ECON 0150 | Economic Data Analysis

The Data Analysis Pipeline for Economists

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

Economists rely on data to build economic models and inform decisionmaking. This course is designed to equip you with the skills to analyze, interpret, and communicate economic data. It 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, and culminating in a capstone project that you will communicate to your peers. We will emphasize practical applications through a systematic approach to data types, exploratory analysis, statistical modeling, and communication of the data. Each week builds progressively toward a capstone project, where students 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. The schedule is only a rough guide and will be updated periodically.

Part 1: Summary Exploratory Data Analysis (Weeks 1 - 3)

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. The appropriate data visualization depends on the data type. Part 1 indroduces 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: Pattern Exploratory Data Analysis (Weeks 4 - 5)

Not only do we care about single variables, but we often also care about the relationships between variables. Like summary EDA, 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, and a simple model of causal relationships.

Part 3: Building Linear Models (Weeks 6 - 7)

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: The General Linear Model (Weeks 8 and 10)

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 also lets us handle multiple predictors of different types - continuous, categorical, and their interactions - while maintaining our core tools of visualization and residual analysis.

Topics include simple regression, categorical predictors, multi-category predictors, multivariate regression.

Part 5: Advanced GLM (Weeks 11 - 13)

Economic variables don't always play by simple rules - they can be binary outcomes, ordered categories, repeated measures, or time series. We extend the General Linear Model to handle these special cases, using our understanding of data types to choose the right modeling strategy for each situation.

Topics include fixed effects, interactions, seasonality, panel methods, time series, and model selection.

Part 6: Communicating with Data (Weeks 14 - 15)

We've built from description to modeling, but insights are only useful if others can understand them. Here we focus on crafting clear narratives about economic relationships, choosing the right visualizations for our findings, and presenting results with slides in ways that are easily understandable.

Course Resources

- **Office Hours**: Come to office hours! :) They are posted on the course page.
- **TA**: The grader for the course is Jacob Stenstrom. He is excellent and can serve as a resource for material and questions about the grading. You can contact him by email dcy14@pitt.edu or drop by his office hours, posted on Canvas.
- **Optional Textbooks**:
  - *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. :)

Key Dates

- Jan 20: MLK Day, no class
- Feb 26/27: Asynchronous class
- Mar 2 - 9: Spring recess

Grades

Your final grade of this course comes from weekly ***Homework***, weekly in-class ***MiniExams***, and a ***Final Capstone Project***. I've designed the course to encourage you to put in the work to develop the core skills covered in the course. In the interest of taking the pressure off any one peice of work, I drop your lowest three Homeworks and MiniExams.

- ***Homework*** (10%): There will be Homework due most Fridays at 5PM. Homework intended to practice applying the skills developed in the course and serves as a roadmap for the capstone project.
  - Homework will primarily be focused on working with data. Students will be asked to use the most recent concepts with a chosen dataset. Some datasets will be provided, but students can experiment with their capstone project by applying the recent concepts to a chosen external dataset. Homework will be graded on appropriateness and will not depend on the chosen dataset. More detailed feedback on the assignments is available in office hours. Extra credit will often be available for students who come for feedback on the assignments.
  - You are strongly encouraged to work in groups but your work must be your own.
- ***MiniExams*** (70%): There will be a MiniExam most weeks during the first 10 minutes of the first class of the week. MiniExams will focus primarily on the material covered the previous week, and may often include a preview question on upcoming material for extra credit. All MiniExams are open-book and open-note. You may not use electronic resources or communicate with others in any way.
- ***Final Capstone Project*** (20%): The final capstone project is made up of two parts: a capstone presentation and a capstone paper. The capstone presentation will be given during the final exam period. The capstone project can either be done individually or with another student. The aim is for you to demostrate the range of skills developed in the class.

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 powerful tool you will need to be familiar with in your life and work. This class does not prohibit use of AI. However, all work must be your own. Copying from an uncited source is a serious issue. That said, do not hesitate to use AI when it can be of use but I would encourage you to treat it as an instructor. Work with it to learn the skills; do not copy.

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.

Notes

1. 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.
2. I also encourage you to use some of the excellent learning tools available for free online for learning Python and using the pandas library.