One-day statistics intensive course IV for soil scientists
on March 10th 2023 (in English)

IV. Multivariate statistics: principal component analysis, partial least squares regression, cluster analyses, regression trees and rule-based models using R for soil scientists

Almost all scientific studies rely to some extent on correct statistical analyses. While statistical software packages for scientists offer great opportunities and provide many powerful tools (e.g., in data mining and exploratory statistics), there are many pitfalls, which may result in wrong or nonreproducible manuscripts. This problem has been known for a long time and has been addressed explicitly in some research fields other than the geosciences. This short course aims to address potential problems in geoscientific studies and to reduce the number of non-reproducible studies.

Typical problematic fields in the area of multivariate statistics may be (I) a lack of knowledge of the opportunities and limitations of multivariate approaches; and (II) insufficient description of the multivariate analyses in publications.

The intensive course IV aims to improve soil scientists' statistical knowledge. A main objective is to reduce the occurrence of the above-mentioned problems in soil science research and publications.

Date: Intensive course IV: March, 10th 2023 9:30 a.m. until 5:00 p.m

Location: The intensive course will be held as Zoom session.

Costs: The costs for the intensive course IV are 80.00€. The price stated here is a final price and must be transferred before the start of the course.

Materials: Lecture notes (more than 100 pages), exercises and model solutions will be provided.

Recommended literature:

- Everitt & Hothorn (2011). An Introduction to Applied Multivariate Analysis with R. Springer.
- Crawley (2012). The R Book. 2nd Ed., Wiley.
- Lantz, B. 2019. Machine Learning with R. Packt Publishing.
- Wehrens (2020). Chemometrics with R. 2nd Ed., Springer.

- Ludwig, B., Wölfel, P., Greenberg, I., Piepho, H.-P., Spörlein, P. 2022. Application of mixed-effects modelling and rule-based models to explain copper variation in soil profiles of southern Germany. Eur. J. Soil Sci. https://doi.org/10.1111/ejss.13258

Lecturers: Prof. Dr. Bernard Ludwig and Dr. Isabel Greenberg

Schedule of the intensive course:
Friday, 10.03.2023

Time	Contents
09:30 - 11:00	Welcome, matrix operations, calculation of eigenvalues and eigenvectors, centering and z-transformation
11:00 - 11:15	Break
11:15 - 12:45	Variance-covariance and correlation, principal component analysis (PCA), calculations, presentations and interpretations
12:45 - 13:45	Lunch break
13:45 - 15:15	Pretreatment of data (use of the Savitzky-Golay filter), principal component regression (PCR) and partial least squares regression (PLSR)
15:15 - 15:30	Break
15:30 - 17:00	Introduction to partitional and hierarchical clustering. Presentation of regression trees and rule-based models.

Registration is open until February 15th 2023. Registration and general queries: Prof. Bernard Ludwig, Kassel University, bludwig@uni-kassel.de

Please note that the intensive course may be cancelled if not enough registrations have been received. In this case or in the event of cancellation due to illness of the lecturer or any other events beyond the control of the lecturer, there will be no claims possible, except for the reimbursement of participation costs.

Additional information on statistical training courses: see

https://www.uni-kassel.de/fb11agrar/en/sections-/-facilities/environmental-chemistry/statistics-courses