

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Πολλές παραμετρικές στατιστικές μέθοδοι ελέγχου υποθέσεων στηρίζονται στην υπόθεση ότι η μεταβλητότητα των δεδομένων περιγράφεται από μία κατανομή συγκεκριμένης μορφής (για παράδειγμα, κανονική ή κάποια άλλη κατανομή που δεν απέχει πολύ από την κανονική). Έτσι, στους ελέγχους υποθέσεων που αναφέρονται στην μέση τιμή ενός πληθυσμού, γίνεται η υπόθεση ότι ο δειγματικός μέσος κατανέμεται κανονικά. Στους ελέγχους υποθέσεων που αναφέρονται στις παραμέτρους του γραμμικού μοντέλου, ο προσθετός που αντιστοιχεί στο τυχαίο σφάλμα θεωρείται ότι κατανέμεται σύμφωνα με την κανονική κατανομή. Τέλος, στους ελέγχους υποθέσεων στο πλαίσιο προβλημάτων ανάλυσης διασποράς, υποθέτουμε ότι η εντός των επιδράσεων μεταβλητότητα περιγράφεται από την κανονική κατανομή. Οποτεδήποτε δεν είναι επιτρεπτή η οποιαδήποτε υπόθεση για την μορφή του πληθυσμού, ο ερευνητής στρέφεται στο Κεντρικό Οριακό Θεώρημα και σε εμπειρικές μελέτες για την θεωρητική και εμπειρική υποστήριξη αυτής της "βολικής" υπόθεσης κανονικότητας προκειμένου να χρησιμοποιήσει μια παραμετρική μέθοδο. Στην πράξη, όμως, εμφανίζονται συχνά περιπτώσεις στις οποίες τα δείγματα είναι μικρά και τα δεδομένα κατανέμονται εμφανώς μη κανονικά. Εμφανίζονται επίσης περιπτώσεις, όπου και αν ακόμα υπάρχει κάποια ένδειξη κανονικότητας, ο ερευνητής που ενδεχομένως έχει βαθιά γνώση του πληθυσμού, είναι επιφυλακτικός στο να κάνει μια τέτοια υπόθεση.

Οι στατιστικοί επενόησαν αρκετές εναλλακτικές τεχνικές για τον ερευνητή ο οποίος διστάζει να υποθέσει κανονικότητα. Οι τεχνικές αυτές δεν απαιτούν συγκεκριμένες υποθέσεις για την μορφή του πληθυσμού από τον οποίο προέρχονται τα δεδομένα και μπορούν να χρησιμοποιηθούν τόσο για μικρά όσο και για μεγάλα δείγματα. Είναι δηλαδή σχεδιασμένες για να μπορούν να χρησιμοποιηθούν ανεξάρτητα από την κατανομή των δεδομένων. Επομένως, οι τεχνικές αυτές μπορούν να χρησιμοποιηθούν τόσο για κανονικούς πληθυσμούς όσο και για μη κανονικούς πληθυσμούς. Για τον λόγο αυτό, οι τεχνικές αυτές ονομάζονται *ελεύθερες κατανομών* ή μη παραμετρικές τεχνικές (*distribution-free* ή *non-parametric techniques*).

Οι μη παραμετρικές τεχνικές είναι πολύ απλές στην χρήση τους εν γένει, και ο σχεδιασμός τους είναι αποτέλεσμα στοιχειωδών θεωρήσεων. Εάν τα δεδομένα ακολουθούν στην πραγματικότητα την κανονική κατανομή, τότε οι μη παραμετρικοί έλεγχοι υποθέσεων δεν είναι το ίδιο ισχυροί όπως οι αντίστοιχοι παραμετρικοί έλεγχοι, οι οποίοι κάνουν χρήση της υπόθεσης της κανονικότητας. Για δεδομένη πιθανότητα σφάλματος τύπου I, οι μη παραμετρικοί έλεγχοι έχουν υψηλότερη πιθανότητα σφάλματος τύπου II. Ένας έλεγχος, ο οποίος αγνοεί πληροφορίες σχετικά με τα δεδομένα, όπως είναι η μορφή της κατανομής τους, δεν αναμένεται να είναι το ίδιο καλός όπως ένας έλεγχος ο οποίος κάνει χρήση αυτής της πληροφορίας. Έτσι, ένας έλεγχος, ο οποίος δεν λαβαίνει υπόψη του ότι τα δεδομένα προέρχονται από μια κανονική κατανομή, δεν αναμένεται να είναι το ίδιο ισχυρός όπως ένας έλεγχος ο οποίος χρησιμοποιεί αυτή την υπόθεση. Από το άλλο μέρος, εάν τα δεδομένα προέρχονται από μη κανονικό πληθυσμό, τότε οι μη παραμετρικοί έλεγχοι έχουν ένα σαφές πλεονέκτημα έναντι των παραμετρικών ελέγχων, οι οποίοι στηρίζονται

στην εσφαλμένη υπόθεση της κανονικότητας των δεδομένων. Η σοβαρότητα του σφάλματος και, επομένως, η ακρίβεια των παραμετρικών ελέγχων εξαρτάται, κατά συνέπεια, από το πόσο εσφαλμένη είναι η υπόθεση της κανονικότητας. Επειδή οι μη παραμετρικοί έλεγχοι στηρίζονται σε ελάχιστες υποθέσεις για τους πληθυσμούς από τους οποίους προέρχονται τα δεδομένα και, πάντως, όχι σε υποθέσεις οι οποίες αναφέρονται στην μορφή των πληθυσμών αυτών, είναι εν γένει πολύ *ευσταθείς (robust)*.

Συνοψίζοντας, οι μη παραμετρικές μέθοδοι α) αποβλέπουν σε ευρύτερα πεδία εφαρμογής λόγω του ότι οι κατανομές στις οποίες αναφέρονται είναι λιγότερο περιορισμένες από ό,τι στα αντίστοιχα παραμετρικά προβλήματα, β) δεν είναι εξίσου ισχυρές με τις αντίστοιχες παραμετρικές μεθόδους και γ) είναι περισσότερο ευσταθείς επειδή ακριβώς δεν επηρεάζονται από την μορφή της κατανομής των δεδομένων. Παρ' όλα αυτά, οι μη παραμετρικές μέθοδοι συχνά είναι σχεδόν το ίδιο αποτελεσματικές όπως οι παραμετρικές μέθοδοι οι οποίες κάνουν αυστηρές υποθέσεις για τον πληθυσμό από τον οποίο προέρχονται τα δεδομένα. Ένα άλλο σημαντικό πλεονέκτημα των μη παραμετρικών μεθόδων είναι ότι μπορούν να εφαρμοθούν σε δεδομένα που είναι ταξινομημένα σε κατηγορίες (κατηγορικά δεδομένα) και τα οποία είναι σε κλίμακα διάταξης ή ακόμα και απλώς σε ονομαστική κλίμακα, ενώ οι παραμετρικές μέθοδοι προϋποθέτουν ακριβείς μετρήσεις. Τέλος, θα πρέπει να σημειωθεί ότι οι μη παραμετρικές μέθοδοι μπορούν να θεωρηθούν ως *προπαρασκευαστικές* για τις παραμετρικές μεθόδους, με την έννοια ότι, η χρησιμοποίηση μιας παραμετρικής μεθόδου, η οποία βασίζεται στην υπόθεση της κανονικότητας, θα πρέπει να

έπεται ενός ελέγχου, με μία μη παραμετρική μέθοδο, της υπόθεσης ότι τα δεδομένα έχουν προέλθει από μια κανονική κατανομή.

1.2 ΧΡΗΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΠΑΚΕΤΩΝ ΣΤΗΝ ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

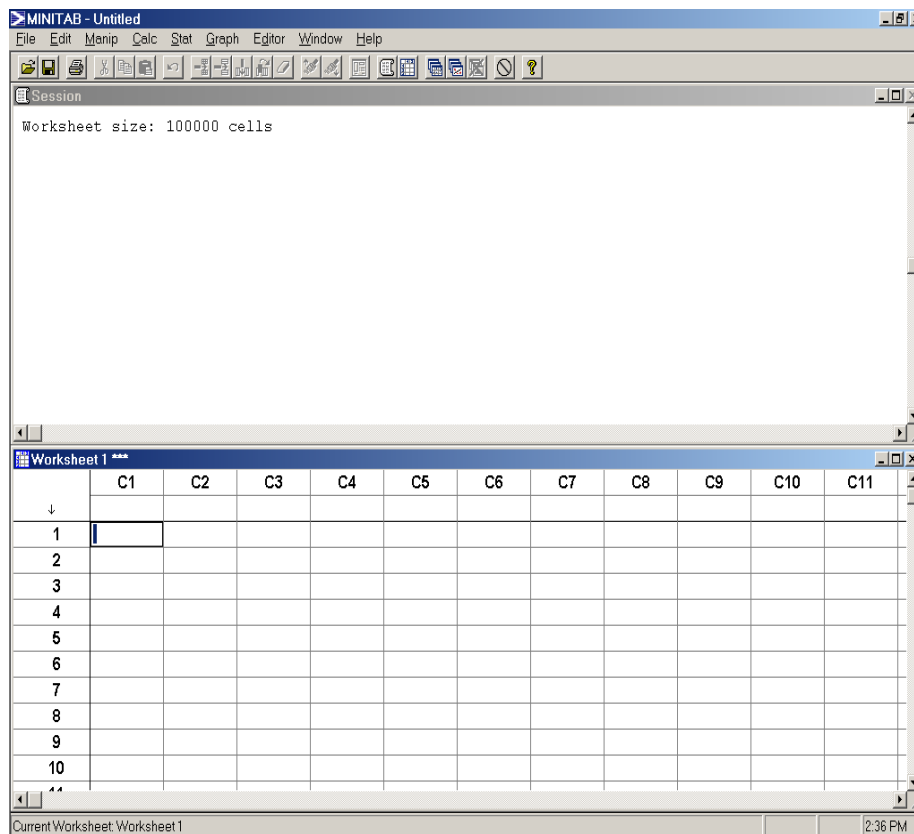
Η χρησιμοποίηση των μεθόδων της Μη Παραμετρικής Στατιστικής διευκολύνεται εξαιρετικά με την εφαρμογή τους μέσω στατιστικών πακέτων που, στον τομέα της στατιστικής ανάλυσης στοιχείων, έχουν εξελιχθεί τα τελευταία χρόνια σε απαραίτητα εργαλεία, απλά και εύκολα στην χρήση τους η οποία σήμερα πια δεν προϋποθέτει γνώση της πληθώρας των εντολών που οι πρώτες εκδόσεις τους απαιτούσαν.

1.2.1 Μη Παραμετρικές Μέθοδοι με το MINITAB

Το πακέτο MINITAB θεωρείται ως ένα από τα ευκολότερα στην χρήση τους στατιστικά πακέτα με πολλές δυνατότητες. Το MINITAB μπορεί να χρησιμοποιηθεί τόσο σε μεγάλα συστήματα όσο και σε συστήματα μικροϋπολογιστών. Η φιλοσοφία του MINITAB είναι στο ότι αποκτά κανείς διέξοδο στην δομή των οδηγιών (command structure) μέσω του καταλόγου επιλογών (μενού, menu) με τον ίδιο τρόπο όπως μέσω πληκτρολογημένων εντολών.

Στα επόμενα, παρουσιάζονται οι βασικές εντολές χειρισμού της έκδοσης 12.0 του πακέτου MINITAB, στην οποία έχει γίνει και η επεξεργασία των λύσεων εκείνων των παραδειγμάτων του βιβλίου που δίνονται με το πακέτο αυτό.

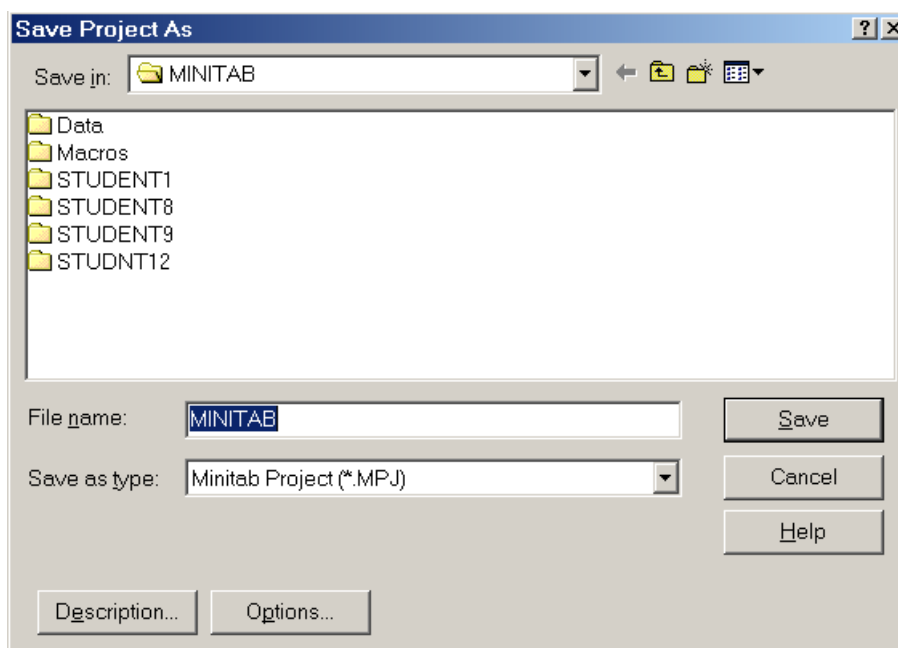
1. Εισαγωγή δεδομένων: Η οθόνη εργασίας του MINITAB χωρίζεται σε δύο τμήματα: Το επάνω τμήμα ονομάζεται **session window** και μοιάζει με ένα λευκό φύλλο χαρτί όπου εμφανίζονται τα αποτελέσματα των αναλύσεων καθώς και τα μηνύματα που μας δίνει το πρόγραμμα. Το κάτω τμήμα ονομάζεται **worksheet** και μοιάζει με ένα φύλλο εργασίας του Excel. Εκεί εισάγονται τα δεδομένα. Το σχήμα που ακολουθεί απεικονίζει την μορφή της οθόνης εργασίας.



Τα δεδομένα καταχωρίζονται σε στήλες. Κάθε στήλη συμβολίζει μία μεταβλητή, ενώ κάθε γραμμή συμβολίζει μία δειγματοληπτική μονάδα, στην οποία έχουμε κάνει μετρήσεις. Στην μη αριθμημένη γραμμή μπορούμε να πληκτρολογήσουμε το όνομα της μεταβλητής.

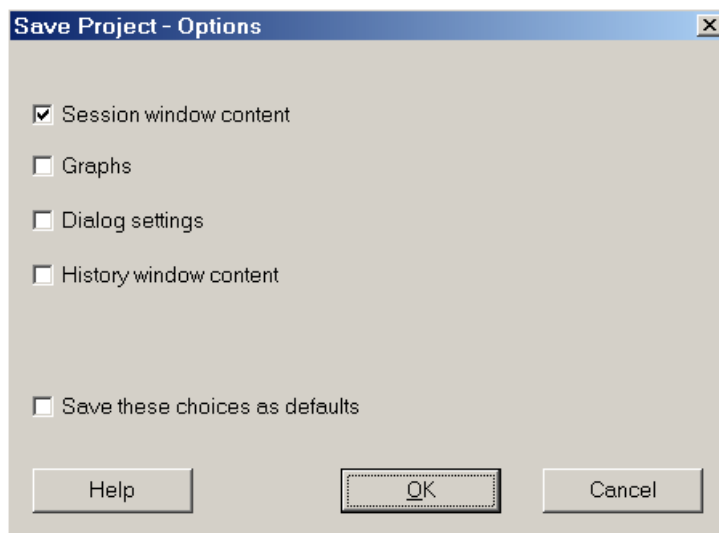
2. Χειρισμός δεδομένων: Ο χειρισμός των δεδομένων είναι πολύ απλός. Μπορούμε να πληκτρολογήσουμε τιμές, να δώσουμε ονόματα στις μεταβλητές, να διαγράψουμε στήλες ή γραμμές όπως και σε άλλα προγράμματα με φύλλα δεδομένων. Για να πληκτρολογήσουμε μία τιμή σε ένα «κελλί», το επιλέγουμε με το ποντίκι και πληκτρολογούμε την τιμή. Μετά πιέζουμε Enter ή επιλέγουμε κάποιο άλλο κελλί. Στην παραπάνω οθόνη είναι επιλεγμένο το κελλί (γραμμή 1, στήλη C1). Για να διαγράψουμε μία γραμμή, επιλέγουμε με το ποντίκι τον αριθμό της, πιέζουμε το δεξί πλήκτρο του ποντικιού και επιλέγουμε **Delete Cells** στον κατάλογο που εμφανίζεται.

Για να αποθηκεύσουμε τα δεδομένα μας, επιλέγουμε **File, Save project**. Αυτό οδηγεί στο επόμενο πλαίσιο διαλόγου.



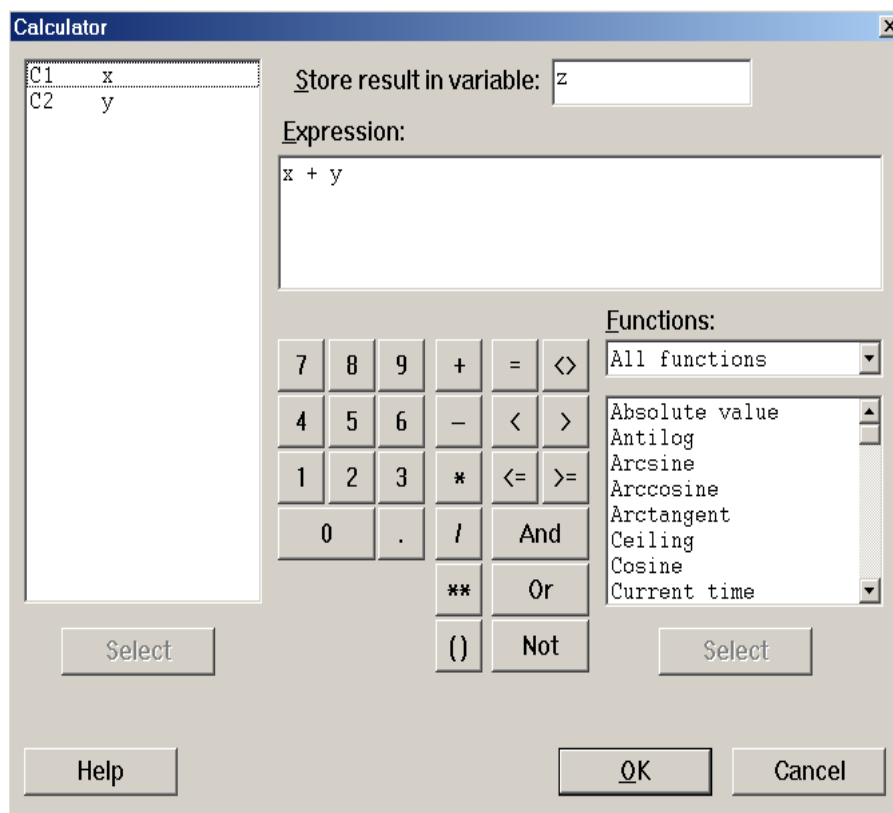
Στο πλαίσιο αυτό επιλέγουμε πού θα αποθηκευθεί το φύλλο δεδομένων και με ποιο όνομα. Επιλέγοντας **Options**, μπορούμε να

δηλώσουμε στο MINITAB πώς θα αποθηκευθούν τα δεδομένα μας. Αυτό γίνεται στο παρακάτω πλαίσιο διαλόγου:



Αν δεν επιλέξουμε καμία από τις τέσσερις διαθέσιμες επιλογές στο πάνω τμήμα του πλαισίου, θα αποθηκευθούν μόνο τα δεδομένα. Είναι καλό να αποθηκεύουμε και το session window, αν αυτό περιέχει αποτελέσματα αναλύσεων, αλλά και γραφήματα που ενδεχομένως υπάρχουν. Αυτό γίνεται επιλέγοντας **Session window content** και **Graphs** αντίστοιχα.

Μπορούμε, επίσης, να κάνουμε απλούς υπολογισμούς με μεταβλητές και να δημιουργήσουμε νέες μεταβλητές: Ας υποθέσουμε ότι έχουμε δύο μεταβλητές με ονόματα x και y και ότι θέλουμε να υπολογίσουμε μία νέα μεταβλητή z που είναι το άθροισμα των δύο πρώτων μεταβλητών. Επιλέγουμε **Calc, Calculator** και προκύπτει το ακόλουθο πλαίσιο διαλόγου.



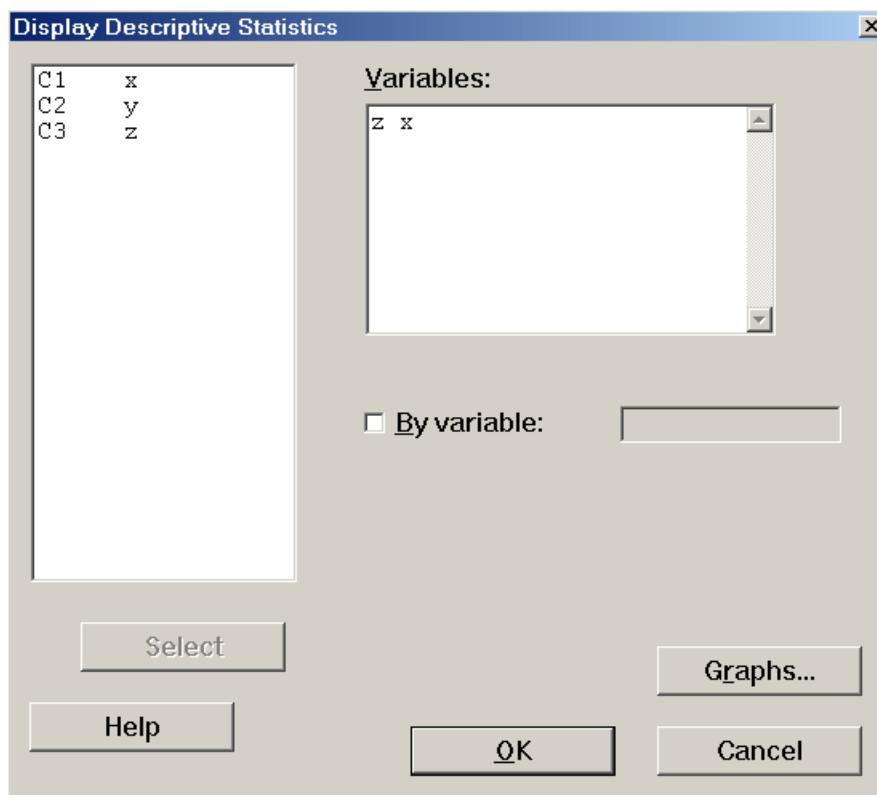
Στο πεδίο **Store result in variable**, δηλώνουμε το όνομα της μεταβλητής, στην οποία θα αποθηκευθεί το αποτέλεσμα των υπολογισμών. Αν αυτή υπάρχει ήδη, οι τιμές της θα αντικατασταθούν από το αποτέλεσμα των υπολογισμών (το MINITAB θα μας ρωτήσει αν επιθυμούμε να γίνει αυτό). Στην αντίθετη περίπτωση, η μεταβλητή θα δημιουργηθεί στην πρώτη κενή στήλη του φύλλου δεδομένων. Στο πεδίο **Expression**, δηλώνουμε τον μαθηματικό τύπο που θέλουμε να υπολογισθεί. Αυτός μπορεί να περιλαμβάνει οποιεσδήποτε μεταβλητές υπάρχουν ήδη στο φύλλο δεδομένων (εμφανίζονται στο αριστερό μέρος του παραθύρου), οποιουσδήποτε μαθηματικούς τελεστές εμφανίζονται στο κέντρο του πλαισίου, καθώς και οποιαδήποτε μαθηματική συνάρτηση υπάρχει στο πεδίο **Functions**. Αφού

πληκτρολογήσουμε την μαθηματική έκφραση, πιέζουμε **OK** και υπολογίζεται η νέα μεταβλητή.


Σημείωση: όταν κάνουμε υπολογισμούς μεταξύ μεταβλητών πρέπει αυτές να έχουν τον ίδιο αριθμό παρατηρήσεων. Σε αντίθετη περίπτωση, το αποτέλεσμα θα έχει τόσες παρατηρήσεις, όσες η μεταβλητή με τις λιγότερες παρατηρήσεις. Οι υπόλοιπες θέσεις, μέχρι να συμπληρωθεί πλήθος παρατηρήσεων ίσο με αυτές της μεταβλητής με τις περισσότερες, θα θεωρηθούν ως ελλείποντες τιμές (missing values).

3. Περιγραφικά στατιστικά μέτρα: Το πακέτο MINITAB παρέχει την δυνατότητα υπολογισμού απλών περιγραφικών στατιστικών μέτρων για μία ή περισσότερες μεταβλητές.

Όλες οι στατιστικές αναλύσεις είναι διαθέσιμες στην επιλογή **Stat**. Για τον υπολογισμό περιγραφικών στατιστικών μέτρων επιλέγουμε κατά σειρά **Stat, Basic Statistics, Display Descriptive Statistics** και οδηγούμεθα στο εξής πλαίσιο διαλόγου.



Στο πεδίο **Variables** δηλώνουμε τις μεταβλητές για τις οποίες θέλουμε να υπολογίσουμε περιγραφικά στατιστικά μέτρα. Το MINITAB μας δίνει όλα τα συνηθισμένα περιγραφικά στατιστικά (μέσο, διακύμανση, τυπική απόκλιση, τεταρτημόρια, κ.λ.π.) αλλά δεν μπορούμε να επιλέξουμε να μας δίνει μόνο ορισμένα από αυτά).

4. Το πλήκτρο Undo: Σε περίπτωση που κάνουμε μία λάθος ενέργεια (π.χ., αν διαγράψουμε μία στήλη με δεδομένα που μας χρειάζονται), έχουμε τρόπο να διορθώσουμε το λάθος μας. Η πίεση του πλήκτρου Undo, δηλαδή του πλήκτρου , αναιρεί την τελευταία ενέργεια που κάναμε πριν το πιάσουμε. Αν το πιάσουμε κι άλλες φορές θα αναιρεθούν η προτελευταία ενέργεια, η αντιπροτελευταία ενέργεια,

κ.ο.κ. Αν το πλήκτρο είναι απενεργοποιημένο (έχει ένα ανοικτό γκρι χρώμα), τότε δεν υπάρχουν ενέργειες που μπορούν να αναιρεθούν.

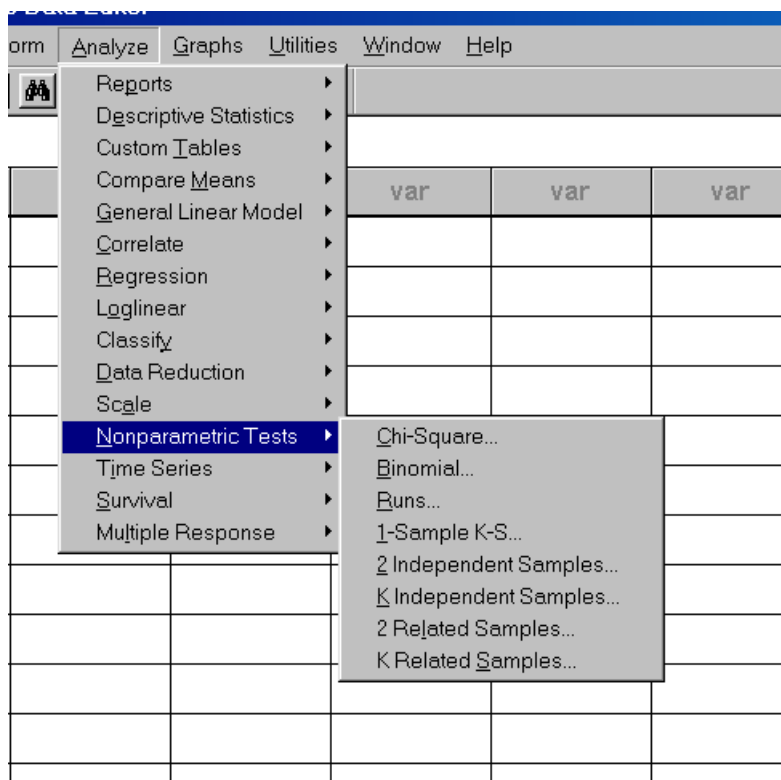
Σημείωση: Μία πράξη Undo δεν ακυρώνεται από ένα άλλο Undo.

1.2.2 Μη Παραμετρικές Μέθοδοι με το SPSS

Το πρόγραμμα SPSS (Statistical Package for the Social Sciences) κυκλοφορεί για πάρα πολλά χρόνια. Ξεκίνησε από τα συστήματα Unix, σε μη γραφικά περιβάλλοντα. Παλαιότερα, ο χρήστης έπρεπε να γνωρίζει τις εντολές του προγράμματος, τις οποίες πληκτρολογούσε σε μία γραμμή εντολών. Οι γραφικές παραστάσεις ήταν υποτυπώδεις και φτιάχνονταν με τη χρησιμοποίηση χαρακτήρων κειμένου.

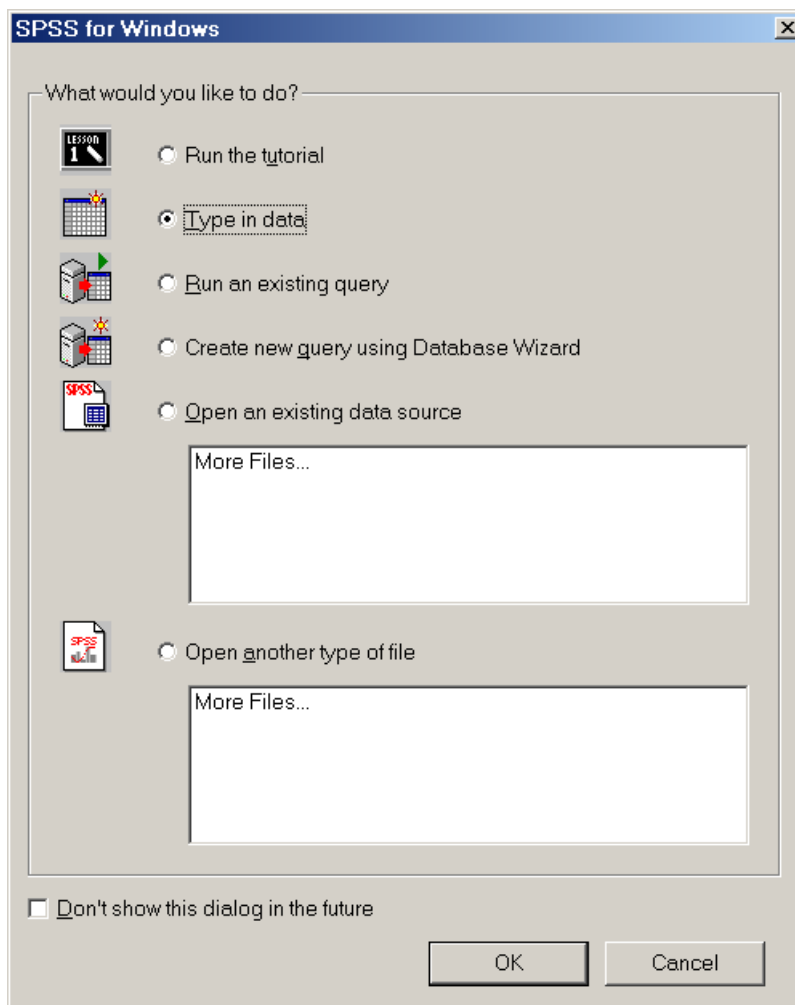
Τώρα πια, στις εκδόσεις για Microsoft Windows, το SPSS έχει αποκτήσει γραφικό περιβάλλον. Η πλειονότητα των διαθέσιμων εντολών είναι προσβάσιμη από ένα σύστημα επιλογών (μενού, menu) όπου ο χρήστης ελάχιστα χρειάζεται να πληκτρολογήσει. Τα γραφικά είναι πια πολυσύνθετα και περίτεχνα. Παράλληλα, υπάρχουν τεράστιες δυνατότητες διαμόρφωσης των γραφικών παραστάσεων κατά την βούλησή μας. Πρέπει να σημειωθεί ότι κάποιος μπορεί, αν το επιθυμεί, να πληκτρολογεί εντολές και να γράφει μέχρι και προγράμματα στην γλώσσα του SPSS. Αυτό, όμως, δεν είναι αναγκαίο να γίνει παρά μόνο για πολύ εξειδικευμένες αναλύσεις.

Οι περισσότεροι μη παραμετρικοί έλεγχοι στο SPSS βρίσκονται στον υποκατάλογο επιλογών (υπομενού, submenu) **Nonparametric Tests** του μενού **Analyze**.



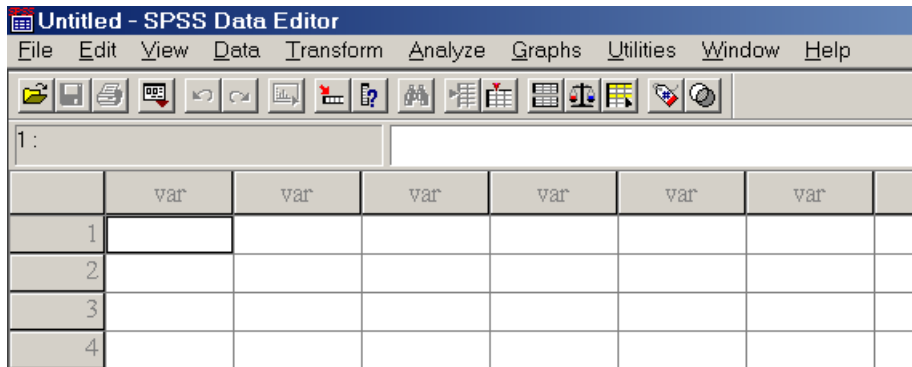
Στην συνέχεια, γίνεται μία συνοπτική παρουσίαση των βασικών εντολών χειρισμού της έκδοσης 10.0 του προγράμματος στατιστικών αναλύσεων SPSS με την οποία έχει γίνει η επεξεργασία των στοιχείων όσων παραδειγμάτων του βιβλίου λύνονται με το πακέτο αυτό.

1. Εισαγωγή δεδομένων. Το SPSS διαθέτει ένα παράθυρο εισαγωγής και εμφάνισης των δεδομένων παρόμοιο με ένα φύλλο του Excel. Κατά την εκκίνηση το πρόγραμμα, ενδέχεται να οδηγηθούμε (ανάλογα με τις ρυθμίσεις εκείνου που χρησιμοποίησε το πρόγραμμα πριν από εμάς) σε ένα πλαίσιο διαλόγου όπως το παρακάτω.



Η δεύτερη επιλογή οδηγεί σ' ένα κενό φύλλο δεδομένων, όπου μπορούμε να πληκτρολογήσουμε δεδομένα. Η προτελευταία επιλογή μας προτρέπει να ανοίξουμε ένα αρχείο δεδομένων και, μάλιστα, εμφανίζει τα αρχεία με τα οποία δουλέψαμε πρόσφατα. Η τελευταία επιλογή μας προτρέπει να ανοίξουμε ένα αρχείο που δημιουργήθηκε με κάποιο άλλο πρόγραμμα. Υπάρχει επίσης και η επιλογή να μην εμφανίζεται αυτό το πλαίσιο διαλόγου στις μελλοντικές εκκινήσεις του προγράμματος.

Επιλέγοντας **Type in data** εμφανίζεται το κενό φύλλο δεδομένων:



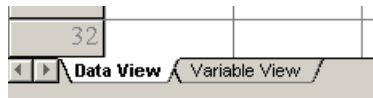
Όπως βλέπουμε, σε κάθε στήλη του φύλλου δεδομένων μπορούμε να αποθηκεύσουμε μια μεταβλητή (π.χ. ένα δείγμα). Κάθε γραμμή αντιπροσωπεύει μια παρατήρηση του δείγματος. Το παραπάνω παράθυρο λέγεται **Data View** και είναι η μία από τις δύο μορφές που μπορεί να έχει το παράθυρο δεδομένων.

Για να εισαγάγουμε δεδομένα, πρέπει, καταρχήν, να δώσουμε ονόματα και ιδιότητες στις μεταβλητές. Οι ιδιότητες καθορίζουν τον τύπο της μεταβλητής, το πλήθος των δεκαδικών ψηφίων της, την σημασία των τιμών της (π.χ. αν πρόκειται για κωδικές τιμές), κ.λ.π. Το όνομα μιας μεταβλητής δεν μπορεί να υπερβαίνει τους 8 χαρακτήρες.

Ας υποθέσουμε, για παράδειγμα ότι έχουμε ένα δείγμα μετρήσεων του ύψους ατόμων, ανδρών και γυναικών, τα οποία κατάγονται από την Ελλάδα, την Ιταλία και την Ισπανία. Για κάθε άτομο, έχουμε συνεπώς τρεις μεταβλητές (φύλο, χώρα και ύψος). Ας υποθέσουμε ότι οι μεταβλητές ονομάζονται **sex**, **country** και **height** αντίστοιχα. Οι μεταβλητές **sex** και **country** είναι κατηγορικές μεταβλητές, τις τιμές των οποίων μπορούμε να ορίσουμε ως εξής: 0

άνδρας, 1 γυναίκα, 0 Ελλάδα, 1 Ιταλία, 2 Ισπανία. Στην συνέχεια, θα δείξουμε πώς εισάγονται το όνομα και η σημασία των τιμών της μεταβλητής country. Η δήλωση των άλλων δύο αφήνεται ως άσκηση.

Για να δηλώσουμε μεταβλητές, πρέπει να μεταβούμε στην άλλη μορφή του παραθύρου δεδομένων που λέγεται **Variable View**. Αυτό γίνεται από το κάτω αριστερά μέρος του παραθύρου όπου εμφανίζονται οι δύο παρακάτω ετικέτες:



Επιλέγοντας την κατάλληλη ετικέτα με το ποντίκι, μεταφερόμεθα στην μορφή του παραθύρου που επιθυμούμε. Το παράθυρο **Variable View** έχει την μορφή:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1								
2								

Κάθε γραμμή αναφέρεται σε μία μεταβλητή και κάθε στήλη σε ένα χαρακτηριστικό των μεταβλητών. Για να δηλώσουμε την μεταβλητή country στο πρώτο κελί της πρώτης γραμμής, απλά πληκτρολογούμε το όνομα της μεταβλητής. Βλέπουμε τότε ότι όλα τα κελιά παίρνουν κάποια προεπιλεγμένη τιμή την οποία εμείς μπορούμε να αλλάξουμε αν θέλουμε:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	country	Numeric	8	2		None	None	8
2								
3								

Για να αλλάξουμε κάποια επιλογή, πηγαίνουμε στο αντίστοιχο κελλί, όπου δίπλα στην τρέχουσα επιλογή εμφανίζεται ένα πλήκτρο (όπως στο πεδίο **Type** στην παραπάνω εικόνα). Πιέζοντας το πλήκτρο, εμφανίζονται όλες οι διαθέσιμες επιλογές για το πεδίο και επιλέγουμε αυτή που θέλουμε.

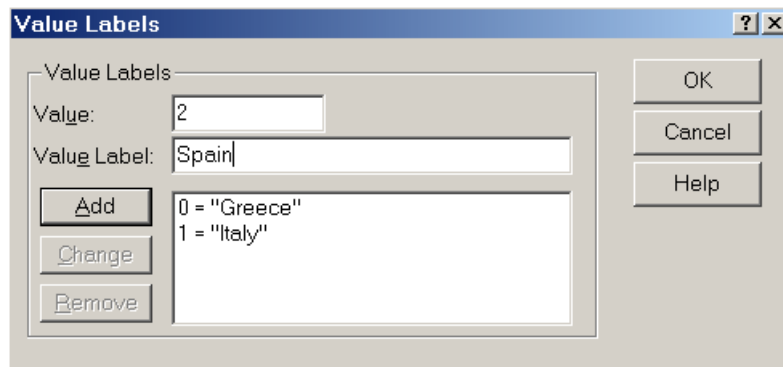
Στο πεδίο **Type**, δηλώνουμε τον τύπο της μεταβλητής. Πιέζοντας το πλήκτρο, εμφανίζεται το ακόλουθο πλαίσιο διαλόγου:

The dialog box titled "Define Variable Type:" contains a list of radio button options: **N**umeric (selected), **C**omma, **D**ot, **S**cientific notation, **D**ate, **D**ollar, **C**ustom currency, and **S**tring. To the right of the list are two input fields: "Width:" with the value "8" and "Decimal Places:" with the value "0". On the far right, there are three buttons: "Continue", "Cancel", and "Help".

Επιλέγουμε **Numeric** γιατί οι τιμές που παίρνει η μεταβλητή country είναι αριθμητικές (κι ας είναι κωδικοί). Στο πεδίο **Width**, που δηλώνει μέχρι πόσα ψηφία μπορούν να είναι οι τιμές, αφήνουμε τον αριθμό 8 και στο **Decimal places**, που δηλώνει το επιτρεπτό πλήθος δεκαδικών ψηφίων, δίνουμε την τιμή 0, αφού η μεταβλητή μας παίρνει μόνο ακέραιες τιμές. Κατόπιν πιέζουμε **OK**. Βλέπουμε ότι με αυτόν τον τρόπο αλλάζουν οι τιμές των πεδίων **Width** και **Decimal Places**

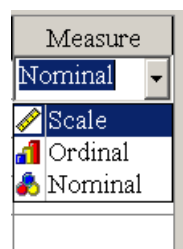
στο **Variable View**. Αν το επιθυμούμε, μπορούμε να αλλάξουμε και μεμονωμένα τις τιμές των πεδίων πηγαίνοντας στα αντίστοιχα κελιά.

Στο πεδίο **Label** εισάγουμε μια επεξηγηματική ετικέτα για την μεταβλητή. Για παράδειγμα, πληκτρολογήστε **country of origin**. Στο πεδίο **Values** δίνουμε όλες τις τιμές που μπορεί να πάρει η μεταβλητή και την σημασία τους (προφανώς μόνο για κατηγορικές μεταβλητές, των οποίων οι τιμές κωδικοποιούνται με αριθμούς). Πηγαίνοντας στο αντίστοιχο κελλί και πιέζοντας το πλήκτρο που εμφανίζεται, προκύπτει το ακόλουθο πλαίσιο διαλόγου:



Στο πεδίο **Value**, εισάγουμε την τιμή της μεταβλητής και στο πεδίο **Value Label**, την σημασία της. Για κάθε ζεύγος τιμής-σημασίας, πιέζουμε **Add** για να εισαχθεί στο πρόγραμμα. Αν κάνουμε αριστερό κλικ σε κάποια μεταβλητή που έχει ήδη εισαχθεί, ενεργοποιείται η επιλογή **Remove** και το ζεύγος εμφανίζεται στις θέσεις **Value** και **Value Label**. Αν πιέσουμε **Remove**, το ζεύγος διαγράφεται, ενώ, αν αρχίσουμε να γράφουμε, είτε στο πεδίο **Value** είτε στο πεδίο **Value Label**, ενεργοποιείται το πλήκτρο **Change** προκειμένου να μας δοθεί η δυνατότητα να αλλάξουμε μια τιμή ή μια ετικέτα. Αφού τελειώσουμε, πιέζουμε **OK**.

Το πεδίο **Missing** δεν μας ενδιαφέρει επί του παρόντος. Στο πεδίο **Align**, δηλώνουμε πώς θα στοιχίζονται στα κελιά οι τιμές της μεταβλητής. Οι διαθέσιμες επιλογές είναι αριστερά, στο κέντρο ή δεξιά. Στο πεδίο **Measure** δηλώνουμε τον τύπο της μεταβλητής. Πιέζοντας το πλήκτρο, προκύπτει ο ακόλουθος κατάλογος:



Scale σημαίνει αριθμητική μεταβλητή (π.χ. ύψος, βάρος, εισόδημα, κ.λ.π). **Ordinal** σημαίνει κατηγορική μεταβλητή, της οποίας οι αριθμητικές τιμές υποδηλώνουν κάποια διάταξη (π.χ. 1 κακό, 2 μέτριο, 3 καλό). **Nominal** σημαίνει κατηγορική μεταβλητή με τις αριθμητικές της τιμές να έχουν κάποια ερμηνεία χωρίς όμως να υποδηλώνουν διάταξη (π.χ. 0 άνδρας, 1 γυναίκα). Για την μεταβλητή **country** επιλέγουμε **Nominal**.

Σημείωση: Ακόμα και μια **scale** μεταβλητή μπορεί να έχει 0 δεκαδικά ψηφία.

Αφού εισαχθούν και οι δύο άλλες μεταβλητές, το παράθυρο δεδομένων στην μορφή **Data View** θα δείχνει ως εξής:

country	sex	height	
	country of origin		

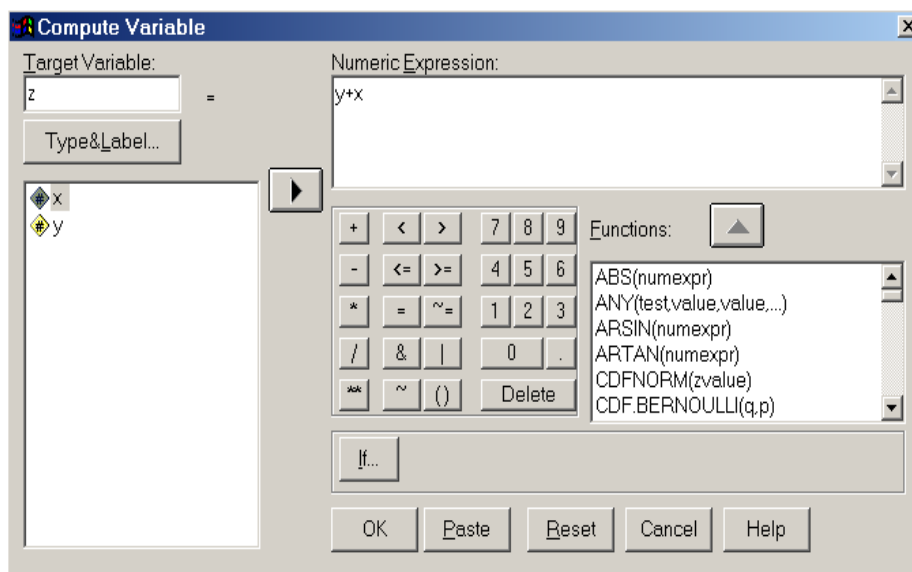
Η σημασία των ετικετών για τις μεταβλητές γίνεται αντιληπτή από το ότι, καθώς περνάμε το ποντίκι πάνω από το όνομα της κάθε μεταβλητής, εμφανίζεται στην οθόνη η ετικέτα της. Τιμές δίνουμε στις μεταβλητές με απλή πληκτρολόγηση στα κελιά του **Data View**.

Με μετάβαση στο **Variable View** μπορούμε να προσθέσουμε μεταβλητές ή να αλλάξουμε ιδιότητες μεταβλητών.

Μπορούμε να διαγράψουμε μεταβλητές κάνοντας δεξί κλικ στο όνομα τους και επιλέγοντας **Clear**, να τις αντιγράψουμε κάνοντας κλικ στο όνομα τους και **Copy, Paste**, και να αποθηκεύσουμε το αρχείο μας για μελλοντική χρήση με **File, Save**. Επίσης μπορούμε να αντιγράψουμε κελιά στο Excel.

Αν εισαγάγουμε τιμές σε κάποια κενή στήλη, το SPSS θα της δώσει αυτόματα ένα όνομα του τύπου **VAR0001**, το οποίο μπορούμε να μεταβάλουμε αργότερα, όπως και όλες τις υπόλοιπες ιδιότητες της μεταβλητής.

Μπορούμε επίσης να κάνουμε απλούς υπολογισμούς με μεταβλητές και να δημιουργήσουμε νέες μεταβλητές. Ας υποθέσουμε ότι έχουμε δύο μεταβλητές με ονόματα **x** και **y** και ότι θέλουμε να υπολογίσουμε μία νέα μεταβλητή **z** που θα είναι το άθροισμα των δύο πρώτων μεταβλητών. Επιλέγοντας **Transform, Compute**, προκύπτει το ακόλουθο πλαίσιο διαλόγου:



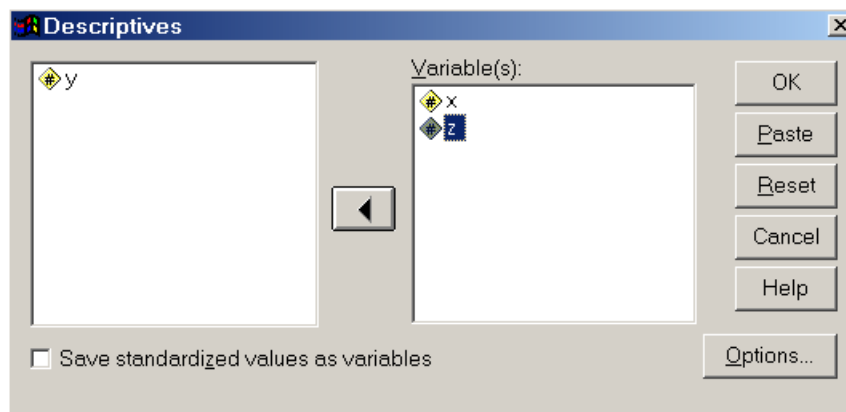
Στο πεδίο **Target Variable**, δηλώνουμε το όνομα της μεταβλητής στην οποία θα αποθηκευθεί το αποτέλεσμα των υπολογισμών. Αν αυτή υπάρχει ήδη, οι τιμές της θα αντικατασταθούν από το αποτέλεσμα των υπολογισμών (το SPSS θα μας ρωτήσει αν επιθυμούμε να γίνει αυτό), ενώ, αν δεν υπάρχει, θα δημιουργηθεί στην πρώτη κενή στήλη του φύλλου δεδομένων. Στο πεδίο **Numeric Expression**, δηλώνουμε τον μαθηματικό τύπο που θέλουμε να υπολογισθεί. Αυτός μπορεί να περιλαμβάνει οποιεσδήποτε μεταβλητές υπάρχουν ήδη στο φύλλο δεδομένων (εμφανίζονται στο αριστερό κάτω μέρος του παραθύρου), οποιουσδήποτε μαθηματικούς τελεστές (εμφανίζονται στο κέντρο του πλαισίου), καθώς και οποιαδήποτε μαθηματική συνάρτηση υπάρχει στο πεδίο **Functions**. Αφού πληκτρολογήσουμε την μαθηματική έκφραση, πιέζουμε **OK** και υπολογίζεται η νέα μεταβλητή.

Σημείωση: Όταν κάνουμε υπολογισμούς μεταξύ μεταβλητών πρέπει αυτές να έχουν τον ίδιο αριθμό παρατηρήσεων. Σε αντίθετη περίπτωση, το αποτέλεσμα θα έχει τόσες παρατηρήσεις όσες η μεταβλητή με τις

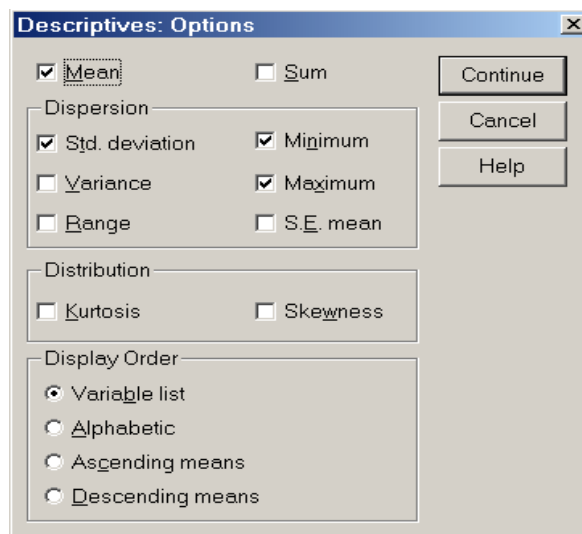
λιγότερες παρατηρήσεις και οι υπόλοιπες θέσεις, μέχρι να συμπληρωθεί πλήθος παρατηρήσεων ίσο με αυτές της μεταβλητής με τις περισσότερες παρατηρήσεις, θα θεωρηθούν ως ελλείποντες τιμές (missing values).

3. Περιγραφικά στατιστικά μέτρα: Θα δείξουμε, τέλος, πώς μπορούμε να υπολογίσουμε απλά περιγραφικά στατιστικά μέτρα για μία ή περισσότερες μεταβλητές.


Όλες οι στατιστικές αναλύσεις είναι διαθέσιμες στην επιλογή **Analyze**. Για τον υπολογισμό περιγραφικών στατιστικών μέτρων, επιλέγουμε **Analyze, Descriptive Statistics, Descriptives** και οδηγούμεθα στο εξής πλαίσιο διαλόγου:



Στο πεδίο **Variable(s)**, δηλώνουμε τις μεταβλητές για τις οποίες θέλουμε να υπολογίσουμε περιγραφικά στατιστικά μέτρα. Για να δηλώσουμε ποια στατιστικά μέτρα θέλουμε, πιέζουμε **Options** και οδηγούμεθα στο ακόλουθο πλαίσιο:



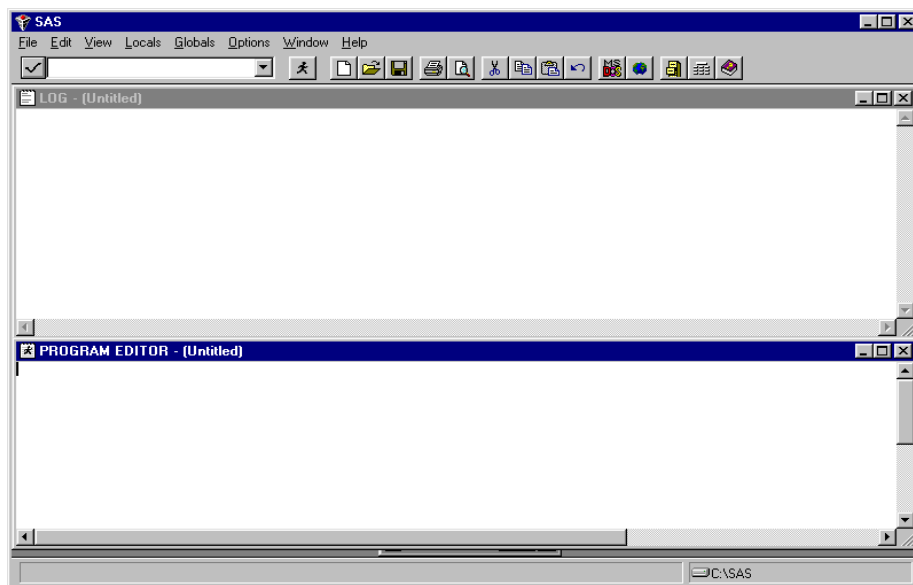
Στο πλαίσιο αυτό, είναι διαθέσιμα προς επιλογή διάφορα στατιστικά μέτρα. Επιλέγοντας αυτά που θέλουμε, πιέζοντας **Continue** και έτσι επιστρέφουμε στο προηγούμενο πλαίσιο όπου πιέζοντας **OK**, εμφανίζονται στην οθόνη τα περιγραφικά στατιστικά μέτρα που ζητήσαμε.

4. Το πλήκτρο Undo: Σε περίπτωση που κάνουμε μία λάθος ενέργεια (π.χ. αν διαγράψουμε μία στήλη με δεδομένα που μας χρειάζονται), έχουμε τρόπο να διορθώσουμε το λάθος μας. Η πίεση του πλήκτρου Undo, δηλαδή του πλήκτρου , αναιρεί την τελευταία ενέργεια που κάναμε πριν το πιάσουμε. Αν το πιάσουμε κι άλλες φορές, θα αναιρεθούν η προτελευταία ενέργεια, η αντιπροτελευταία ενέργεια, κ.ο.κ. Αν το πλήκτρο είναι απενεργοποιημένο (έχει ένα ανοικτό γκρι χρώμα), τότε δεν υπάρχουν ενέργειες που μπορούν να αναιρεθούν. Μία πράξη Undo δεν ακυρώνεται από ένα άλλο Undo.

1.2.3 Μη Παραμετρικές Μέθοδοι με το SAS (6.12)

Το SAS είναι ένα πακέτο για στατιστικές αναλύσεις που δουλεύει κυρίως με εντολές, ενώ οι δυνατότητες που παρέχονται μέσω των διαδικασιών του **menu** είναι περιορισμένες. Αυτό βέβαια απαιτεί γνώση από την πλευρά του χρήστη των στατιστικών μεθόδων (η χρήση του δεν είναι απλή και έτσι δεν μπορεί να χρησιμοποιηθεί από «μη Στατιστικούς» ενώ, παράλληλα, δίδει στο χρήστη την δυνατότητα εκτέλεσης ενός μεγάλου αριθμού εργασιών, μετασχηματισμών και στατιστικών αναλύσεων (θεωρείται ως ένα από τα πιο «ισχυρά» στατιστικά πακέτα αναφορικά με τις δυνατότητες που παρέχονται).

Η κλίση του προγράμματος ανοίγει το παράθυρο που ακολουθεί.

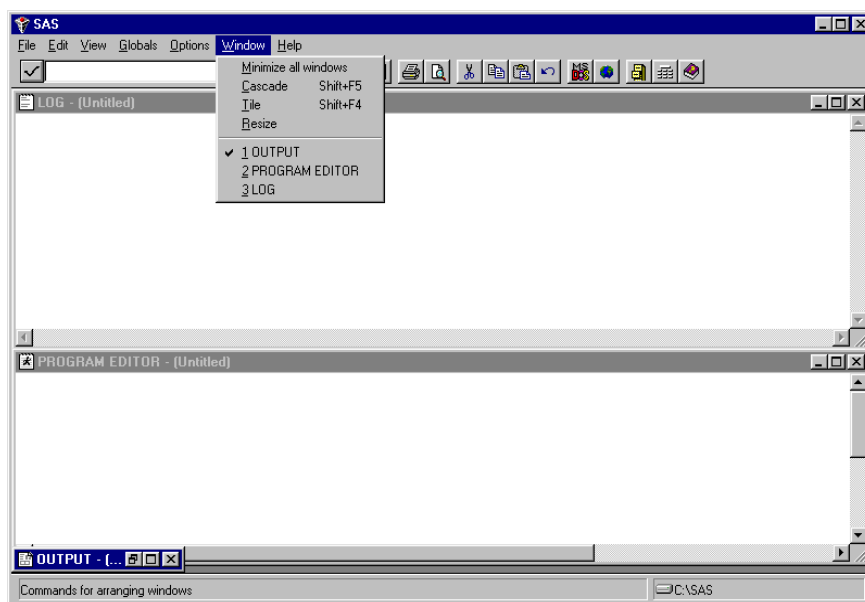


Το παράθυρο **PROGRAM EDITOR** είναι το πεδίο στο οποίο πληκτρολογούμε τις εντολές (θα αναφέρεται στα παρακάτω ως **παράθυρο εντολών**, ενώ στο παράθυρο **LOG** εμφανίζονται σχόλια, επισημάνσεις, απαιτούμενος χρόνος εκτέλεσης κάθε εντολής καθώς και λάθη, μετά την εκτέλεση των εντολών που έχουμε πληκτρολογήσει στο

παράθυρο εντολών. Θα πρέπει να σημειωθεί ότι οι εντολές που έχουν εισαχθεί λανθασμένα δεν δίνουν αποτελέσματα, ακόμη και αν το λάθος βρίσκεται σε κάποιο σημείο υποεντολής (που παρέχει πληροφορίες πέραν αυτών που παρέχει η βασική εντολή).

Για την εκτέλεση των εντολών που έχουμε πληκτρολογήσει, μπορούμε να τις μαρκάρουμε και να επιλέξουμε από το menu, **locals** και **submit**. Θα πρέπει να σημειωθεί ότι η επιλογή **locals** εμφανίζεται στο menu μόνο όταν ο cursor βρίσκεται στο παράθυρο εντολών.

Πίσω από αυτά τα δύο παράθυρα, υπάρχει το παράθυρο **OUTPUT**, το οποίο εμφανίζεται μπροστά στην οθόνη αυτόματα μετά την εκτέλεση κάποιας εντολής και δίνει τα αποτελέσματα. Ο χρήστης επίσης μπορεί να μετακινηθεί οποτεδήποτε σε κάποιο από αυτά τα παράθυρα από την επιλογή **Window** του βασικού menu.



Οι εντολές αποθηκεύονται σε αρχείο που έχει κατάληξη **.sas**, τα σχόλια σε αρχείο με κατάληξη **.log**, ενώ τα αποτελέσματα σε αρχείο με κατάληξη **.lis**.

Ένα αρχείο εντολών SAS αρχίζει με την εντολή

```
OPTIONS linesize= pagesize= ;
```

(συνήθως `OPTIONS linesize=80 pagesize=65 ;`), η οποία ορίζει το μήκος γραμμής και σελίδας στο αρχείο των αποτελεσμάτων, αλλά η εντολή αυτή δεν είναι υποχρεωτική.

Αν θέλουμε να φύγουμε από το SAS μετά την εκτέλεση του προγράμματος, τελειώνουμε με την εντολή

```
endsas;
```

Σημείωση: Πάντα χρησιμοποιούμε «;» στο τέλος κάθε εντολής. (Η παράλειψη αυτού οδηγεί σε εμφάνιση λαθών και η εντολή δεν εκτελείται). Επίσης, μπορούμε να προσθέσουμε σχόλια στο πρόγραμμα SAS χρησιμοποιώντας την εντολή `* comments ;` ή την εντολή `/* comments */`

Τα δεδομένα στο SAS

Τρεις τύποι αρχείων μπορούν να χρησιμοποιηθούν στο βήμα των δεδομένων (**data step**): δεδομένα από αρχείο, δεδομένα που δίδονται στο πρόγραμμα SAS και ήδη υπάρχοντα αρχεία δεδομένων SAS.

Θα δούμε τους τρεις τύπους αρχείων και τις βασικές εντολές διαχείρισης δεδομένων με μια σειρά από παραδείγματα.

1. Εισαγωγή δεδομένων από αρχείο:

Εισάγουμε τα παρακάτω δεδομένα σε αρχείο με όνομα `char.txt`

```
Helen 160 52 22 f
```

```
Mary 170 60 20 f
```

Alexis 185 75 18 m
Daphne 168 50 21 f
George 190 82 21 m
Caterina 173 70 19 f
Petros 178 79 22 m
Costas 181 78 24 m
Michael 180 77 18 m

Στην συνέχεια, πληκτρολογούμε τις εξής εντολές στο πρόγραμμα εντολών.

```
* this program reads and analyzes data from a text file;  
options linesize=80 pagesize=65;
```

```
DATA one;  
INFILE 'char.txt';  
INPUT id $ h w a s $  
;  
class=1;  
city='Athens';
```

```
LABEL h='height'  
      w='weight'  
      a='age'  
      s='gender';  
proc sort;by s;  
proc print; title 'Demographic Gharacteristics';  
proc means; var w h a;  
proc means; var w h a; by s;
```

Σχόλια

- Κατασκευάσαμε το SAS αρχείο δεδομένων **ONE** που περιέχει τα δεδομένα του αρχείου char.txt.
- Χρησιμοποιούμε το σύμβολο \$ μετά το όνομα αλφαριθμητικής μεταβλητής για να δηλώσουμε ότι η μεταβλητή περιέχει χαρακτήρες, όχι αριθμούς (εδώ, οι μεταβλητές **id** και **s**).
- Χρησιμοποιώντας τις εντολές **class=1;** και **city='Athens';**, προσθέτουμε τις μεταβλητές **class** (με τιμή 1 για κάθε άτομο) και **city** (με την αλφαριθμητική τιμή Athens για κάθε άτομο) στο σετ δεδομένων ONE. Παρατήρηση: για να δηλώσουμε αλφαριθμητική τιμή, χρησιμοποιούμε «μονά» εισαγωγικά (π.χ. 'Athens').
- Χρησιμοποιούμε την εντολή **LABEL** για επεξηγηματικά σχόλια στις μεταβλητές.
- Οι εντολές γράφονται είτε με μικρά είτε με κεφαλαία γράμματα.
- Με την λέξη **proc** αρχίζουν όλες οι εντολές που αφορούν στις στατιστικές διαδικασίες που θέλουμε να εκτελέσουμε. Είναι σημαντικό να τονισθεί ότι η δημιουργία νέων μεταβλητών δεν είναι δυνατή όταν έχουμε προχωρήσει σε εντολές στατιστικών διαδικασιών. Για να γίνει δυνατό αυτό, θα πρέπει να επανέλθουμε σε βήμα διαχείρισης δεδομένων (αρχίζοντας και πάλι με τη λέξη **data**).

2. Εισαγωγή δεδομένων στο πρόγραμμα εντολών:

```
* this program reads and analyzes data given in the SAS  
program;  
options linesize=80 pagesize=65;  
DATA one;
```

```

INPUT id $ h w a s $
;
class=1;
city='Athens';
LABEL h='height'
      w='weight'
      a='age'
      s='gender';
CARDS;
Helen 160 52 22 f
Mary 170 60 20 f
Alexis 185 75 18 m
Daphne 168 50 21 f
George 190 82 21 m
Caterina 173 70 19 f
Petros 178 79 22 m
Costas 181 78 24 m
Michael 180 77 18 m
;
run;

proc sort;by s;
proc print; title 'Demographic Gharacteristics';
proc means; var w h a;
proc means; var w h a; by s;

```

Με τις παραπάνω εντολές,

- Δηλώνουμε ότι αρχίζει η εισαγωγή δεδομένων με την εντολή **cards;**
- Χρησιμοποιούμε “;” στο τέλος της εισαγωγής δεδομένων (π.χ. *Michael 180 77 18 m ;*).
- Τελειώνουμε την εισαγωγή δεδομένων με την εντολή **run**.
- Δεν χρησιμοποιούμε την εντολή **INFILE**.

3. Ενοποίηση αρχείων:

Γράφουμε το σετ δεδομένων που ακολουθεί σε αρχείο με όνομα height.txt

Helen 160 f

Mary 170 f

Alexis 185 m

Daphne 168 f

George 190 m

Caterina 173 f

Petros 178 m

Costas 181 m

Michael 180 m

Επίσης, γράφουμε το σετ δεδομένων που ακολουθεί σε αρχείο με όνομα weight.txt

Caterina 70 19

Petros 79 22

Costas 78 24

Michael 77 18

Helen 52 22

Mary 60 20

Alexis 75 18

Daphne 50 21

George 82 21

Στην συνέχεια, στο πρόγραμμα εντολών πληκτρολογούμε τις εντολές:

```
* this program reads and analyzes data from text files and  
creates a merged data set ;
```

```
options linesize=80 pagesize=65;
```

```
DATA one;
```

```
INFILE 'height.txt';
```

```
INPUT id $ h s $
```

```
;
```

```
class=1;
```

```
city='Athens';
```

```
proc sort; by id;
```

```
proc print; title 'HEIGHT';
```

```
DATA two;
```

```
INFILE 'weight.txt';
```

```
INPUT id $ w a
```

```
;
```

```
proc sort; by id;
```

```
proc print; title 'WEIGHT';
```

```
DATA all; merge one(in=aa)two(in=bb); by id;
```

```
if aa and bb;
```

```
LABEL h='height'
```

```
w='weight'
```

```
a='age'
```

```
s='gender';
```

```
proc sort;by s;
```

```
proc print; title 'Demographic Gharacteristics';
```

```
proc means; var w h a;
```

```
proc means; var w h a; by s;
```

Παρατηρήσεις:

- Μπορούμε να διαβάσουμε **περισσότερα από ένα αρχεία** σε ένα μόνο πρόγραμμα SAS.
- Δημιουργήσαμε το σετ δεδομένων SAS all που περιέχει τα δεδομένα των one και two.

4. Ανάγνωση δεδομένων από αρχείο κειμένου ή κατευθείαν από το πρόγραμμα εντολών, όταν οι τιμές δεν διαχωρίζονται με κενό μεταξύ τους.

** This program reads and analyzes data from a text file which are not blank separated;*

```
options linesize=80 pagesize=65;
```

```
DATA one;
```

```
INPUT id $ 1-6 h 7-9 w 10-11 a 12-13 s $ 14
```

```
;
```

```

class=1;
city='Athens';

LABEL h='height'
      w='weight'
      a='age'
      s='gender';
CARDS;
George1908221m
Helen 1605222f
Mary 1706020f
Alexis 1857518m
;
run;
proc print; title 'Demographic Gharacteristics';
proc means; var w h a;

```

Παρατήρηση: Μπορούμε να δηλώσουμε τις στήλες στις οποίες εμφανίζονται οι μεταβλητές (χρήσιμο όταν οι τιμές δεν διαχωρίζονται με κενό).

5. Ανάγνωση δεδομένων όταν οι μεταβλητές δεν δίδονται σε στήλες αλλά οι τιμές των μεταβλητών μιας δειγματικής μονάδας έπονται της προηγούμενης στην ίδια γραμμή.

```

* this program reads and analyzes data from a single line;
options linesize=80 pagesize=65;

```

```

DATA one;

```



```

INPUT h w a @@;
LABEL h='height'
      w='weight'
      a='age';
CARDS;
178 67 21 165 61 24 176 70 19 182 77 25;
run;
proc print; title 'Demographic Characteristics';
proc means; var w h a;
endsas;

```

Παρατήρηση: Χρησιμοποιούμε τα σύμβολα @@ στο τέλος της εντολής INPUT για να δηλώσουμε ότι τα δεδομένα δίνονται σε μια γραμμή.

Χρήσιμες εντολές στο DATA STEP

1. Δημιουργία και τροποποίηση μεταβλητών

Για να τροποποιήσουμε μεταβλητές ή να δημιουργήσουμε νέες, μπορούμε να χρησιμοποιήσουμε κατάλληλες εντολές σε οποιοδήποτε DATA step.

Παραδείγματα: age=age+age*10; (τροποποίηση υπάρχουσας μεταβλητής)

netsales=sales-expenses; (καινούργια μεταβλητή)

```

** δύναμη
* πολ/σμός
/ διαίρεση
+ πρόσθεση
- αφαίρεση

```

Μπορούμε επίσης να χρησιμοποιήσουμε συναρτήσεις του SAS:

Παράδειγμα: avgincome=mean(income1,income2);

2. Εντολή MISSING

Με την εντολή αυτή δηλώνουμε τις αγνοούμενες ή ελλείπουσες τιμές (missing values) του αρχείου δεδομένων. Για παράδειγμα, αν οι αγνοούμενες τιμές αυτές δηλώνονται με το γράμμα **E** στο αρχείο δεδομένων, χρησιμοποιούμε την εντολή

missing E;

3. Εντολή LRECL

Δηλώνει το μέγιστο μέγεθος γραμμής ανά παρατήρηση. Η εντολή χρησιμοποιείται απαραίτητα στην εντολή **INPUT**, αν στο αρχείο δεδομένων υπάρχουν περισσότεροι από 256 χαρακτήρες ανά γραμμή.

INPUT id \$ h w a s \$ lrecl=1000;

4. Εντολή ενοποίησης MERGE

Χρησιμοποιείται για οριζόντια ενοποίηση αρχείων αφού προηγηθεί ταξινόμηση των στοιχείων των αρχείων ως προς μία κοινή μεταβλητή τους. Δηλαδή, αν συμβολίσουμε με **var** την κοινή μεταβλητή των αρχείων, χρησιμοποιούμε πάντα πρώτα την εντολή

proc sort; by var;

- Αν θέλουμε το ενοποιημένο αρχείο να περιλαμβάνει μόνο τις κοινές παρατηρήσεις των αρχείων **one** και **two** πληκτρολογούμε

proc sort; by var;

data all;

merge one(in=aa) two(in=bb); by var;

if aa and bb;

- Αν θέλουμε το ενοποιημένο αρχείο να περιλαμβάνει μόνο τις παρατηρήσεις από το αρχείο **two** που υπάρχουν και στο **one**, πληκτρολογούμε

```
proc sort; by var;  
data all;  
merge one(in=aa) two(in=bb); by var;  
if aa;
```

- Αν θέλουμε το ενοποιημένο αρχείο να περιλαμβάνει μόνο τις παρατηρήσεις από το αρχείο **two** που δεν υπάρχουν στο **one**, πληκτρολογούμε

```
proc sort; by var;  
data all;  
merge one(in=aa) two(in=bb); by var;  
if a ne b;
```

5. Εντολή SET

Χρησιμοποιείται για κάθετη ενοποίηση αρχείων όταν υπάρχουν κοινές μεταβλητές στα αρχεία. Για παράδειγμα, οι εντολές

```
data new;  
set one two;  
run;
```

ενοποιούν κάθετα τα αρχεία **one** και **two**.

Όταν η εντολή χρησιμοποιείται για ένα αρχείο, απλά αντιγράφει τα περιεχόμενά του σε κάποιο άλλο αρχείο. Για παράδειγμα, οι εντολές

```
data new;  
set one;  
run;
```

αντιγράφουν τα περιεχόμενα του αρχείου **one** στο αρχείο **new**.

6. Εντολή DROP

Χρησιμοποιείται όταν θέλουμε να παραλείψουμε μεταβλητές από το αρχείο δεδομένων. Για παράδειγμα,

```
data one;  
set new;  
drop w;  
run;
```

7. Εντολή KEEP

Κρατάμε τις μεταβλητές στο αρχείο. Για παράδειγμα,

```
data one;  
set new;  
keep h s;  
run;
```

8. Εντολή DELETE

Χρησιμοποιείται όταν θέλουμε να διαγράψουμε παρατηρήσεις από το SAS αρχείο δεδομένων. Για παράδειγμα,

```
data one;  
set new;  
if a>19 then delete;  
run;
```

9. Εντολή RENAME

Με την εντολή αυτή μπορούμε να μετονομάσουμε μία μεταβλητή. Για παράδειγμα,

```
data one;
set new;
rename h=height
      s=gender;
run;
```

10. Εντολές IF, THEN και ELSE

Οι εντολές αυτές είναι λογικές εντολές. Η εντολή **IF** χρησιμοποιείται για να δηλωθούν οι προϋποθέσεις κάτω από τις οποίες η εντολή **THEN** θα επιτρέψει την εκτέλεση άλλων εντολών.

Παράδειγμα 1: *data drug1;*

```
set cltrials;
if placebo='YES';
cholchg=chol2-chol1;
glucose=glucose+glucose*.10;
run;
```

Παράδειγμα 2: *if test>=95 and (theme='A' or project='A') then grade='A+';*

Με τις εντολές **IF** και **THEN** μπορούν να χρησιμοποιηθούν και οι εξής λογικοί τελεστές:

Σύμβολο	Σημασία
&	AND
	OR
^ or ~	NOT

Πολλές φορές, αντί να χρησιμοποιούμε μια σειρά από **IF-THEN** εντολές, μπορούμε να χρησιμοποιήσουμε την εντολή **ELSE**, για να δηλώσουμε μια εναλλακτική ενέργεια όταν η συνθήκη στην **IF-THEN** εντολή δεν ικανοποιείται. Για παράδειγμα,

```
if code='1' then type='Fixed';  
else if code='2' then type='Variable';  
else type='Unknown';
```