Στατιστικές Εφαρμογές με τη χρήση του S.P.S.S

Statistical Applications with the use of S.P.S.S

Δρ. Ευστάθιος Δημητριάδης

Καβάλα, 2021

Ευστάθιος Δημητριάδης
Καθηγητής ΔΙ.ΠΑ.Ε.
Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας
Τηλ.: 2510 462304
Email: edimit@mst.ihu.gr

Ο **Δημητριάδης Ευστάθιος** είναι **Καθηγητής** Στατιστικής και Οικονομικών Μαθηματικών στο **Τμήμα Διοίκησης Επιστήμης και Τεχνολογίας του Διεθνούς Πανεπιστημίου της Ελλάδος (ΔΙ.ΠΑ.Ε.)**. Ο Δρ. Δημητριάδης σπούδασε **Μαθηματικά** και συνέχισε με Μεταπτυχιακές σπουδές τόσο στη **Στατιστική και Δημογραφία** όσο και στη **Διασφάλιση Ποιότητας**. Εκπόνησε τη **Διδακτορική του Διατριβή στο Τμήμα Εφαρμοσμένης Πληροφορικής του Πανεπιστήμιου Μακεδονίας**.

Διετέλεσε μέλος της Συγκλήτου και της Επιτροπής Ερευνών του Τ.Ε.Ι Ανατολικής Μακεδονίας και Θράκης και είναι Υπεύθυνος Πρακτικής Άσκησης του Τμήματος. Είναι Επιστημονικός Υπεύθυνος Έργων INTERREG και μέλος Ομάδων εργασίας Ερευνητικών Προγραμμάτων. Συμμετείχε ενεργά σε προγράμματα Erasmus με οργάνωση και καθοδήγηση ομάδων φοιτητών σε Διεθνείς Διαγωνιστικές Εκδηλώσεις με σημαντική επιτυχία.

Τα ερευνητικά του ενδιαφέροντα έχουν ως εξής: Πληροφορική, Εκπαίδευση, Τουρισμός, H.R.M., Στατιστικές εφαρμογές στις Κοινωνικές Επιστήμες. Προς το παρόν διδάσκει τακτικά σε προπτυχιακά μαθήματα του Διεθνούς Πανεπιστημίου της Ελλάδος. Διαθέτει σημαντική και πολυετή εμπειρία διδασκαλίας και σε πολλά μεταπτυχιακά προγράμματα (Ερευνητική Μεθοδολογία, Διαχείριση Έργων και Ποσοτικές μεθόδους στη Λήψη Αποφάσεων). Έχει συμμετάσχει σε πλήθος εθνικών και διεθνών συνεδρίων, τόσο στην Ελλάδα όσο και σε διάφορες χώρες του εξωτερικού (60) και πολλά άρθρα του δημοσιεύονται σε επιστημονικά περιοδικά και επίσημα πρακτικά συνεδρίων (61). Είναι επίσης συγγραφέας ενός επιστημονικού βιβλίου με τίτλο «Στατιστική επιχειρήσεων με εφαρμογές στις S.P.S.S και LISREL».

Το παρόν εγχειρίδιο αποτελείται από επιλεγμένα κεφάλαια του συγγράμματος: Στατιστική επιχειρήσεων με εφαρμογές στις S.P.S.S και LISREL, Εκδόσεις Κριτική, του συγγραφέα Ευσταθίου Δημητριάδη.

Χορηγείται για χρήση, αποκλειστικά και μόνο, των φοιτητών του Μεταπτυχιακού προγράμματος ***Διδακτική των Επιστημών και Σύγχρονες Τεχνολογίες.*** Απαγορεύεται η αναδημοσίευση και η διανομή ολοκλήρου ή μέρους του εγχειριδίου.

Ο Συγγραφέας
Ευστάθιος Δημητριάδης
Καθηγητής

Κεφάλαιο 2

Καταχώριση– Παρουσίαση Δεδομένων

Chapter 2

Data

Introduction –Presentation

2. Εισαγωγή

Μετά την ολοκλήρωση της συλλογής των απαιτούμενων πληροφοριών (δεδομένων), συνήθως με τη χρήση ερωτηματολογίου, ακολουθεί η καταχώρηση των πληροφοριών στο S.P.S.S και στη συνέχεια η παρουσίαση των πρώτων ενδεικτικών αποτελεσμάτων.

2.1 Καταχώριση δεδομένων

Είναι γνωστό, από τη θεωρία της Στατιστικής, ότι κάθε ερώτηση ενός ερωτηματολογίου αποτελεί μια **μεταβλητή (variable or item)** η οποία μπορεί να είναι:

- **Ποσοτική (Quantitative),** όταν παίρνει τιμές από αριθμητικό σύνολο.
- **Ποιοτική (Qualitative),** όταν παίρνει τιμές από μη αριθμητικό σύνολο.

Ανάλογα με τη φύση των χαρακτηριστικών τα οποία θέλουμε να μετρήσουμε χρησιμοποιούμε και διαφορετική **κλίμακα μέτρησης (measurement scale)** με αποτέλεσμα οι δύο βασικές κατηγορίες μεταβλητών να διαχωρίζονται ακόμη περισσότερο.

Έτσι για την **ποσοτική** μεταβλητή έχουμε:

✓ **Αναλογική Κλίμακα (Ratio scale)**, όταν η μεταβλητή παίρνει τιμές οι οποίες είναι αναλογικές. Για παράδειγμα θα μπορούσαμε να αναφέρουμε τους μισθούς των υπαλλήλων, την ηλικία ή το βάρος τους. Δηλαδή, κάποιος που είναι 50 ετών έχει ζήσει διπλάσια χρόνια από κάποιον άλλο 25 ετών.

✓ **Διαστημική Κλίμακα (Interval scale),** όταν η μεταβλητή παίρνει τιμές των οποίων ο λόγος δεν έχει καμία αξία. Αν για παράδειγμα αναφέρουμε ότι η σημερινή θερμοκρασία είναι 30^0 C, δεν

σημαίνει ότι κάνει διπλάσια ζέστη από κάποια άλλη μέρα με θερμοκρασία 15^0 C. Άλλη περίπτωση είναι η βαθμολογία στα διάφορα τεστ GMATT, TOEFL, IQ κ.λπ.

Η **ποσοτική** μεταβλητή διακρίνεται επίσης σε:

- **Συνεχή (Continuous),** όταν παίρνει όλες τις τιμές ενός διαστήματος. Το βάρος ενός ανθρώπου, αλλά και οι μισθοί των υπαλλήλων του Δημοσίου είναι παράδειγμα συνεχών μεταβλητών.
- **Ασυνεχή (Discrete),** όταν παίρνει μεμονωμένες μόνο τιμές ενός διαστήματος. Ο αριθμός των παιδιών μιας οικογένειας, όπως και ο αριθμός των μαθημάτων για την ολοκλήρωση ενός προγράμματος σπουδών είναι ασυνεχείς μεταβλητές.

Για την **ποιοτική** μεταβλητή έχουμε αντίστοιχα:

✓ **Ονομαστική Κλίμακα (Nominal scale),** όταν η μεταβλητή παίρνει τιμές οι οποίες δεν μπορούν να ιεραρχηθούν. Παράδειγμα το χρώμα των ματιών ή το θρήσκευμα είναι μεταβλητές των οποίων οι τιμές δεν μπορούν να ιεραρχηθούν

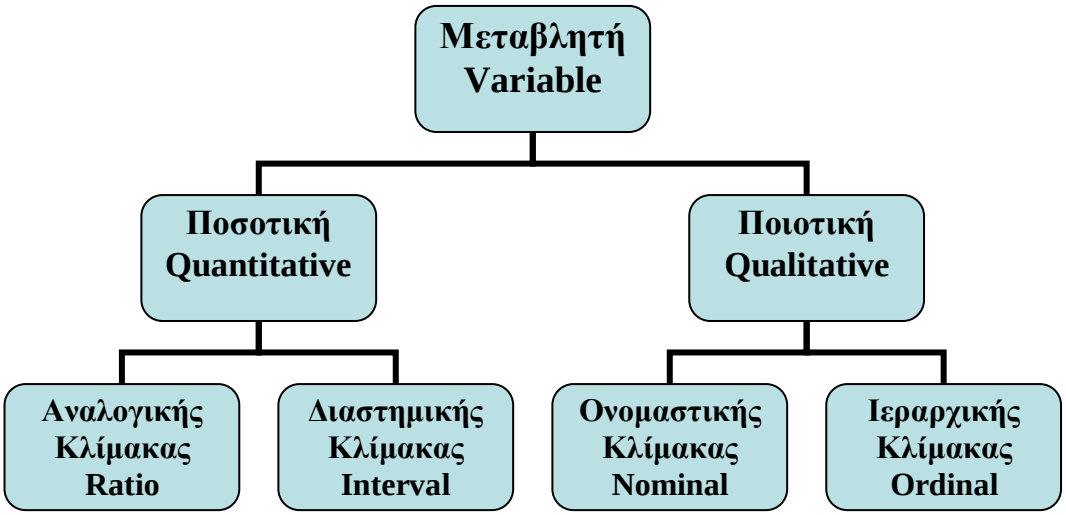
✓ **Κλίμακα Τάξης ή Ιεράρχησης (Ordinal scale)**, όταν η μεταβλητή παίρνει τιμές οι οποίες μπορεί να ιεραρχηθούν. Οι μεταβλητές αυτές ορίζουν ξεκάθαρα την ιεράρχηση μεταξύ των διαφόρων κατηγοριών αλλά οι απόλυτες αποστάσεις μεταξύ των κατηγοριών είναι άγνωστες. Παράδειγμα την κατάσταση της υγείας ενός ασθενούς μπορούμε να τη χαρακτηρίσουμε από κακή έως άριστη με ενδιάμεση κλιμάκωση, όπως επίσης την απόδοση ενός υπαλλήλου στην εργασία του.

!!!! *Σύμφωνα με πολλούς επιστήμονες, η θέση των μεταβλητών ιεραρχικής κλίμακας ως προς την κατάταξή τους σε ποιοτικές ή*

ποσοτικές μεταβλητές είναι ασαφής. Πολλές φορές αντιμετωπίζονται ως ποσοτικές, ενώ άλλοτε αναλύονται με τη χρήση μεθόδων κατάλληλων για ποιοτικές μεταβλητές. Βέβαια, οι μεταβλητές αυτές προσεγγίζουν περισσότερο τις μεταβλητές διαστημικής κλίμακας παρά τις μεταβλητές ονομαστικής κλίμακας.

Για την καταχώρηση των απαντήσεων, λοιπόν, είναι απαραίτητο να εξετάσουμε χωριστά αυτές τις περιπτώσεις, καθώς η μεθοδολογία και η τεχνική που εφαρμόζουμε διαφέρει.

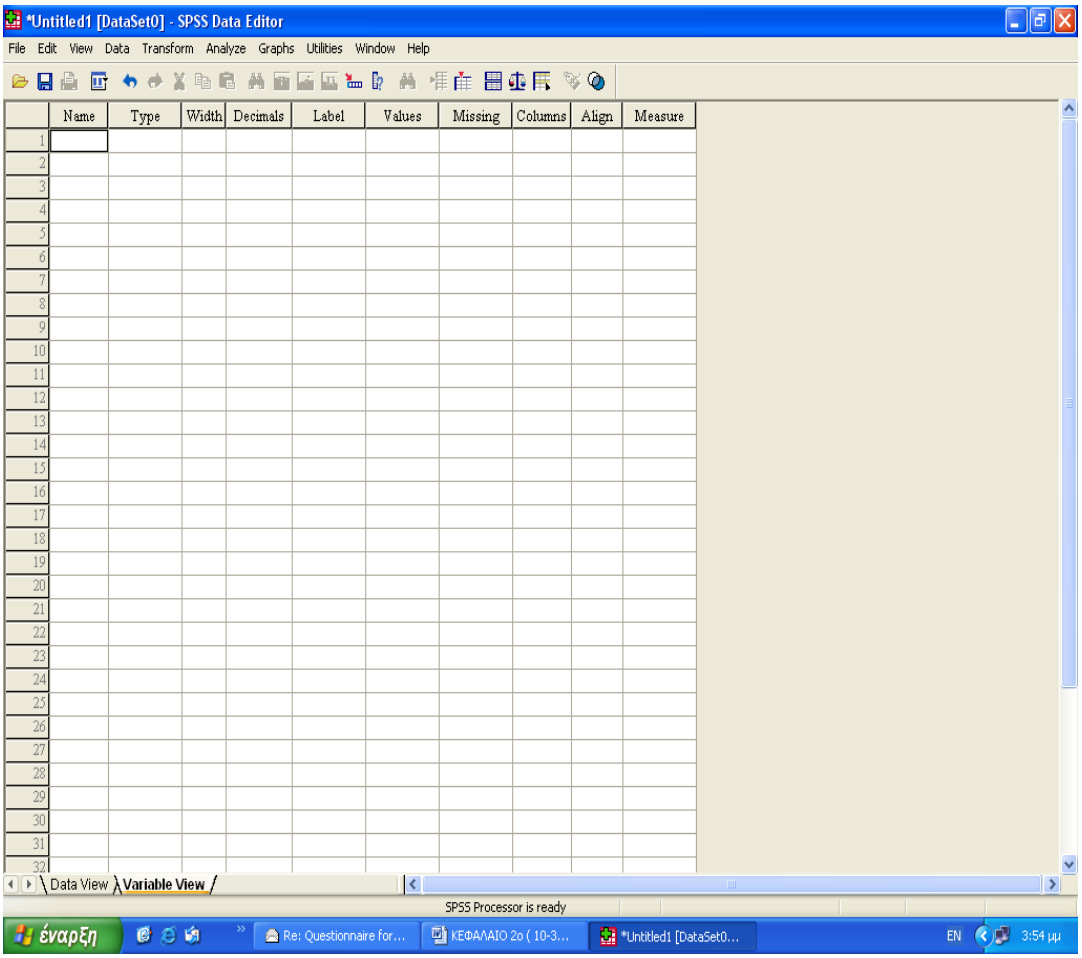
Σχήμα 1: Είδη Μεταβλητών



2.1.1 Καταχώριση ποσοτικών δεδομένων

Όταν τα δεδομένα είναι ποσοτικά, η καταχώρησή τους είναι πολύ απλή.

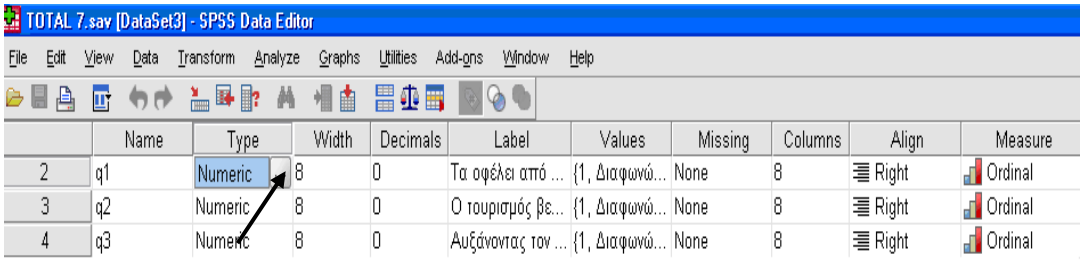
- Πατάμε στο κουμπί ***Variable View*** του **Data Editor** (κάτω αριστερά στην εικόνα 1.1) ή κάνουμε διπλό κλικ στην επικεφαλίδα της στήλης (var) που θέλουμε να εισάγουμε τα δεδομένα και εμφανίζεται η εικόνα 2.1.



Εικόνα 2.1

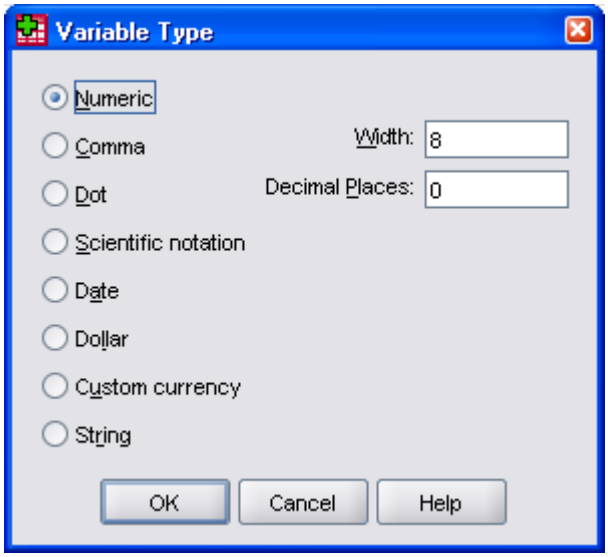
- Στη στήλη ***name (όνομα)*** γράφουμε, κατά προτίμηση με λατινικούς χαρακτήρες, ένα σύντομο τίτλο για την ερώτηση που θα καταχωρήσουμε ή αφήνουμε την αυτόματη αρίθμηση που προτείνει το πρόγραμμα. (π.χ Var00001).

- Πατώντας στη στήλη ***Type (τύπος)*** εμφανίζεται η εικόνα 2.2



Εικόνα 2.2

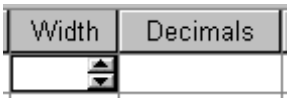
- Με κλικ στο κουμπάκι εμφανίζεται η εικόνα 2.3



Εικόνα 2.3

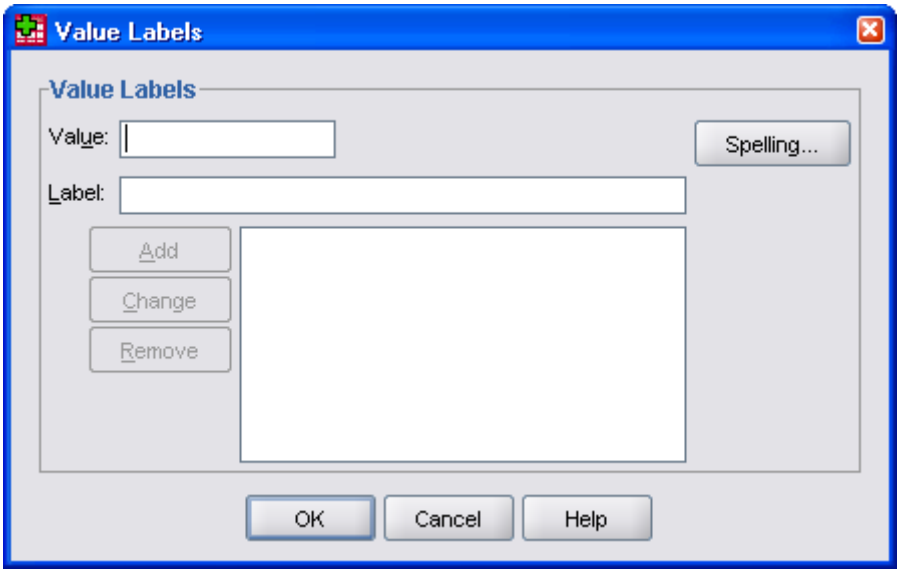
- Επιλέγουμε ***Numeric (αριθμητικό)*** και στη συνέχεια
- Στην ένδειξη ***Width (πλάτος)*** αναγράφουμε το μέγιστο πλήθος των ακέραιων ψηφίων που έχει η μεταβλητή ενώ
- Στην ένδειξη ***Decimal Places (δεκαδικά σημεία)*** το πλήθος των δεκαδικών ψηφίων της μεταβλητής, αν φυσικά υπάρχουν. Ευνόητο είναι ότι για τιμές ακέραιες στη θέση ***Decimal Places*** βάζουμε τον αριθμό 0 ή τίποτα.

Τις ενδείξεις ***Width*** και ***Decimal Places*** μπορούμε να τις συμπληρώσουμε απευθείας από τις στήλες ***Width*** και ***Decimals*** της εικόνας 2.1, οι οποίες παίρνουν τη μορφή που φαίνεται στη συνέχεια, μόλις πατήσουμε μέσα.



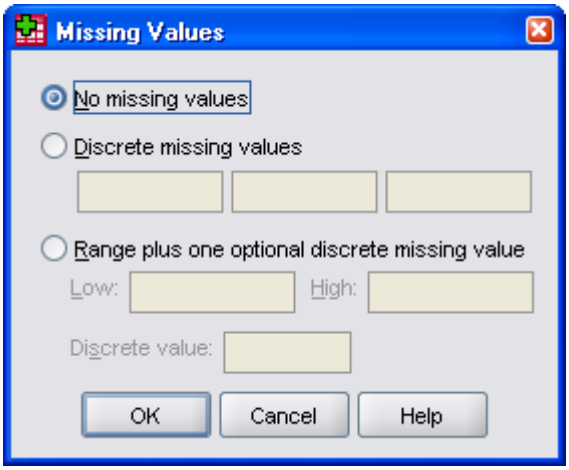
Εικόνα 2.4

- Στη συνέχεια ***OK*** για να επιστρέψουμε στην εικόνα 2.1.
- Στη στήλη ***Label (επικεφαλίδα)***, της εικόνας 2.1, γράφουμε τον τίτλο της ερώτησης, αν θέλουμε με ελληνικά ψηφία, ο οποίος θα εμφανίζεται σαν επικεφαλίδα όταν με το ποντίκι περνάμε πάνω από τη στήλη αλλά και στην παρουσίαση των αποτελεσμάτων.
- Πατώντας στη στήλη ***Values (τιμές)*** εμφανίζεται ένα κουμπάκι όμοιο με αυτό της στήλης ***Type*** το οποίο αν ενεργοποιήσουμε θα εμφανιστεί η επόμενη εικόνα.



Εικόνα 2.5

- Τη φόρμα αυτή σε περίπτωση ποσοτικών δεδομένων δεν τη χρησιμοποιούμε, ενώ αντίθετα είναι πολύ χρήσιμη σε περιπτώσεις ποιοτικών δεδομένων καθώς χρησιμεύει στην κωδικοποίηση αυτών, όπως θα δούμε στη συνέχεια.
- Η στήλη ***Missing (χαμένος ή ελλιπής)***, της εικόνας 2.1 μας δίνει την επόμενη φόρμα.



Εικόνα 2.6

- Στη φόρμα αυτή σημειώνουμε τις τιμές ή τους κωδικούς των τιμών οι οποίες θέλουμε να θεωρούνται ***missing*** στην επεξεργασία και στην παρουσίαση των δεδομένων. Είναι περισσότερο χρήσιμη σε κωδικοποιημένα δεδομένα.

Έτσι με την επιλογή:

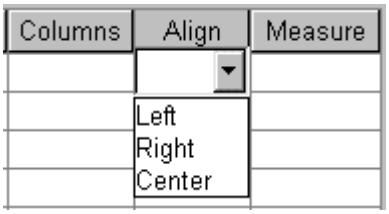
✓ ***No missing values*** δεν θα θεωρείται ***missing*** καμία τιμή παρά μόνο τα κενά κελιά.

✓ ***Discrete missing values*** θα θεωρούνται ***missing*** οι τιμές που θα αναγράψουμε στα τρία παράθυρα.

✓ ***Range plus one optional discrete missing value*** θα θεωρούνται ***missing*** οι τιμές από το παράθυρο ***Low*** μέχρι το παράθυρο ***High*** εκτός της τιμής που θα αναγραφεί στο παράθυρο ***Discrete value***.

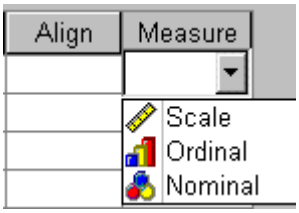
- Στη συνέχεια ***OK*** και επιστροφή στην εικόνα 2.1 από τη στήλη ***Columns*** της οποίας ρυθμίζουμε τον αριθμό των χαρακτήρων οι οποίοι θα χωρούν σε κάθε κελί της στήλης.

- Πατώντας στη στήλη ***Align (ευθυγράμμιση-στοίχιση)*** της εικόνας 2.1 έχουμε την επόμενη φόρμα από την οποία επιλέγουμε ***Left (αριστερά)*** ή ***Right (δεξιά)*** ή ***Center (κέντρο)*** ανάλογα με τη στοίχιση που θέλουμε να έχουν τα δεδομένα μέσα στα κελιά.



Εικόνα 2.7

- Τέλος πατώντας στη στήλη ***Measure (μέτρο)*** της εικόνας 2.1 έχουμε τη φόρμα από την οποία δηλώνουμε το είδος της μεταβλητής. Στην περίπτωση ποσοτικών μεταβλητών επιλέγουμε ***Scale*** ενώ σε περίπτωση ποιοτικών ***Ordinal*** ή ***Nominal*** όπως θα δούμε στη συνέχεια.



Εικόνα 2.8

Αφού ολοκληρώσουμε τη διαμόρφωση της πρώτης μεταβλητής, συνεχίζουμε στις επόμενες γραμμές τη διαμόρφωση των υπολοίπων μεταβλητών μέχρι να ολοκληρωθούν οι ερωτήσεις του ερωτηματολογίου. Στη συνέχεια από το κουμπί ***Data View*** στο κάτω αριστερό μέρος της εικόνας 2.1 επανερχόμαστε στο χώρο καταχώρισης δεδομένων.

Εφαρμογή: Στην ερώτηση: ***For how long do you plan the company's needs for its personnel*?** οι απαντήσεις είναι αριθμοί οι οποίοι εκφράζουν μήνες. Η καταχώριση λοιπόν, αυτών των αριθμών μπορεί να γίνει πολύ απλά με τη μέθοδο που αναπτύξαμε.

Αναλυτικά θα έχουμε:

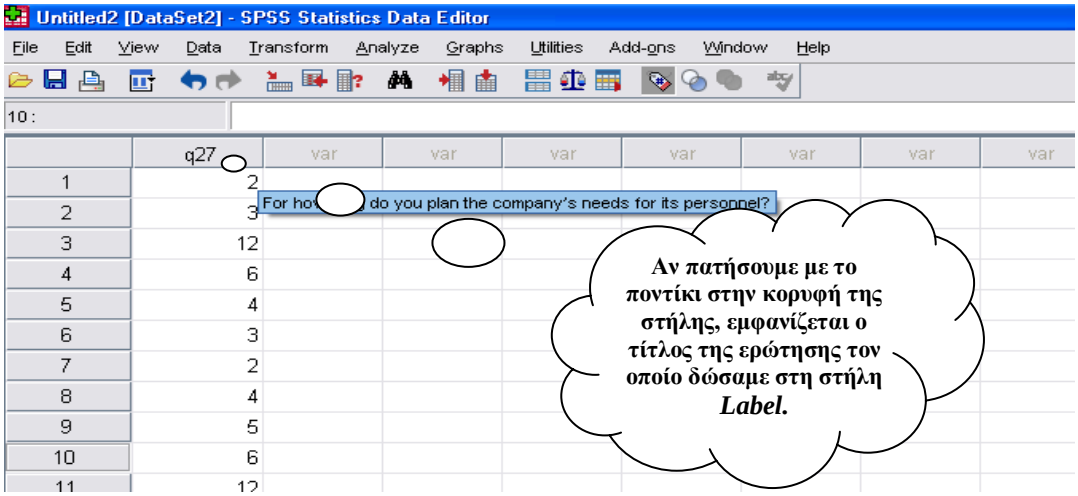
- Διπλό κλικ στην πρώτη στήλη του ***Data Editor*** και εμφάνιση της εικόνας 2.1.

- Στη στήλη ***name*** αναγράφουμε τον αριθμό της ερώτησης π.χ ***Q27***.

- Στη συνέχεια πατώντας στη στήλη ***type*** εμφανίζεται η εικόνα 2.3 στην οποία επιλέγουμε ***numeric***. Στη θέση ***Width*** γράφουμε 2, ενώ στη θέση ***Decimal Places*** γράφουμε 0.

- Στη στήλη ***Labels*** πληκτρολογούμε τον τίτλο της ερώτησης **Q27 «*For how long do you plan the company's needs for its personnel?*»** Κατόπιν ***πατάμε Data View*** και επανερχόμαστε στον ***Data Editor*** έτοιμοι για την καταχώρηση των αριθμητικών δεδομένων.

Αν υποθέσουμε, λοιπόν, ότι έχουμε καταχωρήσει τα δεδομένα η μορφή πλέον του ***Data Editor*** θα είναι η επόμενη:



Εικόνα 2.9

2.1.2 Καταχώριση ποιοτικών δεδομένων σε ονομαστική κλίμακα

Καταχώρηση ποιοτικών δεδομένων σε ονομαστική κλίμακα σημαίνει αναγραφή λέξεων και ίσως ολόκληρων προτάσεων σε κάθε κελί. Αυτό σημαίνει πολύ χρόνο αλλά και κόπο.

Έτσι, για παράδειγμα, στην ερώτηση:

Q3. Which is the form/type of the company?

Οι προβλεπόμενες απαντήσεις είναι οι επόμενες:

Personal

Ordinary partnership

Limited partnership

L.T.D

S.A

Αν λοιπόν θέλουμε να καταχωρήσουμε τις απαντήσεις αυτής της ερώτησης, θα πρέπει κάθε φορά να γράφουμε μία από τις 5 πιο πάνω απαντήσεις.

Η διαδικασία καταχώρησης είναι η επόμενη:

- Πατάμε στο κουμπί ***Variable View*** της εικόνας 1.1 και εμφανίζεται η εικόνα 2.1.
- Στη στήλη ***Name*** γράφουμε τον αριθμό της ερώτησης.
- Από τη στήλη ***Type*** εμφανίζεται η εικόνα 2.3. Επιλέγουμε ***String*** και στην ένδειξη ***Characters*** σημειώνουμε τον αριθμό των χαρακτήρων που θέλουμε να εμφανίζονται στα κελιά. Εδώ ο αριθμός 8 δηλώνει ότι θα εμφανίζονται μόνο 8 χαρακτήρες από κάθε απάντηση. Αν δώσουμε μεγαλύτερο αριθμό θα παρατηρήσουμε ότι αυτόματα αναπροσαρμόζεται και το πλάτος της στήλης.
- Στη στήλη ***Label*** αναγράφουμε τον τίτλο της ερώτησης και
- Από τη στήλη ***Measure*** επιλέγουμε ***Nominal***, εφόσον πρόκειται για ποιοτική μεταβλητή σε ονομαστική κλίμακα.
- Στη συνέχεια με ***Data View*** γυρνάμε στην εικόνα 1.1 και ο ***Data Editor*** είναι έτοιμος για την καταχώρηση των δεδομένων.

!!! *Η μέθοδος αυτή είναι αρκετά κουραστική και χρονοβόρος. Για το λόγο αυτό είναι προτιμότερο, ακόμη και αν πρόκειται για ποιοτικά δεδομένα σε ονοματική κλίμακα να ακολουθούμε τη δεύτερη μέθοδο που περιγράφεται στην επόμενη παράγραφο (στη στήλη Measure επιλέγουμε πάλι Nominal).*

2.1.3 Καταχώριση ποιοτικών δεδομένων σε κλίμακα τάξης

Για την καταχώρηση ποιοτικών δεδομένων σε κλίμακα τάξης (ordinal) μπορούμε να ακολουθήσουμε μία από τις παρακάτω δύο μεθόδους:

Πρώτη Μέθοδος: Είναι πολύ απλή και απαιτεί μία προεργασία πάνω στο ερωτηματολόγιο. Κατά την προεργασία αυτή κωδικοποιούμε την κάθε απάντηση δίνοντας σε καθεμία έναν αριθμό ως κωδικό και στη συνέχεια καταχωρούμε στα κελιά τους κωδικούς, ακολουθώντας ***την ίδια διαδικασία που ακολουθήσαμε και κατά την καταχώρηση των ποσοτικών δεδομένων***.

Η μέθοδος αυτή είναι πολύ απλή, αλλά παρουσιάζει το εξής μειονέκτημα κατά την παρουσίαση των αποτελεσμάτων. Στους διάφορους πίνακες επεξεργασμένων αποτελεσμάτων οι οποίοι θα προκύψουν δεν θα εμφανίζονται οι απαντήσεις αλλά οι κωδικοί αυτών, με αποτέλεσμα να χρειάζεται να ανατρέχουμε στο ερωτηματολόγιο για να θυμηθούμε τι σημαίνει κωδικός 3 ή 5.

Για να παρακάμψουμε αυτή την αδυναμία και να έχουμε άρτια αποτελέσματα είναι καλύτερα να χρησιμοποιούμε τη δεύτερη μέθοδο.

Δεύτερη Μέθοδος: Απαιτεί περισσότερο χρόνο προετοιμασίας αλλά σαφώς είναι πιο πλήρης.

Και στη μέθοδο αυτή είναι απαραίτητη η κωδικοποίηση των απαντήσεων όπως ακριβώς εξηγήσαμε στην προηγούμενη περίπτωση.

Στη συνέχεια εισάγουμε τους κωδικούς στη στήλη που επιλέξαμε ακολουθώντας τα επόμενα βήματα:

- Πατάμε στο κουμπί ***Variable View*** του **Data Editor** (εικόνα 1.1) ή κάνουμε διπλό κλικ στην επικεφαλίδα της στήλης και εμφανίζεται η εικόνα 2.1.
- Στη στήλη ***Name*** γράφουμε τον αριθμό της ερώτησης
- Στη στήλη ***Measure*** επιλέγουμε ***Ordinal***, εφόσον πρόκειται για ***ποιοτική μεταβλητή σε κλίμακα τάξης***.
- Πατάμε στην στήλη ***Type*** και εμφανίζεται η εικόνα 2.3.

✓ Επιλέγουμε ***Numeric*** και στην ένδειξη ***Width*** σημειώνουμε έναν αριθμό ανάλογα με το πλήθος των κωδικών. Στην περίπτωση που το πλήθος των κωδικών είναι μονοψήφιος αριθμός βάζουμε 1, ενώ αν είναι διψήφιος βάζουμε 2.

✓ Την ένδειξη ***Decimal Places*** την αφήνουμε κενή.

- Στη στήλη ***Label*** γράφουμε τον τίτλο της ερώτησης.
- Πατάμε στην στήλη ***Values*** και εμφανίζεται η εικόνα 2.5.

✓ Στο παράθυρο ***Value*** γράφουμε τον κωδικό και στο παράθυρο ***Value Label*** την απάντηση που αντιστοιχεί στο συγκεκριμένο κωδικό.

✓ Πατάμε στην εντολή ***Add*** και εισάγεται στο μεγάλο παράθυρο ο κωδικός μαζί με την απάντηση που αντιστοιχεί σε αυτόν.

✓ Συνεχίζουμε με τον ίδιο τρόπο μέχρι να εισάγουμε όλες τις απαντήσεις με τους κωδικούς τους στο παράθυρο.

!!! *Αν θέλουμε να κάνουμε κάποια διόρθωση, επιλέγουμε το σημείο, διορθώνουμε και στη συνέχεια πατάμε* ***Change*** *στην εικόνα 2.5.*

!!! *Αν θέλουμε να διαγράψουμε κάποια απάντηση, την επιλέγουμε και στη συνέχεια πατάμε* ***Remove*** *στην εικόνα 2.5.*

- Στη συνέχεια πατάμε ***O.K***, επιστρέφουμε στην εικόνα 2.1 και ***Data View*** για την εισαγωγή των δεδομένων στα κελιά με την κωδικοποιημένη τους, βέβαια, μορφή

Τα δεδομένα αυτά θα εμφανίζονται σαν κωδικοί, δηλαδή αριθμοί, αλλά υπάρχει η δυνατότητα να εμφανίζονται και οι απαντήσεις όπως ακριβώς αυτές βρίσκονται στο ερωτηματολόγιο.

Για να πετύχουμε κάτι τέτοιο αρκεί να ακολουθήσουμε τα επόμενα βήματα:

- Πατάμε στο μενού ***View*** και
- Τσεκάρουμε την ένδειξη ***Value Labels***.

Μετά από αυτό, ενώ εμείς θα δίνουμε στον υπολογιστή τους κωδικούς οι οποίοι αντιστοιχούν στις απαντήσεις της ερώτησης, στην οθόνη θα εμφανίζονται ολόκληρες οι απαντήσεις. Για να εμφανίζονται μόνο, οι κωδικοί στο μενού ***View*** πρέπει να μην είναι τσεκαρισμένη η ένδειξη ***Value Labels***.

Εφαρμογή:

Έστω η ερώτηση:

Q1. Ποια είναι η κατάσταση της υγείας σας;

Οι προβλεπόμενες απαντήσεις:

Κάκιστη, Κακή, Μέτρια, Καλή, Άριστη

- Κωδικοποιούμε τις απαντήσεις, δίνοντας τους κωδικούς 1, 2, 3, 4 και 5 στην 1η, 2η, 3η ,4η και 5η απάντηση αντίστοιχα.
- Επιλέγουμε τη στήλη καταχώρησης των απαντήσεων με διπλό κλικ.
- Στη στήλη ***Name*** γράφουμε τον αριθμό της ερώτησης ***Q1***
- Στη στήλη ***Measurement*** σημειώνουμε ***Ordinal***
- Στη στήλη ***Type*** επιλέγουμε ***Numeric*** και στη θέση ***Width*** γράφουμε 1 γιατί το πλήθος των κωδικών είναι μονοψήφιος αριθμός, ενώ στη θέση ***Decimal Places*** γράφουμε 0 ή τίποτα
- Στη στήλη ***Label*** γράφουμε τον τίτλο της ερώτησης:

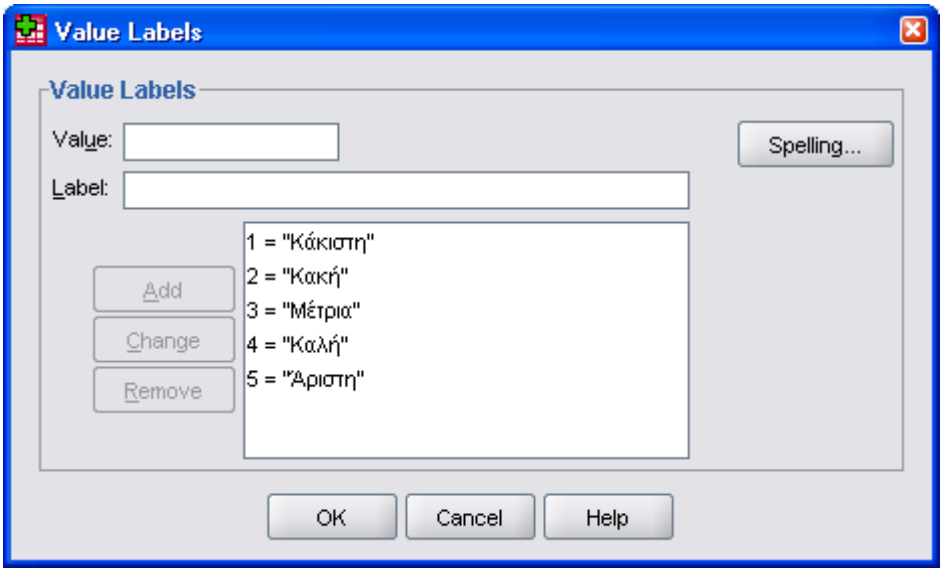
Q1. Ποια είναι η κατάσταση της υγείας σας;

- Πατάμε στη στήλη ***Values*** και

✓ Στην ένδειξη ***Value*** γράφουμε τον κωδικό 1 ενώ στην ένδειξη ***Value Label*** την απάντηση που αντιστοιχεί στο συγκεκριμένο κωδικό, δηλαδή: ***Κακή***.

✓ Πατάμε στην εντολή ***Add*** και εισάγεται στο μεγάλο παράθυρο ο κωδικός μαζί με την απάντηση που αντιστοιχεί σε αυτόν.

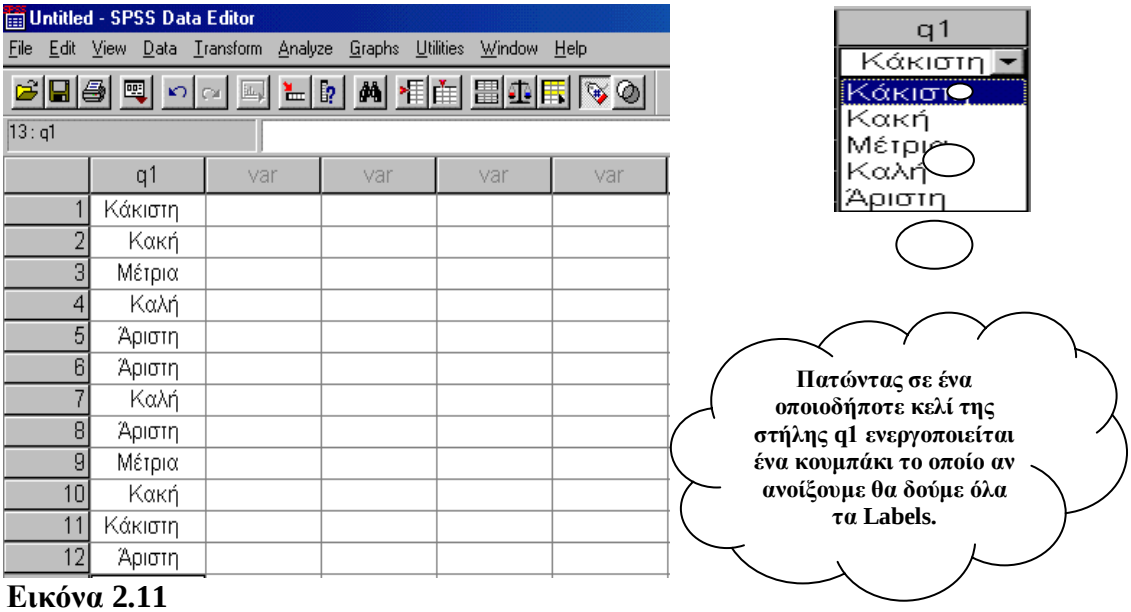
✓ Συνεχίζουμε με τον ίδιο τρόπο μέχρι να εισάγουμε όλες τις απαντήσεις με τους κωδικούς τους στο παράθυρο. Όταν τελειώσουμε θα έχουμε μία εικόνα σαν την επόμενη.



Εικόνα 2.10

- Στη συνέχεια ***O.K***, επιστροφή στην εικόνα 2.1 και από εκεί με ***Data View*** στον ***Data Editor*** για καταχώρηση δεδομένων.

Η μορφή της στήλης στον ***Data Editor*** θα είναι αυτή της εικόνας 2.11.



Εικόνα 2.11

2.1.4 Καταχώριση ερωτήσεων πολλαπλών απαντήσεων (Multiple Responses)

Σε ένα ερωτηματολόγιο υπάρχουν πολλές φορές ερωτήσεις στις οποίες έχουμε περισσότερες από μία επιλογές. Στις περιπτώσεις αυτές δεν είναι δυνατόν μέσα σε ένα κελί να καταχωρήσουμε όλες τις απαντήσεις, όχι γιατί δεν αρκεί ο χώρος, αλλά γιατί δεν είναι δυνατόν να επεξεργαστεί ο υπολογιστής στοιχεία καταχωρημένα με αυτόν τον τρόπο.

Για να καταγράψουμε δεδομένα τέτοιου είδους, μπορούμε να ακολουθήσουμε δύο διαφορετικές μεθόδους τις οποίες θα αναλύσουμε στη συνέχεια με παράλληλη εφαρμογή σε ένα συγκεκριμένο παράδειγμα.

Μέθοδος 1η : Ποιές από τις παρακάτω ιδιότητες των εργαζομένων αξιολογούνται;

Οι απαντήσεις, ήδη κωδικοποιημένες, από τις οποίες θα πρέπει να επιλέξουμε μία ή και περισσότερες είναι οι επόμενες:

1. Η Γνώση της εργασίας
2. Η Ποιότητα της εργασίας
3. Η Ποσότητα της εργασίας
4. Η Υπευθυνότητα και η Εγκυρότητα κατά την εκτέλεση της εργασίας
5. Η Επιμέλεια και η Ακρίβεια
6. Οι Διαπροσωπικές σχέσεις
7. Η Αποτελεσματική χρήση του χρόνου
8. Οι Πρωτοβουλίες

- Από την εικόνα 1.1 επιλέγουμε ***Variable View*** και
- Στη στήλη ***Name*** δίνουμε τον αριθμό της ερώτησης (***d31***).
- Στη στήλη ***Type*** ορίζουμε ***numeric*** –ανεξάρτητα αν πρόκειται για ποιοτική μεταβλητή σε ονομαστική κλίμακα- και
- Στη στήλη ***Label*** δίνουμε τον τίτλο της ερώτησης.

"Ποιές από τις παρακάτω ιδιότητες των εργαζομένων αξιολογούνται;"

- Από τη στήλη ***Values*** εισάγουμε τους κωδικούς με τις αντίστοιχες απαντήσεις (**1=** Η Γνώση της εργασίας, **2=** Η Ποιότητα της εργασίας κ.λπ**),** όπως έχουμε μάθει σε προηγούμενη παράγραφο και κατόπιν
- Ακολουθούμε την ίδια διαδικασία για τη δεύτερη στήλη, όπου δίνουμε τον αριθμό της ερώτησης μαζί με ένα δεύτερο συνθετικό (***d32***) και στη συνέχεια
- Εισάγουμε τους κωδικούς με τις αντίστοιχες απαντήσεις
- Επαναλαμβάνουμε την ίδια διαδικασία μέχρι να τελειώσουν οι στήλες (όσες και οι επιτρεπόμενες επιλογές).

!!! *Συνήθως επιλέγονται τόσες στήλες όσες είναι και οι προβλεπόμενες απαντήσεις στο δεδομένο ερώτημα. Αν όμως δεν υπάρχει κανένα ερωτηματολόγιο στο οποίο σημειώθηκαν όλες οι απαντήσεις, τότε σίγουρα κάποια ή κάποιες από τις στήλες που αρχικά επελέγησαν θα είναι κενές δεδομένων, οπότε για οικονομία στηλών μπορούμε να τις διαγράψουμε.*

!!! *Για να αποφύγουμε να γράψουμε τα* ***Labels*** *τόσες φορές όσες είναι οι στήλες μπορούμε, αφού ολοκληρώσουμε την πρώτη στήλη, να κάνουμε*

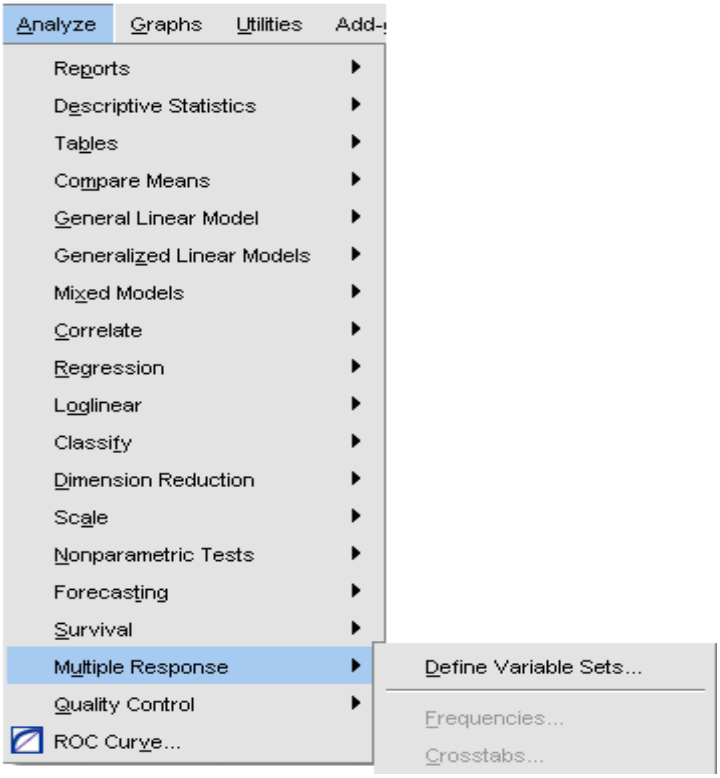
κλικ στην κεφαλίδα της πρώτης στήλης, στη συνέχεια ***Copy*** *και μετά* ***Paste*** *στις επόμενες στήλες.*

- Αρχίζουμε την εισαγωγή των δεδομένων με τον εξής τρόπο:

✓ Αν υπάρχει ***μία μόνον απάντηση***, αυτή καταχωρείται στην ***πρώτη στήλη*** ανεξαρτήτως του κωδικού της.

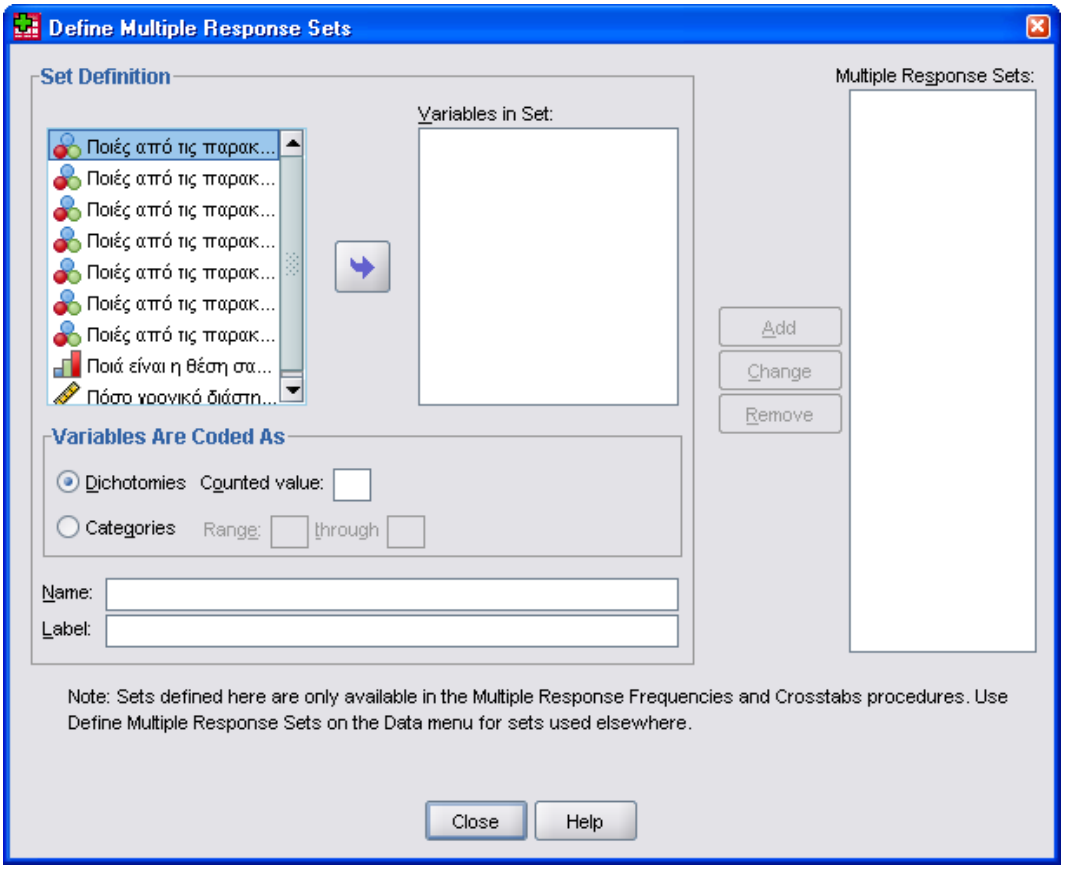
✓ Αν υπάρχουν ***δύο απαντήσεις***, αυτές με τους αντίστοιχους κωδικούς καταχωρούνται στην ***πρώτη*** και ***δεύτερη*** στήλη. Εκείνο που πρέπει να τονιστεί είναι ότι δεν πρέπει η απάντηση με κωδικό 3 να καταχωρηθεί στην $3^{η}$ στήλη, εκτός και αν κάποιος έδωσε την $1^{η}$ την $2^{η}$ και την $3^{η}$ απάντηση, οπότε αναγκαστικά η θέση της είναι η τρίτη στήλη. Στη συνέχεια

- Από το μενού ***Analyze*** επιλέγουμε
- ***Multiple Response*** και εμφανίζεται η εικόνα 2.12.



Εικόνα 2.12

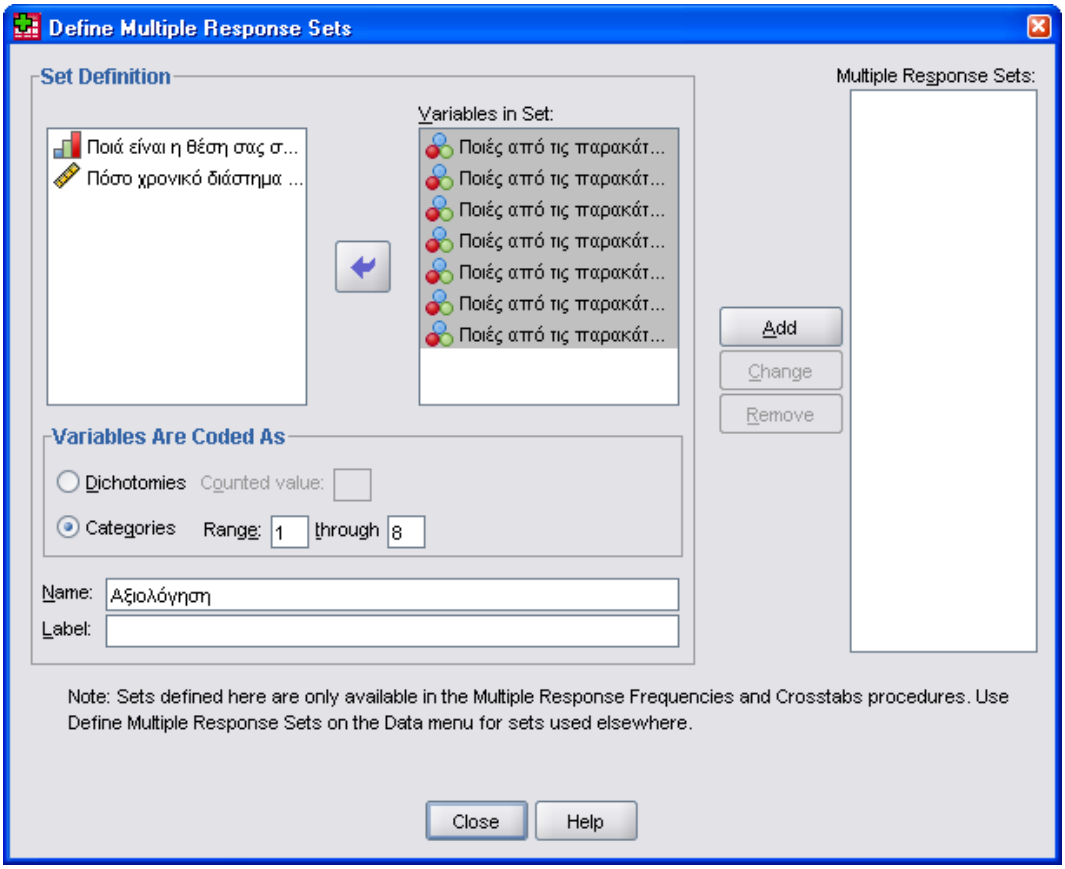
- Επιλέγουμε ***Define Variable Sets***, οπότε εμφανίζεται η επόμενη εικόνα 2.13.



Εικόνα 2.13

- Στο παράθυρο ***Set Definition*** εμφανίζονται όλες οι στήλες του ***Data Editor*** που περιέχουν δεδομένα και όχι μόνον αυτές που θέλουμε να αποτελέσουν set.

- Από αυτές τις στήλες θα επιλέξουμε αυτές οι οποίες θέλουμε να αποτελούν set και με κλικ στο βέλος θα τις μεταφέρουμε δεξιά στο παράθυρο ***Variables in Set***.



Εικόνα 2.14

- Τσεκάρουμε ***Categories*** και στη θέση ***range*** σημειώνουμε στο πρώτο παράθυρο τον κωδικό της πρώτης απάντησης και στο δεύτερο παράθυρο τον κωδικό της τελευταίας απάντησης. Στη συγκεκριμένη περίπτωση οι κωδικοί είναι από 1 έως 8.

- Στη θέση ***Name*** δίνουμε ένα όνομα το οποίο θα είναι πλέον το χαρακτηριστικό του set που δημιουργήσαμε

- Στη συνέχεια ***Add*** και το όνομα μπαίνει στο παράθυρο ***Multiple Response Sets.***

- ***Close*** και επιστροφή στον ***Data Editor***.

Μέθοδος 2η: Έστω η ίδια ερώτηση: Ποιές από τις παρακάτω ιδιότητες των εργαζομένων αξιολογούνται;

Οι απαντήσεις, ήδη κωδικοποιημένες, από τις οποίες θα πρέπει να επιλέξουμε μία ή και περισσότερες είναι οι επόμενες:

1. Η Γνώση της εργασίας
2. Η Ποιότητα της εργασίας
3. Η Ποσότητα της εργασίας
4. Η Υπευθυνότητα και η Εγκυρότητα κατά την εκτέλεση της εργασίας
5. Η Επιμέλεια και η Ακρίβεια
6. Οι Διαπροσωπικές σχέσεις
7. Η Αποτελεσματική χρήση του χρόνου
8. Οι Πρωτοβουλίες

- Από το ***Data View***
- Στη στήλη ***Name*** της πρώτης γραμμής δίνουμε τον αριθμό της ερώτησης (***d11***) και στη συνέχεια ***Label*** = Η Γνώση της εργασίας
- Στη στήλη ***Name***, της δεύτερης γραμμής γράφουμε ***d12*** και ***Label***= Η Ποιότητα της εργασίας
- Στη στήλη ***Name***, της τρίτης γραμμής γράφουμε ***d13*** και ***Label=*** Η Ποσότητα της εργασίας
- Στη στήλη ***Name***, της τρίτης γραμμής γράφουμε ***d14*** και ***Label=*** Η Υπευθυνότητα και η Εγκυρότητα κατά την εκτέλεση της εργασίας
- Στη στήλη ***Name***, της τρίτης γραμμής γράφουμε ***d15*** και ***Label=*** Η Επιμέλεια και η Ακρίβεια

- Στη στήλη ***Name***, της τρίτης γραμμής γράφουμε ***d16*** και ***Label*****=** Οι Διαπροσωπικές σχέσεις
- Στη στήλη ***Name***, της τρίτης γραμμής γράφουμε ***d17*** και ***Label*****=** Η Αποτελεσματική χρήση του χρόνου και τέλος
- Στη στήλη ***Name***, της τρίτης γραμμής γράφουμε ***d18*** και ***Label*****=** Οι Πρωτοβουλίες

!!! *Για όλες τις μεταβλητές στη στήλη* ***Type*** *επιλέξαμε* ***Numeric*** *χωρίς δεκαδικά.*

Αρχίζουμε την εισαγωγή των δεδομένων με τον εξής τρόπο: Αν υπάρχει επιλογή της πρώτης ερώτησης, αυτή καταχωρείται στην πρώτη στήλη με τον κωδικό 1. Αν δεν υπάρχει επιλογή, αυτή καταχωρείται με τον κωδικό 0. Το ίδιο επαναλαμβάνεται και με τις επόμενες στήλες οι οποίες αντιστοιχούν σε διαφορετικές απαντήσεις. Στη συνέχεια

- Από το μενού ***Analyze*** επιλέγουμε ***Multiple Response,*** εμφανίζεται η εικόνα 2.12 και
- ***Define Variable Sets***, οπότε εμφανίζεται η εικόνα 2.13.
- Στο παράθυρο ***Set Definition*** εμφανίζονται όλες οι στήλες του **Data Editor** που περιέχουν δεδομένα και όχι μόνον αυτές που θέλουμε να αποτελέσουν set.
- Από αυτές τις στήλες θα επιλέξουμε αυτές οι οποίες θέλουμε να αποτελούν set και με κλικ στο βέλος θα τις μεταφέρουμε δεξιά στο παράθυρο ***Variables in Set***.
- Επιλέγουμε ***Dichotomies*** και στη θέση ***Counted value*** βάζουμε το 1.
- Στη θέση ***Name*** δίνουμε ένα όνομα το οποίο θα είναι πλέον το χαρακτηριστικό του set που δημιουργήσαμε

- Στη συνέχεια ***Add***, και το όνομα μπαίνει στο παράθυρο ***Multiple Response Sets.***
- ***Close*** και επιστροφή στον ***Data Editor***.

!!! *Τα* ***sets*** *τα οποία δημιουργήθηκαν δεν είναι νέες μεταβλητές (Variables) και για το λόγο αυτό, όταν βγούμε από το αρχείο, θα χαθούν. Ένας τρόπος για να τα έχουμε διαθέσιμα κάθε φορά που μπαίνουμε στο αρχείο χωρίς να χρειάζεται να τα ξαναδημιουργούμε είναι ο επόμενος:*

Μετά τη δημιουργία του ***set*** πηγαίνουμε στο μενού ***Analyze*** και επιλέγουμε ***Multiple Response***. Από ***Frequencies ή Crosstabs***, αφού μεταφέρουμε το ***set*** στο δεξί παράθυρο, κάνουμε ***Paste*** και στη συνέχεια από το μενού ***File*** κάνουμε ***Save us*** και δημιουργούμε ένα αρχείο ***Syntax***. Το αρχείο αυτό είναι πλέον διαθέσιμο και για να πάρουμε τα αποτελέσματα από το ***set*** που θέλουμε, δεν έχουμε παρά να ανοίξουμε το αρχείο ***Syntax*** και από το μενού ***Run*** να επιλέξουμε ***All ή Select.***

!!! *Προσοχή, πρέπει να είναι ανοιχτή η βάση δεδομένων (****Data Editor****) από την οποία δημιουργήθηκε το* ***set.***

2.1.5 Ομαδοποίηση ποσοτικών δεδομένων

Πολλές φορές οι τιμές της μεταβλητής, είναι τόσο πολλές που κρίνεται απαραίτητη η ομαδοποίησή τους. Θα πρέπει στο σημείο αυτό να τονίσουμε ότι η δημιουργία ομάδων έχει σχέση με το πλήθος των τιμών της μεταβλητής και όχι με το πλήθος των υπό μελέτη μονάδων.

Η τεχνική της ομαδοποίησης αναπτύσσεται στη θεωρία της Στατιστικής (βλ. *Περιγραφική Στατιστική Ε. Δημητριάδη, σ.34*) και αποτελεί συνήθη τακτική κυρίως όταν εξετάζουμε οικονομικά μεγέθη.

Πολλές φορές στα ερωτηματολόγια υπάρχουν ήδη οι ομάδες και στην περίπτωση αυτή για την εισαγωγή των δεδομένων ακολουθούμε την επόμενη μέθοδο:

Πρώτη Μέθοδος:

- Κωδικοποιούμε τις ομάδες. Αν δηλαδή οι ομάδες που θα δημιουργήσουμε

είναι οι 100-110, 110-120, 120-130, 130-140, 140-150, τότε οι κωδικοί θα είναι 1, 2, 3, 4 και 5 αντίστοιχα για την κάθε ομάδα.

- Επιλέγουμε με διπλό κλικ τη στήλη στην οποία θα καταγράψουμε τους κωδικούς των δεδομένων
- Δίνουμε το όνομα που θέλουμε στη θέση ***name***
- Επιλέγουμε ***ordinal*** στη στήλη ***measure***

Πατάμε στην στήλη ***Label*** και γράφουμε τον τίτλο της ερώτησης

- Ενεργοποιούμε τη στήλη ***Values*** και στη γνωστή φόρμα γράφουμε τον κωδικό της πρώτης ομάδας στο παράθυρο ***value*** και στη θέση ***Value Label*** την ομάδα.
- Στη συνέχεια ***Add*** για να μπει στο παράθυρο ο κωδικός μαζί με την ομάδα και συνεχίζουμε με αυτό τον τρόπο μέχρι να τελειώσουμε με τις ομάδες.
- Τέλος ***O.K*** επανερχόμαστε στον ***Data Editor***, έτοιμοι για την εισαγωγή των κωδικών που αντιστοιχούν σε κάθε ομάδα.

Άλλες φορές τα ποσοτικά δεδομένα εισάγονται αναλυτικά στον ***Data Editor*** και στη συνέχεια θέλουμε να τα ομαδοποιήσουμε.

Αφού, λοιπόν, αποφασίσουμε τόσο για τον αριθμό των ομάδων, όσο και για τα άκρα των ομάδων αυτών, θα ακολουθήσουμε την επόμενη μέθοδο εισαγωγής δεδομένων στον υπολογιστή.

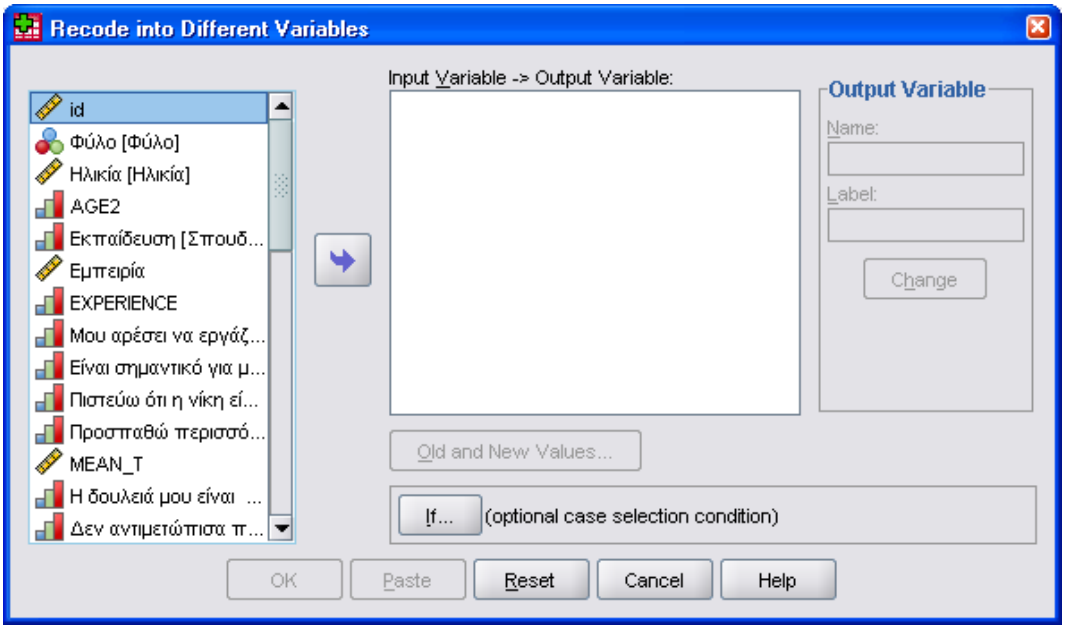
Δεύτερη Μέθοδος:

- Πατάμε στο μενού ***Transform*** και έχουμε την εικόνα 2.15

Εικόνα 2.15

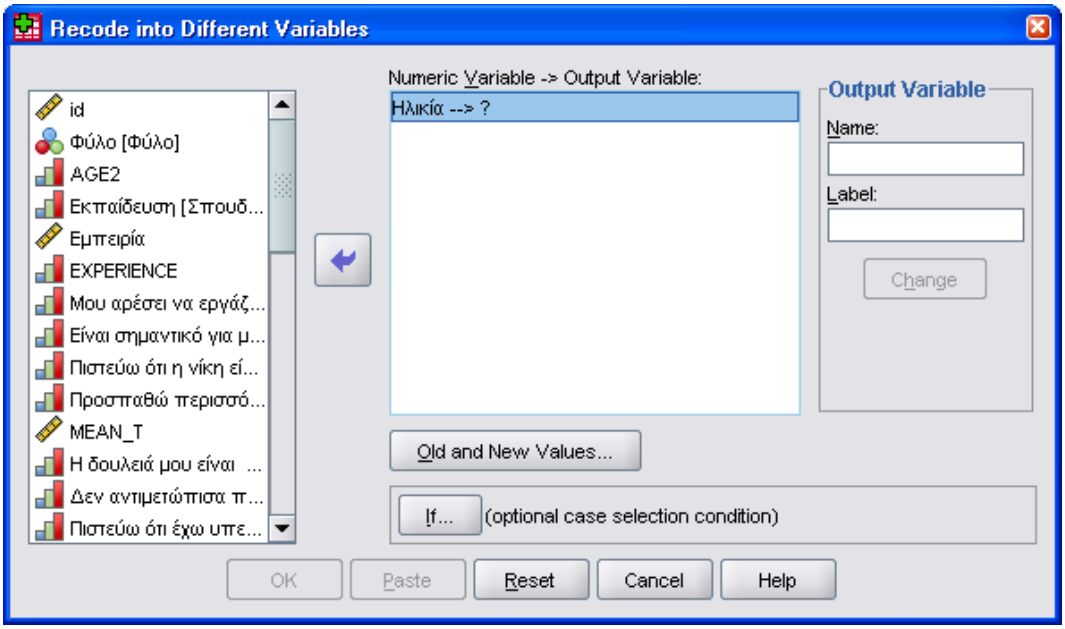
- Επιλέγουμε ***Recode into Same Variables***, αν θέλουμε η ομαδοποίηση να γίνει στην ίδια στήλη ή

- ***Recode into Different Variables***, αν θέλουμε η ομαδοποίηση να γίνει σε διαφορετική στήλη. Έστω ότι επιλέξαμε ***Recode into Different Variables***.



Εικόνα 2.16

- Επιλέγουμε από το παράθυρο αριστερά τη μεταβλητή που θέλουμε και με πάτημα στο ***μαύρο βέλος*** αυτή μεταφέρεται στο παράθυρο δεξιά, όπως φαίνεται στην επόμενη εικόνα 2.17.

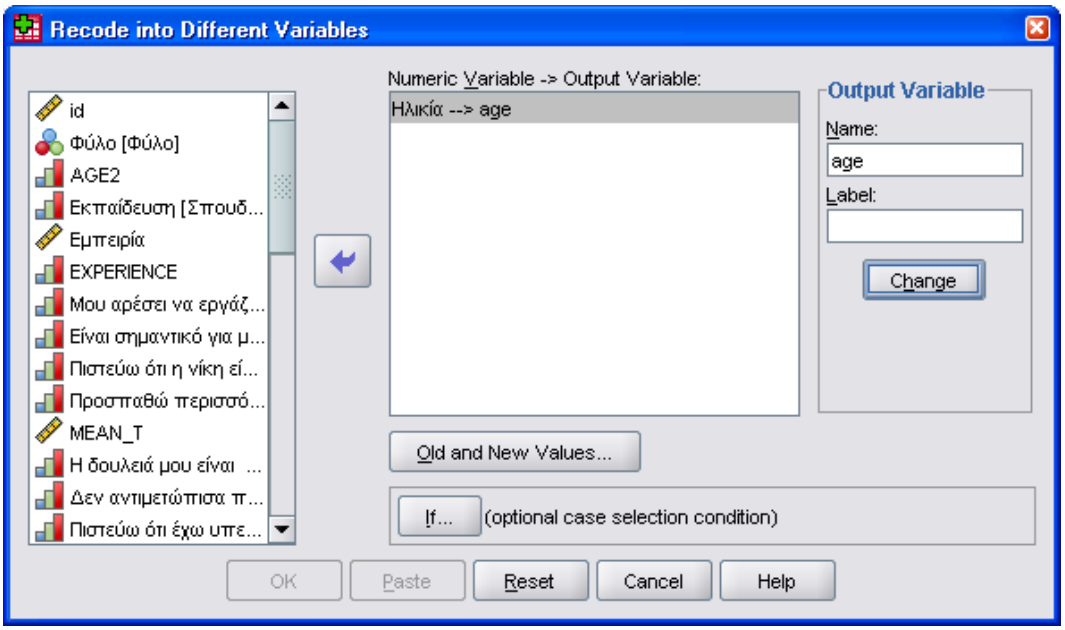


Εικόνα 2.17

- Στη θέση ***Name*** γράφουμε το όνομα που θα δώσουμε στη νέα στήλη και

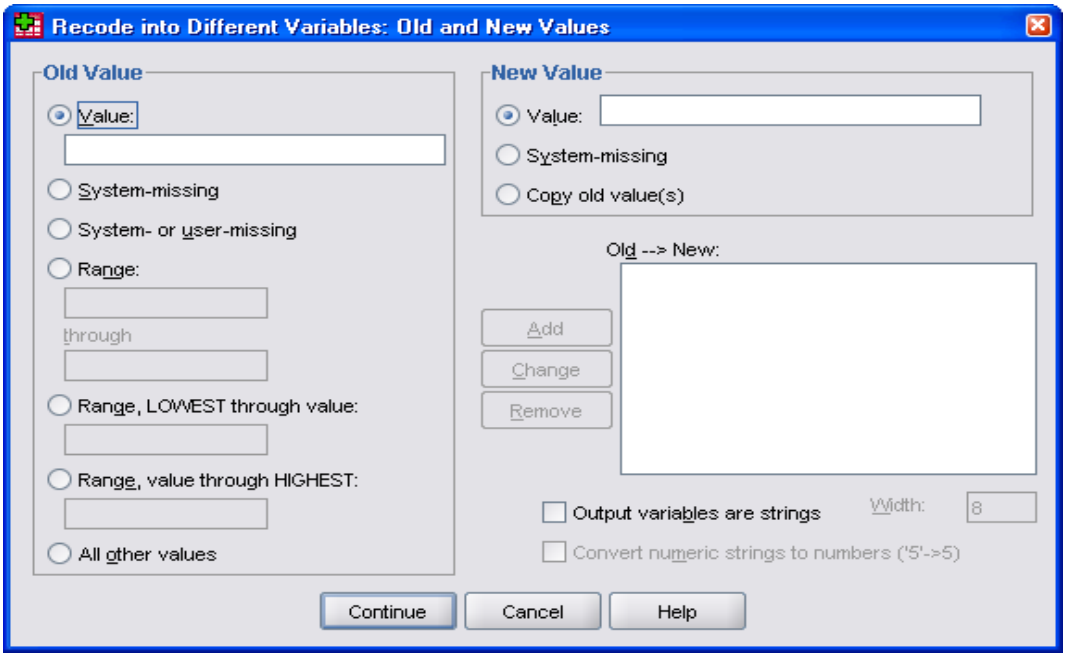
• Πατώντας στην ένδειξη ***Change*** το όνομα που δώσαμε, πηγαίνει δίπλα στο

όνομα της προηγούμενης μεταβλητής στη θέση του ερωτηματικού, όπως φαίνεται στην επόμενη εικόνα 2.18.



Εικόνα 2.18

• Στη συνέχεια πατάμε στην επιλογή ***Old and New Values*** και έχουμε την εικόνα 2.19.



Εικόνα 2.19

- Πατάμε στην επιλογή ***Range (Old Value)*** και στα δύο παράθυρα, γράφουμε

αριστερά και δεξιά το μικρό και το μεγάλο άκρο της πρώτης ομάδας. Για παράδειγμα, στο αριστερό παράθυρο 18 και στο δεξί 25.

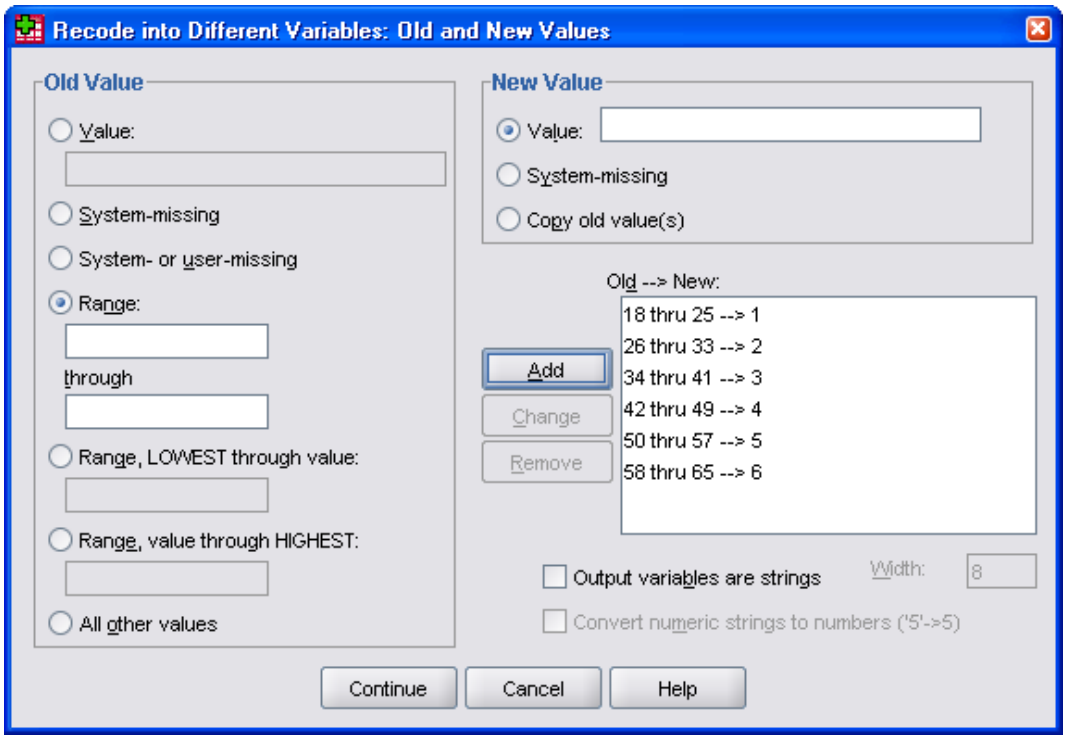
- Πατάμε δεξιά στην ένδειξη ***Value (New Value)*** και δίνουμε τον κωδικό ο οποίος αντικαθιστά την πρώτη ομάδα. Για την ομάδα 18-25 ο κωδικός είναι 1.

!!! *Αν η μεταβλητή είναι συνεχής, τότε οι ομάδες έχουν τη μορφή κλειστών συνεχόμενων διαστημάτων, με αποτέλεσμα το μεγάλο άκρο του κάθε διαστήματος- εκτός του τελευταίου- να συμπίπτει με το μικρό του επόμενου. Σε αυτή την περίπτωση οι τιμές της μεταβλητής που συμπίπτουν με το μεγάλο άκρο ενός διαστήματος - άρα και με το μικρό του επόμενου- προσμετρώνται στο πρώτο διάστημα.*

!!! *Αν η μεταβλητή είναι ασυνεχής, τότε τα όρια του κάθε διαστήματος είναι διακριτά και οι τιμές ανήκουν στο αντίστοιχο διάστημα.*

!!! *Στην περίπτωση που έχουμε ανοιχτή ομαδοποίηση, τότε στην εικόνα 2.19 επιλέγουμε το δεύτερο* ***Range*** *όπου αναγράφουμε στο μοναδικό παράθυρο (****LOWEST through****) τη μεγάλη τιμή του πρώτου ανοιχτού διαστήματος. Στη συνέχεια, στο δεύτερο* ***Range*** *με τα δύο παράθυρα γράφουμε τα όρια των επόμενων διαστημάτων, όπως ήδη έχουμε δει. Για τελευταίο διάστημα χρησιμοποιούμε το τρίτο* ***Range*** *και στο μοναδικό παράθυρο (****through HIGHEST****) αναγράφουμε τη μικρή τιμή του τελευταίου ανοιχτού διαστήματος.*

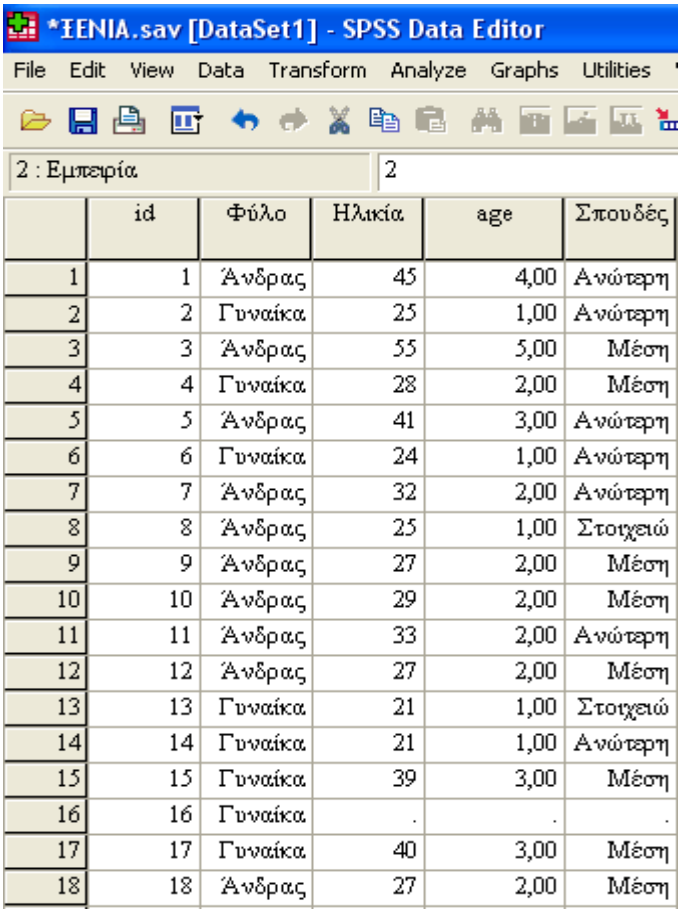
- Στη συνέχεια ***Add*** και εισαγωγή της ομάδας και του κωδικού της στο παράθυρο. Συνεχίζουμε με τον τρόπο αυτό μέχρι να τελειώσουν όλες οι ομάδες και έχουμε μία εικόνα σαν την επόμενη.



Εικόνα 2.20

- Στη συνέχεια ***Continue- O.K*** και εμφάνιση του ***Data Editor*** με μία νέα στήλη, στην οποία πλέον εμφανίζονται οι κωδικοί και όχι οι τιμές της μεταβλητής.

!!! *Επιλέγουμε ομαδοποίηση σε διαφορετική στήλη γιατί με τον τρόπο αυτό πετυχαίνουμε αυτό που θέλουμε (ομάδες) για καλύτερη παρουσίαση και συγχρόνως έχουμε διαθέσιμη και τη στήλη με τα απλά δεδομένα για καλύτερη επεξεργασία.*



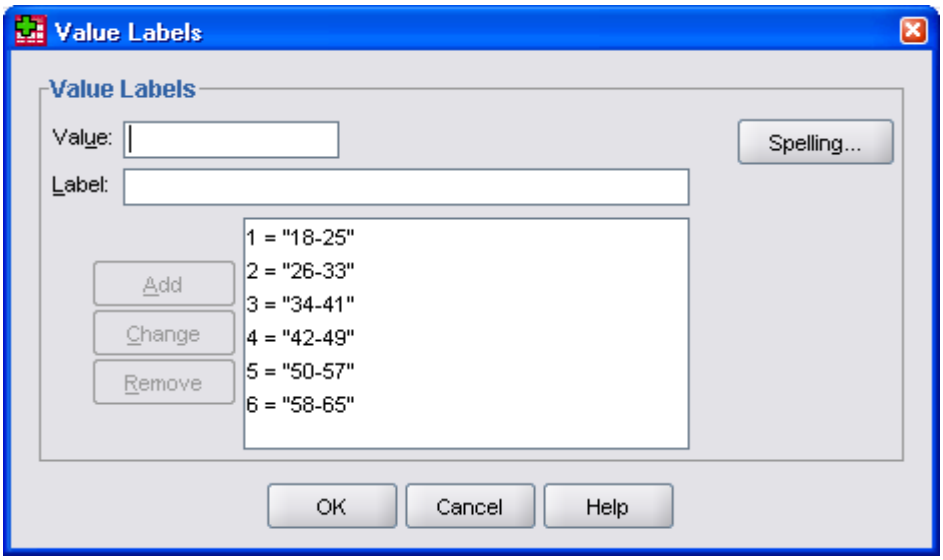
	id	Φύλο	Ηλικία	age	Σπουδές
1	1	Άνδρας	45	4,00	Ανώτερη
2	2	Γυναίκα	25	1,00	Ανώτερη
3	3	Άνδρας	55	5,00	Μέση
4	4	Γυναίκα	28	2,00	Μέση
5	5	Άνδρας	41	3,00	Ανώτερη
6	6	Γυναίκα	24	1,00	Ανώτερη
7	7	Άνδρας	32	2,00	Ανώτερη
8	8	Άνδρας	25	1,00	Στοιχειώ
9	9	Άνδρας	27	2,00	Μέση
10	10	Άνδρας	29	2,00	Μέση
11	11	Άνδρας	33	2,00	Ανώτερη
12	12	Άνδρας	27	2,00	Μέση
13	13	Γυναίκα	21	1,00	Στοιχειώ
14	14	Γυναίκα	21	1,00	Ανώτερη
15	15	Γυναίκα	39	3,00	Μέση
16	16	Γυναίκα	.	.	.
17	17	Γυναίκα	40	3,00	Μέση
18	18	Άνδρας	27	2,00	Μέση

Εικόνα 2.21

Αν τώρα θέλουμε αντί των κωδικών να εμφανίζονται οι ομάδες, όπως ακριβώς τις καθορίσαμε, πρέπει να ακολουθήσουμε την εξής διαδικασία:

- ***Variable view***
- ***Values*** και εμφάνιση της γνωστής εικόνας στην οποία στη θέση ***Value*** γράφουμε τον κωδικό 1, ο οποίος αντιστοιχεί στην ομάδα 18-25 την οποία θα αναγράψουμε στη θέση ***Value Label***.

- ***Add*** και συνεχίζουμε με την ίδια διαδικασία μέχρι να τελειώσουν οι κωδικοί και οι αντίστοιχες ομάδες. Η μορφή της εικόνας θα είναι η επόμενη.



Εικόνα 2.22

- ***O.K*** και εμφάνιση του ***Data Editor*** με τις ομάδες τιμών.

*ΞΕΝΙΑ.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities

2 : Εμπειρία 2

	id	Φύλο	Ηλικία	age	Σπουδές
1	1	Άνδρας	45	42-49	Ανώτερη
2	2	Γυναίκα	25	18-25	Ανώτερη
3	3	Άνδρας	55	50-57	Μέση
4	4	Γυναίκα	28	26-33	Μέση
5	5	Άνδρας	41	34-41	Ανώτερη
6	6	Γυναίκα	24	18-25	Ανώτερη
7	7	Άνδρας	32	26-33	Ανώτερη
8	8	Άνδρας	25	18-25	Στοιχειώ
9	9	Άνδρας	27	26-33	Μέση
10	10	Άνδρας	29	26-33	Μέση
11	11	Άνδρας	33	26-33	Ανώτερη
12	12	Άνδρας	27	26-33	Μέση
13	13	Γυναίκα	21	18-25	Στοιχειώ
14	14	Γυναίκα	21	18-25	Ανώτερη
15	15	Γυναίκα	39	34-41	Μέση

Εικόνα 2.23

2.2 Παρουσίαση δεδομένων (Output-Viewer)

Η παρουσίαση των δεδομένων, μετά την εισαγωγή τους στον ***Data Editor***, μπορεί να γίνει με συγκεντρωτικούς πίνακες κατανομής συχνοτήτων αλλά και γραφήματα. Με τον τρόπο αυτό αποκτάται μια γενικότερη εικόνα των δεδομένων εκ μέρους του ερευνητή, ενώ συγχρόνως γίνεται ευκολότερα κατανοητή η δομή των δεδομένων από άτομα μη ειδικά στην Στατιστική.

Η παρουσίαση των δεδομένων με μορφή πινάκων κατανομής συχνοτήτων αναλύεται στη συνέχεια αυτού του κεφαλαίου, ενώ η δημιουργία γραφημάτων περιγράφεται αναλυτικά στο κεφάλαιο 13.

2.2.1 Πίνακες Συχνοτήτων (Frequency Tables)

Οι πίνακες συχνοτήτων, ειδικότερα για ποιοτικές μεταβλητές για τις οποίες δεν έχουμε τη δυνατότητα υπολογισμού των βασικών στατιστικών μέτρων, είναι πολύ χρήσιμοι.

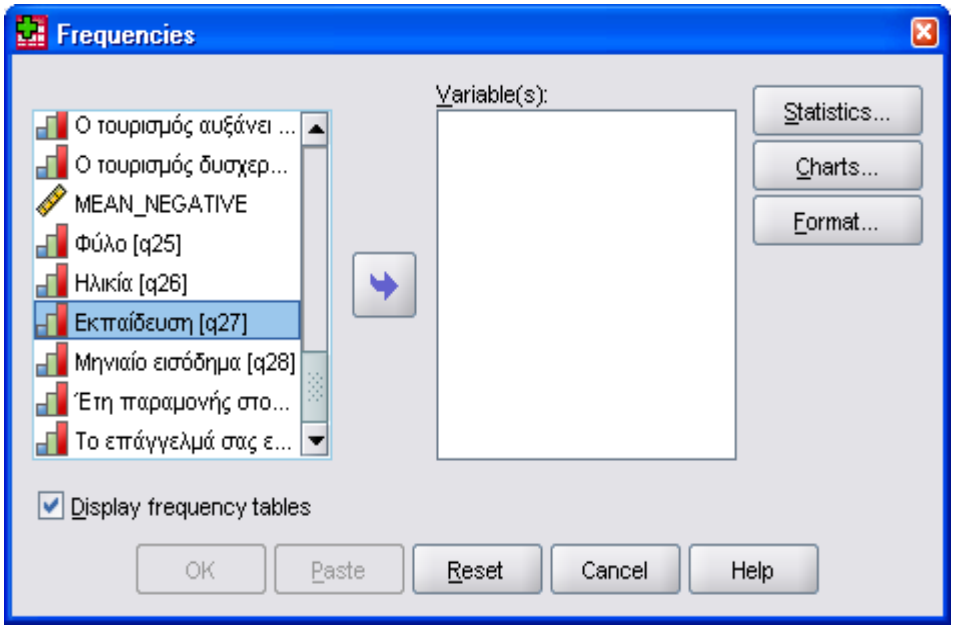
Στη συνέχεια αναλύεται η διαδικασία δημιουργίας απλών αλλά και σύνθετων πινάκων κατανομής συχνοτήτων.

2.2.1.1 Ποσοτική –Ποιοτική Μεταβλητή

Όταν οι μεταβλητής είναι ποσοτικές ή ποιοτικές, μετά την εισαγωγή τους στον ***Data Editor***, για να πάρουμε ένα ***συγκεντρωτικό πίνακα*** ο οποίος αναγράφει:

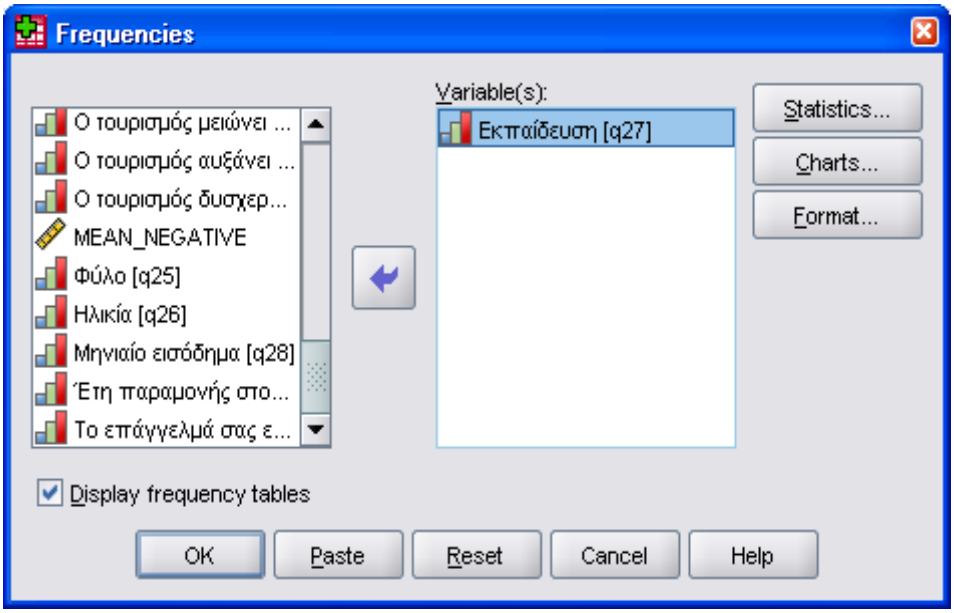
- ✓ ***Τη συχνότητα (Frequency)***
- ✓ ***Το ποσοστό (Percent)***
- ✓ ***Το πραγματικό ποσοστό (Valid Percent)*** και
- ✓ ***Το αθροιστικό ποσοστό (Cumulative percent)*** των τιμών των μεταβλητών, ακολουθούμε την παρακάτω διαδικασία.

- Πατάμε στο μενού ***Analyze*** του ***Data Editor.***
- Επιλέγουμε ***Descriptive statistics*** και στη συνέχεια ***Frequencies*** για να προκύψει η επόμενη εικόνα.



Εικόνα 2.24

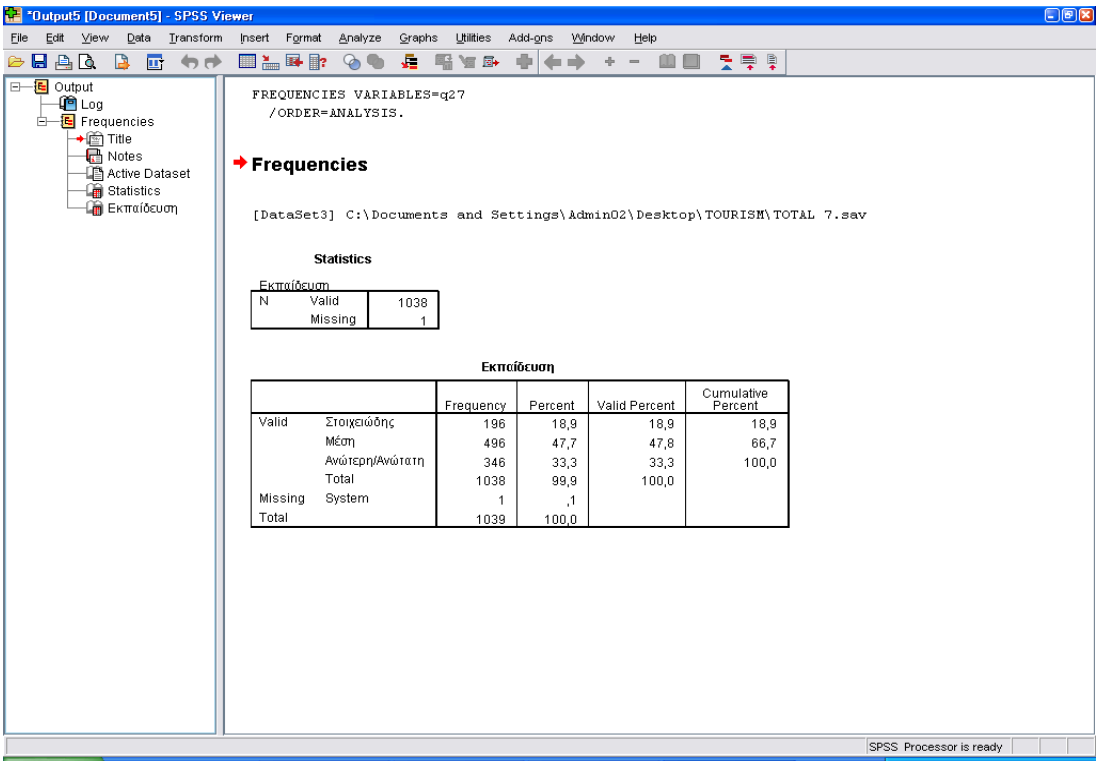
- Στο παράθυρο αριστερά εμφανίζονται όλες οι στήλες του ***Data Editor*** από τις οποίες θα επιλέξουμε, με απλό πάτημα, τη στήλη ή τις στήλες, τα δεδομένα των οποίων θέλουμε να παρουσιάσουμε.
- Αφού κάνουμε την επιλογή, πατάμε στο βέλος έτσι ώστε η στήλη ή οι στήλες που επιλέξαμε να μεταφερθούν στο παράθυρο δεξιά, όπως φαίνεται στην επόμενη εικόνα.



Εικόνα 2.25

- Τσεκάρουμε την ένδειξη ***Display frequency tables*** και στη συνέχεια

- ***O.K*** και εμφάνιση του ***Output*** με την παρακάτω μορφή.



Εικόνα 2.26

Στην κορυφή του πίνακα βλέπουμε τον τίτλο της ερώτησης.

Στην πρώτη στήλη του πίνακα οι ***τιμές*** της μεταβλητής, ενώ στη συνέχεια υπάρχουν οι στήλες των ***συχνοτήτων***, των ***ποσοστών***, των ***πραγματικών ποσοστών*** και των ***αθροιστικών ποσοστών (υπολογίζεται από τις πραγματικές συχνότητες)***.

!!! *Πολλές φορές στους πίνακες κατανομής συχνοτήτων εμφανίζεται η ένδειξη* ***Missing Values.*** *Με τον όρο αυτό αναφέρονται τα κελιά που δεν περιέχουν τιμές. Δηλαδή, αν σε κάποια ερώτηση δεν απαντήσουν κάποια άτομα, τα αντίστοιχα κελιά θα μείνουν κενά. Τα κελιά αυτά αποτελούν* ***missing values****. Επίσης* ***missing values*** *αποτελούν και οι τιμές τις οποίες εμείς ορίσαμε στη στήλη* ***Missing*** *του χώρου εργασίας* ***Variable View****.*

!!! *Οι τιμές* ***percent*** *προκύπτουν διαιρώντας την τιμή* ***Frequency*** *με την τιμή* ***total cases.***

!!! *Οι τιμές* ***Valid percent*** *προκύπτουν διαιρώντας την τιμή* ***Frequency*** *με την τιμή* ***valid cases.***

!!! *Οι τιμές* ***percent*** *και* ***Valid percent*** *είναι ίσες μόνο, όταν δεν υπάρχουν* ***Missing Values.***

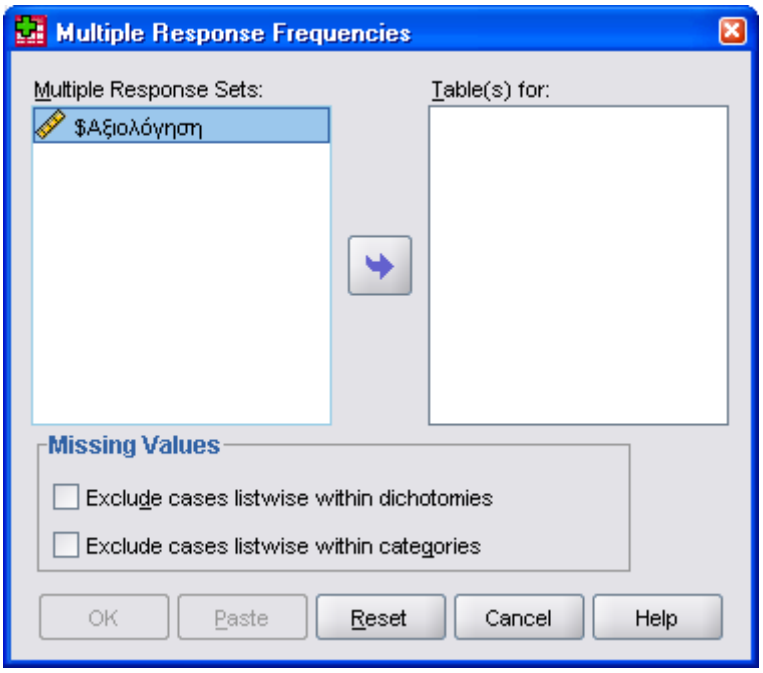
!!! *Οι τιμές* ***Cumulative percent*** *προκύπτουν ξεκινώντας με την πρώτη τιμή της στήλης* ***Valid percent*** *και προσθέτοντας κάθε φορά την επόμενη.*

2.2.1.2 Πολλαπλές απαντήσεις -Multiple Responses

Όταν έχουμε ερωτήσεις πολλαπλών απαντήσεων, για την καταχώρησή τους στον ***Data Edidor***, ακολουθούμε τη διαδικασία της παραγράφου 2.1.4 δημιουργώντας ***Sets***.

Όταν τώρα θέλουμε να παρουσιάσουμε τα αποτελέσματα αυτών των ερωτήσεων, ακολουθούμε την επόμενη διαδικασία.

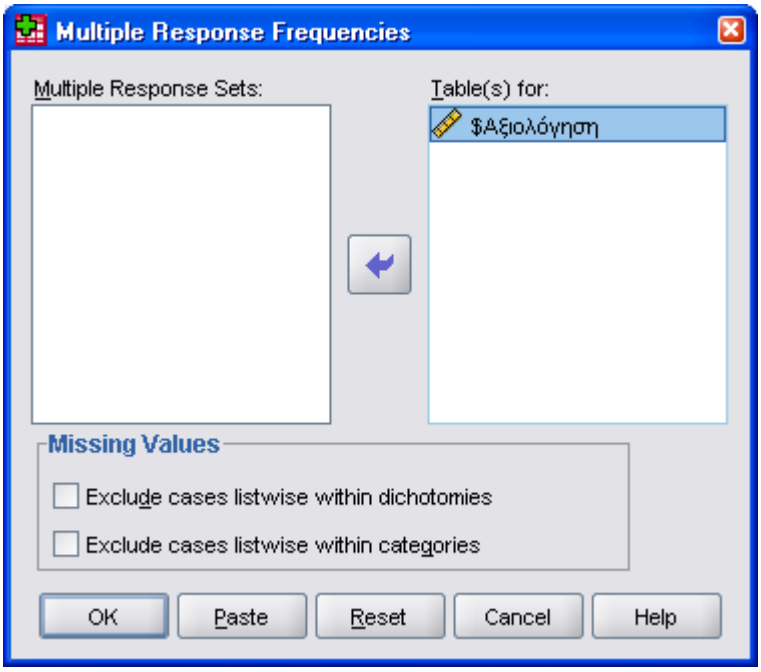
- Από το μενού ***Analyze*** επιλέγουμε ***Multiple Response*** και από εκεί με πάτημα στην επιλογή ***Frequencies*** εμφανίζεται η επόμενη εικόνα.



Εικόνα 2.27

Στο παράθυρο αριστερά φαίνονται όλα τα ***Sets*** που έχουμε δημιουργήσει. Στο παράδειγμά μας υπάρχει μόνο το ***Set «Αξιολόγηση»***.

- Τσεκάρουμε το ή τα ***Sets*** που θέλουμε να παρουσιάσουμε και πατώντας στο βελάκι το ή τα μεταφέρουμε στο παράθυρο δεξιά, όπως φαίνεται στην επόμενη εικόνα.



Εικόνα 2.28

- Πατάμε ***O.K*** και τα αποτελέσματα εμφανίζονται στους πίνακες 2.1 και 2.2.

Πίνακας 2.1: Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$Αξιολόγηση[a]	208	100,0%	0	,0%	208	100,0%

a. Group

Ο πίνακας 2.1 δίνει πληροφορίες σχετικά με το πλήθος των έγκυρων περιπτώσεων (Valid cases) και των κενών (missing) κελιών. Στην συγκεκριμένη περίπτωση οι περιπτώσεις είναι 208 και είναι όλες έγκυρες.

Στον πίνακα 2.2, που ακολουθεί, περιέχονται οι πληροφορίες που ενδιαφέρουν περισσότερο. Έτσι, έχουμε κατά σειρά:

- Τη στήλη με όλες τις διαθέσιμες, προς επιλογή, απαντήσεις
- Το πλήθος N των επιλογών κάθε απάντησης και το αντίστοιχο ποσοστό. Η στήλη ***percent of Responses*** προκύπτει διαιρώντας την τιμή ***N of responses*** με την τιμή ***Total Responses***.

- Η στήλη ***percent of cases*** η οποία προκύπτει διαιρώντας την τιμή ***N of responses*** με την τιμή ***N of valid cases*** του πίνακα 2.1.

Πίνακας 2.2: Frequencies

		Responses		Percent of Cases
		N	Percent	
$Αξιολόγηση[a]	Η γνώση της εργασίας	188	14,9%	90,4%
	Η ποιότητα της εργασίας	194	15,3%	93,3%
	Η ποσότητα της εργασίας	145	11,5%	69,7%
	Η υπευθυνότητα και η εγκυρότητα κατά την εκτέλεση της εργασί	180	14,2%	86,5%
	Η επιμέλεια και η ακρίβεια	157	12,4%	75,5%
	Οι διαπροσωπικές σχέσεις	134	10,6%	64,4%
	Η αποτελεσματική χρήση του χρόνου	131	10,4%	63,0%
	Οι πρωτοβουλίες	135	10,7%	64,9%
Total		1264	100,0%	607,7%

a. Group

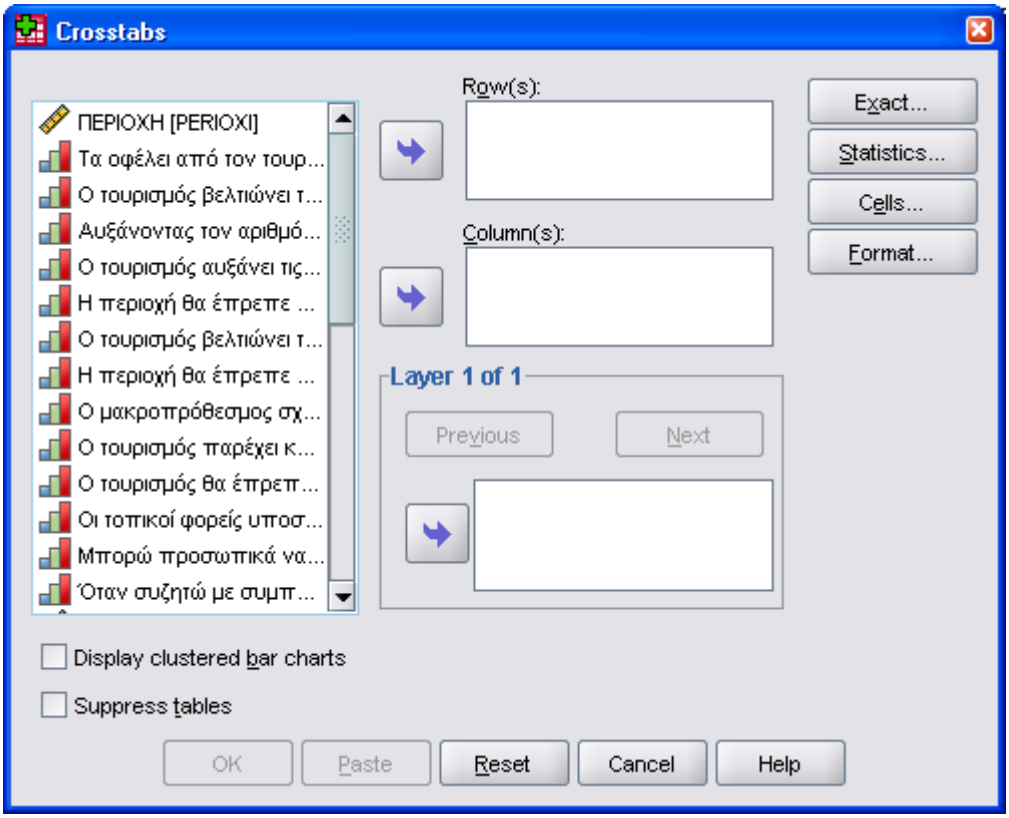
2.2.1.3 Διασταύρωση μεταβλητών- Crosstabs

Ιδιαίτερο ενδιαφέρον παρουσιάζουν στη ***Στατιστική Ανάλυση*** οι πίνακες ***διπλής ή πολλαπλής εισόδου***.

Για τη δημιουργία και παρουσίαση αυτών των πινάκων ακολουθούμε την παρακάτω διαδικασία.

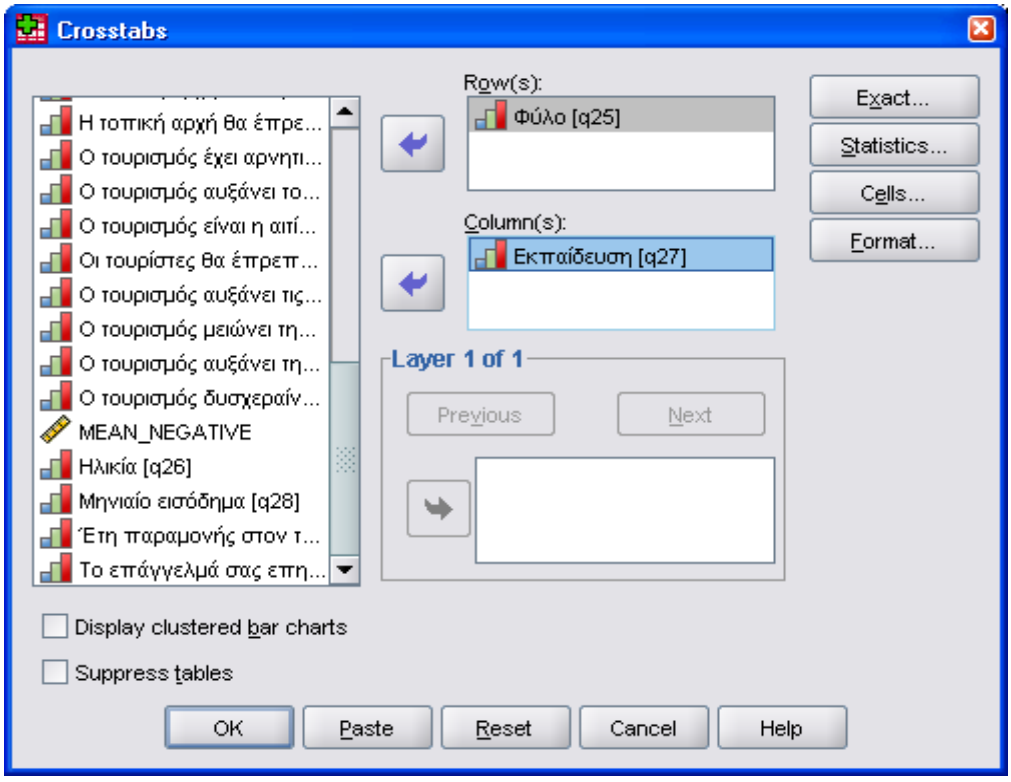
1. **Αν και οι δύο ερωτήσεις είναι απλές:**

- Από το μενού ***Analyze*** επιλέγουμε ***Descriptive Statistics*** και στη συνέχεια ***Crosstabs*** με άμεση εμφάνιση της εικόνας 2.29.



Εικόνα 2.29

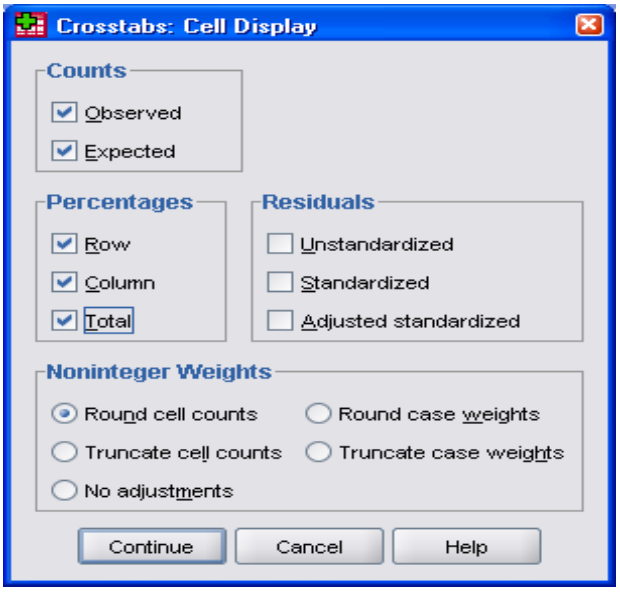
- Από το παράθυρο αριστερά επιλέγουμε την ερώτηση που θέλουμε να αποτελεί τις ***γραμμές (rows)*** του πίνακα και με πάτημα στο βελάκι την περνάμε στο παράθυρο δεξιά. Στη συνέχεια, πάλι από το παράθυρο αριστερά, επιλέγουμε την ερώτηση που θέλουμε να αποτελεί τις ***στήλες (columns)*** του πίνακα και με πάτημα στο βελάκι την περνάμε στο παράθυρο δεξιά. Έτσι πλέον η εικόνα 2.29 έχει πάρει την επόμενη μορφή.



Εικόνα 2.30

!!! *Αν θέλουμε πίνακα περισσοτέρων διαστάσεων, πρέπει στο παράθυρο Previous Layer 1 of 1 να εισάγουμε και άλλες μεταβλητές.*

- Πατάμε στην ένδειξη ***Cells*** και εμφανίζεται η επόμενη εικόνα.



Εικόνα 2.31

Αν στην ένδειξη ***Counts*** τσεκάρουμε:

- ***Observed***, στον πίνακα διπλής εισόδου θα εμφανίζονται μόνον οι ***εμπειρικές*** ή ***παρατηρούμενες*** συχνότητες των διασταυρωμένων απαντήσεων.
- ***Expected*** στον πίνακα διπλής εισόδου θα εμφανίζονται μόνον οι ***θεωρητικές*** ή ***αναμενόμενες*** συχνότητες των διασταυρωμένων απαντήσεων. Οι θεωρητικές τιμές προκύπτουν αν πολλαπλασιάσουμε το άθροισμα της στήλης με το άθροισμα της γραμμής και διαιρέσουμε με το συνολικό μέγεθος του πληθυσμού ή του δείγματος.

Από την ένδειξη ***Percentages*** μπορούμε να επιλέξουμε:

- ***Row***, αν θέλουμε να εμφανίζονται τα ποσοστά των ***γραμμών,***
- ***Column***, αν θέλουμε να εμφανίζονται τα ποσοστά των ***στηλών*** και
- ***Total***, αν θέλουμε να εμφανίζονται τα ***συνολικά*** ποσοστά

Αν στην ένδειξη ***Residuals- κατάλοιπα*** επιλέξουμε:

- ***Unstandardized*** στον πίνακα διπλής εισόδου θα εμφανίζονται τα ***μη τυποποιημένα κατάλοιπα*** (διαφορά εμπειρικών και θεωρητικών συχνοτήτων)
- ***Standardized*** στον πίνακα διπλής εισόδου θα εμφανίζονται τα ***τυποποιημένα κατάλοιπα***
- ***Adj. Standardized*** στον πίνακα διπλής εισόδου θα εμφανίζονται τα ***διορθωμένα τυποποιημένα*** κατάλοιπα.
- Πατάμε ***Continue***, επιστρέφουμε στον εικόνα 2.30 και με ***O.K*** εμφανίζεται ο πίνακας 2.3 με τα αποτελέσματα των επιλογών μας.

Πίνακας 2.3: Crosstabulation Φύλο*Εκπαίδευση

			Εκπαίδευση			
			Στοιχειώδης	Μέση	Ανώτερη/Ανώτατη	Total
Φύλο	Άνδρας	Count	114	325	177	616
		Expected Count	117.1	294.6	204.3	616.0
		% within Φύλο	18.5%	52.8%	28.7%	100.0%
		% within Εκπαίδευση	58.2%	65.9%	51.8%	59.7%
		% of Total	11.1%	31.5%	17.2%	59.7%
	Γυναίκα	Count	82	168	165	415
		Expected Count	78.9	198.4	137.7	415.0
		% within Φύλο	19.8%	40.5%	39.8%	100.0%
		% within Εκπαίδευση	41,8%	34,1%	48,2%	40,3%
		% of Total	8.0%	16.3%	16.0%	40.3%
Total		Count	196	493	342	1031
		Expected Count	196,0	493,0	342,0	1031,0
		% within Φύλο	19,0%	47,8%	33,2%	100,0%
		% within Εκπαίδευση	100,0%	100,0%	100,0%	100,0%
		% of Total	19,0%	47,8%	33,2%	100,0%

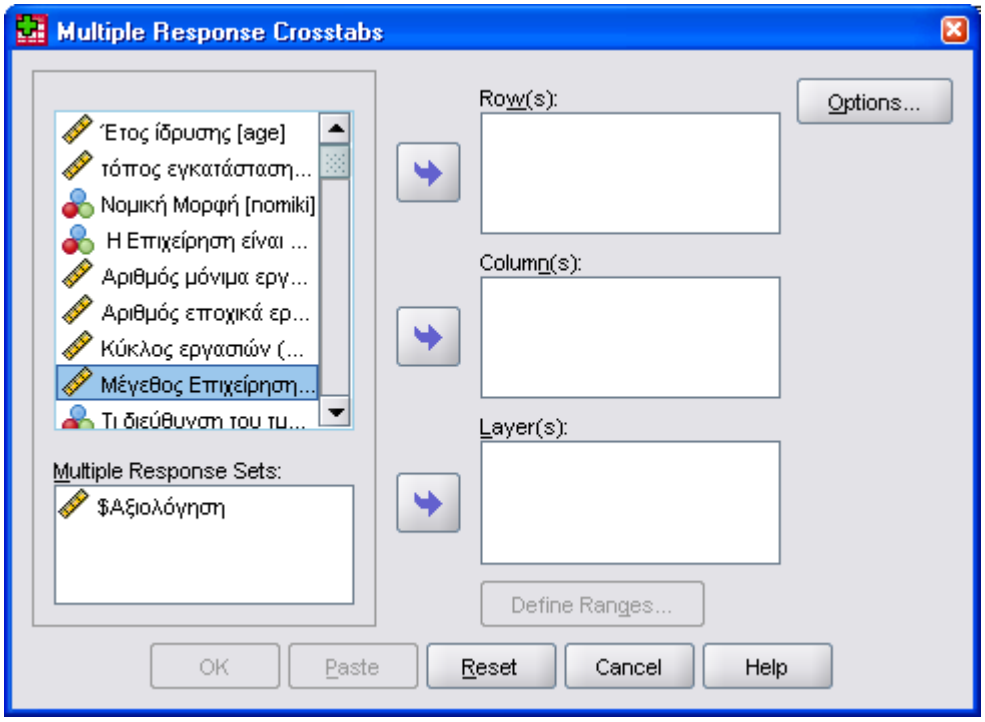
Παρατηρούμε ότι στην κορυφή του πίνακα αναφέρονται οι ***τίτλοι*** των μεταβλητών που διασταυρώσαμε.

Στην πρώτη στήλη εμφανίζονται οι ***τιμές*** της μίας μεταβλητής (φύλο), ενώ στην πρώτη γραμμή οι ***τιμές*** της άλλης μεταβλητής (εκπαίδευση). Στη συνέχεια μέσα στα κελιά υπάρχουν τα αποτελέσματα των δικών μας επιλογών. Δηλαδή ***εμπειρικές συχνότητες, θεωρητικές συχνότητες, ποσοστά γραμμής, στήλης και συνολικά ποσοστά***.

!!! *Η επιλογή Statistics, της εικόνας 2.29, θα αναπτυχθεί στο κεφάλαιο* 4.

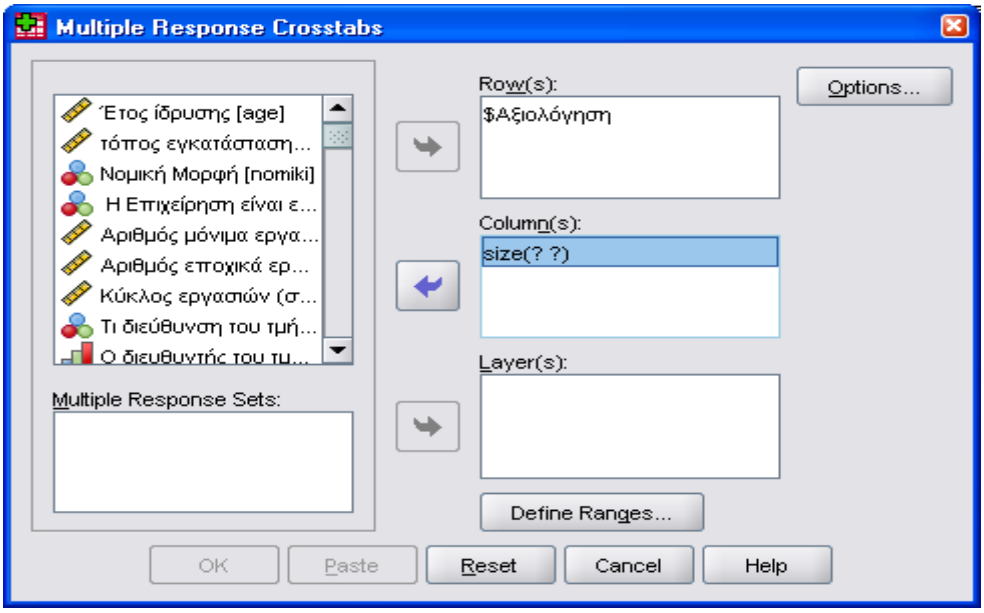
2. Αν η μία τουλάχιστον ερώτηση είναι πολλαπλών απαντήσεων:

- Από το μενού ***Analyze*** επιλέγουμε ***Multiple Response*** και στη συνέχεια
- ***Crosstabs*** και εμφανίζεται η επόμενη εικόνα.



Εικόνα 2.32

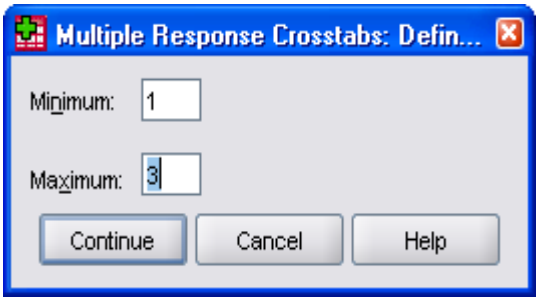
- Από το επάνω παράθυρο επιλέγουμε μία απλή μεταβλητή και την τοποθετούμε στη θέση ***Row*** ή ***Column*** και από το παράθυρο ***Multiple Response Sets*** επιλέγουμε το σετ το οποίο τοποθετούμε στη θέση ***Column*** ή ***Row***. Έστω ότι βάζουμε το σετ ***Αξιολόγηση*** στη θέση ***Row*** και την μεταβλητή ***Μέγεθος επιχείρησης*** στη θέση ***Column***. Μετά από αυτές τις ενέργειες θα έχουμε την επόμενη εικόνα.



Εικόνα 2.33

- Παρατηρούμε ότι στη θέση ***Columns***, δίπλα στον τίτλο της μεταβλητής ***size***, υπάρχει μία παρένθεση με δύο ερωτηματικά. Αυτό θα συμβαίνει πάντα στις απλές μεταβλητές. Συγχρόνως, η ένδειξη ***O.K*** δεν είναι ενεργοποιημένη. Για να ενεργοποιηθεί θα πρέπει:

- Να κάνουμε κλικ στο κουμπί ***Define Ranges*** και να εμφανιστεί η επόμενη εικόνα.

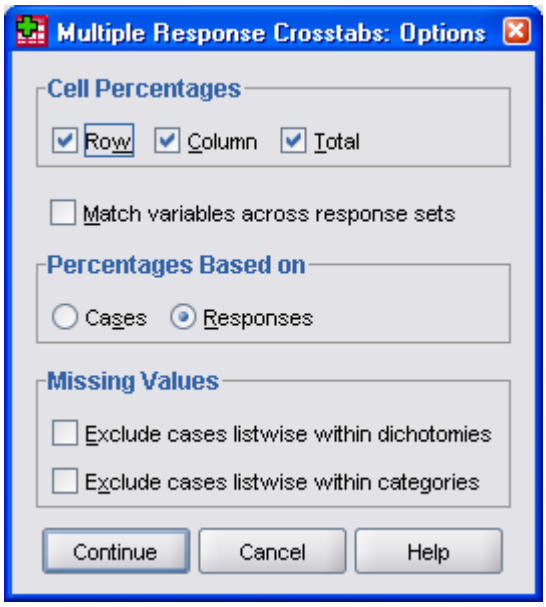


Εικόνα 2.34

- Στη θέση ***Minimum*** βάζουμε τον πρώτο κωδικό της μεταβλητής ***size***, ενώ στη θέση ***Maximum*** βάζουμε τον τελευταίο κωδικό της μεταβλητής ***size***.

- Στη συνέχεια ***Continue*** και γυρίζουμε στην εικόνα 2.33 όπου πλέον η ένδειξη ***O.K*** είναι ενεργοποιημένη.

- Πατάμε στο κουμπί ***Options*** και εμφανίζεται η εικόνα 2.35.



Εικόνα 2.35

- Πρώτα απ' όλα στη θέση ***Percentages Based on:*** επιλέγουμε ***Cases*** ή ***Responses***, ανάλογα με το αν θέλουμε τα αποτελέσματα να στηρίζονται στις ***περιπτώσεις*** (πλήθος ερωτηματολογίων) ή στις ***απαντήσεις*** (πλήθος απαντήσεων). Προτείνεται η επιλογή ***Responses***.

- Στη θέση ***Cell Percentages*** επιλέγουμε:

✓ ***Row*** και /ή

✓ ***Column*** και /ή

✓ ***Total*** ανάλογα με τα ποσοστά που θέλουμε να εμφανίζονται και στη συνέχεια

✓ ***Continue***, επαναφορά στην εικόνα 2.33 και με

✓ ***O.K*** έχουμε τον πίνακα 2.4 με τα αποτελέσματα.

Σε κάθε κελί υπάρχουν 4 αριθμοί. Ο πρώτος από αυτούς εκφράζει τη συχνότητα του κάθε κελιού. Το 39, στο πρώτο κελί, δηλώνει ότι 39 μικρές επιχειρήσεις θεωρούν ότι η γνώση της εργασίας από τους εργαζομένους πρέπει να αξιολογείται. Το 20,9% είναι το ποσοστό γραμμής και προκύπτει διαιρώντας τη συχνότητα του κελιού με το

πλήθος των συνολικών απαντήσεων της γραμμής. Δηλαδή, στη συγκεκριμένη περίπτωση (39/187)*100=20,9%. Το 16% είναι το ποσοστό της στήλης και προκύπτει διαιρώντας τη συχνότητα του κελιού με το σύνολο των απαντήσεων της στήλης (39/244) και τέλος το 3,1% είναι το ποσοστό στο σύνολο των απαντήσεων και προκύπτει διαιρώντας τη συχνότητα του κελιού με το σύνολο των απαντήσεων (39/1259).

Πίνακας 2.4: Crosstabulation Αξιολόγηση*size

			Μέγεθος Επιχείρησης			
			Μικρή (<50)	Μικρομεσαία (50-249)	Μεγάλη (>=250)	Total
a	Η γνώση της εργασίας	Count	39	88	60	187
		% within $Αξιολόγηση	20,9%	47,1%	32,1%	
		% within size	16,0%	14,7%	14,4%	
		% of Total	3,1%	7,0%	4,8%	14,9%
	Η ποιότητα της εργασίας	Count	41	89	63	193
		% within $Αξιολόγηση	21,2%	46,1%	32,6%	
		% within size	16,8%	14,9%	15,1%	
		% of Total	3,3%	7,1%	5,0%	15,3%
	Η ποσότητα της εργασίας	Count	26	74	44	144
		% within $Αξιολόγηση	18,1%	51,4%	30,6%	
		% within size	10,7%	12,4%	10,6%	
		% of Total	2,1%	5,9%	3,5%	11,4%
	Η υπευθυνότητα και η εγκυρότητα κατά την εκτέλεση της εργασί	Count	36	88	56	180
		% within $Αξιολόγηση	20,0%	48,9%	31,1%	
		% within size	14,8%	14,7%	13,4%	
		% of Total	2,9%	7,0%	4,4%	14,3%
	Η επιμέλεια και η ακρίβεια	Count	31	77	48	156
		% within $Αξιολόγηση	19,9%	49,4%	30,8%	
		% within size	12,7%	12,9%	11,5%	
		% of Total	2,5%	6,1%	3,8%	12,4%
	Οι διαπροσωπικές σχέσεις	Count	21	61	52	134
		% within $Αξιολόγηση	15,7%	45,5%	38,8%	
		% within size	8,6%	10,2%	12,5%	
		% of Total	1,7%	4,8%	4,1%	10,6%
	Η αποτελεσματική χρήση του χρόνου	Count	27	62	41	130
		% within $Αξιολόγηση	20,8%	47,7%	31,5%	
		% within size	11,1%	10,4%	9,8%	
		% of Total	2,1%	4,9%	3,3%	10,3%
	Οι πρωτοβουλίες	Count	23	59	53	135
		% within $Αξιολόγηση	17,0%	43,7%	39,3%	
		% within size	9,4%	9,9%	12,7%	
		% of Total	1,8%	4,7%	4,2%	10,7%
Total		Count	244	598	417	1259
		% of Total	19,4%	47,5%	33,1%	100,0%

Κεφάλαιο 3

Βασικά Στατιστικά Μέτρα

Chapter 3

Basic Statistics

3. Εισαγωγή

Η επεξεργασία των δεδομένων συνίσταται, αρχικά, στον υπολογισμό κάποιων βασικών στατιστικών μέτρων (μέση τιμή, τυπική απόκλιση κ.λ.π) και στη συνέχεια, ανάλογα με το είδος των μεταβλητών αλλά και τον σκοπό της έρευνας, επεκτείνεται σε μεθόδους επαγωγικής στατιστικής και πολυμεταβλητής ανάλυσης δεδομένων.

Στην περίπτωση που οι μεταβλητές είναι ***ποσοτικές (διαστημικής ή αναλογικής κλίμακας)***, μπορούν να υπολογισθούν όλα τα ***Βασικά Στατιστικά μέτρα*** με πολύ απλό και σύντομο τρόπο. Επίσης, αυτά μπορούν να υπολογιστούν και στην περίπτωση **μεταβλητών ιεραρχικής κλίμακας**, με δεδομένο ότι αυτές αντιμετωπίζονται συνήθως ως μεταβλητές διαστημικής κλίμακας.

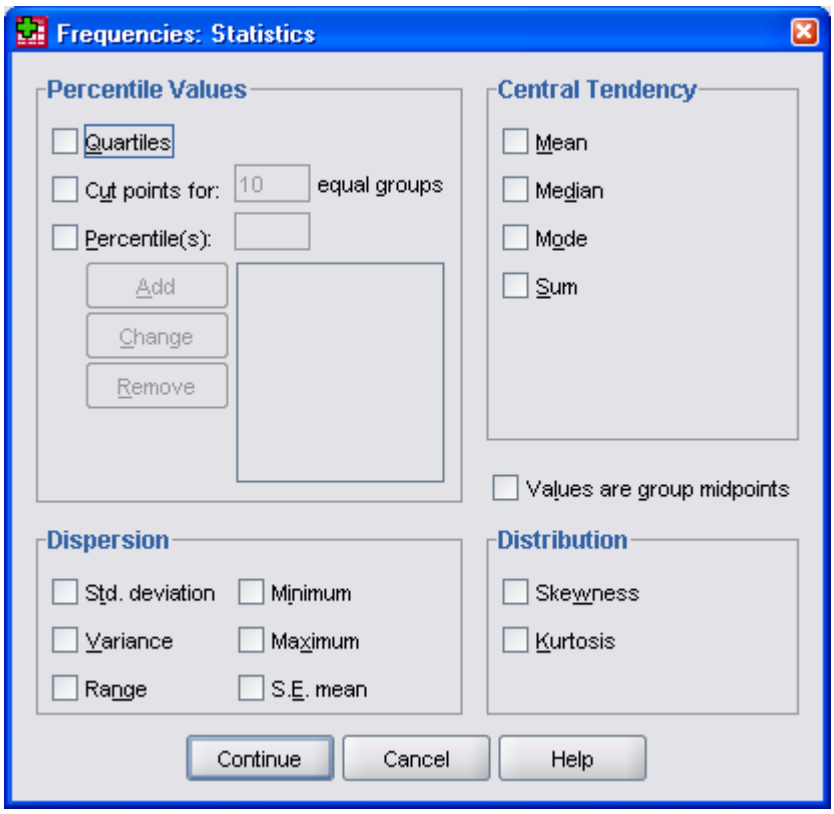
3.1 Βασικά Στατιστικά Μέτρα (Basic Statistics Measures)

Για τον υπολογισμό των βασικών στατιστικών μέτρων η διαδικασία την οποία ακολουθούμε περιγράφεται στη συνέχεια.

- Από το μενού ***Analyze*** επιλέγουμε ***Descriptive Statistics*** και στη συνέχεια
- ***Frequencies***, με αποτέλεσμα να εμφανιστεί η εικόνα 2.24.

- Επιλέγουμε τη μεταβλητή που θέλουμε για επεξεργασία και την μεταφέρουμε στο παράθυρο ***Variable***

- Επιλέγουμε ***Statistics*** στην εικόνα 2.24 και εμφανίζεται το επόμενο πλαίσιο διαλόγου.



Εικόνα 3.1

Το παραπάνω πλαίσιο διαλόγου χωρίζεται σε ***τέσσερα*** κύρια μέρη τα οποία θα αναλυθούν στη συνέχεια.

3.1.1 Εκατοστιαίες Τιμές (*Percentiles Values)*

- ***Quartiles (Τεταρτημόρια).*** Τα τεταρτημόρια χωρίζουν την κατανομή σε 4 μέρη που το καθένα περιέχει το 25% του συνόλου των παρατηρήσεων. Ενδεικτικά, ο τύπος με τον οποίο υπολογίζονται τα τεταρτημόρια για ομαδοποιημένες κατανομές είναι ο επόμενος: $\mathbf{M_{k/4} = X_i + \frac{\delta}{F_i}(\frac{k \cdot N}{4} - \Phi_{i-1})}$, όπου k=1,2,3. Η θέση του τεταρτημόριου

υπολογίζεται από τη σχέση kN/4 και εντοπίζεται στη στήλη των αθροιστικών συχνοτήτων. X_i είναι η μικρότερη τιμή του διαστήματος στο οποίο αντιστοιχεί η θέση, F_i η συχνότητα του ιδίου διαστήματος, δ το πλάτος του διαστήματος και Φ_{i-1} η αθροιστική συχνότητα του προηγούμενου διαστήματος.

- ***Cut points for ν equal groups (Χωρισμός των δεδομένων σε ν ομάδες*** που η κάθε μία περιέχει το ίδιο πλήθος παρατηρήσεων).

- ***Percentiles (Εκατοστημόρια)***. Τα εκατοστημόρια χωρίζουν την κατανομή σε 100 μέρη που το καθένα περιέχει το 1% του συνόλου των πληροφοριών. Αν τσεκάρουμε αυτή την ένδειξη, πρέπει στο παράθυρο να γράψουμε το ποσοστό που θέλουμε και στη συνέχεια να κάνουμε ***Add***. Αυτό μπορεί να επαναληφθεί για περισσότερες φορές. Ο τύπος υπολογισμού για ομαδοποιημένες κατανομές είναι ο επόμενος:

$\mathbf{M_{k/100} = X_i + \frac{\delta}{F_i}\left(\frac{k \cdot N}{100} - \Phi_{i-1}\right)}$, όπου k=1,2,3,…..,99. Η εφαρμογή του τύπου είναι ανάλογη αυτής των τεταρτημόριων.

3.1.2 Μέτρα Κεντρικής Τάσης (Central tendency)

- ***Mean (Αριθμητικός μέσος).*** Αριθμητικός Μέσος ή απλά Μέσος ν αριθμών είναι το πηλίκον του αθροίσματος των ν αριθμών προς ν. Ενδεικτικά: $\overline{X} = \frac{\sum_{i=1}^{\nu} X_i}{\nu}$, για αταξινόμητες παρατηρήσεις και

$\overline{X} = \frac{\sum_{i=1}^{\nu} F_i X_i}{\sum_{i=1}^{\nu} F_i}$ για ταξινομημένες και ομαδοποιημένες παρατηρήσεις.

- ***Median (Διάμεσος).*** Η διάμεσος περιγράφεται ως μία τιμή που χωρίζει το υψηλότερο μισό ενός δείγματος ή ενός πληθυσμού από το

χαμηλότερο μισό. Η διάμεσος ενός πεπερασμένου πλήθους αριθμών μπορεί να βρεθεί με την τακτοποίηση όλων των παρατηρήσεων από τη χαμηλότερη προς την υψηλότερη τιμή και την επιλογή της μεσαίας τιμής. Εάν υπάρχει άρτιο πλήθος παρατηρήσεων, η διάμεσος δεν είναι μοναδική. Έτσι, συχνά παίρνουμε ως διάμεσο τον μέσο των δύο μεσαίων παρατηρήσεων.¶Η Διάμεσος ομαδοποιημένων παρατηρήσεων υπολογίζεται με τη βοήθεια του τύπου: $\mathbf{M_{1/2} = X_i + \frac{\delta}{F_i}(\frac{N}{2} - \Phi_{i-1})}$.

Για τον υπολογισμό της Διαμέσου συμπληρώνουμε τη στήλη των αθροιστικών συχνοτήτων και προσδιορίζουμε τη θέση της από τη σχέση N/2. Βρίσκουμε τις δύο διαδοχικές αθροιστικές συχνότητες που περιέχουν την θέση που βρήκαμε και τραβάμε μία γραμμή μεταξύ αυτών. Επιλέγουμε την ομάδα κάτω από την γραμμή και $\mathbf{X}_i$ η μικρότερη τιμή της ομάδας κάτω από τη γραμμή, δ το πλάτος της ομάδας κάτω από τη γραμμή, F_i η συχνότητα της ομάδας κάτω από τη γραμμή, N/2 η θέση της διαμέσου και Φ_{i-1} η αθροιστική συχνότητα πάνω από τη γραμμή.

- ***Mode (Τύπος ή Σημείο Μέγιστης Συχνότητα).*** Σημείο Μέγιστης Συχνότητας ενός πλήθους τιμών είναι η τιμή με τη μεγαλύτερη συχνότητα. Το Σημείο Μέγιστης Συχνότητας ομαδοποιημένων παρατηρήσεων δίνεται από τον τύπο: $\mathbf{M_0 = X_i + \frac{\delta \cdot \Delta_1}{\Delta_1 + \Delta_2}}$ όπου, $\mathbf{X_i}$ το μικρό άκρο του διαστήματος με τη μεγαλύτερη συχνότητα, $\boldsymbol{\delta}$ το πλάτος του διαστήματος με τη μεγαλύτερη συχνότητα, $\boldsymbol{\Delta_1}$ η διαφορά της συχνότητας της προηγούμενης ομάδας από τη μέγιστη συχνότητα και $\boldsymbol{\Delta_2}$ η διαφορά της συχνότητας της επόμενης ομάδας από τη μέγιστη συχνότητα.

- ***Sum (Άθροισμα).*** Το άθροισμα των τιμών όλων των παρατηρήσεων.

3.1.3 Μέτρα Διασποράς (Dispersion)

- ***Std. Deviation (Τυπική απόκλιση).*** Η τετραγωνική ρίζα της Διακύμανσης
- ***Variance (Διακύμανση).*** Με τον όρο Διακύμανση, εννοούμε τον αριθμητικό μέσο των τετραγώνων των αποκλίσεων όλων των τιμών της μεταβλητής από τον αριθμητικό μέσο. Η Διακύμανση μπορεί να υπολογιστεί με τη χρήση των παρακάτω τύπων: $S^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N}$, όταν έχουμε αταξινόμητα δεδομένα και $S^2 = \frac{\sum_{i=1}^{\nu} F_i \cdot x_i^2}{\sum_{i=1}^{\nu} F_i} - \bar{X}^2$ όταν έχουμε ταξινομημένα ή ομαδοποιημένα δεδομένα.

- ***Range (Εύρος).*** Εύρος είναι η διαφορά της μικρότερης από την μεγαλύτερη τιμή των παρατηρήσεων μιας κατανομής: R=max-min

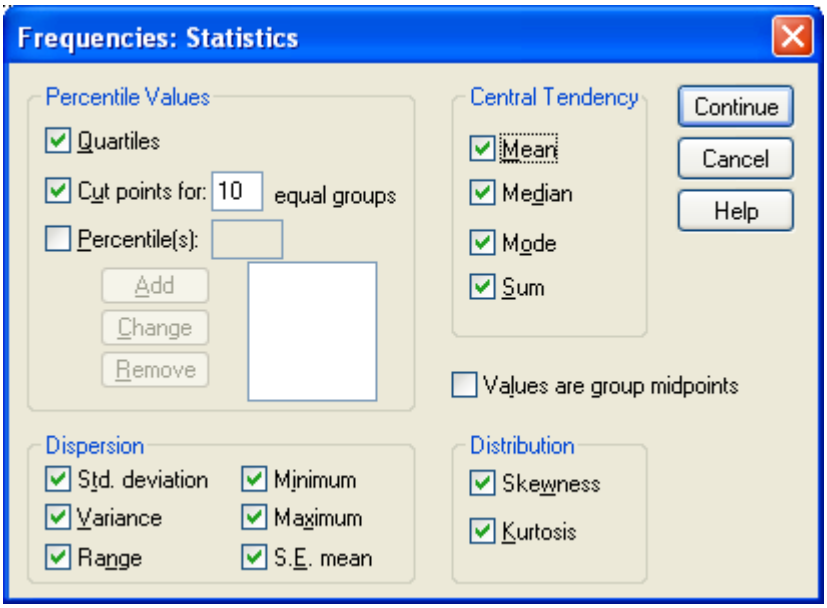
- ***S.E. Mean (Τυπικό σφάλμα αριθ. Μέσου)***

3.1.4 Κατανομή -Distribution

- ***Skewness (Ασυμμετρία).*** Η Ασυμμετρία είναι ένα μέτρο το οποίο μας πληροφορεί για το πόσο συμμετρικά είναι τοποθετημένες οι τιμές μιας μεταβλητής γύρω από τον αριθμητικό μέσο.

- ***Kurtosis (Κύρτωση).*** Η Κύρτωση χαρακτηρίζει κατά πόσο η καμπύλη μίας κατανομής είναι πεπλατυσμένη ή όχι, δεχόμενοι σαν κανονική κύρτωση αυτή της κανονικής καμπύλης.

Τσεκάρουμε τις παραμέτρους που θέλουμε να υπολογίσουμε, ενώ την ένδειξη ***Display Frequency Tables*** του πίνακα 2.24 την ενεργοποιούμε, αν δεν έχουμε ζητήσει ήδη πίνακα κατανομής συχνοτήτων και έχουμε την επόμενη εικόνα.



Εικόνα 3.2

- ***Continue*** και επιστροφή στην εικόνα 2.24.
- ***O.K*** και εμφάνιση του επόμενου πίνακα.

Πίνακας 3.1: Statistics

Ηλικία

N	Valid	132
	Missing	2
Mean		30,14
Std. Error of Mean		,696
Median		28,00
Mode		25[a]
Std. Deviation		8,000
Variance		63,997
Skewness		1,339
Std. Error of Skewness		,211
Kurtosis		2,064
Std. Error of Kurtosis		,419
Range		45
Minimum		15
Maximum		60
Sum		3978
Percentiles	10	22,00
	20	24,60
	25	25,00
	30	25,00
	40	27,00
	50	28,00
	60	29,00
	70	32,00
	75	33,00
	80	36,00
	90	41,70

a. Multiple modes exist. The smallest value is shown

Στον πίνακα αυτό υπάρχουν τα αποτελέσματα της επεξεργασίας των δεδομένων που ζητήσαμε.

!!! *Στην περίπτωση ποιοτικών μεταβλητών ονομαστικής κλίμακας δεν μπορούμε να υπολογίσουμε τα Στατιστικά μέτρα του πίνακα 3.1, παρά μόνο το Σημείο Μέγιστης Συχνότητας.*

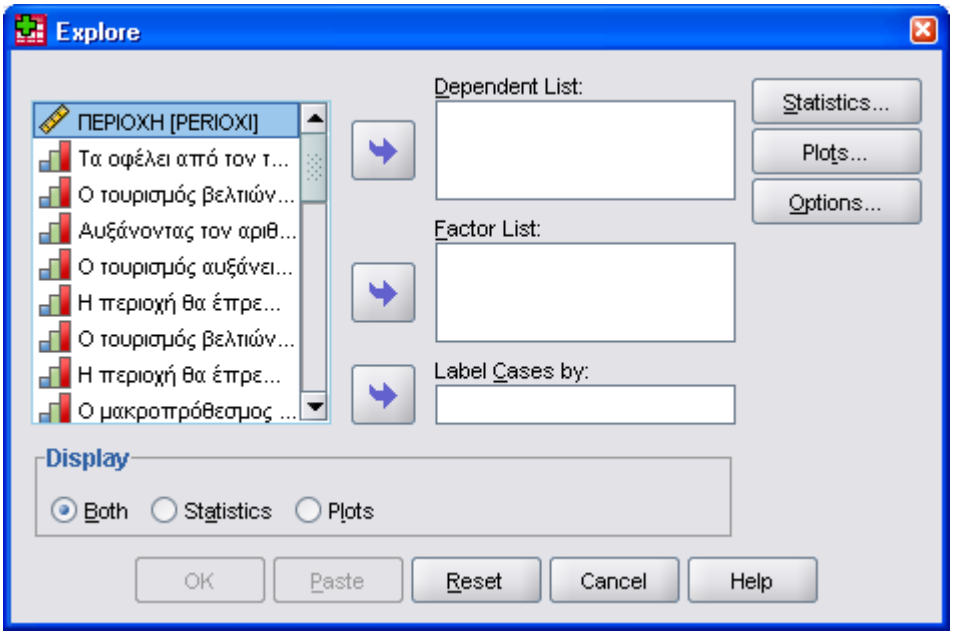
3.2 Διερεύνηση (Explore)

Πολλές φορές καταχωρούμε μία ποσοτική μεταβλητή σε μία στήλη και παράλληλα κάποιες άλλες ποιοτικές μεταβλητές σε άλλες στήλες. Είναι ιδιαίτερα **σημαντικό** και **χρήσιμο** να μπορούμε να έχουμε εύκολα και γρήγορα όλα τα **βασικά στατιστικά μέτρα** της ποσοτικής μεταβλητής, όχι όμως **στο σύνολο των περιπτώσεων, αλλά ομαδοποιημένα με βάση κάποια ποιοτική μεταβλητή**. Με τον τρόπο αυτό θα μπορούμε να κάνουμε συγκρίσεις των βασικών στατιστικών μέτρων της ίδιας μεταβλητής στις διαφορετικές ομάδες.

Έστω, για παράδειγμα, ότι έχουμε καταχωρήσει σε μια στήλη τις **ηλικίες** των κατοίκων από διάφορες περιοχές της Ελλάδας και σε μια άλλη στήλη το όνομα της **περιοχής.** Αν θέλουμε να κάνουμε σύγκριση των βασικών στατιστικών μέτρων (μέση τιμή, τυπική απόκλιση κ.λπ) της ηλικίας, στις διάφορες περιοχές, τότε η διαδικασία ***Explore*** είναι η κατάλληλη.

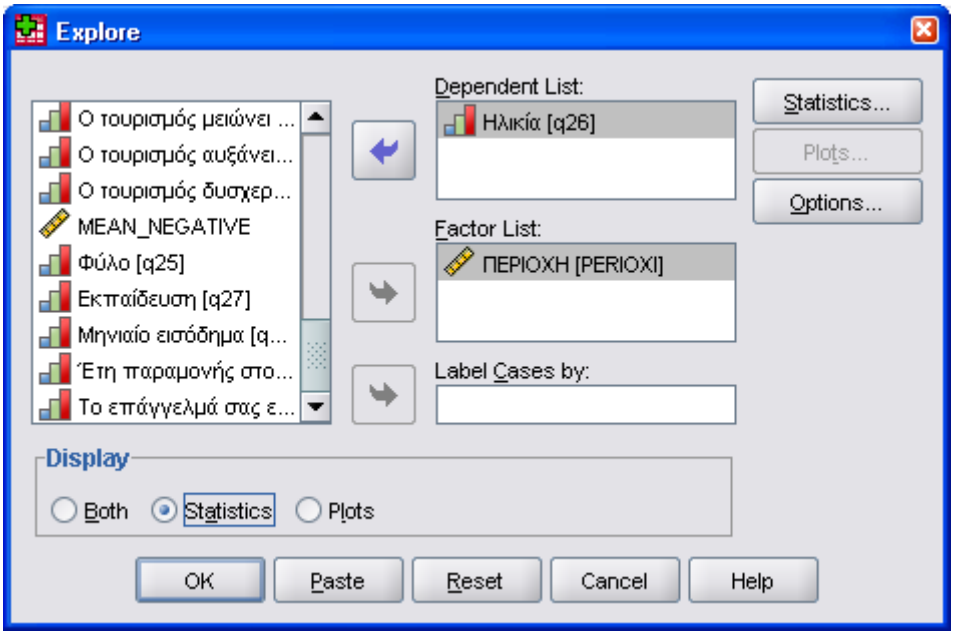
Στη συνέχεια θα δούμε τη διαδικασία την οποία πρέπει να ακολουθήσουμε. Από το μενού ***Analyze*** επιλέγουμε

- ***Descriptive Statistics*** και στη συνέχεια
- ***Explore*** και έχουμε την επόμενη εικόνα



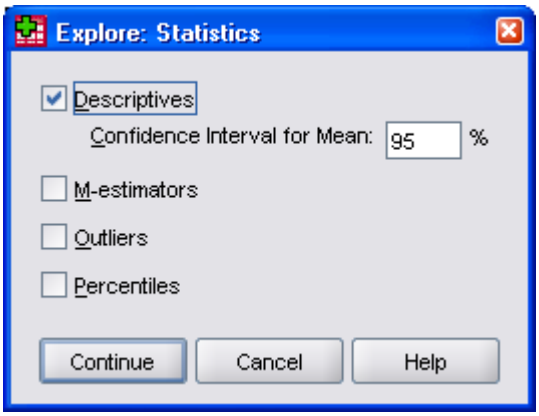
Εικόνα 3.3

- Στο παράθυρο ***Dependent List*** βάζουμε την ποσοτική μεταβλητή (στο παράδειγμά μας ***Ηλικία)***
- Στο παράθυρο ***Factor List*** βάζουμε την ποιοτική μεταβλητή, η οποία θα χωρίσει την ποσοτική μεταβλητή σε ομάδες (στο παράδειγμά μας ***Περιοχή***).
- Από το ***Display*** επιλέγουμε ***statistics***, αν θέλουμε μόνο τα βασικά στατιστικά μέτρα, ***Plots*** αν θέλουμε μόνο γράφημα και ***Both*** αν θέλουμε και τα δύο. Έστω ότι επιλέξαμε ***statistics***. Η εικόνα 3.4 θα έχει πλέον την επόμενη μορφή.



Εικόνα 3.4

- Στη συνέχεια πατάμε στο κουμπί ***Statistics*** και εμφανίζεται το επόμενο πλαίσιο διαλόγου.



Εικόνα 3.5

- Τσεκάρουμε ***Descriptives*** και στο παράθυρο ***Confidence interval for mean*** δίνουμε έναν αριθμό, συνήθως μεταξύ του 95 και του 100. Καθορίζουμε δηλαδή το επίπεδο εμπιστοσύνης του διαστήματος εμπιστοσύνης για τη μέση τιμή.

- Μπορούμε επίσης να τσεκάρουμε ***M-estimators***, για να πάρουμε εκτιμήσεις για τη μέση τιμή.

- ***Qutliers***, για να πάρουμε τις πέντε υψηλότερες και τις πέντε χαμηλότερες τιμές της κάθε ομάδας.
- ***Percentiles***, για να πάρουμε το 5ο, 10ο, 25ο, 50ο, 75ο, 90ο και 95ο εκατοστημόριο.

Στο παράδειγμά μας, έχουμε τσεκάρει μόνο ***Descriptives*** και θέλουμε ένα ***95% διάστημα εμπιστοσύνης*** για τη μέση τιμή των ομάδων.

- Στη συνέχεια πατάμε ***Continue***, επανερχόμαστε στην εικόνα 3.4, από την οποία επιλέγουμε ***O.K*** και έχουμε τα αποτελέσματα, όπως εμφανίζονται στον επόμενο πίνακα.

Πίνακας 3.2: Descriptives

		Ηλικία		
		Περιοχή		
		ΘΑΣΟΣ	ΛΙΜΝΗ ΠΛΑΣΤΗΡΑ	ΝΥΜΦΑΙΟ
Mean		43,88	41,03	38,54
95% Confidence Interval for Mean	Lower Bound	41,91	38,69	36,08
	Upper Bound	45,85	43,38	41,00
5% Trimmed Mean		43,74	40,51	37,92
Median		43,00	40,00	40,00
Variance		149,140	210,959	194,714
Std. Deviation		12,212	14,524	13,954
Minimum		22	18	17
Maximum		70	83	72
Range		48	65	55
Interquartile Range		20	21	22
Skewness		,170	,532	,394
Kurtosis		-,995	-,353	-,485

Κεφάλαιο 4

Έλεγχος Ανεξαρτησίας

Chapter 4

Test of Independence

4. Εισαγωγή

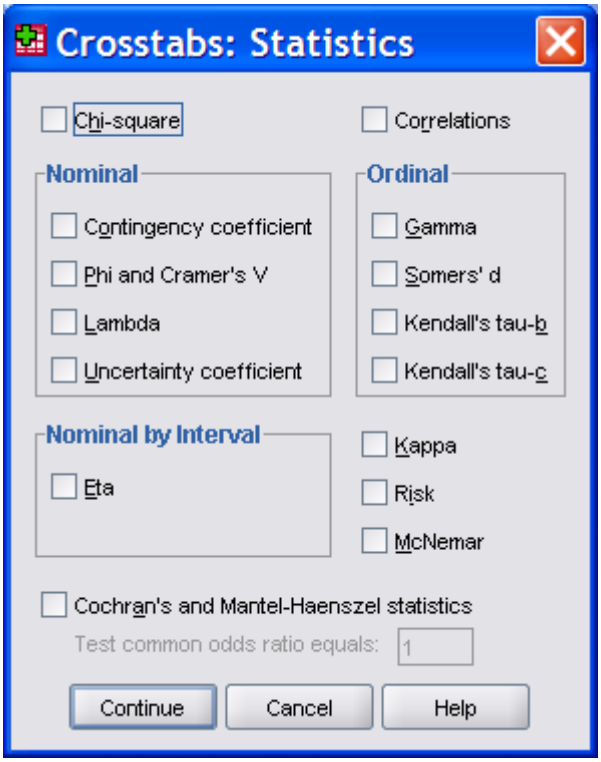
Η διαδικασία ***Cross-tabs*** διαμορφώνει πίνακες διπλής ή πολλαπλής εισόδου και παρέχει μια ποικιλία ελέγχων και μέτρων συσχέτισης μόνο για πίνακες διπλής εισόδου και ειδικά για ποιοτικές μεταβλητές (ονοματικής και ιεραρχικής κλίμακας). Η δομή του πίνακα και η κατηγορία των μεταβλητών καθορίζουν ποιος έλεγχος ή μέτρο πρέπει να χρησιμοποιηθεί.

4.1 Έλεγχος Ανεξαρτησίας- Test of Independence

Στην παράγραφο αυτή θα ασχοληθούμε αναλυτικά με τους ελέγχους ανεξαρτησίας και τα μέτρα συσχέτισης και εναρμόνισης που προσφέρει η διαδικασία ***Cross-tabs***, διαχωρίζοντας τις διάφορες περιπτώσεις ανάλογα με την κατηγορία των μεταβλητών.

4.1.1 Ποιοτικές Μεταβλητές Ονομαστικής κλίμακας (Nominal)

Ανεξαρτήτως των διαστάσεων του πίνακα, μετά την εισαγωγή των μεταβλητών ονομαστικής κλίμακας (π.χ φύλο και επάγγελμα) στα αντίστοιχα παράθυρα, όπως αυτή έχει αναπτυχθεί στην παράγραφο 2.2.1.3, πατάμε στο κουμπί ***Statistics*** της εικόνας 2.29 και έχουμε την εικόνα 4.1.



Εικόνα 4.1

❖ Επιλέγουμε ***Chi-Square*** και με ***Continue*** και ***O.K*** έχουμε τον επόμενο πίνακα σαν παράδειγμα.

Πίνακας 4.1: Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	32,048[a]	2	,000
Likelihood Ratio	33,832	2	,000
Linear-by-Linear Association	1,186	1	,276
N of Valid Cases	184		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 14,84.

!!! Ο πίνακας 4.1 συνοδεύεται συχνά και από μία υποσημείωση η οποία αναφέρει το ποσοστό των κελιών με αναμενόμενη (θεωρητική) συχνότητα μικρότερη του 5. Αν το ποσοστό αυτό είναι μεγαλύτερο του 20% τα τεστ που ακολουθούν δεν πρέπει να θεωρούνται αξιόπιστα. Στο

παρών παράδειγμα κανένα κελί δεν έχει αναμενόμενη συχνότητα μικρότερη του 5.

Οι δείκτες αυτού του πίνακα είναι κατά σειρά:

✓ ***Pearson Chi-Square (χ^2):*** Στη θέση ***Value*** η τιμή του δείκτη, όπως αυτή προκύπτει με αντικατάσταση στον τύπο:

$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - \phi_{ij})^2}{\phi_{ij}}$, όπου f_{ij} οι πραγματικές συχνότητες και ϕ_{ij} οι θεωρητικές ή αναμενόμενες (expected) συχνότητες. Στη θέση ***df*** οι βαθμοί ελευθερίας του ελέγχου, οι οποίοι για το συγκεκριμένο δείκτη είναι ***(R-1)(C-1)***, όπου ***R*** και ***C*** οι διαστάσεις του πίνακα. Στη θέση ***Asymp. Sig*** η ισχύς του ελέγχου. Ο χ^2 έλεγχος ονομάζεται ***έλεγχος ανεξαρτησίας*** και η μ***ηδενική*** υπόθεση του ελέγχου είναι ότι οι μεταβλητές είναι ανεξάρτητες. Η ε***ναλλακτική*** υπόθεση είναι ότι οι μεταβλητές είναι εξαρτημένες. Αν λοιπόν η τιμή ***Sig.χ^2 < 0,05*** απορρίπτουμε την υπόθεση της ανεξαρτησίας, σε επίπεδο σημαντικότητας 5% και δεχόμαστε ότι οι μεταβλητές είναι εξαρτημένες. Αν ***χ^2 = 0*** τότε ***Sig.χ^2 = 1***. Γενικότερα για τον δείκτη χ^2 *του* Pearson θα μπορούσαμε να πούμε ότι εξασφαλίζει λίγες πληροφορίες σχετικά με το πώς είναι συσχετισμένες οι μεταβλητές ή πόσο ισχυρή είναι η σχέση. Επίσης το μέγεθος του χ^2 δεν εξαρτάται μόνο, από τη διαφορά εμπειρικών και θεωρητικών τιμών, αλλά και από το μέγεθος του δείγματος.

!!! *Στο παράδειγμά μας, η τιμή του χ^2 ελέγχου είναι 32,048 και είναι στατιστικά σημαντική καθώς* ***Asymp. Sig****.=0,000. Δηλαδή, οι μεταβλητές φύλο και επάγγελμα είναι εξαρτημένες.*

✓ ***Likelihood ratio:*** Εναλλακτική μορφή του προηγούμενου δείκτη που χρησιμοποιείται για ***λογαριθμο- γραμμικά μοντέλα*** (***log-linear models)***. Για μεγάλα δείγματα οι δύο αυτοί δείκτες δίνουν το ίδιο σχεδόν αποτέλεσμα.

✓ ***Linear-by-Linear Association***: Χρησιμοποιείται μόνο, όταν και οι δύο μεταβλητές είναι ποσοτικές. Σε αντίθετη περίπτωση δεν έχει νόημα. Στην πράξη είναι το τετράγωνο του συντελεστή συσχέτισης του ***Pearson***, πολλαπλασιασμένο με το μέγεθος του δείγματος ελαττωμένο κατά μία μονάδα.

Για μεταβλητές *Nominal* εκτός του δείκτη ***Chi-Square*** μπορούμε να πατήσουμε στη φόρμα ***Statistics*** της εικόνας 4.1 και να επιλέξουμε:

- ***Phi and Cramer's V***
- ***Contingency coefficient***
- ***Lambda and***
- ***Uncertainty coefficient.***

Στη συνέχεια ***Continue***, ***O.K*** και εμφάνιση του επόμενου πίνακα:

Πίνακας 4.2: Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,417	,000
	Cramer's V	,417	,000
	Contingency Coefficient	,385	,000
N of Valid Cases		184	

Οι δείκτες του πίνακα *(Symmetric measures)* είναι κατά σειρά:

✓ ***Phi***: Μέτρο συσχέτισης που παίρνει τιμές από 0 μέχρι 1, μπορεί όμως και να ξεπεράσει την τιμή 1.

✓ ***Cramer's V***: Μέτρο συσχέτισης που παίρνει τιμές μεταξύ του 0 και του 1 και ο τύπος τον οποίο χρησιμοποιούμε για τον υπολογισμό

του είναι: $\mathbf{C = \sqrt{\frac{\chi^2}{\nu(m-1)}}}$, όπου ν το πλήθος των παρατηρήσεων και m η μικρότερη από τις διαστάσεις του πίνακα.

✓ ***Contingency coefficient***: Μέτρο συσχέτισης με τιμές από 0 έως 1, αν και γενικά δεν φτάνει το 1. Για πίνακες πάνω από 4×4 η μέγιστη τιμή του είναι 0,87.

✓ Η στήλη ***Approx.Sig***. είναι η ισχύς του ελέγχου για τις τιμές των παραμέτρων κάτω από τη μηδενική υπόθεση ότι αυτές είναι 0.

!!! *Και οι τρεις δείκτες δίνουν διαφορετικό αποτέλεσμα και είναι 0 αν και μόνο αν ο δείκτης* ***Chi-Square*** *είναι 0.*

!!! *Σε κάθε περίπτωση, τιμές κοντά το 0 δείχνουν ασθενή εξάρτηση, ενώ τιμές κοντά στη μονάδα δείχνουν ισχυρή εξάρτηση.*

Ένας άλλος πίνακας που επίσης εμφανίζεται είναι ο πίνακας ***Directional Measures.***

Πίνακας 4.3: Directional Measures

			Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	,172	,077	2,080	,038
		Επάγγελμα Dependent	,144	,069	1,940	,052
		Φύλο Dependent	,215	,107	1,793	,073
	Goodman and Kruskal tau	Επάγγελμα Dependent	,081	,027		,000[c]
		Φύλο Dependent	,174	,053		,000[c]
	Uncertainty Coefficient	Symmetric	,107	,034	3,143	,000[d]
		Επάγγελμα Dependent	,086	,027	3,143	,000[d]
		Φύλο Dependent	,142	,045	3,143	,000[d]

c. Based on chi-square approximation

d. Likelihood ratio chi-square probability.

Οι δείκτες του πίνακα *(Directional measures)* είναι κατά σειρά:

✓ ***Lambda*:** Μέτρο συσχέτισης που εκφράζει την αναλογική μείωση του λάθους όταν οι τιμές της ανεξάρτητης μεταβλητής χρησιμοποιούνται για να προβλέψουν τις τιμές της εξαρτημένης μεταβλητής.

✓ ***Goodman and Kruskal's tau και Uncertainty coefficient***: Μέτρα συσχέτισης που εκφράζουν την αναλογική μείωση του λάθους όταν οι τιμές μίας μεταβλητής χρησιμοποιούνται για να προβλέψουν τις τιμές άλλης μεταβλητής.

!!! *Οι τιμές και των τριών δεικτών είναι μεταξύ 0 και 1.*

!!! *Η τιμή 0 μας δείχνει ότι η γνώση της ανεξάρτητης μεταβλητής δεν βοηθάει στην πρόβλεψη της εξαρτημένης. Η τιμή 1 δείχνει ότι η γνώση της ανεξάρτητης μεταβλητής καθορίζει τέλεια την εξαρτημένη μεταβλητή.*

!!! *Και οι τρεις αυτοί δείκτες εξετάζουν τις δύο μεταβλητές θεωρώντας την μία εξ αυτών εξαρτημένη, την άλλη ανεξάρτητη και στη συνέχεια το αντίθετο.*

✓ Η στήλη ***Asymptotic Std.Error*** βοηθάει στη δημιουργία διαστημάτων εμπιστοσύνης για τις εκτιμώμενες παραμέτρους.

✓ Η στήλη ***Approx.T*** είναι το πηλίκο των στηλών ***Value*** και ***Asymptotic Std.Error*** και είναι η τιμή την οποία ελέγχουμε για να αποδεχθούμε ή να απορρίψουμε τη μηδενική υπόθεση που εδώ μας λέει ότι η προς μελέτη παράμετρος είναι 0.

✓ Τέλος η στήλη ***Approx. Sig***. είναι η ισχύς του τεστ.

4.1.2 Ποιοτικές μεταβλητές ιεραρχικής κλίμακας (*Ordinal)*

Ανεξαρτήτως των διαστάσεων του πίνακα, αφού περάσουμε τις ιεραρχικής κλίμακας μεταβλητές στις θέσεις ***Row*** και ***Column*** της εικόνας 2.29, πατάμε στο κουμπί ***Statistics*** και από την εικόνα 4.1 επιλέγουμε:

- ***Chi-Square*** και ***Correlation***

Επιπλέον, από την περιοχή ***Ordinal*** επιλέγουμε:

- ***Gamma***
- ***Somers'd***
- ***Kendall's tau-b***
- ***Kendall's tau-c***

Στη συνέχεια ***Continue***, επιστροφή στην εικόνα 2.29, ***O.K*** και έχουμε τους επόμενους πίνακες:

Ο πίνακας ***Chi-Square*** (Πίνακας 4.4) είναι ο ίδιος ακριβώς με τον πίνακα 4.1, τους δείκτες του οποίου εξηγήσαμε ήδη στην παράγραφο 4.1.1.

Πίνακας 4.4: Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26,517[a]	6	,000
Likelihood Ratio	26,353	6	,000
Linear-by-Linear Association	18,378	1	,000
N of Valid Cases	207		

a. 1 cells (8,3%) have expected count less than 5. The minimum expected count is 4,57.

Ο πίνακας ***Symmetric Measures*** (Πίνακας 4.5) περιέχει κατά σειρά τους δείκτες

- ✓ ***Kendall's tau-b***: Μη παραμετρικό μέτρο συσχέτισης για ιεραρχικής κλίμακας ή διατεταγμένες (ranked) μεταβλητές που λαμβάνει υπόψη του τις βαθμίδες.

✓ ***Kendall's tau-c:*** Μη παραμετρικό μέτρο συσχέτισης για ιεραρχικής κλίμακας ή διατεταγμένες (ranked) μεταβλητές που αγνοεί τις βαθμίδες.

!!! *Το πρόσημο του συντελεστή των παραπάνω μέτρων, φανερώνει την διεύθυνση και η απόλυτη τιμή τους την ένταση της σχέσης, με μεγάλες τιμές να δείχνουν ισχυρή σχέση. Παίρνουν τιμές στο διάστημα [-1, +1] αλλά οι τιμές -1 και +1 μπορούν να επιτευχθούν μόνο από τετραγωνικούς πίνακες.*

✓ ***Gamma:*** Συμμετρικό μέτρο συσχέτισης μεταξύ δύο ιεραρχικής κλίμακας μεταβλητών το οποίο παίρνει τιμές στο διάστημα [-1, +1].

✓ ***Spearman Correlation***: Μέτρο συσχέτισης μεταξύ μεταβλητών διατεταγμένων βαθμίδων (rank orders). Οι τιμές του κυμαίνονται από –1 έως 1. Στις θέσεις –1 και 1 έχουμε τέλεια αρνητική ή θετική εξάρτηση και στη θέση 0 καμία εξάρτηση.

✓ ***Pearson's R***: Μέτρο γραμμικής συσχέτισης που χρησιμοποιείται μόνο όταν οι μεταβλητές είναι ποσοτικές. Αναλυτική παρουσίαση του συντελεστή του ***Pearson*** στην παράγραφο 6.1.

!!! *Οι τιμές των παραπάνω μέτρων κυμαίνονται από –1 έως 1. Στις θέσεις –1 και 1 έχουμε τέλεια αρνητική ή θετική εξάρτηση και στη θέση 0 καμία εξάρτηση.*

Πίνακας 4.5: Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	,279	,060	4,573	,000

	Kendall's tau-c	,264	,058	4,573	,000
	Gamma	,433	,088	4,573	,000
	Spearman Correlation	,310	,066	4,662	,000[c]
Interval by Interval	Pearson's R	,299	,066	4,481	,000[c]
N of Valid Cases		207			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Στον πίνακα ***Directional Measures*** (Πίνακας 4.6) έχουμε:

✓ Τον δείκτη ***Somers'd:*** Μέτρο συσχέτισης μεταξύ δύο ιεραρχικής κλίμακας μεταβλητών. Οι τιμές του κυμαίνονται από -1 έως +1.

!!! *Τιμές κοντά στο -1 ή το +1 δείχνουν έντονη εξάρτηση μεταξύ των δύο μεταβλητών ενώ τιμές κοντά στο 0 δείχνουν μικρή ή καθόλου εξάρτηση.*

Πίνακας 4.6: Directional Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Somers' d Symmetric	,279	,060	4,573	,000
	Επίπεδο εκπαίδευσης Dependent	,279	,061	4,573	,000
	Μέγεθος Επιχείρησης Dependent	,279	,059	4,573	,000

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

4.1.3 Ειδικές περιπτώσεις

1. *Eta*: Ο δείκτης αυτός είναι κατάλληλος στην περίπτωση που η εξαρτημένη μεταβλητή είναι διαστημικής κλίμακας και η ανεξάρτητη

μεταβλητή ονομαστικής κλίμακας με περιορισμένο πλήθος επιπέδων. Παίρνει τιμές στο διάστημα [0,1] και τιμές κοντά στο 1 φανερώνουν υψηλό επίπεδο συσχέτισης, ενώ τιμές κοντά στο 0 φανερώνουν έλλειψη συσχέτισης. Απαραίτητη προϋπόθεση η ποιοτική μεταβλητή να κωδικοποιηθεί αριθμητικά.

2. *Kappa:* Μετρά την συμφωνία μεταξύ των εκτιμήσεων δύο αξιολογητών όταν και οι δύο αξιολογούν το ίδιο αντικείμενο. Η τιμή 1 εκφράζει τέλεια συμφωνία ενώ η τιμή 0 δείχνει ότι η συμφωνία δεν είναι καλή. Ο δείκτης Kappa είναι διαθέσιμος μόνο για πίνακες στους οποίους και οι δύο μεταβλητές χρησιμοποιούν την ίδια κατηγορία τιμών και επίσης και οι δύο μεταβλητές έχουν το ίδιο πλήθος κατηγοριών.

3. *Risk:* Δείκτης κατάλληλος για 2 x 2 πίνακες. Μετρά την ένταση της σχέσης μεταξύ της παρουσίας ενός παράγοντα και της πραγματοποίησης ενός γεγονότος. Αν το διάστημα εμπιστοσύνης για τον δείκτη, περιέχει την τιμή 1 δεν μπορείς να υποθέσεις ότι ο παράγοντας σχετίζεται με το γεγονός. Ως μια εκτίμηση, μπορεί να χρησιμοποιηθεί ο λόγος των σχετικών πιθανοτήτων ή ο σχετικός κίνδυνος αν η εμφάνιση του παράγοντα είναι σπάνια.

4. *McNemar*: Ένας μη παραμετρικός έλεγχος για δύο συσχετιζόμενες διχοτομικές μεταβλητές. Ελέγχει για μεταβολές στις απαντήσεις χρησιμοποιώντας την χ^2 κατανομή. Χρήσιμος για πειραματικούς σχεδιασμούς τύπου «πριν και μετά». Για τετραγωνικούς πίνακες μεγαλύτερων διαστάσεων, χρησιμοποιείται ο έλεγχος συμμετρίας ***McNemar*** – ***Bowker***.

5. ***Cochran's and Mantel-Haenszel*:** Οι δείκτες Cochran's and Mantel-Haenszel μπορούν να χρησιμοποιηθούν για τον έλεγχο της ανεξαρτησίας μεταξύ μιας διχοτομικής παραγοντικής μεταβλητής και μιας διχοτομικής μεταβλητής απόκρισης, λαμβάνοντας υπόψη τα πρότυπα συμμεταβλητότητας (covariate patterns) που καθορίστηκαν από μία ή περισσότερες μεταβλητές ελέγχου. Να σημειωθεί ότι ενώ άλλα στατιστικά μέτρα υπολογίζονται στρώμα-στρώμα, το στατιστικό των Cochran και Mantel-Haenszel υπολογίζεται μία φορά για όλα τα στρώματα.

!!! *Για τον υπολογισμό των παραπάνω δεικτών αρκεί να τους επιλέξουμε, στην εικόνα 4.1*.

Κεφάλαιο 5

Σύγκριση Μέσων

Chapter 5

Compare Means

5. Εισαγωγή

Στην εφαρμοσμένη έρευνα, πολλές φορές, προκύπτει η ανάγκη σύγκρισης των μέσων τιμών (***Compare means***) δύο ή περισσοτέρων πληθυσμών. Οι περιπτώσεις είναι πολλές και απαιτείται να γίνεται

διάκριση μεταξύ αυτών. Το S.P.S.S έχει τη δυνατότητα να αντιμετωπίσει τις εξής περιπτώσεις:

- **Means**
- **One –Sample T Test**
- **Independent -Sample T Test**
- **Paired- Samples T Test** και
- **One -Way ANOVA**

Για να εργαστούμε με μία από τις παραπάνω περιπτώσεις, πρέπει από το μενού **Analyze** να τσεκάρουμε **Compare Mean**s και στη συνέχεια να επιλέξουμε αυτήν η οποία είναι κατάλληλη για την ανάλυση.

Στις επόμενες παραγράφους θα αναπτυχθεί αναλυτικά κάθε μία από τις παραπάνω περιπτώσεις.

5.1 Περίπτωση Means

Στην περίπτωση αυτή έχουμε μία ή περισσότερες ποσοτικές μεταβλητές (ή ιεραρχικής κλίμακας ποιοτικές) και μια ή περισσότερες ποιοτικές μεταβλητές. Οι ποιοτικές μεταβλητές είναι αυτές οι οποίες χωρίζουν το αρχικό πλήθος των παρατηρήσεων (δείγμα ή πληθυσμός) σε υποομάδες. Η σύγκριση λοιπόν, γίνεται μεταξύ των μέσων τιμών των υποομάδων οι οποίες δημιουργήθηκαν εξαιτίας της ποιοτικής μεταβλητής.

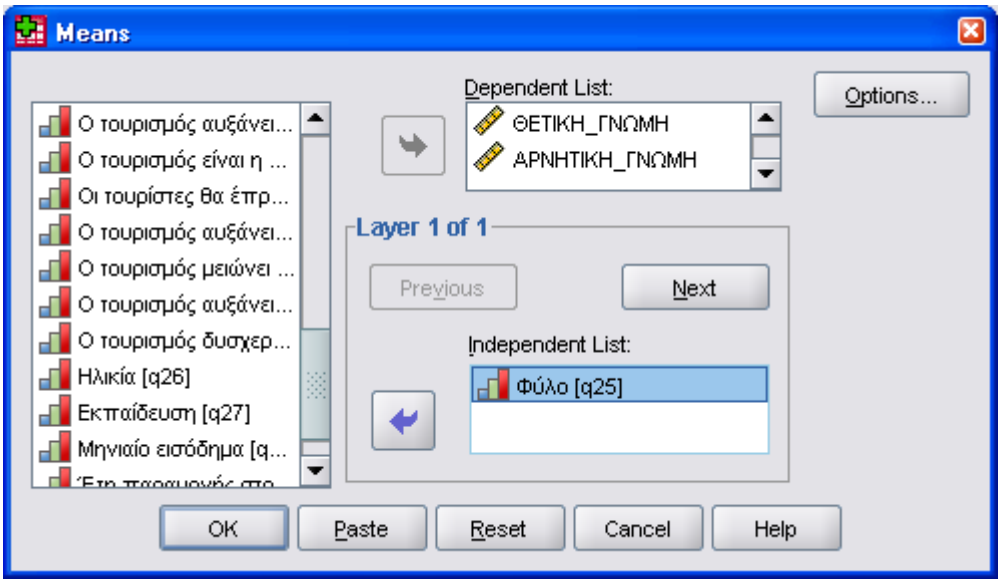
Για την ανάπτυξη της μεθοδολογίας της συγκεκριμένης περίπτωσης θα χρησιμοποιηθεί το επόμενο παράδειγμα.

Παράδειγμα 1: Σε έρευνα η οποία πραγματοποιήθηκε σε 1039 κατοίκους μικρών τουριστικών περιοχών, τους ζητήθηκε να αξιολογήσουν σε μία κλίμακα από 1 έως 5 τις ***θετικές*** και τις

αρνητικές επιπτώσεις του τουρισμού στην περιοχή τους. Θέλουμε να ελέγξουμε αν οι γνώμες των κατοίκων διαφέρουν σημαντικά ανάλογα με φύλο τους.

Για την επίτευξη του στόχου μας αρκεί να ακολουθήσουμε την παρακάτω διαδικασία.

- Από το μενού ***Analyze***, επιλέγουμε ***Compare Means***, στη συνέχεια ***Means*** και έχουμε την επόμενη εικόνα.

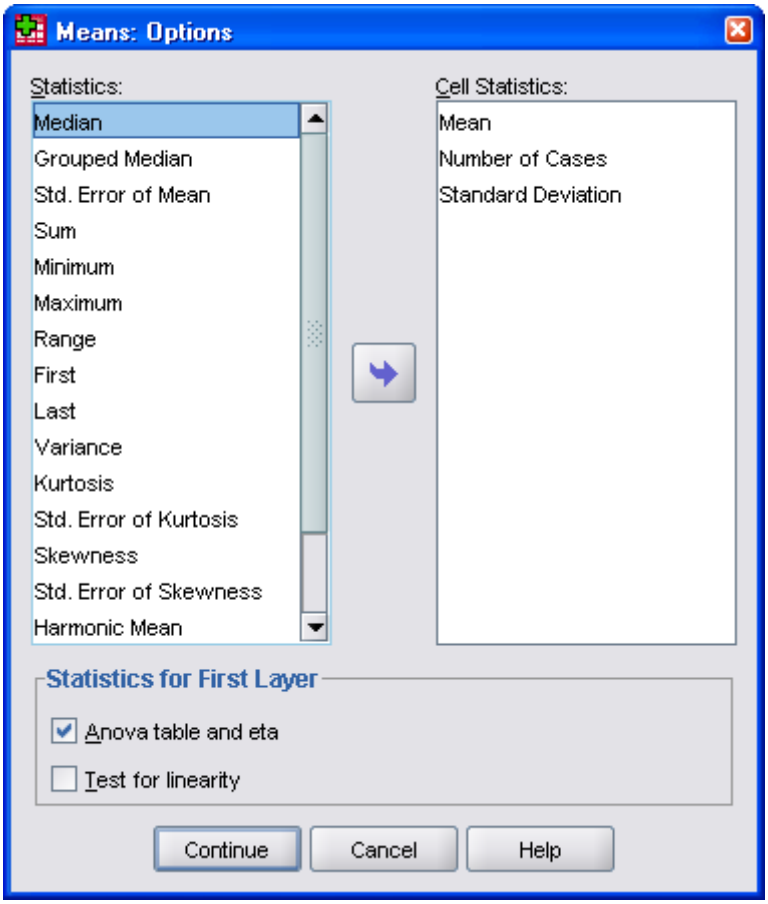


Εικόνα 5.1

- Στη θέση ***Dependent List*** βάζουμε την ή τις εξαρτημένες μεταβλητές (ποσοτικές ή ιεραρχικής κλίμακας ποιοτικές). Στο παράδειγμά μας ***«θετική γνώμη»*** και «***αρνητική γνώμη***».

- Στη θέση ***Independent List*** βάζουμε την πρώτη ποιοτική (ανεξάρτητη) μεταβλητή ***«φύλο»*** και πατάμε **Next** αν θέλουμε να βάλουμε και δεύτερη.

- Στη συνέχεια επιλέγουμε ***Options*** και έχουμε την επόμενη φόρμα



Εικόνα 5.2

- Από το παράθυρο ***Statistics*** επιλέγουμε όσες από τις παραμέτρους θέλουμε να υπολογίσουμε και τις μεταφέρουμε στο παράθυρο ***Cell Statistics.***
- Στη συνέχεια τσεκάρουμε την ένδειξη ***Anova table and eta***.
- Η ένδειξη ***Test for Linearity*** επιλέγεται μόνο όταν η ανεξάρτητη μεταβλητή είναι ιεραρχικής κλίμακας και έχει τουλάχιστον τρία επίπεδα.

Στο συγκεκριμένο παράδειγμα δεν έχει νόημα η επιλογή της ένδειξης ***Test for Linearity***.

- Πατάμε ***continue***, επιστρέφουμε στην εικόνα 5.1 και με ***O.K*** έχουμε τους εξής πίνακες:

Πίνακας ***Report*** ο οποίος δίνει:

- Τους **μέσους** (**means**), την **τυπική απόκλιση** (**Std. Deviation**) και το **πλήθος** των ατόμων (**N**), ανά υποομάδα (φύλο).

Πίνακας 5.1: Report

Φύλο		ΘΕΤΙΚΗ_ΓΝΩΜΗ	ΑΡΝΗΤΙΚΗ_ΓΝΩΜΗ
Άνδρας	Mean	4,0370	3,4829
	N	616	616
	Std. Deviation	,57320	,60651
Γυναίκα	Mean	3,9110	3,3567
	N	416	416
	Std. Deviation	,49910	,63465
Total	Mean	3,9862	3,4320
	N	1032	1032
	Std. Deviation	,54779	,62080

Παρατηρούμε, ότι η άποψη των ανδρών (4,04) σχετικά με τις θετικές επιπτώσεις είναι ελαφρώς καλύτερη από την αντίστοιχη των γυναικών (3,91). Στις αρνητικές επιπτώσεις υπερτερούν εκ νέου οι άνδρες (3,48) των γυναικών (3,36).

Ο πίνακας ***ANOVA***

- Ο πίνακας ***ANOVA*** μας πληροφορεί για το αν υπάρχει σημαντική διαφορά μεταξύ των μέσων τιμών των υποομάδων οι οποίες δημιουργήθηκαν. **Μηδενική υπόθεση του ελέγχου είναι η ισότητα των μέσων τιμών των υποομάδων και διατυπώνεται ως εξής:**

H_0: $\mu_A=\mu_\Gamma$ έναντι της εναλλακτικής υπόθεσης H_1: $\mu_A \neq \mu_\Gamma$

Η μηδενική υπόθεση απορρίπτεται όταν το ***Sig.*** του στατιστικού ***F*** **(*Combined)*** είναι πολύ μικρό **(**συνήθως μικρότερο του 0,05**)**.

Στο παράδειγμά μας έχουμε **Sig. F= 0,000<0,05** για τις θετικές και **Sig. F= 0,001<0,05** για τις αρνητικές επιπτώσεις. Συνεπώς απορρίπτεται η

υπόθεση της μη σημαντικής διαφοράς απόψεων μεταξύ ανδρών και γυναικών.

Πίνακας 5.2: ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
ΘΕΤΙΚΗ_ΓΝΩΜΗ * Φύλο	Between Groups	(Combined)	3,938	1	3,938	13,279	**,000**
	Within Groups		305,438	1030	,297		
	Total		309,375	1031			
ΑΡΝΗΤΙΚΗ_ΓΝΩΜΗ * Φύλο	Between Groups	(Combined)	3,953	1	3,953	10,349	**,001**
	Within Groups		393,382	1030	,382		
	Total		397,334	1031			

Ο πίνακας ***Measures of Association*** μάς δίνει:

- Τους δείκτες ***Eta*** και ***Eta Squared (Eta Squared = Eta^2)*** οι οποίοι λαμβάνονται υπόψη μόνο όταν η ανεξάρτητη μεταβλητή είναι ονομαστικής κλίμακας (Nominal). Η τιμή του δείκτη ***Eta Squared*** προκύπτει αν διαιρέσουμε ***Sum of Squares- between groups / Total-Sum of Squares***, από τον πίνακα ANOVA και εξηγεί **το ποσοστό των μεταβολών της εξαρτημένης μεταβλητής το οποίο οφείλεται στις μεταβολές της ανεξάρτητης μεταβλητής**. Παίρνει τιμές στο διάστημα [0, 1], όπου το 0 δείχνει ότι η ανεξάρτητη μεταβλητή δεν ερμηνεύει κανένα μέρος των μεταβολών της εξαρτημένης, ενώ το 1 δείχνει ότι η ανεξάρτητη μεταβλητή ερμηνεύει το 100% των μεταβολών της εξαρτημένης.

Στο παράδειγμά μας μπορούμε να πούμε ότι **το φύλο** εξηγεί μόνο το 1,3**%** των μεταβολών των θετικών

απόψεων. Επίσης, το **φύλο**, εξηγεί μόνο το 1% των μεταβολών των αρνητικών απόψεων.

Πίνακας 5.3: Measures of Association

	Eta	Eta Squared
ΘΕΤΙΚΗ_ΓΝΩΜΗ * Φύλο	,113	,013
ΑΡΝΗΤΙΚΗ_ΓΝΩΜΗ * Φύλο	,100	,010

Παράδειγμα 2: Στην προηγούμενη έρευνα θέλουμε να ελέγξουμε αν οι γνώμες των κατοίκων διαφέρουν σημαντικά, ανάλογα με το επίπεδο μόρφωσής τους. Σε αυτή την περίπτωση έχουμε πάλι ως εξαρτημένες μεταβλητές την αξιολόγηση των θετικών και των αρνητικών επιπτώσεων του τουρισμού από τους κατοίκους και ως ανεξάρτητη το επίπεδο μόρφωσης. Η μεταβλητή επίπεδο μόρφωσης είναι ιεραρχικής κλίμακας και έχει τρία επίπεδα (Στοιχειώδης, Μέση, Ανώτερη/Ανώτατη).

Ακολουθώντας την προηγούμενη διαδικασία θα πρέπει να βάλουμε στη θέση της εξαρτημένης μεταβλητής την μεταβλητή ***«Εκπαίδευση»*** και στην εικόνα 5.2 να επιλέξουμε επιπλέον την ένδειξη ***Test for Linearity.*** Στη συνέχεια πατάμε Ο.Κ και έχουμε τους πίνακες που ακολουθούν.

Πίνακας 5.1.1: Report

Εκπαίδευση		ΘΕΤΙΚΗ_ΓΝΩΜΗ	ΑΡΝΗΤΙΚΗ_ΓΝΩΜΗ
Στοιχειώδης	Mean	3,7975	3,4746
	N	196	196

	Std. Deviation	,61195	,59607
Μέση	Mean	4,0260	3,5010
	N	496	496
	Std. Deviation	,53204	,60029
Ανώτερη/Ανώτατη	Mean	4,0367	3,3203
	N	346	346
	Std. Deviation	,50895	,64978
Total	Mean	3,9864	3,4358
	N	1038	1038
	Std. Deviation	,54777	,62134

Πίνακας 5.2.1: ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
ΘΕΤΙΚΗ_ΓΝΩΜΗ * Εκπαίδευση	Between Groups	(Combined)	8,648	2	4,324	14,795	,000
		Linearity	5,693	1	5,693	19,476	,000
		Deviation from Linearity	2,956	1	2,956	10,113	,002
	Within Groups		302,508	1035	,292		
	Total		311,156	1037			
ΑΡΝΗΤΙΚΗ_ΓΝΩΜΗ * Εκπαίδευση	Between Groups	(Combined)	7,022	2	3,511	9,239	,000
		Linearity	4,350	1	4,350	11,447	,001
		Deviation from Linearity	2,672	1	2,672	7,031	,008
	Within Groups		393,320	1035	,380		
	Total		400,342	1037			

Στον πίνακα 5.2.1 παρατηρούμε, σε σχέση με τον πίνακα 5.2, ότι υπάρχει ο δείκτης ***Linearity*** ο οποίος εξετάζει αν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών. **Η αρχική (μηδενική) υπόθεση ότι δεν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών, απορρίπτεται όταν τα αντίστοιχα Sig.F είναι πολύ μικρά (<0,05).**

Υπάρχει επίσης ο δείκτης ***Deviation from Linearity*** ο οποίος εξετάζει αν υπάρχει απόκλιση από τη γραμμικότητα και **η αρχική υπόθεση ότι δεν υπάρχει απόκλιση απορρίπτεται όταν τα αντίστοιχα Sig.F είναι πολύ μικρά (<0,05).**

!!! *Οι δείκτες* ***Linearity*** *και* ***Deviation from Linearity*** *έχουν νόημα όταν η ανεξάρτητη μεταβλητή (Independent) είναι* ***ιεραρχικής κλίμακας (ordinal)*** *και έχει* ***τουλάχιστον τρία επίπεδα.***

Πίνακας 5.3.1: Measures of Association

	R	R Squared	Eta	Eta Squared
ΘΕΤΙΚΗ_ΓΝΩΜΗ * Εκπαίδευση	,135	,018	,167	,028
ΑΡΝΗΤΙΚΗ_ΓΝΩΜΗ * Εκπαίδευση	-,104	,011	,132	,018

- Στον πίνακα 5.3.1 έχουμε τούς γνωστούς, από την παλινδρόμηση, δείκτες ***R*** και ***R Squared***, οι οποίοι λαμβάνονται υπόψη μόνο στην περίπτωση που η ανεξάρτητη μεταβλητή είναι ***Ordinal***. Η τιμή του δείκτη ***R Squared*** κυμαίνεται από 0 έως 1 και προκύπτει από τον δείκτη ***R*** στο τετράγωνο.

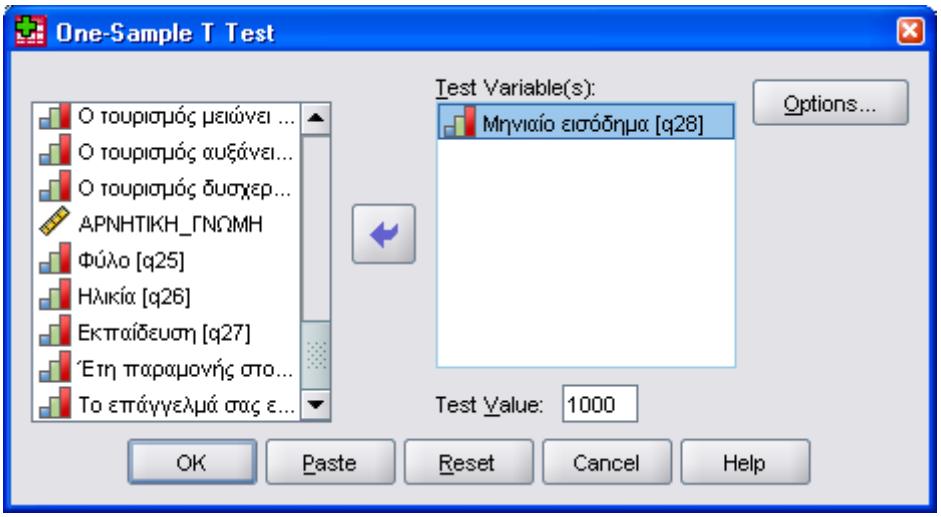
5.2 Σύγκριση μέσης τιμής δείγματος και πληθυσμού

Περίπτωση One sample T-test

Στην περίπτωση αυτή η σύγκριση γίνεται μεταξύ της μέσης τιμής ενός δείγματος και της υποτιθέμενης τιμής του μέσου του πληθυσμού από τον οποίο πήραμε το δείγμα. Η μηδενική υπόθεση του ελέγχου αυτού είναι ότι η μέση τιμή του δείγματος και η μέση τιμή του πληθυσμού είναι ίσες. Για την ολοκλήρωση της διαδικασίας θα χρησιμοποιηθεί το επόμενο παράδειγμα.

Παράδειγμα: Έστω ότι έχουμε καταγράψει το μηνιαίο εισόδημα 426 κατοίκων τριών διαφορετικών περιοχών της Ελλάδας. Θέλουμε να συγκρίνουμε το μέσο μηνιαίο εισόδημα των τριών αυτών περιοχών (δείγμα) με το μέσο μηνιαίο εισόδημα των κατοίκων όλης της Ελλάδας (πληθυσμός), το οποίο ανέρχεται στο ποσό των 1.000 €.

- Από το μενού ***Analyze***, επιλέγουμε ***Compare Means***, στη συνέχεια ***One sample T-test*** και έχουμε την επόμενη εικόνα.

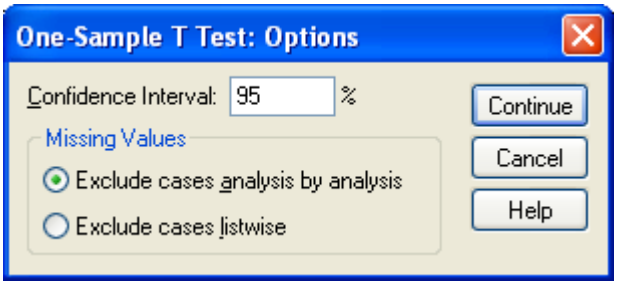


Εικόνα 5.3

- Από το παράθυρο αριστερά επιλέγουμε την ποσοτική μεταβλητή (**Μηνιαίο Εισόδημα**) και τη μεταφέρουμε στο παράθυρο ***Test Variable***.

- Στο παράθυρο ***Test Value*** βάζουμε τη μέση τιμή του πληθυσμού με την οποία θα συγκρίνουμε τη μέση τιμή του δείγματος. Έστω ότι αυτή είναι 1.000 €.

- Πατάμε στην επιλογή ***Options*** και εμφανίζεται η επόμενη εικόνα.



Εικόνα 5.4

- Γράφουμε το ***επίπεδο εμπιστοσύνης*** του διαστήματος εμπιστοσύνης ***(Confidence Interval)*** το οποίο θέλουμε (προκαθορισμένο 95%) και στη συνέχεια ***Continue*** επιστροφή στην εικόνα 5.3 και με O.K έχουμε τους παρακάτω πίνακες.

Ο πίνακας ***One-Sample Statistics*** μας δίνει:

- Το πλήθος των έγκυρων απαντήσεων ***(N)***
- Τον αριθμητικό μέσο ***(mean)***
- Την τυπική απόκλιση αυτών ***(std. Deviation)*** και
- Το τυπικό σφάλμα του μέσου ***(std. Error mean)***.

Πίνακας 5.4: One Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Εισόδημα	417	989,0168	389,17225	19,05785

Ο πίνακας ***One-Sample Test*** μας δίνει:

- Την τιμή του t-test (t =- 0,576)
- Το sig. του t-test (Sig. = 0,565) και

- Τα άκρα του διαστήματος εμπιστοσύνης για τη διαφορά των μέσων τιμών δείγματος και πληθυσμού (Lower- Upper).

Το σημαντικότερο σημείο αυτού του πίνακα είναι η τιμή ***sig. του t-test***, γιατί με βάση αυτήν θα απορριφθεί ή θα γίνει αποδεκτή η **μηδενική υπόθεση της ισότητας των δύο μέσων**. Αν είναι μικρότερη του 0,05 (επίπεδο σημαντικότητας 5%) τότε απορρίπτουμε τη μηδενική υπόθεση, ενώ σε αντίθετη περίπτωση την αποδεχόμαστε.

Στην περίπτωσή μας είναι: 0,565 > 0,05 και συνεπώς δεχόμαστε ότι το μέσο μηνιαίο εισόδημα των κατοίκων του δείγματος δεν διαφέρει σημαντικά από το μέσο μηνιαίο εισόδημα των κατοίκων όλης της χώρας.

Πίνακας 5.5: One Sample Test

	Test Value = 1000					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Εισόδημα	-,576	416	,565	-10,98321	-48,4449	26,4785

5.3 Σύγκριση των μέσων τιμών δύο ανεξάρτητων δειγμάτων

Περίπτωση Independent -Samples T test

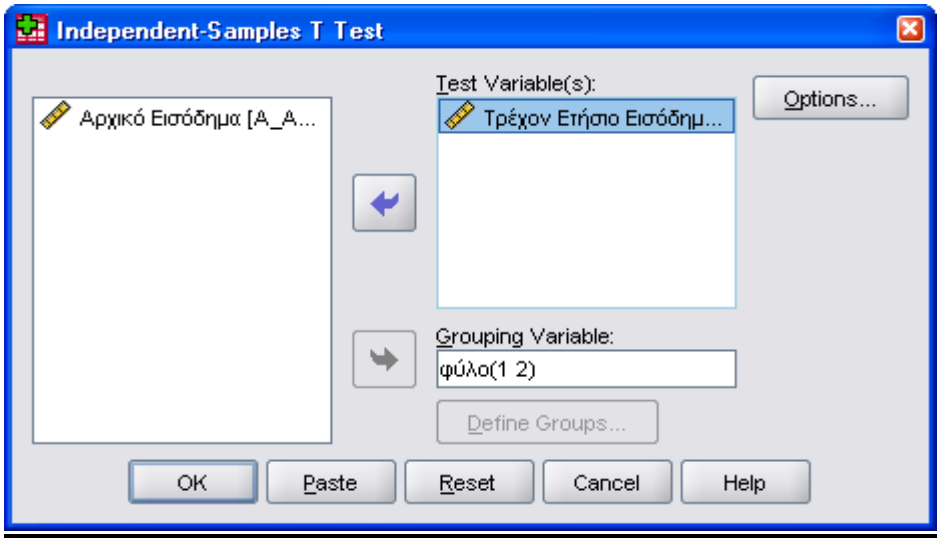
Δύο δείγματα (ή και περισσότερα) χαρακτηρίζονται ως ανεξάρτητα όταν τα δεδομένα και στα δύο τα έχουμε συλλέξει με τη μέθοδο της τυχαίας δειγματοληψίας.

Στην περίπτωση αυτή, μηδενική υπόθεση του ελέγχου είναι η ισότητα των μέσων τιμών των δύο πληθυσμών από τους οποίους πήραμε τα δείγματα.

Θα χρησιμοποιηθούν τα δεδομένα του επόμενου παραδείγματος για να ολοκληρωθεί η διαδικασία.

Παράδειγμα: Καταγράφηκαν τα ετήσια εισοδήματα 258 ανδρών και 216 γυναικών. Για τον έλεγχο της υπόθεσης της ισότητας των μέσων εισοδημάτων ανδρών και γυναικών πρέπει:

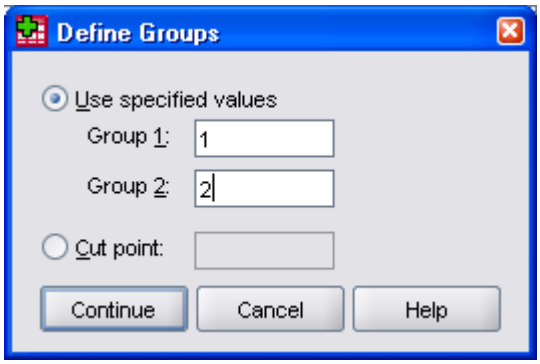
- Από το μενού ***Analyze***, επιλέγουμε ***Compare Means***, στη συνέχεια ***Independent -Samples T test*** και έχουμε την επόμενη εικόνα.



Εικόνα 5.5

- Από το παράθυρο αριστερά επιλέγουμε την ποσοτική μεταβλητή (**Τρέχον Ετήσιο Εισόδημα**) και τη μεταφέρουμε στο παράθυρο ***Test Variable***.

- Στο παράθυρο ***Grouping Variable*** βάζουμε τη ποιοτική μεταβλητή ***Φύλο*** και στη συνέχεια πατάμε στην επιλογή ***Define Groups*** και εμφανίζεται η επόμενη φόρμα:



Εικόνα 5.6

- Στην επιλογή ***Use specified values*** και στις θέσεις ***Group1*** και ***Group 2*** βάζουμε τους κωδικούς των δύο φύλων των οποίων θέλουμε να συγκρίνουμε τις μέσες τιμές. Οι κωδικοί είναι: 1 για τους άνδρες και 2 για τις γυναίκες. ***Αν η μεταβλητή έχει περισσότερα επίπεδα θα πρέπει να επιλέξουμε μόνο τα δύο που θέλουμε να συγκρίνουμε.***

- Η επιλογή ***Cut Point*** είναι χρήσιμη στην περίπτωση που η ποιοτική μεταβλητή (***Grouping Variable***) έχει περισσότερους από δύο κωδικούς. Τότε έχουμε τη δυνατότητα να κάνουμε διαχωρισμό του συνόλου των κωδικών σε δύο μέρη. Αν για παράδειγμα η μεταβλητή αποτελείται από 5 κωδικούς και στο παράθυρο ***Cut Point*** γράψουμε τον αριθμό 3, τότε θα έχουμε για σύγκριση ένα γκρουπ με τους κωδικούς 1 και 2 και ένα γκρουπ με τους κωδικούς 3, 4 και 5.

- Πατάμε ***Continue***, επιστρέφουμε στην εικόνα 5.5, επιλέγουμε ***Options***, εμφανίζεται η εικόνα 5.4, όπου δηλώνουμε το επιθυμητό **διάστημα εμπιστοσύνης** και στη συνέχεια με ***O.K*** έχουμε τους πίνακες που ακολουθούν.

Στον πίνακα **Group Statistics** έχουμε:

- Το πλήθος ***(N)*** των ανδρών και των γυναικών
- Τον μέσο αυτών ***(mean)***
- Την τυπική απόκλιση αυτών ***(std. Deviation)*** και
- Το τυπικό σφάλμα του μέσου ***(std. Error mean)***.

Πίνακας 5.6: Group Statistics

	Φύλο	N	Mean	Std. Deviation	Std. Error Mean
Ετήσιο Εισόδημα	Άνδρας	258	41441,78	19499,214	1213,968
	υναίκα	216	26031,92	7558,021	514,258

Στον πίνακα **Independent Samples Test** έχουμε κατά σειρά:

- Το έλεγχο της **ισότητας των διακυμάνσεων**- ***Levene's test for Equality of variances***. **Μηδενική υπόθεση του ελέγχου η ισότητα των Διακυμάνσεων (Variances) των δύο πληθυσμών (Ομοιογενείς πληθυσμοί) από τους οποίους πήραμε τα δείγματα-*equal Variances assumed*. Αν η τιμή Sig. είναι μικρότερη από το 0,05, απορρίπτουμε αυτή την υπόθεση (σε επίπεδο σημαντικότητας 5%) και δεχόμαστε ότι οι διακυμάνσεις δεν είναι ίσες-*equal Variances not assumed*.** Στον πίνακα 5.7 παρατηρούμε ότι **Sig. = 0,000 < 0,05** και επομένως δεχόμαστε ότι οι διακυμάνσεις των κλάδων δεν είναι ίσες.

- Στη συνέχεια του ίδιου πίνακα έχουμε τον έλεγχο **ισότητας των μέσων**- ***t test for equality of mean***. **Μηδενική υπόθεση του ελέγχου η ισότητα των μέσων τιμών των πληθυσμών. Αν η τιμή Sig.(2 tailed) είναι μικρότερη από το 0,05 απορρίπτουμε αυτή την υπόθεση (σε επίπεδο σημαντικότητας 5%) και δεχόμαστε ότι οι μέσοι δεν είναι ίσοι**. Παρατηρούμε, βέβαια, ότι υπάρχουν δύο τέτοιες τιμές. Εμείς θα επιλέξουμε τη μία από τις δύο με βάση την **αποδοχή** ή την **απόρριψη** της μηδενικής υπόθεσης του προηγούμενου ελέγχου. Αν λοιπόν, με βάση τα αποτελέσματα, έχουμε απορρίψει την υπόθεση της ισότητας των διακυμάνσεων (***equal Variances assumed)***, τότε θα επιλέξουμε την δεύτερη από τις δύο τιμές του **t-test**. Στην περίπτωση όμως που είχαμε δεχθεί την υπόθεση της ισότητας των διακυμάνσεων ***(equal Variances not assumed)*** θα επιλέγαμε τη πρώτη τιμή του **t-test**. Στο παράδειγμά μας έχουμε απορρίψει την ισότητα των διακυμάνσεων και επομένως μας ενδιαφέρει η δεύτερη τιμή, η οποία είναι Sig.(2 tailed)=0,000<0,05. Βασιζόμενοι σε αυτό το αποτέλεσμα

απορρίπτουμε την ισότητα των δύο μέσων. Στον ίδιο πίνακα δίνονται στη συνέχεια:

- Η ***διαφορά των μέσων (mean difference)***
- ***Το τυπικό σφάλμα της διαφοράς (std. Error Difference)*** και
- ***Τα όρια (Lower-Upper) του διαστήματος εμπιστοσύνης της διαφοράς των μέσων (95% Confidence interval of the difference)***.

Πίνακας 5.7: Independent Sample Test

			Ετήσιο Εισόδημα	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances		F	119,669	
		Sig.	,000	
t-test for Equality of Means		t	10,945	11,688
		df	472	344,262
		Sig. (2-tailed)	,000	,000
		Mean Difference	15409,862	15409,862
		Std. Error Difference	1407,906	1318,400
	95% Confidence Interval of the Difference	Lower	12643,322	12816,728
		Upper	18176,401	18002,996

5.4 Σύγκριση των μέσων τιμών δύο εξαρτημένων δειγμάτων

Περίπτωση Paired-Samples t Test

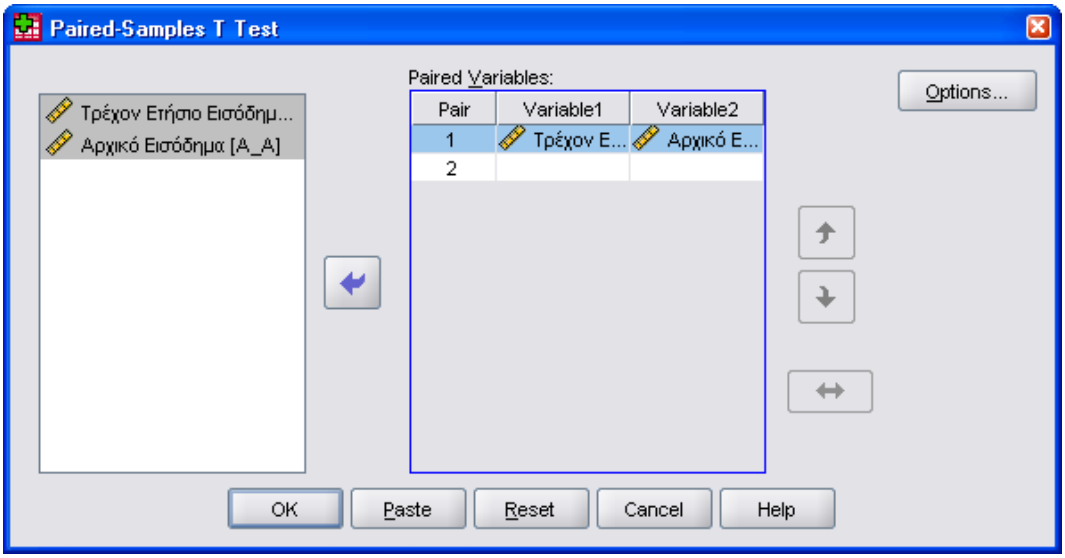
Ενώ στις προηγούμενες περιπτώσεις τα δείγματα ήταν **ανεξάρτητα**, στην παράγραφο αυτή θα εξετάσουμε την περίπτωση δύο **εξαρτημένων δειγμάτων (εξισωμένων κατά ζεύγη)**. Με τον όρο **εξαρτημένα δείγματα** εννοούμε δύο δείγματα όπου τα δεδομένα του πρώτου έχουν επιλεγεί με τη μέθοδο της τυχαίας δειγματοληψίας, ενώ του δεύτερου έχουν επιλεγεί με τέτοιο τρόπο ώστε να είναι ως προς

ορισμένα χαρακτηριστικά **ισότιμα** με τα δεδομένα του πρώτου. Μια συνηθισμένη περίπτωση εξαρτημένων δειγμάτων είναι οι μετρήσεις των πειραματικών μονάδων, για το ίδιο χαρακτηριστικό, **πριν (before)** και **μετά (after)** από μία ενέργεια. Αν μετρήσουμε τους παλμούς κάποιων ατόμων σε ηρεμία και στη συνέχεια μετρήσουμε τους παλμούς των ιδίων ατόμων μετά από μία πορεία 2 χιλιομέτρων έχουμε δημιουργήσεις εξαρτημένα δείγματα. Επίσης, εξαρτημένα δείγματα δημιουργούνται και όταν κάνουμε μέτρηση των ίδιων χαρακτηριστικών με δύο διαφορετικές συσκευές. Η μέτρηση της αρτηριακής πίεσης ατόμων με δύο διαφορετικά πιεσόμετρα ή το ζύγισμα με δύο διαφορετικές ζυγαριές δημιουργεί εξαρτημένα ζεύγη.

Παράδειγμα: Έστω ότι έχουμε καταγράψει τις αρχικές ετήσιες αποδοχές 1.500 εργαζομένων και τις αποδοχές των ιδίων μετά από 3 χρόνια εργασίας. Θέλουμε να ελέγξουμε αν υπάρχει σημαντική διαφορά μεταξύ των αρχικών αποδοχών και των αποδοχών μετά 3 χρόνια.

Για τον έλεγχο της προαναφερθείσας υπόθεσης θα ακολουθήσουμε την εξής διαδικασία:

- Από το μενού ***Analyze***, επιλέγουμε ***Compare Means***, στη συνέχεια ***Paired-Samples t Test*** και από το αριστερό παράθυρο επιλέγουμε τις δύο μεταβλητές οι οποίες αποτελούν τα εξαρτημένα δείγματα (**Τρέχον Ετήσιο Εισόδημα - Αρχικό Εισόδημα**) και τις μεταφέρουμε στο παράθυρο δεξιά. Προκύπτει η επόμενη εικόνα.



Εικόνα 5.7

- Πατάμε στο κουμπί ***Options*** και εμφανίζεται η φόρμα της εικόνας 5.4. Επιλέγουμε το επιθυμητό επίπεδο εμπιστοσύνης και
- Επιστρέφουμε με ***Continue*** στην εικόνα 5.7 και με ***O.K*** έχουμε τους επόμενους πίνακες:

Τον πίνακα 5.8 **Paired Samples Statistics** ο οποίος περιέχει τα γνωστά βασικά στατιστικά μέτρα και για τα δύο δείγματα

Πίνακας 5.8: Paired Samples Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Τρέχον Ετήσιο Εισόδημα	34419,57	1.500	17075,661	784,311
	Αρχικό Εισόδημα	17016,09	1.500	7870,638	361,510

Ο πίνακας **Paired Samples Correlations** μας δίνει:

- το **συντελεστή συσχέτισης** των δύο μεταβλητών-**Correlation**-

και

- το **Sig**. του αντιστοίχου τεστ.

Στο συγκεκριμένο παράδειγμα ο συντελεστής συσχέτισης είναι 0,880 και μπορούμε να πούμε ότι η σχέση των τιμών των δύο μεταβλητών είναι **πολύ έντονη**.

Πίνακας 5.9: Paired Samples Correlations

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Τρέχον Ετήσιο Εισόδημα & Αρχικό Εισόδημα	1.500	,880	,000

Ο πίνακας **Paired Samples Test** δίνει τις ***διαφορές (paired differences)***:

- ***Των μέσων (mean)***
- ***Των τυπικών αποκλίσεων (Std. Deviation)***
- ***Των τυπικών σφαλμάτων των μέσων (std. Error mean)*** και
- ***Τα όρια (Lower-Upper) του διαστήματος εμπιστοσύνης της διαφοράς των μέσων (95% Confidence interval of the difference)***.

Στη συνέχεια δίνει:

- την τιμή του **t -Test** του οποίου **η αρχική υπόθεση, όπως είναι γνωστό, είναι η ισότητα των μέσων τιμών**
- την τιμή **Sig.(2 tailed)**, η οποία συγκρινόμενη με το ορισθέν επίπεδο σημαντικότητας (συνήθως το 5%) μας οδηγεί στην αποδοχή ή την απόρριψη της αρχικής υπόθεσης σύμφωνα με τα γνωστά κριτήρια.

Στο συγκεκριμένο παράδειγμα η τιμή Sig.(2 tailed) = 0,000< 0,001 και συνεπώς απορρίπτουμε την αρχική υπόθεση σε επίπεδο σημαντικότητας 1‰ και συμπερασματικά λοιπόν, μπορούμε να πούμε ότι οι μέσες αποδοχές διαφέρουν σημαντικά.

Πίνακας 5.10: Paired Samples Test

Paired Samples Test

	Pair 1

			Τρέχον Ετήσιο Εισόδημα - Αρχικό Εισόδημα
Paired Differences	Mean		17403,481
	Std. Deviation		10814,620
	Std. Error Mean		496,732
	95% Confidence Interval of the Difference	Lower	16427,407
		Upper	18379,555
t			35,036
df			473
Sig. (2-tailed)			,000

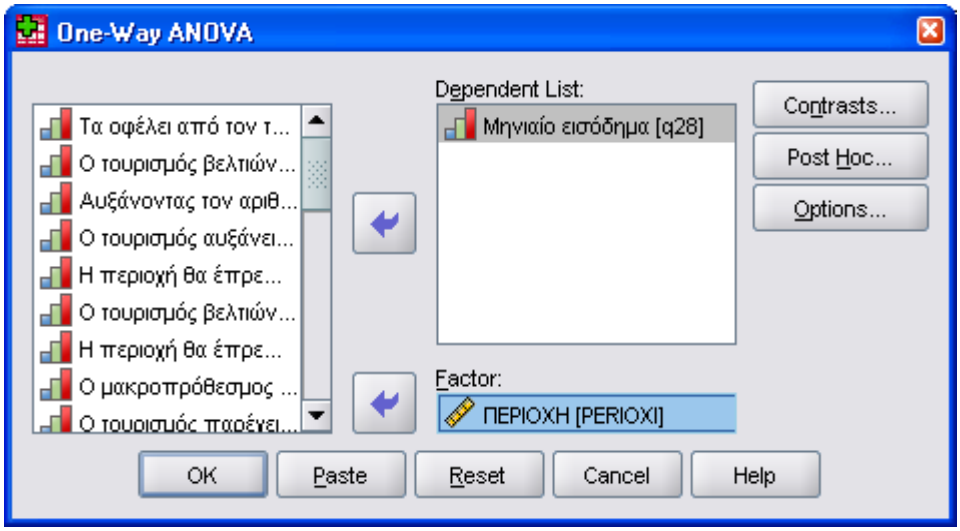
5.5 Σύγκριση των μέσων τιμών πολλών ανεξάρτητων δειγμάτων

Περίπτωση One-Way ANOVA

Η περίπτωση αυτή παρουσιάζει πολλές ομοιότητες με την περίπτωση της παραγράφου 5.3. Θα διαπιστώσουμε όμως, με τη βοήθεια του επόμενου παραδείγματος, ότι δίνει περισσότερες πληροφορίες και εξετάζει πιο διεξοδικά το ίδιο θέμα.

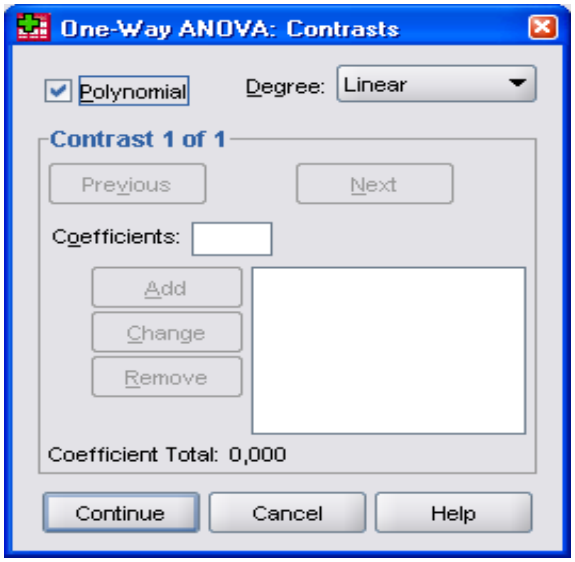
Παράδειγμα: Καταγράφηκαν τα μηνιαία εισοδήματα 288 κατοίκων τριών διαφορετικών τουριστικών περιοχών της Ελλάδας. Για τον έλεγχο της υπόθεσης της ισότητας των μέσων μηνιαίων εισοδημάτων μεταξύ των κατοίκων των τριών περιοχών ακολουθούμε την επόμενη διαδικασία.

- Από το μενού ***Analyze*** επιλέγουμε ***Compare Means***
- στη συνέχεια ***One-Way ANOVA*** και έχουμε την επόμενη εικόνα.



Εικόνα 5.8

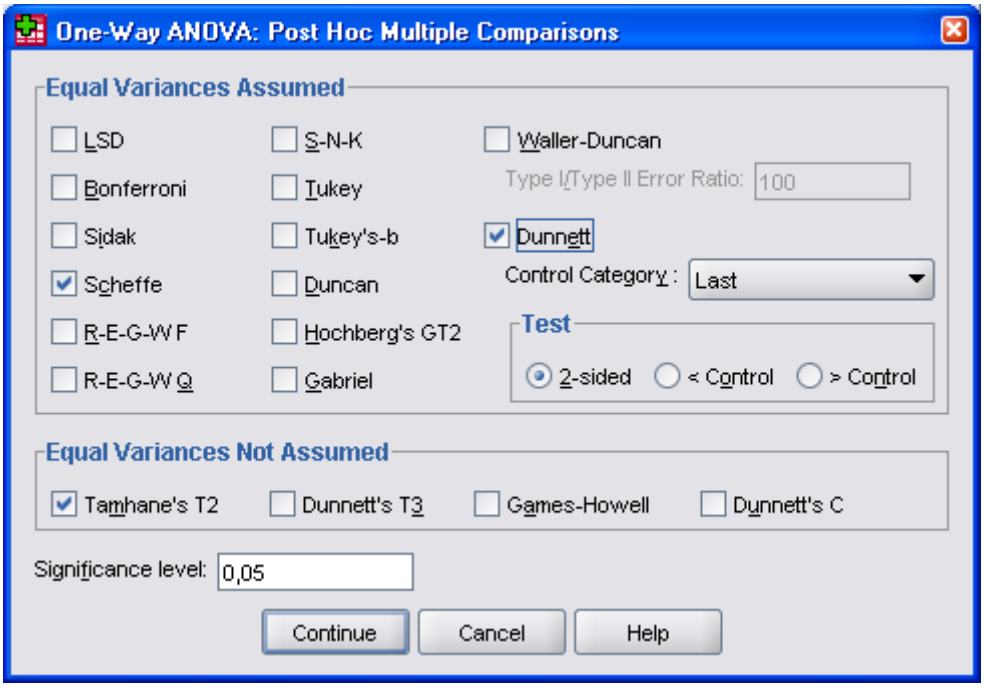
- Στη θέση ***Dependent List*** βάζουμε την ποσοτική μεταβλητή (Εισόδημα)
- Στη θέση ***Factor*** βάζουμε τη μεταβλητή (Περιοχή), η οποία θα χωρίσει σε υποομάδες την ποσοτική μεταβλητή
- Επιλέγουμε ***Contrasts*** και εμφανίζεται η επόμενη φόρμα.



Εικόνα 5.9

- Τσεκάρουμε ***Polynomial*** και στο παράθυρο ***Degree*** επιλέγουμε ***Linear*** ή ***Quadratic*** ή ***Cubic*** ή ***4th*** ή ***5th***

- Με ***continue*** επιστρέφουμε στην εικόνα 5.8, επιλέγουμε ***Post Hoc*** και έχουμε την επόμενη φόρμα



Εικόνα 5.10

- **Στην περιοχή *Equal Variances Assumed* έχουμε τους δείκτες:**
  - ✓ ***LSD***
  - ✓ ***Bonferroni***
  - ✓ ***Sidak*** και
  - ✓ ***Dunnett***

οι οποίοι δίνουν πίνακα **Multiple Comparisons *(πολλαπλών συγκρίσεων)***.

Στην ίδια περιοχή έχουμε επίσης τους δείκτες:

- ✓ ***R-E-G-W-F***
- ✓ ***R-E-G-W-Q***
- ✓ ***S-N-K***
- ✓ ***Tukey's b***

- ✓ ***Duncan***
- ✓ ***Waller-Duncan***

οι οποίοι δίνουν τον πίνακα **Homogeneous Subsets (ομογενείς υποομάδες)**

Τέλος, υπάρχουν και οι δείκτες:

- ✓ ***Scheffe***
- ✓ ***Tukey***
- ✓ ***Hochberg's GT2***
- ✓ ***Gabriel***

οι οποίοι δίνουν και τους δύο προαναφερθέντες πίνακες.

- Στην περιοχή ***Equal Variances not Assumed*** έχουμε κατά σειρά τους δείκτες:
  - ✓ ***Tamhane's T2***
  - ✓ ***Dunnett's T3***
  - ✓ ***Games-Howell***
  - ✓ ***Dunnett's C***

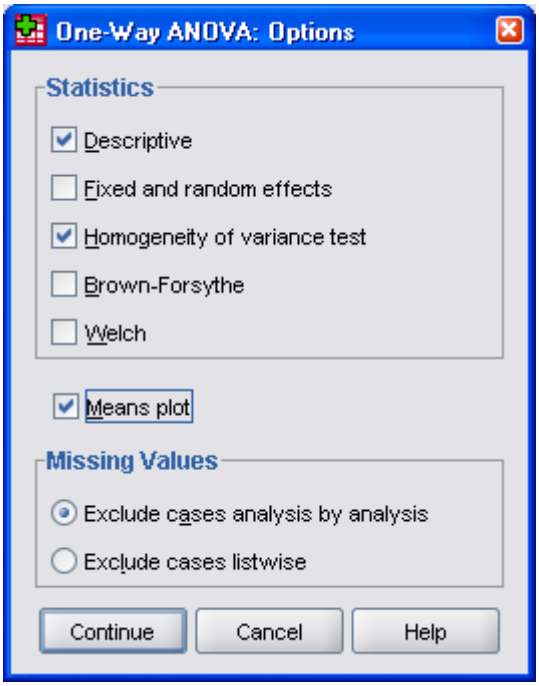
οι οποίοι δίνουν πίνακα **Multiple Comparisons.**

- Τέλος, στο παράθυρο ***Significance Level*** σημειώνουμε το επιθυμητό επίπεδο σημαντικότητας

!!! *Τα αποτελέσματα από τους διάφορους δείκτες, που δίνουν ίδιους πίνακες, διαφέρουν ελάχιστα και έχουμε τη δυνατότητα να επιλέγουμε μόνον αυτούς που θέλουμε.*

Έστω ότι επιλέξαμε τους δείκτες **Sheffe**, **Dunnett και Tamhane's T2** στην εικόνα 5.10.

- Στη συνέχεια με ***Continue*** επιστρέφουμε στην εικόνα 5.8 από την οποία επιλέγουμε ***Options*** και έχουμε τη φόρμα της εικόνας 5.11.



Εικόνα 5.11

- Επιλέγουμε ***Descriptive*** αν θέλουμε τα απλά περιγραφικά μέτρα των υποομάδων
- ***Homogeneity-of-variance*** για να πάρουμε το τεστ ομοιογένειας και
- ***Means plot*** για να έχουμε και τη γραφική απεικόνιση των μέσων
- ***Continue*** και επιστροφή στην εικόνα 5.8, όπου με ***O.K*** παίρνουμε τους παρακάτω πίνακες:

Ο πίνακας ***Descriptives*** δίνει τα απλά στατιστικά μέτρα (**Mean, Std.Deviation, Std.Error, Minimum, Maximum**) καθώς και το διάστημα εμπιστοσύνης των μέσων τιμών των υποομάδων.

Πίνακας 5.11: Descriptives

Εισόδημα

		ΘΑΣΟΣ	ΛΙΜΝΗ ΠΛΑΣΤΗΡΑ	ΝΥΜΦΑΙΟ	Total
N		150	150	117	417
Mean		1056,8000	940,2667	964,6154	989,0168
Std. Deviation		390,48811	380,65635	389,29985	389,17225
Std. Error		31,88322	31,08046	35,99078	19,05785
95% Confidence Interval for Mean	Lower Bound	993,7983	878,8513	893,3311	951,5551
	Upper Bound	1119,8017	1001,6821	1035,8997	1026,4785
Minimum		580,00	580,00	580,00	580,00
Maximum		1700,00	1700,00	1700,00	1700,00

Ο πίνακας ***Test of Homogeneity of Variance*** μας δίνει:

- την τιμή ***Levene Statistic*** και το ***Sig***. αυτού του τεστ. **Η αρχική υπόθεση του τεστ είναι ότι οι υποομάδες έχουν ίσες διακυμάνσεις. Η υπόθεση αυτή απορρίπτεται, σε επίπεδο σημαντικότητας 0,05, αν Sig. < 0,05**.

Στο παράδειγμά μας Sig. = 0,004 < 0,05 και επομένως δεχόμαστε ότι οι διακυμάνσεις των υποομάδων διαφέρουν σημαντικά.

Πίνακας 5.12: Test of Homogeneity of Variance

Εισόδημα

Levene Statistic	df1	df2	Sig.
,692	2	414	,501

Ο πίνακας ***ANOVA*** έχει αυτή τη μορφή και δίνει πολλές πληροφορίες γιατί στην εικόνα 5.9 (**Contrasts**) επιλέξαμε **Polynomial με Degree-Linear**. Αν δεν κάναμε αυτές τις επιλογές, ο πίνακας θα περιείχε μόνο την πρώτη στήλη. Αναλυτικά, λοιπόν, ο πίνακας 5.13 μας δίνει:

- Την τιμή ***F***, της στήλης ***Combined***, η οποία εξετάζει αν υπάρχει διαφορά μεταξύ ενός ή περισσότερων μέσων. **Η αρχική υπόθεση είναι ότι δεν υπάρχει διαφορά μεταξύ των μέσων. Αν Sig.<0,05 απορρίπτεται αυτή η υπόθεση, σε επίπεδο**

σημαντικότητας 0,05 και επομένως τουλάχιστον ένας μέσος διαφέρει από τους υπολοίπους.

- Την τιμή ***F***, της στήλης ***Unweighted***, οποία εμφανίζεται μόνον όταν τα μεγέθη των δειγμάτων δεν είναι ίσα και θεωρεί ότι όλοι οι μέσοι των υποομάδων έχουν το ίδιο βάρος ακόμη και αν τα μεγέθη τους διαφέρουν σημαντικά. **Η αρχική υπόθεση είναι ότι δεν υπάρχει γραμμική σχέση μεταξύ των μέσων. Αν Sig.<0,05 απορρίπτεται αυτή η υπόθεση, σε επίπεδο σημαντικότητας 0,05 και επομένως υπάρχει γραμμική σχέση**.

- Την τιμή ***F***, της στήλης ***Weighted,*** κατά την οποία οι μέσοι των υποομάδων έχουν διαφορετικό βάρος. Ισχύει και εδώ η ίδια υπόθεση με την προηγούμενη.

- Την τιμή ***F***, της στήλης ***Deviation***, για την οποία αναφερθήκαμε στην παράγραφο 5.1 (πίνακας 5.2.1).

Πίνακας 5.13: ANOVA

Εισόδημα

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		1115336	2	557667,928	3,730	,025
	Linear Term	Unweighted	558576,6	1	558576,622	3,736	,054
		Weighted	641420,1	1	641420,137	4,291	,039
		Deviation	473915,7	1	473915,720	3,170	,076
Within Groups			61889961	414	149492,659		
Total			63005297	416			

!!! *Τα αποτελέσματα του πίνακα 5.13, για το συγκεκριμένο παράδειγμα, έχουν ερμηνευτεί στον πίνακα 5.2.1.*

Ο πίνακας ***Multiple Comparisons*** δίνει για τους δείκτες που επιλέξαμε (**Scheffe, Tamhane, Dunnett**):

- ***τις διαφορές του μέσου κάθε υποομάδας από όλες τις άλλες***
- ***την τυπική απόκλιση*** των διαφορών
- ***το Sig.*** των διαφορών
- ***το διάστημα εμπιστοσύνης*** των διαφορών και
- για το δείκτη (**Dunnett**) ***τις διαφορές όλων των υποομάδων*** (**Θάσος, Πλαστήρα**) ***από την τελευταία*** (**Νυμφαίο**)

Πίνακας 5.14: Multiple Comparisons

Dependent Variable: Εισόδημα

	(I) Περιοχή	(J) Περιοχή	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	ΘΑΣΟΣ	ΛΙΜΝΗ ΠΛΑΣΤΗΡ	116,53333*	44,64567	,034	6,8555	226,2112
		ΝΥΜΦΑΙΟ	92,18462	47,68997	,156	-24,9720	209,3412
	ΛΙΜΝΗ ΠΛΑΣΤΗΡ	ΘΑΣΟΣ	-116,53333*	44,64567	,034	-226,2112	-6,8555
		ΝΥΜΦΑΙΟ	-24,34872	47,68997	,878	-141,5053	92,8079
	ΝΥΜΦΑΙΟ	ΘΑΣΟΣ	-92,18462	47,68997	,156	-209,3412	24,9720
		ΛΙΜΝΗ ΠΛΑΣΤΗΡ	24,34872	47,68997	,878	-92,8079	141,5053
Tamhane	ΘΑΣΟΣ	ΛΙΜΝΗ ΠΛΑΣΤΗΡ	116,53333*	44,52566	,028	9,6164	223,4503
		ΝΥΜΦΑΙΟ	92,18462	48,08198	,160	-23,3976	207,7668
	ΛΙΜΝΗ ΠΛΑΣΤΗΡ	ΘΑΣΟΣ	-116,53333*	44,52566	,028	-223,4503	-9,6164
		ΝΥΜΦΑΙΟ	-24,34872	47,55346	,940	-138,6700	89,9725
	ΝΥΜΦΑΙΟ	ΘΑΣΟΣ	-92,18462	48,08198	,160	-207,7668	23,3976
		ΛΙΜΝΗ ΠΛΑΣΤΗΡ	24,34872	47,55346	,940	-89,9725	138,6700
Dunnett t (2-sided)[a]	ΘΑΣΟΣ	ΝΥΜΦΑΙΟ	92,18462	47,68997	,095	-13,2978	197,6670
	ΛΙΜΝΗ ΠΛΑΣΤΗΡ	ΝΥΜΦΑΙΟ	-24,34872	47,68997	,822	-129,8311	81,1337

*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Ο πίνακας ***Homogeneus Subsets*** μάς προτείνει, με δεδομένη την ανισότητα των μέσων των υποομάδων, τις κατηγορίες εκείνες οι οποίες θα μπορούσαν να δημιουργήσουν ενδεχομένως γκρουπ όπου οι μέσες τιμές δεν θα διέφεραν σημαντικά. Έτσι βλέπουμε:

- ένα γκρουπ αποτελούμενο από **Πλαστήρα** και **Νυμφαίο** και με **Sig. = 0,873** (αποδοχή της υπόθεσης της ισότητας των μέσων) και

- ένα γκρουπ αποτελούμενο από τις **Νυμφαίο και Θάσο Sig.= 0,144**.

Πίνακας 5.15: Homogeneous Subsets

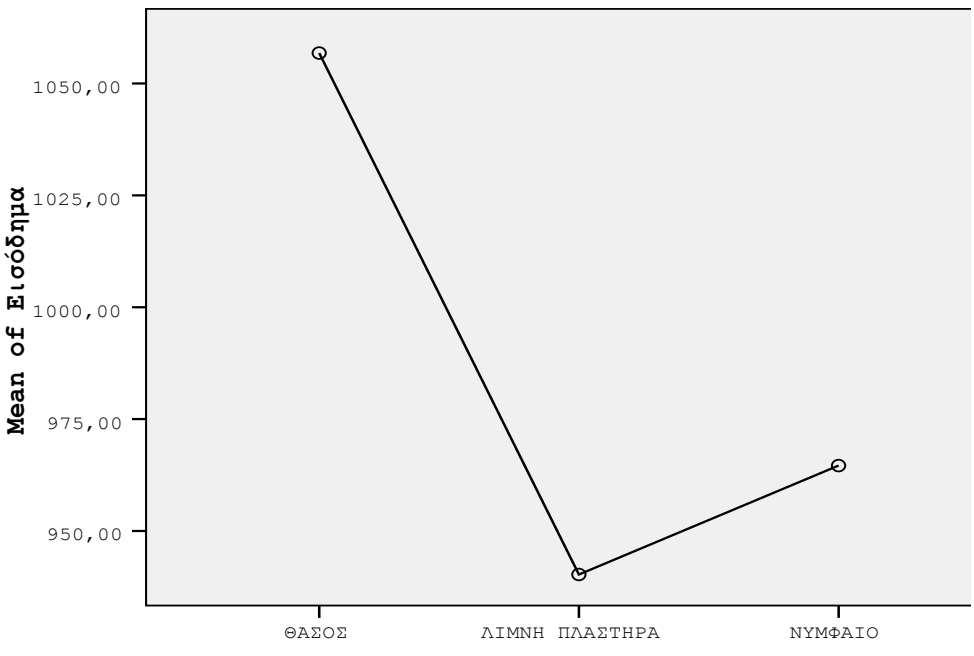
	Περιοχή	N	Subset for alpha = .05	
			1	2
Scheffe[a,b]	ΛΙΜΝΗ ΠΛΑΣΤΗΡΑ	150	940,2667	
	ΝΥΜΦΑΙΟ	117	964,6154	964,6154
	ΘΑΣΟΣ	150		1056,8000
	Sig.		,873	,144

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 137,109.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Στο γράφημα -***Means plots***- που ακολουθεί βλέπουμε την απεικόνιση των τριών μέσων.



Κεφάλαιο 6

Ανάλυση Συσχέτισης - Παλινδρόμησης

Chapter 6

Correlation- Regression Analysis

6. Εισαγωγή

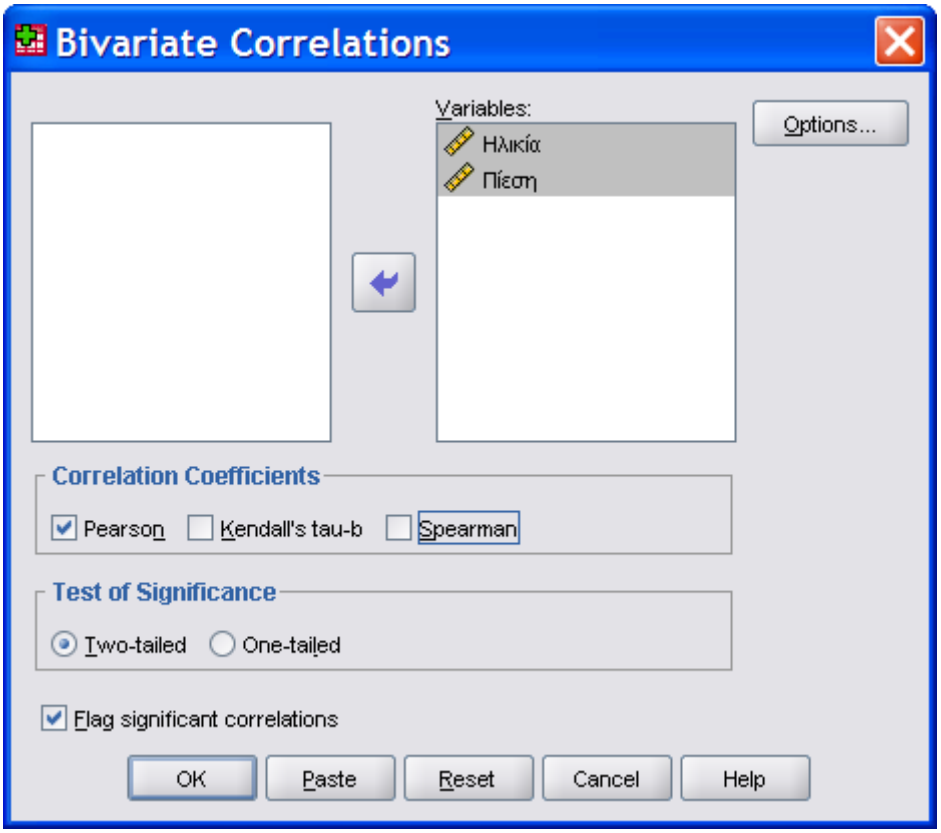
Η αναζήτηση της σχέσης με την οποία θα περιγράψουμε τη συνεργασία μεταξύ δύο ή περισσοτέρων μεταβλητών ιδιοτήτων των μονάδων ενός πληθυσμού μας οδηγεί στη δημιουργία ενός **Μαθηματικού μοντέλου**. Η διαδικασία την οποία χρησιμοποιούμε για τη δημιουργία του **Μοντέλου** αυτού ονομάζεται **Ανάλυση Παλινδρόμησης (Regression Analysis)**. Ο όρος **παλινδρόμηση** χρησιμοποιήθηκε, στη Στατιστική, πρώτα από τον Francis Galton το 1877, ο οποίος παρατήρησε ότι πατέρες χαμηλού αναστήματος αποκτούν αγόρια των οποίων το μέσο ανάστημα μετακινείται προς το μέσο ανάστημα του πληθυσμού.

Ενώ με την Ανάλυση Παλινδρόμησης προσδιορίζουμε τη μαθηματική σχέση η οποία συνδέει δύο ή περισσότερες μεταβλητές, για να μετρήσουμε την ένταση της σχέσης και να προσδιορίσουμε την κατεύθυνσή της χρησιμοποιούμε την **Ανάλυση Συσχέτισης (Correlation Analysis)**. Η ανάλυση συσχέτισης προηγείται της ανάλυσης παλινδρόμησης καθώς με αυτήν ελέγχεται η ύπαρξη σχέσης μεταξύ των μεταβλητών. Αν διαπιστωθεί ικανοποιητική σχέση τότε προχωράμε στην ανάλυση παλινδρόμησης. Για τον υπολογισμό της έντασης της σχέσης χρησιμοποιούμε τους συντελεστές συσχέτισης των **Pearson**, **Spearman** και **Kendall** ανάλογα με το είδος των μεταβλητών.

6.1 Συσχέτιση

Για τον υπολογισμό του συντελεστή συσχέτισης δύο ή περισσότερων ποσοτικών μεταβλητών ακολουθούμε την παρακάτω διαδικασία.

- Από το μενού ***Analyze*** επιλέγουμε ***Correlate*** και στη συνέχεια ***Bivariate***. Εμφανίζεται η επόμενη φόρμα:

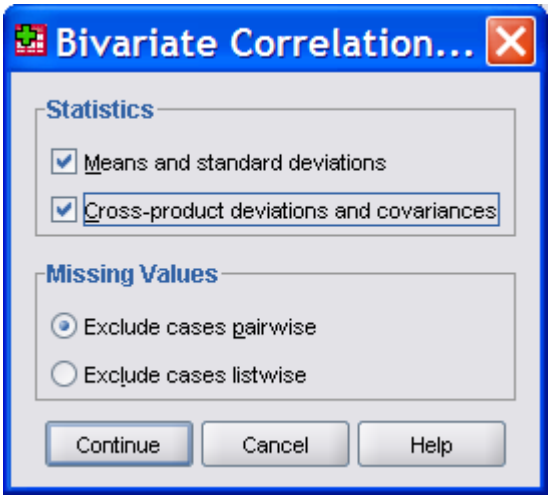


Εικόνα 6.1

- Μεταφέρουμε από το αριστερό παράθυρο στο δεξί τις μεταβλητές που θέλουμε να ελέγξουμε αν σχετίζονται και πόσο και επιλέγουμε ως ***Correlation Coefficient : Pearson***.

- Τσεκάρουμε με √ την επιλογή ***Flag significant correlations*** έτσι ώστε οι συσχετίσεις που είναι στατιστικά σημαντικές να σημειώνονται με ένα ή δύο αστεράκια (*).

- Πατάμε στο κουμπί ***Options*** της φόρμας 6.1 και έχουμε την επόμενη.



Εικόνα 6.2

- Τσεκάρουμε: ***Means and Standard deviations*** και ***Cross-product deviations and covariances,*** στη συνέχεια επιστροφή στην φόρμα 6.1 και με ***O.K*** έχουμε τους παρακάτω πίνακες:

Πίνακας 6.1: Descriptive Statistics

	Mean	Std. Deviation	N
Ηλικία	52,33	11,873	12
Πίεση	14,03	1,508	12

Ο πίνακας 6.1 δίνει:

- Την ***μέση τιμή***, ***την τυπική απόκλιση*** και το ***πλήθος N*** των παρατηρήσεων, για τις δύο μεταβλητές. Στο συγκεκριμένο παράδειγμα έχουμε 12 άτομα με μέση ηλικία τα 52,33 χρόνια και μέση αρτηριακή πίεση 14,03.

Πίνακας 6.2: Correlations

		Ηλικία	Πίεση
Ηλικία	Pearson Correlation	1	,896**
	Sig. (2-tailed)		,000

	Sum of Squares and Cross-products	1550,667	176,467
	Covariance	140,970	16,042
	N	12	12
Πίεση	Pearson Correlation	,896**	1
	Sig. (2-tailed)	,000	
	Sum of Squares and Cross-products	176,467	25,007
	Covariance	16,042	2,273
	N	12	12

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

Στον πίνακα 6.2 έχουμε:

- Τον συντελεστή γραμμικής συσχέτισης ***r*** του Pearson με τιμή 0,896.
- Το επίπεδο σημαντικότητας (Sig.) του συντελεστή συσχέτισης
- Το ***Άθροισμα των Τετραγώνων (sum of squares)*** και τα ***Διαγώνια Γινόμενα (cross- products)***.
- Την ***Συνδιακύμανση (Covariance)*** και
- Το πλήθος N των παρατηρήσεων.

!!!! *Ο συντελεστής γραμμικής συσχέτισης* ***r*** *του Pearson είναι κατάλληλος όταν οι μεταβλητές είναι ποσοτικές σε αναλογική κλίμακα και η σχέση τους γραμμική. Δύο μεταβλητές μπορεί να σχετίζονται τέλεια, αλλά αν η σχέση τους δεν είναι γραμμική ο συντελεστής συσχέτισης του* ***Pearson*** *δεν είναι ο κατάλληλος για να μετρήσει τη σχέση τους.*

!!!! *Ο συντελεστής γραμμικής συσχέτισης* ***r*** *του Pearson παίρνει τιμές στο διάστημα [-1,+1]. Τιμές κοντά στο -1 δείχνουν έντονη αρνητική συσχέτιση κάτι που σημαίνει ότι όσο αυξάνονται οι τιμές της μίας*

μεταβλητής οι τιμές της άλλης ελαττώνονται. Τιμές κοντά στο +1 δείχνουν αντίθετα έντονη θετική συσχέτιση και μεταβολή προς την ίδια κατεύθυνση και των δύο μεταβλητών. Τιμές κοντά στο 0 φανερώνουν έλλειψη σχέσης μεταξύ των μεταβλητών. Στην πράξη, ικανοποιητικές θεωρούνται οι συσχετίσεις με τιμές $|\mathbf{r}| > 0{,}7$.

!!!! *Για τον υπολογισμό του συντελεστή συσχέτισης του Pearson χρησιμοποιούμε τον επόμενο τύπο:* $\mathbf{r} = \dfrac{\sum \mathbf{X_i Y_i} - \nu \overline{\mathbf{XY}}}{\nu \sigma_x \sigma_y}$*,όπου* $\overline{\mathbf{X}}, \overline{\mathbf{Y}}$ *είναι οι αριθμητικοί μέσοι των δύο μεταβλητών,* $\boldsymbol{\sigma_x}, \boldsymbol{\sigma_y}$ *οι τυπικές αποκλίσεις των δύο μεταβλητών και* ***ν*** *το πλήθος των παρατηρήσεων (ζευγών).*

!!!! *Προϋπόθεση για τον υπολογισμό του συντελεστή συσχέτισης του* ***Pearson*** *η* ***διμεταβλητή κανονικότητα (bivariate normal)*** *κάθε ζεύγους μεταβλητών.*

!!!! *Στην περίπτωση που οι μεταβλητές είναι ποσοτικές αλλά δεν μας ενδιαφέρει η σχέση μεταξύ των τιμών των μεταβλητών αλλά η σχέση μεταξύ των* ***τάξεων μεγέθους- Rank Order*** *(σειρά κατάταξης των αρχικών τιμών) των μεταβλητών, χρησιμοποιούμε τους συντελεστές συσχέτισης του* ***Spearman*** *ή του* ***Kendall****.*

!!!! *Τους ίδιους δείκτες χρησιμοποιούμε και όταν οι μεταβλητές είναι* ***διαστημικής κλίμακας ή ιεραρχικής κλίμακας****.*

!!!! *Για τον υπολογισμό του συντελεστή συσχέτισης του* ***Spearman*** $(\boldsymbol{r_s})$ *χρησιμοποιούμε τον τύπο:* $r_s = 1 - \dfrac{6\sum d_i^2}{\nu(\nu^2 - 1)}$, όπου $\mathbf{d_i} = |\mathbf{R_i} - \mathbf{C_i}|$.

Η μεταβλητή $\boldsymbol{R_i}$ *εκφράζει την σειρά κατάταξης των* ***i*** *τιμών της μεταβλητής Χ και η* $\boldsymbol{C_i}$ *την σειρά κατάταξης των* ***i*** *τιμών της μεταβλητής Υ.*

!!!! *Οι συντελεστές συσχέτισης του* ***Spearman*** *και του* ***Kendall****. παίρνουν τιμές στο διάστημα [-1,+1].*

6.2 Γραμμική Παλινδρόμηση

Η **παλινδρόμηση** μπορεί να είναι:

- **Απλή-Simple** (μία μόνον ανεξάρτητη μεταβλητή Χ και μια εξαρτημένη Υ) ή
- **Πολλαπλή–Multiple** (δύο ή περισσότερες ανεξάρτητες μεταβλητές $X_1, X_2,\ldots, X_k$ και μια εξαρτημένη Υ).

Η διερεύνηση της μορφής της παλινδρόμησης είναι το βασικό πρόβλημα το οποίο αρχικά θα πρέπει να επιλυθεί. Είναι δηλαδή απαραίτητο να προσδιορίσουμε αν οι τιμές Χ και Υ προσαρμόζονται καλύτερα σε μία ευθεία, ή παραβολή, ή έλλειψη, ή υπερβολή κ.λ.π. Ο προσδιορισμός αυτός γίνεται με διάφορες θεωρητικές μεθόδους ή ευκολότερα με τη βοήθεια του S.P.S.S.

Ο λόγος που επιβάλλει την εύρεση της **σχέσης** μεταξύ των μεταβλητών και τη δημιουργία του παλινδρομικού μοντέλου είναι η **πρόβλεψη (prediction)** μελλοντικών εξελίξεων και ο **έλεγχος (test)** μελλοντικών γεγονότων.

Η γραμμική παλινδρόμηση είναι από τις πλέον διαδεδομένες μεθόδους ανάλυσης δεδομένων και χρησιμοποιείται για την επίλυση σημαντικών ερευνητικών προβλημάτων. Για την εφαρμογή της είναι απαραίτητο, τα δεδομένα να είναι ποσοτικά ή ποιοτικά σε ιεραρχική (ordinal) κλίμακα

και ο ερευνητής να έχει αποφασίσει ποια μεταβλητή είναι η εξαρτημένη έτσι ώστε οι υπόλοιπες να αποτελέσουν τις ανεξάρτητες.

Για την εξαρτημένη μεταβλητή **Y** (dependent or criterion) και τις ανεξάρτητες μεταβλητές $\mathbf{X_1, X_2, \ldots, X_k}$ (independent or predictor), η εξίσωσης της πολλαπλής γραμμικής παλινδρόμησης η οποία προκύπτει με τη μέθοδο των ελαχίστων τετραγώνων (least square methods) έχει τη γενική μορφή:

$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{1i} + \ldots\ldots + \hat{b}_k X_{ki}$, όπου:

- $\mathbf{i} = 1, 2, \ldots\ldots, \mathbf{v}$ το πλήθος των παρατηρήσεων
- $\hat{Y}_i$: Οι θεωρητικές τιμές της εξαρτημένης μεταβλητής
- $\mathbf{X_{1i}, X_{2i}, \ldots, X_{ki}}$: Οι τιμές των ανεξάρτητων μεταβλητών
- $\hat{b}_0$: Ο σταθερός όρος (constant) και
- $\hat{b}_1, \ldots, \hat{b}_k$: Οι συντελεστές παλινδρόμησης (coefficients). Οι συντελεστές αυτοί, δείχνουν την μέση μεταβολή της εξαρτημένης μεταβλητής όταν η ανεξάρτητη μεταβληθεί κατά μια μονάδα.

Στην περίπτωση που υπάρχει μόνο μια ανεξάρτητη μεταβλητή X, τότε έχουμε **απλή γραμμική παλινδρόμηση** και το μοντέλο περιγράφεται από τη σχέση: $\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$.

Για να υπολογίσουμε τις τιμές των $\hat{b}_0$ και $\hat{b}_1$ αρκεί να λύσουμε το σύστημα των **κανονικών εξισώσεων**:

$$\sum Y_i = v\hat{b}_0 + \hat{b}_1 \sum X_i$$
$$\sum X_i Y_i = \hat{b}_0 \sum X_i + \hat{b}_1 \sum X_i^2 .$$

Μπορούμε επίσης με κατάλληλους μετασχηματισμούς των εξισώσεων του παραπάνω συστήματος να υπολογίσουμε τις τιμές των $\hat{b}_0$ και $\hat{b}_1$ από τις επόμενες σχέσεις: $\hat{b}_1 = \dfrac{\sum X_i Y_i - \nu \bar{X}\bar{Y}}{\sum X_i^2 - \nu \bar{X}^2}$ και $\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$.

Στην περίπτωση που οι ανεξάρτητες μεταβλητές είναι περισσότερες από μια τότε έχουμε **πολλαπλή γραμμική παλινδρόμηση** και οι συντελεστές παλινδρόμησης και ο σταθερός όρος υπολογίζονται με παρόμοιο αλλά ποιο σύνθετο τρόπο. Στην περίπτωση αυτή είναι απαραίτητη η χρήση του S.P.S.S.

Για τον υπολογισμό των συντελεστών παλινδρόμησης και την πρόβλεψη των τιμών της εξαρτημένης μεταβλητής πρέπει να ικανοποιούνται οι παρακάτω υποθέσεις:

- **Γραμμικότητα των μεταβλητών**. Η γραμμικότητα της σχέσης μεταξύ της εξαρτημένης των ανεξάρτητων μεταβλητών εκφράζει τον βαθμό κατά τον οποίο η αλλαγή στις τιμές της εξαρτημένης μεταβλητής είναι σταθερή για τις διάφορες τιμές της ανεξάρτητης μεταβλητής.

- **Σταθερή διακύμανση των διαταρακτικών όρων– λαθών (error terms)**. Η παρουσία άνισων διακυμάνσεων (ετεροσκεδαστικότητα) είναι μια από τις συνηθέστερες αποκλίσεις από τις υποθέσεις.

- **Ανεξαρτησία των διαταρακτικών όρων- λαθών (error terms).** Υποτίθεται ότι κάθε προβλεπόμενη τιμή είναι ανεξάρτητη. Δηλαδή η κάθε θεωρητική τιμή $\hat{Y}_i$ είναι ανεξάρτητη από κάθε άλλη, προηγούμενη ή επόμενη.

- **Κανονικότητα της κατανομής του διαταρακτικού όρου**. Ίσως η συνηθέστερη απόκλιση από τις υποθέσεις είναι η μη

κανονικότητα της εξαρτημένης μεταβλητής ή των ανεξάρτητων μεταβλητών ή και των δύο.

!!! *Για κάθε περίπτωση απόκλισης από τις υποθέσεις, υπάρχουν διορθωτικές κινήσεις. Πολλές φορές όμως συμβαίνει να έχουμε αποκλίσεις σε περισσότερες από μια υποθέσεις και στην προσπάθειά μας να διορθώσουμε την μία, δημιουργούμε πρόβλημα σε άλλη.*

6.2.1 Αξιολόγηση Γραμμικού Μοντέλου

Για την αξιολόγηση του μοντέλου της γραμμικής παλινδρόμησης κάνουμε μια σειρά ελέγχων, οι σημαντικότεροι από τους οποίους είναι οι εξής:

1^{ος} Έλεγχος: Έλεγχος της σημαντικότητας της τιμής **F** (F-κατανομή) του πίνακα **ANOVA** (Analysis of Variance). Η διαμόρφωση του πίνακα παρουσιάζεται στη συνέχεια.

Πίνακας 6.3: ANOVA

	Sum of Squares-SS	**d.f**	**Mean Square-MS**	**F**
Regression	SSR=$\sum(\hat{y}_i - \bar{y})^2$	k	MSR=SSR/k	MSR/MSE
Residual (Error)	SSE=$\sum(y_i - \hat{y}_i)^2$	ν-k-1	MSE=SSE/ν-k-1	
Total	SST=$\sum(y_i - \bar{y})^2$	ν-1		

Στον πίνακα 6.1 έχουμε:

✓ Το άθροισμα των τετραγώνων των διαφορών που οφείλονται στη γραμμή της παλινδρόμησης (***regression***). Είναι οι διαφορές των θεωρητικών τιμών $\hat{y}_i$ από τη μέση τιμή $\bar{y}$ των εμπειρικών τιμών. **SSR=**$\sum(\hat{y}_i - \bar{y})^2$

✓ Το άθροισμα των τετραγώνων των καταλοίπων -σφαλμάτων (***residual or error***), δηλαδή των διαφορών μεταξύ εμπειρικών y_i και θεωρητικών $\hat{y}_i$ τιμών. **SSE=** $\sum(y_i - \hat{y}_i)^2$

✓ Το συνολικό άθροισμα των τετραγώνων των διαφορών (***total***), δηλαδή των διαφορών των εμπειρικών τιμών y_i από τη μέση τιμή $\bar{y}$ αυτών. **SST=** $\sum(y_i - \bar{y})^2$.

Ισχύει: **SST=SSR+SSE**

✓ Τους βαθμούς ελευθερίας (***df) k, ν-k-1 και ν-1*** αντίστοιχα, όπου **k** το πλήθος των ανεξάρτητων μεταβλητών και **ν** το σύνολο των παρατηρήσεων.

✓ Τον μέσο των προηγούμενων αθροισμάτων (***sum. of square/df***) και

✓ Την τιμή του ***F*** κριτηρίου (***MSR/MSE***). Την τιμή ***F*** τη συγκρίνουμε με την τιμή της κατανομής **F** σε δεδομένο επίπεδο σημαντικότητας **α** και με **k** βαθμούς ελευθερίας στον αριθμητή και **ν-k-1** βαθμούς ελευθερίας στον παρονομαστή.

Αν $\mathbf{F > F_{\alpha,\ k,\ \nu-k-1}}$ δεχόμαστε την στατιστική σημαντικότητα των σχέσεων μεταξύ των μεταβλητών και την συνολική καταλληλότητα του μοντέλου.

2ος Έλεγχος: Έλεγχος της τιμής του συντελεστή προσδιορισμού $\mathbf{R^2}$ (coefficient of determination), ο οποίος εκφράζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τις ανεξάρτητες μεταβλητές. Ο συντελεστής προσδιορισμού υπολογίζεται από τη σχέση $\mathbf{R^2=SSR/SST}$. Οι τιμές του κυμαίνονται από 0 έως 1 και τιμές μεγαλύτερες του 0,5 θεωρούνται ικανοποιητικές. Για ασφαλέστερα συμπεράσματα η τιμή του πρέπει να ελέγχεται σε

σχέση με το μέγεθος του δείγματος, το επίπεδο σημαντικότητας και το πλήθος των ανεξάρτητων μεταβλητών. Από τον συντελεστή προσδιορισμού $\mathbf{R^2}$ προκύπτει και ο διορθωμένος συντελεστής προσδιορισμού $\overline{\mathbf{R}}^{\mathbf{2}}$ **(Adjusted R-square)** με τιμές επίσης από 0 έως 1. Ο τύπος υπολογισμού του είναι ο εξής:

$\overline{R}^2 = 1-(1-R^2)\frac{\nu-1}{\nu-k-1}$, όπου **k** το πλήθος των ανεξάρτητων μεταβλητών και **ν** το πλήθος των παρατηρήσεων.

3ος Έλεγχος: Έλεγχος της σημαντικότητας του σταθερού όρου και των συντελεστών παλινδρόμησης σε συγκεκριμένο και καθορισμένο επίπεδο σημαντικότητας α (συνήθως α=5%). Η υπόθεση την οποία ελέγχουμε στη συγκεκριμένη περίπτωση είναι ότι οι συντελεστές είναι 0.

4ος Έλεγχος: Έλεγχος της **αυτοσυσχέτισης** (autocorrelation) με τη χρήση του δείκτη **Durbin-Watson** (τιμές από 0 έως 4). Τιμές κοντά στο 2 δείχνουν ότι το φαινόμενο της αυτοσυσχέτισης δεν είναι έντονο.

5ος Έλεγχος: Έλεγχος της **συγγραμμικότητας-πολυσυγγραμμικότητας** (linearity- multicollinearity) με τη χρήση των δεικτών **Tolerance** (τιμές από 0 έως 1) και **VIF (variance inflation factor)**. Η τιμή του δείκτη **VIF** προκύπτει από τη σχέση **VIF= =1/Tolerance**. Ο δείκτης Tolerance παίρνει τιμές στο διάστημα [0-1] και για μικρές τιμές (κοντά στο 0) η μεταβλητή είναι σχεδόν σε γραμμικό συνδυασμό με τις άλλες ανεξάρτητες μεταβλητές. Ο δείκτης VIF μεγαλώνει όταν ο δείκτης Tolerance μικραίνει. Συνήθως, ένα πρώτο φίλτρο αποτελεί η τιμή 5, ενώ ένα δεύτερο πιο ελαστικό φίλτρο

είναι η τιμή 10. Ανεξάρτητες μεταβλητές με δείκτη VIF μεγαλύτερο του 10 συνιστάται να αποβάλλονται από το μοντέλο.

6.2.2 Δείγμα

Εκτός των ελέγχων που αναφέρθηκαν στην προηγούμενη παράγραφο πρέπει να γίνει έλεγχος και της καταλληλότητας του δείγματος ως προς το μέγεθος και την αντιπροσωπευτικότητα. Το μέγεθος του δείγματος έχει άμεση επίδραση στην καταλληλότητα και την στατιστική ισχύ της παλινδρόμησης. Μικρά δείγματα, με λιγότερες από 20 παρατηρήσεις, είναι κατάλληλα μόνο για απλή παλινδρόμηση. Επίσης δείγματα με 1.000 ή περισσότερες παρατηρήσεις κάνουν τους στατιστικούς ελέγχους υπερβολικά ευαίσθητους, δείχνοντας σχεδόν όλες τις σχέσεις στατιστικά σημαντικές. Με τα μεγάλα δείγματα ο αναλυτής πρέπει να είναι σίγουρος ότι τα κριτήρια της πρακτικής σημαντικότητας ταυτίζονται με αυτά της στατιστικής σημαντικότητας. Το μέγεθος του δείγματος, το επίπεδο σημαντικότητας το οποίο επιλέγεται και το πλήθος των ανεξάρτητων μεταβλητών διαδραματίζουν σημαντικό ρόλο στην δημιουργία στατιστικά σημαντικού δείκτη προσδιορισμού R^2, σε δεδομένο επίπεδο ισχύος. Έτσι, ένα δείγμα 50 ατόμων με 5 ανεξάρτητες μεταβλητές και σε επίπεδο σημαντικότητας 5% απαιτεί R^2 τουλάχιστον 0.23. Το ίδιο μέγεθος δείγματος με 10 ανεξάρτητες μεταβλητές και στο ίδιο επίπεδο σημαντικότητας απαιτεί R^2 τουλάχιστον 0.29. Το μέγεθος του δείγματος επιδρά και στην γενίκευση των αποτελεσμάτων, με την σχέση μεταξύ των παρατηρήσεων και του πλήθους των ανεξάρτητων μεταβλητών. Το ελάχιστα αποδεκτό όριο είναι 5 παρατηρήσεις για κάθε ανεξάρτητη μεταβλητή, ενώ το επιθυμητό είναι συνήθως 15 με 20

παρατηρήσεις για κάθε ανεξάρτητη μεταβλητή. Αν επιτευχθεί αυτή η αναλογία και με δεδομένη την αντιπροσωπευτικότητα του δείγματος τα αποτελέσματα μπορούν να γενικευθούν. Στην περίπτωση που εφαρμόζουμε η μέθοδος της **διαδοχικής επιλογής** ανεξάρτητων μεταβλητών (**Stepwise Regression Method**s) η αναλογία αυξάνεται σε 50 προς 1.

Για την ολοκλήρωση της διαδικασίας της πολλαπλής παλινδρόμησης έχουμε την δυνατότητα επιλογής μεταξύ των μεθόδων ***Enter, Stepwise, Remove, Backward*** *και* ***Forward.***

Συγκεκριμένα θα αναπτυχθεί η μέθοδος **Enter** και η μέθοδος **Stepwise**.

Με τη μέθοδο αυτή ο ερευνητής επιθυμεί τη δημιουργία μοντέλου με όλες τις διαθέσιμες ανεξάρτητες μεταβλητές και στη συνέχεια αξιολογεί το μοντέλο, τους συντελεστές και τους δείκτες οι οποίοι προκύπτουν.

Θα ξεκινήσουμε με ένα απλό παράδειγμα για τον υπολογισμό των απαραίτητων μέτρων έτσι ώστε να έχουμε μια πρώτη και πολύ απλή προσέγγιση του θέματος.

6.3 Διαδικασία Δημιουργίας και Ελέγχου Γραμμικού Παλινδρομικού Μοντέλου

Στην παράγραφο αυτή θα αναπτυχθεί η διαδικασία δημιουργίας και ελέγχου ενός γραμμικού παλινδρομικού μοντέλου με τη χρήση του S.P.S.S. Για το σκοπό αυτό θα χρησιμοποιηθούν δύο συγκεκριμένα παραδείγματα. Το πρώτο παράδειγμα θα βοηθήσει στον υπολογισμό των απαραίτητων δεικτών, για μια πρώτη και πολύ απλή προσέγγιση του θέματος. Το δεύτερο παράδειγμα θα βοηθήσει στην λεπτομερή και

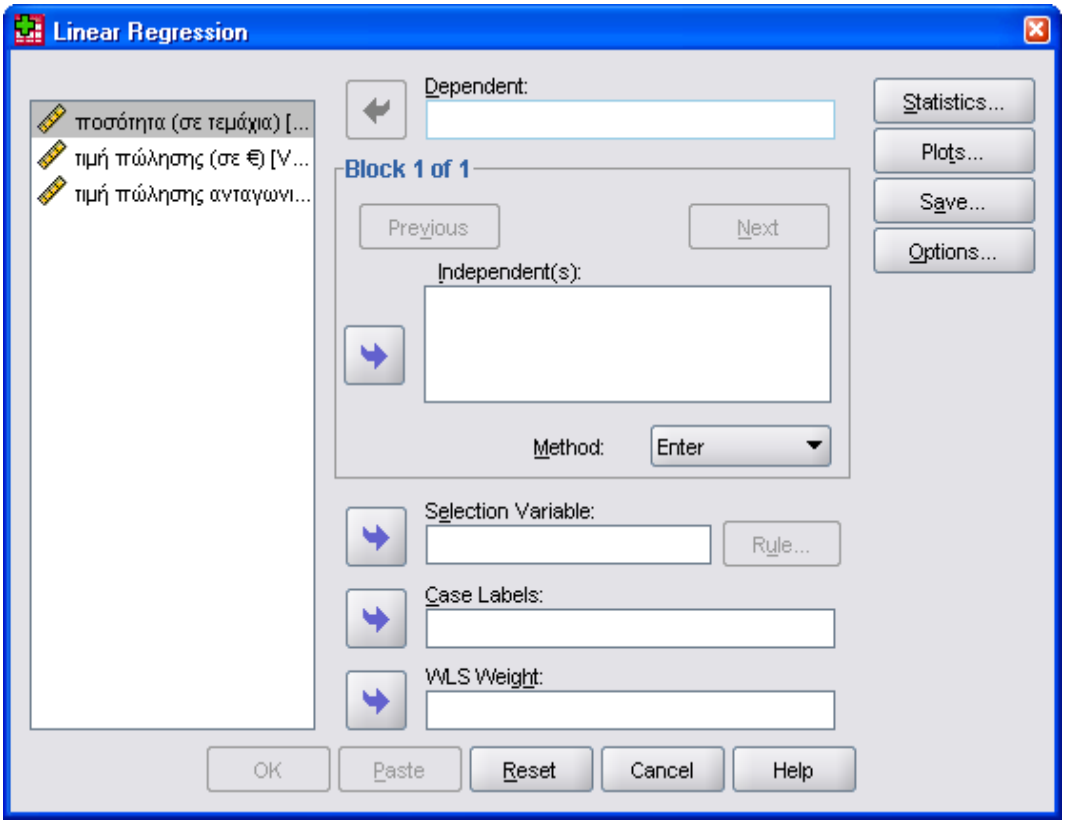
αναλυτική παρουσίαση των δυνατοτήτων όλων των επιλογών οι οποίες είναι διαθέσιμες στο S.P.S.S και οι οποίες είναι απαραίτητες για μια ολοκληρωτική και εμπεριστατωμένη ανάλυση.

6.3.1 Μέθοδος Enter

Με τη μέθοδο αυτή ο ερευνητής επιθυμεί τη δημιουργία μοντέλου με όλες τις διαθέσιμες ανεξάρτητες μεταβλητές και στη συνέχεια αξιολογεί το μοντέλο, τους συντελεστές και τους δείκτες οι οποίοι προκύπτουν.

Παράδειγμα 1^{ο}: Έστω ότι έχουμε τις μεταβλητές **ποσότητα** (η μηνιαία ποσότητα πώλησης συγκεκριμένου προϊόντος, σε τεμάχια), **τιμή πώλησης** (μέση τιμή πώλησης του συγκεκριμένου προϊόντος, ανά τεμάχιο), **τιμή πώλησης ανταγωνιστών** (μέση τιμή πώλησης ιδίου προϊόντος τριών ανταγωνιστικών εταιριών, ανά τεμάχιο). Υποθέτοντας ότι η ποσότητα πώλησης εξαρτάται από την τιμή πώλησης του συγκεκριμένου προϊόντος αλλά και από τη μέση τιμή πώλησης των ανταγωνιστριών εταιριών, θέλουμε να κατασκευάσουμε και να ελέγξουμε το παλινδρομικό μοντέλο.

- Από το μενού ***Analyze*** επιλέγουμε ***Regression*** και στη συνέχεια ***Linear***.

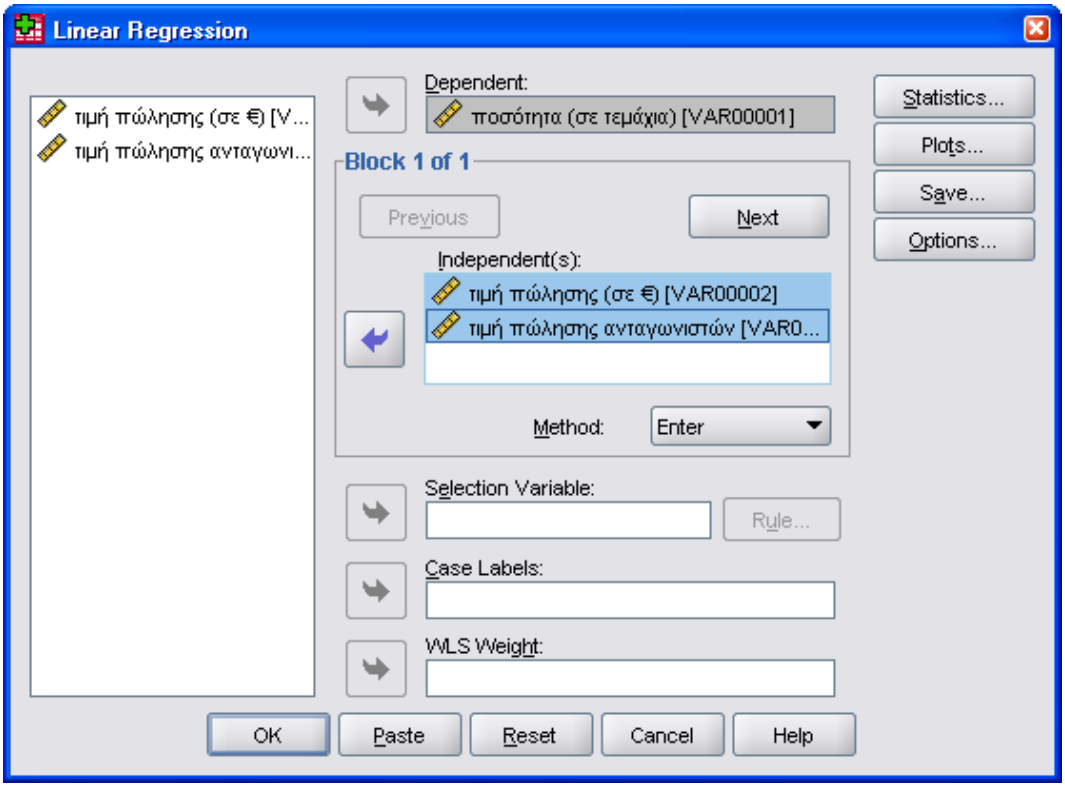


Εικόνα 6.3

Στην πάνω φόρμα, στο μεγάλο παράθυρο αριστερά διακρίνουμε όλες τις μεταβλητές.

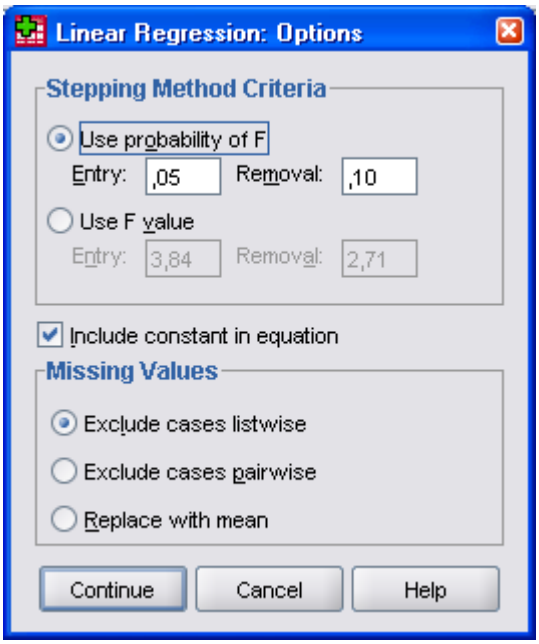
- Επιλέγουμε την εξαρτημένη μεταβλητή (**Ποσότητα**) και τη μεταφέρουμε στο παράθυρο ***dependent***.

- Επιλέγουμε τις δύο ανεξάρτητες μεταβλητές (**τιμή πώλησης, τιμή πώλησης ανταγωνιστών**) και τις μεταφέρουμε στο παράθυρο ***independent***. Η προηγούμενη φόρμα έχει πλέον τη μορφή της εικόνας που ακολουθεί.



Εικόνα 6.4

- Πατάμε στο κουμπί ***Options*** και εμφανίζεται η επόμενη εικόνα



Εικόνα 6.5

Στη φόρμα αυτή έχουμε ουσιαστικά τα ***κριτήρια*** εισόδου-εξόδου των μεταβλητών στο μοντέλο.

- Ένα κριτήριο μπορεί να είναι το ***επίπεδο σημαντικότητας (significance level***) της τιμής ***F (Use probability of F).*** Με βάση το κριτήριο αυτό, η μεταβλητή μπαίνει στο μοντέλο αν το ***επίπεδο σημαντικότητας (significance level)*** για κάθε ***F τιμή*** είναι μικρότερο από την τιμή που δώσαμε στο παράθυρο ***Entry***. Αντίθετα αφαιρείται αν το ***επίπεδο σημαντικότητας (significance level)*** για κάθε ***F τιμή*** είναι μεγαλύτερο από την τιμή που δώσαμε στο παράθυρο ***Removal***. Σε κάθε περίπτωση η τιμή ***Entry*** πρέπει να είναι μικρότερη από την τιμή ***Removal*** και μάλιστα πρέπει να είναι και οι δύο θετικές. Συνήθως, για να βάλουμε περισσότερες μεταβλητές στο μοντέλο μεγαλώνουμε την τιμή ***Entry***, ενώ για να αφαιρέσουμε περισσότερες μεταβλητές μικραίνουμε την τιμή ***Removal***.

- Ένα άλλο κριτήριο μπορεί να είναι η τιμή ***F (Use F Value)***. Με βάση το κριτήριο αυτό η μεταβλητή μπαίνει στο μοντέλο αν κάθε ***F τιμή*** είναι μεγαλύτερη από την τιμή που δώσαμε στο παράθυρο ***Entry***. Αντίθετα αφαιρείται αν κάθε ***F τιμή*** είναι μικρότερη από την τιμή που δώσαμε στο παράθυρο ***Removal***. Σε κάθε περίπτωση, η τιμή ***Entry*** πρέπει να είναι μεγαλύτερη από την τιμή ***Removal*** και μάλιστα πρέπει να είναι και οι δύο θετικές. Συνήθως για να βάλουμε περισσότερες μεταβλητές στο μοντέλο μικραίνουμε την τιμή ***Entry***, ενώ για να αφαιρέσουμε περισσότερες μεταβλητές μεγαλώνουμε την τιμή ***Removal***.

- Η ένδειξη ***Include constant in equation*** τσεκάρεται για να πάρουμε το σταθερό όρο του μοντέλου της παλινδρόμησης και στη

συνέχεια ***Continue***, επιστροφή στην εικόνα 6.4 και με ***O.K*** εμφάνιση των αποτελεσμάτων, όπως αυτά φαίνονται στους επόμενους πίνακες.

Αναλυτικότερα στον πίνακα ***Model Summary*** (Πίνακας 6.4) βλέπουμε:

- Το συντελεστή συσχέτισης (***R)***
- Το δείκτη προσδιορισμού (***R Square)***
- Το διορθωμένο δείκτη προσδιορισμού (***Adjusted R Square)*** και
- Το τυπικό σφάλμα της εκτίμησης (***Std. Error of the estimate***)

Πίνακας 6.4: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,922[a]	,850	,822	1,74048

a. Predictors: (Constant), τιμή πώλησης ανταγωνιστών (σε €), τιμή πώλησης (σε €)

Στον πίνακα ***ANOVA*** (Πίνακας 6.5) έχουμε:

- Το άθροισμα τετραγώνων της παλινδρόμησης (***regression***), το άθροισμα τετραγώνων των σφαλμάτων (***residual***) και το συνολικό άθροισμα τετραγώνων (***total***).
- Τους βαθμούς ελευθερίας ***df (k, v-k-1, v-1)*** αντίστοιχα
- Τον μέσο των προηγούμενων αθροισμάτων (***sum. Of square/df***)
- Την τιμή του ***F*** κριτηρίου (***Mean square Regression/Mean square Residual***) και, τέλος,
- Το ***Sig***. (περιθώριο λάθους της εκτίμησης).

Πίνακας 6.5: ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	188,392	2	94,196	31,095	,000[a]
	Residual	33,322	11	3,029		
	Total	221,714	13			

a. Predictors: (Constant), τιμή πώλησης ανταγωνιστών (σε €), τιμή πώλησης (σε €)
b. Dependent Variable: Ποσότητα (σε τεμάχια)

Τέλος, ο πίνακας ***Coefficients*** (Πίνακας 6.6) δίνει:

- Τους συντελεστές ***b_0, b_1 και b_2 (constant, τιμή αγοράς, τιμή ανταγωνιστών)*** στη στήλη ***B***.
- Το τυπικό σφάλμα αυτών των τιμών στη στήλη ***Std. Error***
- Τις τιμές του ***t-test*** (τιμές $|t| > 1.96$ δείχνουν στατιστικά σημαντικούς συντελεστές παλινδρόμησης, σε επίπεδο σημαντικότητας 5%) και τέλος
- Το ***sig.*** του t-test για τους συντελεστές.

Πίνακας 6.6: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18,874	15,195		1,242	,240
	τιμή πώλησης (σε €)	-,119	,049	-,473	-2,420	,034
	τιμή πώλησης ανταγωνιστών (σε €)	,130	,051	,499	2,554	,027

a. Dependent Variable: Ποσότητα (σε τεμάχια)

Με βάση τα αποτελέσματα του πίνακα 6.6, το συγκεκριμένο μοντέλο θα έχει τη μορφή:

Ποσότητα = 18,874-0,119*τιμή πώλησης+0,130*τιμή πώλησης ανταγωνιστών.

Από τη στήλη **Beta** προκύπτει ότι η μεταβλητή «τιμή πώλησης ανταγωνιστών» επηρεάζει περισσότερο τη μεταβλητή «ποσότητα» γιατί η τυποποιημένη τιμής της είναι μεγαλύτερη, κατ' απόλυτο τιμή, από την αντίστοιχη της μεταβλητής «τιμή πώλησης». Επίσης από τη στήλη των τιμών t και των αντίστοιχων τιμών Sig. προκύπτει ότι ενώ οι συντελεστές των ανεξάρτητων μεταβλητών είναι στατιστικά σημαντικοί σε επίπεδο σημαντικότητας 5% (sig.<0,05), ο σταθερός όρος είναι στατιστικά ασήμαντος (sig.=0,240>0,05).

*!!!! Οι τιμές **Beta** είναι ιδιαίτερα χρήσιμες, όταν οι μονάδες μέτρησης των μεταβλητών είναι διαφορετικές.*

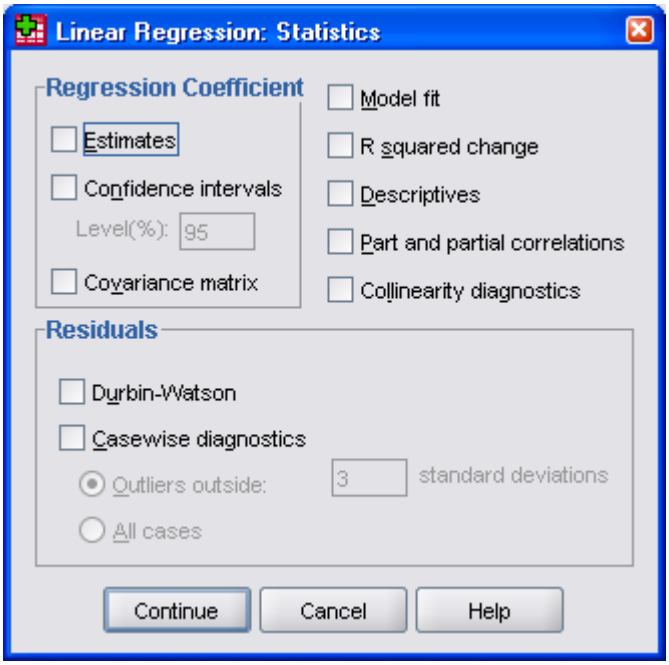
Τα προβλήματα τα οποία έχουμε να αντιμετωπίσουμε στην πολλαπλή παλινδρόμηση είναι συνήθως πολλά και σύνθετα. Για το λόγο αυτό, η απλή και επιφανειακή αντιμετώπιση, όπως στο προηγούμενο παράδειγμα, δίνει μια γενική μόνον εικόνα του προβλήματος. Είναι απολύτως απαραίτητο, πριν οριστικοποιήσουμε το μοντέλο μας, να προβούμε σε πολλούς ελέγχους για να πετύχουμε ισχυρή προβλεπτική ικανότητα με μεγάλο βαθμό αξιοπιστίας. Στο παράδειγμα το οποίο θα ακολουθήσει θα αναλύσουμε τη χρήση και τη χρησιμότητα των προσφερόμενων επιλογών, βήμα-βήμα.

Παράδειγμα 2ο: Έστω οι μεταβλητές ***Τιμή πώλησης*** (σε χιλ. €), ***Εμβαδόν*** (σε m^2), ***Όροφος***, ***Πλήθος μπάνιων*** και ***παλαιότητα*** (σε έτη). Θέλουμε να ελέγξουμε αν και κατά πόσο η τιμή πώλησης ενός διαμερίσματος εξαρτάται από τα τετραγωνικά μέτρα, τον όροφο, τον αριθμό των μπάνιων και την ηλικία του διαμερίσματος. Συγχρόνως, θέλουμε να προσδιορίσουμε αναλυτικότερα την επίδραση της κάθε μεταβλητής στο μοντέλο και να ελέγξουμε την αξιοπιστία των αποτελεσμάτων.

Η διαδικασία είναι κατ' αρχάς η ίδια με αυτή του πρώτου παραδείγματος. Έτσι ξεκινώντας:

- Από το μενού ***Analyze*** επιλέγουμε ***Regression*** και στη συνέχεια ***Linear***.
- Εμφανίζεται εικόνα όμοια με την 6.3, με όλες τις μεταβλητές τις οποίες έχουμε εισαγάγει στον ***Data Editor***.

- Μεταφέρουμε στη θέση ***Dependent*** τη μεταβλητή ***Τιμή πώλησης***
- Στη θέση ***Independents*** τις μεταβλητές ***Εμβαδόν, Όροφος, Πλήθος μπάνιων και παλαιότητα.***
- Στη θέση ***Method*** αφήνουμε την ένδειξη ***Enter***
- Στη θέση ***Case Labels*** θα μπορούσαμε να βάλουμε μια άλλη μεταβλητή, συνήθως μη ποσοτική, όπως για παράδειγμα η περιοχή του διαμερίσματος. Αυτό συνήθως συμβαίνει όταν το πλήθος των τιμών είναι μεγάλο και η μεταβλητή είναι σημαντική.
- Στη συνέχεια πατάμε στο κουμπί ***Statistics*** και εμφανίζεται η επόμενη φόρμα.



Εικόνα 6.6

Θα δούμε αναλυτικά τι δίνει η κάθε επιλογή αυτής της φόρμας, ανά περιοχή.

α. Περιοχή Regression Coefficients

α.1: Από την επιλογή ***Estimates*** προκύπτει ο πίνακας 6.7 **(Coefficients[a])** στον οποίο περιέχονται **ο σταθερός όρος** και **οι συντελεστές μερικής παλινδρόμησης**, οι **τυποποιημένες τιμές** των συντελεστών μερικής παλινδρόμησης (Beta), οι τιμές του ***t-test*** και το ***Sig***., με βάση το οποίο δεχόμαστε ή απορρίπτουμε την μηδενική υπόθεση τη σχετική με την σημαντικότητα των συντελεστών.

Η εξίσωση του συγκεκριμένου μοντέλου θα είναι η ακόλουθη:

Τιμή πώλησης = -6,836+0,803*m^2-7,642*Όροφος+4,590*Μπάνια-1,274*έτη.

Πίνακας 6.7: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-6,836	7,141		-,957	,341
	Εμβαδόν (σε m2)	,803	,050	1,005	16,178	,000
	Όροφος	-7,642	5,000	-,063	-1,528	,130
	Πλήθος Μπάνιων	4,590	4,413	,056	1,040	,301
	Παλαιότητα (σε έτη)	-1,274	,511	-,078	-2,494	,015

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

!!!! *Ελέγχοντας το* ***Sig.*** *των συντελεστών σε όλες τις περιπτώσεις, δεχόμαστε ότι ο σταθερός όρος και οι συντελεστές μερικής παλινδρόμησης του ορόφου και του πλήθους μπάνιων είναι στατιστικά ασήμαντοι.*

α.2: Από την επιλογή ***Confidence intervals*** προκύπτει ο πίνακας 6.8 **(Coefficients[a])** ο οποίος περιέχει τα ***κατώτερα (Lower Bound) και ανώτερα (Upper Bound) όρια του διαστήματος εμπιστοσύνης*** των συντελεστών που υπολογίσαμε στην προηγούμενη ενέργεια.

Πίνακας 6.8: Coefficients

Model		95% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	-21,034	7,363
	Εμβαδόν (σε m2)	,705	,902
	Όροφος	-17,584	2,299
	Πλήθος Μπάνιων	-4,185	13,366
	Παλαιότητα (σε έτη)	-2,289	-,259

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

α.3: Από την επιλογή **Covariance matrix** προκύπτει ο πίνακας 6.9 **(Coefficients Correlations)** στον οποίο έχουμε τους ***συντελεστές συσχέτισης*** και τις ***συνδιακυμάνσεις*** μεταξύ των ανεξάρτητων μεταβλητών.

Πίνακας 6.9: Coefficient Correlations

Model			Παλαιότητα (σε έτη)	Πλήθος Μπάνιων	Όροφος	Εμβαδόν (σε m2)
1	Correlations	Παλαιότητα (σε έτη)	1,000	,513	-,158	-,501
		Πλήθος Μπάνιων	,513	1,000	-,181	-,743
		Όροφος	-,158	-,181	1,000	-,368
		Εμβαδόν (σε m2)	-,501	-,743	-,368	1,000
	Covariances	Παλαιότητα (σε έτη)	,261	1,157	-,403	-,013
		Πλήθος Μπάνιων	1,157	19,479	-3,999	-,163
		Όροφος	-,403	-3,999	25,001	-,091
		Εμβαδόν (σε m2)	-,013	-,163	-,091	,002

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

!!!! *Ο μεγάλος συντελεστής συσχέτισης μεταξύ κάποιων μεταβλητών δημιουργεί πρόβλημα συγγραμικότητας (linearity) ή πολυσυγγραμμικότητας (multicollinearity), η οποία εξηγεί γιατί οι μερικοί συντελεστές είναι στατιστικά ασήμαντοι.*

α.4: Από την επιλογή **Model fit** έχουμε:

- Τον πίνακα 6.10 **(Model Summary)** ο οποίος περιέχει, κατά σειρά, τον ***συντελεστή πολλαπλής συσχέτισης (R),*** το ***δείκτη προσδιορισμού (R Square)***, το ***διορθωμένο δείκτη προσδιορισμού (Adjusted R Square)*** και το ***τυπικό σφάλμα της εκτίμησης (Std. Error of the Estimate).***

Πίνακας 6.10: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,974[a]	,949	,947	18,992

a. Predictors: (Constant), Παλαιότητα (σε έτη), Πλήθος Μπάνιων , Όροφος, Εμβαδόν (σε m2)

!!!! *Παρατηρούμε ότι η τιμή του δείκτη* $\boldsymbol{R^2}$ *είναι πολύ ικανοποιητική και μπορούμε να ισχυριστούμε ότι το 94,9% των μεταβολών της μεταβλητής τιμή πώλησης ερμηνεύεται από τις μεταβολές των μεταβλητών παλαιότητα, πλήθος μπάνιων, όροφος και εμβαδόν, ενώ το υπόλοιπο 5,1% οφείλεται σε τυχαίους και ανερμήνευτους παράγοντες.*

- Τον πίνακα 6.11 ***(ANOVA)*** ο οποίος περιέχει:

- Το άθροισμα τετραγώνων της παλινδρόμησης (***regression***), το άθροισμα τετραγώνων των σφαλμάτων (***residual***) και το συνολικό άθροισμα τετραγώνων (***total***).
- Τους βαθμούς ελευθερίας ***df (k, v-k-1, v-1)*** αντίστοιχα
- Τον μέσο των προηγούμενων αθροισμάτων (***sum. Of square/df***)
- Την τιμή του ***F*** κριτηρίου (***MSR/MSE***) και τέλος
- Το ***Sig***. (περιθώριο λάθους της εκτίμησης).

Πίνακας 6.11: ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	573290,6	4	143322,656	397,353	,000[a]
	Residual	30658,914	85	360,693		
	Total	603949,5	89			

a. Predictors: (Constant), Παλαιότητα (σε έτη), Πλήθος Μπάνιων , Όροφος, Εμβαδόν (σε m2)

b. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

!!!! *Από τον πίνακα 6.11 και τον έλεγχο της τιμής του στατιστικού F προκύπτει η σημαντικότητα των σχέσεων μεταξύ των μεταβλητών και κατά συνέπεια η σημαντικότητα του μοντέλου.*

α.5: Την επιλογή ***R Squared Change*** τη χρησιμοποιούμε όταν πραγματοποιούμε ιεραρχική παλινδρόμηση, περίπτωση την οποία θα εξετάσουμε σε επόμενο παράδειγμα.

α.6: Από την επιλογή ***Descriptives*** παίρνουμε:

➢ Τον πίνακα 6.12 (***Descriptive Statistics***) όπου μάς δίνονται η μέση τιμή (***Mean***) και η τυπική απόκλιση (***Std.Deviation***) όλων των μεταβλητών.

Πίνακας 6.12: Descriptive Statistics

	Mean	Std. Deviation	N
Τιμή πώλησης (σε χιλ. €)	157,45	82,377	90
Εμβαδόν (σε m2)	227,67	103,060	90
Όροφος	1,93	,684	90
Πλήθος Μπάνιων	2,27	1,003	90
Παλαιότητα (σε έτη)	11,20	5,064	90

➢ Τον πίνακα 6.13 (***Correlation***) που μάς δίνει τους ***συντελεστές συσχέτισης*** μεταξύ όλων των μεταβλητών (εξαρτημένης-ανεξαρτήτων) καθώς επίσης και τα ***Sig***. των τεστ για όλους τους συντελεστές συσχέτισης που υπολογίστηκαν.

Πίνακας 6.13: Correlations

		Τιμή πώλησης (σε χιλ. €)	Εμβαδόν (σε m2)	Όροφος	Πλήθος Μπάνιων	Παλαιότητα (σε έτη)
Pearson Correlation	Τιμή πώλησης (σε χιλ.	1,000	,969	,751	,852	,321
	Εμβαδόν (σε m2)	,969	1,000	,800	,846	,415
	Όροφος	,751	,800	1,000	,714	,374
	Πλήθος Μπάνιων	,852	,846	,714	1,000	,109
	Παλαιότητα (σε έτη)	,321	,415	,374	,109	1,000
Sig. (1-tailed)	Τιμή πώλησης (σε χιλ.	.	,000	,000	,000	,001
	Εμβαδόν (σε m2)	,000	.	,000	,000	,000
	Όροφος	,000	,000	.	,000	,000
	Πλήθος Μπάνιων	,000	,000	,000	.	,154
	Παλαιότητα (σε έτη)	,001	,000	,000	,154	.
N	Τιμή πώλησης (σε χιλ.	90	90	90	90	90
	Εμβαδόν (σε m2)	90	90	90	90	90
	Όροφος	90	90	90	90	90
	Πλήθος Μπάνιων	90	90	90	90	90
	Παλαιότητα (σε έτη)	90	90	90	90	90

➢ Τον πίνακα ***Model Summary*** ο οποίος είναι ίδιος με τον πίνακα 6.10.

- Τον πίνακα ***ANOVA*** ο οποίος είναι ίδιος με τον πίνακα 6.11 και
- Τον πίνακα ***Coefficients*** ο οποίος είναι ίδιος με τον πίνακα 6.7.

α.7: Από την επιλογή ***Part and Partial Correlation*** παίρνουμε τον πίνακα 6.14 (***Coefficients***) στον οποίο έχουμε τη συσχέτιση κάθε ανεξάρτητης με την εξαρτημένη μεταβλητή ***(Zero-order)*** και τη συσχέτιση μεταξύ της ανεξάρτητης και της εξαρτημένης μεταβλητής όταν η γραμμική επίδραση από τις άλλες ανεξάρτητες μεταβλητές του μοντέλου έχουν εξαλειφθεί ***(Partial)***.

Πίνακας 6.14: Coefficients

Model		Correlations		
		Zero-order	Partial	Part
1	Εμβαδόν (σε m2)	,969	,869	,395
	Όροφος	,751	-,164	-,037
	Πλήθος Μπάνιων	,852	,112	,025
	Παλαιότητα (σε έτη)	,321	-,261	-,061

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

α.8: Από την επιλογή **Collinearity Diagnostic** παίρνουμε:

- Τον πίνακα 6.15 **(Coefficients)** με τους δείκτες **Tolerance** (ανεκτικότητα) και **VIF (Variance Inflation Factor)**, για την αξιολόγηση της συγγραμμικότητας- πολυσυγγραμμικότητας.

Πίνακας 6.15: Coefficients

Model		Collinearity Statistics	
		Tolerance	VIF
1	Εμβαδόν (σε m2)	,155	6,464
	Όροφος	,347	2,883
	Πλήθος Μπάνιων	,207	4,839
	Παλαιότητα (σε έτη)	,606	1,651

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

!!!! *Η τιμή του δείκτη Tolerance για την μεταβλητή Εμβαδόν είναι λίγο μεγαλύτερη από το αυστηρό όριο 5 και συνεπώς δεν είναι έντονο το φαινόμενο της συγγραμμικότητα..*

➢ Τον πίνακα 6.16 (***Collinearity Diagnostics***) με τον δείκτη ***Eigenvalue (ιδιοτιμή)***, του οποίου οι τιμές οι οποίες πλησιάζουν προς το 0 δείχνουν μεγάλη ***διασυσχέτιση (intercorrelation)***. Στον ίδιο πίνακα έχουμε το δείκτη ***Condition Index***, ο οποίος όταν πάρει τιμές μεγαλύτερες του 15 υπάρχει **πιθανό** πρόβλημα, ενώ για τιμές μεγαλύτερες του 30 υπάρχει **σοβαρό** πρόβλημα ***συγγραμμικότητας***.

Πίνακας 6.16: Collinearity Diagnostics

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions				
				(Constant)	Εμβαδόν (σε m2)	Όροφος	Πλήθος Μπάνιων	Παλαιότητα (σε έτη)
1	1	4,701	1,000	,00	,00	,00	,00	,00
	2	,168	5,287	,03	,01	,00	,06	,34
	3	,085	7,456	,62	,05	,00	,00	,21
	4	,031	12,332	,06	,02	,93	,19	,09
	5	,015	17,693	,29	,92	,07	,75	,36

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

b. Περιοχή Residuals

b.1 Από την επιλογή **Durbin-Watson** προκύπτει:

➢ Ο πίνακας 6.17 (**Model Summary**) στον οποίο έχουμε κατά σειρά το συντελεστή πολλαπλής συσχέτισης (R), το δείκτη προσδιορισμού (R Square), το διορθωμένο δείκτη προσδιορισμού (Adjusted R Square), το τυπικό σφάλμα της εκτίμησης (Std. Error of the Estimate) και τέλος το δείκτη **Durbin- Watson** για την αξιολόγηση της **αυτοσυσχέτισης (autocorrelation)**.

Πίνακας 6.17: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,974[a]	,949	,947	18,992	1,091

a. Predictors: (Constant), Παλαιότητα (σε έτη), Πλήθος Μπάνιων , Όροφος, Εμβαδόν (σε m2)

b. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

!!!! *Ο δείκτης* ***Durbin-Watson*** *(συντελεστής αυτοσυσχέτισης – autocorrelation) με τιμή 1,091 δεν είναι ικανοποιητικός.*

Παίρνουμε επίσης τους πίνακες:

- ***ANOVA*** όπως ο πίνακας 6.11.
- ***Coefficients*** όπως ο πίνακας 6.7 και
- ***Residuals Statistics*** (Πίνακας 6.18) στον οποίο δίνονται ***οι προβλεπόμενες τιμές (predicted values), τα σφάλματα (Residuals), η τυπική απόκλιση των προβλεπόμενων τιμών (Std. Predicted values) και η τυπική απόκλιση των σφαλμάτων (Std. Residuals).***

Πίνακας 6.18: Residual Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	58,60	293,44	157,45	80,259	90
Residual	-34,762	41,237	,000	18,560	90
Std. Predicted Valu	-1,232	1,694	,000	1,000	90
Std. Residual	-1,830	2,171	,000	,977	90

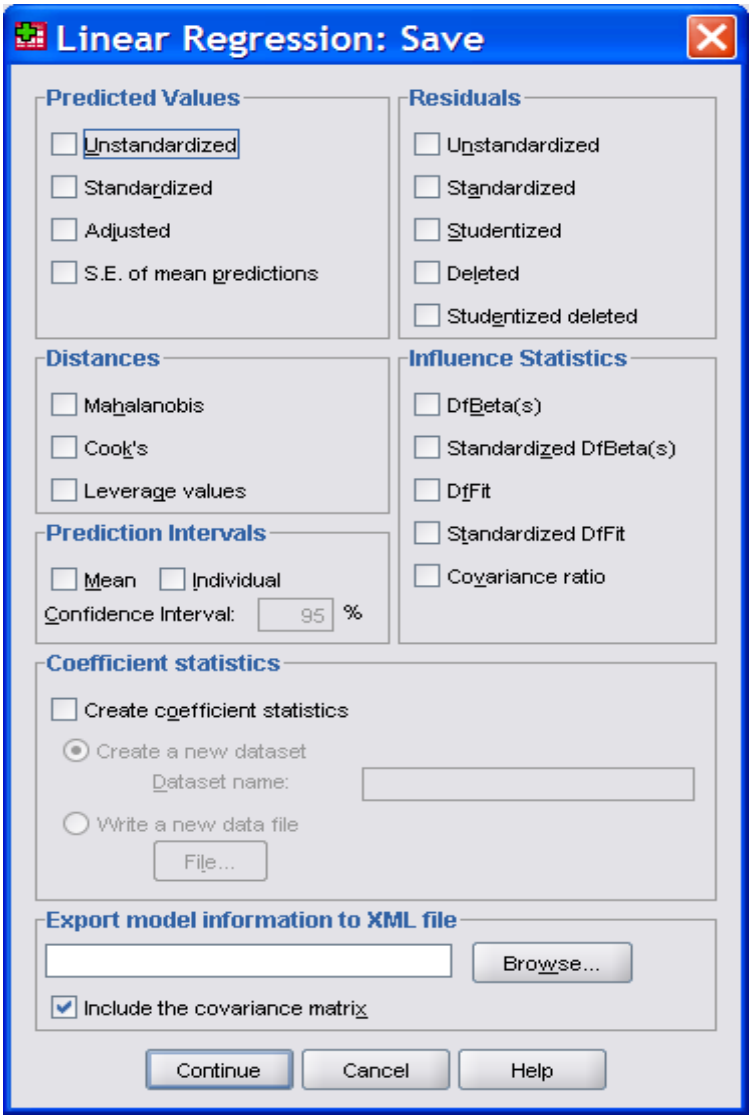
a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

b.2: Από την επιλογή **Casewise Diagnostic** οι πίνακες που προκύπτουν είναι κατά σειρά:

- ***Model Summary***, όπως ο πίνακας 6.10
- ***ANOVA***, όπως ο πίνακας 6.11
- ***Coefficients***, όπως ο πίνακας 6.7 και
- ***Residual Statistics***, όπως ο πίνακας 6.18.

!!!! *Μετά από κάθε επιλογή στην εικόνα 6.6, προκειμένου να πάρουμε τους πίνακες που είδαμε, πατάμε* ***continue*** *και αφού επιστρέψουμε στην εικόνα 6.4 επιλέγουμε O.K.*

Μετά τον υπολογισμό των διαφόρων μέτρων από τη φόρμα ***Statistics*** της εικόνας 6.6, πατάμε στο κουμπί ***Save*** και εμφανίζεται η εικόνα 6.7.



Εικόνα 6.7

Θα δούμε αναλυτικά τα αποτελέσματα των επιλογών μας από την παραπάνω φόρμα.

α. Περιοχή *Predicted Value.*

α.1 Από την επιλογή ***Unstandardized και (ή) Standartized*** έχουμε κατά σειρά τους πίνακες:

- ***Model Summary***, όπως ο πίνακας 6.10
- ***ANOVA***, όπως ο πίνακας 6.11
- ***Coefficients***, όπως ο πίνακας 6.7

- ***Residuals Statistics***, όπως ο πίνακας 6.18 και
- Στο ***Data Editor***, τη στήλη ***Pre_1 (Unstandardized Precicted Value)*** με τις προβλεπόμενες τιμές και (ή) τη στήλη ***Zpr_1 (Standardized Precicted Value)*** με τις **z**-τιμές των προηγουμένων.

α.2 Από την επιλογή **Adjusted και (ή) S.E of Mean predictions** εμφανίζονται οι πίνακες 6.7, 6.10, 6.11 με τα γνωστά αποτελέσματα και

- Ο πίνακας 6.19 (***Residuals Statistics***) με τις ελάχιστες, τις μέγιστες, τις μέσες τιμές και τις τυπικές αποκλίσεις πολλών χρήσιμων δεικτών.
- Στο ***Data Editor***, στήλη ***adj_1 (Adjusted Precicted Value)*** με τις διορθωμένες τιμές και η στήλη ***sep_1 (S.E_of predicted value)*** με τα τυπικά σφάλματα των εκτιμήσεων.

Πίνακας 6.19: Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	58,60	293,44	157,45	80,259	90
Std. Predicted Value	-1,232	1,694	,000	1,000	90
Standard Error of Predicted Value	2,270	6,036	4,341	1,100	90
Adjusted Predicted Value	58,60	293,00	157,37	80,369	90
Residual	-34,762	41,237	,000	18,560	90
Std. Residual	-1,830	2,171	,000	,977	90
Stud. Residual	-1,865	2,242	,002	1,006	90
Deleted Residual	-36,084	43,981	,084	19,675	90
Stud. Deleted Residual	-1,893	2,298	,004	1,018	90
Mahal. Distance	,282	8,002	3,956	2,244	90
Cook's Distance	,000	,067	,012	,018	90
Centered Leverage Value	,003	,090	,044	,025	90

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

b. Περιοχή *Distances*

b.1. Από την επιλογή **Mahalanobis και (ή) Cook's και (ή) Leverage value**: εμφανίζονται οι πίνακες 6.7, 6.10, 6.11 και 6.19 και

- Στο ***Data Editor*** η στήλη ***mah_1 (Mahalanobis distance)***, η στήλη ***coo_1 (Cook's distance)*** και η στήλη ***lev_1 (Centered Leverage value)***.

c. Περιοχή *Residuals*

c.1 Από την επιλογή **Unstandartized και (ή) Standartized**: εμφανίζονται οι πίνακες 6.7, 6.9, 6.11 και 6.18 και

- Στο ***Data Editor*** η στήλη ***res_1 (Unstantardized Residual)*** και (ή) στήλη ***zre_1*** (***Standartized Residual)***.

c.2. Από την επιλογή **Studentized και (ή) Deleted και (ή) Studentized deleted** εμφανίζονται οι πίνακες 6.7, 6.10, 6.11 και 6.18 και

- Στο ***Data Editor*** η στήλη ***sre_1 (Studentized Residuals)*** και (ή) στήλη ***dre_1*** (***Deleted Residual)*** και (ή) στήλη ***sdr_1 (Studentized deleted Residuals).***

d. Περιοχή *Influence Statistics*

d.1 Από την επιλογή **DfBeta(s)** και (ή) **Standardized DfBeta(s)** και (ή) **DfFit** και (ή) **Standardized DfFit** και (ή) **Covariance ratio** εμφανίζονται οι πίνακες 6.7, 6.10, 6.11 και 6.19

- Στο ***Data Editor*** οι στήλες ***bfb0_1 (DFBETA Intercept), bfb1_1*** (***DFBETA X1), bfb2_1 (DFBETA X2), bfb3_1 (DFBETA X3)*** και ***bfb4_1 (DFBETA X4).***
- Οι στήλες ***sdb0_1 (Standartized DFBETA Intercept), sdb1_1*** (***Standartized DFBETA X1), sdb2_1 (Standartized DFBETA X2), sdb3_1 (Standartized DFBETA X3)*** και ***sdb4_1 (Standartized DFBETA X4).***
- Η στήλη ***dff_1 (DFFIT)***

- Η στήλη ***sdf_1 (Standartize DFFIT)*** και
- Η στήλη ***cov_1 (COVRATIO)***

e. Περιοχή Prediction Intervals

e.1. Από την επιλογή **Mean** και (ή) **Individual** εμφανίζονται οι πίνακες 6.5, 6.8, 6.9 και 6.17 και

- Στο ***Data Editor*** οι στήλες ***lmci_1 (95% L CI for Y mean)*** και ***uici_1 (95% U CI for Y mean)*** και

- Οι στήλες ***lici_1 (95% L CI for Y Individual)*** και ***uici_1 (95% U CI for Y Individual)***.

!!! Μετά από κάθε επιλογή στην εικόνα 6.7, προκειμένου να πάρουμε τους πίνακες που είδαμε, πατάμε ***continue*** *και αφού επιστρέψουμε στην εικόνα 6.4 επιλέγουμε O.K.*

!!! *Οι δυνατότητες της επιλογής* ***Option*** *από τη φόρμα 6.4 αναλύθηκαν στο προηγούμενο παράδειγμα*

!!! *Η επιλογή* ***Plots*** *της φόρμας 6.4 μας δίνει τη δυνατότητα να δημιουργήσουμε γραφικές παραστάσεις*

6.3.2 Μέθοδος Stepwise

Το σύνηθες πρόβλημα στην πολλαπλή γραμμική παλινδρόμηση είναι η επιλογή των ανεξάρτητων μεταβλητών X οι οποίες συνεισφέρουν ουσιαστικά στην ερμηνεία της διακύμανσης των τιμών της εξαρτημένης μεταβλητής Y.

Η μέθοδος της **διαδοχικής επιλογής** των ανεξάρτητων μεταβλητών (**Stepwise regression**)**,** την οποία προσφέρει το S.P.S.S, είναι η κατάλληλη για τον εντοπισμό των πλέον σημαντικών

ανεξάρτητων μεταβλητών. Η μέθοδος αυτή, συνήθως χρησιμοποιείται όταν το μοντέλο είναι πειραματικό και δε γνωρίζουμε τη συνεισφορά της κάθε μεταβλητής στην ερμηνευτικότητά του. Με τη μέθοδο της **διαδοχικής επιλογής** δημιουργούνται όλα τα εναλλακτικά μοντέλα τα οποία είναι στατιστικά σημαντικά και έχουμε την δυνατότητα να επιλέξουμε αυτό που ικανοποιεί τα κριτήριά μας.

Για την ανάπτυξη της μεθόδου **Stepwise** θα χρησιμοποιηθεί το δεύτερο παράδειγμα της προηγούμενης παραγράφου.

Όπως σε κάθε περίπτωση:

- Από το μενού ***Analyze*** επιλέγουμε ***Regression,*** στη συνέχεια ***Linear*** και
- Μεταφέρουμε στη θέση ***Dependent*** τη μεταβλητή **Τιμή πώλησης**
- Στη θέση ***Independents*** τις μεταβλητές **Εμβαδόν, Όροφος, Πλήθος μπάνιων** και ***παλαιότητα.***
- Στο παράθυρο ***Method*** πατάμε ***Stepwise.***
- Στη συνέχεια ***O.K*** και εμφανίζονται οι επόμενοι πίνακες.

Ο πίνακας ***Model Summary*** δίνει τις τιμές **R, R Square, Adjusted R Square** και **Std. Error of the Estimate** για τα δύο προτεινόμενα μοντέλα. Παρατηρούμε λοιπόν ότι, το πρώτο μοντέλο αποτελείται από μία μόνο ανεξάρτητη μεταβλητή (Εμβαδόν) και ο δείκτης προσδιορισμού είναι 0,940. Το δεύτερο προτεινόμενο μοντέλο αποτελείται από δύο ανεξάρτητες μεταβλητές (Εμβαδόν και Παλαιότητα) και ο αντίστοιχος δείκτης προσδιορισμού είναι 0,947.

Δηλαδή, η είσοδος της δεύτερης μεταβλητής (Παλαιότητα) βελτίωσε τον δείκτη προσδιορισμού κατά 0,7% μόνο.

Από τον πίνακα αυτό αντιλαμβανόμαστε ότι, οι δύο άλλες μεταβλητές (Πλήθος μπάνιων, Όροφος) είναι περιττό να εισέλθουν στο μοντέλο καθώς η συνεισφορά τους είναι μηδαμινή ενώ συγχρόνως αυτό γίνεται πιο σύνθετο.

Πίνακας 6.20: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,969[a]	,940	,939	20,369
2	,973[b]	,947	,946	19,094

a. Predictors: (Constant), Εμβαδόν (σε m2)

b. Predictors: (Constant), Εμβαδόν (σε m2), Παλαιότητα (σε έτη)

Ο πίνακας ***ANOVA*** μας δίνει τις τιμές **F** και **Sig.** οι οποίες πιστοποιούν τη σημαντικότητα και των δύο προτεινόμενων μοντέλων.

Πίνακας 6.21 : ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	567439,9	1	567439,937	1367,715	,000[a]
	Residual	36509,599	88	414,882		
	Total	603949,5	89			
2	Regression	572230,8	2	286115,406	784,774	,000[b]
	Residual	31718,723	87	364,583		
	Total	603949,5	89			

a. Predictors: (Constant), Εμβαδόν (σε m2)

b. Predictors: (Constant), Εμβαδόν (σε m2), Παλαιότητα (σε έτη)

c. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

Ο πίνακας ***Coefficients*** δίνει το σταθερό όρο, το συντελεστή παλινδρόμησης τις τιμές Beta, t και Sig. των προτεινόμενων, με τη μέθοδο ***Stepwise***, μοντέλων. Έτσι θα έχουμε:

Τιμή πώλησης = -18,935+0,775* Εμβαδόν για το πρώτο μοντέλο και ***Τιμή πώλησης=-8,495+0,807*Εμβαδόν-1,59*Παλαιότητα*** για το δεύτερο μοντέλο.

Παρατηρούμε ότι οι συντελεστές παλινδρόμησης και στα δύο μοντέλα είναι στατιστικά σημαντικοί.

!!! *Αξίζει να αναφερθεί ότι οι μεταβλητές που προτείνεται να μπούνε στο μοντέλο είναι αυτές που στην πλήρη ανάπτυξή του, με τη μέθοδο* ***Enter****, έδιναν συντελεστές παλινδρόμησης στατιστικά σημαντικούς.*

Πίνακας 6.22: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-18,935	5,231		-3,620	,000
	Εμβαδόν (σε m2)	,775	,021	,969	36,983	,000
2	(Constant)	-8,495	5,686		-1,494	,139
	Εμβαδόν (σε m2)	,807	,022	1,010	37,396	,000
	Παλαιότητα (σε έτη	-1,593	,439	-,098	-3,625	,000

a. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

Ο πίνακας ***Excluded Variables*** περιέχει τις ανεξάρτητες μεταβλητές τις οποίες η μέθοδος ***Stepwise*** προτείνει **να μην συμπεριλάβουμε** στο $1^{ο}$ και στο $2^{ο}$ μοντέλο.

Πίνακας 6.23: Excluded Variables

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Όροφος	-,067[a]	-1,539	,127	-,163	,361
	Πλήθος Μπάνιων	,114[a]	2,385	,019	,248	,285
	Παλαιότητα (σε έτη)	-,098[a]	-3,625	,000	-,362	,828
2	Όροφος	-,056[b]	-1,362	,177	-,145	,359
	Πλήθος Μπάνιων	,041[b]	,770	,443	,083	,214

a. Predictors in the Model: (Constant), Εμβαδόν (σε m2)

b. Predictors in the Model: (Constant), Εμβαδόν (σε m2), Παλαιότητα (σε έτη)

c. Dependent Variable: Τιμή πώλησης (σε χιλ. €)

!!! *Όλες οι επιλογές που μπορούμε να κάνουμε από την εικόνα 6.4 και τις φόρμες που προκύπτουν από αυτήν, έχουν αναπτυχθεί στη μέθοδο* ***Enter*** *(παράδειγμα 2ο), είναι ίδιες και στη μέθοδο* ***Stepwise*** *και δεν διαφέρει η ερμηνεία των αποτελεσμάτων τα οποία προκύπτουν από αυτές.*

!!! *Επιλέγοντας* ***Plots*** *από το κουμπί της εικόνας 6.4 μπορούμε να πάρουμε γραφικές παραστάσεις σχετικές με το μοντέλο της παλινδρόμησης.*

6.4 Ιεραρχική Παλινδρόμηση -Hierarchical Regression

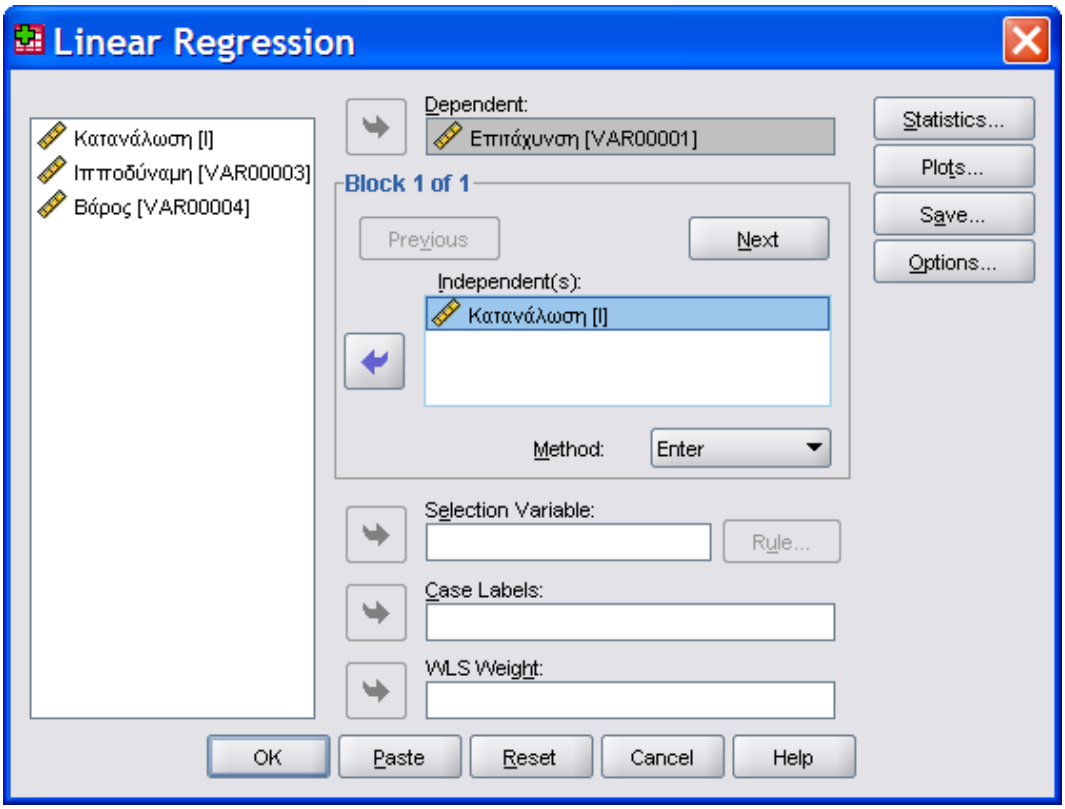
Στην Ιεραρχική πολλαπλή παλινδρόμηση ο ερευνητής καθορίζει όχι μόνο πόσες ανεξάρτητες μεταβλητές θα βάλει στο μοντέλο αλλά και την σειρά με την οποία θα τις βάλει. Συνήθως η σειρά εισαγωγής βασίζεται σε λογικές ή θεωρητικές εκτιμήσεις.

Με τους συντελεστές ***μερικού προσδιορισμού*** μπορούμε να εκτιμήσουμε το ποσοστό της ανερμήνευτης διακύμανσης, από προηγούμενες ανεξάρτητες μεταβλητές, το οποίο θα ερμηνευτεί αν προστεθεί στο υπόδειγμα μία νέα ανεξάρτητη μεταβλητή. Έτσι μπορούμε να επιλέξουμε από μια ομάδα υποψηφίων ανεξάρτητων μεταβλητών αυτές που πραγματικά συνεισφέρουν στη βελτίωση της ερμηνευτικότητας του μοντέλου. Στην πράξη αναζητούμε τις μεταβλητές εκείνες οι οποίες βελτιώνουν σημαντικά την τιμή του δείκτη προσδιορισμού (R^2).

Παράδειγμα: Μετρήσαμε την «***κατανάλωση***» σε λίτρα ανά 100 Km, το «***βάρος***» σε κιλά, την «***ιπποδύναμη***» σε άλογα και την «***επιτάχυνση***» σε δευτερόλεπτα από 0 έως 100 Km/h, 150 αυτοκινήτων διαφόρων τύπων. Θέλουμε να δημιουργήσουμε ένα μοντέλο παλινδρόμησης με το οποίο θα μπορούμε να προβλέπουμε την «*επιτάχυνση*» των αυτοκινήτων στηριζόμενοι στην «*κατανάλωση*», το «*βάρος*» και την «*ιπποδύναμη*» αυτών. Μας ενδιαφέρει επίσης να μελετήσουμε την προσφορά της κάθε ανεξάρτητης μεταβλητής στην ερμηνεία της διακύμανσης της εξαρτημένης μεταβλητής.

Για την πραγματοποίηση των παραπάνω θα πραγματοποιήσουμε **Ιεραρχική Παλινδρόμηση**, με τη χρήση του SPSS, ακολουθώντας την επόμενη διαδικασία:

- Από το μενού ***Analyze*** επιλέγουμε ***Regression,*** στη συνέχεια ***Linear*** και εμφανίζεται η επόμενη εικόνα..

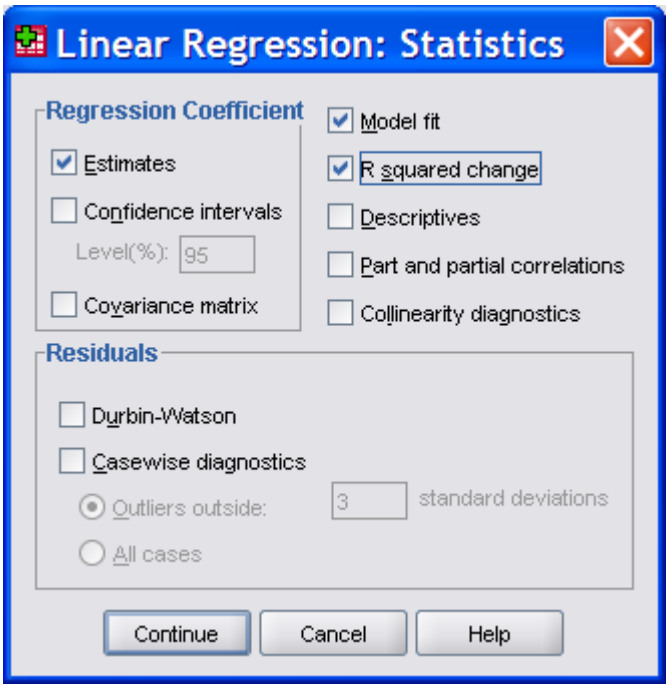


Εικόνα 6.10

- Στο παράθυρο ***Dependent*** βάζουμε τη μεταβλητή «*επιτάχυνση*» και στο παράθυρο ***Independent***, πάνω από το οποίο είναι ενεργοποιημένη η ένδειξη ***Block 1 of 1***, την μεταβλητή η οποία θεωρούμε ότι έχει την μεγαλύτερη συνεισφορά στην ερμηνευτικότητα του μοντέλου. Στη συγκεκριμένη περίπτωση την μεταβλητή «*κατανάλωση*». Στη συνέχεια πατάμε στο κουμπί **Next**, ενεργοποιείται η ένδειξη ***Block 2 of 2***, εισάγουμε τη μεταβλητή «*ιπποδύναμη*» και

τέλος πατάμε πάλι στο κουμπί ***Next***, ενεργοποιείται η ένδειξη ***Block 3 of 3*** και εισάγουμε τη μεταβλητή «*βάρος*».

- Κατόπιν πατάμε στο κουμπί ***Statistics***



Εικόνα 6.9

- Τσεκάρουμε τις ενδείξεις ***Estimates***, ***Model Fit*** και ***R Squared Change***

- Επιλέγουμε ***Continue*** επιστρέφουμε στην προηγούμενη εικόνα και

- ***O.K*** για να έχουμε τα αποτελέσματα τα οποία ζητάμε.

Στον πίνακα ***Model Summary*** το **μοντέλο 1 (Model 1)** το οποίο περιέχει μόνο τη μεταβλητή «*κατανάλωση*» δίνει **R-square=0,394 (Δείκτης Προσδιορισμού)**. Δηλαδή η μεταβλητή «*κατανάλωση*», μόνη της, ερμηνεύει το 39,4% των μεταβολών της εξαρτημένης μεταβλητής «επιτάχυνση».

Το μοντέλο 2 το οποίο περιέχει τις μεταβλητές «κατανάλωση» και «ιπποδύναμη» δίνει **R-square=0,628**. Επομένως οι 2 αυτές μεταβλητές

ερμηνεύουν το 62,8% των μεταβολών της εξαρτημένης μεταβλητής. Έχουμε δηλαδή μια αύξηση της τιμής του **R- square** ίση με 62,8%-39,4%=23,4%. Γίνεται φανερό ότι η μεταβλητή «ιπποδύναμη» η οποία μπήκε στο μοντέλο 2 ερμηνεύει το 23,4% των μεταβολών της εξαρτημένης μεταβλητής, μετά την αφαίρεση των επιδράσεων της μεταβλητής «κατανάλωση».

Στο σημείο αυτό αξίζει να αναλυθεί ο όρος **«δείκτης μερικού προσδιορισμού»** ο οποίος αναφέρθηκε στην αρχή της παραγράφου. Στο 1^{ο} μοντέλο, η μεταβλητή «κ*ατανάλωση*» ερμήνευσε μόνη της το 39,4% των μεταβολών της εξαρτημένης μεταβλητής «επιτάχυνση». Έμεινε δηλαδή ανερμήνευτο το 100%- 39,4% = 60,6% του συνόλου των μεταβολών. Η μεταβλητή η οποία εισήλθε στο 2^{ο} μοντέλο βελτίωσε τη συνολική τιμή του δείκτη προσδιορισμού κατά 23,4%. Αυτό σημαίνει ότι ο δείκτης μερικού προσδιορισμού της μεταβλητής «ιπποδύναμη» είναι: 60,6/23,4 = 38,6%. Με απλά λόγια η μεταβλητή «ιπποδύναμη» ερμήνευσε το 38,6% του υπολοίπου ανερμήνευτου από την μεταβλητή «κ*ατανάλωση*». Αν αντίθετα γνωρίζουμε τον δείκτη μερικού προσδιορισμού μιας μεταβλητής και τον δείκτη προσδιορισμού του μοντέλου χωρίς αυτή τη μεταβλητή, τότε μπορούμε να υπολογίσουμε τον δείκτη προσδιορισμού του μοντέλου και με τις δύο μεταβλητές. Στη συγκεκριμένη περίπτωση γνωρίζαμε ότι ο δείκτης προσδιορισμού με τη μεταβλητή «κ*ατανάλωση*» ήταν 39,4%. Υπήρχε δηλαδή ένα ποσοστό ανερμήνευτων μεταβολών 60,6% . Με δεδομένο 38,6%, δείκτη μερικού προσδιορισμού για την μεταβλητή «ιπποδύναμη», θα έχουμε δείκτη προσδιορισμού:

60,6% (υπόλοιπο ανερμήνευτο από την μεταβλητή κατανάλωση)*38,6%(δείκτης μερικού προσδιορισμού μεταβλητής

ιπποδύναμη)+39,4% (δείκτης προσδιορισμού 1ου μοντέλου) = 62,8% (δείκτης προσδιορισμού 2ου μοντέλου).

Τέλος το μοντέλο 3 **(Model 3)** το οποίο περιέχει και την μεταβλητή «βάρος», δίνει **R-square =0,670.** Συνολικά λοιπόν οι τρεις εξαρτημένες μεταβλητές ερμηνεύουν το 67% του συνόλου των μεταβολών της εξαρτημένης μεταβλητής. Εδώ έχουμε επίσης μια αύξηση της τιμής του **R- square** ίση με 67%-62,8%=4,2% Συνεπώς η μεταβλητή «βάρος» η οποία μπήκε τελευταία στο μοντέλο ερμηνεύει μόνο το 4,2% των μεταβολών της εξαρτημένης μεταβλητής, μετά την αφαίρεση των επιδράσεων των μεταβλητών «κατανάλωση»και «ιπποδύναμη».

Συμπερασματικά μπορούμε να πούμε ότι το 67% του συνόλου των μεταβολών της εξαρτημένης μεταβλητής «επιτάχυνση» ερμηνεύεται από την μεταβλητή «κατανάλωση» (39,4%), τη μεταβλητή «ιπποδύναμη» (23,4%) και τη μεταβλητή «βάρος» (4,2%).

Στον παραπάνω πίνακα και στη στήλη **R- Square Change** βλέπουμε τις μεταβολές του δείκτη **R- Square**, από μοντέλο σε μοντέλο και στη στήλη **F- Change** τις αλλαγές στην τιμή του στατιστικού F. Επίσης, η στήλη **Sig. F Change** μας ενημερώνει για την στατιστική σημαντικότητα των τιμών της στήλης **F- Change**. Σε κάθε περίπτωση που η τιμή της στήλης αυτής είναι μικρότερη του 0,05 η τιμή του **R- Square** είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%. Στο συγκεκριμένο παράδειγμα όλες οι τιμές της στήλης αυτής είναι 0 και συνεπώς όλες οι τιμές του δείκτη **R- Square** είναι στατιστικά σημαντικές.

Ο **Διορθωμένος Δείκτης Προσδιορισμού (Adjusted R- square)** επηρεάζεται από το μέγεθος του δείγματος και το πλήθος των

ανεξάρτητων μεταβλητών. Στο συγκεκριμένο παράδειγμα έχουμε μεγάλο δείγμα με σχετικά λίγες ανεξάρτητες μεταβλητές και συνεπώς η διαφορά μεταξύ δείκτη προσδιορισμού και διορθωμένου δείκτη προσδιορισμού είναι πολύ μικρή.

Πίνακας 6.24: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	,628[a]	,394	,390	2,314	,394	90,457	1	139	,000
2	,793[b]	,628	,623	1,819	,234	86,954	1	138	,000
3	,818[c]	,670	,663	1,721	,041	17,194	1	137	,000

a. Predictors: (Constant), Κατανάλωση

b. Predictors: (Constant), Κατανάλωση, Ιπποδύναμη

c. Predictors: (Constant), Κατανάλωση, Ιπποδύναμη, Βάρος

Στον πίνακα **ANOVA** έχουμε πάλι τα τρία μοντέλα και την τιμή του στατιστικού **F** για κάθε μοντέλο. Η στήλη **Sig.** μας επιτρέπει να αξιολογήσουμε την τιμή του κάθε F ως προς την στατιστική σημαντικότητά του.

Παρατηρούμε ότι και οι τρεις τιμές F είναι στατιστικά σημαντικές (Sig<0,05) και συνεπώς υπάρχουν ικανοποιητικές συσχετίσεις μεταξύ των μεταβλητών του κάθε μοντέλου.

Από τον πίνακα αυτό μπορούμε, επίσης, να υπολογίσουμε άμεσα τις τιμές του **R- square** του κάθε μοντέλου διαιρώντας την τιμή **Regression** της στήλης **Sum of Squares** με την τιμή **Total** της ίδιας στήλης. Έτσι θα έχουμε:

Για το πρώτο μοντέλο **R- square**=484,289/1228,47=0,349

Για το δεύτερο μοντέλο **R- square**=771,946/1228,47=0,628

Για το τρίτο μοντέλο **R- square**=822,854/1228,47=0,670

Πίνακας 6.25: ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	484,289	1	484,289	90,457	,000[a]
	Residual	744,183	139	5,354		
	Total	1228,47	140			
2	Regression	771,946	2	385,973	116,673	,000[b]
	Residual	456,525	138	3,308		
	Total	1228,47	140			
3	Regression	822,854	3	274,285	92,641	,000[c]
	Residual	405,618	137	2,961		
	Total	1228,47	140			

a. Predictors: (Constant), Κατανάλωση

b. Predictors: (Constant), Κατανάλωση, Ιπποδύναμη

c. Predictors: (Constant), Κατανάλωση, Ιπποδύναμη, Βάρος

d. Dependent Variable: Επιτάχυνση

Στον πίνακα **Coefficients** και στη στήλη **Unstandardized Coefficients B** περιέχονται οι μη τυποποιημένοι συντελεστές των μεταβλητών και ο σταθερός όρος του κάθε μοντέλου. Επίσης στη στήλη **Standardized Coefficients Beta** υπάρχουν οι τυποποιημένοι και συνεπώς άμεσα συγκρίσιμοι συντελεστές των μεταβλητών των τριών μοντέλων. Υπάρχει επίσης η στήλη των τιμών **t** και η στήλη **Sig.** η οποία ελέγχει την σημαντικότητα των τιμών **t**. Έτσι σε κάθε περίπτωση που η τιμή **Sig.** είναι μικρότερη του 0,05 η αντίστοιχη τιμή του συντελεστή είναι στατιστικά σημαντική (διάφορη του 0).

Στο μοντέλο 1 το sig. του συντελεστή της μεταβλητής «κατανάλωση» είναι 0 και συνεπώς ο συντελεστής είναι στατιστικά σημαντικός.

Στο μοντέλο 2 το sig. του συντελεστή της μεταβλητής «κατανάλωση» είναι 0,527>0,05 και συνεπώς είναι στατιστικά

ασήμαντος. Αντίθετα το sig. του συντελεστή της μεταβλητής «ιπποδύναμη» είναι 0,000 και είναι στατιστικά σημαντικός.

Τέλος στο μοντέλο 3 το sig. του συντελεστή της μεταβλητής «κατανάλωση» είναι 0,303>0,05 και συνεπώς είναι στατιστικά ασήμαντος. Αντίθετα τα sig. των συντελεστών των μεταβλητών «ιπποδύναμη» και «βάρος» είναι 0,000 υποδηλώνοντας τη στατιστική σημαντικότητά τους.

Πίνακας 6.26: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	20,807	,690		30,167	,000
	Κατανάλωση	-,498	,052	-,628	-9,511	,000
2	(Constant)	20,654	,542		38,078	,000
	Κατανάλωση	,045	,071	,057	,635	,527
	Ιπποδύναμη	-,054	,006	-,839	-9,325	,000
3	(Constant)	19,917	,543		36,674	,000
	Κατανάλωση	-,076	,073	-,096	-1,034	,303
	Ιπποδύναμη	-,070	,007	-1,087	-10,445	,000
	Βάρος	,003	,001	,435	4,147	,000

a. Dependent Variable: Επιτάχυνση

!!!! Αν ο ερευνητής ιεραρχήσει με διαφορετικό τρόπο τη σημαντικότητα των μεταβλητών και επομένως τις εισάγει στο μοντέλο με διαφορετική σειρά, τα ενδιάμεσα αποτελέσματα θα αλλάξουν, το τελικό όμως μοντέλο με όλες τις μεταβλητές θα δίνει τον ίδιο δείκτη προσδιορισμού, όπως και προηγουμένως.

Κεφάλαιο 7
Διερευνητική Παραγοντική Ανάλυση

Chapter 7
Exploratory Factor Analysis

7. Εισαγωγή

Με τον όρο **Παραγοντική Ανάλυση (Factor Analysis)** εννοούμε μία **πολυμεταβλητή (multivariate)** στατιστική μέθοδο αλληλεξάρτησης, της οποίας πρωταρχικός σκοπός είναι να προσδιορίσει τη δομή ενός πίνακα δεδομένων. Διακρίνεται σε

Διερευνητική (Exploratory και **Επικυρωτική ή Επιβεβαιωτική (Confirmatory Factor Analysis) Παραγοντική Ανάλυση**.

Η **Διερευνητική Παραγοντική Ανάλυση** σχετίζεται με την διερεύνηση του τύπου της σχέσης μεταξύ ενός πλήθους μεταβλητών. Στην ουσία η δομή του παραγοντικού μοντέλου ή της υφιστάμενης θεωρίας είναι άγνωστη. Έτσι, η διερευνητική παραγοντική ανάλυση μπορεί να θεωρηθεί ως μια τεχνική η οποία βοηθάει στη δημιουργία θεωρίας. Χρησιμοποιείται όταν δημιουργούμε ένα ερωτηματολόγιο, βασισμένοι στη θεωρία ή και την εμπειρία, με το οποίο θέλουμε να μετρήσουμε κάποιες έννοιες (μεταβλητές ή παράγοντες). Συνήθως οι έννοιες τις οποίες θέλουμε να μετρήσουμε αποτελούνται από πολλές προσδιοριστικές μεταβλητές και η βοήθεια της παραγοντικής ανάλυσης στην δομική αξιοπιστία των εννοιών είναι πολύ σημαντική.

Από την άλλη πλευρά, στην **Επικυρωτική ή Επιβεβαιωτική Παραγοντική Ανάλυση** (Κεφάλαιο 11), η ακριβής δομή του παραγοντικού μοντέλου η οποία βασίζεται σε υφιστάμενη θεωρία θεωρείται δεδομένη. Βοηθάει, στην ουσία, στην επικύρωση των προσδιοριστικών μεταβλητών οι οποίες χρησιμοποιούνται για την μέτρηση συγκεκριμένων δομών (παραγόντων).

7.1 Χρησιμότητα της Διερευνητικής Παραγοντικής Ανάλυσης

Η παραγοντική ανάλυση είναι χρήσιμη στην ανάλυση δεδομένων γιατί:

- Μελετά τη συσχέτιση μεταξύ μεγάλου αριθμού αλληλοσυνδεόμενων μεταβλητών δια μέσου της ομαδοποίησης αυτών σε **παράγοντες (factor)**. Μετά την ομαδοποίηση οι μεταβλητές του

κάθε παράγοντα είναι μεταξύ τους περισσότερο συσχετισμένες σε σχέση με μεταβλητές που ανήκουν σε άλλους παράγοντες. Αντίθετα οι παράγοντες έχουν πολύ μικρό και μηδενικό συντελεστή συσχέτισης.

- Ερμηνεύει κάθε παράγοντα σύμφωνα με τη σημασία των μεταβλητών. Για παράδειγμα 5 ερωτήσεις οι οποίες ανήκουν στον ίδιο παράγοντα μπορεί να μετράνε την αξιοπιστία ενός συστήματος αξιολόγησης εργαζομένων.
- Συγκεντρώνει πολλές μεταβλητές δημιουργώντας λίγους παράγοντες για κάθε έναν από τους οποίους υπολογίζει το σκορ το οποίο σαν νέα μεταβλητή μπορεί να χρησιμοποιηθεί για ***t-τεστ, παλινδρόμηση, ανάλυση διακύμανσης και πολλά άλλα***.

7.2 Προϋποθέσεις χρησιμοποίησης της παραγοντικής ανάλυσης

Οι βασικές προϋποθέσεις για την χρησιμοποίηση της διερευνητικής παραγοντικής ανάλυσης είναι οι εξής:

- Οι μεταβλητές πρέπει να είναι **ποσοτικές** σε οποιαδήποτε κλίμακα μέτρησης (ratio or interval). Μπορεί επίσης να είναι μεταβλητές οι οποίες εκφράζουν το βαθμό ικανοποίησης ή επιθυμίας αρκεί να υπάρχει μία αριθμητική κλίμακα όπου οι χαμηλές τιμές εκφράζουν μικρό βαθμό ικανοποίησης ή επιθυμίας και οι υψηλές τιμές μεγάλο βαθμό ικανοποίησης ή επιθυμίας (μπορεί να συμβαίνει και το αντίθετο). Για παράδειγμα, στο ερώτημα πόσο ικανοποιημένοι είστε από την εξωτερική πολιτική της κυβέρνησης μπορούμε να δώσουμε μία κλίμακα από το 1 έως το 7, όπου το 1 σημαίνει καθόλου ικανοποιημένοι ενώ το 7 δηλώνει απόλυτα ικανοποιημένοι. Σε πολλές περιπτώσεις οι μεταβλητές μπορεί να είναι και **δυαδικές (dummy**

variables). Αν όλες οι μεταβλητές είναι δυαδικές τότε η ***Boolean*** παραγοντική ανάλυση είναι η πλέον κατάλληλη μέθοδος.

- Το μέγεθος του δείγματος να μην είναι μικρότερο των 50 ατόμων και κατά προτίμηση να είναι μεγαλύτερο των 100 ατόμων. Πολλοί ερευνητές προτείνουν ένα ελάχιστο αριθμό 20 ατόμων για κάθε μεταβλητή, ενώ η πλέον αποδεκτή αναλογία είναι 10 άτομα για κάθε μεταβλητή.
- Τα δεδομένα πρέπει να ακολουθούν τη διμεταβλητή κανονική κατανομή για κάθε ζεύγος μεταβλητών και
- Οι παρατηρήσεις να είναι ανεξάρτητες.

!!! *Οι προϋποθέσεις που αναφέρονται παραπάνω είναι αναγκαίες αλλά όχι ικανές. Για την πετυχημένη ολοκλήρωση της παραγοντικής ανάλυσης και την δημιουργία παραγόντων αξιόπιστων είναι απαραίτητο να γίνουν και άλλοι έλεγχοι τους οποίους θα δούμε στη συνέχεια.*

7.3 Τεχνικές Παραγοντικής Ανάλυσης

Η μέθοδος της ανάλυσης παραγόντων αποτελείται από μια μεγάλη ποικιλία τεχνικών οι οποίες εφαρμόζονται κατά περίπτωση. Συγκεκριμένα έχουμε στη διάθεσή μας:

- **Επτά μεθόδους για την εξαγωγή παραγόντων (factor extraction)**.

1. Principal components
2. Unweighted least squares
3. Generalized least squares
4. Maximum likelihood

5. Principal axis factoring
6. Alpha factoring και
7. Image factoring.

- **Πέντε μεθόδους περιστροφής (rotation)**.

1. Varimax
2. Direct Oblimin
3. Quatrimax
4. Equamax
5. Promax.

- **Τρεις μεθόδους για τον υπολογισμό των παραγοντικών σκορ (factor scores).**

1. Regression
2. Barthlett
3. Anderson- Rubin.

7.4 Διαδικασία παραγοντικής ανάλυσης

Στη συνέχεια θα ακολουθήσουμε, βήμα- βήμα, τη διαδικασία της παραγοντικής ανάλυσης, μέσω του S.P.S.S, αναλύοντας τα βασικότερα σημεία αυτής, ερμηνεύοντας κάποιους σημαντικούς πίνακες και δείκτες και πραγματοποιώντας τους απαραίτητους ελέγχους. Στη διαδικασία αυτή θα χρησιμοποιήσουμε ένα συγκεκριμένο παράδειγμα.

Παράδειγμα: Πραγματοποιήθηκε έρευνα σε 208 βιομηχανικές επιχειρήσεις στην Ελλάδα με σκοπό την αξιολόγηση της απόδοσή τους σε διάφορους τομείς. Για το σκοπό αυτό χρησιμοποιήθηκαν 12 μεταβλητές οι οποίες αναφέρονται στη συνέχεια:

$\mathbf{X_1}$: Κόστος ανά μονάδα προϊόντος

$\mathbf{X_2}$: Ποιότητα προϊόντος

X_3: Χρονική ακρίβεια παράδοσης

X_4: Ευελιξία αλλαγής όγκου παραγωγής

X_5: Ταχύτητα εισαγωγής νέων προϊόντων

X_6: Όγκος πωλήσεων

X_7: Ρυθμός μεταβολής όγκου πωλήσεων

X_8: Μερίδιο αγοράς

X_9: Ρυθμός μεταβολής μεριδίου αγοράς

X_{10}: Περιθώρια κέρδους

X_{11}: Απόδοση ίδιων κεφαλαίων

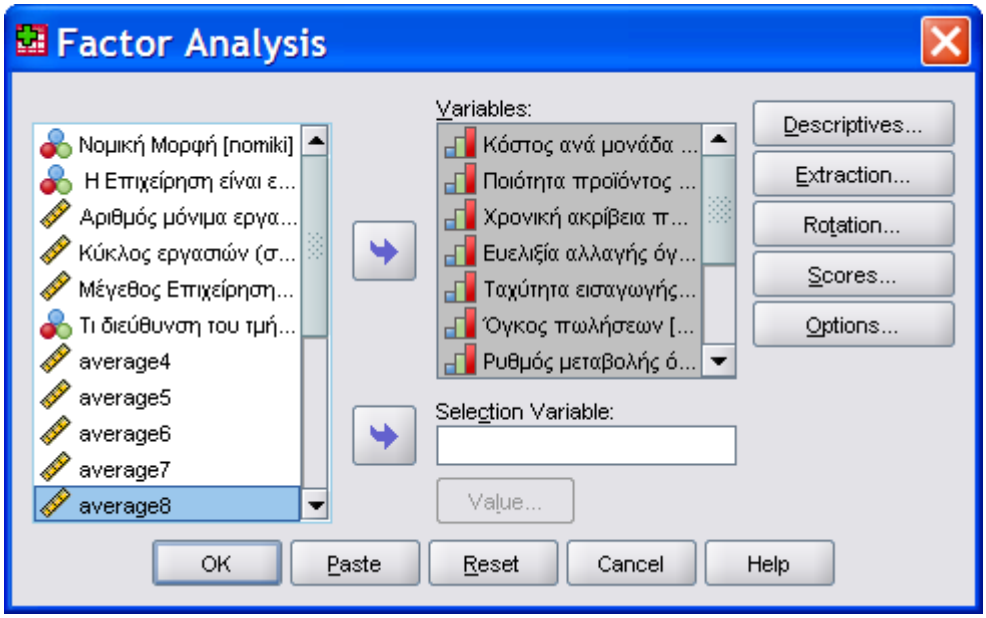
X_{12}: Καθαρά κέρδη μετά φόρων

Οι παραπάνω μεταβλητές χρησιμοποιούνται συνήθως για την μέτρηση της συνολικής απόδοσης των επιχειρήσεων και αξιολογούνται σε 5-βάθμια κλίμακα Likert με την εξής διαβάθμιση: 1= πολύ κάτω από τον μέσο όρο του κλάδου, 2 = κάτω από τον μέσο όρο, 3= στον μέσο όρο, 4= πάνω από τον μέσο όρο και 5= πολύ πάνω από τον μέσο όρο του κλάδου.

Με τη βοήθεια της παραγοντικής ανάλυσης, σε αυτό το παράδειγμα, θα προσπαθήσουμε να ομαδοποιήσουμε τις δώδεκα μεταβλητές έτσι ώστε κάθε ομάδα να αποτελεί μία διαφορετική εννοιολογική υποενότητα στα πλαίσια βέβαια, της γενικής έννοιας της ***«απόδοσης»***.

Η διαδικασία την οποία θα ακολουθήσουμε για την ολοκλήρωση της παραγοντικής ανάλυσης περιγράφεται στη συνέχεια.

- ✓ Από το μενού ***Analyze*** του ***Data Editor***
- ✓ Επιλέγουμε ***Dimension Reduction*** και στη συνέχεια ***Factor***.

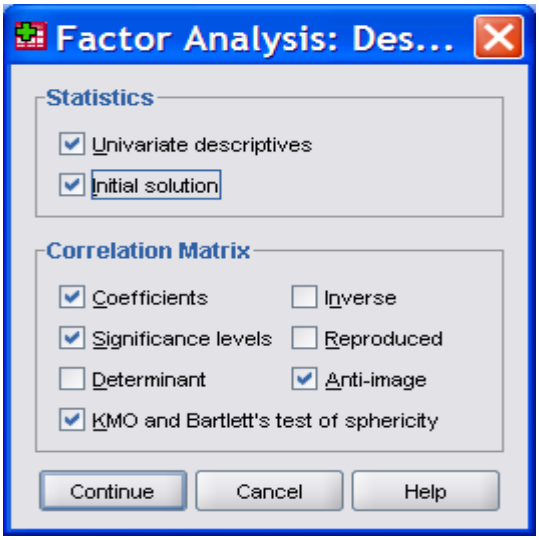


Εικόνα 7.1

✓ Επιλέγουμε τις μεταβλητές οι οποίες θα αποτελέσουν τη βάση της παραγοντικής ανάλυσης (στο παράδειγμά μας τις 12 που περιγράψαμε) και τις περνάμε στο παράθυρο ***Variables***.

!!! *Στη διαδικασία της παραγοντικής ανάλυσης, ακόμη και αν οι βασικές προϋποθέσεις οι οποίες αναφέρθηκαν στην προηγούμενη παράγραφο ικανοποιούνται, είναι απαραίτητο να κάνουμε και άλλους ελέγχους οι οποίοι εξασφαλίζουν την επάρκεια και την καταλληλότητα των δεδομένων και των μεταβλητών για τη δημιουργία των παραγόντων. Για το λόγο αυτό σαν πρώτο βήμα θα πρέπει να είναι η επιλογή των δεικτών εκείνων οι οποίοι ερμηνεύουν τις προηγούμενες προϋποθέσεις.*

✓ Πατάμε στο κουμπί ***Descriptives*** και έχουμε το επόμενο πλαίσιο διαλόγου.

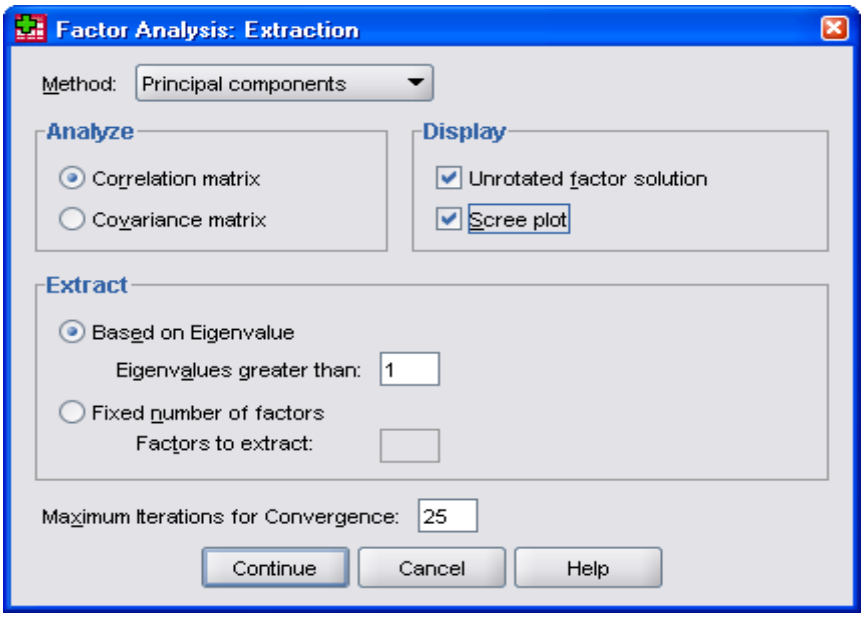


Εικόνα 7.2

✓ Τσεκάρουμε στην περιοχή ***Statistics*** τις ενδείξεις ***Univariate descriptives*** και ***Initial Solution***, στην περιοχή ***Correlation Matrix*** τις ενδείξεις ***Coefficients***, ***Significance leveles***, ***KMO and Batrlett's test of Sphericity*** και ***Anti-image***.

✓ Στη συνέχεια ***Continue***, επιστροφή στην εικόνα 7.1 και

✓ Πατάμε στο κουμπί ***Extraction*** και έχουμε το επόμενο πλαίσιο διαλόγου.



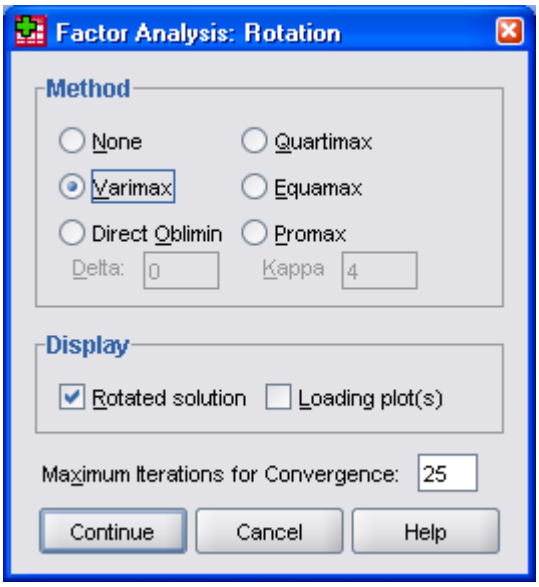
Εικόνα 7.3

✓ Από το πτυσσόμενο παράθυρο ***Method*** επιλέγουμε ***Principal components***.

✓ Από την περιοχή ***Extract*** επιλέγουμε ***Eigenvalues over: 1***.

✓ Από την περιοχή ***Display*** επιλέγουμε ***Unrotated factor solution*** και ***Scree plot***.

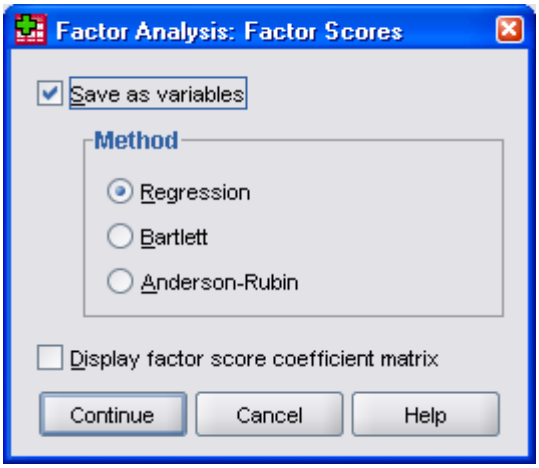
✓ Στη συνέχεια ***Continue*** επιστροφή στην εικόνα 7.1, πάτημα στο κουμπί ***Rotation*** και εμφάνιση του επόμενου πλαισίου διαλόγου.



Εικόνα 7.4

✓ Στην ένδειξη ***Method*** τσεκάρουμε ***Varimax*** και στην περιοχή ***Display*** επιλέγουμε ***Rotated solution***.

✓ Στη συνέχεια ***Continue*** και ***OK*** επιστροφή στην εικόνα 7.1, πάτημα στο κουμπί ***Scores*** και από το επόμενο πλαίσιο



Εικόνα 7.5

✓ Επιλέγουμε ***Save as variables*** και από την περιοχή ***Method*** επιλέγουμε ***Regression***, πατάμε ***Continue*** και επιστρέφουμε στην εικόνα 7.1. Η επιλογή ***Options*** αυτής της εικόνας δεν περιέχει σημαντικές ενέργειες και για το λόγο αυτό στο παρόν σημείο δεν περιγράφεται.

✓ Στη συνέχεια πατάμε ***O.K*** και έχουμε τους πίνακες οι οποίοι ακολουθούν, σαν αποτέλεσμα των επιλογών μας.

Στον πίνακα ***Περιγραφική Στατιστική- Descriptive Statistics*** (πίνακας 7.1) δίνονται τα δύο βασικότερα μέτρα της απλής περιγραφικής στατιστικής, ο ***αριθμητικός μέσος (Mean)*** και η ***τυπική απόκλιση (Std. Deviation)*** για όλες τις μεταβλητές. Συγχρόνως δίνεται και ο αριθμός των παρατηρήσεων για κάθε μεταβλητή ***(Analysis N)***.

Πίνακας 7.1: Descriptives Statistics

	Mean	Std. Deviation	Analysis N
Κόστος ανά μονάδα προϊόντος	3,71	0,879	195
Ποιότητα προϊόντος	4,49	0,645	195
Χρονική ακρίβεια παράδοσης	4,22	0,803	195
Ευελιξία αλλαγής όγκου παραγωγής	3,97	0,879	195
Ταχύτητα εισαγωγής νέων προϊόντων	3,92	0,927	195
Όγκος πωλήσεων	4,09	0,801	195
Ρυθμός μεταβολής όγκου πωλήσεων	3,87	0,775	195
Μερίδιο αγοράς	3,96	0,907	195
Ρυθμός μεταβολής μεριδίου αγοράς	3,74	0,865	195
Περιθώρια κέρδους	3,74	0,935	195
Απόδοση ίδιων κεφαλαίων	3,71	0,898	195
Καθαρά κέρδη μετά φόρων	3,67	1,018	195

Στον πίνακα ***Συσχετίσεων- Correlation Matrix*** (Πίνακας 7.2) και στη γραμμή ***Correlation*** εμφανίζονται οι ***συσχετίσεις όλων των ζευγών των μεταβλητών***. Παρατηρούμε στη διαγώνιο του πίνακα υπάρχει η τιμή 1 η οποία είναι ο συντελεστής συσχέτισης της κάθε μεταβλητής με την ίδια. Οι τιμές κάτω από τη διαγώνιο είναι ίδιες με αυτές πάνω από τη διαγώνιο. Από μία πρώτη ματιά μπορούμε να εντοπίσουμε μεταξύ ποιών μεταβλητών υπάρχει μεγάλη εξάρτηση, ενδεικτικό των παραγόντων που πρόκειται να δημιουργηθούν στη συνέχεια. Στον ίδιο πίνακα, στη γραμμή ***Sig. (1-tailed)***, εμφανίζονται οι σημαντικότητες αυτών των συσχετίσεων. Στο συγκεκριμένο παράδειγμα όλες είναι σημαντικές σε επίπεδο σημαντικότητας 0,05. Αυτό το στοιχείο δημιουργεί τις προϋποθέσεις για περαιτέρω εξέταση της επάρκειας για παραγοντική ανάλυση. Το πλήθος των συσχετίσεων είναι ν(ν-1)/2, όπου ν το πλήθος των μεταβλητών.

Πίνακας 7.2: Correlation Matrix

		χ1.	χ2.	χ3.	χ4.	χ5.	χ6.	χ7.	χ8.	χ9.	χ10.	χ11.	χ12.
Correlation	**X1.**	**1,000**	**,257**	**,278**	**,289**	**,313**	**,529**	**,443**	**,444**	**,404**	**,510**	**,396**	**,436**
	X2.	,257	**1,000**	**,414**	**,418**	**,386**	**,322**	**,306**	**,307**	**,234**	**,101**	**,194**	**,127**
	X3.	,278	,414	**1,000**	**,492**	**,315**	**,290**	**,278**	**,324**	**,318**	**,213**	**,317**	**,295**
	X4.	,289	,418	,492	1,000	**,528**	**,334**	**,342**	**,276**	**,227**	**,166**	**,283**	**,196**
	X5.	,313	,386	,315	,528	**1,000**	**,295**	**,315**	**,321**	**,334**	**,290**	**,355**	**,283**
	X6.	,529	,322	,290	,334	,295	**1,000**	**,635**	**,679**	**,555**	**,404**	**,418**	**,430**
	X7.	,443	,306	,278	,342	,315	,635	**1,000**	**,461**	**,641**	**,414**	**,447**	**,408**
	X8.	,444	,307	,324	,276	,321	,679	,461	**1,000**	**,584**	**,449**	**,466**	**,432**
	X9.	,404	,234	,318	,227	,334	,555	,641	,584	**1,000**	**,484**	**,547**	**,513**
	X10.	,510	,101	,213	,166	,290	,404	,414	,449	,484	**1,000**	**,596**	**,700**
	X11.	,396	,194	,317	,283	,355	,418	,447	,466	,547	,596	**1,000**	**,696**
	χ12.	,436	,127	,295	,196	,283	,430	,408	,432	,513	,700	,696	**1,000**
Sig. (1-tailed)	**X1.**		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	X2.	,000		,000	,000	,000	,000	,000	,000	,000	,080	,003	,038
	X3.	,000	,000		,000	,000	,000	,000	,000	,000	,001	,000	,000
	X4.	,000	,000	,000		,000	,000	,000	,000	,001	,010	,000	,003
	X5.	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000
	X6.	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000
	X7.	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000
	X8.	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000
	X9.	,000	,000	,000	,001	,000	,000	,000	,000		,000	,000	,000
	X10.	,000	,080	,001	,010	,000	,000	,000	,000	,000		,000	,000
	X11.	,000	,003	,000	,000	,000	,000	,000	,000	,000	,000		,000
	χ12.	,000	,038	,000	,003	,000	,000	,000	,000	,000	,000	,000	

Στον πίνακα ***KMO and Bartlett's Test*** (Πίνακας 7.3) ένας δείκτης σύγκρισης του σχετικού μεγέθους των συντελεστών συσχέτισης με τους μερικούς συντελεστές συσχέτισης είναι το στατιστικό ***Kaiser-Meyer-Olkin***. Οι τιμές του δείκτη αυτού κυμαίνονται από 0 έως 1.

Τιμές κοντά στη μονάδα δηλώνουν ότι τα δεδομένα είναι κατάλληλα για παραγοντική ανάλυση. Αντίθετα τιμές κάτω του 0,5 θεωρούνται μη αποδεκτές και δεν συνιστάται η συνέχιση της παραγοντικής διαδικασίας. Στην πράξη τιμές γύρω στο 0,8 θεωρούνται αρκετά καλές για να συνεχίσουμε.

Γενικότερα μπορούμε να πούμε ότι η τιμή του δείκτη ***KMO*** μεγαλώνει όταν:

1. Το μέγεθος του δείγματος μεγαλώνει.
2. Ο μέσος όρος των συσχετίσεων μεγαλώνει.
3. Το πλήθος των μεταβλητών αυξάνει ή
4. Το πλήθος των παραγόντων ελαττώνεται.

Στον ίδιο πίνακα υπάρχει και το ***Bartlett's Test of sphericity*** το οποίο αποφαίνεται για την παρουσία συσχετίσεων μεταξύ των μεταβλητών. Στην ουσία μας δίνει την πιθανότητα ο πίνακας συσχέτισης να έχει σημαντικές συσχετίσεις μεταξύ κάποιων μεταβλητών. Έτσι αν το Sig. του δείκτη αυτού είναι μικρότερο του 0,05 απορρίπτεται η υπόθεση της μη ύπαρξης σημαντικών συσχετίσεων, σε επίπεδο σημαντικότητας 5%.

Στο παράδειγμά μας, από τον πίνακα 7.3, διαπιστώνουμε ότι ο δείκτης ***KMO*** είναι 0,863 κάτι που σημαίνει ότι προτείνεται η συνέχιση της διαδικασίας της παραγοντικής ανάλυσης καθώς και ***sig.*** του δείκτη ***Bartlett's Test of sphericity*** είναι 0, δηλαδή υπάρχουν στον πίνακα συσχετίσεων μεταβλητές οι οποίες συσχετίζονται σε ικανοποιητικό βαθμό.

Στην περίπτωση που ο δείκτης ***K.M.O*** δεν είναι ικανοποιητικός αυτό θα οφείλεται στο ότι οι δείκτες ***M.S.A*** κάποιας ή κάποιων μεταβλητών είναι πολύ χαμηλοί.

Πίνακας 7.3: K.M.O and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,863
Bartlett's Test of Sphericity	Approx. Chi-Square	1068,521
	df	66
	Sig.	,000

Για να ελέγξουμε τον δείκτη ***MSA*** (***Measures of Sampling Adequacy-Μέτρα-Δειγματικής Επάρκειας)*** της κάθε μεταβλητή ξεχωριστά, ανατρέχουμε στον πίνακα ***Anti-image Matrices*** (Πίνακας 7.4). Οι μεταβλητές που έχουν τιμές ***MSA*** μη αποδεκτές θα αποκλειστούν από τη συνέχεια της διαδικασίας και οι υπόλοιπες θα δώσουν ***KMO*** αποδεκτό έτσι ώστε να συνεχιστεί η διαδικασία. Η μείωση των μεταβλητών πιθανόν να οδηγήσει και σε ταυτόχρονη μείωση των παραγόντων, κάτι που θα προκαλέσει αύξηση της τιμής του δείκτη ***KMO***. Στον πίνακα που ακολουθεί θα ελέγξουμε τους ατομικούς δείκτες για κάθε μεταβλητή και θα αποφανθούμε, με το σκεπτικό που ήδη αναπτύξαμε, ποιες μεταβλητές πρέπει να αποκλείσουμε από τη συνέχεια. Βέβαια, επειδή η τιμή του δείκτη Κ.Μ.Ο, στο συγκεκριμένο παράδειγμα, είναι πολύ ικανοποιητική θα διαπιστώσουμε ότι οι τιμές M.S.A όλων των μεταβλητών είναι αποδεκτές.

Στη γραμμή ***Anti-Image Correlation*** του πίνακα 7.4, οι τιμές διαγώνια (έντονη γραφή) είναι οι τιμές του δείκτη ***Measures of Sampling Adequacy (MSA)*** για κάθε μεταβλητή. Ο δείκτης αυτός μας επιτρέπει να εξετάσουμε την καταλληλότητα κάθε μεταβλητής χωριστά για την χρησιμοποίησή της στην ανάλυση. Τιμές κοντά στο 1 είναι ενδείξεις ότι η μεταβλητή είναι πολύ καλή για να χρησιμοποιηθεί ενώ

τιμές μικρότερες του 0,5 φανερώνουν ακαταλληλότητα και συνεπώς οι μεταβλητές αυτές πρέπει να αποκλείονται.

Πίνακας 7.4: Anti-image Matrices

		χ1.	χ2.	χ3.	χ4.	χ5.	χ6.	χ7.	χ8.	χ9.	χ10.	χ11.	χ12.
Anti-image Covariance	**X1.**	,593	-,036	-,030	-,022	-,036	-,108	-,032	-,011	,013	-,134	,009	-,014
	X2.	-,036	,693	-,159	-,068	-,126	-,028	-,052	-,048	,022	,046	,002	,028
	X3.	-,030	-,159	,649	-,205	,036	,027	,026	-,038	-,057	,026	-,019	-,056
	X4.	-,022	-,068	-,205	,553	-,229	-,041	-,066	,011	,069	,027	-,036	,022
	X5.	-,036	-,126	,036	-,229	,624	,030	,012	-,022	-,055	-,036	-,043	,001
	X6.	-,108	-,028	,027	-,041	,030	,375	-,149	-,191	-,009	,030	,021	-,038
	X7.	-,032	-,052	,026	-,066	,012	-,149	,442	,068	-,177	-,032	-,027	,016
	X8.	-,011	-,048	-,038	,011	-,022	-,191	,068	,438	-,113	-,052	-,040	,019
	X9.	,013	,022	-,057	,069	-,055	-,009	-,177	-,113	,425	-,014	-,063	-,034
	X10.	-,134	,046	,026	,027	-,036	,030	-,032	-,052	-,014	,424	-,058	-,171
	X11.	,009	,002	-,019	-,036	-,043	,021	-,027	-,040	-,063	-,058	,432	-,168
	χ12.	-,014	,028	-,056	,022	,001	-,038	,016	,019	-,034	-,171	-,168	,374
Anti-image Correlation	**X1.**	**,926**[a]	-,056	-,048	-,039	-,058	-,229	-,062	-,022	,027	-,268	,019	-,030
	X2.	-,056	**,874**[a]	-,238	-,109	-,192	-,055	-,093	-,088	,040	,086	,003	,055
	X3.	-,048	-,238	**,845**[a]	-,341	,057	,054	,048	-,071	-,109	,050	-,035	-,113
	X4.	-,039	-,109	-,341	**,790**[a]	-,390	-,091	-,134	,022	,142	,055	-,074	,048
	X5.	-,058	-,192	,057	-,390	**,854**[a]	,063	,023	-,042	-,106	-,071	-,083	,002
	X6.	-,229	-,055	,054	-,091	,063	**,841**[a]	-,365	-,472	-,024	,076	,051	-,102
	X7.	-,062	-,093	,048	-,134	,023	-,365	**,854**[a]	,156	-,408	-,073	-,061	,039
	X8.	-,022	-,088	-,071	,022	-,042	-,472	,156	**,860**[a]	-,262	-,121	-,092	,048
	X9.	,027	,040	-,109	,142	-,106	-,024	-,408	-,262	**,883**[a]	-,034	-,147	-,084
	X10.	-,268	,086	,050	,055	-,071	,076	-,073	-,121	-,034	**,864**[a]	-,136	-,431
	X11.	,019	,003	-,035	-,074	-,083	,051	-,061	-,092	-,147	-,136	**,902**[a]	-,417
	χ12.	-,030	,055	-,113	,048	,002	-,102	,039	,048	-,084	-,431	-,417	**,845**[a]

Ο πίνακας ***Communalities- συμμετοχικότητες*** (πίνακας 7.5) περιέχει τις τιμές ***communality (Initial and Extraction)*** οι οποίες εκφράζουν το ποσοστό της μεταβολής της κάθε μεταβλητής το οποίο

ερμηνεύεται από τους παράγοντες. Στον συγκεκριμένο πίνακα υπάρχουν οι τιμές πριν και μετά την επιλογή του πλήθους των παραγόντων της ανάλυσης. Στη μέθοδο ***Principal components –κύριες συνιστώσες*** οι τιμές ***Initial*** είναι πάντα 1. Οι τιμές ***Communalities*** κυμαίνονται από 0 έως 1. Τιμή 0 δηλώνει ότι οι παράγοντες δεν ερμηνεύουν κανένα ποσοστό μεταβολής της μεταβλητής, ενώ τιμή 1 δηλώνει ότι το 100% των μεταβολών της μεταβλητής ερμηνεύεται από τους παράγοντες. Σε άλλες μεθόδους εξαγωγής παραγόντων ***(factor extraction)*** οι τιμές της στήλης ***extraction*** είναι το R^2 της πολλαπλής παλινδρόμησης στην οποία η κάθε μεταβλητή είναι εξαρτημένη, με ανεξάρτητες κάθε φορά όλες τις άλλες.

Πίνακας 7.5: Communalities

	Initial	Extraction
Κόστος ανά μονάδα προϊόντος	1,000	,461
Ποιότητα προϊόντος	1,000	,579
Χρονική ακρίβεια παράδοσης	1,000	,493
Ευελιξία αλλαγής όγκου παραγωγής	1,000	,662
Ταχύτητα εισαγωγής νέων προϊόντων	1,000	,488
Όγκος πωλήσεων	1,000	,578
Ρυθμός μεταβολής όγκου πωλήσεων	1,000	,536
Μερίδιο αγοράς	1,000	,552
Ρυθμός μεταβολής μεριδίου αγοράς	1,000	,612
Περιθώρια κέρδους	1,000	,667
Απόδοση ίδιων κεφαλαίων	1,000	,612
Καθαρά κέρδη μετά φόρων	1,000	,671

Extraction Method: Principal Component Analysis.

Στον πίνακα ***Total Variance Explained-συνολική ερμηνευθείσα διακύμανση*** (πίνακας 7.6) έχουμε πολλές στήλες με σημαντικές πληροφορίες.

✓ Η στήλη ***Component*** απλά δίνει το πλήθος των παραγόντων οι οποίοι είναι όσες και οι μεταβλητές. Στην προκειμένη περίπτωση φτάνει μέχρι το 12, καθώς 12 είναι οι μεταβλητές μας.

✓ Η στήλη ***Total (Initial Eingenvalues)*** δίνει τις ιδιοτιμές–***eingenvalues*** για κάθε παράγοντα και είναι ταξινομημένες κατά τάξη μεγέθους. Το άθροισμα των τιμών αυτών είναι ίσο με το πλήθος των παραγόντων. Στη δική μας περίπτωση είναι 12.

✓ Η στήλη ***% of Variance (Initial Eingenvalues)*** δίνει το ποσοστό της διακύμανσης το οποίο ερμηνεύεται από τον παράγοντα. Είναι το πηλίκο της κάθε ιδιοτιμής της στήλης ***Total*** δια του ***συνολικού πλήθους των παραγόντων***. Στο παράδειγμά μας:
44,430=(5,332/12)*100.

✓ Η στήλη ***Cumulative % (Initial Eingenvalues)*** περιέχει αθροιστικά τα ποσοστά της προηγούμενης στήλης.

✓ Η στήλη ***Total (Extraction of Sums of Squared Loadings)*** δίνει μόνο τους παράγοντες των οποίων οι ιδιοτιμές είναι μεγαλύτερες του 1. Η επιλογή αυτής της τιμής για την ιδιοτιμή έγινε στην εικόνα 7.3. Είναι στην ουσία ένας δείκτης ο οποίος καθορίζει τον αριθμό των παραγόντων οι οποίοι θα προκύψουν από την παραγοντική ανάλυση. Δηλαδή, οι παράγοντες εκείνοι οι οποίοι έχουν ιδιοτιμή μεγαλύτερη από αυτήν που εμείς ορίσαμε θα αποτελέσουν τους τελικούς παράγοντες της ανάλυσης. Με τον τρόπο αυτό ο τελικός αριθμός των παραγόντων δεν είναι εκ των προτέρων γνωστός. Αντίθετα αν για άλλους λόγους και ανεξαρτήτως ιδιοτιμής θελήσουμε ένα συγκεκριμένο αριθμό παραγόντων τότε στην εικόνα 7.3 και στη θέση ***Number of Factors*** αναγράφουμε ένα αριθμό ο οποίος δηλώνει τον επιθυμητό αριθμό παραγόντων. Στο συγκεκριμένο παράδειγμα, οι

παράγοντες οι οποίοι δημιουργούνται με βάση αυτό το κριτήριο είναι δύο.

- Η στήλη ***% of Variance (Extraction Sums of Squared Loadings)*** δίνει το ποσοστό της διακύμανσης το οποίο ερμηνεύεται από τους δύο παράγοντες, με ιδιοτιμή μεγαλύτερη του 1. Στο παράδειγμά μας ο πρώτος παράγοντας περιγράφει το 44,430% των μεταβολών, ο δεύτερος το 13,169%.

- ✓ Η στήλη ***Cumulative % (Extraction Sums of Squared Loadings)*** περιέχει αθροιστικά τα δύο ποσοστά της προηγούμενης στήλης (44,430%+13,169%=57,599%).

!!! *Στη μέθοδο* ***Principal components****, όπως θα έχετε προσέξει, οι τιμές στις περιοχές* ***Initial*** *και* ***Extraction*** *παραμένουν αμετάβλητες*.

- ✓ Η στήλη ***Total (Rotation Sums of Squared Loadings)*** δίνει τους παράγοντες οι οποίοι έχουν ιδιοτιμές μεγαλύτερες του 1 μετά την ***περιστροφή (rotation)***. Παρατηρούμε ότι οι τιμές μετά την περιστροφή έχουν αλλάξει.

- ✓ Η στήλη ***% of Variance (Rotation of Sums of Squared Loadings)*** δίνει το ποσοστό της διακύμανσης το οποίο ερμηνεύεται από τους δύο παράγοντες, με ιδιοτιμή μεγαλύτερη του 1, μετά την περιστροφή.

- ✓ Η στήλη ***Cumulative % (Rotation of Sums of Squared Loadings)*** είναι η στήλη των αθροιστικών ποσοστών της προηγούμενης στήλης. Στην πρώτη γραμμή αναγράφεται το ποσοστό του πρώτου παράγοντα (35,854%) και στην δεύτερη γραμμή προστίθεται σε αυτό και το ποσοστό του δεύτερου παράγοντα (35,854%+21,745%= 57,599%).

Πίνακας 7.6: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5,332	44,430	44,430	5,332	44,430	44,430	4,302	35,854	35,854
2	1,580	13,169	57,599	1,580	13,169	57,599	2,609	21,745	57,599
3	,991	8,256	65,854						
4	,706	5,887	71,741						
5	,671	5,595	77,336						
6	,584	4,866	82,202						
7	,544	4,537	86,739						
8	,443	3,693	90,432						
9	,354	2,948	93,380						
10	,315	2,622	96,002						
11	,268	2,234	98,236						
12	,212	1,764	100,000						

Extraction Method: Principal Component Analysis.

!!! *Ενώ οι τιμές των στηλών* ***Total*** *και* ***% of Variance*** *της περιοχής* ***Rotation of Sums of Squared Loadings*** *είναι διαφορετικές από τις αντίστοιχες της περιοχής* ***Extraction Sums of Squared Loadings****, η τιμή* ***Commulative*** *και στις δύο περιοχές είναι η ίδια. Δηλαδή, συνολικά οι δύο παράγοντες και με τις δύο μεθόδους, ερμηνεύουν ίδιο ποσοστό διακύμανσης.*

!!! *Για τον καθορισμό του πλήθους των παραγόντων μπορούμε να πάρουμε υπόψη μας κάποιο από τα επόμενα τέσσερα κριτήρια.*

- *Το πρώτο κριτήριο βασίζεται στην τιμή της ιδιοτιμής που εμείς καθορίζουμε. Συνήθως πρέπει να είναι μεγαλύτερη του 1. Έτσι κάθε παράγοντες με ιδιοτιμή μεγαλύτερη του 1 θεωρείται σημαντικός, ενώ κάθε παράγοντας με ιδιοτιμή μικρότερη του 1 θεωρείται μη σημαντικός*

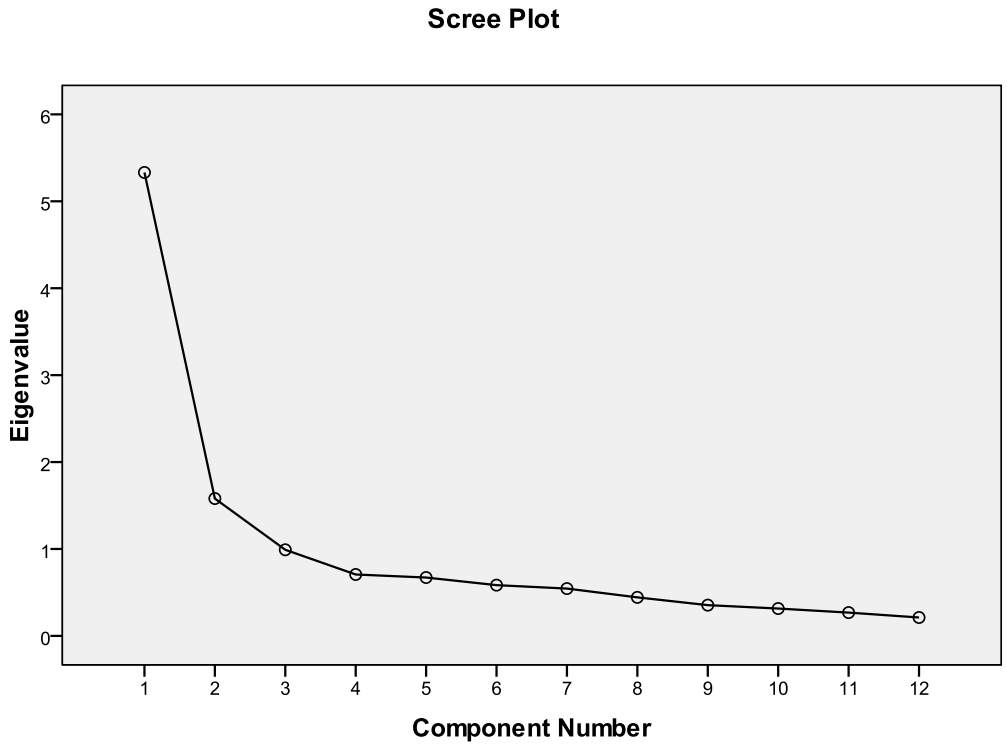
και αγνοείται. Αυτό το κριτήριο συνίσταται όταν οι μεταβλητές είναι από 20 έως 50. Στην περίπτωση που οι μεταβλητές είναι λιγότερες των 20 είναι πιθανό να μας δώσει πολύ μικρό αριθμό παραγόντων, έτσι ώστε να μη μπορούμε να βγάλουμε ασφαλή συμπεράσματα. Αντίθετα στην περίπτωση κατά την οποία έχουμε περισσότερες από 50 μεταβλητές είναι πολύ πιθανό να μας δώσει πάρα πολλούς παράγοντες.

- *Το δεύτερο κριτήριο είναι η εκ των προτέρων βούληση για συγκεκριμένο αριθμό παραγόντων με βάση κάποια δεδομένα. Αν για παράδειγμα ο ερευνητής θέλει να κάνει σύγκριση των αποτελεσμάτων της δικής του έρευνας με κάποια άλλη η οποία έδωσε 5 παράγοντες θα πρέπει και αυτός να ζητήσει 5 παράγοντες.*
- *Το τρίτο κριτήριο είναι το άθροισμα των διακυμάνσεων το οποίο θεωρούμε ικανοποιητικό. Αν δηλαδή θέλουμε το ποσοστό της διακύμανσης το οποίο περιγράφεται από τους παράγοντες να είναι 85% τότε θα επιλέξουμε τόσους διαδοχικούς παράγοντες ώστε να φτάσουμε στο επιθυμητό αποτέλεσμα.*
- *Το τέταρτο κριτήριο στηρίζεται στο* ***Scree Plot- κρημνογράφημα*** *(Σχήμα 1). Με βάση αυτή τη γραφική απεικόνιση, μετά από το σημείο το οποίο η καμπύλη τείνει να γίνει ευθεία πρέπει να απορρίπτονται οι παράγοντες. Σε σχέση με το πρώτο κριτήριο δίνει τουλάχιστον ένα παράγοντα περισσότερο.*

Στο σχήμα που ακολουθεί ***(Scree Plot- κρημνογράφημα)*** βλέποντας την ιδιοτιμή του κάθε παράγοντα, μπορούμε εύκολα να προσδιορίσουμε αυτούς που υπερβαίνουν το 1 και επομένως αποτελούν τους παράγοντες που πληρούν τον περιορισμό που θέσαμε με βάση το πρώτο κριτήριο. Παρατηρούμε επίσης ότι από το σημείο 3 και δεξιά η

καμπύλη γίνεται σχεδόν ευθεία. Θα μπορούσαμε επομένως, βάση του τέταρτου κριτηρίου, να δημιουργήσουμε και τρεις παράγοντες.

Σχήμα 1: Scree Plot



Στον πίνακα ***Component Matrix*** (πίνακας 7.7) έχουμε πλέον τους δύο ***παράγοντες (component 1 and 2)*** και τις αντίστοιχες ***φορτίσεις (loandings)*** των μεταβλητών στους δύο αυτούς παράγοντες. Οι τιμές αυτές κυμαίνονται από -1 έως 1.

Πίνακας 7.7: Component Matrix[α]

	Component	
	1	**2**
Κόστος ανά μονάδα προϊόντος	-,086	,674
Ποιότητα προϊόντος	,464	,602
Χρονική ακρίβεια παράδοσης	,535	,455
Ευελιξία αλλαγής όγκου παραγωγής	,530	,617

Ταχύτητα εισαγωγής νέων προϊόντων	,565	,410
Όγκος πωλήσεων	,759	-,036
Ρυθμός μεταβολής όγκου πωλήσεων	,731	-,035
Μερίδιο αγοράς	,740	-,070
Ρυθμός μεταβολής μεριδίου αγοράς	,759	-,189
Περιθώρια κέρδους	,694	-,430
Απόδοση ίδιων κεφαλαίων	,737	-,262
Καθαρά κέρδη μετά φόρων	,717	-,396

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

!!! *Από την στιγμή που επιλέξαμε ως μέθοδο ορθογώνιας περιστροφής των αξόνων την μέθοδο Varimax ο πίνακας 7.7 δεν έχει μεγάλη σημασία.*

Στον πίνακα ***Rotated Component Matrix*** (πίνακας 7.8) έχουμε τους δύο ***παράγοντες (component 1 and 2)*** με τις αντίστοιχες ***φορτίσεις (loandings)*** των μεταβλητών στους δύο αυτούς παράγοντες, μετά την ορθογώνια περιστροφή. Η περιστροφή έχει ως σκοπό την αύξηση των μεγάλων φορτίσεων και αντίστοιχα τη μείωση των μικρών. Από τον πίνακα αυτό θα καταλήξουμε τελικά στη σύνθεση των παραγόντων. Το κριτήριο το οποίο θα χρησιμοποιήσουμε είναι το μέγεθος της φόρτισης της κάθε μεταβλητής στους παράγοντες. Με απλά λόγια η μεταβλητή ανήκει στον παράγοντα στον οποίο έχει μεγαλύτερη φόρτιση.

Πίνακας 7.8: Rotated Component Matrix[a]

	Component	
	1	**2**
Κόστος ανά μονάδα προϊόντος	,280	,619
Ποιότητα προϊόντος	,080	,756

Χρονική ακρίβεια παράδοσης	,217	,667
Ευελιξία αλλαγής όγκου παραγωγής	,128	,803
Ταχύτητα εισαγωγής νέων προϊόντων	,267	,646
Όγκος πωλήσεων	,666	,367
Ρυθμός μεταβολής όγκου πωλήσεων	,641	,354
Μερίδιο αγοράς	,667	,328
Ρυθμός μεταβολής μεριδίου αγοράς	,746	,237
Περιθώρια κέρδους	,817	-,003
Απόδοση ίδιων κεφαλαίων	,765	,163
Καθαρά κέρδη μετά φόρων	,818	,038

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Στο συγκεκριμένο παράδειγμα, παρατηρούμε ότι οι πέντε μεταβλητές: ***Κόστος ανά μονάδα προϊόντος, Ποιότητα προϊόντος, Χρονική ακρίβεια παράδοσης, Ευελιξία αλλαγής όγκου παραγωγής*** και ***Ταχύτητα εισαγωγής νέων προϊόντων*** ανήκουν στον πρώτο παράγοντα, ενώ οι επτά μεταβλητές: ***Όγκος πωλήσεων, Ρυθμός μεταβολής όγκου πωλήσεων, Μερίδιο αγοράς, Ρυθμός μεταβολής μεριδίου αγοράς, Περιθώρια κέρδους, Απόδοση ίδιων κεφαλαίων*** και ***Καθαρά κέρδη μετά φόρων*** ανήκουν στο δεύτερο παράγοντα.

Σε αυτό το στάδιο μπορούμε να ονομάσουμε τους παράγοντες οι οποίοι δημιουργήθηκαν. Αυτό επιτυγχάνεται με την προσεκτική μελέτη των μεταβλητών οι οποίες απαρτίζουν τον κάθε παράγοντα και σύμφωνα με την θεωρία στην οποία βασίστηκε η επιλογή των μεταβλητών. Προσπαθούμε δηλαδή να βρούμε μία ονομασία- χαρακτηρισμό η οποία καλύπτει όλες τις μεταβλητές του παράγοντα. Στο παράδειγμά μας, ο πρώτος παράγοντας μπορεί εύκολα να ονομασθεί ***«Λειτουργική Απόδοση-Operational Performance»*** και ο δεύτερος

«Χρηματοοικονομική Απόδοση- Financial Performance» καθώς στην πράξη οι μεταβλητές αυτές χρησιμοποιούνται για την μέτρηση της λειτουργικής και της χρηματοοικονομικής απόδοσης των επιχειρήσεων.

!!! *Για να θεωρηθούν σημαντικές οι φορτίσεις σε κάθε παράγοντα θα πρέπει να λάβουμε υπόψη τρία κριτήρια.*

- *Το πρώτο κριτήριο είναι η τιμή της* ***φόρτισης (factor loading)*** *της μεταβλητής. Μπορεί δηλαδή μια μεταβλητή να ανήκει σε ένα παράγοντα γιατί απλώς εκεί έχει μεγαλύτερη φόρτιση. Αυτό όμως δεν σημαίνει κατ' ανάγκη ότι η φόρτιση είναι ικανοποιητική. Ενδεικτικά αναφέρω ότι φορτίσεις ±0,5 και πάνω θεωρούνται σημαντικές, ενώ ως ελάχιστο αποδεκτό όριο θεωρείται η τιμή ±0,3. Επειδή φόρτιση είναι η συσχέτιση του παράγοντα με τη μεταβλητή, το τετράγωνο της φόρτισης δίνει το ποσό της συνολικής διασποράς της μεταβλητής το οποίο οφείλεται στον παράγοντα. Έτσι μία φόρτιση 0,6 δηλώνει ότι το 36% της συνολικής διασποράς της μεταβλητής οφείλεται στον παράγοντα.*

- *Το δεύτερο κριτήριο είναι το μέγεθος του δείγματος σε συνδυασμό με την τιμή της φόρτισης. Για να θεωρήσουμε δηλαδή τη φόρτιση μίας μεταβλητής ικανοποιητική θα πρέπει να ελέγξουμε και το μέγεθος του δείγματος. Χαρακτηριστικά αναφέρω ότι για δείγμα μεγέθους 100 μονάδων απαιτείται φόρτιση 0,55 και πάνω για να θεωρηθεί σημαντική. Όσο βέβαια το μέγεθος του δείγματος μεγαλώνει η φόρτιση που απαιτείται για να θεωρηθεί σημαντική μικραίνει. Έτσι ένα δείγμα μεγέθους 350 μονάδων απαιτεί φόρτιση 0,3.*

- *Το τρίτο κριτήριο είναι το πλήθος των μεταβλητών οι οποίες λαμβάνουν μέρος στην παραγοντική ανάλυση. Αυξάνοντας το πλήθος τους, το επιθυμητό όριο της φόρτισης μικραίνει.*

Ο πίνακας ***Μετατροπής των Συνιστωσών- Component Transformation Matrix*** (πίνακας 7.9) είναι ο εκ περιστροφής πίνακας ο οποίος χρησιμοποιήθηκε για τη μετατροπή των φορτίσεων του πίνακα 7.7, σε αυτές του πίνακα 7.8.

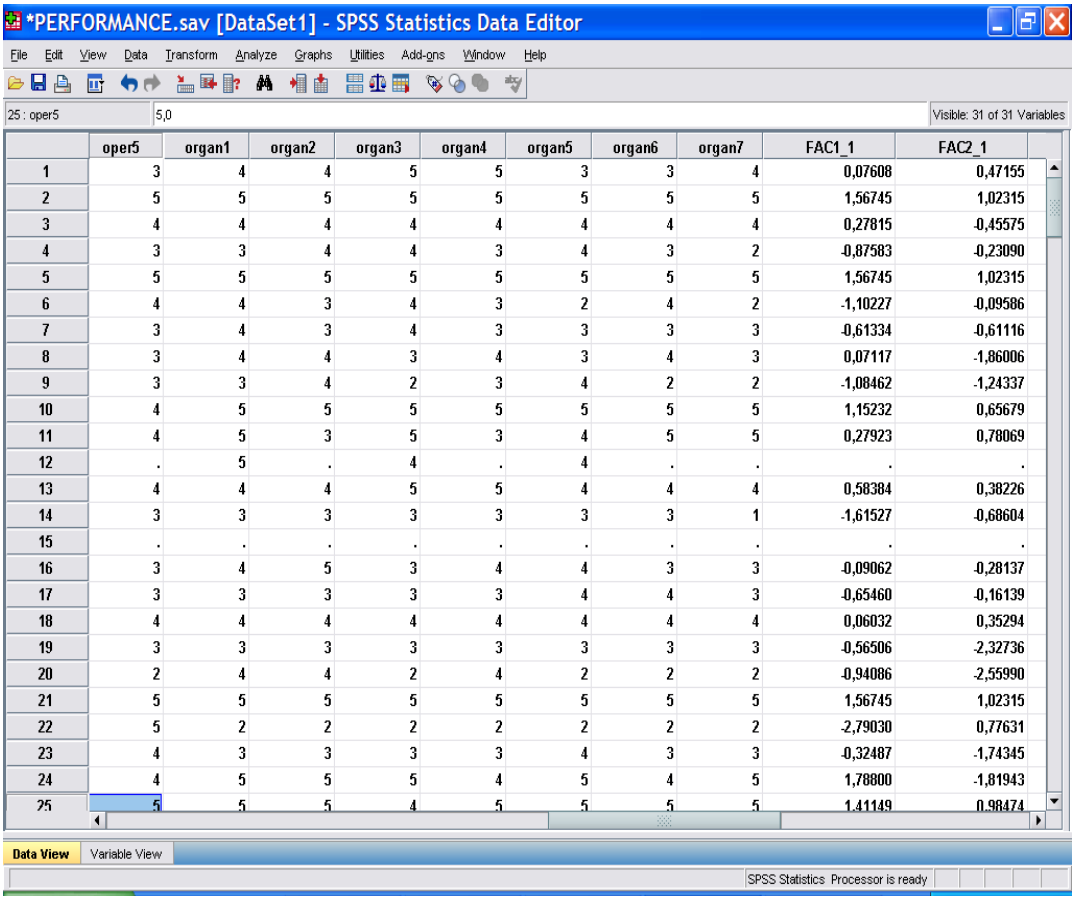
Πίνακας 7.9: Component Transformation Matrix

Component	1	2
1	,852	,524
2	-,524	,852

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Τέλος επιστρέφοντας στο ***Data Editor*** θα παρατηρήσουμε δύο νέες στήλες (Εικόνα 7.6) με ονόματα ***fac1_1*** και ***fac2_1*** οι οποίες είναι τα παραγοντικά σκορ τα οποία ζητήσαμε να εμφανίζονται σαν νέες μεταβλητές, στην εικόνα 7.7, με την επιλογή ***Save as Variables***. Με απλά λόγια για κάθε ***περίπτωση- case***, έχουμε δύο νέες μεταβλητές, ***fac1_1*** και ***fac2_1***, οι οποίες μπορούν να αντικαταστήσουν τις αρχικές μεταβλητές σε άλλες πολυμεταβλητές μεθόδους ανάλυσης όπως ***t-τεστ, παλινδρόμηση, ανάλυση διακύμανσης***.

!!!! *Το χαρακτηριστικό γνώρισμα των παραγοντικών σκορ είναι ότι ο συντελεστής συσχέτισής τους είναι 0.*



	oper5	organ1	organ2	organ3	organ4	organ5	organ6	organ7	FAC1_1	FAC2_1
1	3	4	4	5	5	3	3	4	0,07608	0,47155
2	5	5	5	5	5	5	5	5	1,56745	1,02315
3	4	4	4	4	4	4	4	4	0,27815	-0,45575
4	3	3	4	4	3	4	3	2	-0,87583	-0,23090
5	5	5	5	5	5	5	5	5	1,56745	1,02315
6	4	4	3	4	3	2	4	2	-1,10227	-0,09586
7	3	4	3	4	3	3	3	3	-0,61334	-0,61116
8	3	4	4	3	4	3	4	3	0,07117	-1,86006
9	3	3	4	2	3	4	2	2	-1,08462	-1,24337
10	4	5	5	5	5	5	5	5	1,15232	0,65679
11	4	5	3	5	3	4	5	5	0,27923	0,78069
12	.	5	.	4	.	4	.	.	.	.
13	4	4	4	5	5	4	4	4	0,58384	0,38226
14	3	3	3	3	3	3	3	1	-1,61527	-0,68604
15	.	.	.	.	.	.	.	.	.	.
16	3	4	5	3	4	4	3	3	-0,09062	-0,28137
17	3	3	3	3	3	4	4	3	-0,65460	-0,16139
18	4	4	4	4	4	4	4	4	0,06032	0,35294
19	3	3	3	3	3	3	3	3	-0,56506	-2,32736
20	2	4	4	2	4	2	2	2	-0,94086	-2,55990
21	5	5	5	5	5	5	5	5	1,56745	1,02315
22	5	2	2	2	2	2	2	2	-2,79030	0,77631
23	4	3	3	3	3	4	3	3	-0,32487	-1,74345
24	4	5	5	5	4	5	4	5	1,78800	-1,81943
25	5	5	5	4	5	5	5	5	1,41149	0,98474

Εικόνα 7.6

!!! *Η διαδικασία παραγοντικής ανάλυσης την οποία ακολουθήσαμε είναι η πλέον κλασική και τυπική. Όπως ήδη αναφέρθηκε στην αρχή της παραγράφου το S.P.S.S προσφέρει μία μεγάλη ποικιλία επιλογών σχετικά με μεθόδους εξαγωγής παραγόντων, μεθόδους περιστροφής και εξαγωγής παραγοντικών σκορ. Εκείνο όμως το οποίο πρέπει ο κάθε αναλυτής να προσέξει, ανεξαρτήτως μεθόδου, είναι σχολαστικός έλεγχος των δεικτών εκείνων οι οποίοι μας πληροφορούν για την εγκυρότητα και επάρκεια των αποτελεσμάτων.*

!!! *Η επικύρωση των αποτελεσμάτων της παραγοντικής ανάλυσης είναι πολύ σημαντική στον καθορισμό της δομής των μεταβλητών. Για το λόγο αυτό μπορούμε, αφού πρώτα ολοκληρώσουμε τη διαδικασία*

ακολουθώντας τη μέθοδο που επιλέξαμε, να δοκιμάσουμε στη συνέχεια κάποια άλλη μέθοδο από τις προσφερόμενες έτσι ώστε να αντιπαραθέσουμε τα αποτελέσματα. Πολλές φορές, όταν τα δεδομένα είναι πολλά, συνηθίζεται το σπάσιμο των δεδομένων σε δύο τυχαίου μεγέθους δείγματα (ή και ισομεγέθη) και στη συνέχεια η αντιπαράθεση των αποτελεσμάτων της ανάλυσης των δύο δειγμάτων μεταξύ τους και με τα αρχικά. Με τον τρόπο αυτό ελέγχουμε τη σταθερότητα των αποτελεσμάτων.

7.5 Ανάλυση Αξιοπιστίας (Reliability Analysis)

Μετά την οριστικοποίηση των παραγόντων, με την ανάλυση αξιοπιστίας ελέγχεται η εσωτερική συνοχή του κάθε παράγοντα. Το μέτρο το οποίο χρησιμοποιείται περισσότερο για την εκτίμηση της αξιοπιστίας είναι ο δείκτης ***a*** του ***Cronbach***. Οι τιμές του δείκτη κυμαίνονται από 0 έως 1, ενώ τιμές μεγαλύτερες του 0,7 δείχνουν ικανοποιητική συνοχή και αξιόπιστο παράγοντα. Για τον υπολογισμό του δείκτη χρησιμοποιούμε τον τύπο:

$$\alpha = \frac{\mathbf{v}}{\mathbf{v}-\mathbf{1}}\left(\mathbf{1} - \frac{\sum_{\mathbf{i=1}}^{\mathbf{v}} \boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{Yi}}}{\boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{X}}}\right)$$, όπου **v** είναι το πλήθος των μεταβλητών του παράγοντα, $\boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{X}}$ η διακύμανση του συνόλου των παρατηρηθέντων τιμών ελέγχου και $\boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{Yi}}$ η διακύμανση της **i** μεταβλητής.

Εναλλακτικά, ο τυποποιημένος ***Cronbach's α*** μπορεί επίσης να οριστεί ως: $\alpha = \frac{\mathbf{v}\cdot\bar{\mathbf{r}}}{\mathbf{1}+(\mathbf{v}-\mathbf{1})\cdot\bar{\mathbf{r}}}$, όπου **v** είναι το πλήθος των

μεταβλητών του παράγοντα και $\bar{r}$ η μέση συσχέτιση μεταξύ όλων των μεταβλητών του παράγοντα.

!!! *Ο δείκτης* ***α*** *του* ***Cronbach*** *υπολογίζεται όταν έχουμε τουλάχιστον δύο μεταβλητές και αυξάνεται όταν η συσχέτιση μεταξύ των μεταβλητών του παράγοντα αυξάνει.*

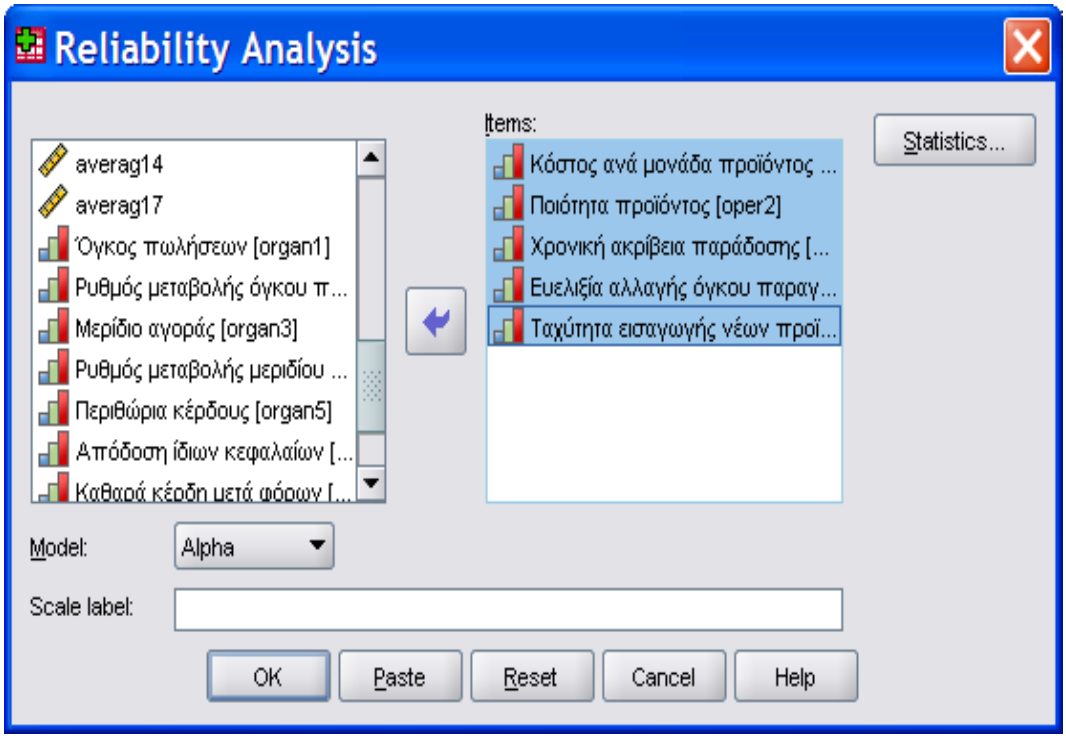
!!! *Τα δεδομένα μπορεί να είναι διχοτομικά (dichotomous), ιεραρχικής κλίμακας ή διαστημικής κλίμακας αλλά οπωσδήποτε κωδικοποιημένα με αριθμούς.*

!!! *Οι παρατηρήσεις θα πρέπει να είναι ανεξάρτητες και κάθε ζεύγος μεταβλητών πρέπει να προσεγγίζει την διμεταβλητή κανονική κατανομή.*

7.5.1 Διαδικασία Ανάλυσης Αξιοπιστίας

Για να υπολογίσουμε τον δείκτη αξιοπιστίας με τη βοήθεια του S.P.S.S ακολουθούμε την παρακάτω διαδικασία:

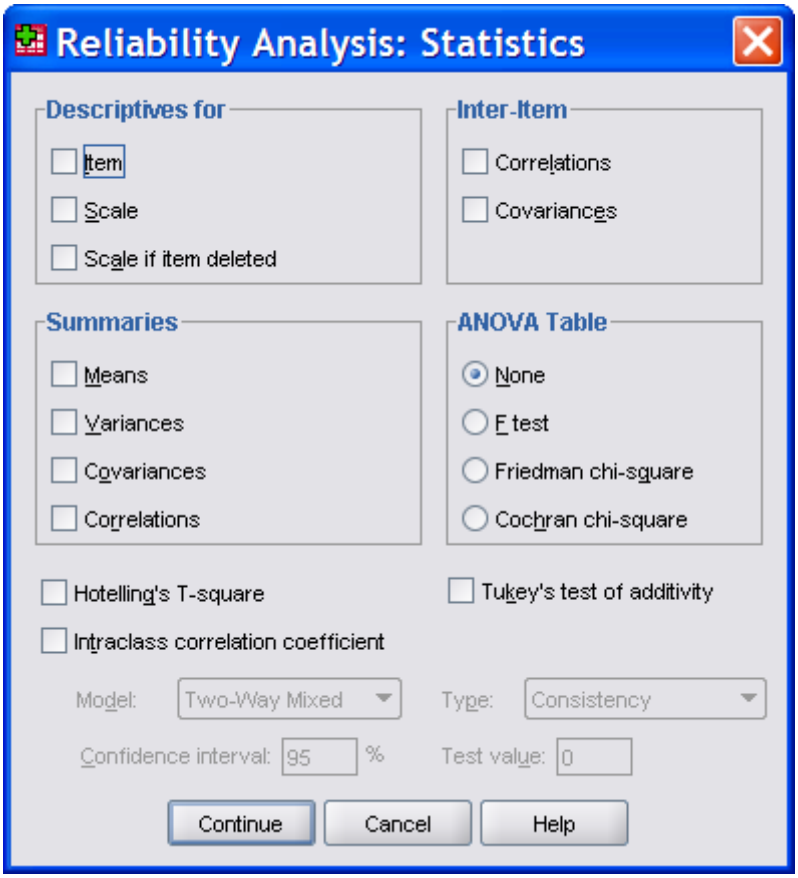
- ✓ Από το μενού ***Analyze*** του ***Data Editor*** επιλέγουμε ***Scale*** και στη συνέχεια ***Reliability Analysis***. Προκύπτει η επόμενη φόρμα.



Εικόνα 7.7

✓ Από το αριστερό παράθυρο, μεταφέρουμε στο δεξί παράθυρο τις μεταβλητές που ανήκουν στον ίδιο παράγοντα. Στο συγκεκριμένο παράδειγμα έχουμε 5 μεταβλητές οι οποίες αποτελούν τον παράγοντα «Λειτουργική απόδοση».

✓ Στη θέση ***Model*** επιλέγουμε ***Alpha*** για να πάρουμε την τιμή του δείκτη ***α*** και στη συνέχεια πατάμε στο κουμπί ***Statistics*** και έχουμε την επόμενη φόρμα.



Εικόνα 7.8

Στην εικόνα 7.8 έχουμε πολλές επιλογές οι οποίες θα αναλυθούν στη συνέχεια.

- Στην περιοχή ***Descriptives for*** οι επιλογές είναι:

1. *Item*. Παράγει περιγραφικά στατιστικά μέτρα για όλες της μεταβλητές.

2. *Scale*. Παράγει περιγραφικά στατιστικά μέτρα για τον παράγοντα.

3. *Scale if item deleted*. Περιέχει συνοπτικά στατιστικά μέτρα συγκρίνοντας κάθε μεταβλητή με τον παράγοντα που αποτελείται από όλες τις υπόλοιπες μεταβλητές. Τα στατιστικά περιέχουν την μέση τιμή του παράγοντα και την διακύμανση αν μία μεταβλητή διαγραφεί από τον παράγοντα, την συσχέτιση μεταξύ των μεταβλητών και του

παράγοντα που αποτελείται από τις υπόλοιπες μεταβλητές και το δείκτη ***α*** αν μία μεταβλητή διαγραφεί από τον παράγοντα.

- Στην περιοχή ***summaries*** οι επιλογές είναι:

1. ***Means.*** Παρέχει συνοπτικά στατιστικά μέτρα για τους μέσους των μεταβλητών. Παρέχει επίσης, την μικρότερη, την μεγαλύτερη και τον μέσο όρο των αριθμητικών μέσων των μεταβλητών, το εύρος και την διακύμανση των μέσων των μεταβλητών και τον λόγο του μεγαλύτερου προς τον μικρότερο αριθμητικό μέσο των μεταβλητών .

2. ***Variances***. Παρέχει συνοπτικά στατιστικά μέτρα για την διακύμανση των μεταβλητών. Αναλυτικότερα, την μικρότερη, την μεγαλύτερη και την μέση διακύμανση των μεταβλητών. Το εύρος και την διακύμανση των διακυμάνσεων των μεταβλητών και τον λόγο της μεγαλύτερης προς την μικρότερη διακύμανση των μεταβλητών.

3. ***Covariances***. Παρέχει συνοπτικά στατιστικά μέτρα για την συνδιακύμανση μεταξύ των μεταβλητών. Αναλυτικότερα, την μικρότερη, την μεγαλύτερη και την μέση συνδιακύμανση. Το εύρος και την διακύμανση των συνδιακυμάνσεων και τον λόγο της μεγαλύτερης προς την μικρότερη συνδιακύμανση των μεταβλητών.

4. ***Correlations***. Παρέχει συνοπτικά στατιστικά μέτρα για την συσχέτιση μεταξύ των μεταβλητών. Αναλυτικότερα, την μικρότερη, την μεγαλύτερη και την μέση συσχέτιση. Το εύρος και την διακύμανση των συσχετίσεων και τον λόγο της μεγαλύτερης προς την μικρότερη συσχέτιση των μεταβλητών.

- Στην περιοχή ***Inter-Item*** οι επιλογές είναι:

1. *Correlations:* Παράγει τον πίνακα συσχετίσεων μεταξύ των μεταβλητών.

2. *Covariances*: Παράγει τον πίνακα συνδιακυμάνσεων μεταξύ των μεταβλητών.

- Στην περιοχή ***ANOVA Table*** οι επιλογές είναι:

1. *None*: Όταν δεν θέλουμε τίποτα από αυτή την περιοχή

2. *F- test*: Παρέχει τον πίνακα ανάλυσης διακύμανσης.

3. *Friedman chi-square*: Παρέχει τον χ^2 συντελεστή ***συμφωνίας-αρμονίας (concordance)*** του ***Friedman*** και του ***Kendall***. Αυτή η επιλογή είναι κατάλληλη για δεδομένα που είναι σε μορφή ***τάξης (form of ranks)***. Ο χ^2 έλεγχος αντικαθιστά τον συνηθισμένο F έλεγχο του πίνακα ANOVA.

4. *Cochran chi-square*: Παρέχει τον δείκτη ***Q*** του ***Cochran***. Αυτή η επιλογή είναι κατάλληλη για δεδομένα διχοτομικά. Το στατιστικό Q αντικαθιστά τον συνηθισμένο F έλεγχο του πίνακα ANOVA.

- Επιπλέον έχουμε τις επιλογές:

1. *Hotelling's T-square*: Παράγει έναν πολυμεταβλητό έλεγχο για την μηδενική υπόθεση της ισότητας των μέσων τιμών όλων των μεταβλητών του παράγοντα.

2. *Tukey's test of additivity:* Παράγει έναν έλεγχο για την υπόθεση της μη ύπαρξης πολλαπλασιαστικής αλληλεπίδρασης μεταξύ των μεταβλητών του παράγοντα.

3. *Intraclass correlation coefficient*: Παράγει μέτρα συνοχής ή συμφωνίας των τιμών μεταξύ των περιπτώσεων.

4. ***Model***: Επιλέγεις το μοντέλο με το οποίο θα υπολογισθούν οι ενδοταξικοί συντελεστές συσχέτισης. Τα διαθέσιμα μοντέλα είναι το ***Two-Way Mixed, Two-Way Random,*** και ***One-Way Random.*** Επιλέγουμε ***Two-Way Mixed*** όταν η επίδραση του ατόμου είναι τυχαία και η επίδραση της μεταβλητής σταθερή, επιλέγουμε ***Two-Way Random*** όταν η επίδραση του ατόμου και η επίδραση της μεταβλητής είναι τυχαία και τέλος επιλέγουμε ***One-Way Random*** όταν η επίδραση του ατόμου είναι τυχαία.

5. ***Type***: Επιλέγεις τον τύπο του δείκτη. Διαθέσιμοι τύποι είναι οι ***συνοχής (Consistency)*** και ***απόλυτης συμφωνίας (Absolute Agreement)***.

6. ***Confidence interval***: Καθορίζεις το επίπεδο του διαστήματος εμπιστοσύνης. Το προκαθορισμένο είναι 95%.

7. ***Test value***: Καθορίζεις την υποθετική τιμή του συντελεστή για τον έλεγχο της υπόθεσης. Αυτή είναι η τιμή με την οποία συγκρίνεται η παρατηρηθείσα (observed) τιμή. Η προκαθορισμένη τιμή είναι το 0.

Αφού κάνουμε τις επιλογές που θεωρούμε απαραίτητες, για τις ανάγκες της ανάλυσης, πατάμε στο κουμπί ***Continue*** της εικόνας 7.8, επιστρέφουμε στην εικόνα 7.7 και με O.K θα έχουμε τους πίνακες που ζητήσαμε.

Συνήθως, ζητάμε τον δείκτη ***α*** του ***Cronbach*** και επιπλέον τσεκάρουμε και ***Scale if item deleted*** στην περιοχή ***Descriptives for*** για να διαπιστώσουμε ποια ή ποιες μεταβλητές δεν συνεισφέρουν ικανοποιητικά στη διαμόρφωση της τιμής του δείκτη ***α***. Μετά από αυτές τις επιλογές έχουμε τους παρακάτω πίνακες:

Τον πίνακα ***Reliability Statistics*** (Πίνακας 7.9) ο οποίος δίνει την τιμή του δείκτη ***a*** του ***Cronbach*** και το πλήθος **N** των μεταβλητών. Στο συγκεκριμένο παράδειγμα ο δείκτης είναι 0,739>0,7 και συνεπώς η εσωτερική συνοχή του παράγοντα «Λειτουργική Απόδοση» είναι ικανοποιητική.

Πίνακας 7.9: Reliability Statistics

Cronbach's Alpha	N of Items
,739	5

Τον πίνακα ***Item-Total Statistics*** (Πίνακας 7.10) ο οποίος στην πρώτη στήλη αναφέρει τις μεταβλητές του παράγοντα.

Στην δεύτερη στήλη δίνεται η μέση τιμή του παράγοντα αν διαγραφεί η μεταβλητή της αντίστοιχης γραμμής.

Στην τρίτη στήλη υπάρχει η διακύμανση του παράγοντα αν διαγραφεί η μεταβλητή της αντίστοιχης γραμμής.

Στην τέταρτη στήλη δίνεται η διορθωμένη συσχέτιση της μεταβλητής προς την συνολική. Όταν οι τιμή σε αυτή την στήλη, για κάποια ή κάποιες μεταβλητές είναι μικρότερη του 0,5 τότε σωστό είναι η μεταβλητή αυτή να αποβάλλεται από τον παράγοντα. Έτσι, θα βελτιωθεί σημαντικά ο δείκτης ***α***. Στο συγκεκριμένο παράδειγμα, αν διαγράψουμε την μεταβλητή ***«Κόστος ανά μονάδα προϊόντος»*** που έχει δείκτη ***Corrected Item- Total Correlation = 0,368<0,5*** ο δείκτης ***α*** θα γίνει 0,746 σαφώς μεγαλύτερος από τον προηγούμενο.

Στην πέμπτη στήλη υπάρχει ο δείκτης ***α*** μετά την διαγραφή της μεταβλητής της αντίστοιχης γραμμής. Παρατηρούμε ότι αν αποβάλλουμε την μεταβλητή που προαναφέραμε, η τιμή του δείκτη α θα γίνει 0,746. Σε κάθε άλλη περίπτωση αποβολής μεταβλητής ο δείκτης θα μειωθεί σε σχέση με την αρχική του τιμή (0,739).

Πίνακας 7.10: Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Κόστος ανά μονάδα προϊόντος	16,63	6,095	,368	,746
Ποιότητα προϊόντος	15,85	6,368	,518	,696
Χρονική ακρίβεια παράδοσης	16,12	5,859	,512	,690
Ευελιξία αλλαγής όγκου παραγωγής	16,36	5,281	,607	,651
Ταχύτητα εισαγωγής νέων προϊόντων	16,41	5,324	,541	,679

!!! Ο δείκτης ***α*** του παράγοντα ***«Χρηματοοικονομική Απόδοση»*** του παραδείγματός μας έδωσε τιμή 0,883 σαφώς καλύτερη από την αντίστοιχη του παράγοντα ***«Λειτουργική Απόδοση»***.