

Lösung zu Kapitel 12: Beispiel 2

In einer Studie zu Charakteristika von Supermärkten wurden 637 Personen in Wales über die Bedeutung mehrerer Eigenschaften von Supermärkten auf einer Skala, die von 1 (wenig wichtig) bis 5 (sehr wichtig) reichte, befragt.

Im Datenfile `smarkt.dat` sind es die Variablen `q08a-q08n, q08q-q08v`.

- Gibt es unter den Befragten Gruppen mit ähnlichen Erwartungen, was ein Supermarkt bieten soll?

Nach dem Einlesen der Daten und Eliminieren wir Beobachtungen mit fehlenden Werten und nehmen nur die Variablen in den neuen Datensatz (`smarkt1`) mit, die in die Clusteranalyse eingehen.

Das Clustern der Variablen erledigen wir durch eine Clusteranalyse der transponierten Datenmatrix.

R

```
> library("cluster")
> smarkt <- read.table("smarkt.dat", header = TRUE)
> attach(smarkt)
> smarkt1 <- na.omit(smarkt[6:25])
> detach(smarkt)
> clustvars <- agnes(t(smarkt1))
> pltree(clustvars, main = "Clustern von Variablen")
```

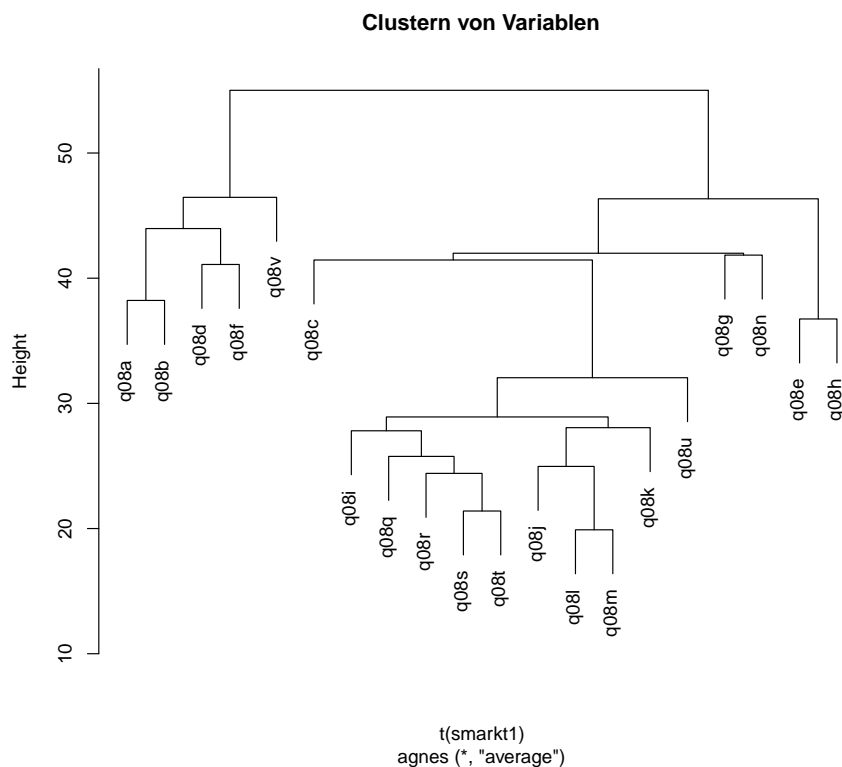


Abbildung 1: Dendrogramm für das Clustern von Variablen

Setzt man im Dendrogramm (► Abbildung 1) in der Höhe von etwa 35 dem Aggregationsprozess ein Ende, erhält man fünf Variablencluster. Wir geben nur den letzten Buchstaben des Variablennamens:

- a, b, d, f, v
- c
- g, n
- e, h
- restliche zehn Variablen