Lund University

School of Economics and Management

STAN46, Statistics: Functional Data Analysis, 7.5 credits

Statistik: Analys av funktionell data, 7,5 högskolepoäng

Second Cycle / Avancerad nivå

Details of approval

The syllabus was approved by The Board of the Department of Statistics on 2018-03-12 to be valid from 2018-03-13, spring semester 2018.

General Information

Language of instruction: English

Main field of studies	*Depth of study relative to the degree requirements*
Statistics	A1N, Second cycle, has only first-cycle course/s as entry requirements

Learning outcomes

Knowledge and understanding

For a passing grade the student shall

- demonstrate understanding of both technical and conceptual aspects of dealing with data treated as samples of functions,
- demonstrate knowledge of methods that goes beyond standard multivariate statistical methodology through dimension free approach that accounts for the order in observed variables,
- demonstrate understanding of the functional data approach which extends from statistical methodology such as the principal component analysis and multivariate regression,
- demonstrate understanding of how to represent data as functions, when the data are inherently too highly dimensional to be effectively treated as a vector, and
- demonstrate understanding of the role which derivative functions play in construction solutions to practical problems.

Competence and skills

For a passing grade the student shall

- demonstrate the ability to efficiently handle data that have an ordering inscribed in them and/or are highly dimensional through functional decomposition, and
- demonstrate skills in a number of algorithmic approaches, with assistance of software packages, both to analyse and visualise functional data.

Judgement and approach

For a passing grade, the student shall

- demonstrate insight into available choices for visualisation and data exploration in the functional setting,
- demonstrate the ability to make assessments of the type of approach that is most suitable for the problem at hand, and
- critically discuss the results of analysis obtained by a particular method.

Course content

In many practical problems originating from areas like geophysical sciences, astronomy, medicine, etc., the observations often correspond to measurements of a continuous process, which is a function of time and/or space. With rapid advances in technology, such measurements are becoming more and more frequent, thus pushing the boundaries of high-dimensional data analysis. In data science, high dimensional problems are often approached through dimension free analysis by treating an observation as a function. This is one way to circumvent the technical aspects identified with analysing high-dimensional data, and thus avoiding the curse of dimensionality. This course is meant to introduce the students to various aspects of dimension free analysis, which obviates the technical and computational hurdles associated with high dimensional data. This way of dealing with high dimensional data is an emerging and rapidly developing field that requires understanding both established methods and newly adopted techniques. The primary objective of this course will be to focus on the application of functional data analysis techniques to real world problems, and thus, mathematical rigour is often traded for adaptability to applications.

Beginning with the basics of the analysis of data that may be considered to be "functions", this course will discuss various visualization and data exploration techniques. Specifically, the course will extensively deal with nonparametric spline smoothing, functional linear models, functional PCA, regularisation methods, analysis involving derivatives, registration and nonlinear smoothing.

Students are required to work on projects to apply the techniques on real world problems. The preferred software for this course will be R and/or Python, however, the students are permitted to use any mathematical software of their liking that have facilities to perform all task in the course (Matlab being one example). Project discussions will enable students to share and compare ideas with each other and to receive specific guidance from the instructors. Efforts will be made to help students to embed real-world problems into mathematical models so that suitable algorithms can be applied with consideration of computational constraints. By surveying special topics, students will be exposed growing range of new methodologies.

Course design

The course is designed as a series of lectures, student presentations, and lab sessions with reports.

Assessment

Grading is based on individual performance, via written assignments, oral presentations as well as group activities.

The University views plagiarism very seriously, and will take disciplinary actions against students for any kind of attempted malpractice in examinations and assessments. Plagiarism is considered to be a very serious academic offence. The penalty that may be imposed for this, and other unfair practice in examinations or assessments, includes suspension from the University.

The examiner, in consultation with Disability Support Services, may deviate from the regular form of examination in order to provide a permanently disabled student with a form of examination equivalent to that of a student without a disability.

Subcourses that are part of this course can be found in an appendix at the end of this document.

Grades

Marking scale: Fail, E, D, C, B, A.

A (Excellent) 85-100 points/percent. A distinguished result that is excellent with regard to theoretical depth, practical relevance, analytical ability and independent thought.

B (Very good) 75-84 points/percent. A very good result with regard to theoretical depth, practical relevance, analytical ability and independent thought.

C (Good) 65-74 points/percent. The result is of a good standard with regard to theoretical depth, practical relevance, analytical ability and independent thought and lives up to expectations.

D (Satisfactory) 55-64 points/percent. The result is of a satisfactory standard with regard to theoretical depth, practical relevance, analytical ability and independent thought.

E (Sufficient) 50-54 points/percent. The result satisfies the minimum requirements with regard to theoretical depth, practical relevance, analytical ability and independent thought, but not more.

F (Fail) 0-49 points/percent. The result does not meet the minimum requirements with regard to theoretical depth, practical relevance, analytical ability and independent thought.

To pass the course, the students must have been awarded the grade of E or higher.

Entry requirements

General prerequisites for the Master's Program in Statistics, and STAN41 or a course in linear algebra with matrix calculus.

Further information

The course will be taught using the fda library in R or pyFDA in Python. The Functional Data Analysis package is also available for Matlab.

Subcourses in STAN46, Statistics: Functional Data Analysis

Applies from V19

1901 Laboratory work, 2,0 hp
Grading scale: Fail, Pass
1902 Written assignment, 2,0 hp
Grading scale: Fail, Pass
1903 Oral presentation, 3,5 hp
Grading scale: Fail, Pass