Department of Statistics, VT 2025

Statistical Theory and Modeling, Master's level, 7.5 hp, ST2601

The course is part of the Master's program in data science, statistics and decision analysis, 120 credits.

Course contents

This course is a continuation of the course *Statistics and Data Analysis for Computer and Systems Sciences, 15 credits*. It provides advanced and expanded knowledge in probability theory, modelling, and statistical inference theory, which forms the foundation for modern machine learning and AI. Theoretical derivations are connected with practical applications in programming languages for data analysis and machine learning. Regression and classification models are a significant part of the course. Prediction-based methods such as cross-validation are used for model selection. The course offers an introduction to the mathematical methods that are essential tools for the course and ends with an introduction to time series analysis.

The following topics are covered:

- **Mathematical methods**: derivatives, integrals, optimization, numerical optimization, vectors, and matrices.

- **Probability theory**: discrete and continuous stochastic variables, density and probability functions, distribution functions, multivariate distributions, multivariate normal distribution, marginal distributions, conditional distributions, independence, expected value, variance, and covariance, functions of stochastic variables, sampling distributions, law of large numbers, central limit theorem.
- **Modelling and prediction**: linear and non-linear regression, dummy variables and interactions, model selection, cross-validation, overfitting, regularization, classification, logistic regression, multinomial logistic regression, Poisson regression.
- **Inference**: point estimations, bias-variance trade-off, maximum likelihood (ML), likelihood theory, numerical optimization for ML estimations, bootstrap.
- **Time series**: trend and seasonality, autocorrelation, autoregressive models.

The examination consists of

- **Exam 1** (exam code 11ST): Statistical theory and modelling, exam.
- **Exam 2** (exam code 12SI): Statistical theory and modelling, home assignment.

Learning outcomes

To pass the course, students should be able to:

Knowledge and comprehension:

- Define and describe theoretical concepts in probability.
- Explain the principles behind inference methods and their use in practical applications.
- Select appropriate models and methods depending on the situation and research question within the course content.

Skills and abilities:

- Formulate and solve problems of advanced nature within probability theory.
- Formulate and solve problems of advanced nature related to estimation, inference, and prediction.
- Perform calculations, simulations, and analyses in a programming language related to the course content.

Evaluation ability and approach:

- Interpret, evaluate, and critically review results with respect to relevant scientific aspects.

Teachers and general information

Teachers	Roll
Mattias Villani	Course responsible, examiner and lecturer
Fasna Kottakkunnan	Exercises and Computer labs
Ralf Xhaferi	Exercises and Computer labs

All teachers have office hours by appointment. Contact us via email to schedule a meeting time (either on campus or via Zoom).

The Statistical Department is located at the newly built Campus Albano, Albanovägen 12, House 4, 6th floor. General information regarding the department can be found on the department website.

All course material - including slides, reading instructions and schedule link - are available on the course page.

Student-specific information, submission of the home assignments and messages during the course are handled on the learning platform Athena.

Course evaluation

After the course is completed, an evaluation of the course is conducted. The course evaluation is used as part of the quality work for the course and as a means of student influence. The evaluation is carried out by sending a survey via email to all registered course participants. The responses from the course participants are compiled and, along with the final report/course evaluation from the course responsible teacher, uploaded to Athena.

Teaching and mandatory participation

The teaching consists of 12 lectures (F1-F12), 8 exercises (Ö1-Ö8) and 3 computer lab sessions (DL1-DL3) according to the schedule. There are also eight scheduled slots for getting help from one of the teaching assistants (jour). Please see the link to TimeEdit on the course page for the current schedule.

Participation in all teaching activities is voluntary, but **strongly recommended**. Learning statistics takes a lot of practice and it will be hard to pass the course without the necessary effort.

Knowledge Check and Examination

Exam 1 is an individual written exam - using both pen and paper and computer - graded on a seven-point scale related to the course objectives:

> **i** Grade Scale for Exam 1
>
> A Excellent
> B Very good
> C Good
> D Satisfactory
> E Sufficient
> Fx Insufficient, requires some additional work
> F Insufficient, requires much more work

Exam 2 is a group assignment, graded on a two-point scale: pass (G) or fail (U).

To pass the course, a minimum grade of E on Exam 1 and a pass grade on Exam 2 are required. The final grade for the course is determined by the grade on Exam 1.

- Students who receive at least an E grade on the exam may not retake the exam for a higher grade.
- Grades of Fx and F on the exam are failing grades and require re-examination. Therefore, students who receive an Fx grade cannot retake the exam for a higher grade.
- Students who receive an Fx or F grade on an exam have the right to undergo at least four additional exams as long as the course is offered to achieve at least an E grade.
- Students who receive an Fx or F grade on two occasions from an examiner have the right to request that a different examiner be appointed to determine the grade at the next exam session. This request must be made in writing to the head of department.
- Two examination opportunities are available for each exam during the current semester.

Exam 1 (11ST): Statistical theory and modelling, exam, 6 credits

- The exam (code 11ST) is an individual written exam, using both pen and paper and computer.
- The exam duration is 5 hours.

- Internet access is not permitted during the exam.
- Personal code from your own solutions to computer labs is allowed as a resource. The personal code file must be uploaded for verification before the exam (as instructed on Athena).
- Collaboration is not allowed during the exam, nor are any aids other than those permitted by the examiner.
- Special accommodations may be allowed upon request to the department's study and career counselor and with the examiner's approval. Contact the study and career counselor well in advance of the exam, preferably no later than three weeks before the exam date.
- Rules governing exams at Stockholm University can be found at: Rules for on-site exams

> **!** Note!
>
> Remember, you must register at least 10 days before the exam. If you have registered correctly, you will receive a confirmation email with an anonymous code. This confirmation serves as your receipt of registration. If you need to re-register for an old course code, you can only do so via email to expedition@stat.su.se. **Failure to register means you cannot take the exam!**

See TimeEdit for scheduled exam sessions.

Exam 2 – Home Assignment, 1.5 credits

- The assignments are group work, with three (3) persons per group. Group division is done on Athena under the "Create Workgroup" document in the Assignments folder.
- The assignment consists of 2 subprojects presented as written reports in R. Instructions for the assignments will be available on Athena at the start of the course.
- Collaboration within the group is allowed, but individual assessment and grading within the group may occur. All group members are responsible for and must be able to account for all parts of the work presented in the reports. Collaboration between groups is permitted, but each group must submit their unique reports. All forms of plagiarism are prohibited, and text matching software may be used as needed. Read Rules and Handling Procedures for Disciplinary Matters.
- Exam 2 is graded Pass (all subtasks in both projects approved) or Fail (at least one subtask fails). If one or more subtasks fail, there is an opportunity for correction during the current semester.

- Note! It is not possible to correct if the submission is only made at the second opportunity. This means that if you miss a submission and instead submit at a later time and fail, you cannot correct the assignment.
- Note! All subtasks in both projects must be completed and approved during the current semester for the entire assignment to be approved. Results from subtasks are not saved and cannot be transferred to future semesters.

Submission Deadlines

- Deadline for submission of home assignment: TBD
- Deadline for possible correction: TBD

Grading Criteria

Exam 1 (11ST), 6 credits

The exam (code 11ST) is an individual written exam, using both pen and paper and computer. The written exam covers material according to the course content.

Exam 1 Grading Criteria

A (Excellent): The student can excellently use concepts within probability and inference theory not necessarily directly covered in the course. They can solve and interpret complex problems concerning random variables, distributions and estimators correctly and well-structured, set up simple statistical models for specific situations, and use R for calculations and data analysis. Requires at least 90% on the written exam.

B (Very Good): The student can very effectively use concepts within probability and inference theory covered in the course. They can solve and interpret complex problems concerning random variables, distributions and estimators correctly and well-structured, set up simple statistical models for some concrete situations, and use R for calculations and data analysis. Awarded for 80-89% on the written exam.

C (Good): The student can effectively use concepts within probability and inference theory covered in the course. They can correctly solve and interpret problems concerning random variables, distributions and estimators, set up simple statistical models for some concrete situations, and use R for calculations and data analysis. Awarded for 70-79% on the written exam.

D (Satisfactory): The student can satisfactorily use concepts within probability and inference theory covered in the course. They can correctly solve and interpret problems concerning random variables, distributions and estimators, set up simple statistical models for some concrete situations, and use R for calculations and data analysis. Awarded for 60-69% on the written exam.

E (Sufficient): The student can sufficiently use concepts within probability and inference theory covered in the course. They can mainly correctly solve and interpret problems concerning random variables, distributions and estimators, set up simple statistical models for some concrete situations, and use R for calculations and data analysis. Awarded for 50-59% on the written exam.

Fx (Fail, additional work required): The student's performance is insufficient according to at least one criterion for E. Equivalent to 40-49% of points on the written exam.

F (Fail, much more work required): The student's performance shows clear deficiencies according to the criteria for E. Equivalent to 0-39% of points on the written exam.

Exam 2 (12ST), 1.5 credits

Grading for the assignment is Pass (G) or Fail (U). The following grading criteria apply:

i Exam 2 (12ST) Grading Criteria

G (Pass): The student has set up appropriate statistical models for given situations, demonstrated sufficient ability to use statistical terminology, utilized R correctly, and presented the results in well-written reports following the instructions. All subtasks should be correctly solved.

U (Fail): The student's performance is insufficient according to at least one criterion for Pass.

Course Literature

The main book for the course is:

- Wackerley, Mendenhall and Scheaffer (2021). Mathematical Statistics with Applications, 7th edition, Cengage.

Other course materials such as supplementary materials, lecture notes, exercises, and instructions for assignments will be posted on the course web page. Links to selected online sources will also be provided there.