

STA 440: Case studies in the practice of statistics

Spring 2021, Online, Duke University
Lecture: T/Th 10:15 – 11:30a; Lab: F 1:45 – 3:00p

Course overview

STA 440 is an intensive applied course that asks you to analyze timely real-world data across diverse domains in a principled, data-driven way. There may be more than one reasonable approach for any given situation, and you may be introduced to new material and techniques you haven't yet seen. Along the way, you'll work with a team of peers, develop critical thinking and communication skills, develop best-practices in version-control and reproducibility, and strive to become a creative and well-rounded practicing statistician.

| Instructor | E-mail | Office Hours |
|----------------|-------------------------|--|
| Yue Jiang | yue.jiang@duke.edu | See Sakai for time and code or by virtual appointment. |
| TAs | | |
| Anna Yanchenko | anna.yanchenko@duke.edu | See Sakai for time and code |
| Carol Wang | zhuoqun.wang@duke.edu | See Sakai for time and code |

Topics covered

This semester's case studies include:

- comparing the safety and efficacy of two drugs in preventing opioid-relapse following rehabilitation,
- evaluating whether associations exist between batting average and physical characteristics of professional baseball players, and
- examining whether there is still evidence of racial disparities in stop-question-frisk events and outcomes despite recent efforts at reform.

In addition, each student will conduct an individual case study project of their own interest and provide meaningful, detailed reviews and critiques of case projects from peers.

To develop your skills as a practicing statistician, throughout the course you will:

- solidify skills in reproducible research and programming, including version-control and collaboration via GitHub,
- critically think about reasonable analyses in the context of recent real-world data,
- express statistical models clearly and correctly,
- develop scientific writing skills by providing clear, concise, data-driven conclusions suitable for allied researchers, and
- effectively and concisely communicate results to peers.

This course recognizes the extraordinary circumstances that many students may be facing and have created the course policies and schedule to help reduce stress and promote effective course mastery. If you find that anything is affecting your wellbeing or academic success and you find it difficult to complete your work, please let me and your Academic Deans know as soon as possible. We are happy to work with you to make sure you succeed in STA 440. You may also find additional information regarding personal emergencies here.

Diversity and inclusion

It is my intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity and in alignment with Duke's Commitment to Diversity and Inclusion. Your suggestions are encouraged and appreciated; please let me know ways to improve the effectiveness of the course for you personally, or for other students or student groups.

Furthermore, I would like to create a learning environment for my students that supports a diversity of thoughts, perspectives and experiences, and honors your identities. To help accomplish this, if you feel like your performance in the class is being impacted by your experiences outside of class, please don't hesitate to come and talk with me. If you prefer to speak with someone outside of the course, your Academic Dean is an excellent resource. I (like many people) am still in the process of learning about diverse perspectives and identities. If something was said in class (by anyone) that made you feel uncomfortable, please talk to me about it.

Academic honesty

Academic honesty is of paramount importance in this class, and all work must be done in accordance with the Duke Community Standard, reproduced as follows:

To uphold the Duke Community Standard:

- I will not lie, cheat, or steal in my academic endeavors;
- I will conduct myself honorably in all my endeavors; and
- I will act if the Standard is compromised.

By enrolling in this course, you have agreed to abide by and uphold the provisions of the Duke Community Standard as well as the policies specific to this course. Cheating or plagiarism on assignments, lying about an illness or absence and other forms of academic dishonesty are a breach of trust with classmates and faculty, violate this Standard, and will not be tolerated; any violations will automatically result in a grade of 0 on the assignment, be reported to the Office of Student Conduct for further action, and potentially a failing (F) course grade depending on the magnitude of the offense.

Occasionally, datasets we are privileged to use in class are confidential and cannot be distributed more broadly or without express permission from the data-granting sponsor. Any unauthorized dissemination or further use of these datasets beyond this class is a violation of the Duke Community Standard.

Reusing code: You are always welcome to use online resources (e.g. StackOverflow) on your case studies. If you use code from an outside source, either directly or as inspiration, you must explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism and a violation of the Duke Community Standard.

On individual assignments, you may not directly share code or write up with other students. On team assignments, you may not directly share code or write up with another team. Unauthorized sharing of the code or write up will be considered a violation for all students involved.

Activities and assessments

Case studies: Each *group* case study (1, 2, and 3) will involve two submissions by the group: an initial submission, consisting of a written report, reproducible code on the GitHub repository, and a recorded oral presentation. This first submission will receive a grade of x . After receipt of comments from the instructor and classmates, groups will have the opportunity to write a response to review and submit a revised written report with the ability to earn up to half of the missing points on that component. Case study 0 will be an individual assignment and have only a single submission and peer review.

Groups will change throughout the semester. Individual contributions to each submission will be assessed, both by group members and by instructor assessment of the GitHub repository commit history. Team members must provide assessments in order to receive credit for an assignment as part of the group's peer evaluation process. Note that an individual team member's grade will be modified if peer evaluations indicate this is appropriate.

Individual peer reviews: For each group case study, you will be asked to formally review and assess reproducibility of two other group's assignments. This feedback will be provided to the groups and will also be assessed as part of the course. The ability to provide thoughtful, constructive feedback is critical in the workplace and is a valued skill. Peer reviewers will get feedback from the report authors regarding the helpfulness of their reviews. If you've never written a peer review before, don't worry – you will be provided resources, guidance, and ample opportunity to practice.

Individual project: Each student will complete an individual project as part of the course. The individual project should use data that have not previously been used by the student in a project, and the analysis should be entirely the student's own work. Any external resources used should be clearly documented. The student may use self-identified data or a resource provided by the instructor.

The individual project involves multiple due dates throughout the semester, to ensure students devote the required time and energy to their effort. These interim submissions will be reviewed by both the instructional team and peers, with the goal of maximizing the quality of the final report. Further details are available on the individual project page.

An additional resource (you can decide whether to opt in) available to you in Spring 2021 is participation in the Duke Reader Project. This is a great way to get help with your writing, which should lead to clearer reports and, I anticipate, higher quality work. There is no penalty for not participating; neither is there any accommodation in grading for those who do or do not participate.

Grade calculation

The grading basis for this class is a traditional letter grade according to the standard university policy. The following table presents the contribution of each component to a student's final grade:

| | |
|----------------------------|-----|
| Case study 00 (individual) | 10% |
| Case study 01 (group) | 15% |
| Case study 01 (group) | 15% |
| Case study 01 (group) | 15% |
| Individual peer reviews | 20% |
| Individual project | 25% |

Additionally, individual extra credit up to 3% of the semester grade will be available, with one individual opportunity granted for each of the three group case studies. A letter grade will be assigned as follows:

| | | | | |
|----|---|----|---|-----|
| 93 | ≤ | A | ≤ | 103 |
| 90 | ≤ | A- | < | 93 |
| 87 | ≤ | B+ | < | 90 |
| 83 | ≤ | B | < | 87 |
| 80 | ≤ | B- | < | 83 |
| 77 | ≤ | C+ | < | 80 |
| 73 | ≤ | C | < | 77 |
| 70 | ≤ | C- | < | 73 |
| 67 | ≤ | D+ | < | 70 |
| 63 | ≤ | D | < | 67 |
| 60 | ≤ | D- | < | 63 |
| 0 | ≤ | F | < | 60 |

I never "curve down." These posted cut points are guaranteed minimums. As well, this course is not graded to a pre-specified distribution (i.e., "curved"); if every student earns a 95 in the course, then every student will receive an A.

Regrade requests and late policy

Regrade requests must be made within two days of when a report is returned. These will be honored if points were tallied incorrectly, or if you feel part of your report is correct, but it was marked wrong (these things do happen!). No regrade will be made to alter the number of points deducted for an issue. When a regrade request is evaluated, if new errors are identified, additional points may be deducted from the grade.

The late policy for assignments is as follows: if work is turned in within 24 hours of the due date/time, then there is no penalty (essentially you have a 24-hour grace period). However, due to the fast-paced nature of this course, no late work will be accepted beyond this grace period. Do not treat the grace period as a "modified deadline."

Accessibility

Duke University is committed to providing equal access to students with documented disabilities. Students with disabilities may contact the Student Disability Access Office (SDAO) to ensure your access to this course and to the program. There you can engage in a confidential conversation about the process for requesting reasonable accommodations both in the classroom and in clinical settings. Students are encouraged to register with the SDAO as soon as they begin the program. Note that accommodations are not provided retroactively.

Where to find course materials

The course website will have an up-to-date course schedule, policies, and links to assignment introductions. Detailed assignment introductions will be made available on each assignment's GitHub repository in the course organization. Recorded lecture and lab session videos will be made available on Sakai via Zoom, with potential brief pre-lecture content videos on Warpwire prior to certain lectures. However, you are strongly encouraged to attend the live session. Announcements will be sent to the class by e-mail, so please check your e-mail regularly.

Course schedule

This schedule may be updated as the semester progresses with all changes documented here. All sessions will be recorded and posted on Sakai. Major case studies and semester milestones are documented on the schedule below. However, regular ungraded mini-assignments will be given and discussed in class and lab to check mastery and identify areas for additional focus. Assignments are due at 11:59 PM US Eastern time unless otherwise noted.

| Case 0: Stop-question-frisk | | | |
|-------------------------------------|------|--------|---|
| Week 00 | Thu, | Jan 21 | Data science ethics |
| Week 01 | Tue, | Jan 26 | How to write a model |
| | Thu, | Jan 28 | Data visualization |
| | Sun, | Jan 31 | Due: Case 0 write-up |
| Case 1: Opioid abuse relapse | | | |
| Week 02 | Tue, | Feb 02 | Survival analysis (1) |
| | Thu, | Feb 04 | Survival analysis (2) Due: Case 0 peer review |
| Week 03 | Tue, | Feb 09 | Clinical trials Due: Project proposal |
| | Thu, | Feb 11 | Case study work day |
| | Sun, | Feb 14 | Due: Case 1 write-up |
| Week 04 | Tue, | Feb 16 | Writing and responding to reviews |
| | Thu, | Feb 18 | Writing statistical analysis plans Due: Case 1 peer review |
| Case 2: MLB batting averages | | | |
| Week 05 | Tue, | Feb 23 | Estimation in linear models (1) Due: Case 1 response/revision |
| | Thu, | Feb 26 | Estimation in linear models (2) |
| Week 06 | Tue, | Mar 02 | The EM algorithm Due: Project intro/data/EDA |
| | Thu, | Mar 04 | Case study work day |
| | Sun, | Mar 07 | Due: Case 2 write-up |
| Week 07 | Tue, | Mar 09 | No classes held |
| | Thu, | Mar 11 | Model selection (1) Due: Case 2 peer review |
| Week 08 | Tue, | Mar 16 | Model selection (2) Due: Case 2 response/revision |

Course schedule (continued)

Case 3: Stop-question-frisk (revisited)

| | | | |
|---------|------|--------|---------------------------------|
| Week 08 | Thu, | Mar 18 | Spatial data and regression (1) |
| Week 09 | Tue, | Mar 23 | Spatial data and regression (2) |
| | | | Due: Project methods/SAP |
| | Thu, | Mar 25 | Case study work day |
| | Sun, | Mar 28 | Due: Case 3 write-up |
| Week 10 | Tue, | Mar 30 | Careers in statistical science |
| | | | Due: Project methods/SAP |
| | Thu, | Apr 01 | Project work day |
| | | | Due: Case 3 peer review |
| | Sun, | Apr 04 | Due: Project write-up |

Project presentations and review

| | | | |
|---------|------|--------|---------------------------------------|
| Week 11 | Tue, | Apr 06 | Project presentations |
| | | | Due: Case 3 response/revision |
| | Thu, | Apr 08 | Project presentations |
| Week 12 | Tue, | Apr 13 | Project presentations |
| | Thu, | Apr 15 | Project presentations |
| Week 13 | Tue, | Apr 20 | Project presentations |
| | Thu, | Apr 22 | Project presentations |
| | Fri, | Apr 23 | Due: Project peer review |
| Week 14 | Sat, | May 01 | Due: Project response/revision |