Multivariate Statistics: Exercise 1

October 10, 2018

Introduction:

Load the data `Auto` from the package `ISLR`. This means that you first need to install the package with

```
install.packages("ISLR")
```

and then load the data with

```
data(Auto,package="ISLR")
```

Look at `?Auto` or at `str(Auto)`. The data contain technical measurements on vehicles, and also include additional information like origin.

- Plot the data matrix in a scatterplot matrix using

  ```
  plot(Auto)
  ```

  What do you conclude? Visualize the `origin` information by using different colors and symbols.

- Inspect the numeric (i.e. continuously scaled) variables individually with exploratory data visualization. Install the package `StatDA`, and load it with

  ```
  load(StatDA)
  ```

  Use the function `edaplot()` to visualize the single variables. What do you conclude?

- Estimate the correlation matrix of the numeric variables using the function `cor()`. The default is to use the Pearson correlation, a measure of the linear relationship. Look at `help(cor)` to try out alternative methods. What are the alternatives doing? Would they be more appropriate than the Pearson correlation? Use the `StatDA` function `CorCompare()` to compare two estimation methods.

- Use the Pearson correlation for the sub-datasets split according to the variable "origin". This gives three correlation matrices, one for each origin. Compare them with `CorCompare()`. What do you conclude?

- Perform an eigenvalue decomposition of the covariance matrix of the numeric variables (all observations) using the function `eigen()`.

  Now first scale the data to column means of zero and variances of one using the function `scale()`, and do the eigenvalue decomposition of the covariance matrix again. Finally, do an eigenvalue decomposition of the (Pearson) correlation matrix. What do you conclude from the results?

- Show – based on the eigenvalue decomposition of the covariance matrix – that Equation (1.7) of the course notes is correct.

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer_1.R* (no word document, no plots). Submit this file to Exercise 1 of our tuwel course (deadline October 9).