

Autorin: Antje Seidel-Schulze, Deutsches Institut für Urbanistik, Berlin

## Präsentation der Ergebnisse von Clusteranalysen

Der folgende Themenkomplex beantwortet die Frage, wie die von Statistikprogrammen automatisch erstellten Tabellen in Clusteranalyseprozeduren interpretiert werden und zeigt auf, welche Möglichkeiten es für die Darstellung von Ergebnissen der Clusteranalyse gibt.

### 1. Interpretation der Ausgabedatei von SPSS

a) Was ist eine Distanz- bzw. Näherungsmatrix?

Die erste Ausgabe, die SPSS produziert, zeigt die Distanz- oder Näherungsmatrix. Sie stellt die Situation vor dem ersten Schritt des Clusterverfahrens dar und zeigt die Abstände zwischen sämtlichen möglichen Paaren, die sich aus der Spalte/Zeile ergeben. Als Abstandsmaß ist von SPSS als Voreinstellung die quadrierte euklidische Distanz vorgegeben. Bei dieser kann es auch belassen werden, da es sich um metrische Daten handelt.

Abbildung 1: Distanzmatrix zeigt Abstand der Fälle zueinander

Fall	Bruess	Liege	antwerp	Prah	Ostrav	6: obenh	ven	Aalbor	Berlin	Dortmu	Esse	:Frankf	Oc	Hamb	Leipzig	Mu
1:Bruessel	,000	1,771	,504	3,121	3,107	5,588	5,935	3,317	7,245	5,061	14,331	8,114	0,610	9,7		
2:Liege	1,771	,000	2,127	3,310	2,251	3,246	3,028	4,693	7,892	7,748	15,892	13,371	0,581	18,2		
3:Antwerpen	,504	2,127	,000	4,571	4,740	5,835	5,547	5,546	8,956	7,896	17,323	10,090	2,279	10,9		
4:Praha	3,121	3,310	4,571	,000	9,141	1,254	9,798	5,701	1,603	3,814	16,832	6,661	5,870	5,0		
5:Ostrava	3,107	2,251	4,740	9,141	,000	0,822	3,475	9,883	2,785	3,199	7,365	7,211	4,827	13,0		
6:Kobenhav	5,588	3,246	6,835	1,254	0,822	,000	7,604	4,497	10,300	7,781	13,992	3,699	4,004	1,4		
7:Aalborg	5,935	3,028	6,547	9,798	3,475	7,604	,000	1,025	2,553	3,219	6,077	3,446	4,308	6,1		
8:Berlin	3,317	4,693	15,546	5,701	9,883	4,497	1,025	,000	3,583	2,626	3,383	6,097	2,750	16,3		
9:Dortmund	7,245	7,892	8,956	1,603	2,785	0,300	2,553	3,583	,000	,350	1,603	3,036	,433	10,3		
10:Essen	6,061	7,748	7,896	3,814	3,199	7,781	3,219	2,626	,350	,000	1,963	2,008	1,078	8,6		
11:Frankfur Ode	4,331	5,892	17,323	3,832	7,365	3,992	5,077	3,383	1,603	1,963	,000	3,642	1,264	12,6		
12:Hamburg	8,114	3,371	10,090	5,661	7,211	3,699	3,446	3,097	3,036	2,008	3,642	,000	4,571	2,8		
13:Leipzig	0,610	0,581	12,279	5,870	4,827	4,004	4,308	2,750	,433	1,078	1,264	4,571	,000	13,6		
14:Munche	9,732	3,203	10,942	5,054	3,031	1,414	3,190	3,359	0,385	3,616	12,601	2,874	3,678	,0		
15:Tallinn	4,594	9,627	7,457	2,317	5,540	2,500	3,223	0,993	6,350	4,607	8,629	3,002	9,695	4,3		

Bei genauer Betrachtung der Werte lässt sich so z.B. sagen (s. Kreise in Abbildung 1), dass der Abstand zwischen den Fällen Antwerpen (Spalte) und Brüssel (Zeile) mit 0.50 sehr gering, der Abstand zwischen Antwerpen und Frankfurt/Oder mit 17,32 sehr groß ist. Anders gesagt – je kleiner die Werte, desto näher liegen Städte beieinander, desto ähnlicher sind sie; je größer die Werte, desto unähnlicher sind Städte.

Die entstehende Matrix ist übrigens symmetrisch aufgebaut, das heißt, jeder Wert ist zweimal vorhanden, so dass eigentlich nur eine Hälfte der Tabelle (oberhalb oder unterhalb der Diagonalen) betrachtet werden muss.

b) Was ist eine Agglomerationstabelle oder Zuordnungsübersicht?

Eine weitere Ausgabedatei und gewissermaßen die erste Übersicht der Clusteranalyse ist die Zuordnungsübersicht (s. Abbildung 2). Diese erhält man in SPSS über das Untermenü „Statistik“. In der Übersicht beschreibt jede Zeile eine Stufe der Clusterbildung.



Man erhält diese Ausgabe standardmäßig in SPSS, kann aber im Menü auch andere Formen der Visualisierung, z.B. das Eiszapfendiagramm wählen. Auch hier wird gezeigt, dass ähnliche Fälle als Nachbarn gruppiert werden.

#### d) Zusammenfassung

Die vorgestellten Tabellen und Übersichten, sind trotz des hohen Informationsgehalts nur bedingt geeignet, um ohne weiteres in Veröffentlichungen verarbeitet zu werden. Sofern die Zielgruppe Fachpublikum ist, dürfte das kein Problem sein. Meist aber ist die Zielgruppe eine breitere Öffentlichkeit insbesondere in Politik und Verwaltung. Für diesen Fall dürfte aus der Auswahl der vorgestellten drei Übersichten einzig das Dendrogramm von einigem Interesse sein.

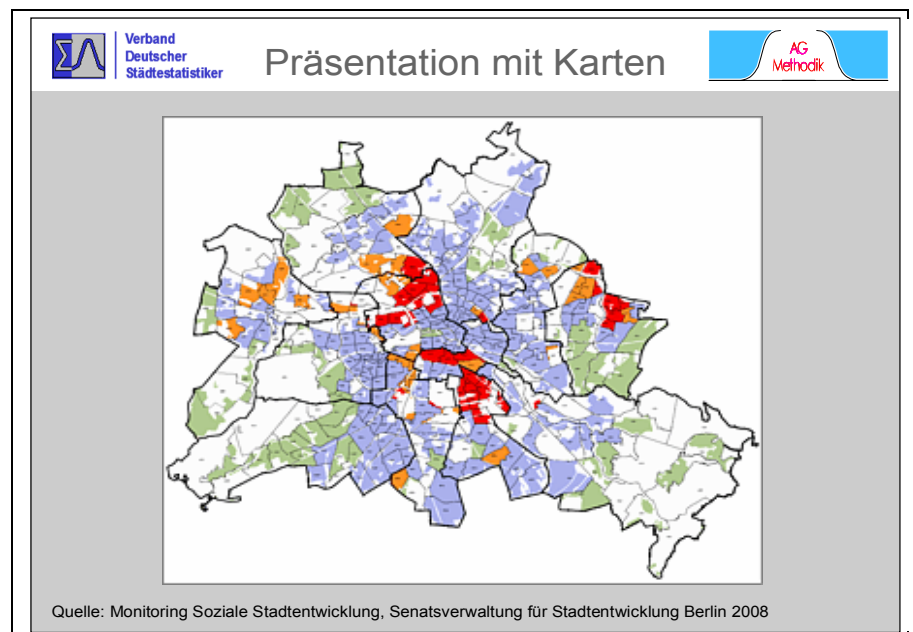
## 2. Präsentation der Ergebnisse

Im ersten Teil des Artikels wurden die Ergebnisausgaben vorgestellt, die das Statistikprogramm SPSS bereitstellt. Um die Ergebnisse einer breiteren Öffentlichkeit zugänglich zu machen, ist es nun noch notwendig, die Ergebnisse etwas leichter „verdaulich“ aufzubereiten und zu interpretieren. Vor allem geht es darum, die mittels Clusteranalyse ermittelten Typen zu charakterisieren.

#### Erstellung von Karten

Sofern es sich um räumliche Daten handelt (was in der vergleichenden Stadtforschung meist die Bezugsebene ist), ist die Visualisierung der Ergebnisse in Karten ein sehr anschauliches Mittel (s. Abbildung 4). Karten haben sowohl für Fachpublikum, als auch für Politik, Verwaltung und Bürger einen hohen Präsentationswert und sollten je nach Gestaltung der Kartenlegenden oder ergänzenden Texte auch einen hohen Informationsgehalt aufweisen.

Abbildung 4: Visualisierung der Ergebnisse mit Karten und begleitenden Texten

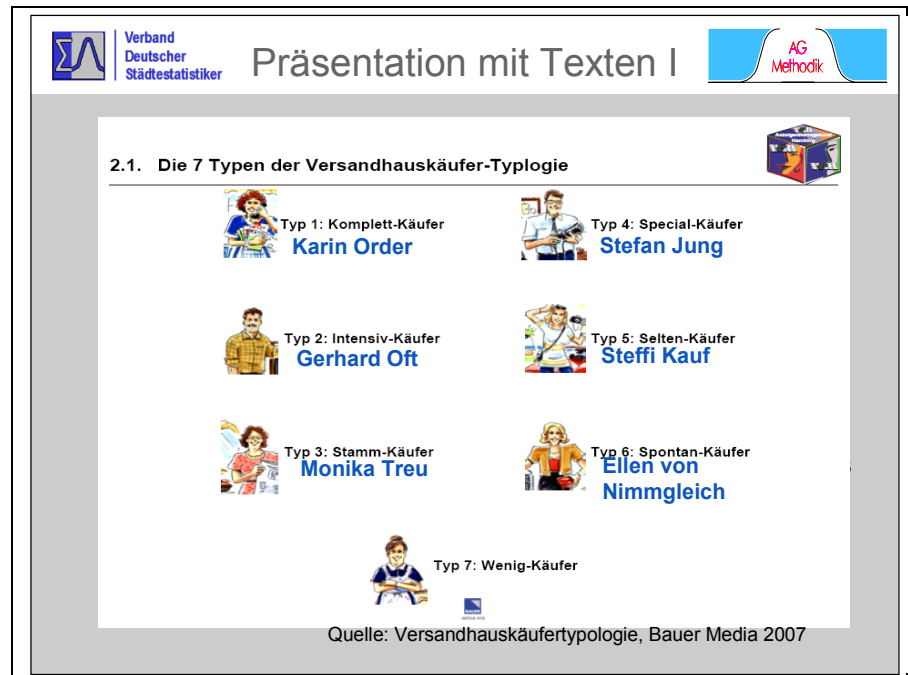


#### Texte

Bei der Verfassung der Texte zur Beschreibung der Cluster sollten möglichst treffende Überschriften für die Cluster oder Typen gefunden werden, was im Fall von Stadt- oder Stadtteiltypologien nicht immer einfach, aber möglich ist. Zur Erläuterung, inwieweit treffende Bezeichnungen bereits Teil der Ergebnispräsentation sind, soll hier (s. Abbildung 5) ein Beispiel für eine Clusteranalyse aus dem Marketingbereich des Versandhaushandels vorgestellt werden.

Hier wurden die Kundinnen und Kunden eines Versandhauses nach demografischen, sozialstrukturellen Merkmalen, Gewohnheiten der Freizeitgestaltung und ihrem Einkaufs- und Bestellverhalten typisiert und für jeden Typ ein imaginärer „Name“ vergeben.

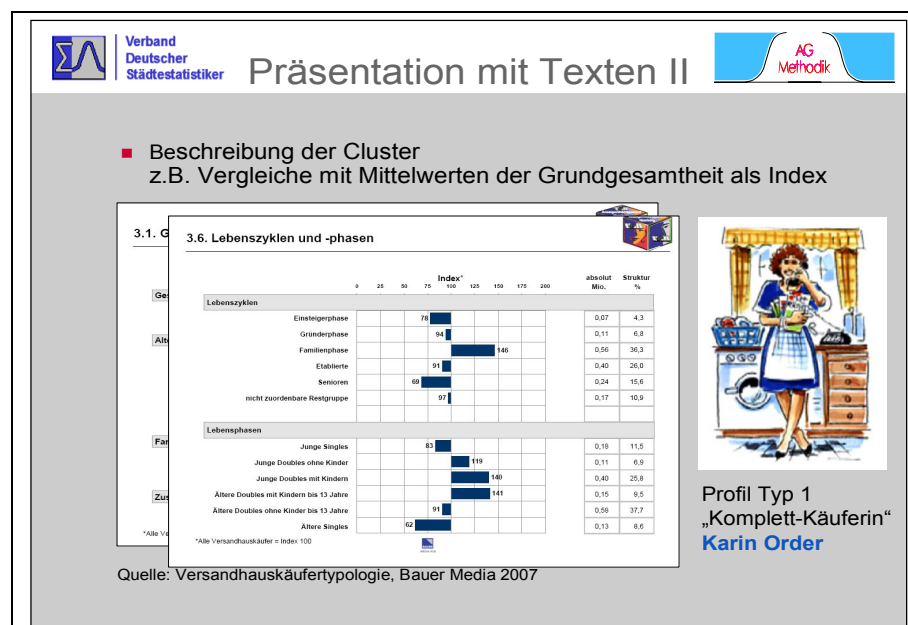
Abbildung 5: Prägnante Kurztitel für Cluster



Diese Namen sind so treffend gewählt, dass sie bereits beim Lesen ein Bild des jeweiligen Käufertypus entstehen lassen und eine Idee über demografische, soziale und kulturelle Merkmale vermitteln.

Diese Merkmale können dann – wie in dieser Typologie auch geschehen – in einem weiteren Schritt als Steckbrief zur Beschreibung der Cluster dienen (s. Abbildung 6). Um die wesentlichen Merkmale der Cluster herauszustellen, bietet sich die Bezugnahme zum Mittelwert der Grundgesamtheit an.

Abbildung 6: Beschreibung der Cluster über Mittelwertvergleiche



Abweichungen der Clustermittelwerte vom Gesamtmittelwert zeigen deutlich, welche Merkmale durch ihre besonders hohe Abweichung vom Gesamtmittelwert für das jeweilige Cluster besonders typisch sind.

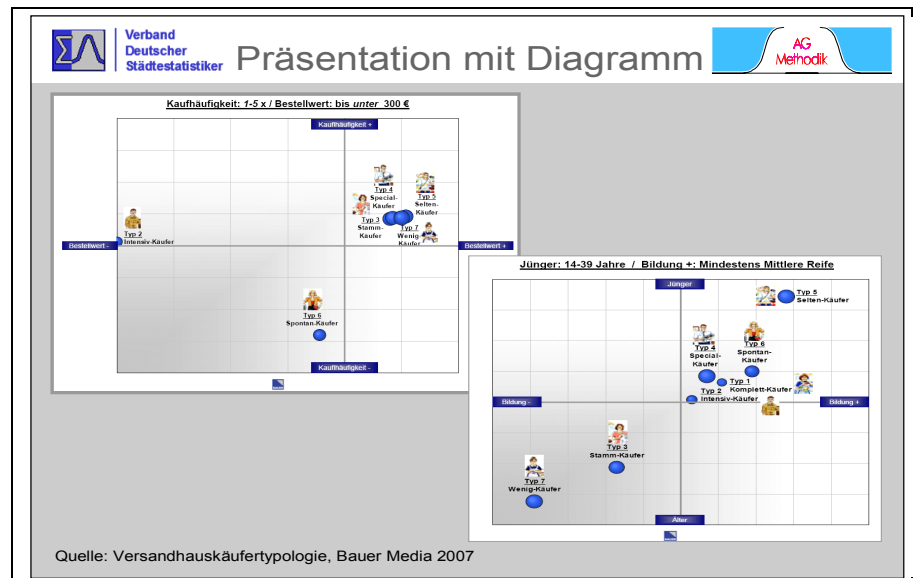
Die Standardabweichung kann dabei zusatzlich berucktigt werden, um die Streuung eines Merkmals innerhalb des Clusters zu berucktigen.

### Diagramme

Auch Diagramme, insbesondere Streudiagramme, eignen sich zur Darstellung von Unterschieden zwischen Gruppen, Typen bzw. Clustern. Auch wenn in einem (zweidimensionalen) Streudiagramm (s. Abbildung 7) nur zwei Merkmale verglichen werden konnen, so zeigen sich die Falle der Cluster aber sehr anschaulich als Punktwolken.

Die Reduzierung der Komplexitat ermoglicht bessere Verstandlichkeit und Sachverhalte konnen einfacher vermittelt werden.

Abbildung 7: Zweidimensionale Abbildung ausgewahlter Merkmale im Streudiagramm



### Zusammenfassung

Die unter Punkt 2 dargestellten Prasentationsformen von Ergebnissen sind fur die offentlichkeit in Politik, Verwaltung und bei Burgern deutlich besser geeignet, als die unter Punkt 1 vorgestellten Tabellen und bersichten. In der Kombination mit kurzen erklarenden Texten lassen sich die Ergebnisse von Clusteranalysen damit auch fur statistisch weniger Interessierte so aufbereiten, dass ihr hoher Informationsgehalt anschaulich vermittelt werden kann.

Autorin:  
Antje Seidel-Schulze  
Deutsches Institut fur Urbanistik – Difu  
Arbeitsbereich Wirtschaft und Finanzen  
Strae des 17. Juni 112  
10623 Berlin  
Tel.: 030 – 3900 1198  
E-mail: Seidel-Schulze@difu.de