BIOL 350 – Foundations of Biostatistics – Fall 2024

Instructor: Dr. Logan W. Cole – E-mail: lcole@geneseo.edu – Phone number: 585-813-6923

Class time: Asynchronous Location: Remote

Office hours: Tuesdays at 4:00 PM - 5:00 PM eastern time

Prerequisite: BIOL 116 or BIOL 118/120 General Biology Laboratory

Course description: Quantitative analysis in the biological sciences involves understanding the scientific method, experimental design, sampling theory, graphs, and data analysis and interpretation. Students will develop and test hypotheses using statistics and graphing techniques in computer-based laboratory exercises. (3 credit hours)

Required textbook: *Introduction to Statistics for the Life and Biomedical Sciences* by Vu and Harrington (2020) 1st edition, OpenIntro, Inc.

Recommended reference: *A Primer in Biological Data Analysis and Visualization Using R* by Hartvigsen (2021) 2nd edition, Columbia University Press.

Date	Day	Chapter	Topic/deadline	Readings/resources
8/26	M	1	Introduction to data	Equation editor in Word R tutorial
9/6	F		Review questions	
9/10	T		Problem set	
		2	Probability	Regan et al. 2002 Phylogeny inference primer Guide to public databases
9/17	T		Review questions	
9/20	F		Problem set	
		3	Distributions of random variables	Misconceptions about the normal distribution
9/27	F		Review questions	
10/1	T		Problem set	
		4	Foundations for inference	Betensky 2019 Halsey 2019 Greenland et al. 2016
10/8	T		Review questions	
10/11	F		Problem set	
		5	Inference for numerical data	Ranganathan et al. 2016
10/25	F		Review questions	
10/29	T		Problem set	
		6	Simple linear regression	
11/5	T		Review questions	
11/8	F		Problem set	
		7	Multiple linear regression	Morrissey & Ruxton 2018
11/15	F		Review questions	
11/19	T		Problem set	
		8	Inference of categorical data	Xu et al. 2010
12/3	T		Review questions	
12/6	F		Problem set	
Research paper				
10/18	F		Data retrieval	Dryad/GenBank
11/11	M		Research paper and analyses	Research paper instructions
12/9	M		Revised paper	

NOTE FOR ALL ASSIGNMENTS

For all questions that require math, show your steps (see the equation editor tutorial if you don't know how to do that in a document). For all questions that require code, show your code (or if you use a program with a graphical user interface, show all decisions made and functions used). For all questions that require the written word, write in complete sentences.

ALL ASSIGNMENTS ARE DUE AT NOON (EASTERN TIME) ON THEIR DUE DATE

REVIEW QUESTIONS

After you read the material (the notes and the corresponding chapter), you will be posed some questions to review the material. These sets will comprise 24% of your grade for this course (24% - 3% for each set of review questions). Individual questions will be graded on the basis of good-faith completion and I will provide extensive feedback for these review questions. So, if you do not address the question or part of your question in your response, you will lose credit.

PROBLEM SETS

These problems sets will comprise the plurality of your grade for this course (40% - 5% for each problem set). These will be graded on correctness. These sets will require a lot of time and will have a relatively short window for completion.

RESEARCH PAPER

A research paper will be a major part of the grade for this class (26%). This research paper will be written in the format of a scientific paper and involves a statistical analysis of data from a public database. This analysis will include hypothesis testing and data visualization, not just descriptive statistics (though descriptive statistics can be used to bolster your arguments). The recommended reference by Hartvigsen may be helpful for data analysis and visualization.

PARTICIPATION

I will pose questions on the discussion board and you will respond – this grade (10% of your course grade) will be based upon good-faith participation on the discussion board, which includes responses in complete sentences that address the question asked. As an online course, it is also important that you respond to emails in a timely manner – stay apprised of your email while this course is running – this will also be considered.

COURSE GRADE BREAKDOWN

REVIEW QUESTIONS (8 sets x 3 points per set)	24%
PROBLEM SETS (8 sets x 5 points per set)	40%
RESEARCH PAPER	26%
PARTICIPATION	10%

LATE WORK AND MAKE-UP POLICY

Unexcused late work will be accepted initially at a 10% grade reduction with a further reduction of 10% for each day late and a maximum deduction of 50% (work can be 'made up' more than four days late at a 50% deduction).

LEARNING OUTCOMES

- Explain why and how biologists use statistics
- Apply experimental design to biological problems
- Evaluate experimental design and statistical analyses used in biological studies
- Identify the extent to which conclusions from statistical analyses can be generalized
- Conduct appropriate statistical analyses on different kinds of biological data
- Use computers to do statistical analyses, particularly using the R programming language, while understanding the limits of such analyses
- Apply modern principles of data management to ensure reproducibility of analyses
- Interpret and communicate the results of statistical analyses as used in biology

Data means little outside a framework in which we interpret it – statistics provides a framework. The origin of modern frequentist statistics, the focus of this course, is deeply intertwined with that of genetics and the modern synthesis of evolutionary biology. Biology is most generalizable when appropriate statistics are applied. Statistics has a pivotal role in giving us comprehensive explanations of the living world that go beyond compelling anecdotes.

This course will require students to use R, a computer programming language for statistical analysis and data visualization. R was touched upon in BIOL 116, 118, and 120, which are prerequisites for this course. We will frequently employ this language to complete assignments. Though there will be instruction on R in this course, as an asynchronous online course, it is incumbent upon the student to work independently to gain some R proficiency. As such, it may benefit the student to reflect upon prior usage of the R language and refresh themselves with material from previous courses or other resources, such as the recommended text.

REQUIRED TEXT

This course will use the first edition of *Introduction to Statistics for the Life and Biomedical Sciences* by Vu and Harrington. Course notes will correspond to this text though will emphasize what I think is important. I have also added some material that will provide more understanding of how statistics are as presented in modern scientific literature. Review questions (and my feedback) will also reflect these emphases. All these components (online discussion, course notes, textbook readings, literature assignments, review questions, and problem sets) constitute the multiple ways of thinking that are conducive to a more complete understanding of statistics. It will be difficult to do well in this course if you do not give all of these components attention.

REQUIRED MATERIALS

1. Computer and reliable internet access.
2. Access to the Brightspace site for this course – check it frequently
3. Several pieces of software
    a. R software (free download)
    b. R studio (free download)
    c. Microsoft Office including Excel, Powerpoint, and Word (free download for Geneseo students)
    d. Adobe Acrobat (free download for Geneseo students) or similar PDF reader
    e. Minitab or SPSS – optional (free download for Geneseo students)
4. The aforementioned required textbook: *Introduction to Statistics for the Life and Biomedical Sciences* by Vu and Harrington (and recommended reference by Hartvigsen)

ADDITIONAL SUPPORT AND ACCOMODATIONS

I am pleased to help you with any issues you have understanding the material for the course. I will be very responsive to emails – I should be able to respond in a day (usually much sooner) and if I don't, please send me a reminder because I probably just missed it. You can also come to my office hours online. If you cannot make those office hours but want to directly ask me some questions or get support in real time, we can also arrange an appointment to meet via Zoom or by telephone.

Additional tutoring (including support for using R) will also be available for BIOL 350 students later in the semester.

This course will honor all legally required accommodations for persons with disabilities, attendance on religious holidays, and leaves of absence for those on duty.

ACCESSIBILITY

All course materials are available on Brightspace and I've tried to make them accessible; if there are any difficulties accessing any materials (including the need for different formats), please let me know as soon as possible and I will do my best to find a solution.

SUNY Geneseo is dedicated to providing an equitable and inclusive educational experience for all students. The Office of Accessibility will coordinate reasonable accommodations for persons with documented physical, emotional, or cognitive disabilities, as well as medical conditions related to pregnancy or parenting. Students with approved accommodations may submit a semester request to renew their academic accommodations. More information on the process for requesting academic accommodations is on the Office of Accessibility Services website. You can also contact the Office of Accessibility Services for questions related to access and accommodation: they are located at Erwin Hall 22 or contacted by telephone at (585) 245-5112 or by email at access@geneseo.edu.

STATEMENT OF EQUITY AND OPEN COMMUNICATION

We recognize that each class we teach is composed of diverse populations and are aware of and attentive to inequities of experience based on social identities including but not limited to race, class, assigned gender, gender identity, sexuality, geographical background, language background, religion, disability, age, and nationality. This classroom operates on a model of equity and partnership, in which we expect and appreciate diverse perspectives and ideas. If anyone is experiencing exclusion, intentional or unintentional aggression, silencing, or any other form of oppression, I encourage open communication with myself and/or the class as a whole.