Multivariate Statistics: Exercise 7

November 27, 2014

Factor analysis:

1. - On the web 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()`.

    - Compute the principal components and show the first two PCs in a biplot.

    - Compare with a biplot of a factor analysis model with two factors. This can be computed by:
    `fa <- factanal(scale(mydata),factors=2,scores="regression")`
    In this function, a maximum-likelihood estimation is carried out to estimate the parameters. What are the differences to PCA (loadings, scores, uniquenesses)?
    With `print(fa)` you can see the variance proportions of the factors. How are these values computed?

    - 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?

2. Use the data set *flu.csv*, available at the web page of our exercises. The data can be loaded with `read.csv()`. They originate from
http://www.google.org/flutrends/data.txt
and describe the estimated weekly activity of influenza in several countries (how does this work?). Delete the first 160 rows (contain missings) and ignore the values for Sweden (zeros). Log-transform the data to approach better symmetry. Compute factor analysis models with an appropriate number of factors, using the maximum-likelihood methods and principal factor analysis. Compare and discuss the results.

Save your (successful) R code together with short documentations and interpretations of results in a text file, named as *Familyname8.R*. Send the file as an email attachment to *mehmet.mert@tuwien.ac.at*, at latest Tuesday (25.11).