

Τμήμα
Μηχανικών
Πληροφορικής τ.ε.
Τεχνολογικό Εκπαιδευτικό Ίδρυμα
Δυτικής Ελλάδας

Πτυχιακή Εργασία

Σύγχρονα Στατιστικά Λογισμικά και οι Εφαρμογές τους

Ντάνης Δημήτριος

Επιβλέπων καθηγητής: Κος Κούγιας Ιωάννης



Αντίρριο 2016

Ευχαριστίες

Για την εκπόνηση της παρούσας πτυχιακής εργασίας, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή του ΤΕΙ Δυτικής Ελλάδας κύριο Ιωάννη Κούγια, για τη συνεργασία και την πολύτιμη συμβολή του στην ολοκλήρωση της.

Μάιος 2016
Δημήτριος Ντάνης

Περιεχόμενα

Εισαγωγή.....	4
1.Στατιστικά Λογισμικά	6
1.1 Γενικά για τα στατιστικά λογισμικά.....	6
1.2 SPSS.....	7
1.3 MINITAB.....	8
1.4 Γλώσσα R.....	9
1.5 SAS.....	10
2.SPSS (statistical package for the social sciences)	11
2.1 Εισαγωγή στο SPSS	11
2.2 Ερωτηματολόγιο.....	11
2.3 Παρουσίαση SPSS σε Windows	15
2.3.1 Εκκίνηση SPSS	15
2.3.2 Data Editor (Variable view)	16
2.4 Βασικές εντολές	22
2.4.1 Εντολή Compute.....	23
2.4.2 Εντολή Count.....	24
2.4.3 Εντολή Recode.....	27
2.4.4 Εντολή Frequencies.....	30
2.4.5 Εντολή Descriptives	35
2.4.6 Εντολή Crosstabs	37
2.5 Γραφήματα.....	40
2.5.1 Pie Chart	42
2.5.2 Histogram.....	43
3.MINITAB	44
3.1 Εισαγωγή στο Minitab.....	44
3.2 Παρουσίαση Minitab σε Windows.....	45
3.2.1 Εισαγωγή δεδομένων	46
3.2.2 Αλλαγή μεταβλητής	47
3.2.3 Επιλογές δεδομένων	48
3.3 Στατιστική Ανάλυση	51
3.3.1 Περιγραφική Στατιστική.....	51
3.3.2 Κατανομή Συχνοτήτων	53
3.4 Γραφικές Παραστάσεις.....	56
3.4.1 Histogram.....	57
3.4.2 Bar Graph	59
3.4.3 Pie Chart	61

4. R.....	63
4.1 Εισαγωγή στην γλώσσα R	63
4.2 Εγκατάσταση της R	63
4.3 Αριθμητικοί Τελεστές.....	65
4.4 Τελεστές Εκχώρησης της R	66
4.5 Βασικές Συναρτήσεις.....	67
4.6 Παρουσίαση R σε Windows	68
4.6.1 Επιλογή Βοήθεια	69
4.6.2 Βασική σύνταξη εντολών.....	72
4.6.3 Αποθήκευση και επανάκτηση αρχείων	73
4.7 Αντικείμενα.....	74
4.7.1 Διανύσματα	76
4.7.2 Δισδιάστατοι Πίνακες.....	78
4.7.3 Πίνακες μεγαλύτερης διάστασης (arrays)	79
4.7.4 Πλαίσια Δεδομένων.....	80
4.7.5 Λίστες	80
4.8 Γραφήματα.....	81
4.8.1 Απλά γραφήματα.....	81
4.8.2 Γραφικές δυνατότητες	82
5. SAS.....	83
5.1 Εισαγωγή στο λογισμικό SAS	83
5.2 Εγκατάσταση του SAS.....	83
5.3 Βασικά του λογισμικού.....	89
5.3.1 Basic Components of SAS	90
5.3.2 Κανονισμοί και εντολές στην SAS	92
5.3.3 SAS Functions.....	94
5.4 Εφαρμογή λογισμικού SAS.....	96
5.4.1 Παράδειγμα Συναρτήσεων	96
5.4.2 Proc Print.....	98
5.4.3 Proc Frequencies and Proc Mean	99
5.4.4 Proc Univariate.....	103
5.5 Graph in SAS.....	106
5.5.1 Bar Chart	106
5.5.2 Pie Chart	108
5.5.3 Bar Chart from Counts	112
6.Συμπεράσματα.....	114
Εικόνες.....	115
Βιβλιογραφία	116
Ηλεκτρονική Βιβλιογραφία.....	117

ΕΙΣΑΓΩΓΗ

Η πτυχιακή αυτή εργασία ασχολείται με τα στατιστικά λογισμικά πακέτα που κυκλοφορούν στην αγορά και χρησιμοποιούνται από μεγάλες επιχειρήσεις, για ιατρικούς, κοινωνικούς, οικονομικούς και εκπαιδευτικούς σκοπούς . Θεωρείται σωστό να γίνει μια αναφορά για τον ορισμό και την ιστορία της Στατιστικής, η οποία βρίσκεται σε μεγάλο βαθμό στην καθημερινότητα μας.

Η λέξη Στατιστική προέρχεται από την αρχαία ελληνική λέξη **στατίζω** που σημαίνει ταξινομώ, τοποθετώ. Η Στατιστική έκανε την εμφάνιση της κατά τις πρώτες δημιουργίες οργανωμένων κοινωνιών. Σε αυτές τις κοινωνίες η Στατιστική χρησιμοποιήθηκε ως πίνακες για την καταγραφή των γεννήσεων, των θανάτων, των προϊόντων, τους φόρους κτλ, μέσα στην κοινωνία. Ένα χαρακτηριστικό παράδειγμα είναι ο κατάλογος των πλοίων με αριθμητικά δεδομένα, των Αχαιών στον Τρωικό Πόλεμο, από τον Όμηρο.

Ως ορισμός η Στατιστική υποδηλώνει τις μαθηματικές μεθόδους που λήφθηκαν με μέτρηση ή απαρίθμηση και τα ίδια αυτά τα δεδομένα έχουν υποστεί κάποιους χειρισμούς.

Η 1^η ιστορικά συλλογή στατιστικών στοιχείων, καταγράφεται το 2238 π.Χ. στην Κίνα, όπου ο Αυτοκράτορας Γιάο έκανε απογραφή του πληθυσμού της χώρας. Τέτοιες απογραφές είχαν πραγματοποιήσει και άλλοι αρχαίοι λαοί, όπως οι Πέρσες , οι αρχαίοι Έλληνες, οι Βαβυλώνιοι, οι Ρωμαίοι και οι Αιγύπτιοι. Συγκεκριμένα στην αρχαία Ελλάδα η συλλογή στατιστικών δεδομένων χρησίμευε στην φορολόγηση και στην στράτευση. Επίσης ο όρος Στατιστική αναφέρεται στα έργα του Σωκράτη και του Αριστοτέλη.

Συστηματική συλλογή δεδομένων πραγματοποιήθηκε κατά την Αναγέννηση και κυρίως στις περιοχές της Βενετίας και Φλωρεντίας, όπου στην συνέχεια επεκτάθηκε σε όλη την Ευρώπη. Τον 11^ο αιώνα, ο Γουλιέλμος ο Κατακτητής πραγματοποίησε μια μεγάλη στατιστική απογραφή, στην οποία κατέγραψε τις μονάδες παραγωγής της Αγγλίας. Στα επόμενα χρόνια ξεκίνησε η απογραφή θανάτων και ασθενειών κυρίως λόγω της ασθένεια πανώλης που κράτησε για περίπου 4 αιώνες. Πολλοί επιστήμονες θεωρούν αφετηρία της Στατιστικής την χρονιά 1663, στην οποία εκδόθηκε το βιβλίο του Τζον Γκράουντ, Φυσικές και Πολιτικές παρατηρήσεις της θνησιμότητας. Η ανάπτυξη του εμπορίου στους επόμενους αιώνες, ανέπτυξε και τις μελέτες των οικονομικών δεδομένων. Σήμερα η Στατιστική θεωρείται αυτοτελή επιστήμη με συγκεκριμένες μεθόδους ανάλυσης.

Η Στατιστική σήμερα έχει ενσωματωθεί σε όλους τους κλάδους των επιστημών όπως σε κοινωνικές, ανθρωπιστικές επιστήμες, σε τεχνολογικές (πληροφορική, μηχανολογία) σε Επιστήμες Υγείας (Ιατρικές, Φαρμακευτικές) και σε επιστήμες Οικονομίας και Διοίκησης (Μάρκετινγκ, Οικονομικά). Για τον σκοπό αυτό έχουν δημιουργηθεί αρκετά στατιστικά λογισμικά για την διευκόλυνση των χρηστών.

Στην πτυχιακή αυτή εργασία θα παρουσιάσω περιληπτικά κάποια στατιστικά λογισμικά πακέτα. Και στην συνέχεια θα τα παρουσιάσω εκτενέστερα σε περιβάλλον Windows.

1 Στατιστικά Λογισμικά Πακέτα

1.1 Γενικά για τα Στατιστικά Πακέτα

Η ραγδαία εξέλιξη της τεχνολογίας και η επιστημονική πρόοδος που έχει επιτευχθεί τα τελευταίες δεκαετίες, οδηγούν στην χρησιμοποίηση των ηλεκτρονικών υπολογιστών στην καθημερινότητα μας. Για την κάλυψη των αναγκών αυτών δημιουργήθηκαν πολλά και διάφορα λογισμικά για όλους τους τομείς.

Με την εξέλιξη όμως αυτή δημιουργήθηκαν και αυξημένες απαιτήσεις για συγκέντρωση, ανάλυση και αναπαράσταση δεδομένων. Για τον λόγο αυτό από την δεκαετία του 1950 και μετά άρχισαν να αναπτύσσονται σε πανεπιστήμια διάφορα στατιστικά λογισμικά πακέτα.

Τα λογισμικά αυτά πακέτα αναπτύχθηκαν σε διάφορες γλώσσες προγραμματισμού όπως είναι η S, η C και η C++. Με την ανάπτυξη της οικονομία και τις απαιτήσεις της βιομηχανίας τα λογισμικά αυτά έγιναν πιο αναγκαία. Με την πάροδο του χρόνου τα πακέτα εξελίχθηκαν, έγιναν συμβατά με τις τρέχουσες τεχνολογίες και είναι πια δυνατόν να τρέχουν σε όλα τα λειτουργικά περιβάλλοντα (Windows, Linux, Unix, Mac Os , MS-Dos). Λογισμικό, όπως η εταιρεία IBM

Στην σημερινή εποχή στατιστικά λογισμικά πακέτα χρησιμοποιούνται παγκοσμίως από μεγάλες εταιρίες, μεγάλους οργανισμούς και πανεπιστήμια. Η ανάγκη για την επεξεργασία και την ανάλυση των δεδομένων που υπάρχει, έχει κάνει παγκοσμίως φήμης οργανισμούς να αγοράσουν η να δημιουργήσουν δικό τους λογισμικό, όπως η εταιρεία IBM.

1.2 SPSS

Το λογισμικό πακέτο SPSS (Statistical Package for the Social Sciences) αναπτύχθηκε από 2 φοιτητές του Πανεπιστημίου Στάνφορντ το 1965, τους Νόρμαν Νίε και Ντέιλ Μπεν.

Η 1^η έκδοση του SPSS δημιουργήθηκε με γλώσσα προγραμματισμού FORTRAN. Το 1967 το λογισμικό πήρε το όνομα του ως έχει σήμερα (Statistical Package for the Social Sciences). Το 1983 κυκλοφορεί η πρώτη έκδοση του SPSS για προσωπικό ηλεκτρονικό υπολογιστή που χρησιμοποιούσε λειτουργικό σύστημα MS-Dos



Εικόνα 1: SPSS logo

Η τελευταία έκδοση του λογισμικού είναι σχεδιασμένη σε αντικειμενοστραφή γλώσσα προγραμματισμού Java.

Η μεγάλη εξέλιξη του λογισμικού, έκανε την εταιρία IBM να αγοράσει το SPSS έναντι του ποσού των 1.26 δις δολάρια. Χρησιμοποιείται σήμερα για στατιστικές αναλύσεις. Κυρίως χρησιμοποιείται στις κοινωνικές και ιατρικές επιστήμες αλλά και στον κλάδο του Marketing



Εικόνα 2: IBM και SPSS

1.3 MINITAB

Το στατιστικό λογισμικό MINITAB δημιουργήθηκε και αναπτύχθηκε στο Πανεπιστήμιο της Πενσυλβανία το 1972 από τους ερευνητές Barbara Ryan, Thomas Ryan και Bryan Joiner. Το λογισμικό αυτό πακέτο δημιουργήθηκε με σκοπό οι φοιτητές του πανεπιστημίου να διδάσκονται την στατιστική και να υλοποιούν στατιστικούς υπολογισμούς μέσω υπολογιστών.



Εικόνα 3: MINITAB logo

Το λογισμικό ανήκει στην ιδιωτική εταιρεία Minitab Inc, η οποία σε σύνδεση με το Minitab προσφέρει Quality Trainer, το οποίο είναι ένα πακέτο eLearning και επίσης προσφέρει και Quality Companion, το οποίο είναι ένα εργαλείο διαχείρισης Six Sigma και Lean Manufacturing.

Σήμερα το στατιστικό λογισμικό πακέτο Minitab χρησιμοποιείται για εκπαιδευτικούς σκοπούς σε πάνω από 40.000 Κολλέγια και Πανεπιστήμια κυρίως στις Ηνωμένες Πολιτείες Αμερικής αλλά και στον υπόλοιπο κόσμο.

1.4 ΓΛΩΣΣΑ R

Η R είναι μια γλώσσα προγραμματισμού αλλά λογισμικό περιβάλλον ταυτόχρονα. Η χρήση της είναι κυρίως για την πραγματοποίηση στατιστικών υπολογισμών και των γραφικών αναπαραστάσεων τους.

Έκανε την εμφάνιση της το 1993 στο Πανεπιστήμιο του Όκλαντ της Νέας Ζηλανδίας από τους Ross Ihaka και Robert Gentleman (το όνομα της οφείλεται στα αρχικά των μικρών ονομάτων των δημιουργών της, Ross και Robert). Αρχικά η R ήταν εκτέλεση της γλώσσας προγραμματισμού S.



Εικόνα 4: R
programming language
logo

Η R αρχικά είχε γραφτεί σε διάφορες γλώσσες προγραμματισμού όπως σε C, FORTRAN και R. Αν και η R έχει command line interface, έχουν ήδη κυκλοφορήσει γραφικά περιβάλλοντα για την διευκόλυνση στην χρησιμοποίησή της. Η R είναι απόκτημα της GNU, αναπτύσσεται από την R Core Team και είναι διαθέσιμη ελεύθερα από την GNU License. Σήμερα η R χρησιμοποιείται κυρίως από στατιστικούς και από data miners για την ανάπτυξη στατιστικών λογισμικών.

1.5 SAS

Το στατιστικό λογισμικό πακέτο SAS (statistical analysis system) δημιουργήθηκε και αναπτύχθηκε στο Πανεπιστήμιο της Βόρειας Καρολίνας από το 1966 έως το 1976. Το πρότζεκτ αυτό, χρηματοδοτήθηκε από το Εθνικό Ινστιτούτο Υγείας των Ηνωμένων Πολιτειών Αμερικής, και είχε ως δημιουργό τον James Barr και τον απόφοιτο φοιτητή James Goodnight. Το 1973 στην ομάδα εντάχθηκε και ο John Sall. Το 1976 ιδρύθηκε η SAS Institute Inc.



Εικόνα 5: SAS logo

Η εταιρεία SAS Institute Inc. στις επόμενες δεκαετίες θα ανακηρυχτεί η πιο γρήγορα αναπτυσσομένη εταιρεία των Η.Π.Α. . Στην συνέχεια ενσωματώθηκαν νέες στατιστικές διαδικασίες και παρουσιάστηκε το JMP από την SAS Institute.

Το λογισμικό SAS χρησιμοποιείται κυρίως για αναλύσεις στα social media και στον τομέα του Marketing.

2 SPSS

2.1 Εισαγωγή στο SPSS

Αναπτύχθηκε το 1965 στο Πανεπιστήμιο Στάνφορντ του Σαν Φρανσίσκο από 2 φοιτητές πολιτικών επιστημών, τον Νόρμαν Νίε και Ντέιλ Μπέν. Λόγω των ελάχιστων και μη ανεπτυγμένων διαθέσιμων λογισμικών που υπήρχαν, οι 2 φοιτητές ανέπτυξαν την ιδέα δημιουργίας ενός νέου προγράμματος. Η ιδέα ξεκίνησε να υλοποιείται με γλώσσα προγραμματισμού FORTRAN, σε έναν υπολογιστή τύπου IBM 7090.

Η 1^η έκδοση έγινε γνωστή το 1967 όπου και το λογισμικό ονομάστηκε Statistical package for the social science (SPSS). Το 1970 πραγματοποιείται η ίδρυση της εταιρίας με έδρα πια το Πανεπιστήμιο του Σικάγο. Στην ομάδα έχει ήδη ενταχθεί και ο Χάντλεϊ Χάλ. Το 1975 κυκλοφορεί η 6^η έκδοση και από τότε μέχρι το 1981 θα κυκλοφορήσουν οι εκδόσεις 7,8 και 9.

Η πρώτη έκδοση για PC κυκλοφόρησε το 1983 και ήταν για την MS-Dos. Μετονομάστηκε σε Superior Performance Software System. Το SPSS 5 είναι η 1^η έκδοση η οποία κυκλοφόρησε για περιβάλλον Windows. Η έκδοση SPSS 16.0 τρέχει σε περιβάλλοντα Windows, Mac και Linux .

Αποκτήθηκε από την IBM το 2009 για το ποσό των 1.26 δις δολάρια. Η τελευταία έκδοση του SPSS, η έκδοση 21 είναι σχεδιασμένη σε αντικειμενοστραφή βάση με την βοήθεια της Java.

2.2 Ερωτηματολόγιο

Η καλύτερη συλλογή δεδομένων αποτελεί ένα ερωτηματολόγιο. Δημιουργούμε ένα ερωτηματολόγιο το οποίο συμπληρώθηκε από 20 άτομα και απαντήσαν σε 8 ερωτήσεις ο καθένας. Με την χρήση του ερωτηματολογίου, θα εκμεταλευτούμε τα δεδομένα του για να δούμε κάποιες βασικές εντολές του

λογισμικού SPSS, καθώς και να δημιουργήσουμε γραφήματα. Τα δεδομένα αυτά θα εισαχθούν στο πρόγραμμα SPSS.

Ερωτηματολόγιο

Φύλο

- Άνδρας
- Γυναίκα

Ηλικία

Η απάντησή σας

Ύψος

Η απάντησή σας

Βάρος

Η απάντησή σας

Εκπαίδευση

- Δημοτικό
- Γυμνάσιο
- Λύκειο
- ΙΕΚ
- ΤΕΙ
- ΑΕΙ
- Μεταπτυχιακό

Επάγγελμα

- Δημόσιος Υπάλληλος
- Ιδιωτικός Υπάλληλος
- Φοιτητής
- Αγρότης
- Επιχειρηματίας
- Άνεργος

Ξένη Γλώσσα

- Αγγλικά
- Γαλλικά
- Ισπανικά
- Γερμανικά
- Άλλη
- Καμία

Καπνιστής

- ΝΑΙ
- ΟΧΙ

Πίνακας δεδομένων

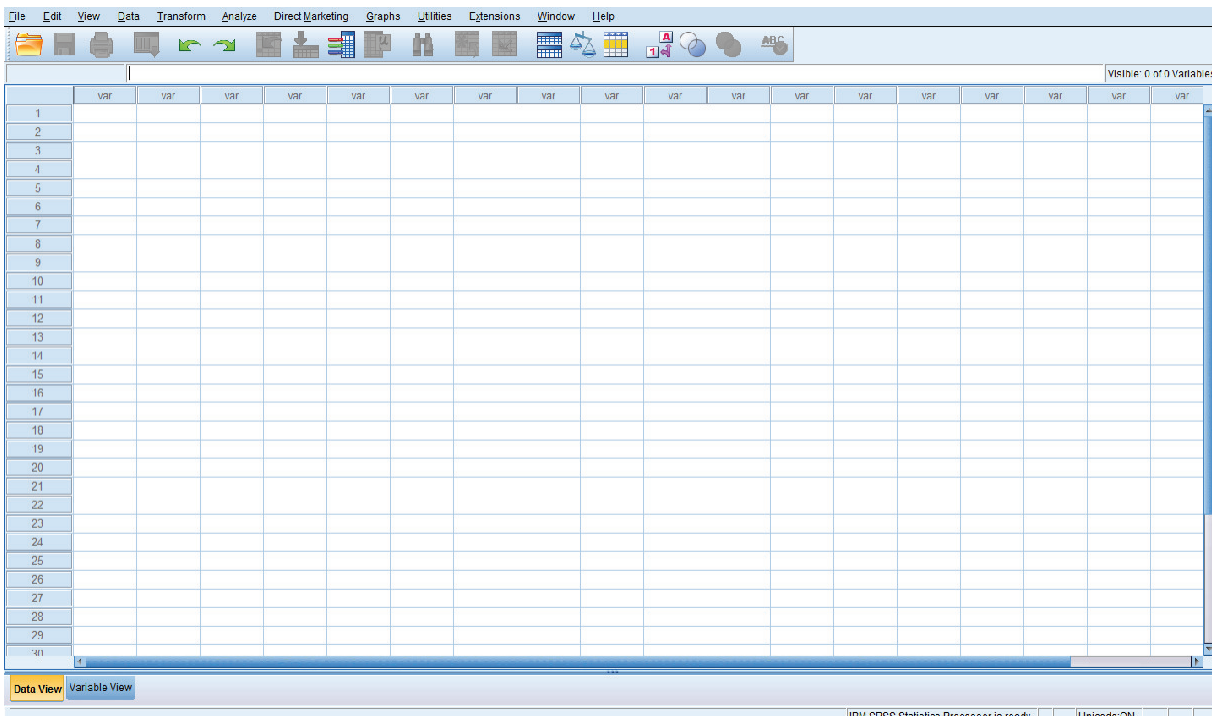
N	Φύλο	Ηλικία	Ύψος	Βάρος	Εκπαίδευση	Επάγγελμα	Γλώσσα	Κάπνισμα
1	Ανδρας	23	1.82 cm	85 kg	Λύκειο	Ιδ Υπάλλ	Αγγλικά	Ναι
2	Γυναίκα	18	1.65 cm	60kg	ΑΕΙ	Φοιτητής	Γαλλικά	Όχι
3	Ανδρας	31	1.75cm	70 kg	Γυμνάσιο	Άνεργος	Καμία	Ναι
4	Ανδρας	29	1.92 cm	88 kg	Μεταπτυχιακ ό	Επιχειρηματί ας	Άλλη	Όχι
5	Γυναίκα	40	1.65 cm	72 kg	ΤΕΙ	Δημ Υπάλλ	Γαλλικά	Όχι
6	Γυναίκα	25	1.65 cm	70 kg	ΙΕΚ	Άνεργος	Ισπανικά	Όχι
7	Ανδρας	31	1.79 cm	95 kg	Δημοτικό	Αγρότης	Καμία	Ναι
8	Ανδρας	19	1.69 cm	70 kg	Γυμνάσιο	Επιχειρηματί ας	Αγγλικά	Όχι
9	Ανδρας	27	1.86 cm	98 kg	ΑΕΙ	Ιδ Υπάλλ	Γερμανικά	Όχι
10	Γυναίκα	50	1.65 cm	65 kg	ΤΕΙ	Ιδ Υπάλλ	Αγγλικά	Ναι
11	Γυναίκα	32	1.69 cm	76 kg	ΑΕΙ	Δημ Υπάλλ	Αγγλικά	Όχι
12	Ανδρας	49	1.73 cm	86 kg	Λύκειο	Άνεργος	Καμία	Όχι
13	Γυναίκα	20	1.75 cm	65 kg	ΤΕΙ	Φοιτητής	Γαλλικά	Όχι
14	Γυναίκα	26	1.71 cm	60 kg	Μεταπτυχιακ ό	Άνεργος	Γερμανικά	Ναι
15	Γυναίκα	38	1.62 cm	69 kg	ΙΕΚ	Επιχειρηματί ας	Αγγλικά	Ναι
16	Ανδρας	25	1.87 cm	79 kg	ΤΕΙ	Ιδ Υπάλλ	Καμία	Όχι
17	Γυναίκα	26	1.64 cm	69 kg	ΙΕΚ	Δημ Υπάλλ	Αγγλικά	Όχι
18	Ανδρας	18	1.68 cm	79 kg	Λύκειο	Αγρότης	Αγγλικά	Όχι
19	Γυναίκα	19	1.63 cm	72 kg	ΤΕΙ	Φοιτητής	Ισπανικά	Όχι
20	Ανδρας	25	1.76 cm	70 kg	ΑΕΙ	Φοιτητής	Γερμανικά	Ναι

Ο παραπάνω πίνακας θα ονομάζεται στο εξής πίνακας δεδομένων. Για να επεξεργαστούμε τα δεδομένα του ερωτηματολπίου στο SPSS θα πρέπει να υπάρχουν σε μορφή πίνακα δεδομένων. Όπως παρατηρούμε ο πίνακας αποτελείται από έναν αριθμό στηλών και γραμμών. Κάθε στήλη είναι μια μεταβλητή ενώ κάθε γραμμή είναι ένα ερωτηματολόγιο.

2.3 Παρουσίαση SPSS σε Windows

2.3.1 Εκκίνηση SPSS

Είμαστε έτοιμοι να ανοίξουμε το πρόγραμμα SPSS και αρχίζουμε να περάσουμε τα δεδομένα από το ερωτηματολόγιο. Κατά την εκκίνηση του προγράμματος εμφανίζεται το παράθυρο του SPSS DATA EDITOR.

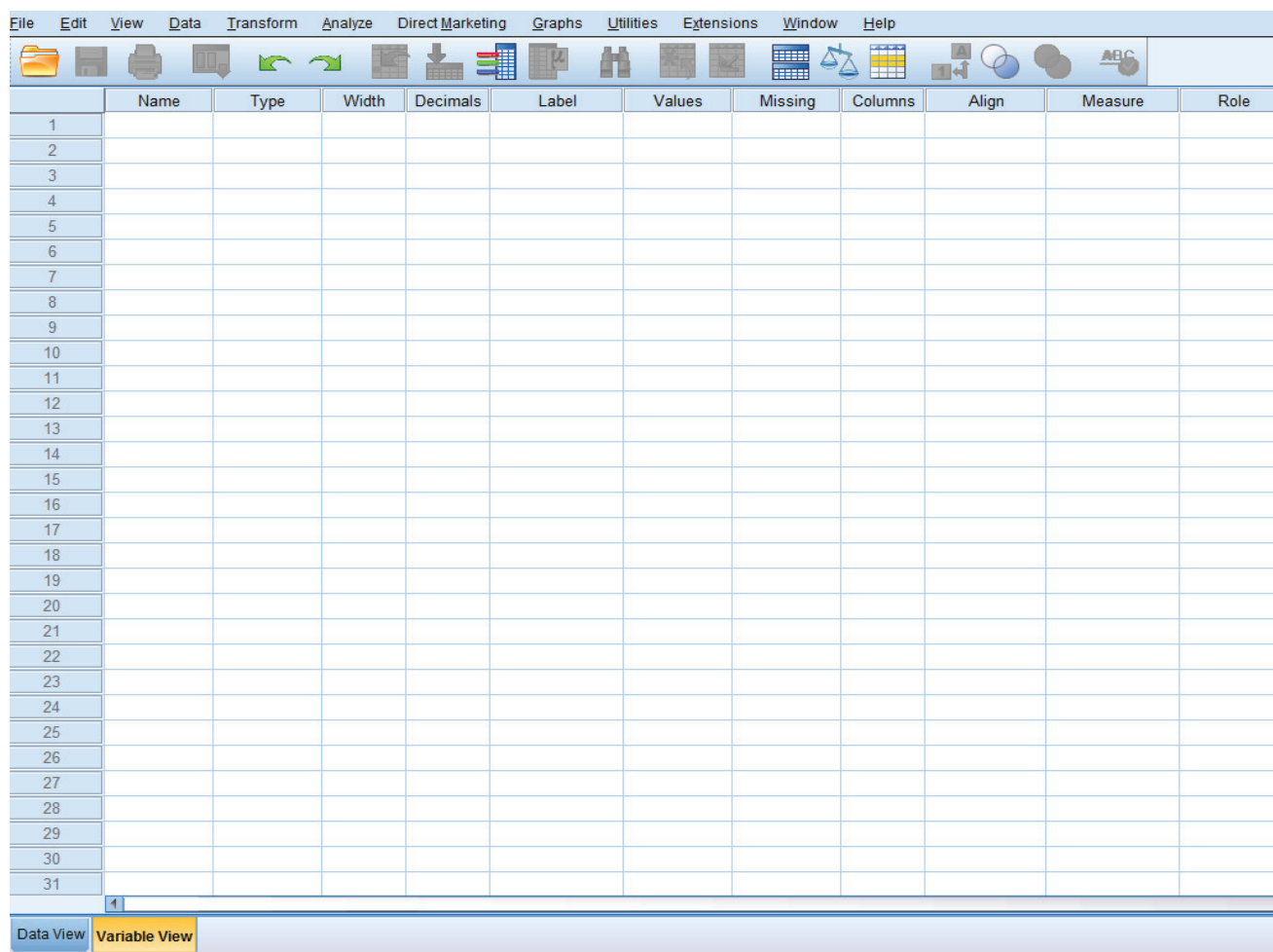


SPSS DATA EDITOR

Το παράθυρο είναι αρχικά κενό. Μοιάζει αρκετά με το Excel. Αποτελείται από ένα πλέγμα γραμμών και στηλών στο οποίο καταχωρούνται με πληκτρολόγηση τα δεδομένα. Επιλέγουμε την **Variable View** για την εισαγωγή των δεδομένων

2.3.2 Data Editor (variable view)

Variable view: Με την επιλογή Variable View στο παράθυρο του DATA EDITOR Εμφανίζονται οι πληροφορίες που ενσωματώνονται στις μεταβλητές που θα ορίσουμε στην συνέχεια. Οι πληροφορίες αυτές αφορούν τα ονόματα των μεταβλητών, την variable label (λεκτική περιγραφή), τι τύπος είναι (π.χ. αλφαριθμητική ή αριθμητική κλπ.), επίπεδο μέτρησης και τις missing values.



DATA EDITOR (VARIABLE VIEW)

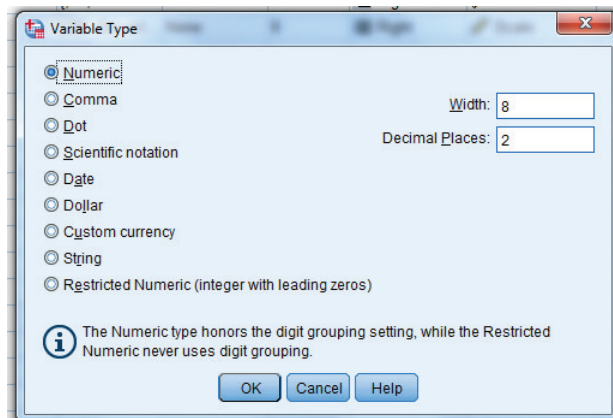
Στην συνέχεια περνάμε όλα τα δεδομένα από το ερωτηματολόγιο στον DATA EDITOR σε Variable View.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Φύλο	Numeric	8	2		{,00, Δεν απ...	None	8	Right	Scale	Input
2	Ηλικία	Numeric	8	2		None	None	8	Right	Scale	Input
3	Ύψος	Numeric	8	2		None	None	8	Right	Scale	Input
4	Βάρος	Numeric	8	2		None	None	8	Right	Scale	Input
5	Έκπαίδευση	Numeric	8	2		{,00, Δεν απ...	None	8	Right	Scale	Input
6	Επάγγελμα	Numeric	8	2		{,00, Δεν απ...	None	8	Right	Scale	Input
7	Ξένη_Γλώσσα	Numeric	8	2		{,00, Δεν απ...	None	8	Right	Scale	Input
8	Καπνιστής	Numeric	8	2		{,00, Δεν απ...	None	8	Right	Scale	Input
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											

Data Editor με δεδομένα

Οι ιδιότητες των μεταβλητών που εμφανίζονται στην επιλογή **Variable View** είναι:

- 1) Το όνομα μιας μεταβλητής (Variable name π.χ. Φύλο, Ηλικία κλπ.)
- 2) Τύπος της μεταβλητής (Type).

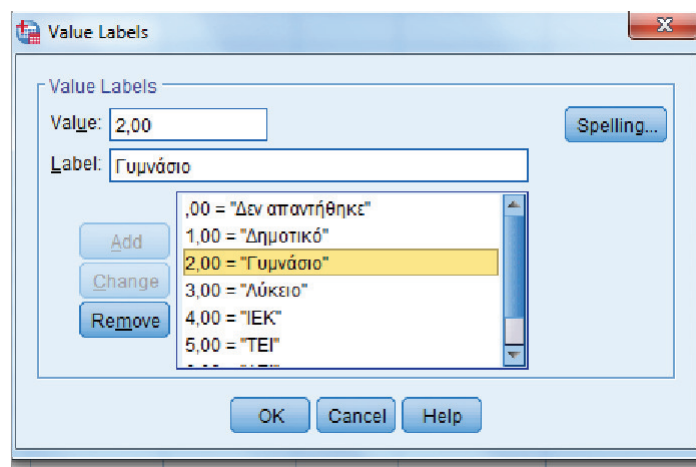


Δήλωση τύπου μεταβλητής

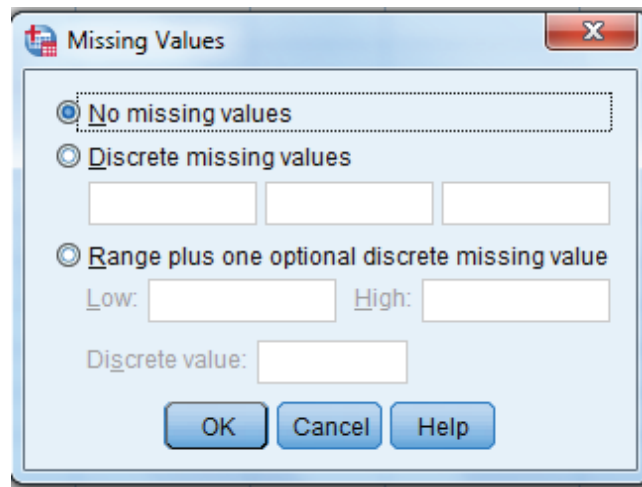
Επιλέγουμε τον τύπο της μεταβλητής που θέλουμε να χρησιμοποιήσουμε.

Decimel: Ο αριθμός δεκαδικών ψηφίων της μεταβλητής.

Values: Η περιγραφή των τιμών μιας μεταβλητής με την βοήθεια ετικετών



Missing Values: Για να ορίσουμε τις μη περασμένες ή απούσες τιμές.



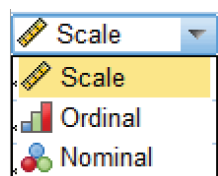
Columns: Το πλάτος της μεταβλητής που καταχωρείται στον Data Editor.

Align: Εδώ ορίζουμε την στοίχιση των τιμών της μεταβλητής στα κελιά.

Measurement Level: Η κλίμακα στην οποία μπορεί να μετρηθεί η μεταβλήτη.

Υπάρχουν 3 κλίμακες:

- **Scale**
- **Ordinal**
- **Nominal**



Στην συνέχεια θα περάσουμε όλα τα δεδομένα στο παράθυρο Data

View. Το αποτέλεσμα θα είναι:

	Φύλο	Ηλικία	Ύψος	Βάρος	Έκπαιδευση	Επάγγελμα	Ξένη_Γλώσσα	Καπνιστής	var	var	var
1	1,00	23	182,00	85,00	3,00	2,00	1,00	1,00			
2	2,00	18	165,00	60,00	6,00	4,00	2,00	2,00			
3	1,00	31	175,00	70,00	2,00	6,00	6,00	1,00			
4	1,00	29	192,00	88,00	7,00	5,00	5,00	2,00			
5	2,00	40	165,00	72,00	5,00	1,00	2,00	2,00			
6	2,00	25	165,00	70,00	4,00	6,00	3,00	2,00			
7	1,00	31	179,00	95,00	1,00	4,00	6,00	1,00			
8	1,00	19	169,00	70,00	2,00	5,00	1,00	2,00			
9	1,00	27	186,00	98,00	6,00	2,00	4,00	2,00			
10	2,00	50	165,00	65,00	5,00	2,00	1,00	1,00			
11	2,00	32	169,00	76,00	6,00	1,00	1,00	2,00			
12	1,00	49	173,00	86,00	3,00	6,00	6,00	2,00			
13	2,00	20	175,00	65,00	5,00	3,00	2,00	2,00			
14	2,00	26	171,00	60,00	7,00	6,00	4,00	1,00			
15	2,00	38	162,00	69,00	4,00	5,00	1,00	1,00			
16	1,00	25	187,00	79,00	5,00	2,00	6,00	2,00			
17	2,00	26	164,00	69,00	4,00	1,00	1,00	2,00			
18	1,00	18	168,00	79,00	3,00	4,00	1,00	2,00			
19	2,00	19	163,00	72,00	5,00	3,00	3,00	2,00			
20	1,00	25	176,00	70,00	6,00	3,00	4,00	1,00			
21											
22											
23											
24											
25											
26											
27											
28											
29											

Data view με τα δεδομένα

Στην συνέχεια από το μενού θα επιλέξουμε **View** και **Value Labels** και το αποτέλεσμα θα είναι το εξής:

*Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

20 : Καπνιστής 1,00

	Φύλο	Ηλικία	Ύψος	Βάρος	Εκπαίδευση	Επάγγελμα	Ξένη_Γλώσσα	Καπνιστής	var
1	Ανδρας	23	182,00	85,00	Λύκειο	Ιδιωικός Υπάλληλος	Αγγλικά	Ναί	
2	Γυναίκα	18	165,00	60,00	ΑΕΙ	Αγρότης	Γαλλικά	Όχι	
3	Ανδρας	31	175,00	70,00	Γυμνάσιο	Ανεργος	Καμία	Ναί	
4	Ανδρας	29	192,00	88,00	Μεταπτυχιακό	Επιχειρηματίας	Άλλη	Όχι	
5	Γυναίκα	40	165,00	72,00	ΤΕΙ	Δημόσιος Υπάλληλος	Γαλλικά	Όχι	
6	Γυναίκα	25	165,00	70,00	ΙΕΚ	Ανεργος	Ισπανικά	Όχι	
7	Ανδρας	31	179,00	95,00	Δημοτικό	Αγρότης	Καμία	Ναί	
8	Ανδρας	19	169,00	70,00	Γυμνάσιο	Επιχειρηματίας	Αγγλικά	Όχι	
9	Ανδρας	27	186,00	98,00	ΑΕΙ	Ιδιωικός Υπάλληλος	Γερμανικά	Όχι	
10	Γυναίκα	50	165,00	65,00	ΤΕΙ	Ιδιωικός Υπάλληλος	Αγγλικά	Ναί	
11	Γυναίκα	32	169,00	76,00	ΑΕΙ	Δημόσιος Υπάλληλος	Αγγλικά	Όχι	
12	Ανδρας	49	173,00	86,00	Λύκειο	Ανεργος	Καμία	Όχι	
13	Γυναίκα	20	175,00	65,00	ΤΕΙ	Φοιτητής	Γαλλικά	Όχι	
14	Γυναίκα	26	171,00	60,00	Μεταπτυχιακό	Ανεργος	Γερμανικά	Ναί	
15	Γυναίκα	38	162,00	69,00	ΙΕΚ	Επιχειρηματίας	Αγγλικά	Ναί	
16	Ανδρας	25	187,00	79,00	ΤΕΙ	Ιδιωικός Υπάλληλος	Καμία	Όχι	
17	Γυναίκα	26	164,00	69,00	ΙΕΚ	Δημόσιος Υπάλληλος	Αγγλικά	Όχι	
18	Ανδρας	18	168,00	79,00	Λύκειο	Αγρότης	Αγγλικά	Όχι	
19	Γυναίκα	19	163,00	72,00	ΤΕΙ	Φοιτητής	Ισπανικά	Όχι	
20	Ανδρας	25	176,00	70,00	ΑΕΙ	Φοιτητής	Γερμανικά	Ναί	
21									
22									
23									
24									
25									
26									
27									
28									
29									

Data View Variable View

Data editor σε εμφάνιση Value Labels

Παραπάνω βλέπουμε τα στοιχεία από τον πίνακα δεδομένων που έχουμε από τις απαντήσεις του ερωτηματολογίου.

Παρακάτω θα δούμε βασικές εντολές του προγράμματος.

2.4 Βασικές Εντολές

Παρακάτω θα αναφέρουμε και θα αναλύσουμε σε εφαρμογή τις 6 βασικές εντολές του λογισμικού. Αυτές είναι: Εντολή Recode, εντολή Compute, εντολή Count, εντολή Frequencies, εντολή Descriptives και εντολή Crosstabs.

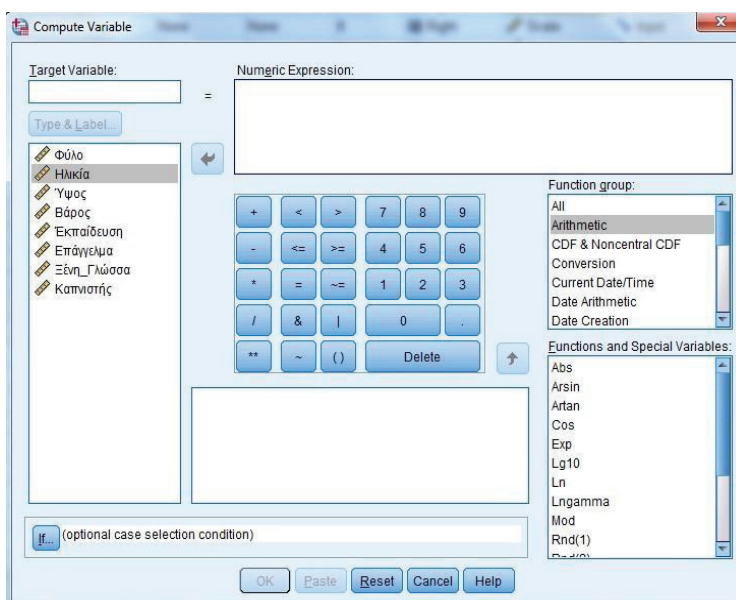
2.4.1 Εντολή Compute

Με την εντολή αυτή μπορούμε να υπολογίσουμε νέες τιμές μιας καινούριας μεταβλητής ή να υπολογίσουμε πάλι τιμές για υπάρχουσες μεταβλητές. Η μεταβλητή μπορεί να είναι αριθμητικού ή αλφαριθμητικού τύπου. Με την επιλογή **if** μπορούμε να δημιουργήσουμε μια συνθήκη για τον έλεγχο περιπτώσεων που θα γίνεται ο υπολογισμός.

Από την γραμμή του μενού επιλέγουμε

Transform
Compute Variable...

Θα ανοίξει το παράθυρο:



Compute Variable

Στην συνέχεια θα δηλώσουμε:

- 1) Το όνομα της μεταβλητής (στο παράδειγμα μας είναι η μεταβλητή, Ηλικία) η μιας νέας μεταβλητής στο πεδίο **Target Value**.
- 2) Και στο πεδίο **Numeric Expression** δηλώνουμε τον ορισμό της μεταβλητής.

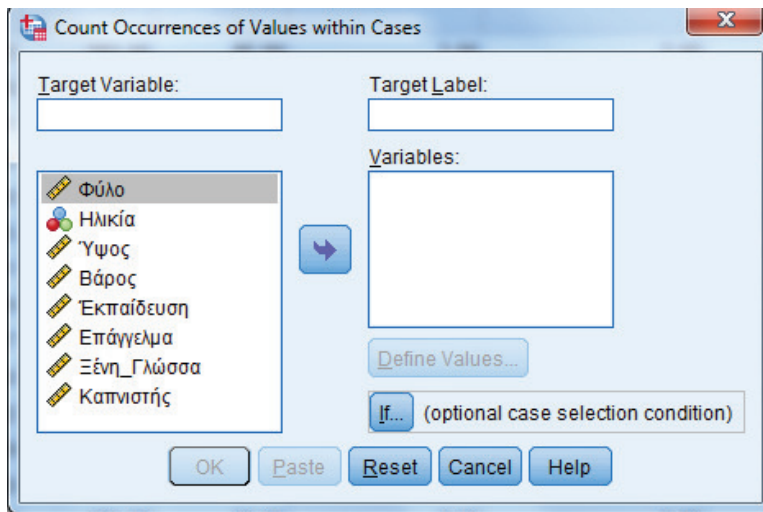
2.4.2 Εντολή Count

Με την εντολή αυτή μπορούμε να υπολογίσουμε τον αριθμό των εμφανίσεων μιας ή περισσότερων τιμών σε μια λίστα μεταβλητών.

Από την γραμμή του μενού επιλέγουμε:

Transform

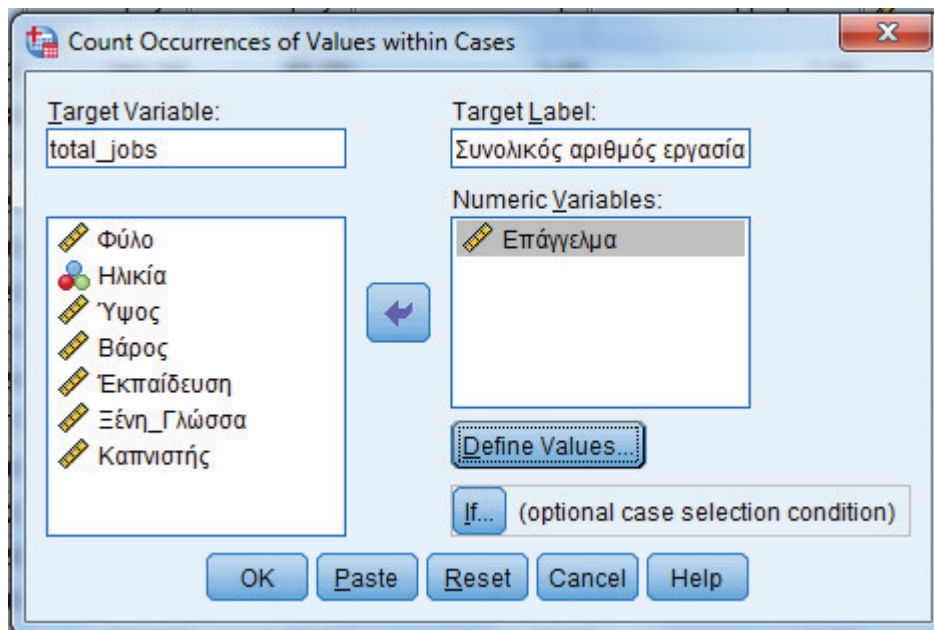
Count Values within Cases...



Στο παράθυρο που εμφανίστηκε στο πεδίο Target Variable θα συμπληρώσουμε την μεταβλητή που θέλουμε, υπάρχουσα ή δημιουργούμε κάποια καινούργια.

Στο πεδίο **Target Label** θα συμπληρώσουμε μια σχετική περιγραφή με το όνομα της μεταβλητής που βάλουμε στο πεδίο **Target Variable**.

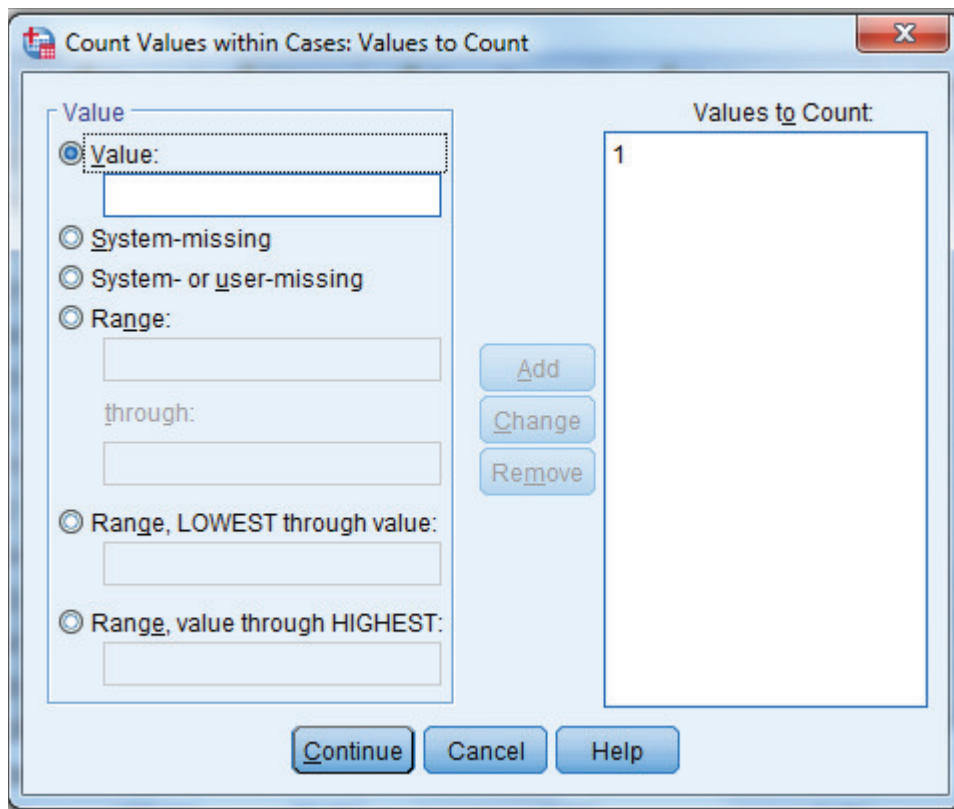
Έστω δημιουργούμε μια καινούρια μεταβλητή total_jobs και την εισάγουμε στο παράθυρο. Επίσης θα εισάγουμε και την αντίστοιχη περιγραφή της μεταβλητής (Συνολικός αριθμός εργασίας). Στην συνέχεια θα μεταφέρουμε τις μεταβλητές που θέλουμε (στην περίπτωση μας την μεταβλητή εκπαίδευση) στο πεδίο Variables.



Εισαγωγή μεταβλητής στο πεδίο Variables

Θα επιλέξουμε την επιλογή **Define Values**.

Στο παράθυρο που θα ανοίξει θα προσθέσουμε την μεταβλητή που θέλουμε, π.χ. το 1 στο **Values to Count** ώστε να υπολογιστούν πόσες αντίστοιχες συγκεκριμένες εργασίες υπάρχουν.



Πρόσθεση μεταβλητών για υπολογισμό

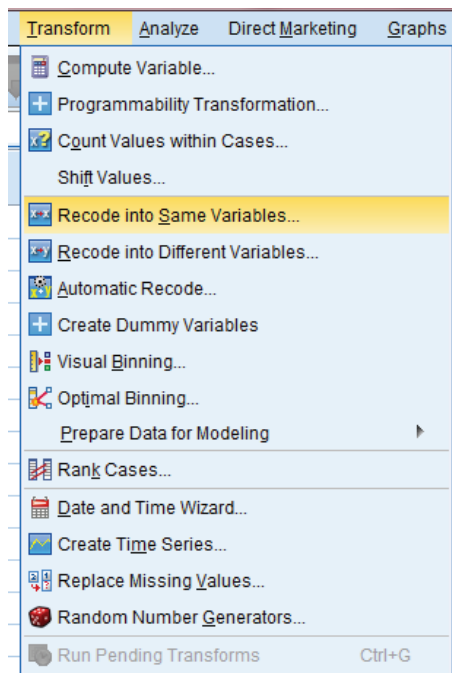
2.4.3 Εντολή Recode

Μέσω της εντολής Recode, οι τιμές μια μεταβλητής για την ανάγκη μιας ανάλυσης μπορούν επανακωδικοποιηθούν. Τα αποτελέσματα που προκύπτουν καταχωρούνται σαν νέες τιμές στην μεταβλητή ή ως τιμές μια καινούριας μεταβλητής. Για την επανακωδικοποίηση στην ίδια μεταβλητή θα δουλέψουμε ως εξής:

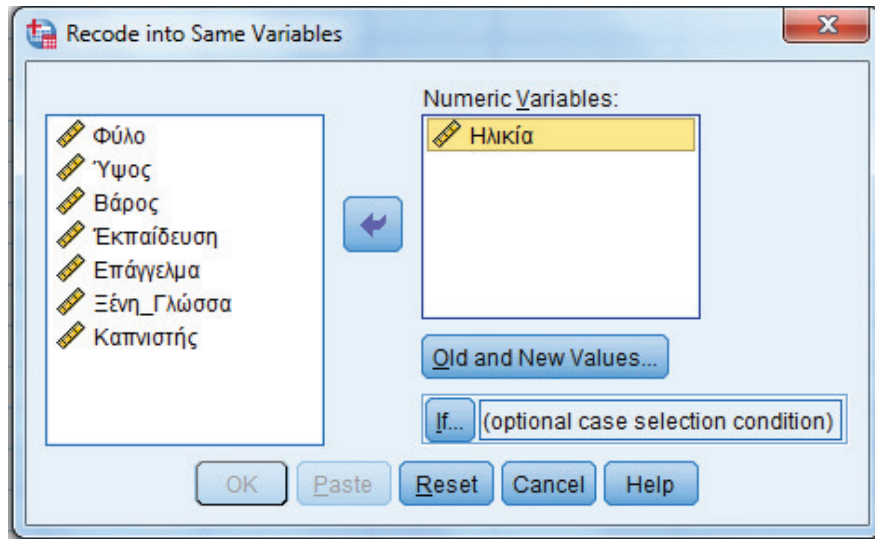
Από την γραμμή του μενού θα επιλέξουμε

Transform

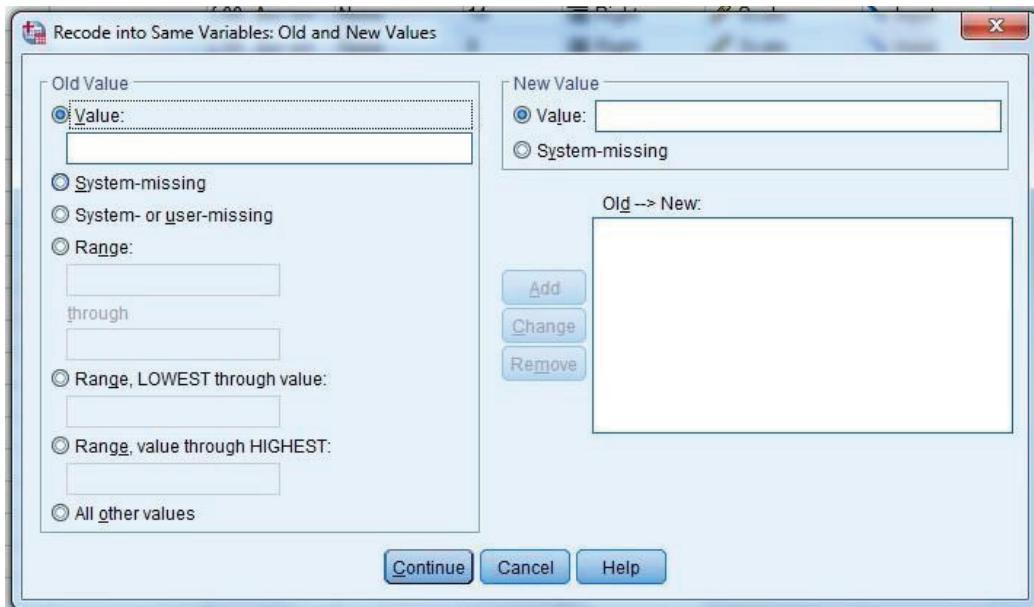
Recode Into same Variables..



Ανοίγει το παράθυρο **Recode Into Same Variable** και επιλέγουμε την μεταβλητή **Ηλικία**.



Στην συνέχεια επιλέγουμε **Old and New Values** και ορίζουμε πως θα γίνουν οι επανακωδικοποιήσεις.

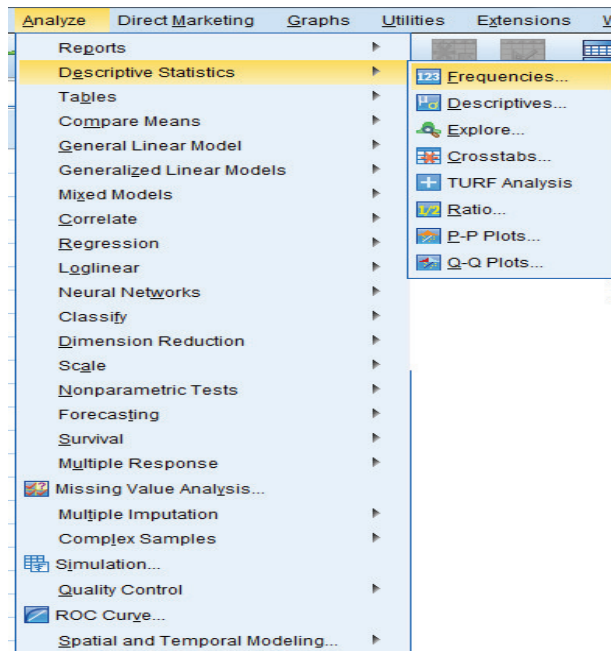


2.4.4 Εντολή Frequencies

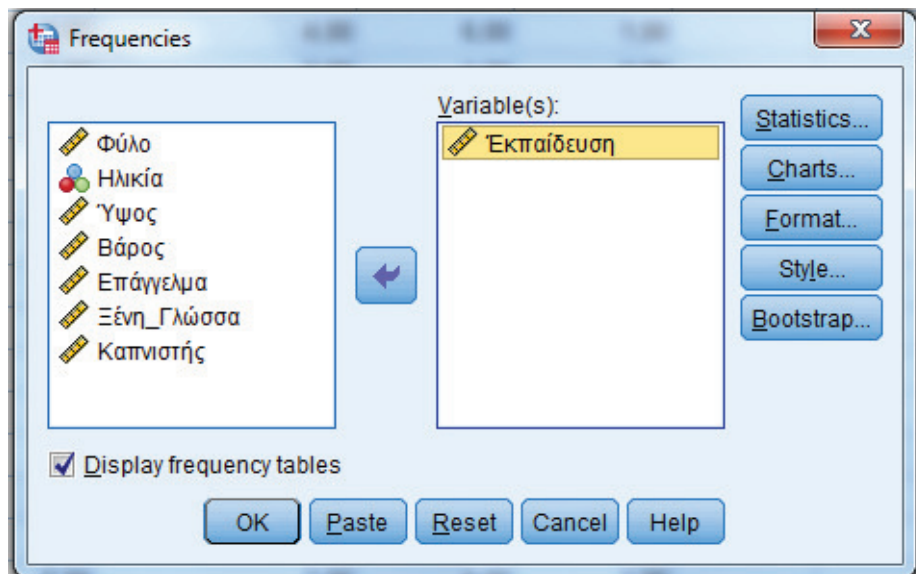
Η εντολή Frequencies παράγει πίνακες συχνοτήτων και διαγραμματικές απεικονίσεις για τις μεταβλητές. Επίσης χρησιμοποιείται για τον υπολογισμό περιγραφικών στατιστικών μέτρων ποσοτικών μεταβλητών. Είναι ο πρώτος έλεγχος σε περίπτωση που θα εντοπιστούν κάποιες λανθασμένες τιμές και θα πρέπει να διορθωθούν για την μετέπειτα στατιστική ανάλυση τους.

Για να χρησιμοποιήσουμε την εντολή Frequencies: Από την γραμμή ου μενού θα πιλέξουμε :

Analyze
Descriptive Statistics
Frequencies



Για παράδειγμα θα επιλέξουμε την μεταβλητή **Εκπαίδευση**.



Μπορούν να επιλεγούν τα πλήκτρα για:

Statistics: Για να υπολογιστούν τα στατιστικά μέτρα των ποσοτικών μεταβλητών.

Charts: Για την δημιουργία ραβδογραμμμάτων, κυκλικών διαγραμμμάτων και ιστογραμμμάτων.

Format: Για τον ορισμό τρόπου διάταξης των τιμών στους πίνακες συχνοτήτων.

Στην συνέχεια επιλέγοντας το Ok δημιουργείται ένα πίνακας αποτελεσμάτων σε ένα νέο αρχείο Output.

		Έκπαίδευση			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Δημοτικό	1	5,0	5,0	5,0
	Γυμνάσιο	2	10,0	10,0	15,0
	Λύκειο	3	15,0	15,0	30,0
	ΙΕΚ	3	15,0	15,0	45,0
	ΤΕΙ	5	25,0	25,0	70,0
	ΑΕΙ	4	20,0	20,0	90,0
	Μεταπτυχιακό	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Στην στήλη **Valid** εμφανίζονται οι τιμές της μεταβλητής.

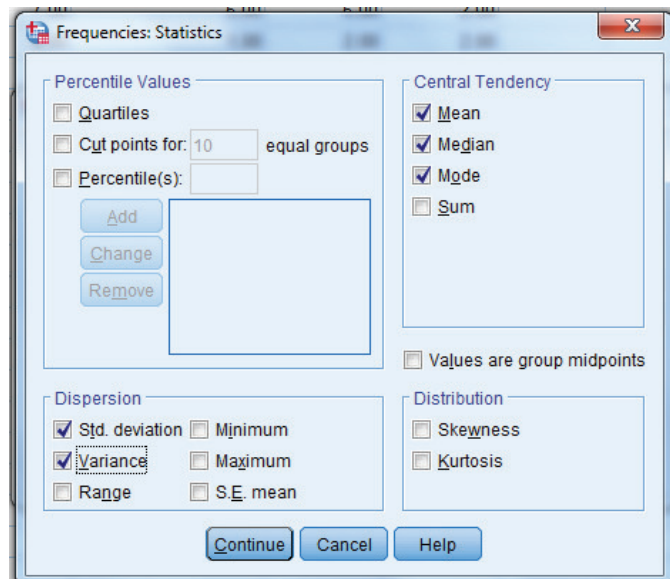
Στην στήλη **Frequency** εμφανίζονται το πόσες φορές εμφανίζεται η κάθε τιμή. Στην στήλη **Percent** εμφανίζεται το πόσο τοις εκατό για κάθε μεταβλητή.

Στην στήλη **Valid Percent** εμφανίζεται το ποσοστό επί τοις εκατό της κάθε τιμής.

Στην στήλη **Comulative Percent** εμφανίζεται το αθροιστικό ποσό.

Για να υπολογίσουμε τα στατιστικά μεγέθη θα επιλέξουμε το πλήκτρο **Statistics** και το πλήκτρο

Charts για την εμφάνιση των αντίστοιχων διαγραμμάτων.



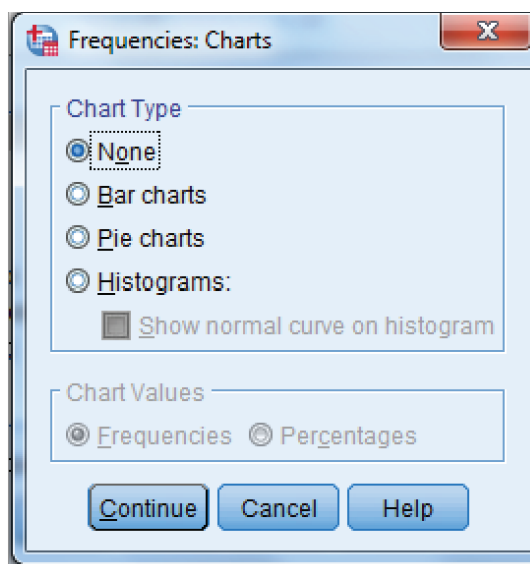
Επιλογή στατιστικών

Frequencies

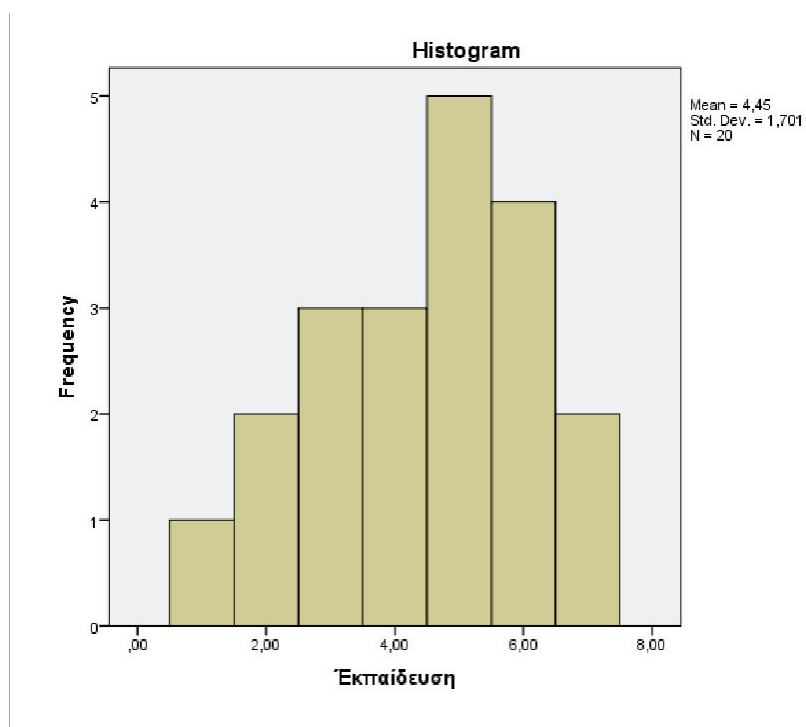
Statistics		
Έκπαίδευση		
N	Valid	20
	Missing	0
Mean		4,4500
Median		5,0000
Mode		5,00
Std. Deviation		1,70062
Variance		2,892

Charts

Από το παρακάτω παράθυρο θα επιλέξουμε τα γραφήματα που θέλουμε να μας εμφανιστούν.



Επιλογή Chart



Histogram

2.4.5 Εντολή Descriptives

Η εντολή **Descriptives** υπολογίζει τα στατιστικά μέτρα για 1 ή περισσότερες ποσοτικές μεταβλητές και παράγει τις τιμές τους. Οι τιμές αυτές αποθηκεύονται σε ένα αρχείο δεδομένων και στην συνέχεια μπορούν να χρησιμοποιηθούν για στατιστική επεξεργασία.

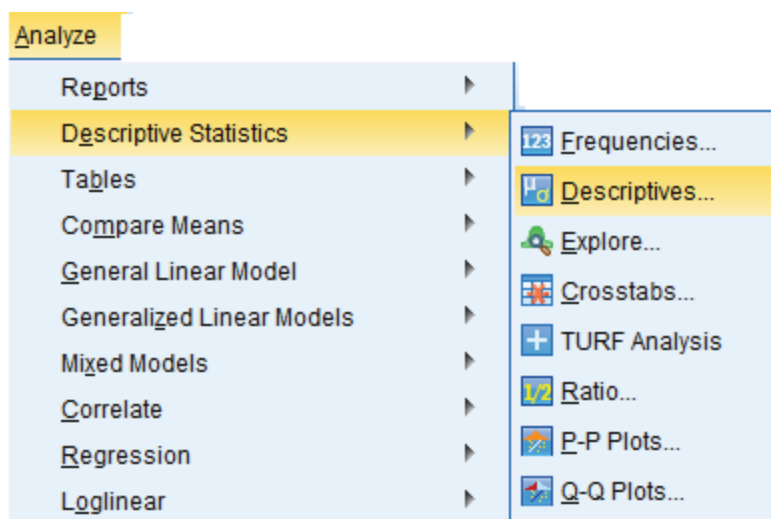
Για να χρησιμοποιήσουμε την εντολή **Descriptives**:

Από την γραμμή του μενού επιλέγουμε:

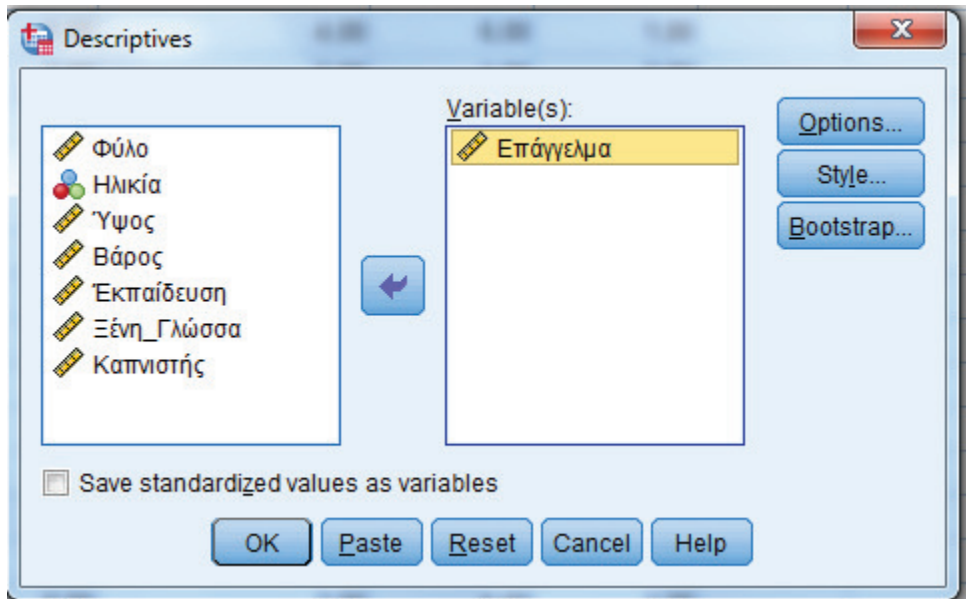
Analyze

Descriptive Statistics

Descriptives...



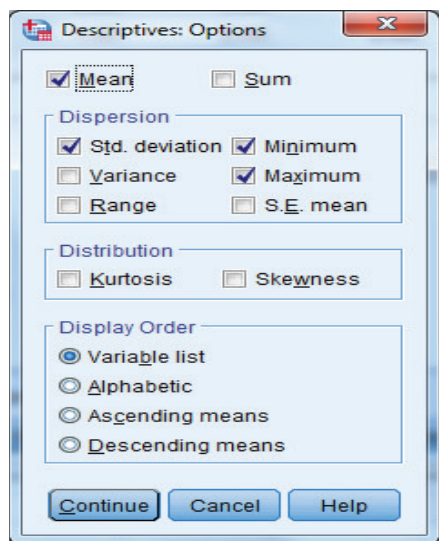
Εμφανίζεται το παράθυρο **Descriptives**



Επιλογή Μεταβλητής

Επιλέγουμε την μεταβλητή **Επάγγελμα** και θα υπολογίσουμε τον μέσο όρο, την τυπική απόκλιση, την ελάχιστη και μέγιστη τιμή για την μεταβλητή.

Επιλέγουμε το πλήκτρο **Options**.



Με τις επιλογές **Continue** και **Ok** δημιουργείται ο παρακάτω πίνακας σε ένα αρχείο Output.

→ Descriptives

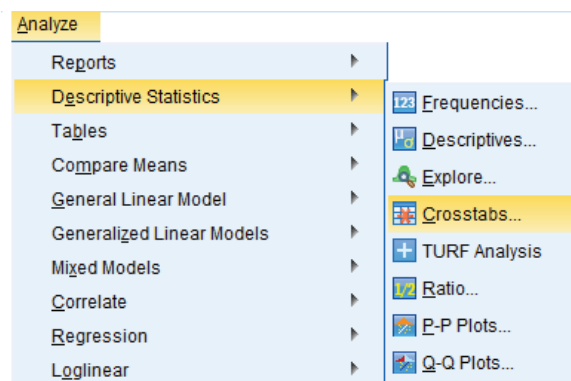
	N	Minimum	Maximum	Mean	Std. Deviation
Επάγγελμα	20	1,00	6,00	3,5500	1,79106
Valid N (listwise)	20				

2.4.6 Εντολή Crosstabs

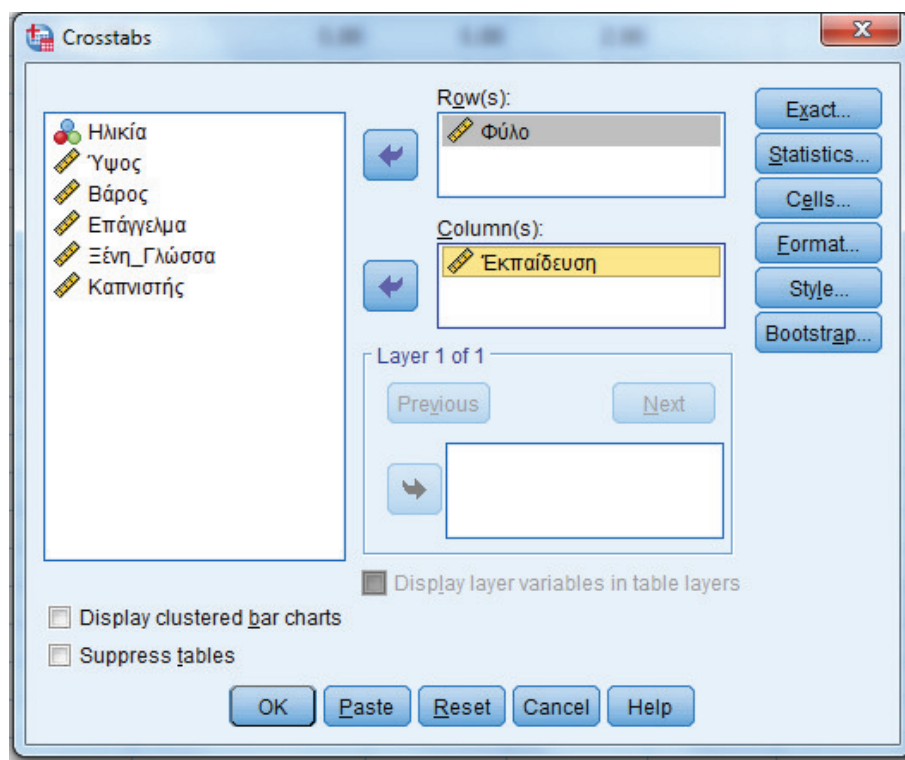
Με την εντολή αυτή παράγουμε διδιάστατους ή πολυδιάστατους πίνακες συχνοτήτων. Πρέπει να ορίσουμε τουλάχιστον 2 μεταβλητές μια για τις γραμμές και μια για τις στήλες του πίνακα που θα δημιουργηθεί.

Για να κατασκευάσουμε έναν πίνακα διπλής εισόδου Από την γραμμή μενού επιλέγουμε:

Analyze
Descriptive Statistics
Crosstabs



Εμφανίζεται το παράθυρο **Crosstabs**.



Επιλογή Μεταβλητής

Για παράδειγμα θα επιλέξουμε τις μεταβλητές **Φύλο** και **Εκπαίδευση**.

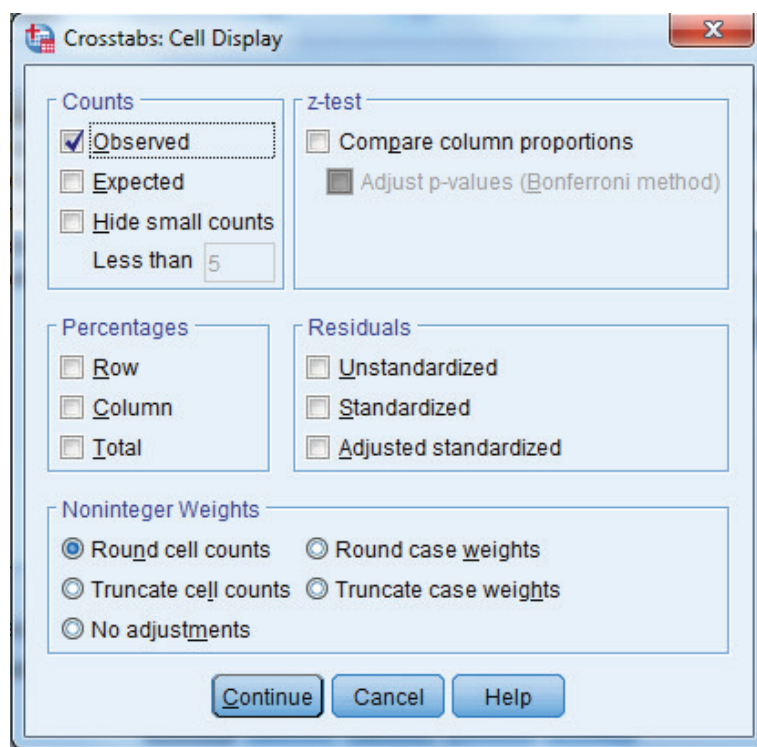
Στον παρακάτω πίνακα βλέπουμε τις αντιστοιχίες της μεταβλητής **Φύλο** με την μεταβλητή **Εκπαίδευση**.

Φύλο * Έκπαίδευση Crosstabulation

Count

		Έκπαίδευση							Total
		Δημοτικό	Γυμνάσιο	Λύκειο	IEK	TEI	AEI	Μεταπτυχιακό	
Φύλο	Άνδρας	1	2	3	0	1	2	1	10
	Γυναίκα	0	0	0	3	4	2	1	10
Total		1	2	3	3	5	4	2	20

Με την επιλογή του πλήκτρου Cells εμφανίζεται ένα παράθυρο στο οποίο μπορούμε να επιλέξουμε τα στοιχεία που θέλουμε να αναγράφονται στον πίνακα που θα εμφανίσουμε.

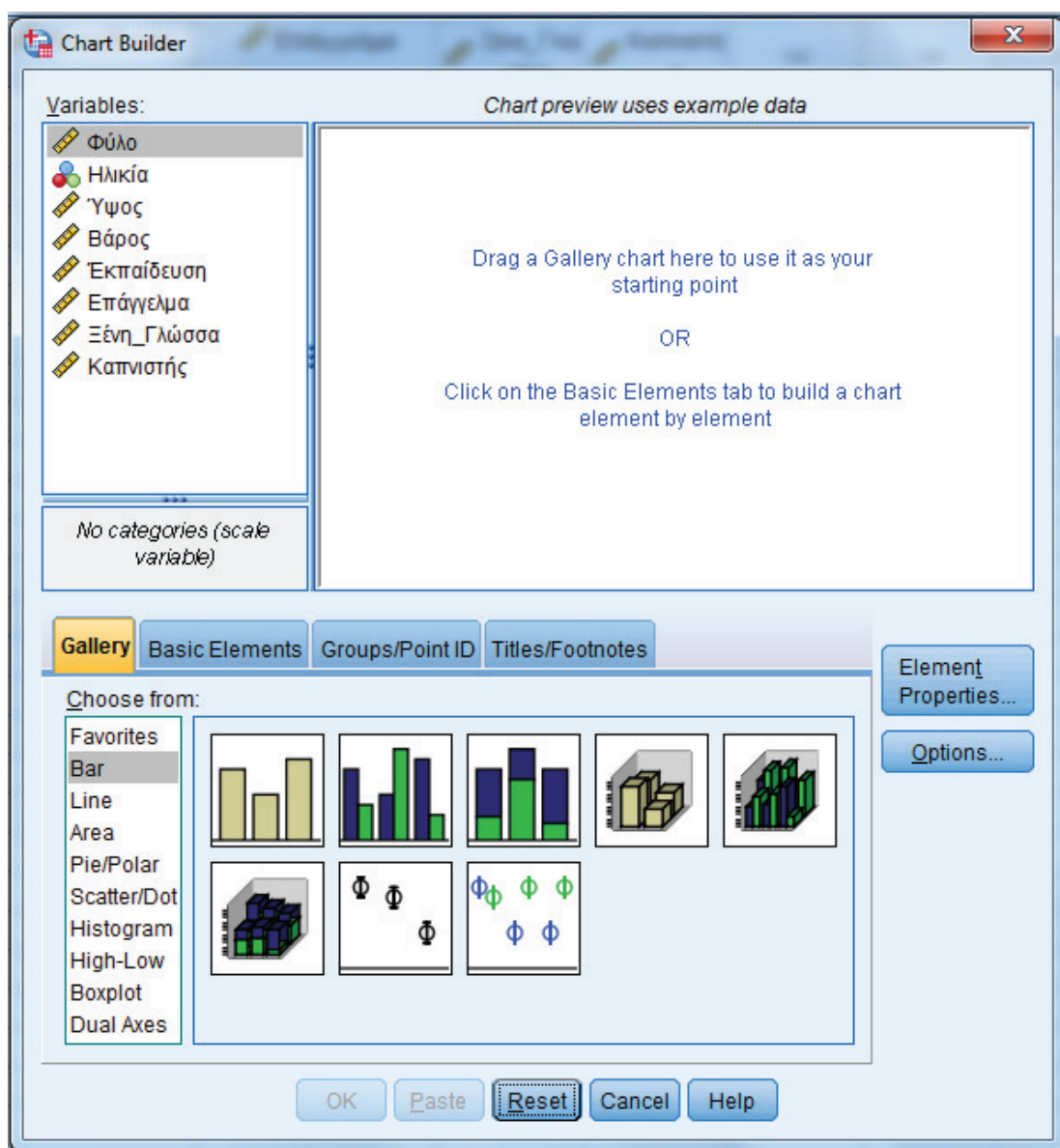


2.5 ΓΡΑΦΗΜΑΤΑ

Για να αναπαραστήσουμε την ανάλυση των δεδομένων θα δημιουργήσουμε γραφήματα.

Για να δημιουργήσουμε γραφήματα, από την γραμμή του μενού επιλέγουμε **Graphs/Chart Builder..**

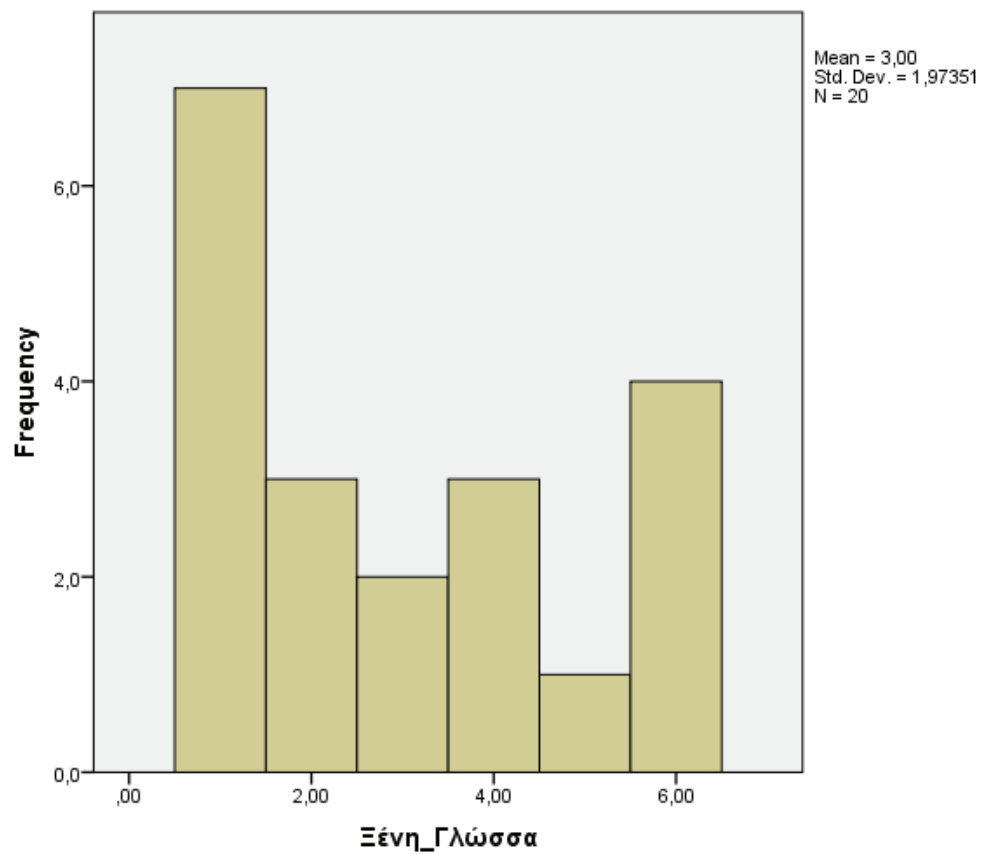
Θα εμφανιστεί το παρακάτω παράθυρο:



Επιλογή Γραφήματος

Στην στήλη **Variables** για παράδειγμα θα επιλέξουμε την μεταβλητή Ξένη Γλώσσα και από την επιλογή **Gallery** θα διαλέξουμε την δυνατότητα **Bar**. Με διπλό κλικ στο πρώτο εικονίδιο θα ανοίξουμε και θα εμφανίσουμε το πρώτο ραβδόγραμμα.

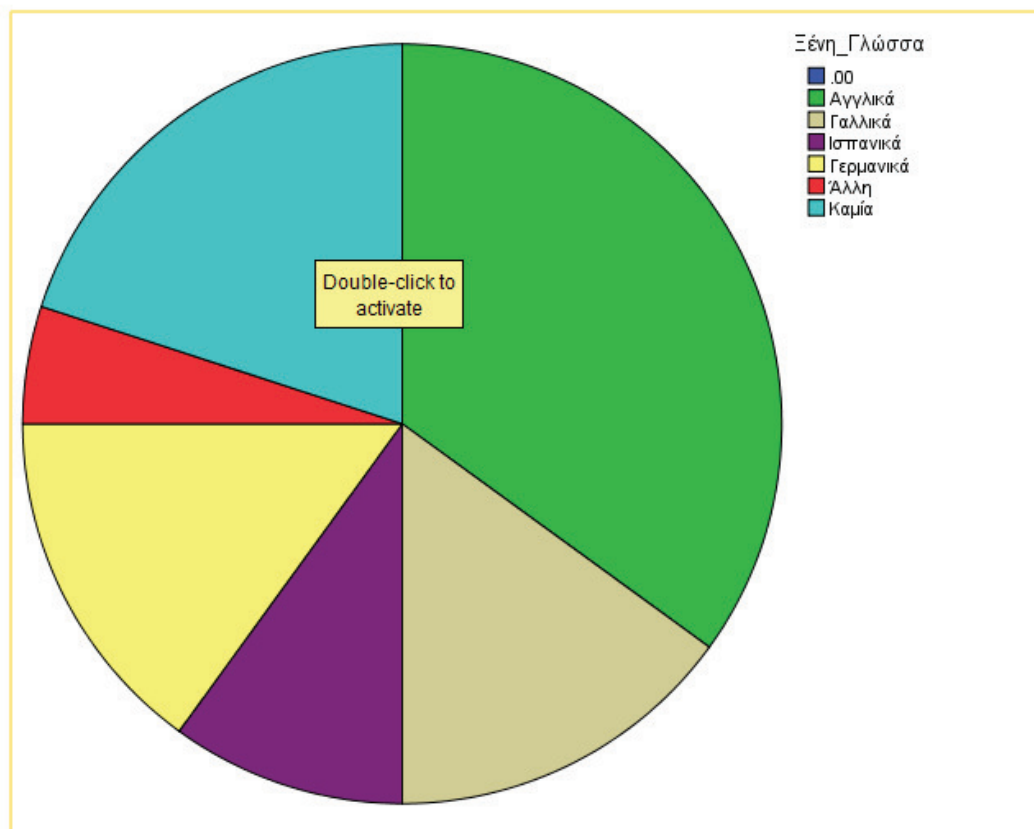
→ **GGraph**



Με διπλό κλικ πάνω στο γράφημα μας, ανοίγει ένα νέο παράθυρο το οποίο μας δίνει την δυνατότητα να μπορούμε να επεξεργαστούμε το γράφημα.

2.5.1 Pie Chart

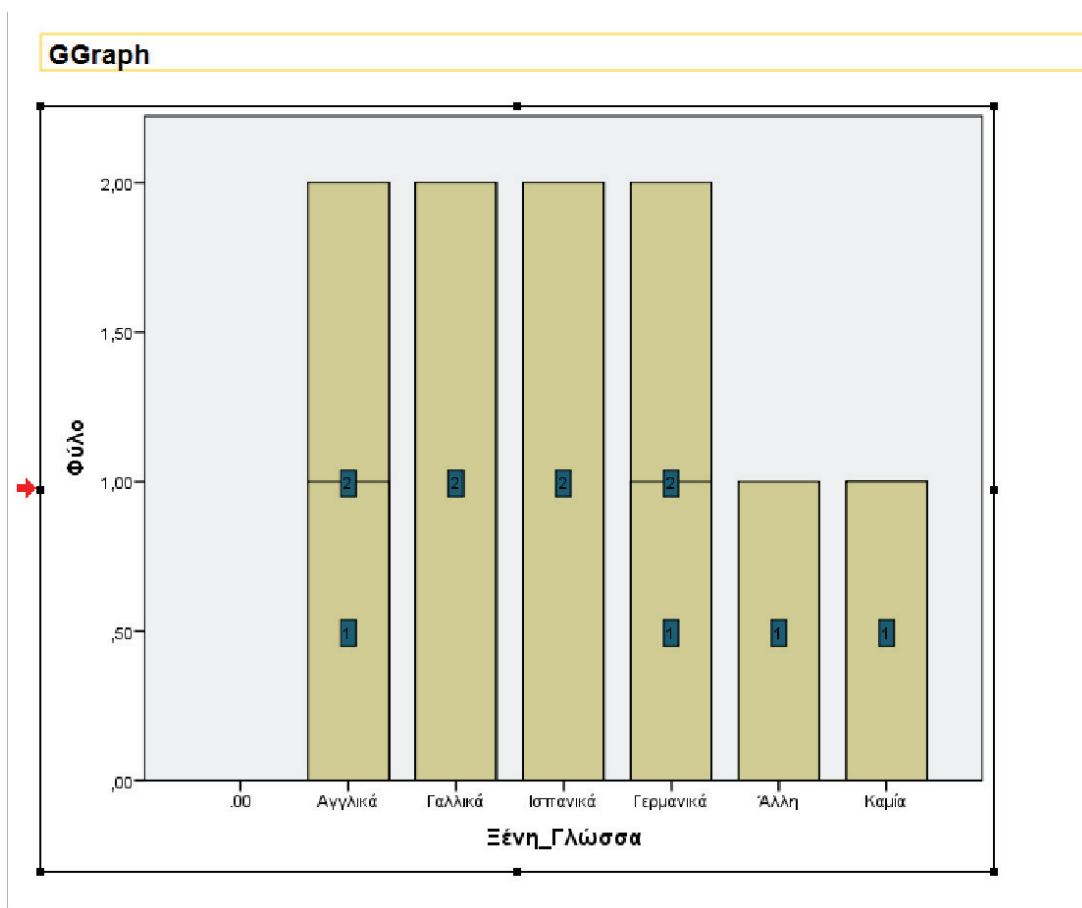
→ **GGraph**



Pie Chart

2.5.2 HISTOGRAM

Συνδυάζουμε τις μεταβλητές Ξένη Γλώσσα και Φύλο.



Histogram

3 MINITAB

3.1 Εισαγωγή στο Minitab

Το **Minitab** είναι ένα στατιστικό λογισμικό το οποίο αναπτύχθηκε στο πανεπιστήμιο της **Πενσυλβανία** το **1972** από τους ερευνητές **Barabara F. Ryan, Thomas A. Ryan** και **Bryan L. Joiner**. Ξεκίνησε ως η light έκδοση του OMNITAB ενός προγράμματος στατιστικής ανάλυσης από την εταιρία NIST. Το 1986 κυκλοφόρησε το documentation του OMNITAB το οποίο δεν έχει αλλάξει από τότε.

Η αρχική ιδέα της ανάπτυξης του από τους δημιουργούς του ήταν, για την εισαγωγή σε ηλεκτρονικούς υπολογιστές μαθήματος στατιστικής στα περισσότερα κολλέγια των **Ηνωμένων Πολιτειών της Αμερικής**.

Στις αρχές της δεκαετίας του 80 το **Minitab** χρησιμοποιήθηκε και από εκπαιδευτικά ιδρύματα και σε χώρες εκτός της Η.Π.Α και στην συνέχεια σε εταιρίες και στην βιομηχανία.

Σήμερα το στατιστικό λογισμικό **Minitab** χρησιμοποιείται σε παραπάνω από 40.000 κολλέγια και Πανεπιστήμια σε όλο τον κόσμο.

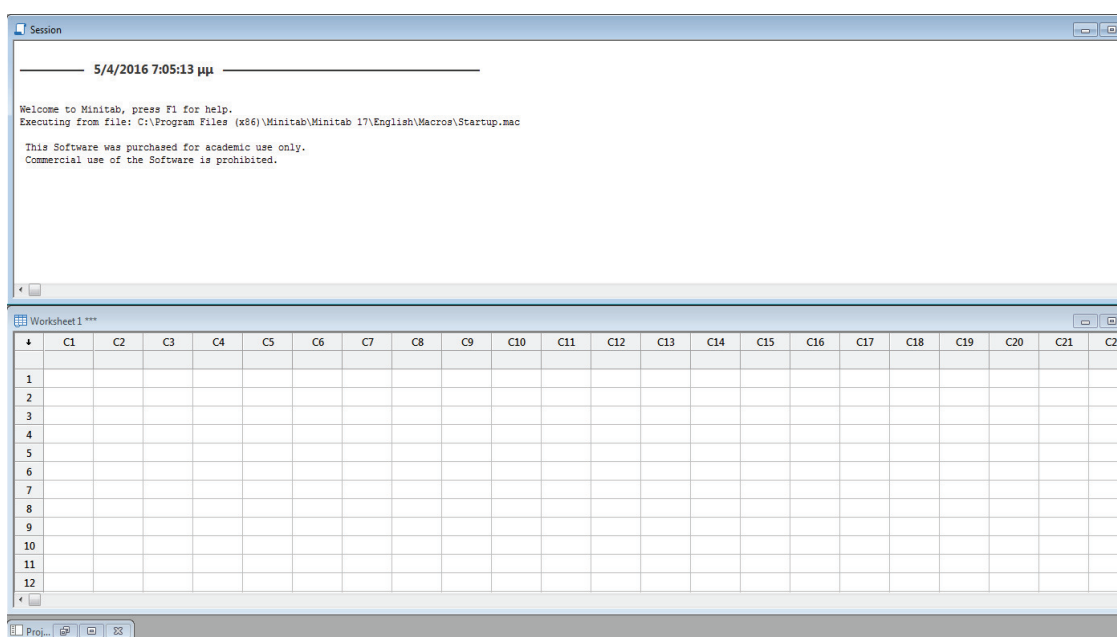
Το **Minitab** ανήκει σε μια ιδιωτική εταιρία την **Minitab Inc.**, η οποία έχει έδρα το κολλέγιο της πολιτείας της Πενσυλβανία. Επίσης **Minitab Inc**, έχει αναπτύξει **Quality Trainer** και **Quality Companion** των οποίων η χρήση του μπορεί να συνδεθεί με το **Minitab**.

Το **Quality Trainer** είναι ένα online e-learning μάθημα που διδάσκει στατιστικές αναλύσεις για βελτίωση ποιότητας.

Το **Quality Companion** χρησιμοποιείται παγκοσμίως για σχεδιασμό και εκτέλεση διάφορων project.

3.2 Παρουσίαση του Minitab σε Windows

Με διπλό κλικ στο εικονίδιο του Minitab θα τρέξουμε το λογισμικό. Κατά την εκκίνηση του προγράμματος εμφανίζεται το παρακάτω παράθυρο:



Παρατηρούμε ότι στο παράθυρο αυτό υπάρχουν 3 μέρη:

Session Window: Στο παράθυρο αυτό γίνεται η εμφάνιση των αναφορών και των μηνυμάτων από τις εντολές που εκτελούνται στο **Minitab**.

Data Window Worksheet 1*:** Στο φύλλο αυτό το οποίο μοιάζει με το γνωστό Excel, εισάγουμε και επεξεργαζόμαστε τα δεδομένα. Κάθε γραμμή θα περιέχει διάφορα δεδομένα για ένα αντικείμενο ή άτομο. Η κάθε στήλη αντιστοιχεί σε μια μεταβλητή.

Project Manager: Βρίσκεται κάτω αριστερά του παραθύρου και η χρήση του είναι να βοηθάει στην αναζήτηση διαφόρων στοιχείων στο λογισμικό.

3.2.1 Εισαγωγή Δεδομένων

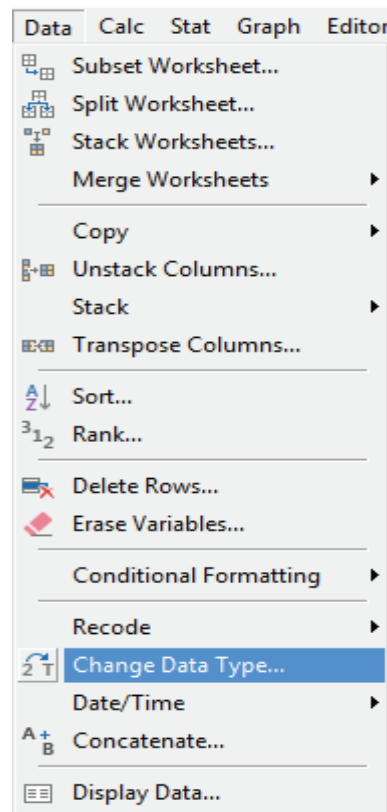
Μπορούμε να εισάγουμε δεδομένα με πληκτρολόγηση, με αντιγραφή και επικόλληση ή και με άνοιγμα ενός αρχείου. Για παράδειγμα θα ανοίξουμε ένα αρχείο το οποίο περιέχεται στο **Minitab**, όπως θα δούμε και απρακάτω. Το αρχείο αυτό είναι μια διαδικασία αποστολής προϊόντων.

Οι μεταβλητές που δημιουργούνται είναι: το time zone (Center), η ημερομηνία και η ώρα που έγινε η παραγγελία (Order), η ημερομηνία και η ώρα άφιξης του προϊόντος (Arrival), οι μέρες που χρειάστηκαν από την παραγγελία μέχρι την άφιξη (Days), η κατάσταση που βρίσκεται το προϊόν (Status) και η απόσταση που έκανε το προϊόν να φτάσει στον προορισμό του (Distance).

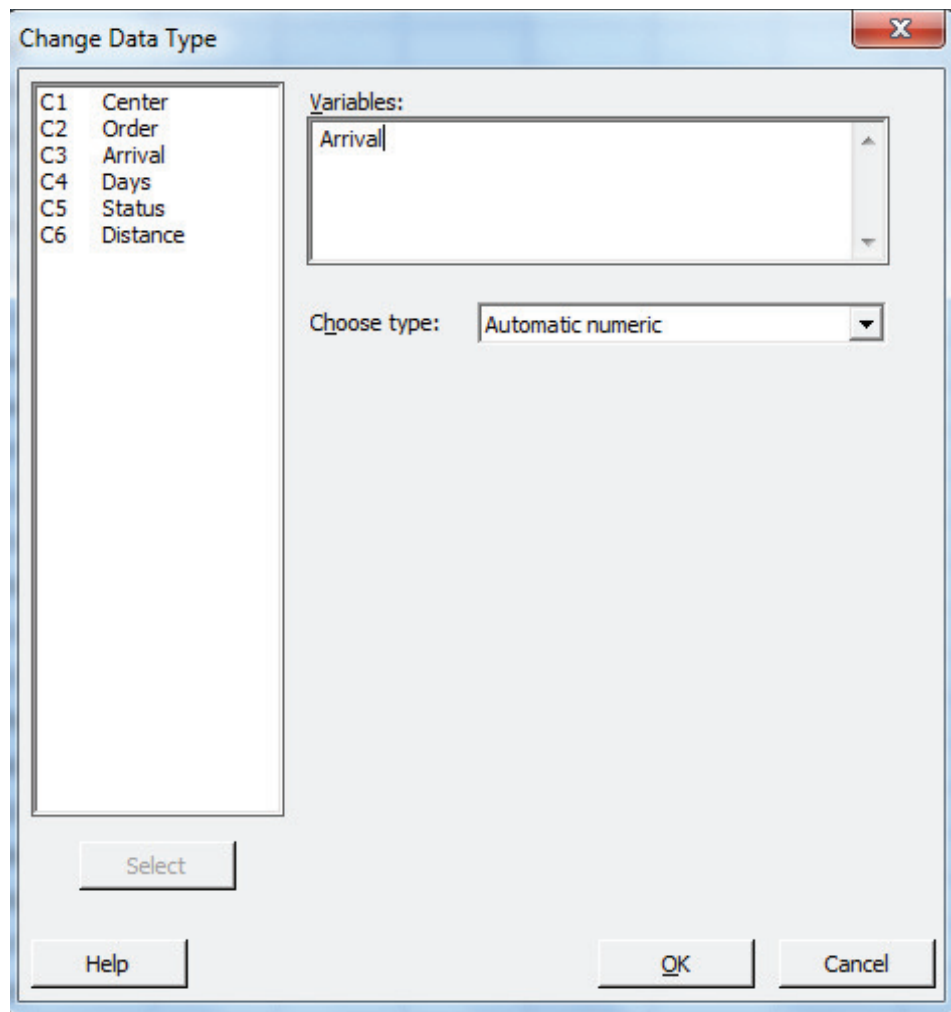
↓	C1-T	C2-D	C3-D	C4	C5-T	C6
	Center	Order	Arrival	Days	Status	Distance
1	Eastern	3/4/2013 8:34	3/8/2013 15:21	4,28264	On time	255
2	Eastern	3/4/2013 8:35	3/7/2013 17:05	3,35417	On time	196
3	Eastern	3/4/2013 8:38	*	* Back order		299
4	Eastern	3/4/2013 8:40	3/8/2013 15:52	4,30000	On time	205
5	Eastern	3/4/2013 8:42	3/10/2013 14:48	6,25417	Late	250
6	Eastern	3/4/2013 8:43	3/9/2013 15:45	5,29306	On time	93
7	Eastern	3/4/2013 8:50	3/8/2013 10:02	4,05000	On time	189
8	Eastern	3/4/2013 8:55	3/9/2013 16:30	5,31597	On time	335
9	Eastern	3/4/2013 8:58	3/9/2013 10:32	5,06528	On time	211
10	Eastern	3/4/2013 9:11	3/8/2013 16:02	4,28542	On time	254
11	Eastern	3/4/2013 9:13	3/9/2013 15:58	5,28125	On time	264
12	Eastern	3/4/2013 9:20	3/9/2013 16:13	5,28681	On time	197
13	Eastern	3/4/2013 9:24	3/8/2013 12:32	4,13056	On time	11
14	Eastern	3/4/2013 9:29	3/8/2013 17:08	4,31875	On time	353
15	Eastern	3/4/2013 9:31	3/7/2013 11:56	3,10069	On time	129
16	Eastern	3/4/2013 9:31	3/10/2013 15:49	6,26250	Late	153
17	Eastern	3/4/2013 9:33	3/8/2013 15:56	4,26597	On time	102
18	Eastern	3/4/2013 9:38	3/9/2013 15:04	5,22639	On time	279
19	Eastern	3/4/2013 9:46	3/9/2013 10:03	5,01181	On time	340
20	Eastern	3/4/2013 9:51	3/10/2013 16:51	6,29167	Late	282
21	Eastern	3/4/2013 9:54	3/8/2013 17:00	4,29583	On time	459
22	Eastern	3/4/2013 9:58	*	* Back order		378
23	Eastern	3/4/2013 10:00	3/8/2013 9:46	3,99028	On time	257
24	Eastern	3/4/2013 10:08	3/8/2013 11:59	4,07708	On time	318
25	Eastern	3/4/2013 10:11	3/9/2013 8:20	4,92292	On time	487
26	Eastern	3/4/2013 10:18	3/6/2013 17:06	2,28333	On time	310

3.2.2 Αλλαγή Μεταβλητής

Σε μερικές περιπτώσεις χρειάζεται να αλλάξουμε τον τύπο της μεταβλητής. Για να το επιτύχουμε αυτό, από το μενού του λογισμικού, θα επιλέξουμε το **Data**, στην συνέχεια το **Change Data Type...**



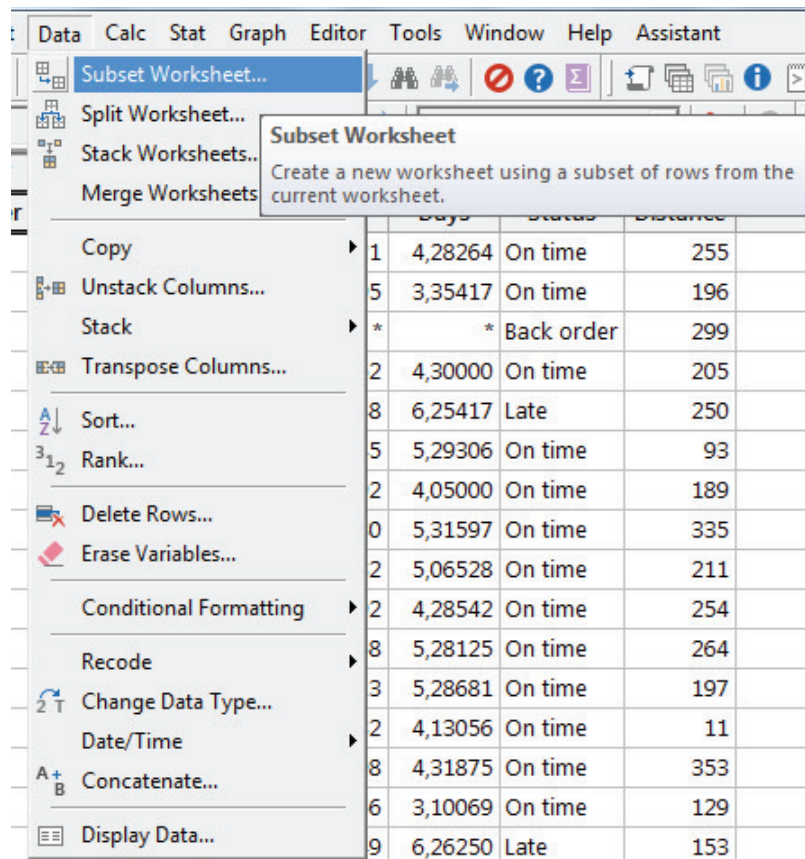
Θα εμφανιστεί το παρακάτω παράθυρο οπού επιλέγοντας την μεταβλητή που θέλουμε να αλλάξουμε τον τύπο της, την εισάγουμε στον κενό κουτί (**Variables**) και από το **Choose Type**, επιλέγουμε την επιθυμητή μεταβλητή.



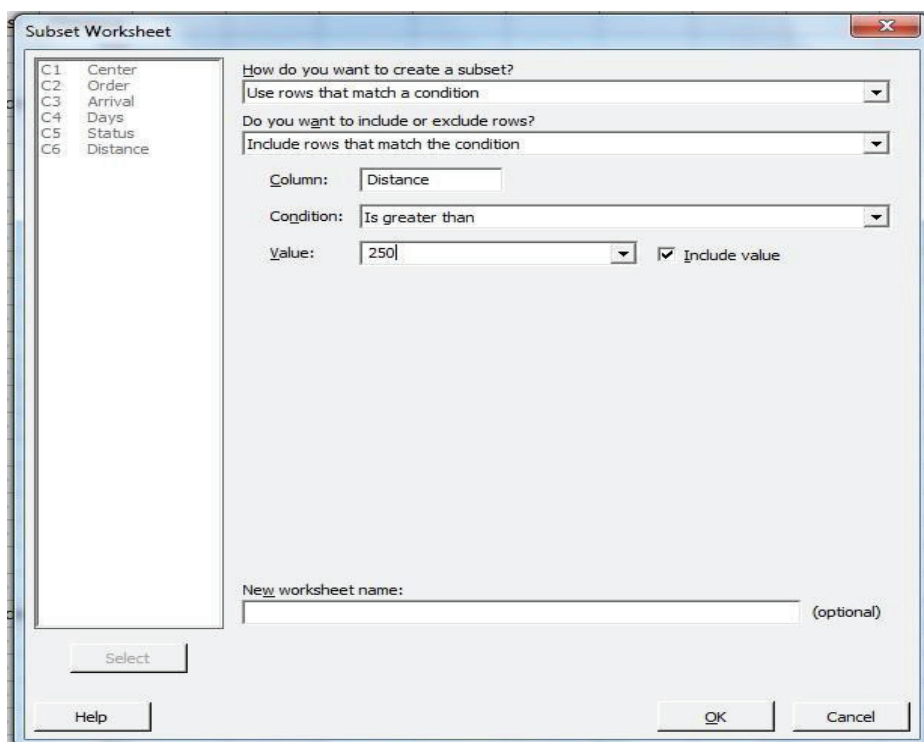
3.2.3 Επιλογές Δεδομένων

Αν θέλουμε να κάνουμε κάποια στατιστική ανάλυση για συγκεκριμένα στοιχεία του συνόλου τα οποία θα ικανοποιούν μια μόνο συγκεκριμένη συνθήκη. Για παράδειγμα θέλουμε ανάλυση για αποστάσεις μεγαλύτερες από **250**.

Από το μενού του λογισμικού θα επιλέξουμε **Data** και στην συνέχεια την επιλογή **Subset Worksheet**.



Αφού το επιλέξουμε θα εμφανιστεί το παρακάτω παράθυρο:



Στο **Column** θα επιλέξουμε την μεταβλητή **Distance** που μας ενδιαφέρει.

Στο **Condition** θα επιλέξουμε **Is greater than**, αφού θέλουμε να είναι μεγαλύτερο από.

Και στην επιλογή **Value** θα γράψουμε την τιμή **250** αφού θέλουμε η μεταβλητή **Distance** να είναι μεγαλύτερη από **250 (Distance>250)**.

Με την επιλογή **OK** θα εμφανίσει το παρακάτω νέο **Worksheet**, όπου όπως θα παρατηρήσουμε όλες οι τιμές του **Distance** είναι πάνω από **250**.

↓	C1-T	C2-D	C3-D	C4	C5-T	C6
	Center	Order	Arrival	Days	Status	Distance
1	Eastern	3/4/2013 8:34	3/8/2013 15:21	4,28264	On time	255
2	Eastern	3/4/2013 8:38	*	*	Back order	299
3	Eastern	3/4/2013 8:42	3/10/2013 14:48	6,25417	Late	250
4	Eastern	3/4/2013 8:55	3/9/2013 16:30	5,31597	On time	335
5	Eastern	3/4/2013 9:11	3/8/2013 16:02	4,28542	On time	254
6	Eastern	3/4/2013 9:13	3/9/2013 15:58	5,28125	On time	264
7	Eastern	3/4/2013 9:29	3/8/2013 17:08	4,31875	On time	353
8	Eastern	3/4/2013 9:38	3/9/2013 15:04	5,22639	On time	279
9	Eastern	3/4/2013 9:46	3/9/2013 10:03	5,01181	On time	340
10	Eastern	3/4/2013 9:51	3/10/2013 16:51	6,29167	Late	282
11	Eastern	3/4/2013 9:54	3/8/2013 17:00	4,29583	On time	459
12	Eastern	3/4/2013 9:58	*	*	Back order	378
13	Eastern	3/4/2013 10:00	3/8/2013 9:46	3,99028	On time	257
14	Eastern	3/4/2013 10:08	3/8/2013 11:59	4,07708	On time	318
15	Eastern	3/4/2013 10:11	3/9/2013 8:20	4,92292	On time	487
16	Eastern	3/4/2013 10:18	3/6/2013 17:06	2,28333	On time	310
17	Eastern	3/4/2013 10:22	3/7/2013 10:44	3,01528	On time	263
18	Eastern	3/4/2013 10:43	3/10/2013 17:02	6,26319	Late	353
19	Eastern	3/4/2013 10:51	3/11/2013 8:16	6,89236	Late	384
20	Eastern	3/4/2013 10:55	3/8/2013 12:15	4,05556	On time	315
21	Eastern	3/4/2013 11:00	*	*	Back order	467
22	Eastern	3/4/2013 11:12	3/8/2013 10:32	3,97222	On time	323
23	Eastern	3/4/2013 11:14	3/6/2013 17:36	2,26528	On time	291
24	Eastern	3/4/2013 11:21	3/6/2013 17:25	2,25278	On time	349

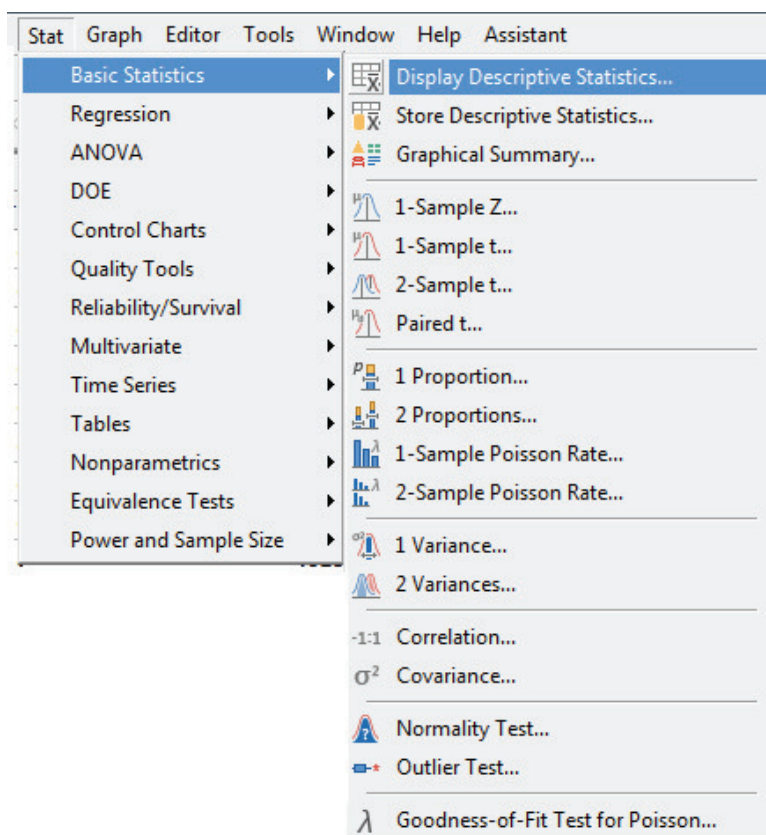
3.3 Στατιστική Ανάλυση

Θα χωρίσουμε την Στατιστική Ανάλυση σε 2 κομμάτια, την Περιγραφική Στατιστική και την Κατανομή Συχνοτήτων.

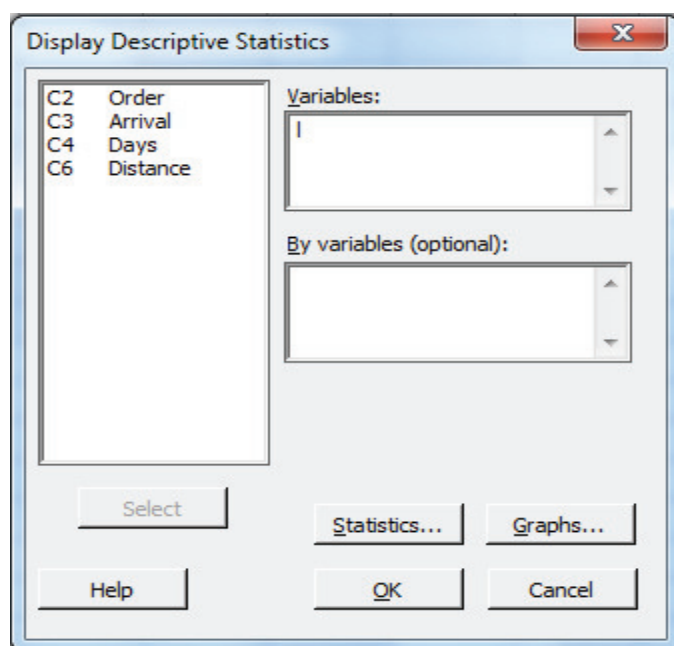
Για να ξεκινήσουμε την **Στατιστική Ανάλυση** από το μενού θα επιλέξουμε την επιλογή **Stat**.

3.3.1 Περιγραφική Στατιστική

Από το μενού αφού επιλέξουμε **Stat**, θα επιλέξουμε **Basic** και στην συνέχεια **Statistics Display Descriptive Statistics...**

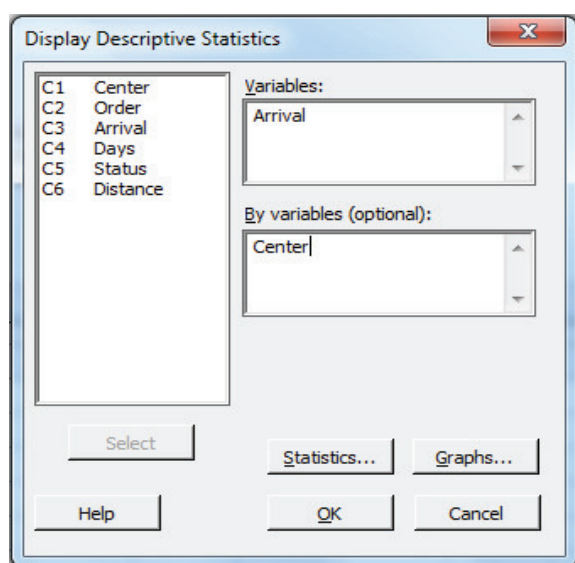


Θα εμφανιστεί το παρακάτω παράθυρο:



Παρατηρούμε ότι έχουν εμφανιστεί μόνο οι μεταβλητές οι οποίες έχουν αριθμητική τιμή.

Θα επιλέξουμε την μεταβλητή **Arrival** και αν μας ζητηθεί μια **By Variable** θα επιλέξουμε την **Center** όπως φαίνεται παρακάτω:



Με την επιλογή **OK** στο **Session Window** θα εμφανιστούν τα αποτελέσματα.

Descriptive Statistics: Arrival

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Arrival	302	17	41341	0,0778	1,35	41338	41340	41341	41342	41345

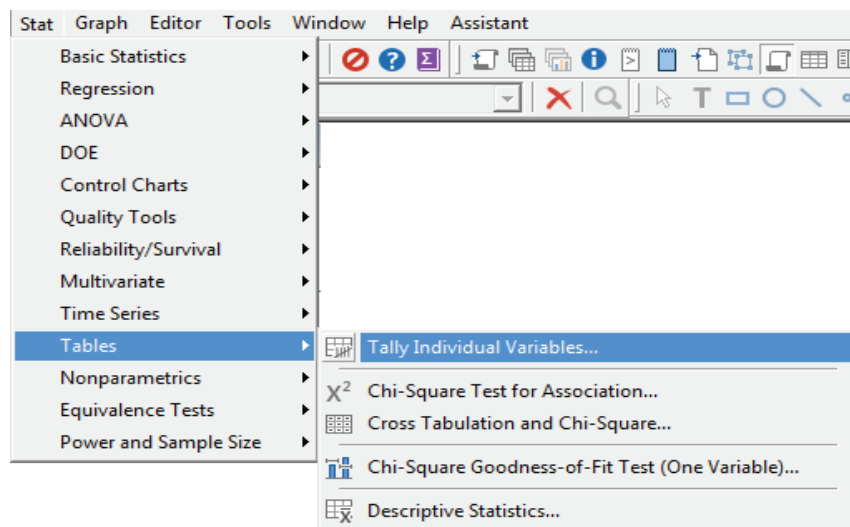
Και με **By Variable** την μεταβλητή **Center** τα αποτελέσματα θα είναι:

Descriptive Statistics: Arrival

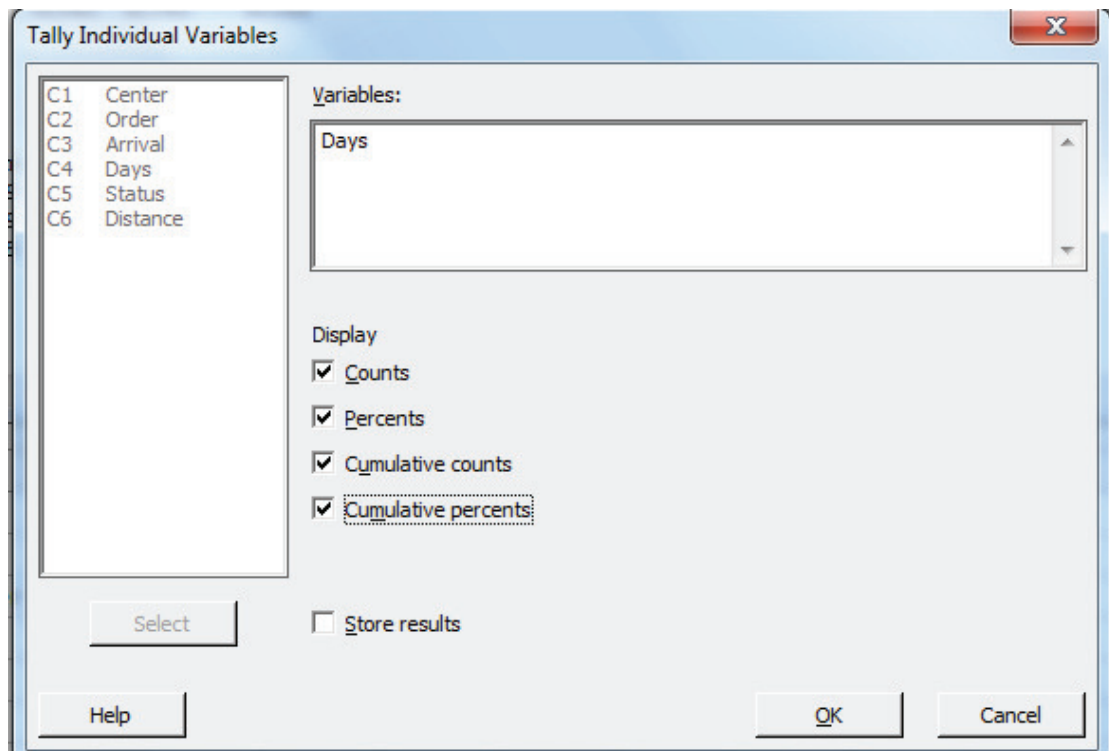
Variable	Center	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Arrival	Central	99	6	41341	0,129	1,29	41339	41341	41341	41342	41345
	Eastern	101	8	41342	0,124	1,24	41339	41341	41342	41343	41345
	Western	102	3	41341	0,107	1,08	41338	41340	41340	41341	41343

3.3.2 Κατανομή Συχνοτήτων

Από το μενού θα επιλέξουμε το **Stat** στην συνέχεια το **Tables** και έπειτα το **Tally Individual Variables...**



Θα εμφανιστεί το παρακάτω παράθυρο:



Επιλέγοντας την μεταβλητή **Days** μέσω αυτής της εντολής μπορούμε να υπολογίσουμε την συχνότητα, την αθροιστική συχνότητα, την σχετική συχνότητα και την αθροιστική σχετική συχνότητα.

Tally for Discrete Variables: Days

Days	Count	Percent	CumCnt	CumPct
0,87083	1	0,33	1	0,33
0,91389	1	0,33	2	0,66
1,03194	1	0,33	3	0,99
1,10347	1	0,33	4	1,32
1,16667	1	0,33	5	1,66
1,25278	1	0,33	6	1,99
1,26736	1	0,33	7	2,32
1,29167	1	0,33	8	2,65
1,67292	1	0,33	9	2,98
1,70833	1	0,33	10	3,31
1,72917	2	0,66	12	3,97
1,73542	1	0,33	13	4,30
1,74583	2	0,66	15	4,97
1,74931	2	0,66	17	5,63
1,77917	1	0,33	18	5,96
1,80694	1	0,33	19	6,29
1,81458	1	0,33	20	6,62
1,82778	1	0,33	21	6,95
1,85972	1	0,33	22	7,28
1,86181	1	0,33	23	7,62
1,88472	1	0,33	24	7,95

Βλέπουμε ένα κομμάτι των αποτελεσμάτων για την μεταβλητή **Days**. Η στήλη **Count** εμφανίζει όλες τις τιμές της μεταβλητής **Days**.

Η στήλη **Percent** την αντίστοιχη συχνότητα εμφάνισης της μεταβλητής **Days**.

Η στήλη **CumCnt** εμφανίζει την αθροιστική συχνότητα της μεταβλητής **Days**.

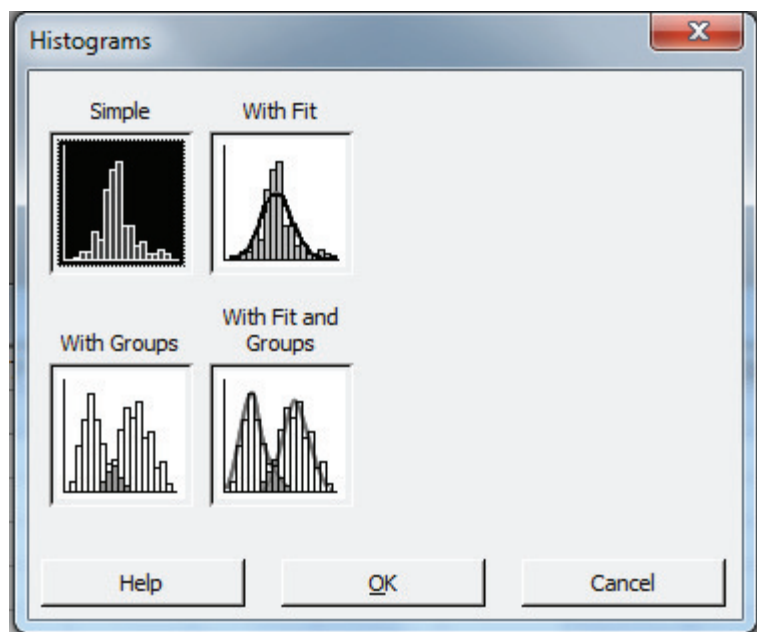
Η στήλη **CumPct** εμφανίζει την σχετική αθροιστική συχνότητα της μεταβλητής **Days**.

3.4 Γραφικές Παραστάσεις

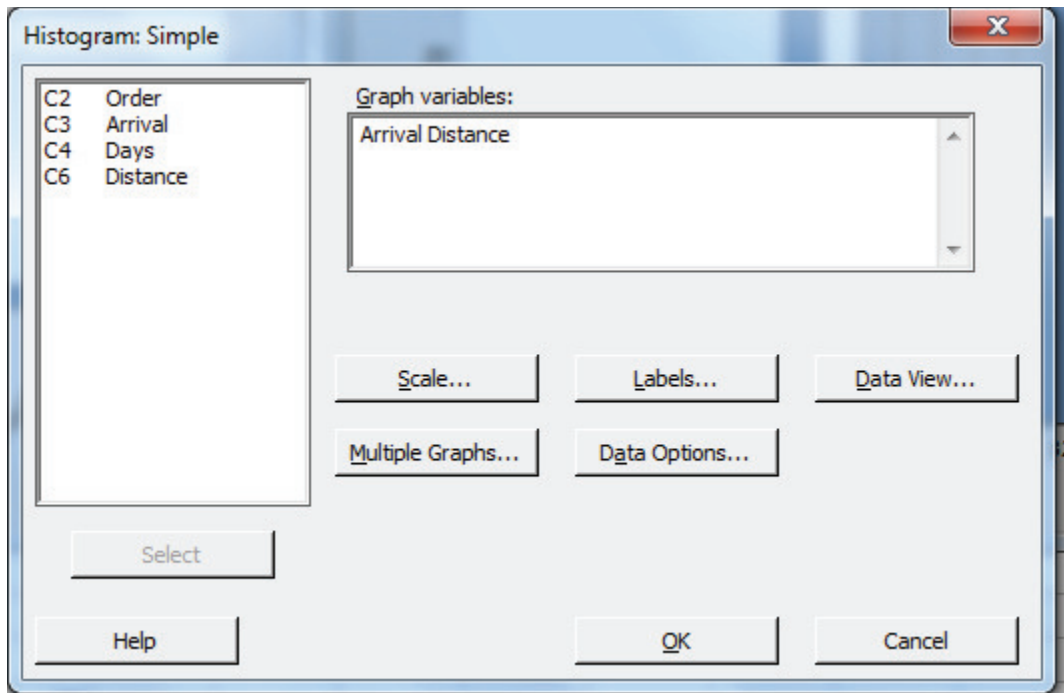
Παρακάτω θα δούμε 3 διαφορετικές γραφικές παραστάσεις. Θα δούμε πώς να δημιουργήσουμε παραστάσεις όπως Histogram, Bar Graph και Pie Chart.

3.4.1 Histogram

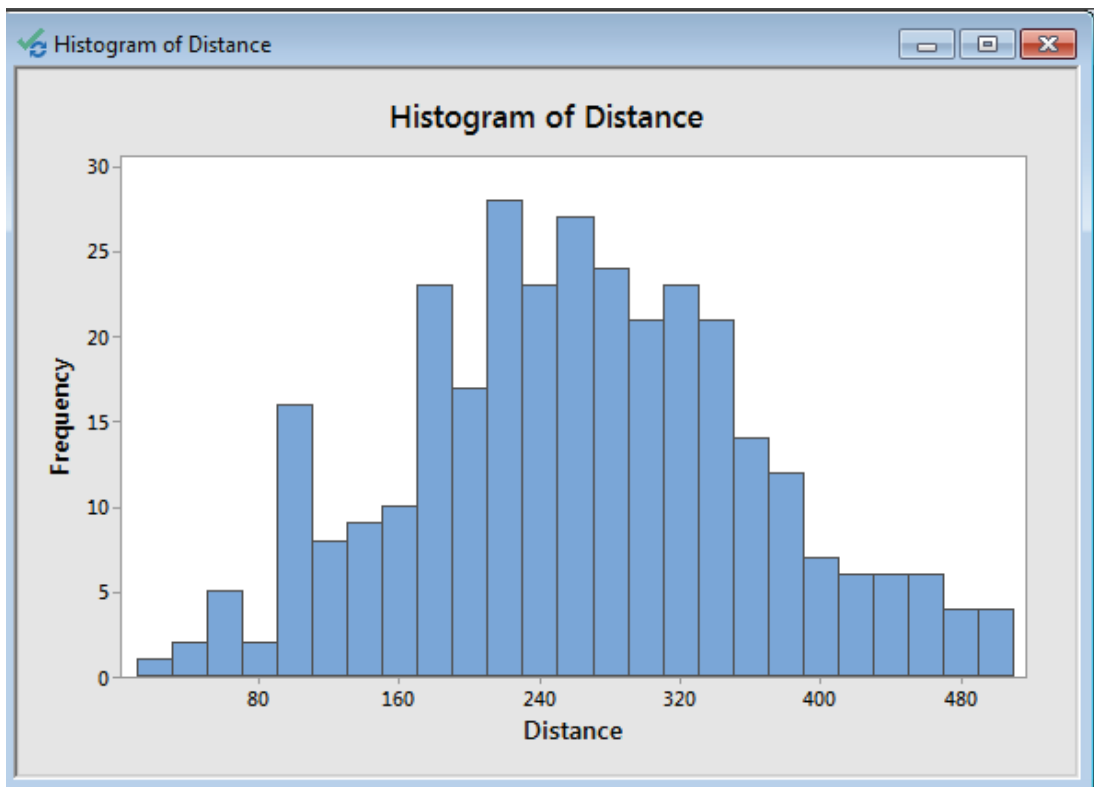
Για την δημιουργία ενός **histogram** , από το μενού θα επιλέξουμε **Graph** και στην συνέχεια **histogram**.

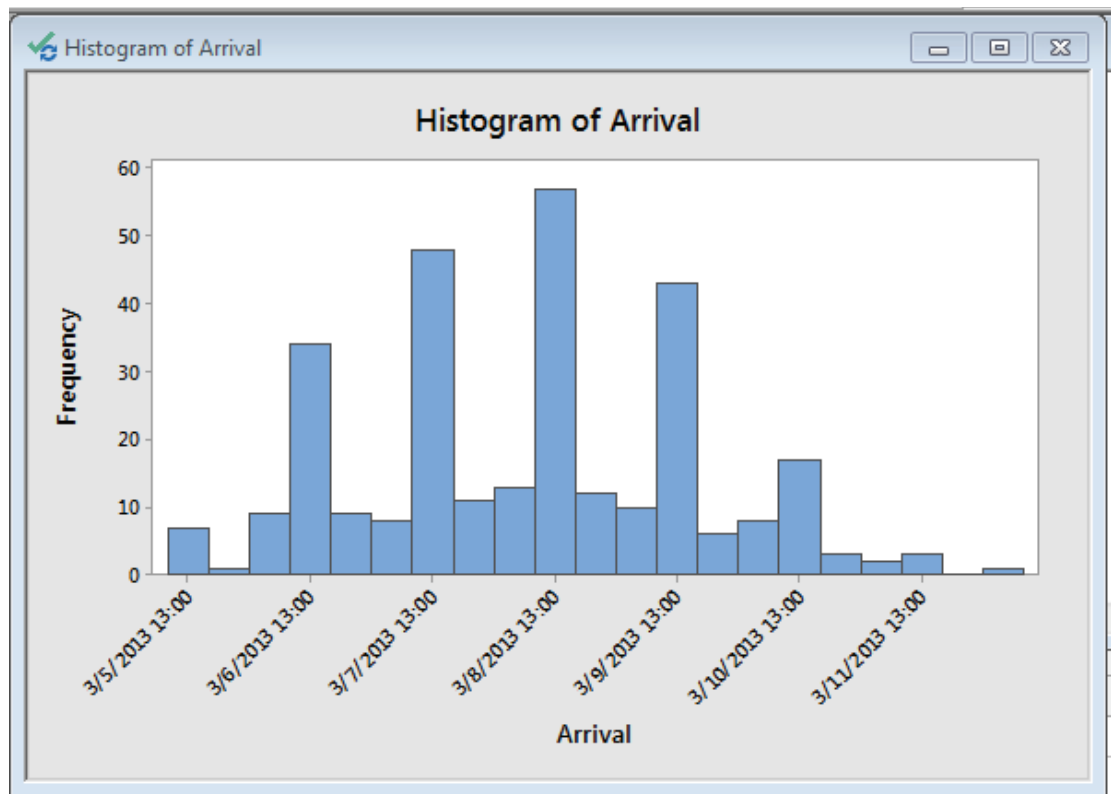


Επιλέγουμε το γράφημα που θέλουμε να χρησιμοποιήσουμε και εμφανίζεται το παρακάτω παράθυρο:



Και στην συνέχεια θα επιλέξουμε τις μεταβλητές που θέλουμε να μας εμφανιστούν.



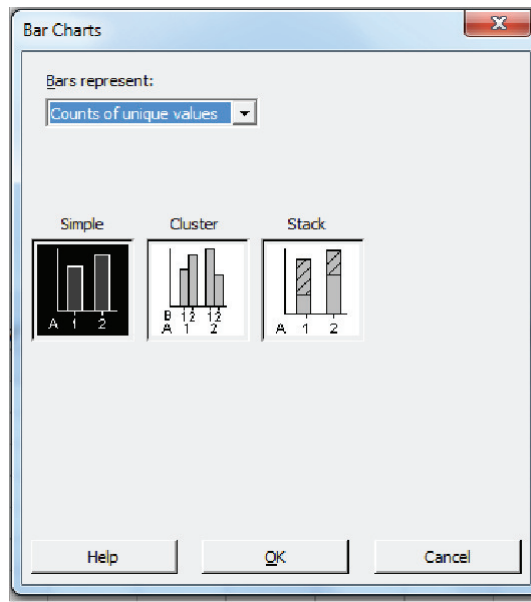


3.4.2 Bar Graph

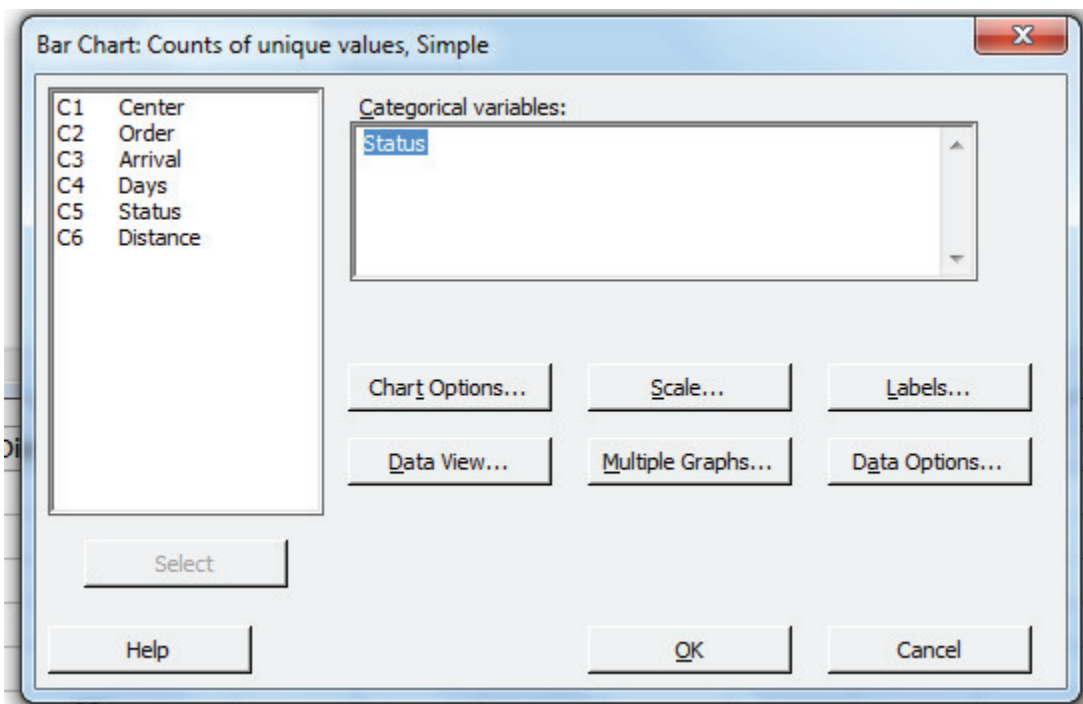
Για την δημιουργία ενός **Bar Chart** , από το μενού θα επιλέξουμε **Graph** και στην συνέχεια

Bar Chart.

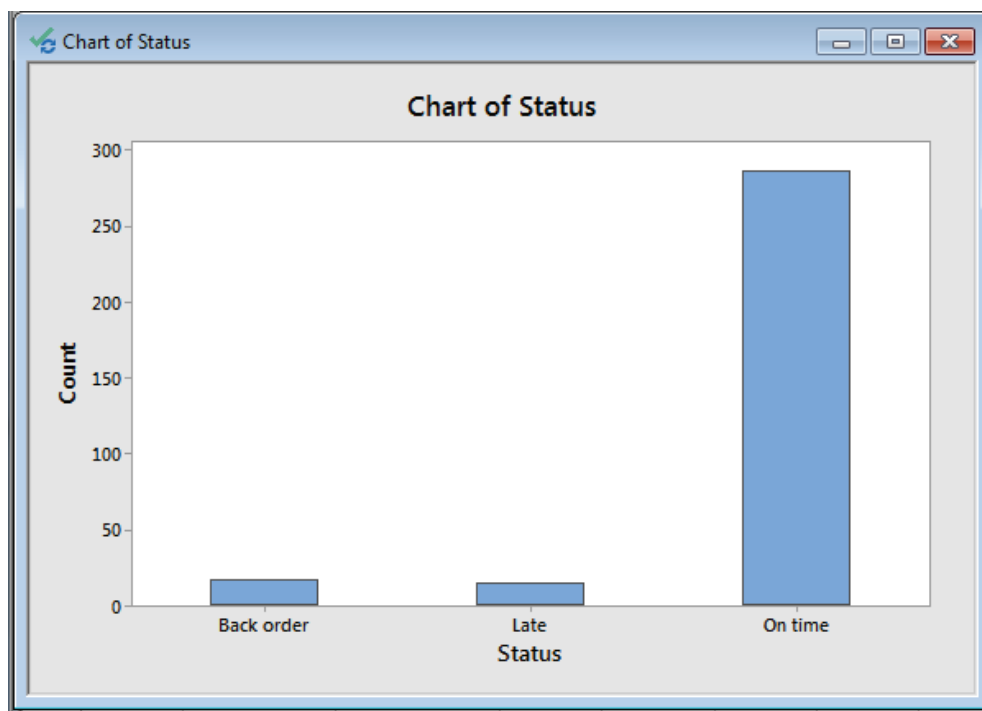
Εμφανίζετε το παρακάτω παράθυρο:



Από το παράθυρο θα επιλέξουμε το γράφημα που θέλουμε και εμφανίζεται το παρακάτω παράθυρο:



Επιλέγουμε την μεταβλητή **Status** και θα μας εμφανίσει :

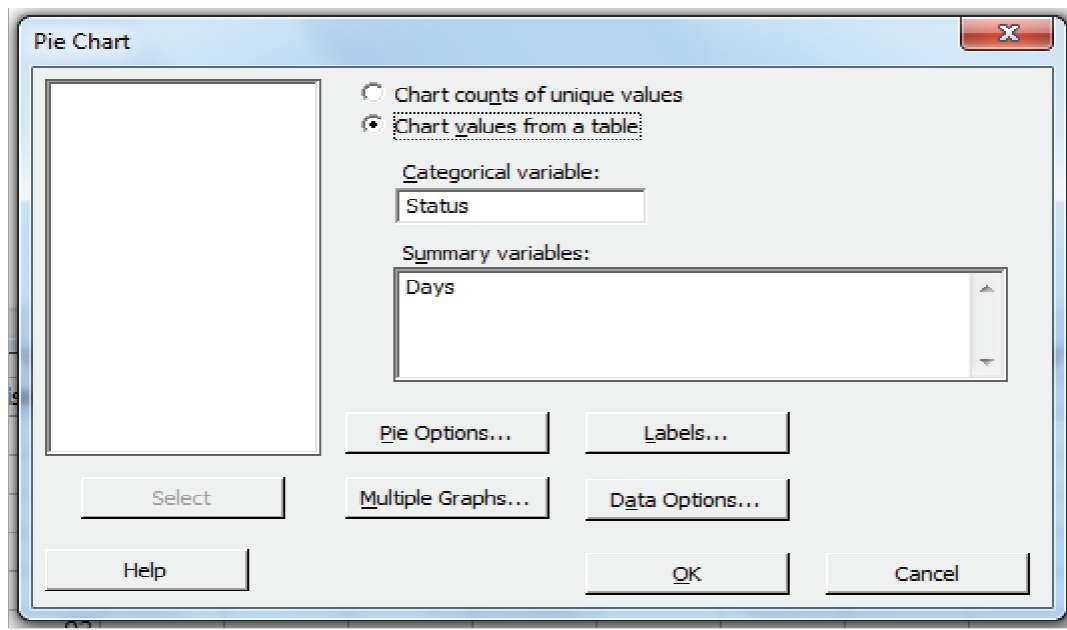


3.4.3 Pie Chart

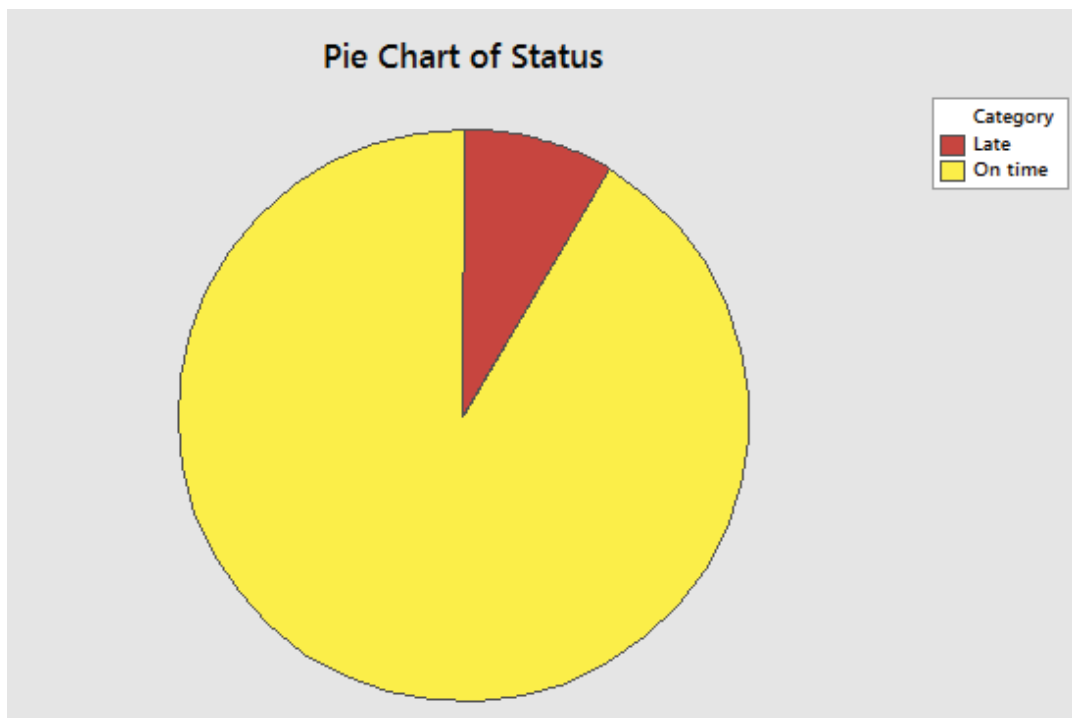
Για την δημιουργία ενός **Pie Chart** , από το μενού θα επιλέξουμε **Graph** και στην συνέχεια

Pie Chart.

Εμφανίζετε το παρακάτω παράθυρο:



Επιλέγουμε την μεταβλητή **Status** και θα μας εμφανίσει :



4 Γλώσσα R

4.1 Εισαγωγή στην γλώσσα R

Το στατιστικό λογισμικό **R** δημιουργήθηκε από τους **Ross Ihaka** και **Robert Gentleman** στα μέσα της δεκαετίας του 90, στο **Πανεπιστήμιο του Όκλαντ**, στην **Νέα Ζηλανδία**. Από το 1997 αναπτύσσεται από την **R Development Core Team**. Είναι έργο της GNU. Το όνομα της προέρχεται από τα αρχικά ονόματα των δημιουργών της (**Ross, Robert**).

Είναι μια γλώσσα προγραμματισμού και επίσης ένα περιβάλλον για ανάλυση δεδομένων, στατιστικούς υπολογισμούς και δημιουργίας γραφημάτων. Μπορεί να θεωρηθεί παρόμοια με την γλώσσα προγραμματισμού S η οποία έχει υλοποιηθεί από την **Bell Laboratories**. Οι διαφορές τους είναι ελάχιστες μιας και η γλώσσα R αποτελείται από ένα μεγάλο μέρος από κώδικα της γλώσσας S.

Η R είναι διαθέσιμη ως ελεύθερο λογισμικό από την **GNU**. Έχει την δυνατότητα αφού μεταγλωττιστεί να εκτελεστεί σε περιβάλλον **UNIX, LINUX, Windows** και **Mac OS**.

4.2 Εγκατάσταση R

Για να εγκαταστήσουμε την R , θα πρέπει να επισκεφτούμε την ιστοσελίδα <https://cran.rstudio.com/> . Κατά την περιήγηση μας στην σελίδα θα διαλέξουμε την έκδοση που θέλουμε για το αντίστοιχο λειτουργικό σύστημα που χρησιμοποιούμε.

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for MacOS X](#)
- [Download R for Windows](#)

Στην συνέχεια αφού θα επιλέξουμε λειτουργικό σύστημα (στην περίπτωση μας Windows), θα επιλέξουμε την επιλογή **base**.

R for Windows

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch). This is what you want if you install R for the first time .
contrib	Binaries of contributed packages (managed by Uwe Ligges)

Στην συνέχεια επιλέγουμε **Download R 3.2.5 for Windows**. Αποθηκεύουμε το αρχείο που θα κατέβει στον υπολογιστή και στην συνέχεια θα πραγματοποιήσουμε την εγκατάσταση του.

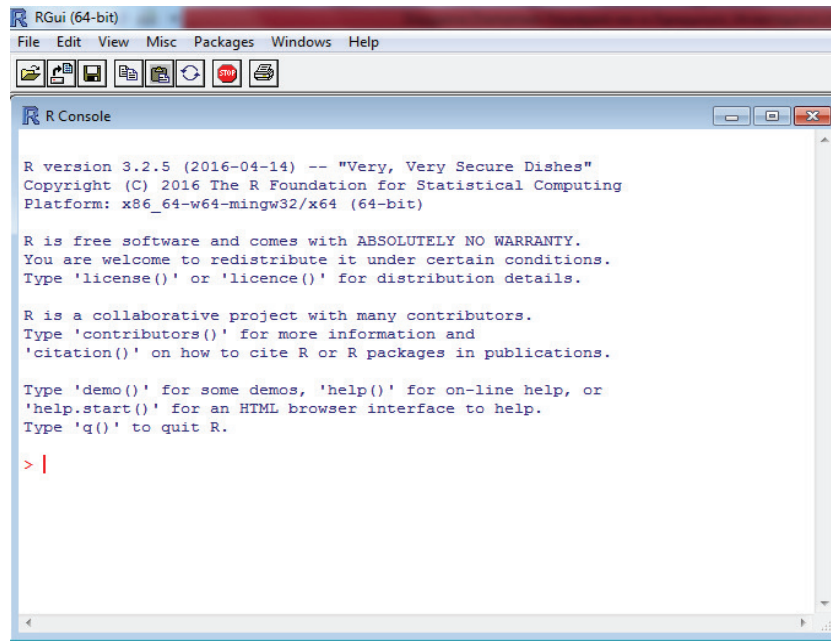
R-3.2.5 for Windows (32/64 bit)

[Download R 3.2.5 for Windows](#) (62 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

Αφού ολοκληρώσουμε την εγκατάσταση του, με διπλό κλικ στο εικονίδιο της R, που βρίσκεται στην επιφάνεια εργασίας, ανοίγουμε το λογισμικό. Θα εμφανιστεί το παράθυρο της εφαρμογής.



4.3 Αριθμητικοί Τελεστές της R

Παρακάτω δίνονται οι αριθμητικοί τελεστές της R.

Σύμβολο	Πράξη
+	Πρόσθεση
-	Αφαίρεση
*	Πολλαπλασιασμός
/	Διαίρεση
^	Ύψωση σε δύναμη
%%/%	Ακέραια διαίρεση
%%	Υπόλοιπο διαίρεσης

```

>3+4 #my first command
[1] 7
>7-2
[1] 5
>14/2
[1] 7
>4*2
[1] 8
>2^3
[1] 8
>17%%3
[1] 5
>8/0
[1] Inf

```

- Οτιδήποτε ακολουθεί μετά από # αγνοείται
- Το Inf δηλώνει άπειρο
- Το NaN δηλώνει ότι η πράξη δεν μπορεί να γίνει.

4.4 Τελεστές Εκχώρησης της R

Παρακάτω δίνονται οι τελεστές εκχώρησης της R.

Τελεστής	Ερμηνεία
<-	Εισάγει το αποτέλεσμα στο αριστερό μέλος της σχέσης
->	Εισάγει το αποτέλεσμα στο δεξί μέλος της σχέσης
<	Μεγαλύτερο από
>	Μικρότερο από
>=	Μεγαλύτερο ή ίσο από
<=	Μικρότερο ή ίσο από
==	Ίσο με
!=	Όχι ίσο με

- Η **R** διαβάζει διαφορετικά τα κεφαλαία γράμματα. Το **x** είναι διαφορετικό αντικείμενο από το **X**.
- Τα αντικείμενα μπορεί να είναι πίνακες, διανύσματα, δεδομένα ή λίστες.
- Αν σε ένα αντικείμενο που ήδη υπάρχει καταχωρήσουμε μια άλλη τιμή, η τιμή αυτή θα αντικαταστήσει την προηγούμενη.

4.5 Βασικές Συναρτήσεις της R

Παρακάτω στον πίνακα θα δούμε κάποιες βασικές συναρτήσεις της R.

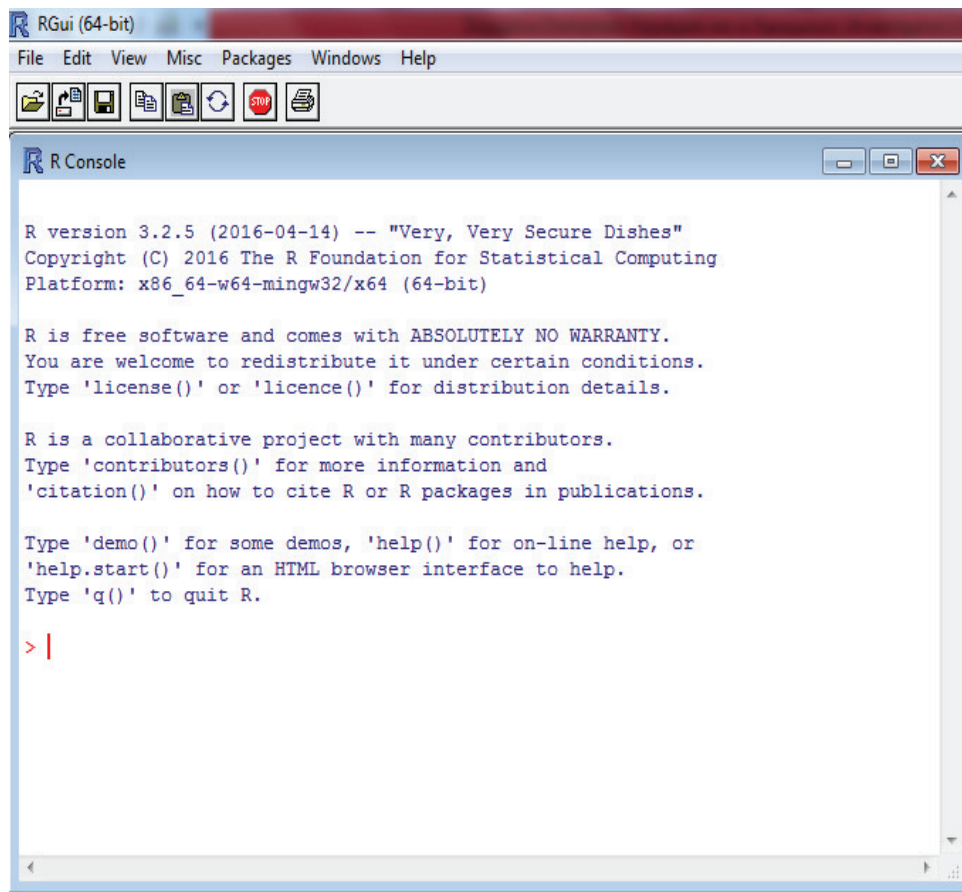
Συνάρτηση	Λειτουργία
sqrt()	Τετραγωνική ρίζα
abs()	Απόλυτη τιμή
log()	Φυσικός λογάριθμος
log2()	Λογάριθμος με βάση 2
exp()	Εκθετική συνάρτηση
cos()	Συνημίτονο
sin()	Ημίτονο
tan()	Εφαπτομένη
acos()	Τόξο συνημίτονου
asin()	Τόξο ημιτόνου
factorial()	Παραγοντικό

- Στην **R** το **pi** δηλώνει το $\pi=3.14...$
- Η συνάρτηση **round()** παίρνει ως τιμές τον αριθμό που θέλουμε να στρογγυλοποιήσουμε και τα δεκαδικά ψηφία που θέλουμε να εμφανίζονται.
- Η συνάρτηση **builtins()** επιστρέφει μια λίστα με όλες τις συναρτήσεις της **R**.
- **Gamma()**: Συνάρτηση γάμμα.

- **Ceiling()**: Μεγαλύτερος ακέραιος.
- **Floor()**: Μικρότερος ακέραιος.

4.6 Παρουσίαση της γλώσσας R σε Windows

Με διπλό κλικ στο εικονίδιο της R που βρίσκεται στην επιφάνεια εργασίας του υπολογιστή μας, τρέχουμε το λογισμικό. Εμφανίζεται το παρακάτω παράθυρο:



```
RGui (64-bit)
File Edit View Misc Packages Windows Help
[Icons]

R Console
R version 3.2.5 (2016-04-14) -- "Very, Very Secure Dishes"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

Για να εκτελέσουμε μια εντολή θα την πληκτρολογήσουμε και στην συνέχεια θα πατήσουμε το πλήκτρο **Enter**.

```
> 5*4/2  
[1] 10  
> |
```

Αν επιλέξουμε από την μπάρα εργαλείων του προγράμματος **Edit > Clear Console**, μας δίνει την δυνατότητα να διαγράψουμε το περιεχόμενο της κονσόλας.

4.6.1 Επιλογή Βοήθειας

Σε περίπτωση που θα χρειαστούμε βοήθεια για την σύνταξη κάποιας συνάρτησης θα πρέπει να πληκτρολογήσουμε στην κονσόλα την συνάρτηση **help** και σε παρένθεση το **όνομα της συνάρτησης** που χρειαζόμαστε ή και επίσης μπορούμε την εντολή : **? όνομα συνάρτησης** .

Για παράδειγμα χρησιμοποιούμε την συνάρτηση **round()**

R Console

```
> help(round)
> ?round
> |
```

Θα ανοίξει ένα παράθυρο βοήθειας το οποίο θα είναι σε **web** μορφή, όπου θα περιέχει την περιγραφή, την χρήση, τα ορίσματα, τις λεπτομέρειες, τις μεθόδους και επίσης παραδείγματα της συνάρτησης **round**.

Rounding of Numbers

Description

`ceiling` takes a single numeric argument `x` and returns a numeric vector containing the smallest integers not less than the corresponding elements of `x`.

`floor` takes a single numeric argument `x` and returns a numeric vector containing the largest integers not greater than the corresponding elements of `x`.

`trunc` takes a single numeric argument `x` and returns a numeric vector containing the integers formed by truncating the values in `x` toward 0.

`round` rounds the values in its first argument to the specified number of decimal places (default 0).

`signif` rounds the values in its first argument to the specified number of significant digits.

Usage

```
ceiling(x)
floor(x)
trunc(x, ...)
```

```
round(x, digits = 0)
signif(x, digits = 6)
```

Arguments

`x` a numeric vector. Or, for `round` and `signif`, a complex vector.

`digits` integer indicating the number of decimal places (`round`) or significant digits (`signif`) to be used. Negative values are allowed (see 'Details').

... arguments to be passed to methods.

Arguments

`x`
a numeric vector. Or, for `round` and `signif`, a complex vector.

`digits`
integer indicating the number of decimal places (`round`) or significant digits (`signif`) to be used. Negative values are allowed (see 'Details').

...
arguments to be passed to methods.

Details

These are generic functions: methods can be defined for them individually or via the `Math` group generic.

Note that for rounding off a 5, the IEC 60559 standard is expected to be used, 'go to the even digit'. Therefore `round(0.5)` is 0 and `round(-1.5)` is -2. However, this is dependent on OS services and on representation error (since e.g. 0.15 is not represented exactly, the rounding rule applies to the represented number and not to the printed number, and so `round(0.15, 1)` could be either 0.1 or 0.2).

Rounding to a negative number of digits means rounding to a power of ten, so for example `round(x, digits = -2)` rounds to the nearest hundred.

For `signif` the recognized values of `digits` are 1...22, and non-missing values are rounded to the nearest integer in that range. Complex numbers are rounded to retain the specified number of digits in the larger of the components. Each element of the vector is rounded individually, unlike printing.

These are all primitive functions.

S4 methods

These are all (internally) S4 generic.

`ceiling`, `floor` and `trunc` are members of the `Math` group generic. As an S4 generic, `trunc` has only one argument.

`round` and `signif` are members of the `Math2` group generic.

Warning

The realities of computer arithmetic can cause unexpected results, especially with `floor` and `ceiling`. For example, we 'know' that `floor(log(x, base = 8))` for `x = 8` is 1, but 0 has been seen on an R platform. It is normally necessary to use a tolerance.

Με την συνάρτηση **example** μπορούμε να εμφανίσουμε παραδείγματα και εφαρμογές μιας συνάρτησης.

Για παράδειγμα: **example(log)**

```
R Console

> example(log)

log> log(exp(3))
[1] 3

log> log10(1e7) # = 7
[1] 7

log> x <- 10^-(1+2*1:9)

log> cbind(x, log(1+x), log1p(x), exp(x)-1, expm1(x))
      x
[1,] 1e-03 9.995003e-04 9.995003e-04 1.000500e-03 1.000500e-03
[2,] 1e-05 9.999950e-06 9.999950e-06 1.000005e-05 1.000005e-05
[3,] 1e-07 1.000000e-07 1.000000e-07 1.000000e-07 1.000000e-07
[4,] 1e-09 1.000000e-09 1.000000e-09 1.000000e-09 1.000000e-09
[5,] 1e-11 1.000000e-11 1.000000e-11 1.000000e-11 1.000000e-11
[6,] 1e-13 9.992007e-14 1.000000e-13 9.992007e-14 1.000000e-13
[7,] 1e-15 1.110223e-15 1.000000e-15 1.110223e-15 1.000000e-15
[8,] 1e-17 0.000000e+00 1.000000e-17 0.000000e+00 1.000000e-17
[9,] 1e-19 0.000000e+00 1.000000e-19 0.000000e+00 1.000000e-19
```

4.6.2 Βασική Σύνταξη Εντολών

Για την σύνταξη των εντολών χρησιμοποιούνται αλφαριθμητικές εκφράσεις. Υπάρχει διαχωρισμός ανάμεσα στα κεφαλαία και στα μικρά γράμματα (το **x** είναι διαφορετικό από το **X**). Οι εντολές μπορεί να είναι εκφράσεις ή και εκχωρίσεις.

Παραδείγματα:

Υπολογισμός μια έκφρασης όπως φαίνεται παρακάτω

```
> sqrt(16) +5
[1] 9
> |
```

Μια εντολή εκχώρησης όπως φαίνεται παρακάτω

```
> x <- 15+3-2
> x
[1] 16
> |
```

Καταχωρούμε μια τιμή για το x και με το πλήκτρο **Enter** καταχωρείται, δεν εμφανίζεται στην κονσόλα εκτός και αν ζητηθεί.

4.6.3 Αποθήκευση και Επανάκτηση Αρχείων

Μας παρέχεται η δυνατότητα να αποθηκεύσουμε τα αντικείμενα που θέλουμε σαν προκαθορισμένο πρότυπο ή σαν ένα απλό αρχείο κειμένου. Η αποθήκευση γίνεται με την εντολή:

```
> save(Dimitris, file="Dimitris.Rdata")
```

Το **Dimitris** είναι το όνομα του αντικειμένου και το **Dimitris.Rdata** είναι το όνομα του αρχείου που θα αποθηκευτεί.

```
> save(Dimitris, file="Dimitris.Rdata", ascii=TRUE)
```

Η παράμετρος που χρησιμοποιείται στην παραπάνω εντολή είναι για την αξιοποίηση αυτού του αρχείου και από άλλα στατιστικά λογισμικά πακέτα. Οπότε αποθηκεύεται σαν ένα απλό αρχείο κειμένου.

Για να **επανακτήσουμε** το προηγούμενο αυτό αντικείμενο, δίνουμε την εντολή:

```
> load ("Dimitris.Rdata")
```

4.7 Αντικείμενα

Τα αντικείμενα δεδομένων μπορούν αν είναι σε διάφορες μορφές στις οποίες μπορούν να αποθηκευτούν δεδομένα στην **R**.

Κάθε αντικείμενο μπορεί να είναι :

- **Πραγματικός αριθμός**
- **Μιγαδικός αριθμός**
- **Δεδομένο λογικής**
- **Δεδομένα χαρακτήρων**

Με την εντολή **class()** μπορούμε να δούμε ποιος είναι ο τύπος ενός αντικειμένου. Επίσης τα αντικείμενα στην R χωρίζονται σε διάφορες δομές:

- **Διανύσματα (vectors)**
- **Δισδιάστατοι Πίνακες (matrices)**
- **Πολυδιάστατοι Πίνακες (arrays)**
- **Πλαίσια Δεδομένων (data frames)**
- **Λίστες (Lists)**

4.7.1 Διανύσματα

Ένα διάνυσμα στην R είναι ένα σύνολο τιμών σε σειρά του ίδιου τύπου.

Ανάλογα με το είδος των αντικειμένων έχουμε:

- **Αριθμητικά Διανύσματα**
- **Διανύσματα Χαρακτήρων**
- **Λογικά Διανύσματα**
- **Διανύσματα Κατηγοριών**

Με την εντολή `c` μπορούμε να προσδιορίσουμε ένα διάνυσμα:

```
> x <- c(1,3,4,5)
> x
[1] 1 3 4 5
> |
```

Μήκος Διανύματος:

```
> x <- c(1,3,4,5)
> length(x)
[1] 4
```

Ελάχιστη / Μέγιστη τιμή

```
> x <- c(1,3,4,5)
> min(x)
[1] 1
> max(x)
[1] 5
```

Άθροισμα και γινόμενο τιμών

```
> x <- c(1,3,4,5)
> sum(x)
[1] 13
> prod(x)
[1] 60
> |
```

Σε περίπτωση κατασκευής διανυσμάτων που χρειάζεται να γίνει επανάληψη κάποιων τιμών, χρησιμοποιούμε την συνάρτηση **rep()**. Η συνάρτηση **rep()** καθορίζει την τιμή των επαναλήψεων που θα γίνουν με όρισμα το **times**. Ή το μέγεθος με όρισμα το **length**.

```
> rep(NA, 5)
[1] NA NA NA NA NA
> rep(x, 3)
[1] 1 3 4 5 1 3 4 5 1 3 4 5
> |
```

4.7.2 Δισδιάστατος Πίνακας

Είναι μια δομή δεδομένων που χρησιμοποιείται για να τακτοποιηθούν οι τιμές κατά γραμμές και κατά στήλες. Για την δημιουργία ενός πίνακα χρησιμοποιούμε την εντολή **matrix()**, και των αριθμό των γραμμών και των στηλών. Η εντολή **dim()** δίνει την διάσταση του πίνακα.

```
> x <- 1:10
> x<- matrix (x, ncol=2)
> x
      [,1] [,2]
[1,]    1    6
[2,]    2    7
[3,]    3    8
[4,]    4    9
[5,]    5   10
> |
```

Για να δημιουργήσουμε διαγώνιους πίνακες χρησιμοποιούμε την εντολή **diag()**.

```
> diag(1:4)
      [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    2    0    0
[3,]    0    0    3    0
[4,]    0    0    0    4
> |
```

Πράξεις Πινάκων

Σύμβολο	Πράξη
%*%	Πολλαπλασιασμός;
T()	Ανάστροφος Πίνακας
Solve()	Αντίστροφος Πίνακας
Eigen()	Ιδιοτιμές και ιδιοδιανύσματα
Det()	Ορίζουσα Πίνακα

4.7.3 Πίνακες μεγαλύτερης διάστασης

Οι πολυδιάστατοι πίνακες, είναι πίνακες με διαστάσεις 3 ή περισσότερες.

Για την δημιουργία τους χρησιμοποιούμε την εντολή **array()** και για να δηλώσουμε τις διαστάσεις τους την εντολή **dim()**.

```
> x<- array(c(1:12,36:48), dim=c(2,3,4))
> x
, , 1
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
, , 2
     [,1] [,2] [,3]
[1,]    7    9   11
[2,]    8   10   12
, , 3
     [,1] [,2] [,3]
[1,]   36   38   40
[2,]   37   39   41
, , 4
     [,1] [,2] [,3]
[1,]   42   44   46
[2,]   43   45   47
> |
```

4.7.4 Πλαίσια Δεδομένων

Τα πλαίσια δεδομένων είναι δισδιάστατοι πίνακες που σε αυτούς γίνεται ταξινόμηση των τιμών κατά την στήλη τους και ανεξάρτητα από τον τύπο της.

Με την εντολή **data.frame()** δημιουργούμε ένα πλαίσιο δεδομένων.

Για παράδειγμα:

```
> Gender<- c('Male', 'Male', 'Female', 'Female')
> Gender
[1] "Male" "Male" "Female" "Female"
> Smoking<-c(F,T,T,F)
> Smoking
[1] FALSE TRUE TRUE FALSE
> Drinking<-c(T,F,T,F)
> Drinking
[1] TRUE FALSE TRUE FALSE
> sample <- data.frame(Gender, Smoking, Drinking)
> sample
  Gender Smoking Drinking
1  Male   FALSE     TRUE
2  Male    TRUE    FALSE
3 Female    TRUE     TRUE
4 Female   FALSE    FALSE
> |
```

4.7.5 Λίστες

Οι λίστες είναι η ανάγκη δημιουργίας για αντικείμενα δεδομένων τα οποία περιέχουν διαφορετικές δομές.

Για να δημιουργήσουμε μια λίστα θα χρησιμοποιήσουμε την εντολή **list()** με παραμέτρους τα αντικείμενα που θέλουμε μαζί με τα ονόματα τους.

```
> group1 <- c(rep(1,10), rep(2,15))
> group2 <- c(22,34,76,67,3,-5,19)

> groups <- list(case=group1, control=group2)
> groups
$case
 [1] 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

$control
 [1] 22 34 76 67 3 -5 19

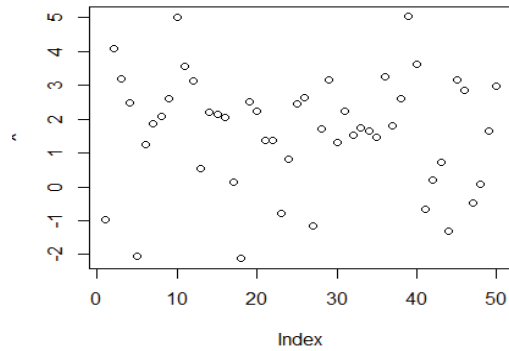
> groups[[2]]
 [1] 22 34 76 67 3 -5 19
> length(groups)
 [1] 2
> mode(groups)
 [1] "list"
> |
```

4.8 ΓΡΑΦΗΜΑΤΑ

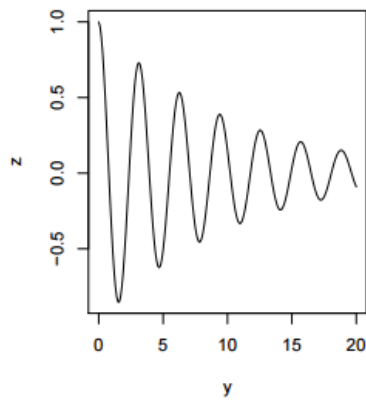
4.8.1 Απλά Γραφήματα

Τα απλά γραφήματα σχετίζονται με τυχαία μονοδιάστατη μεταβλητή. Η βασική εντολή για μια γραφική παράσταση είναι η εντολή **plot**.

```
> x <- rnorm(50, mean=1, sd=2)
> plot(x)
```



```
> y<- seq(0,20, .1)
> z<- exp(-y/10)*cos(2*y)
> plot(y,z, type="l")
```



4.8.2 Γραφικές Δυνατότητες

Οι παρακάτω **γραφικές δυνατότητες** είναι οι πιο χρήσιμες από τις υπόλοιπες.

```
> par(mfrow=c(2,2)) #a 2x2 plot
> plot(1:20,1:20,main="Straight Line")
> hist(rnorm(50), main="Histogram of Normal")
> qqnorm(rt(100,10),main="Samples from t(10)")

~
> plot(density(rnorm(50)),main="Normal Density")
> par(mfrow=c(1,1))
~
```

5 SAS (Statistical Analysis System)

5.1 Εισαγωγή στο λογισμικό SAS

Το στατιστικό λογισμικό SAS (Statistical Analysis System) άρχισε να αναπτύσσεται στο Πανεπιστήμιο της Βόρειας Καρολίνας το 1966 από τους Anthony Barr και James Goodnight και σε λίγα χρόνια εντάχθηκε και ο John Sall. Αρχικά το SAS ξεκίνησε ως ένα project για γεωργικές αναλύσεις, το οποίο χρηματοδοτήθηκε από το Εθνικό Ινστιτούτο Υγείας των Η.Π.Α.. Η ζήτηση για τέτοιου είδους λογισμικά ανεβήκε και έτσι το SAS από το 1976 και μέχρι σήμερα εξυπηρετεί πελάτες από όλων των ειδών τις βιομηχανίες, από φαρμακευτικές εταιρίες και τράπεζες μέχρι και σε ακαδημαϊκές και κυβερνητικές οντότητες.

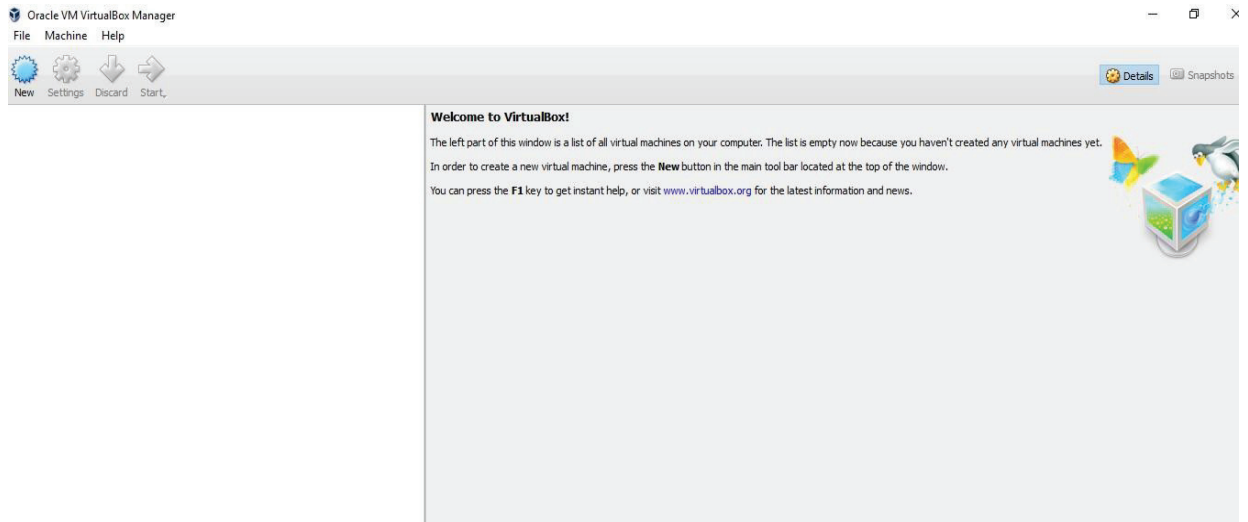
Το λογισμικό SAS είναι ένα ολοκληρωμένο σύστημα από λύσεις λογισμικού που επιτρέπει στον χρήστη τις ακόλουθες εφαρμογές :

- Κατοχύρωση δεδομένων, ανάκτηση και διαχείριση τους
- Εγγραφή έκθεσης και σχεδιασμός γραφημάτων
- Στατιστική και Μαθηματική ανάλυση
- Επιχειρηματικές προβλέψεις και υποστήριξη αποφάσεων
- Έρευνα και διαχείριση project
- Ανάπτυξη Εφαρμογών

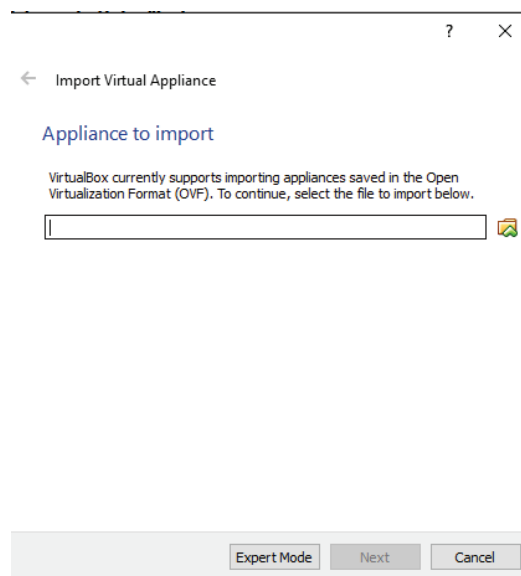
5.2 Εγκατάσταση του SAS

Για να εγκαταστήσουμε το λογισμικό στον υπολογιστή μας, θα επισκευτούμε την ιστοσελίδα της SAS (www.sas.com) για την λήψη του λογισμικού. Μας δίνει την δυνατότητα να επιλέξουμε το λειτουργικό το οποίο χρησιμοποιήσουμε (Windows, OS X, LINUX) στην περίπτωση μας επιλέγουμε Windows. Πρώτα όμως θα χρειαστεί να

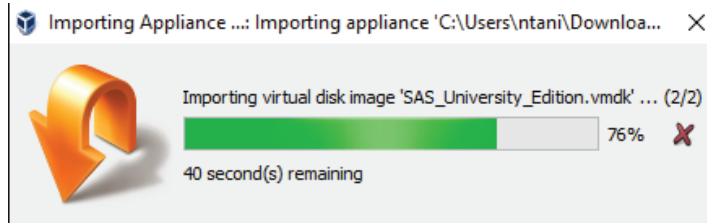
κάνουμε λήψη και εγκατάσταση το Oracle Virtual Box ή το VMware. Αφου εγκαταστήσουμε το Virtual Box θα κάνουμε λήψη και το SAS University Edition. Μετά και την λήψη και του SAS θα ανοίξουμε το Virtual Box. Θα εμφανιστεί το παρακάτω παράθυρο :



Στην συνέχεια θα πρέπει να προσθέσουμε το SAS University Edition στο Virtual Box. Για να γίνει αυτό θα επιλέξουμε το **File** από την γραμμή του menu, στην συνέχεια **Import Appliance...**

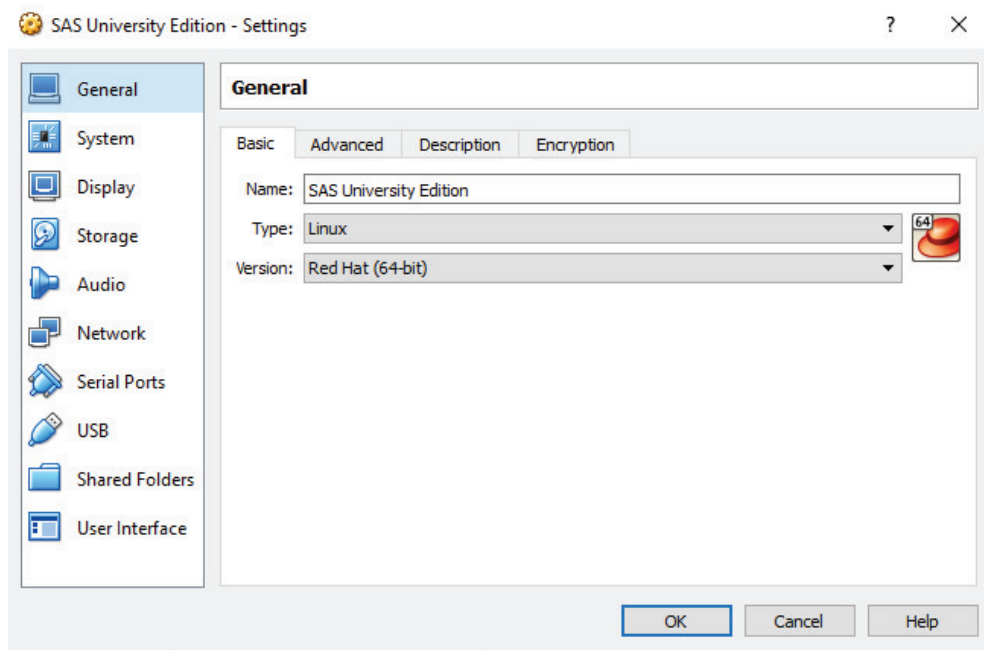


Επιλέγουμε **Expert Mode** και στο παράθυρο που εμφανίζεται επιλέγουμε το αρχείο με κατάληξη .ova, που είναι το αρχείου που κάναμε λήψη για το SAS. Στην συνέχεια στο παράθυρο που είμαστε θα κάνουμε κλικ στην επιλογή **Import**.

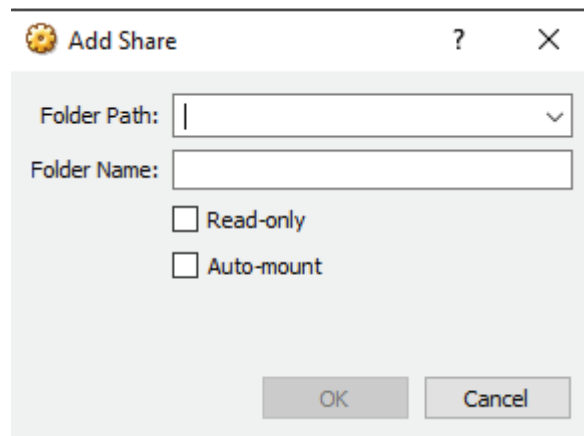


Αφού ολοκληρωθεί η πρόσθεση του αρχείου SAS στο Virtual Box, θα δημιουργήσουμε έναν φάκελο (όπου μας εξυπηρετεί) με ονομασία **SASUniversityEdition** και μέσα σε αυτόν τον φάκελο θα δημιουργήσουμε ακόμη έναν νέο φάκελο με ονομασία **myfolders**.

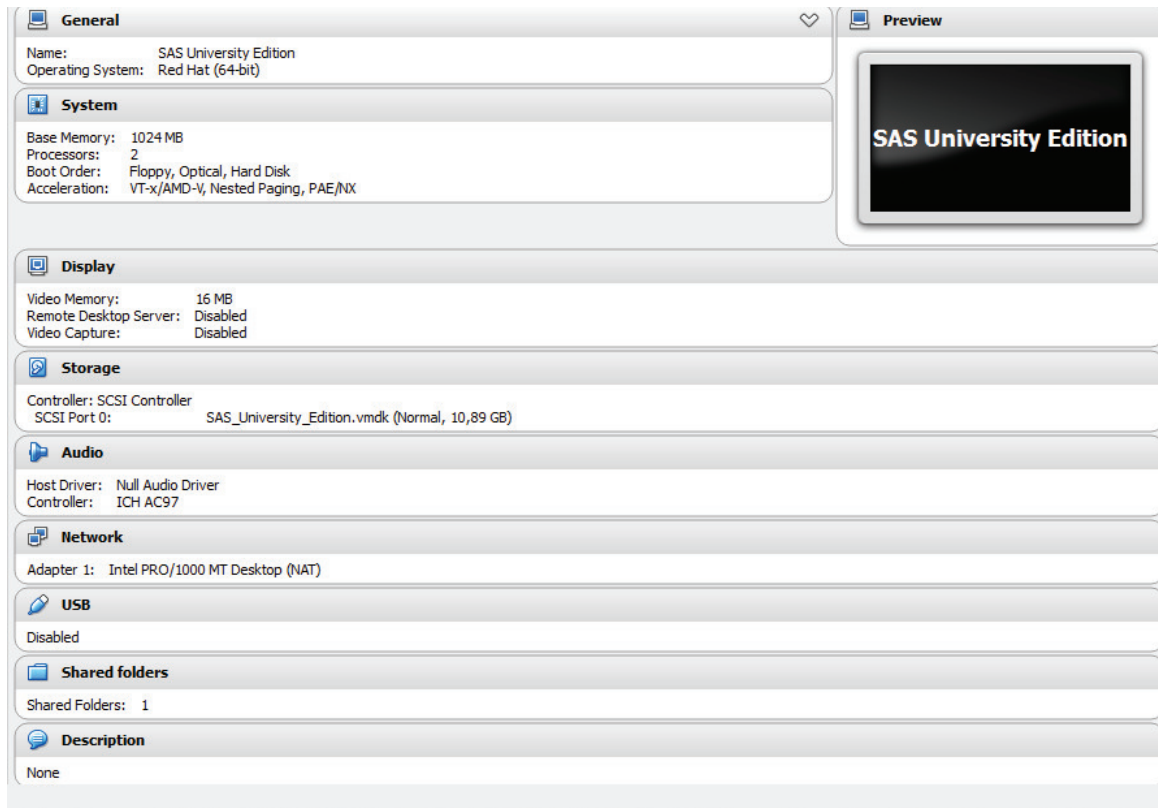
Στην συνέχεια θα ρυθμίσουμε το Virtual Box με το SAS. Στο Virtual Box, στην γραμμή του μενού θα επιλέξουμε **Machine, Settings** και θα μας εμφανίσει το παρακάτω παράθυρο :



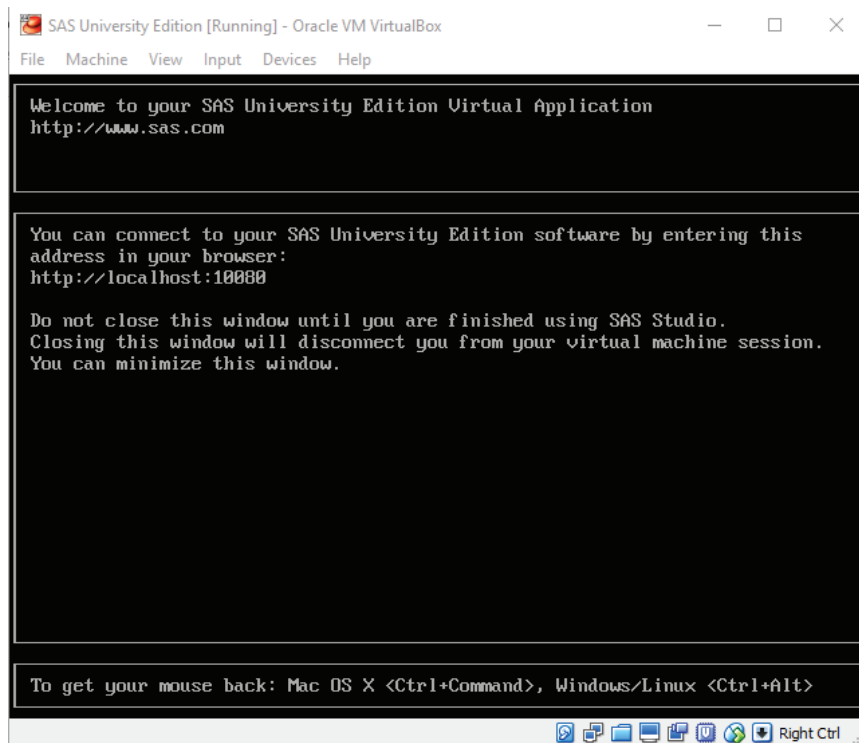
Από την αριστερή στήλη θα επιλέξουμε το **Shared Folders**. Στο νέο παράθυρο που θα εμφανιστεί θα επιλέξουμε στα δεξιά το εικονίδιο με όνομα **add folders**.



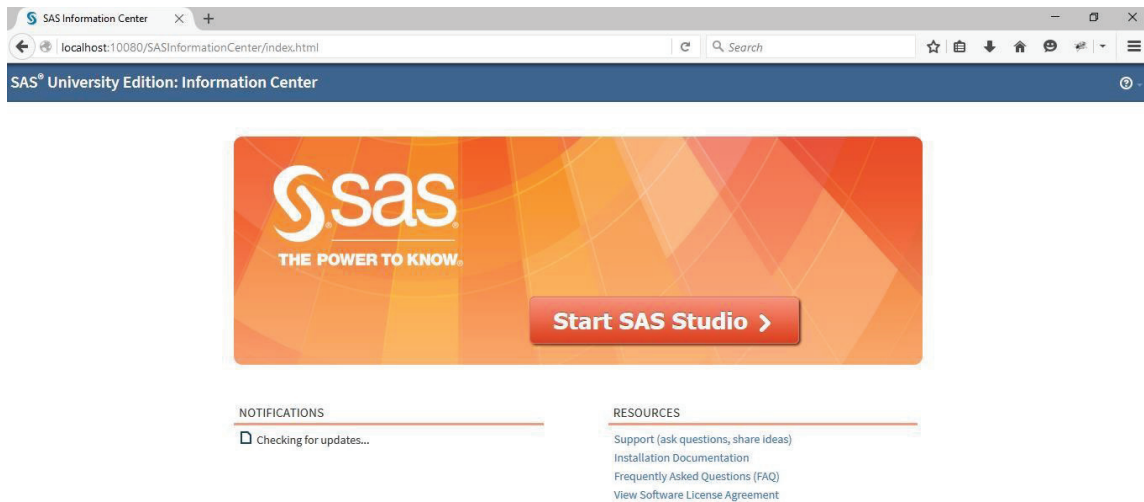
Στο παράθυρο αυτό που εμφανίστηκε, στο πεδίο **Folder Path**, θα βρούμε το μονοπάτι όπου έχουμε αποθηκεύσει τον φάκελο **myfolders**. Θα τον επιλέξουμε και θα κλικάρουμε και την επιλογή **Auto-mount**.



Από την γραμμή του μενού επιλέγουμε **Machine, Start, Normal Start**. Στο παράθυρο που θα εμφανιστεί αφού θα περιμένουμε λίγα δευτερόλεπτα, θα είμαστε έτοιμοι να τρέξουμε το SAS.



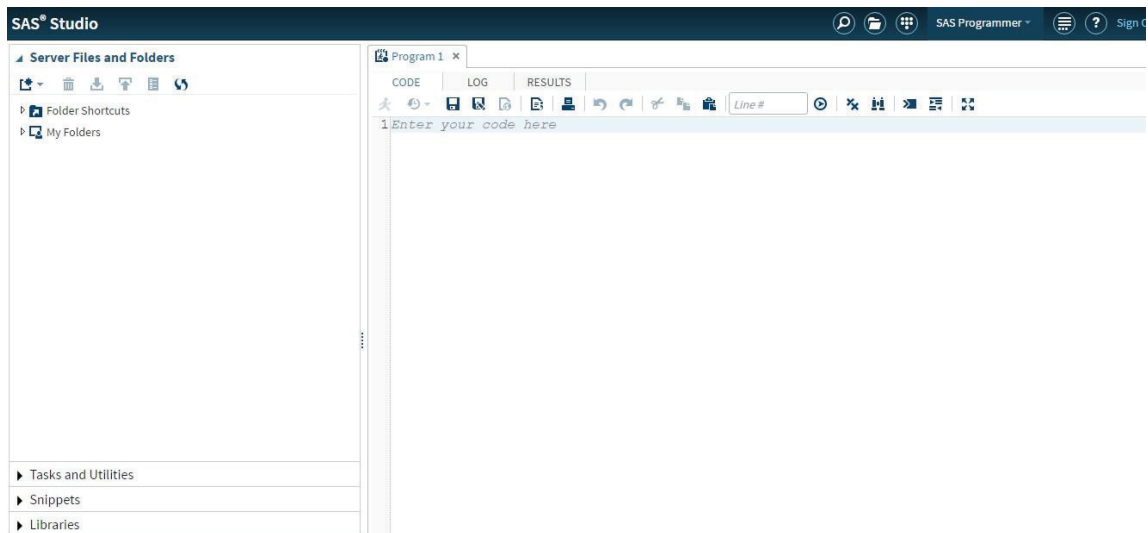
Θα πάρουμε την διεύθυνση που μας δίνεται παραπάνω <http://localhost:10080> και θα την ανοίξουμε σε ένα **browser** που χρησιμοποιούμε.



5.3 Βασικά του λογισμικού

Το λογισμικό SAS αποτελείται από 3 παράθυρα:

- **Code**
Στο παράθυρο αυτό γράφουμε το πρόγραμμα σε μορφή κώδικα
- **Log**
Παράθυρο στο οποίο το SAS εμφανίζει μηνύματα με λάθη τα οποία μπορεί έχουν προκύψει στο πρόγραμμα.
- **Results**
Το παράθυρο στο οποίο εμφανίζονται τα αποτελέσματα.



5.3.1 Basic Components of SAS

- Το λογισμικό περιμένει τα δεδομένα να έχουν μια ειδική δομή. Αυτή η ειδική δομή λέγεται SAS dataset.
- Η δομή του SAS dataset είναι της μορφής πίνακα και περιέχει Μεταβλητές και Παρατηρήσεις.
- Οι σειρές είναι οι Observations (παρατηρήσεις)
- Οι στήλες είναι οι Variables (μεταβλητές)

Ένα πρόγραμμα SAS είναι κατασκευασμένο από 2 μπλοκ:

- DATA Step
- PROC Step

Ένα τυπικό πρόγραμμα SAS αρχίζει με ένα DATA Step για την δημιουργία SAS dataset και μετά περνάει στην PROC Step όπου πραγματοποιείται η ανάλυση.

DATA Step: Είναι μια κατάσταση, η οποία διαβάζει τα δεδομένα, δημιουργεί νέες datasets ή μεταβλητές, και επίσης διαχειρίζεται τις datasets.

PROCEDURES: Κατάσταση, η οποία μπορεί να εκτελέσει στατιστικές αναλύσεις, να δημιουργήσει και να εκτυπώσει εκθέσεις.

Για παράδειγμα:

DATA Step:

```
DATA distance;  
  miles = 23;  
  kilometer = 1.61*miles;  
  RUN;
```

PROC Step:

```
PROC PRINT DATA = distance;  
  RUN;
```

- Οποιοσδήποτε συνδυασμός DATA Step και Procedures μπορεί να χρησιμοποιηθεί.
- Η κατάσταση Run είναι επιθυμητό να χρησιμοποιείται παντού στο πρόγραμμα.

5.3.2 Κανονισμοί και εντολές στην SAS

Η SAS μπορεί να διαβάσει δεδομένα η να δημιουργήσει ένα data set από εξωτερικά αρχεία όπως είναι τα txt και τα csv.

Βασική δομή κώδικα:

DATA dataset-name;

INFILE file-specification<options> <host-options>;

INPUT <specification(s)> <@|@@>;

Run;

INFILE: Η κατάσταση αυτή διαβάζει τα αρχεία που θέλουμε. Ακολουθείται από file-specification<options> <host-options> που δηλώνουν το μονοπάτι που βρίσκεται το αρχείο μας και ο τύπος του αρχείου.

INPUT: Η κατάσταση αυτή προσθέτει, αναθέτει και συνδιάζει τις τιμές με τις αντίστοιχες μεταβλητές.

Για παράδειγμα :

Data newdata;

Infile 'C:/dat.txt';

Input name \$ age;

Run;

Το σύμβολο \$ χρησιμοποιείται για να δηλώσει την τιμή ως character, διαφορετικά η SAS το δηλώνει σαν αριθμητική τιμή.

Μια SAS data set μπορεί επίσης να δημιουργηθεί από την πληκτρολόγηση τιμών στον editor του προγράμματος χρησιμοποιώντας DATALINES.

Για παράδειγμα :

```
7 data newdata;  
8 Input name $ age;  
9 DATALINES;  
10 Mary 24  
11 Suzan 32  
12 ;  
13 Run;
```

Αποτέλεσμα:

Total rows: 2 Total columns: 2

	name	age
1	Mary	24
2	Suzan	32

5.3.3 SAS Functions

Μια SAS συνάρτηση εκτελεί έναν υπολογισμό και χειρίζεται το σύστημα και επιστρέφει μια τιμή. Οι περισσότερες συναρτήσεις χρησιμοποιούν arguments που δίνονται από τους χρήστες, και λίγες αποκτούν arguments από το λογισμικό τους περιβάλλον.

Η σύνταξη της Συνάρτησης :

Function-name (argument-1<...,argument-n>)

Function-name(OF variable-list)

Where

Function name (names the function)

Argument (Μπορεί να είναι όνομα μεταβλητής, σταθερά, η κάποια έκφραση της SAS, συμπεριλαμβάνοντας και κάποια συνάρτηση)

Αριθμητικές Συναρτήσεις:

SUM: Επιστρέφει το άθροισμα των στοιχείων.

SUM(argument,argument,...)

Παράδειγμα ο παρακάτω πίνακας:

Συνάρτηση	Αποτέλεσμα
X1=sum(4,9,3,8);	24
X2=sum(4,9,3,8,.);	24
X3=sum(of x1-x2);	48
X4=sum(of x1-x3, 5);	101
Y1=20; Y2=30; X5=sum(of y:);	50

Συνάρτηση Character:

SUBSTR (αριστερά από =): Αντικαθιστά τις τιμές των χαρακτληρων που περιέχει.

SUBSTR(argument,position<,n>)=characters-to-replace

Arguments

Argument (καθορίζει μια μετβλητή χαρακτήρα).

Position (καθορίζει μια αριθμητική έκφραση, η οποία είναι στην αρχή της θέσης ενός χαρακτήρα).

Characters-to-replace (καθορίζει μια έκφραση χαρακτήρα που θα αντικαταστήσει τα περιόχμενα ενός στοιχείου).

5.4 Εφαρμογή λογισμικού SAS

5.4.1 Παραδείγματα Συναρτήσεων

Παρακάτω θα δούμε παραδείγματα των συναρτήσεων ROUND, CATT και SCAN.

Συνάρτηση ROUND:

Η συνάρτηση στρογγυλοποιεί το πρώτο στοιχείο στο πλησιέστερο πολλαπλάσιο του δεύτερου στοιχείου, ή στον πλησιέστερο ακέραιο εάν το δεύτερο στοιχείο παραληφθεί.

```
1 Data ROUND;
2 IN Submission history;
3 d1=round(NUM,100);
4 d2=round(NUM,10);
5 d3=round(NUM,1);
6 d4=round(NUM,.1);
7 d5=round(NUM,.01);
8 datalines;
9 12345.67890
10 ;
11 proc print data=round;
12 run;
```

Αφού τρέξουμε τον κώδικα θα έχουμε ως αποτέλεσμα :

Obs	NUM	d1	d2	d3	d4	d5
1	12345.68	12300	12350	12346	12345.7	12345.68

Συνάρτηση CATT:

Η συνάρτηση CATTδιαγράφει τα κενά που υπάρχουν μετά από τους χαρακτήρες ενώ τα κένα που είναι πριν τα διατηρεί.

```
1 data _null_;
2   x='  The  Olym';
3   y='pic Arts Festi';
4   z='  val includes works by D  ';
5   a='ale Chihuly.';
6   result=cats(x,y,z,a);
7   put result $char.;
8 run;
```

Έχει ως αποτέλεσμα :

```
The  Olympic Arts Festival includes works by Dale Chihuly.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds
```

Συνάρτηση SCAN:

Η συνάρτηση Scan είναι ένας τρόπος για να ξεχωρίσουμε λέξεις και αριθμούς από μια φράση ή από μια λίστα.

Παρακάτω βρίσκουμε την πρώτη και την τελευταία λέξη σε ένα string

```

1 data firstlast;
2   input String $60.;
3   First_Word = scan(string, 1);
4   Last_Word = scan(string, -1);
5   datalines4;
6 Jack and Jill
7 & Bob & Carol & Ted & Alice &
8 Leonardo
9 ! $ % & ( ) * + , - . / ;
10 ;;;;
11 proc print data=firstlast;
12 run;

```

Αφού τρέξουμε τον κώδικα θα έχουμε ως αποτέλεσμα :

Obs	String	First_Word	Last_Word
1	Jack and Jill	Jack	Jill
2	& Bob & Carol & Ted & Alice &	Bob	Alice
3	Leonardo	Leonardo	Leonardo
4	! \$ % & () * + , - . / ;		

5.4.2 Proc Print

Η Proc Print μπορεί να χρησιμοποιηθεί για την εμφάνιση ενός data set. Για να μας εμφανίσει όλες τις μεταβλητές από ένα data set, χρησιμοποιούμε την παρακάτω σύνταξη:

```

1 Data demographics;
2 Infile '/folders/myfolders/mydata.txt';
3 Input Gender $ Age Weight Height;
4 run;
5 proc print;
6 run;
7

```

Παρακάτω βλέπουμε και το αποτέλεσμα :

Obs	Gender	Age	Weight	Height
1	M	50	78	1.80
2	F	26	59	1.60
3	F	36	62	1.59
4	F	45	73	1.75
5	M	32	88	1.95
6	M	23	73	1.70
7	F	49	63	1.66
8	M	33	72	1.81
9	M	19	96	1.86
10	F	26	55	1.59

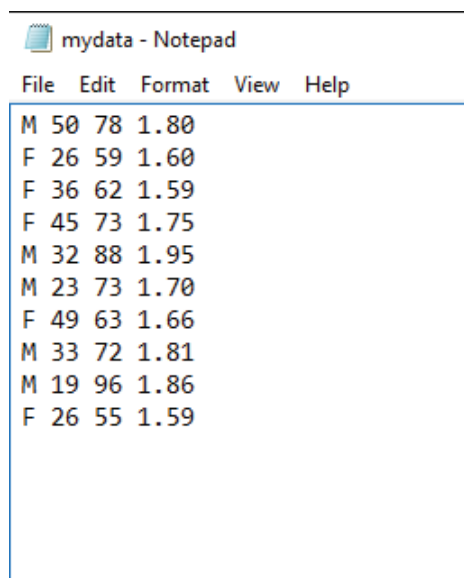
Σε περίπτωση που θέλουμε να είμαστε πιο ακριβείς, για τον λόγο ότι η Proc Print εμφανίζει την data set που δημιουργήθηκε χρονικά τελευταία, στην εντολή θα προσθέσουμε και το όνομα του data που θέλουμε.

```
Proc print data=demographics;
```

```
Run;
```

5.4.3 Proc Frequencies and Proc Means

Στην αρχή θα δημιουργήσουμε ένα αρχείο txt και θα το ονομάσουμε mydata.txt. Το αρχείο αυτό θα περιέχει κάποια δημογραφικά στοιχεία 10 ατόμων.



```
mydata - Notepad
File Edit Format View Help
M 50 78 1.80
F 26 59 1.60
F 36 62 1.59
F 45 73 1.75
M 32 88 1.95
M 23 73 1.70
F 49 63 1.66
M 33 72 1.81
M 19 96 1.86
F 26 55 1.59
```


Η πρώτη στήλη δηλώνει το φύλο, η δεύτερη στήλη δηλώνει την ηλικία, η τρίτη στήλη δηλώνει τα κιλά και η τέταρτη στήλη δηλώνει το ύψος.

Θα προσθέσουμε το αρχείο mydata.txt στο πρόγραμμα SAS

```
1 Data demographics;
2 Infile '/folders/myfolders/mydata.txt';
3 Input Gender $ Age Weight Height;
4 Run;
```

- Στην πρώτη σειρά του κώδικα δίνουμε την ονομασία Data demographics
- Στην δεύτερη σειρά προσθέτουμε το αρχείο στο πρόγραμμα.
- Στην τρίτη σειρά ονομάζουμε τις στήλες
- Τρέχουμε τον κώδικα

Αφού τρέξουμε τον κώδικα θα μας εμφανιστεί το παρακάτω παράθυρο, που θα δούμε το αρχείο txt που προσθέσαμε με τις στήλες όπως τις δηλώσαμε.

	Gender	Age	Weight	Height
1	M	50	78	1.8
2	F	26	59	1.6
3	F	36	62	1.59
4	F	45	73	1.75
5	M	32	88	1.95
6	M	23	73	1.7
7	F	49	63	1.66
8	M	33	72	1.81
9	M	19	96	1.86
10	F	26	55	1.59

Με την συνάρτηση Frequencies έχουμε ως αποτέλεσμα να υπολογίσουμε την συχνότητα, το ποσοστό επί τις εκατό της συχνότητας, την αθροιστική συχνότητα και το ποσοστό επί τις εκατό της αθροιστικής συχνότητας.

```
6 Title "Gender Frequencies";  
7 Proc Freq data=demographics;  
8   Table Gender;  
9 Run;
```

Gender Frequencies

The FREQ Procedure

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	5	50.00	5	50.00
M	5	50.00	10	100.00

Παρακάτω χωρίς να δηλώσουμε το πίνακα Gender θα βγάλουμε την συχνότητα για όλες τις μεταβλητές.

```
6 Title "Gender Frequencies";  
7 Proc Freq data=demographics;  
8  
9 Run;
```

Age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
19	1	10.00	1	10.00
23	1	10.00	2	20.00
26	2	20.00	4	40.00
32	1	10.00	5	50.00
33	1	10.00	6	60.00
36	1	10.00	7	70.00
45	1	10.00	8	80.00
49	1	10.00	9	90.00
50	1	10.00	10	100.00

Weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
55	1	10.00	1	10.00
59	1	10.00	2	20.00
62	1	10.00	3	30.00
63	1	10.00	4	40.00
72	1	10.00	5	50.00
73	2	20.00	7	70.00
78	1	10.00	8	80.00
88	1	10.00	9	90.00
96	1	10.00	10	100.00

Height	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.59	2	20.00	2	20.00
1.6	1	10.00	3	30.00
1.66	1	10.00	4	40.00
1.7	1	10.00	5	50.00
1.75	1	10.00	6	60.00
1.8	1	10.00	7	70.00
1.81	1	10.00	8	80.00
1.86	1	10.00	9	90.00
1.95	1	10.00	10	100.00

Με την συνάρτηση Means μπορούμε να δούμε διάφορα στατιστικά των μεταβλητών.

```

11 Title "Summary Statistics";
12 Proc Means data=demographics;
13     Var Age Weight Height;
14 Run;

```

Summary Statistics

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Age	10	33.9000000	10.9792734	19.0000000	50.0000000
Weight	10	71.9000000	12.9481873	55.0000000	96.0000000
Height	10	1.7310000	0.1242265	1.5900000	1.9500000

5.4.4 Proc Univariate

Η διερευνητική ανάλυση δεδομένων (Exploratory Data Analysis, EDA) είναι ένα σημαντικό πρώτο βήμα σε κάθε τύπο ανάλυσης δεδομένων. Αποτελεσματικό EDA βελτιώνει την ποιότητα της ανάλυσης.

Η Proc Univariate παρότι είναι παρόμοια με την Proc Means. Έχει μεγαλύτερο εύρος υπολογισμού στατιστικών, και συγκεκριμένα χρήσιμη στην εξέταση κατανομής μιας μεταβλητής. Επίσης προσφέρει έναν αριθμό στατιστικών και plots για έναν αριθμό μεταβλητών που είναι χρήσιμα για την EDA.

Παρακάτω βλέπουμε ένα παράδειγμα :

```

1 Data demographics;
2 Infile '/folders/myfolders/mydata.txt';
3 Input Gender $ Age Weight Height;
4 run;
5 proc univariate;
6 var age;
7 run;
8

```

Όπως παρατηρούμε χρησιμοποιήσαμε την Proc Univariate για την μεταβλητή Age. Παρακάτω εμφανίζονται τα αποτελέσματα:

The UNIVARIATE Procedure
Variable: Age

Moments			
N	10	Sum Weights	10
Mean	33.9	Sum Observations	339
Std Deviation	10.9792734	Variance	120.544444
Skewness	0.35643124	Kurtosis	-1.2513921
Uncorrected SS	12577	Corrected SS	1084.9
Coeff Variation	32.3872372	Std Error Mean	3.4719511

Basic Statistical Measures			
Location		Variability	
Mean	33.90000	Std Deviation	10.97927
Median	32.50000	Variance	120.54444
Mode	26.00000	Range	31.00000
		Interquartile Range	19.00000

Tests for Location: $\mu_0=0$				
Test		Statistic	p Value	
Student's t	t	9.763962	Pr > t	<.0001
Sign	M	5	Pr >= M	0.0020
Signed Rank	S	27.5	Pr >= S	0.0020

Quantiles (Definition 5)	
Level	Quantile
100% Max	50.0
99%	50.0
95%	50.0
90%	49.5
75% Q3	45.0
50% Median	32.5
25% Q1	26.0
10%	21.0
5%	19.0
1%	19.0
0% Min	19.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
19	9	33	8
23	6	36	3
26	10	45	4
26	2	49	7
32	5	50	1

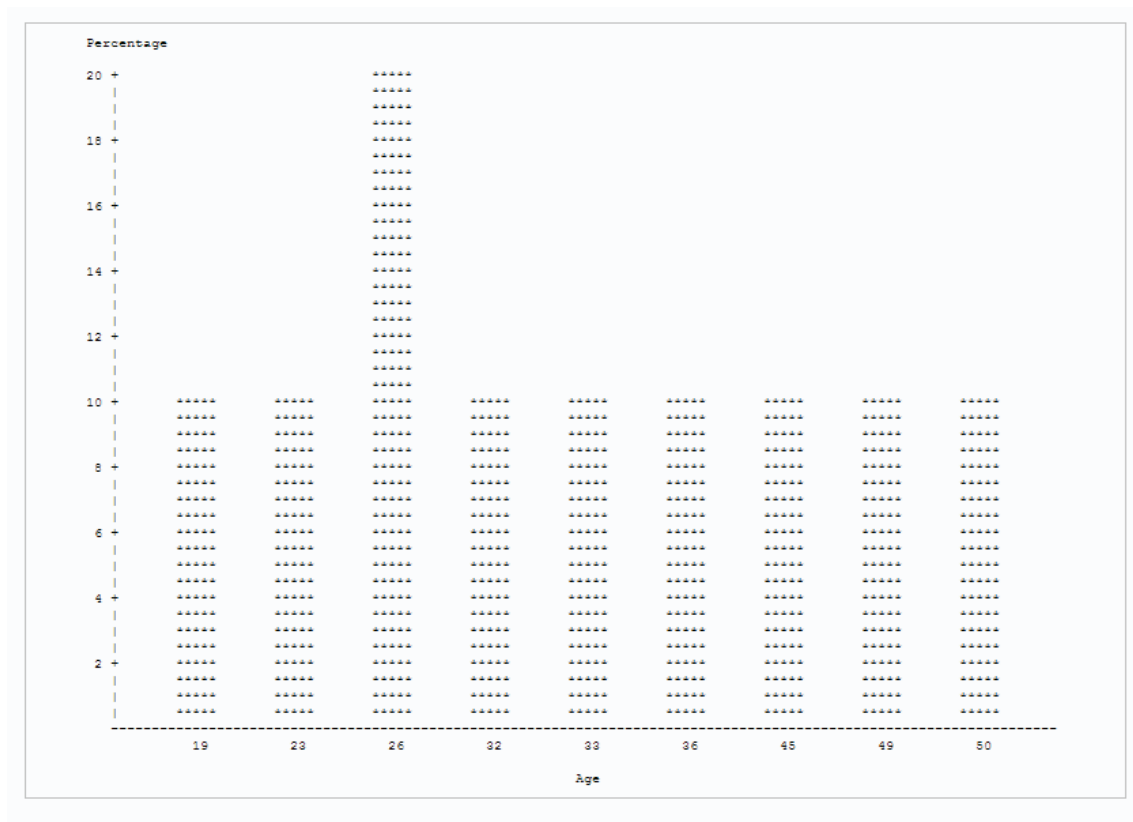
Η Proc Univariate εμφανίζει από μόνη της, τις τιμές από τις 5 υψηλότερες και τις 5 χαμηλότερες. Αν θέλουμε αυτές τις τιμές να τις αναγνωρίσουμε από μια συγκεκριμένη μεταβλητή, θα χρησιμοποιήσουμε την ID.

5.5 Graph in SAS

5.5.1 Bar Chart

Με την proc chart και επιλέγοντας την μεταβλητή που θέλουμε θα εμφανίσουμε ένα γράφημα.

```
1 Data demographics;  
2 Infile '/folders/myfolders/mydata.txt';  
3 Input Gender $ Age Weight Height;  
4 proc chart;  
5 vbar age / discrete type=percent;  
6 run;  
7  
8
```

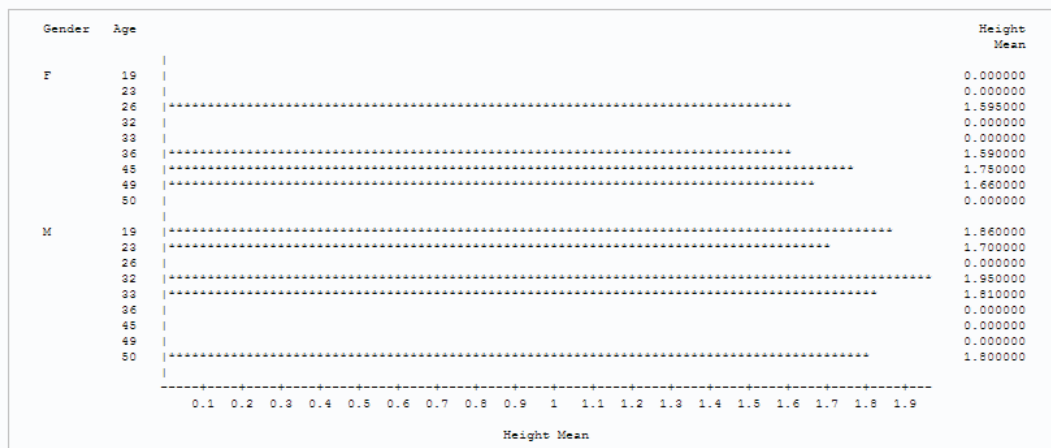


Παρακάτω θα χρησιμοποιήσουμε τις VBAR και HBAR καταστάσεις, οι οποίες θα μας δώσουν ένα ομαδικό bar chart με τις μεταβλητές που θέλουμε.

```

1 Data demographics;
2 Infile '/folders/myfolders/mydata.txt';
3 Input Gender $ Age Weight Height;
4 proc chart ;
5 hbar age / discrete type=mean
6   sumvar=height mean
7   group=gender;
8 run;
9

```

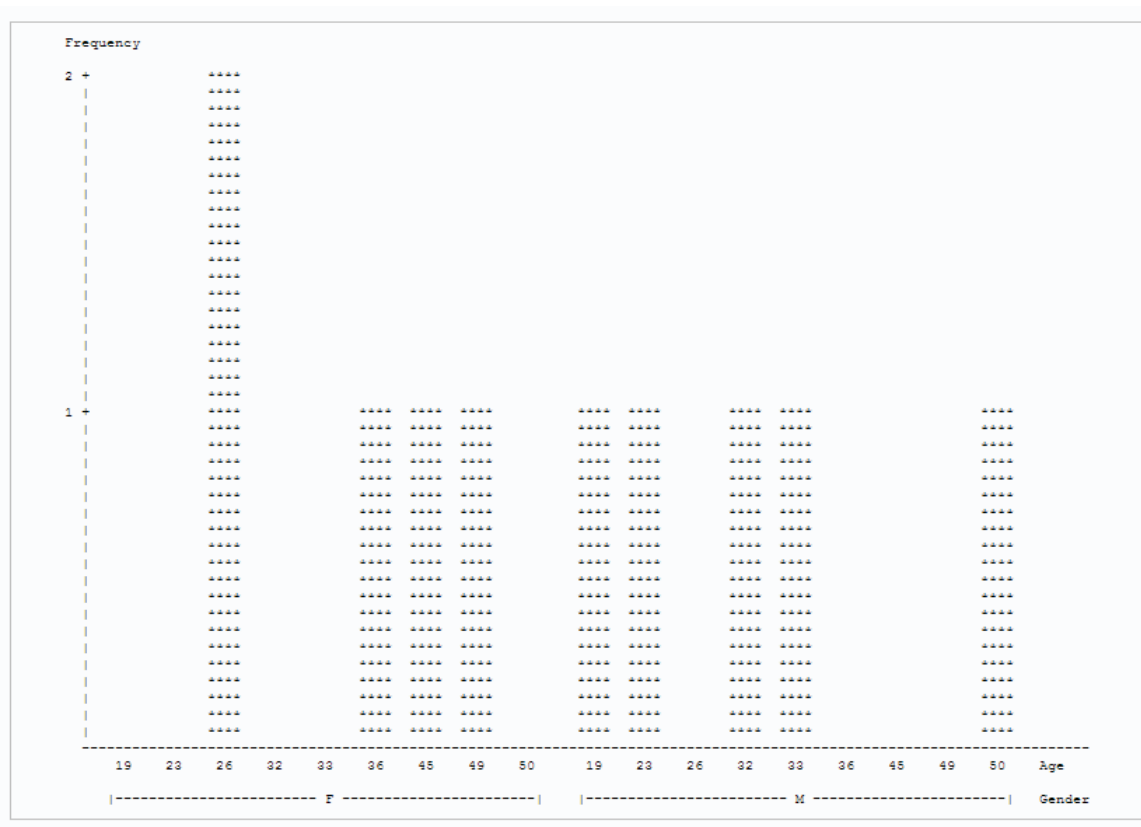


Με vbar statement:


```

1 Data demographics;
2 Infile '/folders/myfolders/mydata.txt';
3 Input Gender $ Age Weight Height;
4 proc chart ;
5 vbar age / discrete type=mean
6 group=gender;
7 run;
8

```



5.5.2 Pie Chart

Θα χρησιμοποιήσουμε ένα άλλο παράδειγμα για την δημιουργία ενός pie chart.

Στην αρχή φτιάχνουμε το γραφικό μας περιβάλλον.

```

1 /* Set the graphics environment */
2 goptions reset=all border cback=white
3         ctext=black htitle=12pt;
4

```

Στην συνέχεια θα δημιουργήσουμε και θα εισάγουμε τα δεδομένα μας.

Τα δεδομένα μας είναι το τμήμα (Dept), η πόλη (Site), Quarter και οι πωλήσεις (Sales).

```

5 /* Create data set TOTALS */
6 data totals;
7     length Dept $ 7 Site $ 8;
8     input Dept Site Quarter Sales;
9     datalines;
10 Parts    Sydney  1 4043.97
11 Parts    Atlanta  1 6225.26
12 Parts    Paris    1 3543.97
13 Repairs  Sydney  1 5592.82
14 Repairs  Atlanta  1 9210.21
15 Repairs  Paris    1 8591.98
16 Tools    Sydney  1 1775.74
17 Tools    Atlanta  1 2424.19
18 Tools    Paris    1 5914.25
19 Parts    Sydney  2 3723.44
20 Parts    Atlanta  2 11595.07
21 Parts    Paris    2 9558.29
22 Repairs  Sydney  2 5505.31
23 Repairs  Atlanta  2 4589.59
24 Repairs  Paris    2 7538.56
25 Tools    Sydney  2 2945.17
26 Tools    Atlanta  2 1903.99
27 Tools    Paris    2 7868.34
28 Parts    Sydney  3 8437.96
29 Parts    Atlanta  3 6847.91
30 Parts    Paris    3 6789.85

```

```

31 Repairs Sydney 3 4426.46
32 Repairs Atlanta 3 5011.66
33 Repairs Paris 3 6510.38
34 Tools Sydney 3 3767.10
35 Tools Atlanta 3 3048.52
36 Tools Paris 3 9017.96
37 Parts Sydney 4 6065.57
38 Parts Atlanta 4 9388.51
39 Parts Paris 4 8509.08
40 Repairs Sydney 4 3012.99
41 Repairs Atlanta 4 2088.30
42 Repairs Paris 4 5530.37
43 Tools Sydney 4 3817.36
44 Tools Atlanta 4 4354.18
45 Tools Paris 4 6511.70
46 ;
47 run;
48

```

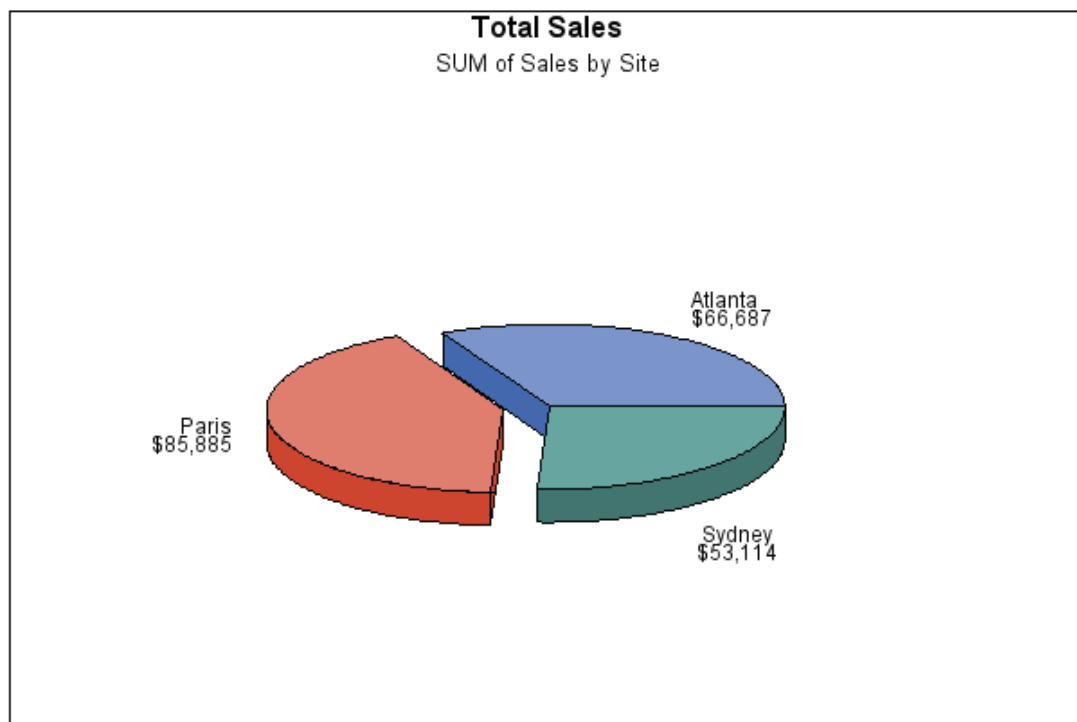
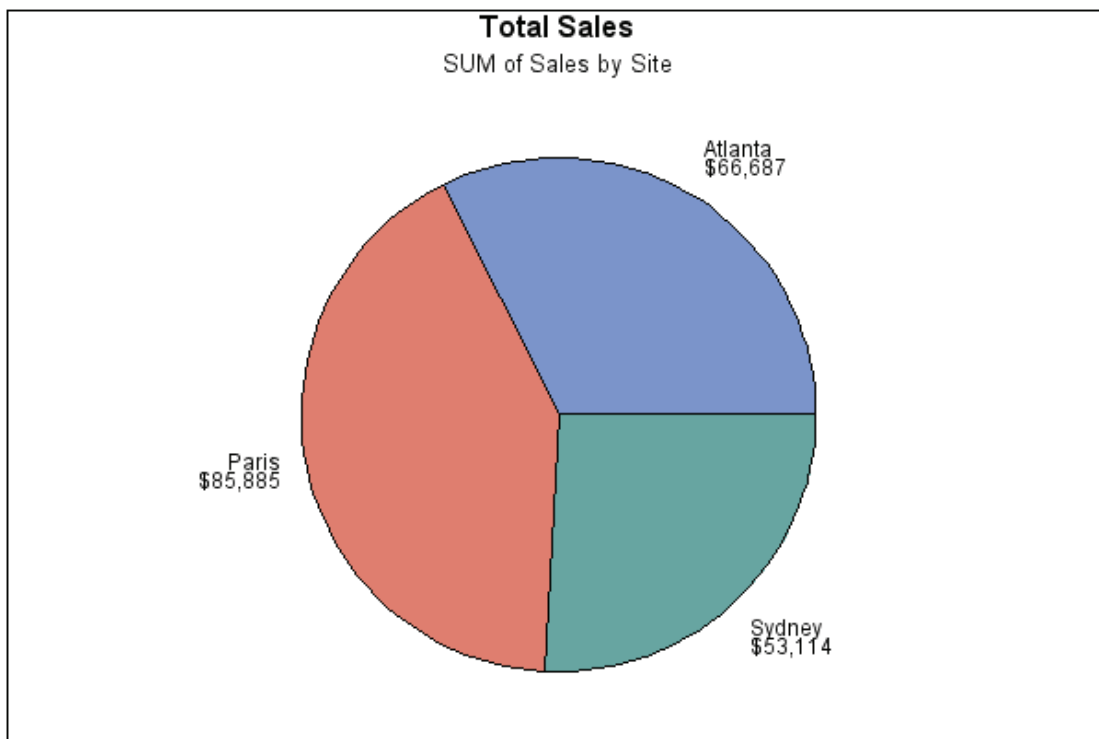
Στην συνέχεια στον κώδικα μας θα φτιάξουμε 2 pie chart. Και τα 2 θα έχουν ως αποτέλεσμα το άθροισμα των πωλήσεων ανά πόλη μόνο που το δεύτερο γράφημα θα εμφανίζεται σε 3d.

```

49 title 'Total Sales';
50
51 /* Generate first pie chart */
52 proc gchart data=totals;
53     format sales dollar8.;
54     pie site / sumvar=sales
55         coutline=black;
56 run;
57 quit;
58
59 /* Generate second pie chart */
60 proc gchart data=totals;
61     format sales dollar8.;
62     pie3d site / sumvar=sales
63         coutline=black
64         explode='Paris';
65 run;
66 quit;
67

```

Αποτελέσματα:

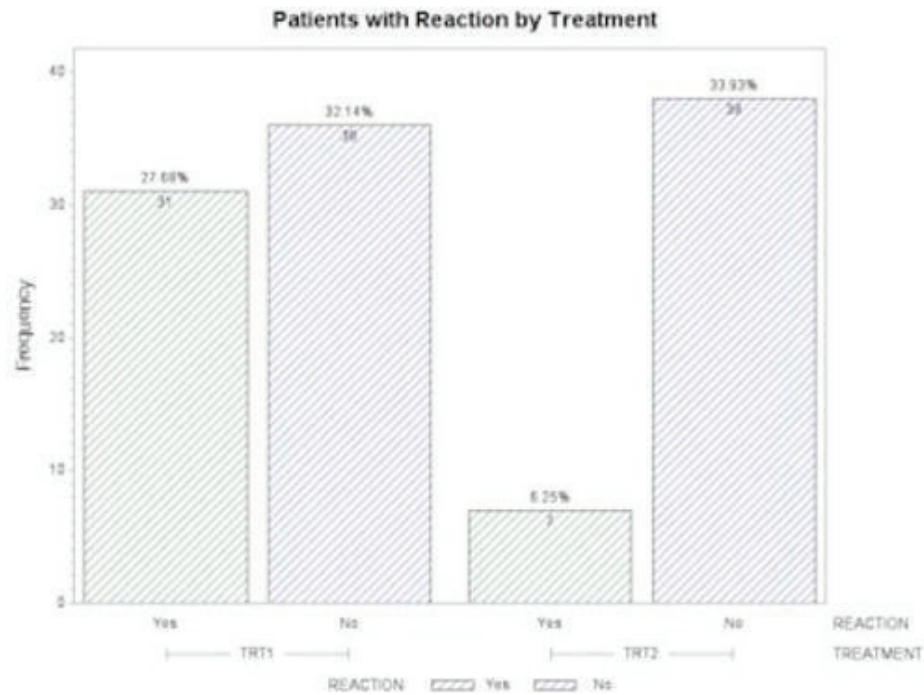


5.5.3 Bar Chart from Counts

Υποθέτουμε ότι έχουμε μετρήσεις από μια ανάλυση και θέλουμε να δημιουργήσουμε ένα bar chart. Σε αυτή τη περίπτωση έχουμε τις μεταβλητές Treatment και Reaction.

- Για Treatment 1 (συμπεριφορά), 31 άνθρωποι Reaction (αντέδρασαν) και 36 όχι (σύνολο 67 άνθρωποι).
- Για Treatment 2, 7 Reaction και 38 όχι (σύνολο 45 άνθρωποι).

```
1 PROC FORMAT;
2 VALUE FMTTREATMENT 1="TRT1" 2="TRT2";
3 VALUE FmtReaction 1="Yes" 2="No";
4 RUN;
5 * Enter Data as counts;
6 data v;input TREATMENT REACTION WT;
7 datalines;
8 1 1 21
9 1 2 36
10 2 1 17
11 2 2 38
12 ;
13 * Set up bar chart;
14 GOPTIONS RESET=ALL;
15 ods html image_dpi=100; *change to 300 for high resolution;
16 pattern1 value=R1 color=green;
17 pattern2 value=R1 color=BLUE;
18 AXIS1 LABEL=(A=90 H=1.5 F="SWISS" "Frequency" C=BLUE);
19 Title "Patients with Reaction by Treatment";
20 proc gchart data=v;
21 vbar REACTION/DISCRETE
22 FREQ=WT
23 INSIDE=FREQ OUTSIDE=PCT
24 coutline=black
25 woutline=1
26 width=20
27 GROUP=TREATMENT
28 SUBGROUP=REACTION
29 RAXIS=AXIS1;
30 ;
31 format TREATMENT FMTTREATMENT. REACTION FmtReaction.;
32 run;
```



Σημειώσεις:

- **Freq=To WT** δηλώνει στο **CHART** να χρησιμοποιήσει την **WT** μεταβλητή (counts, summarized data) αντί ενός στοιχείου ανά καταγραφή.
- Η **label AXIS1** που τοποθετεί το **Y-AXIS** σε γωνία 90 μοιρών.
- Η **INSIDE** και η **OUTSIDE** δηλώνουν αριθμούς και το **PCT** στο γράφημα.
- Η **ODS HTML** μας επιτρέπει να προσδιορίσουμε την ανάλυση, χρησιμοποιούμε 300 για να δημιουργήσουμε υψηλή ανάλυση γραφήματος.

6.ΣΥΜΠΕΡΑΣΜΑΤΑ

Η εργασία αυτή περιγράφει και αναλύει την χρήση των κορυφαίων και ποιοτικότερων διαφόρων στατιστικών λογισμικών πακέτων που κυκλοφορούν σε κλάδους της Πληροφορικής, των Οικονομικών, Κοινωνικών, Πολιτικών και Ιατρικών Επιστημών, καθώς και στον τομέα της Εκπαίδευσης. Η τεχνολογική πρόοδος καθώς και η ανάπτυξη και η διαρκείς διάδοση της πληροφορίας, έχουν σαν αποτέλεσμα την συνεχή ανάπτυξη των στατιστικών λογισμικών.

Τα καλύτερα στατιστικά λογισμικά για το 2016 όπως ανακηρύχτηκαν μετά από ψηφοφορία είναι τα: SPSS, Minitab, R και SAS. Τα έχουμε αναλύσει περιεκτικότερα. Όπως είδαμε τα περισσότερα από αυτά έχουν κάποια κοινά και όμοια χαρακτηριστικά. Παρακάτω θα πραγματοποιήσουμε μια μικρή σύγκριση μεταξύ τους.

Το SPSS είναι γραμμένο σε Java και χρησιμοποιεί την R, Python και την Sax Basic, σε Scripting language. Είναι διαθέσιμο σε Windows, Mac Os και Linux. Επίσης είναι διαθέσιμο σε όλες τις μεθόδους ANOVA (Analysis of Variance) και έχει διαθέσιμα όλα τα υπάρχοντα γραφήματα. Το Minitab είναι διαθέσιμο μόνο για τα Windows. Είναι διαθέσιμο σε όλες τις μεθόδους ANOVA εκτός από την Mixed Model και έχει διαθέσιμα όλα τα υπάρχοντα γραφήματα. Η R είναι γραμμένη σε C, FORTRAN και R και χρησιμοποιεί την R, Python και Perl σε Scripting language. Είναι διαθέσιμη για όλα τα υπάρχοντα λειτουργικά συστήματα. Επίσης είναι διαθέσιμο σε όλες τις μεθόδους ANOVA και έχει διαθέσιμα όλα τα υπάρχοντα γραφήματα. Το λογισμικό SAS είναι γραμμένο σε δική του γλώσσα την SAS language, είναι διαθέσιμο σε Windows, Linux και Unix. Είναι διαθέσιμο σε όλες τις μεθόδους ANOVA και έχει διαθέσιμα όλα τα υπάρχοντα γραφήματα.

Τέλος μέσα από αυτήν την εργασία διακρίνουμε τις ομοιότητες που υπάρχουν στα διάφορα αυτά λογισμικά σε οποιοδήποτε κλάδο επιστημών χρησιμοποιούνται. Με το πέρασμα των χρόνων και με την μεγαλύτερη ακόμα πρόοδος της τεχνολογίας θα οδηγήσει στις αυξημένες απαιτήσεις της πληροφορίας. Περισσότερες πληροφορίες και δεδομένα θα υπάρξουν για συγκέντρωση, ανάλυση και αναπαράσταση τους. Τα στατιστικά λογισμικά θα γίνουν πιο πολύπλοκα, θα αναπτυχθούν και θα είναι έτοιμα να καλύψουν τις νέες ανάγκες.

EΙΚΟΝΕΣ

Εικόνα εξώφυλλου <http://stat.psu.edu/world-campus>

Εικόνα 1 <https://en.wikipedia.org/wiki/SPSS>

Εικόνα 2 http://www.statwks.com/wp-content/uploads/2015/01/ibm_spss_logo.png

Εικόνα 3 <http://www.prweb.com/releases/minitab/17/prweb12521349.htm>

Εικόνα 4 <http://namaristats.com/static/img/langs/logo-r.png>

Εικόνα 5 <http://uwm.edu/software/wp-content/uploads/sites/76/2014/04/sas-logo.jpg>

Εικόνα 6 http://www.statanordic.com/images/stata_logo_med_blue.jpg

Εικόνα 7 http://eits.uga.edu/_resources/files/images/JMP_logo.png

ΒΙΒΛΙΟΓΡΑΦΙΑ

Γναρδέλλης Χαράλαμπος , Ανάλυση δεδομένων με το IBM SPSS Statistics 21 ,
Εκδόσεις Παπαζήσης, Αθήνα 2013

Φουσάκης Δημήτριος, Ανάλυση δεδομένων με χρήση της R, Εκδόσεις Τσότρας ,
Αθήνα 2013

Lawal Bayo , Categorical data analysis with SAS and SPSS applications ,
Εκδόσεις Lawrence Erlbaum, Mahwah, N.J 2003

ΗΛΕΚΤΡΟΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] https://en.wikipedia.org/wiki/Comparison_of_statistical_packages
- [2] <https://en.wikipedia.org/wiki/SPSS>
- [3] <http://www-01.ibm.com/software/analytics/spss/products/statistics/>
- [4] <http://stat-athens.aueb.gr/~akostaki/gr/courses/SPSS14.pdf>
- [5] <http://users.auth.gr/agpapana/SPSS>
- [6] https://www.ucy.ac.cy/hr/documents/2012/SPSS_13.pdf
- [7] [http://www.math.upatras.gr/~adk/lectures/ida/lab1/tutor6\(gr\).pdf](http://www.math.upatras.gr/~adk/lectures/ida/lab1/tutor6(gr).pdf)
- [8] <https://www.minitab.com/en-us/>
- [9] <https://en.wikipedia.org/wiki/Minitab>
- [10] <http://stat-athens.aueb.gr/~akostaki/gr/courses/Minitab12.pdf>
- [11] [https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))
- [12] <https://www.r-project.org/about.html>
- [13] http://www.math.ntua.gr/~fouskakis/Data_Analysis/02.pdf
- [14] <https://cran.r-project.org/doc/contrib/mainfokianoscharalambous.pdf>
- [15] http://stat-athens.aueb.gr/~grstats/notes/r_giagos.pdf
- [16] <http://www.math.auth.gr/sites/default/files/R%20guide%20by%20cmoi.pdf>
- [17] http://www.sas.com/el_gr/home.html
- [18] [https://en.wikipedia.org/wiki/SAS_\(software\)](https://en.wikipedia.org/wiki/SAS_(software))
- [19] <http://www.stata.com/>
- [20] <https://en.wikipedia.org/wiki/Stata>
- [21] http://www.jmp.com/en_us/software.html

[22] [https://en.wikipedia.org/wiki/JMP_\(statistical_software\)](https://en.wikipedia.org/wiki/JMP_(statistical_software))

[23]

[http://nemertes.lis.upatras.gr/jspui/bitstream/10889/5359/3/Nimertis_Alexopoulou\(ekon\).pdf](http://nemertes.lis.upatras.gr/jspui/bitstream/10889/5359/3/Nimertis_Alexopoulou(ekon).pdf)

[24] <http://support.sas.com/>

[25] <http://www.stattutorials.com/>

[26] <http://analytics.ncsu.edu/sesug/2000/p-905.pdf>