Module Description, available in: EN

Advanced Statistical Data Analysis

General Information

Number of ECTS Credits

3

Module code

FTP_AdvStDaAn

Valid for academic year

2019-2020

Last modification

2018-10-31

Responsible of module

Andreas Ruckstuhl (ZHAW, andreas.ruckstuhl@zhaw.ch)

Explanations regarding the language definitions for each location:

- Instruction is given in the language defined below for each location/each time the module is held.
- Documentation is available in the languages defined below. Where documents are in several languages, the percentage distribution is shown (100% = all the documentation).
- The examination is available 100% in the languages shown for each location/each time it is held.

	Berne	Lausanne			Lugano	Zurich		
Instruction						**X** E 100%		
Documentation						**X** E 100%		
Examination						**X** E 100%		

Module Category

FTP Fundamental theoretical principles

Lessons

2 lecture periods and 1 tutorial period per week

Entry level competences

Prerequisites, previous knowledge

Basic knowledge in probability, statistical inference and regression analysis on the level of Devore, Farnum and Doi, “Applied Statistics for Engineers and Scientists”, 2014 Cengage Learning.
User knowledge of R, MATLAB, Python or any other statistical software.

Brief course description of module objectives and content

One of the most used (statistical) models for inferential data analysis is the linear regression model. But it is restricted to a Gaussian distributed response and a linear function for linking the linear combination of predictors with the expected response. Generalized Linear and Additive Models (GLM, GAM) allow us to relax some of these restrictions by specifying a more general set of response distributions and nonlinear link functions. Hence we can analyze a wider variety of real world phenomenon such as counts, binary outcomes proportions and amounts (i.e. non-negative real-valued data). The aim of this modelling approach is to better understand the response outcome induced by the predictors based on the available data,

allowing for better and more informed interpretation of the phenomenon. In the first part, this course will provide an overview over the GLM approach and will detail many benefits and a few pitfalls.
The second part of this course covers the very popular and growing field of Bayesian statistics. We start with the fundamental principles of a Bayesian approach to the analysis of data that allows for a better accounting of uncertainty and more explicit statements of assumptions. We illustrate the basic mathematical framework as well as explanations of philosophy and interpretation. We set up and discuss introducing examples and extend them to more challenging modeling approaches, where we assess the outcome of a parameter in the face of uncertainty of other parameters. Completion of this part will give you an understanding and the ability to perform basic data analyses doing the Bayesian way.

Aims, content, methods

Learning objectives and acquired competencies

The students are able to analyze data by Generalized Linear and Additive Models (GLM and GAM) and understand the benefits that these model approaches offer for the analysis of normally and non-normally distributed response variables. They also perceive the difference between a frequentist and a Bayesian modelling approach. The students acquire a comprehensive overview how the open source statistical environment R is used and are able to perform a data analysis applying the techniques introduced in the course on real data sets.

Contents of module with emphasis on teaching content

First Part (8 weeks):

- Review of the concepts of multiple linear regression analysis with respect to inference, prediction, model evaluation and variable selection. (2 weeks)
- Extending the linear regression model to generalized linear and additive models including logistic, Poisson, and Gamma regression. Revise inference (including robust and modern methods), evaluation and variable selection for such models. (5 weeks)
- Some discussion about the consequences when modelling for predictive or inferential purposes. (1 week)

Second Part (6 weeks):

- Concept of Bayesian statistics, set up a Bayesian model with specifying suitable prior distributions, (3 weeks)
- Make inference in a Bayesian model by simulating from the model directly and by applying MCMC-techniques, e.g. the Gibbs- and Metropolis–Hastings sampling algorithm. (3 weeks)

The contents listed are illustrated with used cases from the industrial and scientific fields. The practical work is done with the open source statistical analysis environment R.

Teaching and learning methods

Lectures and practical work on computer with the statistical analysis environment R.

Literature

Slides and lecture notes will be available in addition to recommended book chapters.

Assessment

Certification requirements

Module does not use certification requirements

Basic principle for exams

As a rule, all the standard final exams for modules and also all repetition exams are to be in written form

Standard final exam for a module and written repetition exam

Kind of exam

written

Duration of exam

120 minutes

Permissible aids

Aids permitted as specified below:

Permissible electronic aids

statistical software R on examination laptop, if technically feasible

Other permissible aids

no restrictions

Special case: Repetition exam as oral exam

Kind of exam

oral

Duration of exam

30 minutes

Permissible aids

No aids permitted