2.6 ANOVA-simultaneous component analysis (ASCA)

ANOVA-Simultaneous Component Analysis (ASCA) is a method that is specifically suited to determine the chemical influences different factors may have. Therefore, it combines ideas from Analysis of Variance (ANOVA) with those of Principal Component Analysis. In these exercises you will use ASCA to describe the chemical defence that plants have against their natural enemies.

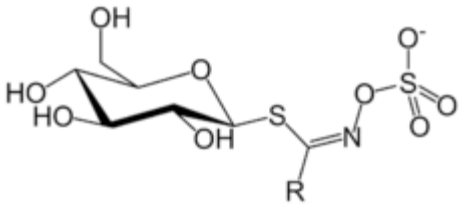

Figure 1 Glucosinolate

Plants in the Cabbage family (*Brassicaceae*) make very specific defence compounds when under attack by herbivores. These compounds are called glucosinolates: about 120 different species exist, varying only in the group R (Figure 1).

To simulate a herbivore attack and thus elicit the plant to make glucosinolates, the plant hormone Jasmonic Acid (JA) was administered to either the roots (Root-induced) or to the leaves (Shoot-induced) of cabbage plants (*B. oleracea*). Then the glucosinolate levels were measured 1, 3, 7 and 14 days after treatment. This measurement was destructive such that on every time-point different plants were analysed. The plants contain 11 different glucosinolates. We want to know now:

- **How the treatment affects the glucosinolate composition.**
- **Whether these changes also vary in time.**

Table 1 Glucosinolates

Variable #	Glucosinolate
1	PRO
2	RAPH
3	ALY
4	GNL
5	GNA
6	4OH
7	GBN
8	GBC
9	4MeOH
10	NAS
11	NEO

Exercises

---

1 . Raw data

a. Describe in words what kinds of chemical variation you expect there to be present in this experiment? Which factors and interactions may be present in the data (go back to ANOVA)? How would you split up the variation in the data caused by the experiment?

b. Inspect the raw data using 'exer.2.6.1'. Which glucosinolates change in time? Which differ between treatments?

---

2. ASCA model

a. Make an ASCA model of the mean-centered data, use 'exer.2.6.2' for this. Discuss how many PCs each submodel needs, taking into account the three criteria described in the lectures.

b. Use 'exer.2.6.3' to make an ASCA model of the plant data. Discuss the different submodels and which information they describe.

c. Compare and discuss the ASCA model with the PCA models you made in the previous session. Which advantages do each of the models have (parsimonity, interpretability)? In which case does it not make sense anymore to make an ASCA model?

---

3. **ANOVA and PCA**

a. Discuss how ASCA can overfit the data. Try to describe an approach by which you could test for this overfit.

b. How could you combine ANOVA and PCA in another way? How would this change your view on the experiment?
