

### Exercise 12

Consider again the data set introduced during the course session, which contains the head sizes of the first two sons of several families. Perform a principal component analysis (PCA) on the full data set, have a look at the coefficients of the principal components (the ‘loadings’), and visualize the variance explained by the PCA.

### Exercise 13

Obtain principal components based on the following correlation matrix, which was computed from physical attributes of 3000 convicted criminals (MacDonnell, 1902).

$$R = \begin{matrix} & \begin{matrix} \text{Head length} \\ \text{Head breadth} \\ \text{Face breadth} \\ \text{Left finger length} \\ \text{Left forearm length} \\ \text{Left foot length} \\ \text{Height} \end{matrix} \end{matrix} \begin{pmatrix} 1.000 & 0.402 & 0.396 & 0.301 & 0.305 & 0.339 & 0.340 \\ 0.402 & 1.000 & 0.618 & 0.150 & 0.135 & 0.206 & 0.183 \\ 0.396 & 0.618 & 1.000 & 0.321 & 0.289 & 0.363 & 0.345 \\ 0.301 & 0.150 & 0.321 & 1.000 & 0.846 & 0.759 & 0.661 \\ 0.305 & 0.135 & 0.289 & 0.846 & 1.000 & 0.797 & 0.800 \\ 0.339 & 0.206 & 0.363 & 0.759 & 0.797 & 1.000 & 0.736 \\ 0.340 & 0.183 & 0.345 & 0.661 & 0.800 & 0.736 & 1.000 \end{pmatrix}$$

Interpret the resulting principal components, and make a proposal on how many components might be appropriate here.

### Exercise 14

Load the data set `measure.txt` into R. It contains chest, waist and hip girth (in inches) as well as the gender of 20 persons. Perform a principal component analysis with an appropriate R function based on the variables `chest`, `waist` and `hips`. Visualize the data in the space spanned by the principal components, by plotting the scores of each individual person on the first two PCs into a coordinate system (scatter plot of the scores). Indicate female and male persons by using different plotting symbols and/or colors. Give an interpretation of the first two principal components.

### Aufgabe 15

Load the data set `jet.txt` into R. It contains different characteristics of fighter jets, namely:

FFD: first flight date, in months after January 1940;  
SPR: specific power, proportional to power per unit weight;  
RGF: flight range factor;  
PLF: payload as a fraction of gross weight of aircraft;  
SLF: sustained load factor;  
CAR: a binary variable that takes the value 1 if the aircraft can land on a carrier and 0 otherwise.

- Visualize the data by using appropriate scatter plots, and indicate jets which have the ability to land on a carrier by a triangle plotting symbol.
- Perform a principal component analysis based on the variables `SPR`, `RGF`, `PLF` and `SLF`. How many components are appropriate here?
- Now visualize the data by plotting the scores of each individual jet on the first two PCs into a coordinate system (scatter plot of scores). Again, indicate those jets with the ability to land on a carrier by a triangle plotting symbol.