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) - Lessons contact or online (total 42 of which 28 central teaching): 32 h - Self-study: 58 h
Module Coordinator	**Name**: Thomas Ott **Phone**: +41 (0)58 934 56 84 **Email**: thomas.ott@zhaw.ch **Address**: ZHAW Life Sciences und Facility Management, Einsiedlerstrasse 31a, 8820 Wädenswil
Lecturers	• Thomas Ott, ZHAW • Lorenzo Tanadini, BFH
Entry requirements	Attending the module "Handling and Visualizing Data" is required. 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	This module introduces <u>exploratory methods</u> and <u>regression models</u> for data analysis. <u>Exploratory part:</u> • 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)

Modelling part:

- Simple linear regression (including transformations)
- Nonparametric regression (regression splines, local regression) and quantile 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)

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

Tentative schedule:

Week	**<1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**>7**
Central		4L	4L	4L	4L	4L	4L	4L	
Local		2L	2L	2L	2L	2L	2L	2L	
Self-study	10h	38h							10h

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 a blend of classical lectures with more interactive teaching approaches such as large group activation methods and small group exercise sessions. Each course day contains an exercise session in which students work on the exercises, can ask questions and get individual support. Some parts of the central teaching can be offered in a distance learning mode. Details will be communicated one month before the start of the module.

Local teaching consists of students actively solving 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	Final written individualised exam (open book, using individual laptop computers to run data analyses) or project-based assignment to be conducted within 5-10 days (100%). Details will be communicated one month in advance.
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	For ZHAW and FHNW: Olten For BFH and HES-SO: Fribourg
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	04.02.2021