

CS 7200: Statistical Methods for Computer Science

Fall 2024

September 16, 2024

Location: Tue and Fri 9:50am-11:30am, Ryder Hall 156

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: Mrs. Sarah Szvetcz, 177 Huntington Ave, 9th floor, szvetcz.s@northeastern.edu
Office hours Wednesdays 4-5pm and Thursdays 11am-12pm, 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, 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. Examples of implementations of statistical methods will be provided in R.

Course web page: <https://ovitek.github.io/CS7200/F24/index.html>

Daily updates on the schedule, handouts and homework assignments will be posted on the course page.

Attendance: Attendance is optional. However, you are responsible for all the material covered in class.

Communication: The course will be using the discussion board Piazza

<https://piazza.com/northeastern/fall2024/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 homeworks during the semester. Extensions to homework deadlines can be obtained if requested **at least 48 hours** before the deadline, and duly justified. Homeworks turned in after the deadline will not receive credit.

Answers to the homework problems should be submitted in a format that is easy to both read and reproduce. Computational problems can be done in any programming language (R preferred), and solutions should include the data file (.rda preferred), the source (.rmd preferred), and the output file (.pdf or .html).

Although some aspects of the homeworks can be discussed with your colleagues and on Piazza, and asking/answering questions on Piazza is encouraged, each homework should be done independently. A homework having any degree of similarity with that of another student (current or past, at Northeastern or outside) is considered plagiarism, and will not be accepted. The homework will be assigned a grade of 0. Additional consequences are described at

<https://osccr.sites.northeastern.edu/academic-integrity-policy/>

Exams: One in-class midterm, and one in-class final exam.

Grades: All grades will be distributed via Canvas.

Re-grades of homeworks and exams: All re-grading requests should be made in writing, within **one week** after receiving the grade. The request should state the specific question 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.

The project grade consists of project proposal (20%), project report (oral 30% and written, 30%), and project review (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/>

Breakdown of the final grade: The final grade is based on a total of 400 points broken down into homeworks (100 pts), midterm (100 pts), project (100 pts), final exam (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 homeworks, the exams and the project, to determine the new grade. The new grade can potentially be lower than the original grade.