Introduction to MATH 615

Last Updated: Wed Jul 20 8:38:44 PM
Back to the [Schedule]

What is this course about

- Developing the skills to conduct statistically valid and reproducible research.
- Understanding how data needs to be structured and formatted for analysis, so you can better prepare data collection methods for future research.
- Practicing the skills to be the boss of your own data without relying on others to "run the numbers" for you.
- Learning basic statistical techniques for a small selection of analysis situations.
- Learning how to do all of this in a reproducible manner to save you headache and time during your research.
- Laying the statistical foundation so you can learn to apply more advanced statistical models as needed, such as those covered in Applied Statistics II (Math 456).

See the syllabus for more detailed learning objectives.

Who am I?

My name is Robin Donatello and my pronouns are she/her. You can address me as "Robin", "Dr. D", or some other respectful title such as "professor".

I have a Doctorate in Public Health (DrPH) Biostatistics from UCLA, but I'm a Chico alum. I double majored in Statistics & Biology, with minor in Chemistry, and a first generation college student who started at Butte College.

My campus life consists of training the next generation of Scientists how to harness the power of Statistics and Data in a responsible and ethical manner, supporting students in their academic adventures through intensive advising and research, leading the Data Science Initiative (DSI) to develop Data Science capacity on our campus, and providing analytical support and statistical consulting for many projects on and off campus.

I am also a research manager at the Center for Healthy communities. I help to manage their research and evaluation team.

I have yet another hat as a PI for Project DA-FANH, a multi-campus collaboration whose goal is to retain and graduate highly qualified students while dismantling barriers and existing stereotypes. We aim to build a pathway for students of color in the intersection of Data Science/Data Analytics within food, agriculture, natural resources and human sciences (FANH) careers. We run workshops and internships for both graduate and undergraduate students.

When I'm not on campus, typically I'm growing food for my family, out adventuring with my dogs, or getting some game time in.

You can learn more about the work that I do on my website

What are all the course materials?

I don't sanitize this class for you by keeping everything in a learning management system like Blackboard or Canvas. In a working environment you have to deal with multiple platforms, multiple accounts and manage multiple locations for files and content. I use the best tools for the job. Here is an overview of all the things we'll be working with in this class. Homework 0 provides a checklist for getting connected and testing out these tools.

- Class website
- Bookmark this page. You will be here a lotl
- Contains details on weekly topics, due dates, links to notes, assignments and additional materials
- Often links will be broken. Typo's happen. Notify me via Discord and I'll get to it asap.

Textbook

- The textbook is used for data, reading and learning content.
- This wil be used in Math 456, and is an excellent reference guide.
- If you're curious, I get about 1% of the sale in royalty. The income is *not* why I collaborated on this edition. I used the 4th edition in grad school and really like teaching out of it.
- **Blackboard Learn (BBL)** is used for recording grades via grading rubric.
- **Google Drive**
- Assignments will be turned in and peer reviewed through Google Drive.
- You will be added to the Math 615 Google drive using your campus email.
- **Discord** will be used for outside class discussions, homework help and general chatter.
  - I will not answer most questions through email.
  - Download either the phone app or the desktop app (I use both). This is mandatory. Do not rely on remembering to log in via the web browser. You will miss important notifications.
- **Lecture notes**
  - Stand alone lecture notes like these.
  - Collaborative lecture notes on Hack MD (more info below)
  - Applied Statistics notebook
  - Most are available as PDF or HTML.
- **Hack MD**
  - Interactive collaborative notes
  - https://hackmd.io/@norcalbiostat/intro_hackmd#/
  - No account needed to view, but account needed for editing (which you will be doing)
- Poll Everywhere is a web-based student response system.
- During class I will display an activity on-screen and you respond on your device (computer or mobile).
- I have already registered you for Poll Everywhere by connecting it to Blackboard.
- Use your student email to log in and use the password reset link to create a new password. Let me know if you can't get registered.

What does this look like in practice?

Here is a typical schedule for student success

- Sunday you check the course website schedule to note important upcoming due dates and to plan accordingly.
- Before Monday class you watch one of the lecture videos and do some reading on that week's topic.
- Monday class: ~25 min lecture on topic for the week, ~50 min actively working on homework
- Before Wednesday class you complete the assigned reading and videos, and prepare answers to the discussion questions
- Wednesday class: ~10 minute topic quiz, ~25 min questions/discussion, ~50 min actively working on homework
- Before Thursday EOD you submit your draft assignment to google drive.
- Between Friday & Saturday noon you provide feedback on 1-2 of your classmates assignments and continue to work on your own assignment using the rubric in BBL as your guide.
- By EOD Sunday you revise your assignment based on your peer's feedback and submit your final assignment for grading

EOD: End of day, or 11:59pm.

> Expect to spend 10-12 hours each week on this class. I will do my best not to move due dates so you can plan accordingly.

Project

- This course will revolve around a data analysis project.
- Individual projects, but you will collaborate with each other through a peer review process.
- All assignments are desgined to support your research.
- Must choose a project out of select data sets.
  - Individual research is typically not developed or robust enough to be demonstrative.
- Project will culminate with the creation and presentation of a research poster.
- More details are on the project page.

Computing and Reproducibility

- No more TI-83, modern statistics is computational based. And I don't mean Excel.
- Big push for open research in the Natural and Social Sciences.
  - Sharing code & data. Sometimes required along with manuscript for publishing.
- Reproducibility. Give someone else access to your data and code, and they can replicate your findings.
  - We will practice this in this class.
  - I practice this by putting all class material online with a cc-by license. (others are free to copy and share my work with acknowledgement)
- Review these Slides on reproducible research in the social sciences.
  - I will not require any measure of version control or open source coding in this class.
- Be mindful about file naming conventions (slide 11). Make a plan and stick with it.
  - https://www.xkcd.com/1459/
- Expect to bring your laptop every day to class.
  - The more reading and content learning done outside of class, the more time for in class analysis and discussion

Software program of choice (SPC)

- This class is not a class on how to use the software program. You will be responsible for learning a lot of the programming language outside of class on your own time. That's part of the process of a Masters program. You are learning how to learn, how to look stuff up, how to do new things. You can't learn all the things in 2 years.
  - There is an entire secton on the help page of the class website that is devoted to programming resources.
- All my lecture notes use R. This entire website is built with R. R is a pioneer in generating reproducible and publishable quality reports.
  - Here's an student-generated example
- I will not dictate which software program you use in this class. Homework 0 has more information on the two historically popular programs for this class; SPSS and R.
- But I will expect you to submit reproducible code. You can point and click your way to an answer, but code must be saved and reusable with minimal changes.
- Be open to new things, there is power in being polyglottal. You can use one language in here and another language for another class or project.
- Your professors in your other classes, or your masters committee may want you to learn a specific language. That should influence your choice.
  - So should your industry. Don't make assumptions, look at job postings and see what they want.
  - I.e. Center for Healthy Communities has a lot of Nutrition faculty/students. They're solidly using R.
  - Political Science & Economics often use Stata. I've met Sociologists that use SAS, Stata, R and SPSS.
- I am fluent in SAS, STATA and R, reasonably read in SPSS but can provide limited support for Python.
- The choice of software you use for this class is up to you, but you will be expected to learn how to navigate the programs outside of class time.
- I highly recommend making a habit to attend Community Coding even if it's just to sit and do your work with your peers.