

# Ανάλυση σε ομάδες με τη χρήση του στατιστικού πακέτου SPSS (Cluster Analysis using SPSS)

---

## Περιεχόμενα:

1. Σκοπός
2. Στάδια ομαδοποίησης
3. Μέθοδοι ομαδοποίησης
4. Παράδειγμα
  - 4.1. Εισαγωγή δεδομένων
  - 4.2. Ομαδοποίηση
  - 4.3. Γραφική αναπαράσταση
  - 4.4. Επιλογή βέλτιστου αριθμού ομάδων
5. Επανάληψη παραδείγματος με άλλες ιεραρχικές μεθόδους ομαδοποίησης

## 1. Σκοπός

---

Σκοπός της ανάλυσης σε ομάδες είναι η ομαδοποίηση δεδομένων σε ομοιογενείς (homogeneous) ομάδες βάση διαφόρων μεταβλητών.

Οι ομάδες εσωτερικά είναι ομοιογενείς ενώ εξωτερικά ανομοιογενείς (heterogeneous).

**Άσκηση 1: Τι ακριβώς εννοούμε με την πρόταση «Οι ομάδες εσωτερικά είναι ομοιογενείς ενώ εξωτερικά ανομοιογενείς (heterogeneous)».**

## 2. Στάδια ομαδοποίησης

---

Ο βασικός σκοπός της ανάλυσης σε ομάδες είναι η κατηγοριοποίηση όμοιων παρατηρήσεων σε ομάδες. Τα βασικά στάδια ομαδοποίησης συνοψίζονται στα πιο κάτω:

- 🔗 Ποσοτικοποίηση της ομοιογένειας. Ποιο μέτρο ομοιογένειας χρησιμοποιούμε;

.....

- 🔗 Επιλογή μεθόδου ομαδοποίησης. Πώς δημιουργούνται οι ομάδες;

.....

- ✚ Καθορισμός του βέλτιστου αριθμού ομάδων. Πόσες ομάδες να συντάξουμε;

.....

- ✚ Επαλήθευση αποτελεσμάτων. Με ποια μέθοδο;

.....

### 3. Μέθοδοι ομαδοποίησης

---

Υπάρχουν δύο μεγάλες κατηγορίες μεθόδων ομαδοποίησης:

- ✚ Ιεραρχικές διαδικασίες: συσσωρευτικές
- ✚ Μη-ιεραρχικές διαδικασίες

#### Ιεραρχικές διαδικασίες

- ✚ Το χαρακτηριστικό αυτών των μεθόδων είναι ότι ακολουθούν συγκεκριμένα βήματα για την ομαδοποίηση των δεδομένων. Σταδιακά δημιουργούνται μεγαλύτερες ομάδες με την ενοποίηση μικρότερων ομάδων που δημιουργήθηκαν σε προηγούμενα στάδια.
- ✚ Στο πρώτο στάδιο ενώνονται οι παρατηρήσεις με τη μικρότερη απόσταση.
- ✚ Στο τελικό στάδιο όλες οι παρατηρήσεις ενοποιούνται, ενώ σε κάθε στάδιο μόνο οι πιο κοντινές παρατηρήσεις ενώνονται.
- ✚ Όλα τα στάδια ομαδοποίησης συνοψίζονται στο Δενδρόγραμμα.
- ✚ Υπάρχουν διαφορετικές ιεραρχικές μέθοδοι ανάλογα με τον ορισμό της απόστασης μεταξύ δύο σημείων.

Οι πιο διαδεδομένες είναι οι:

- ✚ Μέθοδος Απλής Σύνδεσης (Single Linkage)
- ✚ Πλήρης Σύνδεσης (Complete Linkage)
- ✚ Μέσου Συνδέσμου (Average Linkage)
- ✚ Ward's Method
- ✚ Κεντροειδής Μέθοδος (Centroid Method)

**Άσκηση 2: Πως διαφοροποιούνται οι τρεις πρώτες διαδικασίες;**

## 4. Παράδειγμα

---

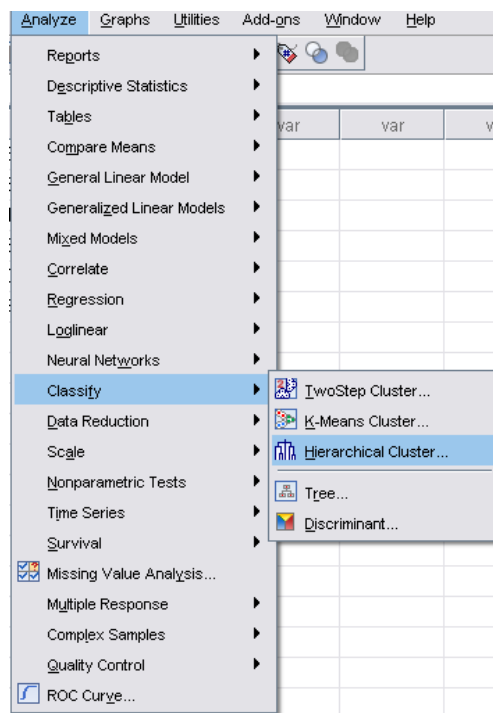
### 4.1. Εισαγωγή δεδομένων

**Άσκηση 3: Εισαγάγετε το πιο κάτω δεδομένα στο SPSS.**

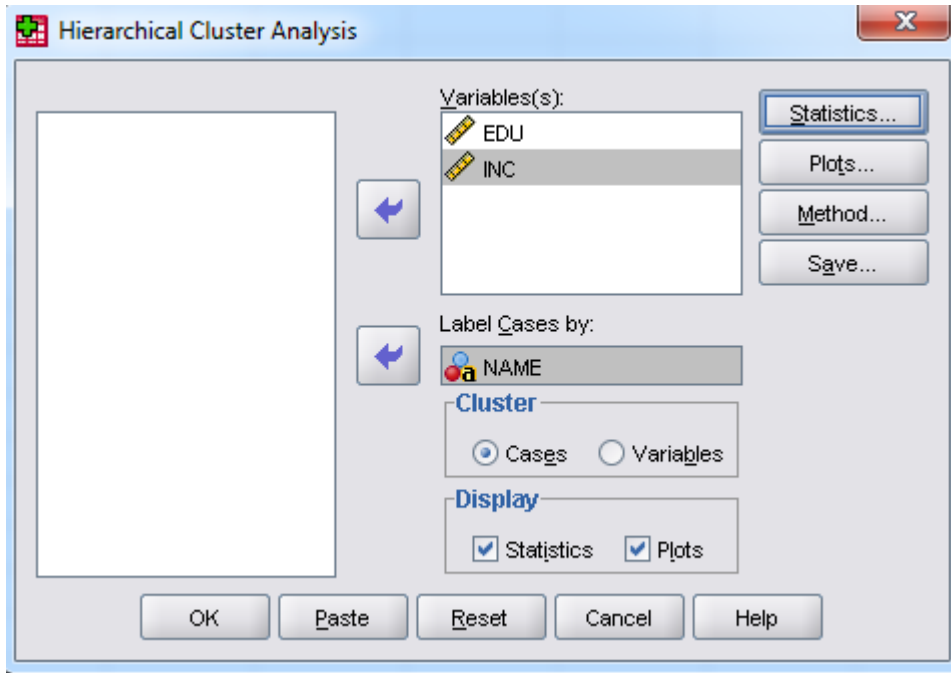
Income	5	6	15	16	25	30
Education	5	6	14	15	20	19
NAME	S1	S2	S3	S4	S5	S6

### 4.2. Ομαδοποίηση

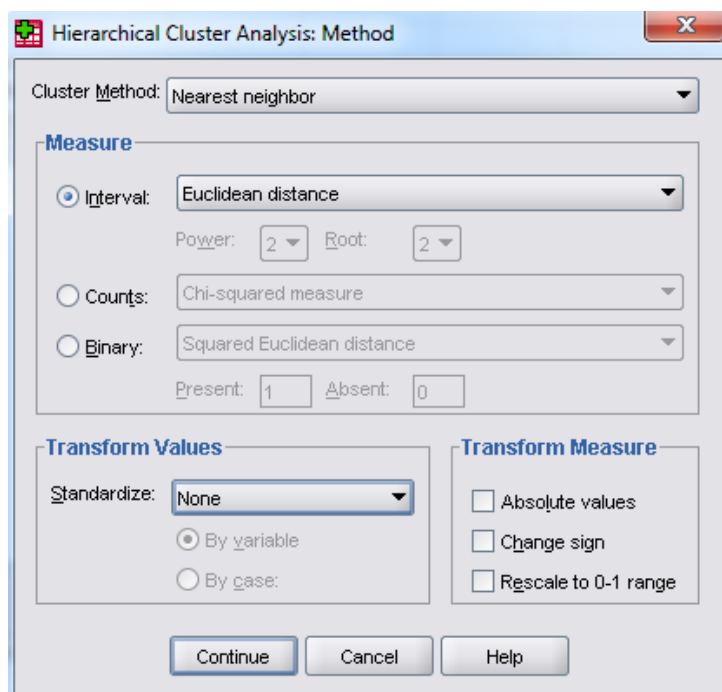
Για τη διαδικασία ομαδοποίησης επιλέγουμε *Analyze* → *Classify* → *Hierarchical Clustering*.



- ✚ Στο παράθυρο που εμφανίζεται επιλέξτε όλες τις μεταβλητές που θέλουμε να διερευνήσουμε. Επιλέξτε το “NAME” σαν **“Label”**.
- ✚ Σημειώστε ότι επιλέγουμε να ομαδοποιήσουμε περιπτώσεις **“Cases”** αντί μεταβλητές **“Variables”**. Αυτό σημαίνει ότι θέλουμε να κατηγοριοποιήσουμε τις παρατηρήσεις (γραμμές) στα δεδομένα μας.

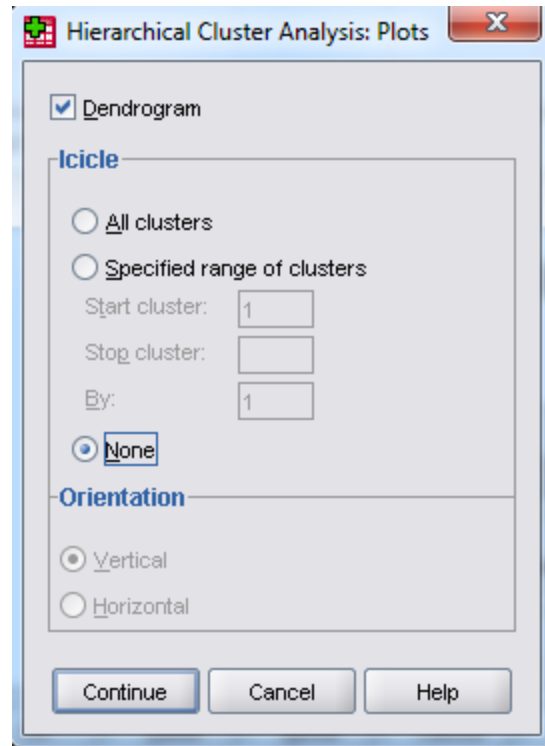


- ✚ Στην επιλογή **Statistics** επιλέξτε την δημιουργία Συσσωρευτικού Σχεδίου (**Agglomeration Schedule**) και Πίνακα Εγγύτητας (**Proximity Matrix**).
- ✚ Στην επιλογή **Method** διαλέγουμε **Nearest Neighbor**. Αυτή η επιλογή είναι συνώνυμη με την μέθοδο Single Linkage. Άλλες επιλογές όπως Furthest Neighbor ή Between-groups linkage είναι συνώνυμα των Complete and Average Linkage, αντίστοιχα.
- ✚ Επιπλέον επιλέξτε **Interval** → **Euclidean Distance**.
- ✚ Στην επιλογή Standardize → None. Εάν απαιτούσε η φύση των δεδομένων την τυποποίηση των μονάδων μέτρησης των μεταβλητών θα επιλέγαμε Z-scores.



### 4.3. Γραφική Αναπαράσταση

Για τη γραφική αναπαράσταση των ομάδων επιλέγουμε τη δημιουργία Δενδρογράμματος *Plots* → *Dendrogram*. Ενώ στην επιλογή *Icicle Plot* → *None*.



**Άσκηση 3: Εξηγήστε τα αποτελέσματα. Τι δείχνεις το Agglomeration Schedules; Ποιος είναι ο βέλτιστος αριθμός ομάδων που δημιουργείται;**

5. Επανάληψη παραδείγματος με άλλες ιεραρχικές μεθόδους ομαδοποίησης

---

**Άσκηση 4: Επαναλάβετε την ομαδοποίηση χρησιμοποιώντας την μέθοδο Πλήρους Σύνδεσης και Μέσης Σύνδεσης. Συγκρίνεται τα αποτελέσματα. Ποιος είναι ο βέλτιστος αριθμός ομάδων που δημιουργείται σε κάθε περίπτωση; Διαφέρει και γιατί;**