Department of Biological Sciences, Binghamton University
Biol 437: Biostatistics
Fall 2018 Syllabus

Lecture – Tuesdays and Thursdays 1:15-2:40 in WH G002

Discussion Sections – Monday 1:10-2:10 (S3 G13); Wednesday 2:20-3:20 (S3 G13)

Instructor Information

Professor: Dr. Tom Powell
Email: powellt@binghamton.edu
Office: Science III 112
Office phone: 777-4439
Office Hours: Monday 3:00- 4:00, Thursday: 3:00 - 4:00; **or by appointment**

TA: Pheobe Deneen
Email: pdeneen1@binghamton.edu
Office: Science III 115
Office Hours: TBD

Text Book (required): Michael Whitlock and Dolph Schluter (2015) *Analysis of Biological Data*, 2nd edition, W.H. Freeman, NY

Other required material: Calculator (scientific or graphing); Access to computer with R

Course information

Course description- All of modern scientific progress requires statistics. Science isn't just an accumulated body of knowledge about the universe to be memorized. Rather, science is a way of knowing; it's how we advance human knowledge through rigorous empirical inquiry. Science, at its core, is about confronting models (hypotheses) of how the world works with data. Statistics provide an objective framework for generating empirical descriptions of the world and for evaluating whether our models are congruent with data.

Advanced statistics is often emphasized in biology curricula in particular because the complexity of biological systems requires us to move beyond trivial methods. Statistics as a subfield of mathematics grew directly out of the need for biologists to grapple with the extraordinary variation in the living world. There is more variation in a single population of Trinidadian guppies (or even an isogenic line of *Drosophila* in the lab) than exists across all of the electrons in the universe. Biological statistics allow us not only to describe this variation but to account for it as we evaluate hypotheses. A robust statistical toolkit is an absolute requirement for any biologist, from ecologists studying the impacts of invasive species to population geneticists studying natural selection in action to cell biologists studying dynein function to epidemiologists studying the factors effecting influenza outbreaks.

This course represents a first step in developing such a toolkit. We will cover ways of objectively describing patterns in data, presenting data graphically is ways that ease understanding, and learn about some of the basic probability building blocks on which statistical tests are built. From there, we will move into applying statistical analyses to be able to make inferences about the world. We will learn how to design experiments in ways that make statistical analysis possible and populate your biostastical toolkit with a series of standard analyses that cover a wide range of possible variables and study designs. You will execute many of these analyes by hand (which greatly reinforces your understanding of how they actually work) and also learn how to implement them in R – the statistical computing framework that has become the standard for the field.

Learning objectives - On the successful completion of Biol437, students will be able to:

1) Use appropriate metrics for describing central tendency and variation in biological data
2) Interpret patterns in visual depictions of data
3) Explain the connection between building statistical models of biological systems and hypothesis testing
4) Design experiments that fit basic statistical models *a priori*
5) Analyze and critique statistical methods presented in examples of primary scientific literature
6) Apply basic statistical methods to test hypotheses with different combinations of categorical and continuous variables with real biological data
7) Demonstrate a basic proficiency for applying these statistical models using the platform R.

Along with these *proximate* goals (those related directly to the curriculum at hand), I also intend for this course to help students progress towards some important *ultimate* goals of science process in education, including 1) an understanding of how biologists use the scientific practice, 2) becoming comfortable with reading the primary scientific literature, and 3) developing an understanding of the integrative nature of modern biological sciences.

Communication policy – I want to be as accessible as possible to students. Although this isn't a particularly large class, I happen to be *extremely busy this semester*. I really wish that I could make myself available to work hands-on with each student on every topic in the course throughout the semester, but that simply isn't feasible with my many other responsibilities this semester. In lieu of being able to meeting with each of you outside of class on a regular basis (**and especially in lieu of having to respond to daily emails from students**), I'm going to propose a multi-pronged strategy for getting you the help that you need to succeed in this class. I urge each of you to take advantage of all of the options available to you. **1) Ask questions in class**; with a moderate sized class like this, I hope that everyone grows comfortable speaking and asking question in class (although I know that won't always been the case). Keep in mind that if you don't understand something, it's almost definitely the case that some of your classmates have the same question. They will certainly appreciate that you took the initiative to ask it in class. **2) Work with your peers** I strongly encourage group studying outside of class. **3) Utilize**

the discussion sections - The whole point of these is to allow you to engage with the course material in a different format that is more conducive to interaction with your peers and instructors **4) Come to office hours** – Both Pheobe and I will be holding regular office hours throughout the semester. Theses time slots are set aside specifically to interact with you outside of class –please take advantage of them. Of course, if the posted times clash with your schedule, don't hesitate to schedule another appointment. **If all else fails – send me an email** – I will endeavor to respond to emails promptly. If responding to your inquiry requires some thought on my end, I will send you a quick email acknowledging that I received your message and when to expect a proper response. I think that reciprocal dialogue is almost always the best way to resolve issues, so I strongly encourage students to utilize the other four methods first.

Grading and assignments

Grades for the semester will be assigned based on the following table:

Category	Percentage of final grade
Test I	17.5
Test II	17.5
Final	17.5
Primary literature assignment	5
Experimental design	5
Discussion/Homework assignments	17.5
Project	20

Exams – There will be three tests throughout the semester. The dates are: **September 27, November 1, and the Final (TBD – I'll update this once the schedule comes out).** The material in the course builds on itself as the semester progresses. In that sense, all of the exams are somewhat cumulative (you won't be able to forget the information from week 2 and expect to do well on the material for week 12). However, Exam II will not directly ask you about material covered in Exam I. Exam III will concentrate *primarily* on material covered since Exam II, but there will be some questions relating directly to **fundamental** material covered since the beginning of the semester.

Exams will involve a combination of **multiple choice, short answer, and problem-based**

questions (you should expect problems to represent the bulk of the time and points on each test).

Project – One major assessment in the course will be a written project in which you apply your developing knowledge of biological statistics to test a hypothesis with real data. Approval of topics and appropriate check-ins on progress will occur during the semester. The schedule and full details of this important assignment will be posted to myCourses and discussed in class. The major difference between the graduate and undergraduate sections of this class will be the scope of this project.

Discussion/homework assignments –During the semester (see schedule), you will have a number of small assignments designed to encourage you to work through problems and apply the statistical tests learned in class. These assignments will involve a combination of hand calculations and the execution of statistical methods in R. You will have the opportunity to work on these during the discussions and afterwards at home. They will be due during your assigned discussion section the following week. Grading for each these assignments will be as described during class.

Primary Literature Assignment As upper division biology students, you need to become comfortable with navigating the primary scientific literature in your field. One assignment during the semester (schedule and details to be posted on myCourses and discussed in class) will involve writing a short description of the statistical methods employed in a real scientific paper of your choosing.

Experimental Design Assignment Experimental design is an integral part of biostatistics; stats can only address hypotheses if the data are derived from experiments that conform to the underlying statistical models. In one brief assignment this semester, you will have the opportunity to design an experiment to address a biological question in which you are interested (schedule and details to be posted on myCourses and discussed in class).

Extra credit – There will be only **one** opportunity for "extra credit" during the semester, and it requires a non-trivial amount of effort on your part. This semester, the biology department's seminar series hosts a number of internal and external speakers who present research on a wide variety of topics. The schedule for the seminar series is currently a work in progress (I know because I'm the coordinator), but I will post the schedule with the relevant speakers in myCourses and give reminders during class as the information becomes available. You can choose to earn extra credit for **one** of these events (of course, you're welcome to come to other talks, but each student only has a chance to get points once). To earn credit, you must write a one page (single-spaced) coherent synopsis of the speaker's research, with particular attention paid to statistical issues (such as what kind of variables were involved, what kind of statistical tests did the speaker mention, what formats did the speaker use to display their data… etc) and include one question about future direction of work presented in the talk. For each speaker, I will have a Turnitin assignment posted on myCourses for ten days after each talk. If you choose to do this, I will add **five points to your lowest exam grade**. I will not offer any additional extra credit opportunities on an individual basis –it's simply not fair to other students.

Course Policies

Official BU policy of credit hours and work expectations – This course is a 4-credit course, which means that students are expected to do at least 12.5 hours of course-related work or activity each week during the semester. This includes scheduled class lecture/discussion meeting times as well as time spent completing assigned readings, studying for test and examinations, participating in discussion sessions, preparing written assignments, and other course-related tasks.

Disability-related Equal Access Accommodations – Students needing accommodations to ensure their equitable access and participation in this course should notify the instructor with an Academic Accommodation Authorization from Binghamton University's Services for Students with Disabilities (SSD) office as soon as they're aware of their need for such arrangements. Please visit the SSD website (www.binghamton.edu/ssd) for more detailed information. The office is located in University Union, 119.

Academic Dishonesty -I fully expect each of you to abide by the University's *Student Academic Honesty Code* (follow link on https://www.binghamton.edu/harpur/faculty/acad-honesty.html) in all of your work connected with this course. Please note that I reserve the right to use plagiarism detection software on any material you turn in. Any infractions will be reported to the Harpur College Academic Honesty Committee and will result in a grade of zero for the assignment.

Late assignments – Late assignments will be accepted for one day after the due date, for a letter grade point deduction. Late assignments will not be accepted more than 24 hours after the due date.

Missed exams – If something occurs that affects your ability to be in class for a scheduled exam (e.g. a *serious* illness or legitimate emergency), the onus is on you to make a good faith effort to make me aware of the issue as **soon as possible**. Informing me that you missed an exam after the fact is not acceptable. Scheduling of make-up exams will be at my discretion, but they must occur as promptly as possible after the scheduled exam date.

Dealing with stress and difficulties this semester – One of the things that has impressed me about BU in my short time here is the university's explicit efforts to provide support for its students. College or grad school can be very stressful. If during this semester, you find yourself under undue personal or academic stress – please reach out for support. The people at this university, myself included, really want you to succeed and care about your well-being. Please don't hesitate to talk to me about any issues that may affect your work in my class. Additionally, I am more than happy to help you in reaching out to any one of a wide variety of campus resources.

Tentative lecture schedule and reading assignments

week	date	subject	reading
1	8/23	Syllabus with TA (Powell away)	
2	8/28	Intro to biological variation	Ch 1. 1-17
	8/30	Visualizing data	Ch 2 25-51

3	9/4	Descriptive statistics I	Ch 3 65-84
	9/6	Descriptive statistics II	
4	9/11	**NO CLASS**	
	9/13	Estimating Parameters & start of Probability	Ch 4 95-107
5	9/18	Probability theory	Ch 5 117-139
	9/20	Basics of hypothesis testing **[END material for test 1]**	Ch 6 149-168
6	9/25	Proportional data	Ch 7 179-193
	9/27	**Test I**	
7	10/2	Frequency distributions	Ch 8 203-224
	10/4	Categorical~Categorical: contingency tests	Ch 9 235-255
8	10/9	The Gaussian distribution: gateway to parametric statistics	Ch 10 273-293
	10/11	**NO CLASS**	
9	10/16	Continuous ~ Categorical I: one-sample t-tests	Ch 11 303 -317
	10/18	Continuous~Categorical II: two-sample t-tests	Ch 12 327-353
10	10/23	Violating parametric assumptions I	Ch 13 369-399
	10/25	Violating parametric assumptions II [**END for test II**)	
11	10/30	Experimental Design I	Ch 14 423-450 & Diamond 1986
	11/1	**Test II**	
12	11/6	Experimental Design II -continued discussion from 10/30- **Please VOTE – if you have to miss class to do so – no worries**	(I know you probably won't read the 10/30 stuff until the 6th)
	11/8	Continuous~Categorical w/ more than two samples: ANOVA	Ch 15 459-486

13	11/13	Continuous~Categorical w/ more than two samples: ANOVA	
	11/15	Continuous~Continuous I – correlations	Ch 16 503-524
14	11/20	Continuous~Continuous II – regression	Ch 17 539-575
	11/22	**NO CLASS**	
15	11/27	Continuous~Continous III – regression	
	11/29	Continuous~Continuous IV- multiple regression	Ch 18 605-625
16	12/4	Continuous~Continuous IV- multiple regression	
	12/6	Whirlwind tour of some Advanced Topics	TBD
		FINAL EXAM – TBD	

Tentative Discussion Section Schedule * = discussion assignment

DISCUSSION	M	W	Topic
1	*9/5	*8/29	**Installing R**
2	*9/12	*9/17	**Biology jobs discussion**
3	9/24	9/26	**Test review**
4	10/1	10/3	**Dataframes***
5	10/8	10/10	**Fundamentals in R**
6	10/15	10/17	**t-tests***
7	10/22	10/24	**Non-parametric tests***
8	10/29	10/31	**Test review**
9	11/5	11/7	**Experimental design assignment***
10	11/12	11/14	**ANOVA***
11	11/9	NO CLASS	**NO DISCUSSION**
12	11/26	11/28	**Regressions***
13	12/3	12/5	**Review/Project help**