GR5010 Quantitative Theory and Methodology for the Social Sciences

Fall 2024
Wednesdays 4:10PM–6:00PM (Section 4)
602 Northwest Corner Building

Mike He
Course Instructor
zh2263@columbia.edu
Office: 528 Building, Room 706 (Teachers College)
Office Hours: Mondays 4–6pm and by appointment

Yin (Linda) Lin
Teaching Assistant
yl5384@columbia.edu
Office: International Affairs Building, Room 270B
Office Hours: Tuesdays 2–4pm

Course Overview

This course – one of the two foundational courses in the QMSS curriculum – is designed as an in-depth introduction to the social sciences and its methodologies. It is intended to give a broad overview so students can intelligently combine ideas in solving real-world problems.

We will focus on the logic and design of social research, beginning with some concepts and topics common to research across the social sciences. We will later move on to understanding the principles behind an array of methodologies used in the social sciences: causal inference, experimentation, observational studies, formal models, surveys, and applied machine-learning techniques. We will analyze their applications using cases drawn from the research literature.

The focus of this course is not on the techniques themselves – you will have ample opportunity to learn that deeply in other courses – but in understanding the logic behind the use of these tools to extract meaningful answers from their applications.

Prerequisites: It is assumed that you have had at least one semester of graduate-level statistics involving linear regression and analysis of variance. Some basic mathematics and algebra will also be assumed.

Course Materials

Required Readings are mandatory and should be completed before each class. They are the theoretical foundation to discuss each week's topic.

Topic Readings are also mandatory and should be completed before each class. These are empirical applications of the methods we will discuss and analyze in class.

Complementary Readings are intended to serve as further (and future) references if you ever want to delve deeper on a particular topic, but are not a requirement to this class.

Thesis Readings are two texts intended to be resources for the crafting of your Thesis Proposal, but will not be discussed in class. Read the assigned chapters in the suggested order/dates:

- Booth, W. C., Colomb, G. G., and Williams, J. M. (2024). *The Craft of Research*. University of Chicago Press, Chicago, IL, fifth edition.
- Weston, A. (2018). *A Rulebook for Arguments*. Hackett Publishing Co, Indianapolis, IN, fifth edition.

Course Dynamics

The class will be a combination of lectures and focused discussion. We will devote the first half of the class to a lecture on the topic assigned for the week, and the second half of the class to analyze one or two Topic Readings which contain applications of the methods under discussion. To kick off these discussions, students will give brief presentations to introduce each reading, followed by a class discussion to analyze in detail these applications of the methods we are studying.

Each **Topic Presentation** should address the following questions:

- What is the research question?
- What is the specific hypothesis under investigation?
- Can it be falsified?
- Why was a particular method chosen?
- What inferences can be drawn from its results?
- What are its limitations?

Each presentation should be accompanied by a one-page summary to be emailed to the entire class at 4PM on the day prior to class. **Before each class, you should have read, thought about, and be prepared to discuss all Required and Topic Readings assigned for the week.**

At the end of the course, you will turn in a 10–12-page **Thesis Proposal**. To ensure that you produce a fully-fledged product, you will be required to hand in **Thesis Proposal Deliverables** (approximately) every two weeks. These will constitute the foundational pieces for your proposal. Feedback will be provided on these deliverables where required. Deliverables should be submitted on Canvas **by 4PM (ET) on the indicated dates**.

Late Submission Policy: Deliverables and your thesis proposal are expected to be submitted on the due date. For every day after the submission date, 10% of the maximum grade will be deducted from the score.

Course Requirements

The final grade of the course will be based of your fulfillment of each of the following requirements:

Attendance and class participation (20%): Students are expected to have read all the required readings before class, participate in class exercises and team assignments, and actively contribute to class discussion. Note that you will not obtain the full 20% unless you actively participate in class.

Topic presentations (15%): Students will be assigned to groups to present an overview of a **Topic Reading**. Make sure to send your one-page summary to the full class by 4PM on the day prior to the class when you are presenting.

Thesis Proposal Deliverables (40%): Students will submit pieces of their proposal every two weeks. Make sure to submit each one of them by 4PM (ET) on the indicated dates.

Thesis proposal (20%): Throughout the course, you will work on a research proposal. This 10–12-page paper will be turned in at the end of the semester.

Group peer review (5%): Towards the beginning of the semester (around Week 4), students will be assigned to teams composed of individuals with different skill sets, and will work with the group on various activities throughout the semester. To hold all group members accountable, you will complete a survey assessing each other's efforts at the end of the semester.

All written work must be original and produced exclusively for this class. You are expected to follow the University's guidelines for the submission of written work.

Statement on Academic Integrity

Columbia's intellectual community relies on academic integrity and responsibility as the cornerstone of its work. Graduate students are expected to exhibit the highest level of personal and academic honesty as they engage in scholarly discourse and research. In practical terms, you must be responsible for the full and accurate attribution of the ideas of others in all of your research papers and projects; you must be honest when taking your examinations; you must always submit your own work and not that of another student, scholar, or internet source, **including generative AI software** (such as ChatGPT). Graduate students are responsible for knowing and correctly utilizing referencing and bibliographical guidelines. When in doubt, consult your professor. Citation and plagiarism-prevention resources can be found at the GSAS page on Academic Integrity and Responsible Conduct of Research.

Failure to observe these rules of conduct will have serious academic consequences, **up to and including dismissal from the university**. If a faculty member suspects a breach of academic honesty, appropriate investigative and disciplinary action will be taken following the Dean's Discipline procedures.

I take academic integrity violations very seriously. All deliverables submitted in the course are automatically submitted to Turnitin, and any instances of suspected plagiarism will be referred to the Columbia University Student Conduct and Community Standards (SCCS) through Academic Integrity Violation Referrals for academic discipline.

Statement on Disability Accommodations

If you have been certified by Disability Services (DS) to receive accommodations, please either bring your accommodation letter from DS to your professor's office hours to confirm your accommodation needs, or ask your liaison in GSAS to consult with your professor. If you believe that you may have a disability that requires accommodation, please contact **Disability Services** at 212-854-2388 or disability@columbia.edu.

Important: To request and receive an accommodation you must be certified by DS.

Note: The information in this syllabus is subject to change.

Course Outline

CLASS 1: INTRODUCTION TO THE COURSE [9/4]

What is this course (and what it is not). Course overview. What is so unique about quantitative methods applied to the social sciences? Why do we need models to understand the world? Why is it useful to have statistical models in the social sciences?

CLASS 2: THE "SCIENCE" OF SOCIAL SCIENCE [9/11]

What is so scientific about social sciences? A look at "the method". Inductive vs. deductive perspectives. Theories, hypothesis, and falsifiability. Links to quantitative methods. Mechanisms.

Required Readings:

- Gelman, A. (2011). Induction and deduction in Bayesian data analysis. *Rationality, Markets and Morals*, 2:67–78.
- Elster, J. (2007). *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge University Press, Cambridge, MA. [Ch 1-2]

Thesis Readings:

- Booth, W. C., Colomb, G. G., and Williams, J. M. (2024). *The Craft of Research*. University of Chicago Press, Chicago, IL, fifth edition. [Ch 1]
- Weston, A. (2018). *A Rulebook for Arguments*. Hackett Publishing Co, Indianapolis, IN, fifth edition. [Ch I-II]

Complementary Readings:

- Popper, K. (2002[1935]). *The Logic of Scientific Discovery*. Routledge, New York, NY.
- Kuhn, T. S. (2012). *The Structure of Scientific Revolution*. University of Chicago Press, Chicago, IL, fourth edition.
- King, G., Keohane, R. O., and Verba, S. (1995). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press, Princeton, NJ.
- Gelman, A. and Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1):8-38.
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34:103-115.

[9/13] – Thesis Proposal Deliverable #1 (research topic) due.

CLASSES 3 | 4: CAUSALITY AND CAUSAL INFERENCE (I & II) [9/18, 9/25]

[9/18] Guest lecture from Dr. Ilan Cerna-Turoff

Causes of effects or effects of causes? The search for causes: from Aristotle to Fisher. The fundamental problem of causal inference. The Neyman-Rubin model.

Required Readings:

- Hernán M. A., Robins J. M. (2020) *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. [Ch 1]
- Holland, P. W. (1980). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945-960.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modelling, decisions. *Journal of the American Statistical Association*, 100(469):322-331.

Thesis Readings:

- Booth, W. C., Colomb, G. G., and Williams, J. M. (2024). *The Craft of Research*. University of Chicago Press, Chicago, IL, fifth edition. [Ch 2]
- Weston, A. (2018). *A Rulebook for Arguments*. Hackett Publishing Co, Indianapolis, IN, fifth edition. [Ch VI-VII]

Complementary Readings:

- Imbens, G. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. Cambridge University Press, Cambridge, MA.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, MA, second edition.
- Sekhon, J. S. (2004). Quality meets quantity: Case studies, conditional probability and counterfactuals. *Perspectives on Politics*, 2(2):281-293.
- Dawid, A. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407-448.
- Page, S. (2006). Path dependence. *Quarterly Journal of Political Science*, 1(1):87-115.

[9/25] – Thesis Proposal Deliverable #2 (hypotheses) due.

CLASS 5: EXPERIMENTS AND RANDOMIZATION [10/2]

Theoretical foundations of experiments. Statistical foundations of experiments. Taxonomy of randomized experiments. Randomized experiments as the golden standard for causal inference. Inference form randomized experiments.

Required Readings:

- Hernán M. A., Robins J. M. (2020) *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. [Ch 2]
- Green, D. and Gerber, A. (2003). The under-provision of experiments in Political Science. *Annals of the American Academy of Political Science*, 589:94-112.
- Angrist, J. D. and Pischke, J.-S. (2015). *Mastering 'Metrics': The Path from Cause to Effect*. Princeton University Press, Princeton, NJ. [Ch 1]

Topic: Voter Turnout

- De la O, A. (2013). Do conditional cash transfers affect electoral behavior? Evidence from a randomized experiment in Mexico. *American Journal of Political Science*, 57(1):1-14.

Thesis Readings:

- Booth, W. C., Colomb, G. G., and Williams, J. M. (2024). *The Craft of Research*. University of Chicago Press, Chicago, IL, fifth edition. [Ch 13, 17]

Complementary Readings:

- Martel García, F. and Wantchekon, L. (2010). Theory, external validity, and experimental
- inference: Some conjectures. *Annals of the American Academy of Political Science*, 628(132-147).
- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.
- Imbens, G. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. Cambridge University Press, Cambridge, MA.
- Morton, R. and WIlliams, K. (2010). *Experimental Political Science and the Study of Causality*. Cambridge University Press, Cambridge, MA.
- Salganik, M. J. (2018). *Bit by bit*. Princeton University Press, Princeton, NJ. [Ch 4]

CLASS 6: OBSERVATIONAL STUDIES [10/9]

How are they different from experiments? Understanding their limitations for causal inference. Statistical tools to cope with non-random assignment of treatments. Inference from observational studies.

Required Readings:

- Hernán M. A., Robins J. M. (2020) *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. [Ch 3]
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2(3):808-840.
- Przeworski, A. (2009). Is the science of Comparative Politics possible? In Boix, C. and Stokes, S. C., editors, *Oxford Handbook of Comparative Politics*. Oxford University Press, New York, NY.

Topic: Selected statistical applications to observational studies

- Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493-505.

Complementary Readings:

- Cochran, W. G. (2015 [1972]). Observational studies. *Observational Studies*, 1(1):126-136.
- Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge University Press, Cambridge, MA.
- Rosenbaum, P. R. (2002). *Observational Studies*. Springer, New York, NY.
- Imai, K., King, G., and Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the royal statistical society: series A (statistics in society)*, 171(2):481-502.
- Winship, C. and Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25(1):659-706.
- Sekhon, J. S. and Titiunik, R. (2012). When natural experiments are neither natural nor experiments. *American Political Science Review*, 106(1):35-57.

[10/9] – Thesis Proposal Deliverable #3 (data description) due.

CLASS 7: MEASUREMENT AND MEASUREMENT ERROR [10/16]

Research design and the research question. Measurements as a function of concepts. Theoretical consequences of measurement error. Statistical consequences of measurement error.

Required Readings:

- Geddes, B. (1990). How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis*, 2(1):131-150.
- Hausman, J. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *Journal of Economic Perspectives*, 15(4):57-67.
- Zeger, S. L., Thomas, D., Dominici F., Samet J. M., Schwartz J., Dockery D., Cohen A. (2000). Exposure Measurement Error in Time-Series Studies of Air Pollution: Concepts and Consequences. *Environmental Health Perspectives*, 108(5):419-426.

Topic: Economic Perceptions

- Michelitch, K., Morales, M. A., Owen, A., and Tucker, J. A. (2012). Looking to the future: Prospective economic voting in 2008 presidential elections. *Electoral Studies*, 31(4):838-851.

Thesis Readings:

- Booth, W. C., Colomb, G. G., and Williams, J. M. (2024). *The Craft of Research*. University of Chicago Press, Chicago, IL, fifth edition. [Ch 5-7, 10]

Complementary Readings:

- Alwin, D. F. (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. John Wiley & Sons, Hoboken, NJ.
- Goertz, G. (2008). Concepts, theories, and numbers: A checklist for constructing, evaluating, and using concepts or quantitative measures. In Box-Steffensmeier, J., Brady, H. E., Collier, D., and Goertz, G., editors*, Oxford Handbook of Political Methodology*. Oxford University Press, New York, NY.
- Jackman, S. (2008). Measurement. In Box-Steffensmeier, J., Brady, H. E., Collier, D., and Goertz, G., editors, *Oxford Handbook of Political Methodology*. Oxford University Press, New York, NY.

CLASS 8: IDENTIFICATION [10/23]

Statistical inference vs. identification. Extrapolation. Selection. Endogeneity. The empirical – and theoretical – problems of endogeneity. Quantitative methods to address identification. Instrumental variables.

Required Readings:

- Angrist, J. D. and Pischke, J.-S. (2015). *Mastering 'Metrics': The Path from Cause to Effect*. Princeton University Press, Princeton, NJ. [Ch 3]
- Morgan, S. L. and Winship, C. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, Cambridge, MA, second edition. [Ch 3]

Topic: Selected statistical applications to address identification

- Miguel, E., Satyanath, S., and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4):725-753.

Complementary Readings:

- Manski, C. F. (1999). *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, MA.
- Aronow, P. M. and Miller, B. T. (2019). *Foundations of Agnostic Statistics*. Cambridge University Press, Cambridge, MA. [Ch 6-7]
- Keele, L. (2015). The discipline of identification. *PS: Political Science & Politics*, 48(1):102-105.
- Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, 15(4):69-85.
- Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2014). Causality and endogeneity: Problems and solutions. In *Oxford Handbook of Leadership and Organizations*. Oxford University Press, New York, NY.
- Angrist, J. D., Imbens, G., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444-455.

[10/23] – Thesis Proposal Deliverable #4 (literature review) due.

CLASSES 9 | 10: SURVEY RESEARCH AND SURVEY METHODOLOGY (I & II) [10/30, 11/6]

What is wrong (and what is right) with polls? The Total Survey Error Paradigm. Respondent selection issues. Response accuracy issues. Survey administration issues.

Required Readings:

- Bautista, R. (2012). An overlooked approach in survey research: Total Survey Error. In *Handbook of Survey Methodology for the Social Sciences*. Springer, New York, NY.
- Kennedy, C., Blumenthal, M., Clement, S., Clinton, J. D., Durand, C., Franklin, C., McGeeney, K., Miringoff, L., Olson, K., Rivers, D., Saad, L., Witt, E., and Wlezien, C. (2017). An evaluation of 2016 election polls in the United States. AAPOR Ad Hoc Committee on 2016 Election Polling.

Topic: Question Wording and Response Scales

- Lundmark, S., Gilljam, M., and Dahlberg, S. (2015). Measuring generalized trust: An examination of question wording and the number of scale points. *Public Opinion Quarterly*, 80(1):26-43.

Topic: Mode of Data Collection

- Sakshaug, J. W., Yan, T., and Tourangeau, R. (2010). Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly*, 74(5):907-933.

Thesis Readings:

- Booth, W. C., Colomb, G. G., and Williams, J. M. (2016). *The Craft of Research*. University of Chicago Press, Chicago, IL, fourth edition. [Ch 11-12, 14-15]
- Weston, A. (2018). *A Rulebook for Arguments*. Hackett Publishing Co, Indianapolis, IN, fifth edition. [Ch VIII]

Complementary Readings:

- Weisberg, H. F. (2009). *The Total Survey Error Approach: A guide to the new science of survey research*. University of Chicago Press, Chicago, IL.
- Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press, Cambridge, MA.
- Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., and Lane, J. (2016). *Big Data and Social Science: A practical guide to methods and tools*. CRC Press, New York, NY.
- De Leeuw, E. D., Hox, J., and Dillman, D. (2012). *International Handbook of Survey Methodology*. Routledge, New York, NY.
- Salganik, M. J. (2018). *Bit by bit*. Princeton University Press, Princeton, NJ. [Ch 3]

[11/6] – Thesis Proposal Deliverable #5 (research strategy) due.

CLASS 11: APPLIED MACHINE LEARNING [11/13]

[11/20] No class

What do we learn from machine learning? ML for inference and prediction. ML applications to social science problems.

Required Readings:

- Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87-106.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3):289-310.

Topic: Text as Data – Estimating Ideology from Texts

- Barberá, P. (2014). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1):76-91.

Topic: Machine Learning and Causal Inference

- Bansak, K., Ferwerda, J., Hainmueller, J., Dillion, A., Hangartner, D., Lawrence, D., and Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, 359:325-329.

Complementary Readings:

- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. Springer, New York, NY, second edition.
- Murphy, K. P. (2012). *Machine Learning: a probabilistic approach*. MIT Press, Cambridge, MA.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3-27.
- Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267-297.
- Manning, C., Raghavan, P., and Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge University Press, Cambridge, MA.
- Kern, H. L., Stuart, E. A., Hill, J., and Green, D. (2016). Assessing methods for generalizing experimental impact estimates to target populations. *Journal of Research on Educational Effectiveness*, 9(1):103-127.

[11/20] – Thesis Proposal Deliverable #6 (research proposal draft) due.

CLASS 12: MODELING HUMAN BEHAVIOR [12/4]

[11/27] Academic Holiday – No class

Why models? Can behavior be modeled? Applied game theory and rational choice paradigm. Formal models to help explain collective behavior.

Required Readings:

- Page, S. E. (2018). *The model thinker*. Basic Books, New York, NY. [Ch 2-4]

Complementary Readings:

- Friedman, M. (2008). The methodology of positive economics. In Hausman, D. J., editor, *The Philosophy of Economics*. Cambridge University Press, Cambridge, MA.
- Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago Press, Chicago, IL.
- Sen, A. (1999). The possibility of social choice. *American Economic Review*, 89(3):349-378.
- Riker, W. H. (1995). The political psychology of rational choice theory. *Political Psychology*, 16(1):23-44.
- Arrow, K. J. (1994). Methodological individualism and social knowledge. *The American Economic Review*, 84(2):1-9.
- Morton, R. B. (1999). *Methods and models: A guide to the empirical analysis of formal models in political science*. Cambridge University Press, Cambridge, MA.
- Osborne, M. J. (2004). *An Introduction to Game Theory*. Oxford University Press, New York, NY.
- Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar, Strauss and Groux, New York, NY.
- Chwe, M. S.-Y. (2014). *Jane Austen: Game Theorist*. Princeton University Press, Princeton, NJ.

[12/11] – FINAL PAPER DUE.