Technology Arts Sciences Cologne
Faculty of Economics, Business and Law
Prof. Dr. Arrenberg
Room 221, Tel. 39 14
jutta.arrenberg@th-koeln.de

Exercises Quantitative Methods
Worksheet: Cluster Analysis

Exercise 12.1 (Martens, Seite 229, Statistische Datenanalyse mit SPSS für Windows)

We want to cluster 40 vehicles due to the values of the variables speed (Geschwindigkeit km/h), cubic capacity (Hubraum in ccm), luggage compartment (Kofferraum in Liter), output (Leistung in kW) and payload (Zuladung in kg).

The greatest jump in the dendrogram provides five cluster. The cluster centers are:

Final Cluster Center

	Cluster				
	1	2	3	4	5
Hubraum	2 723	3 350	1 557	2 385	4 803
Geschwindigkeit	189	246	177	175	204
Kofferraum	3 620	368	638	1 850	1 753
Leistung	113	201	69	94	196
Zuladung	708	425	432	633	579

The memberships are:

Cluster 1 Renault Espace V6; Chrysler Voyager

Cluster 2 Alfa 156 2,5 V6; Alfa Spider 3,0 V6; Audi S8 Allrad; Bentley Arnage; BMW Z3 Coupé; Chrysler 300 M; Ferrari F355; Jaguar XJ 8 3,2; Maserati Quattroporte; Porsche 911 Carrera; Mercedes S 300

Cluster 3 Audi A4 1,8; Citrœn Saxo 1,1i X; Fiat Punto 60 SX; Fiat Coupé 1,8 16V; Ford Ka 1,3; Honda Civic 1,4i; Mazda 323 F 1,5; Mazda MX5; Smart; Opel Corsa 1.0; Opel Vectra CDX; Peugeot 206; Peugeot 406 SV 2,0; Renault Twingo; Volvo C 70 T5; VW Polo SDL; VW Beetle

Cluster 4 Citrœn Xantia 1,9 T; Ford Galaxy 2,3 GLX; Honda Shuttle 2,3i; Mercedes C 220 T; Mitsubishi Pajero; Nissan Patrol 2,8

Cluster 5 BMW 750i; Jeep Grand Cherokee; Puch G 500; Toyota Landcruiser

Exercise 12.2
Please open the file Pisa_00_03_06_09_12.sav

We want to cluster the participating countries of the PISA survey 2012 due to a cluster analysis:

1. Hierarchical cluster analysis (no fixed number of clusters)

2. Hierarchical cluster analysis (no fixed number of clusters) without Shanghai

3. Hierarchical cluster analysis with fixed number of clusters (without Shanghai) The greatest jump in the dendrogram provides the number of clusters.

4. K-means Cluster analysis (without Shanghai)

Exercise 12.3

a) What are the levels of a variable in a hierarchical cluster analysis?

b) What are the levels of a variable in a k-means cluster analysis?

c) A telephone provider wants to cluster the clients into groups due to the values of the following variables of the last month:

- Calls in the provider network
- National calls not in the provider network
- International calls
- Short messages services in the provider network
- National short messages services not in the provider network
- International short messages services

The values of the variables are listed in the file *Telefon.sav*.

1. Please run a hierarchical cluster analysis for all six variables. How do you get the number of clusters?

2. Please run a k-means cluster analysis for all six variables. The number of clusters is the finding of c.1). What are the final cluster centers? Please comment the clusters.

Exercise 12.4
The values of the variables tax per year, mileage in liter pro 100 km, cubic capacity in ccm, horsepower, exhaust fumes in g/km and price in Euro are listed in the file *auto.sav*. The observed vehicles are:

Vehicle	Cluster membership
Audi S4 Avant A8	
Opel Astra 1.5 Turbo	
VW Golf GT 1.4 TSI	
Ford Mondeo 2.0 TDCi	
Mazda 6 2.0 CD	
Opel Vectra 1.9 CDTI	
Peugeot 407 HDI	
Toyota Avensis 2.0 D-4D	
VW Passat 1.9 TDI	
BMW 530i	
Mercedes E350	
BMW X5 4.4i	
Porsche Cayenne S	
Audi Q7 4.2 FSI	
Daewoo Kalos 1.4 SE	
Hyundai Getz 1.3 GLS	
Ford Fiesta 1.4 i	
Daihatsu cuore	
Fiat Panda	
Daihatsu Sirion	
Nissan Micra	
Suzuki Swift	
Ford Mondeo ST220	
Opel vectra GTS	

We want to cluster the vehicles:

a) Please delete the vehicle Porsche 911 Carrera 4S and run hierarchical cluster analysis.

    1. How to determine the number of clusters?

    2. How many clusters?

b) Please delete the vehicle Porsche 911 Carrera 4S and run a k-means cluster analysis. The number of clusters is the result of a).

    1. Please plot the cluster membership of every vehicle in the table above.

    2. Please comment the p-values in the ANOVA-table.

Exercise 12.5

Please open the file *Kriminalität.sav*. Delete the case number 9 "'District of Columbia"'.

a) Run a hierarchical cluster analysis with all variables. What is the number of clusters?

b) Run a k-means cluster analysis with the number of clusters under a). Commet the clusters.

c) Plot a scatterplot of all cases with the first principal component as the x-axis and the second principal component as the y-axis. Mark the clusters.

Exercise 12.1

<div align="center">

Final Cluster Center

	Cluster				
	1	**2**	**3**	**4**	**5**
Cubic Capacity	**2 723**	**3 350**	**1 557**	**2 385**	**4 803**
Top Speed	189	**246**	177	**175**	204
Luggage Compartment	**3 620**	**368**	638	1 850	1 753
Horsepower	113	**201**	**69**	94	196
Payload	**708**	**425**	432	633	579

</div>

Cluster 1: Giants with four wheels

Cluster 2: Runabouts

Cluster 3: Inefficient vehicles

Cluster 4: Worm gear

Cluster 5: Gasoline guzzler

Exercise 12.4   auto.sav

X1=tax, X2=mileage, X3=cubic capacity, X4=horsepower, X5=exhaust fumes, X6=price

Hierarchical cluster analysis

Zuordnungsübersicht

| Schritt | Zusammengeführte Cluster | | Koeffizienten | Erstes Vorkommen des Clusters | | Nächster Schritt |
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	14	24	47,950	0	0	3
2	28	31	51,323	0	0	9
3	14	26	301,334	1	0	12
4	29	32	325,733	0	0	7
5	9	30	419,097	0	0	6
6	9	27	480,528	5	0	9
7	10	29	567,323	0	4	11
8	22	23	587,344	0	0	10
9	9	28	668,006	6	2	11
10	13	22	701,217	0	8	15
11	9	10	726,703	9	7	20
12	14	15	851,306	3	0	13
13	14	25	1020,301	12	0	15
14	18	19	1383,258	0	0	16
15	13	14	1502,127	10	13	21
16	17	18	1944,551	0	14	22
17	11	12	2050,764	0	0	19
18	20	21	3349,607	0	0	20
19	11	16	6388,774	17	0	22
20	9	20	7635,075	11	18	21
21	9	13	8763,850	20	15	23
22	11	17	9982,809	19	16	23
23	9	11	11885,337	21	22	0

Greatest jump of the coefficients in step 18 with coefficient 3349.607 to step 19 with coefficient 6388.774. Recommend number of clusters = n − 18 = 24 − 18 =6.

k-Means cluster analysis

Clusterzentren der endgültigen Lösung

	Cluster					
	1	2	3	4	5	6
Steuern	284,00	247,25	219,50	297,33	80,38	196,00
Verbrauch	9,40	8,15	12,60	19,53	7,96	15,00
Hubraum	4163,00	1793,13	3247,00	4357,33	1279,63	2879,50
PS	344,00	127,00	272,00	336,67	78,00	228,00
Abgas	322,00	162,13	211,00	334,67	143,50	249,50
Grundpreis	55500,00	23339,50	48153,50	67212,67	11204,38	33602,50

Cluster 1 (PS-stark): Audi S4 Avant A8

Cluster 2 (moderater Preis, moderater Verbrauch): Opel Astra 1.5 Turbo, VW Golf GT 1.4 TSI, Ford Mondeo 2.0 TDCi, Mazda 6 2.0 CD, Opel Vectra 1.9 CDTI, Peugeot 407 HDI, Toyota Avensis 2.0 D-4D, VW Passat 1.9 TDI

Cluster 3 (mittlerer Verbrauch, eher großer Hubraum) : BMW 530i, Mercedes E350

Cluster 4 (teuer, hohe Steuern, hoher Verbrauch, großer Hubraum, hoher Abgaswert): BMW X5 4.4i, Porsche Cayenne S, Audi Q7 4.2 FSI

Cluster 5 (Kleinwagen, günstig im Verbrauch und Steuern, niedrige PS, niedriger Abgaswert, günstig in Anschaffung): Daewoo Kalos 1.4 SE, Hyundai Getz 1.3 GLS, Ford Fiesta 1.4 i, Daihatsu cuore, Fiat Panda, Daihatsu Sirion, Nissan Micra, Suzuki Swift

Cluster 6 (starker Verbrauch): Ford Mondeo ST220, Opel vectra GTS