

Μονάδα Διασφάλισης Ποιότητας Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών

Παραδοτέο 3.2.1

Εγχειρίδιο μεθοδολογίας επεξεργασίας δεδομένων αξιολόγησης

ΠΑΚΕΤΟ ΕΡΓΑΣΙΑΣ 3: Ανἀπτυξη 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 Μι

ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ Μετη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ, ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ

Εγχειρίδιο Στατιστικής Επεξεργασίας Δεδομένων

1 Πίνακας περιεχομένων

2		Σκα	οπός.		4					
3		Нπ	ροέλ	ευση των δεδομένων	5					
4	Η δομή η οργάνωση και η τήρηση των δεδομένων6									
5	Το λογισμικό στατιστικής ανάλυσης7									
	5.	1	ΗГ	λώσσα R	8					
6		Ημ	δοθ3ι	δολογία της Ανάλυσης	11					
	6.	1	Ηει	ισαγωγή των δεδομένων (Data Entry)	11					
	6.2	2	Прс	οπαρασκευή των δεδομένων	16					
	6.3	3	Περ	ριγραφικά στατιστικά μέτρα						
	6.3.1 Τα απαιτούμενα στατιστικά μέτρα κατά περίσταση:									
	6.4 Κατασκευή κλιμάκων									
		6.4	.1	Πολυμεταβλητές τεχνικές εισαγωγικά						
		6.4	.2	Factor Analysis						
	6.5	5	Prin	cipal Component Analysis						
		6.5	.1	Εκτέλεση με το SPSS						
	6.6	5	Έλε	γχοι						
		6.6	.1	Έλεγχος κανονικότητας						
		6.6	.2	Συντελεστές συσχέτισης	45					
		6.6	.3	Χ2 έλεγχος ανεξαρτησίας						
		6.6	.4	T-test						
		6.6	.5	Ανάλυση Διακύμανσης						
7		Κα	τασκ	ευή Διαγραμμάτων έλεγχου/Control Charts/Shewart Charts						
8		Κα	τασκ	ευή Διαγράμματος Pareto						
9		Βιβ	λιογ	ραφία	66					
10)	Βιβλιογραφία								

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

δεδομἑνων

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.

МО.ΔΙ.Π. Ε.Κ.Π.А.

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 , και λοιπά ,στατιστικά ή μη πακέτα. Λαμβάνοντας Mathematica 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 γλώσα 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 υπερχείλισης 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.

🗄 modip data table 3a.sav [DataSet1] - SPSS Siatistics Data Editor 🔹 🖻 🗶										
ile <u>E</u> dit ⊻	∕iew <u>D</u> ata <u>⊺</u> r	ransform Analyze Graphs Utilities Add-ons Window Help								
> 📕 🚑	🖬 🔶 🇭	늘 📭 📴 👘 🏥 🔠 🦺 🥅 🗞 🎯 🌑 🛷								
11 : Programma_prophysia Τμήμις Γαλλικής Φιλολογίας Visible: 16 of 16 Variables										
	ID	Programma_proptyxiakon_spoudon	Sxoli	Synolikos_arithmo s_neoeisaxthento n	Eisaxthentes_me _eisagogikes_eks etaseis	Eisaxthentes_me _meteggrafes	Eisaxthentes_n _katataktiries_e 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 & Γεωπεριιβάλλοντος	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 Γλώσαας 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				
~~	1	T / T / A / A / A A A A A A A A A A A A		225	470		•			
ta View	Variabla Viaw									

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 **να***r*.

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:

e <u>E</u> dit ∖	<u>/</u> iew	Data	Iransform	Analyze	<u>G</u> raphs	<u>U</u> tilities	Add-ons Window H	elp					
		b 🔿		? 44	+∭ क	- - -	× 🖓 🌒 😽						
	<u> </u>	Nan	ne	Түре	Wi	dth Deci		Label	Values	Missing	Columns	Align	Measur
1	ID			Numeric	8	0			None	None	8	≡ Right	🛷 Scale
2	Progr	amma_	proptyxi	String	55	0	Πρόγραμμα Προπτυχ	(ιακών Σπουδών	None	None	50	≣ Left	🛷 Scale
3	Sxoli			Numeric	8	0	Σχολή		{1, Θετικών	None	8	署 Right	🖋 Scale
4	Synol	ikos_a	rithmos	Numeric	11	0	Συνολικός Αριθμός ν	:οεισαχθέντων	None	None	11	理 Right	🛷 Scale
5	Eisax	thentes	_me_ei	Numeric	11	0	Εισαχθέντες με : εισ	αγωγικές εξετάσεις	None	None	11	I Right	🛷 Scale
6	Eisax	thentes	_me_m	Numeric	11	0	Εισαχθέντες με :μετε	γγραφές	None	None	11	≡ Right	🛷 Scale
7	Eisax	thentes	_me_ka	Numeric	11	0	Εισαχθέντες με : κατ	ατακτήριες εξετάσεις	None	None	11	I Right	🛷 Scale
8	Eisax	thentes	_apo_all	Numeric	11	0	Εισαχθέντες από άλ	λες κατηγορίες	None	None	11	≡ Right	🔗 Scale
9	Synol	ikos_a	ritmos_fo	Numeric	11	0	Συνολικός αριθμός ε	γεγραμμένων φοιτητών σε όλα	None	None	11	理 Right	🛷 Scale
10	Arithn	nos_ap	ofoiton	Numeric	22	2	Αριθμός αποφοίτων κ	ατά το έτος αναφοράς	None	None	22	≡ Right	🖋 Scale
11	Meso	s_vathr	nos_pty	Numeric	22	2	Μέσος βαθμός πτυχί	ου κατά το έτος αναφοράς	None	None	22	理 Right	🛷 Scale
12	Plhth	os_pty>	iouxon	Numeric	11	0	Πλήθος πτυχιούχων	με βαθμό πτυχίου: 5.0-6.49	None	None	11	I Right	🔗 Scale
13	Plhth	os_pty>	iouxon	Numeric	11	0	Πλήθος πτυχιούχων	με βαθμό πτυχίου:6.5-8.5	None	None	11	I Right	🛷 Scale
14	Plhth	os_pty>	iouxon	Numeric	11	0	Πλήθος πτυχιούχων	με βαθμό πτυχίου:8,51-10	None	None	11	理 Right	🛷 Scale
15	Poso	sto_pou	_denolo	Numeric	22	4	Ποσοστό φοιτητών ετ	τί των εισαγομένων που ΔΕΝ ολ	None	None	22	I Right	🔗 Scale
16	Progr	amma_	spoudon	Numeric	8	0	Είδος Προγράμματος	Σπουδών	{1, Προπτυχ	None	8	I Right	🚮 Ordinal
17	1												
19													
21													
23													
24													
25													
26													
29													
30	1												
31	1												
	•	_	_	_									

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 :

🖬 modip da	ita table 3a.sav [DataSe	et1] - SPSS Statistics Data Editor					
<u>E</u> ile <u>E</u> dit <u>y</u>	<u>√</u> iew <u>D</u> ata <u>T</u> ransform	<u>A</u> nalyze <u>G</u> raphs <u>U</u> tilities Add- <u>o</u> ns <u>W</u> indow <u>H</u> elp					
📂 📕 🚔	📴 👆 秒 🔚 🖬 🛛	i? 🗚 📲 🏦 🖽 🖽 🐃 👒 🙆 🖷 🔻					
	Name	Type Width Deci Lab	el				
1	D	Numeric 8 0					
2	Programma_proptyxi	String 🔛 Variable Tune	πουδώ				
3	Sxoli	Nume					
4	Synolikos_arithmos	Nume 💿 Numeric	έντων				
5	Eisaxthentes_me_ei	Nume Comma Width: 8	ς εξετά				
6	Eisaxthentes_me_m	Nume O Dot Decimal Places: 0					
7	Eisaxthentes_me_ka	Nume O Scientific potation					
8	Eisaxthentes_apo_all	Nume	γορίες				
9	Synolikos_aritmos_fo	Nume	μένων (
10	Arithmos_apofoiton	Nume	πος αν				
11	Mesos_vathmos_pty	Nume Custom currency	το έτο				
12	Plhthos_ptyxiouxon	Nume O String	ι πτυχί				
13	Plhthos_ptyxiouxon	Nume OK Cancel Help	η πτυχί				
14	Plhthos_ptyxiouxon	Nume	πτυχί				
15	Pososto_pou_denolo	Numeric 22 4 Ποσοστό φοιτητών επί των ει	σαγομέ				
	-						

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 δεδομένα αφορούν σε ποσοτικές μετρήσεις (Scale), та 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

Mia 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 :

	Na	me	Туре	Width	Deci	Label	Values	N
1	ID		Numeric	8	0		None	None
1 3 3 4 5 5 7 3 3 3 3 3 3 3 3 4 5 5 6 7 7 8 9 9 1 2 3 4 5 5 6 6 7 7 8 9 9 9 9 9 9 9 9 9 9 9 9 9	Na ID Programn Sxoli Synolikos Eisaxther Eisaxther Eisaxther Eisaxther Synolikos Arithmos Mesos_va Pihthos_r Pihthos_r Programn	me Target Varias pososto ap Type 8 L D D D D D D D D D D D D D D D	<u>Type</u> Numeric <u>Variable</u> <u>Variable</u> <u>ble:</u> <u>ofoton_5_6</u> <u>abel</u> <u>upux Προπτυχρα</u> <u>Sxolij</u> <u>có</u> ς Αριθμός νεος <u>vreς με: εισαγ</u> <u>vreς με: μετεγγ</u> <u>vreς με: μετεγγ</u> <u>vreς με: ματοφ</u> <u>vreς ματοφ</u> <u>vreς με: ματοφ</u> <u>vreς με: ματοφ</u>	Width 8	Deci D Num (Phili e (Phili e e ornalition)	Label gric Expression: hos_phydouxon_me_vethmo_phyxiou_50_me_6.49 / Synolikos_ ola_fa_eti_kata_to_etos_anaforas)*100	Values None	None None None None None None None None
5				ок	E	easte Reset Cancel Help		
6 7								-

Το λευκό κουτάκι που λέγεται Target Variable πρέπει να συμπληρωθεί ένα όνομα. Εκεί θα αποθηκευτεί Jμ 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, μπούμε 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 υπαρχουσών 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 υπαρχουσών 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:

⊻iew Data	Transform	Analyze	<u>G</u> raphs	Utilities /	Add- <u>o</u> ns	<u>VV</u> indov	v t ab⊊∕	<u>-l</u> elp		
N	ame	Type	Wic	th Deci				Labe	el	Value
ID		Numeric	8	0						None
Progr	κατο de into μο μο το το το το το το το Σχολή [Sxoli] Εισαχθέντες μ Εισαχθέντες μ Γιλήθος πτωχ Πολήθος πτωχ Πολήθος πτωχ Γιλήθος πτωχ	Different φοτητυ με : εισ με :μετε με :μετε με :μετε με :μετε με :μετε με :μετε με :μετε φοταν φοτα		Numeric Ye Synolikos Qid and If	riable -> C arithmos_r New Valu (optional c <u>R</u> eset	es	ction o	: on> ? condition)	Cutput Variable	Vone 1, Θετικά None 1, Προπτ

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

Name	Type Width	Deci	Label	
D Recode into	Different Variable	s: Old and Nev	v Values	
Progr E Sxoli Synol © Value: Fisax		New V	alue	
Eisax <u>System-miss</u>	ing	0 Cot	y old value(s)	
Eisax Osystem-ory Eisax ORange: Synol	iser-missing		Old> New:	
Arithr through Meso			ge	
Pihthi ORange, LOW	EST through value:	Remo	ve	
Poso:	nrough niGHEST:] Output variables are strings	<u>W</u> idth: 8
Progr All <u>o</u> ther valu	ues Cont	inue Cance	Convert numeric strings to nu	mbers ('5'->5)

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 θα то Range каі 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 συνεχεία

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 Ανεξαρτήτου αντιστοιχεί. επιλογής αποθήκευσης της μετασχηματισμένης στήλης δεδομένων, αν πατήσουμε την επιλογή If θα εμφανιστεί το παράθυρο της εικόνας :

	Name	Туре	Width De	ci			Label		Values	Mis
		Numeric 8	8 0						None	None
Progr	Recode into	Different Var	iables: If C	ases					Ð	lone
xoli	n 🖉 🖉		0 0	chude ell ceru						lone
ynol	🛷 Πρόγραμμα Γ	Ιροπτυχιακ	05	skula if sees	no antiofica es	a dili anu				lone
isax	🛷 Σχολή (Sxoli)			ciude il case	sausnes co	nation				lone
isax	Συνολικός Αι Εκπινθένσες Α	ριθμός νεο								lone
isax	Εισαχθέντες μ		↔							uone
Isax	🛷 Εισαχθέντες μ	με : κατατα								lone
rithe	💉 Εισαχθέντες ο	από άλλες						Function group:		lone
laco	δυνολικός αρ	χθμός εγγε	+	< >	7 8	9		All		lone
lbtb	Μέστος βαθμά	ίς πτυχίου	-	<= >=	4 5	6		CDF & Noncentral	I CDF	lone
Ibth	🛷 Πλήθος πτυχ	ιούχων με						Conversion		lone
lhth	🛷 Πλήθος πτυχ	ιούχων με						Current Date/Time	•	lone
000	Πλήθος πτυχ	ιούχων με	1	8 1	0			Date Arithmetic		lone
mar	Είδος Ποογοά	ίματος Σ		~0	Dele	te		Eunctions and Spe	ecial Variables:	lone
rogi										ione
- 1										
_										
_										
_										
_										
_										
	L									
			0	ontinue	Cancel	H	lelp			

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 περίσταση:

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

МО.ΔΙ.Π. Ε.К.Π.А.

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 Διασφάλισης 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.

Εγχειρίδιο Στατιστικής Επεξεργασίας Δεδομένων

- 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 (Ανἁλυση 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 пю 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 μn 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 επεξηγηματικού μοντέλου συγκεκριμένο ενόc με 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 Movτέλα (Measurement Models)

МО.ΔΙ.Π. Ε.К.Π.А.

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 περισσοτέρων 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			
ΗΤΕΣ	LES)		Factor Analysis	Latent Trait Analysis	Latent Trait Analysis			
ѱҽѵдометавлн	(LATENT VARIABL	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 ὑπαρξη μιας συσχἑτισης μεταξὑ δυο δεικτών (φανερών) μπορεί να αποτελέσει σοβαρή ἐνδειξη για μια καταγωγή ή επίδραση. Όσο ακόμα παραμένει υπαρκτή μια συσχἑτιση μπορεί να υποτεθεί η ὑπαρξη 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 ΡCA

Η PCA μετασχηματίζει το σύνολο των συσχετιζόμενων x₁, x₂,....x_p σε ένα σύνολο ασυσχέτιστων μεταβλητών y₁, y₂,....y_p (τα οποία καλούνται principal components) έτσι ώστε η y₁ να εξηγεί τη μέγιστη δυνατή διακύμανση από τη συνολική διακύμανση, η y₂ να εξηγεί όσο περισσότερο γίνεται από την υπολειπόμενη διακύμανση κ.λπ.. Το σύνολο των y₁....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....

🖬 Data esot	🖬 Data esoterikis aksiologisis MODIP.sav [DataSet1] - SPSS Statistics Data Editor										
<u>Eile E</u> dit <u>Y</u>	_iew <u>D</u> ata <u>T</u> r	ansform	Analyze Graphs	Utilities	Add-	ons Window Help					
😕 📕 🚔	📴 🔸 🐡	*** 🖬	Reports		•	🐳 🕗 🜑 📲					
	Name	Ту	D <u>e</u> scriptive Stat	tistics	►	Labe					
1	@1_1Ηομά	Numerio	Tables		►	1 Η ομάδα μου γνωρίζει ξεκάθαρα τι					
2	@1_20iɛvép	Numeric	RFM Analysis		►	2 Οι ενέργειες της ομάδας μου βασίζα					
3	@1_30iστó	Numeric	Compare Means	5	►	3 Οι στόχοι της ομάδας μου ευθυγραι					
4	@1_4Ηομά	Numerio	<u>G</u> eneral Linear	Model	►	4 Η ομάδα μου συγκεντρώνει τις απαι					
5	@1_5Ηγνώμ	Numerio	Generali <u>z</u> ed Lin	ear Models	►	5 Η γνώμη κάθε μέλους της ομάδας ι					
6	@1_6Οισυγ	Numerio	Mi <u>×</u> ed Models		►	6 Οι συγκεντρώσεις της ομάδας μου ε					
7	@1_7Στιςσυ	Numeric	<u>C</u> orrelate		►	7 Στις συγκεντρώσεις της ομάδας μοι					
8	@1_8Hαπό	Numeric	Regression		►	8 Η απόδοση της ομάδας μου εξαρτά					
9	@1_9Οιομά	Numeric	Loglinear		•	9 Οι ομάδες που αποτελούν την ΜΟΔ					
10	@1_10Hoµ	Numeric	Neural Net <u>w</u> ork	s	►	10 Η ομάδα μου χρησιμοποιεί συστη					
11	@1_110iði	Numerio	Classi <u>f</u> y		►_	11 Οι διαδικασίες της ομάδας μου ετ					
12	@1_12Hop	Numeric	Dimension Redu	uction	•	🔏 Eactor κτηρ					
13	@1_13Hop	Numeric	Sc <u>a</u> le		•	Correspondence Analysis αρκι					
14	@2_1Υπάρ	Numeric	<u>N</u> onparametric	Tests	► :	🛱: Optimal Scaling εργα					
15	@2_2Ταμέλ	Numerio	Forecasting		• • `	2 Τα μέλη της ομάδας μου συνεργάζ					
16	@2 3Ταμέλ	Numerio	Survival		•	3 Τα μέλη της ομάδας μου ένουν την ι					

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

Εγχειρίδιο Στατιστικής Επεξεργασίας Δεδομένων

			-		,
ΙΗομά	Numeric	11	0	11Η ομάδ	ια μου ννωρίζει ξεκάθαρα τι πρόκειται να κάνει κάθε φορά None
2Οιενέρ	Factor Ana	lysis			× Viene in the second se
3Οιστό.					Variables:
4Ηομά	💑 2_1 Υπάρχ	ει συχνή επι	οινωνία/συνεργα	ασία 📤	1_1 Η ομάδα μου γνωρίζει ξεκάθαρα τι πρόκειτα
5Ηγνώμ	💑 2_2 Τα μέλ	η της ομάδας	μου συνεργάζον	/100	💑 1_2 Οι ενέργειες της ομάδας μου βασίζονται σε Εxtraction
6Οισυγ.	💰 2_3 Τα μέλι	η της ομάδας	μου έχουν την ατ	παρ	💑 1_3 Οι στόχοι της ομάδας μου ευθυγραμμίζονται 🛛 🛛 Rotation
Στιςσυ	💑 2_4 Μεταξί	ύ των μελών	της ομάδας μου υ	πά	Δ 1_4 Η ομάδα μου συγκεντρώνει τις απαραίτητες
3Ηαπό.	δ 2_5 Τα μέλ	.η της ομάδας	μου αληλοϋποσ	πηρ	Δ 1_5 Η γνώμη κάθε μέλους της ομάδας μου λαμ
3Οιομά.	Φ. 2_6 Μεταξι Δ. 2 7 Τα μέλ	υ των μελων ΄ η της ουάδας	της ομασας μου υ 	ττα	Ο Τ_5 ΟΙ συγκεντρωσεις της ομασάς μου είναι άπ Ορτίοπε
10Hou	2_7 π με/	νομαι μπεοή	ι μου οινουν το μα ιανος / τυχεοός π	του.	
110-18	2_9 H συν	εργασία μου μ	με την ΜΟΔΙΠ ΕΚΙ	ΠA	Α 1_9 Οι ομάδες που αποτελούν την ΜΟΔΙΠ συνε
10101	💑 2_10 Alor8i	άνομαι μέλος	μιας ομάδας που	າໜ 🚺 🚺	🧾 💑 1_10. Η ομάδα μου χρησιμοποιεί συστηματικές
2H0µ	💰 2_11 Πιστε	ύω ότι η ΜΟΔ	μΠ ΕΚΠΑ διακρίνα	εται	💑 1_11. Οι διαδικασίες της ομάδας μου επιτελούντ
ізноµ	💑 2_12 Πιστε	ύω ότι η αμοι	3ή μου είναι επαρ	οκή	🢑 1_12 Η ομάδα μου μπορεί να χαρακτηρισθεί ως
ΙΥπάρ.	💑 2_13 Θεωρ	ώ ικανοποιη	τικούς τους χώρο	υς	🂑 1_13 Η ομάδα μου ενημερώνεται επαρκώς για τι
2Ταμέλ.	2_14 Η πα	ρεχομενη πλ	ηροφοριακή και /	VOI	
3Ταμέλ.	2_15 AGO/		αυτό σας στους τ	пар тар	
Μεταξ.	2 15 ASIO	ονήστε τον ε	αυτό σας στους τ	παρ	
Ταμέλ.	💑 2_15 Αξιολ	ογήστε τον ε	αυτό σας στους τ	παρ	
ίΜεταξ.	💰 2_15 Αξιολ	ογήστε τον ε	αυτό σας στους τ	παρ	
7 ταμέλ.	🕹 2_15 Αξιολ	ογήστε τον ε	αυτό σας στους τ	παρ	Selection Variable:
3Αισθά.	💑 3_1 Η Διεύ	θυνση μπορ	εί να χαρακτηρισθ	θεί 📃 🛄	
Энтиха	💑 3_2 Η Διεύ	θυνση καταν	έμει δίκαια τις υπ	ευθ 👻	Value
ΠΑισθ			ОК	Paste	Reset Cancel Help
11 Durga					
	. Hannone		-	2_11 - 11010	· · · · · · · · · · · · · · · · · · ·

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.

$MO.\Delta I.\Pi. \ E.K.\Pi.A.$

~~~~~		P00 P0
την g	μπαο 💫 1.3.Οι στόχοι της ουάδας μο	υ ευθυ
μου	🖬 Factor Analysis: Rotation 🛛 🛛 🔀	ντει τις
οÜΠα		ης ομά
μου	Method	ιάδας ι
√ то		ομάδο
ερός		μου εξ
키니 E		ύν την
:ς <b>π</b> ι	O Direct Oblimin O Promax	τοιεί σι
ιακρ	Detta: 0 Kappa 4	δας μο
:ιεπ		ι χαραι
χώρ	Display	εται ετ
ήκα	Rotated solution	
TOU		
TOU	Maximum Iterations for Convergence: 25	
TOU	mazimam ter ations for convergence. 23	
τους	Continue Cancel Help	
TOU		
τους	; παρ Selection Variable:	

## Αποτελἑσματα:

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

	In	itial Eigon	values	Extract	tion Sums	of Squared	Rotation Sums of Squared			
	111	illai Eigenv	alues		Luauing	5		Luauing	15	
		% of	Cumulative		% of	Cumulative		% of	Cumulative	
Component	Total	Variance	%	Total	Variance	%	Total	Variance	%	
1	5,760	44,310	44,310	5,760	44,310	44,310	3,978	30,599	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	u .						
6	,514	3,953	97,141	u .						
7	,349	2,685	99,826							
8	,023	,174	100,000							
9	8,897E-	6,844E-	100,000							
	17	16		u						
10	-	-2,984E-	100,000							
	3,879E-	18								
	19									

#### **Total Variance Explained**

#### Εγχειρίδιο Στατιστικής Επεξεργασίας Δεδομένων

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 συνυπολογίζεται 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									
		Comp	onent						
	1	2	3	4					
1_1 Η ομάδα μου γνωρίζει ξεκάθαρα	,873								
τι πρόκειται να κάνει κάθε φορά									
1_2 Οι ενέργειες της ομάδας μου	,912								
βασίζονται σε ένα ξεκάθαρο βασικό									
σκοπό της υπηρεσίας									
1_3 Οι στόχοι της ομάδας μου				,866					
ευθυγραμμίζονται με τους στόχους									
της υπηρεσίας									
1_4 Η ομάδα μου συγκεντρώνει τις			,889						
απαραίτητες δεξιότητες και έχει									
επαρκές προσωπικό για να πετύχει									
τους στόχους της									
1_5 Η γνώμη κάθε μέλους της	,829								
ομάδας μου λαμβάνεται υπόψη στην									
λήψη αποφάσεων									
1_6 Οι συγκεντρώσεις της ομάδας		,838							
μου είναι αποτελεσματικές και									
παράγουν έργο									
1_7 Στις συγκεντρώσεις της ομάδας	,902								
μου το κάθε μέλος συμμετέχει ενεργά									
και διατυπώνει ελεύθερα τις απόψεις									
тоџ									
1_8 Η απόδοση της ομάδας μου			,907						
εξαρτάται από την απόδοση του									
εκάστοτε ατομικού έργου									
1_9 Οι ομάδες που αποτελούν την				,868,					
ΜΟΔΙΠ συνεργάζονται ικανοποιητικά									
μεταξύ τους									
1_10 Η ομάδα μου χρησιμοποιεί		,650	,617						
συστηματικές διαδικασίες λήψης									
αποφάσεων και επίλυσης									
προβλημάτων									
1_11 Οι διαδικασίες της ομάδας μου			,539						
επιτελούνται στον συμφωνημένο									
χρόνο									
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		,738				
επαρκώς για τις αλλαγές και τις						
εξελίξεις στο αντικείμενο εργασίας						
της και συμμετέχει σε συνέδρια και						
λοιπές σχετικές δραστηριότητες						
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 εναπομείνασες 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 acronbach, δηλαδή την αξιοπιστία της κλίμακας:

Name	Туре	Width	Decimal	s	Label		Values	Missir
@1_1Ноµά	Numeric	11	0	1_1 Η ομάδα μου γνωρίζει ξεκ	άθαρα τι πρόκειτα	ιι να κάνει κάθε φορό	i None	None
@1_2(	bility Anal	ue ie				να ξεκάθαρο βασικό	σκ None	None
@1_3(	ibility Allat	ys15				με τους στόχους της	υπ None	None
21_4			te te	ems:	Statistics	δεξιότητες και έχει ετ	παρ None	None
01_5F 🕺 1_3	3 Οι στόχοι της	ομάδας μο	•	1_1 Η ομάδα μου γνωρίζει ξεκά		άνεται υπόψη στην λ	ήψη None	None
21 60 🕈 🕂	ι Η ομασα μου	συγκεντρ				τολοπικατικός και παι	οάμ Νορο	None
21 7	5 Οι συγκεντρ 3 Η απόδοση	ασεις της ο της ομάδας		ο 1_5 Η γνωμη καθε μελούς της	🖬 Reliability A	nalysis: Statistics		
21_81 💑 1_1	9 Οι ομάδες π	ου αποτελ		🔓 1_12 Η ομάδα μου μπορεί να χ	CDescriptives f	or	Inter-Item	
21_9 👷 '-	10 Πομασαιμί 11 Οιδιαδικας	τές της ομά			I tem		Correlations	
91_10 了 -	Ι 3Η ομάδα μο	υ ενημερώ			Scale		Covariances	
01_11 🗟 2 ·	Ι Υπάρχει συν	νή επικοιν 🖪	1 L		Scale if item	dalated		
21_12 Modet	Alpha	-			Jocgie in Reini	ucicieu		
21_13 Scale la	bet TEST	NG			Summaries -		ANOVA Table	
2_1					<u>M</u> eans		None	
92_2		ок	Paste	Reset Cancel Help	□ ⊻ariances		◯Etest	
2_31 upen	Numeric		0	2_3 to pend till obroad house	Covariances		O Friedman chi-squar	e
)2_4Μεταξ…	Numeric	11	0	2_4 Μεταξύ των μελών της ομ			Cochran chi-squar	
@2_5Ταμέλ	Numeric	11	0	2_5 Τα μέλη της ομάδας μου			0.000241.011.0444	
02_6Μεταξ…	Numeric	11	0	2_6 Μεταξύ των μελών της ομ	Hotelling's T-s	quare	Tukey's test of additi	vity
02_7Ταμέλ	Numeric	11	0	2_7 Τα μέλη της ομάδας μου	Intraclass cor	relation coefficient		
2_8Αισθά	Numeric	11	0	2_8 Αισθάνομαι υπερήφανος				
2_9Ησυνε	Numeric	11	0	2_9 Η συνεργασία μου με την	Mo <u>d</u> el: Tv	vo-vvay Mixed 👻 1	yge: Consistency	
2_10Aισθ	Numeric	11	0	2_10 Αισθάνομαι μέλος μιας α	Confidence in	iterval: 95 % 1	fest val <u>u</u> e: 0	
)2_11Πιστ	Numeric	11	0	2_11 Πιστεύω ότι η ΜΟΔΙΠ Ε		Cancel	Help	
)2_12Πιστ	Numeric	11	0	2_12 Πιστεύω ότι η αμοιβή μο		cancel	nah	

#### МО.ΔΙ.Π. Ε.К.П.А.

Reliability Statistics						
Cronbach's Alpha	N of Items					
,903	5					

Αποδεκτές τιμές Αξιοπιστίας είναι οι τιμές του a>0,7.

Παρἁλληλα εξετἁζουμε 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 ποιο συχνούς ελέγχους που εμφανίζονται στην περίπτωση της Δ.Π.

## 6.6.1 Έλεγχος κανονικότητας

Αναφέραμε προηγουμένως πως όταν то 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, то P-P Plot каі то **Q-Q** Plot (Επιλέγοντας **Analyze** $\rightarrow$ **Descriptive Statistics** $\rightarrow$ **P-P Plots**  $\dot{\eta}$  **Q-Q Plots**). M $\epsilon$ 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 **Ho**), η εναλλακτική υπόθεση (**Alternative Hypothesis** ή **H**₁), το επίπεδο στατιστικής σημαντικότητας (**a**) και το παρατηρηθέν επίπεδο στατιστικής σημαντικότητας (**p-value** ή **Significance**). Οι υποθέσεις είναι της ακόλουθης μορφής:

## Ηο: Η κατανομή των δεδομένων δε διαφέρει από την κανονική κατανομή

# Η₁: 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** $\rightarrow$ **Nonparametric Tests** $\rightarrow$ **1-Sample K-S** και εμφανίζεται το παράθυρο της εικόνας :

#### Εγχειρίδιο Στατιστικής Επεξεργασίας Δεδομένων

	Name	Туре	Width	Decimals	Label
@1_	1Ноμά	Numeric	11	0	1_1 Η ομάδα μου γνωρίζει ξεκάθαρα τι πρόκειται να κάνει κάθε φ
@1_	🖬 One-	Sample Kolmo	gorov-Smiri	nov Test	βαι
@1_ @1_ @1_ @1_ @1_ @1_ @1_ @1_		Πόσο διαλειτουρ Η ΜΟΔΙΠ υποστ Οι ΟΜΕΑ θεωροι Οι ΟΜΕΑ θεωροι Οι ΟΜΕΑ ανταπι Η αντιμετώπτση Η ΜΟΔΙΠ ανταπι Η συνεργασία τη Τα παραγόμενα	γικό θεωρείτε τ ηρίζει αποτελεσμ ύν αποτελεσμ ύν την ΜΟΔΙΠ ι οκρίνονται έγκυ της ΜΟΔΙΠ απο οκρίνεται έγκυ ις ΜΟΔΙΠ με τις από την ΜΟΔΙ	ο σμ xτικ και φ ρα τ.υ Π ( ▼	Τest Variable List:       έχ         Οργάνωση και Λειτουργεία [Organosi_Leit]       Έχαct         Κλίμα, Ταυότητα, Υποστήριξη [Klima_Taft]       Οptions         Διεύθυνση [Diefthinsi]       τέχ         Τέχ       Πληροφοριακό Σύστημα [piroforiako_systi]       τέχ         ψη       γη
@1_ @1_ @2_	Test D	vistribution	n ential		
@2_ @2_ @2	THERE	Nomenc	ок	Paste	Reset Cancel Help Yio
@2	 5Ταμέλ	Numeric	11	0	2_5 Τα μέλη της ομάδας μου αληλοϋποστηρίζονται και ζητούν ελ
@2	6Μεταξ	Numeric	11	0	2 6 Μεταξύ των μελών της ομάδας μου υπάρχει εμπιστοσύνη

Παρατηρούμε ότι η επιλογή για τον έλεγχο κανονικότητας είναι ήδη προεπιλεγμένος από το SPSS (Normal). Επιλέγοντας **Options** εμφανίζεται ένα άλλο παράθυρο στο οποίο μπορούμε να επιλέξουμε και την εμφάνιση ενός πίνακα με κάποια περιγραφικά μέτρα που αφορούν αυτές τις μεταβλητές. Πατώντας **Exact** θα εμφανιστεί το παράθυρο της εικόνας 28. Το SPSS έχει ως προεπιλογή το **Asymptotic only**. Αυτό σημαίνει ότι θα διεξάγει το τεστ κανονικότητας των Kolmogorov-Smirnov όπως επιλέξαμε άλλωστε. Αν επιλέξουμε την επιλογή που βρίσκεται ακριβώς από κάτω, δηλαδή το **Monte Carlo**, θα ενεργοποιηθούν και τα επόμενα δύο λευκά κουτάκια, το Confidence level και το Number of samples. Με την επιλογή Monte Carlo "ζητάμε" από το SPSS va χρησιμοποιήσει και την τεχνική της προσομοίωσης για να κάνει τον έλεγχο της κανονικότητας. Δε θα επεκταθούμε περισσότερο στην τεχνική της προσομοίωσης, παρά μόνο θα πούμε ότι διεξάγει 10000 (προεπιλογή) τεστ κανονικότητας και για κάθε ένα υπολογίζει την pvalue. Στο τέλος εμφανίζει το μέσο όρο αυτών των 10000 p-value

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



## 6.6.2 Συντελεστές συσχέτισης

Or συντελεστές пои 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 άλλη 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 (<u>σε απόλυτη τιμή πάντα</u>), 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:

Ho: ρ=0 ἡ δεν υπἀρχει γραμμικἡ συσχἑτιση μεταξὑ των δύο μεταβλητών

# Η₁: ρ≠0 ἡ υπἀρχει γραμμικἡ συσχἑτιση μεταξὑ των δὑο 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 :

#### МО.ΔΙ.Π. Ε.К.П.А.

isis MOD	IP.sav [Da	ataSet1]	- SPSS S	tatis	tics I	Data Edit	tor	
ansform	<u>A</u> nalyze	<u>G</u> raphs	Utilities	Add	- <u>o</u> ns	<u>vv</u> indov	/ E	telp
<b>****</b> 🖬	Report	ts		►	<b>S</b>	ے 👟	abc	
Ту	D <u>e</u> scri	iptive Statis	tics	►				
Numeric	Ta <u>b</u> les	\$		►	1 H	ομάδα μ	10U Y	νωρίζει
Numeric	REM A	nalysįs		►	20	ι ενέργει	ες τη	ς ομάδι
Numeric	Compa	are Means		►	30	ιστόχοι	της σ	ομάδας
Numeric	<u>G</u> ener	al Linear M	odel	►	4 H	ομάδα μ	100 0	Γυγκεντρ
Numeric	Gener	ali <u>z</u> ed Line	ar Models	►	<u>5</u> ⊢	Η γνώμη κ	κάθε	μέλους
Numeric	Mi⊻ed	Models		►.	60	ισυγκενι	τρώσ	εις της
Numeric	<u>C</u> orrela	ate		►	52 E	<u>∃</u> i∨ariate…		EIG TR
Numeric	<u>R</u> egre:	ssion		▶	12-3	Pa <u>r</u> tial		μάδα
Numeric	Logline	ear		▶	δ	≥istances.		тоте?
Numeric	Neural	l Net <u>w</u> orks		▶ `	10	Η ομάδο	ίμου	χρησιμ
Numeric	Classi	Íy		►	11	Οι διαδικ	<ασí:	ες της ο
Numeric	Dimen:	sion Reduc	tion	►	12 F	Η ομάδα	μου	μπορεί
Numeric	Sc <u>a</u> le			►	13 F	Η ομάδα	μου	ενημερι
Numeric	<u>N</u> onpa	rametric Te	ests	►	1 Y	πάρχει σ	ruxvŕ	ί επικοι
Numeric	Foreca	asting		►	2 T	α μέλη τ	ης ο	μάδας μ
Numeric	<u>S</u> urviv	al		►	ЗТо	α μέλη τη	ης σμ	ιάδας μι
Numeric	M <u>u</u> ltiple	e Respons	e	►	_4 ∿	4εταξύ τα	ων με	ελών της
Numeric	🔀 Missin	g Value Ar	nal <u>v</u> sis		<u>5</u> T	α μέλη τ	ης ο	μάδας μ
Numeric	Multiple	e Imputatio	<b>-</b>	►	_6 N	4εταξύ τα	ων με	ελών της
Numeric	Comple	ex Sample:	s	►	7 T	α μέλη τ	ης ο	μάδας μ
Numeric	Quality	/ Control		►	<u>8</u> A	ισθάνομ	αιυτ	τερήφαι
Numeric	📶 ROC C	≎ur <u>v</u> e			9 ⊢	Ι συνεργα	ασία	μου με
Numeric	. 11		Ο	2	10	Δισθάνοι		iéžoz III

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.

0	🔛 Bivariate Correlations 🛛 🔀	1
0		1
0	Variables:	
0	💑 4_3 Πόσο απαραίτητο θε 🛆 🔗 Οργάνωση και Λειτουργεία [	
0	🧼 4_4. Πόσο αποτελεσματικ	
0	ου του οιαχειτοργίαστης του	
0	💑 5_2 Οι ΟΜΕΑ θεωρούν α	
0	💑 5_3 Οι ΟΜΕΑ θεωρούν τ	
0	ο 5_4 Οι ΟΜΕΑ ανταποκρι	
0		
0	💑 5_7 Η συνεργασία της Μ	
0	💑 5_8 Τα παραγόμενα από 💌	
0	Correlation Coefficients	
0	Pearson <u>K</u> endall's tau-b <u>S</u> pearman	
0		
0	rest of Significance	
0	<u>I</u> wo-tailed ○ One-tailed	
0	✓ Elag significant correlations	
0		
0	OK <u>P</u> aste <u>R</u> eset Cancel Help	
0	12. Э. П. ООУЕРУАЛИА ВОО ВЕ ПИУ ИЮДИТ ЕКНА ЛИЕГ ПОООПОЕВЕИТАЦІА ОТО ПИОПЕ	

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 **ΟΚ** θα εμφανιστούν τα σχήματα:

	(	Correlations			
		Οργάνωση και	Κλίμα, Ταυότητα,		Πληροφοριακό
		Λειτουργεία	Υποστήριξη	Διεύθυνση	Σύστημα
Οργάνωση και	Pearson Correlation	1	,773**	,344	,238
Λειτουργεία	Sig. (2-tailed)		,005	,300	,482
	Ν	11	11	11	11
Κλίμα, Ταυότητα,	Pearson Correlation	<b>,773</b> **	1	,256	,534
Υποστήριξη	Sig. (2-tailed)	,005		,423	,074
	Ν	11	12	12	12
Διεύθυνση	Pearson Correlation	,344	,256	1	-,038
	Sig. (2-tailed)	,300	,423		,907
	Ν	11	12	12	12
Πληροφοριακό Σύστημα	Pearson Correlation	,238	,534	-,038	1
	Sig. (2-tailed)	,482	,074	,907	
	Ν	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 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 α=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 Χ2 έλεγχος ανεξαρτησίας

Στην προηγούμενη παράγραφο είδαμε πως υπολογίζουμε το συντελεστή γραμμικής συσχέτισης για την περίπτωση ποσοτικών μεταβλητών. Τι γίνεται όμως 44

στην περίπτωση που έχουμε κατηγορικές μεταβλητές; Σε αυτήν την περίπτωση χρησιμοποιούμε τον Χ² έλεγχο ανεξαρτησίας. Η απαιτούμενη κλίμακα μέτρησης των μεταβλητών είναι η ονομαστική, παρόλο που και μεταβλητές με διατακτική κλίμακα μπορούν να χρησιμοποιηθούν. Ο Χ² 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:

## Ηο: υπάρχει ανεξαρτησία μεταξύ των δύο μεταβλητών

# Η₁: δεν υπάρχει ανεξαρτησία μεταξύ των δύο μεταβλητών

Η μηδενική υπόθεση είναι πάντα αυτή που <u>δεν</u>υποθέτει εξάρτηση (υποθέτει ανεξαρτησία μεταξύ των μεταβλητών). Η προϋπόθεση που απαιτείται από τον X² 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-valuenou υπολογίζεται με βάση το ακριβές τεστ του Fisher (**Fisher's exact test**) ή το Monte Carlo.

#### $MO.\Delta I.\Pi. ~ E.K.\Pi.A.$

nsform	Analyze	<u>G</u> raphs	Utilities	Ad	d- <u>o</u> ns	Window	w.	Help	
	Repo	orts		►		📀 🌑	abc	·	
Ту	Desc	riptive Stati:	stics	►	123	Erequenci	ies		
Numeric	Table	es		►	E o	Descriptiv	es		ζει ξεκ
Numeric	RFM	Analys <u>i</u> s		►		Explore			ίδας μ
Numeric	Com	pare Means		►		<u>C</u> rosstabs	s		χς μου
Numeric	Gene	eral Linear M	lodel	►	1/2	Ratio			ντρών
Numeric	Gene	erali <u>z</u> ed Line	ar Models	►	<b>*</b>	P-P Plots.			ιυς τη
Numeric	Mi⊻e	d Models		►	<b>*</b>	Q-Q Plots			ης ομά
Numeric	Corre	elate		►	7	Στις συγι	κεντ	ρώσει	ς της ο
Numeric	Regr	ession		►	8⊦	Ι απόδοι	σητ	ης ομ	άδας μ
Numeric	Logli	near		►	90	Οι ομάδε	ς πο	υαπ	οτελού
Numeric	Neur	al Net <u>w</u> orks	:	►	10	Η ομάδ	αμα	ου χρι	ισιμοπ
Numeric	Clas	si <u>f</u> y		►	11	Οι διαδ	ικασ	ríες τη	ις ομάζ
Numeric	Dime	nsion Reduc	tion	►	12	Η ομάδο	х ро	υ μπο	ρεί να
Numeric	Scale	e		►	13	Η ομάδο	х µо	υ ενημ	ιερώνει
Numeric	Nong	parametric T	ests	►	1 1	(πάρχει	συχ	νήεπι	κοινων
Numeric	Fore	casting		►	2	Τα μέλη	της	ομάδα	ας μου
Numeric	Surv	ival		•	31	Γα μέλη τ	ns o	μάδα	ς μου :
Numeric	Multip	ple Respons	e	►	4	Μεταξύ τ	ιων μ	νώλαι	της ομ
Numeric	Missi	ing Value Ar	nal <u>y</u> sis		5	Τα μέλη	της	ομάδι	ας μου
Numeric	Mulțij	ple Imputatio	n	►	6	Μεταξύ τ	rωvi	νώλαι	της ομ
Numeric	Com	pļe× Sample:	s	►	7	Τα μέλη	της	ομάδα	ας μου
Numeric	Qual	ity Control		►	8	Αισθάνομ	μαιι	υπερή	φανος
Numeric	ROC	Cur <u>v</u> e			9	Η συνεργ	γασί	α μου	μετην
Numeric	: 1	1	0	:	2_10	Αισθάνα	ομαι	μέλο	ς μιας
Numeric	: 1	1	0	:	2_11	Πιστεύω	ο ότι	ηMC	
		-	-				-		



Το ακριβές τεστ του Fisher θα διεξαχθεί μόνο στην περίπτωση που έχουμε 2X2 πίνακες όπως στο παράδειγμα. Σε αυτήν την περίπτωση το Monte Carlo δεν υπολογίζεται. Στην περίπτωση λοιπόν που έχουμε τον πίνακα 2X2 λόγου 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):



#### Bar Chart

#### 6.6.4 T-test

Οι ἐλεγχοι αυτοί εφαρμόζονται για ελέγχους ισότητας μέσων 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 πρώτο 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:

```
Ho: \mu_1 = \mu_2
H<sub>1</sub>: \mu_1 \neq \mu_2
```

όπου μ₁ ο μέσος του πληθυσμού του πρώτου δείγματος και μ₂ 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 :

MOD	IP.sav [D	ataSet1]	- SPSS S	tatis	tics	Data E	ditor			
orm	<u>A</u> nalyze	<u>G</u> raphs	Utilities	Add	l- <u>o</u> ns	Wind	low	<u>H</u> elp		
<b>I</b>	Repor	ts		•		<b>@</b>	abç	·		
Ту	D <u>e</u> scr	iptive Statis	stics	►						Labe
meric	Tables	5		►	13	Η ομάί	δα μο	υ ενημερ	ώνεται ε	παρκα
meric	RFM A	Analysįs		►	1	πάρχε	α συχ	νή επικο	ινωνία/σ	υνεργα
meric	Compa	are Means		•	M	Means				άζι
meric	<u>G</u> ener	al Linear M	lodel	►	t	One- <u>S</u> a	mple T	Test		N C
meric	Gener	ali <u>z</u> ed Line	ar Models	►	a-a	Indepen	dent-S	amples T	Test	pυ
meric	Mi⊻ed	Models		►	a	Paired-S	Sample	s T Test		ττα
meric	Correl	ate		►	F.	<u>One-Wa</u>	ay ANC	VA		pυ
meric	Regre	ssion		►	7	Τα μέλι	ητης	ομάδας	μου δίνα	
meric	Loglin	ear		►	8	Αισθάν	ομαιι	υπερήφο	νος / τυ	χερός
meric	Neura	l Net <u>w</u> orks		►	9	Η συνε	ργασί	αμουμα	ε την ΜΟ	
meric	Classi	17		►	10	Αισθά	ίνομαι	μέλος μ	πας ομά	δας πα
meric	Dimen	sion Reduc	tion	►	11	Πιστεί	ύω ότι	η ΜΟΔ	ІП ЕКПА	Α διακ
meric	Sc <u>a</u> le			►	12	Πιστεί	ύω ότι	η αμοιβ	ή μου εί	ίναι επ
meric	Nonpa	arametric Te	ests	►	13	Θεωρ	ώικα	/οποιητι	κούς του	ις χώρ
meric	Forec	asting		►	14	Ηπαρ	μόχэα	ενη πλημ	ροφορια	κή και
meric	<u>S</u> urviv	/al		►	15	Αξιολι	ογήστ	ε τον εα	υτό σας	στους
meric	M <u>u</u> ltipl	e Respons	e	►	15	Αξιολι	ογήστ	ε τον εα	υτό σας	στους
meric	🚟 Missin	ig Value Ar	nal <u>y</u> sis		15	Αξιολι	ογήστ	ε τον εα	υτό σας	στους
meric	Mul <u>t</u> ipl	e Imputation	n	►	15	Αξιολι	ογήστ	ε τον εα	υτό σας	στους
meric	Comp	ex Samples	s	►	15	Αξιολ	ογήστ	ε τον εα	υτό σας	στους
meric	Quality	y Control		►	15	Αξιολ	ογήστ	ε τον εα	υτό σας	στους
meric	ROC C	Cur <u>v</u> e			1	Η Διεύθ	θυνση	μπορεί	να χαρο	ικτηρις
meric	: 11		0	3	3_2	Η Διεύθ	θυνση	κατανέμ	ιει δίκαι	ατις υ

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** 

Independent Samples Test

Levene's Test for Equality of Variances				t-test for Equality of Means							
					95% 0 Inter Dif					nfidence al of the rence	
		F	Sig.	t	Sig. (2- Mean Std. Error				Lower	Upper	
Οργάνωση και Λειτουργεία	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 pvalue (Sig.) του τεστ του Levene, κοιτάζουμε την πρώτη ή τη δεύτερη γραμμή αποτελεσμάτων. Στην προκειμένη περίπτωση η pvalue 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 a=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:

Ho:  $\mu_1 = \mu_2$ 

```
\mathsf{H}_{_{1}}:\mu_{_{1}}\neq\mu_{_{2}}
```

Επιλέγουμε

Analyze/NonparametricTests/2

Independent Samples και εμφανίζεται το παράθυρο της εικόνας :

gisis MOD	IP.sav [D	ataSet1]	- SPSS S	tatis	rics I	Data Ed	itor		
ransform	Analyze	<u>G</u> raphs	Utilities	Add	ons	<u>W</u> indo [.]	~ !	Help	
🔚 🖬 🛛	Report	ts		►	<b>N</b> (	<b>2 -</b>	abc		
Ту	D <u>e</u> scri	ipti∨e Statis	tics	►					
Numeric	Ta <u>b</u> les	s		►	13	Η ομάδα	χ μου	ι ενημερώνεται	εп
Numeric	REM A	<u>Inalysi</u> s		►	1 Y	πάρχει	συχν	ή επικοινωνία/	συ
. Numeric	Compa	are Means		►	2 Т	α μέλη	της σ	ομάδας μου σι	laar
. Numeric	<u>G</u> ener	al Linear M	odel	►	зт	αμέληι	ης ο	μάδας μου έχι	ouv
. Numerio	Gener	ali <u>z</u> ed Line	ar Models	►	<u>4</u> N	Λεταξύ	των μ	ιελών της ομάζ	5ας
. Numeric	Mi <u>×</u> ed	Models		►	5 Т	α μέλη	της ο	ομάδας μου αλ	۸ŋλ
. Numeric	<u>C</u> orrel	ate		►	<u>6</u> N	Λεταξύ ΄	των μ	ελών της ομάζ	5ας
. Numerio	<u>R</u> egre	ssion		►	T 7	α μέλη	της σ	ομάδας μου δί	vou
. Numerio	Logline	ear		►	8 4	λισθάνοι	μαιυ	περήφανος / τ	υX«
. Numerio	Neura	l Net <u>w</u> orks		►	9 F	Η συνεργ	γασίο	α μου με την Ν	104
. Numeric	Classi	Íy		►	10	Αισθάνι	ομαι	μέλος μιας ομ	άδι
Numerio	Dimen	sion Reduc	tion	►	11	Πιστεύα	υ ότι	η ΜΟΔΙΠ ΕΚΙ	٦А
Numerio	Sc <u>a</u> le			•	12	Πιστεύα	υ ότι	η αμοιβή μου	είνα
Numeric	Nonpa	arametric Te	ests	•	<b>X</b> ²	<u>C</u> hi-Squar	e		Þς
Numerio	Forec	asting		•	0/1	<u>B</u> inomial	-		ĸr
Numerio	<u>S</u> urviv	/al		•	AAAB	<u>R</u> uns			0
Numerio	Multipl	e Respons	e	•	<b>A</b> :	L-Sample	K-S		0
Numerio	😼 Missin	ig Value Ar	nal⊻sis…		<b>28.</b> g	2 Indepen	dent S	Samples	0
Numerio	Multipl	e Imputation	n –	- ►	<u> </u>	< Indeper	ndent :	Samples	0
Numeric	Compl	ex Samples	s	•	<u> </u>	2 Re <u>l</u> ated	Samp	les	0
Numeric	Quality	y Control		•	225	< Related	Samp	oles	0
. Numerio	🖉 ROC C	Cur⊻e		_	<u>1</u> ⊦	Η Διεύθι	νση	μπορεί να χαρ	ακ

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 **ΟΚ** 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 παράγοντα διαφοροποίησης μία μεταβλητή με 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

μετασχηματισμού πάνω στην ποσοτική ή εξαρτημένη είδος 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 καταλοίπων καλό κανονικότητα των θα ήταν είτε va 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 της ομοσκεδαστικότητας περίπτωση αλλά η υπόθεση της κανονικότητας ικανοποιείται, μπορούμε να χρησιμοποιήσουμε το τεστ του Welch ή το τεστ των Brown-Forsythe, τα οποία είναι ανθεκτικά περιπτώσεις ετεροσκεδαστικότητας. σε Θa χρησιμοποιήσουμε τα δεδομένα των αυτοκινήτων για να ελέγξουμε κατά πόσο τα βάρη διαφέρουν από χώρα σε χώρα. Οι μηδενικές υποθέσεις ορίζονται ως εξής:

Ho: μ₁=μ₂=μ₃ Η₁: ἐνα τουλἁχιστον ζεὑγος μἑσων διαφἑρει

# Εγχειρίδιο Στατιστικής Επεξεργασίας Δεδομένων

U	12.8. Δισθανομαι υπεορφανος / τυνεοος που Κοισκομαι στο ΜΙ	INTER LA	None	None	11	= Right
0	📴 One-Way ANOVA	×	None	None	11	≡ Right
0	Dependent List:		None	None	11	≡ Right
Π	💑 5_3 Οι ΟΜΕΑ θεωρ 🔺 🖉 Οργάγωστα και Δειτομο	rasts	None	None	11	I Right
0	δ_4 Οι ΟΜΕΑ αντα	Hoc	None	None	11	= Dight
	💑 5_5 Η αντιμετώπιση	0.000	NUMB	None	11	- Right
U	💑 5_6 Η ΜΟΔΙΠ αντα		None	None	11	= Right
0	💑 5_7 Η συνεργασία τ	One-Way	ANOVA	: Post Hoc Multip	le Compariso	ns
0	💑 5_8 Τα παραγόμεν					
0	🐓 Κλίμα, Ταυότητα, Υπ	Equal Varia	ances A	ssumed		
0	Διεύθυνση (Diefthinsi) Eactor:			<u>S-N-K</u>	Waller-Dun	can
0	Τιληροφοριακό 2001 Ο Δ.2_15 Αξιολογήστε τον	Bonferro	oni	V Tukev	Type I{Type I	Error Ratio: 100
Ω	OK Paste Reset Cancel Help					
-		🗌 Зі́дак		□ Tukey.s-b		
0	2_13 Αξιοκογήστε τον εαύτο σας στους παρακατώ τορείς. [Α	Scheffe		Duncan	Control Categ	gory : Last
U	3_1 Η Διεύθυνση μπορεί να χαρακτηρίσθει ως αποτελεσματικ	<u>R-E-G-W</u>	VF	Hochberg's GT2	_ Test —	
0	3_2 Η Διεύθυνση κατανέμει δίκαια τις υπευθυνότητες στα μέλ	R-E-G-M	vo	Gabriel	2-sided	a ⊖ < Control (
0	3_3 Η Διεύθυνση δίνει το «καλό παράδειγμα» στα μέλη της ο		. ~			0 12 11 0
0	3_4 Η Διεύθυνση συνεργάζεται και στηρίζει σε ατομικό επίπε	- Equal Varia	ances N	ot Assumed ———		
0	3 5 Η Διεύθυνση της υπηρεσίας προάγει την ανάληψη πρωτο	Tamban	70			
0	3.6. Τα μέλη της ομάδας μου καταθέτουν τις απόψεις τους ελ		8512	Dunneus 13	Games-How	
n	3.7. Η Διεύθυνση εξειδικεύει και κάνει σαφείς κάθε φορά του	Signi <u>f</u> icance l	evel: 0,0	5		
0	<ol> <li>2.8. Πάσο σποτολοσιιστικός βοίσκοτο τις αγόρισμος πορβολής</li> </ol>					
0				Continue	ancel	Help
U	3_9 Ποσο γνωστη θεωρειτε την ΜΟΔΙΠ στο σύνολο της πανετ					
_						

#### 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 Κατασκευἡ Διαγραμμἁτων ἑλεγχου/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: (a) 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 ή ονομάζονται ειδικές 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 тην παραγωγή ενός προϊόντος είναι та 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, Τ) 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 διεργασία σε μια (online processmonitoring). 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:

isis MOD	IP.sav [D	ataSet1]	- SPSS S	tatis	rics	Data Ed	itor		
ansform	<u>A</u> nalyze	<u>G</u> raphs	Utilities	Add	ons	<u>Windo</u>	∾ i	<u>H</u> elp	
<b>****</b> [	Repor	ts		►		🍲 🌑	abc		
Ту	D <u>e</u> scr	iptive Statis	stics	►					
Numeric	Ta <u>b</u> les	s		►	13	Η ομάδα	χ μου	ενημε	ρών:
Numeric	REM A	Analys <u>i</u> s		►	1 Y	πάρχει	συχν	ή επικ	οινω
Numeric	Compa	are Means		►	2 '	Γα μέλη	της ο	μάδας	; μοι
Numeric	<u>G</u> ener	al Linear M	odel	►	зт	α μέλη ι	ns of	ιάδας	μου
Numerio	Gener	ali <u>z</u> ed Line	ar Models	►	4	Μεταξύ	ιων μ	ελών τ	ηςο
Numeric	Mi <u>×</u> ed	Models		►	5	Γα μέλη	της ο	μάδας	; μοι
Numeric	<u>C</u> orrel	ate		►	<u>6</u> I	Μεταξύ ΄	ιων μ	ελών τ	ηςο
Numeric	<u>R</u> egre	ssion		►	7 1	Γα μέλη	της ο	μάδας	; μοι
Numeric	Loglin	ear		►	8 /	Αισθάνοι	u i n	περήφ	ανος
Numeric	Neura	l Net <u>w</u> orks		►	9 1	Η συνερν	γασίο	ίμου μ	ıε τr
Numeric	Classi	ÍY		►	10	Αισθάνι	ομαιι	μέλος	μιας
Numeric	Dimen	sion Reduc	tion	►	11	Πιστεύα	υ ότι	η ΜΟΖ	710-1
Numeric	Sc <u>a</u> le			►	12	Πιστεύα	υ ότι	η αμοι	βήμ
Numeric	<u>N</u> onpa	arametric Te	ests	►	13	Θεωρώ	ικαν	οποιητ	ικού
Numeric	Forec	asting		►	14	Η παρε	χόμε	νη πλη	ροφ
Numeric	<u>S</u> urviv	/al		►	15	Αξιολογ	γήστε	: TOV 80	χυτό
Numeric	M <u>u</u> ltipl	e Respons	e	►	15	Αξιολον	γήστε	: TOV 80	χυτό
Numeric	🛃 Missin	ig Value Ar	nal <u>y</u> sis		15	Αξιολον	γήστε	: TOV 80	χυτό
Numerio	Multipl	e Imputatio	n	►	15	Αξιολογ	γήστε	: TOV 80	χυτό
Numeric	Compl	ex Sample:	3	•	15	Αξιολογ	/ήστε	: TOV 80	χυτό
Numeric	Quality	y Control		•	~~	Control Cl	harts	-	UτÓ
Numeric	📶 ROC C	Cur <u>v</u> e				Pa <u>r</u> eto Ch	arts		Vα





Control Chart: Οργάνωση και Λειτουργεία

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 (ἡ ανάλυση Pareto) οφείλει το ὀνομά του στον Ιταλό οικονομολόγο Vilfredo Pareto ο οποίος είναι κυρίως γνωστός για την παρατήρησή του ὀτι 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:

is MODI	s MODIP.sav [DataSet1] - SPSS Statistics Data Editor									
<u>T</u> ransfor	m <u>A</u> l	halyze	<u>G</u> raphs	Utilities	Add	- <u>o</u> ns	<u>W</u> indo	w	<u>H</u> elp	
<b>*</b>	- [	Repor	ts		•	<b>V</b> 9	۵ 🌑	abc	'	
	Ту	D <u>e</u> scr	iptive Statis	stics	►					
; Num	eric	Ta <u>b</u> les	3		•	Οp	όλος μ	ιου σ	την Ο	MEA 1
Num	eric	REM A	Analysįs		•	HN	10ΔІП	EKL	ΙΑ επι	ικεντρό
Num	eric	Compa	are Means		•	НМ	104П	EKL	ΙΑ απ	αντά ό
; Num	eric	<u>G</u> ener	al Linear M	odel	•	Οp	όλος τ	ης Μ	ΙΟΔΙΠ	EKNA
Num	eric	Gener	ali <u>z</u> ed Line	ar Models	•	Ore	απαιτή	σεις	της Μ	ΙΟΔΙΠ
Num	eric	Mi <u>×</u> ed	Models		•	Τα χ	χρονοδ	ιαγρ	άμμα	τα ποι
i Num	eric	<u>C</u> orrel	ate		•	Hε	пкоім	ονία	με το Έ	προσω
Num	eric	<u>R</u> egre	ssion		•	. Or c	τυνεργ	άτες	της Ν	10ΔІП
j Num	eric	Loglin	ear		•	Нσ	υχνή ε	πιко	ινωνία	με τον
) Num	eric	Neura	l Net <u>w</u> orks		►	I_ To	προσ	ωπικ	ό της	ΜΟΔΙΙ
Num	eric	Classi	Íy		►	_н,	σύνταξ	,η τω	ν εκθε	έσεων
r Num	eric	<u>D</u> imen	sion Reduc	tion	►	LH	ηγεσία	της	моΔ	IU EKI
: Num	eric	Sc <u>a</u> le			•	L ET	τισκέπ	тоµо	α την	ιστοσε
Num	eric	<u>N</u> onpa	arametric Te	ests	•	i_ 0i	διαδικ	ασίε	ς αξιο	ολόγησ
Num	eric	Forec	asting		•	_ Αξ	ιοποιή	σατε	: στο ⁻	Τμήμα
Strin	g	<u>S</u> urviv	/al		►	I_ Av	Ναι, τ	τόσο	ι διδά	σκοντε
Num	eric	M <u>u</u> ltipl	e Respons	e	•	<u> _</u> Пó	ίσο χρι	ήσιμ	ες και	επιτυ;
Num	eriq 🔀	🗿 Missin	ig Value Ar	nal <u>y</u> sis		voSte	ep Clus	sterl	Numb	er
Num	eric	Multipl	e Imputatio	n	•	voSte	ep Clus	sterl	Numb	er
Num	eric	Comp	ex Sample:	3	•	voSte	ep Clus	sterl	Numb	er
Num	eric	<u>Q</u> uality	y Control		•	<u>~</u> 0	on <u>t</u> rol C	harts		íα
Num	eri 🙋	ROC	Cur <u>v</u> e			🖬 Pa	a <u>r</u> eto Cł	narts.		
Num	eric	11		5	E	πίδρο	αση στ	ην π	ορεία	του Πα

U 3_ Η ΜΟΔΙΠ ΕΚΠΑ απαντά άμεσα στα ερωτήματα και βρίσκει λύσεις στα προβλήματα τ 4 Ο οάλος τος ΜΟΔΙΠ ΕΚΠΑ δεν είναι ιδιαίτερα πομαντικός για το μέλλον του Ιδρύματο. Π Define Simple Pareto: Sums of Separate Variables  $\left[\times\right]$ ⊻ariables: <u>T</u>itles... ✓ 4_Ο ρόλος της ΜΟΔΙΠ ΕΚΠ...
 ✓ 5_Οι απαιτήσεις της ΜΟΔΙΠ ...
 ✓ 9_Η συχνή επικοινωνία με τ... ελεσματική 🗞 1_ Ο ρόλος μου στη... 📥 2_Η ΜΟΔΙΠ ΕΚΠΑ ε... Options.. ματός μου για ΔΙΠ ΕΚΠΑ είν 15_Οι διαδικασίες αξιολόγη... 🔗 6_ Τα χρονοδιαγράμ... + υμία 7_Η επικοινωνία με .. η συνεργασίο 10_Το προσωπικό ... ήματα που αν 🔗 11_Η σύνταξη των ε.. 12_Η ηγεσία της MO... ιληροφοριακέ ☑ <u>D</u>isplay cumulative line 13_Επισκέπτομαι τ... που σας παρ 4. 17_ Αξιοποιήσατε σ... Panel by όγιο φοιτητών Ro<u>w</u>s: υτικές ημερίδ + 💑 TwoStep Cluster Nu... 💑 TwoStep Cluster Nu... Nest variables (no empty rows) 💑 TwoStep Cluster Nu... -Template-Use chart specifications from: Eile... OK Paste Reset Cancel Help Εγκυρότητα και Λειτουργία





# 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*. Willey 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.