουμε ένα μοντέλο πολλαπλής παλινδρόμησης της εξαρτημένης μεταβλητής σε σχέση με τις ανεξάρτητες X_1 - X_4 , και αποθηκεύουμε τα υπόλοιπα ($RES1$). Στη συνέχεια εφαρμόζουμε πάλι ένα μοντέλο πολλαπλής παλινδρόμησης, όμως πλέον ως εξαρτημένη μεταβλητή θα δηλώσουμε τη μεταβλητή X_5 και ως ανεξάρτητες τις υπόλοιπες X_1 - X_4 . Ομοίως από το μοντέλο αυτό αποθηκεύουμε τα υπόλοιπα ($RES2$).

Στη συνέχεια δημιουργούμε το διάγραμμα διασποράς των υπολοίπων των δύο παλινδρομήσεων. Από το διάγραμμα δεν παρατηρούμε ότι υπάρχει μια κάποια μορφή συσχέτισης. Συνεπώς η μεταβλητή X_5 θα μπορούσε να αφαιρεθεί από το μοντέλο.



Εικόνα 7.209: Διάγραμμα πρόσθετων μεταβλητών για την ανεξάρτητη μεταβλητή «Ωρα της Ημέρας» (X_5)

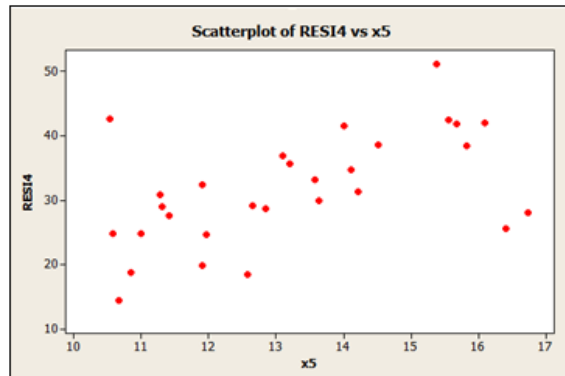
Διάγραμμα Μερικών Υπολοίπων

Σ' αυτή την τεχνική, στο πρώτο βήμα εφαρμόζουμε ένα μοντέλο πολλαπλής παλινδρόμησης της εξαρτημένης μεταβλητής σε σχέση με όλες τις ανεξάρτητες X_1 - X_5 , και αποθηκεύουμε τα υπόλοιπα ($RES3$). Από την προσαρμογή του μοντέλου κρατάμε τον συντελεστή παλινδρόμησης της μεταβλητής που εξετάζουμε:

$$\hat{\beta}_{X_5} = 2,39$$

Εξίσωση 7.81

Στη συνέχεια δημιουργούμε τα μερικά υπόλοιπα ($RESI4$) από τη σχέση: $RESI4 = RESI3 + (2,39 * X_5)$. Στη συνέχεια δημιουργούμε το διάγραμμα διασποράς των μερικών υπολοίπων σε σχέση με τις παρατηρήσεις της υπό εξέταση μεταβλητής (X_5).



Εικόνα 7.210: Διάγραμμα μερικών υπολοίπων για την ανεξάρτητη μεταβλητή «Ωρα της Ημέρας» (X_5)

7.2 Λογιστική Παλινδρόμηση (Logistic Regression)

Στη λογιστική παλινδρόμηση η εξαρτημένη μεταβλητή είναι κατηγορική και δίτιμη. Για αυτήν εξετάζουμε την πιθανότητα (τα ποσοστά) εμφάνισης των δύο κατηγοριών σε σχέση με τις ανεξάρτητες μεταβλητές - παράγοντες. Επειδή σκοπός είναι να εκτιμηθεί η πιθανότητα εμφάνισης ενός συμβάντος, συνεπάγεται ότι οι τιμές που θα πρέπει να προκύπτουν από το γραμμικό υπόδειγμα περιέχονται στο διάστημα $[0,1]$.

Για τον λόγο αυτό υποθέτουμε ότι η μεταβλητή ακολουθεί διωνυμική κατανομή και ότι η σύνδεση της πιθανότητας εμφάνισης του γεγονότος p_i συνδέεται με το γραμμικό υπόδειγμα μέσω της link function:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Εξίσωση 7.82

Όπου:

$$\log\left(\frac{p_i}{1-p_i}\right)$$

Εξίσωση 7.83

Ο λογάριθμος του λόγου σχετικής πιθανότητας. Απολογαριθμίζοντας προκύπτει ότι η πιθανότητα της κατηγορίας της εξαρτημένης μεταβλητής θα είναι:

$$p_i = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}$$

Εξίσωση 7.84

Στο μοντέλο της λογιστικής παλινδρόμησης η εκτίμηση των συντελεστών πραγματοποιείται με τη μέθοδο μέγιστης πιθανοφάνειας (maximum likelihood method) αντί της μεθόδου ελαχίστων τετραγώνων.

Η ερμηνεία τους, όμως, δεν προκύπτει με τον ίδιο τρόπο όπως στη γραμμική παλινδρόμησης, και θα πρέπει να γίνει τροποποίηση, ώστε να εκφραστούν με την κατάλληλη μορφή, δηλ. e^{β} . Τα περισσότερα προγράμματα στον πίνακα των συντελεστών εμφανίζουν και την σχέση e^{β} . Κάθε συντελεστής εκφράζει τη μεταβολή του λογαρίθμου της σχετικής πιθανότητας (Εξίσωση 7.83) για μια μονάδα αύξησης της ανεξάρτητης μεταβλητής.

Έστω ότι έχουμε ένα μοντέλο με μια ανεξάρτητη δίτιμη μεταβλητή (π.χ. το φύλο όπου 0 = γυναίκες και 1 = άνδρες). Οπότε έχουμε:

$$\log\left(\frac{p_A}{1-p_A}\right) = \beta_0 + \beta_1 \Rightarrow \frac{p_A}{1-p_A} = e^{\beta_0 + \beta_1}$$

Εξίσωση 7.85

Είναι η σχετική πιθανότητα πραγματοποίησης του συμβάντος στους άνδρες, και:

$$\log\left(\frac{p_\Gamma}{1-p_\Gamma}\right) = \beta_0 \Rightarrow \frac{p_\Gamma}{1-p_\Gamma} = e^{\beta_0}$$

Εξίσωση 7.86

Είναι η σχετική πιθανότητα πραγματοποίησης του συμβάντος στις γυναίκες. Διαιρώντας τα δύο μέλη λαμβάνουμε τον λόγο συμπληρωματικών πιθανοτήτων (odds ratio) πραγματοποίησης του συμβάντος των ανδρών προς την πραγματοποίηση του συμβάντος των γυναικών, δηλ.:

$$\frac{\frac{p_A}{1-p_A}}{\frac{p_\Gamma}{1-p_\Gamma}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Εξίσωση 7.87

Όπως και στη γραμμική παλινδρόμηση, ομοίως οι συντελεστές του μοντέλου, οι οποίοι αποτελούν σημειακές εκτιμήσεις, υπόκεινται σε έλεγχο για τη σημαντικότητά τους στο μοντέλο. Δηλαδή η υπόθεση που εξετάζεται είναι η:

$$H_0 : \hat{\beta}_i = 0$$

$$H_1 : \hat{\beta}_i \neq 0$$

Εξίσωση 7.88

Ο έλεγχος γίνεται σε $(1-\alpha)\%$ επίπεδο σημαντικότητας και η στατιστική συνάρτηση υπό την H_0 είναι η:

$$\frac{\hat{\beta}_i}{s.e.(\hat{\beta}_i)}$$

Εξίσωση 7.89

Η οποία καλείται Wald, ακολουθεί ασυμπτωτικά την κανονική κατανομή, ενώ το τετράγωνο αυτής ακολουθεί την X^2 -κατανομή με 1 βαθμό ελευθερίας. Η υπόθεση απορρίπτεται αν η τιμή του ελέγχου είναι μεγαλύτερη από $X^2_{1,\alpha/2}$.

Η σύγκριση μοντέλων με σκοπό την επιλογή βέλτιστου μοντέλου, πραγματοποιείται με τη βοήθεια της ελεγχουσυνάρτησης Deviance (που στην ουσία είναι ένα likelihood ratio test) η οποία ορίζεται ως:

$$D = -2\{\hat{l}_2 - \hat{l}_1\}$$

Εξίσωση 7.90

Όπου:

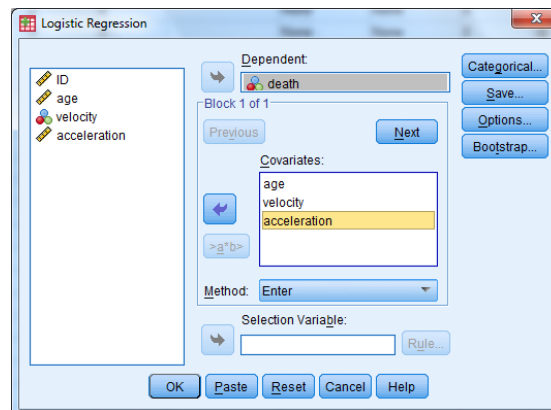
$$\hat{l}_2$$

Η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας του μοντέλου 2.

$$\hat{l}_1$$

Η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας του μοντέλου 1. Επίσης το μοντέλο 2 είναι υποσύνολο (nested) του μοντέλου 1, και ακολουθεί την X^2 κατανομή με $k-p-1$ βαθμούς ελευθερίας.

Έστω ότι θέλουμε να εξετάσουμε την πιθανότητα επιβίωσης του οδηγού σε ένα τροχαίο ατύχημα, σε σχέση με τους παράγοντες «Ηλικία Οδηγού», «Ταχύτητα Οχήματος» και «Επιτάχυνση Οχήματος». Η δυνατότητα προσφέρεται από την επιλογή *Analyze > Regression > Binary Logistic*.



Εικόνα 7.211: Το menu Logistic Regression

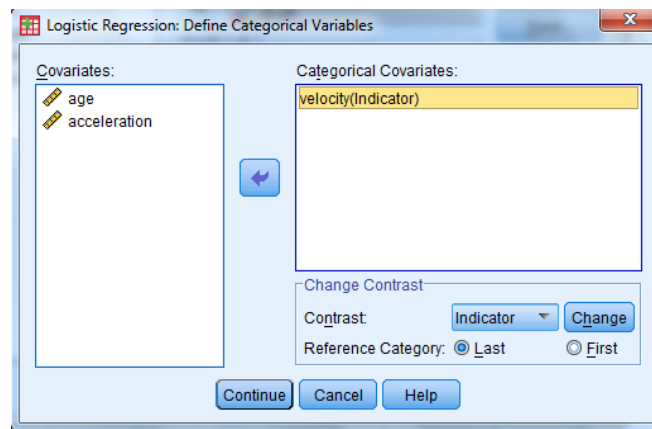
Στο πλαίσιο *Dependent* ορίζουμε την εξαρτημένη δίτιμη μεταβλητή, και στο πλαίσιο *Covariates* εισάγουμε τους προς εξέταση παράγοντες. Στη συνέχεια στο πεδίο *Method* επιλέγουμε τη μέθοδο επιλογής του βέλτιστου μοντέλου. Οι δυνατές επιλογές είναι:

- *Enter*: Μ' αυτή την μέθοδο εισάγονται στο μοντέλο όλες οι μεταβλητές ανεξάρτητα από το αν είναι στατιστικά σημαντικές σε σχέση με την εξαρτημένη μεταβλητή.
- *Forward Selection (Conditional)*: Είναι μια διαδικασία στην οποία οι μεταβλητές εισέρχονται σταδιακά στο μοντέλο. Ο έλεγχος για την εισαγωγή μιας μεταβλητής βασίζεται στη στατιστική σημαντικότητα της μεταβλητής, και ο έλεγχος για την αφαίρεση μιας μεταβλητής βασίζεται στην πιθανότητα του λόγου πιθανοφάνειας βασιζόμενου στην υπό συνθήκη εκτίμηση των παραμέτρων.
- *Forward Selection (Likelihood Ratio)*: Είναι μια διαδικασία στην οποία οι μεταβλητές εισέρχονται σταδιακά στο μοντέλο. Ο έλεγχος για την εισαγωγή μιας μεταβλητής βασίζεται στη στατιστική σημαντικότητα της μεταβλητής, και ο έλεγχος για την αφαίρεση μιας μεταβλητής βασίζεται στην πιθανότητα του λόγου πιθανοφάνειας βασιζόμενου στην εκτίμηση της μέγιστης μερικής πιθανοφάνειας.
- *Forward Selection (Wald)*: Είναι μια διαδικασία στην οποία οι μεταβλητές εισέρχονται σταδιακά στο μοντέλο. Ο έλεγχος για την εισαγωγή μιας μεταβλητής βασίζεται στη στατιστική σημαντικότητα της μεταβλητής, και ο έλεγχος για την αφαίρεση μιας μεταβλητής βασίζεται στην πιθανότητα της στατιστικής Wald.
- *Backward Elimination (Conditional)*: Είναι μια διαδικασία στην οποία οι μεταβλητές αφαιρούνται σταδιακά από το μοντέλο. Ο έλεγχος για την αφαίρεση μιας μεταβλητής βασίζεται στην πιθανότητα του λόγου πιθανοφάνειών υπολογισμένου με βάση την υπό συνθήκη εκτίμηση των παραμέτρων.

- *Backward Elimination (Likelihood Ratio)*: Είναι μια διαδικασία στην οποία οι μεταβλητές αφαιρούνται σταδιακά από το μοντέλο. Ο έλεγχος για την αφαίρεση μιας μεταβλητής βασίζεται στην πιθανότητα του λόγου πιθανοφανειών υπολογισμένου με βάση την εκτίμηση της μέγιστης μερικής πιθανοφάνειας.
- *Backward Elimination (Wald)*: Είναι μια διαδικασία στην οποία οι μεταβλητές αφαιρούνται σταδιακά από το μοντέλο. Ο έλεγχος για την αφαίρεση μιας μεταβλητής είναι βασισμένος στην πιθανότητα της στατιστικής Wald.

Στο πλαίσιο *Selection Variable* μπορούμε να ορίσουμε μια μεταβλητή καθώς και τη συνθήκη μέσω της οποίας θα απομονώσουμε το υποσύνολο των δεδομένων στο οποίο επιθυμούμε να εφαρμόσουμε την παλινδρόμηση.

Από το πλήκτρο *Categorical* ορίζουμε τις κατηγορικές μεταβλητές του μοντέλου με σκοπό τη δημιουργία ψευδομεταβλητών. Όμως, και οι τρεις ανεξάρτητες μεταβλητές είναι συνεχείς, συνεπώς δεν απαιτείται ούτε η δημιουργία ψευδομεταβλητών ούτε κάποια αλλαγή στην κωδικοποίηση των δεδομένων. Ωστόσο, να αναφέρουμε ότι αν για παράδειγμα η μεταβλητή «Ταχύτητα» (velocity) είχε μετατραπεί σε κατηγορική με περισσότερες από δύο κατηγορίες, τότε θα έπρεπε για την εισαγωγή της στο μοντέλο να οριστούν οι αντίστοιχες δείκτριες κατηγορικές, μεταφέροντας τη μεταβλητή στο πλαίσιο *Categorical Covariates* και επιλέγοντας *Contrast = Indicator*.



Εικόνα 7.212: Το menu *Logistic Regression > Define Categorical Variables*

Αναλυτικότερα οι δυνατές επιλογές που προσφέρονται από την εφαρμογή είναι:

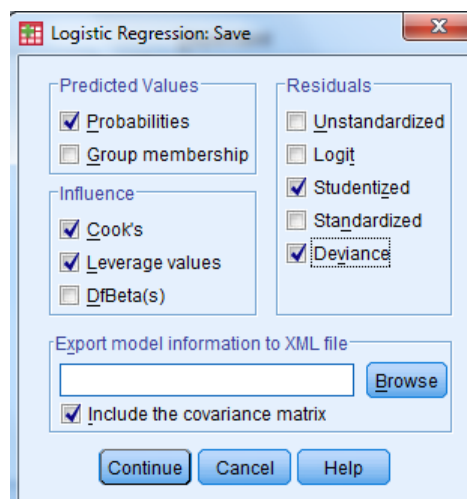
- *Indicator*: Με την επιλογή αυτή γίνεται επανακωδικοποίηση της κατηγορικής, δημιουργώντας ψευδομεταβλητές-δείκτες, όπου θα δηλώνει την παρουσία με την τιμή 1 και την απουσία της κατηγορίας με την τιμή 0. Προφανώς για την εισαγωγή μιας κατηγορικής με k κατηγορίες, απαιτούνται $k-1$ δείκτες. Έτσι η σχετική πιθανότητα της παρουσίας της κάθε κατηγορίας θα συγκρίνεται με τη σχετική πιθανότητα της κατηγορίας αναφοράς.
- *Simple*: Μ' αυτή την επιλογή ομοίως δημιουργούνται ψευδομεταβλητές-δείκτες, όμως αλλάζει ο τρόπος κωδικοποίησης της κατηγορίας αναφοράς. Ομοίως κάθε κατηγορία της ανεξάρτητης μεταβλητής (εκτός από την κατηγορία αναφοράς) θα συγκρίνεται με την κατηγορία αναφοράς.
- *Difference*: Μ' αυτή την επιλογή, κάθε κατηγορία της ανεξάρτητης μεταβλητής εκτός από την πρώτη κατηγορία, θα συγκρίνεται με τη μέση σχετική πιθανότητα των προηγούμενων κατηγοριών. Η μέθοδος είναι επίσης γνωστή με την ονομασία αντίστροφη Helmert.
- *Helmert*: Μ' αυτή την επιλογή, κάθε κατηγορία της ανεξάρτητης μεταβλητής εκτός από την τελευταία κατηγορία, θα συγκρίνεται με τη μέση σχετική πιθανότητα των μεταγενέστερων κατηγοριών.

- *Repeated*: Μ' αυτή την επιλογή, η σχετική πιθανότητα κάθε κατηγορίας της ανεξάρτητης μεταβλητής εκτός από την πρώτη κατηγορία, θα συγκρίνεται με την κατηγορία που προηγήθηκε.
- *Polynomial*: Χρησιμοποιήθηκε για τη δημιουργία ορθογώνιων πολυωνυμικών αντιθέσεων, με σκοπό ελέγχους γραμμικής φύσεως, όπου οι κατηγορίες υποτίθεται ότι ισαπέχουν, ελέγχοντας αν υπάρχουν γραμμικού τύπου μεταβολές στη σχετική πιθανότητα. Πολυωνυμικές αντιθέσεις παρέχονται μόνο για συνεχείς μεταβλητές.
- *Deviation*: Μ' αυτή την επιλογή κάθε κατηγορία της ανεξάρτητης μεταβλητής εκτός από την κατηγορία αναφοράς θα συγκρίνεται με τη συνολική σχετική πιθανότητα.

Σημεία προσοχής

Αν επιλεγούν Deviation, Simple, ή Indicator, θα πρέπει επίσης να οριστεί αν ως κατηγορία αναφοράς θα είναι η πρώτη (First) ή η τελευταία (Last) κατηγορία στην επιλογή *Reference Category*.

Από το πλήκτρο *Save* μπορούμε να αποθηκεύσουμε ως νέες μεταβλητές τις εκτιμώμενες παρατηρήσεις, τα σφάλματα, και άλλα χρήσιμα στατιστικά για τα διαγνωστικά του μοντέλου. Μέσω αυτών αντλούμε τα στοιχεία εκείνα που θα μας οδηγήσουν στην επιλογή του «βέλτιστου» μοντέλου.



Εικόνα 7.213: Το menu *Logistic Regression > Save*

Σχετικά με τις εκτιμώμενες παρατηρήσεις (Predicted Values) έχουμε:

- *Probabilities*: Για κάθε παρατήρηση εκτιμάται η προβλεπόμενη πιθανότητα εμφάνισης του συμβάντος. Το πρόγραμμα λαμβάνει ως συμβάν την κατηγορία με την υψηλότερη τιμή κωδικοποίησης.
- *Predicted Group membership*: Είναι ο κωδικός της κατηγορίας της εξαρτημένης μεταβλητής, στην οποία θα έπρεπε να ανήκει η κάθε παρατήρηση βασισμένη σε διαχωριστική ανάλυση.

Για την εύρεση σημείων επιρροής (Influence) έχουμε:

- *Cook's*: Ένα μέτρο ανάλογο της γραμμικής παλινδρόμησης, που δείχνει το κατά πόσο τα σφάλματα όλων των παρατηρήσεων θα αλλάξουν αν η συγκεκριμένη παρατήρηση αποκλειστεί από τον υπολογισμό των συντελεστών παλινδρόμησης.
- *Leverage Values*: Η σχετική επίδραση της κάθε παρατήρησης στην προσαρμογή του μοντέλου.
- *DfBeta(s)*: Υπολογίζεται η διαφορά στους συντελεστές παλινδρόμησης από τον αποκλεισμό μιας συγκεκριμένης παρατήρησης από την παλινδρόμηση. Οι διαφορές υπολογίζονται για όλους τους όρους του μοντέλου, συμπεριλαμβανομένου και του σταθερού όρου.

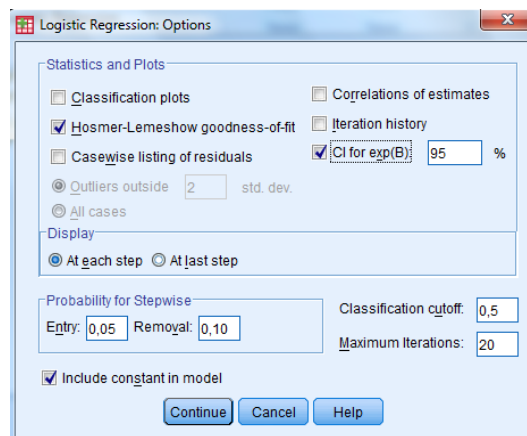
Για την εκτίμηση των σφαλμάτων (Residuals) έχουμε:

- *Unstandardized Residuals*: Μη τυποποιημένα σφάλματα, δηλ. η διαφορά μεταξύ μιας παρατήρησης και της εκτιμώμενης τιμής από το μοντέλο.
- *Logit Residual*: Είναι τα σφάλματα που προκύπτουν αν τα μη τυποποιημένα σφάλματα διαιρεθούν με το γινόμενο $p(1-p)$, όπου p η προβλεπόμενη πιθανότητα.
- *Studentized Residual*: Είναι η μεταβολή στη deviance του μοντέλου, αν η παρατήρηση εξαιρεθεί από την ανάλυση.
- *Standardized Residuals*: Είναι τα τυποποιημένα σφάλματα, ή αλλιώς σφάλματα Pearson, δηλαδή τα μη τυποποιημένα σφάλματα διαιρεμένα με την εκτίμηση της τυπικής τους απόκλισης.

$$\sqrt{p(1-p)}$$

Εξίσωση 7.91

- *Deviance*: σφάλματα βασισμένα στην deviance του μοντέλου.



Εικόνα 7.214: Το menu *Logistic Regression > Options*

Πρόσθετα στατιστικά που προσφέρονται μέσω του πλήκτρου *Options*.

- *Classification plots*: Για την εμφάνιση γραφήματος ταξινόμησης των παρατηρήσεων.
- *Hosmer-Lemeshow goodness-of-fit*: Είναι ένα μέτρο καλής προσαρμογής του μοντέλου, πιο ισχυρό από τη συνήθη στατιστική καλής προσαρμογής (goodness-of-fit), ειδικότερα για μοντέλα με συνεχείς μεταβλητές και μικρό μέγεθος δεδομένων. Βασίζεται στην ομαδοποίηση των παρατηρήσεων σε δεκατημόρια κινδύνου και στη σύγκριση της παρατηρούμενης με την εκτιμώμενη πιθανότητα σε κάθε δεκατημόριο.
- *Correlations of estimates*: Συσχετίσεις των εκτιμήσεων.
- *CI for exp(B)*: Την εμφάνιση διαστημάτων εμπιστοσύνης της σχέσης e^{B_i} .

Τα παραπάνω δύναται να εμφανιστούν είτε σε κάθε στάδιο της ανάλυσης (At each step) είτε μόνο στο τελικό μοντέλο (At last step).

- *Probability for Stepwise*: Σ' αυτό το σημείο ορίζονται τα κριτήρια εισόδου ή εξόδου μιας μεταβλητής στο μοντέλο, όταν εκτελούνται οι αυτόματες διαδικασίες επιλογής μοντέλου (forward, backward ή stepwise). Έτσι, μια μεταβλητή θα εισαχθεί στο μοντέλο αν το επίπεδο σημαντικότητάς της (με βάση το αντίστοιχο κριτήριο εισαγωγής, π.χ. likelihood ratio statistic) είναι μικρότερο από την τιμή εισαγωγής (Entry), και θα αφαιρεθεί αν είναι μεγαλύτερο από την τιμή αποκλεισμού (Removal).

- *Classification cutoff*: Στο σημείο αυτό ορίζεται η τιμή της πιθανότητας με την οποία θα γίνει ο διαχωρισμός / ταξινόμηση των παρατηρήσεων. Παρατηρήσεις των οποίων η προβλεπόμενη πιθανότητα που υπερβαίνει τη δηλωθείσα τιμή χαρακτηρίζεται ως θετική, αλλιώς ως αρνητική.
- *Maximum Iterations*: Ορίζεται το πλήθος των επαναλήψεων του αλγορίθμου που εκτιμά τους συντελεστές του μοντέλου.
- *Include constant in model*: Είναι προεπιλεγμένη παράμετρος προκειμένου στο μοντέλο να περιέχεται και ο σταθερός όρος.

Χρησιμοποιήθηκε το αρχείο δεδομένων *carcrash.sav* (βλ. σύνοψη) όπου περιέχονται δεδομένα ηλιακής ενέργειας $N = 58$ παρατηρήσεων. Έτσι εφαρμόζοντας τη διαδικασία λαμβάνουμε:

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	58	100,0
	Missing Cases	0	,0
	Total	58	100,0
Unselected Cases		0	,0
Total		58	100,0

a. If weight is in effect, see classification table for the total number of cases.

Εικόνα 7.215: Πλήθος δεδομένων

Ο πίνακας αποτυπώνει το πλήθος των έγκυρων και των ελλειπουσών τιμών. Εάν σε κάποια εκ των ανεξάρτητων μεταβλητών υπάρχει ελλείπουσα παρατήρηση τότε η περίπτωση αυτή εξαιρείται από την ανάλυση. Είναι σημαντικό να γίνεται έλεγχος του πλήθους των ελλειπουσών τιμών καθώς η ύπαρξή τους μειώνει το μέγεθος του δείγματος στην ανάλυση.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Εικόνα 7.216: Εσωτερική κωδικοποίηση της εξαρτημένης μεταβλητής

Ο πίνακας μας δείχνει την εσωτερική κωδικοποίηση των τιμών της δίτιμης εξαρτημένης μεταβλητής. Στην περίπτωσή μας η μεταβλητή είχε εξ αρχής κωδικοποιηθεί με τις τιμές 0 και 1, όπου σύμφωνα και με την παραμετροποίηση στο πλήκτρο *Categorical > Reference Category*, 0 σημαίνει την απουσία του συμβάντος (θάνατος) και 1 σημαίνει την εμφάνιση του συμβάντος.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
age	58	19	65	39,78	13,534
velocity	58	40	61	47,91	5,381
acceleration	58	58	268	122,79	40,960
Valid N (listwise)	58				

Εικόνα 7.217: Περιγραφικά μέτρα των ανεξάρτητων μεταβλητών

Η ηλικία των οδηγών κυμαίνεται από 19-65 με μέση ηλικία 40 ετών, οι οποίοι οδηγούσαν με ταχύτητα από 40 - 61 km/h και μέση ταχύτητα τα 50 km/h, ενώ η επιτάχυνση κυμαίνεται από 58-268 m/sec², με μέση επιτάχυνση 123 m/sec².

Variables in the Equation							95% C.I. for EXP(B)		
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1 ^a	age	,171	,043	15,629	1	,000	1,186	1,090	1,291
	velocity	,146	,112	1,700	1	,192	1,158	,929	1,442
	acceleration	,016	,014	1,246	1	,264	1,016	,988	1,046
	Constant	-15,054	5,297	8,077	1	,004	,000		

a. Variable(s) entered on step 1: age, velocity, acceleration.

Εικόνα 7.218: Πίνακας των συντελεστών παλινδρόμησης του μοντέλου

Το μοντέλο που προσαρμόστηκε είναι το:

$$\log\left(\frac{p}{1-p}\right) = -15.054 + 0.171age + 0.146velocity + 0.016acceleration$$

Εξίσωση 7.92

Από το μοντέλο προκύπτει ότι:

- Αύξηση της ηλικίας κατά 1 έτος αναμένεται να αυξήσει κατά 18,6% την πιθανότητα αποβίωσης. Η αύξηση στο odds ratio υπολογίζεται από τη σχέση $(1,186 - 1) * 100 = 18,6\%$. Εκφράζοντας διαφορετικά το αποτέλεσμα έχουμε ότι σε αύξηση της ηλικίας κατά μια δεκαετία (πολλαπλασιάζουμε τον συντελεστή παλινδρόμησης επί 10 και απολογαριθμούμε), η πιθανότητα αποβίωσης του οδηγού είναι 5,53 φορές μεγαλύτερη από την πιθανότητα επιβίωσης.
- Σε αύξηση της ταχύτητας κατά 10 km/h, η πιθανότητα αποβίωσης θα είναι 4,3 φορές μεγαλύτερη από την πιθανότητα επιβίωσης.
- Σε αύξηση της επιτάχυνσης κατά 10 m/sec², η πιθανότητα αποβίωσης θα είναι 1,173 φορές μεγαλύτερη από την πιθανότητα επιβίωσης.

Ωστόσο, από τα παραπάνω μόνο η ηλικία προκύπτει στατιστικά σημαντική για την πιθανότητα αποβίωσης ($p\text{-value} < 0,001$) σύμφωνα με τον έλεγχο Wald. Στις τελευταίες θέσεις του πίνακα προσφέρονται και τα διαστήματα εμπιστοσύνης των odds ratio.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	43,976 ^a	,450	,606

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Εικόνα 7.219: Συντελεστές Προσδιορισμού και λογαριθμο-πιθανοφάνεια

Από τον συντελεστή προσδιορισμού του Nagelkerke R έχουμε ότι περίπου το 61% της μεταβλητότητας της εξαρτημένης μεταβλητής ερμηνεύεται από τις ανεξάρτητες μεταβλητές του μοντέλου.

Classification Table^a

Observed		Predicted		
		death		Percentage Correct
		0	1	
Step 1 death	0	18	6	75,0
	1	4	30	88,2
Overall Percentage				82,8

a. The cut value is ,500

Εικόνα 7.220: Ακρίβεια πρόβλεψης του μοντέλου

Η συνολική ακρίβεια του μοντέλου να προβλέψει αν ένα άτομο που θα εμπλακεί σε ατύχημα, θα πεθάνει ή όχι είναι 82,8%. Από τον παραπάνω πίνακα μπορούμε επίσης να λάβουμε την ευαισθησία $30/34 = 88,2\%$ και την ειδικότητα $18/24=75\%$. Η θετική διαγνωστική αξία (*PPV-Positive Predictive Value*) μπορεί να υπολογιστεί από τη σχέση $30/36=83,3\%$ και η αρνητική διαγνωστική αξία (*NPV-Negative Predictive Value*) από τη σχέση $18/22=81,8\%$.

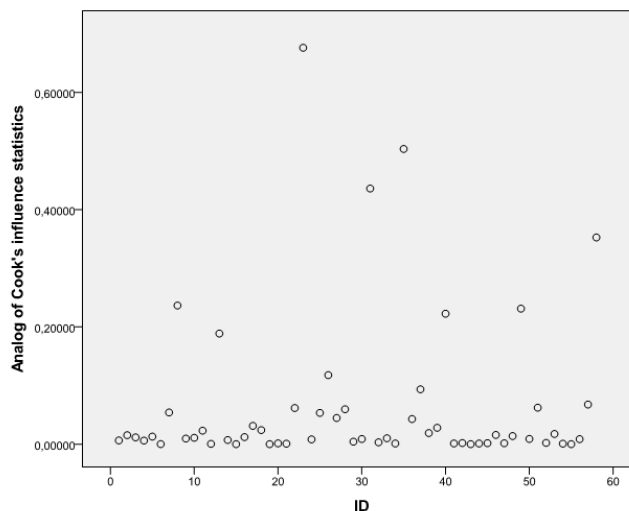
Στην περίπτωση που έχουμε μια νέα υπόθεση, τότε έχοντας τις τιμές του ατόμου στις ανεξάρτητες μεταβλητές, μπορούμε να προβλέψουμε τι θα συμβεί μέσω του μοντέλου που εκτιμήθηκε. Τα παραπάνω προέκυψαν έχοντας ως σημείο διαχωρισμού (cutoff) την προεπιλεγμένη τιμή 0,5 για την οποία προκύπτουν καλά αποτελέσματα. Θα μπορούσαμε όμως να αλλάζουμε το σημείο διαχωρισμού δηλώνοντας όπως είπαμε μια διαφορετική τιμή στο ομώνυμο πεδίο του πλήκτρου *Options*.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	12,289	8	,139

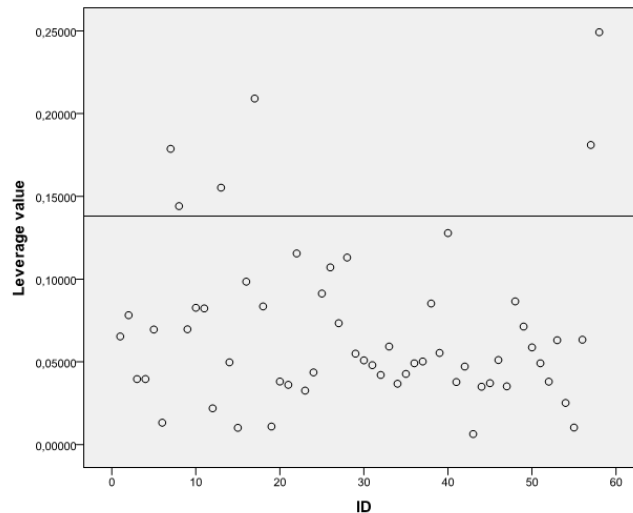
Εικόνα 7.221: Αποτελέσματα ελέγχου καλής προσαρμογής

Ο έλεγχος των Hosmer-Lemeshow είναι ένα στατιστικό τεστ καλής προσαρμογής για μοντέλα λογιστικής παλινδρόμησης, το οποίο αξιολογεί κατά πόσο οι παρατηρούμενες τιμές του λόγου πιθανότητας του γεγονότος (συμβάντος) συμφωνούν με τους αναμενόμενους λόγους σε υποομάδες του πληθυσμού του μοντέλου. Από τον έλεγχο συμπεραίνουμε καλή προσαρμογή του μοντέλου.



Εικόνα 7.222: Αποστάσεις Cook για τον έλεγχο πιθανών σημείων επιρροής.

Όλες οι αποστάσεις Cook είναι μικρότερες της μονάδας, συνεπώς δεν διαπιστώνεται η ύπαρξη σημείων επιρροής.



Εικόνα 7.223: Leverage αποστάσεις για τον έλεγχο πιθανών ακραίων παρατηρήσεων

Από το γράφημα διαπιστώνουμε την ύπαρξη μερικών πιθανών ακραίων παρατηρήσεων:

$$\text{Leverage αποστάσεις} > 0,103 \left[= \frac{2p}{n} = \frac{2 \cdot 3}{58} \right]$$

Εξίσωση 7.93

Correlation Matrix

		Constant	age	velocity	acceleration
Step 1	Constant	1,000	-,610	-,874	,224
	age	-,610	1,000	,221	,264
	velocity	-,874	,221	1,000	-,625
	acceleration	,224	,264	-,625	1,000

Εικόνα 7.224: Πίνακας συσχετίσεων για τον έλεγχο πολυσυγγραμμικότητας στο μοντέλο

Από τον πίνακα των συντελεστών συσχέτισης του μοντέλου διαπιστώνουμε μια κάπως υψηλή συσχέτιση μεταξύ «Ταχύτητας» και «Επιτάχυνσης», συνεπώς ενδεχομένως να υπάρχει πολυσυγγραμμικότητα. Ομοίως και μεταξύ της «Ηλικίας» και της «Ταχύτητας» σε σχέση με τη σταθερά.

ROC Curve

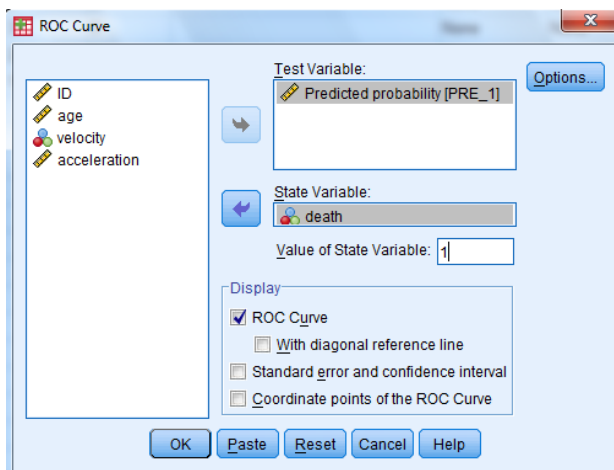
Ένα μέτρο για τον έλεγχο καλής προσαρμογής που χρησιμοποιείται συχνά στη λογιστική παλινδρόμηση είναι η καμπύλη λειτουργικού χαρακτηριστικού δείκτη (*ROC-Receiver Operating Characteristic*).

Η καμπύλη δημιουργείται με τη γραφική αναπαράσταση του πραγματικού θετικού ποσοστού (*TPR-True Positive Rate*), έναντι του ποσοστού ψευδών θετικών (*FPR-False Positive Rate*) σε διάφορα σημεία διαχωρισμού. Το πραγματικό θετικό ποσοστό είναι επίσης γνωστό ως «ευαισθησία», ενώ το ψευδώς θετικό ποσοστό είναι γνωστό ως 1-ειδικότητα (*se-sensitivity*), όπου ως ειδικότητα (*sp-specificity*) ορίζεται το πραγματικά αρνητικό ποσοστό. Η καμπύλη ROC είναι συνεπώς η *se* σε σχέση με την *1-sp*.

Η δυνατότητα προσφέρεται από την επιλογή *Analyze > ROC Curve*.

Στο πεδίο *Test Variable* εισάγουμε τις προβλεπόμενες πιθανότητες που δημιουργήθηκαν ως νέα μεταβλητή μέσα από το πλήκτρο *Save > Probabilities* κατά την εκτέλεση της λογιστικής παλινδρόμησης. Στο πεδίο *State Variable* εισάγουμε την κατηγορική (εξαρτημένη) μεταβλητή, δηλαδή τη μεταβλητή που δείχνει

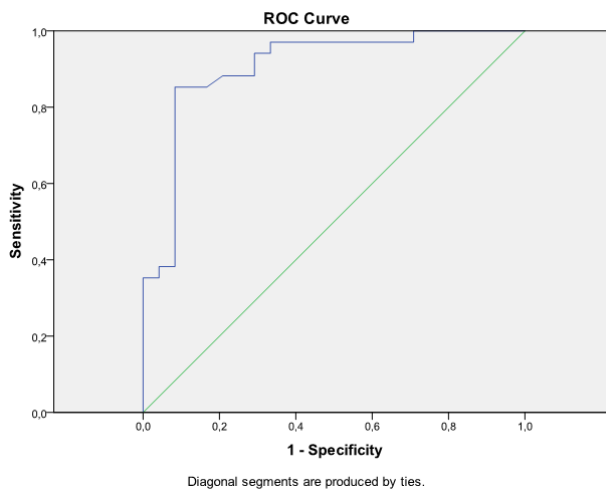
την πραγματική κατηγορία στην οποία ανήκει το άτομο, και στο πεδίο *Value of State Variable* την τιμή που αντιπροσωπεύει την παρουσία του συμβάντος.



Εικόνα 7.225: Το menu ROC Curve

Επιπροσθέτως, μπορούμε να επιλέξουμε την εμφάνιση της διαγώνιας ευθείας *With diagonal reference line*, του τυπικού σφάλματος της εκτίμησης του εμβαδού κάτω από την καμπύλη καθώς και διάστημα εμπιστοσύνης της εκτίμησης του *Standard error and confidence interval*.

Η διαδικασία ολοκληρώνεται με το πλήκτρο *OK* όπου και λαμβάνουμε:



Εικόνα 7.226: Καμπύλη ROC (*Receiver Operating Characteristic curve*)

Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,906	,042	,000	,823	,989

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Εικόνα 7.227: Περιοχή κάτω από την καμπύλη (*Area Under Curve*)

Η περιοχή κάτω από την καμπύλη (area under curve) αντιπροσωπεύει την πιθανότητα ένας διαγνωστικός έλεγχος να διαχωρίσει ορθά ένα τυχαίο ζεύγος ενός ατόμου με την παρουσία του συμβάντος, και ενός ατόμου με την απουσία του συμβάντος. Λαμβάνει τιμές από 0 έως 1, και όσο μεγαλύτερη είναι η τιμή του εμβαδού κάτω από την καμπύλη, τόσο μεγαλύτερη είναι η ακρίβεια του διαγνωστικού ελέγχου.

- Για τιμές $0,90-1 >$ Εξαιρετικό διαχωρισμό.
- Για τιμές $0,80-0,90 >$ Καλό διαχωρισμό.
- Για τιμές $0,70-0,80 >$ Σχετικά καλό διαχωρισμό.
- Για τιμές $0,60-0,70 >$ Φτωχό διαχωρισμό.
- Για τιμές $0,50-0,60 >$ Τότε ο διαγνωστικός έλεγχος αποτυγχάνει να κάνει ορθό διαχωρισμό.
- Για τιμές από $0-0,5 >$ Ο διαχωρισμός λειτουργεί προς την αντίθετη κατεύθυνση.

Η 3^η στήλη του πίνακα περιέχει τις *p-value* για τον έλεγχο της μη καλής ορθής ταξινόμησης του διαγνωστικού ελέγχου, ενώ η 4^η και η 5^η στήλη μας δίνουν το 95% διάστημα εμπιστοσύνης για την πραγματική τιμή του εμβαδού κάτω από την καμπύλη.

Βιβλιογραφικές Αναφορές Κεφαλαίου

Ελληνόγλωσσες

Howitt, D. & Cramer, D. (Επιμέλεια: Σ. Κοντάκος) (2011). *Στατιστική με το SPSS 16.0*. Αθήνα: Εκδόσεις Κλειδάριθμος.

Οικονόμου, Π. & Καρώνη, Χ. (2002). *Στατιστικά μοντέλα με το MINITAB*. Αθήνα: Εκδόσεις Ε.Μ.Π.

Καρώνη, Χ. & Οικονόμου, Π. (2010). *Στατιστικά μοντέλα παλινδρόμησης*. Αθήνα: Εκδόσεις Συμμεών.

Ξενόγλωσσες

Carver, R. & Nash, J. (2011). *Doing Data Analysis with SPSS: Version 18.0*. 5th Edition Easton Cengage Learning.

Coakes, S. J. & Steed, L. G. (1999). *SPSS: Analysis without anguish: Versions 7.0, 7.5, 8.0 for Windows*. Brisbane: Jacaranda Wiley.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Marques de Sa, J. P. (2007). *Applied Statistics Using SPSS. STATISTICA, MATLAB and R*, 2nd Edition, Porto: Universidade de Porto

Montgomery, D. C. & Peck, E. A. (1992). *Introduction to Linear Regression Analysis*. New York: Wiley-Interscience Publication. New York, USA.

Norusis Marija, J. (2002). *SPSS 11.0 Guide to data analysis*. Upper Saddle River New Jersey: Prentice Hall.

IBM Knowledge Center. Retrieved August 30, 2015 from http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/idh_glm_u_sav.htm

ΓΛΩΣΣΑΡΙ

A

ANOVA πίνακας ANOVA table

D

Deviance ελεγχοςυνάρτηση Deviance statistic

Durbin - Watson στατιστική συνάρτηση Durbin - Watson statistic

E

Exit-polls Exit-polls

F

F κατανομή F distribution

F στατιστική συνάρτηση F statistic

Friedman F έλεγχος Friedman F test

H

Hosmer - Lemeshov έλεγχος καλής
προσαρμογής Hosmer - Lemeshov goodness of fit

K

Kolmogorov - Smirnov Kolmogorov - Smirnov

Kruskal - Wallis H έλεγχος Kruskal - Wallis H test

L

Likert κλίμακα Likert scale

Q

Q-Q γράφημα Q-Q Plot

R

r - Pearson συντελεστής συσχέτισης r - Pearson Correlation Coefficient

r - Spearman συντελεστής συσχέτισης r - Spearman Correlation Coefficient

S

Shapiro-Wilks Shapiro-Wilks

T

t κατανομή	t distribution
t στατιστική συνάρτηση	t statistic

W

Wald έλεγχος	Wald test
Wilcoxon έλεγχος	Wilcoxon test

X

X^2 έλεγχος	Chi - square test
---------------	-------------------

A

Αθροιστική Συχνότητα	Cumulative percent
Ακραία τιμή	Outlier
Ακρίβεια	Accuracy
Αλληλεπίδραση	Interaction
Αναγωγισμός	Reductionism
Ανάλυση διακύμανσης	Analysis of variance
Ανάλυση διακύμανσης δύο παραγόντων	Two-way Analysis of Variance
Ανάλυση κατά παράγοντες	Factor Analysis
Αναμενόμενη τιμή	Expected value
Ανεξάρτητη (μεταβλητή)	Independent (Variable)
Αξιοπιστία	Reliability
Απλή γραμμική παλινδρόμηση	Simple linear regression
Απλή τυχαία δειγματοληψία	Simple Random sampling
Απόρριψης περιοχή	Rejection area
Απόσταση Cook	Cook's distance
Αυτοσυσχέτιση	Autocorrelate

B

Βαθμοί ελευθερίας	Degrees of freedom
-------------------	--------------------

Γ

Γραμμική παρεμβολή	Linear Interpolation
Γραμμική τάση	Linear Trend
Γραμμικό διάγραμμα	Line Chart
Γράφημα περιοχής	Area plot

Δ

Δείγμα	Sample
Δειγματοληπτική έρευνα	Sample Survey
Δειγματοληψία ευκαιρίας	Convenience sampling
Δειγματοληψία κατά συστάδες	Cluster sampling
Δειγματοληψία κρίσεως ή σκόπιμη δειγματοληψία	Judgement sampling
Δειγματοληψία της χιονοστιβάδας	Snowball sampling
Δείκτης (μεταβλητή)	Indicator (Variable)
Δημοσκοπήσεις	Polls
Διακριτή (μεταβλητή)	Discrete (Variable)
Διακύμανση	Variance
Διάμεσος	Median
Διασποράς	Scatter/ Dot plot
Διάστημα Εμπιστοσύνης	Confidence Interval
Διατάξιμες Μεταβλητές	Ordinal Variables
Διατομή	Intercept
Διπλή δειγματοληψία	Double sampling

Ε

Εγκυρότητα	Validity
Έλεγχος σημαντικότητας	Test of Significance
Ελεγχοςυνάρτηση	Test statistic
Ελληνική Στατιστική Αρχή	Greek Statistical Society

Εμφωλευμένο μοντέλο	Nested model
Εναλλακτική υπόθεση	Alternative Hypothesis
Ενδοτεταρτημοριακό Εύρος	Interquartile range
Εννοιολόγηση	Conceptualization
Εξαρτημένη (μεταβλητή)	Dependent (Variable)
Επαγωγή	Induction
Επικρατούσα τιμή	Mode
Επίπεδο σημαντικότητας	Level of significance
Επιστημολογία	Epistemology
Έρευνες καταγεγραμμένων στατιστικών	Recorded statistics Surveys
Έρευνες καταγεγραμμένων στατιστικών	Recorded statistics Surveys
Έρευνες Περιεχομένου	Content research
Ερωτηματολόγιο	Questionnaire
Εύρος	Range
H	
Ημική οπτική	Emic perspective
Ητική οπτική	Etic perspective
Θ	
Θετικισμός	Positivism
Θηκόγραμμα	Box - Plot
I	
Ιδιογραφική εξήγηση	Idiographic approach
Ιστόγραμμα	Histogram
K	
Κανονική κατανομή	Normal distribution
Κατηγορική (μεταβλητή)	Nominal (Variable)
Κλίση	Slope

Κυκλικό διάγραμμα	Pie Chart
Λ	
Λογιστική παλινδρόμηση	Logistic regression
Λόγος συμπληρωματικών πιθανοτήτων	Odds ratio
Μ	
Μέγιστης πιθανοφάνειας εκτιμητήρια	Maximum likelihood estimator
Μελέτη κοόρτης	Cohort study
Μελέτη πάνελ	Panel study
Μελέτη τάσης	Trend study
Μερική συσχέτιση	Partial Correlation
Μέσος, μέση τιμή	Mean
Μεταβλητές	Variables
Μετά-έλεγχος	Post-testing
Μέτρα διασποράς	Measures of variability
Μέτρα κεντρικής τάσης	Central tendency measures
Μέτρα κύρτωσης	Kurtosis measures
Μέτρα λοξότητας	Measures of skewness
Μη αντιδραστικές μέθοδοι	Unobtrusive research
Μη διατάξιμες μεταβλητές	Nominal Variables
Μη παραμετρικό	Non parametric
Μηδενική υπόθεση	Null Hypothesis
Μονάδες Ανάλυσης	Units of analysis
Μόγλευση	Leverage
Ν	
Νομοθετική εξήγηση	Nomothetic approach
Ξ	
Ξακρισμένη μέση τιμή	Trimmed Mean

Ξυράφι του Όκαμ

Occam's Razor

O

Οικολογικό σφάλμα

Ecological fallacy

Ομάδα ελέγχου

Control Group

Ομάδες Εστίασης

Focus groups

Ομοσκεδαστικότητα

Homoscedacity

Π

Παλινδρόμηση

Regression

Παράγοντας μεγέθυνσης διασποράς

Variance inflation factor - VIF

Παραγωγή

Deduction

Παραμετρικό

Parametric

Παρατηρήσεις υψηλής επίδρασης

Influential points

Πειραματική ομάδα

Experimental Group

Πειραματικοί σχεδιασμοί

Experimental Designs

Περιγραφική στατιστική

Descriptive Statistics

Πιθανοτική δειγματοληψία

Quota sampling

Πίνακα διακυμάνσεων - συνδιακυμάνσεων

Covariance matrix

Πίνακας συχνοτήτων

Frequency table

Πίνακες συνάφειας

Contingency table

Πληθυσμός

Population

Ποιοτικές μεταβλητές

Qualitive Variables

Πολλαπλή παλινδρόμηση

Multiple regression

Πολυσυγραμμικότητα

Multicollinearity

Ποσοστημόρια

Percentiles

Ποσοστό

Percent

Ποσοτικές μεταβλητές

Quantitive Variables

Προέλεγχος

Pre-testing

P

Ραβδόγραμμα Bar Chart

Σ

Στατιστική συμπερασματολογία Inferential Statistics

Στατιστική συνάρτηση Statistic

Στρωματοποιημένη δειγματοληψία Stratified sampling

Στωικοί φιλόσοφοι Stoic philosophers

Συγγραμμικότητα Collinearity

Συνάρτηση σύνδεσης Link function

Συνδιακύμανση Covariance

Συνεχής (μεταβλητή) Scale (Variable)

Συντελεστής ασυμμετρίας Skewness Coefficient

Συντελεστής κύρτωσης Kurtosis Coefficient

Συντελεστής μεταβλητότητας Coefficient of variation

Συντελεστής παραλλακτικότητας Variability coefficient

Συντελεστής προσδιορισμού Coefficient of determination

Συστηματική δειγματοληψία Systematic sampling

Συσχέτιση Correlation

Σφάλμα Error

Σχετική συχνότητα Relative frequency

Tα

Ταίριασμα Matching

Τεταρτημόρια Quartiles

Τυπική απόκλιση Standard Deviation

Τυπική κανονική κατανομή Standard normal distribution

Τυπικό σφάλμα Standard Deviation

Τυπικό σφάλμα Standard error

Τυχαία μεταβλητή	Random variable
Τυχαιοποίηση	Randomizing
Υπόλοιπο	Residual
Φ	
Φαινομενολογία	Phenomenology
Ψ	
Ψευδομεταβλητή	Dummy (Variable)