ISyE 4031: Regression and Forecasting

Spring 2027, Section T09, CRN 26 101

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Required for BSIE | 3-0-3 credit hours

Class Meetings

- **Days:** Tuesday and Thursday
- Time: 9:30 to 10:45 a.m. Eastern Time
- Location: George Tower/Scheller Tower 0370
- **Delivery:** In person only. There is no remote attendance option.
- **Canvas:** https://canvas.gatech.edu/

Instructor and Teaching Team

Instructor: Prof. Shihao Yang

- Office: George Tower 1213
- Email: shihao.yang@isye.gatech.edu
- Office hours and teaching assistant information will be announced in Canvas.

Prerequisite

ISyE 3030, Basic Statistical Methods.

Official Catalog Description

The current Georgia Tech catalog describes ISyE 4031 as follows:

> Regression analysis: multiple linear regression, diagnostics, and variable selection. Forecasting: exponential smoothing techniques and autoregressive moving average models.

Catalog source: https://catalog.gatech.edu/coursesaz/isye/

Course Description

This section covers regression, basic statistical learning methods related to regression and classification, time series, and forecasting models, with applications in science and engineering. Building on the core course syllabus, some regression review is shortened because related material is covered in ISyE 3030. The released time is used to introduce statistical and machine-learning methods that every BSIE student should encounter.

Required statistical-learning topics include logistic regression, ridge regression, LASSO, principal component analysis, and support vector machines. K-nearest neighbors and K-means may be included as time permits. Forecasting topics include time-series regression, exponential smoothing, and autoregressive integrated moving average models, with seasonal modeling included if time permits.

Course Redesign and Learning Approach

This section uses AI coding agents in statistical computing. Students will use AI coding agents such as Codex, Claude Code, and Google Antigravity to plan, implement, debug, test, and explain statistical analyses. R examples from earlier course materials may be used for comparison.

The statistical ideas remain central. Students are responsible for understanding the models, checking assumptions, validating code and results, and communicating conclusions. An AI-generated answer is not evidence that an analysis is correct.

Learning Outcomes

By the end of the course, students should be able to:

1. Formulate real-life problems using regression, statistical-learning, time-series, and forecasting models.
2. Identify and collect appropriate data, and explain which variables and data structures are useful for a given problem.
3. Fit, diagnose, compare, and interpret regression and statistical-learning models, including logistic regression, regularized regression, PCA, and support vector machines.
4. Fit and assess time-series regression, exponential-smoothing, and ARIMA forecasting models.
5. Use AI coding agents to build reproducible analyses from real data.
6. Critically test AI-generated code, calculations, explanations, and citations.
7. Communicate model assumptions, uncertainty, limitations, and practical conclusions clearly.

Required Course Materials and Technology

The course requires the materials and technology below.

- **Required textbook:** Bowerman, O'Connell, and Koehler, Forecasting, Time Series, and Regression: An Applied Approach, 4th edition, ISBN-13 978-0-534-40 977-7. Abbreviated as FTR.
- **Recommended statistical-learning reference:** Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, An Introduction to Statistical Learning: with Applications in R, 2nd edition. Abbreviated as ISL. The electronic edition is available free at https://www.statlearning.com/. The Python edition available from the same site may also be used.
- **Laptop access:** Students need regular access to a laptop that they can bring to class and on which they have permission to install and configure course software.
- **AI coding-agent access:** Students need access to at least one supported local AI coding agent. No particular commercial vendor is required. Some tools offer free or educational access; supported setup and backup options will be maintained in Canvas. The course cannot purchase or reimburse individual subscriptions.

Assigned readings and any additional required files will be identified in Canvas.

Technology and AI-Agent Workflow

1. **Local AI coding agents:** The primary AI workflow uses agents that can work directly with project files, the terminal, packages, tests, Git, and generated outputs. Supported options may include Codex, Claude Code, Kiro, Google Antigravity, Cursor, and other suitable tools.
2. **Python:** The primary language for regression, statistical learning, forecasting, data analysis, and visualization.
3. **Git and GitHub:** GitHub repositories will be used for version control and selected course materials. GitHub Codespaces may be used for selected repositories or as a backup environment, but it is not the primary AI-agent environment.
4. **Setup guidance:** Account, authentication, installation, and backup-access instructions will be maintained in Canvas. Students who lack reliable laptop access, administrator rights, or access to a supported agent should contact the instructor privately as early as possible.
5. **Student responsibility:** Students must understand submitted work, verify all results, protect credentials and private data, and disclose AI use as required below.

Course Outline

Part I: Statistical and Regression Foundations

- Statistical review, including t-tests, F-tests, and p-values
- Review of simple and multiple linear regression
- Model diagnostics and residual analysis
- Variable selection and model building

Part II: Statistical Learning

- Logistic regression and classification metrics

- Ridge regression and LASSO
- Model tuning, cross-validation, and comparison
- Principal component analysis
- Support vector machines, including margins, soft-margin classification, kernels, tuning, and interpretation
- Optional topics as time permits: K-nearest neighbors and K-means

Part III: Forecasting

- Time-series regression
- Exponential smoothing
- ARIMA models
- Optional topic as time permits: seasonal ARIMA modeling

Students are responsible for material covered in class even if it is not explicitly listed in this outline.

Evaluation and Grading

Component	Weight
Homework	10%
Midterm exams (2, in person; 12.5% each)	25%
Course challenge: milestones and leaderboard results	15%
Final challenge presentation and oral defense	25%
In-class quiz participation	25%
Total	**100%**

Homework

- Homework will generally be assigned weekly or every other week and posted in Canvas.
- The lowest homework score will be dropped. After that drop, each remaining homework assignment contributes equally to the homework grade.
- Late submissions are not accepted unless an accommodation or prior written arrangement applies.
- Students may work in approved pairs when an assignment permits it. One submission may be made for the pair, but both names must be listed clearly.
- Submit through the Canvas assignment or designated course repository, as specified by the assignment.
- Requests for regrading must be submitted within one week after a grade is returned.

Midterm Exams

- Two midterms are planned as in-person, proctored, closed-book exams. Each midterm contributes 12.5 percentage points to the course grade.
- Two double-sided pages of notes are permitted unless Canvas states otherwise.
- Each exam will occur during the normal 75-minute class period.
- Midterm 1: date to be announced in Canvas. This exam will emphasize statistical review, simple and multiple linear regression, diagnostics, transformations, variable selection, and validation.
- Midterm 2: date to be announced in Canvas. This exam will emphasize ridge regression, LASSO, logistic regression, classification metrics, principal component analysis, support vector machines, tuning, and model comparison.
- A make-up or alternative exam for any other reason requires a prior arrangement and appropriate written documentation, except in an emergency.

Course Challenge: Milestones and Leaderboard Results

- Teams will usually contain two to four students, depending on enrollment.
- The instructor will specify a common data-analysis or forecasting challenge, similar in structure to a Kaggle competition. Teams may not substitute a self-selected project.

- Teams will submit predictions or other challenge results at announced milestones and receive feedback through a course leaderboard.
- This component is based on valid milestone submissions, progress demonstrated across the challenge, and final leaderboard results. The scoring thresholds and milestone schedule will be published with the challenge.
- No written project report is required.
- Team members normally receive the same challenge-results grade, subject to documented participation concerns.

Final Presentation and Oral Defense

- Each team will give a final oral presentation during one of the last two scheduled class meetings.
- The presentation should explain the team's modeling decisions, validation strategy, results, limitations, and use of AI coding agents.
- The instructor will ask questions of individual team members as an oral defense of the work. Every student must be able to explain, test, and critique the team's analysis and code.
- The presentation and team-level communication account for 10 percentage points. Individual responses during the oral defense account for 15 percentage points.
- No separate written report is required.
- Because individual understanding is evaluated, students on the same team may receive different grades for this component.

Final Assessment Policy

- The final course assessment is the final challenge presentation and individual oral defense, scheduled during the last two scheduled class meetings. Presentation dates will be announced in Canvas.
- The final leaderboard submission is due before the presentation period; the date will be announced in Canvas.
- There is no separate written final examination during the Registrar's final-examination period.
- If an Institute closure or other official requirement changes this plan, the revised assessment schedule will be announced promptly in Canvas.

In-Class Quiz Participation

- Formal attendance will not be recorded or graded.
- Short quizzes and check-ins will occur frequently during class and will be the basis of the participation grade.
- Each quiz contributes equally to this component. A timely, good-faith submission earns participation credit; quiz correctness is used primarily for feedback unless a quiz explicitly states otherwise.
- Students are responsible for following the announced procedure for documented absences or approved accommodations.

Attendance and Participation Policy

Formal attendance is not recorded or graded. However, short in-class quizzes and check-ins are frequent and form the participation grade described above. Students are responsible for course material and announcements missed during an absence. For Institute-approved absences, religious observances, illness, personal emergencies, voting, approved accommodations, or other circumstances covered by Georgia Tech policy, notify the instructor in writing as early as reasonably possible and follow the documentation process that applies. Excused absences will be handled without academic penalty through an appropriate alternative when needed.

Grading Scale

Letter	Percentage
A	90 to 100%
B	80 to 89%
C	70 to 79%
D	60 to 69%
F	Below 60%

Collaboration Policy

Discussion and the exchange of ideas are essential to academic work. Students are encouraged to consult with classmates when an assignment permits collaboration. Each student must still understand the work, be able to explain it, and follow the assignment's submission rules. Sources, including books, articles, websites, lectures, classmates, and AI tools, must be acknowledged when they materially contribute to the work.

Generative AI Policy

AI tools are an intended part of the learning process in this section. They may support brainstorming, model selection, code generation, debugging, testing, revision, visualization, and exploration. Their use does not transfer responsibility away from the student.

Unless an assignment states otherwise, submitted work must include a brief **AI Usage Statement** describing:

- Which AI tools were used
- When and for what purposes they were used
- Representative prompts or questions
- How the output was checked, revised, or rejected
- How the AI interaction shaped the submitted work

Students must be able to explain, test, and revise all submitted analyses. AI may not be used to fabricate data, results, citations, or sources; conceal a lack of understanding; impersonate another person; or bypass an assignment's learning objective.

AI output can be inaccurate, incomplete, biased, or insecure. Students must check calculations, code, claims, references, data handling, and model assumptions. Use of AI must comply with the Georgia Tech Academic Honor Code.

Academic Integrity

Students are responsible for knowing and following the Georgia Tech Academic Honor Code. Violations may be referred to the Office of Student Integrity and may result in serious course or Institute consequences.

Students should also review the Student-Faculty Expectations.

Acceptable Student Conduct

Students are expected to contribute to a respectful, inclusive, and professional learning environment. Communicate civilly, allow others to participate, avoid disruptive behavior, and treat classmates and the teaching team with respect. Conduct in class, on Canvas, in course repositories, and in AI-assisted work must follow the Student Code of Conduct (https://policylibrary.gatech.edu/student-life/student-code-conduct) and Georgia Tech's Student-Faculty Expectations.

Disability Accommodations

Georgia Tech provides appropriate academic accommodations for students with disabilities. Students seeking accommodations should contact the Office of Disability Services and provide the instructor with an approved accommodation letter as early as possible. Students are welcome to contact the instructor to discuss how approved accommodations will be implemented in this course.

Course Communications and Syllabus Changes

Canvas is the official channel for course announcements, assignments, submission instructions, grades, due dates, and schedule changes. Students are responsible for checking Canvas regularly and keeping Canvas notifications current. Topic pacing and assignment details may be adjusted as the redesigned course develops. Any syllabus or schedule change will be communicated promptly through Canvas, and the current syllabus PDF will be maintained in Canvas and at https://syllabus.gatech.edu/.