

SI 544 - Introduction to Statistics and Data Analysis

Fall 2020: Fridays 10am-noon via Zoom

Note: Some syllabus details may be subject to change.

Last updated: 9/3/20, 11:33am
(All times are Ann Arbor local time.)

Instructor: Dr. Chris Teplovs (cteplovs@umich.edu)

Office hours: Fridays, 2:30-3:30pm, via [Zoom](https://umich.zoom.us/j/95803250378) (<https://umich.zoom.us/j/95803250378>)

GSI: Shriti Raj and Linfeng Li

Office hours: Monday, 8:30 pm - 9:30 pm, via [Zoom](#)

Thursday, 8:30 pm - 9:30 pm, via [Zoom](#)

Tuesday, 11:00 am - noon, via [Zoom](#)

Communication

The best way to get answers to your questions is to use Piazza (via Canvas). You'll find that you'll get the fastest responses, often from your peers, by using Piazza.

If you have questions about course material, homework, or labs, please feel free to come and talk with us during office hours. Your GSIs should be your first choice for technical questions; conceptual questions are best directed to Dr. Teplovs.

Personal matters should be communicated via email to Dr. Teplovs.

Quick Links

[Course Description](#)

[Learning Objectives](#)

[Competency:](#)

[Literacy:](#)

[Awareness:](#)

[Textbooks](#)

[Schedule/Readings](#)

[Class format](#)

[Pre-recorded Lecture Videos](#)

[Discussion Sections](#)

[Attendance](#)

[A note about scheduling conflicts](#)

[Readings](#)

[Giving and Receiving Assistance](#)

[Google/StackOverflow/etc.](#)

[In-class Notebooks](#)

[Homework Assignments](#)

[Quizzes](#)

[A note about programming experience](#)

[Grading](#)

[Late Policy](#)

[Original Work](#)

[Accommodations for Students with Disabilities](#)

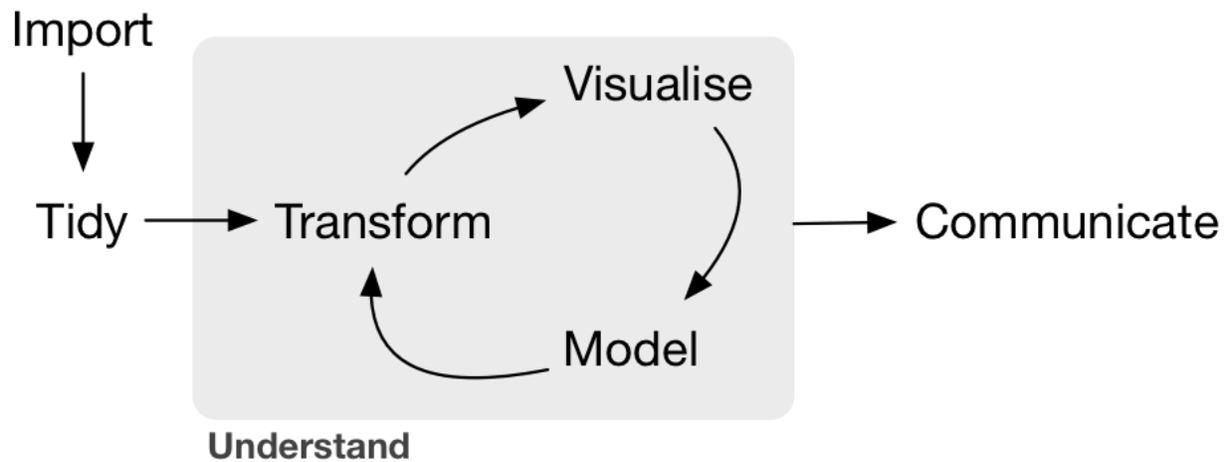
[Student Mental Health and Wellbeing](#)

Course Description

The purpose of this class is to provide students with a basic understanding of statistics, that is the ability to describe various populations and draw inferences about them. It should also sharpen individual intuition about how to read data, interpret data, and judge others' claims about data. Specifically, skills to be developed include the abilities to:

1. characterize population data intuitively for oneself and others,
2. draw conclusions and novel inferences from population data,
3. check assumptions of others' claims and debug their putative "facts", and
4. look for correlations while controlling for confounding effects.

Overall, the chief objective is to demystify the process of inquiry into what isn't known by following this data science pipeline:



Learning Objectives

Competency:

- Importing and tidying data
- Data visualization
- Summary statistics
- Probability
- Simple linear regression
- Linear regression using categorical variables

Literacy:

- Confidence intervals
- Hypothesis testing
- Chi-squared tests
- Regression diagnostics
- Interactions
- Multiple regression
- Using the R programming language

Awareness:

- Confidence Bands and Prediction Intervals

Textbooks

There is one free, open-source online textbook that we will be using throughout the course: <https://moderndive.com/>. The textbook is tightly coupled to the course and we will be following it quite closely. It is in your best interest to read the relevant chapters before class. You are not expected to purchase the printed textbook.

Schedule/Readings

Please see the [Schedule of Topics and Readings](#).

SI 544 FA 2019 Schedule of Topics & Readings

Week	Friday	Topic	Readings (Moderndive Book Chapter(s) and/or other materials) to be completed <i>before</i> Lecture	In-class Notebook Due? (Saturday 12:00pm (noon))	Problem Set Due (Friday 10am) (Note: Problem sets are released one week before they're due)
1	9/4	Introduction to Statistics, Data Science, and the course	None	No	No
2	9/11	Visualization	1 & 2	Yes	No
3	9/18	Data Wrangling & Tidying	3 & 4	Yes	No (change proposed 9/11)
4	9/25	Descriptive Statistics & Basic Probability	3, 8.5.2	Yes	Yes (PS#1) (change proposed 9/11)
5	10/2	Regression I	5	Yes	Yes (PS#2)
6	10/9	Regression II	5 (again)	Yes	No
7	10/16	Multiple Regression	6	Yes	Yes (PS#3)

8	10/23	Sampling & Confidence Intervals	7 & 8	Yes	No
9	10/30	Hypothesis Testing (t-test & ANOVA)	9	Yes	No
10	11/6	Inference for Regression	10	Yes	Yes (PS#4)
11	11/13	Storytelling with Data	11	Yes	No
12	11/20	"Traditional" statistics	TBA	No	No
Week 13: Thanksgiving recess - no classes					
14	12/4	Wrap-up		No	Yes (PS#5)

Class format

We meet synchronously via Zoom once a week in lecture on Fridays and once a week in discussion sections. During the Friday synchronous sessions, activities will include live coding, question-and-answer sessions, and time dedicated to applying principles learned from recorded lectures and readings to real-world data. These sessions will be recorded and made available shortly after the class ends. Students will be expected to hand in their work from the synchronous sessions within 24h of the end of the class.

Pre-recorded Lecture Videos

You will have access to pre-recorded video lectures approximately a week prior to the topics being covered in class. We had planned to set things up so that many of these videos would contain embedded quizzes, which provide both an opportunity to check your understanding as well as contribute points to your overall score in the course. Based on what we've seen in the first week of classes, we will need to modify our plans to accommodate technical problems with the MiVideo system.

Discussion Sections

Discussion sections are designed to give you additional opportunities to learn about applying the tools and techniques covered in the lecture material and synchronous sessions to novel datasets and scenarios. They are not meant to be remedial sessions for poor lectures (that's Chris's problem to deal with). You will meet weekly with approximately 23-25 other students and your GSI via Zoom. We would like you to attend only your assigned section. If you need to miss a discussion for a good reason, please contact your GSI as soon as possible to make arrangements.

Attendance

Attendance in the class and participation in lecture and discussion sections is mandatory. Repeatedly missing class (> 3 times, unexcused) or failing to participate will likely lead to a failing grade.

A note about scheduling conflicts

Some of you have approached me asking if you can still take this course if you have a time conflict with another one. Whereas SI 544 is a remote course this semester due to COVID, we will meet via Zoom from 10:00am to noon on Fridays starting on September 4th and ending on the week before the Thanksgiving break. We will resume our online meetings after the break. As mentioned above, students will be expected to hand in their work from the synchronous sessions within 24h of the end of the class so you will need to commit to doing the work on your own later that same day. The challenge for you will be to complete the work without having the opportunity to interact with your peers and the teaching team.

My general sense is that missing the Friday sessions will make the class more challenging than it has to be, but the decision is ultimately yours and I'll try to support you no matter what you choose to do.

Readings

There are readings assigned almost every week. These are intended to supplement the synchronous classes. You will get much more out of class if you read these before you attend. To encourage you to complete the readings before class, most classes will start with a brief quiz based on the readings. You can accumulate points toward your overall score in the course by completing these quizzes.

Giving and Receiving Assistance

SI106 has this policy and I think it's appropriate here as well (slightly modified from the SI106, W14 syllabus). Learning statistical methods is often challenging. We are going to cover a wide range of topics in the course and we will move quickly between topics. Because it is our goal for you to succeed in the course, we encourage you to get help from anyone you like.

All that said: *You* are responsible for learning the material, and you should make sure that all of the assistance you are getting is focused on gaining knowledge, not just on getting through the assignments. If you receive too much help and/or fail to master the material, you will crash and burn later in the semester.

The final submission of each exercise must be in your own words. If you receive assistance on an assignment, please indicate the nature and the amount of assistance you received. If the assignment is computer code, add a comment indicating who helped you and how. Any excerpts from the work of others must be clearly identified as a quotation, and a proper citation provided (e.g., in the comments of the code if it is a code fragment you have borrowed).

If you are a more advanced student and are willing to help other students, please feel free to do so. Just remember that your goal is to help teach the material to the student receiving the help. It is acceptable for this class to ask for and provide help on an assignment via the Q&A site, *including posting (short!) code fragments*. Just don't post complete answers. If it seems like you've posted too much, one of the instructional staff will contact you to let you know, so don't worry about it. When in doubt, err on the side of helping your fellow students. To reiterate, the collaboration policy is as follows. Collaboration in the class is allowed (and even encouraged) for assignments – you can get help from anyone as long as it is clearly acknowledged. Collaboration or outside help is not allowed on exams, though you will be allowed to use some materials that you bring with you. Use of solutions from previous semesters is not allowed. The authorship of any assignments must be in your own style.

Google/StackOverflow/etc.

Use these! Seriously... make an effort to discover the answers on your own by using the Web. You'll be surprised how often you can find related code to help you. **Give us the URLs of where you found help and even if you don't totally solve the problem, if it looks like you were heading the right direction in finding the solution we'll give you some credit!**

In-class Notebooks

Our synchronous "lecture" meetings are centered around co-constructed R markdown notebooks. I write these notebooks to provide an opportunity for you to practice applying the techniques we are learning. At the end of each synchronous class you are expected to turn in your notebook with the completed exercises. Whereas you may work in groups during the hands-on segments, you will be required to submit your own notebooks for credit.

Homework Assignments

Homework assignments (Problem Sets) will be due no less than five days after they are released (see late policy below). Our goal is to release the homework assignments by 6pm Fridays. You can expect that some of the material in the Problem Sets will be covered in class so it's in your interest to do as much of the class work as possible and make sure you understand the material. The goal of the assignments is to make sure that you understand how to apply the material in class to practical/real-world problems. We will try to provide you with real or realistic data and problems as much as possible -- things you would actually encounter if you were doing data exploration in a professional context.

See above for information on Giving/Receiving help. Copying wholesale (and/or failing to acknowledge your source) will be considered cheating and will be turned over to the academic advising office. Note that I am extremely good at catching people doing this. If you're good enough to hide the fact that you did it, you probably are better off just doing the assignment.

Quizzes

There will be multiple opportunities to accumulate points via quizzes. Quizzes will come in several different forms: in-video quizzes embedded in the recorded lecture videos, beginning-of-class online quizzes, and discussion quizzes. These are meant to be opportunities to demonstrate that you have done the preparatory work required to benefit from the synchronous times we have together. They are designed to be easy if you've done the required prep (and difficult if you haven't). Note that there will be more points available via quizzes than you can accumulate. You can collect a maximum of 100 points, but there will be some number greater than 100 available to you. The exact number is to be determined.

A note about programming experience

In this course we will use a programming environment called 'R'. For those of you without programming experience this might invoke a sense of dread and fear. My assumption is that none of you has any programming experience and that this is all new to you. Whereas this might be a bit boring for those of you who have previous experience programming, I would rather make sure that no one is left behind. You will not be thrown into a new programming environment and left to learn it yourself or fail. The GSIs and I will work with you to make sure you get the basics down in the first few weeks of the course. If you're struggling with programming concepts, please know that you are probably not alone and that the teaching team is happy to help you... but you have to reach out and let us know!

Grading

In-class notebooks (10 notebooks): 250 points
Discussion section work (10 meetings): 250 points
Problem sets (5) : 400 points
Quizzes: up to 100 points

In-class notebooks are completed during the lecture period and must be submitted within 24 hours of the end of class.

Discussion participation is an assessment of the degree to which you participate in discussion section activities including preparatory work, if any.

Problem sets give you the opportunity to apply concepts covered in class to new situations and datasets.

Quizzes test basic knowledge of material covered in the readings and videos. We anticipate making a total of approximately 120 points available, of which you can accumulate up to a maximum of 100 points, so don't worry if you mess up a few quizzes.

The mapping to final grades is:

950	A+
900	A
870	A-
850	B+
800	B
770	B-
750	C+
700	C

650	C-
620	D+
600	D
570	D-
<570	E

My advice to you is to not get too hung up on grades. Yes, your GPA is important. The best things that you can do to ensure a good grade in the course is to show up for class, read the assigned material before class, review your notebooks after each class, and try to have fun with the material. If you do all that the grades should take care of themselves.

Late Policy

I realize that the occasional crisis might screw up your schedule enough to require a bit of extra time in completing a course assignment. Thus, I have instituted the following late policy that gives you a limited number of flexible “late day” credits.

You have three (3) free late days to use during SI 544. One late day equals exactly one 24-hour period after the due date of the assignment (including weekends). No fractional late days: they are all or nothing. Once you have used up your late days **25% penalty** for each subsequent 24h period after the deadline that an assignment is late. For example, if the due date is 10am Friday, with no late days left, penalties would be:

Before 10am Saturday: 25% deduction

Before 10am Sunday: 50% deduction

Before 10am Monday: 75% deduction

After 10am Monday: 100% deduction

You don't need to explain or get permission to use late days, and we will track them for you. In cases where late days can be assigned in multiple ways (e.g. you have only one late day left but hand in two late assignments) we will always allocate late days in a way that maximizes your grade. Note that resubmissions after the deadline will be counted as late submissions.

Original Work

See above for information on Giving/Receiving help. Unless otherwise specified in an assignment, all submitted work must be your own, original work. You may discuss general approaches with others on individual assignments, but may not copy code or answers wholesale and must indicate on your turned-in assignment who you worked with and how. Any excerpts from the work of others must be clearly identified as a quotation, and a proper citation provided. Any violation of the School's policy on Academic and Professional Integrity will result in severe penalties, which might range from failing an assignment, to failing a course, to being expelled from the program, at the discretion of the instructor and the Senior Associate Dean for Academic Affairs.

One caveat in regards to this course or any other related class. You may not submit the same work to both classes for your final project. We are aware of the students in both and while it is ok to broadly tackle the same problem, the work you turn in for both classes must be significantly different. If you're not sure what that means, come talk to us.

Accommodations for Students with Disabilities

If you think you need an accommodation for a disability, please let us know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way we teach may be modified to facilitate your participation and progress. As soon as you make us aware of your needs, we can work with the Office of Services for Students with Disabilities (SSD) to help us determine appropriate accommodations. SSD (734-763-3000; <http://www.umich.edu/sswd/>) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. We will treat any information you provide as private and confidential.

Student Mental Health and Wellbeing

The University of Michigan is committed to advancing the mental health and wellbeing of its students, while acknowledging that a variety of issues, such as strained relationships, increased anxiety, alcohol/drug problems, and depression, directly impacts students' academic performance.

If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and <https://caps.umich.edu/> during and after hours, on weekends and holidays or through its counselors physically located in schools on both

North and Central Campus. You may also consult University Health Service (UHS) at (732) 764-8320 and <https://www.uhs.umich.edu/mentalhealthsvcs> , or for alcohol or drug concerns, see www.uhs.umich.edu/aadresources .

For a more comprehensive listing of the broad range of mental health services available on campus, please visit: <http://www.umich.edu/~mhealth/>