Solution Sheet 6

The full R code will be made available in a separate file on the course homepage.

1. **a)** The significant covariates are
- for Nardstri: the intercept (0.036) and pH10 (0.04)
- for Caluvulg: N10 (0.034) and C10 (0.017)
- for Festrubr: P10 (0.026)

b) Observation #72 is the most prominent outlier; moreover, the distribution of the residuals is skewed.

c) After observation #72 has been omitted, the significant covariates are now
- for Nardstri: the intercept (0.02) and pH10 (0.04)
- for Caluvulg: N10 (0.001) and C10 (0.0002)
- for Festrubr: none (P10 has p-value 0.093)

The partial correlation between the square root-transformed counts of `Nardstri` and `Caluvulg` is -0.24, i.e., controlling for the soil chemistry variables, the abundance of the two species is negatively correlated, indicating that the two species compete.

d) The p-values obtained here are somewhat smaller than those obtained by use of the (square root-transformed) counts of `Nardstri`, `Caluvulg` and `Festrubr`. The residual plots do not show any dependence of the residuals on fitted values or on covariates; the residuals appear to be normally distributed; moreover, the partial correlations are smaller than in **c)**, which is expected since the principal components are uncorrelated by construction.

2. **a)** Due to outliers, the variable `to` (toluol) substantially influences the first and the second principal component. Its influence on the other principal components is practically nil.

b) By normalizing, the effects of the outliers of the variable `to` on the first two principal components is attenuated.

c) For p principal components, the convex hull of the points should have at most $p + 1$ vertices. With each additional component, a new vertex should become apparent—which holds true at least in the beginning.