Regression Analysis

Statistical Foundations for Machine Learning and Financial Applications

Course Numbers:
16:960:563 — Regression Analysis
16:958:563 — For FSRM Program

Semester: Fall 2026
Instructor: Prof. Yifan Hu
Department of Statistics, Rutgers University

Course Description

Regression is a fundamental tool in statistics and a foundation of modern machine learning. This course develops regression methods from both statistical and machine-learning perspectives, with emphasis on estimation, inference, prediction, model evaluation, and generalization.

Topics include linear and logistic regression, model diagnostics and selection, regularization, nonlinear and nonparametric regression, dimension reduction, and resampling methods. Connections to modern machine learning—including prediction versus inference, overfitting, cross-validation, high-dimensional prediction, and regularization—will be emphasized.

Financial applications such as CAPM, factor models, portfolio/index tracking, and interest-rate modeling will be used throughout the course. **No prior knowledge of finance is required.**

Prerequisites

Level IV Statistics, for example Statistics 960:401 or 960:484, or equivalent preparation.

Students should be comfortable with basic probability and statistics, linear algebra, and programming. **R or Python** may be used for computational work.

Tentative Schedule

Date	Topic	Applications and ML Connections
Sep. 1	Statistics and Linear Algebra Review; Regression as a Learning Problem	Estimation vs. prediction; supervised learning; financial data

Date	Topic	Applications and ML Connections
Sep. 8	**No Class — University Policy**	Monday classes meet
Sep. 15	Simple Linear Regression	Least squares, inference; CAPM
Sep. 22	Multiple Linear Regression	Multifactor and Fama–French models
Sep. 29	Statistical Inference and Model Diagnostics	Inference, residual analysis, influential observations
Oct. 6	Prediction, Generalization, and Model Selection	Training/test error, feature selection, cross-validation
Oct. 13	Regularization and High-Dimensional Regression	Ridge, Lasso; portfolio/index tracking
Oct. 20	**Midterm Quiz**	Closed-book + LLM-assisted components
Oct. 27	LLM Hands-On Tutorial	Statistical reasoning, coding, debugging, and verification
Nov. 3	Logistic Regression and Classification	Probabilistic prediction and classification
Nov. 10	Nonlinear Regression	Financial and interest-rate applications
Nov. 17	Nonparametric Regression	Flexible prediction and bias–variance tradeoff
Nov. 24	Principal Components and Dimension Reduction	PCA, latent factors, financial factor models
Dec. 1	Bootstrap, Resampling, and Connections to Modern ML	Uncertainty quantification, model evaluation, review
Dec. 8	**Final Quiz**	Closed-book + LLM-assisted components

The schedule is tentative and may be adjusted depending on the pace of the course.

Grading

- **Midterm Quiz:** 30%
- **Final Quiz:** 40%
- **Course Project:** 30%
- **Bonus Homework:** up to 25%

Midterm Quiz — 30%

The midterm consists of:

- **2-hour closed-book component**
- **1-hour LLM-assisted component**

During the closed-book component, no notes, external resources, or AI tools are permitted.

During the subsequent LLM-assisted component, students may use LLM tools to **revisit and complete problems they were unable to solve during the closed-book component**, as well as to solve **new questions designed for AI-assisted problem solving**.

Students are responsible for critically evaluating and verifying all AI-generated reasoning and answers.

Final Quiz — 40%

The final Quiz will be held during the **last class meeting on December 8** and consists of:

- **1.5-hour closed-book component**
- **1.5-hour LLM-assisted component**

During the closed-book component, no notes, external resources, or AI tools are permitted.

During the subsequent LLM-assisted component, students may use LLM tools to **revisit and complete problems they were unable to solve during the closed-book component**, as well as to solve **new questions designed for AI-assisted problem solving**.

The LLM-assisted portion evaluates not only whether students obtain correct answers, but also whether they can use AI tools effectively while critically checking their outputs.

Course Project — 30%

Each student will carry out an individual data-analysis project using regression or related methods.

Projects may involve **financial data, machine-learning applications, or another substantive application area**. The emphasis is on appropriate modeling, analysis, evaluation, and interpretation.

Bonus Homework — Up to 25%

Challenging problems will be assigned as bonus homework.

Students are encouraged to use AI tools extensively when working on these problems, while critically evaluating and verifying all AI-generated work.

Texts and References

There is no single required textbook. Lecture notes and supplemental materials will be provided.

Useful references include:

- Hastie, Tibshirani, and Friedman, *The Elements of Statistical Learning*.
- James et al., *An Introduction to Statistical Learning*.
- Faraway, *Linear Models with R* / *Linear Models with Python*.
- Lai and Xing, *Statistical Models and Methods for Financial Markets*.

Collaboration and Academic Integrity

Students are encouraged to discuss general concepts and ideas with classmates, but submitted work must reflect each student's own understanding unless collaboration is explicitly permitted.

Rutgers University's Academic Integrity Policy applies to all work in this course.

Artificial Intelligence Tools

Modern generative-AI tools can be valuable for programming, debugging, exploring ideas, and checking understanding, but they can also produce convincing mathematical, statistical, and factual errors.

The October 27 class will provide hands-on instruction in the effective and responsible use of LLM tools for statistical analysis, programming, and model evaluation.

Students are responsible for the correctness of everything they submit and should critically verify AI-generated results.