

Homework 3

Due Wednesday, March 6

1. The dataset **SwissBankNotes** contains several variables measured on 200 Swiss banknotes, 100 of which are genuine and 100 counterfeit. The six numerical variables are: length of the banknote, height of the banknote measured on the left, height of the banknote measured on the right, distance of inner frame to the lower border, distance of inner frame to the upper border, and length of the diagonal.
 - (a) Carry out a principal component analysis (standardizing the variables first). What's the cumulative proportion of the total variance explained by the first two components?
 - (b) From the component loadings (the eigenvectors), interpret the first two principal components.
 - (c) Draw a scatterplot of the first two component scores (that is, ξ_{i1} s vs ξ_{i2} s), using different symbols or colors for genuine and counterfeit banknotes (in R you specify plot symbols with the option `pch` in `plot()`.) Are the two groups well separated?
2. The dataset **turtles** contains variables measured on 48 painted turtles, 24 of which are female and 24 male. The three numerical variables are length, width, and height of the carapace.
 - (a) Take logarithms of the variables and standardize them. Then carry out a principal component analysis. Find the cumulative proportions of the total variance explained by the first two components.
 - (b) From the component loadings, interpret the first two principal components.
 - (c) Draw a scatterplot of the first two component scores, using different symbols or colors for males and females. Are the two groups well separated in the plot?