Biol 8008 Grad course in Data Analysis W2022 Syllabus

Instructor: Dr. Dennis Higgs

email: dhiggs@uwindsor.ca

Course objectives:

This semester we will be focusing on the basics of biological data statistical analysis as well as the basics of how to use R. Both of these topics will be pitched at an introductory level for graduate students that have had limited exposure to both topics and wish to progress to using these tools for their own research. At the completion of the course students should be familiar with the basic types of tests commonly used for analysis of biological data and how to pick an appropriate test but the course is not designed to be an in-depth training in biostatistics theory. Similarly, at the end of the course students should feel comfortable with using R basics and using already developed R packages for their own uses as well as being able to self-learn more advanced R topics but the course is not an in-depth dive into more advanced aspects of R. The course will be an interactive environment of trading tips and approaches for all. We will start with one week on stats and then one week on how to use R; subsequent weeks will be more of a mix of statistical and R topics together to give practical tips on how to carry out analyses and plot data.

There are excellent resources available for statistics and R so it is no longer necessary to purchase textbooks for topics covered in the course. I recommend R for Data Science (https://r4ds.had.co.nz/index.html) and An Introduction to R (https://alexd106.github.io/intro2R/) as excellent resources to get started. A former graduate student in the Mennill lab, Kiirsti Owen (twitter: @KiirstiO) also put together a basic RAdvent Calendar a couple years ago that is a wonderful and simple introduction to R, so reach out to her if you are interested. This is also referenced in the "Getting started" document by Syed that I will post. Because R is open source there are a growing number of online resources available as well.

Course Time and Location: All meetings will be Mondays from 11-1. Initially all will be on Teams (Teams link) but once we are allowed back on campus I will find a classroom. Those that have already told me they are not on campus this semester can join the in-class portion virtually. virtual and synchronous.

Approximate Course Schedule

Date	Topic	Outside resource	Assignments
Jan 24	Intro to distributions and P. Basic tests	Head et al. 2015 Stephens et al. 2006	In-class discussion during week 2
Jan 31	Brief discussion on week 1. Intro to R- basics Data Manipulation in R	https://r4ds.had.co.nz/transform.html Broman & Woo 2018 Dpylr cheatsheet (to be provided) Syed tutorial	1 (Due Feb 7)
Feb 7	Hypothesis development and testing; how to pick a test	Fudge & Turko 2020	
Feb 14	Data visualization & ggplot	https://moderndive.com/2-viz.html	2 visualization (due Feb 21)
Feb 21-25	**Reading Week**		
Feb 28	Discuss visualization assignment Deeper dive into t-tests & ANOVA & non-parametrics	TBD	In class discussion
Mar 7	Contingency Tables & GLMM	Bolker et al. 2008	3 (due Mar 14)
Mar 14	Linear vs. non-linear Regression	TBD	
Mar 21	Repeated measures	TBD	4 (due Mar 28)
Mar 28	Presentations on how to analyse		5 (due Apr 4)
Apr 4	Principal components & Discrim functions		

Grading scheme: Because it is a course we do have to give grades. The grading scheme is below, with more information on assignments below.

Item	Due date	Percent of total
Participation	Throughout semester	20
Assignment 1	Feb 7	10
Assignment 2	Feb 21	15
Assignment 3	Mar 14	15
Assignment 4	Mar 28	20
Assignment 5	Apr 4	20

Assignments: To keep you engaged and to give you feedback there will be 5 assignments throughout the course as detailed below: More detail and data files will be provided for each assignment closer to the due date.

Assignment 1. The basics of R: This is a 2-part assignment. In part A you will build a piece of R code that will do the following:

Load a file into Rstudio (I will give you the csv file to use); show headers in the file; Calculate min, quartiles, median, mean and max for each integer variable in a data frame; use 'mutate" to form a new variable and save this to a new file; return a value for the 60th column in the 5th row of the new dataset; rearrange the new file to sort by 1 variable and save it to a new file.

In part B you will take a piece of R code provided to you that is intentionally faulty and you will fix the R code to properly run.

Assignment 2. Data visualization: I will provide you with a dataset and the variables I wish to see plotted but then you choose the best way to plot the data of interest in R. Submit graph to class with an explanatory legend. Present to class and everyone gives input.

Assignment 3. Data analysis: I will provide you with a number of realistic scenarios and you will decide the best way to analyse these data and discuss the how the data could be interpreted given the test you chose. Note: you will not run the tests on these data, just think about what is the best test to run.

Assignment 4. Data analysis and presentation: You will be given a large data set and you will then make several figures that best describe these data as well as run appropriate statistical analyses, all in R. You will write the data analysis portion of the methods section and plot figures and report a results section with appropriate statistical information; as would be seen in a typical journal article in your field.

Assignment 5. Constructive criticism: You will be given a copy of Assignment 4 from one of your classmates. You will give this assignment a professional review, as if you were a reviewer of a journal article.

A note on doing your own work. Given these extraordinary circumstances this semester all your assignments will be completed offline. I encourage you to work together on these assignments and use whatever outside resources you need to complete your work as this is the way science works in real life anyway. Anything you turn in for grading must be in your own words and be original thought, and all University rules on plagiarism will be strictly followed (https://www.uwindsor.ca/academic-integrity/358/avoiding-plagiarism#:~:text=As%20defined%20in%20the%20University,proper%20acknowledgement%2C%20representing%20these%20as). I do want to encourage collaboration however so please discuss with your classmates any aspects you choose to before putting it into your own words.