ECON 0150 | Economic Data Analysis

Part 1: Foundations and Summary EDA (Weeks 1-3)

Making decisions with data involves understanding the data. Data is easiest to understand when visualized appropriately. The appropriate data visualization depends on the data type.

Week 1: Course Foundations

Day 1: Course Introduction

- Data analysis is the systematic process of examining, cleaning, transofrming, and modeling data to uncover meaningful patters and insights.
- Economic data analysis is the data analysis that economists do.
- There are a few key types of economic data: Categorical (nominal/ordinal) vs. Continuous (interval/ratio).
- The class has two parallel tracks: weekly miniexams and weekly project milestones.
- Show an example of the workflow.

Week 2: Descriptive Analysis of Categorical Variables

Day 1: Nominal Categorical Variables

- Numerical Summaries
  - Frequency tables and counts
  - Proportions and percentages
  - Mode
  - Entropy and variation ratio
- Visualization
  - Bar plots (vertical vs horizontal)
  - Dot plots for counts
  - Grouping strategies for many categories
- Special Cases
  - Binary variables
  - Missing categories
- Examples: Industry codes, geographic regions, binary outcomes

Day 2: Ordinal Categorical Variables

- Numerical Summaries
  - Ordered frequency distributions
  - Median and quartiles
  - Cumulative frequencies
  - Rank-based measures
- Visualization
  - Ordered bar plots
  - Diverging bar plots
  - Likert plots
  - Cumulative frequency plots
- Examples: Education levels, credit ratings, satisfaction scores

Week 3: Descriptive Analysis of Continuous Variables

Day 1-2: Continuous Variables

- Numerical Summaries
  - Central tendency (mean, median, mode)
  - Spread (variance, SD, IQR, range)
  - Shape (skewness, kurtosis)
- Visualization
  - Histograms (bin width choices)
  - Density plots
  - Box plots and violin plots
  - Strip plots and rug plots
- Data Transformations
  - Log
  - Normalization
  - Elasticity
- Examples: Prices, wages, test scores

Part 2: Pattern EDA (Weeks 4-5)

Often we care about the relationships between variables. Like descriptive EDA, the relationships between variables is easiest to understand when visualized. And like descriptive EDA, the most effective data visualization depends on the variable types.

Week 4: Relationships Between Variables

Day 1: Categorical × Categorical Relationships

- Numerical Summaries
  - Two-way frequency tables
  - Row and column percentages
  - Chi-square statistics
  - Measures of association (Cramer's V, etc.)
- Visualization
  - Grouped bar plots (clustered vs stacked)
  - Mosaic plots
- Examples: Education × Employment, Industry × Region

Day 2: Continuous × Continuous Relationships

- Numerical Summaries
  - Correlation coefficients (Pearson, Spearman)
  - Covariance
  - Robust correlation measures
- Visualization
  - Scatterplots and variations (alpha, size)
  - Reference lines (45°, y=x)
  - Smoothed trends (LOWESS)
- Special Cases
  - Non-linear relationships
  - Heteroscedasticity
  - Time series patterns
- Examples: Income × Education, Price × Quantity

Week 5: Relationships Between Mixed Types

Day 1: Categorical × Continuous Relationships

- Visualization
  - Box plots by group
  - Violin plots
  - Overlaid density plots
  - Strip plots with jitter
- Numerical Summaries
  - Group means and medians

    - Group standard deviations
- Special Cases
    - Unequal group sizes
    - Heterogeneous variances
    - Multiple grouping variables
- Examples: Wages by Industry, Prices by Region

Day 2: Complex Relationships

- Multiple Continuous Variables
    - Correlation matrices
    - Partial correlations
    - Scatterplot matrices
- Multiple Categorical Variables
    - Three-way tables
    - Nested relationships
    - Hierarchical structures
- Examples: Economic indicators over time by region

Day 2 (alt): Causality and DAGs

- Alternatively, this may be the place to introduce causality. We've explored relationships, but we need to be careful not to confuse correlation with causation. We can't always disentangle the two, but we can build some models to help clarify our thinking.

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.

Week 6: Variation Around Simple Models

Day 1: Understanding Sampling Variation

- Distribution of Deviations
    - Deviations from means/medians
    - Sampling multiple means
    - Distribution of sample means
- Visualizing Sampling Distributions

    - Histograms of sample means
    - Standard error intuition
    - Simulation approaches
- Connection to CLT
    - Why means are normal
    - Role of sample size
    - Standardizing deviations
- Examples: Average wages, Mean household income

Day 2: Testing with Deviations

- From Deviations to Tests
    - One-sample case
    - Null distributions
    - t-statistics as standardized deviations
    - P-hacking: everything we do is provisional
    - p-hacking can't be fixed with statistics but with ethics and replication
        - Research integrity
        - Independent replication
        - Pre-registration of analyses
        - Publication of null results
    - open science movement
- Running Tests through Regression
    - Mean as simplest model
    - Residuals in mean-only model
    - t-test through regression
- Visual Understanding
    - Plotting null distributions
    - Visualizing p-values
    - Power through visualization
- Examples: Testing average returns, Price deviations from target

Week 7: Introduction to Linear Relationships

Week 7: Adding Slope (No Intercept)

Day 1: Understanding Slope

- Slope Through Origin

- Ratio interpretation
- Deviations from proportional relationships
- Visualizing Slope
  - Scatterplots through origin
  - Residuals from proportional fit
  - Comparing different slopes
- Testing Slopes
  - Null hypothesis for slope
  - t-test for slope
  - Confidence intervals
- Examples: Returns to scale, Proportional relationships

Day 2: Working with Slope Models

- Model Building
  - When to force through origin
  - Impact on residuals
  - R^2 in through-origin models
- Diagnostics for Slope-Only
  - Residual patterns
  - Leverage and influence
  - Prediction intervals
  - p-hacking
- Special Cases
  - Negative slopes
  - Zero slopes
  - Very steep slopes
- Examples: Production functions, Cost relationships

Part 4: The General Linear Model (Weeks 8-10)

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

Week 8: Adding the Intercept

Day 1: Why Intercepts Matter

- Understanding Intercepts
  - Baseline effects
  - Shifting the line
  - Impact on slope interpretation
- Visual Impact
  - How intercepts change fit
  - Residual patterns with/without intercept
  - Comparing models
- Testing Intercepts
  - Why test the intercept
  - Joint tests with slope
  - Economic meaning
  - p-hacking
- Examples: Fixed costs, Base salary plus commission

Day 2: Complete Linear Model

- Putting It All Together
  - Full model interpretation
  - Changes in predictions
  - Understanding all parameters
- Comprehensive Diagnostics
  - Full residual analysis
  - Influence measures
  - Model selection
  - p-hacking
- Real Applications
  - When simpler models work
  - When you need both terms
  - Economic interpretations
- Examples: Wage equations, Demand functions

Week 9: Categorical Predictors in GLM

Day 1: Binary Predictors

- Dummy Variables
  - Interpretation as group differences

    - Connection to t-tests
    - Multiple reference groups
- Visual Understanding
    - Parallel lines
    - Group means visualization
    - Residual patterns by group
- Examples: Gender wage gaps, Treatment effects

Day 2: Multi-Category Predictors

- Multiple Dummies
    - Reference category choice
    - ANOVA through regression
    - Fixed effects interpretation
- Ordered Categories
    - Linear vs dummy coding
    - Testing for ordered effects
    - Polynomial contrasts
- Examples: Industry effects, Education levels

Week 10: Complex Relationships in GLM

Day 1: Interactions

- Continuous × Continuous
    - Changing slopes
    - Centering predictors
    - Visualization techniques
- Categorical × Continuous
    - Different slopes by group
    - ANCOVA models
    - Choosing reference levels
- Categorical × Categorical
    - Interpreting dummy interactions
    - Visualization strategies

Day 2: Non-linear Relationships

- Continuous Predictors
    - Polynomial terms

Module 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.

Week 11: Advanced Applications

Day 1: Panel Data Methods

- Fixed Effects Models
  - Entity fixed effects
  - Time fixed effects
  - Two-way fixed effects
- Random Effects
  - When to use random effects
  - Mixed models introduction
  - Nested structures

Day 2: Model Selection

- Variable Selection
  - By data type
  - Information criteria
  - Cross-validation
- Diagnostics by Type
  - Categorical predictors
  - Continuous predictors
  - Mixed models

Week 12: Time Series and Limited Dependent Variables

Day 1: Time Series Methods

- Time Series Fundamentals
  - Autocorrelation
  - Trends and seasonality
  - Dynamic models

Day 2: Limited Dependent Variables

- Binary Outcomes
  - Logistic regression

- Interpretation
- Diagnostics

Week 13: Advanced Categorical Methods

Day 1: Ordered Outcomes

- Ordered Logistic Models
  - Parallel lines assumption
  - Interpretation
  - Testing assumptions

Day 2: Multinomial Outcomes

- Unordered Categories
  - Multinomial models
  - Multiple base categories
  - Visualization strategies

Module 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 in ways that are easily understandable and can help inform decisionmaking.

Focus Areas:

- Choosing visualizations by data type
- Technical writing
- Advanced visualization
- Presentation skills
- **Focus**: Identifying economic questions and understanding the data analysis workflow.
- **Key Topics**:
  - Types of economic questions.
  - The role of data in economics.
  - Overview of data analysis workflows.
- **Skills Developed**:
  - Framing actionable economic questions.
  - Differentiating data types and their applications.
- **Capstone Connection**: Formulate the primary question for the capstone project.

Weekly Schedule

Week	Part	Class Dates	Topics	Key Activities
1	1: Descriptive EDA	Jan 8	Introduction and Foundations	Course overview, economic data types, project introduction
2		Jan 13, Jan 15	Understanding Variables	Categorical and continuous variable analysis
3		Jan 20, Jan 22	EDA: Descriptives	Summarizing data visually and numerically
4	2: Pattern EDA	Jan 27, Jan 29	EDA: Bivariate Relationships	Exploring relationships between variables
5		Feb 3, Feb 5	EDA: Multivariable Relationships	Heatmaps, faceting, and interaction effects
6	3: Patterns and Models	Feb 10, Feb 12	Understanding Uncertainty	Probability, variability, and confidence intervals
7		Feb 17, Feb 19	Regression Basics	Linear regression, diagnostics, and interpretation
8	4: General Linear Model	Feb 24, Feb 26	Advanced Regression and Hypotheses	ANOVA, fixed effects, and interaction terms
9		Mar 4, Mar 6	Causal Analysis	DAGs, causal inference, and exploratory methods
10		Mar 10 - Mar 12	Spring Recess	No classes
11		Mar 17, Mar 19	Refining Analysis	Residual analysis, advanced diagnostics
12		Mar 24, Mar 26	Storytelling with Data	Crafting narratives and designing visuals
13		Mar 31, Apr 2	Professional Communication	Writing reports, integrating visuals
14		Apr 7, Apr 9	Capstone Project	Refining presentations and reports
15		Apr 14, Apr 16	Capstone Presentations	Delivering final project presentations