Introduction to MATH 615

Last Updated 2019-06-16 12:51:58

Contents

What is this course about **1**

Course resources **1**

Descriptive vs. Inferential Statistics **2**

Project **2**

Assignment Submission **2**

Computing and Reproducibility **2**

Software program of choice (SPC) **3**
SPSS . . . 3
R . . . 3
SAS? STATA? Python? . . . 4

Organizing your working directory **4**

Last Updated: Sun Jun 16 12:51:58 PM
Back to the [Schedule]

What is this course about

Course resources

- **Class Website**
  - https://norcalbiostat.github.io/MATH615/
    - Landing page for announcements
    - Details on weekly topics can be found on the schedule
    - Includes links to notes, assignments and additional materials
  - Often links will be broken. Typo's happen. Notify me via slack and I'll get to it asap.
  - The syllabus covers course details such as grading, office location and classroom policies.
- **Blackboard Learn (BBL)** is used for recording grades
- **Google Drive** – Assignments will be turned in and peer reviewed through Google Drive.
- **Textbook: Practical Multivariable Analysis (PMA5)**
  - The textbook is used for data, reading and learning content.
    - Great long term resource, but a new edition will be coming out next year.
    - I've provided a draft for select chapters in the 6th edition.
- **Slack** will be used for outside class discussions, homework help and general chatter.
  - I will not answer programming questions through email.
  - Download either the phone app or the desktop app (I use both). This is mandatory.

  - I don't think push notifications are not automatically turned on. You must opt in.
- **Lecture notes**
  - Combination of Applied Statistics notebook, and stand alone lecture notes like these.
  - Available as PDF or HTML.

Descriptive vs. Inferential Statistics

- Two main phases of Statistics.
  - Also called Exploratory and Confirmatory
- Passion Driven Statistics
  - Backbone theory behind the class. Read this PDF to get a sense of how statistics fits into science.
  - The videos can be used as an additional learning tool.
- Google `data analysis lifecycle` and look at images. What sense do you get?

Project

- This course will revolve around a data analysis project.
- Individual projects, but you will collaborate with each other
- 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.
- More details are on the project page.

Assignment Submission

See the course FAQ page.

Computing and Reproducibility

- No more TI-83, modern statistics is computational based.
- Big push for open research in the Natural 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 learning that on your own or in another class.
- 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.
- 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 reusible with minimal changes.
- Be open to new things, there is power in being polyglottal.
- The first few weeks will be ramp up time.

SPSS

- Purchase v24 or v25 from http://www-03.ibm.com/software/products/en/spss-stats-gradpack for $50 for 6mo rental.
- Point and click, but can save code and write scripts.
- Stand alone program. No integration. Licenses are not cheap.
- Will be used again in NSFC 600 (no exp necc for that class either)
- On campus resources: From the desk of David Philhour (BSS)
  - Open computer labs in Butte Hall (207, 211) with many open lab hours.
  - Tutoring center in AJH108 run by Dr. Penelope Kuhn.
  - Check availability of Psyc dept lab in Modoc 224
- Off Campus resources
  - Kent State University Tutorials: https://libguides.library.kent.edu/SPSS/home
  - UCLA Institute for Digital Research and Education: https://stats.idre.ucla.edu/spss/
  - Recommended selection of YouTube videos https://www.youtube.com/results?search_query=andy+field+spss+tutorials

R

- Free. Installation Instructions here: https://norcalbiostat.netlify.com/post/software-overview/
- Harder up front, more powerful in the end.
- Seamless integration with a multitude of other scientific analysis and reproducible reporting mechanisms.
- Becoming much more popular in all scientific fields of study. One of the primary languages for Data Science.
- Google at diagram of the `tidyverse` (a suite of functions in R). Compare it to the images of the data analysis life cycle. What sense do you get?
- Need some motivation?
  - https://www.psychologicalscience.org/observer/why-you-should-become-a-user-a-brief-introduction-to-r
  - https://osf.io/j28w7/
  - https://www.youtube.com/watch?v=jn_3N_o2d6Q
- On campus resources
  - Introduction to R (MATH 130) 1 unit CR/NC
  - Data Science Initiative workshops, talks, open drop in analysis time.
- Off Campus resources (a few)
  - Chico R Users Group
    - Meetup
    - Google l-serv
  - Data camp
  - Quick-R

  - Cookbook for R
  - R Examples Repository (This site was also built using R Markdown, is open source and a fabulous example of reproducible research!)

SAS? STATA? Python?

- SAS has only now working on literate and integrated programming by using Jupyter notebooks and SAS University Edition (free)
- Stata has a few user written packages that allow for the integration of LaTeX or markdown into your code document.
- Python is the other primary language for Data Science. I am in the process of learning Python but the capabilities are very great. If you're thinking this route be sure to use Jupyter notebooks.

Organizing your working directory

Using a consistent folder structure across your projects will help keep things organized, and will also make it easy to find/file things in the future. This can be especially helpful when you have multiple projects. In general, you may create directories (folders) for **scripts**, **data**, and **documents**.

You need to choose a naming convention for your class folder and stick with it. Recommended options are:

- ALL CAPS (MATH615)
- no caps (math615)
- snake_case (math_615)
- CamelCase (Math615)

Call this working directory `math615`, and create the four subfolders: `data`, `scripts`, `documents`.

You will put all files related to this class in here. For example lecture notes go in the document folder, homework code files in the `scripts` folder, and data in, you guessed it, the `data` folder.

This means when you download a file, right click and "Save as" or "Save target as" and **actively choose** where to download this file. Do not let files live in your downloads folder.

Your working directory should now look similar to this: