

Multivariate Statistics: Exercise 8

November 28, 2018

Factor analysis:

On the TUWEL page of our exercises you can find the data set *cardata.csv*, originating from

<http://archive.ics.uci.edu/ml/datasets/Automobile>

Load the data with `read.csv("cardata.csv")` into R. Use only the continuous variables (not the factor variables in columns 1 to 9, 15, 15, 18). Note that the data contain missing values (NA), which can be identified by `is.na()`. The observations containing missings can be excluded with `na.omit()`. Should some variables be transformed first before applying PCA or factor analysis?

1. Compute the principal components and show the first two PCs in a biplot. Try to interpret the first two PCs. How much variance do they explain? Show a scree plot to decide for an appropriate number of PCs.
2. Compute a factor analysis model using an appropriate number k of factors. This can be done by:

```
fa <- factanal(scale(mydata), factors=k, scores="regression")
```

In this function, a maximum-likelihood estimation is carried out to estimate the parameters.

 - (a) What is the maximum value of k you could consider?
 - (b) What are the differences to PCA (loadings, scores)? What are the uniquenesses?
 - (c) Show loadings and scores in a biplot and compare with the PCA biplot.
 - (d) How can you interpret the first two factors?
 - (e) With `print(fa)` you can see the variance proportions of the factors. How are these values computed?
3. In `library(StatDA)` you can find the function `pfa()` for principal factor analysis. What is the difference to the method `factanal()`, and how do the results differ?

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer_8.R* (no word document, no plots). Submit this file to Exercise 8 of our tuwel course (deadline November 27).