

Clusteranalyse – Softwareeinsatz anhand eines konkreten Beispiels mit SPSS

(Thomas Nirschl, Stadt Nürnberg)

Gliederung:

1. Vorüberlegungen und vorbereitende Schritte
2. Clusteranalyse mit SPSS

Beispiel: *Kleinräumige Sozialraumanalyse der Stadt Nürnberg*

3. Ergebnis und Bewertung



Vorüberlegungen und vorbereitende Schritte

Hintergrund:

- pragmatisch: letzte Sozialraumanalyse aus dem Jahr 2004
- Zunahme der sozialräumlichen Differenzierungsprozesse \Rightarrow Nachfrage nach „verdichteten“ Informationen ist groß
- Ergebnisse einer neuen Sozialraumanalyse (inkl. Gebietstypisierungen) sollen in das stadtinterne Monitoring einfließen
- Kommunalwahl 2008 \Rightarrow Analyse des Wahlverhaltens in den sozialen Stimmbezirkstypen (*Vorwahl-, Nacht- und Ergebnisheft*)
- Ausgangspunkt für weitere Analysen \Rightarrow z.B. Wanderungen aus bzw. nach Sozialräumen

Vorüberlegungen und vorbereitende Schritte

Anforderungen:

Aufgabe:

Kompromiss zwischen möglichst vielfältigen bzw. umfassenden Daten einerseits und maximaler Kleinräumigkeit andererseits muss gefunden werden

Begriffsdefinition „Sozialraum“:

„vom Menschen strukturierter und geprägter, physisch klar abgrenzbarer Raum“


Ansatz:

explorative quantitative Raumanalyse \Rightarrow Existenz eines vorher festgelegten Untersuchungsraums


Vorüberlegungen und vorbereitende Schritte

Ziel:

Analyse der Verteilung der Merkmalsausprägungen
ausgewählter Daten



Strukturierung und Klassifizierung städtischer Teilgebiete



Bereitstellung einer umfassenden, kleinräumigen und
möglichst realitätsgetreuen Informationsbasis über die
Lebensräume, sozialen Milieus, Wohnformen und
ökonomischen Belastungen in den verschiedenen
Stadtvierteln

Voruberlegungen und vorbereitende Schritte

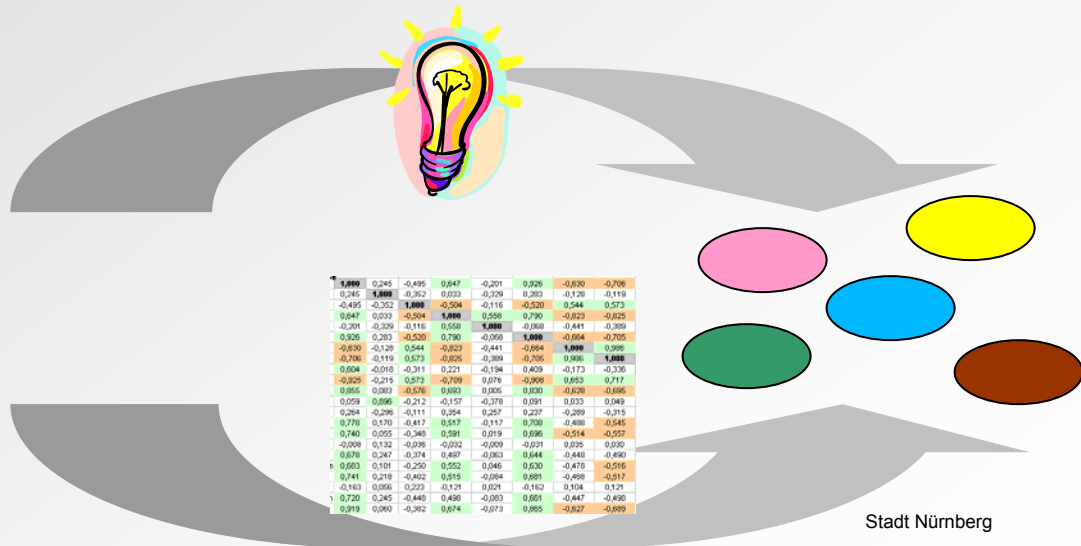
Methodischer Ansatz:

Hierarchisch-agglomerative Clusteranalyse

Idee:

Gruppenbildung anhand ausgewahlter Variablen

⇒ logisch-konzeptionelle Voruberlegungen, Korrelationsanalysen



Vorüberlegungen und vorbereitende Schritte

Methodischer Ansatz:

Wichtig:

Die EDV nimmt dem Anwender (noch) nicht die ganze Arbeit ab !



Einige zentrale Fragen müssen vorab geklärt werden: z.B.:
„Was will ich wie und mit welchen Daten analysieren?“



Vorüberlegungen und vorbereitende Schritte

Software:



SPSS 15 (seit 2008: v.16):

- eigentliche Clusteranalyse
- Projektdatei
- Syntax inkl. Kommentierung



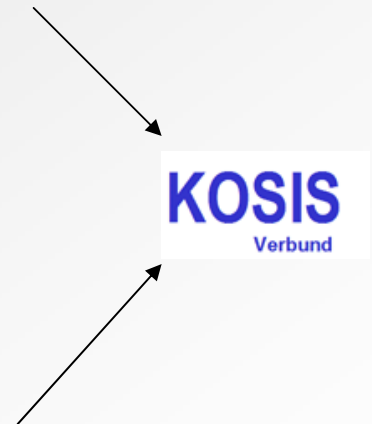
Excel 2000:

- Darstellung der Ergebnisse
- Interpretationshilfe(n)




ArcView 3.3:

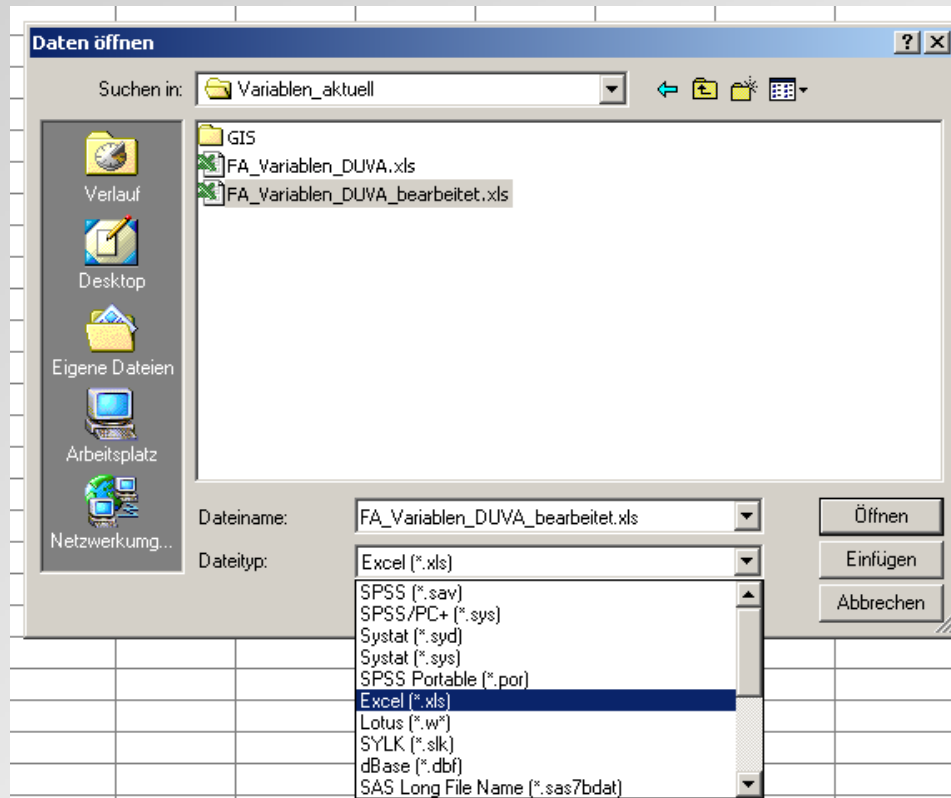
- kartographische Darstellung der Ergebnisse
- räumliche Überprüfung (Expertenwissen)



Clusteranalyse mit SPSS

1.  Erzeugen der Grunddaten (.xls, .csv, .dbf ...)

2. 



Einlesen der Grunddaten je
nach Format uber

- „Datei offnen“ (.xls, .dbf
o.a.),
- „Textdaten lesen“ (.csv,
.txt o.a.)



Clusteranalyse mit SPSS

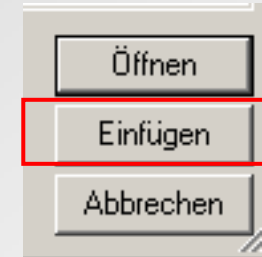
3.

```
Sozialraumanalyse2007_FaktorenundClusteranalyse.sps - SPSS Syntax-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Ausführen Fenster Hilfe

'1:\Projekte\ZST_Sozialtypisierung\Projekt_Sozialraumanalyse 2007\Analyse\Variablen_aktuell\FA_Vari-
'ablen_DUVA_bearbeitet.xls'
/SHEET=name 'AlleVariablen_SPSS'
/CELLRANGE=full
/READNAMES=on
/ASSUMEDSTRWIDTH=32767.
DATASET NAME DatenSet3 WINDOW=FRONT.

**A.2: Umwandlung des numerischen Distrikts in eine Stringvariable**.
string dis(a3).
if distrikt<100 dis= concat("0",ltrim(trim(string(distrikt,3)))).
if distrikt>= 100 dis= ltrim(trim(string(distrikt,3))).
EXECUTE.

**A.3: Zur Berechnung von Summen und Anteilen Umkodieren der Missing Values in "0"-Werte**.
RECODE
dtschomigr dtschmmigr ausl wdb3insg wd3b5insg wd5b10insg wd10b20insg
wdm20insg agu3insg ag3b6insg ag6b10insg ag10b15insg ag15b18insg ag18b25insg
ag25b30insg ag30b45insg ag45b60insg ag60b65insg ag65b75insg ag75b80insg
ag80b85insg agm85insg wdb3migr wd3b5migr wd5b10migr wd10b20migr wdm20migr
agu3migr ag3b6migr ag6b10migr ag10b15migr ag15b18migr ag18b25migr
ag25b30migr ag30b45migr ag45b60migr ag60b65migr ag65b75migr ag75b80migr
ag80b85migr agm85migr wdb3dtsch wd3b5dtsch wd5b10dtsch wd10b20dtsch
wdm20dtsch agu3dtsch ag3b6dtsch ag6b10dtsch ag10b15dtsch ag15b18dtsch
ag18b25dtsch ag25b30dtsch ag30b45dtsch ag45b60dtsch ag60b65dtsch
ag65b75dtsch ag75b80dtsch ag80b85dtsch agm85dtsch hh1insg hh2insg hh3insg
hh4insg hh5uminsg hh1dtsch hh2dtsch hh3dtsch hh4dtsch hh5umdtsch hh1ausl
hh2ausl hh3ausl hh4ausl hh5umausl hht_einpers hht_ekkkwp hht_ekkmewp
hht_emekkw p hht_ememewp hht_pnelkkwp hht_pnelkkmewp hht_pnelmekwp
hht_pnelmekmewp hht_bmekkw p hht_bmekmewp hht_sonst saldo1000dtsch
saldo1000ausl saldo1000insg umzug1000dtsch umzug1000ausl umzug1000insg alg1
alose alosedtsch aloseausl aloseu25 aloselz bedgem bedgem 1k bedgem 2umk
persinbedgem ehb ehbalose ehballerz (MISSING=0) .
EXECUTE.
```



Hinweis:

Kommandos über den Befehl „*Einfügen*“ zunächst in ein sog. **Syntaxfenster** kopieren und als **.sps-Datei** abspeichern. Dies hat den Vorteil, dass die vollzogenen Schritte dokumentiert und v.a. auch kommentiert werden.

Zusätzlich kann dadurch – bei gewissen Vorkenntnissen – auch in die Befehlsstruktur von SPSS eingegriffen werden.



Clusteranalyse mit SPSS

4.

Name	Typ	Spalte	
1	distrikt	Numerisch	11
2	dtschomigr	Numerisch	11
3	dtschmigr	Numerisch	11
4	ausl	Numerisch	11
5	wdb3insg	Numerisch	11
6	wd3b5insg	Numerisch	11
7	wd5b10insg	Numerisch	11
8	wd10b20insg	Numerisch	11
9	wdm20insg	Numerisch	11
10	agu3insg	Numerisch	11
11	ag3b6insg	Numerisch	11

Nach den Berechnungen und dem Labeln von Variablen ist ein Blick auf die Korrelationen zwischen den Variablen oft hilfreich.

Mit „Analysieren-Korrelation-Distanzen...“ gelangt man zu folgendem Dialogfenster, ber welches die entsprechenden Einstellungen zur Ausgabe einer Korrelationsmatrix getatigt werden knnen.

5.

Mae: Unahnlichkeiten ahnlichkeiten

Mae... Pearson-Korrelation

Ma: Intervall: Pearson-Korrelation Binar: Russel und Rao

Standardisieren: Keine Nach Variablen Nach Fallen

Mae transformieren: Absolutwerte Vorzeichen andern Auf Bereich 0-1 skalieren

Clusteranalyse mit SPSS

6.

Ausgabe3 [Dokument3] - SPSS Viewer

Datei Bearbeiten Ansicht Daten Transformieren Einfügen Format Analysieren Grafiken Extras Fenster Hilfe

Ausgabe
Log
Ähnlichkeiten
Titel
Anmerkungen
Arbeitsdatei
Verarbeitete Fälle
Näherungsmatrix

	Migranten	Einwohner mit Hauptwohnsitz insgesamt	Erwerbsfähige 15 bis unter 65 Jahre	Kinder- und Jugendliche unter 15 Jahre	Senioren über 65 Jahre	Wohndauer unter 5 Jahre insgesamt
Migranten	1,000	,827	,850	,843	,597	,908
Einwohner mit Hauptwohnsitz insgesamt	,827	1,000	,993	,950	,889	,931
Erwerbsfähige 15 bis unter 65 Jahre	,850	,993	1,000	,940	,846	,945
Kinder- und Jugendliche unter 15 Jahre	,843	,950	,940	1,000	,781	,903
Senioren über 65 Jahre	,597	,889	,846	,781	1,000	,705
Wohndauer unter 5 Jahre insgesamt	,908	,931	,945	,903	,705	1,000
Wohndauer bis 5 Jahre Deutsch	,702	,924	,923	,849	,768	,926
Wohndauer bis 5 Jahre Migranten	,981	,798	,825	,821	,535	,924
Haushalte insgesamt	,833	,982	,981	,901	,855	,947
Haushalte Deutsch	,717	,970	,956	,866	,913	,878
Haushalte Ausländer	,902	,686	,731	,684	,378	,847
Haushalte mit Kind(ern)	,826	,968	,957	,992	,820	,897
Haushalte Alleinerziehend	,890	,912	,917	,922	,719	,920
Bedarfsgemeinschaft mit Kind(ern)	,947	,744	,766	,801	,493	,846
Anteil Migranten	,559	,148	,187	,225	-,068	,331
Jugendquotient	,133	,042	,024	,253	-,085	,070
Seniorenquotient	-,165	,075	,013	-,042	,410	-,105
Anteil Wohndauer bis 5						

In einem weiteren Fenster – dem sog. **Output** – erscheint das Ergebnis.

Hinweis:

Das Arbeiten im Output ist nicht immer komfortabel. Eine Ausgabe z.B. nach Excel ist zu empfehlen. Dort können dann ohne größeren Aufwand Formatierungen o.ä. erfolgen.



Clusteranalyse mit SPSS

7.

*Sozialraumanalyse2007_Faktoren.sav [DatenSet1] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

15 : distrikt 32

	distrikt	dtschomigr	dtschmmigr	ausl	wdb3insg	wd3b5insg
1	10	152	16	43	104	29
2	11	400	43	135	222	91
3	12	342	54	167	280	83
4	13	91	9	43	64	18
5	14	225	24	40	126	28
6	15	50	6	25	33	12
7	16	228	24	121	199	42
8	17	232	42	86	158	47
9	18	420	63	338	424	99
10	19	725	101	272	520	138
11	20	291	44	92	159	52
12	21	520	56	129	239	101
13	30	10	0	12	14	3
14	31	481	156	278	375	99
15	32	7	0	8	2	0
16	40	1419	465	1217	1360	409
17	41	1054	276	1053	999	342
18	42	1250	295	1206	1045	372
19	50	822	170	347	483	159
20	51	1709	187	335	752	261
21	52	608	133	305	362	133
22	53	232	20	41	71	29
23	54	322	19	37	96	29

Distrikte mit weniger als 50 Einwohnern werden von der Clusteranalyse ausgeschlossen. Zudem werden vier weitere Distrikte als extreme Ausreisser identifiziert und ebenfalls bei der Analyse nicht berucksichtigt.

Es gehen somit 270 Distrikte in die Clusteranalyse ein.



Clusteranalyse mit SPSS

Variablenübersicht:

Soziodemographische Merkmale	Sozioökonomische Merkmale	Physiognomische Merkmale
<ul style="list-style-type: none"> - Migrantenanteil - Anteil Haushalte deutsch an allen HH - Anteil Haushalte ausländisch an allen HH - Anteil Kinder von Migranten - Mobilität insg. je 1.000 - Anteil Senioren ohne Senioren in Heimen - Jugendquote - Anteil HH mit Kind(ern) an allen HH - Anteil 1Personenhaushalte an allen HH - Anteil 5u.m.-Personenhaushalte an allen HH 	<ul style="list-style-type: none"> - Arbeitslosenanteil - Anteil Arbeitsloser Ausländer - Anteil Bedarfsgemeinschaften an allen HH - Anteil Bed.gem. mit Kind(ern) an allen HH m.K. - Anteil SGBII-Empfänger an den Erwerbsfähigen - Personen in Bedarfsgemeinschaften an allen Einwohnern 	<ul style="list-style-type: none"> - Anteil Ein- und Zweifamilienhäuser an allen Wohngebäuden - Bebauungsdichte



Clusteranalyse mit SPSS

8.

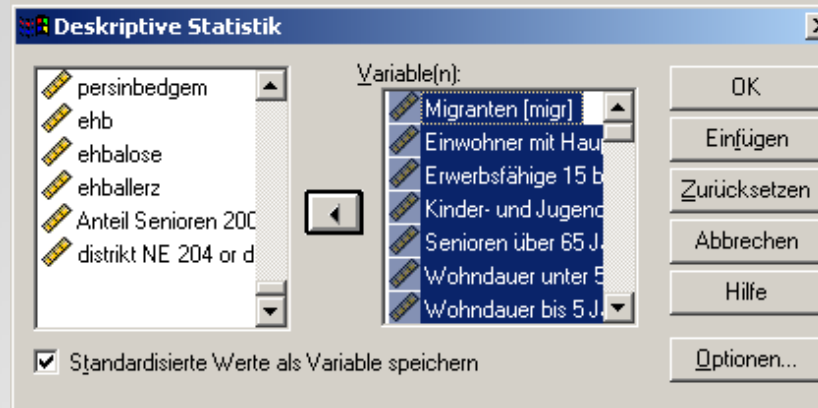
distrikt	dtschomigr
1	10
2	11
3	12
4	13
5	14
6	15
7	16
8	17
9	18
10	19

z-Transformation aller Variablen:

Basieren Variablen auf unterschiedlichen Maeinheiten kann es bei der Clusteranalyse zu einer Vergroerung der Distanzen zwischen den Merkmalsauspragungen kommen.

Losung: Standardisierung der Variablen

9.



Von jedem Wert wird der MW der betreffenden Variablen abgezogen und die Differenz anschließend durch σ dividiert.

Die standardisierten Variablen erhalten somit einen Mittelwert von „0“ und eine Standardabweichung von „1“*.

* Werte mit einem pos. Vorzeichen sind als – im Vergleich zu allen anderen Werten – uberdurchschnittlich ausgepragt zu interpretieren, wahrend Werte mit neg. Vorzeichen als unterdurchschnittlich ausgepragt zu verstehen sind.

Clusteranalyse mit SPSS

10.

Microsoft Excel - FA_Analyse_SPSS_KorrelationsmatrizenZWert00Ausreisser.xls

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster Acrobat

	A	B	C	D	E	F	G	H	I	J	K	L
		Z-Wert: Anteil Migranten	Z-Wert: Jugendquotient	Z-Wert: Seniorenquotient	Z-Wert: Anteil Wohndauer bis 5 Jahre	Z-Wert: Anteil Wohndauer bis 5 Jahre Deutsch	Z-Wert: Anteil Wohndauer bis 5 Jahre Migranten	Z-Wert: Anteil Wohndauer mehr als 5 Jahre insgesamt	Z-Wert: Anteil Wohndauer mehr als 20 Jahre Deutsch	Z-Wert: Anteil Wohndauer mehr als 20 Jahre Migranten	Z-Wert: Anteil Haushalte Deutsch an allen HH	Z-Wert: Anteil Haushalte an allen HH
1												
2	Z-Wert: Anteil Migranten	1,000	-0,495	0,647	-0,201	0,926	-0,630	-0,706	0,604	-0,925	0	
3	Z-Wert: Jugendquotient	0,245	1,000	-0,352	0,033	-0,329	0,283	-0,128	-0,119	-0,018	-0,215	0
4	Z-Wert: Seniorenquotient	-0,495	-0,352	1,000	-0,504	-0,116	-0,520	0,544	0,573	-0,311	0,573	-0
5	Z-Wert: Anteil Wohndauer bis 5 Jahre	0,647	0,033	-0,504	1,000	0,558	0,790	-0,823	-0,825	0,221	-0,709	0
6	Z-Wert: Anteil Wohndauer bis 5 Jahre Deutsch	-0,201	-0,329	-0,116	0,558	1,000	-0,068	-0,441	-0,389	-0,194	0,076	0
7	Z-Wert: Anteil Wohndauer bis 5 Jahre Migranten	0,926	0,283	-0,520	0,790	-0,068	1,000	-0,664	-0,705	0,409	-0,908	0
8	Z-Wert: Anteil Wohndauer mehr als 5 Jahre insgesamt	-0,630	-0,128	0,544	-0,823	-0,441	-0,664	1,000	0,986	-0,173	0,653	-0
9	Z-Wert: Anteil Wohndauer mehr als 20 Jahre Deutsch	-0,706	-0,119	0,573	-0,825	-0,389	-0,705	0,986	1,000	-0,336	0,717	-0
10	Z-Wert: Anteil Wohndauer mehr als 20 Jahre Migranten	0,604	-0,018	-0,311	0,221	-0,194	0,409	-0,173	-0,336	1,000	-0,541	0
11	Z-Wert: Anteil Haushalte Deutsch an allen HH	-0,925	-0,215	0,573	-0,709	0,076	-0,908	0,653	0,717	-0,541	1,000	-0
12	Z-Wert: Anteil Haushalte Auslander an allen HH	0,855	0,083	-0,576	0,693	0,005	0,830	-0,628	-0,695	0,554	-0,965	1
13	Z-Wert: Anteil Haushalte mit Kind(ern) an allen HH	0,059	0,896	-0,212	-0,157	-0,378	0,091	0,033	0,049	-0,103	-0,019	-0
14	Z-Wert: Anteil HH Alleinerziehende an allen HH mit Kind(ern)	0,264	-0,296	-0,111	0,354	0,257	0,237	-0,289	-0,315	0,228	-0,269	0
15	Z-Wert: Arbeitslosenanteil	0,778	0,170	-0,417	0,517	-0,117	0,708	-0,488	-0,545	0,464	-0,777	0
16	Z-Wert: Anteil Arbeitsloser Auslander	0,740	0,055	-0,348	0,591	0,019	0,696	-0,514	-0,557	0,380	-0,736	0
17	Z-Wert: Anteil Arbeitsloser unter 25 Jahre	-0,008	0,132	-0,036	-0,032	-0,009	-0,031	0,035	0,030	0,018	0,033	-0
18	Z-Wert: Anteil Bedarfsgemeinschaften an allen HH	0,678	0,247	-0,374	0,497	-0,063	0,644	-0,448	-0,490	0,362	-0,669	0
19	Z-Wert: Anteil Bedarfsgemeinschaften mit Kind(ern) an allen HH m	0,683	0,101	-0,250	0,552	0,046	0,630	-0,478	-0,516	0,345	-0,622	0
20	Z-Wert: Anteil SGBII-Empfanger an den Erwerbsfahigen	0,741	0,218	-0,402	0,515	-0,084	0,681	-0,468	-0,517	0,412	-0,736	0
21	Z-Wert: Anteil Alleinerziehender SGBII-Empfanger an allen EHB	-0,163	0,056	0,223	-0,121	0,021	-0,162	0,104	0,121	-0,125	0,193	-0
22	Z-Wert: Personen in Bedarfsgemeinschaften an allen Einwohnern	0,720	0,245	-0,448	0,498	-0,083	0,661	-0,447	-0,498	0,414	-0,717	0
23	Z-Wert: Anteil Kinder von Migranten	0,919	0,060	-0,382	0,674	-0,073	0,865	-0,627	-0,689	0,526	-0,856	0
24	Z-Wert: Anteil 1HH Deutsch an allen HH	0,065	-0,661	0,035	0,411	0,588	0,059	-0,353	-0,349	0,062	-0,078	0
25	Z-Wert: Anteil SunHH Deutsch an allen HH	-0,412	0,489	0,012	-0,404	-0,168	-0,362	0,315	0,350	-0,286	0,381	-0
26	Z-Wert: Anteil SunHH Auslander an allen HH	0,685	0,397	-0,452	0,398	-0,191	0,620	-0,383	-0,447	0,479	-0,706	0
27	Z-Wert: Anteil 1HH Auslander an allen HH	0,646	-0,129	-0,552	0,628	0,159	0,637	-0,559	-0,610	0,440	-0,803	0
28	Z-Wert: Anteil 1 HH an allen HH	0,294	-0,596	-0,176	0,574	0,547	0,286	-0,501	-0,516	0,215	-0,363	0
29	Z-Wert: Anteil SunHH an allen HH	0,237	0,762	-0,281	0,043	-0,300	0,274	-0,077	-0,080	0,039	-0,246	0
30	Z-Wert: Anteil Ein- und Zweifamilienhuser	-0,660	0,194	0,307	-0,607	-0,204	-0,578	0,616	0,658	-0,399	0,646	-0
31	Z-Wert: Anteil Neubauten	-0,127	0,275	-0,188	0,094	0,179	-0,020	-0,249	-0,189	-0,265	0,125	-0
32	Z-Wert: Anteil Altbauwohnungen an allen Wohnungen	0,336	0,255	-0,324	0,208	-0,025	0,269	-0,229	-0,277	0,342	-0,413	0
33	Z-Wert: Anteil an Wohnflache ((Wfl.+Nfl.))	-0,224	0,070	0,342	-0,423	-0,263	-0,314	0,328	0,329	-0,087	0,321	-0
34	Z-Wert: Bebauungsdichte ((Wfl.+Nfl.)/Gebfl.)	0,426	-0,335	-0,315	0,503	0,286	0,393	-0,466	-0,490	0,256	-0,446	0
35	Z-Wert: Wohnflache pro Einwohner	-0,225	0,072	0,344	-0,424	-0,264	-0,315	0,327	0,329	-0,092	0,323	-0
36	Z-Wert: Mobilitat Auslander je 1000 2006	-0,142	0,067	-0,183	0,039	0,119	-0,040	0,016	0,043	-0,166	0,005	0
37	Z-Wert: Mobilitat insg. je 1000 2006	0,510	-0,137	-0,477	0,719	0,377	0,586	-0,599	-0,612	0,228	-0,639	0
38	Z-Wert: Wanderungssaldo insg. je 1000 2006	0,162	-0,136	-0,233	0,382	0,291	0,245	-0,193	-0,195	0,062	-0,217	0
39	Z-Wert: Wanderungssaldo Auslander je 1000 2006	-0,058	0,044	-0,051	-0,053	-0,070	-0,012	0,103	0,108	-0,058	-0,002	0
40	Z-Wert: Anteil Senioren 2006	-0,495	-0,352	1,000	-0,504	-0,116	-0,520	0,544	0,573	-0,311	0,573	-0
41	Z-Wert: Anteil Senioren ohne Senioren in Heimen	-0,550	-0,290	0,800	-0,686	-0,327	-0,584	0,753	0,771	-0,297	0,630	-0
42	Dies ist eine ahnlichkeitsmatrix											
43												
44												
45												

Die Korrelationsmatrix der z-transformierten Variablen wird dann aus dem Output nach Excel kopiert. Mittels bedingter Formatierungen konnen dann positive und negative Korrelation sichtbar gemacht werden.



Clusteranalyse mit SPSS

11.

	distrikt	dtschomigr	dtschmmigr	ausl	wdb3insg	wd3b5insg	wd5b10insg	wd10b20insg	wdm20insg	agu3insg	ag3b6insg	ag6b10insg
1	10	152	104	222	158	424	520					
2	11	400										
3	12	342										
4	13	91										
5	14	225										
6	15	50										
7	16	228										
8	17	232										
9	18	420										
10	19	725										

„Analysieren-Klassifizieren-Hierarchische Cluster ...“ führt schließlich zum Verfahren der Clusteranalyse.

In diesem Beispiel wird ein hierarchisch-agglomeratives Verfahren zur Analyse herangezogen.

12.

z-transformierte Variablen

Raumbezug, hier: nach Distrikten

Clusteranalyse mit SPSS

13.

Methode...

Hierarchische Clusteranalyse: Methode

Cluster-Methode: Ward-Methode

Ma

Intervall: Quadrierter Euklidischer Abstand
Exponent: 2 Wurzel: 2

Hufigkeiten: Chi-Quadrat-Ma

Binar: Quadrierte Euklidische Distanz
Vorhanden: 1 Nicht-vorhanden: 0

Werte transformieren

Standardisieren: Keine

Nach Variablen
 Nach Fallen

Mae transformieren

Absolutwerte
 Vorzeichen andern
 Auf Bereich 0-1 skalieren

Die „*Cluster-Methode*“ bezeichnet den Algorithmus, der die Entfernung von Clustern zueinander bestimmt

⇒ Zusammenfassung von Objekten auf Basis der Distanzen

Die Wahl eines geeigneten ahnlichkeits- bzw. Distanzmaes erfolgt im Dialog „*Ma*“ (hier: quadrierter euklidischer Abstand).

14.

Speichern...

Hierarchische Clusteranalyse: Neue Variablen speichern

Cluster-Zugehorigkeit

Keine

Einzelne Losung
Anzahl der Cluster: 6

Bereich von Losungen
Mindestanzahl der Cluster:
Hochstanzahl der Cluster:

Sollten die Variablen noch nicht z-transformiert sein, dann konnte man die Werte hier standardisieren.

Unter „*Neue Variable speichern*“ erfolgen die Einstellungen fur die Zahl der Cluster (hier: 6)

Clusteranalyse mit SPSS

15.

Statistik...

Hierarchische Clusteranalyse: Statistik

Zuordnungsubersicht Weiter

Distanz-Matrix Abbrechen

Cluster-Zugehorigkeit Hilfe

Keine

Einzelne Losung
Anzahl der Cluster:

Bereich von Losungen
Mindestanzahl der Cluster:

Hochstanzahl der Cluster:

16.

Diagramm...

Hierarchische Clusteranalyse: Diagramme

Dendrogramm Weiter

Eiszapfen Abbrechen

Alle Cluster

Angegebener Clusterbereich
Start-Cluster:

Stop-Cluster:

Schritt:

Keine Hilfe

Orientierung

Vertikal

Horizontal

Fall	1:010	2:011	3:012	4:013
1:010	,000	26,733	23,759	49,139
2:011	26,733	,000	20,270	31,659
3:012	23,759	20,270	,000	22,738
4:013	49,139	31,659	22,738	,000
5:014	39,886	21,710	46,337	48,841
6:015	46,888	25,191	41,131	42,723
7:016	79,183	55,716	45,512	78,260
8:017	49,085	19,771	36,940	34,369
9:018	73,259	39,443	36,330	36,540
10:019	40,635	16,494	34,461	33,838
11:020	35,034	19,892	36,187	44,739
12:021	51,914	22,280	55,496	48,440
14:031	40,014	22,285	28,187	24,165

Die Distanzmatrix zeigt die Distanzen vor dem Zusammenfuhren von Fallen zu Clustern („Ausgangsdistanzmatrix“)



Clusteranalyse mit SPSS

17.

1 : distrikt	Zantsen	CLU6_1	
1	-.56781	1	
2	-.92204	1	
3	-1,46815	1	
4	-1,46815	1	
5	-.11026	1	
6	-1,86666	1	
7	,55392	1	
8	-.75968	1	
9	-.99584	2	
10	-1,39435	1	

Nachdem die Cluster-Zugehorigkeit in eine Variable geschrieben wurde, kann nun damit begonnen werden, die Cluster zu beschreiben.

18.

„Analysieren-Mittelwertevergleichen-Mittelwerte...“ liefert neben MW auch Min, Max, σ etc. fur die ermittelten Cluster und dient somit der Interpretation der Ergebnisse.

Alternativ kann auch eine separate und nach Clustern aggregierte Datei ausgegeben werden.

Ergebnis und Bewertung

Clusteranalyse mit SPSS

Vorteile	Nachteile
<ul style="list-style-type: none">- umfassendes (Standard-) Software-Paket mit vielen Funktionen- Importfunktion von Daten; Export der Ergebnisse (graphisch)- Dokumentation, Kommentierung und anderungen sind ber die Syntax sehr gut mglich- umfangreiche Sekundarliteratur- fr KOSIS-Mitglieder vergnstigt	<ul style="list-style-type: none">- Anschaffungspreis (z.B. 5 Netzlizenzen ca. € 3.000)- z.T. mit Funktionen und Optionen „berladen“- Bedienung i.d.R. nicht intuitiv (Einarbeitung bzw. Schulung ntig)- mit neuen Versionen verschwinden alte gewohnte Funktionen oder werden umbenannt

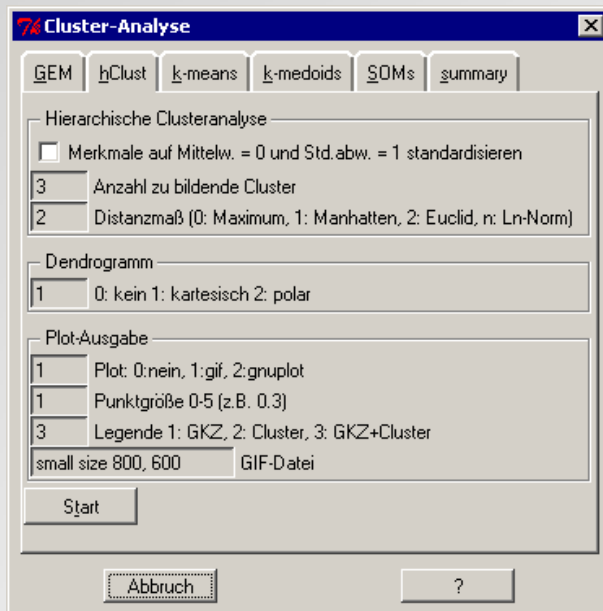


„Cluster“-Modul in neuer SIKURS-Version

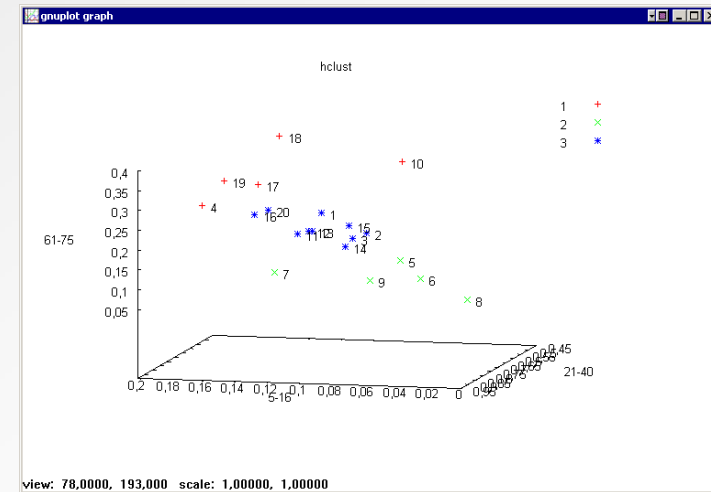
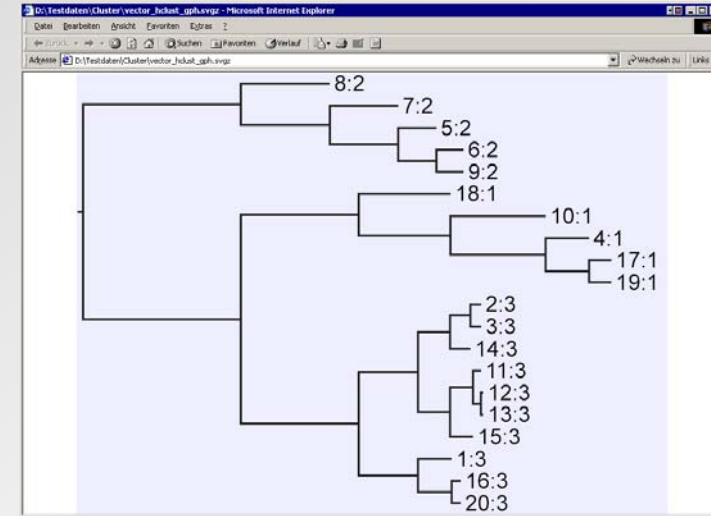
Ziel: demographisches Verhalten bei Prognosen differenzierter abbilden

Relevante Raten und Quoten werden nicht mehr auf die Gesamtstadt bezogen

⇒ Typenbildung (Alter, Geschlecht, Bevölkerungsgruppen) ⇒ anschließende Clusteranalyse in SIKURS berücksichtigt die strukturellen Unterschiede in einer Stadt stärker.



Mehr Informationen unter <http://www.sikurs.de/>



Stadt Nürnberg

Amt für Stadtforschung und Statistik

Nürnberg