

PPOL 5140: Quantitative Analysis and Empirical Methods (L1)

Fall, 2026

Division of Public Policy
The Hong Kong University of Science and
Technology

INSTRUCTOR: Dr. Yatang Lin

PHONE: 23587625

EMAIL: linyt@ust.hk

OFFICE HOURS: by appointment

SECTION: L1 (1732)

LECTURE TIME: Friday 12:30 PM-3:20 PM

LECTURE VENUE: Rm 3, 30/F, Tower 1, Millennium

TEACHING ASSISTANT: Zining Wang

EMAIL: zwangpe@connect.ust.hk

COURSE DESCRIPTION

This course introduces students to essential statistical concepts and methods and their applications in public policy analysis, using real-world data and hands-on data manipulation. This course provides an introduction to research design, measurement, sampling, survey research, descriptive statistics, probability theory, and statistical inference, with an emphasis on the ways in which they are applied to practical policy questions. Students will learn valuable tools for describing, analyzing, and presenting data and research findings. They will also learn the importance of judgment in drawing inferences from data. The course has a strong hands-on lab component, in which students use real datasets to practice data management, statistical analysis, visualization, and interpretation.

GOALS

After completing this course, you will be able to:

1. Familiarize with main research methods commonly used in public policy research and evaluation;
2. Select and apply appropriate statistical methods to a policy question, evaluate what different research designs can establish, and present the findings in a way that is accessible to the public and policymakers;
3. Critically consume policy studies/papers/reports in which statistical analysis is used;
4. Understand basic statistical concepts and methods used in social science research;
5. Become proficient in the use of Stata to analyze real data — importing, cleaning, reshaping, merging, graphing, and estimating — with an AI assistant as your co-pilot;

TEXTBOOKS AND MATERIALS

Required Textbooks

- *Research Methods for Public Administrators* by Elizabethann O'Sullivan and Gary Raymond Rassel, 6th Edition, 2016.
- *Statistical Methods for the Social Sciences* by Alan Agresti, 5th Edition, 2017.
- *Introduction to Econometrics*, by James H. Stock and Mark W. Watson, 4th Edition, 2020

Other Useful Readings

- *Introduction to the Practice of Statistics* by Moore, McCabe and Craig. W.H. Freeman, 9th Edition, 2017
- *How to Lie with Statistics* by Huff and Geis. W. W. Norton & Company, 1993.
- *The Craft of Research* by Booth, W. C., G. G. Colomb, and J. M. Williams, The University of Chicago Press, 2003.
- *The Drunkard's Walk: How Randomness Rules Our Lives* by Leonard Mlodinow, Pantheon Books, 2008.
- *Naked Statistics: Stripping the Dread from the Data* by Charles Wheelan, W.W. Norton & Company, 2013.

Supplementary Reading Materials

You may be asked to read a few research articles as exemplary applications of the methods covered in the course. They will be posted in the electronic format at the course website.

Statistical Software Packages

One of the objectives of the course is to help you gain proficiency using statistical software packages. You are required to use Stata to conduct statistical work in this course. While you may find it beneficial to learn other packages such as SAS or SPSS independently, all coursework must be completed using Stata. The latest version of Stata is installed on PCs in the Computer Barn A Teaching Area. You may purchase Stata BE using a student discount at the following website: [Stata Student Pricing](#) (USD 48 for 6 months, USD 94 for an annual license, and USD 225 for a perpetual license).

In addition to Stata, you may use an approved AI assistant through a web interface or an optional desktop coding environment. AI may help you understand concepts, develop and debug Stata code, interpret output, and improve written explanations. No particular AI provider, API account, or desktop configuration is required unless separately announced.

ASSESSMENT

The assessments of the course will be composed of four parts.

40% — Problem sets

5% — Class participation and engagement

15% — Empirical project

40% — Final exam

Problem sets (40%): 4 problem sets (10 points each)

Problem sets will give you hands-on experience with the analytic techniques introduced in class. They will be posted on the course website, as will suggested answers. Problem sets should be graded primarily on serious effort and completeness, with correctness used as a secondary criterion to support learning. We will give each problem set at least one week before it is due. You must submit the hand- or type-written solutions to TA before the beginning of the class on the due day. Problem sets not received before

the deadline will be considered late. No late submissions will be accepted except in emergencies.

Empirical Project (15%)

The empirical project provides students with an opportunity to apply course concepts to a policy-relevant question using real data, emphasizing research design, statistical analysis, interpretation, reproducibility, and clear communication of findings.

Submission — Prepare a 10–15 page report with integrated tables and graphs, plus a reproducible Stata do-file and log file.

AI exploration is permitted throughout the project. You may use an approved AI assistant to understand the assigned datasets, develop and debug code, interpret output, improve presentation, and revise your writing. No prescribed prompts or AI transcripts are required. All numerical claims must be generated from the assigned data and reproduced by your submitted do-file, and you must be responsible for the output.

Submit the report, one Stata do-file that reproduces every reported result, one complete log file, and a brief AI-use disclosure.

Class participation and engagement (5%)

We strongly believe that student participation can substantially enrich the learning experience for both the students and the instructor. In this spirit, class participation is encouraged. Effective class participation requires that you do the assigned readings before coming to class. You are encouraged to ask questions all the time during the class. Asking and answering questions on the Canvas discussion board will also count towards participation.

Final exams (40%)

The final exam will be closed book/notes and a formula sheet will be provided. Calculators may be used, but statistical functions on them may not. Calculators that allow text storage are not permitted.

TOPICS AND READINGS

** denotes an advanced reading where you should only skim the non-technical portions. All readings are on the course website.*

Week 1 — Fri 4 September

Course Overview and Introduction to Quantitative Policy Analysis

Course goals, expectations, assessment structure, responsible AI use, reproducible workflow, introduction to Stata, and the role of quantitative evidence in public policy analysis.

Week 2 — Fri 11 September

Research Design and Causality

Research questions and hypotheses; association versus causality; potential outcomes; randomized and natural experiments.

PS1 assigned after class.

Week 3 — Fri 18 September

Describing Data and Probability

Variables and measurement; descriptive statistics; visualization; basic probability; random variables; expected value and variance.

Week 4 — Fri 25 September

Distributions, Sampling, and Statistical Inference

Normal and binomial distributions; sampling designs; sampling distributions; standard errors and the central limit theorem.

Tutorial 1: web-AI and Stata verification; establish a project workspace; descriptive statistics and visualization.

PS2 assigned after class.

Week 5 — Fri 2 October

Estimation and Hypothesis Testing

Point and interval estimation; confidence intervals; significance tests; Type I and Type II errors; power.

Tutorial 2: optional desktop-AI setup

PS1 due

Week 6 — Fri 9 October

Comparing Two Groups and Proportions

Comparing two means; paired and independent samples; two-way tables; chi-square tests; inference for proportions.

Tutorial 3: policy experiments and two-group comparisons; PS1 support.

Week 7 — Fri 16 October

Comparing Several Means: ANOVA

One-way and two-way ANOVA; omnibus tests; assumptions; main effects and interactions.

Tutorial 4: one-way ANOVA.

PS2 due, PS3 assigned after class.

Week 8 — Fri 23 October

Simple Linear Regression

Linear relationships; OLS estimation; coefficient interpretation; uncertainty; association versus causal interpretation.

No tutorial. Chung Yeung Festival

Week 9 — Fri 30 October

Multiple Linear Regression

Controls; omitted-variable bias; indicator variables; interactions; model comparison and joint tests.

Tutorial 6: regressions. PS2 support

Week 10 — Fri 6 November

Regression Diagnostics and Robust Inference

Functional form; outliers and leverage; heteroskedasticity; robust standard errors; model diagnostics and presentation.

Tutorial 7: guided download, import, and inspection using public-data files.

PS3 due.

Week 11 — Fri 13 November

Panel Data, Difference-in-Differences,

Cross-sectional, time-series, and panel data; fixed effects; difference-in-differences; identifying assumptions.

Tutorial 8: clean, document, and conduct a basic analysis of the guided-download data; PS3 support.

PS4 assigned after class.

Week 12 — Fri 20 November

Binary Outcomes, and Regression Discontinuity

linear probability, logit, and probit models; regression discontinuity; thresholds, bandwidths, continuity,

and manipulation checks.

Tutorial 9: AI-assisted analysis, panel data, matching

Week 13 — Fri 27 November

AI, Machine Learning, Causal Inference, Big Data, Ethics, Communication, and Review

Prediction versus causal inference; model validation; AI-assisted analysis; matching and sensitivity to modeling choices; big-data opportunities and limits; algorithmic bias; responsible AI; communicating quantitative evidence; cumulative review.

Tutorial 10: final project support, communication of results, and cumulative review.

PS4 due Friday 27 Friday 2026, 4:00 pm.

Empirical Project due Friday 11 December 2026, 4:00 pm.