STAT 3202 - Introduction to Statistical Inference for Data Analytics

Autumn 2018 - The Ohio State University

Course Staff

Instructor: David Dalpiaz

- Email: `dalpiaz.14@osu.edu`
- Office: 114 Pomerene Hall
- Office Hours:
    - Monday: 3:00 PM - 4:00 PM
    - Thursday: 5:00 PM - 6:00 PM

Teaching Assistant: Guowei Li

- Email: `li.7230@osu.edu`
- Office: DALC, Pomerene Hall 151
- Office Hours:
    - Wednesday: 11:30 AM - 1:40 PM
    - Friday: Friday 11:30 AM - 1:40 PM

Grader: Ha Khanh Nguyen

- Email: `nguyen.1833@osu.edu`
- Office: DALC, Pomerene Hall 151
- Office Hours:
    - Tuesday: 9:10 AM – 11:20 AM
    - Thursday: 11:30 AM – 2:50 PM

Location and Time

The course will consist of two 80-minute lectures and one 55-minute recitation per week for a total of 4 credit hours.

- **Lecture:** Wednesday and Friday, 9:35 AM - 10:55 AM, Pomerene Hall (PO) 0250
- **Recitation:** Monday, one of:
    - 9:10 AM - 10:05 AM, 209 W 18th Ave (EA) 0295
    - 10:20 AM - 11:15 AM, 209 W 18th Ave (EA) 0295

Prerequisites

A grade of C- or better in Stat 3201: Introduction to Probability for Data Analytics, or permission of instructor.

Course Materials

Readings

- **Required Textbook:** *Mathematical Statistics with Applications, 7th edition*, by Wackerly, Mendenhall, and Scheaffer, Brooks/Cole, Cengage Learning, 2008.

- Additional reading material may be posted to the course website.

Computing

The R programming language (`https://www.r-project.org/`) and RStudio IDE (`https://www.rstudio.com/`) will be used for labs, projects, and some homework assignments. This software is available in the recitation classrooms. Both R and RStudio are free software that you are encouraged to install on your personal machines. Your TA will help you learn to use R for statistical analysis during recitation, but you should also expect to put in time outside of recitation doing data analysis with R for homework.

Website

- `https://go.osu.edu/stat3202`
- `https://carmen.osu.edu`

Course Description

The course covers foundational inferential methods for learning about populations from samples, including point and interval estimation, and the formulation and testing of hypotheses. Statistical theory is introduced to justify the approaches. The course emphasizes challenges that arise when applying classical ideas to big-data, partially through the use of computational and simulation techniques.

Learning Outcomes

Upon successful completion of the course, students will be able to...

1. Describe the role of a parameter in a statistical model and its relationship to observed data.
2. Use data to estimate and describe uncertainty about the parameters of a statistical model.
3. Translate scientific hypotheses about a population into mathematical statements about parameters in a statistical model.
4. Formulate statistical procedures to test a hypothesis about parameters in a statistical model, and interpret the results in both statistical and application-specific terms.
5. Explain the difference between statistical and practical significance in massive data settings.
6. Appreciate the effect of missing data on statistical inference.
7. Evaluate and compare different statistical procedures for answering the same question.

Grading

Assessments

There will be four types of assessments used in the course; homework, labs, projects, and exams. For each individual assignment or exam, specific policies and directions will be included with the release of an assignment, or in advance of an exam.

- **Homework:** There will be a total of 10 homework assignments, one for each topic of the course. The lowest two homework scores will be dropped from the recorded grades. Except for the first week, homework will occur regularly during any week there is not an exam or a project presentation, subject to minor variation with advance notice. Homework will be due at the beginning of class on the stated due date. Grading will be based on a combination of completion, correctness, and presentation. For full credit, final answers must be appropriately justified. The eight highest homework scores will account for a total of 20% of the overall grade, that is, 2.5% each.

- **Labs:** Lab exercises using `R` and R Markdown will be carried out in approximately half of the recitation sessions. These lab exercises will be documented using R Markdown and turned in via Carmen, and together account for 10% of the overall grade. Due dates will be announced as labs are released.

- **Projects:** There will be two data analysis projects, which will be completed in groups of 3-4 students. Each group will be provided with a data set and several questions related to it. The first project will focus on modeling and parameter estimation; the second project will focus on applying and extending the statistical methods learned in class. Students will present their work in class (10 minute presentations) and also submit a write-up of their results using R Markdown (approximately 3 pages of text, plus any additional figures and tables). Each project will be worth 10% of the overall grade.

- **Midterm Exams:** There will be two in-class midterm exams. Statistical tables will be provided as needed. Calculators may be used on the exams, however no devices with communication abilities are allowed, in particular, mobile phones may not be used. You may use one 8.5 x 11 inch sheet of paper (both sides), with whatever facts, formulas, or explanations you find helpful, for each exam. Each midterm will determine 15% of the overall grade. Like homework, to obtain full credit correct answers must be supported by appropriate justifications, calculations, or explanations. Incorrect answers supported by substantially correct calculations and explanations will receive proportionally appropriate partial credit. Correct answers, unsupported by calculations, explanation, or algebraic work will receive *no credit*. Midterm I will be given on Friday, September 21. Midterm II will be given on Friday, November 2.

- **Final Exam:** A cumulative final examination that accounts for 20% of the overall grade will be given during the University determined examination period of Thursday, December 13th, 8:00 AM - 9:45 AM. Except for the time allotted, the final exam policies are the same as those for the midterm exams.

Grade Components

Type	Percentage
Homework	20
Lab	10
Exam I	15
Exam II	15
Project I	10
Project II	10
Final Exam	20

Grading Scale

A	A-	B+	B	B-	C+	C	C-	D+	D	D-
93%	90%	87%	83%	80%	77%	73%	70%	67%	63%	60%

Course Policies

Email Communications

In order to protect your privacy, all course e-mail correspondence must be done through a valid Ohio State University name.## account. The subject line of the email must begin with [**STAT 3202**] followed by an informative message.

If an email is sent between 9:00 AM Monday and 11:59 PM Thursday, and you follow the above directions,

the staff will make an effort to respond within 24 hours. Homework questions sent the same day a homework is due run the risk of not receiving a response before the homework is due. Plan accordingly and consider office hours as a more reliable alternative.

Data Analytics Learning Center

Graduate teaching assistants (GTAs) for STAT 3201, **3202**, 3301, 3302, 3303 and 4620 will hold their office hours in the Data Analytics Learning Center (DALC) in Pomerene 151. The hours during which the GTA and grader for our course will hold office hours in PO 151 can be found at the top of the syllabus. You can meet with the GTAs for our course in the DALC during their office hours to discuss questions you have about the course material, homework assignments, `R`, etc.

You are welcome to stop by the DALC when it is open even if it is not currently being staffed by the GTA for our course, e.g. if you are looking for a place to study or work on an assignment for one of the supported courses. If the DALC is staffed by a GTA for another Statistics course when you stop by, they will help you if possible, but may not be able to answer all of your questions.

A complete list of hours during which the DALC will be staffed by GTAs for Statistics Department courses can be found at `https://data-analytics.osu.edu/dalc`.

In rare situations due to last minute emergencies, the GTA assigned to the DALC may not be able to attend their office hours. If the DALC is closed when the schedule indicates it should be open, we recommend waiting for a few minutes. If no one shows up in a reasonable amount of time, please email your instructor to let us know about the problem. You can also contact your GTA to see about arranging a make-up time to meet.

Grade Disputes

If you feel a homework, exam, lab, or project was graded incorrectly, you have **one week** from the date it was returned to discuss it with the course instructor. After one week, grading is final except for exceptional circumstances. You may not simply ask for a re-grade, but instead must justify to the staff why the grading was done incorrectly. Also, by disputing any grading, you agree to allow the course staff to review the entire assignment or exam for other errors missed during grading.

Makeup Exams

If you absolutely need a makeup exam and have a *valid excuse*, please contact the instructor to make the necessary arrangements as soon as possible. However, you must give notice in advance in such a situation. A make-up exam could potentially be perceived to be more difficult than a regularly scheduled exam since the instructor cannot guarantee that two exams containing different questions have exactly equal degrees of difficulty. Any make-up exam must be taken within a week of the missed exam. Exceptions to this policy will be permitted only in extreme situations such as serious injury immediately prior to an exam or severe illness requiring hospitalization.

Attendence

You are expected to attend all lectures and recitations. Failure to do so may not have a direct effect on your course grade, but will likely have a significant indirect effect. Any known or potential extracurricular conflicts should be discussed in person with the instructor during the first week of class, or as soon as they arise.

Academic Misconduct

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct `http://studentlife.osu.edu/csc/`.

Special Accommodations

The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: `slds@osu.edu`; 614-292-3307; `slds.osu.edu`; 098 Baker Hall, 113 W. 12th Avenue.

Policy Changes and Updates

Within reason and with advance notice, the instructor reserves the right to make any changes that are considered academically advisable. Such changes, if any, will be announced in class. Please note that it is your responsibility to attend the class and keep track of the proceedings. These changes are likely to contain minor changes to the schedule, however no changes will be made to the exam dates.

The Extended Syllabus

For some thoughts on teaching philosophy, some explanation of policies, and some general tips for success, please see The Extended Syllabus.

Topics and Readings

Topic	Title	Reading
T0	Introduction to Statistical Inference	Ch. 7, Sec. 8.1
T1	Point Estimation I	Sec. 8.2 - 8.3
T2	Point Estimation II	Sec. 9.3 - 9.5
T3	Point Estimation III	Sec. 9.6 - 9.7
T4	Confidence Intervals	Sec. 8.5 - 8.9
T5	Bootstrap	Handout
T6	Hypothesis Testing	Sec. 10.1 - 10.9
T7	Nonparametric Testing	Sec. 15.1 - 15.4
T8	Linear Statistical Models I	Sec. 11.1 - 11.4
T9	Linear Statistical Models II	Sec. 11.5 - 11.7
T10	Bayesian Inference	Sec. 16.1-16.4

Important Dates

Item	Due Date
Exam I	Friday, 9/21
Project I, Presentation	Wednesday, 10/3
Project I, Report	Friday, 10/5
Exam II	Friday, 11/2
Project II, Presentation	Wednesday, 11/28
Project II, Report	Friday, 11/30
Final Exam	Thursday, 12/13