

Factor analysis and clustering of a mixture of quantitative and categorical forest biometric and soil data

Kitikidou, K. and Arambatzis, N.

Dimokritos University of Thrace, Aristotle University of Thessaloniki, Greece

Analyzing and clustering units described by a mixture of sets of quantitative and categorical variables is a relevant challenge. Principal components analysis was used to include these two types of variables in order to study the correlations of a large number of forest variables by grouping the variables in factors. Data were extracted from the management study of the university forest of Taxiarchis (Northern Greece). The variables used in the analysis were: site quality, exposure, petrification, soil, altitude, incline, age, crown diameter, tree height, section size and timber yield. The principal components method of factors extraction (PCA) begins by finding a linear combination of variables (a component) that accounts for as much variation in the original variables as possible. It then finds another component that accounts for as much of the remaining variation as possible and is uncorrelated with the previous component, continuing in this way until there are as many components as original variables. A few components accounted for most of the variation, and these components were used to replace the original variables. Thus, this method reduced the number of variables in the data file. Further on, hierarchical cluster analysis (HCA) was used for detecting groupings in data. The objects in these groups were cases (forest sections). HCA revealed natural groupings within the forest sections that were not apparent otherwise.