Biol 250: Biological Data Analysis

Spring 2019, Tuesday/Thursday 2:30-3:45, ISC 115

Dr. Suann Yang (instructor)
Email: yang@geneseo.edu
Office hours: Tu 1:30-2:30, W/F 10:30-11:20 (or by appointment) in ISC 256

Ms. Natalie Braun (teaching assistant)
Email: nb14@geneseo.edu
Help sessions: M 5:30-6:30; Location TBA

Intended learning outcomes

A successful student in this course will be able to:

1. Design experiments to study biological phenomena
2. Apply the principles of modern data management and best practices in research (including ethics)
3. Explain how and why biologists use statistics
4. Use computers to do statistical analysis, particularly using the R programming language
5. Conduct the proper statistical analysis of different types of biological data
6. Interpret, present, and communicate the results of statistical tests and analyses of data using the conventions of the biology discipline

Required course materials

Always bring these to class:

1. *A Primer in Biological Data Analysis and Visualization Using R* by Gregg Hartvigsen
2. A laptop with the following software
    - R software (free download, https://cran.r-project.org/)
    - RStudio (free download, https://www.rstudio.com/products/RStudio/#Desktop)
    - Microsoft Excel (free to all Geneseo students, https://wiki.geneseo.edu/display/cit/Software+at+Geneseo)

Piazza account: We will be using Piazza for asking and answering questions related to using R. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Forums like Piazza are exactly how members of the programming community gets help fast! Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com. Find our Piazza class page on our Canvas site.

Course format

Learning to do statistics requires just that - *doing statistics.* In general, expect to complete a guided background reading and assignment to prepare for each class period. After delving into the topics further at the beginning of each class, be ready to get started on tackling these topics on your own (with help, of course!) for the remainder of each class period. Work that is started during the class period should be finished as homework unless otherwise indicated.

I will post reading and other assignments, and other materials on our course Canvas site, such as documents to print out and bring to class. Most assignments will be collected electronically using Canvas, and due by the date and time published for each assignment.

Grading scheme

Item	Percent	Notes
Daily assignments	40	Class prep, problem sets, worksheets, etc.
Quizzes	40	7 total, all are cumulative and require the use of R
Collaborative project	20	Includes deadlines throughout semester & peer evaluation
Statistics in the Media		Regular extra credit opportunity (limit of 7)

1. **Assignments:** To help you master each topic, I will assign worksheets, problem sets, and other activities on a regular basis. These will be worth a range of point values (0 to ~50), and the proportion of the points you earn out of the total (~500) will be used to calculate your overall assignment score. Not all assignments will be graded, but if you do all the assignments and put in a quality effort on each assignment, I will add 0.5% to your final percent calculation.
2. **Quizzes:** Every two weeks, you will take a quiz during the first part of class on Thursdays. See the schedule for exact dates. Each quiz begins with written questions to be answered on paper. After handing this in, the quiz proceeds to a section that is completed on your laptop, using R.
3. **Collaborative Data Analysis Project:** In this project, you will select, explore, and analyze a publicly available dataset using the skills learned in class. Some class time will be devoted to developing, discussing, and working on your project. There are multiple components and deadlines throughout the semester: i) proposal (preliminary and revised), ii) cleaning and processing the data, iii) exploratory graphs, iv) analysis report, v) final presentation, and iv) peer evaluations throughout the semester.
4. **Statistics in the Media:** You can find statistics everywhere! In this extra credit assignment, you will find an example of statistics usage in the media, and summarize and critique it. There will be seven windows of opportunity for you to complete this assignment for extra credit points (submit via Canvas). The article, blog post, etc. must be published within each window of time to be valid. Each example must be unique; only the first student to turn in a particular example will get the credit.

Final course grades will be determined as shown here:

A 93.3 - 100%	**B+ 86.6 - 89.99%**	**C+ 76.6 - 79.99%**	**D 60.0 - 69.99%**
A- 90.0 - 93.29%	B 83.3 - 86.59%	C 73.3 - 76.59%	E <59.99%
	B- 80.0 - 83.29%	C- 70.0 - 73.29%	

Attendance and deadlines policy

Please be on time and prepared for class. You must inform me of absences immediately by sending me an email and bringing me documentation if applicable. Excused absences must be made up within one week of your return to class; no make-ups for assignments or quizzes are allowed for unexcused absences. Assignments submitted after the deadline will not be accepted.

Student Code of Conduct

Plagiarism and academic dishonesty. Plagiarism and other forms of academic dishonesty (cheating, turning in another student's work as your own) is not tolerated at SUNY Geneseo. Working with other students on individual assignments is fine, but you must each produce *original* written answers or code (no copying and pasting). Identical or highly similar responses

suggest answers are being copied and ALL students involved will be held responsible. Check with me or Natalie if you are not sure where the line between collaboration and copying stands on any assignment. Evidence of plagiarism and/or academic dishonesty is grounds for a score of '0' on any assignment and further action including notifying the department chair, which can result a report filed with the Dean of Students.

Copyright notice. All materials used in this course, including lectures, slides, and handouts, are copyrighted. Reproduction and/or distribution of materials is prohibited without author consent. This includes, but is not limited to, transferring files to websites such as StudyBlue and Course Hero, storing old tests in sorority/fraternity test banks, and passing on assignments to friends who may take the course in the future. Be aware that UUP (Union of University Professionals, the union representing faculty on this campus) is seeking to take legal action against groups who violate copyright, and that posting or selling copies of materials to such groups may put a student in legal jeopardy.

For full details of the Student Code of Conduct, please see the Student Handbook (https://www.geneseo.edu/handbook/student-code-conduct).

Accommodations

SUNY Geneseo will make reasonable accommodations for persons with documented physical, emotional, or cognitive disabilities. Accommodations will be made for medical conditions related to pregnancy or parenting. Requests for accommodations including letters or review of existing accommodations should be directed to Ms. Heather Packer in the Office of Disability Services in Erwin Hall 22 or disabilityservices@geneseo.edu or 585-245-5112. Students with letters of accommodations should submit a letter to each faculty member at the beginning of the semester and discuss specific arrangements. Additional information on the Office of Disability Services is available at www.geneseo.edu/dean_office/disability_services.

Proposed weekly schedule

The below outline of our semester is what we are aiming for. I may adjust the topics and dates during the semester; be assured I will inform you of any changes in a timely fashion.

Date	***Topic***	***Notes***
Th: 1/24/19	R intro/refresher	
Tu: 1/29/19	Biological research design	
Th: 1/31/19*	Biological research design, continued	
Tu: 2/5/19	Biological research design, continued	
Th: 2/7/19	Getting to know the data	
Tu: 2/12/19[a]	Getting to know the data, continued	[a]Prelim project proposals due
Th: 2/14/19*	Getting to know the data, continued	
Tu: 2/19/19[b]	Getting to know the data, continued	[b]Peer review of prelim proposals
Th: 2/21	Getting the data ready to analyze	
Tu: 2/26/19[c]	Getting the data ready to analyze, continued	[c]Explore project dataset
Th: 2/28/19*	Hypothesis testing	
Tu: 3/5/19	Hypothesis testing, continued	
Th: 3/7/19	Goodness of fit and contingency tests	
Tu: 3/12/19[d]	Goodness of fit and contingency tests, cont	[d]Final project proposals due
Th: 3/14/19*	Tests of difference: two samples	
3/18 to 3/22/19	Spring Break	
Tu: 3/26/19[e]	Tests of difference: two samples, continued	[e]Project team peer eval #1
Th: 3/28/19	Tests of difference: more than two samples	
Tu: 4/2/19	Tests of difference: more than two samples, continued	
Th: 4/4/19*	Tests of relationship: correlation	
Tu: 4/9/19	Tests of relationship: correlation, continued	
Th: 4/11/19	Tests of relationship: regression	4/12-14: SY out of town
Tu: 4/16/19	Tests of relationship: regression, continued	
Th: 4/18/19*	Tests of relationship: regression, continued	
Tu: 4/23/19	Tests of relationship: regression, continued	
Th: 4/25/19	Projects[§]	
Tu: 4/30/19	Projects[§]	
Th: 5/2/19*	Projects[§]	
Tu: 5/7/19[f]	Projects[§]	[f]Project analysis report due
Tu: 5/14[g] (3:30-6 PM)	Project presentations	[g]Project team peer eval #2

*Quiz at beginning of class on these dates

[§]Additional topics may be scheduled depending on what projects need