

# Statistical data analysis

## Multivariate confirmation analysis

### Introduction

The aim of this tutorial is to get familiar with multivariate confirm analysis. As such, the objective of confirmation analysis is to test hypotheses or theories model. These models could be based on a previous empirical research. For this purpose, you will use multivariate analysis of variance (MANOVA) to test your hypothesis about given data.

### 1 Input data

In this tutorial, we will work with a dataset that aims to the quality of potatoes growing in Oregon (for more details see [1]). Each potato is determined by its size, area, holding temperature, holding period, and cooking method. Overall quality consists of three aspects: texture score, flavor score, and moistness score. In this context, the properties of potatoes could be seen as independent variables; on the other hand, the quality of potatoes are dependent variables. For loading of the dataset, you can use a prepared function **Load** in *potato.R* file.

### 2 MANOVA

There are four main theoretical issues to be considered before running MANOVA. Not surprisingly, MANOVA has similar assumptions to ANOVA but extended to the multivariable case:

- **Independence:** Observations should be statistically independent.
- **Random sampling:** Data are randomly sampled from the population of interest.

- **Multivariate normality:** Dependent variables are multivariate normally distributed within each group of the independent variables, which are categorical.
- **Homogeneity of covariance matrices:** The population covariance matrices of each group are equal.

Take the initial assumptions granted. The assumption of multivariate normality can be tested using R with a test known as the Shapiro test implemented in **mshapiro.test** in *mvnrmtest* package. The assumption of equality of covariance matrices is often tested using Box's test that is implemented in **BoxMTest** function in *boxTest.R* file or as a **BoxM** function in *biotools* package.

### 3 Step by Step

You should go through the following steps:

1. Formalize your MANOVA hypothesis.
2. Visualize your data using *plot*, *boxplot*, *etc.* If data have more than two dimensions use a dimensionality reduction method.
3. Check the assumptions for MANOVA model.
4. Use MANOVA to test your hypothesis.
5. Discuss the obtained results (the meaning of the individual statistics, comparison with the visual analysis ad 2, practical implications), compare with the results of simpler MANOVA alternatives (e.g. repeated ANOVA).

### 4 Submission Form

Submit task as ZIP file containing all files necessary to run your code (try to test via Rscript) and write a short report (2-4 pages). The report should contains a definition of the task, interesting implementation details, and your answers for the required tasks.

## References

- [1] Mackey, Stockman: *Cooking Quality of Oregon-Grown Russet Potatoes*, American Potato Journal, pp. 395–407, 1958.