STT 380: Probability & Statistics for Data Science

Michigan State University, Fall, 2020

Course Syllabus

Instructor Information:

Name: Leonard C. Johnson

Office: Work from Home. All meetings will be on zoom

Contacts: via Email - john2772@msu.edu

Office hours: Fri 3:00 – 5:00 p.m. or by Appointment (**On zoom**: link provided on D2L)

Teaching Assistant (Grader) Information:

Name: Sang Kyu Lee

Contacts: via Email - leesa111@msu.edu

Class Information:

Weekly Meetings: Mons & Weds; 3:00 p.m. – 4:50 p.m. **Place**: msu.zoom.us (Link on d2l)

Course Modality: Online – synchronous

Class Attendance: You are expected to attend all lectures throughout the semester. Because of the unprecedented nature of online classes because of the COVID-19 pandemic, I will drop up to 4 missed classes during the semester. If you miss a class for any reason, you will be responsible for materials covered, homework and deadlines for that class period. You can watch the recorded video of the lectures via D2L.

Webpage: All course materials and Homework will be posted on D2L or the department's webpage. Please download course materials from D2L before you attempt to read them. The D2L viewer does not work well with power point slides. Class recorded video will also be posted. Submitted work must be uploaded on D2L in PDF format only.

Required Textbook:

- Probability and Statistics for Data Science, by Carlos Fernandez-Granda. The textbook is freely available online at Carlos Fernandez-Granda's website
- **Prerequisite**: ((MTH 234 or concurrently) or (MTH 254H or concurrently) or (LB 220 or concurrently)) and (MTH 314 or concurrently) and STT 180, or equivalent courses.
- **Required Technology**: Access to high speed internet, TI calculator, R Software, CamScanner App (Andriod or Iphone), Webcam (for exam purposes).

Additional Readings (Not Required):

- Introduction to Statistical Thought by *Michael Lavine.* Available online
- Probability and Statistics for Data Science MATH + R + DATA, *Norman Matloff*, CRC Press

Course Description:

The course will cover Chapters 1-4, and Chapters 6-12 of the above-required textbook. We will also cover some special family of Distributions (Gamma, Beta, Exponential, etc)

Fundamental concepts and methods in probability and statistics from a data science perspective. We will make use of computers where they will help to improve our understanding of fundamental concepts of the course, or where they will perform tedious calculations quickly. We will use applications from a statistical package (mostly R).

Communication amid COVID

In remote learning environments, it is critical to communicate clearly and frequently with your instructors, especially if you encounter a prolonged period where you are unable to engage in course content. We have anticipated that some students may encounter interruptions to their studies for any number of reasons (e.g., illness, need to provide medical or child care, sustained loss of internet, etc.) and have plans in place accordingly.

Learning Continuity: If you must miss classes' or an exam because of ill health or other reasons, get in touch with me as soon as possible by email and make an appointment for a Zoom meeting. It will not be okay to come to me AFTER the exam about such issue. I will try to do my best to help if I am informed prior to missing an exam.

Course Continuity: In the unlikely event of my absence because of health reasons, the Department of Statistics has contingency plans to maintain continuity of the course. In such a case, substitute instructors will take over. Relevant information about substitute instructors etc. will be announced via email and posted on the class website on D2L.

Grading Policies:

Good News: Every student will start the class with a 4.0! I am giving everyone 1% of the grades free. I call this a **Kickstarter point**.

The following grading scale is for the remaining 99% of the final assessed grade in the class:

- **30%** of your grade will be determined by 2 in class midterm exams (15% each).
- **30%** of your grade will be determined by the homework.
- **25%** of your grade will be determined by the final exam

- **10%** of your grade will be determined by the quizzes
- **4%** of your grade will be determined by your class attendance, opening survey and syllabus quiz

Percentage	90-100	80-89	70-79	65-69	60-64	55-59	50-54	0-49
Grade	4.0	3.5	3.0	2.5	2.0	1.5	1.0	0.0

Homework:

There will be **seven (7) homework assignments** during the semester. The homework with your lowest score will be dropped. All homework are due on D2L by the specified due time & date. All homework will be graded and the due date will be posted on D2L when they are assigned. **Late homework will not be accepted and there will be no make-up for homework**. **Five (5) short quizzes** will be given as the instructor sees fit. There will be no announcements for upcoming quizzes. The quiz with your lowest score will be dropped.

Examinations:

There will be two midterm exams and a final exam. You will have enough time to complete your exam and submit them on D2L. All exams will be open notes, open book. You may be required to use zoom camera on the day of your exams. Please see our tentative schedule for exam dates.

How to submit homework and exam on D2L?

You can do this in one of three easy ways:

1. If you have a printer and scanner, you can solve your homework or exam (write in ink or type) and scan it then upload on D2L as pdf.
2. You can use **CamScanner App** to take pictures of your work (write in black ink) with your phone, convert to pdf and upload on D2L.
3. You can work on your ipad and save it as pdf, then upload on D2L

We will practice this with our opening survey on the first day of class.

How can I get help with my work?

You can get help in several ways. The first is to ask question in class or during office hours. You can also email questions to your class instructor or the grader.

Other way to get help is to visit the Statistics Learning Center (SLC) via zoom. There will be graduate students waiting during the listed times to help you with homework problems and general statistics questions. Please make use of the opportunity offered freely by the STT department.

Course Policies:

Classroom (Zoom) Civility

Please turn off your speaker when entering a zoom lecture. You can ask questions via zoom chat or you may stop me during lectures to ask questions if you think everyone might benefit from your question.

Feedbacks and Communications

- All course announcements will be made during video lectures or using D2L. Please check D2L regularly.
- Please DO NOT reply to mass emails sent via D2L, you should create an email with new subject line in case you wish to respond to mass emails.
- The best way to contact me is via my department's email : john2772@msu.edu . Please give me up to 24 hours for email response. Emails sent over the weekend may have a longer waiting time.

Academic Integrity and Honesty

The Department of Statistics and Probability adheres to the policies of academic honesty as specified in the General Student Regulations 1.0, Protection of Scholarships and Grades, and in the All-University of Integrity of Scholarship and Grades which are included in Spartan Life: Student Handbook and Resource Guide. Additional information can be found at https://honorcode.msu.edu

ADA:

- To arrange for accommodation, students with disabilities should contact the Resource Center for People with Disabilities.
- Resource Center for Persons with Disabilities(RCPD): Telephone: 355-9642; Website: http://www.rcpd.msu.edu
- Once you obtain a RCPD VISA please contact the instructor via email to arrange for a meeting to discuss your accommodation options.

Disclaimer

The instructor reserves the right to make small or large scale changes to this syllabus as he sees academically appropriate without any consultation. If and when such changes occur, the instructor will make a formal announcement to the class via multiple media.

IMPORTANT DATES FOR FALL, 2020

Wed, 09/02	Class begins
Mon, 09/07	Labor day (No class)
Wed, 09/09	Open adds end (8:00pm)
Mon, 09/28	Last day to drop with refund (8:00pm)
Wed, 10/14	Midterm exam 1
Wed, 10/21	Middle of the semester
Wed, 10/21	Last day to drop with no grade reported (8:00pm)
Wed, 11/18	Midterm exam 2
Fri, 12/11	Class ends
Mon, 12/14	Final Exams

Tentative Weekly Schedule

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each week, and also as a study guide before each exam, and at the end of the semester.

Week 00, 09/02 – 09/04: Syllabus review, reopening survey and Data Science and Statistics, Intro to R

Week 01, 09/07 – 09/11: Probability space, conditional probability and Independence (**Mon 09/07, No class**)

Week 02, 09/14 - 09/18: Discrete and Continuous Random Variables

Week 03, 09/21 - 09/25: Joint and Marginal distribution

Week 04, 09/28 - 10/02: Covariance and Independence (Review)

Week 05, 10/05 - 10/09: Expectation, Variance and Covariance

Week 06, 10/12 - 10/16: Random Samples (iid) & **Exam 1 (Wed, 10/14, 3:00 – 4:15 p.m., zoom)**

Week 07, 10/19 - 10/23: Convergence, CLT and Law of large numbers

Week 08, 10/26 - 10/30: Summarizing data; Frequentist statistics

Week 09, 11/02 - 11/06: Intro to Bayesian Statistics

Week 10, 11/09 - 11/13: Intro to Bayesian Statistics (cont'd)

Week 11, 11/16 - 11/20: Hypothesis testing & **Exam 2 (Wed, 11/18, 3:00 – 4:15 p.m., zoom)**

Week 12, 11/23 - 11/27: Hypothesis testing (cont'd)

Week 13, 11/30 - 12/04: Linear Regression, MLE and Inference

Week 14, 12/07 - 12/11: Linear Regression, MLE and Inference (Cont'd)

Week 15, Monday 12/14: Final Exam (03:00 – 5:00 p.m., In class, zoom)