COURSE PLAN – USN

Course plan – part of PhD program in Person-Centred Health Care

* Course name (Norwegian)

Modellering av data

* Course name (English)

Modeling data

* Course code

PHDMOD500

* Course level

PhD

* Scope (credits)

10 ECTS

* Language of instruction

English

* Number of semesters

2 – course will be taught over a period of 4 to 6 months

* Course summary

The course provides a broad introduction to modelling data using the free software R. The first part of the course will be aimed at learning basics of R, that are necessary for manipulating data, including a description of the various data types, how to manipulate them, how to input and output data, how to visualize data through graphics and how to integrate these into a report. Subsequently, the theory of several models will be explored and their application to data, using R. Initial topics will concentrate on linear models (e.g., ANOVA and linear regression), but then explore several extensions of them that allow describing and testing more varied data sets. These include mixed-effects models, nonlinear models and generalized linear and additive models. We will consider how to choose among these for a given data set and how to implement and evaluate them in R. Advantages of R are that most modelling functions have a common (or similar) type of interface, so changing models is fairly trivial most of the time and that the model evaluations are based upon a common approach, so that there is considerable transfer in the techniques to use from one model to the next. This makes it easier to concentrate on understanding the differences in the models rather than on how to implement them. Ideally, students will be led to think about the nature of their data and the type of questions that they are asking as guides to what kind of models to select.

* Required prerequisite knowledge

Students should have had a prior course in introductory statistics to be familiar with basic statistical concepts and tests. Notions from linear algebra (vector and matrix operations) and calculus (differentiation and integration) will be helpful.

* Learning outcome

Knowledge

- knowledge of basic data types in R, how to manipulate them, inputting and outputting data, visualizing data
- familiarity with the typical interface of modelling functions in R and the method functions required to analyse and interpret the results
- familiarity with the linear model and several extensions to it that permit characterizing diverse types of data
- knowledge of the concept of reproducible research, i.e., preparation of a report that includes text, data, code and analyses in a form that can be reproduced by another party

Skills

- ability to set-up and visualize data for modelling
- ability to critically choose among and compare different models of the data
- ability to evaluate models through simulation
- experience with a markdown language to produce reports for reproducible research

General competence

- can choose a model to fit to a data set, fit the model, and critically compare it with other models and evaluate its goodness of fit
- can critically justify choice of statistical models that are relevant to his/her research project
- basis for further studies in data modelling and statistics

* Learning activities

The course will be taught over a period of three months with three 2-3 day seminars with self-studies in between. Teaching and learning activities include lectures, discussions and workshops.

Supervised professional training

n/a

* Participation/compulsory work requirements

Compulsory attendance: students have to participate in at least 80 % of the lectures, discussions and workshops to be allowed to take the exam. Attendance will be registered.

* Examination

Oral presentation supplemented by a report outlining the analyses of data from his/her research. The report should contain a thorough analysis using the tools learned in the course.

* Examination support material

All

* Course evaluation

Feedback from students is essential in our efforts to ensure and further improve the high quality of the courses.

Other information

Students are expected to work on their own laptops throughout the course.

* Literature (reading list):

Required literature

A. de Vries & J. Meys (2015) R for Dummies, 2nd Edition, John Wiley & Sons.
A. J. Dobson & A. G. Barnett (2008) An Introduction to Generalized Linear Models, 3rd Edition. Chapman and Hall/CRC.

Supplementary literature:

N. Matloff (2011) The Art of R Programming. No Starch Press.
G. James, D. Witten, T. Hastie & R. Tibshirani (2013) An Introduction to Statistical Learning: with Applications in R, Springer.
W. N. Venables & B. D. Ripley (2002) Modern Applied Statistics with S. Fourth Edition. Springer
S. N. Wood (2006) Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.
K. Knoblauch & L. T. Maloney (2012) Modeling Psychophysical Data in R, Springer.
J. K. Kruschke (2014) Doing Bayesian Data Analysis, 2nd Edition, Academic Press.
J. C. Pinheiro, & D. M. Bates (2000) Mixed-Effects Models in S and S-PLUS, Springer.

* Approved course plan (date, name, title)