ECON 210: Economic Statistics, Fall 2019
Lecture: TR 9:30 am – 10:45 am in SDL 110
Lab: Wednesday 9:05 am – 9:55 am (W) in LIB 140, 10:10 am – 11:00 am (V) in LIB 140

Prof. Akhil Rao
Email: akhilr@middlebury.edu
Course software: Canvas, Middfiles, STATA, PollEverywhere
Office location: Warner Hall 502A
Office hours: Tuesday 11am–noon & 3pm–4pm, Wednesday 1pm–2pm, Thursday noon–1pm, and by appointment

This syllabus will guide our class, but it is not set in stone. I will update it throughout the semester to ensure that it reflects our goals and progress.
Last updated: October 21, 2019

Executive summary:

Exams: **10/9 (lab) & 10/10 (lecture)** and **11/6 (lab) & 11/7 (lecture)** in class. No rescheduling.

Research project: **12/6** (Canvas).

Course grading: about 75% weight on exams and research project, 25% weight on assignments and quizzes.

This course moves fast — come prepared, and ask questions. A particularly helpful question to ask: "I'm not following — could you repeat that?" (It's very helpful when you're feeling generally confused but not quite sure what you're confused by, a state I often find myself in.)

Description and goals

What's this course about? Answer: "Using models to ask questions, and data to answer them."

A lot of people think about research and statistical methods as something *other* people do. The goal of this course is for everyone to develop the capability to understand research other people have done, and to conduct independent research of your own. I want everyone in the class to have the confidence and knowledge to ask important questions and to generate scientific evidence. *I strongly believe that regardless of your background, by the end of this course, you will be able to do research.*

Course Objectives:

By the end of this course, you should be able to do the following:

- Examine and describe data
- Understand the connection between sample statistics and random variables
- Understand the difference between correlation and causation

- Appreciate the central limit theorem (the most important theorem in this class!)
- Write a model to frame a question
- Estimate a multivariable linear regression
- Do hypothesis testing
- Start doing empirical research
- Use a statistical programming language (Stata or R)

Textbook and readings:

We will use two textbooks in this class, both available for free online. Both are required: I will be assigning readings and problems from them.

- OpenIntro Stats 4th edition (OS4) `https://www.openintro.org/stat/textbook.php`

  This will be our main reference, though we will not follow it exactly. I will note specific pages and sections to focus on before and after each lecture. It is the closest to a textbook for this class.

- Causal Inference: The Mixtape v1.7 (CIM) `http://scunning.com/cunningham_mixtape.pdf`

  This will be a secondary reference. primarily for Stata code examples. It will be much more helpful in ECON 211, but it will still be useful here in giving you a sense of how the tools we're learning will be applied.

- I recommend you check out Andrew Gelman's blog (`https://statmodeling.stat.columbia.edu/`). Andrew is a political scientist and statistician at Columbia who blogs quite a bit about the practice and sociology of statistics in the social sciences. Some of the content is well beyond the scope of our class, but there's a lot there which is applicable to us and approximately at our level.
- Other readings: posted as needed

When in doubt, **check the syllabus.** Email me if it doesn't answer your question(s). Please allow me 48 hours to respond, but don't hesitate to follow up if I have not replied.

Course policies

Office hours: There are four hours of scheduled office hours per week (in the beginning of the semester we will make sure that all of you can make at least one scheduled time). Please come to my office hours with questions about problem sets, exams, research, advising, etc.

Extensions: Out of fairness to all students, extensions will generally not be granted without a dean's excuse and will not be granted retroactively. In the case of extenuating circumstances, you should have your dean contact me *before* the deadline. If you miss a scheduled exam and take a make-up exam, this score will be multiplied by 50% (i.e. a score of 100 will count as a 50). For the final research project, each hour that it is late will result in a 10/24 ppt deduction in the final grade.

Emails: In the subject line **please start with "Econ 210: "** — I use email filters to prioritize among the many emails I receive each day, and including this in the subject line will ensure I see your email quickly. I recommend you use email only for questions requiring a brief response and to set up meetings outside of my office hours.[1] For anything substantive (i.e. problem sets, research, advising, etc.), please visit me at office hours. Being able to tackle complicated questions/topics in person is much easier than via email. Sometimes my inbox gets busy and I lose track of things — if I have not replied in 48 hours, please re-send your email (replying to the email with no text is sufficient). I typically do not check my email after 6pm.

Honor Code: Middlebury Honor Code is described at go/honorcode.

Classroom etiquette: I expect you to show respect for each other: show up on time, stay focused, do not use mobile devices and laptops unless approved, and do not hold side conversations. If you need to use a laptop to take notes, you must clear it with me in advance.

Grading policy: Given the sheer amount of grading, a typical problemt set will only have a select number of questions graded. These will typically be selected at random. If you would like your online quiz/exam regraded, you will need to submit a concise written statement (no longer than one page) to explain why. The entire online quiz/exam will be regraded, so there is a possibility that regrading will result in fewer points awarded. Requests for regrades must be submitted within one week of the online quiz/exam being returned. All other assignments (problem sets, in-class quizzes, reading summaries) will not be regraded because they are worth so little. Grades for these assignments are final.

Accommodations: Students who have Letters of Accommodation in this class are encouraged to contact me as early in the semester as possible to ensure that such accommodations are implemented in a timely fashion. If you are requesting extended time accommodations you must schedule your proctored exam with me *at least one week in advance*. For those without Letters of Accommodation, assistance is available to eligible students through Student Accessibility Services. Please contact Jodi Litchfield or Courtney Cioffredi, the ADA Coordinators, for more information: Courtney Cioffredi can be reached at ccioffredi@middlebury.edu or 802-443-2169 and Jodi Litchfield can be reached at litchfie@middlebury.edu or 802-443-5936. All discussions will remain confidential.

I reserve the right to modify the policies listed above.

[1] That said, I understand you may wish to pose a more substantive question over email. I may respond if the answer is easy to type out; else, I will likely ask you to come to my office.

Other notes

- I love meeting with students. But! Please make an appointment if you want to meet outside of office hours. A quick email is fine. Friday is my "research day", so I try not to schedule appointments then, and may be slow to respond.
- This course may seem daunting — we'll be covering complicated concepts, often at a very abstract level. Don't panic! With practice, repetition, and patience, it will come together. I'm here to help – don't hesitate to ask questions.
- On asking questions: sometimes a concept feels fuzzy or like it isn't quite clicking, but it's hard to frame a question precisely. Don't worry! Ask anyway. We may not get to the bottom of it right away, but we can make progress. Questions about material also tend to be correlated across students, so by asking you often provide a public good to your classmates.
- This material goes deep. I may at times limit our inquiry to keep us moving along. If you're curious, I'm happy to talk in more detail in office hours. Think of statistics as a very big onion, and this course as pulling the first few layers off — you can chop those layers up and cook with them, plenty of recipes won't need any more! — but there's a lot of onion left.
- Statistical programming proficiency is an important outcome of this class. **I strongly encourage you to begin with Stata**. CIM has many Stata code examples, and I strongly encourage you to seek them out and run them even if they aren't assigned and even when you don't fully understand the underlying math. OS4 has many R code examples, if you'd rather try your hand at that. For our purposes, R and Stata have equivalent capabilities. When necessary for assignments or exposition, or when requested by you, I will translate OS4 code from R to Stata. If you would like to use R in addition to Stata I am happy to support you, but come talk to me first. When possible, try to translate code from the other language to the one you've chosen to work with.

Course advice

The course is front-loaded, so we will be covering a lot of different concepts early on in the semester. Based on previous student experiences, here's some advice on how to best position yourself for success in the course:

- Plan ahead and do not procrastinate! Review the course outline and calendar and plan your schedule accordingly.
- Each week, dedicate days and times when you work solely on Econ 210 in a distraction free environment. For example, on your calendar, block out Tuesdays and Thursdays from 8:30-10:30 am to work on Econ 210.
- When you work on Econ 210:
  - Focus on this class and avoid multitasking (i.e. checking email, working on another course at the same time, etc . . .).
  - Use Freedom, Leechblock, SelfControl, or other apps that will block the internet. Use these apps on your phone so that you cannot check social media, email, etc.
- If you are interested in discovering more learning techniques and strategies, read up on Cal Newport. He's a successful computer science professor at Georgetown who has interviewed a number of top undergraduate students.

Problem sets, exams, paper, and participation

Problem sets: Problem sets will be posted on Canvas the week before they are due. You are expected to follow the Honor Code throughout the course. You are NOT permitted to use any materials that I or other professors have distributed to students who took ECON 210 in previous semesters; this includes solutions to previous problem sets and exams. Using previously distributed material is a violation of the honor code. I expect you to turn in your own work on all assignments and acknowledge the sources you consult and use. You are encouraged to work in study groups, discuss the problems with each other, and come to my office hours. That said, please ensure that whatever you submit in this course is your own work. Directly copying someone else's work, ideas, or answers without attribution is an Honor Code violation and results in serious consequences.

Reading summaries: Readings will be posted on Canvas the week before they are due. I expect you to read them and write half-to-one page summaries. These are very low-stakes assignments: I am trying to give you some context for what we're learning and encourage you to reflect on it a little, not to evaluate your writing abilities. Don't worry about formatting or summarizing all the details — I just want to get a sense of what you thought while reading and what stood out to you. They will be graded on a binary scale: full points if you submit on time, no points if you do not. There will be no extensions for reading summaries.

Two exams: We will have two in-class exams on **Wednesday 10/9 (lab) & Thursday 10/10 (lecture)** and **Wednesday 11/6 (lab) & Thursday 11/7 (lecture)**. There will be no make-up exams. Any absence on an exam day will result in the other exam being re-weighted to include the missed exam.

Research project: A major goal of this class is developing the ability to conduct independent research. To that end, we will have a research project wherein you ask and answer a question using statistics. This will include posing a question, clearly explaining how you will use statistics to answer it, acquiring relevant data, using the data to generate a first-pass answer to the question, and detailing what assumptions you are making in interpreting your results and what additional data you would need to strengthen your conclusions. The project will be due **Friday 12/6** on Canvas. I will have a detailed assignment sheet outlining my expectations for the research project after the first exam.

For the project, I expect you to take full advantage of the many resources on offer at Middlebury while remaining true to the Honor Code. This means I fully support you in (for example) going to the writing center and working closely with people there to hone your writing, but you must ensure that you are polishing your own ideas and not someone else's. Similarly, I encourage you to go to the library and work with the librarians to find more resources, but be sure to cite what you use and provide your own insights. *Where you use someone else's ideas, you must cite them.*[2] I highly recommend you reach out to **Ryan Clement**, the Economics reference librarian. Ryan is a wealth of knowledge about economic data and literature. His email is rclement@middlebury.edu, and his website is at go.middlebury.edu/ryan.

I am not particular on citation format — anything reasonable will do so long as you are consistent. The main purpose of the citations is so that you can be clear about where you are providing a new contribution and where you are building on the work of others, and so that I can easily find the sources you are referencing. I *strongly* encourage you to use some type of citation manager, such as the built-in manager in Word, Zotero, or BibTeX. It seems a shame to spend your scarce time and resources manually tracking and formatting citations in 2019.

[2] Interestingly, self-plagiarism is a thing — you must also cite yourself if you are using an idea you developed in another work. In general I would like you to write an original paper, but if you are working on a related assignment for another class and would like to cross-reference or otherwise combine them, come talk to me.

Grading

The problem sets, paper, exams, and participation carry the following weights:

Problem sets and reading summaries	10%
In-class quizzes	5%
Online quizzes	10%
First exam	25%
Second exam	25%
Research project	25%

Your letter grade will be based on the weighted average of these elements. The distribution of final grades will dictate the precise cut-offs of your letter grade. If I do need to scale, I will only scale 'upward' so that the scale can only help your grade.

Given the sheer amount of grading, a typical problem set will only have a select number of randomly-chosen questions graded. Your grade for a problem set will be a weighted combination of completeness and correctness, with more weight on completeness.

If you would like your online quiz/exam regraded, you will need to submit a concise written statement (no longer than one page) to explain why. The entire online quiz/exam will be regraded, so there is a possibility that regrading will result in fewer points awarded. Requests for regrades must be submitted within one week of the online quiz/exam being returned.

All other assignments (problem sets, in-class quizzes) will not be regraded because they are worth so little. Grades for these assignments are final.

Course schedule and reading list

The schedule of classes, assignments, and readings is given below. All readings are to be done **prior** to class. They are posted on Canvas. I reserve the right to make changes but will provide ample notification if any changes are made. Weeks marked "heavy" will be particularly busy, and I recommend you start working on the materials due for these classes (problem sets, assignments, and readings) in the week prior.

1. **Week 1, 9/10–9/12: Introduction to statistics and statistical programming**
   - Textbook sections: OS4 p9–15 (1.1–1.2.2), CIM p13–22 (Introduction)
   - Session 1: Course structure; logon to PollEverywhere, MiddFiles, Canvas; *PS1 assigned*
     - Reading to summarize: History of probability part 1 — Pre-history to 1600
   - Lab: Intro to statistical programming
   - Session 2: Sets, summations, and random variables; *PS1, RS1 due; PS2 assigned*
     - Reading to summarize: History of probability part 2 — 17th century France
   - Friday: Quiz 1
2. **Week 2, 9/17–9/19: Data and randomness** *(This week is heavy)*
   - Textbook sections: OS4 p16–17 (1.2.3), CIM p23–30 (up to Bayes Rule) and p38–40
   - Session 1: Probabilities and distributions (Normal distribution, expectations & means, conditioning); *PS2, RS2 due; PS3 assigned*
     - Reading to summarize: On the origin of the Normal Curve
   - Lab: Data analysis (histograms, summary stats, etc)
   - Session 2: More on distributions (variances, covariances, correlations, etc); *PS3 due; PS4 assigned*
     - Reading to summarize: On the origins of graphs in statistics, The invention of correlation
   - Friday: Quiz 2
3. **Week 3, 9/24–9/26: Simulations, LLN, and CLT** *(This week is heavy)*
   - Textbook sections: OS4 pages 82–86, 96–101
   - Session 1: Probabilities and the law of large numbers; *PS4, RS3 due*
     - Reading to summarize: History of probability — Jacob Bernoulli and the law of large numbers
   - Lab: Simulating the LLN (random number generation, forvalues, scalars/matrices, svmat)
   - Session 2: More on random variables (transformations, conditional expectations, Bayes' Rule); *PS5, RS4 due*
     - Reading to summarize: Thomas Bayes and "inverse probability"
   - Friday: Quiz 3
4. **Week 4, 10/1–10/3: Questions, models, and answers** *(This week is heavy)*
   - Textbook sections: OS4 pages 181–192, CIM pages 67–73
   - Session 1: ATEs, linear models *PS6, RS5 due*
   - Lab: Simulating the CLT (forvalues, subscripting)
   - Session 2: DAGs, causal assumptions, and open backdoors

- Friday: Quiz 4

5. **Week 5, 10/8–10/10: Exam week**
   - Textbook sections:
   - Session 1: Review; *PS7 due*
   - Lab: **Lab exam**
   - Session 2: **Lecture exam**
6. **Week 6, 10/15–10/17: Introduction to research** *(This week is heavy)*
   - Textbook sections:
   - Session 1: Standard errors and confidence intervals, hypothesis testing *PS8 due*
     - Reading to summarize: Split-Apply-Combine
   - Lab: Working with data (merging, cleaning)
   - Session 2: Regression and estimation; *PS9, RS7 due*
     - Reading to summarize: On the origin of Least-Squares, Statistics moves from the physical to social sciences
7. **Week 7, 10/22–10/24: Causality and design**
   - Textbook sections:
   - Session 1: *(Midterm recess — no class)*
   - Lab: Split-Apply-Combine workflows in Stata (bysort, gen/egen, statsby)
     - Reading to summarize: Compared with...
   - Session 2: Research design (natural and controlled experiments, assumptions); *PS10, RS8 due*
8. **Week 8, 10/29–10/31: Signal and noise** *(This week is heavy)*
   - Textbook sections:
   - Session 1: Inferential errors (Type 1 & 2, Type S & M); *PS11 due*
     - Readings to summarize: Let's stop talking about published research findings being true or false; The statsby strategy
   - Lab: Inferential errors
   - Session 2: Power, p-values, and the garden of forking paths; *PS12 due*
9. **Week 9, 11/5–11/7: Exam week**
   - Textbook sections:
   - Session 1: Review; *PS13, RS9 due*
   - Lab: **Lab exam**
   - Session 2: **Lecture exam**
10. **Week 10, 11/12–11/14: Multiple regression and bias**
    - Textbook sections:
    - Session 1: Multiple regression and omitted variable bias (open backdoors); *PS14 due*
    - Lab: Coefficient stability and data mining
      - Reading to summarize:

- Session 2: Multiple regression and bad control bias (colliders); *PS15 due*

11. **Week 11, 11/19–11/21: Heterogeneity and uncertainty**
    - Textbook sections:
    - Session 1: Modeling heterogeneity *PS16 due*
    - Lab: Principles of data visualization
    - Session 2: Modeling uncertainty
12. **Thanksgiving recess: 11/26–12/2**
13. **Week 12, 12/3–12/5: Reflections**
    - Textbook sections:
    - Session 1: Research conference; *PS17 due*
    - Lab: Research workshop
    - Session 2: Reflections on statistics

Last updated: October 21, 2019

FALL TERM, 2019

	M	TU	W	TH	F
1	9/9 **Classes Begin**	9/10	9/11	9/12	9/13
2	9/16	9/17	9/18	9/19 Clifford Symposium	9/20 Clifford Symposium **End Add Period**
3	9/23	9/24	9/25	9/26	9/27 Family Weekend
4	9/30	10/1	10/2	10/3	10/4
5	10/7	10/8	10/9	10/10	10/11 Homecoming **End Drop Period**
6	10/14	10/15	10/16	10/17	10/18 **Classes End 4:15**
7	10/21 **Mid-Term Recess**	10/22 **Mid-term Recess**	10/23 **Classes Resume**	10/24	10/25
8	10/28	10/29	10/30	10/31	11/1
9	11/4	11/5	11/6	11/7	11/8
10	11/11	11/12	11/13	11/14	11/15
11	11/18	11/19	11/20	11/21	11/22
BREAK	11/25	11/26 **Classes End 4:15**	11/27 **Thanksgiving Recess**	11/28 **Thanksgiving Recess**	11/29 **Thanksgiving Recess**
12	12/2 **Classes Resume**	12/3	12/4	12/5	12/6 **Classes End 4:15**
EXAM	12/9 **Reading Day**	12/10 **Exam Period Begins**	12/11 Final Exams	12/12 Reading Day	12/13 Final Exams through Sunday 12/15 @ 10pm

For more information visit our Web page! http://www.middlebury.edu/~learn

Office of Learning Resources • Center for Teaching, Learning, and Research • X3131