

MULTIVAR STATS IN ECOL AND GENETICS — EXERCISES, SHEET 6

Exercise 1 This exercise trains you in reading biplots. Consider the correlation biplot and the distance biplot of the height and weight data of the lecture.

- First execute the commands from the script to produce the biplots yourself. You find the file `HeightShoeWeight.txt` on the course homepage.
- In the distance biplot 'Gewicht' and 'Groesse' are orthogonal to each other. Does that mean that they are uncorrelated? You might want to confirm your answer with the `cor()` command.
- From the two biplots read off the approximate (Euclidian) distance between data point '211' and data point '103'.
- Is the data point '211' a student with normal weight, overweight or underweight?
- Guess the sex of student '121'.

Exercise 2 Let us continue as in Exercise 1. Consider the two biplots of the height and weight data of the lecture.

- By looking at the biplots, order the students '104', '106' and '226' by increasing weight. Hint: You need to use projections.
- The number of students used in the pca was $N = 226$. Looking at the lines in the correlation biplot how well (in %) is weight represented by the first two principal components.
- Due the projection in the plane of the first two principal components, the angle between lines in the correlation biplot might not well represent the correlation. Use the command `cor()` to find out which of the correlations between `weight`, `shoe` and `height` is represented poorly.
- By looking at the biplots, order the students '98', '103' and '106' by increasing shoe sizes.

Exercise 3 Download the data set 'EWU.txt' from the web page. The variables X1, X2, X3 and X4 have been explained in the lecture.

- Apply PCA to the data set and produce the distance biplot and the correlation biplot.
- Sort the countries *GB*, *D*, *I* and *IRL* according to their public dept level in 1997. Which of these countries have a debt level below the average?
- Apply all three rules of thumb from the lecture in order to decide which of the principal components are important.

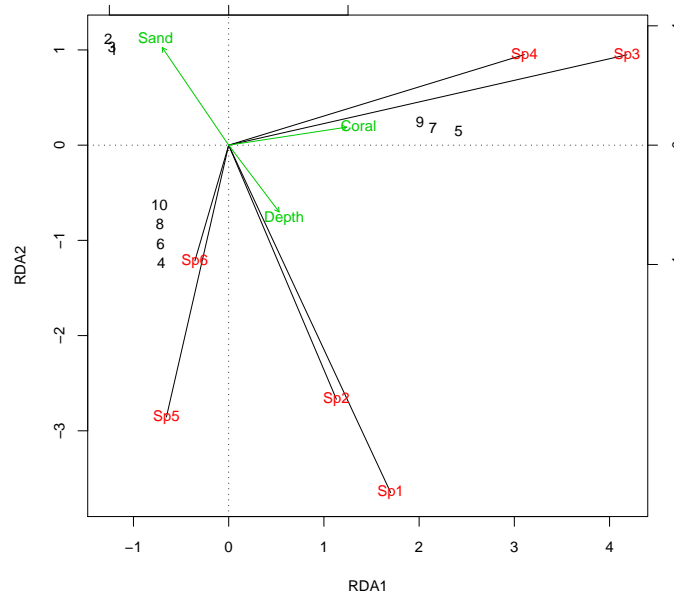
We will get practice in reading RDA triplots on the next exercise sheet. Here we do the first steps in RDA.

Exercise 4 Download the data set 'artificialFishes.txt' from the web page.

- Execute the R commands from the lecture and produce the triplots (do not forget to load the library `vegan` first).
- Play a bit around with `rda()`. For example try `scale=TRUE`, look in the help pages `?rda`, `?plot.rda` or plot the the returned RDA object with `plot()` and different arguments to the option `display=`.
- Here is a different way to produce a triplot:

```
plot(myrda, type="n", scaling=2)
segments(x0=0,
         y0=0,
         x1=scores(myrda, display="species", scaling=2)[,1],
         y1=scores(myrda, display="species", scaling=2)[,2])
text(myrda, display="sp", scaling=2, col=2)
text(myrda, display="bp", scaling=2,
     row.names(scores(myrda, display="bp")), col=3)
text(myrda, labels=fishes$Site, display="sites", scaling=2)
```

Note that `text(rda.obj)` calls `text.cca(rda.obj)`. Learn from these commands and try to produce the respective distance biplot which looks as follows



Exercise 5 Download the data set 'RIKZGroups.txt' from the web page. Which of the variables shall be the response variables and which shall be the explanatory variables? Apply redundancy analysis and produce the distance triplot and the correlation triplot. You might also want to try the commands from the last exercise on this data set.