Cluster Analysis of Car Data: a Trial Run for France 1989

Brian Everitt

The 'counts' data for 15 cars in France 1989, were first converted to proportions using the row totals. Two 'distance' matrices were then constructed:

a) 1-Sum of the corresponding upper and lower triangle elements in the proportion matrix.

b) Distance matrix from (a) + Abs (diff of corresponding upper and lower triangle elements in the proportion matrix).

Complete linkage clustering was applied to each distance matrix with the results indicated by the two figures. Changes in the structure of the two dendrograms reflect the asymmetry aspect of the sales of the cars. For example, in the first figure ROV and VW appear together relatively early in the dendrogram reflecting that a fair proportion of Rovers and/or VWs are exchanged. In the second diagram the connection disappears because of the high asymmetry of the ROV/VW relationship i.e. about 10% of Rovers are exchanged for VWs, and only about 1% of VWs are exchanged for Rovers.

The most stable groups appear to be those involving Renault, Peugeot, Citroen, Fiat and BMW, Mercedes, Saab and Volvo.

For interest, the 2-dimensional solutions from classical MDS are also included.

Plot 1

coord 2

Lada

Merc
BMW

FGM Peu

Saab   Rov Fiat
Cit
Ren

VW Alfa

Volv   Seat

-0.4   -0.2   0.0   0.2

coord 1

Plot 2

coord 2

Rov
Lada
Seat

Fiat   Alfa

Ford GM   Saab

Cit   VW   Volv

Ren
Peu

BMW
Merc

-0.3 -0.2 -0.1 0.0 0.1 0.2 0.3

coord 1