University of Applied Sciences Cologne
Faculty of Economics, Business Administration and Law
Prof. Dr. Arrenberg
Room 221, Tel. 39 14
jutta.arrenberg@fh-koeln.de

Exercises Quantitative Methods

Worksheet: Principal Components Analysis

Exercise 11.1

PISA 2012 is the programme's 5th survey. It assessed the competencies of 15-year-olds in reading, mathematics and science (with a focus on mathematics) in 65 countries and economies.

Around 510 000 students between the ages of 15 years 3 months and 16 years 2 months participated in the assessment, representing about 28 million 15-year-olds globally.

The students took a paper-based test that lasted 2 hours. The tests were a mixture of open-ended and multiple-choice questions that were organised in groups based on a passage setting out a real-life situation. A total of about 390 minutes of test items were covered. Students took different combinations of different tests. They and their school principals also answered questionnaires to provide information about the students' backgrounds, schools and learning experiences and about the broader school system and learning environment.

The performance in mathematics, reading and science is listed in the file:

Pisa_00_03_06 _09_12.sav

Which country performed the best in 2012?

a) Please give an answer to this question due to a principal components analysis. Extract only one principal component.

b) Please give an answer to this question due to the average value of all three scores.

c) Please compare the two rankings under a) and b).

Exercise 11.2

Please open the file *Auto.sav* and run a principal components analysis with the variables Steuern (tax), Verbrauch (fuel consumption), Grundpreis (basic price), Hubraum (cylinder capacity), PS (horsepower), Abgas (exhaust fumes.)

a) Extract two principal components and plot a scatterplot of all cars.

b) Extract one principal component and give an order of the cars.

Solution 11.1

a) There is only one eigenvalue greater than 1

Component	Eigen value	Variance %
1	2,940	98,007
2	0,041	1,364
3	0,019	0,629

With the extraction of only one principal component we can explain 98.007% of the total variance.

The values of the extracted principal component are the following linear combination:

$0{,}336 \cdot$ standardized mathematics literacy
$+0{,}336 \cdot$ standardized reading literacy
$+0{,}338 \cdot$ standardized science literacy

The weights 0.336 and 0.338 differ rarely.

The ranking is:
Shanghai, Singapur, Hongkong, Südkorea, Japan, Taipei, Finnland, Estland, Liechtenstein, Kanada, Macao, Polen, Niederlande, Schweiz, Irland, Vietnam, Deutschland, Australien, Neuseeland, Belgien, Großbrit., Österreich, Frankreich, Tschechien, Slowenien, Dänemark, Norwegen, Lettland, USA, Italien, Luxemburg, Spanien, Portugal, Ungarn, Island, Litauen, Kroatien, Schweden, Russland, Israel, Slowakei, Griechenland, Türkei, Serbien, Zypern, Ver. Arab. E., Rumänien, Bulgarien, Thailand, Chile, Costa Rica, Mexiko, Kasastan, Montenegro, Malaysia, Uruguay, Brasilien, Jordanien, Tunesien, Argentinien, Albanien, Kolumbien, Indonesien, Katar, Peru.

b) The ranking is:
Shanghai, Singapur, Hongkong, Südkorea, Japan, Taipei, Finnland, Estland, Liechtenstein, Macao, Kanada, Polen, Niederlande, Schweiz, Vietnam, Irland, Deutschland, Australien Belgien, Neuseeland, Großbrit., Österreich, Tschechien, Frankreich, Slowenien, Dänemark, Norwegen, Lettland, USA, Italien, Luxemburg, Spanien, Portugal, Ungarn, Island, Litauen, Kroatien, Schweden, Russland, Israel, Slowakei, Griechenland, Türkei, Serbien, Zypern, V. Arab. E., Rumänien, Bulgarien, Thailand, Chile, Costa Rica, Mexiko, Kasastan, Montenegro, Malaysia, Uruguay, Brasilien, Jordanien, Tunesien, Argentinien, Albanien, Kolumbien, Indonesien, Katar, Peru.

c) There are only a few differences in the two ranking lists due to the almost identical weights 0.336 and 0.338 of the linear combination:

1. difference: permutation of the sequence of Canada and Macao

2. difference: permutation of the sequence of Ireland and Viet Nam

3. difference: permutation of the sequence of New Zealand and Belgium

4. difference: permutation of the sequence of France and Czech Republic

Solution 11.2

a) There is only one eigenvalue greater than 1:

Component	Eigenvalue	Variance %
1	4,859	80,975
2	0,717	11,945
3	0,303	5,046
4	0,070	1,168
5	0,032	0,534
6	0,020	0,331

We have to extrat two principal components, if at least 90% of the total variance should be explained with the principal components.

b) There is only one eigenvalue greater than 1, so we can also extract only one principal component. In the scatterplot of a) we can identify four clusters:

1. Cluster: Porsche Cayenne, Audi Q7, BMW X5, Porsche)11 Carrera, Audi S4

2. Cluster: Mercedes E350, Ford Mondeo ST220, Opel vectra GTS, BMW 530i

3. Cluster: Ford Mondeo 2.0, Mazda 6, Toyota Avensis, Opel Astra 1.5 Turbo, Opel Vectra 1.9 CDTI, VW Passat 1.9 TDI, VW Golf GT, Peigeot 407

4. Cluster: Daewoo Kalos, Nissan Micra, Ford Fiesta, Suzuki Swift, Hyundai Getz, Daihatsu Sirion, Fiat Panda, Daihatsu curore