

ECON 0150 | Economic Data Analysis

The Economist's Data Analysis Skillset

Taylor Weidman | 4702 Posvar Hall | taylorjweidman@pitt.edu

Economists rely on data to build economic models of decisionmaking. This course is designed to equip you with the skills to do this analysis, interpretation, and communication of economic data, and may be the most important class you take in college. We will focus on understanding statistical tools and applying them to data, rather than on either the theoretical foundation of the tools or on a simple use of formulas.

The course follows a central arc, beginning with a toolset for understanding data, building a highly flexible model for inference, culminating in a capstone project that you will communicate to your peers. We will emphasize practical applications through a systematic approach to exploratory analysis, statistical modeling, and communication of the data. Each part builds progressively toward a final project, where you will address an economic question using the skills developed in the course. We will consider applications in a variety of areas not strictly limited to economics, although economic data will be the main focus. The course is proceeds in six parts.

Part 1: Exploring Variables

Making decisions with data involves understanding what the data says. Most data cannot be understood simply by looking at a spreadsheet. Data is easiest to understand when visualized appropriately, and the appropriate visualization depends on the data type. Part 1 introduces approaches for summarizing data with figures and tables using python and spreadsheets.

Topics include Python, Excel, visualizing and summarizing nominal categorical variables, ordinal categorical variables, continuous variables, and data transformations.

Part 2: Exploring Relationships

Not only do we care about single variables, but we also care about the relationships between variables. Like single variables, the relationships between variables is easiest to understand when visualized. And like summary EDA, the most effective data visualization depends on the variable types. Part 2 introduces approaches for exploring the relationships between variables using python and spreadsheets.

Topics include relationships between categorical variables, between continuous variables, and relationships between mixed types.

Part 3: Univariate General Linear Models

But how do we know if these patterns represent real relationships or just random noise? To determine whether there is a true underlying relationship, we use deviations from means to build intuition about sampling variation and the Central Limit Theorem. This lets us test our observed patterns using the most basic form of the General Linear Model.

Topics include random variables, sampling, normal distribution, confidence intervals, hypothesis testing, and correlation.

Part 4: Bivariate General Linear Model

In Part 3, we used regression with a single predictor to test whether patterns in our data were real or just noise. But economic relationships are rarely so simple. The General Linear Model not only lets us test simple hypotheses, but it allows us to test relationships between variables.

Topics include simple regression with numerical and categorical predictors.

Part 5: Multivariate General Linear Model

Economic variables don't always play by simple rules. We often have relationships that differ by group or depend on multiple variables. We extend the core ideas of the General Linear Model in Part 3 and Part 4 to handle more complex cases by including control variables, and learn how to build the right model for each situation.

Topics include fixed effects, control variables, interactions, and model selection.

Part 6: Communicating with Data

Parts 1 - 5 have gone from data to description to statistical model. But insights from data are only useful if others can understand them. In Part 6 we focus on developing clear narratives about economic relationships, choosing the right visualizations for our findings, and presenting results with slides in ways that are easily understandable.

Your Work

No matter your grade in the class, you belong in my classroom. All my assessments are designed to push you to do excellent work and convince both you and me of your critical thinking skills with as little unpleasantness as possible. I've deliberately designed the class with no busywork, trick questions, or high-stakes exams. And in return I expect you to work hard.

MiniExams and the Final Project make up the largest part of your final grade. The best predictor of success on MiniExams and the Final Project is the quality of your practice along the way. The class is setup to give you this practice. We will start each topic by introducing a concept, working together with that concept in class in an *Exercise*, then you will work more with the concept at home in *Homework*. At the end of each of the first five parts we will come together for a *MiniExam* where you'll show me you understand the concepts and you'll submit a *Project* where you'll show me you can work with the concepts. At the end of the semester you will build on your five part Projects with a *Final Project*. Together this gives you all the core concepts in order with structured practice in a way that should feel comfortable and intuitive.

Exercises (10%) are done together in class, practicing the concepts covered in each topic. Three will be automatically dropped. Exercises are intended to give you guided practice, due the day of class, and are graded primarily for effort.

Homeworks (10%) are due most Fridays at 5PM, one assignment per topic, with typically multiple topics per week. You may email me before the due date for a no-questions-asked extension through midnight on Sunday. Like Exercises, Homework is intended to give you a considerable amount of low-stakes guided practice and is graded primarily for effort. Group work is *strongly* encouraged, but your work should be your own. Three will be automatically dropped.

MiniExams ($1 \times 20\% + 1 \times 15\% + 1 \times 10\% + 1 \times 5\% + 1 \times 0\% = 50\%$) are designed for you to demonstrate competency and mastery. Five non-cumulative MiniExams will take place during the first part of Tuesday class (*nearly*) every other week. MiniExams will focus primarily on the material covered the current part, and may include a preview question on upcoming material for extra credit. Your top two Miniexam scores will each count as 15% of your grade. Your third highest MiniExam scores will each count as 10% of your grade. Your fourth and fifth highest Miniexam scores will each count as 5% of your grade. Your lowest Miniexam score will be dropped.

Your highest grades will count for a relatively large portion of your grade. This has the effect of lowering the stakes but also means you cannot simply add up the point totals to see how you are doing in the class. Your final grade will be based on the weighted average of four (out of five) MiniExam scores.

During MiniExams:

1. You may use material from your notes and the textbook.
2. You may not use any electronic material, other than a calculator.
3. Your exam should be your own and is subject to the Academic Conduct Code (below). If you are **at all unclear** about a particular piece of material with respect to these two guidelines, it is your responsibility to clarify its appropriateness with the instructor. Please email with any questions about what constitutes an academic violation. This is taken very seriously.

Final Project (30%): The final project is made up of two parts: a final presentation and a final paper. We will build up this project gradually. At the end of each part, you will submit a Part Project, applying the concepts in that part to a dataset of your choice. The final presentation will be given during the final exam period. The final paper will be due at the end of finals week. The final project can either be done individually or with another student. The aim is for you to demonstrate the range of skills developed in the class.

Attendance is taken on TopHat after the add/drop period and counts for an *Extra* 1% of your grade (notice grades add up to 101%). This means that if your score from assignments in the class adds up to 92% and you've attended all classes, you will receive a 93% in the class. This is very good.

Missed work cannot be made up post-hoc. Contact me before the due date to apply for an extension if you anticipate submitting late work.

Course Resources

There are many resources available to help you master the material. Some of your time should be spent on reviewing class material, some doing exercises, and some in discussion with me and your colleagues in class. Data analysis is best learned through practice, so put in the time and reps to master it. Class resources include:

- The **Course Page** holds the course together, where you'll find:
 - Lecture notes and recordings
 - Exercises
 - Homework
 - Demos
 - MiniExam information
- **Canvas** is used primarily for announcements.

- **Gradescope** will be the portal to collect and return assignments.
- **Office Hours** are a great resource so come see me or the TA! Times are listed on Canvas.
- The **Teaching Assistant** is a highly trained economist in the graduate department. They will hold office hours and grade some of your work. More information will be posted soon.
- **Textbooks** (*optional*): These textbooks are not required but may prove helpful.
 - *Data Visualization and Analysis in R*, by Dustin Fife, available freely online at <https://datavisualizations.tech/>. The book provides straightforward explanations of data analysis and visualization principles.
 - *How Charts Lie: Getting Smarter About Visual Information*, by Alberto Cairo, W. W. Norton & Company, 2019. The book provides a critical look at how data visualization can mislead and how to become a more informed consumer of visual information.
 - *Analysis of Economic Data, 2nd Edition*, by Gary Koop, it is very good and may serve you as an optional complement to lecture materials when reviewing course concepts. It is an older version of the text so it is easy to find using a simple search.

Software

The course will develop skills in Excel and Python. No prior experience is required. :)

Disability Services

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services (DRS), 140 William Pitt Union, (412) 648-7890, drsrecep@pitt.edu, (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.

Email Policy

Don't hesitate to email me or the TA with questions. When you do, be considerate of a few things. First, I will try to respond to you as quickly as possible. It's easiest for me to respond if your email has a direct question. I get a considerable number of emails, so during busy times it may take longer than a day or two to get back to you. Second, I do not check emails outside of regular working hours, so if you email me at night or on the weekend, I likely won't see your email until later. This is a good practice I'd encourage for you too.

AI Policy

Artificial intelligence of all kinds is an increasingly useful tool you will need to be familiar with in your life and work. This class does not prohibit use of AI. However, all your work must be your own. Copying from any uncited source is a serious issue. That said, do not hesitate to use AI when it can be of use as an instructor. Work with it to learn the skills; do not copy; cite your sources.

Academic Conduct Code

Students in this course will be expected to comply with the [University of Pittsburgh's Policy on Academic Integrity](#). Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy.

All students are expected to adhere to the standards of the University of Pittsburgh's Policy on Academic Integrity. Any violation is a serious matter. Any student engaged in cheating, plagiarism, or other acts of academic dishonesty would be subject to disciplinary action, may 1) receive a zero on an assessment, or 2) receive an F in the class.

Final Note

I care about your progress over the semester and will do everything I can to help you succeed. Please don't hesitate to stop by office hours or email me if you have any questions or concerns, or even if you just want to chat about the course.