

STAT 113: INTRODUCTION TO STATISTICS (FALL 2020)

CONTACT AND LOGISTICS

Instructor. Colin Reimer Dawson (*he/him*)

Email. cdawson at the college domain

Course Website. <http://colindawson.net/stat113/>

Course Slack Workspace. <https://stat113f2020.slack.com> (join via invite link to email)

Class Zoom ID: (see email/Blackboard)

Hybrid Structure. Once the college resumes in-person instruction, Monday and Friday classes will be held in King 106 for those able to attend, with live Zoom attendance available for students who are not on campus or otherwise unable or uncomfortable attending in person. This is a large lecture hall with enough capacity for everyone enrolled to physically attend while maintaining social distance.

Lab Days. Half the class will be assigned to attend a virtual lab section on Wednesdays via Zoom, where small groups of 2-3 students will work together in “breakout rooms”, with the other half attending on Thursdays. By splitting the class this way, it will keep the number of breakout rooms to 5-7 at a time, allowing me to circulate.

“Flipped” Classes. There will be one additional “asynchronous” class each week consisting of a prerecorded video lecture posted between Monday and Friday, which will be accompanied by a short list of discussion prompts to be submitted via the class Slack workspace.

Subject to Change Statement. The meeting modalities may change, either on an ongoing basis (if the college’s policies change), or potentially for the occasional class here or there, depending on instructional objectives.

Office Hours (via Zoom).

Mondays 1:30-2:30pm (drop-in group office hour)

Wednesdays 11:30am-12:20pm (by appt)

Thursdays 11:30am-12:30pm (drop-in group office hour)

Fridays 1:30-2:30pm (by appt)

Date: Last Revised September 13, 2020.

GENERAL LEARNING OUTCOMES

By the end of this course you should be equipped to...

1. **critically evaluate statistical claims** presented in popular media
2. **recognize and avoid common errors** in interpreting the reported results of a statistical analysis
3. **determine what analyses are appropriate** for answering a particular question
4. **use standard statistical software** to obtain basic **numerical and graphical summaries** of a dataset
5. use standard statistical software to **carry out appropriate procedures of statistical inference**, and **interpret the results in the context** of the motivating real world problem
6. **communicate the findings and limitations** of a statistical analysis to an audience without specialized training in statistics

EMPHASES AND MISCONCEPTIONS

This is a **statistics** course and *not* a **math course**.

The point of the course **is not**

- learning how to calculate things
- memorizing terms and definitions
- proving mathematical properties of statistical tools

It **is** about

- understanding what data is (and isn't!) telling you
- why it is (and isn't) telling you that
- communicating about that process

Students taking this course have often reported that it feels like an unusual hybrid between a “STEM” course and a humanities course, requiring both quantitative and qualitative thinking, as much of the assigned work consists of **written sentences of English explaining thought processes and justifying conclusions**, while relatively little consists of manipulating mathematical equations.

Some of the *concepts* in statistics can be quite subtle, even while the calculations involved are relatively straightforward. We will devote most of our energy (in class and on assignments) to **understanding those nuances, and examining the thought processes** involved.

WHO SHOULD TAKE THIS COURSE

This is intended as a first course in statistics, and does not assume any previous exposure to the subject. It is likely to be a useful course even if you have taken statistics in high school, since we will be delving deeper into the concepts than most high school AP classes; however, **if you have a score of 3 or above on the AP stats exam or have a strong math background, you should take STAT 205 instead**, which starts in the same place but moves at a faster pace.

GRADING

A focus on grades can get in the way of learning, as jumping through hoops needed to achieve a desired grade tends to impair deep thinking. That said, **as long as they take a back seat to a desire to genuinely understand and grow**, grades can sometimes provide a useful bit of concrete feedback, and external motivation when one's internal motivation flags (often due to busyness).

The grading system in this course is non-traditional, and is designed with the goal of **retaining the formative and motivational role grades can play, while minimizing their evaluative role**.

See the handout called "Grading System" for a description of how this works. Be forewarned, **if you are someone who focuses heavily on monitoring your quantitative performance and comparing it to your peers, and stresses out over imperfection, we are likely to frustrate each other**.

However, if you like to keep your focus on thorough understanding and growth, and embracing and learning from mistakes, my hope is that you will find that the grading atmosphere in this class supports that approach.

COMPUTING

Throughout the semester, the lab component of the class will introduce computational techniques (using the statistical language R) to offload the tedious work of counting, arithmetic, plotting, etc.

The primary goal of the computing tools is to free up mental energy for a focus on concepts. That said, **being able to translate your goals into terms explicit enough for a computer is an important skill** in itself, and you will likely devote a significant chunk of time and effort to doing that.

Although the details and syntax of the commands you learn in this class are just one way of doing what they do, and will likely be superseded in the future, my hope is that you will **learn to think in computational terms**, an incredibly useful skill in this day and age well beyond this class.

I do not expect students to come in with any programming skills, and will provide instruction from the ground up, but **you will need to be willing to learn a little programming**, and to refer back to the reference material I provide. You should not expect to internalize everything the first time, and you will never be asked to produce computer code without having access to reference material, but **making an effort to understand some of the internal logic of the commands you are typing will allow you to generalize across examples**, reducing the cognitive load required to figure out what to do, much as knowing some phonics makes writing easier than it would be if you had to memorize every word individually.

MATERIALS

Textbook. The textbook is *Statistics: Unlocking the Power of Data*, by Lock, Lock, Lock, Lock and Lock (it's a "family textbook"), or "Lock 5" for short. Section numbers and exercises refer to the second edition, which is an improvement, in my opinion, over the first.

Note: We will be using an electronic platform called Wiley Plus that supplements the textbook with some interactive exercises and other resources. This is new this semester, so there will probably be some kinks that need to be worked out. But you will need to purchase a version of the book that grants access to the WileyPlus platform. You can either get electronic access only, which has ISBN 9781119491286, or the version that has electronic access plus physical pages, which has ISBN 9781119600657. Both of these should be associated with the class via the bookstore, but you only need one or the other.

Software. We will use the free statistical computing environment RStudio, which is an interface to the free and open-source language R. Once I set up an account for you, you can access the software via Oberlin's RStudio server via a web browser (rstudio.oberlin.edu).

The software is free, and so if you want you can also install R and RStudio on your personal machine (www.r-project.org and www.rstudio.com, respectively); however, this will require a bit more management on your part, and you will still need to log in to the server version to submit assignments and access feedback and solution sets, so unless you have some prior experience with computing I recommend sticking with the browser interface for now to minimize confusion.

The R language is one of the most popular computing tools used by practicing statisticians and data scientists, and so although statistical reasoning is the main goal of the course, competence in R and written presentation of results is a valuable goal in itself, especially if you are going to be working in a field that uses quantitative data.

MISCELLANY

Electronic Devices in Class. I strongly discourage the use of electronic devices during class when attending in person unless it is necessary, as it inevitably fragments attention and may distract others.

For note-taking, take hand-written notes if you are physically able to do so: there is evidence that writing by hand improves your cognitive processing of the material.

Communication Outside Class Time. I have set up a Slack group for communication related to the course. If you are taking the course you will get an invite at your oberlin.edu email address. **I am likely to respond more quickly if you message me there rather than via email;** however, don't hesitate to follow up if you don't hear from me within a day or two, as sometimes things slip through the cracks.

If you have a question or comment that other students might be interested in, I encourage you to **post to one of the classwide channels** rather than PMing me. You might even get a faster response from one of your peers than from me!

I will try to respond to most questions posted on class days by the next class day. **If you need to ask me about something due the following morning, don't wait until the night before!** I have family and parenting responsibilities in the evenings and on weekends, and cannot necessarily respond to messages outside normal "business hours".

Accommodations. If you require accommodations to do your best work in this class, please let me know as early as possible, and consult as well with the Office of Disability Services (ODS). By college policy, **all requests for accommodation require documentation from ODS.** That said, I have tried to structure the course in a way that automatically accommodates the most common situations, especially during pandemic conditions: there are no timed in-class assessments, for example.

Honor Code. The Oberlin College Honor Code formalizes the idea that **all work that you submit is your own and that you have given credit to the ideas and work of others when you incorporate them.** You will be asked to write and sign the honor pledge on each graded assignment that you hand in. The honor pledge reads: “I have adhered to the Honor Code in this assignment.”

What it means to adhere to the honor code depends on context. For each assignment type, I describe what it means to follow the honor code on that assignment below.

More information about the honor code can be found on the web at the Dean of Students site:

<https://www.oberlin.edu/dean-of-students/student-conduct/academic-integrity/students>

I assume mutual trust as the default stance, and will not try to police your work; however, when violations are discovered they will be treated seriously and referred to the honor code committee. The most common violation I have unfortunately run across is unauthorized collaboration on takehome quizzes or exams, so **make sure you understand what, if any collaboration is licensed on any given assignment before you begin.**

YOUR RESPONSIBILITIES

Readings. There will be daily reading assignments posted on the website. **It is important to do the readings,** preferably before but at least soon after material is discussed in class. Many of the concepts in this course need to be engaged with multiple times to be understood.

Attendance and Participation. I expect you to attend class (whether in person or virtually) nearly always, and participate actively. There is no formal attendance and participation component to the grade, but I will take attendance and participation into account when making judgment calls about borderline final grades.

Discussion Prompts. There will be a lecture video posted in the middle of each week with some associated short discussion questions (this takes the place of one of the scheduled class meetings, as the class will be divided in half between Wednesdays and Thursdays for labs). You will be asked to post your answers to these questions to the class Slack workspace after viewing the video. These responses are not graded for correctness; but I will read through them as a way to gauge how folks are experiencing the content, and what areas might need more attention in class.

Homework and Labs. There will be (roughly) weekly assignments, posted on the course website (not Blackboard), consisting of a mix of short electronic “self-check” questions to be completed on the WileyPlus site, a small number of more in depth problems (typically from the textbook) to be handed in for feedback, and one or two exercises related to the lab material for that week.

The self-check problems are tracked automatically by WileyPlus and can be attempted as many times as needed. The textbook and lab problems are due electronically on Monday nights by midnight and will be graded for completion and correctness.

Honor Code: Self check problems should be done individually, but I encourage you to collaborate with each other on the textbook and lab problems, with the following caveats:

1. You must each write your own solutions and code
2. You may not copy each other’s words or commands
3. You must indicate in your submission what other students you worked with
4. You should not submit anything that you do not understand (this last criterion is admittedly fairly subjective; but the idea is that you should interrogate the structure of what you writing or typing and not just imitate an example by shallow pattern-matching)

Quizzes. There will be a short takehome quiz distributed electronically on Mondays, due back by class time on Fridays. We will go over the quiz answers together on Fridays in class, and late submissions will not be accepted. However, replacement quizzes will be available: see the policy on reassessments in the “Grade Policy” handout.

Honor Code: Although they have a take-home format, you should treat quizzes as though you were doing them in class. This means that

1. **You may not collaborate with each other** on quizzes. **There have been several instances of cheating via collaboration in the past — it is easy to do, but it is a violation of the honor code, and will be reported as such.**
2. **You may not consult your textbook, notes, or other sources**, but calculators can be used (though won't be needed often)
3. The quiz must be completed in the designated time frame. They are designed to take about 10-15 minutes, but **you may take up to 30 minutes in one contiguous block**. Effectively this means everyone has built-in “double time”.

Exams. There will be two (takehome) exams at roughly the 1/3 and 2/3 of the way through the semester (see below for dates), and a takehome final exam during finals week.

Exams may only be made up in the event of a serious emergency.

Honor Code:

1. **You may not collaborate with each other on exams**; but you are encouraged to ask me any questions you have (I may not answer all of them). **There have been several instances of cheating via collaboration in the past — it is easy to do, but it is a violation of the honor code, and will be reported as such.**
2. You may not consult source material, but calculators and a **single-sided, handwritten 8 1/2" × 11" note sheet will be allowed**. The requirement that your sheet be handwritten is there to encourage you to condense and process material while creating your sheet, so that creating the note sheet is itself a study opportunity.

Term Project. There will be a data analysis project with milestones throughout the semester, to be done in pairs or threes, culminating in a final writeup at the end of the semester.

The purpose of these projects is to identify a question about a topic of interest to you, find or collect data that can shed light on your question, and use the concepts and techniques you learn in class to try to answer the question.

In past semesters, many projects involve collecting data from a sample of peers via interview; however, due to pandemic conditions making interviews infeasible, projects this semester will need to use either publicly available data, or data not collected

from humans.

Honor Code: The same criteria apply here as for homework and lab assignments (see above). In addition,

1. You must give credit and cite sources for any data or ideas that did not originate within your team. This includes paraphrases as well as direct quotations.
2. All/both members of the group must make approximately equal contributions at each stage of the project.

MAJOR DATES

Monday 9/07	NO CLASS (Labor Day)
Monday 9/28	NO CLASS (Yom Kippur)
Friday 10/09	EXAM 1 Due
Monday 10/19	Exam 1 Graded and Returned
Friday 10/23	Project Proposal Due
Monday 10/26	Project Proposals Evaluated
Friday 11/06	EXAM 2 Due
Monday 11/16	Exam 2 Graded and Returned
Friday 11/20	Project Pilot Analysis Due
Wednesday 11/25	NO CLASS (Thanksgiving Break)
Thursday 11/26	NO CLASS (Thanksgiving Break)
Friday 11/27	NO CLASS (Thanksgiving Break)
Monday 11/30	Feedback on Pilot Project Returned
Monday 12/07	Project Final Writeup Due
Thursday 12/10	Project Feedback Returned
Saturday 12/12	Final Exam Due (9am section only)
Monday 12/14	Final Exam Due (10am section only)

SCHEDULE OF TOPICS

See the “Schedule” tab on the course website (<http://colindawson.net/stat113/schedule/>).

Once you have read this document and the accompanying handout on the Grading System in this course, fill out the “syllabus quiz” here: <https://forms.gle/cjTjj2GY6Jb2EcnS6> (you can refer to the document while taking it)