Κεφάλαιο 4

Μετασχηματισμοί δεδομένων

Το SPSS διαθέτει ισχυρές εντολές για το μετασχηματισμό των δεδομένων ενός αρχείου, καθώς και για τη δημιουργία νέων μεταβλητών. Η ανάγκη για μετασχηματισμούς προκύπτει συχνά στη στατιστική ανάλυση, εφόσον είναι σπάνια η περίπτωση, οι μεταβλητές ενός αρχείου να είναι απόλυτα συμβατές με οποιουδήποτε τύπου επεξεργασία. Για παράδειγμα, πολύ συχνά χρειάζεται να συμπτυχθούν οι κατηγορίες μιας κατηγορικής μεταβλητής ή είναι αναγκαίο να μετασχηματιστούν οι τιμές μιας άλλης ή ακόμη να παραχθούν νέες μεταβλητές συνδυάζοντας τις τιμές άλλων ήδη υπαρχόντων μεταβλητών. Οι διαδικασίες μετασχηματισμού δεδομένων μπορούν να χρησιμοποιηθούν είτε μέσω των μενού είτε ακόμη με τη βοήθεια της γλώσσας εντολών που διαθέτει το SPSS (command syntax). Στη δεύτερη περίπτωση, οι δυνατότητες που παρέχει το σύστημα είναι πολύ ισχυρότερες από αυτές που έχει διά μέσου των μενού, εφόσον υπάρχει μια σειρά από ισχυρές εντολές δομημένου προγραμματισμού, με τη βοήθεια των οποίων μπορούν να μετασχηματίζονται αρχεία, να παράγονται νέες δομές δεδομένων αλλά και να συντάσσονται νέες στατιστικές διαδικασίες (procedures) κατ' επιλογή του χρήστη.

Επανακωδικοποίηση μεταβλητών – Recode

Οι τιμές μιας μεταβλητής μπορούν να επανακωδικοποιηθούν για τις ανάγκες μιας ανάλυσης μέσω της εντολής Recode. Τα αποτελέσματα της επανακωδικοποίησης μπορούν να καταχωρηθούν είτε ως νέες τιμές στην ίδια τη μεταβλητή είτε ως τιμές μιας νέας μεταβλητής που θα ορίσει ο χρήστης.

Επανακωδικοποίηση στην ίδια μεταβλητή – Recode into Same Variables

Η επανακωδικοποίηση στην ίδια τη μεταβλητή αφορά την αντικατάσταση των τιμών της μεταβλητής με άλλες ή την αντικατάσταση ενός συνόλου συνεχών τιμών της μεταβλητής (range) με μια άλλη τιμή. Οι μεταβλητές που μπορούν να επανακωδικοποιηθούν είναι αριθμητικού ή αλφαριθμητικού τύπου. Είναι δυνατό να επανακωδικοποιηθούν πολλές μεταβλητές με την ίδια διαδικασία Recode εφόσον είναι όλες του ιδίου τύπου (δηλαδή ή όλες αριθμητικές ή όλες αλφα-

ριθμητικές). Για την επανακωδικοποίηση στην ίδια τη μεταβλητή εργαζόμαστε ως εξής:

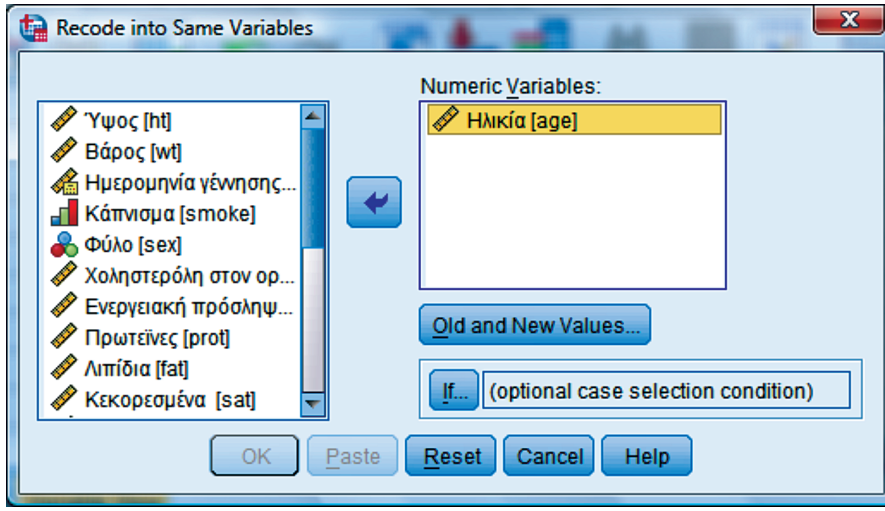
▶ Από τη γραμμή των μενού επιλέγουμε:
Transform
Recode into Same Variables...

Ανοίγει το πλαίσιο διαλόγου Recode into Same Variables (Σχήμα 4.1).

▶ Επιλέγονται οι μεταβλητές που θα επανακωδικοποιηθούν. Αν επιλεγούν πολλές μεταβλητές ταυτόχρονα, πρέπει να είναι όλες του ιδίου τύπου.

▶ Επιλέγουμε Old and New Values, και, στο υποπλαίσιο διαλόγου που ανοίγει, ορίζουμε πώς θα γίνουν οι επανακωδικοποιήσεις (Σχήμα 4.2). Οι επανακωδικοποιήσεις μπορούν να αφορούν ένα υποσύνολο μόνο από τις παρατηρήσεις ενός αρχείου, οπότε πρέπει να πατηθεί το πλήκτρο If.

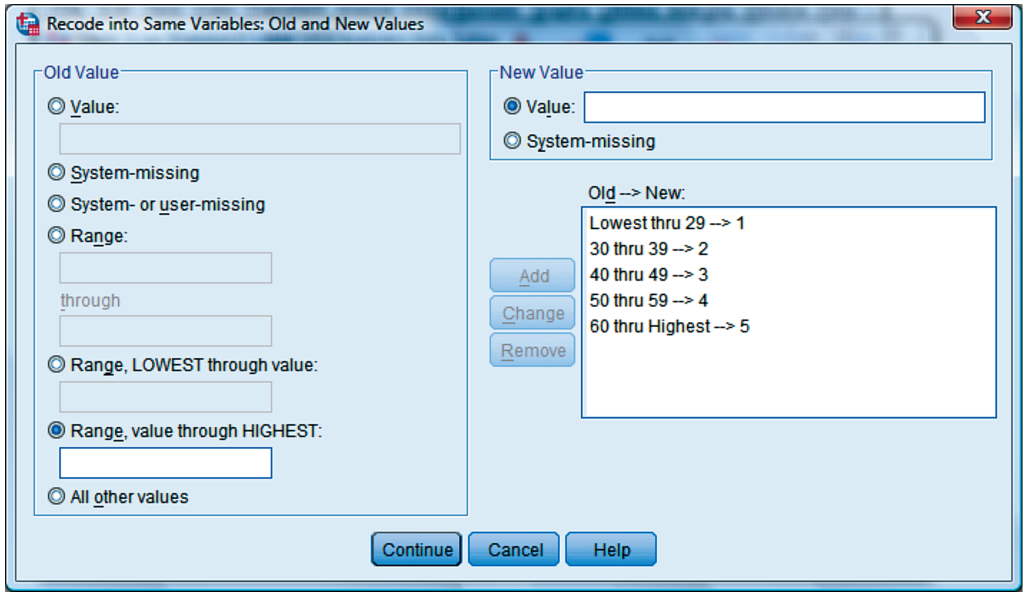
Σχήμα 4.1 Πλαίσιο διαλόγου Recode into Same Variables



Επανακωδικοποίηση στην ίδια μεταβλητή: Old and New Values

Στο υποπλαίσιο διαλόγου Old and New Values ορίζονται οι τιμές που πρόκειται να επανακωδικοποιηθούν. Οι τιμές που θα οριστούν στα πεδία του υποπλαισίου πρέπει να αναφέρονται σε μεταβλητές του ιδίου τύπου.

Σχήμα 4.2 Υποπλαίσιο διαλόγου Old and New Values



Τα πεδία που συμπληρώνονται στο υποπλαίσιο Old and New Values είναι τα εξής:

Old Value. Εδώ ορίζονται οι τιμές που θα επανακωδικοποιηθούν. Οι επανακωδικοποιήσεις που μπορούν να γίνουν, αφορούν απλές τιμές (Value), εύρος τιμών (Range) ή missing values μιας μεταβλητής. Για αλφαριθμητικές μεταβλητές δεν έχει νόημα να ζητηθεί επανακωδικοποίηση εύρους τιμών ή system-missing values (διότι δεν ορίζονται). Στην περίπτωση του εύρους τιμών, το ανώτερο όριο του διαστήματος του εύρους, καθώς και οι user-missing values που βρίσκονται εντός του εύρους, συμπεριλαμβάνονται στις τιμές που θα επανακωδικοποιηθούν.

New Value. Εδώ ορίζεται κάθε φορά η τιμή που θα αντικαταστήσει μια παλιά τιμή ή ένα εύρος παλιών τιμών. Η νέα τιμή που θα αντικαταστήσει τις παλιές μπορεί να είναι και η system-missing value.

Old → New. Στη λίστα αυτή εμφανίζονται οι επανακωδικοποιήσεις που πρόκειται να γίνουν. Κάθε φορά, ορίζοντας μια παλιά και μια νέα τιμή και πατώντας το πλήκτρο Add, η αντίστοιχη επανακωδικοποίηση εμφανίζεται στο εσωτερικό της λίστας. Μπορούμε να αλλάξουμε μια επανακωδικοποίηση ή να την απομακρύνουμε από τη λίστα. Οι επανακωδικοποιήσεις είναι διατεταγμένες εντός της λίστας με βάση τις παλιές τιμές, με την εξής σειρά εμφάνισης: απλές τιμές, εύρος τιμών, missing values, υπόλοιπες τιμές. Σε περίπτωση αλλαγής ή α-

πομάκρυνσης μιας επανακωδικοποίησης, η λίστα επαναδιατάσσεται αυτομάτως σύμφωνα με τη σειρά που αναφέρθηκε.

Στο υποπλαίσιο διαλόγου του Σχήματος 4.2 γίνεται η επανακωδικοποίηση της ηλικίας σε ομάδες ηλικιών. Οι ηλικιακές ομάδες που ορίζονται είναι: ≤ 29, 30-39, 40-49, 50-59 και ≥ 60.

Επανακωδικοποίηση σε διαφορετική μεταβλητή – Recode into Different Variables

Η επανακωδικοποίηση σε διαφορετική μεταβλητή αφορά την αντικατάσταση των τιμών μιας μεταβλητής με άλλες, μόνο που τώρα οι νέες τιμές που δημιουργούνται, καταχωρούνται σε μια νέα μεταβλητή που ορίζει ο χρήστης, ενώ η αρχική μεταβλητή παραμένει αναλλοίωτη. Για αυτού του τύπου τις επανακωδικοποιήσεις ισχύουν οι εξής κανόνες:

- Μπορούν να επανακωδικοποιούνται και αριθμητικές και αλφαριθμητικές μεταβλητές.
- Είναι δυνατό να επανακωδικοποιηθούν πολλές μεταβλητές με την ίδια διαδικασία Recode εφόσον είναι όλες του ιδίου τύπου (δηλαδή ή όλες αριθμητικές ή όλες αλφαριθμητικές).
- Μπορεί μια αριθμητική μεταβλητή να επανακωδικοποιείται σε αλφαριθμητική και αντίστροφα.

Πρακτικά, για την επανακωδικοποίηση μιας μεταβλητής σε άλλη εργαζόμαστε ως εξής:

▶ Από τη γραμμή των μενού επιλέγουμε:
Transform
Recode into Different Variables...

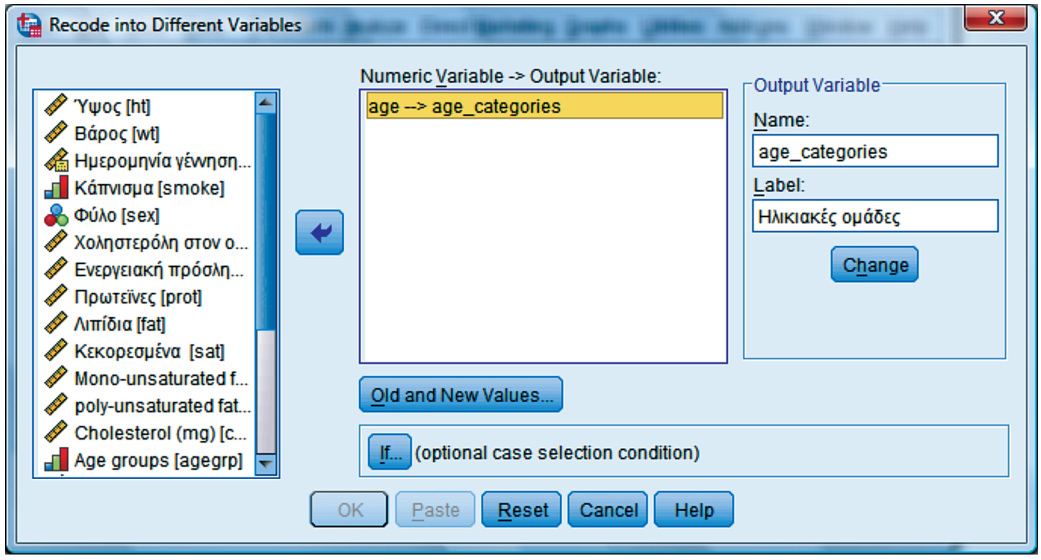
Ανοίγει το πλαίσιο διαλόγου Recode into Different Variables (Σχήμα 4.3).

▶ Επιλέγονται οι μεταβλητές που θα επανακωδικοποιηθούν και εισάγονται στην τελική λίστα (Numeric Variable *ή* String Variable → Output Variable). Αν επιλεγούν πολλές μεταβλητές ταυτόχρονα πρέπει να είναι όλες του ιδίου τύπου.

▶ Για κάθε μεταβλητή που πρόκειται να επανακωδικοποιηθεί, εισάγεται το όνομα και το variable label της νέας μεταβλητής στην οποία θα καταχωρηθούν οι επανακωδικοποιημένες τιμές. Στη συνέχεια, πατάμε το πλήκτρο Change.

▶ Επιλέγουμε Old and New Values, και στο υποπλαίσιο διαλόγου που ανοίγει, ορίζουμε πώς θα γίνουν οι επανακωδικοποιήσεις (Σχήμα 4.4). Οι επανακωδικοποιήσεις μπορούν να αφορούν ένα υποσύνολο μόνο από τις παρατηρήσεις ενός αρχείου, οπότε πρέπει να πατηθεί το πλήκτρο If.

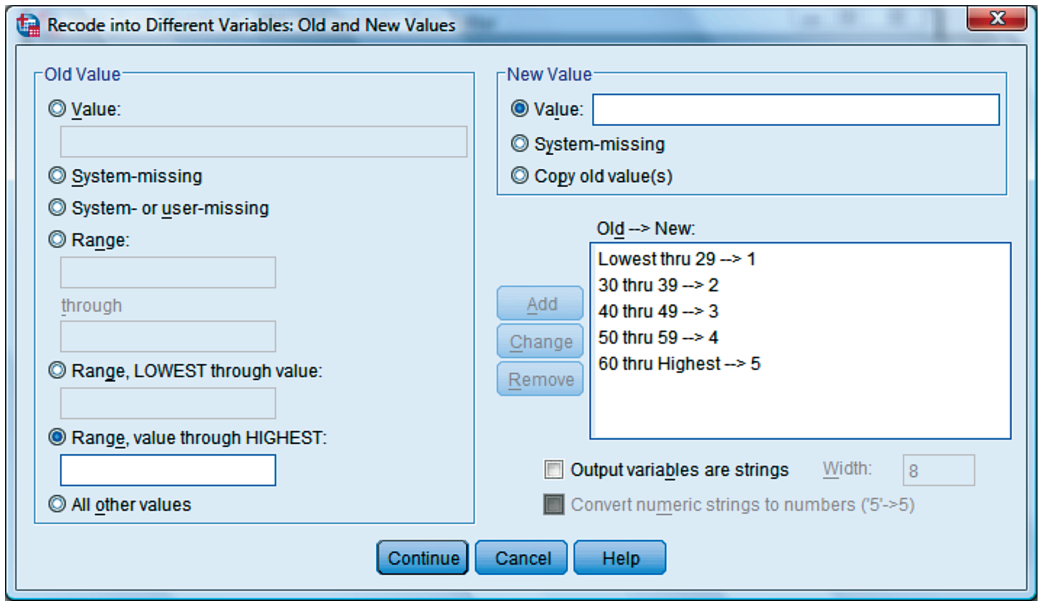
Σχήμα 4.3 Πλαίσιο διαλόγου Recode into Different Variables



Επανακωδικοποίηση σε διαφορετική μεταβλητή: Old and New Values

Στο υποπλαίσιο διαλόγου Old and New Values ορίζονται οι τιμές που πρόκειται να επανακωδικοποιηθούν διαμέσου της εντολής Recode into Different Variables.

Σχήμα 4.4 Υποπλαίσιο διαλόγου Old and New Values



Τα πεδία που συμπληρώνονται στο υποπλαίσιο Old and New Values είναι τα εξής:

Old Value. Εδώ ορίζονται οι τιμές που θα επανακωδικοποιηθούν. Οι επανακωδικοποιήσεις που μπορούν να γίνουν, αφορούν απλές τιμές (Value), εύρος τιμών (Range) ή missing values μιας μεταβλητής. Για αλφαριθμητικές μεταβλητές δεν έχει νόημα να ζητηθεί η επανακωδικοποίηση ενός εύρους τιμών τους ή η επανακωδικοποίηση των system-missing values (διότι δεν ορίζονται). Στην περίπτωση που η επανακωδικοποίηση αφορά εύρος τιμών, το ανώτερο όριο του διαστήματος του εύρους, καθώς και οι user-missing values που βρίσκονται εντός του εύρους, συμπεριλαμβάνονται στις τιμές που θα επανακωδικοποιηθούν.

New Value. Εδώ ορίζεται κάθε φορά η τιμή που θα αντικαταστήσει μια παλιά τιμή ή ένα εύρος παλιών τιμών. Η νέα τιμή που θα αντικαταστήσει τις παλιές μπορεί να είναι και η system-missing value. Ειδικότερα, ισχύουν τα εξής:

- Αν πρόκειται να επανακωδικοποιηθεί μια αριθμητική μεταβλητή σε μια μεταβλητή αλφαριθμητικού τύπου, πρέπει να δηλωθεί επίσης η επιλογή Output variables are strings.

- Οι τιμές της αρχικής (παλιάς) μεταβλητής που δεν έχουν οριστεί στο πεδίο Old value, δεν θα συμπεριληφθούν στις τιμές της νέας μεταβλητής και στη θέση τους θα μπουν system-missing values. Προκειμένου ορισμένες τιμές της αρχικής μεταβλητής να συμπεριληφθούν στη νέα χωρίς επανακωδικοποίηση, δηλώνεται η επιλογή All other values στις παλιές τιμές και η επιλογή Copy old value(s) στις νέες τιμές.

Old → New. Στη λίστα αυτή εμφανίζονται οι επανακωδικοποιήσεις που πρόκειται να γίνουν. Κάθε φορά, ορίζοντας μια παλιά και μια νέα τιμή και πατώντας το πλήκτρο Add, η αντίστοιχη επανακωδικοποίηση εμφανίζεται στο εσωτερικό της λίστας. Μπορούμε να αλλάξουμε μια επανακωδικοποίηση (πατώντας το πλήκτρο Change) ή να την απομακρύνουμε από τη λίστα (πατώντας το πλήκτρο Remove). Οι επανακωδικοποιήσεις είναι διατεταγμένες εντός της λίστας με βάση τις παλιές τιμές, με την εξής σειρά εμφάνισης: απλές τιμές, εύρος τιμών, missing values, υπόλοιπες τιμές. Σε περίπτωση αλλαγής ή απομάκρυνσης μιας επανακωδικοποίησης, η λίστα επαναδιατάσσεται αυτομάτως σύμφωνα με τη σειρά που αναφέρθηκε.

Στο υποπλαίσιο διαλόγου του Σχήματος 4.4 γίνεται η επανακωδικοποίηση της μεταβλητής *age* (Ηλικία) στη μεταβλητή *age_categories* (Ηλικιακές Ομάδες). Οι ηλικιακές ομάδες που ορίζονται είναι: ≤ 29, 30-39, 40-49, 50-59 και ≥ 60.

Πλαίσιο διαλόγου If Cases

Αν οι επανακωδικοποιήσεις αφορούν ένα υποσύνολο μόνο από τις παρατηρήσεις ενός αρχείου, τότε ο ορισμός του υποσυνόλου αυτού γίνεται από το πλήκτρο If. Πατώντας το πλήκτρο If εμφανίζεται το πλαίσιο διαλόγου If Cases (Σχήμα 4.5). Το πλαίσιο αυτό διαλόγου (που στην προκειμένη περίπτωση αφορά την εντολή Recode into Same Variables) είναι ένα γενικό πλαίσιο το οποίο εμφανίζεται κάθε φορά κατά την επιλογή των παρατηρήσεων ενός αρχείου με βάση μια συνθήκη. Για κάθε παρατήρηση του αρχείου, η συνθήκη ελέγχεται ως προς την ικανοποίησή της ή μη.

■ Αν η συνθήκη ικανοποιείται, τότε η παρατήρηση επιλέγεται. Αντίθετα, αν η συνθήκη δεν ικανοποιείται, η παρατήρηση δεν επιλέγεται.

■ Η έκφραση της συνθήκης είναι λογικού τύπου και μπορεί να περιλαμβάνει ονόματα μεταβλητών, αριθμούς, αριθμητικές πράξεις, συναρτήσεις, καθώς και σχεσιακούς ή λογικούς τελεστές.

■ Οι χρησιμοποιούμενοι σχεσιακοί τελεστές είναι οι εξής:

>	*μεγαλύτερο*,	<	*μικρότερο*,
>=	*μεγαλύτερο ή ίσο*,	=<	*μικρότερο ή ίσο*,
=	*ίσο*,	~=	διάφορο.

Οι σχεσιακοί τελεστές μπορούν να επιλέγονται από το πληκτρολόγιο της αριθμομηχανής που υπάρχει εντός του πλαισίου και να επικολλούνται στην έκφραση της συνθήκης. Μπορούν επίσης να πληκτρολογούνται απευθείας από το πληκτρολόγιο του υπολογιστή.

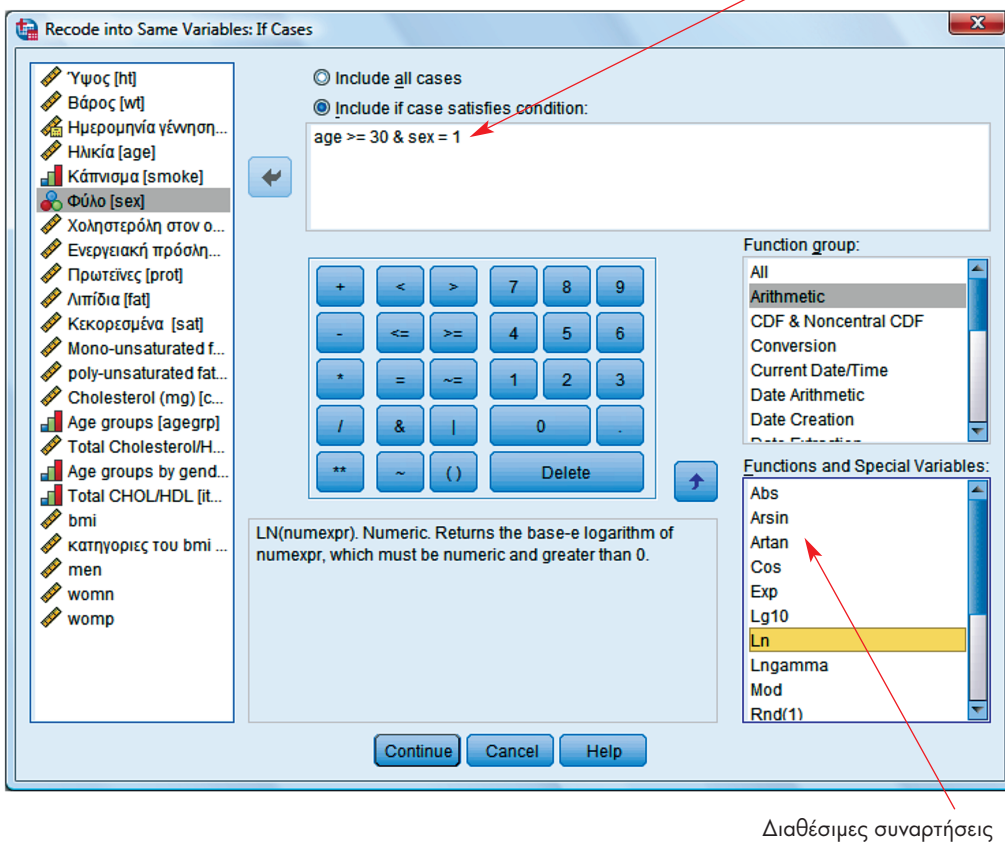
■ Οι λογικοί τελεστές είναι η διάζευξη και η σύζευξη. Η διάζευξη (δηλαδή το OR) εμφανίζεται στο πληκτρολόγιο της αριθμομηχανής με το σύμβολο |, ενώ η σύζευξη (AND) με το σύμβολο &.

■ Οι συναρτήσεις είναι καταχωρημένες κατά ομάδες στη λίστα Function group, ενώ κάθε φορά που επιλέγουμε μια ομάδα, εμφανίζονται στο πεδίο Functions and Special Variables οι συναρτήσεις που αυτή περιλαμβάνει. Κάνοντας κλικ στο όνομα μιας συνάρτησης μπορούμε να πάρουμε πληροφορίες για τη λειτουργία της στο πεδίο που εμφανίζεται στο κέντρο του πλαισίου διαλόγου (στο Σχήμα 4.5 έχει επιλεχθεί να εμφανιστεί η λειτουργία της συνάρτησης Ln). Όταν μια συνάρτηση εισαχθεί στο πεδίο διατύπωσης της συνθήκης, οι παράμετροί της εμφανίζονται με ερωτηματικά και πρέπει να οριστούν από το χρήστη.

■ Η διατύπωση της συνθήκης που πρέπει να ικανοποιείται, καταχωρείται στο πεδίο που εμφανίζεται στο άνω κεντρικό μέρος του πλαισίου διαλόγου. Τα ονόματα των μεταβλητών μπαίνουν στο πεδίο της συνθήκης είτε από τη λίστα

των μεταβλητών που υπάρχει εντός του πλαισίου, είτε με απ' ευθείας πληκτρολόγηση των ονομάτων τους από το πληκτρολόγιο του υπολογιστή (συνιστάται η πρώτη δυνατότητα).

Σχήμα 4.5 Πλαίσιο διαλόγου If Cases



Υπολογισμός τιμών μεταβλητής – Compute Variable

Με την εντολή Compute Variable υπολογίζονται οι τιμές μιας νέας μεταβλητής –ή επαναϋπολογίζονται οι τιμές μιας ήδη υπάρχουσας– με βάση τους μετασχηματισμούς των τιμών άλλων μεταβλητών. Ειδικότερα, ισχύουν τα εξής:

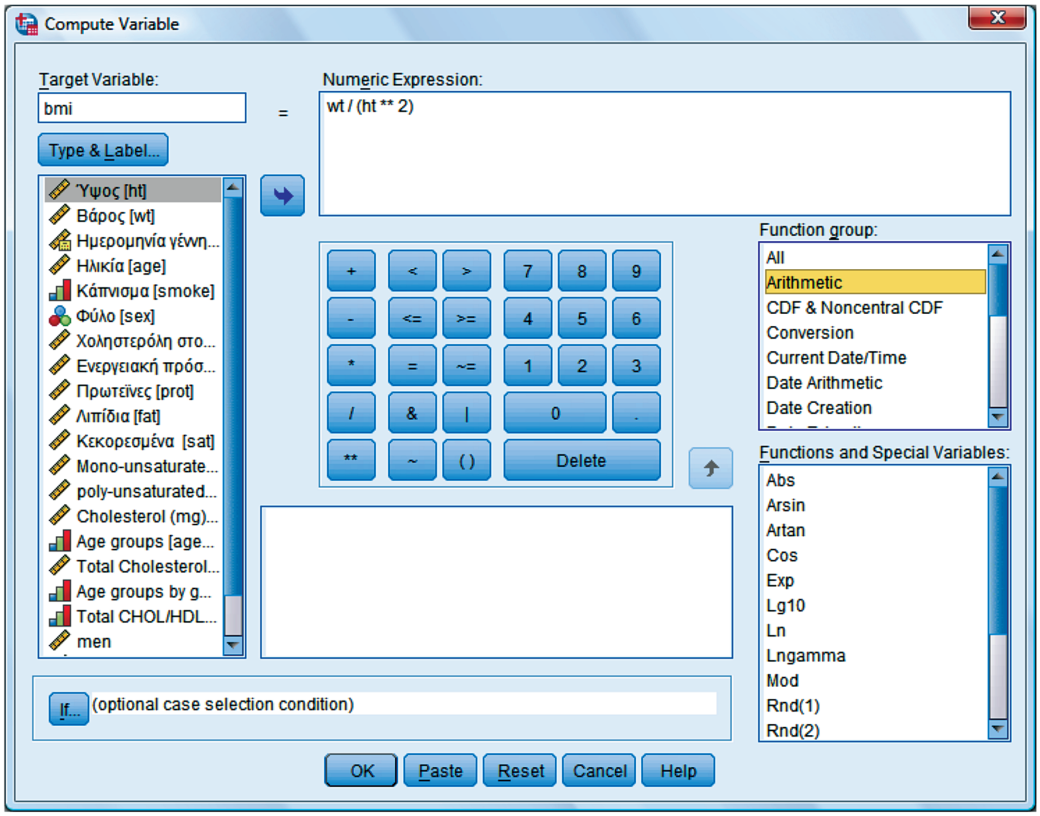
- Η μεταβλητή της οποίας θα υπολογιστούν οι τιμές, μπορεί να είναι είτε αριθμητικού είτε αλφαριθμητικού τύπου.

- Εφόσον πρόκειται να υπολογιστούν οι τιμές μιας νέας μεταβλητής μπορεί να οριστεί ο τύπος και το variable label της μεταβλητής.

■ Ο υπολογισμός των τιμών της μεταβλητής μπορεί να γίνει επιλεκτικά για ένα υποσύνολο μόνο των δεδομένων το οποίο ορίζεται με μια λογική συνθήκη (μέσω του πλήκτρου If).

■ Για τους υπολογισμούς των τιμών μιας μεταβλητής υπάρχουν διαθέσιμες από το SPSS περισσότερες από 70 συναρτήσεις διαφόρων τύπων: αριθμητικές συναρτήσεις, στατιστικές συναρτήσεις, αλφαριθμητικές συναρτήσεις, συναρτήσεις ημερομηνίας, συναρτήσεις κατανομών, συναρτήσεις τυχαίων μεταβλητών και συναρτήσεις για το χειρισμό των missing values.

Σχήμα 4.6 Πλαίσιο διαλόγου Compute Variable



Πρακτικά, για τον υπολογισμό των τιμών μιας μεταβλητής εργαζόμαστε ως εξής:

▶ Από τη γραμμή των μενού επιλέγουμε:
Transform
Compute Variable...

Ανοίγει το πλαίσιο διαλόγου Compute Variable (Σχήμα 4.6).

▶ Στο πεδίο Target Variable του πλαισίου πληκτρολογείται το όνομα της μεταβλητής. Μπορεί να είναι το όνομα μιας νέας μεταβλητής ή μιας ήδη υπάρχουσας.

▶ Οι τιμές της μεταβλητής ορίζονται με τη βοήθεια μιας αριθμητικής έκφρασης η οποία καταχωρείται στο πεδίο Numeric Expression. Η αριθμητική αυτή έκφραση μπορεί να πληκτρολογηθεί είτε απ' ευθείας από το πληκτρολόγιο του υπολογιστή, είτε να εισαχθεί με τη βοήθεια της αριθμομηχανής που υπάρχει στο πλαίσιο διαλόγου.

■ Οι αριθμητικές πράξεις που μπαίνουν στην έκφραση των τιμών της μεταβλητής είναι οι τέσσερις γνωστές αριθμητικές πράξεις, καθώς και η ύψωση ενός αριθμού σε δύναμη. Τα αντίστοιχα σύμβολα εμφανίζονται στο πληκτρολόγιο της αριθμομηχανής που υπάρχει εντός του πλαισίου. Είναι: πρόσθεση (+), αφαίρεση (-), πολλαπλασιασμός (*), διαίρεση (/). Η ύψωση σε δύναμη δηλώνεται με δύο αστεράκια ακολουθούμενα από την αντίστοιχη δύναμη, π.χ. varx**3 σημαίνει την ύψωση των τιμών της *varx* στην τρίτη. Κατά τον υπολογισμό της αριθμητικής έκφρασης, η σειρά εκτέλεσης των πράξεων είναι: ύψωση σε δύναμη, πολλαπλασιασμοί - διαιρέσεις και τέλος προσθέσεις και αφαιρέσεις. Όπου υπάρχουν αμφιβολίες ως προς τη σειρά εκτέλεσης των πράξεων καλό είναι να χρησιμοποιούνται παρενθέσεις.

■ Η έκφραση μπορεί να περιλαμβάνει συναρτήσεις οι οποίες εισάγονται από τη λίστα των συναρτήσεων του πλαισίου διαλόγου. Οι συναρτήσεις είναι καταχωρημένες κατά ομάδες στη λίστα Function group, ενώ κάθε φορά που επιλέγουμε μια ομάδα, εμφανίζονται στο πεδίο Functions and Special Variables οι συναρτήσεις που αυτή περιλαμβάνει. Κάνοντας κλικ στο όνομα μιας συνάρτησης μπορούμε να πάρουμε πληροφορίες για τη λειτουργία της στο πεδίο που εμφανίζεται στο κέντρο του πλαισίου διαλόγου. Όταν μια συνάρτηση εισαχθεί στο πεδίο Numeric Expression οι παράμετροί της εμφανίζονται με ερωτηματικά και πρέπει να οριστούν από το χρήστη.

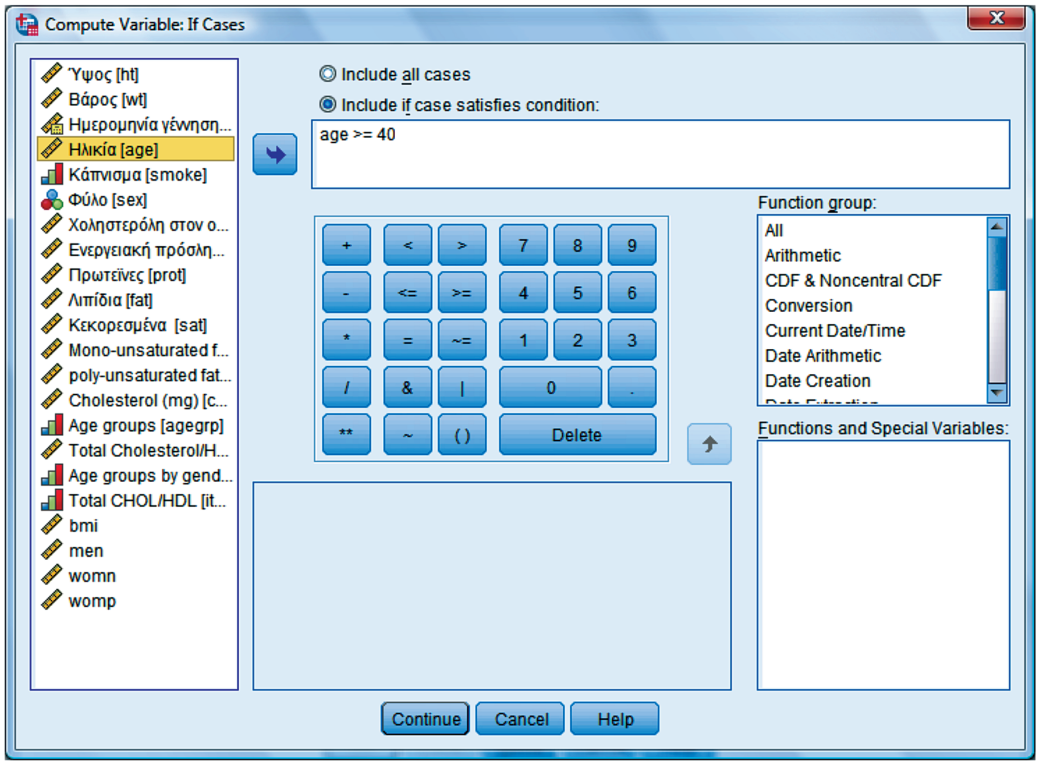
■ Αν στην αριθμητική έκφραση υπάρχουν σταθερές αλφαριθμητικού τύπου πρέπει να περικλείονται σε απλές ή διπλές αποστρόφους (' ', " ").

■ Οι αριθμοί πρέπει να εισάγονται στην έκφραση με αγγλοσαξονική μορφή έχοντας ως δεκαδικό σύμβολο την τελεία (.).

■ Αν πρόκειται να υπολογιστεί μια νέα μεταβλητή αλφαριθμητικού τύπου, πρέπει πρώτα να δηλωθεί ο τύπος της από την επιλογή Type & Label.

■ Αν ο υπολογισμός των τιμών της μεταβλητής αφορά ένα υποσύνολο μόνο από τις παρατηρήσεις ενός αρχείου, τότε ο ορισμός του υποσυνόλου αυτού γίνεται από το πλαίσιο διαλόγου If Cases (Σχήμα 4.7). Η λειτουργία του πλαισίου είναι παρόμοια με του αντίστοιχου πλαισίου που είδαμε στην εντολή Recode.

Σχήμα 4.7 Πλαίσιο διαλόγου Compute Variable: If Cases



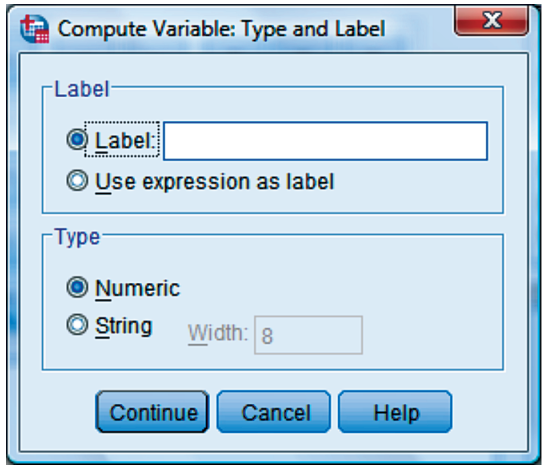
Υπολογισμός τιμών μεταβλητής: Type and Label

Κατά τον υπολογισμό των τιμών μιας νέας μεταβλητής, αυτή ορίζεται κατ' αρχήν ως αριθμητική. Αν πρόκειται να υπολογιστούν οι τιμές μιας νέας μεταβλητής αλφαριθμητικού τύπου, τότε πρέπει, πριν τη διατύπωση της έκφρασης του υπολογισμού της, να δηλωθεί ο τύπος της (type), καθώς και το πλήθος των χαρακτήρων που θα περιλαμβάνει (width). Αυτό γίνεται μέσα από το υποπλαίσιο διαλόγου Type and Label (Σχήμα 4.8). Το υποπλαίσιο αυτό μπορεί να χρησιμοποιείται και κατά τον υπολογισμό μεταβλητών αριθμητικού τύπου αλλά μόνο για τον ορισμό των labels αυτών. Στο υποπλαίσιο Type and Label περιλαμβάνονται τα εξής πεδία:

Label. Εδώ δηλώνεται –επιλεκτικά από το χρήστη– ένα variable label για τη νέα μεταβλητή μήκους μέχρι 120 χαρακτήρων. Εναλλακτικά, μπορεί να χρησιμοποιηθούν ως label, οι πρώτοι 110 χαρακτήρες της έκφρασης υπολογισμού της μεταβλητής (αυτοί που εμφανίζονται στο πεδίο Numeric Expression του πλαισίου Compute Variable).

Type. Στο πεδίο αυτό δηλώνεται ο τύπος και το πλήθος των χαρακτήρων της μεταβλητής (εφόσον πρόκειται για αλφαριθμητική μεταβλητή). Δεν μπορούν να χρησιμοποιούνται αλφαριθμητικές μεταβλητές για να καταχωρηθούν τιμές που προκύπτουν από αριθμητικούς υπολογισμούς.

Σχήμα 4.8 Υποπλαίσιο διαλόγου Type and Label



Χειρισμός των missing values στις συναρτήσεις

Ο χειρισμός των missing values κατά τον υπολογισμό των τιμών μιας μεταβλητής γίνεται με διαφορετικό τρόπο, ανάλογα με το αν οι τιμές της μεταβλητής προκύπτουν από απλές αριθμητικές πράξεις ή υπολογίζονται με τη χρήση συναρτήσεων. Συνήθως, κατά τον υπολογισμό των τιμών μέσω αριθμητικών πράξεων, όταν έστω και μία από τις εμπλεκόμενες μεταβλητές έχει missing values, τότε στην τελική μεταβλητή μπαίνουν επίσης missing values. Π.χ. στην έκφραση:

varx =var1 + var2 + var3,

αν έστω και μία από τις μεταβλητές var1, var2 και var3 έχει missing values σε μια παρατήρηση (case), τότε στην αντίστοιχη παρατήρηση της υπολογιζόμενης μεταβλητής *varx*, θα μπει επίσης missing value.

Διαφορετικός είναι ο χειρισμός των missing values όταν γίνεται χρήση συναρτήσεων. Για παράδειγμα στην έκφραση:

SUM(var1, var2, var3)

το αποτέλεσμα θα είναι missing value μόνο αν **και** οι τρεις μεταβλητές var1, var2 και var3, έχουν missing values.

Στην περίπτωση των στατιστικών συναρτήσεων, μπορεί ο χρήστης να ορίζει κάθε φορά τον ελάχιστο αριθμό των παραμέτρων μιας συνάρτησης, οι οποίες πρέπει να μην έχουν missing values, ώστε το αποτέλεσμα να μην είναι επίσης missing value. Αυτό μπορεί να γίνει πληκτρολογώντας μετά το όνομα της συνάρτησης μια τελεία μαζί με τον ελάχιστο αριθμό των απαιτούμενων παραμέτρων. Π.χ. στην έκφραση:

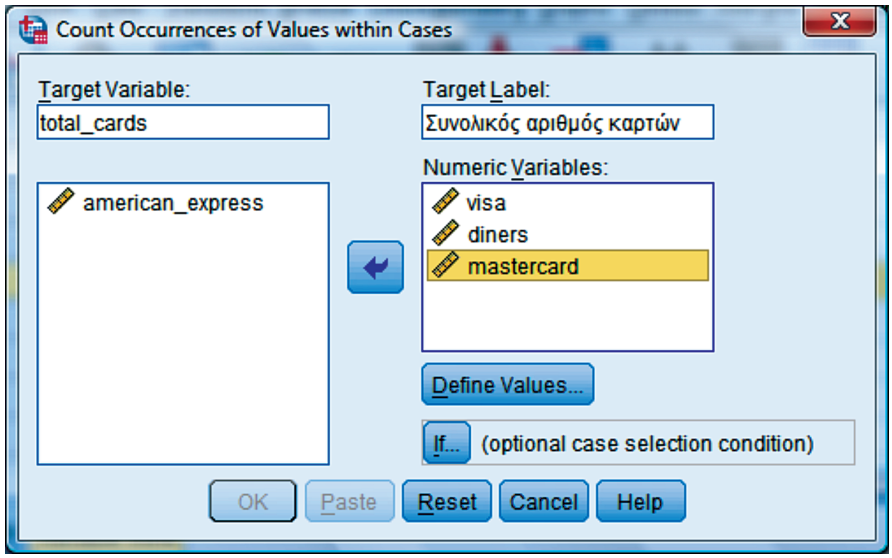
SUM.2(var1, var2, var3)

το αποτέλεσμα για μια παρατήρηση δεν θα είναι missing value, μόνο αν δύο τουλάχιστον από τις μεταβλητές var1, var2, var3 που μπαίνουν ως παράμετροι στη συνάρτηση του αθροίσματος, δεν έχουν missing values.

Συχνότητα εμφάνισης τιμών σε μία παρατήρηση – Count Occurrences of Values within Cases

Με την εντολή αυτή μπορεί να υπολογιστεί για κάθε παρατήρηση ενός αρχείου ο αριθμός των εμφανίσεων μίας ή περισσοτέρων τιμών σε μία λίστα μεταβλητών του αρχείου. Για παράδειγμα, σε μια έρευνα καταναλωτικών συνηθειών υπάρχει μια σειρά από ερωτήσεις που αναφέρονται στις πιστωτικές κάρτες που χρησιμοποιούν οι ερωτώμενοι. Στο ερωτηματολόγιο της έρευνας υπάρχει μια λίστα πιστωτικών καρτών από τις οποίες οι συμμετέχοντες επιλέγουν αυτές που έχουν στην κατοχή τους, απατώντας με *ναι* ή *όχι* (1 ή 2) στην ερώτηση της αντίστοιχης κάρτας. Στην περίπτωση αυτή, μπορεί να υπολογιστεί μια νέα μεταβλητή η οποία θα μετράει τον αριθμό των πιστωτικών καρτών που κατέχει κάθε ερωτώμενος. Ο υπολογισμός της νέας μεταβλητής μπορεί να γίνει μέσω της επιλογής Count του μενού Transform. Το αντίστοιχο πλαίσιο διαλόγου που ανοίγει ονομάζεται Count Occurrences of Values within Cases (Σχήμα 4.9).

Σχήμα 4.9 Πλαίσιο διαλόγου Count Occurrences of Values within Cases



Για να κάνουμε χρήση της εντολής Count εργαζόμαστε ως εξής:

▶ Από τη γραμμή των μενού επιλέγουμε:
Transform
Count Values within Cases...

Ανοίγει το πλαίσιο διαλόγου Count Occurrences of Values within Cases (Σχήμα 4.9).

▶ Πληκτρολογείται το όνομα της νέας μεταβλητής στο πεδίο Target Variable.

▶ Επιλέγονται οι μεταβλητές για τις οποίες θα υπολογιστεί ο αριθμός των εμφανίσεων μίας ή περισσοτέρων τιμών τους (στο παράδειγμα του Σχήματος 4.9 οι μεταβλητές είναι οι πιστωτικές κάρτες).

▶ Από την επιλογή Define Values δηλώνεται η τιμή ή οι τιμές που θα καταμετρηθούν μέσα στη λίστα των μεταβλητών (Σχήμα 4.10).

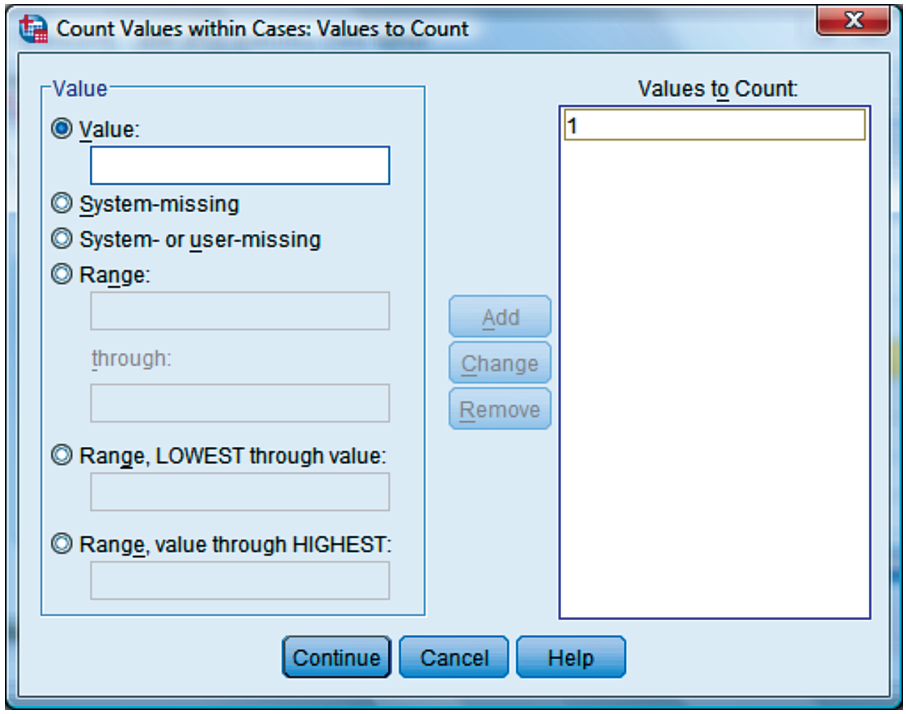
Προαιρετικά, μπορεί η καταμέτρηση αυτή να γίνει για ένα υποσύνολο μόνο των παρατηρήσεων του αρχείου. Στην περίπτωση αυτή, η επιλογή του υποσυνόλου των παρατηρήσεων γίνεται μέσα από το γενικό πλαίσιο If Cases, το οποίο ανοίγει πατώντας το πλήκτρο If (για τον τρόπο λειτουργίας του πλαισίου If Cases βλ. την αντίστοιχη παράγραφο).

▶ Στο υποπλαίσιο διαλόγου που ανοίγει πατώντας το πλήκτρο Define Values,

μπορούν να δηλωθούν απλές τιμές, εύρος τιμών ή missing values (system ή user-missing). Στην περίπτωση που η δήλωση αφορά εύρος τιμών, το ανώτερο όριο του διαστήματος του εύρους, καθώς και οι user-missing values που βρίσκονται εντός του εύρους, συμπεριλαμβάνονται στις τιμές που θα καταμετρηθούν. Όταν δηλώνεται μία ή περισσότερες τιμές προς καταμέτρηση, αυτές εισάγονται στο πεδίο Values to Count αφού πρώτα πατηθεί το πλήκτρο Add. Οι τιμές αυτές μπορούν να διορθώνονται ή να απομακρύνονται από τη λίστα των τιμών πατώντας τα πλήκτρα Change και Remove αντίστοιχα. Στο παράδειγμα των πιστωτικών καρτών, η τιμή που δηλώνεται για την καταφατική απάντηση στην ερώτηση της κατοχής μιας κάρτας είναι η τιμή 1 (= *ναι*) (Σχήμα 4.10).

Κατά την εκτέλεση της εντολής Count, η τιμή της τελικής μεταβλητής (target variable) αυξάνει κατά 1 κάθε φορά που σε μία από τις επιλεγμένες μεταβλητές συναντάται η τιμή ή οι τιμές που έχουν δηλωθεί στο πεδίο Values to Count. Αν για τη λίστα των μεταβλητών έχουν δηλωθεί πολλές τιμές, τότε η τιμή της τελικής μεταβλητής αυξάνει κατά 1 κάθε φορά που συναντάει μια από τις δηλωθείσες τιμές μέσα στη λίστα των μεταβλητών.

Σχήμα 4.10 Υποπλαίσιο διαλόγου που ανοίγει πατώντας το πλήκτρο Define Values



Σχετικές θέσεις των τιμών μιας μεταβλητής – Rank Cases

Η εντολή Rank Cases δημιουργεί νέες μεταβλητές οι οποίες περιέχουν τις σχετικές θέσεις (ranks) των τιμών άλλων αριθμητικών μεταβλητών. Οι σχετικές θέσεις προκύπτουν διατάσσοντας τις τιμές των αρχικών μεταβλητών και θέτοντας στις νέες μεταβλητές τη σειρά διάταξης των αρχικών τιμών (η μικρότερη τιμή παίρνει την τιμή 1, η αμέσως επόμενη την τιμή 2 κ.ο.κ.). Εναλλακτικά, αντί των σχετικών θέσεων των τιμών μπορεί στις νέες μεταβλητές να περιέχονται οι κλασματικές σχετικές θέσεις (fractional ranks), ή οι εκατοστιαίες θέσεις (percentiles) ή άλλα μέτρα που βασίζονται στις σχετικές θέσεις των αρχικών τιμών. Οι νέες μεταβλητές αποκτούν αυτομάτως κατά τη στιγμή του υπολογισμού τους ονόματα και labels. Τα ονόματα αυτά είναι συνδυασμός των ονομάτων των αρχικών μεταβλητών και του αντίστοιχου μέτρου που υπολογίζεται κάθε φορά. Π.χ όταν υπολογίζονται οι απλές σχετικές θέσεις της μεταβλητής *varx*, η νέα μεταβλητή έχει ως όνομα το όνομα της αρχικής μεταβλητής με το γράμμα R να προηγείται (*Rvarx*). Επιπλέον, στο output του SPSS δημιουργείται μια λίστα με τις αντιστοιχήσεις των ονομάτων των αρχικών και των νέων μεταβλητών, καθώς και των labels αυτών.

Η εντολή Rank Cases δίνει στο χρήστη τις εξής επιπλέον δυνατότητες:

- Να ορίζει τον τρόπο διάταξης των αρχικών τιμών για τον υπολογισμό των σχετικών θέσεων. Σύμφωνα με την αρχική ρύθμιση του SPSS οι τιμές διατάσσονται κατά αύξουσα σειρά και μετά υπολογίζονται οι σχετικές θέσεις αυτών. Εναλλακτικά, θα μπορούσε να ζητηθεί από το χρήστη η διάταξη να γίνει κατά φθίνουσα σειρά (στην τελευταία περίπτωση η μεγαλύτερη τιμή θα πάρει ως rank την τιμή 1, η αμέσως μικρότερη την τιμή 2 κ.ο.κ.).

- Ο υπολογισμός των σχετικών θέσεων μπορεί να μη γίνεται συνολικά για όλες τις παρατηρήσεις του αρχείου αλλά για συγκεκριμένες υποομάδες αυτών. Οι υποομάδες αυτές ορίζονται διά μέσου μίας ή περισσοτέρων κατηγορικών μεταβλητών που υπάρχουν στο αρχείο των δεδομένων. Σε αυτές τις περιπτώσεις, οι σχετικές θέσεις υπολογίζονται με διάταξη των τιμών των αρχικών μεταβλητών στο εσωτερικό της κάθε υποομάδας.

Πρακτικά, για τον υπολογισμό των σχετικών θέσεων μιας μεταβλητής εργαζόμαστε ως εξής:

▶ Από τη γραμμή των μενού επιλέγουμε:
Transform
Rank Cases...

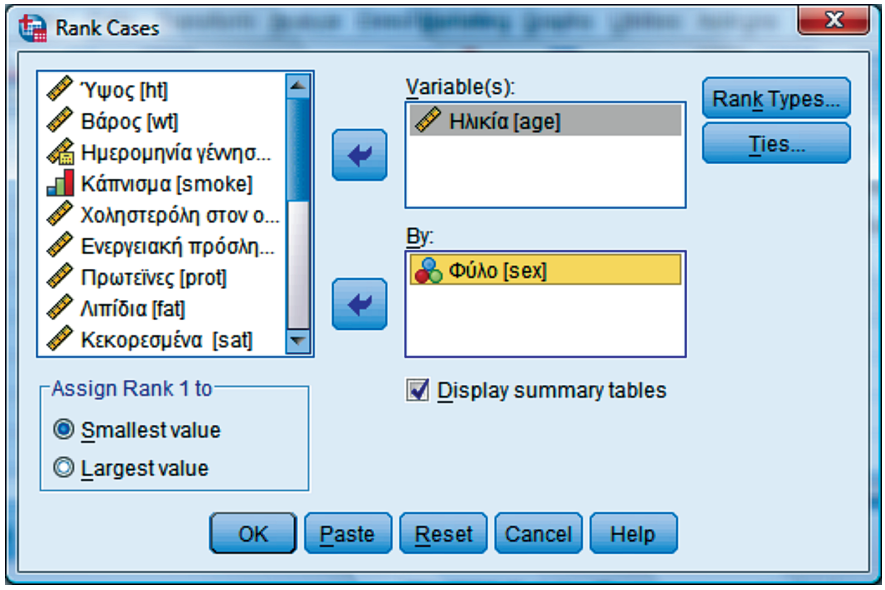
Ανοίγει το πλαίσιο διαλόγου Rank Cases (Σχήμα 4.11).

▶ Επιλέγουμε μία ή περισσότερες μεταβλητές προς διάταξη. Οι μεταβλητές πρέπει να είναι αποκλειστικά αριθμητικού τύπου.

▶ Η διάταξη των τιμών των αρχικών μεταβλητών μπορεί να γίνει κατά αύξουσα ή φθίνουσα σειρά. Για το λόγο αυτόν επιλέγουμε αντίστοιχα τα πλήκτρα: Assign Rank 1 to Smallest value (είναι η αρχική ρύθμιση του SPSS) ή Assign Rank 1 to Largest value.

▶ Επιπλέον, στις νέες μεταβλητές οι σχετικές θέσεις μπορούν να υπολογίζονται στο εσωτερικό συγκεκριμένων υποομάδων του αρχείου που ορίζονται μέσω κατηγορικών μεταβλητών. Οι μεταβλητές αυτές πρέπει να εισάγονται στη λίστα με την ένδειξη By.

Σχήμα 4.11 Πλαίσιο διαλόγου Rank Cases

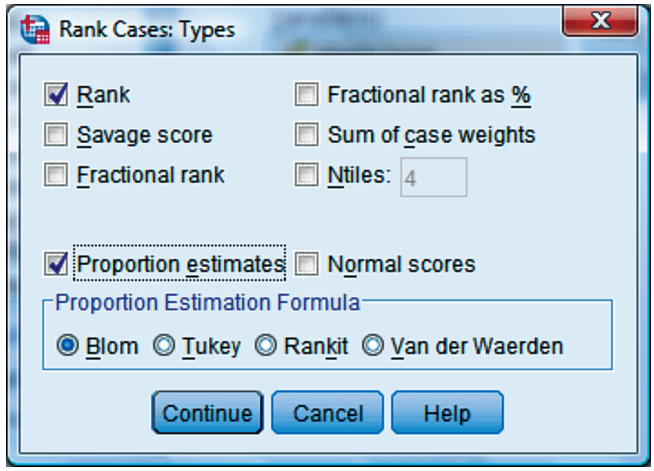


Μέσω του πλαισίου διαλόγου που εμφανίζεται στο Σχήμα 4.11, υπολογίζεται μία νέα μεταβλητή για την *Ηλικία* [*age*] με όνομα *Rage*, η οποία περιλαμβάνει τις σχετικές θέσεις της ηλικίας στο εσωτερικό του κάθε φύλου. Αν δεν είχε δηλωθεί η μεταβλητή *Φύλο* για τον προσδιορισμό των δύο υποομάδων, οι σχετικές θέσεις θα υπολογίζονταν για το σύνολο των παρατηρήσεων του αρχείου.

Τρόπος υπολογισμού των σχετικών θέσεων: Types

Η εντολή Rank Cases δίνει τη δυνατότητα στο χρήστη να ορίσει ο ίδιος τον τρόπο υπολογισμού των σχετικών θέσεων για μια μεταβλητή. Οι μέθοδοι υπολογισμού των σχετικών θέσεων περιλαμβάνονται στο υποπλαίσιο διαλόγου Rank Types (Σχήμα 4.12). Μεταξύ αυτών υπάρχουν οι απλές σχετικές θέσεις (Rank), οι κλασματικές σχετικές θέσεις (Fractional rank), οι εκατοστιαίες θέσεις (Fractional rank as %), τα Savage (exponential) scores κλπ.

Σχήμα 4.12 Υποπλαίσιο διαλόγου Rank Cases: Types

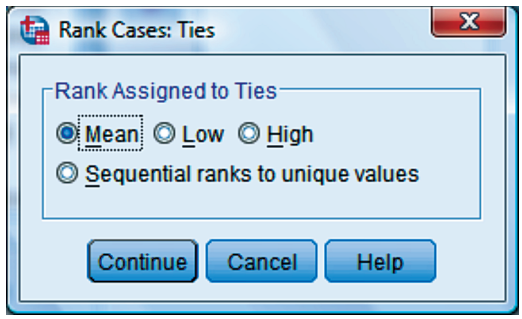


Μπορεί να επιλεχθούν όποιες από τις παραπάνω μεθόδους επιθυμεί ο χρήστης, δημιουργώντας ταυτόχρονα ισάριθμες νέες μεταβλητές.

Ισοβαθμίες: Ties

Ο τρόπος χειρισμού των ισοβαθμιών κατά τη διάταξη των σχετικών θέσεων μιας μεταβλητής μπορεί επίσης να οριστεί από το χρήστη. Ο ορισμός γίνεται από το υποπλαίσιο διαλόγου Ties (Σχήμα 4.13).

Σχήμα 4.13 Υποπλαίσιο διαλόγου Rank Cases: Ties



Ο τρόπος χειρισμού των ισοβαθμιών, μέσω των διαφόρων επιλογών που εμφανίζονται στο υποπλαίσιο Ties, γίνεται αντιληπτός από τον παρακάτω πίνακα:

Τιμή	Mean	Low	High	Sequential
0	3	1	5	1
0	3	1	5	1
0	3	1	5	1
0	3	1	5	1
0	3	1	5	1
1	6,5	6	7	2
1	6,5	6	7	2
3	8	8	8	3

Στην επιλογή Mean, οι ισοβαθμίες υπολογίζονται ως η μέση θέση των τιμών που ισοβαθμούν. Για παράδειγμα, κατά τον υπολογισμό των θέσεων των πέντε πρώτων ίσων τιμών, η μέση θέση είναι (1+2+3+4+5)/5 = 15/5=3. Η επόμενη θέση λογίζεται αρχικά ως η 6η στη συνολική διάταξη (διότι έχουν ήδη προηγηθεί πέντε αριθμοί), επειδή όμως έχουμε πάλι δύο τιμές που ισοβαθμούν, αντικαθίστανται από τη μέση θέση τους, δηλαδή (6+7)/2=6,5. Στις επιλογές Low και High οι ισοβαθμίες αντικαθίστανται αντίστοιχα από τη μικρότερη και μεγαλύτερη αρχική θέση των τιμών που ισοβαθμούν. Τέλος, στην επιλογή Sequential οι ισοβαθμίες παίρνουν όλες την ίδια θέση, ενώ οι θέσεις αυξάνουν πάντα κατά ένα.

Αυτόματη επανακωδικοποίηση – Automatic Recode

Όταν οι κωδικοί των κατηγοριών μιας κατηγορικής μεταβλητής δεν είναι διαδοχικοί ακέραιοι, υπάρχει το ενδεχόμενο, για ορισμένες στατιστικές διαδικασίες, να επιβαρύνεται ο χρόνος εκτέλεσής τους ή να αυξάνουν οι απαιτήσεις τους σε μνήμη. Επιπλέον, υπάρχουν διαδικασίες οι οποίες δεν επιδέχονται αλφαριθμητικές μεταβλητές, ενώ υπάρχουν άλλες οι οποίες απαιτούν, οι κατηγορικές μεταβλητές που υπεισέρχονται σε αυτές, να είναι απαραίτητα κωδικοποιημένες σε διαδοχικούς ακέραιους (π.χ. τα επίπεδα ενός παράγοντα σε μια διαδικασία της ανάλυσης διακύμανσης).

Η εντολή Automatic Recode επανακωδικοποιεί τις τιμές μιας ή περισσοτέρων μεταβλητών (αριθμητικών ή αλφαριθμητικών) και τις μετατρέπει σε διαδοχικούς ακέραιους αριθμούς ξεκινώντας από το 1. Κατά την εκτέλεση της εντολής, δημιουργούνται νέες μεταβλητές (έχοντας ως τιμές διαδοχικούς ακέραιους) με variable labels και value labels ίδια με αυτά των αρχικών μεταβλητών. Για τις τιμές των αρχικών μεταβλητών που δεν έχουν οριστεί value labels, οι επανακω-

δικοποιημένες τιμές των νέων μεταβλητών έχουν ως value labels τις ίδιες τις αρχικές τιμές. Κάθε φορά που εκτελείται η εντολή Automatic Recode, στο αρχείο αποτελεσμάτων του SPSS καταχωρείται ένας πίνακας που δείχνει τις παλιές και τις νέες τιμές, καθώς και τα αντίστοιχα value labels. Ειδικότερα, για την εντολή Automatic Recode ισχύουν τα εξής:

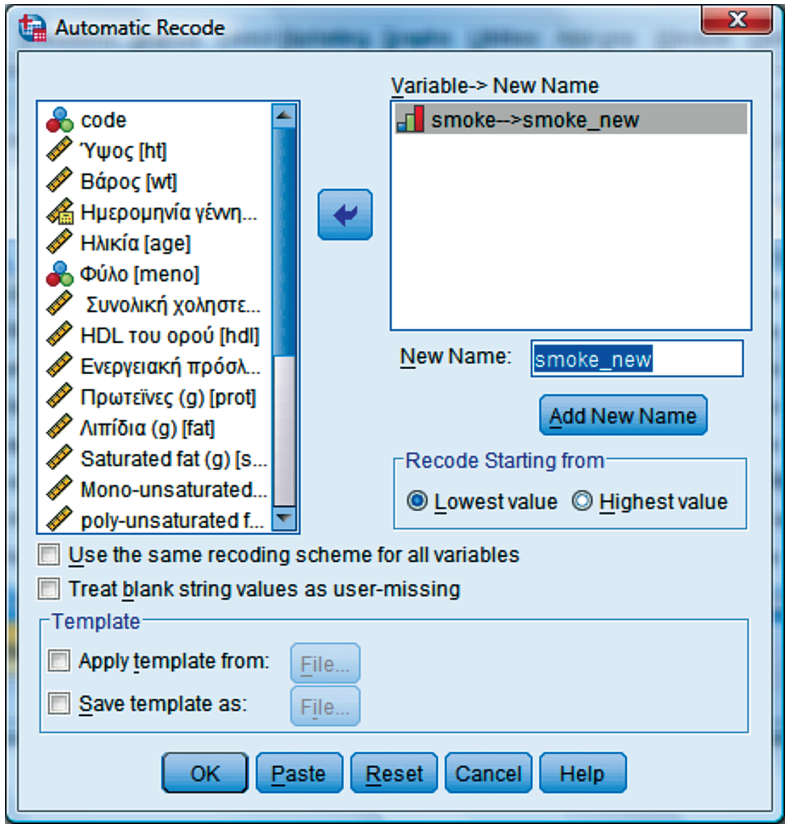
■ Οι αλφαριθμητικές μεταβλητές επανακωδικοποιούνται με βάση την αλφαβητική διάταξη των χαρακτήρων τους, με τα κεφαλαία γράμματα να προηγούνται κατά τη διάταξη των αντίστοιχων μικρών γραμμάτων. Οι system-missing values παραμένουν ως έχουν κατά την επανακωδικοποίηση, ενώ οι user-missing values επανακωδικοποιούνται επίσης σε user-missing values, έχοντας ως τιμές τούς αμέσως μεγαλύτερους ακέραιους αριθμούς από τη μεγαλύτερη τιμή της νέας κωδικοποίησης. Επιπλέον, κατά την επανακωδικοποίηση, οι user-missing values διατηρούν στη νέα μεταβλητή την αρχική τους διάταξη. Για παράδειγμα, αν η αρχική μεταβλητή έχει 8 έγκυρες τιμές, η χαμηλότερη user-missing value της αρχικής μεταβλητής, θα κωδικοποιηθεί στη νέα με την τιμή 9, η αμέσως μεγαλύτερη με την τιμή 10 κ.ο.κ.

■ Παρέχεται η δυνατότητα στο χρήστη να χρησιμοποιήσει το ίδιο σχήμα αυτόματης κωδικοποίησης για όλες τις μεταβλητές που έχουν επιλεγεί προς επανακωδικοποίηση (Use the same recoding scheme for all variables). Αν επιλεγεί η συγκεκριμένη δυνατότητα, τότε όλες οι μεταβλητές οι οποίες θα επανακωδικοποιηθούν, πρέπει να είναι του ιδίου τύπου (ή αριθμητικές ή αλφαριθμητικές). Επιπλέον, οι τιμές όλων των μεταβλητών που επανακωδικοποιούνται, διατάσσονται σε ενιαία διάταξη και, με βάση τη σειρά που θα προκύψει, ορίζονται οι διαδοχικοί ακέραιοι που θα τις αντικαταστήσουν (οπότε στην περίπτωση αυτήν ενδέχεται, για ορισμένες μεταβλητές, οι ακέραιοι αριθμοί που θα προκύψουν από την επανακωδικοποίηση να μην είναι διαδοχικοί).

■ Μέσω της επιλογής Treat blank string values as user-missing, δίνεται η δυνατότητα, κατά την επανακωδικοποίηση των αλφαριθμητικών μεταβλητών, οι κενές τιμές τους να μετατρέπονται σε user-missing values, με τιμή τον αμέσως επόμενο ακέραιο αριθμό από τη μεγαλύτερη έγκυρη τιμή της νέας κωδικοποίησης.

■ Δίνεται η δυνατότητα να αποθηκεύεται ένα σχήμα κωδικοποίησης υπό τη μορφή αρχείου ώστε να είναι δυνατόν να επαναχρησιμοποιηθεί σε άλλες μεταβλητές ή άλλα αρχεία δεδομένων. Η δυνατότητα αυτή παρέχεται από την επιλογή Template. Επιλέγοντας Save template as μπορούμε να αποθηκεύσουμε ένα σχήμα κωδικοποίησης που έχουμε ορίσει, ενώ με την επιλογή Apply template from, μπορούμε να χρησιμοποιήσουμε ένα ήδη αποθηκευμένο σχήμα κωδικοποίησης.

Σχήμα 4.14 Πλαίσιο διαλόγου Automatic Recode



Για την αυτόματη επανακωδικοποίηση μιας μεταβλητής, εργαζόμαστε ως εξής:

- Από τη γραμμή των μενού επιλέγουμε:
  Transform
  Automatic Recode...

Ανοίγει το πλαίσιο διαλόγου Automatic Recode (Σχήμα 4.14).

- Επιλέγονται οι μεταβλητές που θα επανακωδικοποιηθούν.
- Για κάθε μεταβλητή που πρόκειται να επανακωδικοποιηθεί, εισάγουμε ένα όνομα και επιλέγουμε Add New Name.

Επανακωδικοποίηση μεταβλητών μέσω της διαδικασίας Visual Binning

Συχνά, στις στατιστικές αναλύσεις, είναι αναγκαίο μια συνεχής μεταβλητή να επανακωδικοποιηθεί σε κατηγορική. Για παράδειγμα, η ηλικία πολλές φορές χρειάζεται να μετασχηματιστεί σε κατηγορική μεταβλητή ενώ και το εισόδημα σε πολλές αναλύσεις είναι απαραίτητο να επανακωδικοποιείται σε κατηγορίες εισοδήματος. Επιπλέον, σε ορισμένες περιπτώσεις είναι αναγκαίο να συμπτυχθούν οι κατηγορίες μιας κατηγορικής μεταβλητής σε ένα μικρότερο αριθμό κατηγοριών.
Επανακωδικοποιήσεις αυτού του τύπου μπορούν να γίνουν μέσω της διαδικασίας Visual Binning. Η διαδικασία αυτή βοηθάει βήμα-βήμα το χρήστη να επανακωδικοποιήσει τις μεταβλητές που τον ενδιαφέρουν με το βέλτιστο τρόπο.

Ειδικότερα, μέσω της διαδικασίας Visual Binning ο χρήστης μπορεί:

- Να δημιουργήσει νέες κατηγορικές μεταβλητές επανακωδικοποιώντας τις τιμές άλλων συνεχών μεταβλητών. Για παράδειγμα, μπορεί να δημιουργήσει μια κατηγορική μεταβλητή με τις ηλικιακές ομάδες, χρησιμοποιώντας τις τιμές μιας συνεχούς μεταβλητής για την ηλικία.

- Να δημιουργήσει νέες κατηγορικές μεταβλητές οι οποίες θα προκύψουν από τη σύμπτυξη των κατηγοριών άλλων κατηγορικών μεταβλητών. Για παράδειγμα, μια μεταβλητή τύπου Likert με πέντε κατηγορίες: “πολύ ικανοποιημένος/η”, “ικανοποιημένος/η”, “ουδέτερος/η”, “δυσαρεστημένος/η”, “πολύ δυσαρεστημένος/η”, μπορεί να μετατραπεί σε μια άλλη με τρεις κατηγορίες: “ικανοποιημένος/η”, “ουδέτερος/η”, “ δυσαρεστημένος/η”.

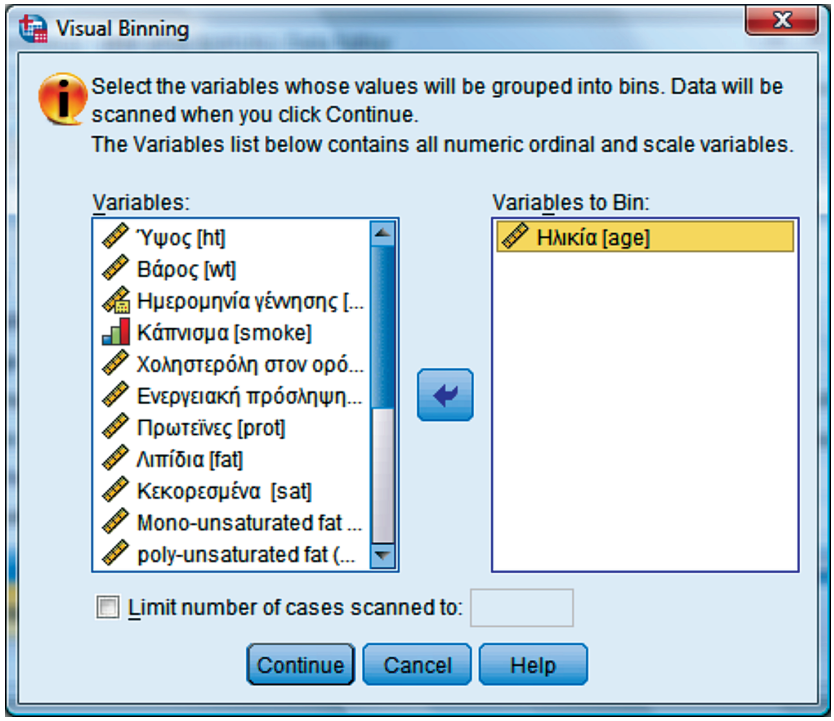
Για να κάνουμε χρήση της διαδικασίας Visual Binning, προκειμένου να μετατρέψουμε μία ή περισσότερες αρχικές μεταβλητές σε νέες κατηγορικές, εργαζόμαστε ως εξής:

▶ Από τη γραμμή των μενού του Data Edtor επιλέγουμε:
Transform
Visual Binning...

Ανοίγει το πρώτο από τα πλαίσια διαλόγου της διαδικασίας Visual Binning (Σχήμα 4.15). Στο πλαίσιο αυτό ο χρήστης επιλέγει τις αρχικές μεταβλητές από τις οποίες θα δημιουργηθούν οι νέες κατηγορικές. Η διαδικασία είναι ουσιαστικά μια διαδικασία συνένωσης (ομαδοποίησης) των τιμών των αρχικών μεταβλητών ώστε να δημιουργηθούν οι κατηγορίες των νέων. Επειδή η λειτουργία της διαδικασίας στοχεύει στη βέλτιστη κατηγοριοποίηση των αρχικών μεταβλητών (αναφορικά με την επιλογή των ορίων των διαστημάτων που θα δη-

μιουργηθούν) είναι απαραίτητο να «διαβαστούν» πρώτα οι τιμές των αρχικών μεταβλητών ώστε να παραχθούν τα ιστογράμματά τους. Η διαδικασία αυτή μπορεί να πάρει ορισμένο χρόνο εφόσον ο αριθμός των παρατηρήσεων του αρχείου είναι μεγάλος (αρκετές χιλιάδες παρατηρήσεων). Για το λόγο αυτό, αν το επιθυμεί ο χρήστης, μπορεί να περιορίσει τις παρατηρήσεις που θα χρησιμοποιηθούν για την παραγωγή των ιστογραμμάτων, δηλώνοντας τον ακριβή αριθμό τους μέσω της επιλογής Limit number of cases scanned to. Η επιλογή αυτή βρίσκεται στη βάση του πλαισίου διαλόγου.

Σχήμα 4.15 Πλαίσιο διαλόγου Visual Binning

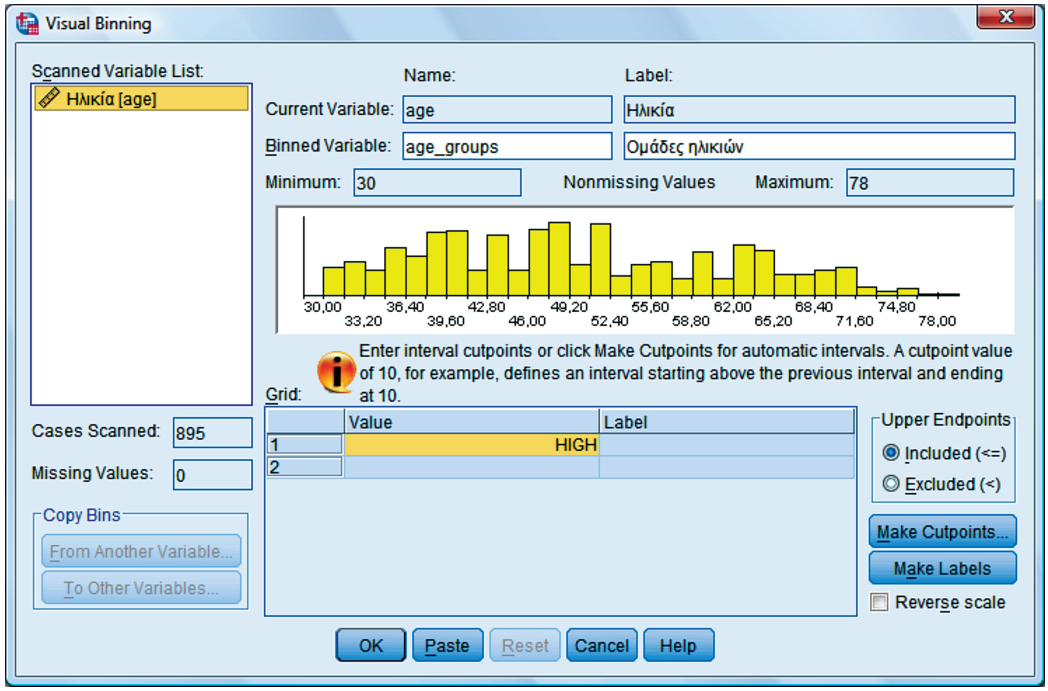


▶ Οι μεταβλητές που πρόκειται να επανακωδικοποιηθούν εισάγονται στη λίστα Variables to Bin. Στη συνέχεια επιλέγουμε Continue.

Στο παράδειγμα του Σχήματος 4.15, η μεταβλητή που θα επανακωδικοποιηθεί είναι η *Ηλικία* [*age*].

Ανοίγει το κύριο πλαίσιο διαλόγου της διαδικασίας Visual Binning (Σχήμα 4.16).

Σχήμα 4.16 Κύριο πλαίσιο διαλόγου της διαδικασίας Visual Binning



▶ Στο νέο πλαίσιο, επιλέγουμε ξανά τη μεταβλητή ή τις μεταβλητές που θα επανακωδικοποιηθούν κάνοντας κλικ στο όνομά τους. Για κάθε μία μεταβλητή που επιλέγουμε, εμφανίζεται εντός του πλαισίου το ιστόγραμμά της. Στην περίπτωση του παραδείγματός μας, η μεταβλητή που επιλέγεται είναι η *Ηλικία* [*age*].

▶ Στο πεδίο Binned Variable Name πληκτρολογείται για κάθε μεταβλητή που θα επανακωδικοποιηθεί, το όνομα της μεταβλητής που θα προκύψει, ενώ στο πεδίο Binned Variable Label εισάγεται το variable label της νέας μεταβλητής. Στο παράδειγμά μας, *age_ groups* και *Ομάδες ηλικιών* αντίστοιχα.

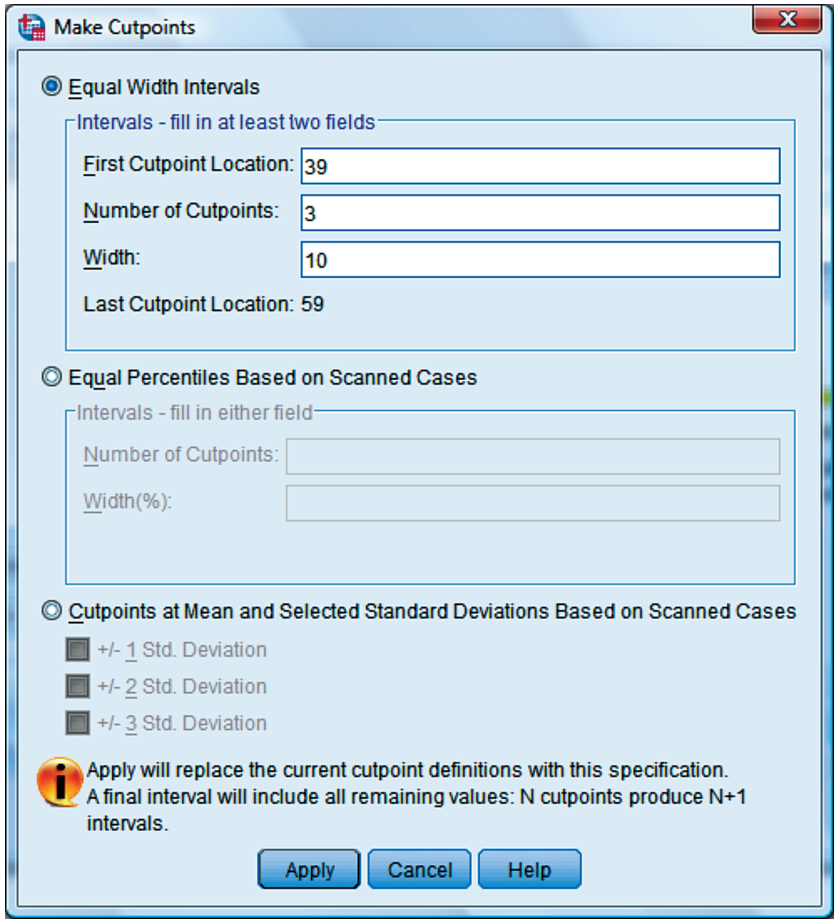
▶ Στη συνέχεια επιλέγουμε Make Cutpoints.

▶ Στο νέο πλαίσιο διαλόγου που εμφανίζεται (Σχήμα 4.17), επιλέγουμε Equal Width Intervals. Από τα τρία πεδία της συγκεκριμένης επιλογής είναι απαραίτητο να δηλωθούν τουλάχιστον δύο. Τα πεδία της επιλογής αφορούν το ανώτερο όριο (cutpoint) του πρώτου διαστήματος της νέα μεταβλητής (First Cutpoint Location), τον αριθμό των ορίων που θα χρησιμοποιηθούν για τον προσδιορισμό των διαστημάτων (Number of Cutpoints) και το εύρος των διαστημάτων (Width). Αν ο χρήστης συμπληρώσει τα δύο από τα τρία αυτά

πεδία, το SPSS υπολογίζει με βέλτιστο τρόπο και το τρίτο. Σε περιπτώσεις που είναι σαφές για το χρήστη ποιες είναι οι τιμές των τριών πεδίων, πρέπει να συμπληρώνονται και οι τρεις. Ο τελικός αριθμός των διαστημάτων (κατηγοριών) που θα προκύψει, θα είναι ίσος με τον αριθμό των δηλωθέντων ορίων συν ένα.

Στο παράδειγμα της ηλικίας (Σχήμα 4.17), δηλώνονται και το ανώτερο όριο του πρώτου διαστήματος (=39) και ο αριθμός των ορίων (=3) και το εύρος των διαστημάτων (=10).

Σχήμα 4.17 Πλαίσιο διαλόγου Make Cutpoints



Εκτός της επιλογής Equal Width Intervals, στο πλαίσιο διαλόγου Make Cutpoints υπάρχουν οι εξής επιπλέον δυνατότητες κατασκευής διαστημάτων:

■ Διαστήματα με περίπου ίσο αριθμό παρατηρήσεων (Equal Percentiles Based on Scanned Cases).

■ Διαστήματα με κεντρικό σημείο τη μέση τιμή της αρχικής μεταβλητής και όρια μία, δύο ή και τρεις τυπικές αποκλίσεις από αυτήν (Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases). Στην περίπτωση αυτή, μπορούν να δηλωθούν μέχρι τρεις επιλογές (+/- 1 Std. Deviation, +/- 2 Std. Deviation, +/- 3 Std. Deviation), οπότε ο μέγιστος αριθμός των διαστημάτων που μπορούν να προκύψουν είναι οκτώ.

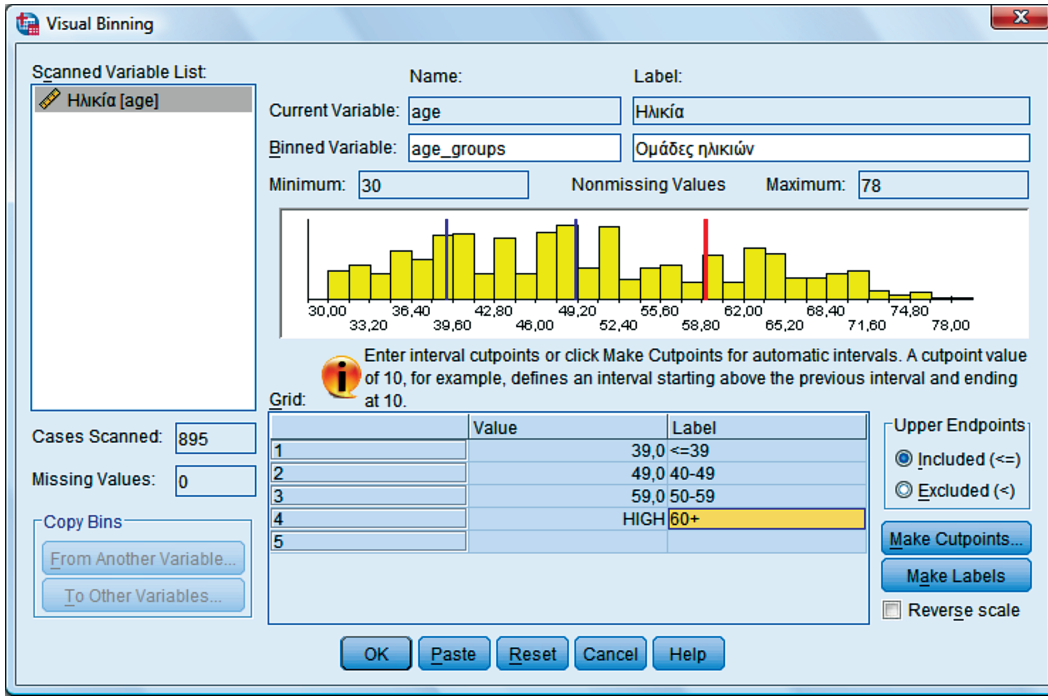
▶ Αφού συμπληρωθούν τα πεδία του πλαισίου Make Cutpoints, επιλέγουμε Apply.

Επανεμφανίζεται το κύριο πλαίσιο διαλόγου Visual Binning (Σχήμα 4.18). Τα ανώτερα όρια των διαστημάτων που θα δημιουργηθούν, εμφανίζονται στο εσωτερικό του πίνακα που υπάρχει στο κέντρο του πλαισίου. Επιπλέον, εμφανίζονται κατακόρυφες διαχωριστικές γραμμές εντός του ιστογράμματος της μεταβλητής, δηλώνοντας τα ανώτερα όρια των διαστημάτων. Αν το επιθυμεί ο χρήστης, μπορεί να τροποποιεί δυναμικά τα όρια των διαστημάτων «τραβώντας» (dragging and dropping) δεξιά ή αριστερά τις διαχωριστικές γραμμές επί του ιστογράμματος. Επίσης, μπορεί να απομακρύνει μια διαχωριστική γραμμή «τραβώντας» την εκτός του ιστογράμματος και ελαττώνοντας με αυτόν τον τρόπο τον αριθμό των διαστημάτων. Πρέπει να σημειωθεί ότι τα διαστήματα μπορούν επίσης να τροποποιηθούν αλλάζοντας τα όριά τους στο εσωτερικό του πίνακα όπου εμφανίζονται.

Η αρχική (default) ρύθμιση του SPSS προβλέπει τα ανώτερα όρια των διαστημάτων να συμπεριλαμβάνονται στα διαστήματα που θα δημιουργηθούν. Το διαπιστώνουμε από το «πατημένο» πλήκτρο Included (<=) στην επιλογή Upper Endpoints. Αν θέλουμε τα ανώτερα όρια να μη συμπεριλαμβάνονται στα διαστήματα που θα προκύψουν, μπορούμε να πατήσουμε το πλήκτρο Excluded (<).

▶ Αν επιθυμούμε να δημιουργηθούν value labels για τις νέες κατηγορικές μεταβλητές, επιλέγουμε Make Labels, πάλι από το κύριο πλαίσιο διαλόγου της διαδικασίας Visual Binning. Η τελευταία αυτή επιλογή δημιουργεί ως value labels ακριβώς τα διαστήματα που έχουν οριστεί π.χ. «< =39» ή «40-49» κλπ. Τα labels μπορούν να τροποποιηθούν από το χρήστη διορθώνοντάς τα ή αλλάζοντάς τα απ' ευθείας από τον πίνακα όπου εμφανίζονται.

Σχήμα 4.18 Κύριο πλαίσιο διαλόγου της διαδικασίας Visual Binning με τα ανώτερα όρια των διαστημάτων και τα αντίστοιχα value labels αυτών



Αντικατάσταση των missing values

Η παρουσία των missing values συχνά δημιουργεί προβλήματα στις στατιστικές αναλύσεις (ειδικά στις πολυμεταβλητές και στις αναλύσεις των χρονολογικών σειρών) με αποτέλεσμα να τίθεται ως απαραίτητη ανάγκη η αντικατάστασή τους. Η αντικατάσταση των missing values μπορεί να γίνει με διάφορους τρόπους μέσω της επιλογής Replace Missing Values. Με τη βοήθεια της επιλογής αυτής, δημιουργούνται νέες μεταβλητές οι οποίες διαφέρουν από τις αρχικές μόνο ως προς τις missing values οι οποίες έχουν αντικατασταθεί από έγκυρες τιμές. Οι νέες μεταβλητές που δημιουργούνται, έχουν ονόματα σχηματιζόμενα από τους πρώτους έξι χαρακτήρες των αρχικών μεταβλητών ακολουθούμενα από χαμηλή παύλα (underscore) και έναν αύξοντα αριθμό. Για παράδειγμα, η αντικατάσταση των missing values για την μεταβλητή *age* θα δημιουργήσει μια νέα μεταβλητή με όνομα *age_1.* Οι νέες μεταβλητές διατηρούν τα value labels των αρχικών μεταβλητών (εφόσον βέβαια αυτά υπάρχουν).

Για την αντικατάσταση των missing values μιας μεταβλητής εργαζόμαστε ως εξής:

- Από τη γραμμή των μενού επιλέγουμε:
  Transform
  Replace Missing Values...

Ανοίγει το πλαίσιο διαλόγου Replace Missing Values (Σχήμα 4.19).

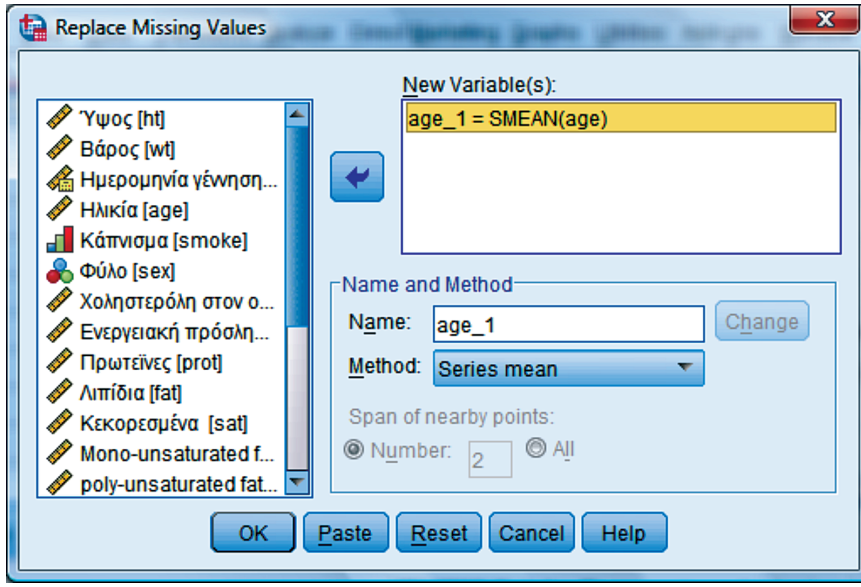
▶ Επιλέγεται η μέθοδος μέσω της οποίας θα εκτιμηθούν οι missing values της μεταβλητής.

▶ Επιλέγεται η μεταβλητή ή οι μεταβλητές για τις οποίες θα εκτιμηθούν οι missing values.

Προαιρετικά ο χρήστης μπορεί:

- Να εισάγει ονόματα για τις νέες μεταβλητές που θα δημιουργηθούν, αλλάζοντας αυτά που εξ' ορισμού δίνει το SPSS.
- Να αλλάξει τη μέθοδο εκτίμησης για μία από τις επιλεγμένες μεταβλητές.

Σχήμα 4.19 Πλαίσιο διαλόγου Replace Missing Values



Μέθοδοι εκτίμησης των Missing Values

Οι διαθέσιμες από το SPSS μέθοδοι για την αντικατάσταση των missing values είναι οι εξής:

Series mean. Η μέθοδος αυτή, η οποία είναι και η default μέθοδος του SPSS, αντικαθιστά τις missing values μιας μεταβλητής με τη μέση τιμή όλων των υπόλοιπων έγκυρων τιμών της μεταβλητής.

Mean of nearby points. Αντικαθιστά τις missing values με τη μέση τιμή των έγκυρων τιμών που βρίσκονται πέριξ αυτών. Ο αριθμός των έγκυρων παρατηρήσεων, άνω και κάτω των missing values, βάσει των οποίων θα γίνει ο υπολογισμός της μέσης τιμής, ορίζεται από την επιλογή Span of nearby points. Η επι-

λογή αυτή ενεργοποιείται μόνο εφόσον ο χρήστης επιλέξει ως μέθοδο υπολογισμού των missing values τη μέση τιμή ή τη διάμεσο των γειτονικών σημείων.

Median of nearby points. Αντικαθιστά τις missing values με τη διάμεσο των έγκυρων τιμών που βρίσκονται πέριξ αυτών. Ο αριθμός των έγκυρων παρατηρήσεων, άνω και κάτω των missing values, βάσει των οποίων θα γίνει ο υπολογισμός της διαμέσου, ορίζεται από την επιλογή Span of nearby points.

Linear interpolation. Αντικαθιστά τις missing values χρησιμοποιώντας τη μέθοδο της γραμμικής παρεμβολής. Για την εκτίμηση μιας missing value, η μέθοδος χρησιμοποιεί την τελευταία τιμή πριν την missing value και την πρώτη μετά την missing value. Αν η πρώτη και η τελευταία τιμή της μεταβλητής έχουν επίσης missing values τότε η missing value δεν αντικαθίσταται.

Linear trend at point. Αντικαθιστά τις missing values μιας μεταβλητής εκτιμώντας τες με τη βοήθεια ενός υποδείγματος γραμμικής παλινδρόμησης, με εξαρτημένη μεταβλητή την ίδια τη μεταβλητή και ανεξάρτητη τη θέση κάθε τιμής της μεταβλητής στο αρχείο (δηλαδή έναν δείκτη με τιμές από το 1 μέχρι το *n*, όπου *n* ο αριθμός των παρατηρήσεων του αρχείου). Οι missing values εκτιμώνται από το υπόδειγμα μέσω της θέσης που κατέχουν στο αρχείο (δηλαδή μέσω των τιμών του δείκτη).

Πρέπει να σημειωθεί ότι εκτός της πρώτης μεθόδου, όπου οι missing values υπολογίζονται ως η μέση τιμή όλων των έγκυρων παρατηρήσεων, οι υπόλοιπες, λόγω της φύσης του υπολογισμού τους, μπορούν να χρησιμοποιηθούν κυρίως σε περιπτώσεις χρονοσειρών (η μεταβλητή, δηλαδή, της οποίας υπολογίζονται οι missing values να είναι μια χρονοσειρά).

Αλγόριθμοι παραγωγής τυχαίων αριθμών - Random Number Generators

Η επιλογή Random Number Generators δίνει τη δυνατότητα στο χρήστη να επιλέγει τον αλγόριθμο (γεννήτρια) παραγωγής τυχαίων αριθμών που θα χρησιμοποιηθεί σε μια διαδικασία λήψης ενός τυχαίου δείγματος. Επίσης, δίνει τη δυνατότητα να δηλώνει ο χρήστης την τιμή εκκίνησης (starting point) του αλγόριθμου, κατά την παραγωγή της σειράς των τυχαίων αριθμών. Το SPSS διαθέτει δύο διαφορετικούς αλγόριθμους για το σκοπό αυτό:

SPSS 12 Compatible. Ο αλγόριθμος αυτός χρησιμοποιείτο από τις προγενέστερες εκδόσεις του SPSS μέχρι και την έκδοση 12. Αν ο χρήστης θέλει να αναπαραγάγει ένα τυχαίο δείγμα, ακριβώς όπως αυτό είχε ληφθεί από παλαιότερες εκδόσεις του λογισμικού, μπορεί, επιλέγοντας τον συγκεκριμένο αλγόριθμο και δίνοντας την τιμή εκκίνησής του (starting point ή seed value), να το αναπαραγάγει επακριβώς.

Mersenne Twister. Ένας νέος αλγόριθμος παραγωγής τυχαίων αριθμών είναι διαθέσιμος στο SPSS από την έκδοση 13 και μετά. Συνιστάται η χρήση αυτού του

αλγόριθμου, εφόσον δεν ενδιαφέρει το χρήστη η αναπαραγωγή αποτελεσμάτων που έχουν προκύψει από προηγούμενες εκδόσεις. Ορισμένες διαδικασίες του SPSS έχουν δικές τους γεννήτριες παραγωγής τυχαίων αριθμών.

Για να κάνουμε χρήση της επιλογής Random Number Generators εργαζόμαστε ως εξής:

▶ Από τη γραμμή των μενού επιλέγουμε:
Transform
Random Number Generators...

Ανοίγει το πλαίσιο διαλόγου Random Number Generators (Σχήμα 4.20).

▶ Από τις επιλογές Active Generator επιλέγουμε το είδος του αλγόριθμου που θα χρησιμοποιήσουμε, αν θα είναι δηλαδή ο SPSS 12 Compatible ή θα είναι ο Mersenne Twister.

▶ Από τις επιλογές Active Generator Initialization ορίζουμε την τιμή εκκίνησης (starting point) του αλγόριθμου που θα χρησιμοποιηθεί. Η τιμή αυτή αλλάζει κάθε φορά που πρόκειται να χρησιμοποιηθεί μια διαδικασία λήψης ενός τυχαίου δείγματος. Επιλέγοντας το πλήκτρο Fixed Value και ορίζοντας στο πεδίο Value την ακριβή τιμή εκκίνησης ενός αλγόριθμου, μπορούμε να αναπαραγάγουμε επακριβώς ένα τυχαίο δείγμα που είχε ληφθεί με προηγούμενη χρήση του ίδιου σημείου εκκίνησης.

Σχήμα 4.20 Πλαίσιο διαλόγου Random Number Generators

