EH6105: Quantitative Methods

Steven V. Miller

Fall 2022

E-mail: steven.miller@ekohist.su.se
Office Hours: *on student request*
Office: Södra Huset A 916

Web: eh6105.svmiller.com
Class Hours: *check schedule*
Class Room: Södra Huset E 347 (Lab)

Course Description

The aim of the course is to provide the students with basic knowledge about quantitative methods in the social sciences, though the topical focus will privilege applications in economic history and international relations. The course will cover both descriptive statistics (i.e. summarizing and visualizing information) and inferential statistics (i.e. making statements about a population based on a sample of it). The major methodological focus will be ordinary least squares (OLS) regression, the basic workhorse of applied quantitative methods, though the course will take care to discuss matters of quantitative research design more broadly. Upon completion of the course, students should be able to read, interpret and assess quantitative research at a fundamental level. The class will serve as foundation training for a student's ability to design and perform a quantitative study, e.g., write a master thesis using quantitative methods.

Learning Outcomes

1. Read and critically evaluate research in economic history and international relations where quantitative methods are used.
2. Formulate statistical models and test hypotheses
3. Solve basic problems by reference to probability theory, correlation analysis, and regression.
4. Use statistical software and present results from analyses of source material.
5. Demonstrate an awareness of ethical aspects of research.

Course Policies

This section of the syllabus details multiple policies that will be implemented in this class through the semester.

Required Readings

Books

Gujarati, Damodar. 2014. *Econometrics by Example*. 2nd ed. Bloomsbury Academic.

Kellstedt, Paul M., and Guy D. Whitten. 2013. *The Fundamentals of Political Science Research*. 2nd ed. New York, NY: Cambridge University Press.

Both books are available for purchase at Akademibokhandeln and can also be borrowed from the library.

Articles

Be advised that additional articles may be assigned for individual lectures. Where applicable, this information will be communicated under Planeringar in Athena. The student is responsible for regularly checking Athena for updates. All articles should be available from the university library. Students who are unfamiliar with how to use the library, but know how to Google things, may find the following link (on my blog) to be useful.

- Set Up Google Scholar to Find Class Readings on Your Syllabus

Other Resources

Quantitative methods is a unique subset of social science pedagogy where you could truly teach yourself if you had a working internet connection. Unlike other topics/fields in the discipline, social scientists who teach students about research design and methodology are apt to share their notes, lectures slides, and discuss stuff on public social media (e.g. Twitter, Mastodon). Coding-related inquiries are even more abundant on a place like Stack Overflow. You could truly "just Google" your way through a lot of this material if you were diligent and receptive to consuming as much information as you can find.

Students may also find some of my previous teaching material to be useful. I have plenty of material on my blog (`svmiller.com/blog`) from which I may draw inspiration this semester. Lecture slides and code for my undergraduate methods class and my graduate-level methods class from my previous employer are also online as well.[1] Lecture slides and lab sessions will almost assuredly borrow, quite liberally, from these past courses.

Lectures and Computer Labs

This course consists of nine lectures and five computer labs. The nine lectures, for which attendance is assumed but not formally required, are scattered across different dates, times, and rooms. The computer labs, for which attendance is *mandatory*, are all in Södra Huset E 347 and at the same time of day. Each lab day will have two groups: Group A and Group B. Group A's lab session is from 9 a.m. to 11 a.m. each lab day. Group B's lab session is from 11 a.m. to 1 p.m. each lab day. Students may select into either group, though the instructor's preference is to treat Group A as a beginner's group and Group B as a more advanced group. The latter group will ideally proceed through a lab session at a quicker pace. The instructor will inform students of their group selection before the first lab session.

Software for Computer Labs

The lab sessions themselves will take place in rooms with computers in them for the student to use, though there is a (reasonable, implicit) assumption that the student has a personal computer. Tablets are not advised for these purposes as it is difficult to install the required third-party software needed for this course. Tablets will also typically lack the kind of memory and processing power for computational uses like this.

Lab sessions and problem sets (more in the next section) will all be done in the R programming language. Students should download this free software programming language at cran.r-project.org and install it on their personal computer. Binaries are available for Windows and Mac (even Linux, if that is the weapon of choice for the student).

[1] See `posc3410.svmiller.com` and `post8000.svmiller.com` for more.

- The R scripts I provide are designed to work on the student's computer with minimal maintenance. I will make this clear in each particular script.
- I *strongly* encourage students to contact me to learn about the language. I will assume that not discussing R with me means the student is fluent with the software.
- Consider getting a graphical user interface (GUI) front-end for R to learn it better. I recommend RStudio, available for free at posit.co. Do note there is a paid option of Rstudio that you *do **not** want.* The paid version is for servers. You want the basic open source integrated development environment (IDE). *This is **free***. Do not pay for anything related to R or Rstudio since you do not need whatever product is available for purchase.

I published a beginner's guide to using R in 2014 when I first started to teach courses that forced students to use the R programming language. I have since streamlined the R requirements for this class, making that guide somewhat dated. You will need to install the following packages, which I illustrate here with the R commands to install them.

```
install.packages("tidyverse")       # for most things workflow
# ^ This is a huge installation. It should take a while.
install.packages("peacesciencer")  # some peace science data
install.packages("stevedata")       # for toy data sets to use in-class
install.packages("stevemisc")       # for some helper functions
install.packages("stevetemplates") # for preparing reports
install.packages("lmtest")          # for model diagnostics
```

The aforementioned R packages are not exhaustive of all packages a student may use this semester, and I reserve the right to ask students to install additional packages along the way (though these requests will ideally be rare). I will make this clear in each lab session and problem set.

Assessment

Course assessment consists of five problems sets in the R programming language along with a written final exam. The problem sets revolve around some R scripts I will provide. The questions in these scripts will tell students to answer questions based on the output of a function in the script, or prompt students to extend the script in some manner (i.e. to induce some programming skills from the student). Problem sets must be typed in either a Word processor or, preferably, LaTeX. Each problem set is graded on a 0-8 scale, resulting in a maximum of 40 points students will earn toward their grade. There is an additional final exam, worth 60 points, that will be given at the end of the semester. This results in 100 total points for the semester.

Computer lab sessions are mandatory and students must attend one of the two sessions that day to receive credit for attendance. The instructor will take attendance by making students sign their name on attendance sheet, making it the student's responsibility to ensure their signature is next to their name by the end of the lab session in order to receive credit. Failure to attend a lab session—even under extraordinary circumstances—will require a complementary assignment to be completed *in addition to* the problem set associated with the lab session. The nature of this assignment will be at the instructor's discretion. Do note that missing more than one lab, under any circumstance, will result in the student's inability to finish the course.

Do note the nature of the problem sets and the exam preclude students from sharing answers. Such a form of cooperation will be considered plagiarism. All problem sets and the exam should be submitted through Athena

before the 8 p.m. deadline each day. Resist the urge to assume your laptop's time is the system's time, and that submission at the literal last minute is an acceptable strategy.

Grading Criteria

Table 1 provides a summary of the distribution of possible grades for the semester, along with a description of what each grade suggests in terms of the learning outcomes.

Table 1: A Description of the Grading Outcomes for the Semester

Grade	Point Range	Summary	Description
F	0-39	FAIL	The student does not reach the requirements for E.
Fx*	40-49	INSUFFICIENT	The student does not reach the requirements for E, but get the opportunity to provide supplementary work to reach E.
E	50-59	ADEQUATE	The student demonstrates a limited theoretical knowledge and ability to apply some of the expected learning outcomes. The student can formulate and test some basic statistical hypotheses, solve problems using probability theory, correlation, and regression analysis. The student also demonstrates knowledge of ethical aspects in quantitative research and show a limited command of R to present quantitative data descriptions and analyses.
D	60-69	SATISFACTORY	The student demonstrates a satisfactory theoretical knowledge and limited ability to apply several expected learning outcomes. The student can formulate and test basic statistical hypotheses, solve problems using probability theory, correlation, and regression analysis. The student also demonstrates knowledge of ethical aspects in quantitative research and show a limited command of R to present quantitative data descriptions and analyses
C	70-79	GOOD	The student demonstrates a good theoretical knowledge and ability to apply a majority expected learning outcomes. The student can formulate and test statistical hypotheses, solve problems using probability theory, correlation, and regression analysis. The student also demonstrates awareness of ethical aspects in quantitative research and can use R to present quantitative data descriptions and analyses.
B	80-89	VERY GOOD	The student demonstrates a very good theoretical knowledge and ability to apply all expected learning outcomes. The student can, in an independent manner formulate and test statistical hypotheses, solve problems using probability theory, correlation and regression analysis. The student also demonstrates good awareness of ethical aspects in quantitative research and shows good command of R to present quantitative data descriptions and analyses.

Table 1: A Description of the Grading Outcomes for the Semester *(continued)*

Grade	Point Range	Summary	Description
A	90-100	EXCELLENT	The student demonstrates an exceptionally good theoretical knowledge and ability to apply all expected learning outcomes. The student can, in a critical and independent manner formulate and test statistical hypotheses, solve problems using probability theory, correlation and regression analysis. The student also demonstrates and excellent awareness of ethical aspects in quantitative research and shows very good command of R to present quantitative data descriptions and analyses.

* A complementary assignment given by the examiner must be handed in within a week to get the grade E.

Plagiarism

Students must be conscientious about how they use the material they read and how they summarize them for in-class assignments. Simply put, the use of other people's materials without attribution is plagiarism. The translation of someone else's materials into another language, without attribution, is also plagiarism. A verbatim copy of someone else's words, with selective changes of just a few words, is also plagiarism when done without attribution. This form of academic dishonesty is considered cheating and exposes the student to a failing grade and potential disciplinary measures on top of that. Any student who is caught cheating or disrupting academic activities can be suspended from lectures and exams for a period of up to six months. The Vice-Chancellor or the Disciplinary Council decides whether the student is to be subject to any disciplinary measures. Please consult the Student Support portal at Stockholm University for more information about your rights, and your responsibilities as students.

Language

The instructor does not yet speak Swedish, but is working on it. The seminars, labs, and all course material will be in English. All assignments must be completed in English as well.

Course Evaluation

Students can evaluate the instructor and the course through a written evaluation, which will be distributed at the end of the course. The course evaluation itself is voluntary, but strongly encouraged as this important form of feedback shapes course development in the department. Student responses are 100% anonymous.

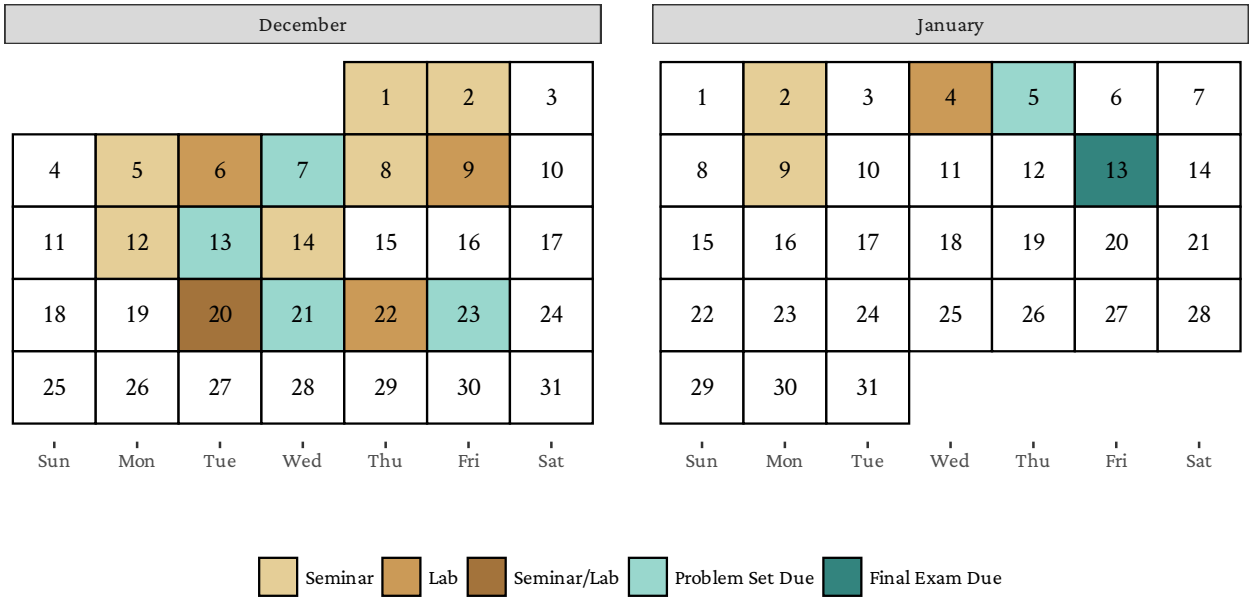

Figure 1: A Calendar for EH6105 (Quantitative Methods, Fall 2022)

Class Schedule

The following serves as the schedule for this course. Each section heading communicates the topic and, in the case of the seminars, what readings from the Kellstedt and Whitten (2013) [KW] and Gujarati (2014) [G] should be done *before* the seminar. Additional article readings will be posted to Athena in due course, when applicable. All computer labs will be in Södra Huset E 347. Information about the time and location of the seminars follow the section heading and description of the seminar. Students should have this information in TimeEdit as well. In the event of a miscommunication about the time and location of a seminar, the information on TimeEdit should be considered correct and the instructor will issue a clarification on Athena.

Seminar 1, 01/12: Introduction (KW, chp. 1)

This will be an introductory lecture introducing students to the course and its requirements.

Time and location: Södra Huset B 413, 11:00-13:00

Seminar 2, 02/12: Research Design (KW, chps. 2-4)

This lecture will discuss basic elements of a research design, including the conceptualization of topics and operationalization of concepts into variables.

Time and location: Södra Huset D 299, 12:00-14:00

Seminar 3, 05/12: Descriptive Statistics (KW, chp. 6)

This lecture will introduce students to some basic descriptive statistics, focusing on quantitative assessments of central tendencies in data and the variation in variables.

Time and location: Södra Huset B 413, 10:00-12:00

Computer Lab 1, 06/12: Introduction to R

This lab will introduce students to the R programming language and Rstudio. This will be more of an orientation than anything, though students will ideally also learn about reading data as well as some basic data manipulation.

- Group A: 09:00-11:00
- Group B: 11:00-13:00

Seminar 4, 08/12: Probability and Sampling (Chp. 7)

This is a lecture of two halves. In the first part, students will get a crash course on what is probability and some rudimentary aspects of probability theory. In the second part, students will learn about the logic of random sampling from a population in order to make inferential statements about the population in question.

Time and location: Södra Huset D 299, 11:00-13:00

Computer Lab 2, 09/12: Descriptive Statistics

This lab will go over some basics of descriptive statistics. Topics will be multiple, though the basic takeaway will implore students to first look at their data before analyzing it. Descriptive statistics is also good diagnostics.

- Group A: 09:00-11:00
- Group B: 11:00-13:00

Seminar 5, 12/12: Hypothesis Testing (KW, Chp. 8)

This lecture will dovetail the previous lecture on probability and sampling to its application in evaluating hypotheses.

Time and location: Södra Huset E 371, 12:00-14:00

Seminar 6, 14/12: Bivariate OLS (KW, Chp. 9)

Here, we introduce students to the basics of ordinary least squares (OLS) regression in the case of two variables. This may seem daunting to students to start, but students who remember the slope-intercept equation from their secondary education should see the basic intuition here.

Time and location: Södra Huset B 497, 15:00-17:00

Computer Lab 3, 20/12: OLS Regression

This lab will go over basic OLS regression, focusing on the simple bivariate case.

- Group A: 09:00-11:00
- Group B: 11:00-13:00

Seminar 7, 20/12: Extending OLS Regression (KW, Chp. 10; G, chps. 1-3)

Once students get the basics of OLS in the bivariate case, extending it becomes a somewhat trivial matter. Topics here will include multiple regression, interactions, fixed effects, and more.

Time and location: Södra Huset B 413, 15:00-17:00

Computer Lab 4, 22/12: OLS Regression 2

This lab will extend OLS regression to multiple regression, including so-called control variables, fixed effects, and more. What emerges is still the familiar "rise over run" of the slope-intercept equation, at least as students should interpret it.

- Group A: 09:00-11:00
- Group B: 11:00-13:00

Seminar 8, 02/01: Testing Assumptions (KW, chp. 11; G, chps. 4, 5, 7)

All models are valid if and only if the assumptions of the model hold. This lecture will expand on the basic assumptions of OLS and identify cases where OLS is a questionable or even invalid model for a given application.

Time and location: Södra Huset B 413, 13:00-15:00

Computer Lab 5, 04/01: Testing Assumptions

This lab will over the basic diagnostic tests of the OLS model, and what you should do if your given model fails one of those diagnostic tests.

- Group A: 09:00-11:00
- Group B: 11:00-13:00

Seminar 9, 09/01: Causality, and Its Issues (KW, chps. 3-4)

Our final lecture will talk about the deceptive difficulty that comes in trying to assess neat, causal relationships in a world governed by endogeneity.

Time and location: Södra Huset E 487, 13:00-15:00