

Multivariate Statistics: Exercise 10

December 12, 2018

Correlation analysis:

Use the data set *movies* from the `library(ggplot2movies)`. See help file for a detailed description.

1. **Multiple correlation analysis:** Compute the multiple correlation between the variable *rating* and the matrix consisting of the variables *year*, *length*, *budget*, *votes*. Delete observations containing missing values.
 - (a) Compute the multiple correlation coefficient. How can you interpret the coefficients of the linear predictor function?
 - (b) Is the multiple correlation coefficient significantly different from zero? Would it be advisable to transform some variables? How about the significance if you omit the variable *votes*?
 - (c) Use the function `CCAgrid` from the R package `ccaPP` – see help. Set the argument `method="pearson"` and compare the results with those from above. Use `method="spearman"` and compare with the previous results. What is the methodological difference?
 - (d) Use the function `permTest` from the `library(ccaPP)`. This function is performing a permutation test for uncorrelatedness, by permuting the observations of the first input. How and why does this work? What is the outcome? Compare with the result in (b).
2. **Canonical correlation analysis:** Compute the canonical correlation between the matrices consisting of the variables *year*, *length*, *budget*, *rating*, *votes* and the variables *Action*, *Animation*, *Comedy*, *Drama*, *Documentary*, *Romance*, *Short*. Delete observations with missings.
 - (a) Use the function `cancor()` – see help. Center and scale the data (why?). How strong is the linear relationship? How can you interpret the linear combinations for the X and Y data?
 - (b) Plot the first two canonical variables φ_1 and η_1 against each other. What can you see in the plot? What seems to be strange here?
 - (c) Apply point 1.(c) here. Plot the canonical variables φ_1 and η_1 .
 - (d) Perform a permutation test, see 1.(d). Make efficient use of the function.

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer_10.R* (no word document, no plots). Submit this file to Exercise 10 of our *tuwel* course (deadline December 11).