

**Two-day statistics intensive courses I and II for soil scientists  
from March 2<sup>nd</sup> until March 3<sup>rd</sup> 2023 (in English)**

**I. Fundamentals of descriptive and inferential statistics and introduction to R  
(March 2<sup>nd</sup>)**

**II. Exploratory statistics: statistical modelling and regressions using R for soil  
scientists (March 3<sup>rd</sup>)**

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.

Examples of such problems could be (I) research without hypotheses; (II) inappropriate experimental design; (III) a lack of understanding of pseudoreplication; (IV) an inappropriate handling of outliers; (V) missing inspections of conditions for hypothesis testing; (VI) an insufficient description of statistical analyses in publications; (VII) a lack of knowledge of the importance of residual inspections; (VIII) a lack of understanding of the differences between a minimal adequate model and a maximal model; (IX) a lack of knowledge of the differences between a calibration, a cross-validation and a validation of a model; and (X) a lack of understanding of important special topics such as polynomial regressions and model comparisons.

The intensive courses I and II aim to improve soil scientists' statistical knowledge and include an introduction to R. A main objective is to reduce the occurrence of the above-mentioned problems in soil science research and publications.

**Date:** Intensive course I: March 2<sup>nd</sup> 2023, 9:30 a.m. until 5:00 p.m.

Intensive course II: March 3<sup>rd</sup> 2023, 9:30 a.m. until 5:00 p.m.

**Location:** The intensive courses will be held as Zoom session.

**Costs:** The costs for the intensive courses are 80.00€ for each of the courses I and II; thus 160.00€ for the two days (courses I and II). The prices stated here are final prices and must be transferred before the start of the course.

**Materials:** Lecture notes (more than 200 pages), exercises and model solutions will be provided.

**Recommended literature:**

- Crawley (2012). The R Book. 2<sup>nd</sup> Ed., Wiley.
- Welham et al. (2014). Statistical Methods in Biology. Design and Analysis of Experiments and Regression, CRC.
- Glaz and Yeater (2020). Applied Statistics in Agricultural, Biological, and Environmental Sciences, Wiley.

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

**Schedule of the intensive courses:****Thursday, 02.03.2023**

Time	Contents
09:30 - 11:00	Welcome; introduction to descriptive statistics (measures of central tendency and of variation, normal distribution and other important distributions, central limit theorem, 95%-confidence interval); introduction to inferential statistics (hypothesis testing, fundamental tests, type I and II errors)
11:00 - 11:15	Break
11:15 - 12:45	Planning of experiments, experimental designs and data analysis (observational study vs. randomized controlled experiment, CRD, RCBD, replication, pseudo-replication and hypothesis testing)
12:45 - 13:45	Lunch break
13:45 - 15:15	Introduction to R (data types, vectors, lists and data frames, reading of data), tests of normality and of variance homogeneity, parametric (various t-tests) and non-parametric tests (Wilcoxon signed rank test, matched-pairs Wilcoxon test, Mann-Whitney U-test); exercises using R with practical examples
15:15 - 15:30	Break
15:30 - 17:00	Spearman and Pearson correlation methods, chi-square tests and exercises using R with practical examples: Correlations

**Friday, 03.03.2023**

Time	Contents
09:30 - 11:00	Welcome, inferential statistics: regressions, lack of fit test and residual inspections
11:00 - 11:15	Break
11:15 - 12:45	Exploratory statistics: statistical modelling (model types and model simplification)
12:45 - 13:45	Lunch break
13:45 - 15:15	Exploratory statistics: steps of model simplification and exercises using R
15:15 - 15:30	Break
15:30 - 17:00	Transformations (square-root and logarithmic), dealing with variability and predictions (confidence and prediction bands; validation and cross-validation), polynomial and non-linear regressions including exercises in R

**Registration is open until February 15<sup>th</sup> 2023. Registration and general queries: Prof. Bernard Ludwig, Kassel University, [bludwig@uni-kassel.de](mailto:bludwig@uni-kassel.de)**

Please note that the intensive courses 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/fb11agr/en/sections/-/facilities/environmental-chemistry/statistics-courses>**