

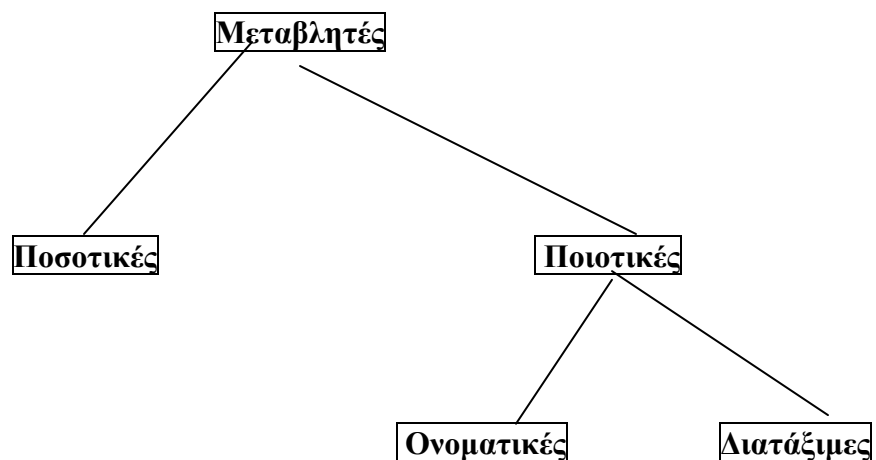
# ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ

## Εισαγωγή και αποθήκευση δεδομένων-Τα βασικά του S.P.S.S.

### 1.1. Εισαγωγή δεδομένων στο S.P.S.S.

Για τη διεξαγωγή μίας στατιστικής μελέτης απαιτείται αρχικά η διατύπωση και ο σχεδιασμός του προς επίλυση προβλήματος. Έπειτα συλλέγονται, είτε από διαθέσιμες πηγές είτε με τη διεξαγωγή έρευνας με ερωτηματολόγιο, τα δεδομένα που αφορούν το προς επίλυση πρόβλημα. Το επόμενο καθοριστικό βήμα, για τη μετέπειτα ανάλυση των δεδομένων, την ερμηνεία των αποτελεσμάτων της ανάλυσης και την εξαγωγή συμπερασμάτων, είναι η καταχώρηση των δεδομένων που θα συλλέξουμε σε κάποιο λογισμικό πακέτο ανάλυσης δεδομένων.

Από τα παραπάνω γίνεται κατανοητό ότι η καταχώρηση των προς ανάλυση δεδομένων στο λογισμικό επιδρά στη μετέπειτα ανάλυση και στην εξαγωγή των συμπερασμάτων. Οι ερωτήσεις που διατυπώνονται σε ένα ερωτηματολόγιο αντιστοιχούν τις περισσότερες φορές σε μία μεταβλητή. Γνωρίζουμε από τη θεωρία της Στατιστικής ότι οι μεταβλητές διακρίνονται σε ποσοτικές και ποιοτικές, οι οποίες με τη σειρά τους διακρίνονται σε διατάξιμες και ονοματικές (βλέπε Ζωγράφος, 2003, σελ. 7-15).

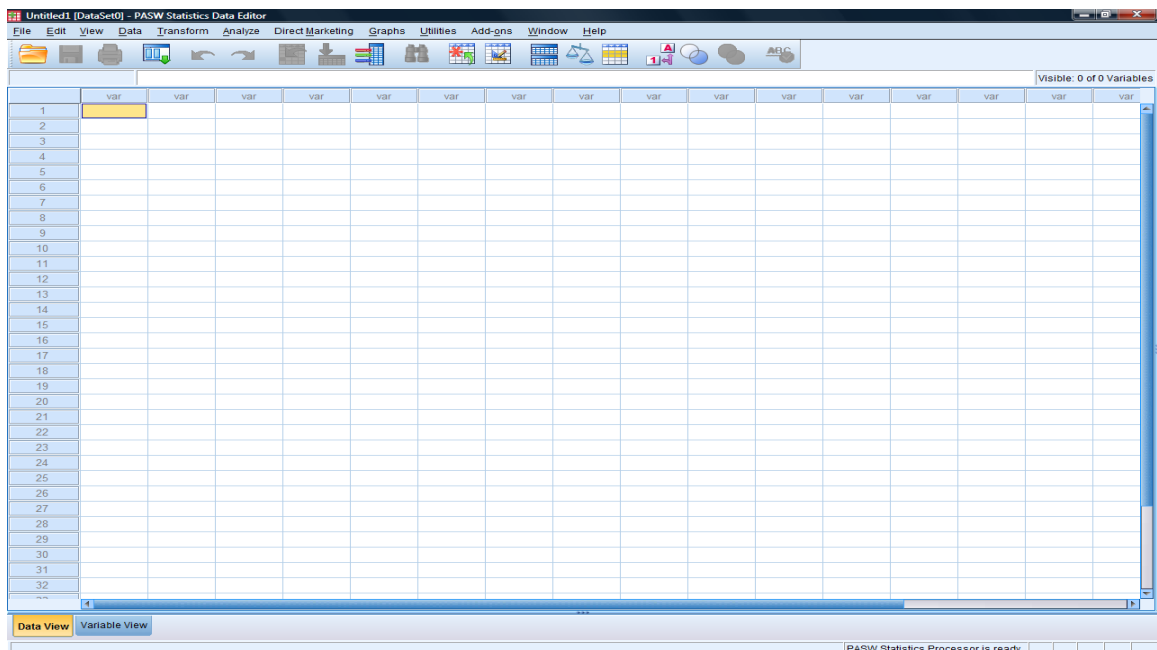


Ποσοτική μεταβλητή είναι εκείνη που μπορεί να μετρηθεί (έχει δηλ. αριθμητικές τιμές). Παραδείγματα ποσοτικών μεταβλητών είναι η ηλικία, το βάρος, το ύψος, η αξία μίας μετοχής, ο δείκτης νοημοσύνης κ.ά.

Ποιοτική μεταβλητή είναι εκείνη που περιγράφει χαρακτηριστικά του πληθυσμού που μεταβάλλονται κατά ποιότητα ή είδος, αλλά όχι κατά μέγεθος. Τέτοιες μεταβλητές είναι το φύλο, η διαγωγή ενός μαθητή, το χρώμα των ματιών-μαλλιών, η στάση υπέρ ή κατά ενός νομοσχεδίου κ.ο.κ. Από αυτές εκείνες που παρέχουν τη δυνατότητα διάταξης ονομάζονται διατάξιμες (π.χ. η διαγωγή ενός μαθητή).

### 1.1.1 Εισαγωγή δεδομένων από το πληκτρολόγιο

Το S.P.S.S. ή όπως αλλιώς έχει μετονομαστεί σε PASW Statistics, έχει ενσωματωμένο έναν ιδιαίτερα εύχρηστο τρόπο εισαγωγής δεδομένων. Όταν ενεργοποιούμε το λογισμικό εμφανίζεται το παρακάτω παράθυρο διαλόγου του Data Editor.



Το πρώτο βήμα θα γίνει με την ενεργοποίηση του Data Editor. Η ενεργοποίηση αυτή επιτυγχάνεται επιλέγοντας **File→New→Data**, ενώ στη συνέχεια θα πρέπει να ληφθεί υπόψη ότι οι οριζόντιες γραμμές αντιστοιχούν στις  $n$  περιπτώσεις-πειραματικές μονάδες, ενώ οι κατακόρυφες στήλες στις  $p$  υπό μελέτη μεταβλητές.

Στη συνέχεια, μέσω ενός παραδείγματος θα περιγραφεί πως εισάγουμε ποσοτικά και πως ποιοτικά δεδομένα στο S.P.S.S..

**Παράδειγμα 1.1 (Ζωγράφος, 2003, σελ. 13-15)**

Έστω ότι επιλέγεται ένα τυχαίο δείγμα 35 παιδιών προσχολικής ηλικίας. Για κάθε παιδί εξετάζεται ο δείκτης νοημοσύνης του, το ύψος του, ο χρόνος σε δευτερόλεπτα που διανύει τα 100 μέτρα, η συμπεριφορά του και η οικονομική κατάσταση της οικογένειάς του. Τα δεδομένα παρουσιάζονται στον πίνακα που ακολουθεί όπου στη στήλη Φύλο Α= Αγόρι, Θ= Κορίτσι, στη στήλη Διαγωγή Α= Κοσμιωτάτη και Β= Κοσμία, στη στήλη Οικονομική Κατάσταση Α=0-450, Β= 450-600, Γ=600-900 και Δ= 900 ευρώ και άνω.

Φύλο	Διαγωγή	Οικον. Κατάσταση	IQ	Ύψος	Χρόνος
A	B	B	111	95	22
Θ	A	Γ	90	98	25
Θ	A	Γ	90	92	18
Θ	A	Γ	90	104	19
A	A	A	104	85	21
A	A	B	72	96	20
Θ	B	B	105	89	21
A	A	Δ	93	103	22
A	A	Γ	99	110	18
A	A	B	93	85	27
A	A	B	84	94	30
Θ	A	A	95	98	21
Θ	A	Γ	93	96	24
A	A	Δ	78	99	26
Θ	A	B	108	83	19
Θ	B	B	100	87	27
A	A	A	81	85	25
A	A	Γ	77	97	24
A	A	Γ	67	96	23
A	A	A	100	107	28
A	A	B	104	102	29
Θ	A	Δ	111	106	19
Θ	A	Δ	122	95	20
A	A	B	99	82	28
Θ	A	B	108	94	31
Θ	A	A	126	90	19
A	B	A	90	90	23
A	A	Γ	110	96	32
A	A	Γ	117	87	27
Θ	A	A	119	97	24
Θ	A	Δ	105	90	22
A	B	B	100	107	18
Θ	A	A	75	92	25
A	A	Γ	96	98	30
Θ	A	Δ	81	95	23

Αρχικά προσδιορίζονται και διακρίνονται σε ποσοτικές και ποιοτικές οι μεταβλητές που χρησιμοποιούνται. Έπειτα καταχωρούνται τα δεδομένα του πίνακα στο S.P.S.S..

Οι μεταβλητές που χρησιμοποιούνται στο παράδειγμα αυτό αναφέρονται στο φύλο, τη διαγωγή, την οικονομική κατάσταση της οικογένειας (ποιοτικές μεταβλητές), το δείκτη νοημοσύνης και το ύψος ενός παιδιού, καθώς και το χρόνο σε δευτερόλεπτα που κάθε παιδί διανύει απόσταση 100 μέτρων (ποσοτικές μεταβλητές).

### **1. Εισαγωγή δεδομένων ποσοτικής μεταβλητής**

Όταν τα δεδομένα είναι ποσοτικά, η εισαγωγή τους είναι πολύ απλή και εύκολη. Για να ξεκινήσουμε την εισαγωγή των δεδομένων της ποσοτικής μεταβλητής επιλέγουμε ένα από τα υπάρχοντα κελιά (εμφανίζεται ένα μαύρο πλαίσιο στο επιλεγθέν κελί), πληκτρολογούμε την αντίστοιχη τιμή και μετά πατάμε Enter. Κατά αυτόν τον τρόπο η αριθμητική τιμή εισέρχεται στην προκαθορισμένη θέση. Επαναλαμβάνουμε τη διαδικασία αυτήν και για τα  $n$  κελιά που δημιουργούνται από τις  $n$  γραμμές και την στήλη. Με τον τρόπο αυτό θα έχουμε εισάγει τα δεδομένα μιας ποσοτικής μεταβλητής. Η ίδια διαδικασία υλοποιείται και για τις υπόλοιπες στήλες-ποσοτικές μεταβλητές. Επομένως, κάθε γραμμή των δεδομένων αντιστοιχεί σε ένα εκ των 35 υποκειμένων και η κάθε στήλη σε μία εκ των μεταβλητών (π.χ. σε μία ερώτηση ή σε ένα υποερώτημα).

### **2. Εισαγωγή δεδομένων ποιοτικής μεταβλητής**

Η εισαγωγή δεδομένων ποιοτικής μεταβλητής διαχωρίζεται στην καταγραφή τους σε μορφή είτε αριθμητικών δεδομένων (που είναι επικρατέστερο να γίνεται) είτε χαρακτήρων. Στη συνέχεια θα αναφέρουμε τον πρώτο τρόπο και ως παρατήρηση θα δοθεί ο δεύτερος τρόπος στο βήμα 4.

Η εισαγωγή δεδομένων ποιοτικής μεταβλητής σε μορφή αριθμητικών δεδομένων προϋποθέτει μία προεργασία πάνω στο ερωτηματολόγιο. Η εν λόγω προεργασία περιλαμβάνει την αντιστοίχιση κωδικών (αριθμητικών τιμών) σε όλες τις πιθανές κατηγορίες-απαντήσεις κάθε ποιοτικής μεταβλητής. Μέσω της κωδικοποίησης αυτής κάθε απάντηση-τιμή της ποιοτικής μεταβλητής αντιστοιχεί σε έναν κωδικό. Στη συνέχεια εισάγουμε στα κελιά τους κωδικούς αυτούς ακολουθώντας την ίδια διαδικασία που

περιγράφηκε και κατά την καταχώρηση των ποσοτικών δεδομένων. Για το λόγο αυτό κάνουμε τη σύμβαση ότι για τη μεταβλητή Φύλο θα κατοχυρώνουμε την τιμή 1 όταν έχουμε Α=Αγόρι και την τιμή 2 όταν είναι Θ=Κορίτσι. Με το ίδιο σκεπτικό, για τη Διαγωγή θα κατοχυρώνουμε την τιμή 1 όταν έχουμε Α και την τιμή 2 όταν είναι Β, ενώ για την Οικονομική Κατάσταση θα εισάγουμε την τιμή 1=Α, 2=Β, 3=Γ και 4=Δ. Προκύπτει το ακόλουθο σύνολο δεδομένων στο S.P.S.S. (δίνεται τμήμα του).

	VAR00001	VAR00002	VAR00003	VAR00004	VAR00005	VAR00006
1	1,00	2,00	2,00	111,00	95,00	22,00
2	2,00	1,00	3,00	90,00	98,00	25,00
3	2,00	1,00	3,00	90,00	92,00	18,00
4	2,00	1,00	3,00	90,00	104,00	19,00
5	1,00	1,00	1,00	104,00	85,00	21,00
6	1,00	1,00	2,00	72,00	96,00	20,00
7	2,00	2,00	2,00	105,00	89,00	21,00
8	1,00	1,00	4,00	93,00	103,00	22,00
9	1,00	1,00	3,00	99,00	110,00	18,00
10	1,00	1,00	2,00	93,00	85,00	27,00
11	1,00	1,00	2,00	84,00	94,00	30,00
12	2,00	1,00	1,00	95,00	98,00	21,00
13	2,00	1,00	3,00	93,00	96,00	24,00
14	1,00	1,00	4,00	78,00	99,00	26,00
15	2,00	1,00	2,00	108,00	83,00	19,00
16	2,00	2,00	2,00	100,00	87,00	27,00
17	1,00	1,00	1,00	81,00	85,00	25,00
18	1,00	1,00	3,00	77,00	97,00	24,00
19	1,00	1,00	3,00	67,00	96,00	23,00
20	1,00	1,00	1,00	100,00	107,00	28,00
21	1,00	1,00	2,00	104,00	102,00	29,00
22	2,00	1,00	4,00	111,00	106,00	19,00
23	2,00	1,00	4,00	122,00	95,00	20,00
24	1,00	1,00	2,00	99,00	82,00	28,00
25	2,00	1,00	2,00	108,00	94,00	31,00
26	2,00	1,00	1,00	126,00	90,00	19,00
27	1,00	2,00	1,00	90,00	90,00	23,00
28	1,00	1,00	3,00	110,00	96,00	32,00
29	1,00	1,00	3,00	117,00	87,00	27,00
30	2,00	1,00	1,00	119,00	97,00	24,00
31	2,00	1,00	4,00	105,00	90,00	22,00
32	1,00	2,00	2,00	100,00	107,00	18,00

### 3. Ονομασία μεταβλητών

Αφού εισάγουμε τα δεδομένα, διαπιστώνουμε ότι το S.P.S.S. εξ ορισμού ονομάζει τις μεταβλητές →VAR00001, VAR00002 κ.ο.κ. Κάτι τέτοιο φυσικά δεν είναι καθόλου εύχρηστο. Θα ορίσουμε σε κάθε μεταβλητή το όνομα που εμείς επιθυμούμε έχοντας ως γνώμονα τους ακόλουθους κανόνες. Το όνομα κάθε μεταβλητής:

- α) μπορεί να έχει το πολύ 64 χαρακτήρες στις συνήθεις γλώσσες (αγγλικά, γαλλικά, ελληνικά)
- β) είναι μοναδικό,

- γ) πρέπει να ξεκινά με γράμμα ή με έναν από τους χαρακτήρες @, #, ή \$,
- δ) δεν πρέπει να περιέχει σημεία στίξης πλην της τελείας, αστεράκια καθώς και κενά,
- ε) δεν μπορεί να περικλείονται οι λέξεις All, Ne, Eq, To, Le, Lt, By, Or, Gt, And, Not, Ge, With,
- στ) μπορεί να γραφεί τόσο με μικρά όσο και με κεφαλαία γράμματα,
- ζ) καλό θα ήταν να μην έχει ως τελευταίο χαρακτήρα την τελεία και την κάτω παύλα, τέλος
- η) δεν επιτρέπεται να ξεκινά με το σύμβολο \$ ονομασία μεταβλητής που ορίζεται από το χρήστη, ενώ είναι επιτρεπτή για παράδειγμα η ακόλουθη ονομασία A.\$@#1.

Η αλλαγή ενός ονόματος (μετονομασία) επιτυγχάνεται επιλέγοντας τη μεταβλητή που θέλουμε να επεξεργαστούμε και κάνοντας διπλό κλικ στο όνομα της.

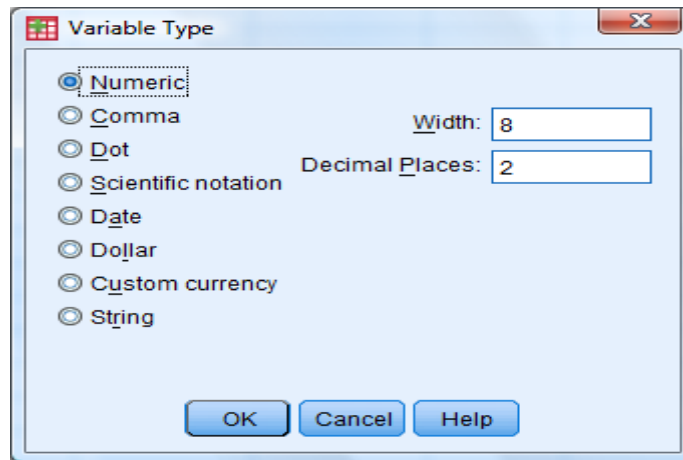
**Σχόλιο:** Εναλλακτικά μπορεί να επιλεγεί το πλαίσιο Variable View (εμφανίζεται στο κάτω αριστερό άκρο της προηγούμενης εικόνας) .

Στο παράθυρο **Variable View** και στο πλαίσιο **Name** εισάγουμε τα κωδικοποιημένα ονόματα των μεταβλητών μας, έστω Sex, Diagogi, Status, Iq, Ipsos, Time. Επιπλέον, στο πεδίο Label δηλώνεται η πλήρης περιγραφή του ονόματος της μεταβλητής που βοηθά στην καλύτερη παρουσίαση των αποτελεσμάτων μας. Κατά αυτό τον τρόπο δηλώνουμε την ονομασία που θα εμφανίζεται στους πίνακες των αποτελεσμάτων των αναλύσεων που θα ακολουθήσουν π.χ. Φύλο, Διαγωγή, Οικον. Κατάσταση, Δείκτης Νοημοσύνης, Ύψος, Χρόνος σε δευτερόλεπτα.

#### **4. Καθορισμός του τύπου της μεταβλητής**

Στο πλαίσιο **Variable Type** το λογισμικό με βάση τις τιμές που πληκτρολογούμε καθορίζει αυτόματα τον τύπο της μεταβλητής, έχοντας ως προεπιλογή να τις εμφανίζει αριθμητικές (numeric) με 2 δεκαδικά ψηφία (Decimals Places) και συνολικό μήκος (δηλώνεται στο πλαίσιο Width) 8 θέσεων. Για τον υπολογισμό του μήκους μίας μεταβλητής λαμβάνονται υπόψη το πρόσημο, το ακέραιο μέρος, η δεκαδική τελεία

καθώς και το δεκαδικό μέρος της. Σε περίπτωση που τα δεδομένα μας είναι τέτοια που παραβιάζονται αυτές οι προεπιλογές πρέπει να τις τροποποιήσουμε κατάλληλα.



Άλλοι δυνατοί τύποι δεδομένων είναι οι ακόλουθοι:

**Comma.** Αριθμητικές τιμές που έχουν το κόμμα «,» ανά τρεις θέσεις και την τελεία «.» ως υποδιαστολή, π.χ. 6,900.38.

**Dot.** Αριθμητικές τιμές που έχουν τελεία «.» ανά τρεις θέσεις και το κόμμα «,» ως υποδιαστολή, π.χ. 6.900,38.

**Scientific notation.** Ποσοτική μεταβλητή της οποίας οι αριθμητικές τιμές γράφονται σε επιστημονική μορφή. Για παράδειγμα στη μορφή 123, 1.23E2, 1.23D2, 1.23E+2, και 1.23+2.

**Date.** Ημερομηνίες.

**Dollar.** Τιμές δολαρίου, με ή χωρίς το σύμβολο (\$), με την τελεία ως υποδιαστολή και το κόμμα ανά τρεις θέσεις, π.χ. 7,355.38 δολάρια.

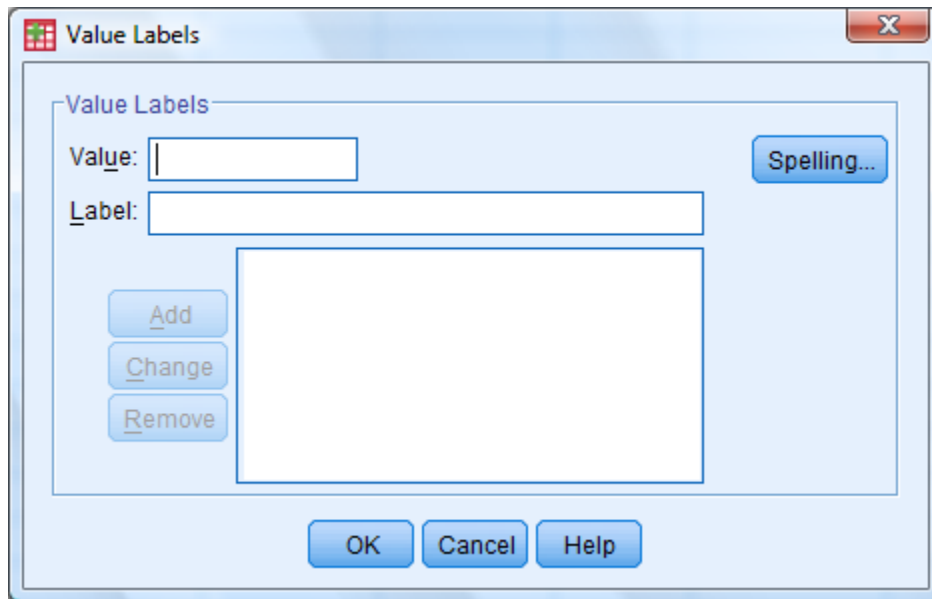
**Custom currency.** Αριθμητικές τιμές των οποίων ο τρόπος εμφάνισης καθορίζεται από το χρήστη.

**String.** Μεταβλητή που δεν είναι αριθμητική και επομένως δε χρησιμοποιείται στους υπολογισμούς.

**Παρατήρηση:** Είναι δυνατή και η εισαγωγή δεδομένων με μη αριθμητικά δεδομένα αρκεί να δημιουργηθεί μία μεταβλητή, η οποία μέσω της καρτέλας του Variable Type του Variable View να οριστεί ως αλφαριθμητική, δηλαδή ως String.

## **5. Ετικέτες τιμών μιας μεταβλητής**

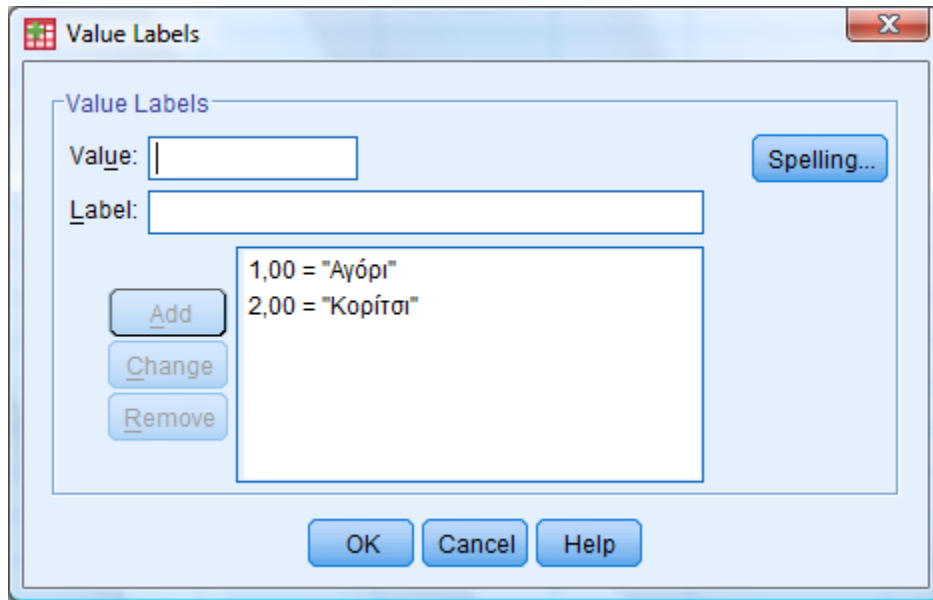
Για να μην ανατρέχουμε συνεχώς στο ερωτηματολόγιο προκειμένου να θυμηθούμε τι σημαίνει ο κάθε κωδικός είναι χρήσιμο όλοι οι κωδικοί των μεταβλητών της μελέτης να καταγράφονται στο πλαίσιο **Values** του παραθύρου Variable View. Επομένως, στο πεδίο αυτό ουσιαστικά εισάγουμε στο λογισμικό τις συμβάσεις τις οποίες κάναμε κατά την καταχώρηση των δεδομένων. Αυτό επιτυγχάνεται κλικάροντας το κάτω δεξί άκρο του κελιού, το οποίο σχηματίζεται από την μεταβλητή και τη στήλη Values. Προκύπτει τότε το παράθυρο διαλόγου:



Εισάγουμε μία-μία τις τιμές της, συνηθέστερα, κατηγορικής μεταβλητής στο πλαίσιο **Value** πληκτρολογώντας την τιμή που αντιπροσωπεύει την κάθε κατηγορία της μεταβλητής, και έπειτα στο πλαίσιο **Value Label** δίνεται η περιγραφή της π.χ. Αγόρι.



Επιλέγουμε το Add και επαναλαμβάνουμε την παραπάνω διαδικασία έως ότου προστεθεί κάθε άλλη δυνατή τιμή και ονομασία της κατηγορικής μεταβλητής. Όταν ολοκληρωθεί η παραπάνω διαδικασία πατάμε το πλήκτρο OK.



Η παραπάνω διαδικασία επαναλαμβάνεται για όλες τις ποιοτικές μεταβλητές του παραδείγματός μας. Αν ακολουθηθεί η παραπάνω διαδικασία θα προκύψει το ακόλουθο παράθυρο.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	sex	Numeric	8	2	Φύλο	{1,00, Αγόρι}..	None	8	Right	Scale
2	diagogi	Numeric	8	2	Διαγωγή	{1,00, Α}...	None	8	Right	Scale
3	status	Numeric	8	2	Οικ.Κατασταση	{1,00, Α}...	None	8	Right	Scale
4	iq	Numeric	8	2	Δείκτης Νοημ	None	None	8	Right	Scale
5	ipsos	Numeric	8	2	Υψος	None	None	8	Right	Scale
6	time	Numeric	8	2	Χρόνος σε δε	None	None	8	Right	Scale
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										

## 6. Επιπλέον δυνατότητες από το παράθυρο Variable View

Στο πλαίσιο **Missing Values** καθορίζουμε τις τιμές των ελλιπών τιμών μίας μεταβλητής. Για το σκοπό αυτό το λογισμικό μας δίνει τις ακόλουθες επιλογές:

α) No missing values (προεπιλογή). Δεν θεωρείται ελλιπής τιμή καμία παρατήρηση εκτός αυτών με τα κενά κελιά.

β) Discrete missing values. Δηλώνονται στα τρία πλαίσια οι 3 διαφορετικές τιμές που όταν καταγράφονται κατά την εισαγωγή των δεδομένων θα σημαίνουν ελλιπή τιμή π.χ. -1, -99, 0. Η επιλογή αυτή είναι ιδιαίτερα χρήσιμη αν για παράδειγμα θέλουμε να γίνει διαχωρισμός μεταξύ των ελλιπών τιμών που εμφανίζονται λόγω του ότι: (i) η

συγκεκριμένη ερώτηση δεν υποβλήθηκε στον ερωτώμενο, (ii) ο ερωτώμενος δεν απάντησε στο συγκεκριμένο ερώτημα και (iii) ο ερωτώμενος είχε αποχωρήσει από την έρευνα ήδη.

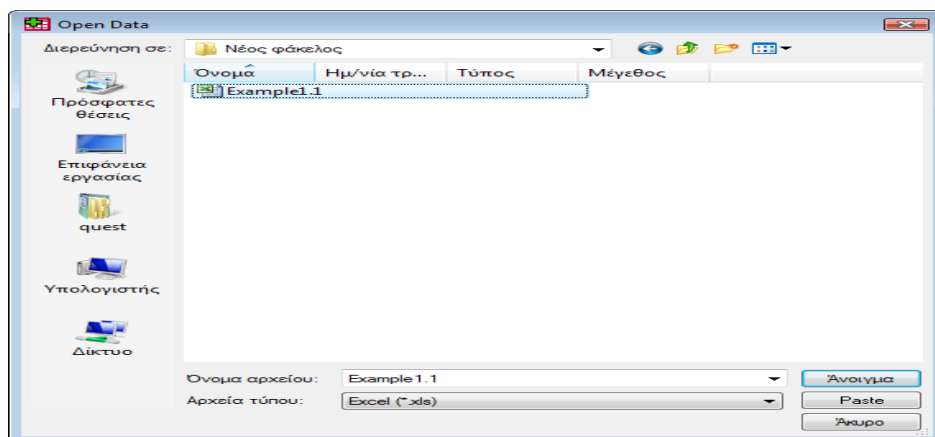
γ) Range plus one optional discrete missing. Δηλώνεται ένα διάστημα, με κάτω και άνω άκρο τις δηλωθείσες τιμές στα πλαίσια Low και High αντίστοιχα, καθώς και μία διακριτή τιμή. Κάθε τιμή που καταγράφεται εντός του διαστήματος καθώς και η διακριτή τιμή λαμβάνεται ως ελλιπής.

Στο πλαίσιο Columns καθορίζουμε το μήκος κάθε στήλης, ενώ στο πλαίσιο Align επιλέγουμε την επιθυμητή στοίχιση των δεδομένων εντός των κελιών (αριστερά, δεξιά, στο κέντρο). Τέλος, από το πλαίσιο Measure καθορίζουμε το είδος της μεταβλητής (Scale=ποσοτική, Ordinal=Διατάξιμη, Nominal=Ονοματική).

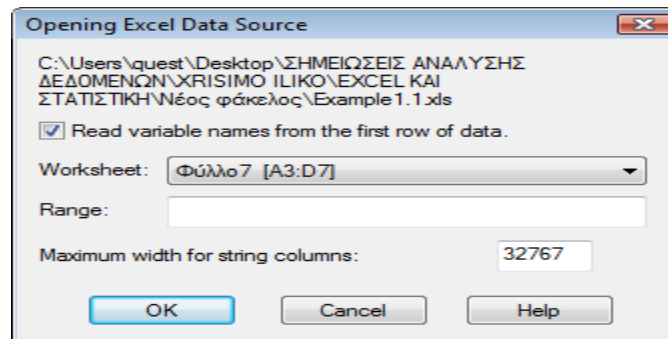
### 1.1.2 Εισαγωγή δεδομένων από αρχείο του Microsoft Excel

Έστω ότι θέλουμε να εισάγουμε στο S.P.S.S τα δεδομένα ενός αρχείου Excel π.χ. του Example1.1.xlsx ( η κατάληξη xls ή xlsx είναι κοινή για αρχεία του Excel, ανάλογα με την έκδοση του). Αυτό επιτυγχάνεται ως εξής:

1. Από το αρχικό παράθυρο διαλόγου του Data View του S.P.S.S. επιλέγουμε File→Open→Data και επιλέγουμε το αρχείο Excel που αναζητούμε. Ζητούμε να ανοίξουμε (Open) αυτό το αρχείο, σύμφωνα και με όσα φαίνονται στην παρακάτω εικόνα



2. Τότε προκύπτει το ακόλουθο παράθυρο διαλόγου που μας επιτρέπει να καθορίζουμε πότε τα ονόματα των μεταβλητών περιέχονται ή όχι στην πρώτη γραμμή του φύλλου εργασίας του Excel (Read variable names from the first row of data). Επιπλέον, καθορίζουμε τα δεδομένα που θέλουμε να μεταφερθούν στο S.P.S.S (δηλώνονται στο Range).



Επιπλέον, είναι επιτρεπτό να καθορίσουμε επιπρόσθετα και ποια φύλλα του αρχείου μας θέλουμε να μετατρέψουμε σε αρχείο δεδομένων του S.P.S.S (δηλώνονται στο Worksheet).

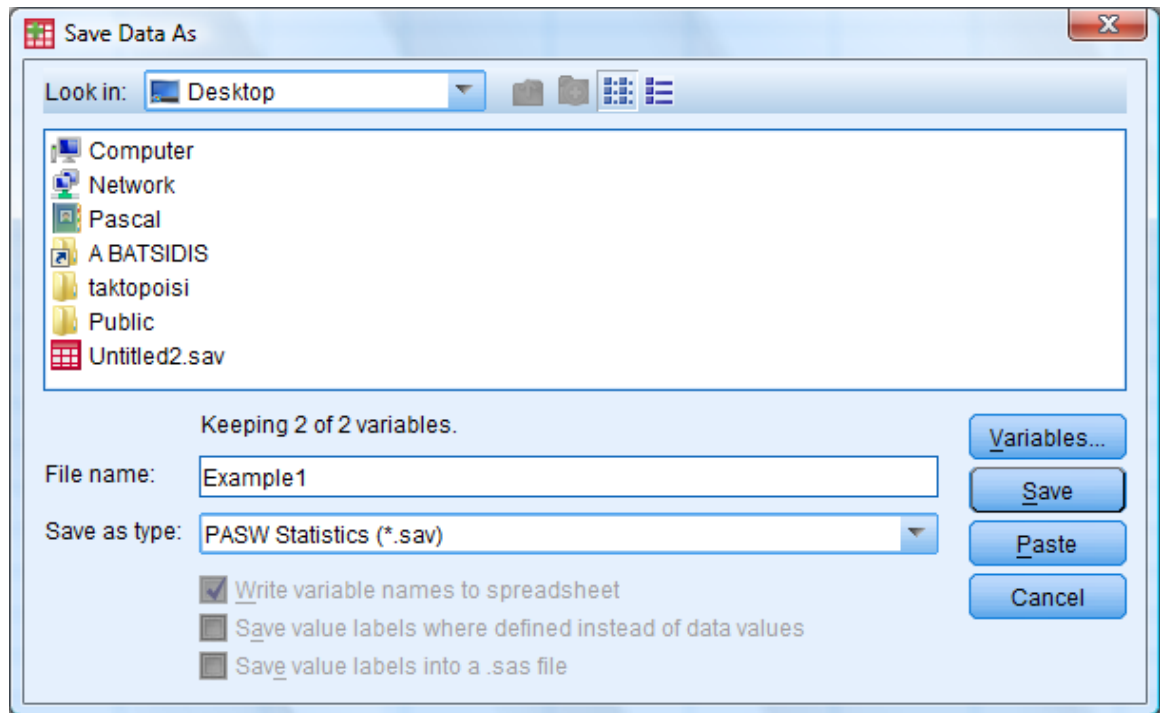
## Παρατήρηση

Αν στο αρχείο Excel τα ονόματα των μεταβλητών δηλώνονται στην πρώτη γραμμή των δεδομένων τότε επιλέγοντας το πλαίσιο Read variable names from first row of data αν οι ονομασίες αυτές δεν συμφωνούν με τους κανόνες ονομασίας των αρχείων του S.P.S.S μετατρέπονται σε τέτοιες έτσι ώστε να ικανοποιούνται και οι αυθεντικές ονομασίες αποθηκεύονται στο πλαίσιο Variables labels.

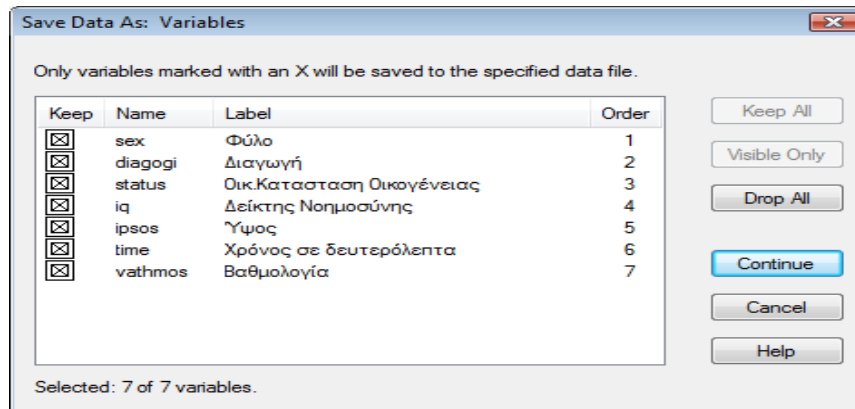
## 1.2 Αποθήκευση Δεδομένων

Η αποθήκευση των δεδομένων επιτυγχάνεται, όπως σε όλα τα προγράμματα, με τις επιλογές **File→Save** για υπάρχον αρχείο δεδομένων και **File→Save as**, όταν

πρόκειται για ένα νέο αρχείο δεδομένων. Στη δεύτερη περίπτωση δηλώνεται και το όνομα του νέου αρχείου καθώς και σε ποιο φάκελο δεδομένων του υπολογιστή θα αποθηκευτεί. Έτσι, το αρχείο του παραδείγματος αποθηκεύτηκε με το όνομα example\_1.sav (sav είναι η κατάληξη των αρχείων δεδομένων του S.P.S.S.) στο φάκελο Data.



Το λογισμικό του S.P.S.S. δίνει τη δυνατότητα τόσο εξ ολοκλήρου αποθήκευσης ενός συνόλου δεδομένων (μέσω της επιλογής Αποθήκευση ή Save) όσο και ενός τμήματος αυτού (μέσω της επιλογής Variables). Ειδικότερα, κλικάροντας στο πλαίσιο Variables προκύπτει το παρακάτω παράθυρο διαλόγου όπου από τη στήλη Keep επιλέγονται οι μεταβλητές τα δεδομένα των οποίων επιθυμούμε να αποθηκευτούν για τη συνέχιση της στατιστικής ανάλυσης



### 1.3 Μετασχηματισμός και επανακωδικοποίηση δεδομένων

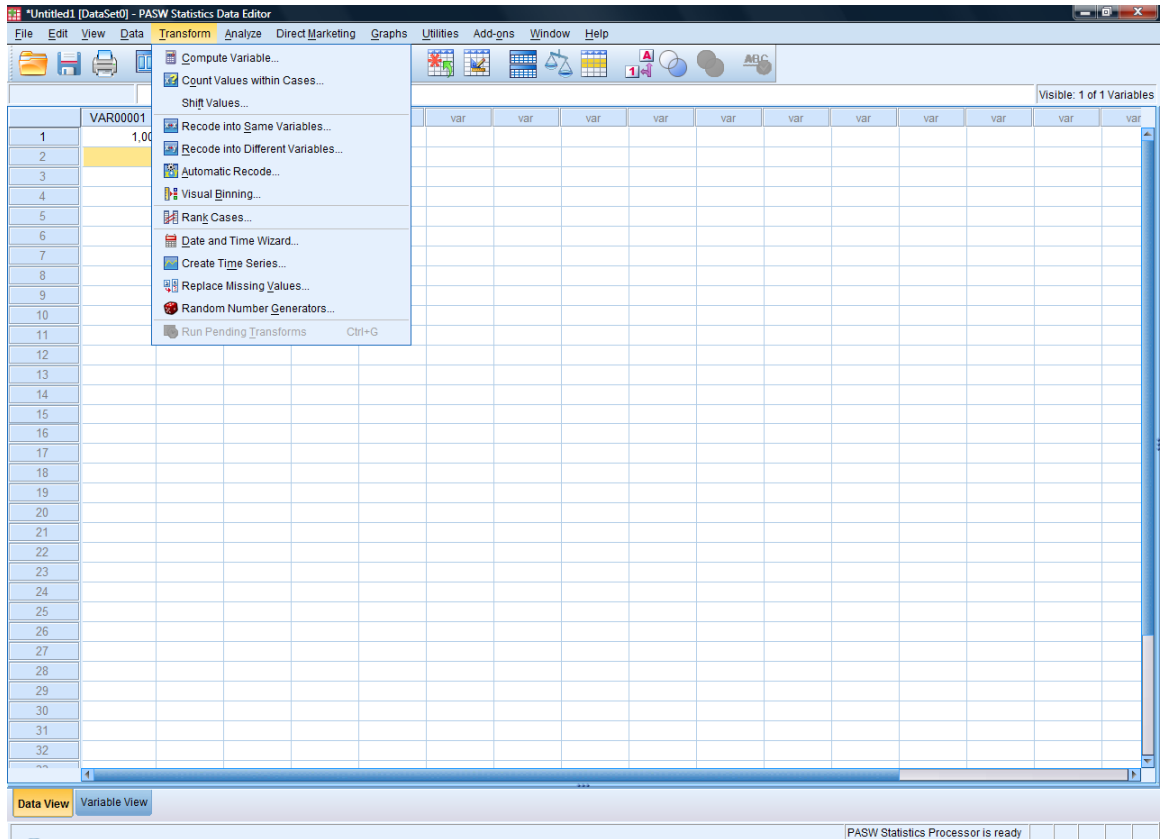
Μερικές φορές σε μία ανάλυση, κρίνεται απαραίτητος ο μετασχηματισμός των δεδομένων π.χ. για την επίτευξη της κανονικότητας όπως θα δούμε σε επόμενο εδάφιο, ή η επανακωδικοποίηση των δεδομένων π.χ. για τη συγχώνευση γειτονικών κελιών στους πίνακες συνάφειας. Ας δούμε περιληπτικά πως υλοποιούνται στο S.P.S.S. μέσω απλών παραδειγμάτων

#### Μετασχηματισμός δεδομένων

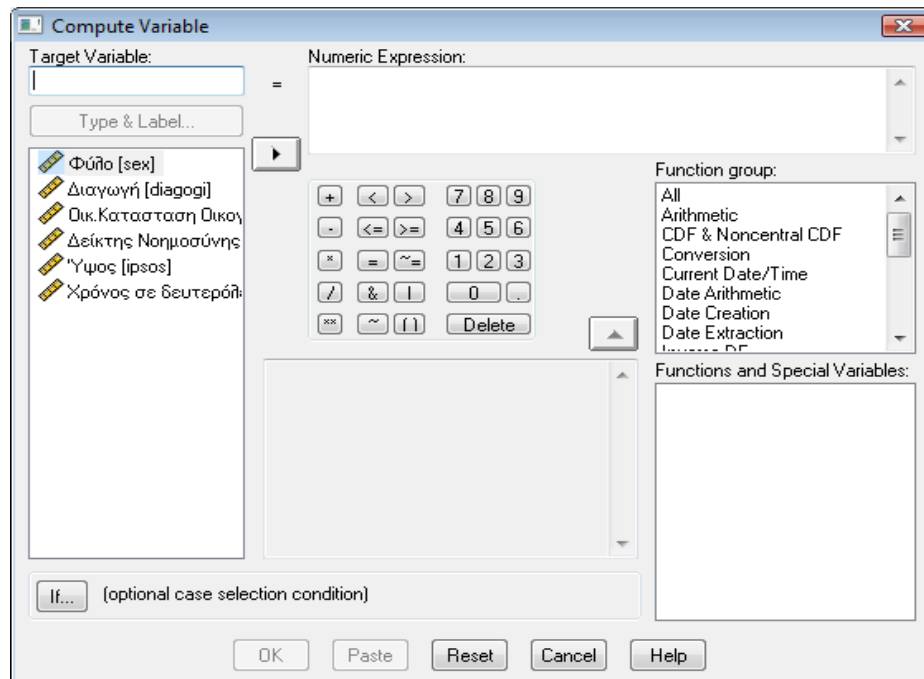
Για κάθε παιδί του Παραδείγματος 1.1. καταγράφεται ο χρόνος σε δευτερόλεπτα που διανύει τα 100 μέτρα καθώς και το ύψος του σε εκατοστά. Ζητείται να δημιουργηθεί μία νέα μεταβλητή που θα μετρά τον προαναφερθέντα χρόνο με μονάδα μέτρησης το λεπτό.

Έχοντας αποφασίσει τον επιθυμητό μετασχηματισμό ακολουθούμε τα ακόλουθα βήματα:

1. Επιλέγουμε από το βασικό μενού **Transform→Compute**



2. Στο παράθυρο **Compute Variable** στο πλαίσιο **Target Variable** δηλώνουμε το όνομα της νέας μεταβλητής, έστω TimeMinutes, ενώ έπειτα από το πλαίσιο **Type & Label** δίνεται η δυνατότητα να ορίσουμε πιο λεπτομερή περιγραφή της.



Στο πλαίσιο Numeric Expression σχηματίζουμε τον κατάλληλο μετασχηματισμό κάνοντας χρήση του Calculator Pad (αν χρειαστεί να χρησιμοποιήσουμε δεκαδικούς τότε κάνουμε χρήση της τελείας) και των συναρτήσεων που δίνονται στο πλαίσιο **Function Group**. Στο πλαίσιο αυτό μεταξύ άλλων έχουμε την ακόλουθη ομαδοποίηση των συναρτήσεων:

α) All: δίνονται όλες οι συναρτήσεις σε αλφαβητική σειρά διάταξης.

β) Arithmetic: δίνονται αριθμητικές συναρτήσεις όπως η απόλυτη τιμή (Abs), το συνημίτονο (Cos), το ημίτονο (Sin), ο δεκαδικός λογάριθμος (Lg10), ο φυσικός λογάριθμος (Ln), η τετραγωνική ρίζα (Sqrt) κ.ά.

γ) CDF and Noncentral CDF: δίνονται οι τιμές των αθροιστικών συναρτήσεων κατανομών (cdf=cumulative distribution function) και των μη κεντρικών αθροιστικών συναρτήσεων κατανομών ειδικών, γνωστών κατανομών όπως η διωνυμική, η εκθετική, η κανονική, η μη κεντρική t κατανομή κ.ά.

δ) Inverse DF: δίνει την τιμή της κατανομής για την οποία η αθροιστική συνάρτηση κατανομής είναι ίση με προκαθορισμένη πιθανότητα.

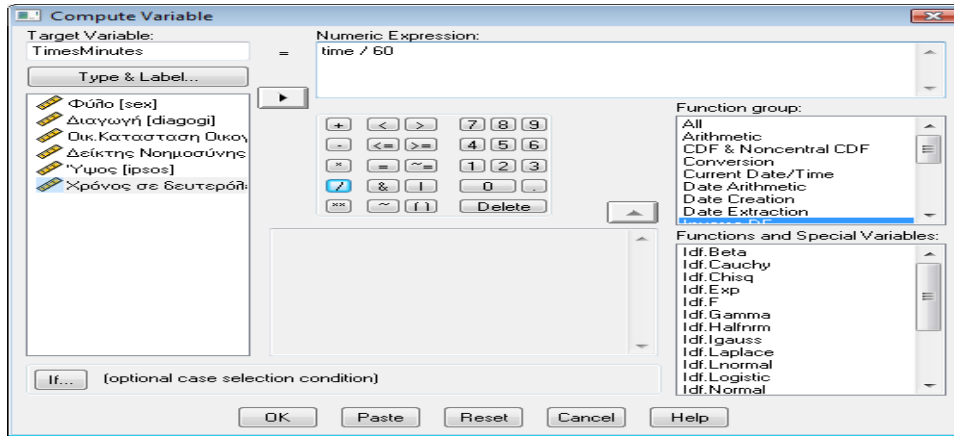
ε) PDF and NonCentral PDF: μας δίνει την τιμή της συνάρτησης πυκνότητας πιθανότητας ή της συνάρτησης πιθανότητας για γνωστές τιμές των παραμέτρων της κατανομής σε προκαθορισμένη τιμή.

στ) Random Numbers: δημιουργεί μία στήλη δεδομένων που αποτελούν ένα τυχαίο δείγμα από διάφορους πληθυσμούς π.χ. από έναν εκθετικό ή κανονικό πληθυσμό.

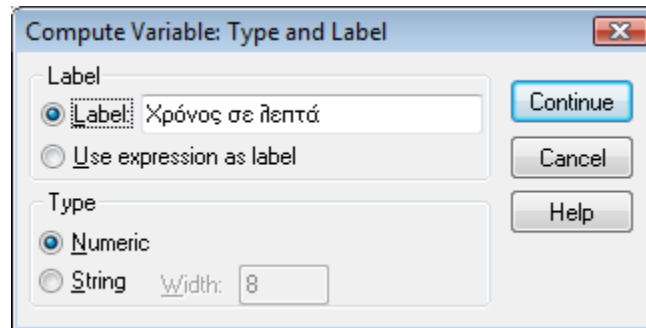
ζ) Statistical: δίνονται στατιστικές συναρτήσεις όπως είναι η μέση τιμή, η τυπική απόκλιση, η διακύμανση, ο συντελεστής μεταβλητότητας κ.ά.

Για το παράδειγμά μας έχουμε:



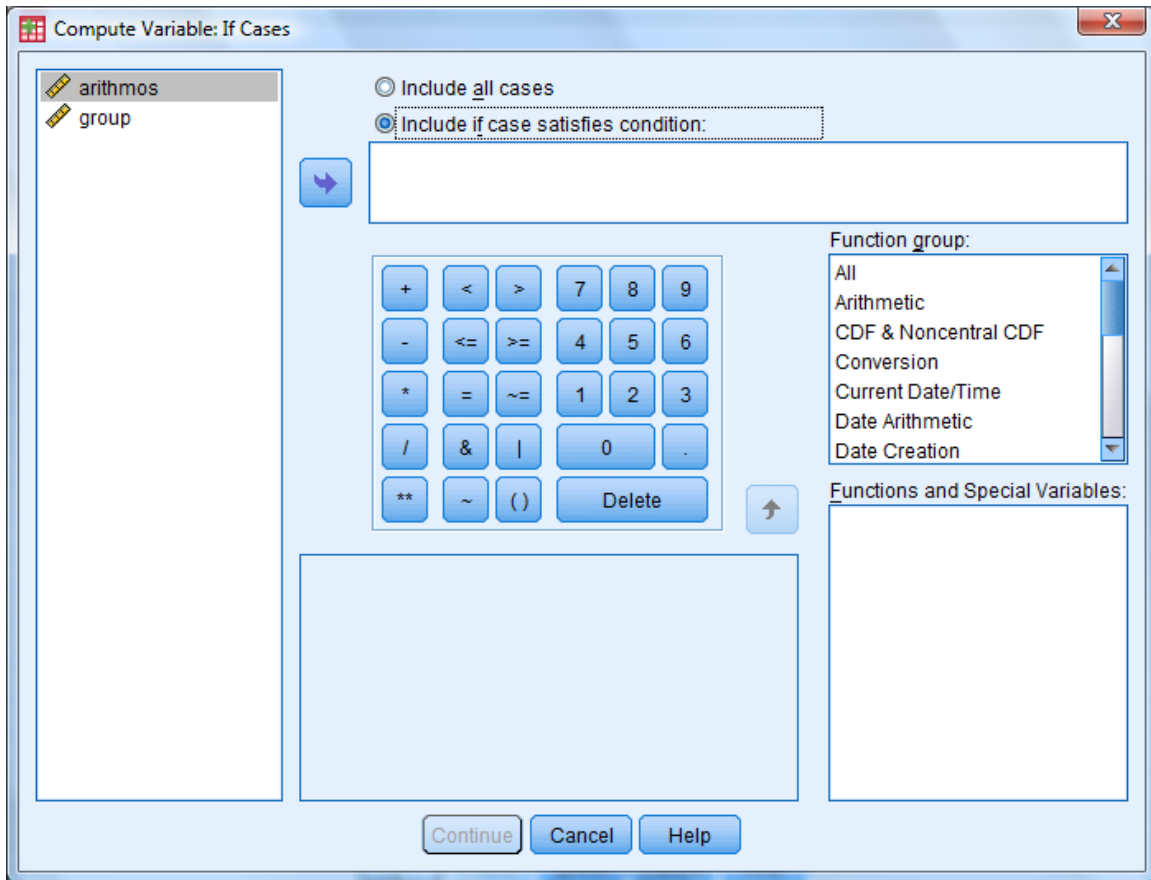


Από το πλαίσιο Type & Label οδηγούμαστε σε ένα παράθυρο διαλόγου όπου μπορούμε να δηλώσουμε το όνομα και τον τύπο της μεταβλητής που δημιουργήσαμε. Ειδικότερα, από το πεδίο Label είτε δίνουμε τη νέα ονομασία είτε χρησιμοποιούμε τη μαθηματική έκφραση ως ονομασία (Use expression as label), ενώ στο πεδίο Type δηλώνουμε αν η νέα μεταβλητή είναι αριθμητική ή αλφαριθμητική (string).



Τέλος, προχωρούμε στην ενεργοποίηση της επιλογής **If...(optional case selection condition)** αν η ύπαρξη τιμών της νέας μεταβλητής εξαρτάται από την ικανοποίηση ή όχι μίας συνθήκης ή έκφρασης μίας άλλης μεταβλητής.

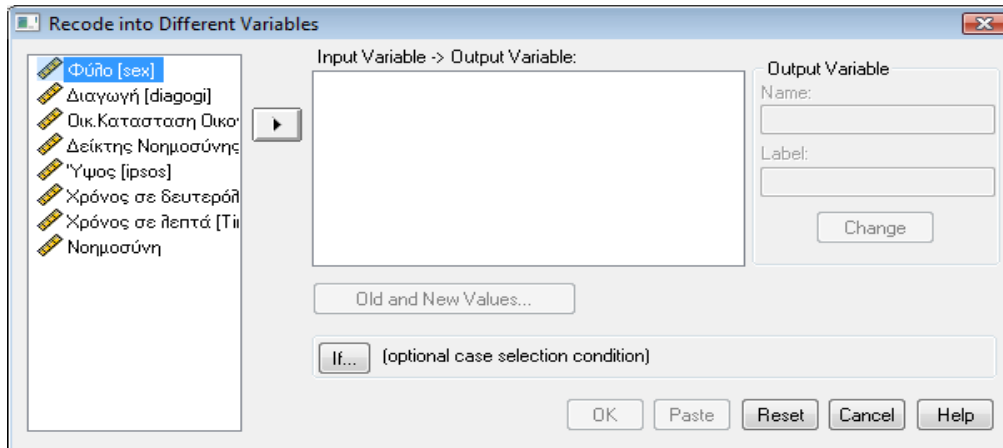
**Σχόλιο:** Εφόσον θέλουμε να χρησιμοποιηθούν οι πειραματικές μονάδες που ικανοποιούν κάποια συνθήκη επιλέγουμε το πλαίσιο Include if case satisfies condition και στο πλαίσιο που ακολουθεί δηλώνεται, σχηματίζεται η επιθυμητή συνθήκη όπως φαίνεται στην παρακάτω εικόνα. Αφού δηλωθεί η αναγκαία συνθήκη πατάμε **Continue**.



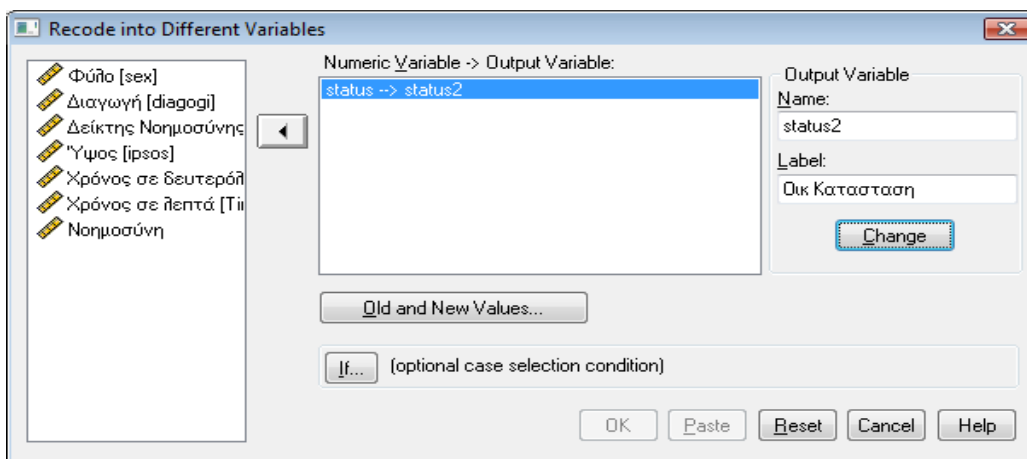
Πατώντας το Continue επανερχόμαστε στο προηγούμενο παράθυρο διαλόγου και εφόσον η νέα μεταβλητή έχει δημιουργηθεί πατάμε το πλαίσιο OK.

### **Επανακωδικοποίηση Μεταβλητών**

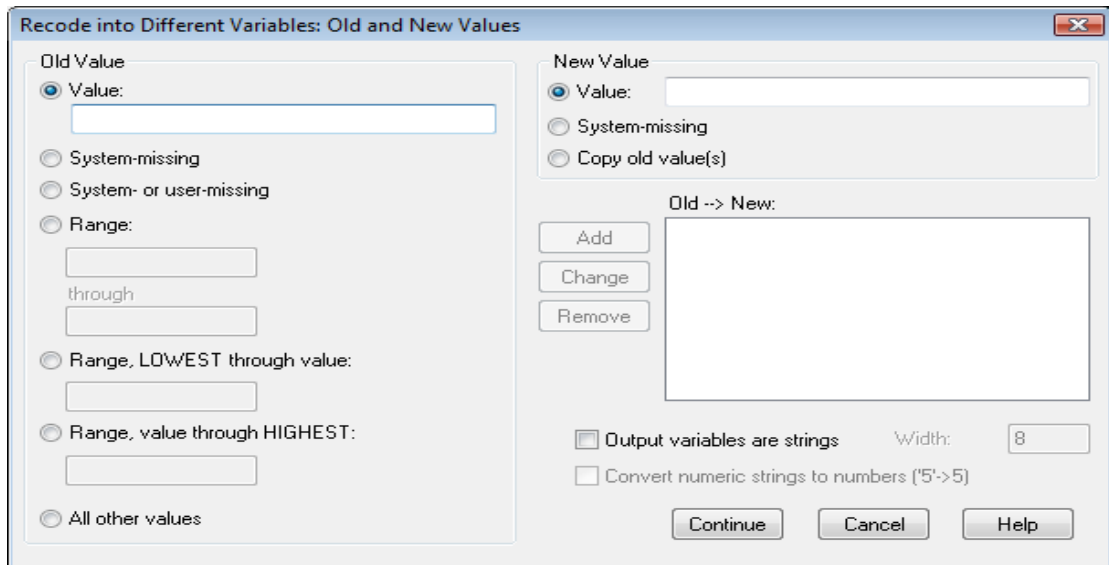
Πολλές φορές εκτός από το μετασχηματισμό των δεδομένων μίας μεταβλητής κρίνεται σκόπιμη η επανακωδικοποίηση της. Αυτό είναι προτιμότερο να γίνεται με τη διαδικασία Recode→Into Different Variables. Τότε ενεργοποιείται ένα νέο παράθυρο διαλόγου, στο οποίο τοποθετούμε στο πλαίσιο Input Variable→Output Variable τη μεταβλητή προς επανακωδικοποίηση καθώς και το όνομα της νέας το οποίο δηλώνεται στο πλαίσιο Output Variable. Μία πλήρης περιγραφή της νέας μεταβλητής δηλώνεται στο πλαίσιο Output Variable Label. Η αλλαγή επιτυγχάνεται πατώντας το πλαίσιο Change και δηλώνοντας την επιθυμητή επανακωδικοποίηση στο παράθυρο διαλόγου που προκύπτει.



Ας δούμε την εν λόγω διαδικασία με χρήση του Παραδείγματος 1.1. Έστω ότι θέλουμε μία νέα κωδικοποίηση για τη μεταβλητή Οικονομική Κατάσταση, σύμφωνα με την οποία όσοι ανήκουν στις κατηγορίες 1 και 2 θα ανήκουν σε μία νέα κατηγορία και όσοι στις 3 και 4 σε μία άλλη. Δηλαδή, όσων παιδιών οι οικογένειες έχουν εισόδημα από 0-600 Ευρώ αποτελούν μία κατηγορία, ενώ τα υπόλοιπα μία άλλη. Ονομάζουμε τη νέα μεταβλητή status2 με ετικέτα Οικ. Κατάσταση.



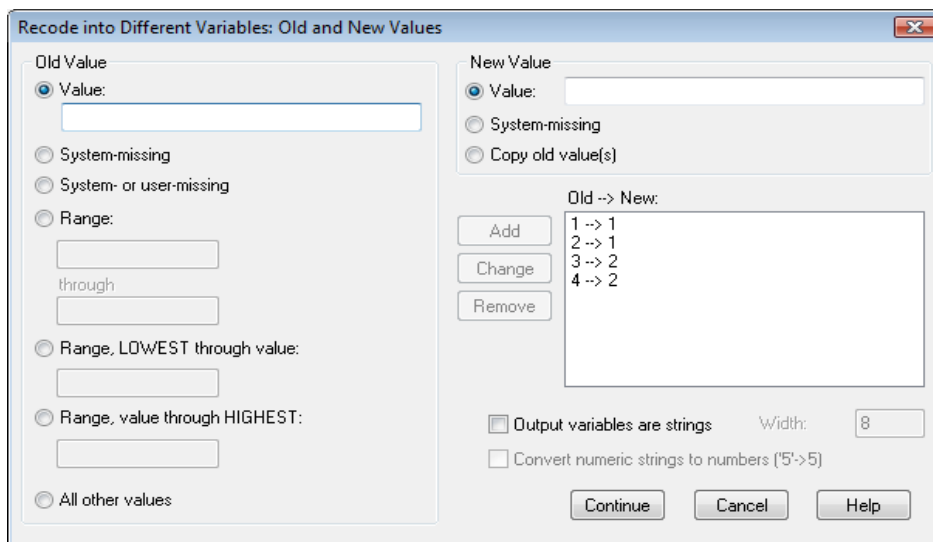
Από την επιλογή Old and New Values οδηγούμαστε σε ένα παράθυρο διαλόγου που μας επιτρέπει την επανακωδικοποίηση μίας ποιοτικής μεταβλητής αλλά και την κωδικοποίηση μίας ποσοτικής σε ποιοτική.



Θα πρέπει να έχουμε όμως υπόψη δύο πολύ βασικούς κανόνες:

- Δεν θα πρέπει να υπάρχουν τομές στις κατηγορίες π.χ. τα διαστήματα [25,50], [50,75] δεν μπορούν να αποτελούν δύο νέες κατηγορίες μίας υπάρχουσας ποσοτικής μεταβλητής.
- Θα πρέπει να υπάρχουν κατάλληλες κατηγορίες για κάθε τιμή των αρχικών δεδομένων π.χ. οι κατηγορίες 25-49, 50-75 δε πρέπει να χρησιμοποιηθούν αν στα αρχικά δεδομένα υπάρχουν τιμές μεταξύ 49 και 50.

Έτσι για το παράδειγμά μας, θα πρέπει να δηλώσουμε ότι οι παλιές τιμές 1 και 2 αντιστοιχούν σε μία νέα κατηγορία με τιμή έστω 1, ενώ οι παλιές 3 και 4 σε μία νέα, με τιμή έστω 2. Έτσι θα πρέπει να προκύψει το ακόλουθο παράθυρο διαλόγου



Πληκτρολογούμε Continue και OK.

**Σχόλιο:** Στο αριστερό μέρος της παραπάνω εικόνας (πλαίσιο Old Value) παρατηρούμε ότι δίνονται οι ακόλουθες επιλογές:

α) Value: πληκτρολογείτε μία παλιά τιμή και αντιστοιχίζεται είτε σε μία νέα τιμή (δηλώνεται στο πλαίσιο Value του New Value) είτε σε ελλιπή τιμή της νέας μεταβλητής (δηλώνεται στο πλαίσιο System missing του New Value).

β) System missing: δηλώνουμε πως θα επανακωδικοποιηθούν οι ελλιπείς τιμές-κενά κελιά.

γ) System-or user missing: δηλώνουμε πως θα επανακωδικοποιηθούν οι ελλιπείς τιμές τόσο του συστήματος όσο και αυτές που έτσι κατοχυρώθηκαν από το χρήστη.

δ) Range: δηλώνεται πως θα επανακωδικοποιηθεί ένα εύρος, διάστημα τιμών το κάτω και άνω άκρο του οποίου δίνεται στα πλαίσια που ακολουθούν.

ε) Range, lowest through value: δηλώνεται πως θα επανακωδικοποιηθούν οι τιμές από την μικρότερη ως αυτή που δηλώνεται στο πλαίσιο που ακολουθεί.

στ) Range, value through highest: δηλώνεται πως θα επανακωδικοποιηθούν οι τιμές από αυτή που δηλώνεται στο πλαίσιο που ακολουθεί ως τη μεγαλύτερη.

ζ) All other values: δηλώνεται πως θα επανακωδικοποιηθούν όλες οι υπόλοιπες τιμές.

