



Μονάδα Διασφάλισης Ποιότητας Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών

Παραδοτέο 3.2.1

Εγχειρίδιο μεθοδολογίας επεξεργασίας δεδομένων αξιολόγησης

ΠΑΚΕΤΟ ΕΡΓΑΣΙΑΣ 3: Ανάπτυξη μεθοδολογικής υποδομής συστήματος αξιολόγησης

ΥΠΟΕΡΓΟ 1: ΟΡΓΑΝΩΣΗ ΛΕΙΤΟΥΡΓΙΑΣ ΜΟΔΙΠ

Τίτλος Πράξης: Μονάδα Διασφάλισης Ποιότητας ΕΚΠΑ

Άξονας Προτεραιότητας 2: Αναβάθμιση της ποιότητας της εκπαίδευσης και
προώθηση της κοινωνικής ενσωμάτωσης στις 3 περιφέρειες Σταδιακής Εξόδου



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ, ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

1 Πίνακας περιεχομένων

2	Σκοπός.....	4
3	Η προέλευση των δεδομένων.....	5
4	Η δομή η οργάνωση και η τήρηση των δεδομένων	6
5	Το λογισμικό στατιστικής ανάλυσης	7
5.1	Η Γλώσσα R.....	8
6	Η μεθοδολογία της Ανάλυσης	11
6.1	Η εισαγωγή των δεδομένων (Data Entry).....	11
6.2	Προπαρασκευή των δεδομένων	16
6.3	Περιγραφικά στατιστικά μέτρα.....	21
6.3.1	Τα απαιτούμενα στατιστικά μέτρα κατά περίπτωση:.....	26
6.4	Κατασκευή κλιμάκων	27
6.4.1	Πολυμεταβλητές τεχνικές εισαγωγικά.....	27
6.4.2	Factor Analysis	30
6.5	Principal Component Analysis.....	33
6.5.1	Εκτέλεση με το SPSS.....	35
6.6	Έλεγχοι.....	42
6.6.1	Έλεγχος κανονικότητας	42
6.6.2	Συντελεστές συσχέτισης	45
6.6.3	χ^2 έλεγχος ανεξαρτησίας	49
6.6.4	T-test	52
6.6.5	Ανάλυση Διακύμανσης.....	56
7	Κατασκευή Διαγραμμάτων έλεγχου/Control Charts/Shewart Charts.....	59
8	Κατασκευή Διαγράμματος Pareto.....	63
9	Βιβλιογραφία	66
10	Βιβλιογραφία	66

2 Σκοπός

Ο σκοπός του εγχειριδίου αυτού είναι να περιγράψει το σύνολο της απαιτούμενης μεθοδολογικής υποδομής που απαιτείται, για την εφαρμογή στατιστικών μεθόδων στην αξιολόγηση και την διασφάλιση ποιότητας του ΕΚΠΑ.

Συγκεκριμένα στο εγχειρίδιο αυτό περιγράφονται όλα τα στάδια που λαμβάνουν χώρα από την συλλογή των δεδομένων την οργάνωσή τους, την καταχώρησή τους, την ανάλυσή τους με διάφορες μεθόδους, την παρουσίαση των αποτελεσμάτων και την συμπερασματολογία.

3 Η προέλευση των δεδομένων

Τα δεδομένα τα οποία των οποίων η επεξεργασία περιγράφεται μπορεί να είναι:

- Αριθμητικά δεδομένα τα οποία αφορούν στοιχεία φοιτητών
- Αριθμητικά δεδομένα που αφορούν την λειτουργία των τμημάτων με οικονομικά στοιχεία
- Αριθμητικά δεδομένα που αφορούν την λειτουργία των τμημάτων με λειτουργικά στοιχεία
- Αριθμητικά και αλφαριθμητικά δεδομένα που έχουν να κάνουν με την έρευνα του προσωπικού του Πανεπιστημίου
- Αλφαριθμητικά δεδομένα τα οποία προέρχονται από αξιολογήσεις
- Δεδομένα τα οποία προέρχονται από κάθε λογής ερωτηματολόγια (ικανοποίησης φοιτητών, αξιολογήσεις κλπ)
- Κάθε άλλη πηγή δεδομένων που δύναται να προσφέρει δεδομένα προς εξαγωγή χρήσιμης πληροφορίας

4 Η δομή η οργάνωση και η τήρηση των δεδομένων

Τα δεδομένα τα οποία συλλέγονται από κάθε μορφής πηγή θα πρέπει να καταχωρούνται απαραίτητα και σε ηλεκτρονική μορφή. Αυτό είναι απαραίτητο όχι μόνο για την περαιτέρω επεξεργασία τους αλλά και για την ευελιξία στην διακίνηση και την φύλαξή τους.

Τα αλφαριθμητικά δεδομένα εκ των πραγμάτων πρέπει να τηρούνται σε μια ιεραρχική βάση δεδομένων με ιεράρχηση η οποία θα προκύπτει:

- από την πηγή προέλευσή τους
- το περιεχόμενό τους
- την ημερομηνία δημιουργίας τους

Για τις ιεραρχικές βάσεις δεδομένων προτείνεται η δημιουργία ηλεκτρονικού ευρετηρίου με υπερσύνδεση (ή εναλλακτικά με id) στο αρχείο δεδομένων.

Τα αριθμητικά δεδομένα μπορούν να οργανωθούν είτε με την παραπάνω μορφή είτε με την μορφή σχεσιακών βάσεων δεδομένων με κατάλληλα κλειδιά μεταξύ των πινάκων.

Η μορφή τήρησής τους σχετίζεται τόσο με το περιεχόμενο τους όσο και με την αρχική τους δομή.

Η διατήρηση των αριθμητικών δεδομένων σε ηλεκτρονική μορφή προτείνεται να γίνεται σε μορφές αρχείων οι οποίες να είναι συμβατές με τα περισσότερα λογισμικά ανάγνωσής και επεξεργασίας των δεδομένων αυτών. Ενδεικτικά προτείνονται οι μορφές: csv,tsv,dat,txt,xls,xlsx. Εξειδικευμένες μορφές όπως: *.sav, *.r κλπ δεν προτείνονται.

5 Το λογισμικό στατιστικής ανάλυσης

Τα λογισμικά στατιστικής ανάλυσης τα οποία προτείνονται είναι δυο κατευθύνσεων:

- Στην πρώτη κατεύθυνση εντάσσονται τα στατιστικά πακέτα. Ένα τέτοιο πακέτο θα πρέπει να είναι εύχρηστο, ενδεχομένως και ευρύτατα γνωστό ώστε να μπορεί να υποστηριχθεί από αρκετούς δυνητικούς χρήστες. Ένα τέτοιο πακέτο είναι στατιστικό πακέτο IBM/SPSS με προτεινόμενες εκδόσεις τις v.17 ή v.20.

- Στην δεύτερη κατεύθυνση λογισμικών εντάσσονται λογισμικά στατιστικής ανάλυσης τα οποία είναι ποιο εξειδικευμένα και με περισσότερες δυνατότητες από το SPSS και με χαρακτηριστικά τα οποία απαντούν στις ιδιαιτερότητες των εφαρμογών που απαιτεί η διασφάλιση ποιότητας σε ένα τόσο μεγάλο και πολύπλοκο οργανισμό όπως το Πανεπιστήμιο Αθηνών.

Το κοινό χαρακτηριστικό των γνωστών εμπορικών λογισμικών είναι ότι είναι «κλειστής μορφής» (closed form - compiled) λογισμικά τα οποία διαχειρίζονται από κάποιο μενού και προσφέρουν μια εκτέλεση της εκάστοτε μεθόδου σε κάθε κλήση του λογισμικού. Όλα τα λογισμικά αυτά είναι ακατάλληλα για την περίπτωση όπου απαιτούνται πολλαπλές εκτελέσεις με δυναμικά μεταβαλλόμενες παραμέτρους σε κάθε εκτέλεση στα πλαίσια αλγορίθμων που έχουν αναπτυχθεί. Για την υλοποίηση των ποιο περίπλοκων αλγορίθμων που περιγράφηκαν απαιτείται η υλοποίηση κώδικα στα πλαίσια κάποιου προγραμματιστικού περιβάλλοντος όπου η κλήση της εκάστοτε μεθόδου θα υπάρχει σαν δυνατότητα ,με την μορφή κάποιας συνάρτησης , ταυτόχρονα με όλο το υπόλοιπο πλήθος των εντολών του προγραμματιστικού περιβάλλοντος.

Έτσι θα δίνεται η δυνατότητα για συγγραφή προγραμμάτων, τα οποία με βρόχους θα εκτελούν πολλαπλές κλήσεις της

επιθυμητής μεθόδου, ταυτόχρονα δε θα μπορούν να χειρίζονται τις απαιτήσεις που προκύπτουν από την στοχαστική φύση των μεταβλητών αυτών. Το αν θα μπορεί το λογισμικό αυτό να δίνει σε μορφή εκτελέσιμου αρχείου (*.exe) το εκάστοτε πρόγραμμα (δηλαδή αν θα είναι compiler) ή απλά θα το τρέχει μεταφράζοντάς το από κάποια γραμμή εντολών (δηλαδή αν θα είναι interpreter) είναι κάτι αδιάφορο στο τρέχων επίπεδο.

Μια ακόμα σημαντική παράμετρος είναι η πολυμορφία και η πολυπλοκότητα των δεδομένων: Τα δεδομένα πολλές φορές συνιστώνται από μια πληθώρα πινάκων (π.χ. αξιολόγηση μαθημάτων από φοιτητές) αλλά και μια πληθώρα τύπων δεδομένων.

Με αυτές λοιπόν τις προδιαγραφές είναι ελάχιστα τα λογισμικά τα οποία είναι κατάλληλα και ουσιαστικά πρόκειται για γλώσσες προγραμματισμού. Επειδή μάλιστα δεν είναι επιθυμητή η δυνατότητα του "compilation" ,με τα παραπάνω φωτογραφίζεται η ανάγκη και κάποια γλώσσα σε μορφή interpreter . Επιπλέον επειδή υπάρχει η ανάγκη η γλώσσα αυτή να έχει και σαφή στατιστικό προσανατολισμό τότε οι επιλογές γίνονται ακόμα λιγότερες. Οι διαθέσιμες επιλογές είναι οι γλώσσες : Matlab, S-Plus , R , Mathematica και λοιπά ,στατιστικά ή μη πακέτα. Λαμβάνοντας υπόψη και την ευρύτητα της εφαρμογής αλλά και την ύπαρξη διεθνούς κοινότητας η οποία «γράφει» στην γλώσσα αυτή (και σε θεματολογία διασφάλισης ποιότητας), προτείνεται η γλώσσα R.

Παρακάτω θα γίνει αναφορά στην γλώσσα R. Για το πακέτο SPSS δεν θα γίνει αναφορά δεδομένου ότι θα δειχθεί λεπτομερέστερα αργότερα πως εκτελούνται οι επιθυμητοί έλεγχοι με το πακέτο αυτό.

5.1 Η Γλώσσα R

Η γλώσσα R (The R Project for Statistical Computing 1996) R είναι μια γλώσσα και το σχετικό περιβάλλον, για στατιστικούς υπολογισμούς και γραφικά. Είναι ένα έργο GNU(ελεύθερη χρήση) η οποία είναι παρόμοια με την γλώσσα S, τη γλώσσα και το περιβάλλον που αναπτύχθηκε στα Bell Laboratories (πρώην AT & T, Lucent Technologies τώρα) από τον John Chambers και την ομάδα του. Η R μπορεί να θεωρηθεί ως μια διαφορετική εκτέλεση της S. Υπάρχουν κάποιες σημαντικές διαφορές, αλλά πολλοί κώδικες γραμμένοι για S εκτελούνται σχεδόν αμετάβλητοι από την R.

Η R είναι διαθέσιμη ως ελεύθερο λογισμικό, σύμφωνα με το Ίδρυμα Ελεύθερου Λογισμικού της GNU General Public License σε μορφή πηγαίου κώδικα. Τρέχει σε μια μεγάλη ποικιλία από πλατφόρμες UNIX και παρόμοια συστήματα (συμπεριλαμβανομένων των FreeBSD και Linux), Windows και MacOS.

Η γλώσσα R εμφανίζει μια σειρά από πλεονεκτήματα όπως:

- Είναι υλοποιημένη σε κώδικα γλώσσας C με αποτέλεσμα να είναι πολύ γρήγορη ειδικά σε επαναληπτικούς βρόγχους.
- Διαχειρίζεται την μνήμη του εκάστοτε υπολογιστικού συστήματος και θέτει η ίδια ως διαθέσιμη εκμεταλλευόμενη μνήμη Ram το 50% της διαθέσιμης μνήμης. Αυτό προστατεύει το υπολογιστικό σύστημα από προβλήματα σε περιπτώσεις ελαττωματικού κώδικα ή υπερχειλίσης μνήμης κλπ
- Παρέχει ένα ευέλικτο περιβάλλον εργασίας με πλήθος από έτοιμες συναρτήσεις
- Είναι εύκολη στην εγκατάσταση και δεν καταναλώνει σημαντικά πόρους του συστήματος (σε αντίθεση π.χ. με το Matlab)
- Σήμερα ένα τεράστιο πλήθος ερευνητών ασχολείται με την R αποτελώντας μια τεράστια κοινότητα στα πλαίσια της οποίας δημοσιεύονται πακέτα λογισμικού για διάφορα προβλήματα που ενσωματώνονται στην R επεκτείνοντας έτσι τις δυνατότητές της .Αυτό είναι τεράστιας σημασίας πλεονέκτημα.

- Είναι εφικτό, για προχωρημένου επιπέδου χρήστες, να γραφεί κώδικας σε γλώσσα R και αυτός να ενσωματωθεί εκ των υστέρων σε κάποιο λογισμικό που είναι γραμμένο σε γλώσσα C (αλλά και το αντίστροφο) και να τρέχει κανονικά σαν μεταγλωττισμένο εκτελέσιμο αρχείο.

Η γλώσσα R είναι σήμερα η *de facto* γλώσσα για την στατιστική έρευνα αλλά και για τις εφαρμογές.

Περισσότερες πληροφορίες για τον τρόπο κλήσης του πακέτου μπορούν να βρεθούν είτε στις παρακάτω παραγράφους είτε στην ιστοσελίδα της γλώσσας R.

6 Η μεθοδολογία της Ανάλυσης

6.1 Η εισαγωγή των δεδομένων (Data Entry)

Η εισαγωγή δεδομένων στο SPSS γίνεται από το παράθυρο Data Editor. Το παράθυρο Data Editor του SPSS είναι αυτό που φαίνεται στην παρακάτω εικόνα.

The screenshot shows the SPSS Statistics Data Editor window with the following data table:

ID	Programma_proptyxiakon_spondon	Sxoli	Synolikos_arithmos_neo eisaxthenton	Eisaxthentes_meisagogikes_etaseis	Eisaxthentes_meisagogikes_etaseis	Eisaxthentes_n_katataktiries_etaseis
1	20 Τμήμα Μαθηματικών	1	392	261	110	
2	23 Τμήμα Πληροφορικής & Τηλεπικοινωνιών	1	294	173	93	
3	18 Τμήμα Φυσικής	1	224	144	60	
4	19 Τμήμα Χημείας	1	180	88	80	
5	21 Τμήμα: Βιολογίας	1	143	100	24	
6	22 Τμήμα Γεωλογίας & Γεωπεριβάλλοντος	1	119	85	27	
7	7 Τμήμα Φιλολογίας	2	517	326	142	
8	9 Τμήμα Φιλοσοφίας, Παιδαγωγικής και Ψυχολογίας	2	452	323	123	
9	8 Τμήμα Ιστορίας & Αρχαιολογίας	2	386	213	149	
10	11 Τμήμα Αγγλικής Γλώσσας και Φιλολογίας	2	336	281	6	
11	12 Τμήμα Γαλλικής Φιλολογίας	2	232	219	0	
12	13 Τμήμα Γερμανικής Γλώσσας και Φιλολογίας	2	131	122	1	
13	10 Πρόγραμμα Ψυχολογίας Τμήμα Φ.Π.Ψ.	2	122	111	3	
14	14 Τμήμα Ιταλικής & Ισπανικής Γλώσσας και Φιλολογίας	2	114	85	0	
15	15 Τμήμα Θεατρικών Σπουδών	2	107	61	27	
16	16 Τμήμα Μουσικών Σπουδών	2	73	56	0	
17	17 Τμήμα Τουρκικών & Σύγχρονων Ασιατικών Σπουδών	2	69	57	0	
18	1 Τμήμα Θεολογίας	3	417	377	0	
19	2 Τμήμα Κοινωνικής Θεολογίας	3	.	.	.	
20	6 Τμήμα Ιατρικής	4	305	172	61	
21	26 Τμήμα Νοσηλευτικής	4	231	166	41	
22	25 Τμήμα Φαρμακευτικής	4	161	111	1	
23	24 Τμήμα Οδοντιατρικής	4	127	89	3	
24	3 Τμήμα Νομικής	5	686	404	165	
25	5 Τμήμα Πολιτικής Επιστήμης & Δημόσιας Διοίκησης	5	394	273	84	
26	4 Τμήμα Οικονομικών	5	368	321	10	
27	30 Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης	6	451	202	232	

Η γραμμή τίτλου είναι η μπλε γραμμή που φαίνεται στο πάνω μέρος του παραθύρου. Το μενού επιλογών είναι παρόμοιο με αυτό που συναντάται στο MS Office. Είναι η σειρά που φαίνεται κάτω από τη γραμμή τίτλου και περιλαμβάνει τις εξής επιλογές του παραθύρου: **File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window** και **Help**. Η γραμμή εργαλείων βρίσκεται κάτω από το μενού επιλογών και αποτελείται από εικονίδια χρήσιμα για λειτουργίες που χρησιμοποιούνται συχνά, όπως αποθήκευση, εκτύπωση, άνοιγμα κάποιου αρχείου.

Οι γραμμές κύλισης βρίσκονται στα δεξιά και στο κάτω μέρος του παραθύρου και μας βοηθάνε να μετακινηθούμε πάνω-κάτω και δεξιά-αριστερά. Στο κάτω μέρος του παραθύρου εμφανίζεται ένα μήνυμα που λέει **SPSS Processor is ready**. Η γραμμή αυτή στην οποία εμφανίζεται αυτό το μήνυμα είναι η γραμμή κατάστασης. Όταν το SPSS διεξάγει κάποιον υπολογισμό, έχει μία διεργασία σε εξέλιξη, ή τερματίσει μία οποιαδήποτε διεργασία θα εμφανίζεται το αντίστοιχο μήνυμα. Στο πάνω μέρος των κελιών εμφανίζεται το όνομα των κελιών. Από τη στιγμή που δεν έχουμε δώσει ονόματα στα κελιά απλά εμφανίζετε η λέξη **var**.

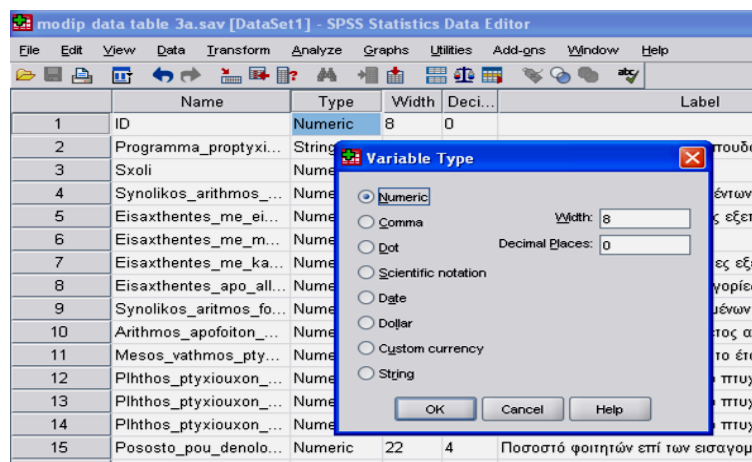
Επίσης στο αριστερό κάτω μέρος του παραθύρου υπάρχουν δύο επιλογές. Η μία είναι η **Data View** και η άλλη είναι η **Variable View**. Αυτή τη στιγμή είναι επιλεγμένη η πρώτη επιλογή. Δηλαδή το παράθυρο που μπορούμε να περάσουμε δεδομένα είναι αυτό που φαίνεται στην οθόνη. Τα δεδομένα τα περνάμε κατά τον ίδιο τρόπο με το Excel, δηλαδή κάθετα αν πρόκειται για τιμές της ίδιας μεταβλητής. Πληκτρολογούμε τον αριθμό και πατάμε **Enter** ή **Tab**.

Αν επιλέξουμε τη δεύτερη επιλογή θα εμφανιστεί το παράθυρο της εικόνας:

	Name	Type	Width	Deci...	Label	Values	Missing	Columns	Align	Measure
1	ID	Numeric	8	0		None	None	8	Right	Scale
2	Programma_proptyxi...	String	55	0	Πρόγραμμα Προπτυχιακών Σπουδών	None	None	50	Left	Scale
3	Sxoli	Numeric	8	0	Σχολή	{1, Θετικών ...	None	8	Right	Scale
4	Synolikos_arithmos_...	Numeric	11	0	Συνολικός Αριθμός νεοεισαχθέντων	None	None	11	Right	Scale
5	Eisaxthentes_me_ei...	Numeric	11	0	Εισαχθέντες με : εισαγωγικές εξετάσεις	None	None	11	Right	Scale
6	Eisaxthentes_me_m...	Numeric	11	0	Εισαχθέντες με : μεταγγραφές	None	None	11	Right	Scale
7	Eisaxthentes_me_ka...	Numeric	11	0	Εισαχθέντες με : κατατακτήριες εξετάσεις	None	None	11	Right	Scale
8	Eisaxthentes_apo_all...	Numeric	11	0	Εισαχθέντες από άλλες κατηγορίες	None	None	11	Right	Scale
9	Synolikos_arithmos_fo...	Numeric	11	0	Συνολικός αριθμός εγγεγραμμένων φοιτητών σε όλα ...	None	None	11	Right	Scale
10	Arithmos_apofoiton_...	Numeric	22	2	Αριθμός αποφοίτων κατά το έτος αναφοράς	None	None	22	Right	Scale
11	Mesos_vathmos_pty...	Numeric	22	2	Μέσος βαθμός πτυχίου κατά το έτος αναφοράς	None	None	22	Right	Scale
12	Pilthos_ptyxiouxon_...	Numeric	11	0	Πλήθος πτυχιούχων με βαθμό πτυχίου: 5.0-6.49	None	None	11	Right	Scale
13	Pilthos_ptyxiouxon_...	Numeric	11	0	Πλήθος πτυχιούχων με βαθμό πτυχίου:6.5-8.5	None	None	11	Right	Scale
14	Pilthos_ptyxiouxon_...	Numeric	11	0	Πλήθος πτυχιούχων με βαθμό πτυχίου:8,51-10	None	None	11	Right	Scale
15	Pososto_pou_denolo...	Numeric	22	4	Ποσοστό φοιτητών επί των εισαγομένων που ΔΕΝ ολ...	None	None	22	Right	Scale
16	Programma_spoudon	Numeric	8	0	Είδος Προγράμματος Σπουδών	{1, Προπτυχ...	None	8	Right	Ordinal

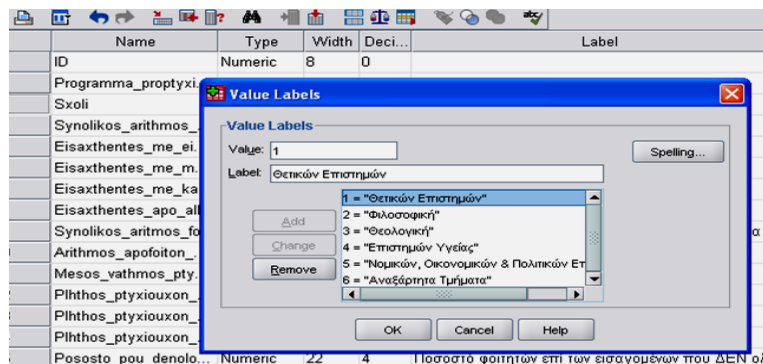
Η πρώτη στήλη έχει τίτλο **Name**. Στα κελιά της πρώτης στήλης δίνουμε τα ονόματα των στηλών των δεδομένων που βρίσκονται στο Data Editor. Έτσι, στο πρώτο κελί αντιστοιχεί το όνομα της πρώτης στήλης των δεδομένων, στο δεύτερο, αντιστοιχεί το όνομα της δεύτερης στήλης και ούτω καθεξής.

Αν επιλέξουμε τη δεύτερη στήλη (**Type**) τότε θα εμφανιστεί το παράθυρο της εικόνας :



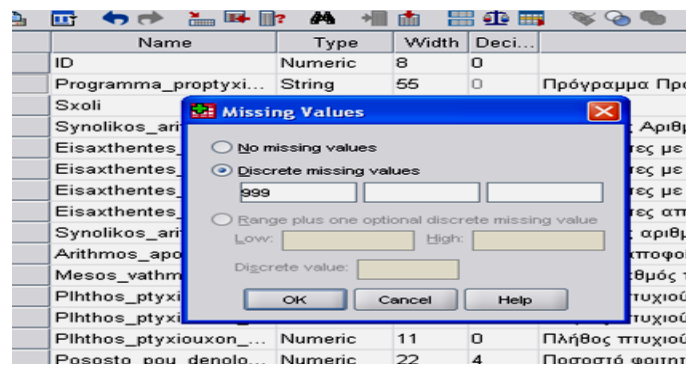
Αυτό μας δίνει τη δυνατότητα να επιλέξουμε τον τύπο των δεδομένων για κάθε στήλη του Data Editor. Ξανά το πρώτο κελί αντιστοιχεί στα δεδομένα της πρώτης στήλης του Data Editor, το δεύτερο κελί στα δεδομένα της δεύτερης στήλης και ούτω καθεξής. Η επιλογή **Numeric** είναι προεπιλεγμένη από το πακέτο διότι τα δεδομένα μας είναι τις πιο πολλές φορές αριθμητικά. Αν επιλέξουμε **String** τότε τα δεδομένα μας θα είναι σε μορφή χαρακτήρα ή απλά θα είναι γράμματα. Με το **Width** ορίζουμε το μέγιστο πλήθος των ψηφίων που θα έχουν τα αριθμητικά δεδομένα και με το **Decimal Places** ορίζουμε το πλήθος των ψηφίων που θα βρίσκονται δεξιά της υποδιαστολής. Το τρίτο και το τέταρτο κελί του παραθύρου της εικόνας αναφέρονται στο πλήθος των ψηφίων αριστερά και δεξιά της υποδιαστολής, όπως ήδη αναφέραμε. Το κελί με τίτλο **Label** είναι η ετικέτα των δεδομένων. Αν δηλώσουμε ονόματα στα δεδομένα μας στους πίνακες που θα εμφανίζονται με τα αποτελέσματα από το SPSS θα εμφανίζονται οι στήλες των δεδομένων με τα ονόματα τους. Αν όμως δηλώσουμε και ετικέτες ή μόνο ετικέτες στις στήλες των δεδομένων θα εμφανίζονται τα αποτελέσματα όπου κάθε στήλη δεδομένων θα έχει για όνομα αυτό που έχουμε ορίσει στις ετικέτες των στηλών. Το τι θα εμφανίζεται μπορεί να επιλεγεί από το χρήστη από την υποεπιλογή **Options** που βρίσκεται μέσα στην επιλογή **Edit**.

Το κελί **Values** οδηγεί το παράθυρο της εικόνας:



Έστω ότι έχουμε ποιοτικά δεδομένα και θέλουμε να τα περάσουμε στο SPSS. Για να γίνει αυτό κωδικοποιούμε εξαρχής τα δεδομένα δίνοντας τους τιμές για κάθε μία κατηγορία.

Στο κελί **Missing** θα ορίσουμε τις χαμένες παρατηρήσεις. Για παράδειγμα κάποιοι εκ των ερωτηθέντων σε ένα ερωτηματολόγιο δεν έχουν απαντήσει σε όλες τις ερωτήσεις. Το παράθυρο που θα εμφανιστεί είναι αυτό που βρίσκεται στην εικόνα :

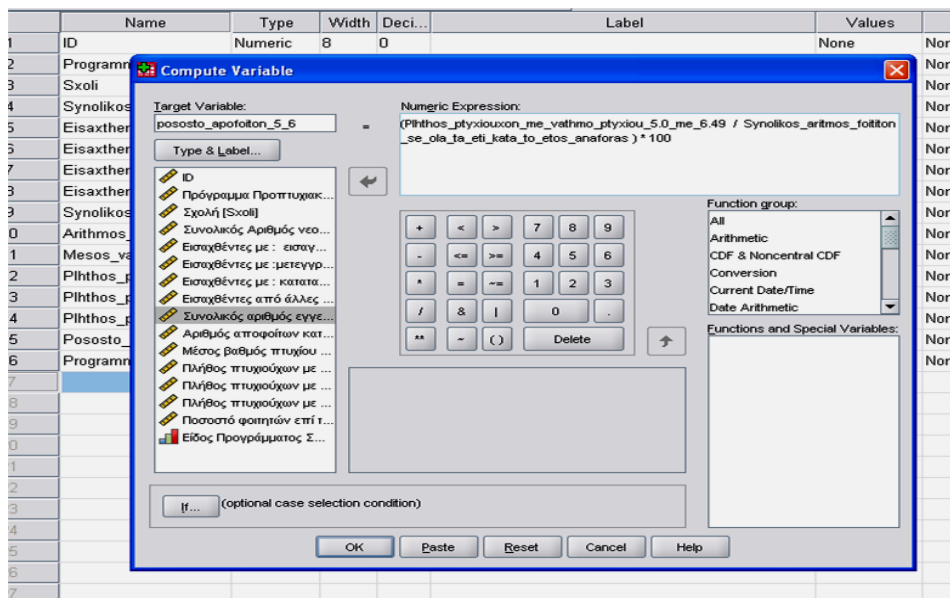


Πρέπει να προσέξουμε πως θα ορίσουμε τις χαμένες τιμές. Για παράδειγμα σε ένα ερωτηματολόγιο που οι απαντήσεις είναι σε κλίμακα Likert από 1 έως 5 τις στις χαμένες τιμές θα δώσουμε έναν αριθμό που δε βρίσκεται μεταξύ 1 και 5. Θα πρέπει να είναι δηλαδή ένας αριθμός που δε συναντάται στα δεδομένα της κάθε στήλης ξεχωριστά. Το κελιά **Columns** και **Align** αναφέρονται στο μέγεθος της στήλης και στη στοίχιση των δεδομένων στην κάθε στήλη. Τέλος, το τελευταίο κελί αναφέρεται στον τύπο των δεδομένων. Αν τα δεδομένα αφορούν σε ποσοτικές μετρήσεις (**Scale**), διατεταγμένες (**Ordinal**) ή ονομαστικές (**Nominal**).

6.2 Προπαρασκευή των δεδομένων

Συχνά χρειάζεται να μετασχηματίσουμε τα δεδομένα ή να δημιουργήσουμε μια νέα μεταβλητή βασιζόμενοι κάποιες υπάρχουσες μεταβλητές

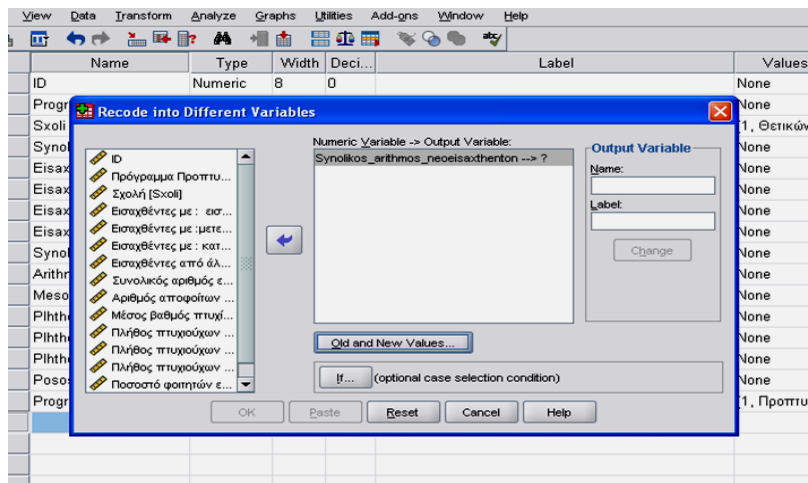
Μία πολύ χρήσιμη επιλογή από το μενού επιλογών είναι αυτή της μετατροπής των δεδομένων (**Transform**). Η πρώτη εντολή που εμπεριέχεται στην επιλογή **Transform** είναι η **Compute Variable**. Επιλέγοντας αυτήν την εντολή θα εμφανιστεί το παράθυρο της εικόνας :



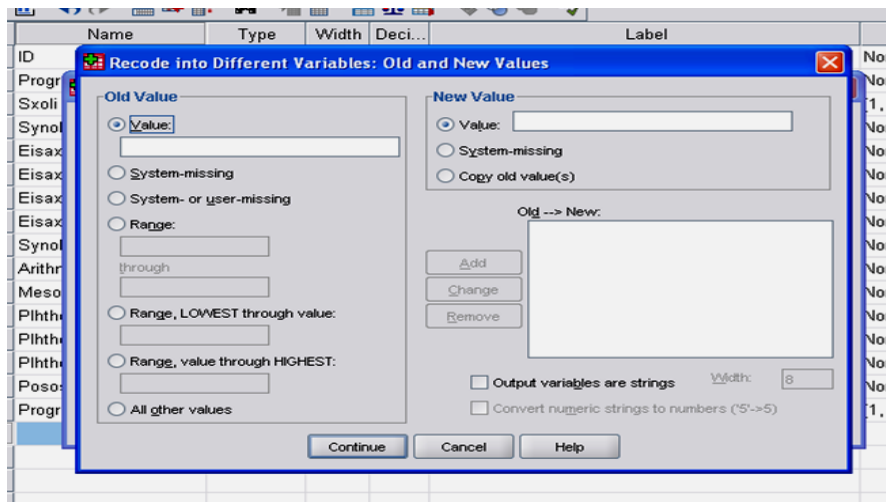
Το λευκό κουτάκι που λέγεται **Target Variable** πρέπει να συμπληρωθεί με ένα όνομα. Εκεί θα αποθηκευτεί η μετασχηματισμένη στήλη δεδομένων. Η μετασχηματισμένη στήλη μπορεί είτε να αποθηκευτεί στην ίδια στήλη είτε σε διαφορετική. Περνώντας τις στήλες από το αριστερό κουτάκι στο κουτάκι που λέγεται **Numeric Expression** ορίζουμε τις στήλες οι οποίες θα μετασχηματιστούν. Το κουτάκι **Function group** περιέχει διάφορα είδη συναρτήσεων όπως μαθηματικές, στατιστικές, μετατροπής και άλλες. Για κάθε είδος συναρτήσεων που επιλέγουμε, στο κουτάκι που βρίσκεται ακριβώς από κάτω βλέπουμε τις διαθέσιμες

συναρτήσεις. Αυτές είναι συναρτήσεις που μας βοηθάνε στο μετασχηματισμό των δεδομένων. Βέβαια, μπορούμε να γράψουμε μία δική μας συνάρτηση μετατροπής η οποία δε βρίσκεται στη λίστα με τις ήδη υπάρχουσες συναρτήσεις.

Η εντολή επανακωδικοποίησης είναι εντολή κωδικοποίησης των ήδη υπάρχουσών στηλών δεδομένων. Ένα παράδειγμα χρησιμοποίησης αυτής της εντολής είναι το εξής: έστω ότι έχουμε συλλέξει ηλικίες ατόμων από 20 έως 90+ έτη. Αντί να δουλεύουμε με τις ηλικίες αυτές καθαυτές θέλουμε να τις κατηγοριοποιήσουμε σε ομάδες ηλικιών έστω 7 τον αριθμό. Μπορούμε να επιλέξουμε είτε να σώσουμε τις ομάδες ηλικιών στη στήλη των ήδη υπάρχουσών ηλικιών (οπότε θα χαθούν οι ηλικίες), είτε σε μία άλλη στήλη. Θα επιλέξουμε δηλαδή είτε **Recode into Same Variables**, είτε **Recode into Different Variables** αντίστοιχα. Αν επιλέξουμε να σώσουμε τη νέα στήλη των ομάδων ηλικιών στην ίδια στήλη των ηλικιών, διαγράφοντας ουσιαστικά τις ηλικίες θα εμφανιστεί το παράθυρο της εικόνας:



Και στις δύο περιπτώσεις θα πρέπει να περάσουμε τη στήλη των δεδομένων που θέλουμε να μετασχηματίσουμε από το αριστερό



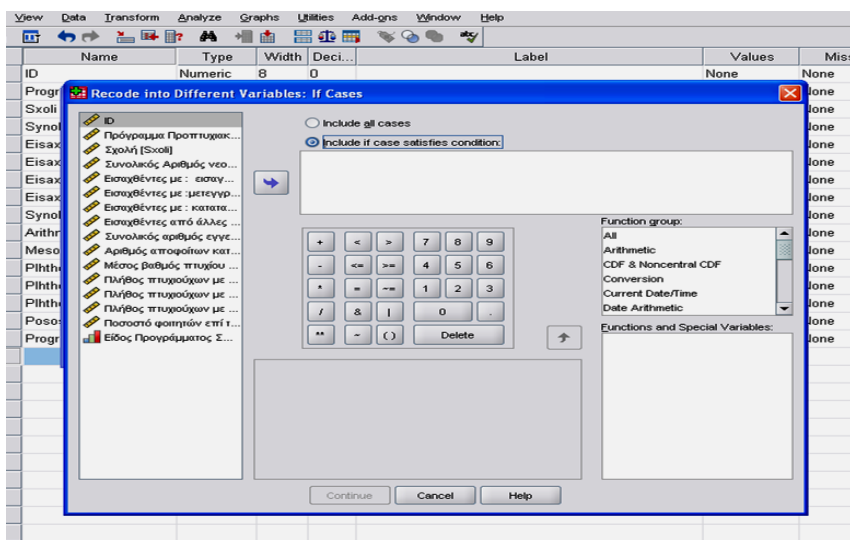
στο δεξιό λευκό κουτάκι. Μόλις το κάνουμε αυτό θα ενεργοποιηθεί η επιλογή **Old and New Values...** που βρίσκεται κάτω από το δεξιό κουτάκι. Επιλέγοντας αυτήν την επιλογή εμφανίζεται το παράθυρο της εικόνας :

Στην περιοχή **Old Value** που βρίσκεται στο αριστερό μέρος του παραθύρου θα επιλέξουμε να πληκτρολογήσουμε τις τιμές των δεδομένων που θα μετασχηματίσουμε. Στο παράδειγμα με τις ηλικιακές ομάδες θα επιλέξουμε το **Range** και θα πληκτρολογήσουμε στα δύο λευκά κουτάκια που θα ενεργοποιηθούν το εύρος των τιμών. Με αυτόν τον τρόπο δηλώνουμε το εύρος των τιμών που θέλουμε να μετασχηματίσουμε. Υπάρχουν άλλες δύο επιλογές που μπορούμε να ορίσουμε εύρη ή διαστήματα τιμών. Είτε από τη χαμηλότερη τιμή έως κάποια τιμή, είτε από κάποια τιμή έως την υψηλότερη. Στη συνέχεια πηγαίνουμε στο δεξιό μέρος του παραθύρου στην περιοχή **New Value**. Στο λευκό κουτάκι που βρίσκεται δεξιά της προεπιλογής **Value** θα πληκτρολογήσουμε τη νέα τιμή για τη συγκεκριμένη ηλικιακή ομάδα. Αφού ξεκινήσαμε με την πρώτη ηλικιακή ομάδα θα βάλουμε τον αριθμό 1. Εν συνέχεια

θα πατήσουμε το κουτάκι **Add** για να καταχωρηθεί η αλλαγή στο SPSS.

Μόλις το κάνουμε αυτό θα εμφανιστεί στο μεγάλο λευκό κουτάκι η καταχωρημένη αλλαγή. Συνεχίζουμε κατά τον ίδιο τρόπο και για τις υπόλοιπες ηλικιακές ομάδες. Αν έχουμε επιλέξει οι νέες κωδικοποιημένες τιμές να αποθηκευτούν στη στήλη με τις ήδη υπάρχουσες τιμές, δε χρειάζεται να κάνουμε τίποτα άλλο.

Στη συνέχεια θα πατήσουμε **Change** ώστε να γίνει η αλλαγή στο όνομα. Αφού τελειώσουμε θα εμφανιστεί στο SPSS Data Editor μία νέα στήλη που θα περιέχει τις κωδικοποιημένες τιμές της αρχικής στήλης. Και στις δύο περιπτώσεις το SPSS αντιστοιχεί στις ήδη υπάρχουσες τιμές τις νέες τιμές ανάλογα με το διάστημα στο οποίο βρίσκονται. Για τις ηλικίες δηλαδή από 20 έως και 29 θα αντιστοιχήσει την τιμή 1 (η ηλικία 30 θα συμπεριληφθεί στη δεύτερη ομάδα, όπως και κάθε άνω άκρο των κλάσεων ή ομάδων). Για τις ηλικίες από 30 έως 39 θα αντιστοιχήσει την τιμή 2. Καλό θα είναι σε αυτό το σημείο να επιλέξουμε στο Data Editor να εμφανίσει το παράθυρο Variable View, για να καθορίσουμε την κάθε κωδικοποιημένη τιμή που ανήσυχη. Στο παράδειγμα με τις ηλικίες επιλέγοντας **Values** και με τη διαδικασία η οποία έχει ήδη περιγραφεί να ορίσουμε την κάθε τιμή σε ποια ηλικιακή ομάδα αντιστοιχεί. Ανεξαρτήτου επιλογής αποθήκευσης της μετασχηματισμένης στήλης δεδομένων, αν πατήσουμε την επιλογή **If** θα εμφανιστεί το παράθυρο της εικόνας :



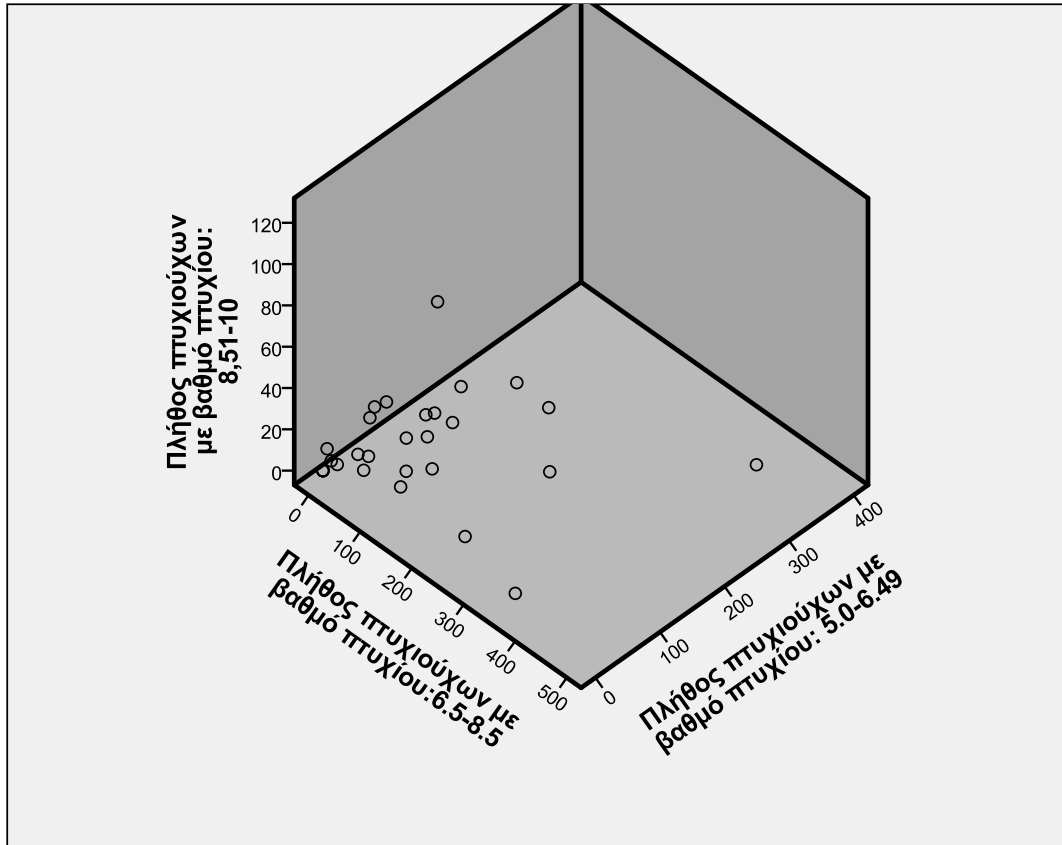
Με αυτήν την επιλογή κωδικοποιούμε μόνο τις τιμές που ικανοποιούν κάποια λογική συνθήκη. Εδώ όπως και στο παράθυρο της εικόνας 10 πρώτα επιλέγουμε τη συνάρτηση από τη λίστα των συναρτήσεων, την ανεβάζουμε πάνω με το βελάκι και μετά περνάμε τη στήλη με την οποία θα δουλέψουμε δεξιά.

6.3 Περιγραφικά στατιστικά μέτρα

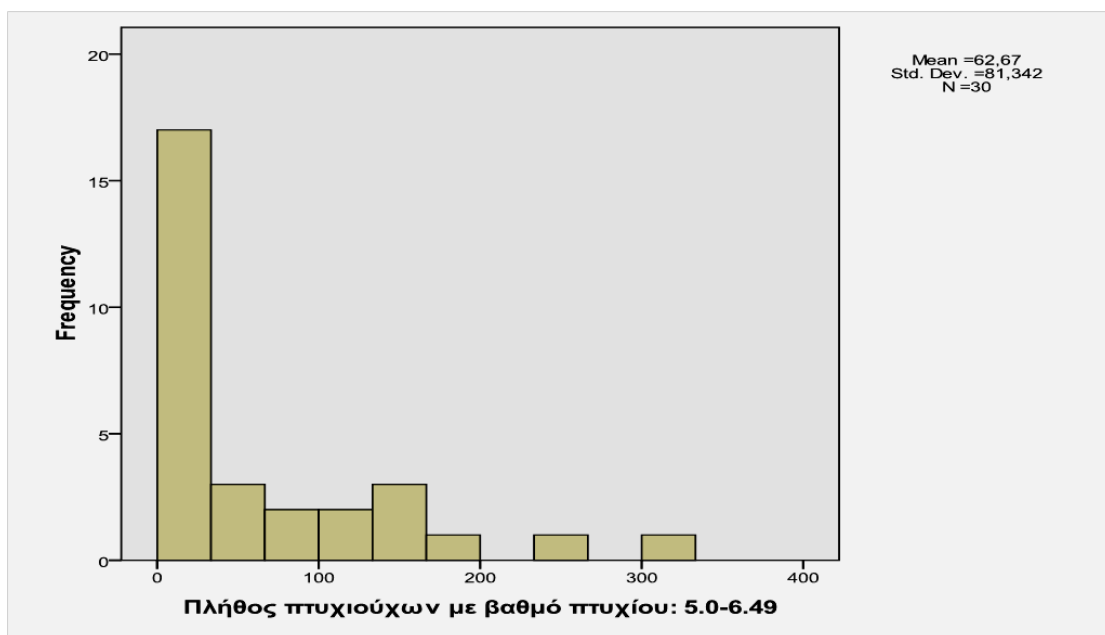
Στην αμέσως επόμενη παράγραφο πρόκειται να γίνει παρουσίαση και εφαρμογή διαφόρων τεχνικών και αλγορίθμων και εξαγωγή των σχετικών αποτελεσμάτων. Τα αποτελέσματα από αυτές τις τεχνικές, στις περισσότερες περιπτώσεις, είναι τεράστια σύνολα από δεδομένα κατάλληλα για επεξεργασία. Πάνω στα αποτελέσματα αυτά θα πρέπει να γίνει στην συνέχεια στατιστική αναφορά και συμπερασματολογία. Για τον λόγο αυτό κρίνεται σκόπιμο να έχει γίνει αναφορά νωρίτερα σε όλο το φάσμα των στατιστικών ελέγχων που είναι διαθέσιμοι και το είδος της συμπερασματολογίας που μπορεί να οικοδομηθεί από αυτούς τους ελέγχους.

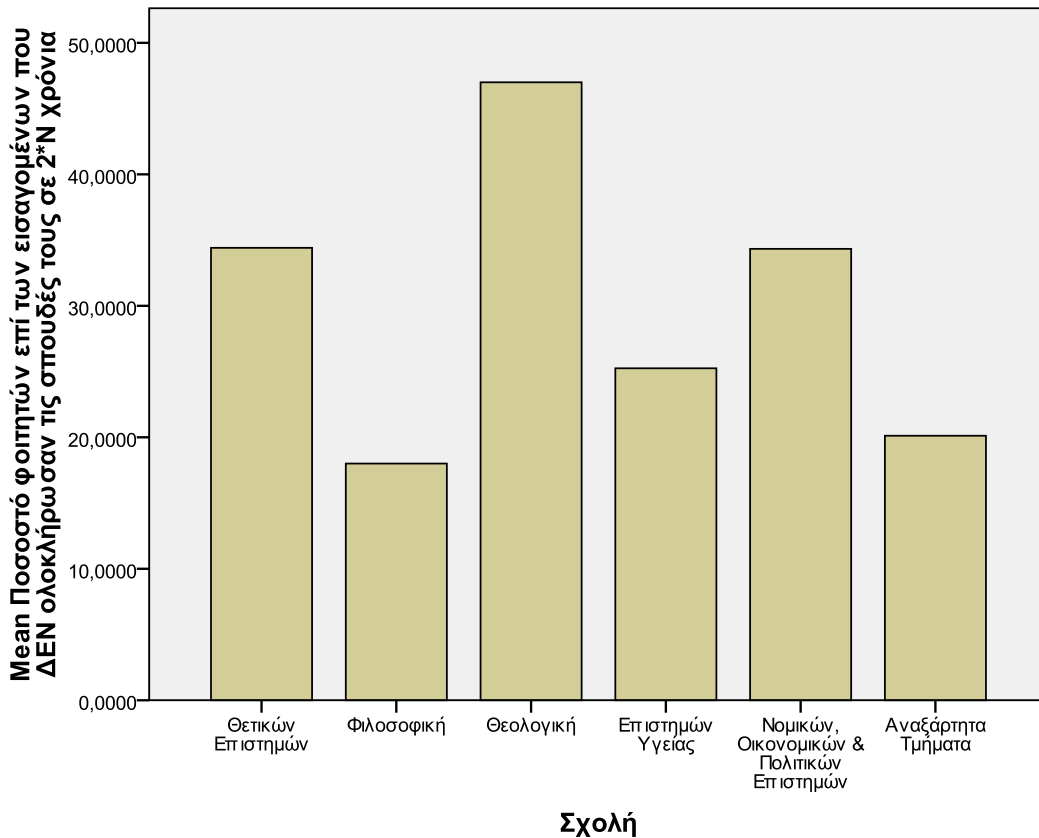
Η συμπερασματολογία αυτή μπορεί να αφορά την εξαγωγή απλών περιγραφικών μέτρων (δείκτες) ή ποιο σύνθετους ελέγχους και εκτιμήσεις παραμέτρων. Ουσιαστικά από την στιγμή που λαμβάνεται με κάποιο τρόπο το απαραίτητο πλήθος δεδομένων από τις επαναλήψεις του εκάστοτε αλγόριθμου τότε είναι διαθέσιμη σχεδόν όλη η γκάμα της στατιστικής (δηλαδή εκατοντάδες δείκτες και έλεγχοι κλπ) και δεν είναι εφικτό ούτε θεμιτό να εξαντληθεί το θέμα σε αυτό το σημείο. Συνεπώς θα παρουσιαστούν κάποια από τα διαθέσιμα εργαλεία της στατιστικής που μπορούν να έχουν εφαρμογή νόημα και ερμηνεία στην περίπτωση της Δ.Π.

➤ **Διάγραμμα διασποράς(scatter plot):** Είναι η απλούστερη διαγραμματική απεικόνιση των αποτελεσμάτων. Ένα τέτοιο διάγραμμα ,όταν το πλήθος των μετρήσεων είναι μεγάλο, αναμένεται να χάνει σε διακριτική ικανότητα γιατί γεμίζει με ολόκληρες σκούρες περιοχές των οποίων η πυκνότητα δεν είναι εφικτό να εκτιμηθεί.

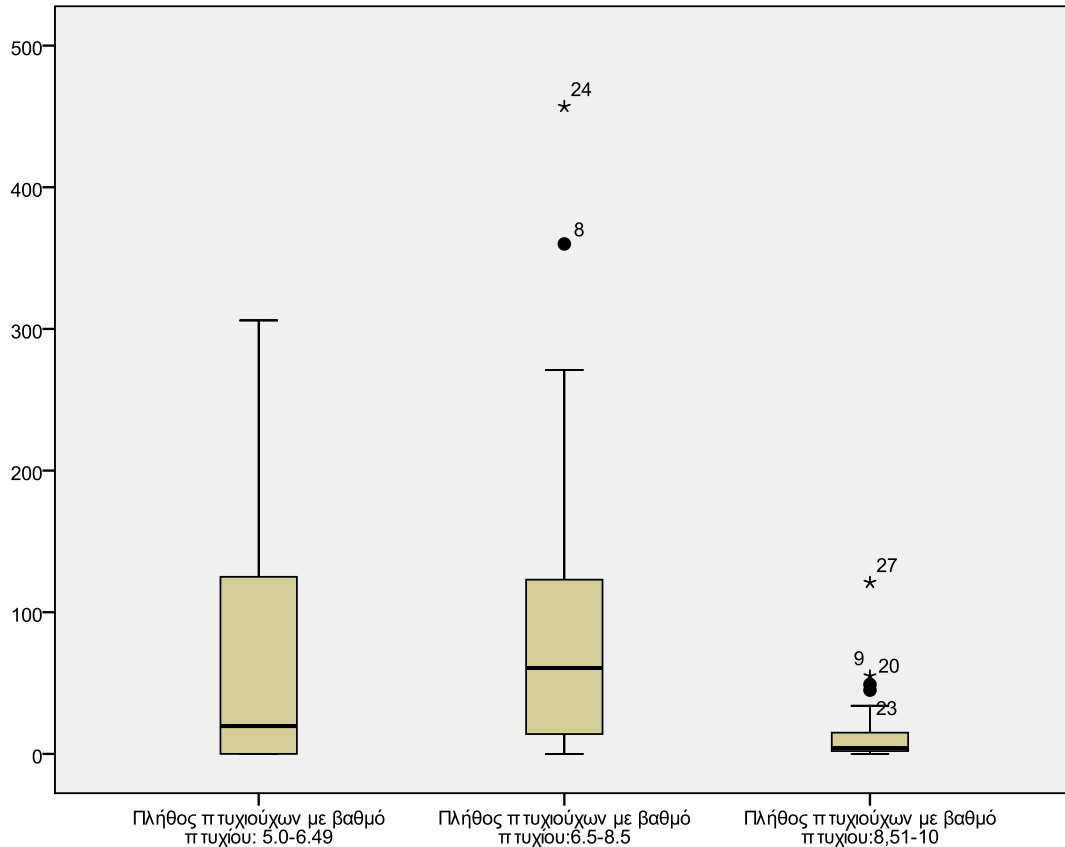


➤ **Ιστόγραμμα:** Είναι ένα γράφημα το οποίο είναι κατάλληλο για να αποκτήσει κανείς μια αίσθηση του τρόπου που κατανέμεται μια τυχαία μεταβλητή. Έχει νόημα μόνο για ποσοτικές μεταβλητές.

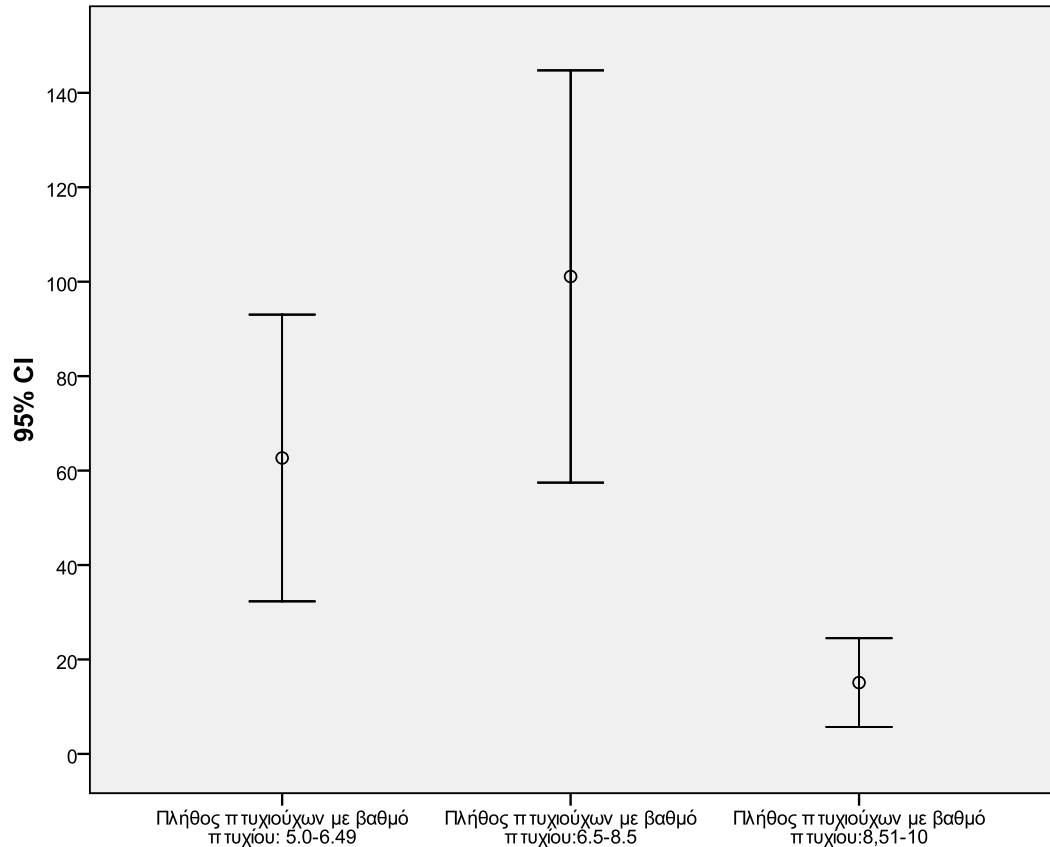




➤ **Θηκόγραμμα (Box plot):** Είναι ένα διάγραμμα το οποίο δείχνει το εύρος που έχουν τα τεταρτημόρια ενός συνόλου δεδομένων. Από το θηκόγραμμα μπορεί να υποθέσει κανείς την μορφή της κατανομής των δεδομένων με πολύ «χονδρικό» τρόπο. Σίγουρα καλύτερο για αυτή την εκτίμηση είναι το ιστόγραμμα αλλά το θηκόγραμμα έχει ένα μεγάλο πλεονέκτημα σε σχέση με το ιστόγραμμα: είναι εφικτό σε ένα διάγραμμα να τοποθετηθούν τα θηκογράμματα πολλών μεταβλητών ταυτόχρονα και να γίνει σύγκριση-κάτι που δεν μπορεί να γίνει με το ιστόγραμμα όταν οι μεταβλητές είναι πολλές.



➤ **Διάγραμμα σφαλμάτων (errors plot):** Το διάγραμμα αυτό απεικονίζει την μέση τιμή για κάθε μεταβλητή (μπορούν σε αυτό το διάγραμμα να απεικονισθούν πολλές μεταβλητές ταυτόχρονα) και ταυτόχρονα ένα διάστημα εμπιστοσύνης για αυτή την μεταβλητή. Το διάστημα εμπιστοσύνης αυτό εκτιμάται με βάση την διασπορά και υπό την υπόθεση της κανονικότητας. Για αυτό τον λόγο σπάνια μπορεί να βασισθεί κανείς σε αυτό.



Τα παραπάνω είναι διαγράμματα και δίνουν μια πρώτη εποπτική και χρήσιμη εικόνα των δεδομένων. Ακολουθούν μια σειρά από περιγραφικά μέτρα που μπορούν να εξαχθούν:

© **Μέση τιμή:** Είναι μέτρο θέσης .Ειδικά στην περίπτωση των σκορ της DEA η μέση τιμή πρέπει να συνεκτιμάται με την διακύμανση και/ή με το ιστόγραμμα γιατί είναι πιθανό μια μονάδα να μην εμφανίζει ποτέ το σκορ το οποίο εμφανίζεται ως μέση τιμή του σκορ της μονάδας αυτής.

© **Διακύμανση :**Είναι μέτρο της διασποράς των τιμών μιας μεταβλητής. Είναι ένας δείκτης που από μόνος του δεν «λέει πολλά » αλλά είναι παράγοντας συνεκτίμησης ,ε την μέση τιμή.

© **Εύρος:** Είναι ένας απλός δείκτης θέσης και διασποράς μπορεί να αποκαλύπτει πολλά κυρίως όταν είναι μικρό σε τιμή .

Ένα τυπικό παράδειγμα πίνακα με περιγραφικά στατιστικά:

Statistics		
Μέσος βαθμός πτυχίου κατά το έτος αναφοράς		
N	Valid	27
	Missing	4
Mean		5,1953
Median		6,7100
Std. Deviation		3,16523
Variance		10,019
Minimum		,00
Maximum		8,24

Άλλοι δείκτες όπως η συμμετρία λοξότητα κλπ μάλλον προσθέτουν πολυπλοκότητα παρά επεξηγηματική αξία.

6.3.1 Τα απαιτούμενα στατιστικά μέτρα κατά περίπτωση:

Για ποσοτικά δεδομένα που ακολουθούν την κανονική κατανομή:

Μέτρα: Μέση τιμή, τυπική απόκλιση, διακύμανση

Διαγράμματα: Ιστόγραμμα

Για ποσοτικά δεδομένα που δεν ακολουθούν την κανονική κατανομή:

Μέτρα: Διάμεσος, εύρος, ακραίες τιμές, ενδοτεταρτημοριακά διαστήματα.

Διαγράμματα: Ιστόγραμμα

Για κατηγορικά δεδομένα:

Μέτρα: πίνακας συχνοτήτων

Διαγράμματα: Ραβδόγραμμα συχνοτήτων, ή πίτα.

6.4 Κατασκευή κλιμάκων

Με τον όρο κατασκευή κλιμάκων εννοούμε τον υπολογισμό μιας «κρυφής» μεταβλητής (μη άμεσα παρατηρήσεις-latent) με χρήση άλλων μεταβλητών οι οποίες είναι άμεσα ή ποιο άμεσα παρατηρήσιμες.

Τυπικές περιπτώσεις τέτοιων μεταβλητών –σχετικές με Διασφάλιση Ποιότητας-είναι η ικανοποίηση ενός χρήστη από μια υπηρεσία, η ποιότητα μιας άυλης υπηρεσίας, κλπ.

Τέτοιες τεχνικές ονομάζονται τεχνικές μείωσης της διαστασιμότητας και ανήκουν στο επιστητό της πολυμεταβλητής ανάλυσης. Η ποιο αντιπροσωπευτική τεχνική της κλάσης αυτών των τεχνικών είναι η Ανάλυση Παραγόντων-Factor Analysis και η Ανάλυση Κυρίων Συνιστωσών – Principal Component analysis.

Οι τεχνικές αυτές και η εφαρμογή τους αναλύονται παρακάτω.

6.4.1 Πολυμεταβλητές τεχνικές εισαγωγικά

Το αντικείμενο της πολυμεταβλητής ανάλυσης είναι όπως ειπώθηκε η ταυτόχρονη ανάλυση πολλών μεταβλητών οι οποίες μπορεί να είναι συνεχείς, διακριτές ή κατηγορικές. Σκοπός της ανάλυσης είναι η ερμηνεία της διακύμανσης και των συνδιακυμάνσεων του συνόλου των μεταβλητών, η γραφική ερμηνεία των δεδομένων και η μείωση της διάστασης των δεδομένων.

Για τον σκοπό αυτό έχουν αναπτυχθεί τις τελευταίες (κυρίως) δεκαετίες μια σειρά από μεθοδολογίες και προσεγγίσεις τις οποίες θα μπορούσαμε να κατηγοριοποιήσουμε(Μουστάκη 2006) ως εξής¹:

Γραφικές προσεγγίσεις:

1. Cluster Analysis (Ανάλυση ομάδων)

¹ Η απόδοση στην Ελληνική γλώσσα δεν είναι διαθέσιμη για όλες τις μεθοδολογίες, οπότε παρατίθεται σε παρένθεση μόνο όποια είναι ευρέως αποδεκτή.

2. Multidimensional scaling
3. Correspondence analysis

Μαθηματικές προσεγγίσεις:

1. Principal Component Analysis (Ανάλυση κυρίων συνιστωσών)

Μοντελοποιημένες Προσεγγίσεις:

1. Latent Variable models
2. Structural equation models

1.1.1.1 Τύποι Μεταβλητών – Επίπεδα Μέτρησης

Το αντικείμενο της πολυμεταβλητής ανάλυσης είναι όπως ειπώθηκε η ταυτόχρονη ανάλυση πολλών μεταβλητών οι οποίες μπορεί να είναι συνεχείς, διακριτές ή κατηγορικές. Σκοπός της ανάλυσης είναι η ερμηνεία της διακύμανσης και των συνδιακυμάνσεων του συνόλου των μεταβλητών, η γραφική ερμηνεία των δεδομένων και η μείωση της διάστασης των δεδομένων.

Για τον σκοπό αυτό έχουν αναπτυχθεί τις τελευταίες (κυρίως) δεκαετίες μια σειρά από μεθοδολογίες και προσεγγίσεις τις οποίες θα μπορούσαμε να κατηγοριοποιήσουμε (Μουστάκη 2006) ως εξής²:

Γραφικές προσεγγίσεις:

4. Cluster Analysis (Ανάλυση ομάδων)
5. Multidimensional scaling
6. Correspondence analysis

Μαθηματικές προσεγγίσεις:

2. Principal Component Analysis (Ανάλυση κυρίων συνιστωσών)

² Η απόδοση στην Ελληνική γλώσσα δεν είναι διαθέσιμη για όλες τις μεθοδολογίες, οπότε παρατίθεται σε παρένθεση μόνο όποια είναι ευρέως αποδεκτή.

Μοντελοποιημένες Προσεγγίσεις:

3. Latent Variable models
4. Structural equation models

Οι μεταβλητές οι οποίες χρησιμοποιούνται στην πολυμεταβλητή ανάλυση μπορούν επίσης να κατηγοριοποιηθούν με βάση κάποια χαρακτηριστικά τους. Έτσι υπάρχουν τα εξής είδη μεταβλητών:

Ανάλογα με το αν είναι ποσοτικές ή όχι:

Ποσοτικές μεταβλητές (Metrical Variables): συνεχείς ή διακριτές Π.χ. τα αποτελέσματα (σκορ) εξετάσεων, χρηματικά ποσά, κ.λ.π.

Κατηγορικές μεταβλητές (Non Metrical Variables): είναι οι μεταβλητές που δεν εκφράζουν κάποια ποσότητα, αλλά αντιπροσωπεύουν μια κατηγορία. Χωρίζονται σε Ordinal ή Nominal ανάλογα αν η τιμή μιας μεταβλητής έχει κάποια διάταξη ή όχι.

Ανάλογα με το αν παρατηρούνται άμεσα ή όχι:

Παρατηρούμενες μεταβλητές: (Manifest Variables, observed variables, Indicator, Items). Όλα αυτά είναι παρατηρούμενες μεταβλητές.

Ψευδομεταβλητές: (Latent Variables, unobserved variables, factors, Components)

Όλες αυτές κατασκευάζονται δευτερογενώς από τις παρατηρούμενες.

Οι κατηγορικές μεταβλητές ανάλογα με το πλήθος και το είδος των επιπέδων που μπορούν να εμφανιστούν μέσα σε αυτές χωρίζονται σε:

Δίτιμες: Διχοτομικά δεδομένα: απαντήσεις του τύπου «σωστό – λάθος» ή «ναι – όχι».

Πολύτιμες με διάταξη (ordinal): όταν τα διάφορα επίπεδα της μεταβλητής εκφράζουν μια κλιμακούμενη ένταση στο μέγεθος που μετράνε. Π.χ. όταν υπάρχει απάντηση ενός ασθενούς για την κλινική του κατάσταση η οποία χαρακτηρίζεται από πολλά επίπεδα ταξινομημένα προς το καλύτερο ή χειρότερο.

Πολύτιμες χωρίς διάταξη (nominal): όταν τα διάφορα επίπεδα της μεταβλητής δεν εκφράζουν μια κλιμακούμενη ένταση στο μέγεθος που μετράνε. Π.χ. επιλογή ανάμεσα σε χρώματα ή σε λέξεις που μπορεί να εκφράζουν καλύτερα ένα νόημα κ.λπ.

6.4.2 Factor Analysis

Η ανάλυση παραγόντων είναι η παλαιότερη και πιο διαδεδομένη μέθοδος από μια κλάση μεθόδων οι οποίες είναι γνωστές με το όνομα "latent variable methods". Η γενική ιδέα και στη factor analysis (FA) είναι ότι υπάρχουν κάποιες μη παρατηρούμενες μεταβλητές οι οποίες βρίσκονται σε λανθάνουσα μορφή και ο στόχος είναι να γίνει ανάκτηση της τιμής αυτών των μεταβλητών μέσω της μέτρησης κάποιων άλλων παρατηρήσιμων και μετρούμενων (observable) μεταβλητών οι οποίες έχουν μια συσχέτιση από τις μη μετρούμενες. Υπάρχει μια πολύ στενή σχέση μεταξύ της PCA και της FA. Σε διάφορα βιβλία (Basilevsky 2004, 351-360) και στατιστικά πακέτα (SPSS) οι δύο μέθοδοι αντιμετωπίζονται κάτω από ένα κοινό πλαίσιο αναφοράς. Η βασική ιδέα και ειδοποιός διαφορά σε σχέση με την PCA είναι η εισαγωγή ενός επεξηγηματικού μοντέλου με συγκεκριμένο πλήθος μεταβλητών (το οποίο πλήθος πρέπει να είναι καθορισμένο από πριν τρέξει η μέθοδος) με συγκεκριμένη επίσης προσχηματισμένη ερμηνεία για το ρόλο των μεταβλητών. Μερικές από τις πιο κλασικές εφαρμογές της μεθόδου αφορούν τις επιστήμες του μάρκετινγκ της επιχειρησιακής έρευνας, της ψυχομετρίας κ.λ.π.

Μετρητικά Μοντέλα (Measurement Models)

Πολλές θεωρίες στις συμπεριφοριστικές, στις οικονομικές και κοινωνικές επιστήμες είναι τυποποιημένες και βασισμένες πάνω σε θεωρητικά οικοδομήματα τα οποία είναι δύσκολο να μετρηθούν ή να παρατηρηθούν. Αυτό συμβαίνει συχνά και στο χώρο της υγείας όσο και γενικότερα στον χώρο της παροχής υπηρεσιών καθόσον η υπηρεσία εξ ορισμού είναι κάτι δύσκολα μετρήσιμο ως άυλη ποσότητα.

Η μέτρηση μιας δομής ή μιας ποσότητας μπορεί να επιτευχθεί με χρήση ενός ή περισσότερων δεικτών (π.χ. ερωτηματολόγια κ.λ.π.) Η χρήση των μετρητικών μοντέλων είναι να περιγράψουν πόσο καλά οι παρατηρούμενοι δείκτες πρόσφεραν ένα μετρητικό εργαλείο για την προς μέτρηση δομή. Είναι επίσης γνωστοί ως latent variables (οι δείκτες συχνότερα αλλά πιο σπάνια και τα μοντέλα).

Τα μετρητικά μοντέλα συνήθως προτείνουν τρόπους με βάση τους οποίους οι παρατηρούμενες μετρήσεις μπορούν να βελτιωθούν. Αυτό αφορά συνήθως τη διαδικασία της αναπαραγωγής των μετρήσεων. Σε μερικές περιπτώσεις ένα μοντέλο μπορεί να ερμηνεύεται από μια μόνο ψευδομεταβλητή, αλλά τις περισσότερες φορές όμως είναι πολυμεταβλητά από τη φύση τους, οπότε απαιτούν περισσότερες από μια ψευδομεταβλητές.

Στόχοι Της Μεθόδου

Σε αναλογία με τη PCA και εν συντομία αναφέρονται οι στόχοι και τα ενδιάμεσα βήματα της factor analysis:

- Κατασκευή κλίμακας
- Μελέτη των σχέσεων μεταξύ ενός συνόλου παρατηρούμενων δεικτών και προσδιορισμός των παραγόντων που επηρεάζουν αυτούς τους δείκτες
- Μείωση των διαστάσεων

- Προσαρμογή μιας ψευδομεταβλητής (μοντέλο) με ένα ή περισσότερους παράγοντες.
- Αξιολόγηση των μετρήσεων με βάση τις προσδιορισμένες διαστάσεις (προσδιορισμένες από τις ψευδομεταβλητές)

Ψευδομεταβλητά Μοντέλα

Όπως αναφέρθηκε και νωρίτερα υπάρχει μια κλάση μεθόδων που είναι βασισμένες πάνω σε μοντέλα. Μια συνοπτική παρουσίαση και κατηγοριοποίηση τους γίνεται στον πίνακα 10.

Όπως φαίνεται από τον πίνακα, η Factor Analysis αφορά την περίπτωση όπου τόσο οι παρατηρούμενες όσο και οι μη παρατηρούμενες μεταβλητές είναι ποσοτικές.

Η ανάλυση αυτή μπορεί να χρησιμοποιηθεί τόσο εξερευνητικά, για να οικοδομηθεί μια θεωρία για τα παρατηρούμενα δεδομένα, όσο και επιβεβαιωτικά για να επαληθευτεί μια προϋπάρχουσα θεωρία όταν προϋπάρχει κάποια θεωρία σχετικά με τα δεδομένα (Exploratory Latent Variable Analysis, Confirmatory Latent Variable Analysis).

		ΦΑΝΕΡΕΣ ΜΕΤΑΒΛΗΤΕΣ (MANIFEST VARIABLES)			
		Μετρικές	Μετρικές	Κατηγορικές	Μικτές
ΨΕΥΔΟΜΕΤΑΒΛΗΤΕΣ (LATENT VARIABLES)	Μετρικές	Factor Analysis	Latent Trait Analysis	Latent Trait Analysis	
	Κατηγορικές	Latent Profile Analysis	Latent Class Analysis	Latent Class Analysis	

Όλα τα Ψευδομεταβλητά μοντέλα είναι παρόμοια με τη γραμμική παλινδρόμηση. Η παλινδρόμηση μπορεί να εννοηθεί ανάμεσα στις φανερές μεταβλητές και στις ψευδομεταβλητές.

Οι υποθέσεις σχετικά με τις κατανομές διατυπώνονται όσον αφορά τα κατάλοιπα της παλινδρόμησης ή τα σφάλματα και έτσι επιτρέπεται να εξαχθούν συμπεράσματα.

Ο στόχος είναι να αντιστραφούν οι συσχετίσεις της παλινδρόμησης και να γίνουν γνωστές οι ψευδομεταβλητές όταν δίνονται οι φανερές μεταβλητές. Εφόσον δεν μπορούν να παρατηρηθούν οι ψευδομεταβλητές, ο ερευνητής μπορεί μόνο να μάθει γι' αυτές έμμεσα εξάγοντας αυτή τη γνώση από τις παρατηρούμενες μεταβλητές.

Μερικές φορές μπορεί να συμβεί κάποιες φανερές μεταβλητές να εξαρτώνται από την ίδια ψευδομεταβλητή.

Η ύπαρξη μιας συσχέτισης μεταξύ δυο δεικτών (φανερών) μπορεί να αποτελέσει σοβαρή ένδειξη για μια καταγωγή ή επίδραση. Όσο ακόμα παραμένει υπαρκτή μια συσχέτιση μπορεί να υποτεθεί η ύπαρξη μιας επιπλέον κοινής πηγής επίδρασης.

Πολύ συχνά ένα μοντέλο μπορεί να αναπαρασταθεί με ένα διάγραμμα διαδρομών το οποίο αναπαριστά τις ψευδομεταβλητές και τις εξαρτώμενες (συσχετισμένες) παρατηρούμενες μεταβλητές:

6.5 Principal Component Analysis

Η ανάλυση κυρίων συνιστωσών (Principal Component Analysis - PCA) είναι μια από τις πιο διαδεδομένες μεθόδους της πολυμεταβλητής ανάλυσης. Η PCA ανήκει στις μαθηματικές τεχνικές της πολυμεταβλητής ανάλυσης. Ουσιαστικά πρόκειται για έναν ορθογώνιο μετασχηματισμό από τον R^p χώρο των παρατηρούμενων μεταβλητών σε έναν R^q χώρο με $p < q$.

Σκοποί – Στόχοι Της Principal Component Analysis

Ο στόχος αρχικά είναι να αντικατασταθούν οι p -συσχετισμένες μεταβλητές από ένα μικρότερο αριθμό μη συσχετισμένων μεταβλητών, όπου όμως θα βρίσκεται η σχεδόν συνολική πληροφορία των αρχικών δεδομένων. Με άλλα λόγια ο σκοπός είναι να μειώσει τη διαστασιμότητα των αρχικών δεδομένων με την ελάχιστη μείωση της διαθέσιμης σε αυτά πληροφορίας.

Για το σκοπό αυτό ερμηνεύεται η προκύπτουσα δομή των δεδομένων στο όριο των πιο σημαντικών principal components.

Ο όρος principal components (κύριες συνιστώσες) αφορά τις μη παρατηρούμενες, αλλά υπαρκτές μεταβλητές οι οποίες είναι λιγότερες στο πλήθος από τις παρατηρούμενες και επηρεάζουν άμεσα τις παρατηρούμενες.

Η πληροφορία μετριέται με βάση τη συνολική διακύμανση των μεταβλητών, ενώ η εξήγηση των components γίνεται με το συγκεκριμένο κάθε φορά παράδειγμα. Αν οι αρχικές μεταβλητές είναι ασυσχέτιστες, τότε κρατώνται όλες καθόσον αυτό σημαίνει ότι κάθε μια αρχική παρατηρούμενη μεταβλητή είναι και ένα component.

Μεθοδολογία PCA

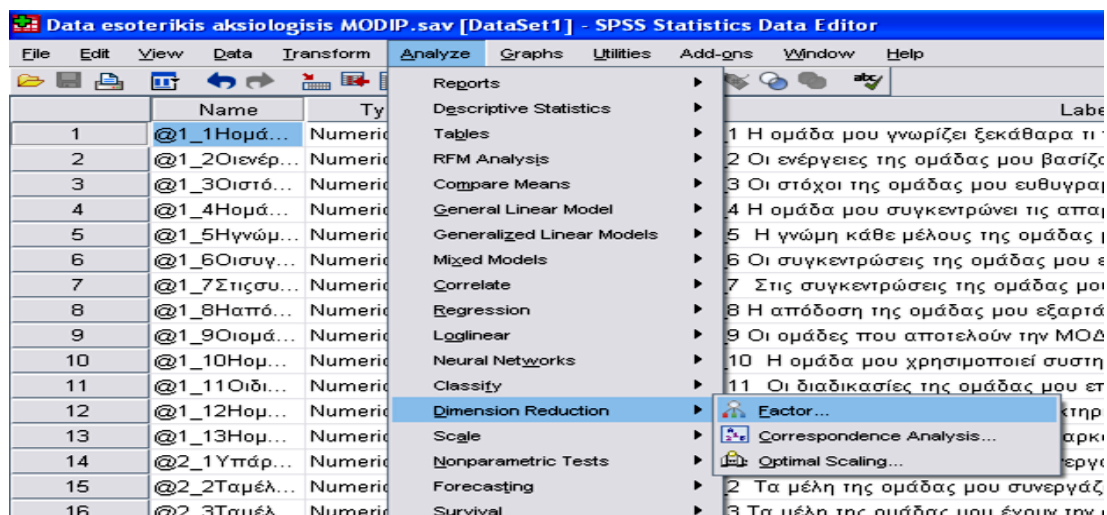
Η PCA μετασχηματίζει το σύνολο των συσχετιζόμενων x_1, x_2, \dots, x_p σε ένα σύνολο ασυσχέτιστων μεταβλητών y_1, y_2, \dots, y_p (τα οποία καλούνται principal components) έτσι ώστε η y_1 να εξηγεί τη μέγιστη δυνατή διακύμανση από τη συνολική διακύμανση, η y_2 να εξηγεί όσο περισσότερο γίνεται από την υπολειπόμενη διακύμανση κ.λπ.. Το σύνολο των y_1, \dots, y_p μεταβλητών εξηγεί το σύνολο της διακύμανσης:

Πριν χρησιμοποιηθεί η μεθοδολογία PCA, κανονικοποιείται στη μονάδα η διακύμανση των x : ώστε έτσι κάθε ένα x να συμμετέχει ακριβώς το ίδιο στην εξήγηση της συνολικής διακύμανσης:

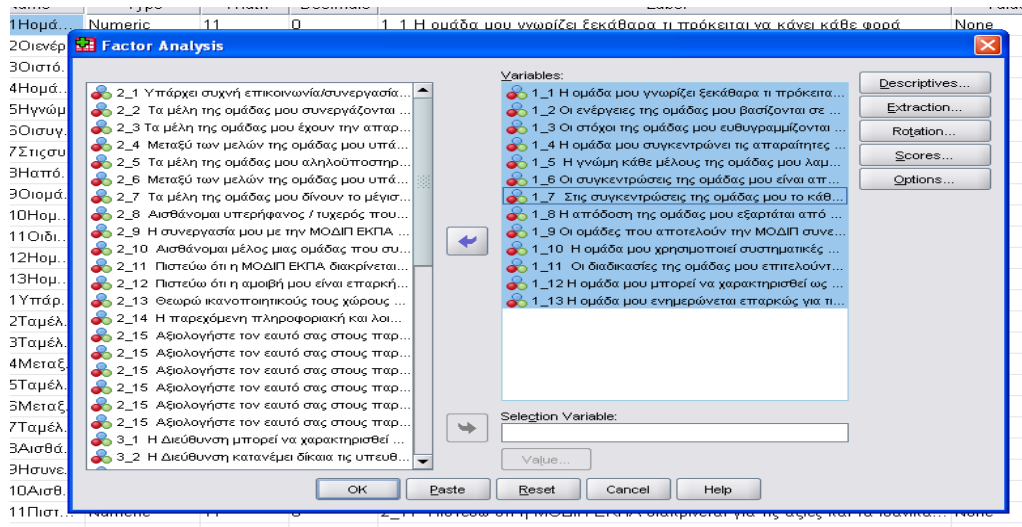
Κανονικοποιείται η διακύμανση αλλά και η μέση τιμή αφαιρώντας από όλα τη μέση τιμή ώστε να έχει προκύψει τελικά τυποποιημένη κανονική κατανομή $N(0,1)$.

6.5.1 Εκτέλεση με το SPSS

Επιλέγουμε από το μενού analyze....

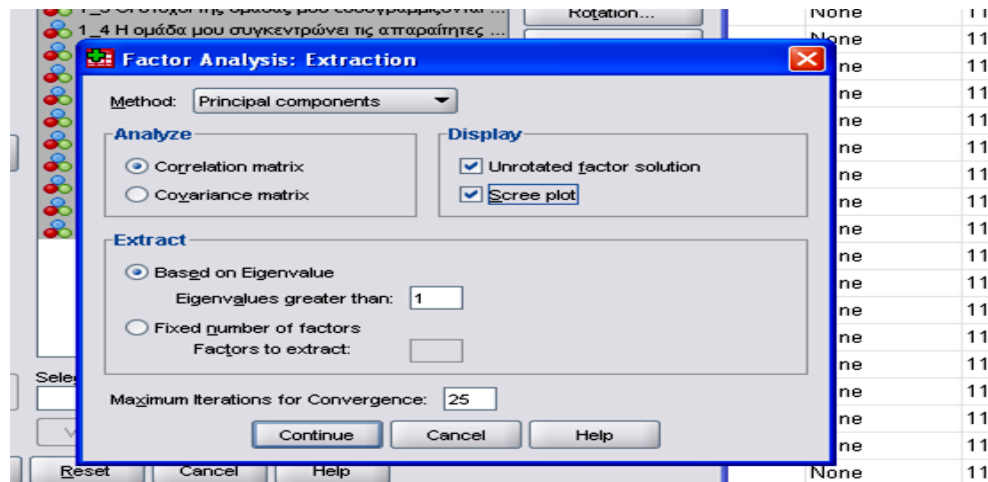


Στην συνέχεια επιλέγουμε το σύνολο των μεταβλητών τις οποίες θέλουμε να διερευνήσουμε ως προς την ύπαρξη παραγόντων:

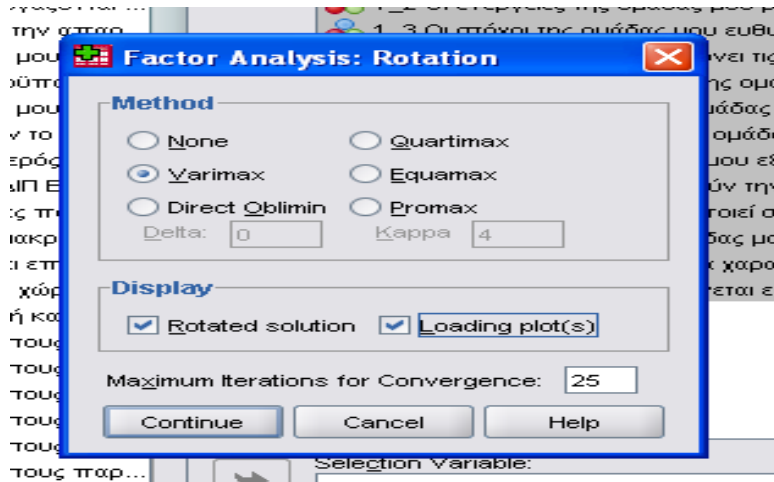


Κατόπιν παραμετροποιούμε την ανάλυση:

Η προεπιλεγμένη μέθοδος είναι η PCA. Επιλέγουμε ανοικτό πλήθος παραγόντων και διαλέγουμε ως κατώφλι για την επιλογή ενός παράγοντα την τιμή της μοναδιαίας ιδιοτιμής:



Επιλέγουμε κάποια μέθοδος περιστροφής (για το ταίριασμα των φορτίσεων). Προτεινόμενη μέθοδος: Varimax.



Αποτελέσματα:

Από τα αποτελέσματα πάντα συνεκτιμούμε τον πίνακα της ερμηνευμένης διακύμανσης μαζί με το scree plot:

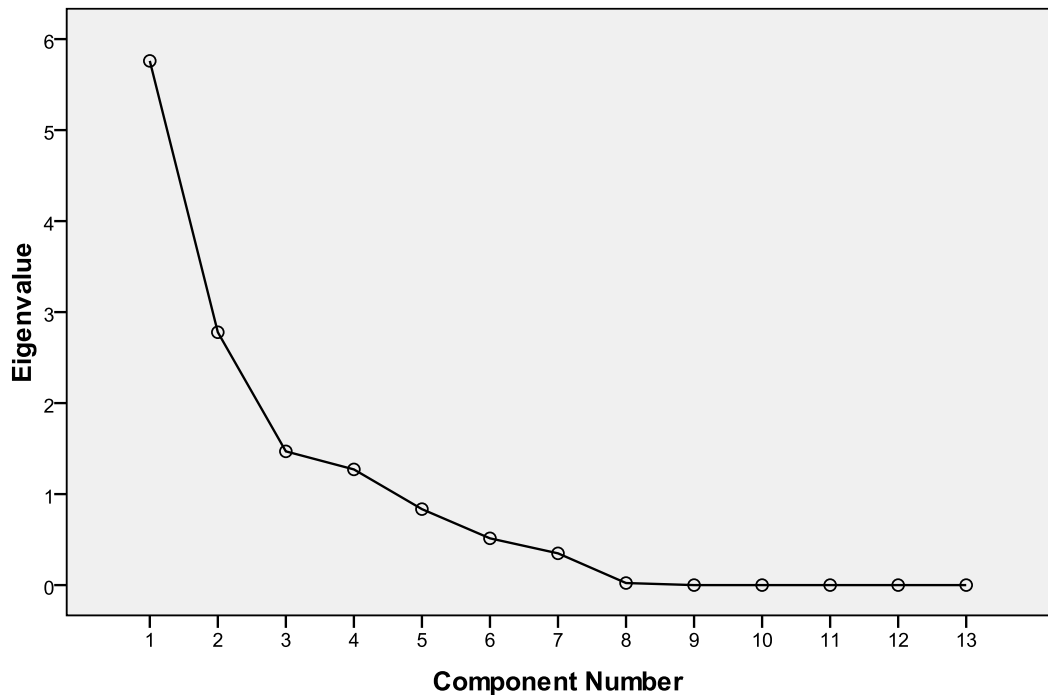
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
	1	5,760	44,310	44,310	5,760	44,310	44,310	3,978	30,599
2	2,779	21,374	65,683	2,779	21,374	65,683	2,659	20,451	51,050
3	1,470	11,305	76,988	1,470	11,305	76,988	2,528	19,446	70,496
4	1,271	9,778	86,766	1,271	9,778	86,766	2,115	16,271	86,766
5	,835	6,422	93,188						
6	,514	3,953	97,141						
7	,349	2,685	99,826						
8	,023	,174	100,000						
9	8,897E-17	6,844E-16	100,000						
10	-3,879E-19	-2,984E-18	100,000						

11	-7,935E-	100,000						
	1,032E-	16						
	16							
12	-1,465E-	100,000						
	1,904E-	15						
	16							
13	-4,498E-	100,000						
	5,848E-	15						
	16							

Extraction Method: Principal Component Analysis.

Scree Plot



Η κρίσιμη απόφαση έχει να κάνει με το αν θα επιλέξουμε το προτεινόμενο πλήθος παραγόντων, αν θα το μειώσουμε ή αν θα το αυξήσουμε.

Πέρα των δυο πινάκων, θα πρέπει να συνοπολογίζεται το γεγονός ότι οι παράγοντες που θα προκύψουν θα πρέπει να έχουν ερμηνευτική υπόσταση και χρηστική αξία. Το γεγονός αυτό μπορεί

να ελεγχθεί κατά βάση από τον πίνακα όπου δίνονται οι φορτίσεις μετά την στροφή:

Rotated Component Matrix ^a				
	Component			
	1	2	3	4
1_1 Η ομάδα μου γνωρίζει ξεκάθαρα τι πρόκειται να κάνει κάθε φορά	,873			
1_2 Οι ενέργειες της ομάδας μου βασίζονται σε ένα ξεκάθαρο βασικό σκοπό της υπηρεσίας	,912			
1_3 Οι στόχοι της ομάδας μου ευθυγραμμίζονται με τους στόχους της υπηρεσίας				,866
1_4 Η ομάδα μου συγκεντρώνει τις απαραίτητες δεξιότητες και έχει επαρκές προσωπικό για να πετύχει τους στόχους της			,889	
1_5 Η γνώμη κάθε μέλους της ομάδας μου λαμβάνεται υπόψη στην λήψη αποφάσεων	,829			
1_6 Οι συγκεντρώσεις της ομάδας μου είναι αποτελεσματικές και παράγουν έργο		,838		
1_7 Στις συγκεντρώσεις της ομάδας μου το κάθε μέλος συμμετέχει ενεργά και διατυπώνει ελεύθερα τις απόψεις του	,902			
1_8 Η απόδοση της ομάδας μου εξαρτάται από την απόδοση του εκάστοτε ατομικού έργου			,907	
1_9 Οι ομάδες που αποτελούν την ΜΟΔΙΠ συνεργάζονται ικανοποιητικά μεταξύ τους				,868
1_10 Η ομάδα μου χρησιμοποιεί συστηματικές διαδικασίες λήψης αποφάσεων και επίλυσης προβλημάτων		,650	,617	
1_11 Οι διαδικασίες της ομάδας μου επιτελούνται στον συμφωνημένο χρόνο			,539	
1_12 Η ομάδα μου μπορεί να	,567	,610		

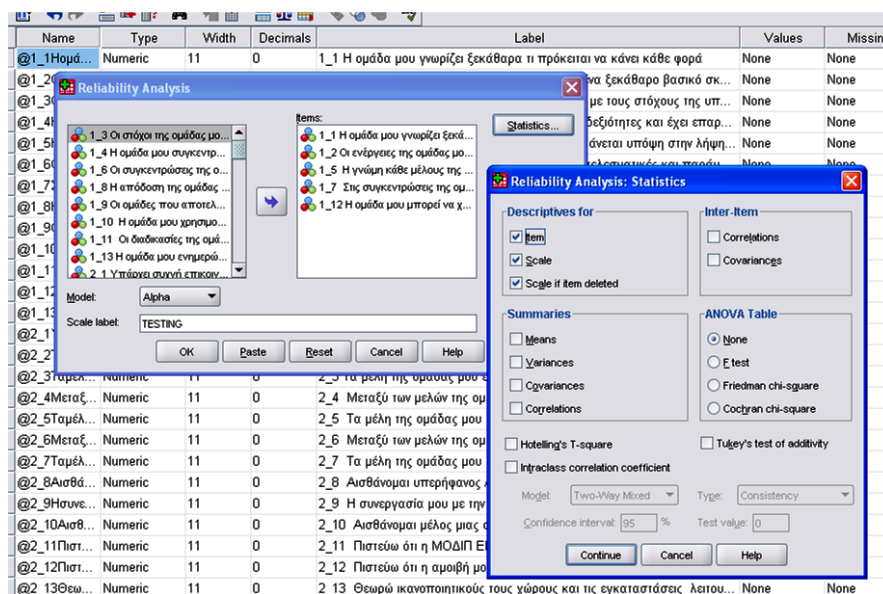
χαρακτηρισθεί ως ευέλικτη στην λήψη αποφάσεων και την αντιμετώπιση προβλημάτων				
1_13 Η ομάδα μου ενημερώνεται επαρκώς για τις αλλαγές και τις εξελίξεις στο αντικείμενο εργασίας της και συμμετέχει σε συνέδρια και λοιπές σχετικές δραστηριότητες			,738	
Extraction Method: Principal Component Analysis.				
Rotation Method: Varimax with Kaiser Normalization.				
a. Rotation converged in 7 iterations.				

Η μεθοδολογία ερμηνείας και ονοματοδοσίας:

- Κρατάμε φορτίσεις μεγαλύτερες του 0,5
- Ερμηνεύουμε με βάση τις εναπομείνουσες φορτίσεις
- Αν μια στήλη δεν έχει καθαρή ερμηνεία, ή πολλές μεταβλητές ή αδυναμία ερμηνείας ή κλπ την απορρίπτουμε.

Κατόπιν :

Για κάθε κλίμακα (στήλη) υπολογίζουμε τον συντελεστή α -cronbach, δηλαδή την αξιοπιστία της κλίμακας:



Reliability Statistics

Cronbach's Alpha	N of Items
,903	5

Αποδεκτές τιμές Αξιοπιστίας είναι οι τιμές του $\alpha > 0,7$.

Παράλληλα εξετάζουμε και το κριτήριο Keizer/Meier/Olkin.

Εξάγουμε και τον πίνακα "scale if item deleted" από τον οποίο ενδεχομένως να προκύψει ανάλογα με την εκτίμησή μας διόρθωση στο πλήθος των μεταβλητών.

Αν το πλήθος είναι μικρότερο ή ίσο του τρία, τότε συνεπικουρούμενοι από σχετικές τιμές του στατιστικού "scale if item deleted" προτείνεται να εισάγονται επιπλέον μεταβλητές προκειμένου να βελτιώνεται η σφαιρικότητα Keizer/Meier/Olkin η οποία μοιραία είναι χαμηλή για μικρό πλήθος ερωτήσεων.

6.6 Έλεγχοι

Παρακάτω θα αναφερθούν μερικοί από τους πιο συχνοί ελέγχους που εμφανίζονται στην περίπτωση της Δ.Π.

6.6.1 Έλεγχος κανονικότητας

Αναφέραμε προηγουμένως πως όταν το ιστόγραμμα συχνοτήτων των ποσοτικών μεταβλητών έχει το σχήμα “καμπάνας”, τότε λέμε ότι τα δεδομένα ακολουθούν την κανονική κατανομή ή κατανέμονται κανονικά. Το ιστόγραμμα όμως δεν είναι “ικανό” να μας απαντήσει στη ερώτηση αν είναι κανονικά τα δεδομένα ή αν προέρχονται από μία κανονική κατανομή με ένα μέσο και μία διακύμανση. Μπορούμε να κατασκευάσουμε δύο γραφήματα με το SPSS, το **P-P Plot** και το **Q-Q Plot** (Επιλέγοντας **Analyze**→**Descriptive Statistics**→**P-P Plots** ή **Q-Q Plots**). Με αυτά τα γραφήματα ελέγχουμε οπτικά την ύπαρξη κανονικότητας στα δεδομένα. Όσο πιο κοντά στην ευθεία είναι τα σημεία του σχήματος τόσο πιο πολλές είναι οι ενδείξεις ότι τα δεδομένα ακολουθούν την κανονική κατανομή. Το μάτι όμως πάλι μπορεί να “πέσει έξω” και να ξεγελαστούμε. Για αυτό το λόγο καταφεύγουμε σε τεστ κανονικότητας για να απαντήσουμε στην προηγούμενη ερώτηση.

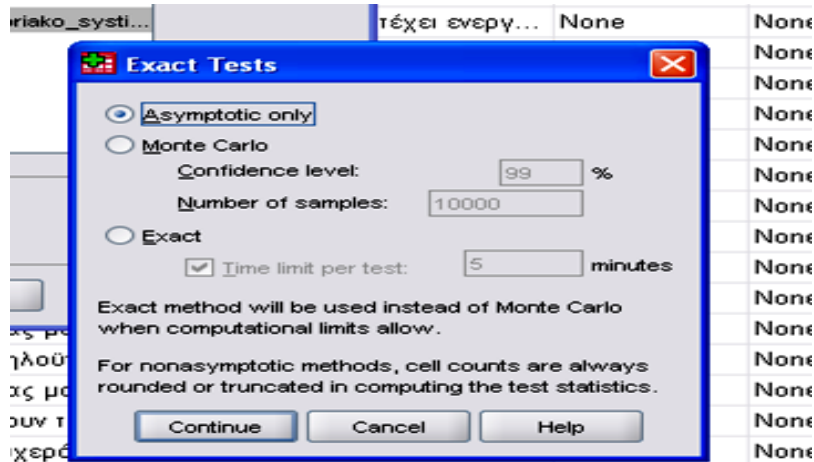
Ο έλεγχος κανονικότητας υπάγεται σε μία ευρύτερη οικογένεια ελέγχων, τη λεγόμενη «έλεγχοι υποθέσεων». Όταν ακούμε για ελέγχους υποθέσεων μας έρχονται πολλά πράγματα στο μυαλό. Κάποια από αυτά είναι η μηδενική υπόθεση (**Null Hypothesis** ή **H₀**), η εναλλακτική υπόθεση (**Alternative Hypothesis** ή **H₁**), το επίπεδο στατιστικής σημαντικότητας (**α**) και το παρατηρηθέν επίπεδο στατιστικής σημαντικότητας (**p-value** ή **Significance**). Οι υποθέσεις είναι της ακόλουθης μορφής:

H₀: Η κατανομή των δεδομένων δε διαφέρει από την κανονική κατανομή

H₁: Η κατανομή των δεδομένων διαφέρει από την κανονική κατανομή

Για τη διεξαγωγή των ελέγχων υποθέσεων χρησιμοποιούνται κάποιοι μαθηματικοί τύποι, που καλούνται ελεγχοσυναρτήσεις. Με βάση το αποτέλεσμα τους οδηγούμαστε στο συμπέρασμα ότι η μηδενική υπόθεση απορρίπτεται ή όχι. Στη συγκεκριμένη περίπτωση η μηδενική υπόθεση την οποία θέλουμε να ελέγξουμε είναι ότι τα δεδομένα ακολουθούν την κανονική ή ότι προέρχονται από ένα πληθυσμό που ακολουθεί την κανονική κατανομή. Η εναλλακτική είναι ότι τα δεδομένα δεν ακολουθούν την κανονική κατανομή. Το επίπεδο στατιστικής σημαντικότητας ορίζεται συνήθως ίσο με 0.05 ή 5%. Το παρατηρηθέν επίπεδο στατιστικής σημαντικότητας ορίζεται ως η πιθανότητα η τιμή του ελέγχου (ελεγχοσυνάρτησης) να πάρει μία τιμή τόσο ακραία ή περισσότερο ακραία από αυτή που πήρε στο συγκεκριμένο δείγμα κάτω από τη μηδενική υπόθεση. Αν η p-value είναι μικρότερη του 0.05, τότε λέμε ότι η μηδενική υπόθεση απορρίπτεται. Αν η p-value είναι μεγαλύτερη ή ίση του 0.05, τότε λέμε ότι η μηδενική υπόθεση δεν απορρίπτεται. Το SPSS εμφανίζει τις τιμές των παρατηρηθέντων επιπέδων στατιστικής σημαντικότητας και τις ονομάζει **(Asymptotic) Significances**. Ο λόγος που χρειαζόμαστε την κανονικότητα των δεδομένων, είναι για να έχουν ισχύ κάποιες στατιστικές τεχνικές που θα χρησιμοποιήσουμε όπως οι έλεγχοι υποθέσεων για τους μέσους, η γραμμική παλινδρόμηση, η ανάλυση διακύμανσης κ.ά. Ας δούμε τώρα στο SPSS πως θα διεξάγουμε ελέγχους κανονικότητας. Πατάμε **Analyze→Nonparametric Tests→1-Sample K-S** και εμφανίζεται το παράθυρο της εικόνας :

και ένα 99% διάστημα εμπιστοσύνης για τον μέσο όρο αυτών των p-value βασισμένο προφανώς στις 10000 p-value..



6.6.2 Συντελεστές συσχέτισης

Οι συντελεστές που θα παρουσιαστούν παρακάτω, αναφέρονται στη γραμμικής φύσεως σχέση που μπορεί να συνδέει τις δύο μεταβλητές. Οι τιμές που μπορεί να πάρει ένας συντελεστής συσχέτισης είναι από -1 έως +1. Αρνητικές τιμές του συντελεστή γραμμικής συσχέτισης δύο μεταβλητών σημαίνει ότι έχουμε την ύπαρξη αρνητικής γραμμικής συσχέτισης. Δηλαδή, οι μεγαλύτερες τιμές της μίας μεταβλητής τείνουν να αντιστοιχούν στις μικρότερες τιμές της άλλης μεταβλητής. Θετικές τιμές του συντελεστή γραμμικής συσχέτισης είναι ένδειξη θετικής γραμμικής συσχέτισης μεταξύ των δύο μεταβλητών. Δηλαδή, οι μεγαλύτερες τιμές της μίας μεταβλητής τείνουν να αντιστοιχούν στις μεγαλύτερες τιμές της άλλης μεταβλητής. Τιμές κοντά στο μηδέν αποτελούν ένδειξη ότι δεν υπάρχει στατιστικά σημαντική γραμμική συσχέτιση μεταξύ των δύο μεταβλητών. Όσο πιο μεγάλες είναι οι τιμές του συντελεστή, ή όσο πιο κοντά βρίσκονται στη μονάδα (σε απόλυτη τιμή πάντα), τόσο πιο ισχυρή είναι η γραμμική συσχέτιση μεταξύ τους.

Οι πιο γνωστοί συντελεστές γραμμικής συσχέτισης είναι οι συντελεστές του **Pearson**, του **Spearman** και του **Kendall**. Η μηδενική και η εναλλακτική υπόθεση εδώ είναι οι εξής:

$H_0: \rho=0$ ή δεν υπάρχει γραμμική συσχέτιση μεταξύ των δύο μεταβλητών

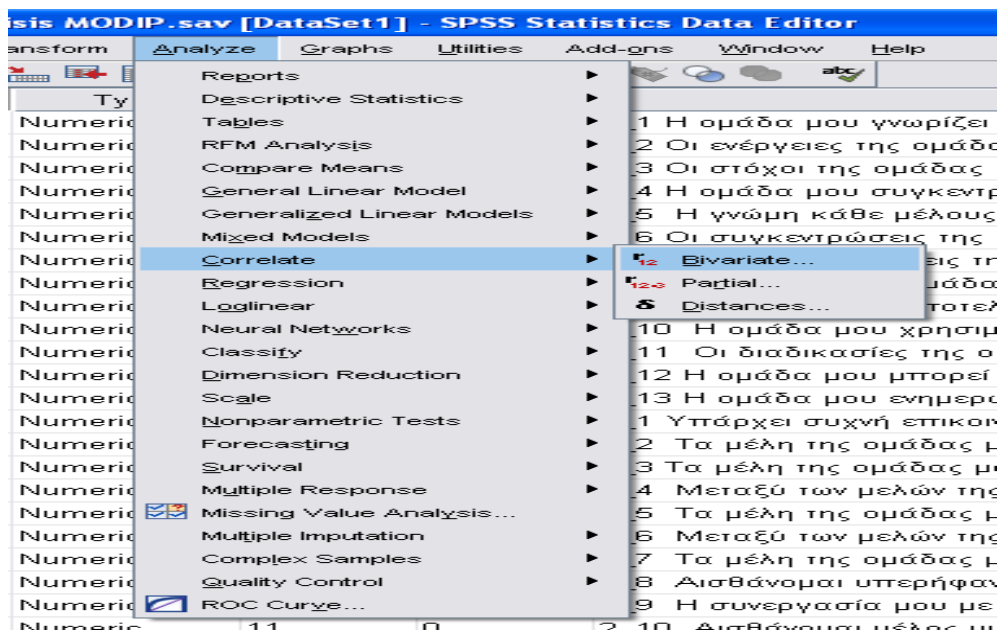
$H_1: \rho \neq 0$ ή υπάρχει γραμμική συσχέτιση μεταξύ των δύο μεταβλητών

Ο συντελεστής συσχέτισης του Pearson "χρειάζεται" την υπόθεση της κανονικότητας των δεδομένων, σε αντίθεση με τους άλλους δύο που δεν "χρειάζονται" την υπόθεση της κανονικότητας των δεδομένων. Βέβαια, για μεγάλα δείγματα, μεγέθους 30 παρατηρήσεων και πάνω και όσο το μέγεθος του δείγματος μεγαλώνει η θεωρία μας λέει ότι οι τιμές των συντελεστών "πλησιάζουν" η μία την άλλη.

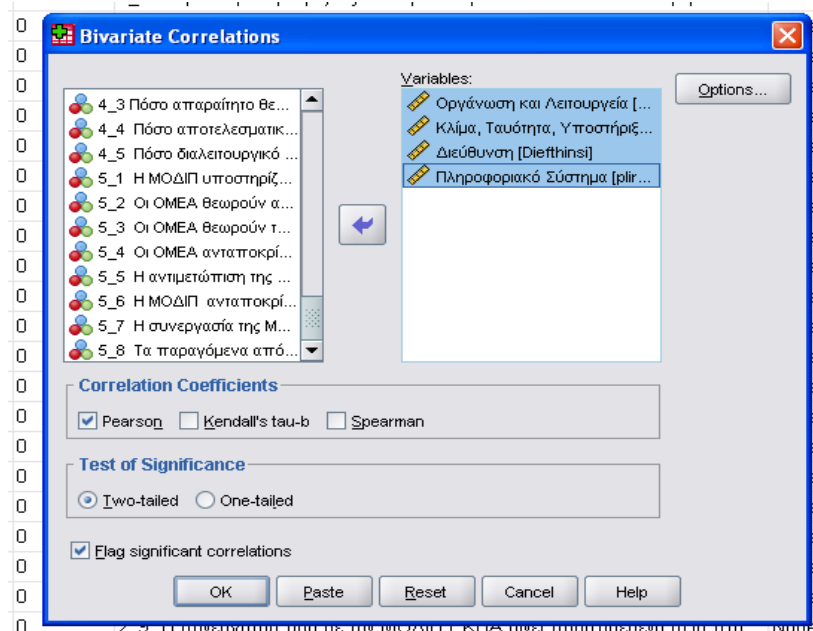
Η κύρια διαφορά των συντελεστών είναι ότι ο συντελεστής του Pearson υπολογίζεται με βάση τα δεδομένα, ενώ οι άλλοι δύο υπολογίζονται με βάση τις τάξεις μεγέθους των δεδομένων.

Ειδικότερα, ο συντελεστής του Spearman είναι ο συντελεστής του Pearson στην ουσία υπολογισμένος για τις τάξεις μεγέθους των δεδομένων. Το γεγονός λοιπόν ότι οι συντελεστές του Spearman και του Kendall υπολογίζονται με βάση τις τάξεις μεγέθους των δεδομένων είναι που επιτρέπει την ελευθερία ως προς τη μη ικανοποίηση της κανονικότητας των μεταβλητών.

Για να υπολογίσουμε τους τρεις αυτούς συντελεστές συσχέτισης στο SPSS επιλέγουμε τα εξής: **Analyze/Correlate/Bivariate** και εμφανίζεται το παράθυρο της εικόνας :



Στο δεξιό κουτάκι πρέπει να περάσουμε τουλάχιστον δύο μεταβλητές, διότι οι συντελεστές συσχέτιση υπολογίζονται για ζεύγη μεταβλητών. Οπότε αν περάσουμε περισσότερες από δύο μεταβλητές, θα υπολογιστούν οι συντελεστές γραμμικής συσχέτισης για όλα τα ζεύγη των μεταβλητών.



Βλέπουμε από την εικόνα ότι μόνο ο συντελεστής του Pearson είναι επιλεγμένος. Αν θέλουμε να εμφανιστούν και οι άλλοι δύο συντελεστές απλά τους επιλέγουμε. Παρατηρήστε ότι στο κάτω

αριστερό μέρος του παραθύρου είναι επιλεγμένη μία επιλογή (**Flag significant correlations**). Η επιλογή **Options** μας δίνει τη δυνατότητα εμφάνισης των μέσων, των τυπικών αποκλίσεων και των πληθών των τιμών για κάθε μεταβλητή. Πατώντας **OK** θα εμφανιστούν τα σχήματα:

Correlations

		Οργάνωση και Λειτουργία	Κλίμα, Ταυτότητα, Υποστήριξη	Διεύθυνση	Πληροφοριακό Σύστημα
Οργάνωση και Λειτουργία	Pearson Correlation	1	,773**	,344	,238
	Sig. (2-tailed)		,005	,300	,482
	N	11	11	11	11
Κλίμα, Ταυτότητα, Υποστήριξη	Pearson Correlation	,773**	1	,256	,534
	Sig. (2-tailed)	,005		,423	,074
	N	11	12	12	12
Διεύθυνση	Pearson Correlation	,344	,256	1	-,038
	Sig. (2-tailed)	,300	,423		,907
	N	11	12	12	12
Πληροφοριακό Σύστημα	Pearson Correlation	,238	,534	-,038	1
	Sig. (2-tailed)	,482	,074	,907	
	N	11	12	12	12

** . Correlation is significant at the 0.01 level (2-tailed).

Βλέπουμε ότι για κάποιες τιμές των συντελεστών γραμμικής συσχέτισης υπάρχουν δύο αστεράκια. Αυτό γίνεται μέσω της επιλογής **Flag significant correlations**. Οι συντελεστές συσχέτισης που υπολογίστηκαν για αυτά τα ζεύγη μεταβλητών ανίχνευσαν κάποιες στατιστικά σημαντικές συσχετίσεις μεταξύ όλων των ζευγών μεταβλητών. Κάτω από κάθε τιμή του συντελεστή συσχέτισης εμφανίζεται μία p-value (**Sig. (2-tailed)**). Η p-value που έχει υπολογιστεί για κάθε συντελεστή ξεχωριστά αναφέρεται στον έλεγχο της υπόθεσης ότι στο συγκεκριμένο ζεύγος μεταβλητών δεν υπάρχει γραμμική συσχέτιση (δηλαδή ότι ο συντελεστής

συσχέτισης για το ζεύγος είναι ίσος με το μηδέν). Αφού το παρατηρηθέν επίπεδο στατιστικής σημαντικότητας είναι μικρότερο του 0.05, συμπεραίνουμε ότι αυτή η υπόθεση απορρίπτεται σε $\alpha=0.05$. Άρα υπάρχει στατιστικά σημαντική γραμμική συσχέτιση μεταξύ του ζεύγους. Στην περίπτωση που η p-value είναι μικρότερη του 0.01, τότε ο συντελεστής συσχέτισης εμφανίζεται με δύο αστεράκια αντί για μόνο ένα. Προσέξτε και το μήνυμα που υπάρχει κάτω από κάθε πίνακα που εξηγεί τι σημαίνουν τα δύο αστεράκια.

Σε αυτό το σημείο καλό θα ήταν να αναφέρουμε ότι ο συντελεστής του Kendall μπορεί να χρησιμοποιηθεί και στην περίπτωση που έχουμε κατηγορικές μεταβλητές οι οποίες όμως είναι υποχρεωτικά σε κλίμακα διάταξης. Είναι δηλαδή διατακτικές κατηγορικές μεταβλητές. Ακόμα να αναφέρουμε ότι με το συντελεστή γραμμικής συσχέτισης ελέγχουμε αν σε ένα ζεύγος μεταβλητών υπάρχει γραμμική συσχέτιση μόνο. Δηλαδή μπορεί να υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών, αλλά όχι γραμμικής φύσεως. Σε αυτήν την περίπτωση αυτή η σχέση που συνδέει τις δύο μεταβλητές δεν μπορεί να ανιχνευτεί με το συντελεστή γραμμικής συσχέτισης. Οπότε προσοχή στην ερμηνεία που δίνουμε στο συντελεστή συσχέτισης. Να υπενθυμίσουμε επίσης ότι η λογική με την οποία απορρίπτουμε ή όχι μία υπόθεση είναι πάντα η ίδια. Αν το παρατηρηθέν επίπεδο στατιστικής σημαντικότητας είναι μικρότερο του 0.05 η υπόθεση απορρίπτεται. Στην αντίθετη περίπτωση δεν απορρίπτεται.

6.6.3 χ^2 έλεγχος ανεξαρτησίας

Στην προηγούμενη παράγραφο είδαμε πως υπολογίζουμε το συντελεστή γραμμικής συσχέτισης για την περίπτωση ποσοτικών μεταβλητών. Τι γίνεται όμως

στην περίπτωση που έχουμε κατηγορικές μεταβλητές; Σε αυτήν την περίπτωση χρησιμοποιούμε τον χ^2 έλεγχο ανεξαρτησίας.

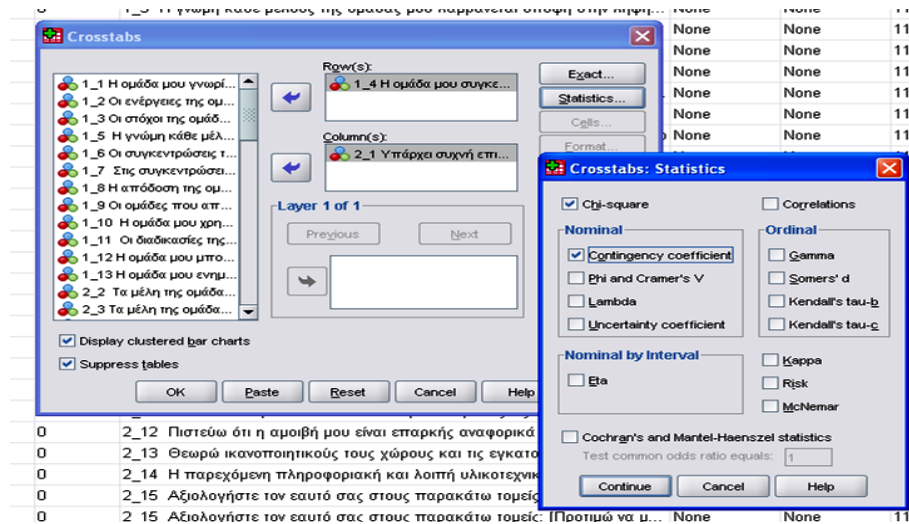
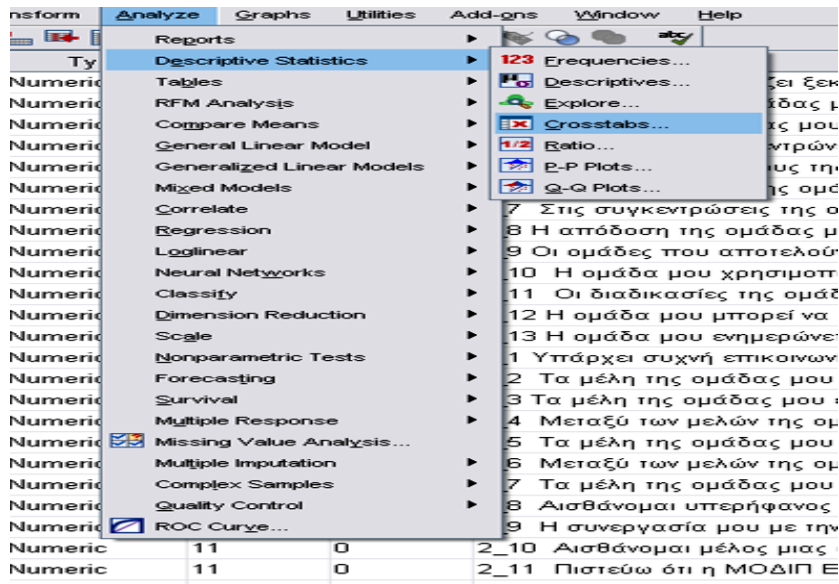
Η απαιτούμενη κλίμακα μέτρησης των μεταβλητών είναι η ονομαστική, παρόλο που και μεταβλητές με διατακτική κλίμακα μπορούν να χρησιμοποιηθούν. Ο χ^2 έλεγχος ανεξαρτησίας χρησιμοποιείται για τον έλεγχο της υπόθεσης ότι δύο κατηγορικές μεταβλητές είναι ανεξάρτητες μεταξύ τους.

Οι κατηγορικές μεταβλητές μπορούν να έχουν οσαδήποτε επίπεδα (ή κατηγορίες), αρκεί βέβαια η κάθε μία να έχει τουλάχιστον δύο επίπεδα. Όπως θα δούμε παρακάτω όταν διεξάγουμε αυτόν τον έλεγχο ανεξαρτησίας με το SPSS, θα εμφανίζεται και ένας πίνακας. Αυτός ο πίνακας θα περιέχει τις συχνότητες εμφάνισης όλων των δυνατών συνδυασμών ζευγών των επιπέδων των κατηγορικών μεταβλητών. Οι υποθέσεις σε αυτήν την περίπτωση είναι οι εξής:

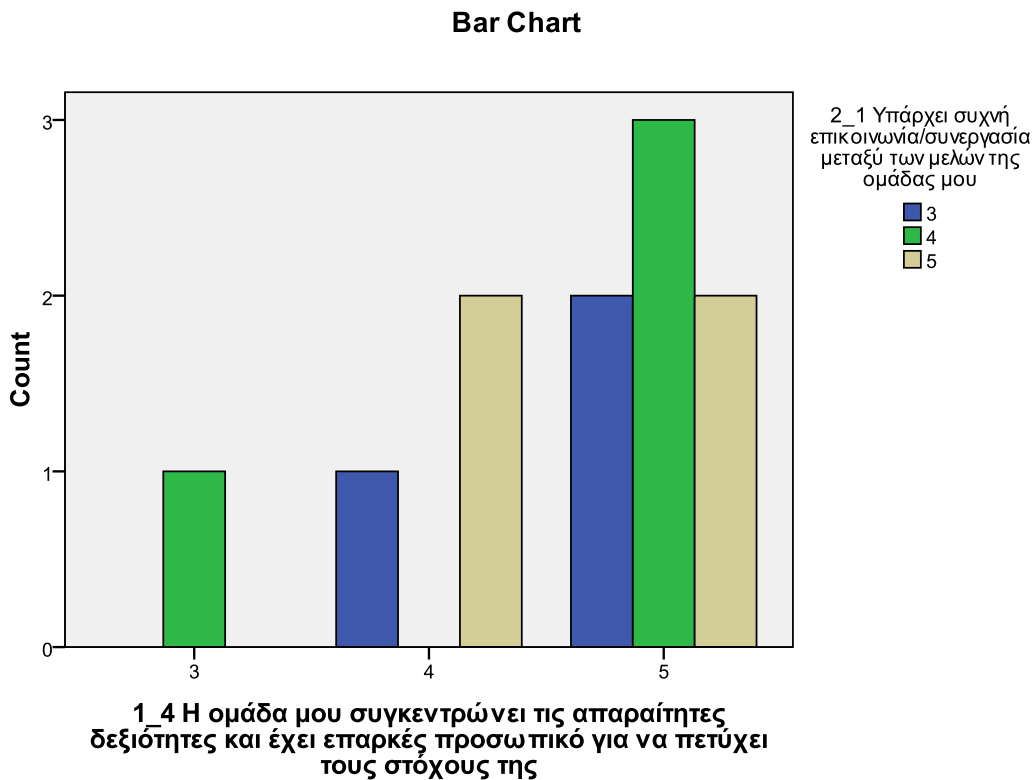
H_0 : υπάρχει ανεξαρτησία μεταξύ των δύο μεταβλητών

H_1 : δεν υπάρχει ανεξαρτησία μεταξύ των δύο μεταβλητών

Η μηδενική υπόθεση είναι πάντα αυτή που δεν υποθέτει εξάρτηση (υποθέτει ανεξαρτησία μεταξύ των μεταβλητών). Η προϋπόθεση που απαιτείται από τον χ^2 έλεγχο ανεξαρτησίας είναι οι συχνότητες των κελιών να είναι τουλάχιστον ίσες με 5. Το SPSS χρησιμοποιεί το άλλο είδος υπόθεσης που θέλει τις αναμενόμενες συχνότητες των κελιών να είναι τουλάχιστον ίσες με 5. Ένα αποδεκτό ποσοστό κελιών που θα έχουν συχνότητες μικρότερες του 5 είναι το 25%, δηλαδή το πολύ ένα στα τέσσερα κελιά να έχει μία τιμή μικρότερη του 5 χωρίς να μειώνεται σημαντικά η αποτελεσματικότητα του τεστ. Αυτό ισχύει βέβαια και για πίνακες που έχουν περισσότερα κελιά. Αν αυτή η υπόθεση δεν ικανοποιείται, τότε κοιτάζουμε την p -value που υπολογίζεται με βάση το ακριβές τεστ του Fisher (**Fisher's exact test**) ή το Monte Carlo.



Το ακριβές τεστ του Fisher θα διεξαχθεί μόνο στην περίπτωση που έχουμε 2Χ2 πίνακες όπως στο παράδειγμα. Σε αυτήν την περίπτωση το Monte Carlo δεν υπολογίζεται. Στην περίπτωση λοιπόν που έχουμε τον πίνακα 2Χ2 λόγω χάριν όπως εδώ έτοιμο, τότε πρέπει να πληκτρολογήσουμε τα δεδομένα στο SPSS Data Editor. Μέσα στις παρενθέσεις έχουμε τοποθετήσει κάποιους αριθμούς (0 και 1) για να μας διευκολύνουν στο να περάσουμε τα δεδομένα στο SPSS. Πρέπει να προσέξουμε ώστε ο κάθε συνδυασμός γραμμής και στήλης να περιέχει τον αριθμό του κελιού που πρέπει. Περνώντας τα δεδομένα στο SPSS θα έχουν την εξής μορφή (σημασία έχει ο κάθε συνδυασμός γραμμής και στήλης να περιέχει το σωστό αριθμό):



6.6.4 T-test

Οι έλεγχοι αυτοί εφαρμόζονται για ελέγχους ισότητας μέσω μεταξύ δύο δειγμάτων τα οποία είναι ανεξάρτητα (η μπορούμε να υποθέσουμε ότι είναι ανεξάρτητα). Πρέπει όμως πρώτα να προηγηθεί μία διαδικασία, να καταχωρήσουμε σε μία στήλη τις μετρήσεις που αφορούν στα δύο δείγματα και σε μία άλλη στήλη να δηλώσουμε το δείγμα από το οποίο προέρχεται η κάθε μεταβλητή. Για παράδειγμα το 1 θα δηλώνει τις τιμές της μεταβλητής που προέρχονται από το πρώτο δείγμα και με 2 τις τιμές που προέρχονται από το δεύτερο δείγμα. Τα μεγέθη των δύο δειγμάτων δεν χρειάζεται να είναι ίσα. Δεν πρέπει να ξεχνάμε ότι έχουμε πάλι την υπόθεση της κανονικότητας που πρέπει να ικανοποιείται (για το t τεστ), εκτός και αν έχουμε μεγάλο μέγεθος δείγματος. Αυτό που θέλουμε να ελέγξουμε είναι αν οι μέσοι των πληθυσμών από τους

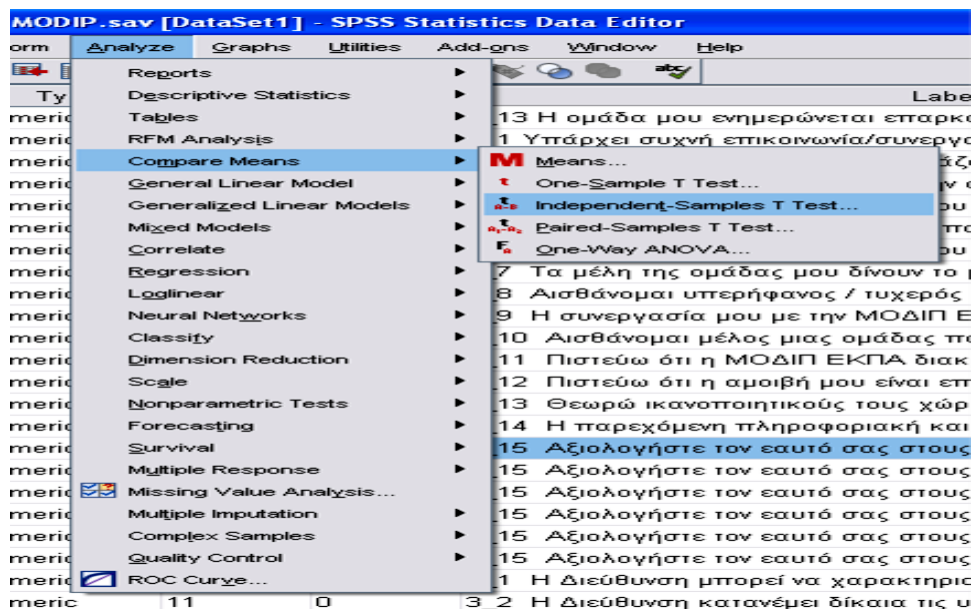
οποίος προέρχονται τα δείγματα διαφέρουν. Οι υποθέσεις διαμορφώνονται ως εξής:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

όπου μ_1 ο μέσος του πληθυσμού του πρώτου δείγματος και μ_2 ο μέσος του πληθυσμού του δεύτερου δείγματος.

Θα εργαστούμε ως εξής. Επιλέγουμε **Analyze/Compare Means/Independent-Samples T Test** και θα εμφανιστεί το παράθυρο της εικόνας :



Περνάμε τη στήλη η οποία περιέχει τις μετρήσεις και για τα δύο δείγματα στο λευκό κουτάκι δεξιά (**Test variable(s):**). Και τη στήλη που δηλώνει οι παρατηρήσεις σε ποιο δείγμα ανήκουν στο κάτω λευκό κουτάκι (**Grouping Variable:**) Δεν μπορούμε όμως να πατήσουμε **OK** ακόμη. Πρέπει να δώσουμε στο SPSS να καταλάβει ποιά μεταβλητή θα χρησιμοποιηθεί για το διαχωρισμό των τιμών της εξαρτημένης μεταβλητής. Αυτό θα γίνει πατώντας **Define Groups**

		Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference		
		F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Οργάνωση και Λειτουργία	Equal variances assumed	.	.	2,600	9	,029	1,12051	,43090	,14576	2,09527
	Equal variances not assumed	1,12051	.	.	.

Ο παραπάνω πίνακας έχει δύο γραμμές αποτελεσμάτων, η πρώτη αναφέρεται στην περίπτωση που μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων και η δεύτερη στην περίπτωση που δεν μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Ο πίνακας είναι χωρισμένος σε δύο κατηγορίες αποτελεσμάτων, η μία αφορά το **Levene** για την ισότητα των διακυμάνσεων και η άλλη περιέχει τα αποτελέσματα του t τεστ που επιλέξαμε να κάνουμε. Όπως αναφέραμε, ο πίνακας έχει δύο γραμμές αποτελεσμάτων, το αν θα κοιτάξουμε την πρώτη ή τη δεύτερη γραμμή αποτελεσμάτων του t τεστ θα μας το "πει" το τεστ του Levene. Το τεστ του Levene ελέγχει την υπόθεση της ισότητας των δύο διακυμάνσεων και υπολογίζει μία p-value. Αν η p-value είναι μικρότερη του 0.05, απορρίπτεται η υπόθεση της ισότητας των διακυμάνσεων. Στην αντίθετη περίπτωση δεν απορρίπτεται. Επομένως, ανάλογα με την p-value (Sig.) του τεστ του Levene, κοιτάζουμε την πρώτη ή τη δεύτερη γραμμή αποτελεσμάτων. Στην προκειμένη περίπτωση η p-value είναι μικρότερη του 0.05, άρα δεν μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Επομένως θα κοιτάξω τη δεύτερη

γραμμή αποτελεσμάτων του πίνακα. Η p-value για τον έλεγχο της ισότητας των δύο μέσων είναι ίση με μηδέν (Sig. (2-tailed)). Άρα η μηδενική υπόθεση απορρίπτεται, δηλαδή οι μέσοι των δύο πληθυσμών από τα οποία προήλθαν τα δύο δείγματα διαφέρουν στατιστικά σημαντικά σε επίπεδο στατιστικής σημαντικότητας $\alpha=5\%$ πάντα. Το 95% διάστημα εμπιστοσύνης για τη διαφορά των πραγματικών μέσων είναι διαφορετικό για την περίπτωση που δεν μπορούμε να υποθέσουμε ισότητα των διακυμάνσεων.

Το αντίστοιχο μη παραμετρικό ανάλογο του t τεστ είναι το τεστ των Mann-Whitney-Wilcoxon. Η διαδικασία την οποία πρέπει να κάνουμε για εκτελέσουμε αυτό το μη παραμετρικό τεστ είναι ίδια με προηγουμένως (συνένωση των δύο μεταβλητών σε μία στήλη κ.λ.π.). Οι υποθέσεις σε αυτήν την περίπτωση ορίζονται ως εξής:

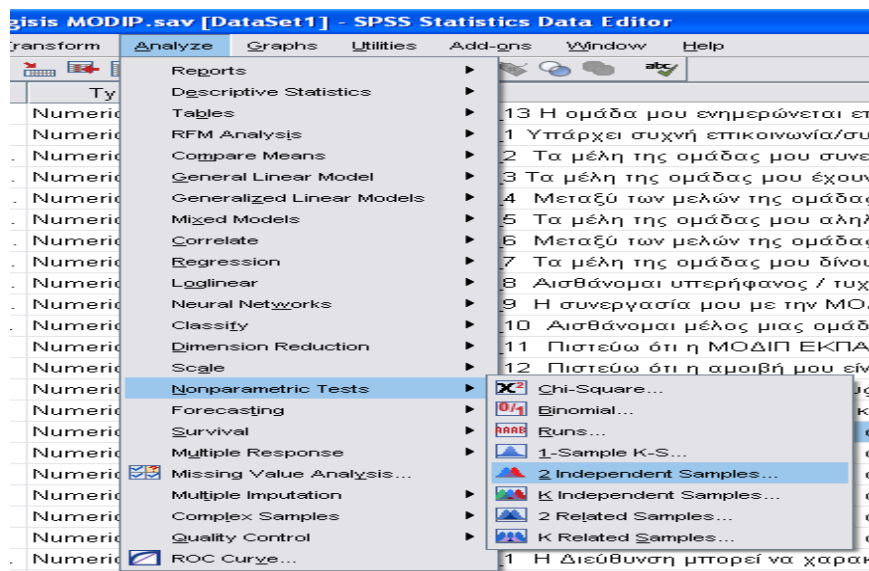
$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Επιλέγουμε

Analyze/NonparametricTests/2

Independent Samples και εμφανίζεται το παράθυρο της εικόνας :



Η διαδικασία είναι ακριβώς η ίδια με προηγουμένως όσον αφορά στη στήλη που δηλώνει τα δείγματα (**Define Groups**). Πατώντας **Exact** επιλέγουμε να εμφανιστούν τα αποτελέσματα του

Monte Carlo ελέγχου. Μόλις τελειώσουμε με όλες τις επιλογές πατάμε **OK** και εμφανίζονται στο Output δύο πίνακες, ο πρώτος περιέχει πληροφορίες για τις τάξεις μεγέθους και τα μεγέθη των δύο δειγμάτων και ο δεύτερος είναι αυτός που δίνει το αποτέλεσμα του ελέγχου.

6.6.5 Ανάλυση Διακύμανσης

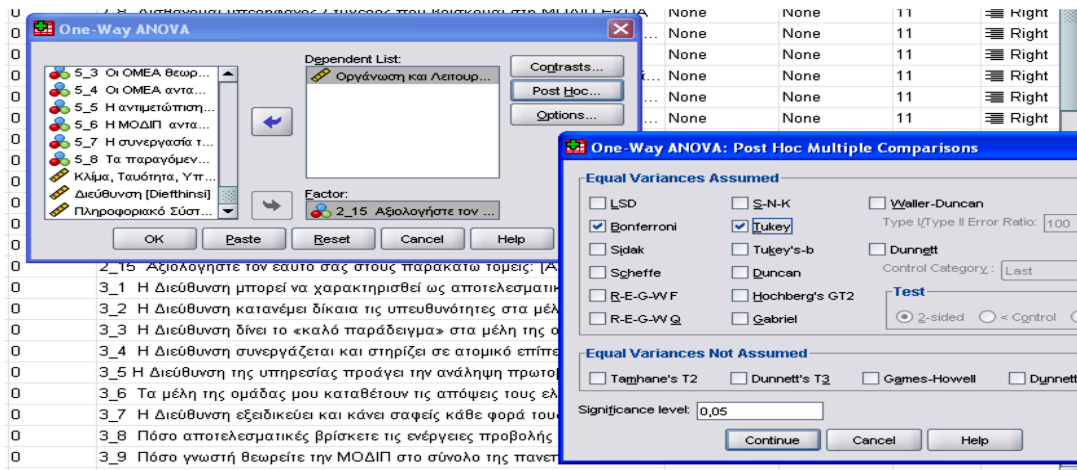
Στην περίπτωση που έχουμε δείγματα που προέρχονται από τρεις πληθυσμούς τους μέσους των οποίων θέλουμε να συγκρίνουμε χρησιμοποιούμε την τεχνική της ανάλυσης διακύμανσης κατά ένα παράγοντα. Μία εναλλακτική μέθοδος είναι χρησιμοποιήσουμε το t τεστ που είδαμε σε όλα τα πιθανά ζεύγη δειγμάτων. Με αυτόν τον τρόπο όμως μειώνουμε αισθητά την πιθανότητα να ισχύουν και οι δύο έλεγχοι ταυτόχρονα. Η ανάλυση διακύμανσης διατηρεί την πιθανότητα αυτή σταθερή με ένα μόνο τεστ. Άλλη έκφραση του προβλήματος είναι ο έλεγχος ισότητας των μέσων για μία ποσοτική μεταβλητή με παράγοντα διαφοροποίησης μία κατηγορική μεταβλητή με τρία επίπεδα. Άρα ελέγχουμε αν η κατηγορική μεταβλητή ή παράγοντας επηρεάζει την ποσοτική μεταβλητή.

Οι υποθέσεις όμως εφαρμογής της μεθόδου είναι πιο αυστηρές σε σχέση με την περίπτωση των δύο δειγμάτων. Είναι ίδιες με την περίπτωση της απλής γραμμικής παλινδρόμησης, δηλαδή κανονικότητα, ανεξαρτησία και ομοσκεδαστικότητα των καταλοίπων. Όταν λέμε ομοσκεδαστικότητα των καταλοίπων εννοούμε ότι τα κατάλοιπα που δημιουργούνται να έχουν ίσες διασπορές για κάθε επίπεδο του παράγοντα. Στην περίπτωση που δεν ισχύουν οι υποθέσεις για τα κατάλοιπα υπάρχουν στατιστικές τεχνικές (οι οποίες προσφέρονται από το SPSS) και μας βοηθάνε να εξαγάγουμε συμπεράσματα. Μπορούμε όμως και να χρησιμοποιήσουμε κάποιο

είδος μετασχηματισμού πάνω στην ποσοτική ή εξαρτημένη μεταβλητή. Τα αποτελέσματα της ανάλυσης διακύμανσης κατά ένα παράγοντα δε χάνουν την ισχύ τους όταν έχουμε μικρές αποκλίσεις από την κανονικότητα. Αν όμως δεν μπορούμε να υποθέσουμε κανονικότητα των καταλοίπων καλό θα ήταν είτε να μετασχηματίσουμε τις τιμές της εξαρτημένης μεταβλητής είτε να χρησιμοποιήσουμε το τεστ των **Kruskal-Wallis**, το μη παραμετρικό ανάλογο τεστ της ανάλυσης διακύμανσης κατά ένα παράγοντα. Στην ουσία πρόκειται για την ανάλυση διακύμανσης κατά ένα παράγοντα βασισμένο στις τάξεις μεγέθους των τιμών της εξαρτημένης μεταβλητής. Η μηδενική υπόθεση όμως που ελέγχεται με αυτό το τεστ αφορά στην ισότητα των διαμέσων και η υπόθεση που κάνουμε για τη χρήση του τεστ είναι ότι οι κατανομές των τιμών της εξαρτημένης μεταβλητής, που δημιουργούνται για κάθε επίπεδο του παράγοντα έχουν το ίδιο σχήμα. Στην περίπτωση που δεν ισχύει η περίπτωση της ομοσκεδαστικότητας αλλά η υπόθεση της κανονικότητας ικανοποιείται, μπορούμε να χρησιμοποιήσουμε το τεστ του **Welch** ή το τεστ των **Brown-Forsythe**, τα οποία είναι ανθεκτικά σε περιπτώσεις ετεροσκεδαστικότητας. Θα χρησιμοποιήσουμε τα δεδομένα των αυτοκινήτων για να ελέγξουμε κατά πόσο τα βάρη διαφέρουν από χώρα σε χώρα. Οι μηδενικές υποθέσεις ορίζονται ως εξής:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_1 : ένα τουλάχιστον ζεύγος μέσων διαφέρει



ANOVA

Οργάνωση και Λειτουργεία

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,078	2	,539	2,490	,152
Within Groups	1,515	7	,216		
Total	2,592	9			

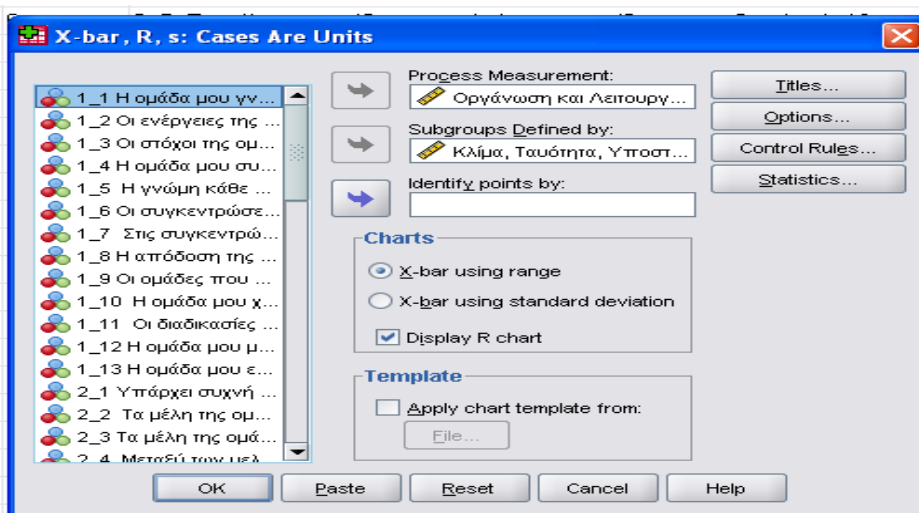
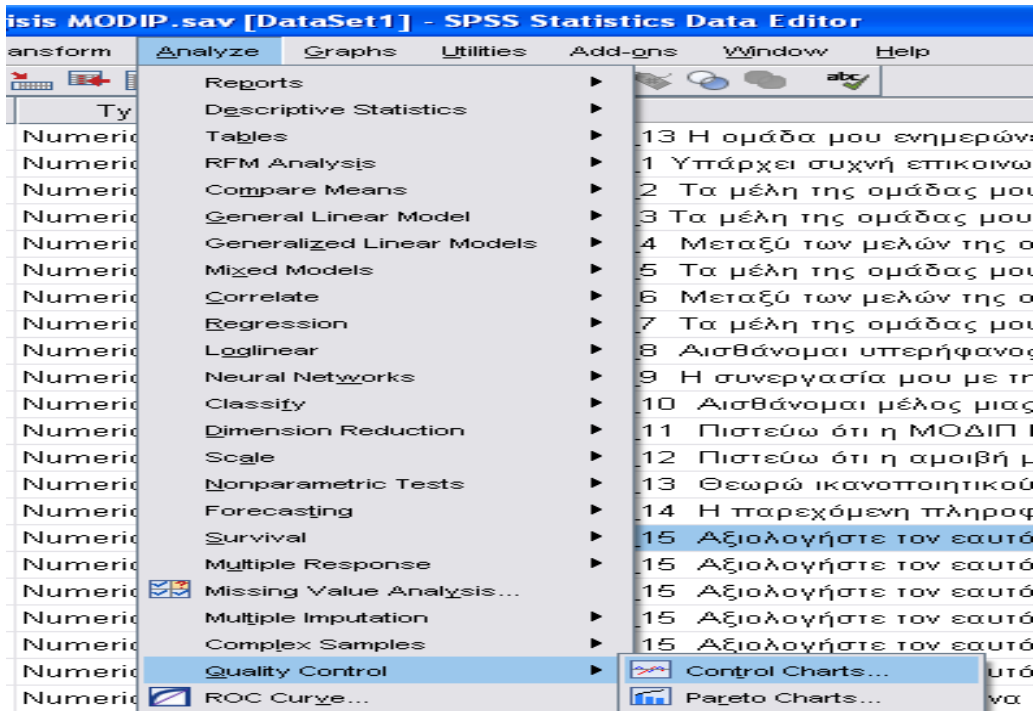
7 Κατασκευή Διαγραμμάτων έλεγχου/Control Charts/Shewart Charts

Σε κάθε παραγωγική διεργασία, ανεξάρτητα από το πόσο καλά σχεδιασμένη είναι και το πόσο προσεκτικά επιβλέπετε και συντηρείται, θα υπάρχει πάντα μια μορφή φυσικής μεταβλητότητας που θα τη συνοδεύει. Δηλαδή, όσο καλά ρυθμισμένα και να είναι τα μηχανήματα, όσο ικανοί και να είναι οι χειριστές των μηχανημάτων, όσο ικανοποιητική και να είναι η πρώτη ύλη, ποτέ δύο παραγόμενα προϊόντα δεν θα είναι τα ίδια (θα υπάρχει κάποιο μετρήσιμο μέγεθος του προϊόντος του οποίου η τιμή θα είναι διαφορετική στα δύο προϊόντα). Αυτή η φυσική μεταβλητότητα είναι το αθροιστικό αποτέλεσμα πολλών μικρών αιτιών οι οποίες αναφέρονται ως κοινές ή συνήθης ή τυχαίες αιτίες μεταβλητότητας (common or chance causes of variation). Η φυσική μεταβλητότητα είναι συνήθως μικρή σε μέγεθος και δεν μπορεί να αποδοθεί σε ελέγξιμους παράγοντες. Μια διεργασία(σύστημα) η οποία λειτουργεί μόνο με την παρουσία φυσικής μεταβλητότητας λέμε ότι είναι εντός(στατιστικού) ελέγχου διεργασία (in (statistical) control process), ή ότι λειτουργεί σε ευσταθή κατάσταση (stable state). Όμως σε μια διεργασία μπορεί να εμφανίζονται περιστασιακά και άλλες μορφές μεταβλητότητας οι οποίες δεν οφείλονται σε τυχαίες αιτίες αλλά αφορούν τη συστηματική αλλαγή στο επίπεδο κάποιου ή κάποιων παραγόντων που καθορίζουν την ποιότητα του προϊόντος. Αυτές οι μορφές μεταβλητότητας οφείλονται συνήθως στους ακόλουθους λόγους: (α) λανθασμένα ρυθμισμένες μηχανές,(β) λάθη των χειριστών των μηχανημάτων, και (γ) κακής ποιότητας ή ελαττωματική πρώτη ύλη. Η μεταβλητότητα που οφείλεται στους παραπάνω λόγους είναι σε μέγεθος πολύ μεγαλύτερη της φυσικής μεταβλητότητας και η παρουσία της οδηγεί συνήθως σε μη αποδεκτά επίπεδα λειτουργίας

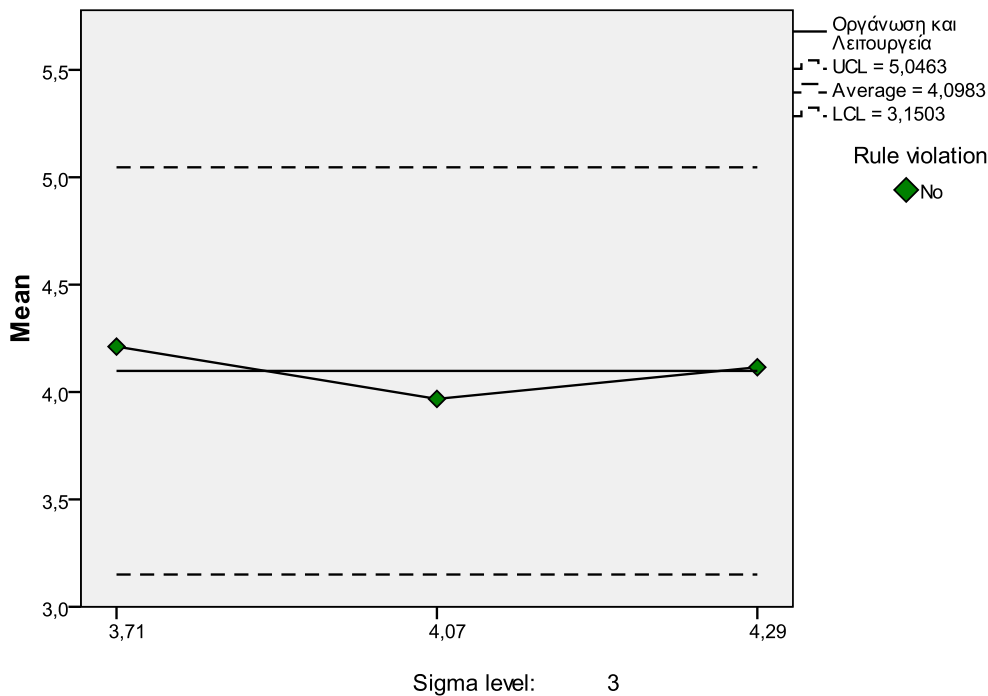
της παραγωγικής διεργασίας. Αυτή η μεταβλητότητα αναφέρεται ως ειδική μεταβλητότητα και οι αιτίες που οδηγούν σε αυτή ονομάζονται ειδικές ή προσδιορισμένες αιτίες μεταβλητότητας (special or assignable causes of variation). Μια διεργασία (σύστημα) η οποία λειτουργεί με την παρουσία ειδικής μεταβλητότητας λέμε ότι είναι εκτός (στατιστικού) ελέγχου διεργασία (out of (statistical) control process) ή ότι λειτουργεί σε ασταθή κατάσταση (unstable state). Άμεσα συνδεδεμένη έννοια με την παραγωγή ενός προϊόντος είναι τα όρια προδιαγραφών (specifications limits) των ποιοτικών χαρακτηριστικών (quality characteristics) του προϊόντος τα οποία καθορίζονται στη φάση σχεδιασμού του. Αυτά είναι το κάτω και το άνω όριο προδιαγραφών (lower and upper specification limits, LSL and USL) και εντός αυτών των ορίων πρέπει να βρίσκονται οι τιμές του ποιοτικού χαρακτηριστικού για κάθε παραγόμενο προϊόν προκειμένου να είναι ποιοτικά αποδεκτό. Επίσης, στη φάση σχεδιασμού του προϊόντος ορίζεται και μια επιθυμητή τιμή για το ποιοτικό χαρακτηριστικό που ονομάζεται τιμή στόχος (target value, T) που είναι συνήθως το μέσο του διαστήματος [LSL, USL]. Κάτω από συνθήκες φυσικής μεταβλητότητας η συντριπτική πλειοψηφία των τιμών του ποιοτικού χαρακτηριστικού στα παραγόμενα προϊόντα βρίσκεται εντός των ορίων των προδιαγραφών. Όμως κάτω από συνθήκες ειδικής μεταβλητότητας δεν ισχύει. Το κύριο αντικείμενο του Στατιστικού Ελέγχου Διεργασιών είναι η έγκαιρη ανίχνευση της εμφάνισης ειδικών αιτιών μεταβλητότητας σε μια διεργασία έτσι ώστε να προχωρήσουμε σε έρευνα και να προβούμε στις απαραίτητες διορθωτικές ενέργειες προτού κατασκευαστούν αρκετά προϊόντα μη συμμορφωμένα με τις προδιαγραφές. Τα διαγράμματα ελέγχου (control charts) είναι μια τεχνική που χρησιμοποιείται ευρέως για την ανίχνευση σε πραγματικό χρόνο της εμφάνισης ειδικών αιτιών μεταβλητότητας σε μια διεργασία (online process monitoring). Ειδικά για το προηγούμενο παράδειγμα θα μας

ενδιέφερε η κατασκευή διαγραμμάτων ελέγχου για την παρακολούθηση της μέσης τιμή και της διακύμανσης (ή της τυπικής απόκλισης) του ποιοτικού χαρακτηριστικού μέσω τυχαίων δειγμάτων από τα παραγόμενα προϊόντα.

Η κατασκευή ενός διαγράμματος ελέγχου με το SPSS Γίνεται ως εξής:



Control Chart: Οργάνωση και Λειτουργία



Για να είναι αποτελεσματικός ο Στατιστικός Έλεγχος Διεργασιών θα πρέπει να συνοδεύεται απαραίτητα με ένα εκτός ελέγχου πρόγραμμα δράσης (out of control action plan, OCAP) το οποίο θα πρέπει να ενεργοποιείται κάθε φορά που το διάγραμμα ελέγχου παρέχει ενδείξεις εμφάνισης ειδικών αιτιών μεταβλητότητας στη διεργασία. Το OCAP περιγράφει την ακολουθία των ενεργειών που πρέπει να γίνουν με σκοπό την εξάλειψη των ειδικών αιτιών μεταβλητότητας σε μια διεργασία και ο σχεδιασμός του απαιτεί τη συνεργασία ατόμων από διάφορα τμήματα της επιχείρησης.

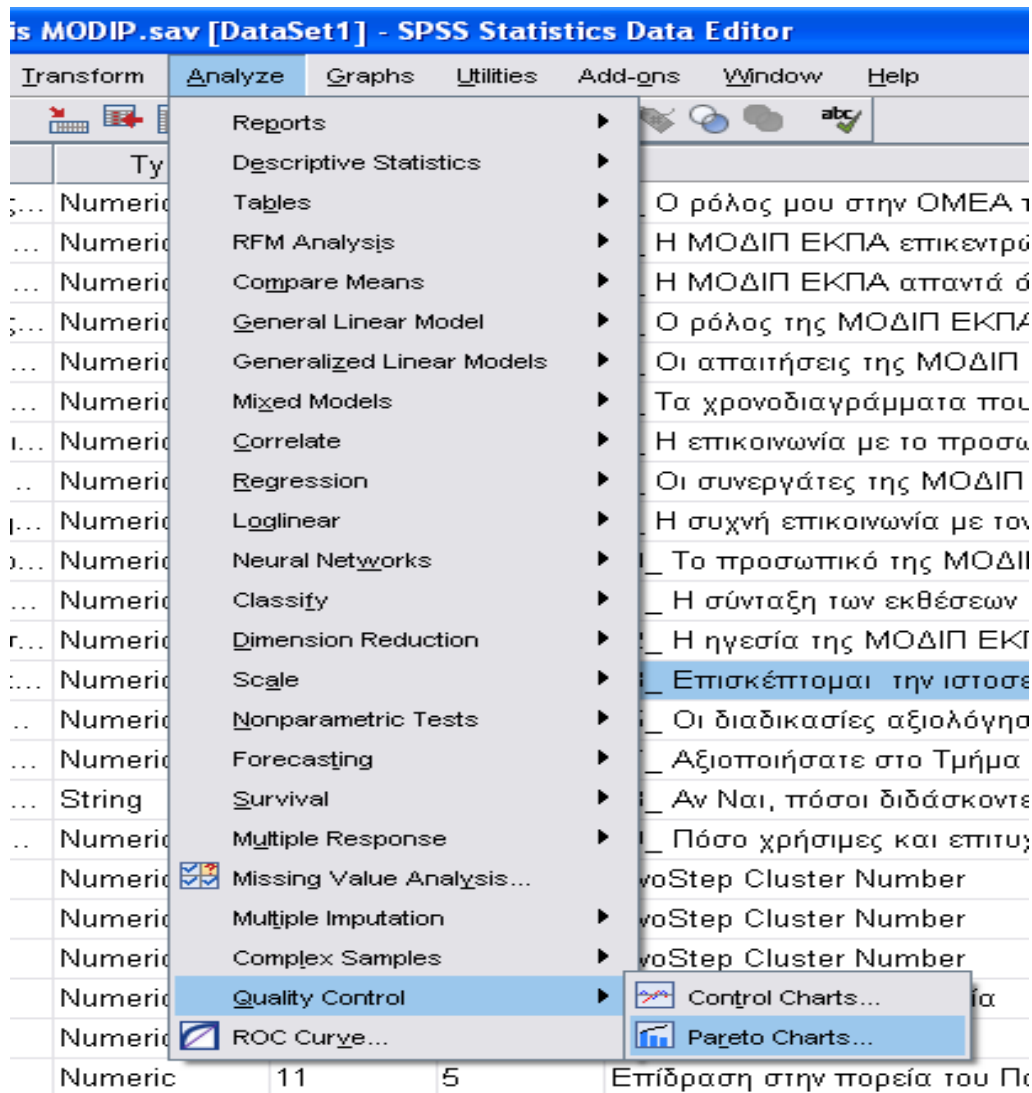
8 Κατασκευή Διαγράμματος Pareto

Το **διάγραμμα Pareto** (ή ανάλυση *Pareto*) οφείλει το όνομά του στον Ιταλό οικονομολόγο *Vilfredo Pareto* ο οποίος είναι κυρίως γνωστός για την παρατήρησή του ότι η κατανομή του 80%90% του πλούτου της χώρας του ήταν συγκεντρωμένη στο 20%10% του πληθυσμού (κανόνας 80/20, αρχή Pareto) .Συχνά εμφανίζεται η αρχή Pareto με την εναλλακτική της μορφή: «Το 80% των προβλημάτων σχετίζονται με το 20% των αιτιών» . Έτσι η ανάλυση Pareto έχει ως σκοπό να διαχωρίσει τις σημαντικές πλευρές ενός προβλήματος από τις λιγότερες σημαντικές.

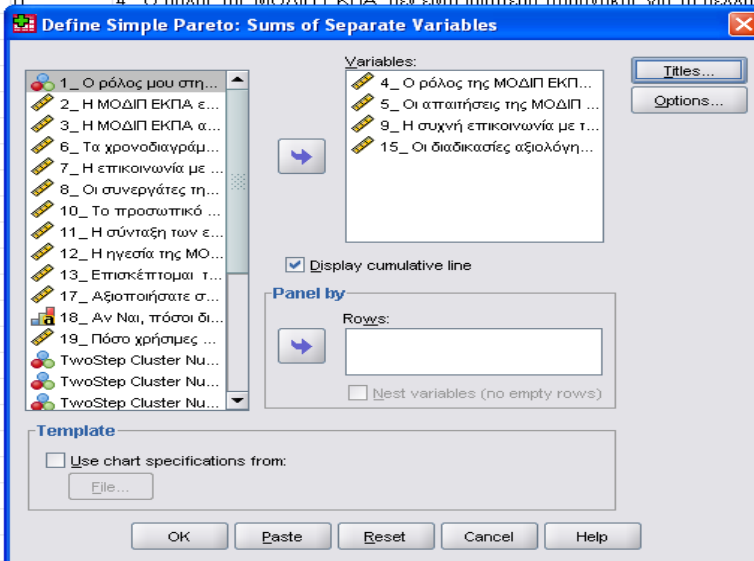
Τα βασικά βήματα που ακολουθούνται για την ανάλυση *Pareto* είναι τα ακόλουθα:·

- Καταγραφή όλων των στοιχείων·
- Μέτρηση των στοιχείων·
- Διάταξη των στοιχείων·
- Δημιουργία αθροιστικών κατανομών·
- Σχεδίαση διαγράμματος *Pareto*·
- Ερμηνεία του διαγράμματος *Pareto*

Η εξαγωγή ενός διαγράμματος Pareto από το πακέτο SPSS έχει ως εξής:



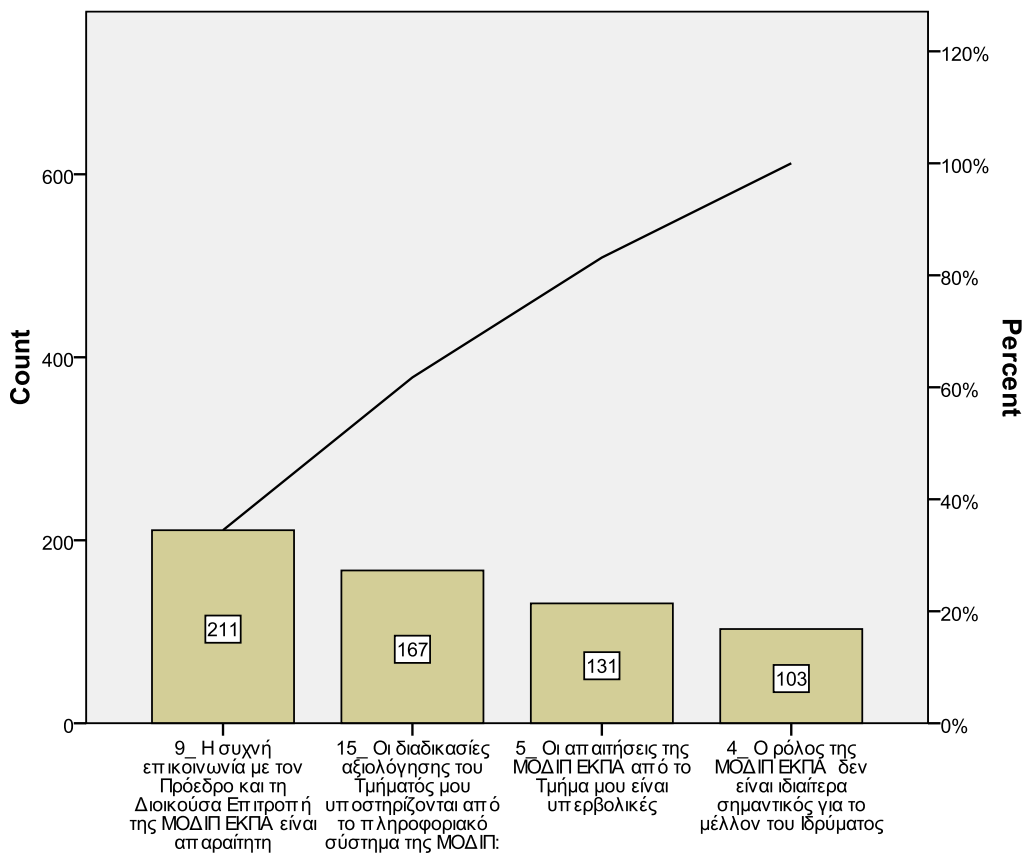
U 3_ Η ΜΟΔΙΠ ΕΚΠΑ απαντά άμεσα στα ερωτήματα και βρίσκει λύσεις στα προβλήματα τ
 P 4_ Ο ρόλος της ΜΟΔΙΠ ΕΚΠΑ δεν είναι ιδιαίτερα σημαντικός για το μέλλον του Ιδρύματος



ελεασματική
 ματός μου για
 ΔΙΠ ΕΚΠΑ εί
 μία
 η συνεργασί
 ήματα που αν
 πληροφοριακό
 που σας παρ
 όγιο φοιτητών
 ατικές ημερίδ

2 Εγκυρότητα και Λειτουργία

Ενώ ένα ενδεικτικό διάγραμμα έχει ως εξής:



9 Βιβλιογραφία

10 Βιβλιογραφία

- Bartholomew, D., Steele, F., Moustaki, E., & Galbraith, J. (2004). *The Analysis and Interpretation of Multivariate Data for Social Scientists*. Chapman & Hall/CRC.
- Basilevsky, A. (2004). *Statistical Factor Analysis and Related Methods: Theory and Applications*. New York: Wiley-Interscience.
- Billingsley, P. (1995). *Probability and Measure*. Willey Series in Statistics.
- Casella, G., & Berger, R. (2001). *Statistical Inference - Second edition*. Duxbury Advanced Series.
- Efron, B., & Tibshirani, R. (1993). *An Introduction to Bootstrap*. New York: Chapman Hall, Inc.
- Montgomery, D. C. (2008). *Introduction to Statistical Quality Control*. Wiley J. New York.
- Powers, C. H. (1987). *Vilfredo Pareto*. California: Sage Publications.
- Robert, C., & Casella, G. (2005). *Monte Carlo Statistical Methods*. New York: Springer.
- The R Project for Statistical Computing*. (1996). Retrieved 05 01, 2009, from <http://www.r-project.org/>
- Αντζουλάκος, Π. (2008). *Στατιστικός Έλεγχος Ποιότητας*. Πανεπιστήμιο Πειραιά.
- Κολυβά-Μαχαίρα, Φ., & Μπόρα-Σέντα, Ε. (1995). *Στατιστική*. Θεσσαλονίκη: Εκδόσεις Ζήτη.
- Μουστάκη, Ε. (2006). *Εισαγωγή στην Πολυμεταβλητή Ανάλυση - Σημειώσεις Μεταπτυχιακών Παραδόσεων, Τμήμα Στατιστικής Ο.Π.Α*. Αθήνα: Οικονομικό Πανεπιστήμιο Αθηνών.
- Τσιότρας, Γ. (2008). *Βελτίωση Ποιότητας*. Αθήνα: Μπένου Ε.