

Contents

- 0.1 Linkage methods 2
 - 0.1.1 Ward’s method 2
 - 0.1.2 Centroid method 2

The hierarchical clustering procedure attempts to identify relatively homogeneous groups of cases (or variables) based on selected characteristics. For example: cluster television shows into homogeneous groups based on viewer characteristics. In hierarchical clustering, an algorithm is used that starts with each case (or variable) in a separate cluster and combines clusters until only one is left.

To cluster cases you need to identify variables you wish to be considered in creating clusters for the cases. The variables to be used for cluster formation are here: picture quality (5 measures), reception quality (3 measures), audio quality (3 measures), ease of programming (1 measure), number of events (1 measure), number of days for future programming (1 measure), remote control (3 measures), and extras (3 measures). Pass these in the Variable(s) box.

Cluster Method: Choose the procedure for combining clusters. The default procedure is called the between-group linkage. SPSS computes the smallest average distance between all group pairs and combines the two groups that are closest. The procedure begins with as many clusters as there are cases (here: 21). At step one, the two cases with the smallest distance between them are clustered. Then SPSS computes distances once more and combines the two that are next closest. After the second step you will have either 18 individual cases and one cluster of 3 cases, or 17 individual cases and two clusters of two cases each. The process continues until all cases are grouped into one large cluster. Measure: Indicate what method is used for distance measuring, the default is Squared Euclidean distance.

0.1 Linkage methods

- Single linkage (minimum distance)
- Complete linkage (maximum distance)
- Average linkage

0.1.1 Ward's method

- Compute sum of squared distances within clusters
- Aggregate clusters with the minimum increase in the overall sum of squares

0.1.2 Centroid method

The distance between two clusters is defined as the difference between the centroids (cluster averages)