

STAT 113: INTRODUCTION TO STATISTICS (SPRING 2022)

CONTACT AND LOGISTICS

Instructor. Colin Reimer Dawson (*they/them*)

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

Course Slack Workspace. <https://stat113s2022.slack.com> (join via invite link on the website)

Class Zoom ID: (see email/Blackboard)

Hybrid Structure. Monday, Wednesday and Friday classes will be held in person in Severance 132A. Thursday is the lab day, and **we will meet on Zoom** for this component of the course (I find this works better than meeting in a lab or sitting around a classroom with laptops).

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.

Mondays 3:30-4:20pm (King 204, drop-in, but appts have priority)

Wednesdays 9:00-9:50am (King 204, by appt only)

Thursdays 11:15-1:00pm (King 203, casual group office hour/lunch/workspace)

Fridays 3:30-5:00pm (King 204, drop-in, but appts have priority)

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

Date: Last Revised February 18, 2022.

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, especially now and increasingly so each year, 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, 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. Assigned readings and self-practice exercises will be from the textbook *Introduction to Modern Statistics* by Mine Çetinkaya-Rundel and Johanna Hardin. Electronic versions (both HTML and PDF) are available for free, or you can order a paperback version for \$20. Both options are available here: <https://openintro.org/book/ims/>.

I will also make use of some exercises and web applets from *Statistics: Unlocking the Power of Data* by the Lock family, but you do not need to buy it unless you want a supplementary resource (it's a good book, albeit pricey).

The topic flow in the course will occasionally not match the order of topics in the book, as I only recently switched over from using the Lock book as the main text and many of my materials still follow that sequence.

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 you can optionally 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 there are reasons why that is necessary for you, as it inevitably fragments attention and may distract others.

For note-taking, **take hand-written notes that supplement the slides and handouts and put things in your own words** if you are physically able to do so: there is evidence that writing by hand improves your cognitive processing of the material. All of the materials I present will be available on the course website, so you don't need to copy that down verbatim.

Communication Outside Class Time. I have set up a Slack workspace for communication related to the course. You can join via the link posted on the course website (requires an 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!

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; 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 exams. **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, which reinforce and supplement what we discuss in class. Many of the concepts in this course need to be engaged with multiple times to be understood. **It is ok, and expected, if things don't necessarily click the first time**

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

Discussion Prompts. I will sometimes ask you to respond to some self-check or discussion questions via a Google Form after the relevant class. 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 homework assignments, posted to the RStudio server and to the course website (not Blackboard). These consist of a mix of short self-check questions, more in-depth analyses, and problems on the weekly labs that have a computing component.

Not all of the homework problems will be collected and even fewer will be graded, but some will. The graded problems should be written up, saved as a pdf, and uploaded to a designated folder on the RStudio server. Labs have their own document format and will already be on the server, so just need to be saved in the right folder. We will go over how to do all of this during the first labs. Feedback and grades will be returned to an analogous folder on the server.

Honor Code: I encourage you to collaborate with each other on homework and lab problems, with the following caveats:

1. You must each write your own solutions and, when applicable, code
2. You may not copy each other's specific 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 several short takehome quizzes that reinforce content covered on the previous homework, and occasionally revisit particularly difficult topics from earlier in the semester. During weeks when there is a quiz, it will be posted as a pdf on the RStudio server, usually around mid-week, and due electronically in the same manner by the end of the week (Friday afternoon around 5pm).

I don't have a formal late policy, but the relevant cutoff points other than the official deadline are when I send the submitted assignments to the grader (I do this by running a script on the server, sometime shortly after the deadline), and when I post solutions (usually Sundays).

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.
2. **Quizzes are closed-book/notes unless otherwise indicated.**
3. Calculators can be used (though won't be needed often)
4. There is no formal time limit, but quizzes are intended to be completable for most people in about 15 minutes; I suggest blocking out 30 minutes.

Exams. There will be two cumulative exams, one around week 5 and another around week 11. There will be an optional final exam due during finals week, which serve as an opportunity to demonstrate mastery of concepts that you need a second or third crack at; in other words, it can be used to replace grades from previous exams or quizzes. All of them (like the quizzes) are untimed takehome assignments.

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. The exams are closed-book like the quizzes 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. The required final assignment is a data collection and analysis project and writeup (around 5-7 pages, generally, perhaps a bit more depending on the number and size of graphs). This is intended to be done in groups of 2-3 students.

The purpose of these projects is to identify a question about a topic of interest to you, design and carry out a procedure to collect relevant data that can shed light on your question, and use the concepts and techniques you learn in class to try to answer the question by analyzing the data.

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 for any data or ideas that did not originate within your team. This includes paraphrases as well as direct quotations.
2. All members of the group must make **approximately equal contributions at each stage** of the project.

MAJOR DATES AND TIMES

Mondays 11:59 PM	Homeworks Due Electronically
Fridays 5:00 PM	Takehome Quizzes Due Electronically
Friday 04/01, 5:00 PM	Takehome Exam 1 Due
Friday 04/22, 5:00 PM	Clear Your Final Project Topic With Me
Friday 05/06, 5:00 PM	Takehome Exam 2 Due
Friday 05/13, 5:00 PM	Group Project "Pilot Phase"/Draft Writeup
Tuesday 05/31 (7-9pm)	5-7 Minute Group Project Presentations; Final Writeup and Takehome

SCHEDULE OF TOPICS

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