![Stockholm University logo]

**Course Description**

**Selected statistical methods with applications,** ST3201, 7.5 hp

**1. Course content**
The course is intended to broaden students' knowledge in statistics, provide good examples of current statistical method and give inspiration for the subject of the bachelor thesis.
The course comprises 3 selected statistical areas/themes: text analysis, handling missing data with imputations and survival analysis. For each area/theme, the course relates theory to practice by combining lectures, practical exercises and programming. A central element is the students' independent work with a number of problems (case studies), which are presented in both written and oral presentations.

**2. Intended learning outcomes**
To achieve a passing grade, the student must, for all methods covered in the course, be able to:

*Knowledge and comprehension*:
- describe definitions and central concepts
- explain relevant theory
- discuss the pros and cons of different methodological considerations.

*Intellectual and practical skills*:
- solve method-specific theoretical problems
- practically apply the selected methods
- perform calculations and analyzes in R
- present the method area and applicable problem solutions in writing and orally.

*Transferable skills and critical and independent thinking*:
- interpret, evaluate and critically review results with regard to relevant scientific aspects.

**3. Teaching format**

Teaching consists of lectures and seminars (F1-F12), and computer exercises (C1-C3).
Attendance at the seminars is compulsory. Absence from one or more seminars can be compensated by attending a recapitulation seminar. The course examiner will inform about the date for this seminar via Athena.

Submission of home assignments before seminars and oral presentations at seminars are mandatory elements of the course.

Teaching takes place in English. The course is given on campus, but there may be elements of digital teaching (this will be announced it via Athena).

## Part 1: **Text mining**

Nowadays, there are vast quantities of unstructured textual information available, for example, from emails, medical journals, social media activities, movie recommendations and web server logs. It is of utmost importance e.g. for companies to be able to analyze such an information and make it quantifiable, in order to see trends and remain competitive.

This part provides an introduction to quantitative methods for analyzing text. Students will learn how to retrieve the text from the original source, process the text data, analyze and summarize the results of text mining experiments.

For successful completion of this part students should be able to: describe and explain basic concepts related to text mining, apply text mining methods to practical problems using software R, discuss the results from text mining experiments.

Teaching consists of 3 lectures (L1 – L3), one seminar (L4) and one computer-session (C1) according to separate schedule. Lecture notes and other relevant material can be handed-out (or made available in Athena) in connection with the lectures and computer session.

**Literature:** research papers and other material provided by instructor, the material will be available via Athena at course start.

**Responsible (examiner) for part 1**

Tatjana von Rosen: Room A4603, Bldg. 4, 6th floor (Albano Campus (Alvanovägen 12), E-mail: tatjana.vonrosen@stat.su.se. Consultation hours: In connection with lectures or by appointment.

## Part 2: **Handling missing data with imputations**

This part provides an introduction to handling missing data with imputations. The following topics will be considered: reasons for missing data, types of missing data, and methods to dealing with missing data (simple and multiple imputation methods). Additionally, in the computer session students will learn how to implement the imputation techniques using R.

For successful completion of this part students should be able to: describe and explain basic concepts related to missing data and imputation techniques, and be able to handle missing data of different types using software R.

Teaching consists of 3 lectures (L5 – L7), one seminar (L8) and one computer-session (C2) according to separate schedule. Lecture notes and other relevant material can be handed-out (or made available in Athena) in connection with the lectures and computer session.

**Literature:** research papers and other material provided by instructor, the material will be available via Athena at course start.

**Responsible (examiner) for part 2**

Tatjana von Rosen: Room A4603, Bldg. 4, 6th floor (Albano Campus (Alvanovägen 12), E-mail: tatjana.vonrosen@stat.su.se. Consultation hours: In connection with lectures or by appointment.

<u>Part 3</u>**: Analysis of survival data**

This part provides introduction to models and methods used in the analysis of survival (duration) data - with applications in the social sciences. Relevant R-procedures will also be covered in the computer session. Preliminary topics that will be covered include: Censoring and other special features of survival data; functions of survival time; Kaplan-Meier and life-table estimation of survival functions; log-rank test for comparison of survival functions; Cox proportional hazards models; parametric survival models.
For successful completion of this part students should be able to:
describe and explain basic concepts, functions, and distributions for survival data; compute and compare survival functions for different groups; and model associations between survival functions and explanatory variables using R.

Teaching consists of 4 lectures (L9 - L12) and one computer-session (C3) according to separate schedule. Lecture notes and other relevant material can be handed-out (or made available in Athena) in connection with the lectures and computer session.

**Literature:** Moore, D. F. (2016), *Applied Survival Analysis Using R*. Springer Verlag. Available online via SU-library
**Responsible (examiner) for part 3**
Gebrenegus Ghilagaber: Room A4602, Bldg. 4, 6th floor (Albano Campus (Alvanovägen 12), E-mail: Gebre@stat.su.se. Consultation hours: In connection with lectures or by appointment.


**4. Examination and assessment criteria**

a) The course is continuously examined through individual written assignments and oral presentations of these during the mandatory seminars. The examination consists of an assignment and an oral presentation per theme of the course (2-4 themes depending on the course occasion). Each written submission is scored according to the grading criteria and these scores are added together for a total score (see point d) below)

b) Grading of the course is according to a seven-point goal-related grading scale: A = Excellent, B = Very good, C = Good, D = Satisfactory, E = Adequate, Fx = Inadequate, F = Totally Inadequate.
Both Fx and F are failing grades and require re-examination.

c) The written grading criteria will be communicated to the students at the latest at the start of the course.
d) For a passing grade on the course, a pass is required on **all** assignments and oral presentations as well as full attendance at the seminars. Weighted grade for the course is determined by a weighted average of the points (percentage of maximum points) on the assignments. Standard rounding rules apply. If there are special reasons, the examiner can grant a student exemption from the obligation of one or more compulsory parts. The student then needs to compensate with another kind of task. Examination assignments that are not submitted on time will not be graded.

Students who have received a grade of E or higher may not undergo a renewed examination for a higher grade.
Possibility of supplementing the grade Fx up to a passing grade is not allowed on this course.