

Econ 123:

Intermediate Data Analysis and Econometrics

Yale University, Spring 2023

Updated on: January 8, 2023

– Preliminary Syllabus and Subject to Change –

ADMINISTRATIVE

LECTURES

- Time: Tuesdays and Thursdays, 9am - 10:15am.
- Location: [WLH 117 - William L. Harkness Hall 117](#)

INSTRUCTOR

Professor Edward Vytlačil (edward.vytlacil@yale.edu)

- Office Hours
 - Time: Mondays, 3-4:30pm
 - Location: [30 Hillhouse Ave](#), Room C149

TEACHING FELLOW

Ken Jung (ken.jung@yale.edu)

Sections: TBD

Office Hours: TBD

WEBPAGES

- Main webpage at: <https://yale.instructure.com/courses/83774>
 - Quizzes, problem sets, exams, zoom links for lectures and sections, recorded lectures, and other materials will be posted to the course's webpage.
- Discussion board: [ed discussion](#)
 - Please post questions to Piazza rather than emailing myself or the TA whenever possible.

ABOUT THE COURSE

This course builds upon Econ 117: Introduction to Data Analysis and Econometrics. As with Econ 117, this course focuses on applied examples from economics, and motivates the methodology with the applications. It includes more theoretical depth than Econ 117. It also includes more advanced **R** coding, such as functional programming in **R**, and additional statistics and econometrics topics not covered in Econ 117 such as computationally intensive methods for inference (e.g., randomization inference, bootstrap inference), instrumental variable methods, maximum likelihood estimation, limited dependent variable models, and panel data models. The course will make connections to economic theory, including to economics models of market entry and discrimination. The course includes an empirical project, and is designed to prepare students to conduct academic-quality empirical research including for senior theses. Relative to a traditional econometrics course, this course will focus more on coding and more on interpreting/analyzing/presenting empirical results, though it will also include theory.

The topics covered will include:

1. Asymptotics
2. Inference, including both finite sample and asymptotic inference, with a focus on computationally intensive methods for inference such as randomization inference and the bootstrap;
3. Multiple Linear Regression, including series regression, and including machine learning approaches to model selection;
4. Selection, Endogeneity, and Instrumental Variables;
5. Maximum Likelihood Estimation and Limited Dependent Variable Models, including for binary choice models.

In most econometrics classes, mathematical methods are introduced and then applied to a few examples. This class turns that around. We will focus on substantive questions first, and then introduce mathematical methods that will help us answer them. By the end of the class, you will have acquired several concrete skills. Specifically, you will:

1. Understand the strengths and weaknesses of different methods.
2. Be able to choose appropriate methods to answer real-world questions.
3. Understand the math behind methods like linear regression and hypothesis testing.
4. Understand the intuition behind these methods.
5. Be able to apply these methods to analyze real data with a powerful statistical analysis package (R)

The methodology covered in this course is broadly applicable throughout the social sciences, and numerous applications will be discussed, though we will focus on substantive applications in economics, particularly in the following areas:

1. Finance – Asset diversification, CAPM;
2. Development Economics, including evaluating conditional cash transfer programs;
3. Education Economics and early childhood development;
4. Industrial Organization, including models of market entry and market structure;
5. Discrimination in loans, the labor market, and the use of police force, including testing between statistical discrimination and taste-based discrimination.

Pre-Requisites

The prerequisites for this course are introductory microeconomics (Econ 108, Econ 110, Econ 115, or equivalent) and Introduction to Data Analysis and Econometrics (Econ 117), or permission of instructor based on having taken equivalent courses. Familiarity with single variable calculus is also required. We recommend that you take introduction to macroeconomics (Econ 116) and intermediate microeconomics (Econ 121 or 125) before taking this course or concurrently with this course, though those courses are not required. Some basic knowledge of linear algebra and multivariate calculus would be helpful but is not required.

This year, students who have taken S&DS 220 or 230 instead of Econ 117, and who have taken all other course requirements, are allowed into the course and should be able to do well in the course. However, the material for this course is based on the presumption that the students have taken Econ 117, and, for some topics, students who have taken S&DS 220 or 230 instead of Econ 117 may find that they need to spend extra time studying material that may be new to them while being review to students who have taken Econ 117.

GRADES

Your grade will be based on the following components:

1. Online Quizzes (10%)

You will have short on-line quizzes approximately every other week. The quizzes will be posted on the [course webpage](#) on Fridays, will remain live for 48 hours once posted, and you will have one hour to complete the quiz once you start it. It is expected that you will not need the full hour to complete the quizzes. Quizzes will primarily focus on (simple) theoretical questions while having some questions related to coding in R. The quizzes are open book/open notes, but you cannot collaborate with other students on the quizzes or discuss the quizzes with other students until after their solutions are posted. The lowest quiz score will be dropped. My tentative plan is for there to be a total of 7 quizzes: 4 quizzes before the midterm, and an additional 3 quizzes after the midterm.

2. Problem Sets (25%)

Your problem sets will be primarily empirical and based on academic research papers, with you coding in R to analyze real data, though they will also have theoretical questions and sometimes will involve simulations rather than real data.

You may work in groups of up to four people on the problem sets, but you must turn in your own individual assignment, and must indicate on your submission the other members of your group. If you don't have a group, but would like us to help match you to other students, please let the instructor or the TA know.

The problem sets will be posted to, and must be submitted to, the [course webpage](#). You are required to use the same [problem set submission procedure](#) as was used in Econ 117. Due dates and times are **strict**, any problem set submitted late will not be accepted without a note from your residential dean emailed to me before the due date. The teaching assistant will grade a randomly selected subset of the problems from each assignment, check for

completion of the other parts of the assignment, and post complete solutions after the problem sets are due. Your problem set with the lowest grade will be dropped.

My tentative plan is for there to be a total of 6 problem sets: 4 problem sets before the midterm, and an additional 2 problem sets after the midterm.

3. Midterm (20%) Exam

You will have an in-class midterm, date TBD. The midterm will focus on theoretical questions and interpreting empirical results, while also having some questions related to use of **R**. Midterm will be open book/open notes.

4. Final (25%) Exam

There will be a final exam. The final exam will be open book/open notes. The final will focus on theoretical questions and interpreting empirical results, while also having some questions related to use of **R**. The final exam will on Tuesday, May 9, 2023 at 2pm.

5. Empirical Project (20%)

There is a required empirical project for the course. This project is your opportunity to use the tools you learn to answer a question you come up with and that you care about. You may work in groups of up to four students. Your project should address a research question of your choosing, and apply the methods from this course to a relevant data set. There are two deliverables for the project, a detailed research proposal and then a final term paper. The detailed research proposal is due on April 4. The final term paper should be approximately 15 to 25 pages, including all figures and tables, and is due on May 10. More detailed guidelines for the empirical project can be found [here](#), and suggested data sources are [here](#).

Errors in Grading

If students believe that there has been a mistake in their grading, the student must prepare a written statement describing in detail the mistake, which then should be emailed to the teaching fellow. If the student still believes a mistake has been made after hearing back from the teaching fellow, they may submit to me a written statement highlighting in detail the mistake and the response from the TF.

Changing assigned grades is extremely unlikely and reserved for clear errors made by myself and the teaching fellow. Re-grading will not be considered unless submitted in writing via email as described above.

LECTURES AND IN-CLASS LABS

Classes will alternate between lectures and in-class labs:

Lectures: Lectures will be in-person. Lecture slides will be posted on the [course webpage](#), but are not designed as a substitute for attending lecture.

Labs: In-class labs will be hands-on labs, where we live-code with **R** to apply methods you learned in lecture to analyze real data and answer real research questions. The labs will be designed to directly help you with your problem sets.

SOFTWARE

As with Econ 117, much of the course work in Econ 123 will involve analysis of data using **R**, an open source implementation of the object-oriented programming language **S**. It is widely used by applied statisticians and its libraries implement a wide variety of statistical and graphical techniques with applications to a range of disciplines, such as the agricultural and biological sciences, genetics, neuroscience and economics. **R** can be downloaded from <https://cran.r-project.org>.

I will incorporate chunks of **R** code into my lecture slides, and I will use **R** with you in the in-class labs. The TF will help you with **R** in sections. The on-line program documentation for **R** is excellent, and there are also many excellent and free **R** references available online, for example, *Econometrics in R* by G. Farnsworth that is available for free.

Code written for **R** can be run on many computational platforms with or without a graphical user interface. However, while **R** has a command line interface by default, there are several graphical front-ends available. A relatively new and popular GUI is offered by **RStudio**, a free and open-source integrated development environment (IDE) for **R**. It contains many useful features that we think make using and learning **R** easier and it is therefore recommended. You can download RStudio (Desktop) for free [here](#). I will use RStudio for the in-class labs.

You may find it helpful to review the [R resources page](#) from Econ 117. I also recommend [Hands-On Programming with R](#) by Garrett Golemund and [faster: Fast Lane to Learning R!](#) by Norm Matloff.

It is hard if not impossible to learn to code without actually practicing the coding. I recommend working through [Project 1](#) and [Project 2](#) of [Hands-On Programming with R](#) by Garrett Golemund. I strongly recommend doing so if you are new to coding in **R**.

TEXTBOOKS

There is no required textbook for this course.

While there is no required textbook

ACCEPTABLE USE POLICY

You are free to use any published materials (e.g., a textbook), in preparing Econ 123 assignments or for learning the material more generally. Similarly, you are free to use online resources such as stackoverflow questions or R tutorials. You are also strongly encouraged to work with others in your class. This is particularly helpful for learning to program. Each person must turn in their own assignment.

The use of any solution materials prepared in a previous year for Econ 117, 123, or 131, other than materials distributed this academic year by the course faculty, is **strictly prohibited** and constitutes cheating. This includes 1) any notes, spreadsheets, or handouts distributed by me in a prior term of Econ 117, 123, or 131; and 2) any notes, solutions, or spreadsheets prepared by former students of Econ 117, 123, or 131, in either written or electronic form.

This policy means you should not solicit or use solutions to previous years' problem sets. The reason for this policy is that access to previous year's materials can create serious inequities between fellow students, and jeopardize the integrity of the academic environment. **Any potential violation of this policy will be reported.**

We take cheating and plagiarism VERY seriously. Every class you have ever take probably states that cheating will not be tolerated, but we mean it.

Cheating or plagiarism will result in a 0 on the assignment and will be reported to the department. You are welcome to work together in groups up to 4, but you are required to submit your own write-up and your own code. Please take precautions to avoid putting the Teaching Assistants or myself in a situation where we are forced to decide if two documents are "too similar". As future researchers, consultants, bankers, entrepreneurs, etc, learning to do honest work in a timely manner is more important than getting everything correct. If you are uncertain, please add proper citation. For example, if you relied heavily on a group-member's code for one part of an assignment, then you should make a footnote highlighting this fact. This may result in a slightly lower grade, but as long as proper credit is clearly given, it does not constitute cheating. The one exception to this rule is using past material from any previous version of this course.

Redistribution of Materials

Do not redistribute any of these materials without written permission.