UNIVERSITY OF GOTHENBURG

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DIT862 Statistical Methods for Data Science, 7.5 credits

Statistiska metoder för Data Science, 7,5 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Computer Science and Engineering on 2018-02-02 and was last revised on 2018-09-27 to be valid from 2018-11-05, autumn semester of 2018.

Field of education: Science 100%
Department: Department of Computer Science and Engineering

Position in the educational system

The course is compulsory within the Applied Data Science Master's Programme. It is also a single subject course at the University of Gothenburg.

The course can be part of the following programme: 1) Applied Data Science Master's Programme (N2ADS)

Main field of studies	*Specialization*
Data Science	A1F, Second cycle, has second-cycle course/s as entry requirements

Entry requirements

To be eligible to the course, the student should have a Bachelor's degree in any subject, or have successfully completed 90 credits of studies in computer science, software engineering, or equivalent. Specifically, the course requires the following:

- at least 15 credits of successfully completed courses in programming,
- one of the courses DIT851 Introduction to Data Science, 7.5 credits, or DIT856 Applied Mathematical Thinking, Master course, 7.5 credits, alternatively at least 7.5 credits of mathematics.

Applicants must prove knowledge of English: English 6/English B or the equivalent level of an internationally recognized test, for example TOEFL, IELTS.

Learning outcomes

After completion of the course the student should be able to:

Knowledge and understanding

- account for basic notions of probability theory and statistical theory, with an emphasis on applications and experiments in data science
- give examples of how probabilistic models are applied in data science applications

Competence and skills

- compute descriptive statistics and visualizations using statistical software libraries
- implement a number of probabilistic models and apply them in data science applications
- apply statistical tests for evaluating data science applications

Judgement and approach

- justify which type of statistical method is applicable for the most common types of experiments in data science applications
- discuss advantages and drawbacks of different types of probabilistic models that can be applicable for a given data science application
- discuss how the misuse of statistical techniques can lead to dubious conclusions

Course content

The course gives an introduction to the theory of probability and statistics, data analysis using descriptive statistics and data visualization, and applications of probabilistic modeling in data science.

In the course, the following broad areas will be covered:

- data analysis including descriptive statistics and data visualization
- probability theory including basic probability calculations, random variables, distributions
- statistical methods including point and interval estimates, hypothesis testing, regression
- probabilistic models in data science applications, for instance, Naive Bayes classifiers and topic models for text or Hidden Markov Models for sequences

Sub-courses

1. **Take-home examination** *(Hemtentamen),* 4 credits
   Grading scale: Pass with Distinction (VG), Pass (G) and Fail (U)
2. **Assignments** *(Inlämningsuppgifter),* 3.5 credits

Grading scale: Pass with Distinction (VG), Pass (G) and Fail (U)

Form of teaching

Lectures, exercise sessions, computer lab sessions.

Language of instruction: English

Assessment

The course is examined by an individual written take-home exam, as well as mandatory written assignments submitted as written reports, some of which will be carried out individually and others in groups of normally 2-4 students.

There will be non-obligatory individual assignments which grant bonus points for the written exam. These bonus points are valid for the whole academic year.

Late submission of the take-home exam results in the grade Fail (U), unless special reasons exist. A failed take-home exam is reexamined by a new take-home exam.

If a student, who has failed the same examined component twice, wishes to change examiner before the next examination, a written application shall be sent to the department responsible for the course and shall be granted unless there are special reasons to the contrary (Chapter 6, Section 22 of Higher Education Ordinance).

In cases where a course has been discontinued or has undergone major changes, the student shall normally be guaranteed at least three examination occasions (including the ordinary examination) during a period of at least one year from the last time the course was given.

Grades

The grading scale comprises: Pass with Distinction (VG), Pass (G) and Fail (U).
A Pass grade (G) for the entire course requires at least a Pass grade for all sub-courses.

To be awarded Pass with Distinction (VG) for the full course, the grade VG must be obtained on both the sub-courses.

Course evaluation

The course is evaluated through meetings both during and after the course between teachers and student representatives. Further, an anonymous questionnaire is used to ensure written information. The outcome of the evaluations serves to improve the course by indicating which parts could be added, improved, changed or removed.

Additional information

Course literature to be announced the latest 8 weeks prior to the start of the course.

The course replaces the course DIT861, 7.5 credits. The course cannot be included in a degree which contains DIT861. Neither can the course be included in a degree which is based on another degree in which the course DIT861 is included.