

Syllabus BIOSC 1120: Biostatistics

BIOSC 1120: Biostatistics - Fall 2018

Instructor

Dr. Nathan Brouwer

Office: A351 Langley Hall ("Bridge")

E-mail: nlb24@pitt.edu

Website: https://brouwern.github.io/BIOSC_1120/index.html

Twitter: @lobrowR

Office Hours

12:00 - 1:00 PM Tuesday & Thursday OR email me at nlb24@pitt.edu to request an appointment

Class Meeting Times & Location

Thursdays 9:30AM - 12:00PM

08/27/2018 - 12/07/2018

170 Crawford Hall

Pre-requisites

The general pre-requisites for this course are completion of:

- 1 year of intro biology
- 1 upper division bio course
- 1 intro stats course.

Specifically:

- Completion of the 2nd semester of Foundation of Biology, eg BIOSC 0160 (Foundations of Bio. 2), 0716 (Foundations 2 - Honors), 0191, 0180, BIOL 0102 or 0120 (General Biology 2)
- 1 advanced biology course, eg BIOSC 0350-Genetics, 0355, 0370-Ecology, 0371, 1000-Biochem, 1810-Macromolecular Structure & Function, 0203, 1430-Ecophysiology, 1515.
- Intro stats: STAT 1000
- Minimum of a "C" for all courses

Check the [course catalog](https://psmobile.pitt.edu/app/catalog/listCatalog) for further details: <https://psmobile.pitt.edu/app/catalog/listCatalog>

Catalog Description

Biostatistics is the application of statistical principals to the design and analysis of biological studies. This course covers the methods, software, theory, and philosophy used in contemporary biostatistics. Examples will be drawn from across biological disciplines, including ecology, physiology, genomics, cell, and molecular biology. Topics covered include basic data science and visualization, hypothesis testing and model building, analysis of variance (ANOVA), linear regression, multivariate statistics (PCA, MANOVA), power analysis, generalized linear models (GLM), basic random effects models, and experimental design. Emphasis is placed on actively using the software R to apply these methods to real biological data. (Class Number: 30894)

Course description

The purpose of this course will be to provide advanced undergraduate and graduate students a solid background in the core concepts of applied biological statistics, the use of the software R for data analysis, and the use of RMarkdown for documenting reproducible analyses. The course will review basic statistical concepts, thoroughly cover contemporary linear regression and analysis of variance (ANOVA), and introduce the advanced concepts of generalized linear models (GLMs) such as logistic regression and mixed effect models. Students will work together on small group projects, and also develop an independent project to practice these methods or to begin using more advanced approaches.

R is the current *lingua franca* of many users and developers of statistical techniques and its use will therefore play a central role in the course from the first day of class. Basic statistical concepts will be reviewed by carrying out mathematical calculations step-by-step both visually in Excel and using code *R*. Moreover, the application and interpretation of all techniques will be presented with direct reference to the necessary *R* code and the resulting *R* output. Each class will be a mix of lecture, discussion and guided *R* exercises, with an emphasizes on hands-on interactions with simulation exercises and and interactive tutorials Throughout the course important aspects of experimental design will be emphasized, such as true replication vs. pseudoreplication, random sampling, and blocking. These topics will be used to introduce the basics of mixed models. The course will also teach the skills necessary to manage, summarize, and graphically display data.

Course Objectives

[These need to be re-written]

1. To apply the scientific method and hypothesis testing to experimental design
2. To implement appropriate experimental designs to biological investigation
3. To assess experimental results with appropriate statistical tests
4. To interpret statistical output, both biologically and statistically
5. To present results of experimental designs using computer software packages
6. To differentiate among the variety of experimental designs and successfully apply an appropriate statistical test
7. Data science & reproducible research

Class meetings

Each lecture will begin promptly at 9:30 a.m.; **attendace will be taken**. Please arrive on time so that disruptions are kept to a minimum. Class time will consist of some lecture, but will mostly consist of discussion and hands-on computer work. Preparation, attendance and participation are therefore essential for success in this course. Students should come prepared to think, ask questions, take notes, and discuss ideas with the class community.

Texts & Readings

Reading the assigned texts will be essential for success in this course. **Students should do all readings and associated assignments before coming to class.**

Required Textbook

Motulsky, H. 2018. Intuitive biostatistics: a nonmathematical guide to statistical thinking. 4th edition. Oxford, New York, NY. [Book website](#) [Amazon](#)

- The required text can be purchased through the campus bookstore or from a variety of sources online.

I am developing a **reading guide** for Motulsky (2018). The goal of the reading guide is to:

- highlight the terms, ideas, concepts, figures, graphs etc which are most important to consider
- indicate sections that can be skipped or are optional
- indicate where I disagree with Motulsky or feel that he could be clarified
- provide ideas, suggestions and hints for the completion of the Reading Reinforcement assignments
- briefly introduce the R code used to analyze the datasets Motulsky discusses

Reading the *Reading Guide* is not required but is highly recommended. Content from your reading assignments will be integrated into the Guide and suggestions for making it a useful resource are highly welcome.

Required papers

Additional reading materials used for the course will be posted on online. A bibliography containing many of these papers appears at the end of the syllabus.

Supplemental text books

The following books will be referred to at times in the course or are highly recommended supplements. These resources will be available from the library, scans will be passed out, or available for free online.

Whitlock, C. & Schluter, D. 2014. The analysis of biological data, 2nd Edition. Roberts and Company Publishers. ISBN 978-1936221486. [Book website](#)
whitlockschluter.zoology.ubc.ca/ [Amazon](#)

- I will be using some sections from this excellent undergraduate text. A copy will be on reserve in the library. Used copies of the 1st edition can be found on amazon.com for <\$40.

Beckerman, Childs & Petchey 2017. [Getting started with R: An introduction for biologists, 2nd ed.](#)

- This is a highly readable guide to the basics for R which emphasizes the use of ggplot2 and dplyr. I would use this for the course if the book was longer. The 1st edition is available online via the library and I am trying to get the 2nd edition. Available on amazon.com for \$30. **Don't get the 1st edition - its outdated.**

Crawley, M. The R Book 2nd Edition.

- This is probably the most comprehensive R book out there (also probably the thickest). Crawley is an ecologist and uses many biological examples. I wish I had bought this when I first started learning R. However, the book is not up to date with regards to graphics. You should be able to find the pdf online. (Don't buy/print the pdf of the 1st edition ed, its layout is supposed to be horrible!)

Gelman & Hill. Data Analysis Using Regression and Multilevel/Hierarchical Models

- Excellent intro chapters on regression, logistic regression, and mixed effects models. This book is where many people learn how to use mixed models, especially Bayesian models.

Supplemental papers

Optional readings that highlight or expand upon course concepts will occasionally be referenced or handed out. These readings may be useful for review but understanding of their content will not be explicitly assessed. They can also be used to complete the Reading Reinforcement Assignments (RRAs). I will make it clear if the content of a reading is required or optional. A bibliography containing many of these papers appears at the end of the syllabus.

Online materials & Social media.

This course will rely heavily on online materials, including online open-access materials I have written for the course, online R communities, and social media. Use of Twitter will be required for some assignments. I prefer to keep all online discussion on Twitter but would be happy to make a Facebook group for the class also.

Course website

The [course website](https://brouwern.github.io/BIOSC_1120/index.html) at https://brouwern.github.io/BIOSC_1120/index.html will be the central repository for all online and digital materials for the course.

Online workbooks & reading guides

I am developing several online resources for this course. These are all hosted via my GitHub account at <https://github.com/brouwern>

- [R for Ecological Data Science](https://github.com/brouwern/BOOK_R_Ecological_Data_Science): github.com/brouwern/BOOK_R_Ecological_Data_Science
- This is the primary workbook and lab manual for the R exercises used in this course
- A [Reading Guide to Motulsky's Intuitive Biostatistics](https://brouwern.github.io/IBSguide/): brouwern.github.io/IBSguide/
- A Reading Guide to Whitlock & Schluter's Analysis of Biological Data

R Packages & Interactive Tutorials

I am developing several R packages that contain datasets, interactive tutorials, and useful code.

- wildlifeR: brouwern.github.io/wildlifeR/
- datasets & exercises related to wildlife ecology
- abdcpanion: github.com/brouwern/abdcpanion
- datasets & interactive tutorials related to Whitlock and Schulters Analysis of Biological Data
- ibscompanion
- datasets & interactive tutorials related to Motulsky's Intuitive Biostatistics

Methods of Evaluation & Assessment

NOTE: Method of evaluation and points are subject to revision. Students will be notified verbally and via email if there are any changes to the syllabus.

Students will be evaluated through participation, weekly pre-class assignments, problem sets, group assignments, and an independent project related to their own work. Each element is summarized below. See the attached table for a breakdown of the points and the individual assignment handouts for further details.

In-class Participation

Attendance is required for all class sessions and students are required to stay for the entire session. Attendance will be recorded and participation (eg asking or answering questions during discussion, asking questions about your R work etc) will be scored each session. If you cannot attend a class or must leave early due to special circumstances please contact me in advance if possible so we can make plans for you to complete the necessary work.

Required office hours

Students are required to meet with the professor at least 2 times to discuss their progress in the course and/or how the course relates to other aspects of their studies or research. This can be done during office hours or over lunch or coffee; a good conversation in the hallway will likely also suffice. If necessary it can be done over the phone or Skype. Upon approval extensive email conversation can substitute for 1 meeting.

Weekly pre-class assignments

Coming prepared is essential for making class time as productive as possible. **Students must read all assigned readings.** Additionally, pre-class assignments will be due each week. Assignments will be typically be based on chapters from the assigned text, *Intuitive Biostatistics*. These assignments are summarized below. See the individuals handouts for each assignment for complete details.

Reading-reinforcement assignments (RRAs)

Most weeks students will be required to complete an assignment to help you engage with and extend the concepts discussed in the course text *Intuitive Biostatistics*. These assignments should generally be easy to complete while reading the book and students will be able to select from different options that best suit their interests. Two of these assignments, however, will require some research and writing. See the “**Reading Reinforcement Assignments (RRAs)**” handout for a full outline of all the assignments and the individual handouts for each assignment.

Pre-lab Computation Tutorials & Computation Challenges

For several chapters of *Intuitive Biostatistics* I will create a visual walkthrough which demonstrates how to carry out the major computations in a spreadsheet program and/or R. Students will be required to complete these tutorials prior to class and complete a worksheet while they work on the tutorials which is due at the beginning of class each week. Computation exercises will cover:

- Binomial confidence intervals (IBS: Ch 4)
- Binomial sampling distributions (ABD: Ch6; IBS: Ch 27)
- 2-sample t-tests (IBS: Ch 30)
- Chi-squared test (IBS: Ch 28)
- Correlation (IBS: Ch 32)
- Regression (IBS: Ch 33)

Prelab “waRmup” exercises

During weeks with no pre-lab computation tutorials/challenges, I will assign short “prelab” exercises which emphasize R code and the interpretation of R output. These prelabs will be due at the beginning of class.

In-class worksheet

Most weeks extensive time will be spent in-class working through simulations and tutorials in R. Each simulation or tutorial will have an accompanying worksheet. Students will have 1 week from when it was begun to complete these worksheets.

Twitter assignment

Twitter will be used for communication in the course. Students will be required to create an account and tweet out questions and other info relevant to the course.

Independent project

Each student will identify a dataset and carry out an independent investigation, culminating in a draft of short scientific paper focused on the methods and results section, and appropriate graphs, tables and appendices. The project will be split into 3 parts.

1. Early in the semester students will be required to submit a proposal outlining the dataset they wish to use and the scientific question of interest.
2. By mid-term students will submit an outline of the introduction, a methods section, and initial descriptive statistics and exploratory graphs of their data.
3. The final paper will consist of a complete scientific manuscript and annotated R code for all analyses.

Though the paper needs to have all of the relevant parts, the introduction and discussion need only briefly outline the biological questions and conclusions. Students are encouraged to use data from their own research or obtain data from their adviser. I can also discuss with students data sets that I have from my forest ecology and avian ecology work. Students are encouraged to work collaboratively in determining the appropriate analysis approaches and to help each other with their R code, but each student must analyze a separate data set. Key to this assignment will be that students work on dataset that allows them to use the focal methods of this class and can be completed. I will work with students to determine if their datasets are appropriate; if necessary we will focus the analysis or dataset to facilitate this. For example, if your dataset only requires a t-test to complete then we will likely find a way to extend or augment the data, question, or analysis so you practice using the full breadth of methods used in the course. In contrast, if the data requires a generalized linear mixed model with repeated measures, we will likely find a way to simplify it.

Grade scale

- A = 93% and above
- A- = 90-92.9%
- B+ = 87-89.9%
- B = 83-86.9%
- B- = 80-82.9%
- C+ = 77-79.9%
- C = 73-76.9%
- C- = 70-72.9%
- D = 60-69.9%
- F = 59.9% and below

Course policies

Email:

Email will be the primary, official form of communication outside of class.

- Email from me to you: You must check your email regularly. If you use a personal email address please arrange to have your university emails forwarded there. For example, I have my University email forwarded to my main email address, brouwern@gmail.com. I will try to include BIOSTATSS in the subject.
- Email from you to me: In the subject line please remember to include BIOSTATS-[your name]-[what its about] For example BIOSTAT-Jude Brouwer-birthweight outliers. During the week I will try to answer emails the day they are received. On weekends my responses may be delayed.

Please note that the University of Pittsburgh Email Communications Policy states: *“Each student is issued a University email address (username@pitt.edu) upon admittance. This email address may be used by the University for official communication with students. Students are expected to read email sent to this account on a regular basis. Failure to read and react to University communications in a timely manner does not absolve the student from knowing and complying with the content of the communications. The University provides an email forwarding service that allows students to read their email via other service providers (e.g., Hotmail, AOL, Yahoo). Students that choose to forward their email from their pitt.edu address to another address do so at their own risk. If email is lost as a result of forwarding, it does not absolve the student from responding to official communications sent to their University email address. To forward email sent to your University account, go to <http://accounts.pitt.edu>, log into your account, click on Edit Forwarding Addresses, and follow the instructions on the page. Be sure to log out of your account when you have finished.”*

G grades

If you wish to petition for a G grade, you must submit a request for this grade in writing to Dr. Brouwer, and you must document your reason(s). You will be required to make arrangements for the specific tasks that you must complete to remove the G grade. Remember that G grades, according to SAS guidelines, are to be given only when students who have been attending a course and have been making regular progress are prevented by circumstances beyond their control from completing the course after it is too late to withdraw. If you miss the final exam, you may receive a G grade if the above conditions are met.

Course Social Media Accounts:

Twitter use is required for this course, at least during class meeting. I can create a Facebook group associated with the course if there is interest.

Late work:

Late homework and assignments will be penalized 10% per day. Assignments are due when class starts and points will be deducted for any assignments handed in after class/lab begins.

Missed exams:

There are no exams :)

Withdrawal

Students are expected to do all assigned work and stay current in their studies. If circumstances arise that prevent a student from staying current with the material, the student should consider withdrawing from the course.

Please consult the University calendar for withdrawal dates.

Academic Integrity

Cheating and/or plagiarism will not be tolerated in this course. The minimum penalty will be a failing grade on the exam/homework/assignment/lab. You may also be barred from completing the course with a passing grade, and referred to the university administration for further disciplinary measures. In addition, your major department will be notified of the offense. Please consult the University academic integrity policy: www.as.pitt.edu/faculty/policy/integrity.html

Cheating takes many forms, including:

1. Any attempt to gain improper advantage in an academic evaluation (exam)
2. Academic impersonation (taking a test or doing an assignment for another person)
3. Plagiarism, including failing to properly cite previously published materials, including websites.
4. Improper research practices
5. Dishonesty in publication

Disability resources

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services (DRS), 140 William Pitt Union, (412) 648-7890, drsrecep@pitt.edu, (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course. As a reminder, students must schedule testing center services 72 business hours prior to the exam. Students who miss the deadline must take the exam in class and will not receive any additional time or accommodations.

Grading

See the attached point breakdown.

Extra credit & Make up assignments

There will be no extra credit offered in this class and no make-up assignments.

Statement on classroom recording

To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the Instructor, and any such recording properly approved in advance can be used solely for the student's own private use.

Extra credit

There will be no extra credit offered in this class.

Collaboration among students

Science is a collaborative effort and students are encouraged to discuss all assignments and to work together. However, all assignments must reflect your own work and contain your own articulation of answers and/or computations. Copying homework answers will be considered plagiarism and cheating.

Submitting assignments via email

Unless approved in advance all homework and other assignments must be submitted as a hard copy. In some cases homework will be submitted digitally with prior approval.