Module title	**Modelling and Exploration of Multivariate Data**
Code	D3
Degree Programme	Master of Science in Life Sciences
Workload	3 ECTS (90 student working hours) - Asynchronous and synchronous distance learning, decentralized teaching: 32 h - Self-study: 58 h (10 h self-study before module starts)
Module Coordinator	**Name**: Prof. Dr. Yulia Sandamirskaya **Phone**: 079 7278966 **Email**: yulia.sandamirskaya@zhaw.ch **Address**: ZHAW Life Sciences und Facility Management, Schloss 1, 8820 Wädenswil
Lecturers	• Yulia Sandamirskaya
Entry requirements	Advanced knowledge of R (level D1) is required. Attending the module "Handling and Visualizing Data" is highly recommended. Prior to the module, additional mandatory preparatory reading, exercises and other material (videos, tests) will be made available to facilitate students preparation for the module. Students are advised to start five weeks before the module with the required preparatory work;
Learning outcomes and competences	After completing the module, students will be able to: • explore multivariate data by means of suitable visualisation and dimensionality reduction techniques • explore and describe the structure of multivariate data using clustering • explore and describe time series data on the basis of suitable visualisations and analysis methods analogue to multivariate data analysis • interpret, visualise and communicate the results of the analyses • use multiple regression models to answer research questions, understand their advantage over univariate methods; fit these models with R and quantify the fit of the model, describe the limitations of precision and reliability of inferential results; test the model assumptions; apply counter measures in case of problems with model assumptions • use elementary nonparametric regression methods to estimate the shape of not necessarily linear regression curves, understand the advantages and limitations of such flexible methods and apply related tools • perform elementary model selection and understand associated problems; test hypotheses, construct confidence and prediction intervals • identify typical pitfalls and amend these problems • understand typical statements in empirical research articles.
Module contents	The module introduces regression methods for data analysis and exploratory methods for multivariate data. Regression part: • Simple linear regression (including transformations) • Nonparametric regression (regression splines, local regression)

	• Multiple linear regression (including regression diagnostics) • Model selection (linked to hypothesis tests and p values) and inference (especially confidence intervals, prediction intervals) • Model diagnostics: assessment the validity of the model assumptions, reflect on the tools used to do this assessment • Possible strengths and limitations of parametric models (link to the exploratory part) Multivariate part: • Basic plots to characterise and visually inspect multivariate data and time series data • Dimensionality reduction techniques (principal component analysis, multi-dimensional scaling) • Clustering methods (k-means clustering and related approaches, hierarchical clustering, evaluation methods) Both parts: • Interpretation and visualisation of the results using suitable graphical representations of the data and the results (e.g. 3D scatter plots with regression surface or biplots) • Scientific reporting of the results, backtranslation from statistical methods to answer the original research questions to the data
Teaching / learning methods	In the weeks before module start, students are expected to do preparatory work to level prior knowledge. The workload is expected to be roughly 10 hours. The students receive preparatory and/or follow-up self-study work for each course day. The self-study consists e.g. of preparatory reading/videos, follow up exercises, examining case studies, etc. Central teaching is offered in a distance learning mode, consisting of asynchronous material such as videos and live consultation sessions. Details will be communicated one month before the start of the module. Local teaching consists of physical presence sessions where students actively solve exercises together with the local teachers. These exercises are meant to deepen the understanding of the material, give an opportunity to practice, provide extensions etc. The main type of tasks will be case studies which illustrate and exemplify the application of the material from central teaching to real life data sets and real problems. All the course contents come with comprehensive lecture notes and additional videos for an individual study and/or online learning.

Assessment of learning outcome	Project-based assignment. Details about the project will be communicated one month in advance. Students have the opportunity to earn bonus points during the local sessions. Attendance of at least 50% of the local sessions and active participation in solving tasks during the session or submission of all weekly exercises is required to get a 10% bonus (+0.5 to the grade) to the points received in the project. These points are required to achieve the maximal mark of 6.
Format	7-weeks
Timing of the module	For ZHAW and FHNW: Autumn semester, CW 45-51 For BFH and HES-SO: Spring semester, CW 15-21
Venue	online / decentralized teaching at respective school
Bibliography	Material will be provided on Moodle.
Language	English
Links to other modules	This module builds on module D1 “Handling and Visualising Data” and complements the module D2 “Design and Analysis of Experiments”.
Comments	Material treated during local teaching is relevant for the exam. Students have to make sure that an updated version of R is installed. Details will be communicated in advance.
Last Update	04.04.2024