LUND UNIVERSITY

Graduate School

FACULTY OF SOCIAL SCIENCES

GRADUATE SCHOOL METHODS COURSES

SIMM61

Quantitative Data Analysis in R

Version 1.0 – September 2021

1. WELCOME

Contact info

Graduate School
e-mail: master@sam.lu.se
Home page: graduateschool.sam.lu.se
Facebook: tinyurl.com/LUgradschoolFB

Student Union
Home page: samvetet.org

Lund University
Home page: http://lunduniversity.lu.se
The university is on Youtube, Facebook and Twitter

Welcome to the Autumn term's course *Quantitative Data Analysis in R.*

The aim of this course is for the student to develop an understanding of key concepts and principles guiding the use of quantitative methods, relate the use of quantitative methods to social science theory building and assessment, acquire practical skills with regard to the performance of statistical analysis and visualisation in R, and develop the ability to independently and critically assess quantitative research.

The student formulates a research question that includes a hypothesised relationship in relation to social scientific theories on a particular theme, and that can be addressed using an available dataset. During the course different techniques for processing, visualising and analysing data in R will be introduced and the student works on answering their own research question using the tools presented to them in the lectures, seminars and computer labs. The student also learns to assimilate and evaluate existing quantitative social science research as it is presented in scientific journals and/or reports.

Moreover, some of the multivariate statistical techniques most commonly used within the social sciences are presented and practiced in R. The focus lies on the relationship between research questions and different multivariate statistical techniques. The teaching includes theory and practice in analytical methods.

Formal learning outcomes for the course

Upon completion of the course, the student shall be able to:

Knowledge and Understanding

- demonstrate an understanding of concepts and principles associated with quantitative methods;
- demonstrate an understanding of the role of quantitative methods in the social sciences;
- demonstrate knowledge of multivariate statistical techniques frequently used within the social sciences;
- demonstrate an understanding of the kinds of research questions that each technique can be used to address;
- demonstrate knowledge on how to conduct applied quantitative analysis in relation to social scientific theories on a particular theme.

Competence and skills

- independently and with proficiency, be able to formulate and examine a hypothesis relevant for quantitative forms of analysis;
- independently and with proficiency, be able to operationalise theoretical concepts at different levels of analysis for quantitative analysis;
- independently and with proficiency, select an appropriate method, interpret the outcome and report the results for a given research question;
- exemplify skills in performing an analysis using the different techniques covered in the course, including but not limited to simple, multiple and multilevel linear regression, and logistic regression;
- independently and with proficiency, demonstrate a working knowledge of the R programming language.

Judgement and approach

- understand and reflect on texts (reports or scientific papers) where the argument is based on statistical analysis in a knowledgeable, independent and theoretically informed way;
- critically and independently reflect on theoretical and methodological aspects of such analysis;
- be able to independently and critically reflect on the relationship between research questions and statistical techniques;
- be able to independently and critically reflect on, and make informed decisions with regard to, core methodological issues in the context of the application of the statistical techniques taught in the course.

Assessment

Overview

The assessment of the course is based on an individual paper (7.5 credits) that includes a literature review in relation to social scientific theories on a particular theme and an applied analysis in R. In addition, multivariate statistical techniques are examined separately in lab reports (7.5 credits). The concepts on which the lab reports are based are introduced in conjunction with the respective lecture.

Grades

Marking scale: Fail, E, D, C, B, A.

The grade for a non-passing result is Fail. The student's performance is assessed with reference to the learning outcomes of the course. For the grade of E the student must show acceptable results. For the grade of D the student must show satisfactory results. For the grade of C the student must show good results. For the grade of B the student must show very good results. For the grade of A the student must show excellent results. For the grade of Fail the student must have shown unacceptable results.

At the start of the course, students are informed about the learning outcomes stated in the syllabus and about the grading scale and how it is applied on the course.

Grading guidelines for Individual paper and Lab reports

The individual paper and each lab report will be graded separately according to the guidelines below.

The overall course grade consists of the average grade of the individual paper – which will account for 50% of the course grade – and the lab reports. For a grade of Pass on the entire course, the student must have been awarded at least E on all assessments for which the grading scale A–E+Fail applies.

If any of the lab reports gets a failed grade, the final grade is a fail for the whole Lab reports segment of the course, and the reports need to be revised and re-submitted at a re-exam deadline. If all lab reports get a passing grade separately, the grades of the are averaged, producing a final grade for the lab reports segment of the course. As stated above, if the stage 1 submission warrants a passing grade then the grade will be based on the stage 2 submission. If the stage 1 submission gets a failed grade, the submission 2 will not be evaluated at this submission (but it can be used in a re-submission later).

For lab reports:

F – The report does not satisfy the criteria to get an E, and/or the report is missing crucial required elements that are required according to the lab assignment description.

E – The student understands the basic fundaments of the analysis and can perform the analysis using a statistical software. The report includes a correct presentation of at least one analysis related to the assignment. The dataset has been checked for errors, and eventual problems have been corrected. Where this is required in the assignment, the assumptions of the statistical test have been addressed.

D – Everything in E. The student has a more extended knowledge of the analysis, and can correctly perform analyses using appropriate methods and strategies, and interpret the results. The report includes a more advanced analysis than E. The results are correctly discussed.

C – Everything in D. The student has a more thorough understanding of the analyses, suitable tables and or graphs (if appropriate) are provided for correct interpretation of the results, and the syntax is shared in a public repository (e.g. GitHub). The report includes comprehensive analyses that are presented clearly and concisely.

B – Everything in C. The student has a deeper understanding of the analyses, is able to perform correct analyses and report this in a clear way. The report is free from errors regarding the interpretation of models and discussion of results. The report shows a good understanding of the analyses.

A – Everything in B. The student has the ability to do and report the analyses in line with published research reports. The report is formulated in an outstanding way.

Re-examination opportunities

The course includes opportunities for assessment at a first examination, a re-sit close to the first examination and a second re-sit for courses that have ended during that school year. Two further re-examinations on the same course content are offered within a year of the end of the course. After this, further reexamination opportunities are offered but in accordance with the current course syllabus.

Cheating and Plagiarism

The analyses and lab reports should be performed individually. It is not allowed to write assignments together with someone else, or to copy any part of the individual paper or lab report or the analysis code of anyone else, not even structural elements. If it is obvious from the assignments, for example because there are exactly the same errors, or exactly the same sentences in the text, then all with similar reports will fail. Make sure that none of your classmates read your text or drafts before submitting it. It is never permissible to include text directly from other sources without clearly specifying which sections were copied and from where (i.e., plagiarism and citing material without crediting the author is not permissible), and if this is done the assignment will fail and the plagiarism will be reported.

Please see appendix I in this guide for more information.

Surveys and Survey Results

Surveys are an important part of course management, as we base future course discussions on the results. The Graduate School Board (including all student representatives) are able to see all survey reports and survey results will also be visible on the course Canvas page once published. But everything in the end hinges on you – please do take the time to answer the survey when it is sent out so we get solid response rates!

Your teachers

Shai Mulinari (course coordinator) is Associate Professor and Senior Lecturer at the Department of Sociology. He undertakes research on the use and regulation of new biomedical technologies, health and healthcare inequalities, and pharmaceutical industry practices, regulation and transparency, using both quantitative and qualitative methods. He is the Graduate School Methods Director.

Shai Mulinari
(course coordinator)
shai.mulinari@soc.lu.se

Anna Kallos is a PhD candidate at the Department of Sociology. Her research concerns part-time work by teenagers in full-time education. Drawing on both quantitative data and interviews, she investigates the labor market participation, work experiences and employment conditions of upper secondary school students in Sweden.

Anna Kallos
anna.kallos@soc.lu.se

Zoltan Kekecs is an assistant professor at ELTE in Hungary at the Institute of Psychology, and a part time Associate Professor at Lund University at the Institute of Psychology. His research is related to the application of psychological techniques in medical contexts, and to meta-science. He is a methodologist and member of the Data and Methods Committee of the Psychological Science Accelerator, a network of over 400 research labs aiming to conduct large, cross-cultural psychological research.

Zoltan Kekecs
zoltan.kekecs@psy.lu.se

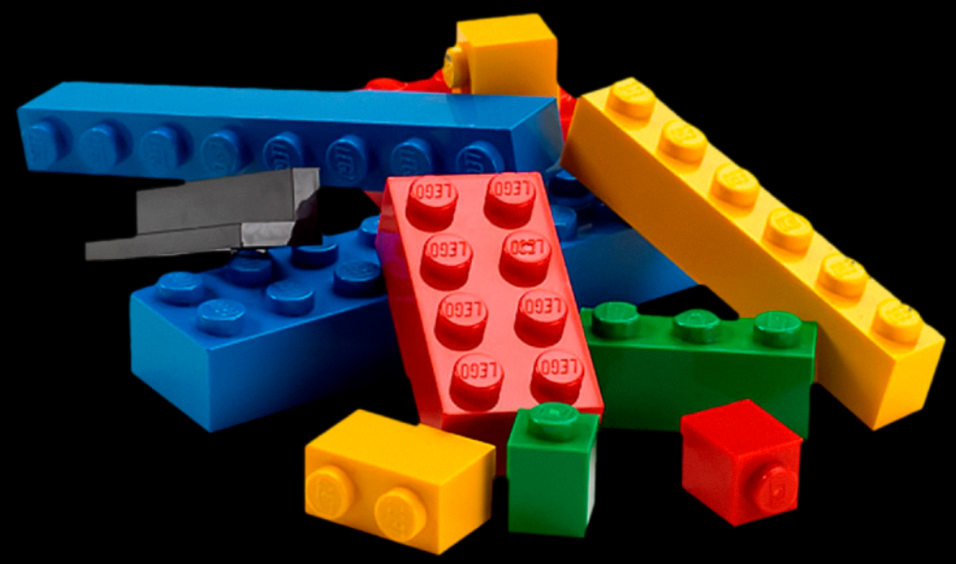
Joost van de Weijer works as a methodologist at Lund University Humanities Lab. As part of this employment, he assists students and researchers in planning and implementing experiments and analyzing the results. He teaches an introductory and a follow-up course in the statistical analysis of experimental data using R.

Joost van de Weijer
joost.van_de_weijer@humlab.lu.se

COURSE RESOURCES

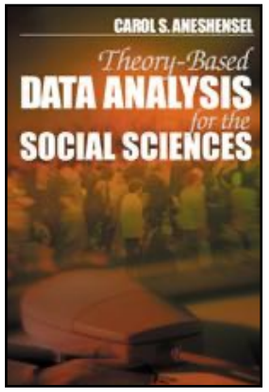
In this section we present the course literature and other course resources. This section is to help you to orient yourself in different types of readings and their functions in the course.

If download links fail, articles will be locatable via LUBSearch



Aneshensel, C. S. (2012). *Theory-based data analysis for the social sciences.* SAGE Publications, Inc.

From the blurb: The advent of complex and powerful computer-generated statistical models has greatly eroded the former prominence of social theory in data analysis, replacing it with an emphasis on statistical technique. To correct this trend, Carol S Aneshensel presents a method for bringing data analysis and statistical technique into line with theory. She approaches this task by first providing an overview that explains the connection between data analysis, statistical technique and theory. This section includes a description of the elaboration model for analyzing the empirical association between two variables by adding a `third variable' to the analysis. The author then introduces a new concept into this model, the focal relationship. This concept is the one cause-and-effect type of relationship of primary significance that is indispensable to the entire theory. Building upon the focal relationship as the cornerstone for all subsequent analysis, two analytic strategies are developed to establish its internal validity: - An exclusionary strategy to eliminate alternative explanations for the focal relationship using control and other independent variables to rule out spuriousness and redundancy, respectively; and, - An inclusive strategy to demonstrate that the focal relationship fits within an interconnected set of relationships predicted by theory using antecedent, intervening and consequent variables. Using real examples of social research, the author demonstrates the use of this approach for two common forms of analysis, multiple linear regression and logistic regression. Whether learning data analysis for the first time or adding new techniques to your repertoire, this book provides an excellent basis for theory-based data analysis.



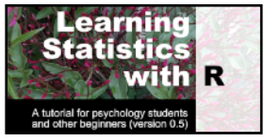
472 pages

ISBN 9781412986342

Publisher info link

Navarro, D. (2019). *Learning statistics with r: A tutorial for psychology students and other beginners: version 0.6.1.*

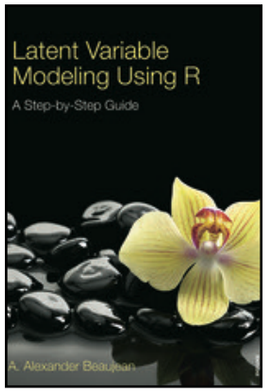
From the blurb: *Learning Statistics with R* covers the contents of an introductory statistics class, as typically taught to undergraduate psychology students, focusing on the use of the R statistical software. The book discusses how to get started in R as well as giving an introduction to data manipulation and writing scripts. From a statistical perspective, the book discusses descriptive statistics and graphing first, followed by chapters on probability theory, sampling and estimation, and null hypothesis testing. After introducing the theory, the book covers the analysis of contingency tables, t-tests, ANOVAs and regression. Bayesian statistics are covered at the end of the book.



Available here

Beaujean, A. A. (2014). *Latent variable modeling using R. A step by step guide.* New York: Routledge.

From the blurb: This step-by-step guide is written for R and latent variable model (LVM) novices. Utilizing a path model approach and focusing on the lavaan package, this book is designed to help readers quickly understand LVMs and their analysis in R. The author reviews the reasoning behind the syntax selected and provides examples that demonstrate how to analyze data for a variety of LVMs. Featuring examples applicable to psychology, education, business, and other social and health sciences, minimal text is devoted to theoretical underpinnings. The material is presented without the use of matrix algebra. As a whole the book prepares readers to write about and interpret LVM results they obtain in R.



218 pages

ISBN 9781315869780

Publisher info link

Poldrack, R. A. (2020) *Statistical Thinking for the 21st Century.*

From the blurb: The goal of this book is to the tell the story of statistics as it is used today by researchers around the world. It's a different story than the one told in most introductory statistics books, which focus on teaching how to use a set of tools to achieve very specific goals. This book focuses on understanding the basic ideas of statistical thinking — a systematic way of thinking about how we describe the world and use data make decisions and predictions, all in the context of the inherent uncertainty that exists in the real world. It also brings to bear current methods that have only become feasible in light of the amazing increases in computational power that have happened in the last few decades. Analyses that would have taken years in the 1950's can now be completed in a few seconds on a standard laptop computer, and this power unleashes the ability to use computer simulation to ask questions in new and powerful ways.

The book is also written in the wake of the reproducibility crisis that has engulfed many areas of science since 2010. One of the important roots of this crisis is found in the way that statistical hypothesis testing has been used (and abused) by researchers (as I detail in the final chapter of the book), and this ties directly back to statistical education. Thus, a goal of the book is to highlight the ways in which current statistical methods may be problematic, and to suggest alternatives.

FORTHCOMING

298 pages

Available here

Course Resources – Book Chapters

1. Fox, J. & Weisberg, Sanford. (2019). *An R companion to applied regression.* (Third edition). Thousand Oaks, California: SAGE Publications, Inc., **Chapters 6-8**. ISBN: 9781544336473
2. Revelle, W. (2009). *An introduction to psychometric theory with applications in R.* **Chapter 6**. Available online at: https://www.personality-project.org/r/book/Chapter6.pdf
3. Wickham H and Grolemund G. *R for Data Science.* **Chapters 1-5**. Available online at: https://r4ds.had.co.nz

Course Resources – Recommended Reading

Dalpiaz, D. (2021). *Applied Statistics with R.* Chapter 17. Logistic regression. Available online at http://daviddalpiaz.github.io/appliedstats/

Clark, M. (2021). *Mixed Models with R.* Available online at https://m-clark.github.io/mixed-models-with-R/

Dudek, B. (2019). *Confirmatory Factor Analysis with R.* Available online: https://shiny.rit.albany.edu/stat/cfa1test/

Kabacoff, R. (2011). *R in Action.* Chapter 14. Principal components and factor analysis. Shelter Island, NY, USA: Manning publications.

Kaplan, David (2017). *Statistical modelling: A fresh approach.* Available online at https://dtkaplan.github.io/SM2-bookdown/

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R.* Springer Science & Business Media

Course Resources – Resources

In this course, we use R for statistical analysis. To participate in the course, you will need to use R and RStudio, which is an advanced and practical user interface for R.

There are many great resources available online that can help you learn and use R.

DataCamp

One great place where you can get started is DataCamp. DataCamp is an online learning platform that teaches coding through short videos and bite-sized interactive coding exercises. You will have access to DataCamp courses for free while you are attending our course thanks to the DataCamp for the Classroom program.

A guide to install R and RStudio:

https://www.datacamp.com/community/tutorials/installing-R-windows-mac-ubuntu

Tutorial on the basics of R:

https://learn.datacamp.com/courses/introduction-to-r

Tutorial on the basic tidyverse functions (tidyverse is a great package collection for data wrangling):

https://learn.datacamp.com/courses/introduction-to-the-tidyverse

Statistical Thinking for the 21st Century

If you would rather learn from a book, why not check out "Statistical Thinking for the 21st Century" by Poldrack. This is a great book that covers the basics of R through interesting examples. This book is freely available online.

Poldrack, R. A. (2020) *Statistical Thinking for the 21st Century.* Available online at http://statsthinking21.org/

RStudio Cloud

If you don't want to install R on your system or you want a stable cloud-based alternative, you can also use RStudio via RStudio Cloud. Aside from allowing you to run R in a browser, RStudio Cloud also lets you share project with others and collaborate with others, and it also provides a seamless Git integration.

https://rstudio.cloud/

StatQuest

Check out the videos from "StatQuest: An epic journey through statistics and machine learning"

https://statquest.org/video-index/

COURSE OVERVIEW

A detailed description of the course content, including work tasks.

Your course at a glance

TIME	COURSE ACTIVITY
Week 1	MULTIPLE LINEAR REGRESSION IN THE SOCIAL SCIENCES
	Lecture 1 | Shai Mulinari Introduction - Theory-based data analysis
	Lecture 2 | Shai Mulinari Multiple linear regression I
	Lecture 3 | Shai Mulinari Multiple linear regression II
	Lecture 4 | Shai Mulinari Multiple linear regression III
Week 2	**INTRODUCING R**
	Lecture 5 + Lab | Joost van de Weijer Introduction to R I
	Lecture 6 + Lab | Joost van de Weijer Introduction to R II
	Lecture 7 + Lab | Joost van de Weijer Introduction to R III
	Lecture 8 + Lab | Joost van de Weijer Introduction to R IV
	Lecture 9 + Lab | Joost van de Weijer Introduction to R V
Week 3	**MULTIPLE LINEAR REGRESSION LABS AND INDIVIDUAL WORK**
	Seminar + Workshop | Shai Mulinari & Anna Kallos Introducing individual paper and datasets
	Lab | Anna Kallos Multiple linear regression I
	Lab | Anna Kallos Multiple linear regression II
Week 4	**MULTIPLE LINEAR REGRESSION LABS AND INDIVIDUAL WORK**
	Lab | Anna Kallos Multiple linear regression III
	Lab | Anna Kallos Multiple linear regression IV
NB. Regularly check the course lesson plan online for potential schedule alterations and to locate relevant classrooms	

TIME	COURSE ACTIVITY
Week 5	INDIVIDUAL WORK ON PAPER AND HAND-IN
Week 6	**MORE ADVANCED MULTIVARIATE TECHNIQUES**
	Lecture 10 + Demonstration | Zoltan Kekecs Logistic regression and special predictors
	Lab + Consultation | Zoltan Kekecs Logistic regression and special predictors
	Lecture 11 + Demonstration | Zoltan Kekecs Mixed effects models I
Week 7	**MORE ADVANCED MULTIVARIATE TECHNIQUES**
	Lecture 12 + Demonstration | Zoltan Kekecs Mixed effects models II
	Lab + Consultation | Zoltan Kekecs Mixed effects models
	Lecture 13 + Demonstration | Zoltan Kekecs Dimension reduction techniques
Week 8	**MORE ADVANCED MULTIVARIATE TECHNIQUES**
	Lab + Consultation | Zoltan Kekecs Dimension reduction techniques
	WINTER BREAK
Week 9	**MORE ADVANCED MULTIVARIATE TECHNIQUES**
	Lecture 14 + Demonstration | Zoltan Kekecs Structural equation modeling (online)
	Lab + Consultation | Zoltan Kekecs Structural equation modeling (online)
Week 10	**FINALISING THE COURSE**
	Extra **Lab + Consultation** | Zoltan Kekecs All topics (online)
Jan 14th 2022 16.00	**Deadline**: Hand in lab reports
Feb 11th 2022 16.00	**Deadline**: First re-examination
Aug 19th 2022 16.00	**Deadline**: Second re-examination
NB. Regularly check the course lesson plan online for potential schedule alterations and to locate relevant classrooms	

Course details

Lecture 1: Introduction: Theory-based data analysis

(lecture) | *Teacher:* Shai Mulinari

Introduction to the course. A key concept of the course, the "focal relationship", is covered. We learn about the principles of quantitative analysis and statistical inference.

Primary reading

Aneshensel, Carol S (2012). Chapters 1-3

Lecture 2: Multiple linear regression I

(lecture) | *Teacher:* Shai Mulinari

We learn how to construct and interpret a multiple regression analysis, following Aneshensel's elaboration model.

Primary reading

Aneshensel, Carol S (2012). Chapters 4-6

Lecture 3: Multiple linear regression II

(lecture) | *Teacher:* Shai Mulinari

We learn more about how to evaluate and interpret spuriousness, redundancy and mediation, based on Aneshensel's elaboration model.

Primary reading

Aneshensel, Carol S (2012). Chapters 7-9

Lecture 4: Multiple linear regression III

(lecture) | *Teacher:* Shai Mulinari

We learn about antecedent and consequent variables and about conditional relationships, based on Aneshensel's elaboration model.

Primary reading

Aneshensel, Carol S (2012). Chapters 10-11 + 13

Lecture 5 with Lab: Introduction to R I

(lecture & lab) | *Teacher:* Joost van de Weijer

Topics: R-studio working environment: Scripts. Structure of commands. Help function. Plot. Installing packages, introducing tidyverse. Customizing the environment. Arithmetic functions with R, including exponentiation and logarithm. Scientific notation. Rounding.

Primary reading

Navarro, D (2019). Chapter 3

Lecture 6 with Lab: Introduction to R II

(lecture & lab) | *Teacher:* Joost van de Weijer

Topics: Variable types, and variable type conversion/coercion. Useful functions: mean, range, standard deviation, variance, IQR, length, unique, seq, table. Data import from external sources, including reading Internet data, local text files, Excel worksheets.

Primary reading

Navarro, D (2019). Chapter 4;

Wickham, H and Grolemund G (2017). Chapter 1

Lecture 7 with Lab: Introduction to R III

(lecture & lab) | *Teacher:* Joost van de Weijer

Topics: Dealing with missing values. Logical operators and logical vectors. Making selections from data.

Primary reading

Navarro, D (2019). Chapter 3;

Wickham, H and Grolemund G (2017). Chapter 5

Lecture 8 with Lab: Introduction to R IV

(lecture & lab) | *Teacher:* Joost van de Weijer

Topics: Plotting: boxplot, scatterplot, line chart, histogram, barchart. ggplot2 package. Conditional plots.

Primary reading

Navarro, D (2019). Chapter 6;

Wickham, H and Grolemund G (2017). Chapter 3

Lecture 9 with Lab: Introduction to R V

(lecture & lab) | *Teacher:* Joost van de Weijer

Topics: Frequency tables, introducing the linear model.

Primary reading

Navarro, D (2019). Chapter 15

Seminar and Workshop: Introducing individual paper and datasets

(seminar & workshop) | *Teacher:* Shai Mulinari & Anna Kallos

Students formulate the research question (involving a focal relationship) that they will answer through their work on their individual paper. Datasets are presented.

What happens if you fail to attend this event?

What to hand in: A one page summary of your project including the following information: Preliminary title; project idea and justification; preliminary hypotheses; description of datasets and variables; sketch of analysis plan.

How to hand in: Canvas

When to hand in: 48 hours after the seminar

Lab: Multiple linear regression I

(lab) | *Teacher:* Anna Kallos

We learn how to construct and interpret a multiple regression analysis in R, following Aneshensel's elaboration model.

Primary reading

Poldrack, R. A. (2020).

Lab: Multiple linear regression II

(lab) | *Teacher:* Anna Kallos

We learn how to construct and interpret a multiple regression analysis in R, following Aneshensel's elaboration model.

Primary reading

Poldrack, R. A. (2020).

Lab: Multiple linear regression III

(lab) | *Teacher:* Anna Kallos

We learn how to construct and interpret a multiple regression analysis in R, following Aneshensel's elaboration model.

Primary reading

Poldrack, R. A. (2020).

Lab: Multiple linear regression IV

(lab) | *Teacher:* Anna Kallos

We learn how to construct and interpret a multiple regression analysis in R, following Aneshensel's elaboration model.

Primary reading

Poldrack, R. A. (2020).

Lecture 10 with Demonstration: Logistic regression and special predictors

(lecture & demo) | *Teacher:* Zoltan Kekecs

We learn how to build more complex regression models in R, including how to model non-linear relationships and interactions between predictors. We also learn how to predict binary outcomes with a regression model (logistic regression). Model diagnostics of these models will also be covered.

Primary reading

Fox, J & Weisberg, S (2019). Chapters 6 and 8.

Also recommended: Dalpiaz, D (2021). Chapter 17.

Lab and Consultation: Logistic regression and special predictors

(lab & consultation) | *Teacher:* Zoltan Kekecs

We will practice the above.

Lecture 11 with Demonstration: Mixed effects models I

(lecture & demo) | *Teacher:* Zoltan Kekecs

We learn how the linear regression framework can be applied to situations where data is clustered and when we only have observations from a limited number of the clusters out of the many clusters existing (e.g. the observations are clustered in classrooms, schools, districts, or municipalities). This and the subsequent class covers important concepts and terminologies related to linear mixed models (a.k.a. multilevel models) and how to formulate these models using R. We also learn how to make a decision between multiple plausible models. Model diagnostics of mixed models will also be covered.

Primary reading

Fox, J & Weisberg, S (2019). Chapters 7 and 8

Also recommended: Clark, M (2021).

Lecture 12 with Demonstration: Mixed effects models II

(lecture & demo) | *Teacher:* Zoltan Kekecs

See above.

Primary reading

Fox, J & Weisberg, S (2019). Chapters 7 and 8

Also recommended: Clark, M (2021).

Lab and Consultation: Mixed effects models

(lab & consultation) | *Teacher:* Zoltan Kekecs

We will practice the above.

Lecture 13 with Demonstration: Dimension reduction techniques (Principal Component Analysis and Exploratory Factor Analysis)

(lecture & demo) | *Teacher:* Zoltan Kekecs

What if we have so many variables that prevent us from building sensible models? How can we explore the underlying variables governing multiple different types of behaviors or economic or social phenomenon? Dimension reduction techniques are designed to solve these issues. Principal component analysis is used to create sensible groups from correlated variables while factor analysis is designed to explore the latent (not observed) factors that are responsible for the correlation of multiple correlated variables. We learn how to use these techniques, how to implement these models in R, and what are the limitations involved in using these techniques. We will also learn how to check their assumptions.

Primary reading

Revelle, W. (2009). Chapter 6.

Also recommended: Kabacoff, R. (2011) Chapter 14

Lab and Consultation: Dimension reduction techniques

(lab & consultation) | *Teacher:* Zoltan Kekecs

We will practice the above.

Lecture 14 with Demonstration: Structural equation modeling (online)

(lecture & demo) | *Teacher:* Zoltan Kekecs

This class will introduce the basics of SEM. SEM is a statistical framework that gives the researcher more freedom about the structure of the models they propose. SEM models don't just estimate a single dependent value from the predictors. Rather, in SEM you can model a system of relationships. This allows us to test theory-driven models of observable phenomenon. In this introductory class we will focus on the fundamental concepts underlying SEM models, and build some basic SEM models that demonstrate some of the unique utility that SEM provides.

Primary reading

Beaujean, A. A. (2014). Chapters 2-3.

Dudek, B. (2019).

Lab and Consultation: Structural equation modeling (online)

(lab & consultation) | *Teacher:* Zoltan Kekecs

We will practice the above.

Extra lab and Consultation: All topics (online)

(lab & consultation) | *Teacher:* Zoltan Kekecs

During this final lab and consultation session students will have a chance to ask any remaining questions and fine-tune and finalize lab reports for stage 2 submission. This session is not topic specific, so students can ask questions about any of the topics covered in the course.

Assignment Instructions

1. Individual paper (7.5 credits)

Detailed instructions for the individual paper will be provided during the course on Canvas. The datasets are introduced during week 3. The paper shall include a literature review in relation to social scientific theories on a particular theme selected by the student and an applied multivariate regression analysis in R based on Aneshensel's elaboration model connected to the selected theme, including relevant R code with annotations.

Use Times New Roman 12 pt font, 1.5 line spacing, with 2.5 cm left and right margins. Use APA formatting guidelines where possible, including in the reporting of statistical results (e.g. report $p < .001$ instead of $p = 0.0000231$). You are encouraged to present correctly labelled figures and tables.

The paper is to be submitted via Canvas on **December 3rd.**

2. Lab reports (7.5 credits)

Other multivariate statistical techniques are examined separately in lab reports. The concepts on which the lab reports are based are introduced in conjunction with the respective lecture.

Submission guidelines for the Lab reports

There are four lab reports, one for each of the following topics: logistic regression, mixed models, dimension reduction, and structural equation modeling. These lab reports should be completed according to the instructions published separately for each of these lab assignments on Canvas. The instructions listed below apply to all reports.

All lab reports are submitted via Canvas. A single document should be submitted, containing all lab reports. Upload the R code you used during solving the lab assignment to GitHub in a public repository. Each lab's R code should be uploaded to GitHub as a separate file! (If the R code for multiple reports is included in a single file on GitHub, it cannot be sent out for code co-pilot review, so you will not get a feedback from others). The submitted lab report document needs to include a link to the GitHub repository containing the analysis codes.

Lab reports are submitted in two stages.

In the stage 1 submission, the lab report is submitted, including a link to the repository containing the analysis codes.

The day after the deadline for stage 1 submission, the R codes are distributed among the students for peer review. Each student will get a set of analysis codes that they are asked to review. We call this code co-pilot.

As a code co-pilot you should review the codes you get and give feedback about the code based on certain criteria, such as: did the code run on your system, did you find any coding errors, is the code properly annotated with comments so that it is possible with someone who knows the goal of the project to follow what is being done in the code, did you find any mistakes in the analytical approach etc. (The feedback should not contain extensive code sections, mostly just text about the errors/mistakes and suggestions on how to correct them.)

The feedback from the code co-pilots will be sent to the original authors of the codes. Students can use this feedback to revise their original submissions (both the code and the lab report) for the stage 2 submission. (Only those who submitted code co-pilot feedback will get the feedback on their own code written up by the other students. Those who do not submit the code co-pilot feedback by the deadline will not get feedback on their own code either.) The lab reports at stage 2 need to be written up and submitted **with tracked changes**, so that it is apparent what was changed compared to the stage 1 submission. Any update to the R code from stage 1 to stage 2 should be made on the same file on GitHub that is linked in the stage 1 submission (instead of uploading a separate file). GitHub tracks changes automatically as long as you **do not delete the file.**

Importantly, the final grade of the lab report is derived in the following way: if the stage 1 submission warrants a passing grade then the grade will be based on the stage 2 submission. If the stage 1 submission gets a failed grade, the submission 2 will not be evaluated at this submission (but it can be used in a re-submission later). So even though there is room for improvement in the stage 2 submission, **the stage 1 submission should be a complete work which can be evaluated on its own (not a draft)**, because a passing grade can only be achieved if both stage 1 and stage 2 submissions warrant a passing grade.

Style guidelines for the Lab reports

The number of pages for the reports is restricted to two A4 pages per lab report included in the document. This means that if the document contains all four lab reports, you can write 8 pages in total. The title page, tables, figures, references, and appendices are not included in the page limit, so the final document is likely to be longer than 8 pages in the end. Use Times New Roman 12 pt font, 1.5 line spacing, with 2.5 cm left and right margins.

For everything else, use APA formatting guidelines where possible, including in the reporting of statistical results (e.g. report $p < .001$ instead of $p = 0.0000231$). You are encouraged to present correctly labelled figures and tables.

The main text should include a very brief introduction, a comprehensive result section (this can be separated into a methods and a results section), and a short discussion section. You don't have to include every table or figure obtained in the statistical software when doing your analysis, only those that are relevant for your report. Report everything that is listed in the 'What to report' sections in the lab assignment instructions for the specific report, and use the 'What to discuss' sections for guidance about what should be included in the discussion section.

APPENDIX I

ACADEMIC WRITING AND PLAGIARISM

Academic honesty

Academic honesty means that you as an author are responsible for your work and that you must be able to support the statements you make. Likewise, citation and referencing must be done correctly and it is never allowed to copy, fabricate or manipulate your data. This means that everything you hand in has to be made and written by you and nobody else. If that is not the case you can be accused of plagiarism, a serious offence. The penalties for plagiarism at LU are for example suspension between 2 weeks and 6 months.

Plagiarism – and how to avoid it

If you copy, paraphrase or translate materials from websites, or library or other sources in your written assignments or thesis without giving full and proper credit to the original author(s), you are committing plagiarism. Accusations concerning plagiarism are taken very seriously and the consequences for your academic career and professional future may be disastrous, involving not only the loss of credit for courses in which the offence occurred, but even suspension for a certain time from your degree programme, not to mention having to live with a lingering reputation for dishonesty. Submitting the work of others as if it were your own is unacceptable. Plagiarism must be understood and avoided at all costs.

Students should expect to have their papers checked for plagiarism electronically. Whenever you use the words or ideas of others, fair academic practice requires that you identify your sources fully and accurately. Simply mentioning an author's work at the beginning of a paper does not mean that you are then free to copy or paraphrase from that work; specific references must be given each time you quote or paraphrase. The fair use of evidence from primary and secondary sources is the basis of academic discourse, and abuse of this fairness undermines the very nature of scholarly research. Although plagiarism is not always illegal (since copyright laws usually presume a financial motive), it is nevertheless a form of intellectual theft and fraud. By committing plagiarism you show disrespect for the fundamental values of the academic community.

If you find yourself in doubt about quotations or your use of sources, it is always a good idea to provide full information.

To learn more about LU policy about Academic honesty visit LUB's page on Academic conduct: libguides.lub.lu.se/mastersprogrammes/academicwriting

Tech system note

Urkund is an automated plagiarism control system used throughout the university. It is integrated in Canvas, and will warn you if its pattern-matching algorithms has been detected something suspect (warnings will appear in Canvas when you prepare to download student assignment texts.

APPENDIX II

PROCESSING STUDENT COMPLAINTS

It is actually relatively rare, but it does happen that students complain about what happens in a course to the point when it is hard to know what to do. The Faculty has set up a common process for these occasions, so both students and teachers know the options. In this appendix we present the faculty guidelines in full.

Processing of complaints from students concerning first and second cycle education at the Faculty of Social Sciences

The present document describes the processing of education-related complaints from students at the Faculty of Social Sciences.

Before students proceed with a complaint, they should find out what rules apply in various situations. Students' rights and obligations at Lund University (LU) are described in the List of students' rights (see link below). For example, the list describes what applies to the study environment, course syllabi and timetables, exams and assessment, degree projects and course evaluation. Another important document that governs education is the relevant course syllabus. It is also possible to obtain information by contacting the study advisor at the department.

Students with a complaint can primarily turn to the relevant lecturer/course director or to the programme director. In many cases the problem can be solved closest to where it arose. For further processing of a complaint, please see the flow chart below.

At LU there is a student representative to whom students with a complaint can turn for support and help. The student representative is not part of the University administration, but an independent party whose role is to support and guide the students' unions and the students in their case. The students can also obtain support and advice from the Social Sciences Students' Union. Support from the student representative or the Social Sciences Students' Union does not require membership in the students' union.

The flow chart below aims to clarify the work flow and contact people in cases of student complaints at the Faculty of Social Sciences. The fundamental principle is that a case is to be processed promptly, documented and registered according to the usual procedures. All student complaints that become cases are to be registered at LU (official document).

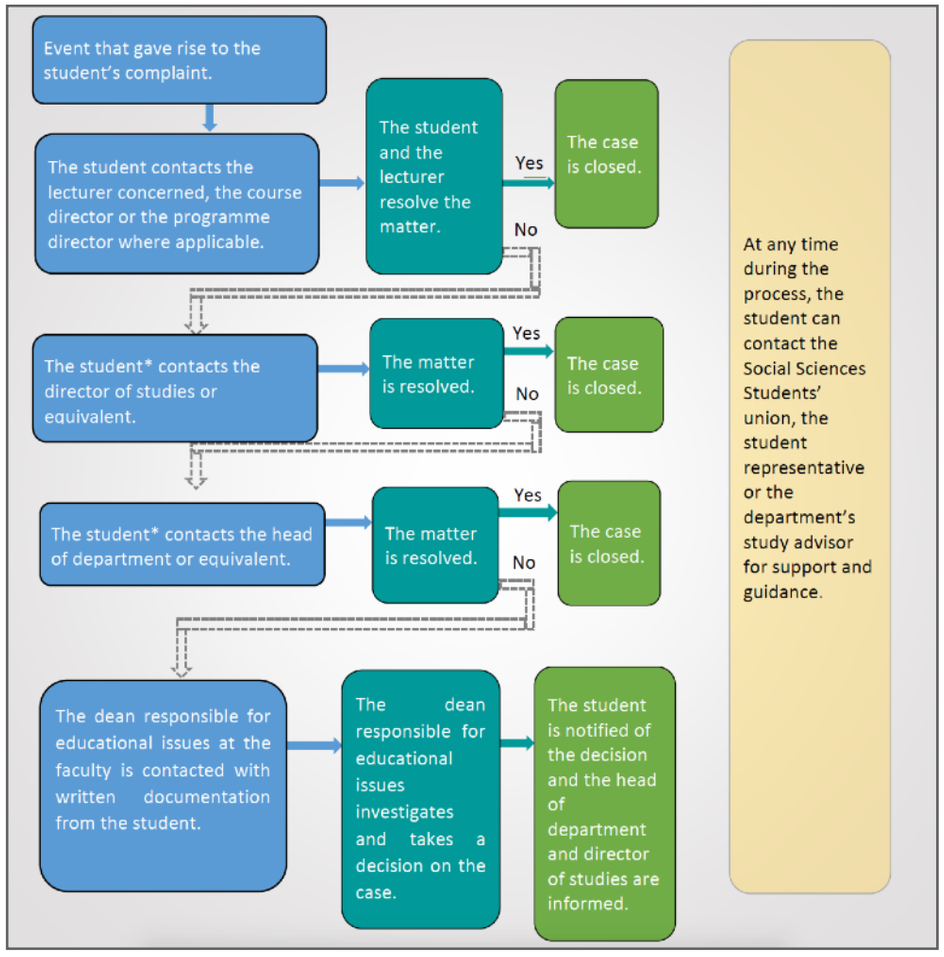
The description of the procedure does not prevent a student from appealing a decision pursuant to Chapter 12 of the Higher Education Ordinance (see below) or reporting LU to the Swedish Higher Education Authority. At LU, it is also possible to turn directly to the vice-chancellor according to guidelines approved on 12 March 2015 (see link below).

The procedure description/flow chart does *not* cover:

- Cases dealing with discrimination or harassment (pursuant to the Discrimination Act 2008:567 and the Work Environment Act 1977:1160). Information on where to turn for these issues is available separately (see link below).
- Cases that concern Chapter 12 of the Higher Education Ordinance: assessment of qualifications and admission, approved leave from studies, deferred entry, credit transfer of previous studies, requests for exemption from study components and applications for degree certificates. If the decision on such matters goes against the applicant, he or she can apply to the Higher Education Appeals Board. Information on how to do this is to be attached to the decisions.
- Disciplinary matters, that are to be processed by the vice-chancellor/disciplinary board (pursuant to Chapter 10 Section 3 of the Higher Education Ordinance).
- Changes to grading decisions (pursuant to information approved on 2 December 2015, see link below).

The present document is to be published on each department's website and information about the document should be disseminated to new students at the Faculty of Social Sciences in connection with course/programme introductions. The document was produced in collaboration with the Social Sciences Students' Union.

Processing of students' complaints at the Faculty of Social Sciences



* The lecturer or the director of studies concerned can also choose to take unresolved issues to the next level.

Relevant links

List of rights for students at Lund University
www.lunduniversity.lu.se/sites/www.lunduniversity.lu.se/files/list-of-rights-lund-university.pdf

Guidelines on handling complaints from students concerning first, second and third cycle studies at Lund University (LU central document regulating these matters). Document approved on 12 March 2015.

www.staff.lu.se/sites/staff.lu.se/files/guidelines-on-handling-complaints-from-students-concerning-first-second-and-third-cycle-studies-at-lund-university.pdf

How to process cases of discrimination or harassment
www.staff.lu.se/employment/work-environment-and-health/health-and-wellness/victimisation-and-harassment

Changes to grading decisions (official document approved on 2 December 2015).
sam.lu.se/internt/sites/sam.lu.se.internt/files/information_om_andring_av_betyg_-_2015-12-02.pdf

APPENDIX III

GRADUATE SCHOOL: A BRIEF HISTORY

An innovative organisational solution to the problem of managing and exploring interdisciplinarity is now a teenager, and an established part of the Faculty of Social Sciences.

A brief history

Graduate School's story began with a push for internationalisation at Lund University prompted primarily by Sweden's adoption of the *Bologna Process* regulations. In 2004, Sweden began the process of reforming the preexisting higher education structure to follow a common European model. The Bologna Process inspired a number of new developments here at the Faculty of Social Sciences. The Faculty Leadership sought to create two-year Master's programmes in accordance with Bologna regulations as well as creating international programmes and courses on the faculty level, and it was decided that the Faculty of Social Sciences should create international master programmes at the faculty level. There already were two international master programmes in existence at the faculty – Welfare Policies and Management and International Development and Management, but those belonged to the Political Science and Human Geography departments respectively. Coordinating master programmes at the faculty level was something that had not been done before.

An advisory board comprised of representatives, usually Directors of Study from nearly every subject at the faculty, was assembled to decide which subject areas should be chosen to become international programmes and courses that might best serve the needs and interests of Social Sciences students. The response to the proposed additions was positive, particularly from departments with lower student rates. A common, faculty level master programme could be more cost effective to run than one at a single department and could even offer courses in theory and method to not only its own programme students but also to students in smaller master programmes elsewhere within the faculty, thereby allowing departments to offer a wider variety of programmes to students.

Developing Interdisciplinarity

While the intention for the programmes to be international was a primary focus from the start, the interdisciplinary aspect of the proposed programmes came later.

The advisory board discussed the issue of how to create a faculty-wide, interdisciplinary master programme at length and decided that such programmes should be theory-based, designed to focus on a major – a primary field of study within the programme subject – and also require applicants to meet the eligibility requirements for their major. Fulfilling major requirements in one field on the bachelor's and subsequently the master's level would then allow a graduate to have the possibility to continue to a PhD.

11 different programme topics were suggested and of those, three were ultimately selected and are still the backbone of Graduate School today: the MSc Programmes in *Development Studies, Global Studies*, and *Social Studies of Gender*. These would be led by a Director of Studies with individual Programme Directors for each of the three programmes and a board made up of the departments participating in the interdisciplinary cooperation. Once the subject areas were decided upon, the advisory board for deciding upon faculty-level international master's education became the steering committee for the three new programmes. Among those in that committee was Kjell Nilsson, who became the first Director of Studies of Graduate School. Franz-Mikael Rundquist would become the Programme Director for Development Studies, Catarina Kinnvall the Director for Global Studies, and Sara Goodman the Director for Social Studies of Gender.

The name "Graduate School" was decided upon, with the intention that the name should communicate its offerings to international students, and to indicate that international master level programmes and courses as well as a few international PhD courses were available there.

Graduate School welcomed its first programme students in the Autumn of 2007. Located in the Eden building, Graduate School was made up of its Director of Studies Kjell Nilsson, two administrative staff, and 9 students in Social Studies of Gender, 26 students in Global Studies, and 23 students in Development Studies.

Although the general opinion towards the newly created international, interdisciplinary programmes and courses was enthusiastic, some at the faculty were still unsure about the idea of international programmes, particularly with regards to having to teach courses in English. Initially, Graduate School sought to incentivise potentially reluctant teachers to lecture on its courses by offering them a few more teaching hours, but as time went by Graduate School was able to find more and more teachers who simply enjoyed working with international students and teaching in English.

Director of Studies Kjell Nilsson's ability to network within the faculty, garner support for and subsequently structure three unique, ambitious interdisciplinary master programmes helped to bring the concept of Graduate School to life. He and the steering committee set the stage for the next level of development for the organisation. In this period, Kristina Jönsson became the new programme director for Development Studies.

In September 2010, Lena Örnberg took the reins as Graduate School Director of Studies. The numbers of programme students had decreased since the programmes' first year, which led to some criticism as to the perceived success of the interdisciplinary programmes. Lena sought to improve both the student experience as well as numbers of students in the programmes by placing emphasis on student events and administrative structure. Teaching and administrative staff would have increased contact, such as at teaching team wrap-up meetings at the end of courses, to create more cohesion between the two groups and to relieve teaching staff of unnecessary administrative tasks. The number of students began to grow and an additional third full time administrative position was added.

Finding (and Creating) a Physical Home

It was at this time that Graduate School moved from the Eden building to Gamla Kirurgen. There the programme would have its own classrooms and study area, separate from other departments. This fostered a feeling of "home" and a sense of belonging among Graduate School students. Events like programme introduction day, potlucks, fika, and information lunches that include both students and staff bring class cohorts together and familiarise them with staff, so students know who to turn to when in need of support.

Seeking to further improve structure and processes, the Graduate School team traveled to the University of Amsterdam in Spring 2011 to meet with colleagues there working with their interdisciplinary Master Programme in International Development Studies. While comparing programme structure and administrative processes with their Amsterdam colleagues, the Graduate School team were somewhat surprised (and pleased) to discover that their Dutch counterparts were impressed by Graduate School's thoroughness in interdisciplinarity. The difference was that the interdisciplinary focus was not limited to the makeup of the student body or the teachers – even the courses were interdisciplinary, down to mixed, interdisciplinary teaching teams on a single course. University of Amsterdam staff thought mixing teaching teams was incredibly ambitious and would not be possible at their university. Lena later remarked that this difference was a testament to the efforts made by the original steering committee that made a truly interdisciplinary Graduate School possible. This practice of interdisciplinary teaching teams continues at Graduate School today and is seen as a strength by staff and students alike.

A Maturing Organisation

By the time Lena left her post as Director of Studies in late 2014, student numbers had risen dramatically and a place in a Graduate School programme became highly sought after by international students. Around that time Lena left, programme directors Kristina Jönsson (Development Studies) and Sara Goodman (Social Studies of Gender) stepped down from their posts. Karin Steen took over for Development Studies and Rebecca Selberg took over for Social Studies of Gender. In 2017, Rebecca stepped down and the role has now been taken on by Marta Kolankiewicz.

After Lena's departure, the remaining admin team members successfully managed programme admissions until Mikael Sundström was installed as the new Director of Studies in the spring of 2015. Since then, Graduate School has looked for complementing ways to develop, further increasing its reach by way of communications material and processes and improved overall quality of courses, particularly methods courses. Programme and course guides and the very handbook you are reading now have been designed, reworked and reformulated to provide comprehensive information with a unique, signature style. Students are kept up to date with a bi-weekly *Newsflash* email with an overview of upcoming important Graduate School information as well as interesting events and activities around the faculty and the university.

In the last five years we have also been placing extra focus on our theory and methods courses offerings. A *Methods Director* position (currently held by Shai Mulinari after a productive stint by our current programme director Chris Swader) has been introduced to keep track of and develop the various courses in theory of science and methods. The aim is to further develop the quality, design, and variety of the method courses that are offered to Graduate School students as well as many other master and PhD students. In addition, we have set about documenting all available theory and method courses at the Faculty of Social Sciences, providing a clearer overall picture of the state of theory and method courses at the faculty.

A New Growth Period

In 2018 two momentous decisions were rendered. First, Graduate School would become the new home of the *Middle Eastern Studies* programme from 2019, with Rola El-Husseini as the designated Programme Director.

Second, Graduate School was to develop a brand new master programme, labelled *MSc in Social Scientific Data Analysis (SSDA)*, slated to start in 2021. Chris Swader is the designated Programme Director for the SSDA.

When these developments have concluded, Graduate School will have grown from 180 full-time student equivalents (*Helårsstudent*, HÅS) to 280!

Graduate School – Our House!

Graduate School is housed in what is now known as "the old surgery clinic" (Gamla Kirurgen). Our two lecture halls (236 & 240) used to be ten-bed wards with an observation room (238) and pantry (237) sandwiched in-between. From the observation room, nurses could keep a watchful eye on recovering patients through two windows that have since been removed. The Student Lounge still has a vaguely religious look to it, and was indeed used as a church room in the past.

In 1868, the house we now inhabit finally opened for business as Lund's main open surgery clinic. The famous and prolific architect Helgo Zettervall designed the building's late gothic style, and although it has undergone substantial renovations in 1905, 1928 and 1978, many of his original ideas remain intact. The most notable changes in the intervening years was probably the installation of many more windows than Zettervall had opted for, and the wing extensions to increase floorspace.

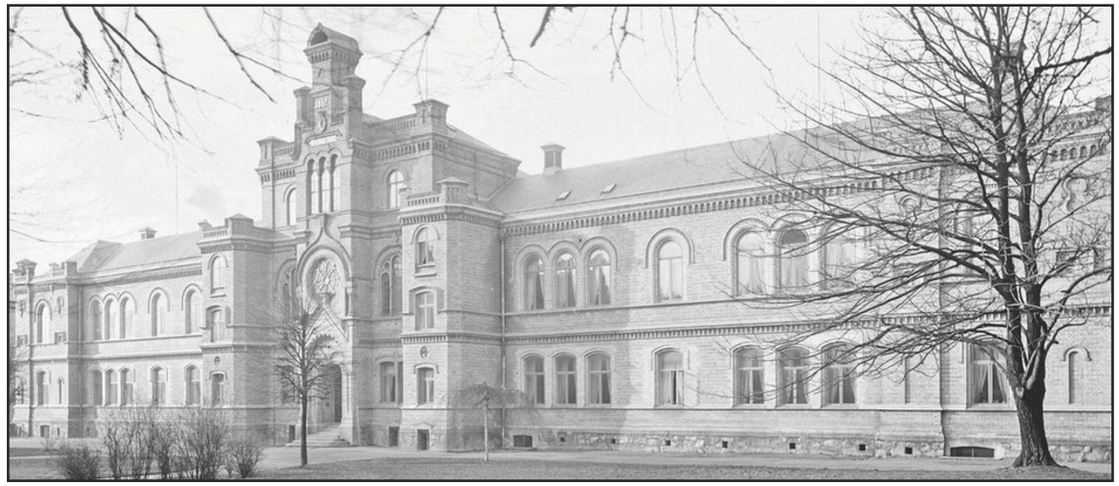
Inside, changes have been much more far-reaching. Among other things, what is now the stairwell in the third floor used to be the very heart of the building as it housed the central operation theatre.

When the hospital moved to its current location in the 1970s, the old buildings were transferred to Lund University which urgently needed more space. The open surgery clinic itself was handed over in 1972, and was at that point listed as an architectural heritage structure to prevent potentially intrusive changes (this status was removed in 2005).

Helgo Zettervall (1831–1907)

Renowned architect who designed the open surgery clinic along with many other buildings around Lund, including the main university building

Over the years, the building has housed a range of University units, notably the "UB3" University Library branch on the top floor. Today it is predominantly a social science building, with the central Faculty Administration, the International Office, Graduate School and the School of Journalism as main anchors. The 150-year old is still going strong!



Picture of the surgical clinic by Per Bagge in 1906. Reproduction: University Library, Lund University.

Important!

We ask all students to take note of the emergency exits (see below), and to be prepared to guide students to the assembly area at the *front* of the building in case of a fire or other hazard demanding evacuation.