Stockholm University

Gebrenegus Ghilagaber

STOCKHOLM UNIVERSITY
Department of Statistics
Spring Semester 2025

24 February 2025

Course Description for **Generalized Linear Models: Theory and Applications,** 7.5 **H**igher **E**ducation **C**redits (HEC) at advanced-level, Course Code: ST5601. Spring 2025.

1. COURSE CONTENTS

The course deals with statistical models that extend the linear modelling framework to response variables that may not be normally distributed. These include binary response (proportions of success or failure records), multinomial responses such as responses in Likert scale (such as responses to questionnaire items), discrete responses such as counts, etc. Focus will be on both the theoretical (logical and mathematical) foundations of the techniques and their applications to real-life data using computer programs. Prior knowledge in probability & inference theory, calculus, matrix algebra, and command of computer programs like R is assumed. Topics that will be covered in more details include:

- Review of Likelihood Theory and Inference
- Exponential Family and Generalized Linear Model Theory
- Linear Models for Continuous Responses
- Logit Models for Binary, Ordinal, and Sequential Responses
- Poisson Models for Count Data and Log-linear Models for Contingency Tables
- Models for Longitudinal and Clustered Data

2. LEARNING GOALS

After completing the course, the student is expected to:

Knowledge and comprehension:
- understand and describe the principles for generalizing linear models using the theory of exponential family of distributions
- understand and describe likelihood-based inference.
- select suitable generalized model, link function and method of analysis in various situations

Intellectual and practical skills:
- solve problems and perform computation and analyses in generalized linear models both analytically and through programming in R.
- apply inference with and evaluation of the generalized linear models covered in the course

- *Critical thinking and appraisal*:
- interpret, evaluate and critically review empirical results in the light of relevant scientific aspects,
- resonate on potential limitations and sources of error in a given analysis.

3. TEACHING FORMAT

Teaching consists of 12 lectures (L1 - L12) and 6 computer-sessions (C1-C6) according to schedule. More detailed instructions on these will be provided during the first lecture (L1) and/or first computer session (C1). The course is given at full-pace on campus though some lecture notes may be available in Athena in connection with the lectures and computer sessions.

4. EXAMINATION AND GRADING

Performance in the course is assessed through individual examination, 6 hec. (exam-code 11 BT) and hand-in assignment that can be done in groups of two students, 1.5 hec (exam-code 11 BI). The examination is graded according to a seven-scale grading system: 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. The hand-in assignment is graded as G = Passed or U = Failed. Individual performances within the group-work (hand-in assignment) will be assessed.

The individual examination (max 50 points) consists of two parts: Part I: a written examination (max 30 points) that covers the material of the course, Part II: a written examination (max 20 points) on practical data analysis using software R. The minimum points required to pass the course are 20 points for Part I and 10 points for Part II.

Further requirements for the Part II:

- The student should properly use the theory of Generalized Linear Models when conducting data analysis.
- The student should be able to correctly use statistical terminology and clearly present and interpret the obtained results.
- The statistical software used should be R. • The results should be reported in the form of a written report.

The final grade of the individual examination is based on the sum of points for Part I and Part II, according to the criteria given in Table 1.

The following criteria are used for assessing the hand-in assignment:

- **Pass:** The student can properly apply the theory of Generalized Linear Models when conducting the task specified in the Home Assignment. The student can identify and use an appropriate statistical model for data analysis and argue for its adequacy. The student can in a satisfactory way present in a written form his/her findings, interpret them using proper graphics, descriptive statistics, significance test etc., explain important statistical concepts relevant to the task, discuss limitations and possible error sources in the analysis. The student can correctly use statistical software R and summarize his/her work in a well-written report using a correct statistical language.

- **Fail:** The student cannot correctly apply the theory of Generalized Linear Models when conducting the task specified in the Home Assignment. The student cannot identify and/or use an appropriate statistical model for data analysis and argue for its adequacy. The student is not able in a satisfactory way to present his/her findings, the student cannot interpret the obtained results using proper graphics, descriptive statistics, significance test etc. The student cannot explain important statistical concepts relevant to the task, and/or discuss limitations and possible error sources in the analysis. The student cannot correctly use statistical software R. The student is not able to summarize his/her work in a well-written report using a correct statistical language.

- A grade of E or higher is required in order to pass the course.
- Students who have attained the grade Fx or F on an individual examination are entitled to at least four additional individual examinations to achieve at least grade E as long as the course is offered.
- Students who have received at least grade E on an individual examination may not retake another examination in an attempt to achieve a higher grade.
- Students who have received the grade Fx or F on individual an examination on two occasions by the same examiner have the right to request that a different examiner be appointed to set the grade of the examination on the next possible occasion. The request must be in writing and sent to the head of the department. The examination denotes all compulsory elements of the course.

For a passing grade on the course, a minimum grade of E on an individual examination and a grade of G on Home Assignment is required. The combined grade for the entire course is equated to the grade for the individual examination. Examination assignments that are not submitted on time will not be graded.

Often the student should sign up for the written exam within 10 days before the exam, but special cases can change this, so this should be checked by the student. Information that the students need to register for exams in Ladok and deadline for registration is posted in Ladok under each course. The student can sign up for the exam 30 days in advance. If that is done correctly an anonymous code is generated immediately on the screen. If there is trouble doing this contact the student expedition (expedition@stat.su.se). **If you have not signed up you cannot take the exam!**

5. COURSE LITERATURE AND OTHER TEACHING MATERIALS

- Dobson, A. J. and Barnett, A. G. (2018). An Introduction to Generalized Linear Models (Fourth Edition). Boca Raton, FL: Chapman and Hall/CRC.
- Software: R
- Some additional material may be handed out (or made available in Athena) during the course.

6. COURSE INSTRUCTORS AND EXAMINER

- Gebrenegus Ghilagaber (Lectures & Examination): Room A4602 (Bldg. 4, Albanovägen 12, 6th floor, Albano Campus). Tel. 08-162983, E-mail: Gebre@stat.su.se
  Consultation hours: In connection with Lectures or by appointment.

- Fasna Kottakkunnan (Computer Sessions): (Bldg. 4, Albanovägen 12, 6th floor, Albano Campus). E-mail: Fasna.Kottakkunnan@stat.su.se
  Consultation hours: In connection with Computer Sessions or by appointment.

Table 1: Grade criteria for the individual examination:

Grade	Criteria (Requirements)
A	**Excellent:** The student can correctly describe and use the theory of generalized linear models as it is considered in the course. Moreover, the student can apply this theory to practical problems that have not been directly discussed in the course. The student can clearly present correct solutions to problems relevant for the course and use a correct statistical language. The grade corresponds to a total of **47.5 - 50** points of the examination score.
B	**Very Good:** The student can correctly describe and use the theory of generalized linear models as it is considered in the course. The student can apply this theory to practical problems of a type that has been discussed in the course. The student can clearly present correct solutions to problems relevant for the course and use a correct statistical language. The grade corresponds to a total of **45 - 47** points of the total examination score.
C	**Good:** The student can correctly describe and use the theory of generalized linear models as it is considered in the course. Moreover, the student can apply this theory to practical problems of a type that has been discussed in the course. The student can present correct solutions to problems relevant for the course and use a correct statistical language. The grade corresponds to a total of **40 - 44.5** points of the total examination score.
D	**Satisfactory:** The student can correctly describe and use the theory of generalized linear models as it is considered in the course. The student can apply this theory to a majority of practical problems that have been discussed in the course. The student can present correct solutions to a majority of problems relevant for the course and use a satisfactory statistical language. The grade corresponds to **35 - 39.5** points of the total examination score.
E	**Adequate (Sufficient):** The student can correctly describe and use the theory of generalized linear models as it is considered in the course. The student can apply this theory to a majority of practical problems that have been discussed in the course. The student can present satisfactory solutions to problems relevant for the course and use a satisfactory statistical language. The grade corresponds to **30 - 34.5** points of the total examination score.
Fx	**Inadequate (Insufficient):** The student cannot correctly describe and use the theory of generalized linear models as it is considered in the course. The student cannot present satisfactory solutions to problems relevant for the course or cannot use a satisfactory statistical language. The grade corresponds to **20 - 29.5** points of the total examination score.
F	**Totally Inadequate (Totally Insufficient):** The student cannot describe and use the theory of generalized linear models as it is considered in the course. The student cannot present satisfactory solutions to problems relevant for the course or cannot use a statistical language. The grade corresponds to **0 - 19.5** points of the total examination score.

7. ARTIFICIAL INTELLIGENCE (AI)-TOOLS AND PLAGIARISM

- Use of AI-tools is allowed for purposes of knowledge acquisition and practice but not to produce material for any form of the examination.

- All forms of plagiarism are prohibited and these include even texts generated by AI-tools.

- Use of AI-tools is allowed for purposes for improving a self-written original text is not allowed

- Software for text-matching and detectors for AI-generated text are used by the department. In the link below, you can read more about Stockholm University's guidelines for disciplinary matters (the English version can be added if/when the need arises):

  https://medarbetare.su.se/vart-su/styrning/regelboken/utbildning/regler-och-handlaggningsordning-for-disciplinarenden