

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ ΣΤΑΤΙΣΤΙΚΟΥ ΠΑΚΕΤΟΥ PSPP



Σύνταξη - Επιμέλεια

Ευστάθιος Αντωνίου
Αναπληρωτής καθηγητής

Βασίλης Κώστογλου
Καθηγητής

Ιανουάριος 2017

Περιεχόμενα

| | | |
|-------|---|----|
| 1 | ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ..... | 3 |
| 1.1 | Περιγραφή των τιμών μιας μεταβλητής..... | 3 |
| 1.1.1 | Ποιοτικές μεταβλητές..... | 3 |
| 1.1.2 | Ποσοτικές μεταβλητές..... | 6 |
| 1.2 | Ανάλυση σε υποσύνολα του δείγματος..... | 10 |
| 1.2.1 | Η Διαδικασία SPLIT FILE..... | 10 |
| 1.2.2 | Η διαδικασία SELECT CASES..... | 12 |
| 1.2.3 | Η Διαδικασία EXPLORE..... | 15 |
| 1.3 | Δημιουργία νέων μεταβλητών..... | 17 |
| 1.3.1 | Η διαδικασία RECODE..... | 17 |
| 1.3.2 | Η διαδικασία COMPUTE..... | 21 |
| | Περιγραφή / συσχέτιση των τιμών δύο μεταβλητών..... | 24 |
| 1.3.3 | Συσχέτιση μεταξύ δύο ποιοτικών μεταβλητών..... | 24 |
| 1.3.4 | Συσχέτιση μεταξύ δύο ποσοτικών μεταβλητών..... | 27 |
| 2 | ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ..... | 29 |
| 2.1 | Έλεγχος υποθέσεων για τη μέση τιμή ενός δείγματος..... | 29 |
| 2.2 | Έλεγχος υποθέσεων για τη διαφορά των μέσων τιμών δύο ανεξάρτητων πληθυσμών..... | 33 |
| 2.3 | Σύγκριση μέσων τιμών σε ζευγάρια παρατηρήσεων..... | 36 |
| 2.4 | Έλεγχος ανεξαρτησίας / ομοιογένειας μεταξύ δύο ποιοτικών μεταβλητών..... | 38 |
| 3 | ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ..... | 42 |
| 3.1 | Απλή παλινδρόμηση..... | 42 |
| 3.2 | Πολλαπλή παλινδρόμηση..... | 45 |
| 4 | ΠΑΡΑΡΤΗΜΑ - Εργαστηριακή Άσκηση Υπόδειγμα..... | 48 |

1 ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Μερικές χρήσιμες διαδικασίες για τη σύνοψη, ταξινόμηση και παρουσίαση των πρωτογενών δεδομένων ενός δείγματος σε εύληπτη μορφή. Τα παραδείγματα που παρουσιάζονται προέρχονται από την άσκηση υπόδειγμα που παρατίθεται στο παράρτημα.

1.1 Περιγραφή των τιμών μιας μεταβλητής

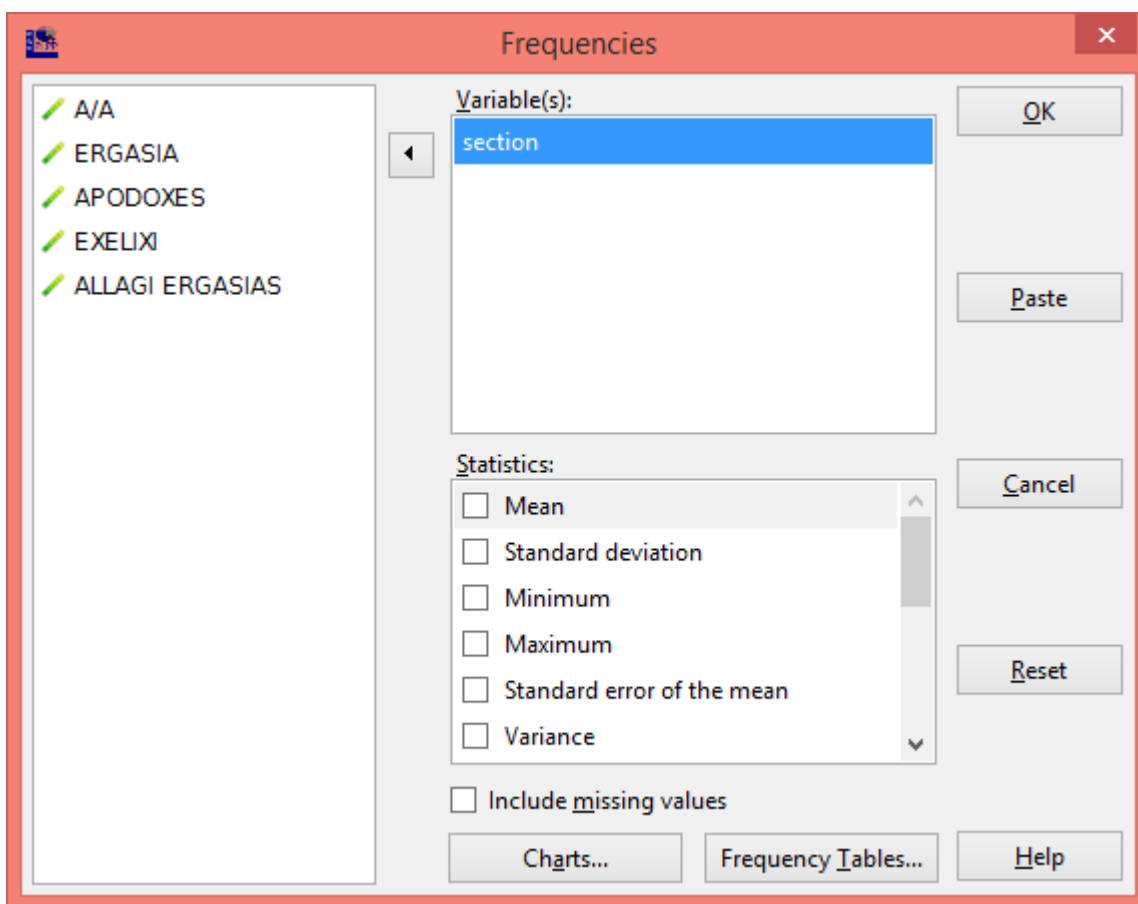
1.1.1 Ποιοτικές μεταβλητές

Οι μέθοδοι σύνοψης και παρουσίασης ποιοτικών δεδομένων περιορίζονται στους πίνακες συχνοτήτων και τις γραφικές παραστάσεις. Με τη διαδικασία **Frequencies** μπορούμε να επιτύχουμε άμεσα την κατασκευή τους.

Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας:

Analyze – Descriptive Statistics – Frequencies

εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Frequencies**:

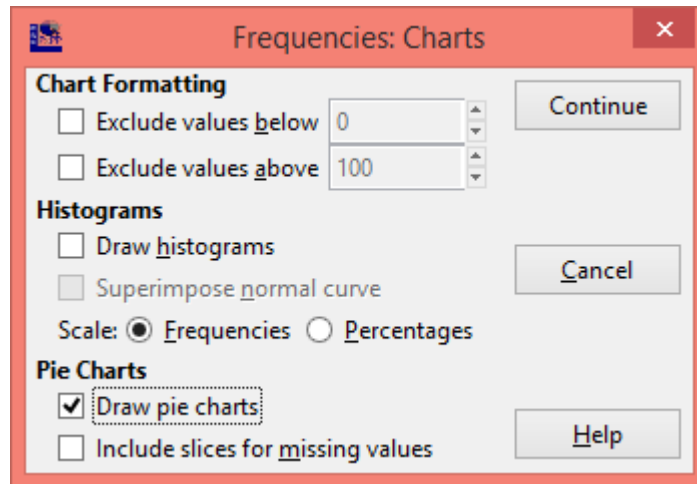


Επιλέγουμε τη μεταβλητή (ή τις μεταβλητές) που θέλουμε να περιγράψουμε και τη μετακινούμε (με τη χρήση του βέλους) στο πλαίσιο variable(s).

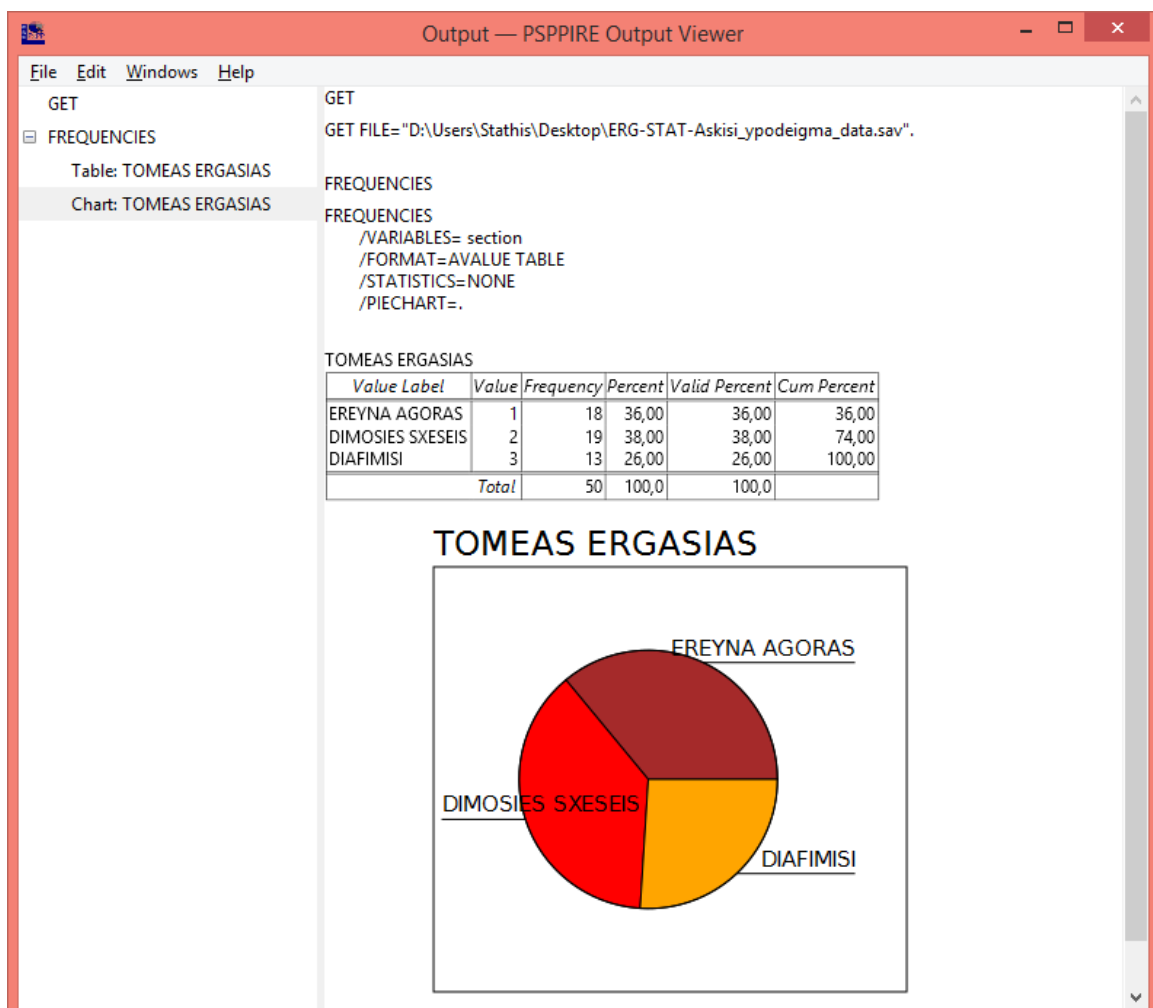
Παρατηρούμε ότι η επιλογές **Mean** (μέση τιμή), **Standard deviation** (τυπική απόκλιση), **Minimum** (ελάχιστο) και **Maximum** (μέγιστο) είναι εξ ορισμού ενεργοποιημένες στη λίστα

Statistics. Αν η πρόθεση μας είναι να αναλύσουμε ποιοτικές μεταβλητές, τα παραπάνω στατιστικά δεν έχουν νόημα, οπότε προτείνεται να απενεργοποιηθούν.

Αν επιθυμούμε την κατασκευή κάποιου γραφήματος θα πρέπει να ενεργοποιήσουμε την επιλογή **Charts...** που βρίσκεται στο ίδιο παράθυρο διαλόγου:



Το κατάλληλο διάγραμμα για την περιγραφή ποιοτικών δεδομένων είναι το κυκλικό διάγραμμα (**pie chart**). Επιλέγοντας το OK ή/και το Continue παίρνουμε τα αποτελέσματα της διαδικασίας σε ένα αρχείο αποτελεσμάτων, όπως το παρακάτω:



Ο πίνακας συχνοτήτων αποτελείται από 6 στήλες:

- Στην πρώτη στήλη εμφανίζονται οι ετικέτες των τιμών της μεταβλητής που αναλύεται.
- Στην δεύτερη εμφανίζονται οι ίδιες οι τιμές της μεταβλητής.
- Στην τρίτη και τέταρτη στήλη εμφανίζονται αντίστοιχα, οι συχνότητες (**Frequency**) εμφάνισης κάθε τιμής της μεταβλητής και τα ποσοστά (**Percent**). Τα ποσοστά αυτά είναι ποσοστά επί του συνολικού μεγέθους του δείγματος συμπεριλαμβανομένων των ελλειπουσών τιμών.
- Στην πέμπτη στήλη εμφανίζονται τα έγκυρα ποσοστά (**Valid Percent**), τα οποία είναι ποσοστά επί του συνολικού μεγέθους του δείγματος εξαιρουμένων των ελλειπουσών τιμών.
- Στην έκτη στήλη εμφανίζονται τα αθροιστικά ποσοστά (**Cum Percent**). Το ποσοστό που αντιστοιχεί σε μια τιμή της μεταβλητής είναι το συνολικό ποσοστό που αντιστοιχεί στην τιμή αυτή και σε όλες τις προηγούμενες τιμές.

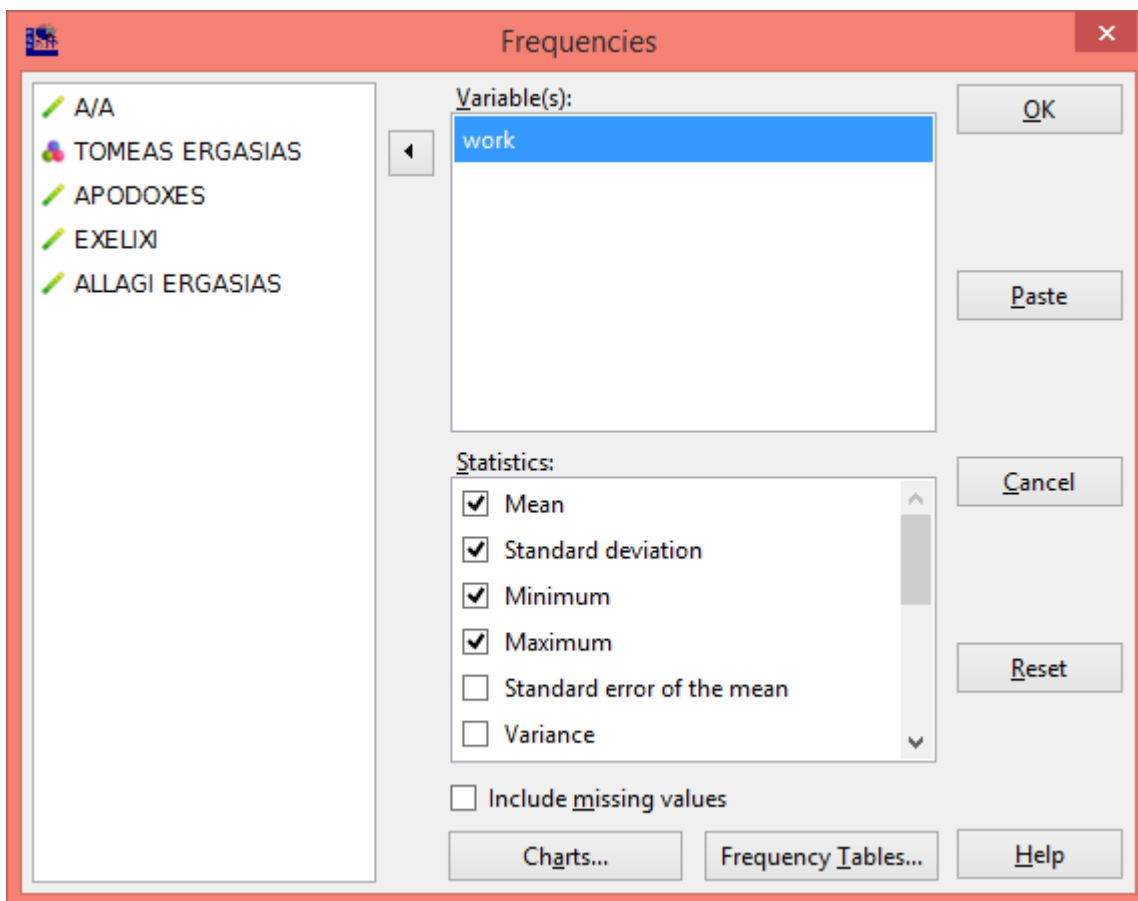
1.1.2 Ποσοτικές μεταβλητές

Με τη διαδικασία **Frequencies** μπορούμε να περιγράψουμε και ποσοτικά χαρακτηριστικά, υπολογίζοντας κατάλληλα στατιστικά μετρά (statistics) και κατασκευάζοντας ιστογράμματα (histogram) για την γραφική παρουσίασή τους.

Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας

Analyze – Descriptive Statistics – Frequencies

εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Frequencies**.



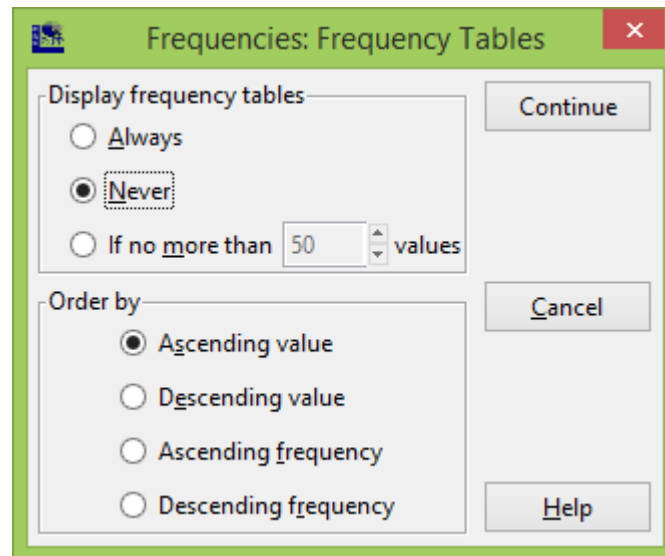
Επιλέγουμε τη μεταβλητή (ή τις μεταβλητές) που θέλουμε να περιγράψουμε και τη μετακινούμε (με τη χρήση του βέλους) στο πλαίσιο variable(s).

Για τον υπολογισμό στατιστικών μέτρων επιλέγουμε από τη λίστα **Statistics** που βρίσκεται στο ίδιο παράθυρο διαλόγου τα στατιστικά μέτρα που θέλουμε να υπολογιστούν.

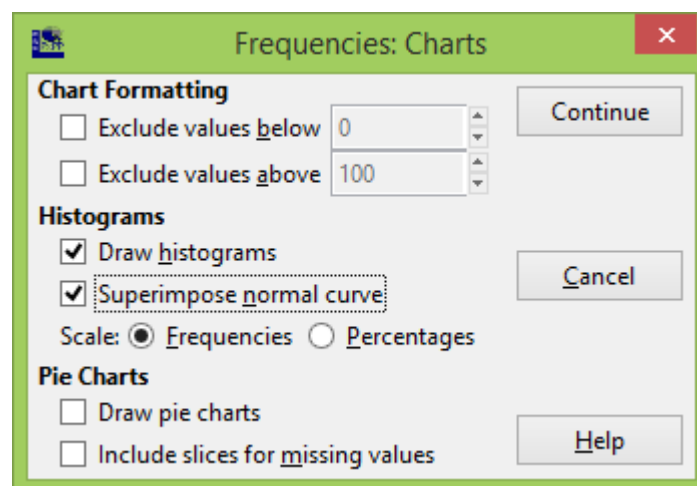
Μερικές Από Τις Επιλογές Στατιστικών Μέτρων

| Central Tendency – Κεντρική Τάση | Dispersion – Διασπορά | Distribution - Κατανομή |
|----------------------------------|--------------------------------------|-------------------------|
| Mean – Αριθμητικός Μέσος | Standard Deviation – Τυπική Απόκλιση | Skewness – Ασυμμετρία |
| Median – Διάμεσος | Variance – Διακύμανση | Kurtosis - Κύρτωση |
| Mode – Επικρατούσα Τιμή | Range – Εύρος | |
| Sum – Άθροισμα | | |

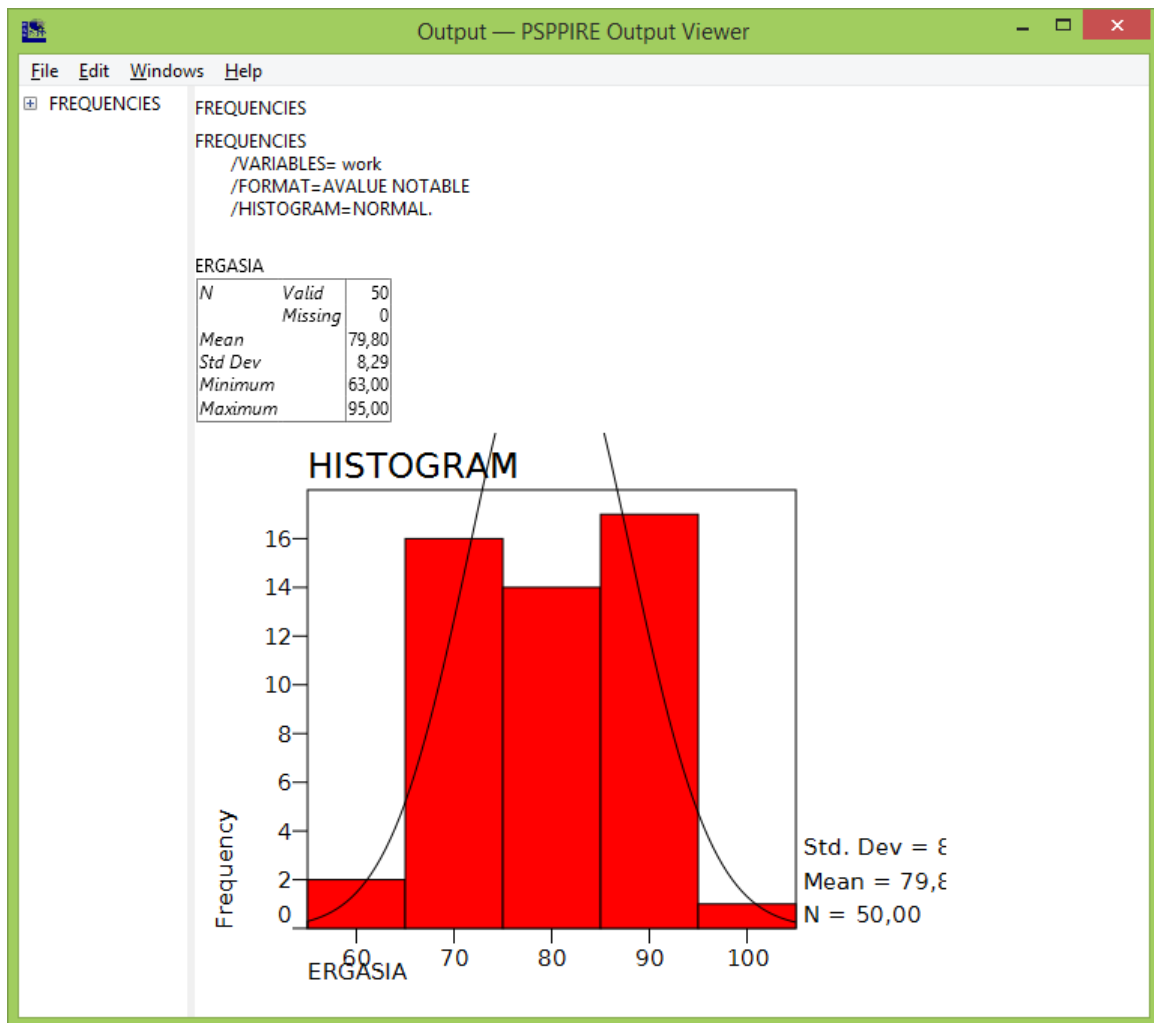
Στην περίπτωση που οι διαφορετικές τιμές που παίρνει η μεταβλητή είναι πολλές είναι σκόπιμο να αποφύγουμε την ανάλυση συχνοτήτων, καθώς το αποτέλεσμα της ανάλυσης δεν θα είναι εύκολα ερμηνεύσιμο. Για να αποφύγουμε τη δημιουργία του πίνακα συχνοτήτων επιλέγουμε το πλήκτρο **Frequency Tables...** και στο νέο παράθυρο διαλόγου που εμφανίζεται, επιλέγουμε **Never** στην ομάδα επιλογών **Display Frequency Tables**.



Αν επιθυμούμε την κατασκευή κάποιου γραφήματος θα πρέπει να ενεργοποιήσουμε την επιλογή **Charts...** που βρίσκεται στο ίδιο παράθυρο διαλόγου και να επιλέξουμε **Draw histograms**. Η επιλογή **Superimpose normal curve** (υπέρθηση της καμπύλης της κανονικής κατανομής) είναι προαιρετική.



Το αποτέλεσμα της παραπάνω διαδικασίας εμφανίζεται στο παράθυρο εξόδου το οποίο έχει την παρακάτω μορφή:



Στον πίνακα με τίτλο ERGASIA (ετικέτα της μεταβλητής work), εμφανίζονται διαδοχικά τα πλήθη των έγκυρων (Valid) και των ελλιπούσών (Missing) περιπτώσεων, η μέση τιμή (Mean), η τυπική απόκλιση (Std Dev), η ελάχιστη (Minimum) και μέγιστη τιμή (Maximum).

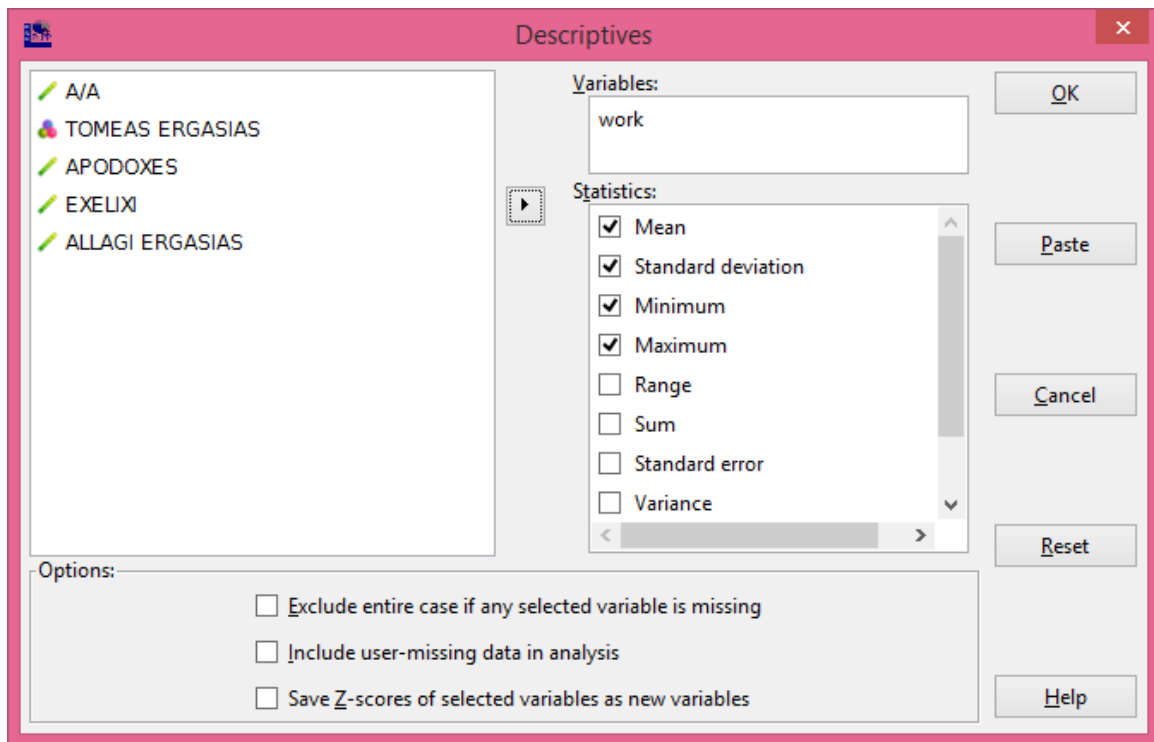
Στη συνέχεια ακολουθεί το ιστόγραμμα της μεταβλητής, όπου αυτόματα έχει γίνει διαχωρισμός των περιπτώσεων σε πέντε κλάσεις.

Εναλλακτικά, αν ενδιαφερόμαστε αποκλειστικά για τον υπολογισμό των στατιστικών μιας ποσοτικής μεταβλητής μπορούμε να χρησιμοποιήσουμε τη διαδικασία **Descriptives**.

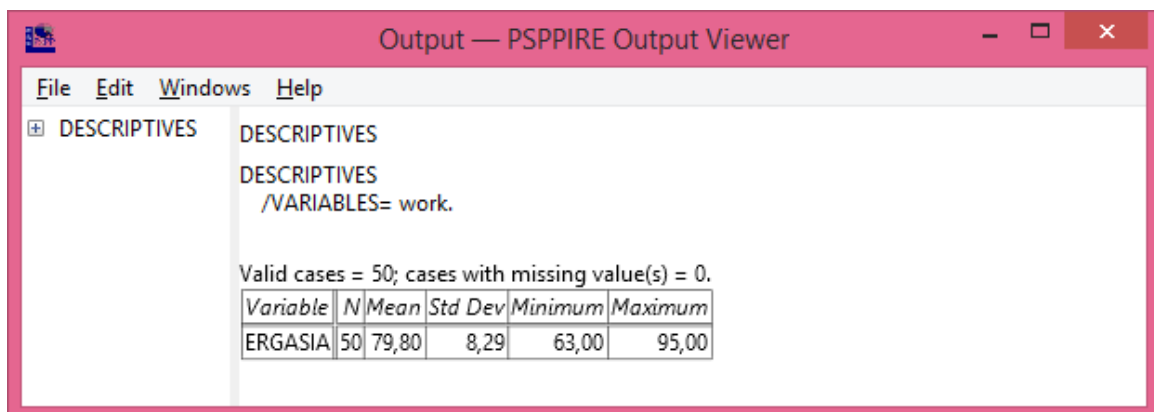
Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας

Analyze – Descriptive Statistics - Descriptives

εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Descriptives**:



Επιλέγοντας, την ή τις ποσοτικές μεταβλητές των οποίων τα στατιστικά επιθυμούμε να υπολογίσουμε καθώς και τα αντίστοιχα στατιστικά από την λίστα, πετυχαίνουμε το ακόλουθο αποτέλεσμα:



Στον πίνακα που προκύπτει εμφανίζονται τα επιλεγμένα από το χρήστη στατιστικά για την υπό ανάλυση μεταβλητή.

1.2 Ανάλυση σε υποσύνολα του δείγματος

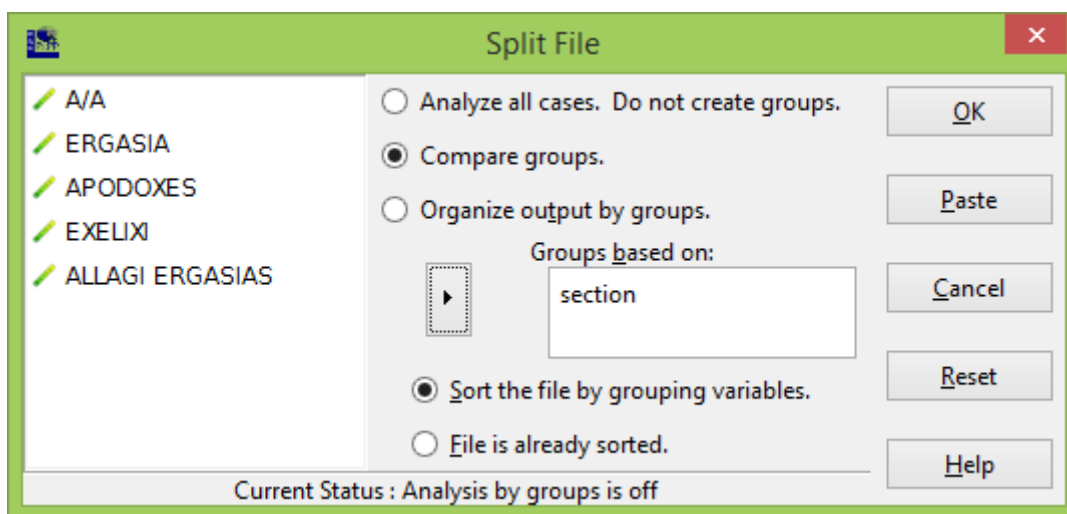
1.2.1 Η Διαδικασία SPLIT FILE

Στην περίπτωση που επιθυμούμε να περιγράψουμε μια μεταβλητή στις διάφορες κατηγορίες μιας μεταβλητής (π.χ. να περιγράψουμε το βαθμό ικανοποίησης των εργαζομένων μιας επιχείρησης από το μισθό τους, ξεχωριστά για κάθε τμήμα της εταιρείας), θα πρέπει πρώτα να διαχωρίσουμε το δείγμα μας με βάση τις κατηγορίες της μεταβλητής αυτής.

Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας

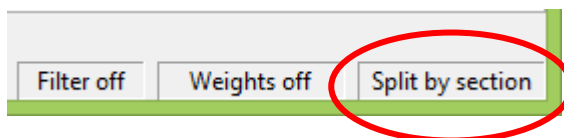
Data – Split File

εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Split File**:



Ενεργοποιούμε την επιλογή **Compare Groups** ή **Organize output by groups** (η διαφορά μεταξύ των δύο είναι μόνο ως προς την παρουσίαση των αποτελεσμάτων) και μετακινούμε στο πλαίσιο **Groups Based On** την μεταβλητή βάση της οποίας θα διαχωριστεί το δείγμα (π.χ. το section).

Επιλέγοντας το OK επανερχόμαστε στο αρχείο δεδομένων όπου οι περιπτώσεις (οι γραμμές) έχουν ταξινομηθεί με βάση τις τιμές της μεταβλητής που επιλέξαμε προηγουμένως. Επιπλέον, παρατηρούμε ότι στην κάτω αριστερή γωνία του παραθύρου της εφαρμογής PSPP, εμφανίζεται πλέον η ένδειξη “Split by section”.



Παρατήρηση: Η μεταβλητή που χρησιμοποιείται για τον διαχωρισμό του δείγματος είναι συνήθως ποιοτική. Στην περίπτωση που επιλεγεί μια ποσοτική μεταβλητή αυτή θα πρέπει να παίρνει λίγες διαφορετικές τιμές αλλιώς τα αποτελέσματα της ανάλυσης δεν θα είναι επαρκώς χρήσιμα.

Στη συνέχεια μπορούμε να περιγράψουμε οποιαδήποτε άλλη μεταβλητή με τη διαδικασία **Frequencies** όπως παρουσιάσαμε προηγουμένως. Για παράδειγμα, αν αναλύσουμε τη μεταβλητή changejob το αποτέλεσμα θα είναι όπως στο σχήμα που ακολουθεί:

Output — PSPPIRE Output Viewer

File Edit Windows Help

FREQUENCIES

FREQUENCIES
/VARIABLES= changejob
/FORMAT=AVALUE TABLE
/STATISTICS=NONE.

Variable Value Label
section 1 EREYNA AGORAS

ALLAGI ERGASIAS

| Value Label | Value | Frequency | Percent | Valid Percent | Cum Percent |
|---------------|-------|-----------|---------|---------------|-------------|
| No chance | 1 | 11 | 61,11 | 61,11 | 61,11 |
| Slight chance | 2 | 5 | 27,78 | 27,78 | 88,89 |
| High chance | 3 | 2 | 11,11 | 11,11 | 100,00 |
| <i>Total</i> | | 18 | 100,0 | 100,0 | |

Variable Value Label
section 2 DIMOSIES SXESEIS

ALLAGI ERGASIAS

| Value Label | Value | Frequency | Percent | Valid Percent | Cum Percent |
|---------------|-------|-----------|---------|---------------|-------------|
| No chance | 1 | 12 | 63,16 | 63,16 | 63,16 |
| Slight chance | 2 | 4 | 21,05 | 21,05 | 84,21 |
| High chance | 3 | 3 | 15,79 | 15,79 | 100,00 |
| <i>Total</i> | | 19 | 100,0 | 100,0 | |

Variable Value Label
section 3 DIAFIMISI

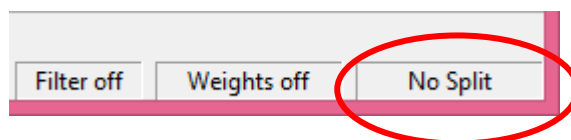
ALLAGI ERGASIAS

| Value Label | Value | Frequency | Percent | Valid Percent | Cum Percent |
|---------------|-------|-----------|---------|---------------|-------------|
| No chance | 1 | 7 | 53,85 | 53,85 | 53,85 |
| Slight chance | 2 | 3 | 23,08 | 23,08 | 76,92 |
| High chance | 3 | 3 | 23,08 | 23,08 | 100,00 |
| <i>Total</i> | | 13 | 100,0 | 100,0 | |

Αν επιθυμούμε να καταργήσουμε το διαχωρισμό των περιπτώσεων, επιλέγουμε και πάλι

Data – Split File

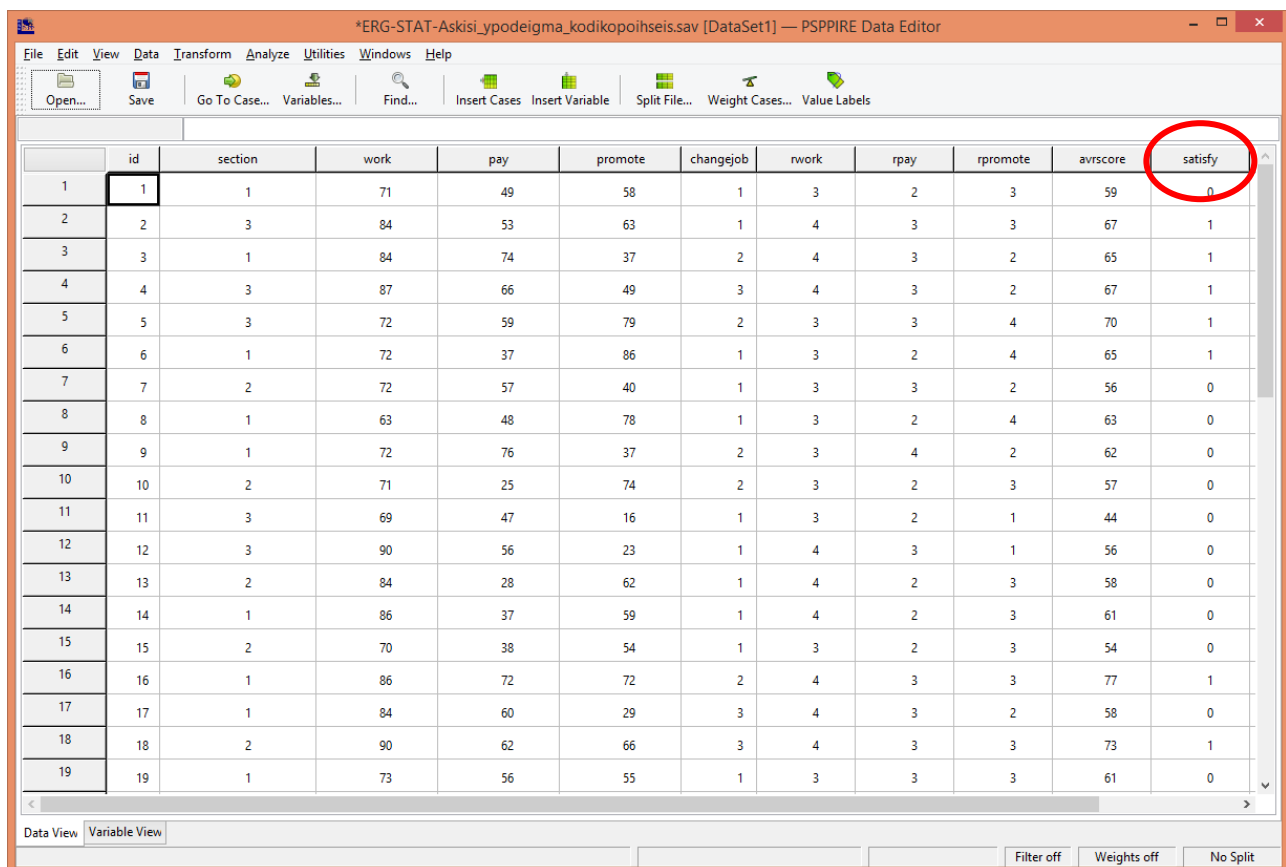
και στο παράθυρο διαλόγου που ακολουθεί επιλέγουμε **Analyze all cases. Do not create groups.** Η κατάργηση της ομαδοποίησης γίνεται ορατή παρατηρώντας την ένδειξη στην κάτω δεξιά γωνία του παραθύρου η οποία γίνεται πλέον:



1.2.2 Η διαδικασία SELECT CASES

Σε ορισμένες περιπτώσεις ζητείται να αναλυθεί ένα μέρος των δεδομένων του δείγματος που δίνεται «φιλτράροντας» το σύνολο των περιπτώσεων μας βάσει κάποιου συγκεκριμένου κριτηρίου. Για την επιλογή ενός συγκεκριμένου υποσυνόλου βάσει κάποιων κριτηρίων, πρέπει να υπάρχει ή να δημιουργηθεί (βλέπε παρακάτω τις διαδικασίες **RECODE** και **COMPUTE**) μια μεταβλητή η οποία παίρνει τιμές 0 ή 1. Οι τιμές 0 και 1 σηματοδοτούν αντίστοιχα τον αποκλεισμό ή τη συμμετοχή της αντίστοιχης περίπτωσης στην επιλογή.

Για παράδειγμα στο στιγμιότυπο που ακολουθεί έχει δημιουργηθεί η μεταβλητή **satisfy**, η οποία παίρνει την τιμή 1 αν η αντίστοιχος εργαζόμενος έχει μέση βαθμολογία (**avrscore**) μεγαλύτερη ή ίσο το 65, οπότε θεωρείται ικανοποιημένος, ενώ παίρνει την τιμή 0 στην αντίθετη περίπτωση:



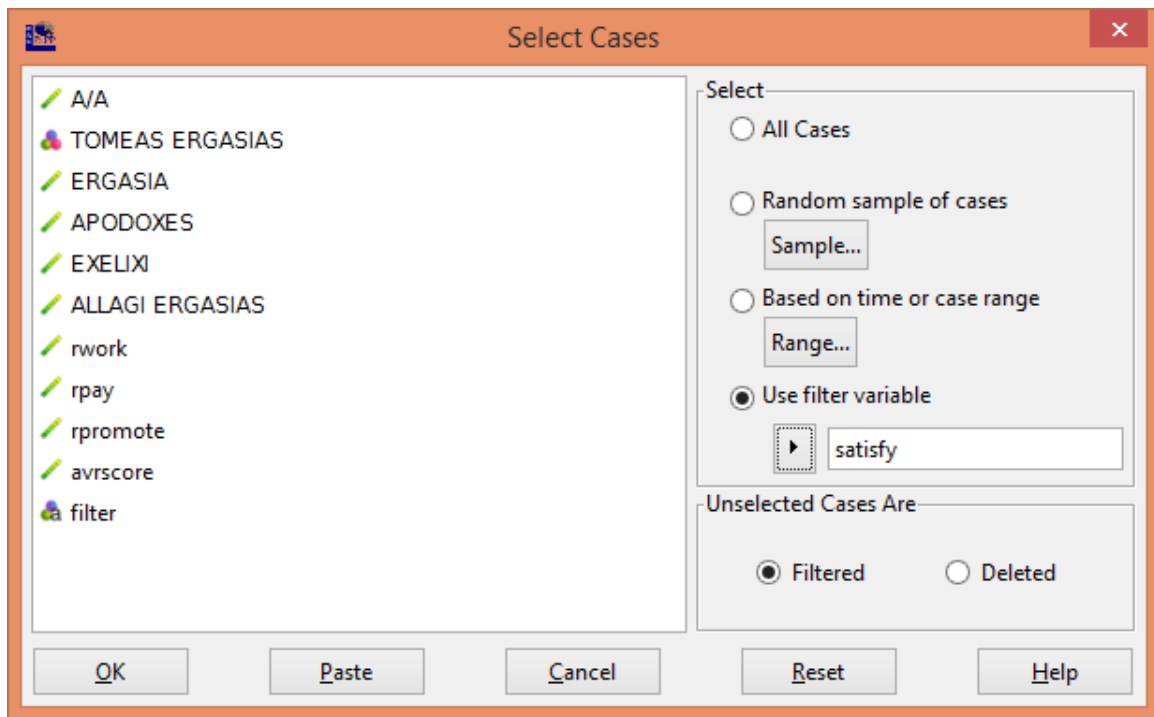
The screenshot shows the SPSS Data Editor window with the following data:

| | id | section | work | pay | promote | changejob | rwork | rpay | rpromote | avrscore | satisfy |
|----|----|---------|------|-----|---------|-----------|-------|------|----------|----------|---------|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | 3 | 2 | 3 | 59 | 0 |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | 4 | 3 | 3 | 67 | 1 |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | 4 | 3 | 2 | 65 | 1 |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | 4 | 3 | 2 | 67 | 1 |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | 3 | 3 | 4 | 70 | 1 |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | 3 | 2 | 4 | 65 | 1 |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | 3 | 3 | 2 | 56 | 0 |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | 3 | 2 | 4 | 63 | 0 |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | 3 | 4 | 2 | 62 | 0 |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | 3 | 2 | 3 | 57 | 0 |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | 3 | 2 | 1 | 44 | 0 |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | 4 | 3 | 1 | 56 | 0 |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | 4 | 2 | 3 | 58 | 0 |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | 4 | 2 | 3 | 61 | 0 |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | 3 | 2 | 3 | 54 | 0 |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | 4 | 3 | 3 | 77 | 1 |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | 4 | 3 | 2 | 58 | 0 |
| 18 | 18 | 2 | 90 | 62 | 66 | 3 | 4 | 3 | 3 | 73 | 1 |
| 19 | 19 | 1 | 73 | 56 | 55 | 1 | 3 | 3 | 3 | 61 | 0 |

Έχοντας στη διάθεση μας μια τέτοια μεταβλητή, επιλέγουμε από τη βασική ράβδο προτιμήσεων του λογισμικού:

Data – Select Cases

οπότε εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Select Cases**:

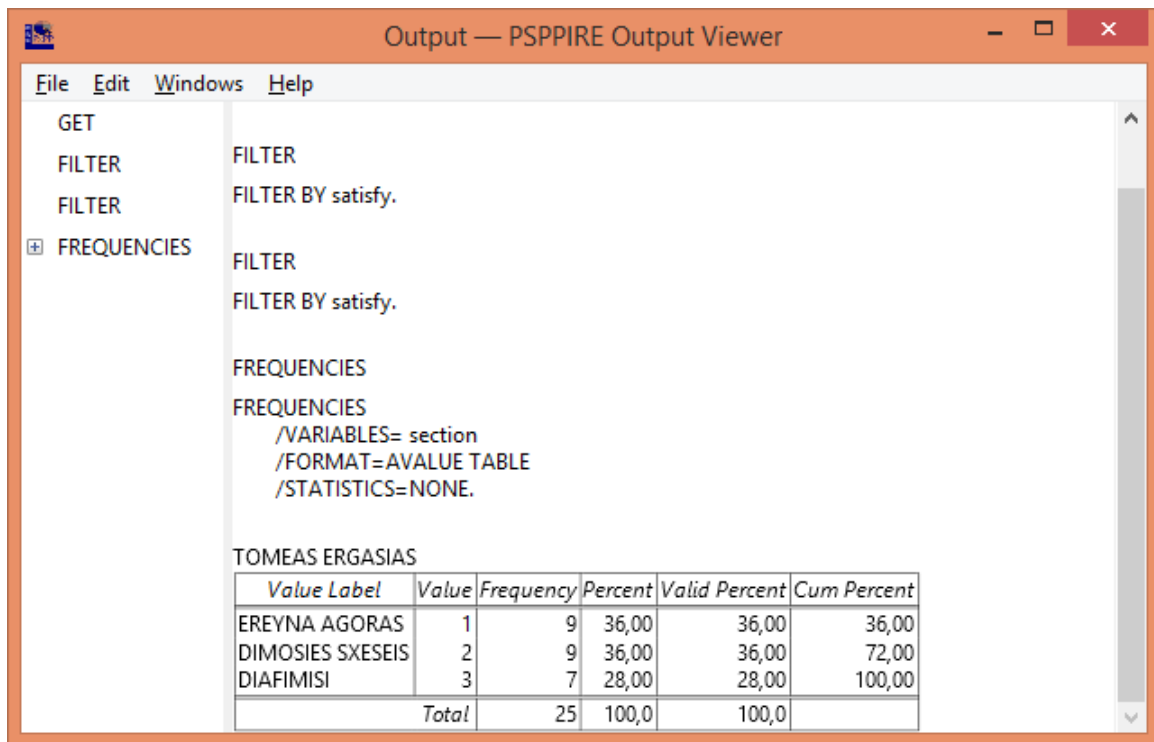


Στο πλαίσιο αυτό επιλέγουμε τη μεταβλητή **AsiaOrAfrica** που επιθυμούμε να λειτουργήσει σαν φίλτρο για τα δεδομένα μας, οπότε το αποτέλεσμα στην κεντρική οθόνη του PSPP θα είναι:

| | id | section | work | pay | promote | changejob | rwork | rpay | rpromote | avrscore | satisfy |
|----|----|---------|------|-----|---------|-----------|-------|------|----------|----------|---------|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | 3 | 2 | 3 | 59 | 0 |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | 4 | 3 | 3 | 67 | 1 |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | 4 | 3 | 2 | 65 | 1 |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | 4 | 3 | 2 | 67 | 1 |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | 3 | 3 | 4 | 70 | 1 |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | 3 | 2 | 4 | 65 | 1 |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | 3 | 3 | 2 | 56 | 0 |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | 3 | 2 | 4 | 63 | 0 |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | 3 | 4 | 2 | 62 | 0 |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | 3 | 2 | 3 | 57 | 0 |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | 3 | 2 | 1 | 44 | 0 |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | 4 | 3 | 1 | 56 | 0 |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | 4 | 2 | 3 | 58 | 0 |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | 4 | 2 | 3 | 61 | 0 |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | 3 | 2 | 3 | 54 | 0 |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | 4 | 3 | 3 | 77 | 1 |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | 4 | 3 | 2 | 58 | 0 |
| 18 | 18 | 2 | 90 | 62 | 66 | 3 | 4 | 3 | 3 | 73 | 1 |
| 19 | 19 | 1 | 73 | 56 | 55 | 1 | 3 | 3 | 3 | 61 | 0 |

Η ενεργοποίηση του φίλτρου μπορεί να γίνει αντιληπτή στο χρήστη από τη στήλη της αρίθμησης των περιπτώσεων, όπου οι περιπτώσεις που δεν έχουν επιλεγεί εμφανίζονται διαγραμμένες με μια διαγώνια γραμμή και από την ένδειξη **“Filter by satisfy”** στη μπάρα κατάστασης στην κάτω δεξιά γωνία του παραθύρου της εφαρμογής.

Οποιαδήποτε ανάλυση πραγματοποιηθεί στη συνέχεια θα αγνοήσει τα δεδομένα που δεν έχουν επιλεγεί. Για παράδειγμα μπορούμε να δημιουργήσουμε ένα πίνακα συχνοτήτων για τις τομείς εργασίας των εργαζομένων οι οποίοι θεωρούνται ικανοποιημένοι από την απασχόλησή τους στην εταιρεία:



The screenshot shows the SPSS Output Viewer window. The left pane lists several 'FREQUENCIES' operations. The right pane displays the output for the last operation, which is a frequency table for the variable 'TOMEAS ERGASIAS'. The table includes columns for Value Label, Value, Frequency, Percent, Valid Percent, and Cum Percent. The data is as follows:

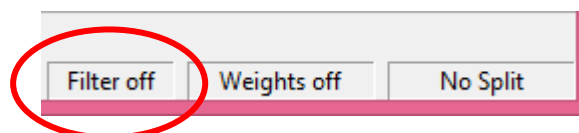
| Value Label | Value | Frequency | Percent | Valid Percent | Cum Percent |
|------------------|-------|-----------|---------|---------------|-------------|
| EREYNA AGORAS | 1 | 9 | 36,00 | 36,00 | 36,00 |
| DIMOSIES SXESEIS | 2 | 9 | 36,00 | 36,00 | 72,00 |
| DIAFIMISI | 3 | 7 | 28,00 | 28,00 | 100,00 |
| Total | | 25 | 100,0 | 100,0 | |

Παρατηρείστε ότι η μέτρηση της σχετική συχνότητας γίνεται επί του συνόλου των 25 επιλεγμένων ικανοποιημένων εργαζομένων και όχι επί των 50 εργαζομένων που περιέχονται στο αρχείο.

Αν επιθυμούμε να καταργήσουμε το φίλτρο που δημιουργήσαμε, επιλέγουμε και πάλι

Data – Select Cases

και στο παράθυρο διαλόγου που ακολουθεί επιλέγουμε **Select - All cases**. Η κατάργηση του φίλτρου γίνεται ορατή παρατηρώντας την ένδειξη στην κάτω δεξιά γωνία του παραθύρου η οποία γίνεται πλέον:



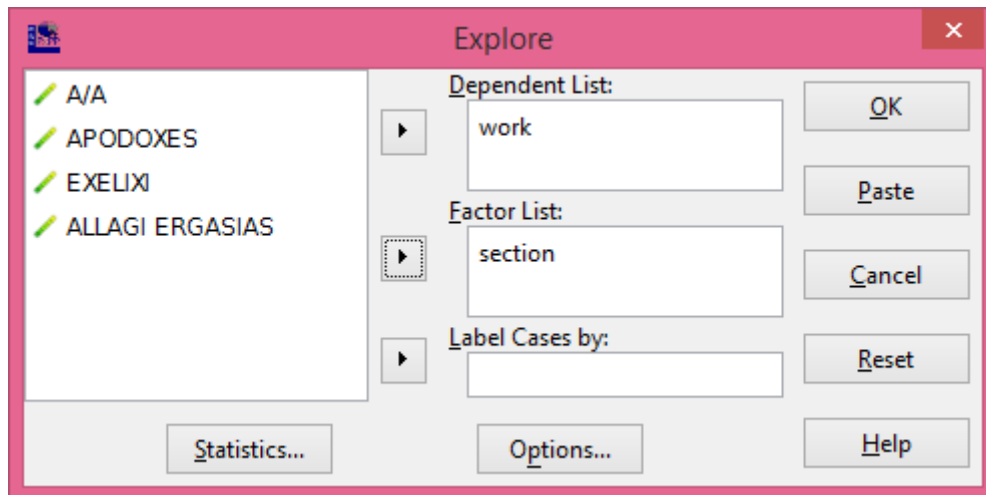
1.2.3 Η Διαδικασία EXPLORE

Με τη διαδικασία **Explore** μπορούμε να επιτύχουμε την πιο πλούσια και πλήρη περιγραφική στατιστική των παρατηρήσεων μιας ποσοτικής μεταβλητής στις διάφορες κατηγορίες κάποιας ποιοτικής.

Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας

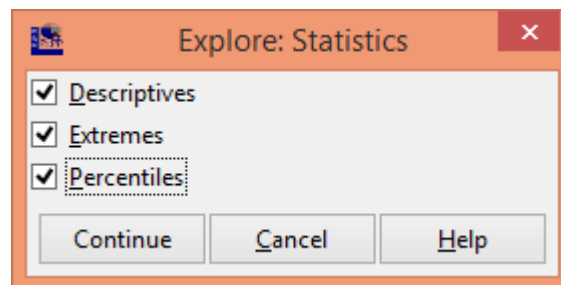
Analyze – Descriptive Statistics – Explore

εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Explore**.



Μετακινούμε την ποσοτική μεταβλητή (work) που θέλουμε να περιγράψουμε στο πλαίσιο **Dependent List** και την ποιοτική μεταβλητή (section) στο πλαίσιο **Factor List**.

Στη συνέχεια επιλέγουμε το πλήκτρο **Statistics...** και στο νέο παράθυρο που αναδύεται επιλέγουμε το είδος της ανάλυσης που θα πραγματοποιηθεί:



Οι τρεις επιλογές αντιστοιχούν στις παρακάτω αναλύσεις:

- **Descriptives** – Υπολογισμός Στατιστικών για τη μεταβλητή (μέση τιμή, τυπική απόκλιση, μέγιστο, ελάχιστο, κ.λπ.),
- **Extremes** – Υπολογισμός των πέντε μεγαλύτερων και μικρότερων τιμών, καθώς και της θέσης τους στο δείγμα.
- **Percentiles** – Υπολογισμός των ποσοστημορίων.

Όλοι οι παραπάνω υπολογισμοί πραγματοποιούνται τόσο για ολόκληρο το δείγμα, όσο και ξεχωριστά για κάθε τιμή της ποιοτικής μεταβλητής, την οποία προσθέσαμε στη λίστα **Factor List**.

Το αποτέλεσμα της παραπάνω διαδικασίας είναι η δημιουργία της εξόδου που ακολουθεί:

Output — PSPPIRE Output Viewer

File Edit Windows Help

EXAMINE

Table: Case Processing Summary
Table: Extreme Values
Table: Percentiles
Table: Descriptives
Table: Case Processing Summary
Table: Extreme Values
Table: Percentiles
Table: Descriptives

EXAMINE
/VARIABLES= work
BY section
/STATISTICS = DESCRIPTIVES EXTREME
/PERCENTILES
/MISSING=LISTWISE.

Case Processing Summary

| | Cases | | | | | |
|---------|-------|---------|---------|---------|-------|---------|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| ERGASIA | 50 | 100% | 0 | 0% | 50 | 100% |

Extreme Values

| | Case Number | Value |
|-------------------|-------------|-------|
| ERGASIA Highest 1 | 13 | 95 |
| | 2 | 10 |
| | 3 | 42 |
| | 4 | 23 |
| | 5 | 17 |
| Lowest 1 | 4 | 63 |
| | 2 | 30 |
| | 3 | 14 |
| | 4 | 41 |
| | 5 | 22 |

Percentiles

| | | Percentiles | | | | | | |
|---------|----------------|-------------|-------|-------|-------|-------|-------|-------|
| | | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| ERGASIA | HAverage | 64,55 | 70,10 | 72,00 | 82,00 | 87,00 | 90,00 | 91,80 |
| | Tukey's Hinges | | | 72,00 | 82,00 | 87,00 | | |

Descriptives

| | Statistic | Std. Error |
|----------------------------------|-------------|------------|
| ERGASIA Mean | 79,80 | 1,17 |
| 95% Confidence Interval for Mean | Lower Bound | 77,44 |
| | Upper Bound | 82,16 |
| 5% Trimmed Mean | 79,92 | |
| Median | 82,00 | |
| Variance | 68,69 | |
| Std. Deviation | 8,29 | |
| Minimum | 63,00 | |
| Maximum | 95,00 | |
| Range | 32,00 | |
| Interquartile Range | 15,00 | |
| Skewness | -,17 | ,34 |
| Kurtosis | -1,05 | ,66 |

Case Processing Summary

| | Cases | | | | | |
|-----------------|-------|---------|---------|---------|-------|---------|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| TOMEAS ERGASIAS | | | | | | |

Σημαντικός είναι ο υπολογισμός **διαστημάτων εμπιστοσύνης** για το μέσο ενός πληθυσμού (**95% confidence interval for mean (lower bound, upper bound)**) που παρουσιάζεται μαζί με τα στατιστικά μέτρα.

1.3 Δημιουργία νέων μεταβλητών

1.3.1 Η διαδικασία RECODE

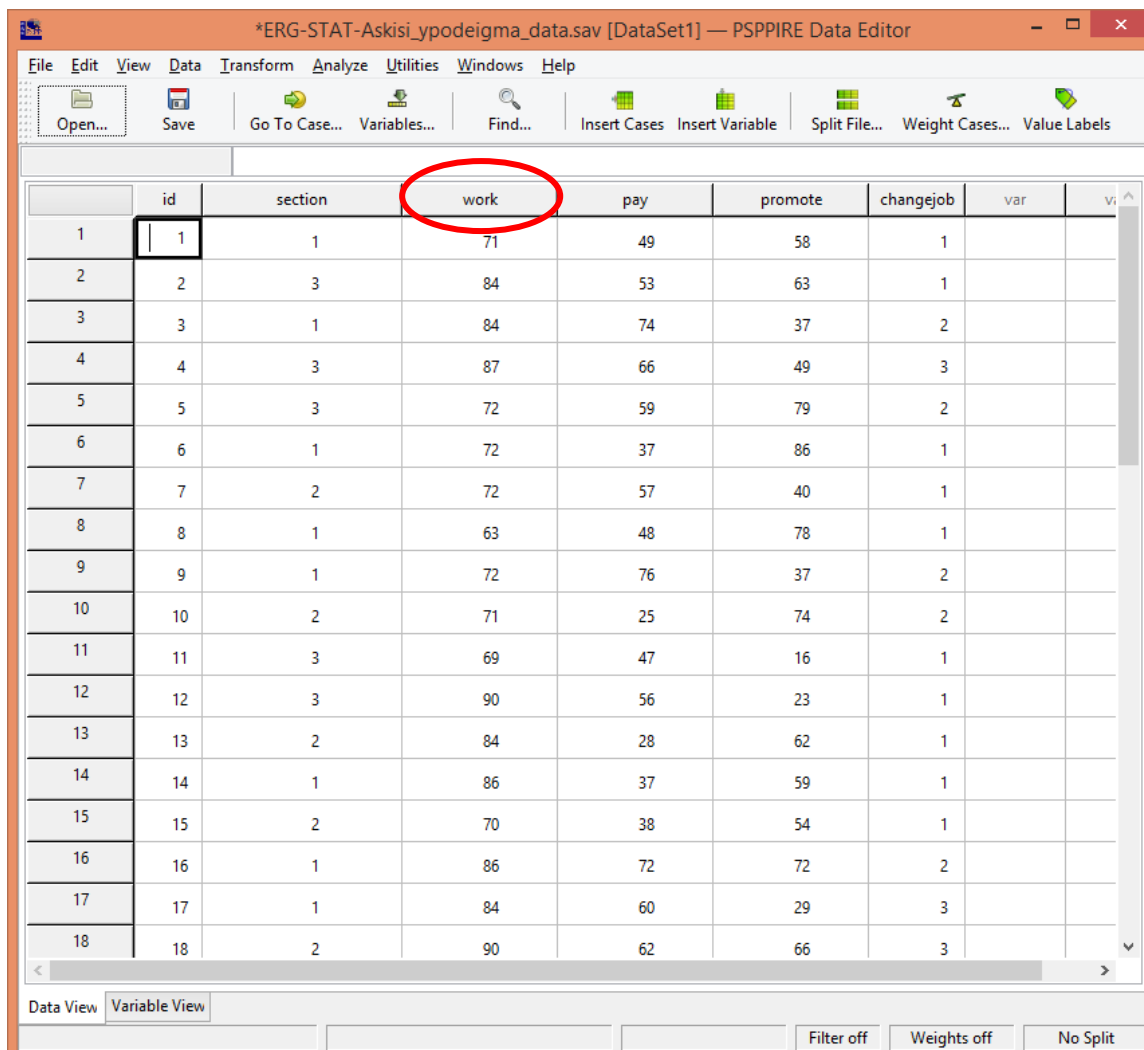
Σε αρκετές περιπτώσεις απαιτείται η δημιουργία νέων μεταβλητών χρησιμοποιώντας τη διαδικασία επανακωδικοποίησης, βάσει κάποιων διακριτών κανόνων. Η διαδικασία αυτή είναι διαθέσιμη σε δύο παραλλαγές στο PSPP. Αν επιθυμούμε η νέα μεταβλητή που θα προκύψει να αντικαταστήσει την αρχική, επιλέγουμε από τη βασική ράβδο προτιμήσεων του λογισμικού:

Transform – Recode into Same Variable...

ενώ αν θέλουμε να δημιουργήσουμε μια νέα μεταβλητή διατηρώντας την αρχική ανέπαφη, επιλέγουμε:

Transform – Recode into Different Variable...

Επειδή η δεύτερη διαδικασία εμπεριέχει και την πρώτη, θα περιγράψουμε τη δεύτερη. Στο παράδειγμα που εμφανίζεται παρακάτω η μεταβλητή **work** περιέχει την αξιολόγηση των συνθηκών εργασίας κάθε εργαζομένου στην κλίμακα 0-100:

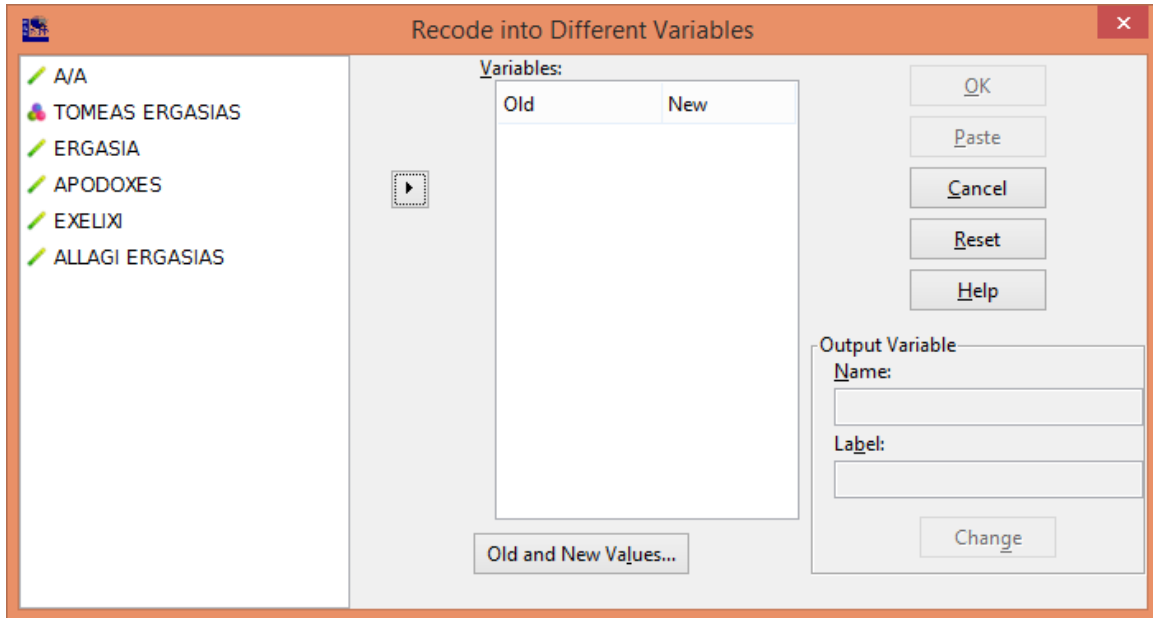


| | id | section | work | pay | promote | changejob | var | vi |
|----|----|---------|------|-----|---------|-----------|-----|----|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | | |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | | |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | | |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | | |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | | |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | | |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | | |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | | |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | | |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | | |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | | |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | | |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | | |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | | |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | | |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | | |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | | |
| 18 | 18 | 2 | 90 | 62 | 66 | 3 | | |

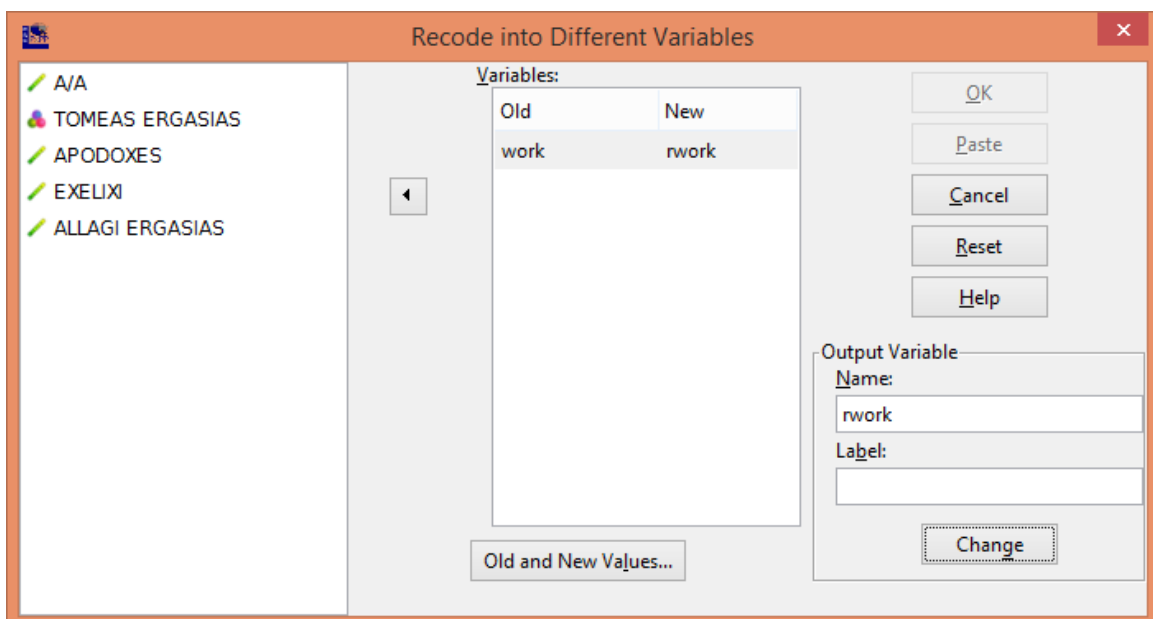
Θέλουμε να δημιουργήσουμε μια νέα μεταβλητή, την οποία ονομάζουμε **rwork**, η οποία θα επανακωδικοποιεί την αρχική μεταβλητή **work** σύμφωνα με τον παρακάτω κανόνα:

- $rwork = 1$ για $work < 25$
- $rwork = 2$ για $25 \leq work < 50$
- $rwork = 3$ για $50 \leq Population < 75$
- $rwork = 4$ για $75 \leq Population$

Επιλέγουμε: **Transform – Recode into Different Variable...**, οπότε εμφανίζεται το παράθυρο διαλόγου **Recode into Different Variable**.



Επιλέγουμε από την αριστερή στήλη τη μεταβλητή βάσει της οποίας θα γίνει η επανακωδικοποίηση και τη μεταφέρουμε χρησιμοποιώντας το πλήκτρο με το βελάκι στη στήλη Old της κεντρικού πλαισίου Variables. Επιλέγουμε τη γραμμή που μεταφέρθηκε στο πλαίσιο Variables και στη συνέχεια πληκτρολογούμε το όνομα της νέας μεταβλητής που θέλουμε να δημιουργήσουμε στο πλαίσιο Output Variable – Name. Για να γίνει τελικά η αντιστοίχιση πιέζουμε το πλήκτρο Change στο ίδιο πλαίσιο. Ολοκληρώνοντας την αντιστοίχιση έχουμε την παρακάτω εικόνα:



Για να ολοκληρωθεί η διαδικασία της αντιστοίχισης των τιμών των δύο μεταβλητών, επιλέγουμε το πλήκτρο **Old and New Values...**. Στο πλαίσιο διαλόγου που εμφανίζεται δημιουργούμε τη ζητούμενη αντιστοίχιση των τιμών:

Recode into Different Variables: Old and New Values

Old Value

Value:

System Missing

System or User Missing

Range:

through

Range, LOWEST thru value

Range, value thru HIGHEST

All other values

New Value

Value: 4

System Missing

Copy old values

Add

Edit

Remove

| Old | New |
|--------------|-----|
| LOWEST - 24 | 1 |
| 25 - 49 | 2 |
| 50 - 74 | 3 |
| 75 - HIGHEST | 4 |

Output variables are strings Width: 0

Convert numeric strings to numbers ('5' -> 5)

Continue Cancel Help

Η τελική αντιστοίχιση εμφανίζεται στη λίστα κανόνων στο δεξιό μέρος του πλαισίου. Για τη δημιουργία των κανόνων επιλέγουμε μια από τις επτά επιλογές στο πλαίσιο Old Value, συμπληρώνουμε την επιθυμητή νέα τιμή στο πεδίο Value, στο πλαίσιο New Value και πιέζουμε το πλήκτρο Add.

Στο συγκεκριμένο παράδειγμα για τη δημιουργία των κανόνων αντιστοίχισης χρησιμοποιήσαμε τις παρακάτω επιλογές:

- Για να αντιστοιχίσουμε τις περιπτώσεις με $work < 25$ στην τιμή 1, επιλέξαμε στο πλαίσιο Old Value, Range, LOWEST thru value και συμπληρώσαμε την τιμή 24, ενώ στο πεδίο Value του πλαισίου New Value συμπληρώσαμε την τιμή 1.
- Για να αντιστοιχίσουμε τις περιπτώσεις με $25 \leq work < 50$ στην τιμή 2, επιλέξαμε στο πλαίσιο Old Value, Range και συμπληρώσαμε τις τιμές 25 και 49, ενώ στο πεδίο Value του πλαισίου New Value συμπληρώσαμε την τιμή 2.
- Για να αντιστοιχίσουμε τις περιπτώσεις με $50 \leq work < 75$ στην τιμή 3, επιλέξαμε στο πλαίσιο Old Value, Range και συμπληρώσαμε τις τιμές 50 και 74, ενώ στο πεδίο Value του πλαισίου New Value συμπληρώσαμε την τιμή 3.
- Για να αντιστοιχίσουμε τις περιπτώσεις με $75 \leq work$ στην τιμή 4, επιλέξαμε στο πλαίσιο Old Value, Range, value thru HIGHEST και συμπληρώσαμε την τιμή 75, ενώ στο πεδίο Value του πλαισίου New Value συμπληρώσαμε την τιμή 4.

Έχοντας ολοκληρώσει την παραπάνω αντιστοίχιση, επιλέγουμε Continue και στο προηγούμενο πλαίσιο διαλόγου το πλήκτρο OK. Το αποτέλεσμα είναι η δημιουργία της νέας μεταβλητής **rwork** όπως φαίνεται στο στιγμιότυπο που ακολουθεί:

*ERG-STAT-Askisi_yodeigma_data.sav [DataSet1] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

Open... Save Go To Case... Variables... Find... Insert Cases Insert Variable Split File... Weight Cases... Value Labels

| | id | section | work | pay | promote | changejob | rwork |
|----|----|---------|------|-----|---------|-----------|-------|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | 3,00 |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | 4,00 |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | 4,00 |
| 4 | 4 | 3 | 87 | 66 | 49 | | 4,00 |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | 3,00 |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | 3,00 |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | 3,00 |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | 3,00 |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | 3,00 |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | 3,00 |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | 3,00 |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | 4,00 |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | 4,00 |
| 14 | 14 | 1 | 86 | 37 | 59 | | 4,00 |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | 3,00 |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | 4,00 |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | 4,00 |
| 18 | 18 | 2 | 90 | 62 | 66 | 3 | 4,00 |

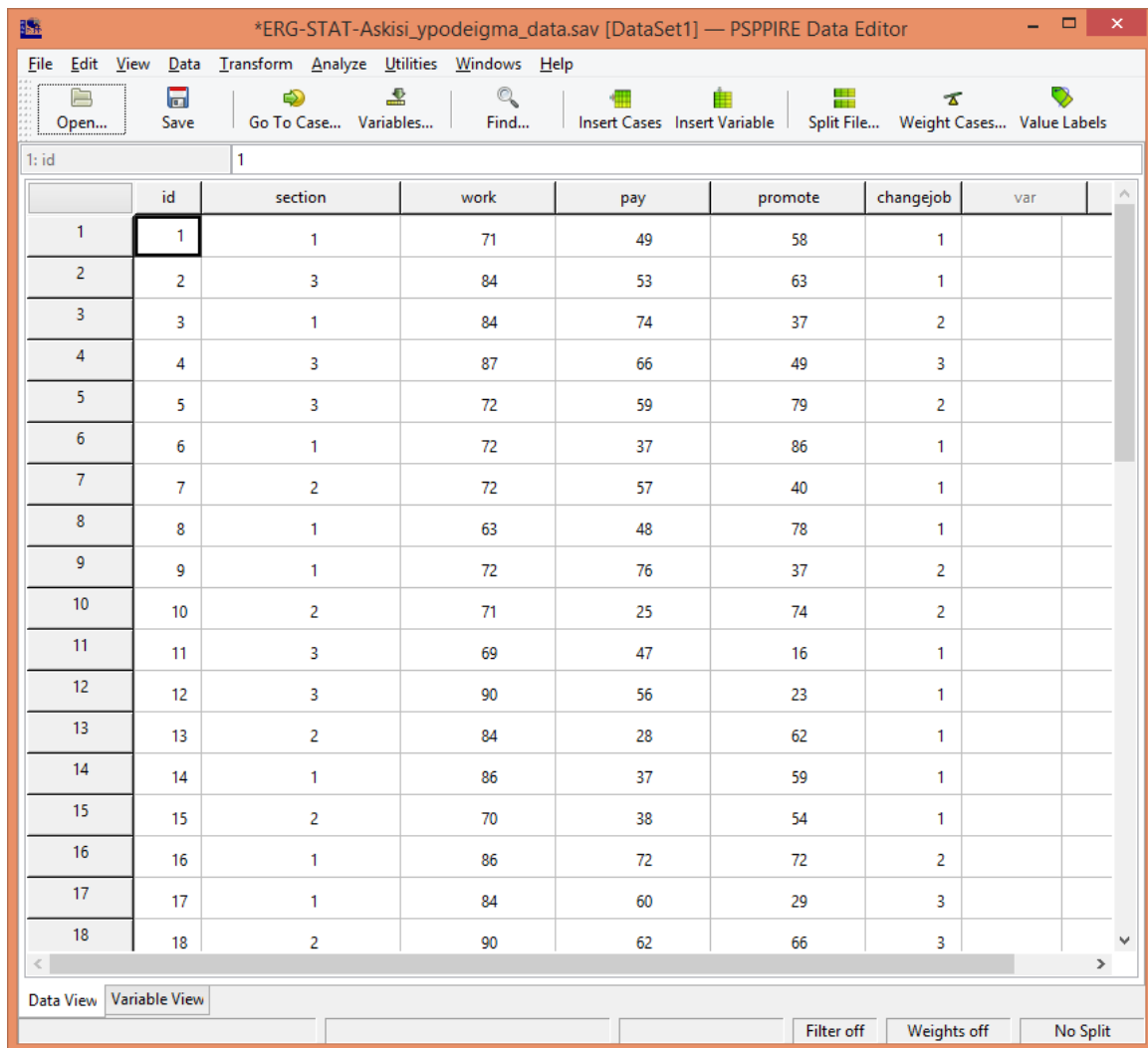
Data View Variable View

Filter off Weights off No Split

1.3.2 Η διαδικασία COMPUTE

Σε ορισμένες περιπτώσεις το ζητούμενο είναι να δημιουργηθεί μια νέα μεταβλητή η οποία προκύπτει ως αποτέλεσμα της εφαρμογής μια συνάρτησης σε μια ή περισσότερες από τις υπάρχουσες μεταβλητές. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας της διαδικασία **COMPUTE**.

Στο παράδειγμα που εμφανίζεται στο παρακάτω στιγμιότυπο, οι μεταβλητές **work**, **pay** και **promote**, εκφράζουν αντίστοιχα την αξιολόγηση που έδωσε ο κάθε ερωτώμενος σχετικά με τις συνθήκες εργασίας, τις αποδοχές και τις δυνατότητες εξέλιξης στην κλίμακα 0-100.



The screenshot shows the SPSS Data Editor window for a dataset named '*ERG-STAT-Askisi_ypodeigma_data.sav'. The main window displays a data grid with 18 rows and 8 columns. The columns are labeled 'id', 'section', 'work', 'pay', 'promote', 'changejob', and 'var'. The 'id' column contains values from 1 to 18. The 'section' column contains values 1, 3, 1, 3, 3, 1, 2, 1, 1, 2, 3, 3, 2, 1, 2, 1, 1, 2. The 'work' column contains values 71, 84, 84, 87, 72, 72, 72, 63, 72, 71, 69, 90, 84, 86, 70, 86, 84, 90. The 'pay' column contains values 49, 53, 74, 66, 59, 37, 57, 48, 76, 25, 47, 56, 28, 37, 38, 72, 60, 62. The 'promote' column contains values 58, 63, 37, 49, 79, 86, 40, 78, 37, 74, 16, 23, 62, 59, 54, 72, 29, 66. The 'changejob' column contains values 1, 1, 2, 3, 2, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 2, 3, 3. The 'var' column is empty. The window also shows a menu bar with 'File', 'Edit', 'View', 'Data', 'Transform', 'Analyze', 'Utilities', 'Windows', and 'Help'. Below the menu bar is a toolbar with icons for 'Open...', 'Save', 'Go To Case...', 'Variables...', 'Find...', 'Insert Cases', 'Insert Variable', 'Split File...', 'Weight Cases...', and 'Value Labels'. At the bottom of the window, there are tabs for 'Data View' and 'Variable View', and buttons for 'Filter off', 'Weights off', and 'No Split'.

| | id | section | work | pay | promote | changejob | var |
|----|----|---------|------|-----|---------|-----------|-----|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | |
| 18 | 18 | 2 | 90 | 62 | 66 | 3 | |

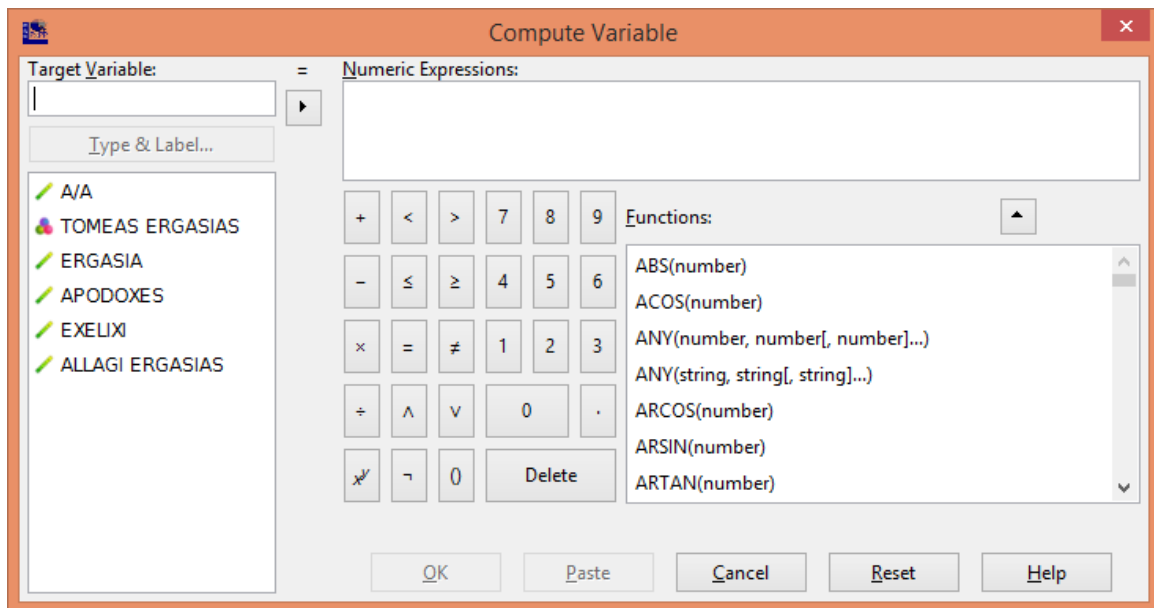
Το ζητούμενο είναι να δημιουργηθεί μια νέα μεταβλητή, με το όνομα **avrscore**, η οποία θα εκφράζει τη μέση τιμή των τριών παραπάνω μεταβλητών. Δηλαδή, η νέα μεταβλητή θα υπολογίζεται βάσει του τύπου:

$$\text{avrscore} = \frac{\text{work} + \text{pay} + \text{promote}}{3}$$

Για να δημιουργήσουμε τη νέα μεταβλητή στο παραπάνω αρχείο του PSPP, θα χρησιμοποιήσουμε τη διαδικασία **COMPUTE**. Επιλέγουμε από τη βασική ράβδο προτιμήσεων του λογισμικού

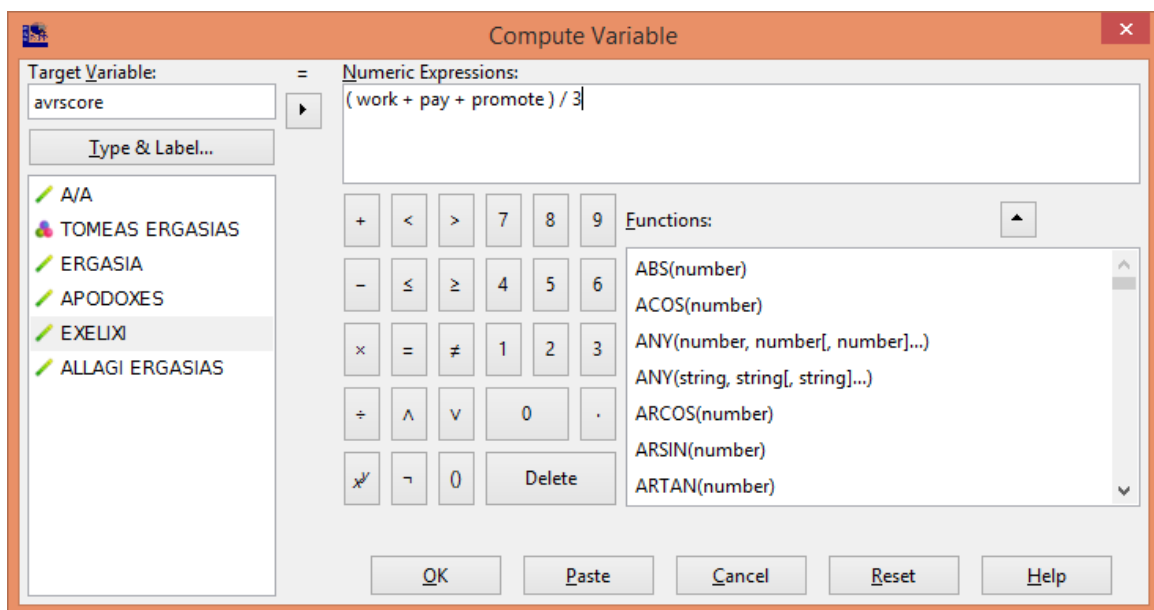
Transform – Compute...

οπότε εμφανίζεται το παρακάτω πλαίσιο διαλόγου:



Στο πεδίο **Target Variable** συμπληρώνουμε το όνομα της νέας μεταβλητής που θα δημιουργήσουμε, δηλαδή **avrscore**, ενώ προαιρετικά επιλέγοντας το πλήκτρο **Type & Label...** μπορούμε να ορίσουμε τον τύπο και την ετικέτα της. Στο πεδίο **Numeric Expressions** πληκτρολογούμε την αριθμητική έκφραση που παράγει την τιμή της νέας μεταβλητής συναρτήσει των μεταβλητών που εμπλέκονται στο ζητούμενο υπολογισμό.

Στην περίπτωσή μας το πλαίσιο διαλόγου θα συμπληρωθεί ως εξής:



Σημειώστε ότι μπορούμε να «συναρμολογήσουμε» την αριθμητική έκφραση που θέλουμε να δημιουργήσουμε, επιλέγοντας τις μεταβλητές που εμπλέκονται από τη λίστα που εμφανίζεται κάτω από το πλήκτρο **Type & Label...** Αντίστοιχα, μπορούμε να αξιοποιήσουμε οποιαδήποτε από τις υπάρχουσες συναρτήσεις από τη λίστα **Functions**, εισάγωντας τες με το αντίστοιχο πλήκτρο με το βελάκι.

Το αποτέλεσμα της εφαρμογής της διαδικασίας εμφανίζεται στο στιγμιότυπο που ακολουθεί:

*ERG-STAT-Askisi_yodeigma_data.sav [DataSet1] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

Open... Save Go To Case... Variables... Find... Insert Cases Insert Variable Split File... Weight Cases... Value Labels

1: id 1

| | id | section | work | pay | promote | changejob | avrscore |
|----|----|---------|------|-----|---------|-----------|----------|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | 59,33 |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | 66,67 |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | 65,00 |
| 4 | 4 | 3 | 87 | 66 | 49 | | 67,33 |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | 70,00 |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | 65,00 |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | 56,33 |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | 63,00 |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | 61,67 |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | 56,67 |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | 44,00 |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | 56,33 |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | 58,00 |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | 60,67 |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | 54,00 |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | 76,67 |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | 57,67 |
| 18 | 18 | 2 | 90 | 62 | 66 | 3 | 72,67 |

Data View Variable View

Filter off Weights off No Split

Περιγραφή / συσχέτιση των τιμών δύο μεταβλητών

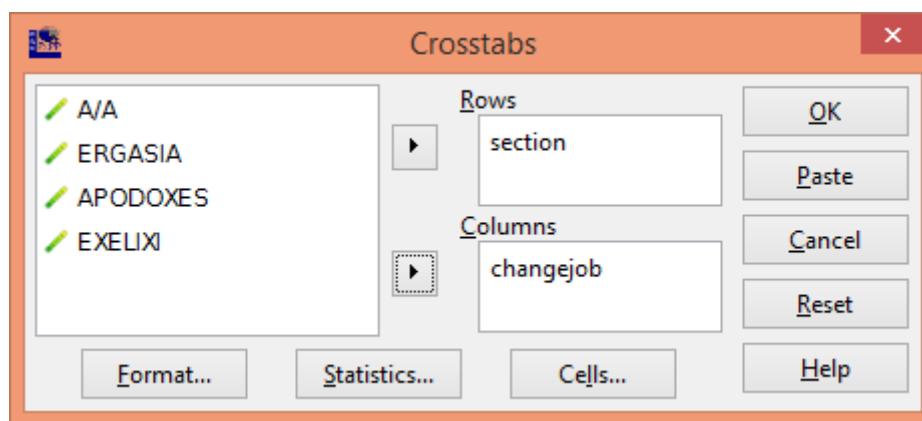
1.3.3 Συσχέτιση μεταξύ δύο ποιοτικών μεταβλητών

Η βασική μέθοδος παρουσίασης δύο ποιοτικών χαρακτηριστικών είναι η κατασκευή της κοινής κατανομής συχνοτήτων (πίνακας συνάφειας) και ο υπολογισμός των αντίστοιχων ποσοστών.

Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας

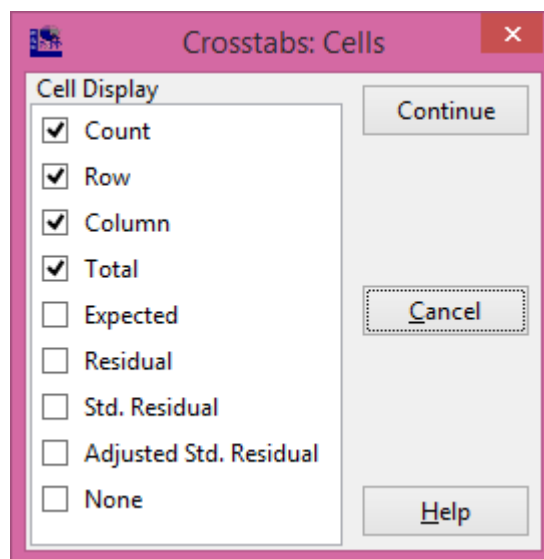
Analyze – Descriptive Statistics – Crosstabs

εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Crosstabs**:



Επιλέγουμε την μεταβλητή, τις κατηγορίες της οποίας θέλουμε να έχουμε στις γραμμές του πίνακα συνάφειας και τη μετακινούμε στο πλαίσιο **Rows**. Επιλέγουμε την άλλη μεταβλητή και τη μετακινούμε στο πλαίσιο **Columns**.

Εφόσον επιθυμούμε επιλέγουμε το πλήκτρο **Cells...** και στο παράθυρο που εμφανίζεται επιλέγουμε το συγκεντρωτικό μέγεθος που θέλουμε να υπολογιστεί σε κάθε κελί του πίνακα:



Τα συνηθέστερα από αυτά είναι:

- **Count:** Πλήθος περιπτώσεων στο αντίστοιχο κελί.
- **Row:** Ποσοστό επί της γραμμής
- **Column:** Ποσοστό επί της στήλης
- **Total:** Ποσοστό επί του συνόλου

Το αποτέλεσμα αυτής της διαδικασίας είναι ο πίνακας που εμφανίζεται στο παρακάτω στιγμιότυπο:

Output — PSPPIRE Output Viewer

File Edit Windows Help

CROSSTABS

CROSSTABS
 /TABLES= section BY changejob
 /FORMAT=AVALUE TABLES PIVOT
 /STATISTICS=CHISQ
 /CELLS=COUNT ROW COLUMN TOTAL.

Summary.

| | Cases | | |
|-----------------------------------|-----------|-----------|-----------|
| | Valid | Missing | Total |
| | N Percent | N Percent | N Percent |
| TOMEAS ERGASIAS * ALLAGI ERGASIAS | 50 100,0% | 0 0,0% | 50 100,0% |

TOMEAS ERGASIAS * ALLAGI ERGASIAS [count, row %, column %, total %].

| TOMEAS ERGASIAS | ALLAGI ERGASIAS | | | Total |
|------------------|--------------------------------------|--------------------------------------|-------------------------------------|--|
| | No chance | Slight chance | High chance | |
| EREYNA AGORAS | 11,00 61,11% 36,67% 22,00% | 5,00 27,78% 41,67% 10,00% | 2,00 11,11% 25,00% 4,00% | 18,00 100,00% 36,00% 36,00% |
| DIMOSIES SXESEIS | 12,00 63,16% 40,00% 24,00% | 4,00 21,05% 33,33% 8,00% | 3,00 15,79% 37,50% 6,00% | 19,00 100,00% 38,00% 38,00% |
| DIAFIMISI | 7,00 53,85% 23,33% 14,00% | 3,00 23,08% 25,00% 6,00% | 3,00 23,08% 37,50% 6,00% | 13,00 100,00% 26,00% 26,00% |
| Total | 30,00 60,00% 100,00% 60,00% | 12,00 24,00% 100,00% 24,00% | 8,00 16,00% 100,00% 16,00% | 50,00 100,00% 100,00% 100,00% |

Chi-square tests.

| Statistic | Value | df | Asymp. Sig. (2-tailed) |
|------------------------------|-------|----|------------------------|
| Pearson Chi-Square | ,97 | 4 | ,91 |
| Likelihood Ratio | ,96 | 4 | ,92 |
| Linear-by-Linear Association | ,44 | 1 | ,50 |
| N of Valid Cases | 50 | | |

- Η πρώτη τιμή που εμφανίζεται σε κάθε ένα από τα κελιά του πίνακα συνάφειας είναι η απόλυτη συχνότητα, δηλαδή το πλήθος των περιπτώσεων, που αντιστοιχεί στις τιμές των δύο μεταβλητών που εξετάζονται. Π.χ. στο κελί που αντιστοιχεί στη διασταύρωση των τιμών “EREYNA AGORAS” και “No Chance”, η τιμή 11 αντιστοιχεί στο πλήθος των εργαζομένων που εμφανίζουν τα δύο αυτά χαρακτηριστικά.
- Η δεύτερη τιμή σε κάθε κελί, δίνει το ποσοστό (%) των περιπτώσεων με τα συγκεκριμένα χαρακτηριστικά, επί του συνόλου των περιπτώσεων της γραμμής στην οποία βρισκόμαστε. Π.χ. στο κελί που αντιστοιχεί στη διασταύρωση των τιμών “EREYNA AGORAS” και “No Chance”, το ποσοστό 61.11%, προκύπτει από τον υπολογισμό $11/18 * 100$.
- Ομοίως, η τρίτη τιμή σε κάθε κελί δίνει το ποσοστό (%) των περιπτώσεων με τα συγκεκριμένα χαρακτηριστικά, επί του συνόλου των περιπτώσεων της στήλης στην οποία βρισκόμαστε. Π.χ. στο κελί που αντιστοιχεί στη διασταύρωση των τιμών “EREYNA AGORAS” και “No Chance”, το ποσοστό 36.67%, προκύπτει από τον υπολογισμό $11/30 * 100$.

- Τέλος, η τέταρτη τιμή σε κάθε κελί, δίνει το ποσοστό (%) των περιπτώσεων με τα συγκεκριμένα χαρακτηριστικά, επί του γενικού συνόλου των περιπτώσεων. Π.χ. στο κελί που αντιστοιχεί στη διασταύρωση των τιμών “EREYNA AGORAS” και “No Chance”, το ποσοστό 22.00%, προκύπτει από τον υπολογισμό $11/50 * 100$.

Περισσότερα στοιχεία σχετικά με την ερμηνεία των αποτελεσμάτων του τελευταίου πίνακα στο παραπάνω στιγμιότυπο της εξόδου, αναφέρονται στην ενότητα 2.4.

Παρατήρηση: Η διαδικασία **Crosstabs** μπορεί να χρησιμοποιηθεί και στην περίπτωση των ποσοτικών μεταβλητών των οποίων οι τιμές είναι λίγες ή έχουν κωδικοποιηθεί προηγουμένως και αντιστοιχηθεί σε διαστήματα τιμών.

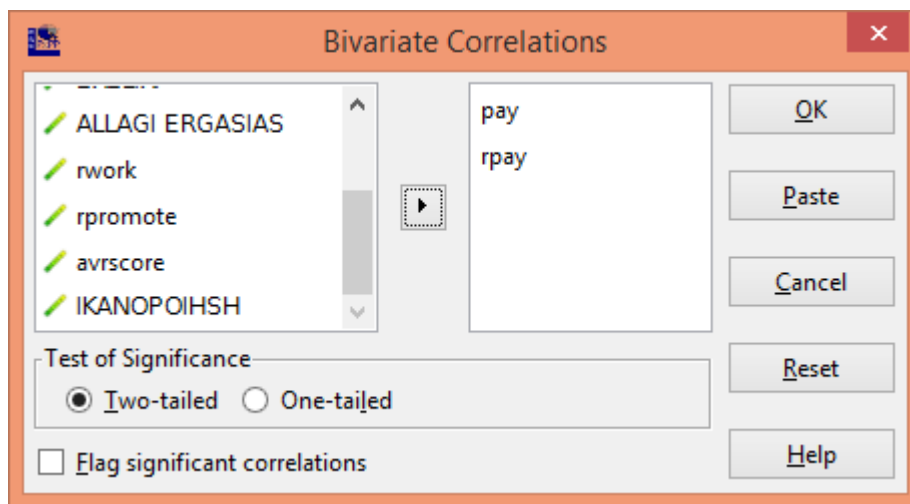
1.3.4 Συσχέτιση μεταξύ δύο ποσοτικών μεταβλητών

Ο υπολογισμός των περιγραφικών στατιστικών μέτρων για τον εντοπισμό της φύσης και της έντασης της σχέσης μεταξύ δύο ποσοτικών μεταβλητών πραγματοποιείται με τον υπολογισμό του συντελεστή γραμμικής συσχέτισης του **Pearson** (διαδικασία **Correlate**) και με την κατασκευή του διαγράμματος διασποράς (διαδικασία **Scatter**).

Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας

Analyze – Bivariate Correlation

εμφανίζεται το πλαίσιο διαλόγου της διαδικασίας **Correlations**.



Επιλέγουμε τις μεταβλητές των οποίων τη σχέση αναζητούμε και τις μετακινούμε στο πλαίσιο variable(s). Μπορούμε να μετακινήσουμε και περισσότερες από δύο μεταβλητές. Στην περίπτωση αυτή οι υπολογισμοί θα γίνουν για όλους τους συνδυασμούς ανά δύο μεταβλητές.

Παρατήρηση: Ο συντελεστής γραμμικής συσχέτισης του **Pearson** (**r**) παίρνει τις τιμές:

$$-1 \leq r \leq +1$$

| -1 | 0 | +1 |
|------------------------------------|-----------------------------|----------------------------------|
| Τέλεια αρνητική γραμμική συσχέτιση | Μηδενική γραμμική συσχέτιση | Τέλεια θετική γραμμική συσχέτιση |

Όσο το r βρίσκεται πιο κοντά στο +1 (-1), τόσο πιο ισχυρή θετική (αρνητική) συσχέτιση υπάρχει.

Όσο το r βρίσκεται πιο κοντά στο 0, τόσο πιο ασθενής συσχέτιση υπάρχει.

Συνήθως, θεωρούμε ότι η συσχέτιση είναι:

| | | |
|---------------------------|------|---------------------|
| Ισχυρή έως πολύ ισχυρή, | όταν | $ r > 0,7$ |
| Μέτρια έως ικανοποιητική, | όταν | $0,5 < r < 0,7$ |
| Ασθενής έως μέτρια, | όταν | $ r < 0,5$ |

Το αποτέλεσμα της παραπάνω διαδικασίας εμφανίζεται στο αρχείο εξόδου που ακολουθεί:

Output — PSPPIRE Output Viewer

File Edit Windows Help

☑ CORRELATIONS

CORRELATIONS
CORRELATION
/VARIABLES = pay rpay
/PRINT = TWOTAIL SIG.

Correlations

| | | APODOXES | rpay |
|----------|---------------------|----------|------|
| APODOXES | Pearson Correlation | 1,00 | ,89 |
| | Sig. (2-tailed) | | ,00 |
| | N | 49 | 49 |
| rpay | Pearson Correlation | ,89 | 1,00 |
| | Sig. (2-tailed) | ,00 | |
| | N | 49 | 49 |

Σημείωση: Ο συντελεστής συσχέτισης του Pearson υπολογίζεται επίσης κατά την εφαρμογή της διαδικασίας Crosstab, αν στο παράθυρο διαλόγου **Crosstab: Statistics**, επιλεγεί το στατιστικό **Corr.**

2 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

Όταν ζητείται να πραγματοποιήσουμε ένα στατιστικό έλεγχο υπόθεσης σε επίπεδο σημαντικότητας α , τότε εφαρμόζουμε την παρακάτω διαδικασία:

1. Διατυπώνουμε τη **μηδενική υπόθεση**, η οποία συμβολίζεται με H_0 , καθώς και την άρνηση της η οποία ονομάζεται **εναλλακτική υπόθεση** που συμβολίζεται με H_1 ή H_a .
2. Επιλέγουμε το **επίπεδο σημαντικότητας α** , το οποίο εκφράζει την πιθανότητα να κάνουμε λάθος απορρίπτοντας την υπόθεση H_0 , ενώ αυτή είναι αληθής. Η συνηθέστερη επιλογή είναι να θέσουμε $\alpha = 0.05$. Η πιθανότητα $1-\alpha$ λέγεται ονομάζεται **επίπεδο εμπιστοσύνης** του ελέγχου.
3. Από τα δεδομένα του δείγματος μας υπολογίζουμε το στατιστικό το οποίο ενδείκνυται από την θεωρία για το εκάστοτε είδος ελέγχου, καθώς και την αντίστοιχη στάθμη σημαντικότητας **sig** που αντιστοιχεί στην τιμή του στατιστικού που υπολογίστηκε.
4. Αποφασίζουμε κατά πόσο αποδεχόμαστε ή όχι τη μηδενική υπόθεση βάσει των παρακάτω:

Αν **sig. $\geq \alpha$** αποδοχή της H_0
Αν **sig. $< \alpha$** απόρριψη της H_0

Σημειώστε ότι οι υπολογισμοί του βήματος 3, γίνονται αυτόματα από το PSPP επιλέγοντας κατά περίπτωση τον στατιστικό έλεγχο που επιθυμούμε.

2.1 Έλεγχος υποθέσεων για τη μέση τιμή ενός δείγματος

Ο έλεγχος αυτός αφορά τις περιπτώσεις όπου θέλουμε να ελέγξουμε, αν η μέση τιμή μιας μεταβλητής διαφέρει στατιστικά σημαντικά από μία δεδομένη τιμή.

Παράδειγμα: Θα θέλαμε να ελέγξουμε κατά πόσο η μέση αξιολόγηση των συνθηκών εργασίας, δηλαδή η μέση τιμή της μεταβλητής **work**, στο παρακάτω αρχείο είναι ίση με $\mu_0=78$.

| | id | section | work | pay | promote | changejob | var | vi |
|----|----|---------|------|-----|---------|-----------|-----|----|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | | |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | | |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | | |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | | |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | | |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | | |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | | |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | | |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | | |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | | |

Στατιστικές Υποθέσεις

$H_0 : \mu = \mu_0$ Η μέση τιμή **δεν** διαφέρει στατιστικά σημαντικά από την δεδομένη τιμή

$H_a : \mu \neq \mu_0$ Η μέση τιμή διαφέρει στατιστικά σημαντικά από την δεδομένη τιμή

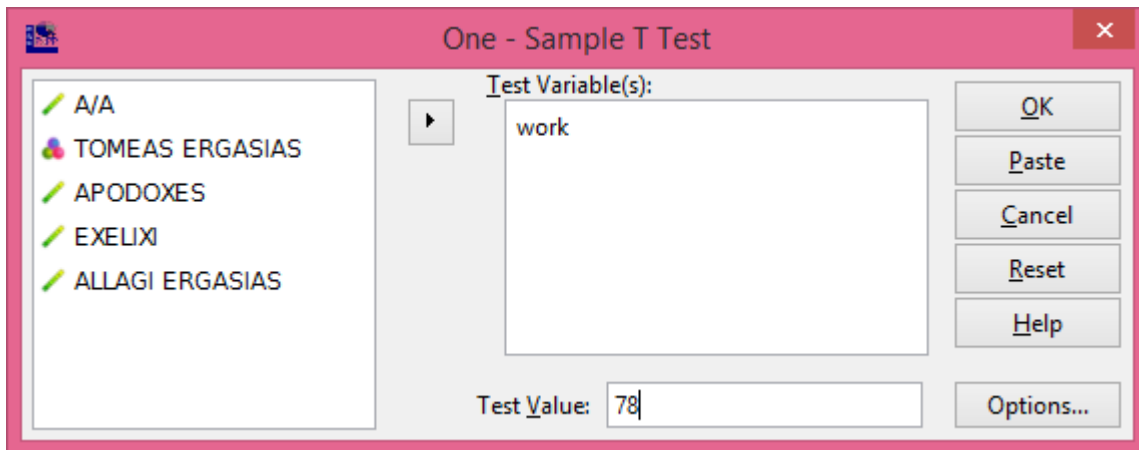
Στατιστικός Έλεγχος

Το είδος του ελέγχου που χρησιμοποιείται σε αυτές τις περιπτώσεις είναι γνωστό ως **t-test** του student.

Λαδκασία στο PSPP

Analyze – Compare Means – One-Sample T Test

Στο παράθυρο που ανοίγει, επιλέγουμε και μεταφέρουμε τη μεταβλητή που μας ενδιαφέρει (work) στο πλαίσιο **Test Variable(s)** ενώ στο πλαίσιο **Test Value** πληκτρολογούμε την υπό έλεγχο τιμή **78**.



Αποτελέσματα

The screenshot shows the 'Output — PSPP Output Viewer' window. It displays the results of a T-TEST. The command used is: T-TEST /TESTVAL=78 /VARIABLES= work /MISSING=ANALYSIS /CRITERIA=CIN(0.95). The output is divided into two main sections: 'One-Sample Statistics' and 'One-Sample Test'.

One-Sample Statistics

| | N | Mean | Std. Deviation | S.E. Mean |
|---------|----|-------|----------------|-----------|
| ERGASIA | 50 | 79,80 | 8,29 | 1,17 |

One-Sample Test

| | Test Value = 78,000000 | | | | | |
|---------|------------------------|----|-----------------|-----------------|---|-------|
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| ERGASIA | 1,54 | 49 | ,13 | 1,80 | -,56 | 4,16 |

Ερμηνεία Αποτελεσμάτων

Στον πρώτο πίνακα εμφανίζονται τα βασικά στατιστικά του δείγματος για τη μεταβλητή work. Παρατηρήστε ότι ο δειγματικός μέσος (Mean) είναι 79,80.

Στον δεύτερο πίνακα δίνονται τα αποτελέσματα του ελέγχου για τη μεταβλητή work. Συγκεκριμένα:

- **t = 1,54**, είναι η τιμή του στατιστικού που χρησιμοποιείται στο t-test.
- **df = 49**, είναι οι βαθμοί ελευθερίας (degrees of freedom) που χρησιμοποιούνται στον έλεγχο. Για το συγκεκριμένο t-test, τίθεται $df = n - 1$ (όπου n το μέγεθος του δείγματος).
- **Sig. = 0,13**, είναι η στάθμη σημαντικότητας του στατιστικού t, βάσει της οποίας παίρνουμε την τελική απόφαση.
- **Mean Difference = 1,80**, είναι η διαφορά του δειγματικού μέσου από την ελεγχόμενη μέση τιμή μ_0 , δηλαδή στην περίπτωση μας **79,80 – 78,00**.
- **95% Confidence interval of the Difference**, (95% διάστημα εμπιστοσύνης για τη διαφορά) δίνει τα εύρος της διαφοράς του μέσου από το μ_0 , έτσι ώστε ο έλεγχος να δίνει στάθμη σημαντικότητας για το στατιστικό t, $sig > 1 - 0,95 = 0,05$ (βλ. περισσότερες λεπτομέρειες στην παρατήρηση παρακάτω).

Επειδή το επίπεδο σημαντικότητας του ελέγχου **sig = 0,13**, είναι μεγαλύτερο του $\alpha = 0,05$, η μηδενική υπόθεση γίνεται αποδεκτή.

Συμπέρασμα

Η μέση τιμή της αξιολόγησης των συνθηκών εργασίας από τους εργαζομένους δεν διαφέρει στατιστικά σημαντικά από την τιμή $\mu_0=78$ ($t = 1,54$, $df = 49$, $sig. = 0,13$).

Παρατήρηση (ερμηνεία του 95% διαστήματος εμπιστοσύνης)

Αξίζει να δείξουμε πως μπορούμε να αξιοποιήσουμε την πληροφορία του 95% διαστήματος εμπιστοσύνης της διαφοράς που υπολογίστηκε κατά τον προηγούμενο έλεγχο. Αν επιλέξουμε ως τιμή μ'_0 , έτσι ώστε:

$$\mu_0 + \text{Lower} \leq \mu'_0 \leq \mu_0 + \text{Upper}$$

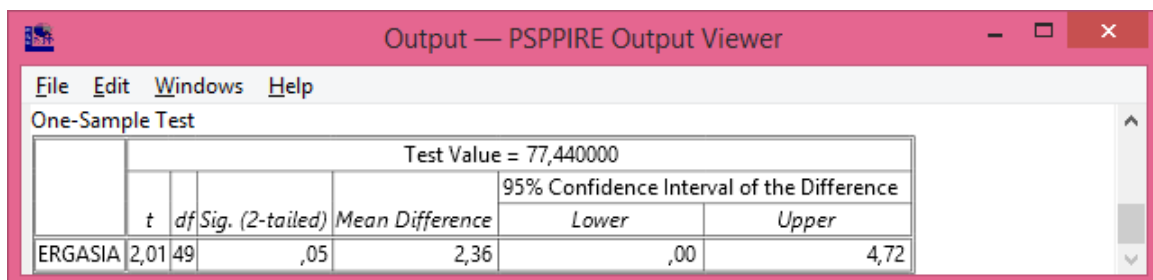
δηλαδή στην περίπτωση μας

$$78 - 0,56 \leq \mu'_0 \leq 78 + 4,16$$

$$77,44 \leq \mu'_0 \leq 82,16$$

τότε η υπόθεση $H_0: \mu = \mu'_0$ θα γινόταν αποδεκτή σε στάθμη σημαντικότητας $\alpha = 1 - 0,95 = 0,05$.

Για παράδειγμα ο έλεγχος $H_0: \mu = 77,44$, δίνει



| One-Sample Test | | | | | | |
|------------------------|------|----|-----------------|-----------------|---|-------|
| Test Value = 77,440000 | | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| ERGASIA | 2,01 | 49 | ,05 | 2,36 | ,00 | 4,72 |

ενώ αντίστοιχα ο έλεγχος $H_0: \mu = 82,16$, δίνει

| One-Sample Test | | | | | | |
|------------------------|-------|----|-----------------|-----------------|---|-------|
| Test Value = 82,160000 | | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| ERGASIA | -2,01 | 49 | ,05 | -2,36 | -4,72 | ,00 |

Παρατηρούμε ότι και στις δύο παραπάνω περιπτώσεις η στάθμη σημαντικότητας είναι **sig.= 0,05**, ακριβώς ίση δηλαδή με το επίπεδο σημαντικότητας του ελέγχου α .

Για οποιαδήποτε τιμή ελέγχου μ_0 στο διάστημα **[77,44, 82,16]**, όπως η τιμή $\mu_0 = 78$ που ελέγχθηκε αρχικά, η υπόθεση H_0 γίνεται αποδεκτή σε επίπεδο σημαντικότητας $\alpha = 0,05$.

Αν εκτελέσουμε τον ίδιο έλεγχο για κάποια τιμή εκτός του παραπάνω διαστήματος, π.χ. θέτοντας $H_0: \mu = 77$, το αποτέλεσμα είναι

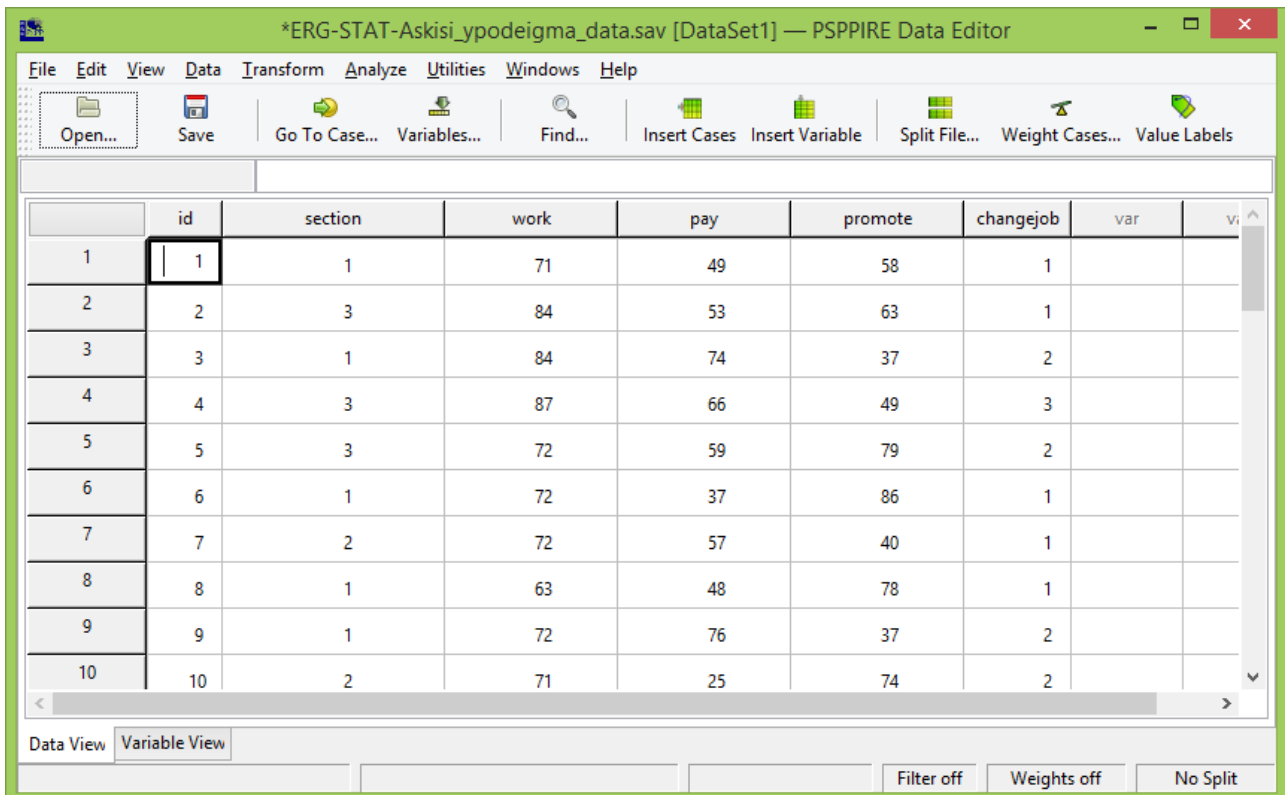
| One-Sample Test | | | | | | |
|------------------------|------|----|-----------------|-----------------|---|-------|
| Test Value = 77,000000 | | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| ERGASIA | 2,39 | 49 | ,02 | 2,80 | ,44 | 5,16 |

από όπου, λόγω του ότι **sig = 0,02 < 0,05**, συμπεραίνουμε ότι η υπόθεση $H_0: \mu = 77$ απορρίπτεται σε επίπεδο σημαντικότητας $\alpha = 0,05$.

2.2 Έλεγχος υποθέσεων για τη διαφορά των μέσων τιμών δύο ανεξάρτητων πληθυσμών

Ο έλεγχος αυτός αφορά τις περιπτώσεις όπου θέλουμε να ελέγξουμε αν η μέση τιμή μιας μεταβλητής διαφέρει ή όχι σε δύο ανεξάρτητους πληθυσμούς.

Παράδειγμα: Θα θέλαμε να ελέγξουμε κατά πόσο η μέση αξιολόγηση των συνθηκών εργασίας, δηλαδή η μέση τιμή της μεταβλητής **work**, διαφέρει ή όχι στους εργαζομένους που απασχολούνται στην έρευνα αγοράς (**section = 1**) σε σχέση με τους συναδέλφους τους στις δημόσιες σχέσεις (**section = 2**).



| | id | section | work | pay | promote | changejob | var | vi |
|----|----|---------|------|-----|---------|-----------|-----|----|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | | |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | | |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | | |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | | |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | | |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | | |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | | |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | | |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | | |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | | |

Στατιστικές Υποθέσεις

$H_0 : \mu_1 = \mu_2$ Οι μέσες τιμές των δύο πληθυσμών **δεν** διαφέρουν στατιστικά σημαντικά

$H_a : \mu_1 \neq \mu_2$ Οι μέσες τιμές των δύο πληθυσμών **διαφέρουν** στατιστικά σημαντικά

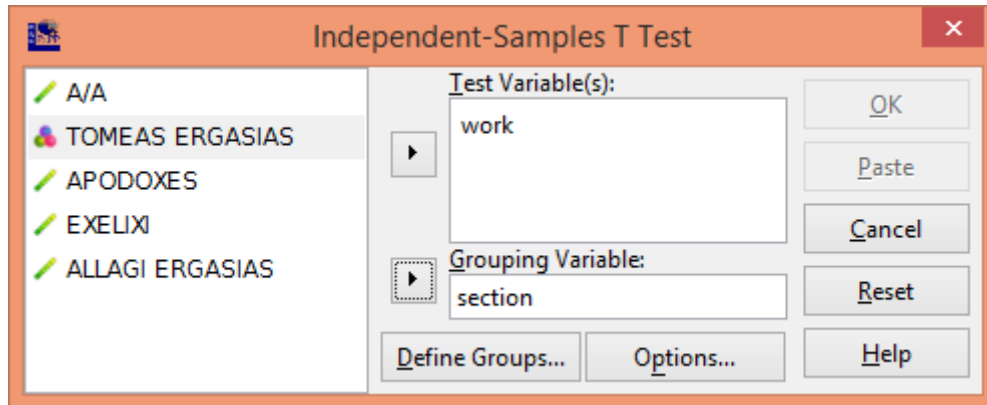
Στατιστικός Έλεγχος

Το είδος του ελέγχου που χρησιμοποιείται σε αυτές τις περιπτώσεις είναι όπως και στην περίπτωση του ελέγχου της μέσης τιμής ενός δείγματος το **t-test**. Επειδή όμως τα δείγματα προέρχονται από διαφορετικούς πληθυσμούς, για την εφαρμογή του t-test απαιτείται προηγουμένως να ελεγχθεί κατά πόσο οι δύο πληθυσμοί έχουν κοινή διασπορά. Για να επιτευχθεί αυτό πριν το t-test εκτελείται μια δοκιμασία η οποία ονομάζεται **Levene's test**, μέσω της οποίας αποφασίζουμε σε ένα δεδομένο επίπεδο σημαντικότητας κατά πόσο οι δύο πληθυσμοί εμφανίζουν ίδια διασπορά ή όχι. Ανάλογα με την απόφαση αυτή, επιλέγεται το κατάλληλα διαφοροποιημένο t-test, βάσει του οποίου ελέγχουμε την ισότητα των μέσων τιμών. Η όλη διαδικασία υπολογισμού των στατιστικών είναι αυτοματοποιημένη στο PSPP.

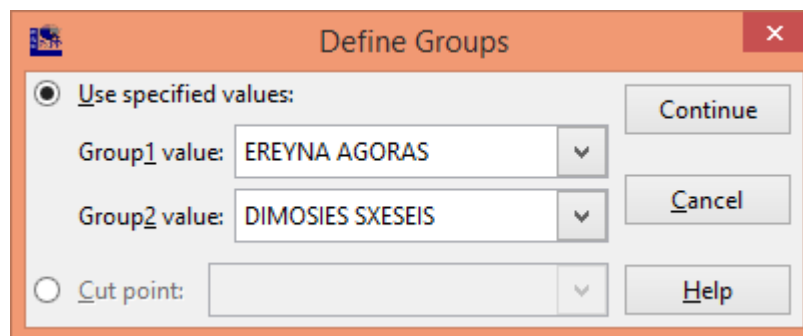
Λιαδικασία

Analyze – Compare Means – Independent-Sample T-Test

Επιλέγουμε τη μεταβλητή, της οποία θέλουμε να ελέγξουμε τη μέση τιμή, στο πλαίσιο **Test Variable** (work) και τη μεταβλητή, βάση των τιμών της οποίας ορίζονται οι δύο ανεξάρτητοι πληθυσμοί, στο πλαίσιο **Grouping Variable** (section).



Παρατηρούμε ότι ενεργοποιείται το κουμπί **Define Groups...** το οποίο και επιλέγουμε. Στο παράθυρο που ανοίγει στα πλαίσια **Group1** και **Group2** δίνουμε αντίστοιχα τις τιμές της μεταβλητής (section) που προσδιορίζουν τους δύο πληθυσμούς. Στο πλαίσιο Group1 δίνουμε την τιμή που αντιστοιχεί στην έρευνα αγοράς και στο Group2 την τιμή που αντιστοιχεί στις δημόσιες σχέσεις.



Στο παράθυρο Define Groups δίνεται και μια ακόμη δυνατότητα προσδιορισμού των δύο πληθυσμών, δηλώνοντας μια οριακή τιμή για τη μεταβλητή Grouping Variable. Για παράδειγμα, θα μπορούσαμε να προσδιορίσουμε δύο πληθυσμούς εργαζομένων ανάλογα με το αν μια ποσοτική μεταβλητή είναι μεγαλύτερη ή μικρότερη κάποιας συγκεκριμένης τιμής.

Αποτελέσματα

Output — PSPPIRE Output Viewer

File Edit Windows Help

T-TEST
T-TEST /VARIABLES= work
/GROUPS=section(1,2) /MISSING=ANALYSIS
/CRITERIA=CIN(0.95).

Group Statistics

| TOMEAS ERGASIAS | N | Mean | Std. Deviation | S.E. Mean |
|-----------------------|----|-------|----------------|-----------|
| ERGASIA EREYNA AGORAS | 18 | 80,67 | 9,47 | 2,23 |
| DIMOSIES SXESEIS | 19 | 79,58 | 7,86 | 1,80 |

Independent Samples Test

| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|-----------------------------|---|------|------------------------------|-------|-----------------|-----------------|-----------------------|---|-------|
| | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | Lower | Upper |
| ERGASIA | ,42 | ,52 | ,38 | 35,00 | ,71 | 1,09 | 2,87 | -4,74 | 6,92 |
| Equal variances assumed | | | | | | | | | |
| Equal variances not assumed | | | ,38 | 33,11 | ,71 | 1,09 | 2,87 | -4,75 | 6,93 |

Ερμηνεία Αποτελεσμάτων

Στον πρώτο πίνακα εμφανίζονται τα στατιστικά των δύο δειγμάτων. Στο παράδειγμα μας οι δειγματικοί μέσοι της μεταβλητής work, στους εργαζόμενους στην έρευνα αγοράς και τις δημόσιες σχέσεις είναι 80,67 και 79,58 αντίστοιχα.

Ο έλεγχος υποθέσεων για την διαφορά μέσων τιμών σε δύο ανεξάρτητους πληθυσμούς (t-test) πραγματοποιείται σε δύο φάσεις. Τα αποτελέσματα αυτών των ελέγχων παρουσιάζονται στο δεύτερο πίνακα των αποτελεσμάτων.

Φάση 1: Σύγκριση διασπορών των δύο πληθυσμών

$$H_0 : \sigma_1^2 = \sigma_2^2$$

Οι διακυμάνσεις των δύο πληθυσμών **δεν** διαφέρουν σημαντικά (equal variances assumed)

$$H_a : \sigma_1^2 \neq \sigma_2^2$$

Οι διακυμάνσεις των δύο πληθυσμών **διαφέρουν** σημαντικά (equal variances not assumed)

Ο στατιστικός έλεγχος που πραγματοποιείται είναι του Levene (F-test). Για το παραπάνω παράδειγμα **F = 0,42 sig. = 0,52**. Επειδή **sig. > 0,05** αποδεχόμαστε τη μηδενική υπόθεση, δηλαδή ότι οι διακυμάνσεις δεν διαφέρουν (equal variances assumed) και συνεχίζουμε να διαβάζουμε την πρώτη γραμμή του ίδιου πίνακα. Αν απορρίπταμε τη μηδενική υπόθεση τότε θα συνεχίζαμε να διαβάζουμε τη δεύτερη γραμμή.

Φάση 2: Σύγκριση μέσων των δύο πληθυσμών

Για το παραπάνω παράδειγμα **t = 0,38, df = 35, sig. = 0,71**. Επειδή **sig. > 0,05** αποδεχόμαστε τη μηδενική υπόθεση, δηλαδή ότι οι μέσοι των δύο πληθυσμών δεν διαφέρουν σημαντικά.

Συμπέρασμα

Η μέση αξιολόγηση των συνθηκών εργασίας δεν διαφέρει στατιστικά σημαντικά μεταξύ των εργαζομένων στην έρευνα αγοράς και στις δημόσιες σχέσεις (**t = 0,38, df = 35, sig. = 0,71**).

2.3 Σύγκριση μέσων τιμών σε ζευγάρια παρατηρήσεων

Ο έλεγχος αυτός αφορά τις περιπτώσεις όπου θέλουμε να ελέγξουμε αν οι μέσες τιμές δύο μεταβλητών διαφέρουν ή όχι όταν οι αντίστοιχες μετρήσεις προέρχονται από ζεύγη παρατηρήσεων, από τον ίδιο πληθυσμό.

Παράδειγμα

Θα θέλαμε να ελέγξουμε κατά πόσο η μέση αξιολόγηση των συνθηκών εργασίας διαφέρει από τη μέση αξιολόγηση των αποδοχών όλων των εργαζομένων.

Στατιστικές Υποθέσεις

$H_0 : \mu_1 = \mu_2$ Οι μέσες τιμές των δύο πληθυσμών **δεν** διαφέρουν στατιστικά σημαντικά

$H_a : \mu_1 \neq \mu_2$ Οι μέσες τιμές των δύο πληθυσμών **διαφέρουν** στατιστικά σημαντικά

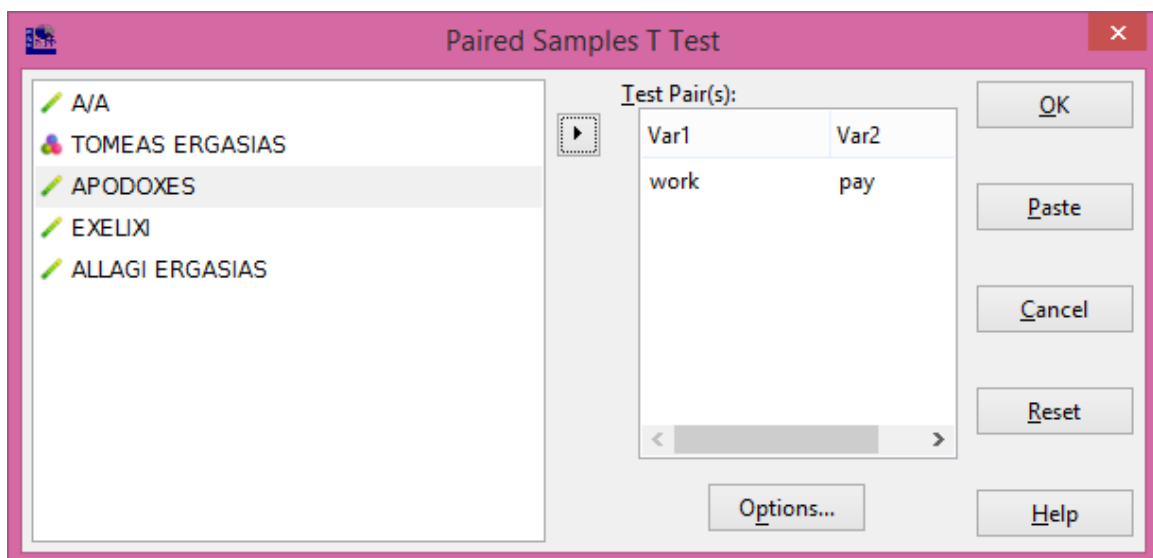
Στατιστικός Έλεγχος

Όπως και στις δύο προηγούμενες περιπτώσεις ελέγχων για τη μέση τιμή, θα χρησιμοποιηθεί το **t-test** του **student**.

Διαδικασία

Analyze – Compare Means – Paired-Samples T-Test

Επιλέγουμε τις δύο μεταβλητές, των οποίων θέλουμε να ελέγξουμε τη μέση τιμή τους και τις μεταφέρουμε στο πλαίσιο **Test Pairs** (work – pay).



Αποτελέσματα

Output — PSPPIRE Output Viewer

File Edit Windows Help

T-TEST
T-TEST
PAIRS = work WITH pay (PAIRED)
/MISSING=ANALYSIS
/CRITERIA=CIN(0.95).

Paired Sample Statistics

| | Mean | N | Std. Deviation | S.E. Mean |
|----------------|-------|----|----------------|-----------|
| Pair 0 ERGASIA | 80,12 | 49 | 8,05 | 1,15 |
| APODOXES | 54,69 | 49 | 14,81 | 2,12 |

Paired Samples Correlations

| | N | Correlation | Sig. |
|---------------------------|----|-------------|------|
| Pair 0 ERGASIA & APODOXES | 49 | ,24 | ,10 |

Paired Samples Test

| | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---------------------------|--------------------|----------------|-----------------|---|-------|-------|----|-----------------|
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | Lower | Upper | | | |
| Pair 0 ERGASIA - APODOXES | 25,43 | 15,07 | 2,15 | 21,10 | 29,76 | 11,81 | 48 | ,00 |

Ερμηνεία Αποτελεσμάτων

Στον πρώτο πίνακα παρουσιάζονται κάποια στατιστικά των δύο μεταβλητών των οποίων η μέση τιμή τίθεται υπό σύγκριση. Η μέση τιμή της αξιολόγησης των συνθηκών εργασίας είναι $\mu_1 = 80,12$, ενώ η μέση αξιολόγηση των αποδοχών $\mu_2 = 54,69$.

Στο δεύτερο πίνακα παρουσιάζεται ο συντελεστής συσχέτισης των δύο μεταβλητών (**Correlation = 0,24**). Το γεγονός ότι ο συντελεστής συσχέτισης είναι θετικός αποτελεί ένδειξη ότι η επιλογή της μεθόδου της σύγκρισης μέσω τιμών από ζευγάρια παρατηρήσεων είναι δικαιολογημένη.

Από τον τρίτο πίνακα εξάγουμε τα τελικά μας συμπεράσματα, καθώς παρατηρούμε ότι **t = 11,81**, **df = 48**, **sig. = 0,00**. Επειδή **sig. < $\alpha = 0,05$** η μηδενική υπόθεση απορρίπτεται.

Συμπέρασμα

Η μέση αξιολόγηση των συνθηκών εργασίας διαφέρει από τη μέση αξιολόγηση των αποδοχών των εργαζομένων σε επίπεδο σημαντικότητας **$\alpha = 0,05$ t = 11,81, df = 48, sig. = 0,00**).

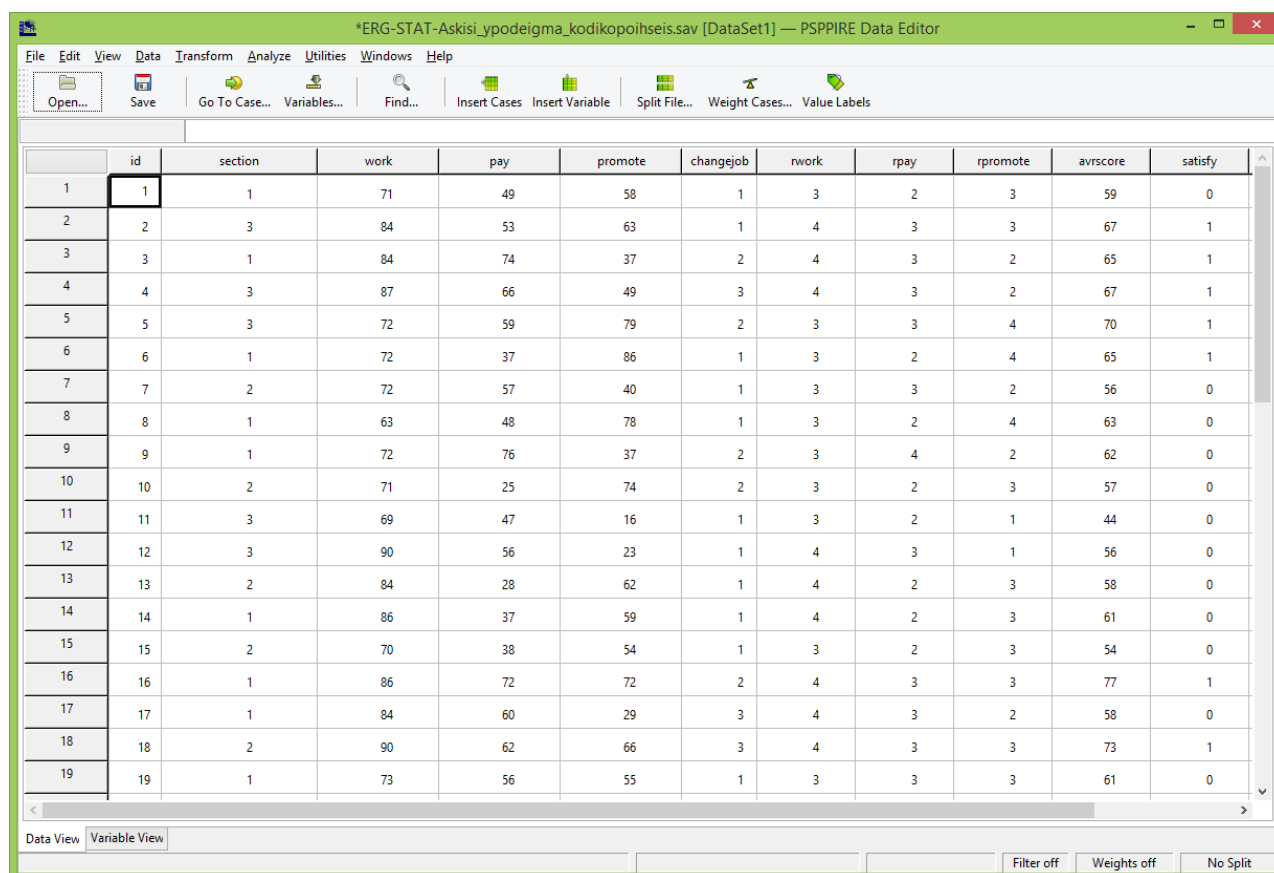
2.4 Έλεγχος ανεξαρτησίας / ομοιογένειας μεταξύ δύο ποιοτικών μεταβλητών

Ο έλεγχος αυτός αφορά τις περιπτώσεις όπου θέλουμε να ελέγξουμε αν δύο ποιοτικές μεταβλητές είναι ανεξάρτητες μεταξύ τους ή όχι. Ισοδύναμα, με την ίδια διαδικασία μπορούμε να ελέγξουμε κατά πόσο μια ποιοτική μεταβλητή παρουσιάζει ομοιογένεια ως προς κάποια ιδιότητα του πληθυσμού.

Παράδειγμα:

Θα θέλαμε να ελέγξουμε κατά πόσο ο τομέας εργασίας των εργαζομένων (**section**) και η μεταβλητή που εκφράζει την ικανοποίηση του εργαζομένου από την εργασία του (**satisfy**), είναι ανεξάρτητα μεταξύ τους χαρακτηριστικά.

Με τον ίδιο έλεγχο, θα μπορούσαμε να ελέγξουμε κατά πόσο η ικανοποίηση των εργαζομένων παρουσιάζει ομοιογένεια, εξεταζόμενη στους τρεις τομείς εργασίας.



| | id | section | work | pay | promote | changejob | rwork | rpay | rpromote | avrscore | satisfy |
|----|----|---------|------|-----|---------|-----------|-------|------|----------|----------|---------|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | 3 | 2 | 3 | 59 | 0 |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | 4 | 3 | 3 | 67 | 1 |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | 4 | 3 | 2 | 65 | 1 |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | 4 | 3 | 2 | 67 | 1 |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | 3 | 3 | 4 | 70 | 1 |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | 3 | 2 | 4 | 65 | 1 |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | 3 | 3 | 2 | 56 | 0 |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | 3 | 2 | 4 | 63 | 0 |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | 3 | 4 | 2 | 62 | 0 |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | 3 | 2 | 3 | 57 | 0 |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | 3 | 2 | 1 | 44 | 0 |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | 4 | 3 | 1 | 56 | 0 |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | 4 | 2 | 3 | 58 | 0 |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | 4 | 2 | 3 | 61 | 0 |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | 3 | 2 | 3 | 54 | 0 |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | 4 | 3 | 3 | 77 | 1 |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | 4 | 3 | 2 | 58 | 0 |
| 18 | 18 | 2 | 90 | 62 | 66 | 3 | 4 | 3 | 3 | 73 | 1 |
| 19 | 19 | 1 | 73 | 56 | 55 | 1 | 3 | 3 | 3 | 61 | 0 |

Στατιστικές Υποθέσεις

Αν ο έλεγχος αφορά την ανεξαρτησία δύο ποιοτικών χαρακτηριστικών του δείγματος, τότε οι υπό έλεγχο υποθέσεις διατυπώνονται ως εξής:

H₀: Ο τομέας εργασίας και η ικανοποίηση του εργαζομένου είναι ανεξάρτητα χαρακτηριστικά μεταξύ τους.

H_a: Ο τομέας εργασίας και η ικανοποίηση του εργαζομένου δεν είναι ανεξάρτητα χαρακτηριστικά μεταξύ τους.

Αντίστοιχα, αν ο έλεγχος αφορά την ομοιογένεια ενός χαρακτηριστικού ως προς κάποιο άλλο, τότε οι υπό έλεγχο υποθέσεις είναι:

H₀: Η ικανοποίηση των εργαζομένων παρουσιάζει ομοιογένεια, ως προς τους τρεις τομείς εργασίας.

H_a: Η ικανοποίηση των εργαζομένων δεν παρουσιάζει ομοιογένεια, ως προς τους τρεις τομείς εργασίας.

Στατιστικός Έλεγχος

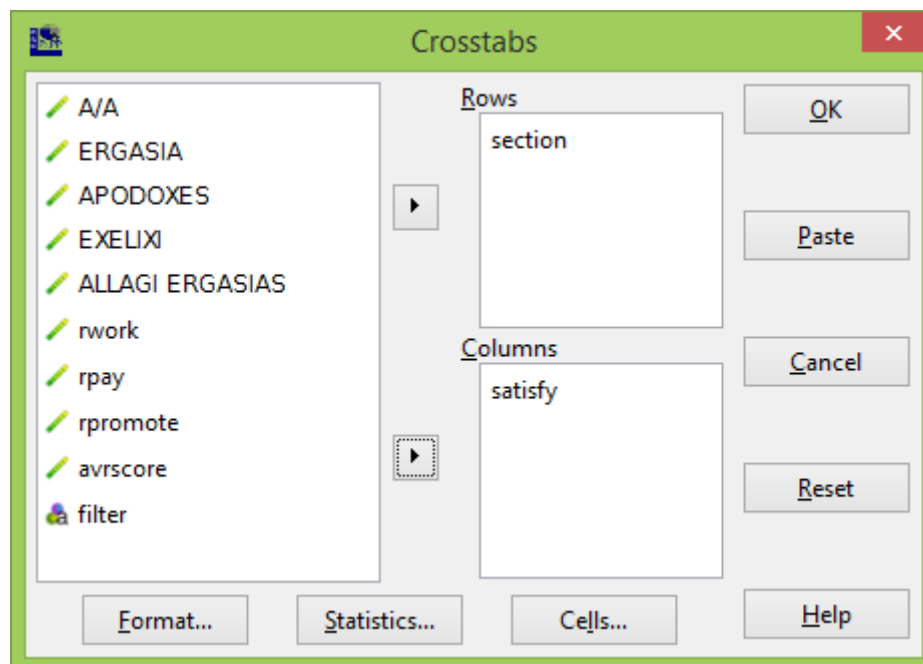
X² –test (chi-square test)

Ο έλεγχος X² εντοπίζει τυχόν διαφορές που υπάρχουν στην κατανομή των τιμών της μιας μεταβλητής στις τιμές της άλλης. Δυστυχώς, δεν μας παρέχει καμία πληροφορία ούτε για την ένταση ούτε για την αιτία της σχέσης που υπάρχει μεταξύ των δύο μεταβλητών (όταν υπάρχει).

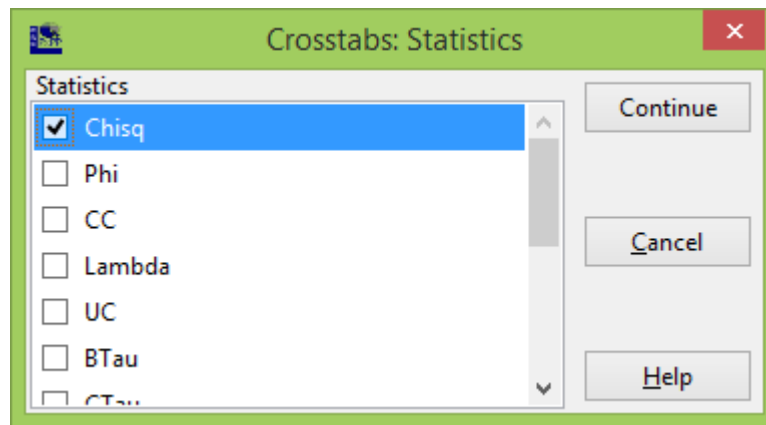
Διαδικασία

Analyze – Descriptive Statistics – Crosstabs

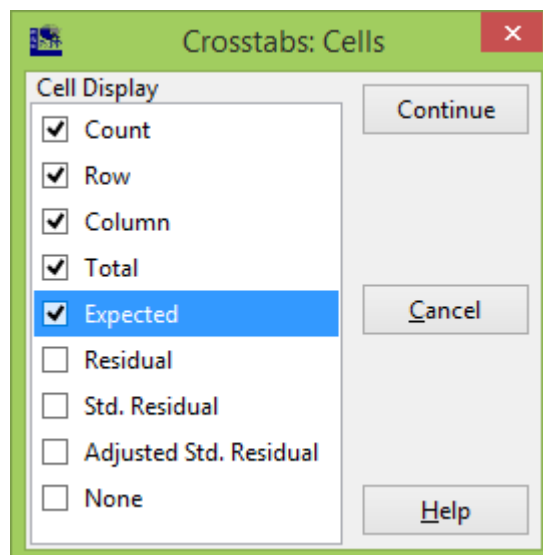
Μεταφέρουμε τη μία μεταβλητή στο πλαίσιο **Row(s) (section)** και την άλλη μεταβλητή στο πλαίσιο **Column(s) (satisfy)**. Το αποτέλεσμα της διαδικασίας δεν επηρεάζεται από το ποια μεταβλητή θα μεταφερθεί σε ποιο πλαίσιο, παρά μόνο στην εμφάνιση κάποιων αποτελεσμάτων.



Στη συνέχεια, ενεργοποιούμε το κουμπί “**Statistics**” και στο παράθυρο που ανοίγει επιλέγουμε το “**Chisq**” (Chi square).



Στη συνέχεια επιλέγουμε **Continue** και στο αρχικό παράθυρο **Crosstabs**, επιλέγουμε το κουμπί **Cells**. Στο παράθυρο διαλόγου **Cells** είναι σκόπιμο να επιλέξουμε την επιλογή **Expected** από τη λίστα **Cell Display**, ώστε να εμφανιστούν στον πίνακα συνάφειας οι αναμενόμενες συχνότητες για κάθε κελί. Οι αναμενόμενες συχνότητες είναι οι συχνότητες που θα έπρεπε θεωρητικά να εμφανίζονται σε κάθε κελί αν πραγματικά οι υπό έλεγχο μεταβλητές ήταν ανεξάρτητες.



Ο λόγος που προτείνεται η ενεργοποίηση της εμφάνισης της αναμενόμενης συχνότητας στα κελιά του πίνακα συνάφειας, είναι ότι σύμφωνα με τη θεωρία για να είναι αξιόπιστος ο έλεγχος ανεξαρτησίας χ^2 , θα πρέπει οι αναμενόμενες συχνότητες να μην είναι μικρότερες του 5.

Τελικά, επιλέγοντας το κουμπί **OK** στο κεντρικό διάλογο **Crosstabs**, εκτελείται η διαδικασία **Crosstabs** στην οποία θα ενσωματώνονται το αποτέλεσμα του ελέγχου ανεξαρτησίας χ^2 .

Αποτελέσματα

Output — PSPPIRE Output Viewer

File Edit Windows Help

Summary.

| | Cases | | | | | |
|------------------------------|-------|---------|---------|---------|-------|---------|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| TOMEAS ERGASIAS * IKANOPOIHS | 48 | 96,0% | 2 | 4,0% | 50 | 100,0% |

TOMEAS ERGASIAS * IKANOPOIHS [count, row %, column %, total %, expected].

| TOMEAS ERGASIAS | IKANOPOIHS | | Total |
|------------------|--|--|---|
| | 0 | 1 | |
| EREYNA AGORAS | 9,00 8,63 50,00% 39,13% 18,75% | 9,00 9,38 50,00% 36,00% 18,75% | 18,00 ,00 100,00% 37,50% 37,50% |
| DIMOSIES SXESEIS | 9,00 8,63 50,00% 39,13% 18,75% | 9,00 9,38 50,00% 36,00% 18,75% | 18,00 ,00 100,00% 37,50% 37,50% |
| DIAFIMISI | 5,00 5,75 41,67% 21,74% 10,42% | 7,00 6,25 58,33% 28,00% 14,58% | 12,00 ,00 100,00% 25,00% 25,00% |
| Total | 23,00 47,92% 100,00% 47,92% | 25,00 52,08% 100,00% 52,08% | 48,00 100,00% 100,00% 100,00% |

Chi-square tests.

| Statistic | Value | df | Asymp. Sig. (2-tailed) |
|------------------------------|-------|----|------------------------|
| Pearson Chi-Square | ,25 | 2 | ,88 |
| Likelihood Ratio | ,25 | 2 | ,88 |
| Linear-by-Linear Association | ,17 | 1 | ,68 |
| N of Valid Cases | 48 | | |

Ερμηνεία Αποτελεσμάτων

Ο έλεγχος X^2 βασίζεται στο στατιστικό του Pearson και γι' αυτό διαβάζουμε την πρώτη γραμμή του τελευταίου πίνακα.

Το επίπεδο σημαντικότητας του ελέγχου **0,88** που είναι μεγαλύτερο του $\alpha = 0,05$.

Επίσης, παρατηρούμε ότι σε όλα τα κελιά του δεύτερου πίνακα, οι αναμενόμενες συχνότητες (δεύτερο νόμμο σε κάθε κελί του πίνακα Crosstabs) είναι **μεγαλύτερες του 5**, άρα το **αποτέλεσμα του ελέγχου μπορεί να θεωρηθεί αξιόπιστο**.

Συνεπώς, η μηδενική υπόθεση γίνεται αποδεκτή.

Συμπέρασμα

Ο τομέας εργασίας και η ικανοποίηση του εργαζομένου είναι ανεξάρτητα χαρακτηριστικά μεταξύ τους ($X^2 = 0.25$, $df = 2$, $sig. = 0.88$).

Εναλλακτικά, αν ο έλεγχος αφορούσε την ομοιογένεια του χαρακτηριστικού ικανοποίηση από την εργασία, ως προς τον τομέα εργασίας, τότε το συμπέρασμα μας θα ήταν:

Η ικανοποίηση των εργαζομένων παρουσιάζει ομοιογένεια, ως προς τους τρεις τομείς εργασίας ($X^2 = 0.25$, $df = 2$, $sig. = 0.88$).

3 ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Η Γραμμική Παλινδρόμηση αποτελεί μία στατιστική μέθοδο, η οποία αποσκοπεί στον προσδιορισμό ενός μαθηματικού μοντέλου για την περιγραφή της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών, το οποίο θα μπορούσε εν δυνάμει να χρησιμοποιηθεί και ως ένα εργαλείο πρόβλεψης των τιμών της μίας μεταβλητής.

3.1 Απλή παλινδρόμηση

Θα θέλαμε να προσδιορίσουμε ένα μαθηματικό μοντέλο το οποίο να περιγράφει το βαθμό ικανοποίησης ενός εργαζομένου (**avrscore**), σε σχέση με ένα από τα επιμέρους χαρακτηριστικά τους.

Λογική Υπόθεση

Ο βαθμός ικανοποίησης ενός εργαζομένου (**avrscore**) μπορεί να περιγραφεί επαρκώς («εξαρτάται») από το βαθμό ικανοποίησης του σε σχέση με τις αποδοχές (**pay**) και συνεπώς θα επιθυμούσαμε να προσδιορίσουμε ένα μαθηματικό μοντέλο το οποίο να συσχετίζει τα δύο μεγέθη.

| | id | section | work | pay | promote | changejob | rwork | rpay | rpromote | avrscore |
|----|----|---------|------|-----|---------|-----------|-------|------|----------|----------|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | 3 | 2 | 3 | 59 |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | 4 | 3 | 3 | 67 |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | 4 | 3 | 2 | 65 |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | 4 | 3 | 2 | 67 |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | 3 | 3 | 4 | 70 |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | 3 | 2 | 4 | 65 |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | 3 | 3 | 2 | 56 |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | 3 | 2 | 4 | 63 |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | 3 | 4 | 2 | 62 |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | 3 | 2 | 3 | 57 |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | 3 | 2 | 1 | 44 |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | 4 | 3 | 1 | 56 |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | 4 | 2 | 3 | 58 |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | 4 | 2 | 3 | 61 |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | 3 | 2 | 3 | 54 |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | 4 | 3 | 3 | 77 |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | 4 | 3 | 2 | 58 |

Επιλογή του Μαθηματικού Μοντέλου

Το γραμμικό μοντέλο:

$$Y = \alpha + \beta X + \varepsilon$$

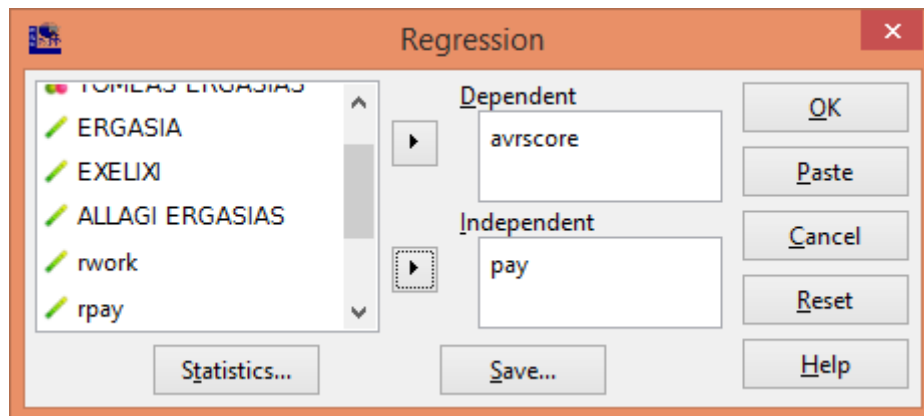
εξαρτημένη μεταβλητή ανεξάρτητη μεταβλητή σφάλμα

όπου, α , β και ε πραγματικοί αριθμοί

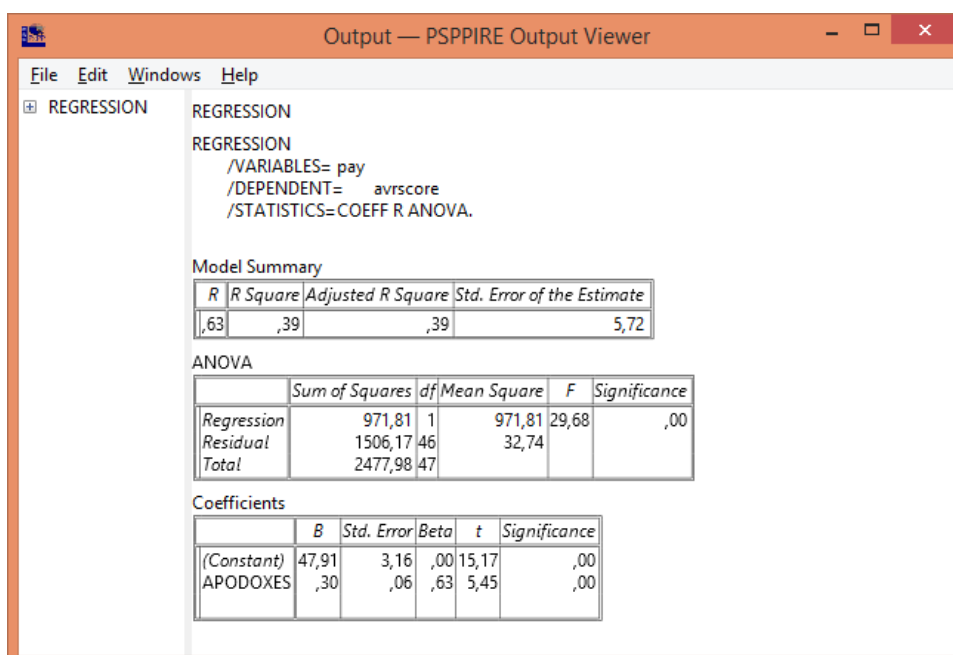
Διαδικασία

Analyze – Regression – Linear

Μεταφέρουμε τη μεταβλητή την οποία μελετούμε (ή / και θέλουμε να κάνουμε πρόβλεψη των τιμών της) στο πλαίσιο **Dependent** (**avrscore**) και την μεταβλητή, την οποία θα χρησιμοποιήσουμε για να ερμηνεύσουμε τις τιμές της πρώτης, στο πλαίσιο **Independent** (**pay**).



Αποτελέσματα



REGRESSION

REGRESSION
/VARIABLES= pay
/DEPENDENT= avrscore
/STATISTICS= COEFF R ANOVA.

Model Summary

| R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-----|----------|-------------------|----------------------------|
| ,63 | ,39 | ,39 | 5,72 |

ANOVA

| | Sum of Squares | df | Mean Square | F | Significance |
|------------|----------------|----|-------------|-------|--------------|
| Regression | 971,81 | 1 | 971,81 | 29,68 | ,00 |
| Residual | 1506,17 | 46 | 32,74 | | |
| Total | 2477,98 | 47 | | | |

Coefficients

| | B | Std. Error | Beta | t | Significance |
|------------|-------|------------|------|-------|--------------|
| (Constant) | 47,91 | 3,16 | ,00 | 15,17 | ,00 |
| APODOXES | ,30 | ,06 | ,63 | 5,45 | ,00 |

Ερμηνεία Αποτελεσμάτων

Πρώτος Πίνακας

Ο δείκτης R-square ($R^2 = 0.39$) εκφράζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τη διακύμανση των τιμών της ανεξάρτητης μεταβλητής. Δηλαδή στο παράδειγμα, το 39% της διακύμανσης του συνολικού βαθμού ικανοποίησης των εργαζομένων, ερμηνεύεται από τη διακύμανση των βαθμού ικανοποίησης από τις αποδοχές τους. Ο συντελεστής αυτός ονομάζεται συντελεστής προσδιορισμού και υποδεικνύει την ποιότητα προσαρμογής της εξίσωσης παλινδρόμησης στα δεδομένα.

Δεύτερος Πίνακας

Ο δεύτερος πίνακας είναι ο πίνακας ανάλυσης διασποράς (ANOVA = Analysis Of Variance). Εξετάζοντας τη τιμή **Significance** στη γραμμή **Regression** του πίνακα, βλέπουμε ότι η στάθμη σημαντικότητας του μοντέλου της παλινδρόμησης είναι $0.00 < 0.05$, άρα το μοντέλο που εφαρμόστηκε εξηγεί στατιστικά την εξαρτημένη μεταβλητή.

Τρίτος Πίνακας

Το μαθηματικό μοντέλο το οποίο προκύπτει, σύμφωνα με τον τρίτο πίνακα, είναι το ακόλουθο:

$$(\text{avrscore}) = 47.91 + 0.30 (\text{pay}) + \varepsilon$$

Έλεγχος του συντελεστή παλινδρόμησης β :

- $H_0 : \beta = 0$
- $H_a : \beta \neq 0$

Εφαρμογή του t-test στη γραμμή APODOXES του πίνακα:

Significance = 0.00 < 0.05, συνεπώς **απορρίπτεται η μηδενική υπόθεση**.

Άρα, ο συντελεστής παλινδρόμησης β είναι στατιστικά σημαντικά διάφορος του μηδέν ($t = 15.17$, $\text{sig} = 0.00$), συνεπώς ο η αξιολόγηση των αποδοχών (**pay**) ερμηνεύει στατιστικά σημαντικά τον βαθμό ικανοποίησης ενός εργαζομένου (**avrscore**).

Δεδομένου ότι ο συντελεστής $\beta = 0.30$ είναι στατιστικά σημαντικός, η τιμή του ερμηνεύεται ως εξής: όταν ο βαθμός ικανοποίησης από τις αποδοχές αυξάνεται κατά μία μονάδα, τότε ο συνολικός βαθμός ικανοποίησης αναμένεται να αυξηθεί κατά 0.30 μονάδες.

3.2 Πολλαπλή παλινδρόμηση

Θα θέλαμε να προσδιορίσουμε ένα μαθηματικό μοντέλο το οποίο να περιγράφει το βαθμό ικανοποίησης ενός εργαζομένου (**avrscore**) σε σχέση με περισσότερα από ένα από τα επιμέρους χαρακτηριστικά τους.

Λογική Υπόθεση

Ο βαθμός ικανοποίησης ενός εργαζομένου (**avrscore**) μπορεί να περιγραφεί επαρκώς («εξαρτάται») από το βαθμό ικανοποίησης του σε σχέση με τις συνθήκες εργασίας (**work**), τις αποδοχές (**pay**) και τις προοπτικές εξέλιξης (**promote**), οπότε θα επιθυμούσαμε να προσδιορίσουμε ένα μαθηματικό μοντέλο το οποίο να συσχετίζει τα τέσσερα μεγέθη.

| | id | section | work | pay | promote | changejob | nwork | rpay | rpromote | avrscore |
|----|----|---------|------|-----|---------|-----------|-------|------|----------|----------|
| 1 | 1 | 1 | 71 | 49 | 58 | 1 | 3 | 2 | 3 | 59 |
| 2 | 2 | 3 | 84 | 53 | 63 | 1 | 4 | 3 | 3 | 67 |
| 3 | 3 | 1 | 84 | 74 | 37 | 2 | 4 | 3 | 2 | 65 |
| 4 | 4 | 3 | 87 | 66 | 49 | 3 | 4 | 3 | 2 | 67 |
| 5 | 5 | 3 | 72 | 59 | 79 | 2 | 3 | 3 | 4 | 70 |
| 6 | 6 | 1 | 72 | 37 | 86 | 1 | 3 | 2 | 4 | 65 |
| 7 | 7 | 2 | 72 | 57 | 40 | 1 | 3 | 3 | 2 | 56 |
| 8 | 8 | 1 | 63 | 48 | 78 | 1 | 3 | 2 | 4 | 63 |
| 9 | 9 | 1 | 72 | 76 | 37 | 2 | 3 | 4 | 2 | 62 |
| 10 | 10 | 2 | 71 | 25 | 74 | 2 | 3 | 2 | 3 | 57 |
| 11 | 11 | 3 | 69 | 47 | 16 | 1 | 3 | 2 | 1 | 44 |
| 12 | 12 | 3 | 90 | 56 | 23 | 1 | 4 | 3 | 1 | 56 |
| 13 | 13 | 2 | 84 | 28 | 62 | 1 | 4 | 2 | 3 | 58 |
| 14 | 14 | 1 | 86 | 37 | 59 | 1 | 4 | 2 | 3 | 61 |
| 15 | 15 | 2 | 70 | 38 | 54 | 1 | 3 | 2 | 3 | 54 |
| 16 | 16 | 1 | 86 | 72 | 72 | 2 | 4 | 3 | 3 | 77 |
| 17 | 17 | 1 | 84 | 60 | 29 | 3 | 4 | 3 | 2 | 58 |

Επιλογή του Μαθηματικού Μοντέλου

Το γραμμικό μοντέλο:

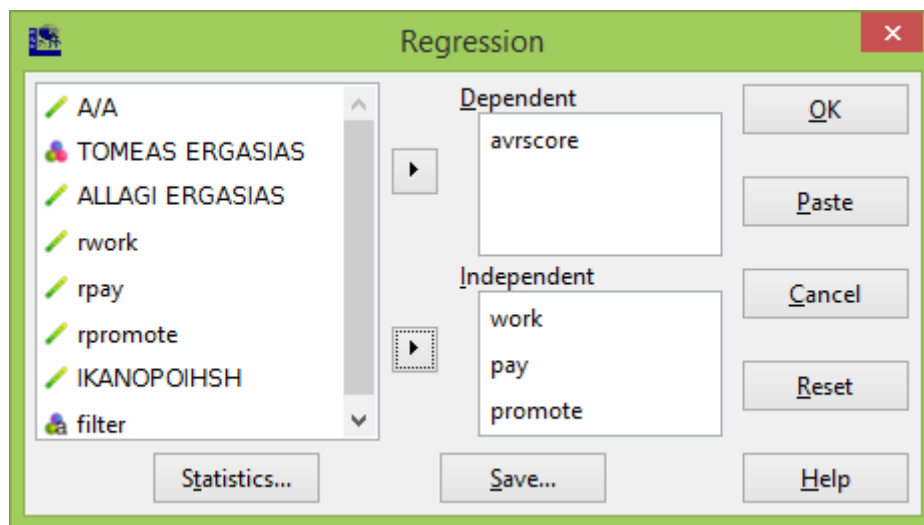
$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Εξαρτημένη Μεταβλητή Ανεξάρτητες Μεταβλητές Σφάλμα

Διαδικασία

Analyze – Regression – Linear

Μεταφέρουμε τη μεταβλητή την οποία μελετούμε (ή / και θέλουμε να κάνουμε πρόβλεψη των τιμών της) στο πλαίσιο **Dependent (avrscore)** και τις μεταβλητές, τις οποίες θα χρησιμοποιήσουμε για να ερμηνεύσουμε τις τιμές της πρώτης, στο πλαίσιο **Independent (work, pay, promote)**.



Αποτελέσματα

Output — PSPPIRE Output Viewer

File Edit Windows Help

REGRESSION
REGRESSION
/VARIABLES= work pay promote
/DEPENDENT= avrscore
/STATISTICS= COEFF R ANOVA.

Model Summary

| R | R Square | Adjusted R Square | Std. Error of the Estimate |
|------|----------|-------------------|----------------------------|
| 1,00 | 1,00 | 1,00 | ,27 |

ANOVA

| | Sum of Squares | df | Mean Square | F | Significance |
|------------|----------------|----|-------------|----------|--------------|
| Regression | 2474,76 | 3 | 824,92 | 11276,82 | ,00 |
| Residual | 3,22 | 44 | ,07 | | |
| Total | 2477,98 | 47 | | | |

Coefficients

| | B | Std. Error | Beta | t | Significance |
|------------|------|------------|------|--------|--------------|
| (Constant) | -,74 | ,45 | ,00 | -1,63 | ,11 |
| ERGASIA | ,34 | ,01 | ,38 | 67,66 | ,00 |
| APODOXES | ,34 | ,00 | ,69 | 121,39 | ,00 |
| EXELIXI | ,33 | ,00 | ,74 | 133,00 | ,00 |

Ερμηνεία Αποτελεσμάτων

Πρώτος Πίνακας

Ο δείκτης R-square ($R^2 = 1.00$) εκφράζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τη διακύμανση των τιμών της ανεξάρτητης μεταβλητής. Δηλαδή στο παράδειγμα, το 100% της διακύμανσης του συνολικού βαθμού ικανοποίησης των

εργαζομένων ερμηνεύεται από τη διακύμανση των αξιολογήσεων των συνθηκών εργασίας, των αποδοχών και των προοπτικών εξέλιξης. Αυτό είναι αναμενόμενο αν θυμηθούμε ότι η μεταβλητή **avrscore** είναι ο μέσος όρος των τριών επιμέρους αξιολογήσεων των εργαζομένων.

Δεύτερος Πίνακας

Η ερμηνεία των στοιχείων του δεύτερου πίνακα είναι ανάλογη με αυτή του παραδείγματος της απλής παλινδρόμησης. Εξετάζοντας τη τιμή **Significance** στη γραμμή **Regression** του πίνακα, βλέπουμε ότι η στάθμη σημαντικότητας του μοντέλου της παλινδρόμησης είναι $0.00 < 0.05$, άρα το μοντέλο που εφαρμόστηκε εξηγεί στατιστικά την εξαρτημένη μεταβλητή.

Τρίτος Πίνακας

Το μαθηματικό μοντέλο το οποίο προκύπτει, σύμφωνα με τον τρίτο πίνακα, είναι το ακόλουθο:

$$(\text{avrscore}) = -0.74 + 0.34 (\text{work}) + 0.34 (\text{pay}) + 0.33 (\text{promote}) + \varepsilon$$

Έλεγχος των συντελεστών παλινδρόμησης β :

- $H_0 : \beta = 0$
- $H_a : \beta \neq 0$

Εφαρμογή του t-test:

work Significance = $0.00 < 0.05$

pay Significance = $0.00 < 0.05$

promote Significance = $0.00 < 0.05$

Συνεπώς, οι τρεις επιμέρους αξιολογήσεις των εργαζομένων ερμηνεύουν στατιστικά σημαντικά την συνολική αξιολόγηση.

Παρατηρείστε ότι οι συντελεστές β που αντιστοιχούν στις ανεξάρτητες μεταβλητές **work**, **pay** και **promote** είναι περίπου ίσοι με $1/3$, όσο δηλαδή αναμένεται δεδομένου ότι η μεταβλητή **avrscore** είναι ο μέσος όρος των τριών επιμέρους αξιολογήσεων των εργαζομένων.

4 ΠΑΡΑΡΤΗΜΑ - Εργαστηριακή Άσκηση Υπόδειγμα

Έρευνα για τη μελέτη της ικανοποίησης από την εργασία

Στόχος της έρευνας αυτής ήταν να προσδιοριστεί σε ποιο βαθμό τα στελέχη της εταιρίας είναι ικανοποιημένα από την εργασία τους. Η έρευνα βασίστηκε στη συμπλήρωση ειδικών ερωτηματολογίων από τα στελέχη της εταιρίας που εργάζονται σε τρία διαφορετικά τμήματα του τμήματος marketing: στην έρευνα αγοράς, στις δημόσιες σχέσεις και στη διαφήμιση. Οι υπάλληλοι απάντησαν σε μία σειρά ερωτήσεων και στη συνέχεια αξιολόγησαν τις συνθήκες εργασίας τους, τις αποδοχές τους και τις δυνατότητες που έχουν για εξέλιξη στην εταιρία, στην κλίμακα 1-100 (με Άριστα = 100), όπως φαίνεται από τα πρωτογενή δεδομένα της έρευνας που δίνονται στον πίνακα της επόμενης σελίδας.

Οι Μεταβλητές της έρευνας

(1) Αύξων Αριθμός (A/A):

Ο αύξων αριθμός κάθε υπαλλήλου. Δεν αποτελεί μεταβλητή, αλλά χρησιμοποιείται προαιρετικά για λόγους ελέγχου των δεδομένων μας.

(2) Τομέας Εργασίας υπαλλήλου:

Εδώ έχουμε μία ποιοτική μεταβλητή με τρεις τιμές: "Έρευνα Αγοράς", "Δημόσιες Σχέσεις" και "Διαφήμιση". Στη μεταβλητή θα δώσουμε το κωδικό όνομα **section** και οι τιμές θα παρασταθούν αντίστοιχα με τους αριθμούς 1, 2 και 3 (ονομαστική κλίμακα).

(3) Αξιολόγηση Συνθηκών Εργασίας:

Είναι ο βαθμός αξιολόγησης των συνθηκών εργασίας και είναι μία ποσοτική συνεχής μεταβλητή. Σύμφωνα μάλιστα με την κλίμακα που δώσαμε, μπορεί να πάρει τιμές σε ένα εύρος 1-100. Θα χρησιμοποιήσουμε τον κωδικό **work** (αναλογική κλίμακα).

(4) Αξιολόγηση Αποδοχών:

Είναι ο βαθμός αξιολόγησης των αποδοχών του υπαλλήλου και είναι όπως και η προηγούμενη μία ποσοτική συνεχής μεταβλητή, που παίρνει τιμές στο διάστημα 1-100. Θα δώσουμε στη μεταβλητή τον κωδικό **pay** (αναλογική κλίμακα).

(5) Αξιολόγηση Δυνατοτήτων Εξέλιξης:

Είναι ο βαθμός αξιολόγησης των δυνατοτήτων που ο κάθε υπάλληλος εκτιμά πως έχει για περαιτέρω εξέλιξη μέσα στην εταιρία και είναι μία ποσοτική συνεχής μεταβλητή, που παίρνει τιμές στο διάστημα 1-100. Θα δώσουμε στη μεταβλητή τον κωδικό **promote** (αναλογική κλίμακα).

Παρατηρούμε ότι σε δύο περιπτώσεις έχουμε χαμένες τιμές. Στις περιπτώσεις με A/A 29 και 33 λείπει η αξιολόγηση των δυνατοτήτων εξέλιξης και η αξιολόγηση των αποδοχών αντίστοιχα. Οι δύο αυτές τιμές θα μπορούσαν να συμβολιστούν με τον αριθμό "9999". Αυτό γιατί οι μεταβλητές **promote** και **pay** έχουν τιμές τριψήφιους αριθμούς, οπότε ένας τετραψήφιος αριθμός θα ξεχωρίζει. Στο PSPP θα μπορούσαμε να πληκτρολογήσουμε στη θέση των χαμένων αυτών τιμών την **παύλα** (-), αφού το πρόγραμμα διακρίνει και αποδίδει στο σύμβολο αυτό την ύπαρξη χαμένων τιμών. Θα χρησιμοποιηθεί αυτός ο δεύτερος τρόπος κωδικοποίησης.

Πρωτογενή δεδομένα της έρευνας

| A/A | Τομέας εργασίας | Αξιολόγηση συνθηκών εργασίας | Αξιολόγηση αποδοχών | Αξιολόγηση δυνατοτήτων εξέλιξης |
|-----|-----------------|------------------------------|---------------------|---------------------------------|
| 1 | 1 | 71 | 49 | 58 |
| 2 | 3 | 84 | 53 | 63 |
| 3 | 1 | 84 | 74 | 37 |
| 4 | 3 | 87 | 66 | 49 |
| 5 | 3 | 72 | 59 | 79 |
| 6 | 1 | 72 | 37 | 86 |
| 7 | 2 | 72 | 57 | 40 |
| 8 | 1 | 63 | 48 | 78 |
| 9 | 1 | 72 | 76 | 37 |
| 10 | 2 | 71 | 25 | 74 |
| 11 | 3 | 69 | 47 | 16 |
| 12 | 3 | 90 | 56 | 23 |
| 13 | 2 | 84 | 28 | 62 |
| 14 | 1 | 86 | 37 | 59 |
| 15 | 2 | 70 | 38 | 54 |
| 16 | 1 | 86 | 72 | 72 |
| 17 | 1 | 84 | 60 | 29 |
| 18 | 2 | 90 | 62 | 66 |
| 19 | 1 | 73 | 56 | 55 |
| 20 | 1 | 94 | 60 | 52 |
| 21 | 2 | 84 | 42 | 66 |
| 22 | 2 | 85 | 56 | 64 |
| 23 | 3 | 88 | 55 | 52 |
| 24 | 3 | 74 | 70 | 51 |
| 25 | 2 | 71 | 45 | 68 |
| 26 | 2 | 88 | 49 | 42 |
| 27 | 1 | 90 | 27 | 67 |
| 28 | 1 | 85 | 89 | 46 |
| 29 | 3 | 79 | 59 | |
| 30 | 3 | 72 | 60 | 45 |
| 31 | 2 | 88 | 36 | 47 |
| 32 | 2 | 77 | 60 | 75 |
| 33 | 2 | 64 | | 61 |
| 34 | 2 | 87 | 51 | 57 |
| 35 | 2 | 77 | 90 | 51 |
| 36 | 3 | 71 | 36 | 55 |
| 37 | 3 | 75 | 53 | 92 |
| 38 | 2 | 74 | 59 | 82 |
| 39 | 3 | 76 | 51 | 54 |
| 40 | 1 | 95 | 66 | 52 |
| 41 | 2 | 89 | 66 | 62 |

| | | | | |
|----|---|----|----|----|
| 42 | 2 | 85 | 57 | 67 |
| 43 | 1 | 65 | 42 | 68 |
| 44 | 1 | 82 | 37 | 54 |
| 45 | 1 | 82 | 60 | 56 |
| 46 | 3 | 89 | 80 | 64 |
| 47 | 2 | 74 | 47 | 63 |
| 48 | 2 | 82 | 49 | 91 |
| 49 | 1 | 90 | 76 | 70 |
| 50 | 1 | 78 | 52 | 72 |

Κωδικοποίηση μεταβλητών

| A/A | section | work | pay | promote |
|-----|---------|------|-----|---------|
| 1 | 1 | 71 | 49 | 58 |
| 2 | 3 | 84 | 53 | 63 |
| 3 | 1 | 84 | 74 | 37 |
| 4 | 3 | 87 | 66 | 49 |
| 5 | 3 | 72 | 59 | 79 |
| 6 | 1 | 72 | 37 | 86 |
| 7 | 2 | 72 | 57 | 40 |
| 8 | 1 | 63 | 48 | 78 |
| 9 | 1 | 72 | 76 | 37 |
| 10 | 2 | 71 | 25 | 74 |
| 11 | 3 | 69 | 47 | 16 |
| 12 | 3 | 90 | 56 | 23 |
| 13 | 2 | 84 | 28 | 62 |
| 14 | 1 | 86 | 37 | 59 |
| 15 | 2 | 70 | 38 | 54 |
| 16 | 1 | 86 | 72 | 72 |
| 17 | 1 | 84 | 60 | 29 |
| 18 | 2 | 90 | 62 | 66 |
| 19 | 1 | 73 | 56 | 55 |
| 20 | 1 | 94 | 60 | 52 |
| 21 | 2 | 84 | 42 | 66 |
| 22 | 2 | 85 | 56 | 64 |
| 23 | 3 | 88 | 55 | 52 |
| 24 | 3 | 74 | 70 | 51 |
| 25 | 2 | 71 | 45 | 68 |
| 26 | 2 | 88 | 49 | 42 |
| 27 | 1 | 90 | 27 | 67 |
| 28 | 1 | 85 | 89 | 46 |
| 29 | 3 | 79 | 59 | - |
| 30 | 3 | 72 | 60 | 45 |
| 31 | 2 | 88 | 36 | 47 |
| 32 | 2 | 77 | 60 | 75 |

| | | | | |
|----|---|----|----|----|
| 33 | 2 | 64 | - | 61 |
| 34 | 2 | 87 | 51 | 57 |
| 35 | 2 | 77 | 90 | 51 |
| 36 | 3 | 71 | 36 | 55 |
| 37 | 3 | 75 | 53 | 92 |
| 38 | 2 | 74 | 59 | 82 |
| 39 | 3 | 76 | 51 | 54 |
| 40 | 1 | 95 | 66 | 52 |
| 41 | 2 | 89 | 66 | 62 |
| 42 | 2 | 85 | 57 | 67 |
| 43 | 1 | 65 | 42 | 68 |
| 44 | 1 | 82 | 37 | 54 |
| 45 | 1 | 82 | 60 | 56 |
| 46 | 3 | 89 | 80 | 64 |
| 47 | 2 | 74 | 47 | 63 |
| 48 | 2 | 82 | 49 | 91 |
| 49 | 1 | 90 | 76 | 70 |
| 50 | 1 | 78 | 52 | 72 |

Ενδεικτικά ερωτήματα

1. Περιγραφική στατιστική

Αρχικά, παρατηρούμε ότι στο δείγμα μας υπάρχει μία ποιοτική μεταβλητή (section) και τρεις ποσοτικές συνεχείς (work, pay και promote).

Η περιγραφική στατιστική θα γίνει με τις διαδικασίες Frequencies για τη μεταβλητή section και τη διαδικασία Descriptives για τις μεταβλητές work, pay και promote.

2. Κωδικοποίηση και μετασχηματισμός δεδομένων

Οι εντολές μετασχηματισμού και επιλογής δεδομένων θα εφαρμοστούν στα δεδομένα ως εξής:

(α) Να κωδικοποιηθούν οι μεταβλητές work, pay και promote σε 4 ομάδες, ως εξής:
 $1 - 24 \rightarrow 1$, $25 - 49 \rightarrow 2$, $50 - 74 \rightarrow 3$ και $75 - 100 \rightarrow 4$.

Να οριστούν οι νέες μεταβλητές rwork, rpay και rpromote, οι οποίες θα δεχτούν τις κωδικοποιημένες τιμές.

(β) Να υπολογιστούν για τον κάθε εργαζόμενο στο δείγμα ένας μέσος όρος αξιολόγησης της επιχείρησης $((\text{work} + \text{pay} + \text{promote})/3)$ ο οποίος θα εκφράζεται με ακέραιο αριθμό 0 – 100 και θα καταχωρηθεί σε μία νέα μεταβλητή avscore.

(γ) Να χαρακτηριστεί ο κάθε εργαζόμενος ικανοποιημένος από την εταιρεία του αν έχει δώσει μέση αξιολόγηση από 65 και πάνω, δημιουργώντας μία νέα μεταβλητή satisfy με τιμές 1 (ικανοποιημένος) και 0 (μη ικανοποιημένος).

(δ) Να επιλεγούν για ανάλυση μόνο εκείνα τα δεδομένα που αναφέρονται στους υπαλλήλους που εργάζονται στην έρευνα αγοράς.

3. Έλεγχοι υποθέσεων για τη μέση τιμή

(α) Μπορούμε να ισχυριστούμε με κίνδυνο σφάλματος 5%, ότι η μέση τιμή της αξιολόγησης συνθηκών εργασίας (work) είναι 78;

(β) Μπορούμε να ισχυριστούμε σε επίπεδο σημαντικότητας 5%, ότι η μέση τιμή της αξιολόγησης συνθηκών εργασίας (work) είναι ίδια στους εργαζόμενους που εργάζονται στον τομέα της έρευνας αγοράς και στις δημόσιες σχέσεις;

(γ) Μπορούμε να ισχυριστούμε σε επίπεδο σημαντικότητας 5%, ότι οι μέσες τιμές της αξιολόγησης συνθηκών εργασίας (work) και της αξιολόγησης των αποδοχών (pay) είναι ίσες;

4. Έλεγχοι ανεξαρτησίας - ομοιογένειας

(α) Να ελεγχθεί σε επίπεδο σημαντικότητας 5%, κατά πόσο ο τομέας εργασίας των εργαζομένων (section) και η μεταβλητή που εκφράζει την ικανοποίηση του εργαζομένου από την εργασία του (satisfy), είναι ανεξάρτητα μεταξύ τους χαρακτηριστικά.

(β) Να ελεγχθεί σε επίπεδο σημαντικότητας 5%, κατά πόσο η ικανοποίηση των εργαζομένων (satisfy) παρουσιάζει ομοιογένεια, εξεταζόμενη στους τρεις τομείς εργασίας (section).

5. Γραμμική Παλινδρόμηση

(α) Να εκτιμηθεί η εξίσωση της γραμμικής παλινδρόμησης, η οποία περιγράφει την εξάρτηση της μεταβλητής που εκφράζει τον μέσο όρο αξιολόγησης της επιχείρησης (avrscore), από την αξιολόγηση των αποδοχών (pay).

(β) Να εκτιμηθεί η εξίσωση της πολλαπλής γραμμικής παλινδρόμησης, η οποία περιγράφει την εξάρτηση της μεταβλητής που εκφράζει τον μέσο όρο αξιολόγησης της επιχείρησης (avrscore), από τις τρεις επιμέρους αξιολογήσεις των συνθηκών εργασίας (work), των αποδοχών (pay) και των δυνατοτήτων εξέλιξης (promote).