

## 9 Multivariate Statistics

### 9.1 Introduction

### 9.2 Principal Component Analysis

In our example, the 1st and 2nd principal components contain 99.17% and 0.83% of the total variance in the data, respectively,

```
labels = []
for i in range(0,30+1):
    labels.append(str(i))
minerals = (['Min1', 'Min2', 'Min3'])
```

However, we use the original data  $x$  for `PCA()`.

We ignore  $PC_3$  at this point and concentrate on  $PC_1$  and  $PC_2$ .

This requires the second output of `PCA()`, which contains the principal component scores:

### 9.3 Independent Component Analysis (by N. Marwan)

### 9.4 Discriminant Analysis

### 9.5 Cluster Analysis

```
plt.title('Euclidean Distance Between Pairs of Samples',
          fontsize
```

### 9.6 Multiple Linear Regression

and observe some differences between the data and the regression plane if we rotate the graph in three dimensions using the `computer mouse`.

### 9.7 Aitchison's Log-Ratio Transformation

Next, we explore an example presented in an extended abstract by Pawlowsky-Glahn and Egozcue (2013). In their article (in German), the authors show spurious correlations and countermeasures in the sense of J. Aitchison in a three-component system consisting of *Sand*, *Lehm*, *Ton* (in English: sand, loam, and clay), with a dilution effect caused by *Wasser* (in English: water). According to the authors, we have the samples in the rows of the arrays, and the columns are *Sand*, *Lehm*, *Ton*, and *Wasser* in the data set of Scientist A, but they are only *Sand*, *Lehm*, and *Ton* in the data set of Scientist B. Both Scientist A and Scientist B obtain the same values for *Sand*, *Lehm*, and *Ton*, but after normalizing them to one, *Sand* and

*Lehm* are positively correlated in the analysis of Scientist A and negatively correlated in the analysis of Scientist B.

### **Recommended Reading**

### **Figure captions**