School of Economics and Management

STAN48, Statistics: Programming for Data Science, 7.5 credits

Statistik: Programmering för data science, 7,5 högskolepoäng

Second Cycle / Avancerad nivå

Details of approval

The syllabus was approved by The Board of the Department of Statistics on 2021-11-29 to be valid from 2022-08-29, autumn semester 2022.

General Information

Second cycle level course in statistics. The course is recommended in a Master's degree in statistics. The course may also be taken as a single subject course or within other Master's programmes at Lund University.

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

The course is a modern version of computational statistics education. It shifts from mainly methodological foundations toward programming skills that allow effective implementation of the methodology in one of the leading programming languages in the field of data science: such as R or Python.

Knowledge and understanding

For a passing grade, the student shall

- be able to explain the interplay between theory and its implementation in an algorithm,
- be able to clarify the conceptualization of the data in a form of a mathematical model involving stochastic components,
- be able to formalize the concept of algorithmic efficiency, and
- be able to identify pros and cons using different computing packages such as R and Python in a concrete statistical problem.

Competence and skills

For a passing grade, the student shall

- be able to independently write an algorithm in selective programming languages such as R and/or Python that implements a given statistical computation method,
- be able to download and manage massive data from different sources and download them into various computing environments,
- be able to carry out a complete analysis of data within a framework of given methodology and report in writing the outcomes,
- be able to improve inefficiently written algorithms and demonstrate advantages of proposed solutions,
- be able to choose a programming package that is most suitable for the problem at hand, and
- be able to communicate to a non-professional the meaning of statistical data analysis outcomes and interpret the way the statistical software represents them.

Judgement and approach

For a passing grade, the student shall

- be able to assess her/his approach and critically discuss and defend chosen solutions,
- be able to document her/his work and progress using annotation and documentation standards, so others can use and continue the work, and
- be able to distribute tasks of a general data analysis problem between the members of her/his group.

Course content

The course presents modern statistical computing as viewed in data science through implementations in popular computing platforms such as R and Python. It covers the following topics:

1. The programming environment: libraries for statistics
2. Working with Data Frames, Arrays, and Matrices
3. Methods for Generating Random Variables
4. Monte Carlo Integration, Inference and Variance Reduction
5. Bootstrap and Resampling
6. Bayesian computing
7. Markov Chain Monte Carlo Methods
8. Optimization and other numerical methods

Course design

The course is designed as a set of lectures, exercises, and labs. It is concluded with the project presentation sessions that serve as an exam. The lab projects and quizzes associated with the lab sessions and the solving problems sessions, respectively, and all contribute to the final grade.

Assessment

The exam is made from the final project and its public presentation which accounts for 50% of the final grade, the lab sessions 25%, and the quizzes 25%. Both quizzes

and the lab sessions must be taken to get a passing grade on each of these two components.

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 maybe 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

90 credits in Statistics, or a total of 90 credits i Mathematics, Programming, Mathematical Statistics and Statistics, of which at least 45 credits in Statistics or Mathematical Statistics, or the equivalent.

Subcourses in STAN48, Statistics: Programming for Data Science

Applies from H22

2201 Project, 3,5 hp
Grading scale: Fail, Pass
2202 Laboratory sessions, 2,0 hp
Grading scale: Fail, Pass
2203 Quizzes, 2,0 hp
Grading scale: Fail, Pass