

Lösung zu Kapitel 10: Beispiel 1

Supermarkt

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 (siehe auch Fallbeispiel 24).

Im Datenfile **superm.sav** sind es die Variablen **q08a - a08n**, **q08q - q08v**.

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

Für diesen Datensatz wurde aufgrund der sehr hohen Fallzahl der Two-Step Clusteralgorithmus angewandt. Die Variablen wurden alle als metrisch aufgefasst. Es wurden insgesamt 20 Variablen eingespielt.

PASW

Analysieren

Klassifizieren

Two-Step-Clusteranalyse...

entsprechende Variablen ☐ Stetige Variable(n):

(hier q08a bis q08v)

unter ☐ Diagramme...

Rang der Wichtigkeit für Variablen ☒

1 Plot pro Variable durch Markieren von Nach Cluster

1 Plot pro Cluster durch Markieren von Nach Variablen

Clusterverteilung

	N	% der Kombination	% der Gesamtsumme
Cluster 1	316	63,6%	49,6%
Cluster 2	181	36,4%	28,4%
Kombiniert	497	100,0%	78,0%
Ausgeschlossene Fälle	140		22,0%
Gesamtwert	637		100,0%

Abbildung 1: Clusterverteilung.

In den Daten konnten zwei Cluster von Gruppen mit ähnlichen Erwartungen, was ein Supermarkt bieten soll, gefunden werden (► Abbildung 1). Cluster 1 enthält dabei die meisten Beobachtungen. Auf die Darstellung der Clusterzentroide wurde in diesem Fall aus Platzgründen verzichtet (20 Variablen, das bedeutet es müssten 20 Clusterzentroide angeführt werden).

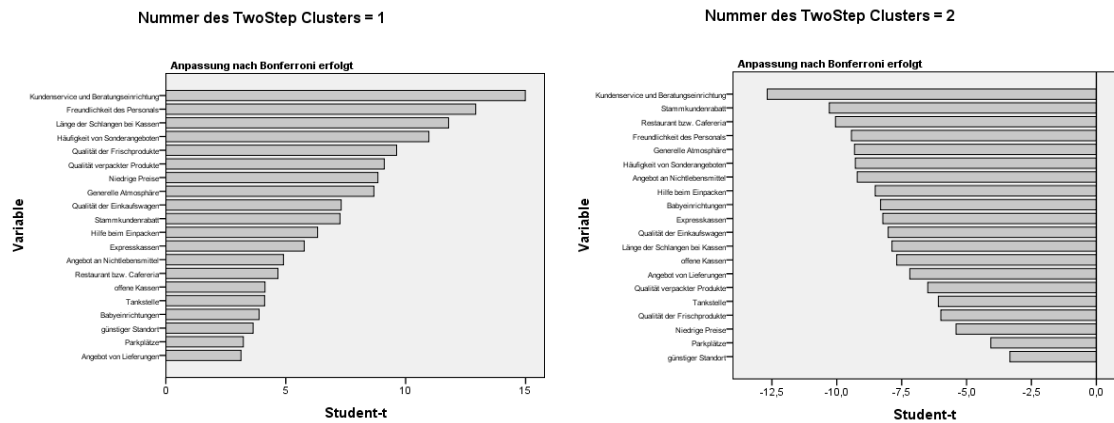


Abbildung 2: Wichtigkeit nach Variablen, Cluster 1 und Cluster 2.

Anscheinend sind für die Bildung der beiden Cluster in etwa die selben Variablen von hoher Wichtigkeit, nur in entgegengesetzter Richtung (► Abbildung 2). In Cluster 1 sind die Antworten überdurchschnittlich hoch (auf der Likertskala bedeutet dies *sehr wichtig*) und in Cluster 2 überdurchschnittlich niedrig (also *wenig wichtig*). Die größte Bedeutung für die Clusterbildung kann man wohl der Variable Kundenservice und Beratungseinrichtung zusprechen (► Abbildung 3).

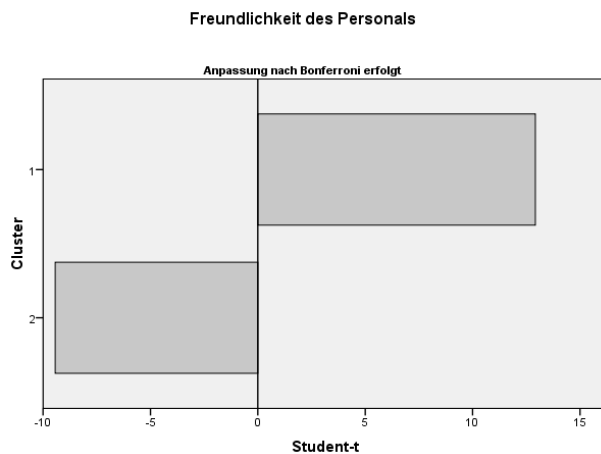


Abbildung 3: Wichtigkeit nach Cluster, Kundenservice und Beratungseinrichtung.

Nun wird der Clusteralgorithmus ein zweites Mal ohne die Variable Kundenservice und Beratungseinrichtung durchlaufen. Man sieht deutlich wie sich die beiden Cluster ändern, die Polung der Wichtigkeit der Variablen für die Clusterbildung ist nun genau entgegengesetzt und auch die Clusteraufteilung ist nun viel ausgewogener.

Clusterverteilung

	N	% der Kombination	% der Gesamtsumme
Cluster 1	254	51,0%	39,9%
Cluster 2	244	49,0%	38,3%
Kombiniert	498	100,0%	78,2%
Ausgeschlossene Fälle	139		21,8%
Gesamtwert	637		100,0%

Abbildung 4: Clusterverteilung, ohne Kundenservice und Beratungseinrichtung.

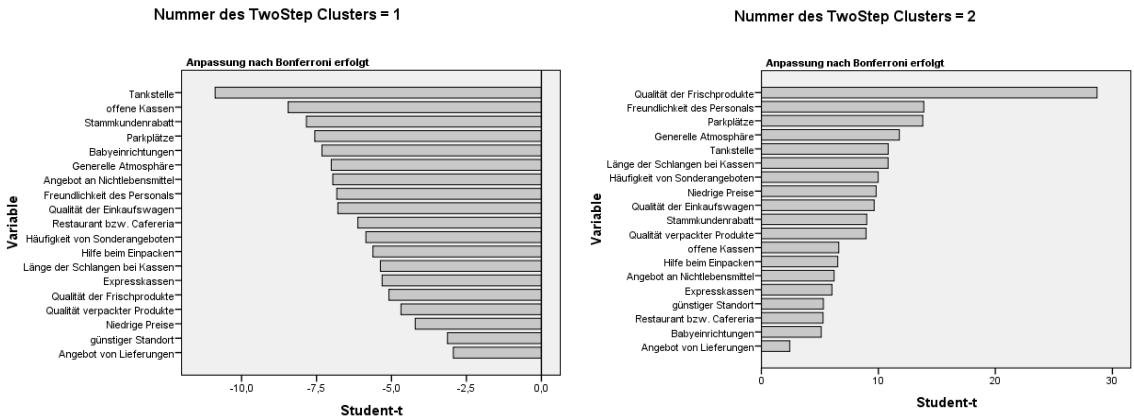


Abbildung 5: Wichtigkeit nach Variablen, Cluster 1 und Cluster 2 ohne Kundenservice und Beratungseinrichtung.