CS 7200: Statistical Methods for Computer Science

Fall 2025

September 4, 2025

Location: Tue and Fri 9:50am-11:30am, Snell Library 049

Instructor: Olga Vitek, 177 Huntington Ave, 9th floor, o.vitek@northeastern.edu
Office hours Tue 11:30-12:30 and Fri 11:30-12:30, or by appointment.

Teaching assistant: Mr. Karna Mendonca, 177 Huntington Ave, 9th floor, mendonca.k@northeastern.edu
Office hours Tue 2-3pm and Thu 2-3pm, or by appointment.

Goals of the course: The course introduces methods of statistical inference, useful in any area of science that collects and analyzes data. The course discusses the methodological foundations, as well as issues of practical implementation and use. It includes a term project involving method development or evaluation, implementation and/or work with real-life investigations.

The course uses the context of linear regression models to discuss the following topics:

- **Univariate associations:** measures of association, basics of frequentist statistical inference (point estimation, sampling distributions, confidence intervals, prediction intervals, hypothesis testing), and benefits and limitations of p-values. If time allows, we will contrast frequentist inference with Bayesian inference for univariate associations.
- **Multivariate associations:** focus on multicollinearity and multiplicity of testing.
- **Causal inference in designed experiments:** foundations of statistical design and analysis of experiments, allocation of experimental resources to the study to maximize efficiency of the conclusions (such as factorial and randomized block designs). If time allows, we will introduce linear mixed effects models and response surface exploration.
- **Causal inference in observational studies:** graphical models, adjustments for confounders, interventions and counterfactual inference.

At the end of the course the students will be able to (1) recognize the problems of inferential nature and understand the underlying principles, (2) use statistical inference to design experiments and analyze data, and appropriately document the process, and (3) draw valid conclusions supported by the experimental design and data analysis, and clearly present the results.

Pre-requisite: Proficiency in linear algebra, probability and programming languages such as R or Python.

Software: Students can work on homework assignments and projects in R or Python. In-class examples of implementations of statistical methods will be provided in R.

Course web page: https://ovitek.github.io/CS7200/F25/index.html
Daily updates on the schedule, handouts and homework assignments will be posted on the course page.

Communication: The course will be using the discussion board Piazza
https://piazza.com/northeastern/fall2025/cs7200 You are encouraged to ask and answer questions on the discussion board. All important announcements will be made through Piazza. Once the course begins, course-related email inquiries will be left unanswered.

Textbook: The main textbook is
Kutner, Nachtsheim, Neter & Li (2005). *Applied Linear Statistical Models*, 5th Ed, McGraw-Hill.

Additional texts will be posted dynamically on Piazza.

Homework: Expect roughly 4 homework assignments during the semester. Solutions to the homeworks will be posted at the same time as the assignments. The homeworks constitute your own independent work. They do not need to be submitted, and will not be graded. They will not be part of the final grade. However do not hesitate to make use of Piazza and of the office hours for homework-related questions.

Quizzes: Expect roughly 12 quizzes assigned on random days during the class. The quizzes will be closed-book, paper and pencil, and will cover the recently discussed topics. They will take 5-10 minutes to complete. We will go over answers to the quizzes once the answers are submitted. Two lowest quizz grades of the semester will be dropped when calculating the final grade.

Exams: One in-class midterm, and one in-class final exam. You are welcome to bring to each exam one page of your personal notes with content of your choice. All re-grading of the exam requests should be made in writing, within **one week** after receiving the grade. The request should state the specific problem that needs to be re-graded, as well as a short (1-2 sentences) explanation of why re-grading is necessary. The new grade can potentially be lower than the original grade.

Project: At the end of the semester the students will perform a group project working with a real-world problem. See a separate document detailing project guidelines.

The project grade consists of project proposal (20%), interim project report (20%), final project report (oral 20% and written, 20%), and review of a project report prepared by another group (20%).

Projects having any degree of similarity with work by any other group, or with any other document (e.g., found online) is considered plagiarism, and will not be accepted. The minimal consequence is that all the group members will receive the project score of 0, and the best possible overall course grade will be C. Additional consequences are described at
`https://osccr.sites.northeastern.edu/academic-integrity-policy/`

AI policy: You are welcome to use AI for brainstorming the projects, and implementing the technical details. While it is ok to use AI for writing the project report, this may introduce mistakes and is therefore not recommended. You cannot use AI for reviewing the project prepared by another group.

Attendance: Since a lot of work happens in-person during the class, you are responsible for all the material discussed in your absence. There is no remote zoom access option, no recordings of the lectures, and no make-up options for the missed lectures and in-class quizzes.

Grades: All grades will be distributed via Canvas.

Breakdown of the final grade: The final grade is based on a total of 400 points broken down into quizzes (100 pts), midterm (100 pts), final exam (100 pts), and project (100 pts).

The final letter grades will follow the usual scale:

90-100% = A-range (i..e, A+, A or A-)
80-89% = B-range (i.e., B+, B or B-)
70-79% = C-range (i.e., C+, C or C-)
60-69% = D
0-59% = F
The cutoffs for '+' and '-' grades will be determined at the end of the semester, at the discretion of the instructor. This scale is subject to change at any time, at the discretion of the instructor.

Academic integrity policy: Academic integrity policy described on this website is strictly enforced
`https://osccr.sites.northeastern.edu/academic-integrity-policy/`

Changes to final course grade: Changes to the final course grade should be requested in writing, within **one week** after receiving the final course grade. The request should contain a technical explanation of

why re-grading is necessary. If the request is justified, the instructor will regrade **all the submissions**, including all the quizzes, the exams and the project, to determine the new grade. The new grade can be lower than the original grade.