Introduction

Francisco Villamil

Applied Quantitative Methods II

IC3JM, Spring 2026

Course overview

- This is the second part of the quantitative methods sequence

Course overview

- This is the second part of the quantitative methods sequence
- Focus on **applying** statistical tools in practice

Course overview

- This is the second part of the quantitative methods sequence
- Focus on **applying** statistical tools in practice
- Less theory, more hands-on work with data

Course overview

- This is the second part of the quantitative methods sequence
- Focus on **applying** statistical tools in practice
- Less theory, more hands-on work with data
- Goal: go from research question to answer

What will you learn?

- How to choose the right model for your question

What will you learn?

- How to choose the right model for your question
- How to interpret and visualize model results

What will you learn?

- How to choose the right model for your question
- How to interpret and visualize model results
- How to evaluate whether a model is appropriate

What will you learn?

- How to choose the right model for your question
- How to interpret and visualize model results
- How to evaluate whether a model is appropriate
- How to work with different types of data (panel, spatial, etc.)

What will you learn?

- How to choose the right model for your question
- How to interpret and visualize model results
- How to evaluate whether a model is appropriate
- How to work with different types of data (panel, spatial, etc.)
- Best practices in computing and reproducibility

Course structure

Feb 5	Introduction
Feb 12-19	¡2¿
Feb 26	¡3¿
Mar 5	¡4¿
Mar 12-19	¡5¿
Mar 26 & Apr 9	¡6¿
Apr 16	¡7¿
Apr 23	Project presentations
Apr 30	Advanced topics

Course structure

Feb 5	Introduction
Feb 12-19	¡2¿
Feb 26	¡3¿
Mar 5	¡4¿
Mar 12-19	¡5¿
Mar 26 & Apr 9	¡6¿
Apr 16	¡7¿
Apr 23	Project presentations
Apr 30	Advanced topics

Course structure

Feb 5	Introduction
Feb 12-19	¡2¿
Feb 26	¡3¿
Mar 5	¡4¿
Mar 12-19	¡5¿
Mar 26 & Apr 9	¡6¿
Apr 16	¡7¿
Apr 23	Project presentations
Apr 30	Advanced topics

Course structure

Feb 5	Introduction
Feb 12-19	¡2¿
Feb 26	¡3¿
Mar 5	¡4¿
Mar 12-19	¡5¿
Mar 26 & Apr 9	¡6¿
Apr 16	¡7¿
Apr 23	Project presentations
Apr 30	Advanced topics

Course structure

Feb 5	Introduction
Feb 12-19	¡2¿
Feb 26	¡3¿
Mar 5	¡4¿
Mar 12-19	¡5¿
Mar 26 & Apr 9	¡6¿
Apr 16	¡7¿
Apr 23	Project presentations
Apr 30	Advanced topics

Course structure

Feb 5	Introduction
Feb 12-19	¡2¿
Feb 26	¡3¿
Mar 5	¡4¿
Mar 12-19	¡5¿
Mar 26 & Apr 9	¡6¿
Apr 16	¡7¿
Apr 23	Project presentations
Apr 30	Advanced topics

Course structure

Feb 5	Introduction
Feb 12-19	¡2¿
Feb 26	¡3¿
Mar 5	¡4¿
Mar 12-19	¡5¿
Mar 26 & Apr 9	¡6¿
Apr 16	¡7¿
Apr 23	Project presentations
Apr 30	Advanced topics

Evaluation

- Problem sets (20%)
  - → Started in class, finished at home
  - → Short deadlines
- Proposal presentation and peer review (10% + 10%)
- Final essay (30%)
  - → Small research note (max 3,000 words)
  - → Original data analysis using R
- Exam (30%)

The Big Picture

The research process

Theory $\longleftrightarrow$ **Data Generating Process** $\longleftrightarrow$ **Data**

- Theories make claims about how the world works

The research process

$$\textbf{Theory} \longleftrightarrow \textbf{Data Generating Process} \longleftrightarrow \textbf{Data}$$

- Theories make claims about how the world works
- These claims imply certain patterns in data

The research process

Theory $\longleftrightarrow$ **Data Generating Process** $\longleftrightarrow$ **Data**

- Theories make claims about how the world works
- These claims imply certain patterns in data
- We observe data and try to learn about the underlying process

What is a Data Generating Process (DGP)?

- The rules that govern how data comes to exist

What is a Data Generating Process (DGP)?

- The rules that govern how data comes to exist
- Includes:

What is a Data Generating Process (DGP)?

- The rules that govern how data comes to exist
- Includes:
  - $\rightarrow$ The social or political process we study

What is a Data Generating Process (DGP)?

- The rules that govern how data comes to exist
- Includes:
  - → The social or political process we study
  - → How observations end up in our dataset

What is a Data Generating Process (DGP)?

- The rules that govern how data comes to exist
- Includes:
  - → The social or political process we study
  - → How observations end up in our dataset
- We never observe the DGP directly

What is a Data Generating Process (DGP)?

- The rules that govern how data comes to exist
- Includes:
  - → The social or political process we study
  - → How observations end up in our dataset
- We never observe the DGP directly
- We use statistical models to make inferences about it

Why do we need statistics?

- Our theories deal with processes, not just data

Why do we need statistics?

- Our theories deal with processes, not just data
- Data is a window into the underlying process

Why do we need statistics?

- Our theories deal with processes, not just data
- Data is a window into the underlying process
- Statistics helps us:

Why do we need statistics?

- Our theories deal with processes, not just data
- Data is a window into the underlying process
- Statistics helps us:
  - → Separate signal from noise

Why do we need statistics?

- Our theories deal with processes, not just data
- Data is a window into the underlying process
- Statistics helps us:
  - $\rightarrow$ Separate signal from noise
  - $\rightarrow$ Quantify uncertainty

Why do we need statistics?

- Our theories deal with processes, not just data
- Data is a window into the underlying process
- Statistics helps us:
  - → Separate signal from noise
  - → Quantify uncertainty
  - → Make valid inferences

Sources of uncertainty

- **Sampling uncertainty**: We observe a sample, not the population

Sources of uncertainty

- **Sampling uncertainty**: We observe a sample, not the population
- **Theoretical uncertainty**: Our theories are simplifications

Sources of uncertainty

- **Sampling uncertainty**: We observe a sample, not the population
- **Theoretical uncertainty**: Our theories are simplifications
- **Fundamental uncertainty**: Some processes are inherently random

Sources of uncertainty

- **Sampling uncertainty**: We observe a sample, not the population
- **Theoretical uncertainty**: Our theories are simplifications
- **Fundamental uncertainty**: Some processes are inherently random

Sources of uncertainty

- **Sampling uncertainty**: We observe a sample, not the population
- **Theoretical uncertainty**: Our theories are simplifications
- **Fundamental uncertainty**: Some processes are inherently random

- All of these create “noise” in our data

Sources of uncertainty

- **Sampling uncertainty**: We observe a sample, not the population
- **Theoretical uncertainty**: Our theories are simplifications
- **Fundamental uncertainty**: Some processes are inherently random

- All of these create "noise" in our data
- Statistical models help us deal with this noise

The logic of statistical inference

- **Probability theory**: Given a known process, what data will we see?

The logic of statistical inference

- **Probability theory**: Given a known process, what data will we see?

The logic of statistical inference

- **Probability theory**: Given a known process, what data will we see?
- **Statistical inference**: Given observed data, what can we learn about the process?

The logic of statistical inference

- **Probability theory**: Given a known process, what data will we see?
- **Statistical inference**: Given observed data, what can we learn about the process?

The logic of statistical inference

- **Probability theory**: Given a known process, what data will we see?
- **Statistical inference**: Given observed data, what can we learn about the process?
- We're doing the reverse: from data back to process

Review: Key Concepts from AQMSS-I

The regression model

The most common tool in social science:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y: outcome we want to explain

The regression model

The most common tool in social science:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y: outcome we want to explain
- X: explanatory variable(s)

The regression model

The most common tool in social science:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y: outcome we want to explain
- X: explanatory variable(s)
- β: coefficients (what we estimate)

The regression model

The most common tool in social science:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y: outcome we want to explain
- X: explanatory variable(s)
- β: coefficients (what we estimate)
- ε: error term (what we can't explain)

What regression tells us

- Regression estimates **conditional expectations**

What regression tells us

- Regression estimates **conditional expectations**
- "What is the average Y for units with a given value of X?"

What regression tells us

- Regression estimates **conditional expectations**
- "What is the average Y for units with a given value of X?"

What regression tells us

- Regression estimates **conditional expectations**
- "What is the average Y for units with a given value of X?"

- The slope β_1 tells us:

What regression tells us

- Regression estimates **conditional expectations**
- "What is the average Y for units with a given value of X?"
- The slope β_1 tells us:
  - $\rightarrow$ How much Y changes, on average

What regression tells us

- Regression estimates **conditional expectations**
- "What is the average Y for units with a given value of X?"

- The slope β_1 tells us:
  - $\rightarrow$ How much Y changes, on average
  - $\rightarrow$ When comparing units that differ by 1 in X

Descriptive vs. Causal interpretation

- **Descriptive**: How do units with different X values compare?
  $\rightarrow$ "People with more education earn more, on average"
- **Causal**: What happens if we change X for a given unit?
  $\rightarrow$ "If we give someone more education, they will earn more"
- Same coefficient, very different claims!

The challenge of causal inference

- Causal effects are about **counterfactuals**

The challenge of causal inference

- Causal effects are about **counterfactuals**
- "What would have happened if things were different?"

The challenge of causal inference

- Causal effects are about **counterfactuals**
- "What would have happened if things were different?"

The challenge of causal inference

- Causal effects are about **counterfactuals**
- "What would have happened if things were different?"

- The problem: we can't observe counterfactuals

The challenge of causal inference

- Causal effects are about **counterfactuals**
- "What would have happened if things were different?"

- The problem: we can't observe counterfactuals
- We need strategies to infer them

The challenge of causal inference

- Causal effects are about **counterfactuals**
- "What would have happened if things were different?"

- The problem: we can't observe counterfactuals
- We need strategies to infer them
- This will be a recurring theme throughout the course

What makes a good analysis?

- Clear research question

What makes a good analysis?

- Clear research question
- Appropriate data for the question

What makes a good analysis?

- Clear research question
- Appropriate data for the question
- Right statistical model for the data

What makes a good analysis?

- Clear research question
- Appropriate data for the question
- Right statistical model for the data
- Correct interpretation of results

What makes a good analysis?

- Clear research question
- Appropriate data for the question
- Right statistical model for the data
- Correct interpretation of results
- Honest about limitations and uncertainty

Looking ahead

- Next session: Applied regression in depth

Looking ahead

- Next session: Applied regression in depth
- How to set up a regression analysis

Looking ahead

- Next session: Applied regression in depth
- How to set up a regression analysis
- How to interpret coefficients correctly

Looking ahead

- Next session: Applied regression in depth
- How to set up a regression analysis
- How to interpret coefficients correctly
- Common pitfalls and how to avoid them

For next week

- Read Urdinez & Cruz (2020), chapters 1-5
- Review your notes on OLS from AQMSS-I
- Start Problem Set 1

- Check Aula Global for additional materials

Questions?