D. Metzler, N. Becker SS 2013

MULTIVAR STATS IN ECOL AND GENETICS — EXERCISES, SHEET 7

Exercise 1 To get a feeling for correspondence analysis (CA) we first look at self-generated data from which we can guess the CA biplot before plotting.

- Define the 3 by 3 matrix

```
Y <- matrix(c(0,0,1,0,1,0,1,0,0),nrow=3)
```

  Assume that this is the abundancy matrix of 3 species at 3 sites. How many individuals of which species have been found at which site? Having answered that, guess what the positions of the species relative to the positions of the sites are (of course it is more difficult to guess the positions of the sites relative to each other). Load the library vegan (The command `cca()` is in this package). Now apply CA to this matrix by executing `myca <- cca(Y)`. Produce the site conditional biplot and the species conditional biplot for `myca`. Do the biplots match your expectations about the correspondence between sites and species? If not, think why you could have guessed from the matrix what the correspondence is.

- Define the matrix

```
Y <- matrix(c(1,0,0,1,1,0,0,0,1),nrow=3)
```

  Again guess the relative positions of sites and species first in the species conditional biplot and then in the site conditional biplot. Done? Then produce the biplots and check your guess.

- Now the same for the matrix

```
Y <- matrix(c(1,0,0,0,1,1,0,0,1,1,1,0,1,1,1,1),nrow=4)
```

  First guess, then look at the biplots.

Exercise 2 Download the file `MexicanPlants.txt` from the web page.

- Use correspondence analysis to visualize the data. To avoid to many variables, only plot the families `grcyn, com, co, vi, ac ,cy, grresto` and `ru` and only the first 20 sites.
- Which family would you expect to find at which site with high probability?
- The dependency of grass on some environmental factors is rather linear than unimodal. Let's see whether PCA does a better job. Apply PCA to the same data set (and same variables) and compare the PCA biplots with the respective CA biplots.

Exercise 3 We continue with the Mexican plant data.

- Apply canonical correspondence analysis (CCA) to the Mexican plant data. Choose the families mentioned in the last exercise as species variables. Furthermore choose the variables `ALTITUDE`,`CATTLEINTENSITY`, `MAXVEGHEIGHT`, `Pasture` and `PLAGUE` as explanatory environmental variables. Use the whole dataset and not only the first 20 sites.
- Produce then the two triplots. What do you observe in the triplots? Which of the grass families prefers which environmental conditions? (A numeric answer is not possible. Your answer should be comparative.)

Exercise 4 In this exercise you shall reflect the assumptions 'linear dependence' and 'unimodal dependence'. Recall the following data sets from the last lectures and exercises. For each data set and each pair of response variable and explanatory variable argue whether the dependence is 'reasonably' linear or 'reasonably' unimodal. Note that there is sometimes no correct answer. It all depends on how you justify your answer.

- Artificial fish data (lecture on RDA)
- RIKZ data (lecture on RDA)
- Mexican plant data (lecture on CA and CCA)

Considering all pairs of variables is too much work. Pick from every data set interesting examples.