Faculty of Health, Science and Technology
Common

Syllabus

Multivariate Statistical Modeling

Course Code: 7FHV002
Course Title: Multivariate Statistical Modeling
Multivariat statistisk modellering
Credits: 7.5 ECTS
Degree Level: Doctoral (or Post-doctoral)

Course Approval
The syllabus was approved by the Faculty of Health, Science and Technology, 3 May 2017 and is valid from the autumn semester 2017.

Language of instruction
English

Prerequisites
The student should be currently enrolled in any doctoral program or post-doctoral program. The admission priority may be given to the students from the department of health sciences.

Learning Outcomes
Upon completion of the course, students should be able to:

1. Demonstrate the advanced levels of understanding of theories of parametric and non-parametric statistical modelling for research on population-based epidemiological data.
2. Analyze data for research projects to clarify the relationship among multiple variables with proper statistical models.
3. Compare and evaluate strengths and limitations of different statistical methods implemented in published studies, particularly in the context of multiple predictors and outcomes.

Course Content
This course will introduce statistical techniques to identify patterns of data distributions as well as to quantify the relationships among multiple factors. The course will be consisted of three parts. All three parts will take place in the campus.

1. Part I is composed of a series of lectures on the theories of several statistical methods as follows:

    - **Stepwise regression models**
    - **Least Absolute Shrinkage and Selection Operator (LASSO)**
    - **Structural equation modeling**
    - **Random forest algorithm**

2. Part II is to provide hands-on experience with data analysis through mandatory workshops. The instructor will provide a dataset consisted of clinical and genetic data from a public database. The students are also allowed to analyze their own data set. The students are expected to complete a project with a focus on variable and model selection and complete individual reports. We will choose three statistical models to demonstrate and allow the student to practice them at the workshop sessions.

    - **Stepwise regression model – using SPSS software**
    - **LASSO – using SPSS software**
    - **Random forest algorithm – using R package**

3. Part III is to provide a better understanding of strengths and limitations of different statistical models related to variable selection through literature reading. The students will also be expected to conduct presentations for their own projects with a focus on (1) interpretation of results and (2) strengths and limitations of their chosen statistical models.

Reading List
See separate document.

Examination (Assessment Test)
The grading/evaluation will be based on two components. First, the students are expected to take the written examination (in the form of multiple-choice questions). The purpose of the written examination is to evaluate the individual level of understanding of statistical modelling related to variable selection (Learning outcome 1). The results of the written examination will account for 50% of the final grade. Another part of assignment in the form of oral presentation will account for 50% of the final grade. Each student will present her or his analysis results and receive the on-site feedback. The instructor/examiner will evaluate her or his performance to see if the individual student has achieved Learning Outcomes 2 and 3.

Grades
One of the grades: Fail (U) or Pass (G) will be awarded.

Quality Assurance
The course convenor has a duty to encourage a continuous dialogue on learning processes and goal fulfilment. A written evaluation is carried out at the conclusion of the course combined with a joint student-teacher discussion of all aspects commented on. The result of the evaluation is collated and made available in accordance with *The Higher Education* Ordinance, Chapter 1, § 14.

Course Certificate
Course certificate is issued on request.

Faculty of Health, Science and Technology
Common

Reading List

Multivariate Statistical Modeling

Valid from autumn semester 2017

Course Code: 7FHV002
Course Title: Multivariate Statistical Modeling
Multivariat statistisk modellering
Credits: 7,5 ECTS
Degree Level: Doctoral

Benjamin A. Goldstein, Eric C. Polley, and Farren B. S. Briggs (2011). Random Forests for Genetic Association Studies. Stat Appl Genet Mol Biol. 2011 Jan 1; 10(1): 32. Published online 2011 Jul 12. doi: 10.2202/1544-6115.1691. PMCID: PMC3154091. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3154091/pdf/sagmb1691.pdf

Motulsky HJ (2015). Common misconceptions about data analysis and statistics. *Pharmacol Res Perspect*. 2015 Feb;3(1):e00093. doi: 10.1002/prp2.93. Epub 2014 Dec 2.

Ratnet B. (2010). Variable selection methods in regression: Ignorable problem, outing notable solution, J*ournal of Targeting, Measurement and Analysis for Marketing*, Vol. 18, 1, 65–75 (**required**).

Rosner B. (2010). *Fundamentals in Biostatistics*, 7th Edition, chapters 1-10 (optional), chapter 11 (**required**), Boston, Massachusetts, USA: Brooks/Cole

Sun, Z. et al. (2013). Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. *Environ Health.* 2013 Oct 4;12(1):85. doi: 10.1186/1476-069X-12-85.

Some other scientific articles will be added.