GEORGIA INSTITUTE OF TECHNOLOGY

School of Public Policy

PUBP 3120: Statistical Analysis For Public Policy

Term/Dates: TERM DATE
Class Time: Tuesdays & Thursdays, 12:30 pm - 1:45 pm
Location: CLASSROOM
Instructor: Ezra G. Goldstein (eggoldstein@gatech.edu)
Office Hours: T & R 2:30-3:30 or Zoom (via appointment)

This syllabus contains an overview of the nature of the course, the expectations for students, and helpful resources offered by Georgia Tech. Note that this syllabus is subject to change at the discretion of the instructor with reasonable notice (> 24 hours) provided to students. SUBJECT TO CHANGE.

Course Description

What do we know and how do we know it? This course introduces students to empiricism, the foundations of probability, and the concepts of statistics and econometrics. The primary emphasis will be on assembling the basic statistical toolkit for applied policy analysis and form a basis for asking and answering empirical questions. Students will be well equipped to carry forward the lessons of this class towards data translation, policy evaluation and analysis, and causal inference. The course is intended for students without prior knowledge of introductory statistics and those with a wide range of methodological backgrounds (e.g., mathematics, econometrics, and computer science).

Course Goals

The course will develop skills students can use to critically evaluate empirical questions and evidence, especially as they relate to public policy. To do so, the course introduces topics including, but not limited to:

- Probability Theory
- Descriptive Statistics
- Inferential Statistics
- Hypothesis Testing
- Regression
- Causal Inference (if time allows)

By the end of the course, it is expected that you will be able to:

- Identify and explain basic concepts in probability and statistics.
- Critically digest and describe both observational and experimental data.
- Interpret and critique key issues in public policy.
- Not be fooled by sensational headlines or videos on TikTok.

Texts and Prerequisites

Prerequisite(s): No prior courses are required. A strong grasp of algebra is recommended. Prior experience with any programming or coding language is suggested, but not necessary.

Required Reading: At least one textbook is required in this course as material meant to supplement the lectures. Readings, and their suggested completion timeline, will appear on the syllabus and occasionally be referenced in course update messages. There are two options for textbooks, each with their own strengths and weaknesses.

For a more detailed and technical supplement, I suggest Stock and Watson. This text spends considerably more space on the mechanics of regression and can be a useful source moving forward, however it provides only a brief overview of many basic concepts in statistics (e.g., probability, inference, and sampling). For a textbook that spends more space on the basics of statistics, I suggest Agresti and Finlay. This textbook provides many nice sociological applications, but it spends quite a lot of space on extraneous material. Moreover, its discussion of regression is lackluster and some of material is dated. Note that any edition of these books will suffice, please find one that suits your budget.

I also strongly suggest a copy of Wheelan's Naked Statistics. It is an incredibly approachable book that details the intuition of statistics. You can find it on amazon for quite cheap.

Required Text(s):

Introduction to Econometrics, 4th Edition
Authors: James Stock and Mark Watson; ISBN-13: 978-0134461991

Statistical Methods for the Social Sciences, 5th Edition
Authors: Alan Agresti and Barbara Finlay; ISBN-13: 978-0134507101

Naked Statistics
Author: Charles Wheelan; ISBN-13: 978-0393347777
This short book is suggested for the course. Readings will be suggested periodically to support the material in the course.

Computing Requirements

Students are expected to have regular access to a personal computer that meets Georgia Tech's minimum hardware and software specifications. Please review the Georgia Tech Required Computer Ownership policy for the current specifications. A computer is needed to access Georgia Tech and course websites, complete DataCamp coding courses, and complete programming exercises in RStudio.

Grade Distribution & Scale

The course consists of several assessments totaling 1,000 points.

Assessment	Points

Assessment	Points
Participation (DataCamp & Class Tracking)	150 (15%)
Four Computer Quizzes	200 (20% total / 50 points each)
Midterm Exam	300 (30%)
Final Exam	350 (35%)
Total	1,000 (100%)

Grading Scale:
>=900: A 800-899: B 700-799: C 600-699: D <=599: F

Assessment Descriptions

Participation: In-class participation and engagement will be assessed via online channels (such as Canvas or DataCamp tracking) or in-class via Kahoot! Games or random attendance checks. Participation as defined in this section accounts for 150 points (15% of your overall grade).

DataCamp: Throughout the course, students will learn productive tools to engage directly with data and produce sample statistics. To do so, our course will rely on DataCamp, a platform for programming, data science, and other modern coding skills. Students will be required to participate in semi-weekly DataCamp courses meant to bring students up to speed with R, a free statistical software.

Each DataCamp course is built upon prior courses, so it is important that students remain active with their DataCamp assignments. Students will receive credit for their DataCamp course participation by uploading certificates of completion on the Canvas page before the due date. *Late submissions will be accepted and eligible for reduced credit only for DataCamp*. Credit will be reduced by one percentage point for each hour late (e.g., if a student submits their certificate 12 hours late, they will earn 88 percent of the total available points).

Quizzes: Four computer quizzes will be assigned throughout the semester, each worth a total of 50 points (5% each, for a total of 20% of your overall grade). Although each Quiz will be different, the structure is the same. A Quiz consists of an at-home component and an in-class component. For instance, students may complete a problem set at home or prepare a printout of relevant material from a provided dataset (e.g., a histogram, a table of means, or regression outputs) using R. Students may then bring this printed output to class and use it to answer a set of interpretation questions under exam-like conditions. The four quizzes are explicitly designated as:

- Quiz 1: Probability Theory
- Quiz 2: Descriptive Statistics
- Quiz 3: Inference
- Quiz 4: Regression

Quiz dates will be announced with a week lead time. Makeup quizzes will not be provided unless a university-approved excuse is given.

Midterm Exam: The midterm exam will cover material from the first half of the semester and will count for 300 points (30% of your overall grade). The exam will be taken during the class period and materials (such as course notes, computers, or smart phones) will not be allowed. Calculators (including graphing calculators) can be used during the exam. The midterm will take place during class on Thursday, October 8.

Final Exam: The final exam will cover material throughout the semester, but will focus more on material we cover after the midterm, and will count for 350 points (35% of your overall grade). The exam will be taken during the assigned final exam period and materials (such as course notes, computers, or smart phones) will not be allowed. Calculators (including graphing calculators) can be used during the exam. The final will take place during the Institute final exam period, December 10-17.

Course Policies, Expectations, and Guidelines

Attendance: Although attendance is not required, it is strongly suggested. Participation, quizzes, and exams will be assessed in class, and no alternatives will be provided unless an Institute-approved exception is provided. For guidance on approved Institute activities and policies surrounding religious or unexpected absences, see the published guidance on approved exceptions.

Technology Policy: Laptops, smart phones, and tablets are permitted in class for the purposes of taking notes, recording lecture audio, and participating in class activities only. If you intend to use your device during class, please sit behind pen-and-paper students so that your screen does not distract the class. I have a two-strike policy for class interruption from digital audio: if your device makes noise in class, you must silence it immediately. If there is another infraction, you will surrender your device until the end of the class period. Please be respectful of your fellow students.

Large language models are a reality of today's learning environment. In my view, a student graduating from the Georgia Institute of Technology must learn to iterate with technology and incorporate cutting-edge tools into their workflow. As a result, I anticipate many students will find AI-tools to be useful in accomplishing code-based assignments. Still, be careful not to rely on generative models for answers to interpretation questions. Keep in mind that a significant portion of your course grade will be assessed without additional materials.

Collaboration Policy: I encourage you to work and study together. In my view, education is a collaborative effort from instructor to student, and student to student. However, keep in mind that quizzes and exams must be taken individually.

Office Hours: My office hours are a time dedicated to YOU! Please take advantage of them. I am here to help you succeed, not just in this course, but as a student here at Georgia Tech. Office hours are reserved to answer questions regarding course material, ask about empirical public policy, and think about big-picture questions that challenge society. I will always make myself available if you have a question about course material. If the office hours listed above are in conflict with your own schedule, I can meet (via appointment) over Zoom.

Institutional Policies

Academic Integrity: Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. Review Georgia Tech's Honor Code and the student Code of Conduct. Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities: If you are a student with learning needs that require special accommodation, contact the Office of Disability Services (404-894-2563) as soon as possible to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Student-Faculty Expectations Agreement: At Georgia Tech, we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. The Student-Faculty Expectations articulate some basic expectations that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

Campus Resources for Students: If necessary, there are a variety of resources available to you that can help you succeed in this class and at Georgia Tech. For help with academic issues, please see the Georgia Tech Library, the Communication Center, and the Center for Academic Success. For mental and emotional help, please reach out to the Counseling Center, the Division of Student Life, and the Women's Resource Center. For a full list of suggested student resources, see this document published by the Center for Teaching and Learning.

Core IMPACTS Statement: This is a Core IMPACTS course that is part of the Social Sciences area. Core IMPACTS refers to the core curriculum, which provides students with essential knowledge in foundational academic areas. This course will help master course content, and support students' broad academic and career goals.
• This course should direct students toward a broad Orienting Question: How do I understand human experiences and connections?
• Completion of this course should enable students to meet the following Learning Outcomes: Students will effectively analyze the complexity of human behavior, and how historical, economics, political, social or geographic relationships develop, persist or change.
• Course content, activities and exercises in this course should help students develop the following Career-Ready Competencies: Intercultural Competence, Perspective-Taking, Persuasion

Course Schedule and DataCamp Integration Outline

The weekly coverage may change depending on the progress of the class. Students must keep up with the assigned reading and DataCamp modules. Class meetings are Tuesdays and Thursdays. Weekly content is subject to change.

Week	Lecture Content	DataCamp Module	Stock & Watson Reading	Agresti & Finlay Reading	Assessments & Quiz Dates
1	Course syllabus; big picture: What is empirical policy analysis? Describing a sample.	1. Introduction to R	Ch. 1	Ch. 1; Ch. 2	
2	Probability theory: single variables and discrete distributions. Expectations and variance.	2. Intermediate R	Ch. 2.1-2.2	Ch. 3.1-3.3; Ch. 4.1-4.2	
3	Probability theory: continuous distributions and the standard normal. Joint distributions.	3. Introduction to the Tidyverse	Ch. 2.3-2.4	Ch. 4.2-4.7	
4	Conditional distributions, covariance and correlation, independence.	4. Data Visualization with ggplot2	Ch. 2.4-2.7		Thu Q1: Expectations, probability, conditional means
5	Descriptive statistics, deceptive statistics, sampling distributions.	5. Data Manipulation with dplyr; Introduction to Statistics in R, Ch. 1	Ch. 3.1-3.3	Ch. 5	
6	Central limit theorem and law of large numbers, statistical inference, and hypothesis testing.	6. Introduction to Statistics in R, Chs. 2-4	Ch. 3.4-3.6	Ch. 6	Thu Q2: Descriptive statistics, histograms, scatter plots
7	Tuesday: No class - Fall Break. Thursday: Midterm Exam.	None			Midterm Exam: Thursday, October 8
8	Comparing two groups and introduction to randomized controlled trials.	7. Sampling in R, Chs. 1-3		Ch. 7	
9	Bivariate relationships, correlation, and mechanics of OLS.	8. Hypothesis Testing, Ch. 1; Introduction to Regression in R, Ch. 1	Ch. 4	Ch. 9.1-9.3	Thu Q3: Statistical inference
10	Regression with one variable.	9. Introduction to Regression, Chs. 2-3	Ch. 5	Ch. 9.4-9.7	
11	Multivariate relationships and omitted-variable bias.	10. Intermediate Regression in R, Ch. 1	Ch. 6.1-6.3	Ch. 10; Ch. 11.1-11.4	
12	Multiple-regression inference, R-squared, and model fit.	10. Intermediate Regression in R, Chs. 2-3	Ch. 6.4-6.8	Ch. 13	
13	Categorical predictors: dummies and interaction terms.		Ch. 8		Thu Q4: Regression interpretation
14	Thanksgiving Break - no course meetings.	None			No class Tuesday or Thursday
15	Research design: internal validity, external validity, and causal inference.	None	Ch. 9		
16	Final exam review and course wrap-up.	None			Final instructional class meeting: Tuesday, December 8. Final exams: December 10-17.