

Statistical Models in Political Analysis

Marco Steenbergen & Benjamin Schlegel

Fall 2022

Instructors: Marco Steenbergen / Benjamin Schlegel
E-Mail: steenbergen@ipz.uzh.ch / kontakt@benjaminschlegel.ch
Vorlesungszeit: Mo. 10:15 - 12:00
Place: AFL-F-121 and Zoom

Important: Please register for the course in the module booking tool. Otherwise we will not be able to give you credit points for the course.

Course Links

- [Course Website](#)
- [Zoom](#)

Office Hours/R-Helpdesk

Benjamin Schlegel: Wednesday, 14:30 – 16:30 or by appointment (H-343)

Marco Steenbergen: Appointment via Microsoft Teams

Course Content and Goals

In this course we look at statistical models and inference. After a recap of linear regression, we will cover multilevel models, binary

dependent variables, ordinal and nominal dependent variables, limited dependent variables, count models, and event duration models. These are among the most widely used models in the social sciences. A major emphasis of the course is on interpretation, visualization, and implementation using R.

By the end of the course, you will have accomplished the following goals:

1. Understand the most important models for social-scientific research and when to use them.
2. Improve your R programming skills.
3. Be able to read and understand papers using the models discussed in the course.
4. Apply and interpret the models in the context of your own research.

Course Format

This course will be conducted slightly differently from the usual format. In lieu of lectures, the course is divided into Q&A and lab sessions. All of the materials that would regularly be taught in a lecture are presented in a course textbook. The Q&A sessions, which will be conducted via Zoom, are meant to address any questions that may arise from the book chapters, including questions about the exercises in the book. As such they require that you prepare questions, although we shall also rely a bit on the Socratic teaching method. The Q&A sessions will be conducted by Marco Steenbergen.

The lab sessions are meant to show you how to estimate, evaluate, and interpret statistical models in R. While each book chapter contains examples of R code, more opportunities for practicing R will be offered in the lab sessions. These will be conducted by Benjamin Schlegel. Unlike the Q&A sessions, the lab sessions take place in-class, in AFL-F-121.

Important:

1. It is absolutely ESSENTIAL that you read the book. You neglect this advice at your own risk.
2. The Zoom sessions take place on October 3, October 17, October 31, November 14, and November 28.

3. The in-class sessions take place on September 19, September 26, October 10, October 24, November 7, November 21, December 5, December 12, and December 19.

Prerequisites

Basic knowledge of R and statistics is required. This includes basic knowledge of descriptive statistics, data visualization, probability theory, and hypothesis testing, as well as skills in opening and manipulating data in R.

Course Materials

Textbook

Steenbergen, Marco R., and Benjamin Schlegel. 2022. *Statistical Modeling for Social Scientists: An Introduction Using R*. Zurich: Department of Political Science.

Software

In this course we will work with the free and open source statistical program R. Because of its flexibility, versatility, and graphical capabilities, R is becoming the standard programming language for political science and is also increasingly used in the private sector. R is available for all major operating systems.

Course Requirements

The course requires a portfolio consisting of:

1. Homework: 3 assignments contribute 40% to the final grade.
2. Exam: A final exam contributes 30% of the final grade
3. Data Essay: A data essay contributes 30% of the final grade.

Homework

The homework assignments are very close to the examples in the textbook and are meant to give you an opportunity to practice your R skills and interpretation. These assignments should be absolved in groups of 2 students (no exceptions). The deadlines are:

- Homework 1: Start on October 24, 2022 12:00 (Deadline: November 3, 2022 08:00 (in the morning!))
- Homework 2: Start on November 2, 2022 12:00 (Deadline: December 1, 2022 08:00 (in the morning!))
- Homework 3: Start on December 5, 2022 12:00 (Deadline: December 15, 2022 08:00 (in the morning!))

Exam

The exam is online and open-book and will be conducted on **January 16, 10.15-11.45**. It assesses some of the same skills as the homework assignments, the difference being that the exam is to be completed individually. Additionally, the exam covers theory and interpretation.

Data Essay

The data essay is an opportunity for you to analyze a data set of your choice. We shall make several data sets available for this purpose. For a particular data set, your task is to select the appropriate modeling technique, to implement the model in R, and to interpret the results. The due date for the data essay is on **January 3, 2023, at 08:00** (in the morning).

Grading

Each assignment has a score. An average score of 60% is needed to pass the course. Note that you may not skip parts of the portfolio and compensate them with other parts.

Schedule

Week 1—September 19—Course Introduction

A discussion of the course, its contents, logistics, and requirements. Instructors: Steenbergen & Schlegel. Location: AFL-F-121.

Week 2—September 26—LAB: Reviewing R

Refreshing your memory cannot hurt, so today we shall be reviewing the fundamentals of R. Instructor: Schlegel. Location: AFL-F-121.

Week 3—October 3—Q&A: Modeling and Regression

Today's Q&A covers the idea of (statistical) modeling, the classical normal linear regression model, maximum likelihood, likelihood ratio and Wald tests, as well as model comparison and evaluation. Instructor: Steenbergen. Location: [Zoom](#).

Readings: Steenbergen & Schlegel, Chapters 1-2, as well as the preface and notation.

Week 4—October 10—LAB: Regression

How does one estimate a linear regression model in R? How does one obtain various regression diagnostics? You will learn all of this today. Instructor: Schlegel. Location: AFL-F-121.

Week 5—October 17—Q&A: Hierarchical Linear Modeling

So far, we have assumed that the observations are conditionally independent. In many contexts, this is not realistic. For example, if citizens live in the same country, then their cultural heritage could induce correlations that cannot be fully captured through predictor variables. Rather than viewing this as a violation of an assumption, we can also view it as an opportunity to incorporate contextual effects into our models. Hierarchical linear models allow for these contextual effects to be modeled. Today's Q&A is all about hierarchical linear modeling. Instructor: Steenbergen. Location: [Zoom](#).

Readings: Steenbergen & Schlegel, Chapter 3.

Week 6—October 24—LAB: Hierarchical Linear Modeling

Today, there will be plenty of opportunity to estimate hierarchical linear models using R. Instructor: Schlegel. Location: AFL-F-121.

Week 7—October 31—Q&A: Binary Dependent Variables

Up until now, the outcome variables have been continuous and unbounded (at least in theory). Many outcome variables in the social sciences, however, are binary in nature. Think about the states of war and peace in conflict research, for example. These kinds of outcome variables need special models such as binary logit and probit analysis. In today's Q&A, we discuss these models and their interpretation. Instructor: Steenbergen. Location: [Zoom](#).

Readings: Steenbergen & Schlegel, Chapter 4.

Week 8—November 7—LAB: Binary Dependent Variables

Today's lab will give you the opportunity to practice estimating, visualizing, and interpreting binary logit and probit models. Instructor: Schlegel. Location: AFL-F-121.

Week 9—November 14—Q&A: Ordinal and Nominal Dependent Variables

Binary dependent variables are examples of categorical outcome variables. There are many other categorical outcome variables, however. When more than two outcomes are possible, we can rely on ordinal logit and multinomial logit models for ordered and unordered categories, respectively. Today's Q&A explores these models and their intricacies. Instructor: Steenbergen. Location: [Zoom](#).

Readings: Steenbergen & Schlegel, Chapters 5-6.

Week 10—November 21—LAB: Ordinal and Nominal Dependent Variables

Today's lab is all about estimating, visualizing, and interpreting ordinal and multinomial logit models. Instructor: Schlegel. Location: AFL-F-121.

Week 11—November 28—Q&A: Limited Dependent Variables and Event Counts

Sometimes, an outcome variable is continuous on principle, but we observe only part of the distribution. We call such a variable a limited dependent variable. At other times, the outcome variable is a count, for instance, the number of tweets sent by a political party in a particular month. Today's Q&A concerns models for limited dependent variables and event counts, including Heckman selection, truncated regression, Tobit, Poisson, and negative binomial models. Instructor: Steenbergen. Location: [Zoom](#).

Readings: Steenbergen & Schlegel, Chapters 7-8.

Week 12—December 5—LAB: Limited Dependent Variables and Event Counts

In today's lab, you will have the opportunity to estimate different models for limited dependent variables and event counts. Instructor: Schlegel. Location: AFL-F-121.

Week 13—December 12—Q&A: Event Duration Models

Social scientists are often interested in time, specifically the amount of time that expires before some event takes place. An example is the amount of time that expires before a government collapses. As the name implies, event duration models take durations as their outcome variables and try to understand why events occur sooner in some instances and later in others. In today's Q&A we discuss several of those models. Instructor: Steenbergen. Location: AFL-F-121.

Readings: Steenbergen & Schlegel, Chapter 9.

Week 14—December 19—LAB: Event Duration Models

In today's lab, we practice the estimation and interpretation of both parametric and semi-parametric event duration models. Instructor: Schlegel. Location: AFL-F-121.