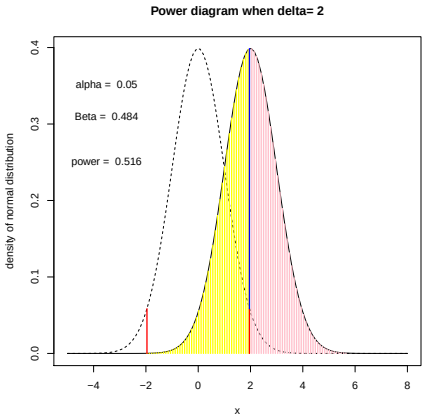
1 BIOL 611 Syllabus: Biometry

1.1 Overview and Learning Objectives

This course is intended as a rigorous introduction into statistical inference for biological data. The course will introduce and cover examples of probability distributions, descriptive statistics, and parametric and non-parametric hypothesis tests. Students should also become versed in automated data handling and basic elements of experimental design.

I expect that by the end of the semester each student should be able to take a dataset and along with explanation of the origin of those data, they should be able to summarize the data graphically and using descriptive statistics. As important, students should be able to construct and conduct hypothesis tests associated with the biological questions that motivate the data.

Finally, the basic principles of reproducible research and analysis will be introduced and reinforced throughout the semester.



1.2 Instructor

Allan Strand
Grice 216 (Lab: Grice 209)
Phone 953-9189
email: `stranda@cofc.edu`

1.3 Text and Web

I am not requiring a text for this course. However, I can recommend Gotelli and Ellison (2004) as an easy to read primer (I've used this book in previous semesters). For an encyclopedia of standard statistics, try Sokal and Rohlf (1995). Dalgaard (2002) will be mostly useful in the lab portion of the course. I'd recommend that you get a copy of this text (though it is also available for free online through the library). Lecture notes should also be available on the web at `http://linum.cofc.edu`. This web-site is password protected.[1] In addition to lecture notes, other course material will be available here.

1.4 Online accounts

See the technology guideline distributed for this course

[1] The username is: bio611 and the password is: see intro email

1.5 Computer software

Two statistical analysis packages are available in the computer room at Grice. The first is 'JMP', produced as a 'point and click' tool by the SAS Institute. The second is R, a freely available implementation of the 'S+' statistical language developed at Bell Labs. Class examples and lab exercises will use R. Many of you may have already used SigmaPlot as well. It is a nice graphics program, though R has more flexibility and reproducibility. Finally, both Excel and LibreOffice.org (essentially identical to OpenOffice.org) 'calc' are available on these computers.

We will be using R this semester. I encourage you to install R and Rstudio on your own computer, but I will provide an online R environment for everyone in the class.

You should also have easy access to a spreadsheet program. Excel is fine, but google sheets is a nice alternative that is available to anyone with a web-browser.

1.6 Labs

Labs will implement concepts discussed in lecture using concrete examples in R. Every Tuesday and Thursday lab sections will meet from 10:15=1:15. It is important that the class is divided evenly across the two sections. Students will be confirmed in lab sections the first day of class.

1.7 Assignments

1.7.1 Exams

There will be two cumulative exams. Given the type of course that Biometry is by necessity, these will be take-home exams. You are on you honor to work *completely alone* on these exams.

1.7.2 Problem sets

There will be two problem sets assigned this semester. They will require you to take the concepts in class and apply them in a relatively low-pressure, relaxed(!?!) environment. I will provide a series of questions, usually associated with data. Expect these problem sets to be challenging and to require independent thought. They will not consist of regurgitation-type questions. At the same time, I fully intend to provide you with the tools and skills needed to solve these problems. Finally, unlike the exams, I expect and encourage you to collaborate with your peers on these problems. I do expect you to each perform and report the analyses individually, however.

1.7.3 Final Projects

You will be required to take an empirical dataset and (re)analyze the data to address a biological process of their choosing. I expect that most students will derive data from the

primary literature, although if you can find a dataset from your advisor, potential advisor, or elsewhere, that would be great. Projects will be presented using suitable visual aids (Powerpoint may seem easiest, but not the only way to go). All students will write critiques of each project presentation. These critiques will comprise 20% of the project total. Due to the size of the class, presentations will be limited to 10 minutes each.

1.7.4 Percent Weights

Assignment	Weight
Problem sets (total)	40
Exams 1 & 2	30
Project	20
Participation	10

1.8 Attendence

It is rather ridiculous that this language has to be included in the syllabus of a graduate class: The difficulty of this class will increase exponentially as you miss classes. Furthermore, if you miss either lecture or lab, please do not expect the instructor to "re-lecture" at your convenience.

Having said that, you are all professionals who are seeking training. It is really up to you how much effort you choose to put into the course; I will not be taking attendance.

Accommodating Disabilities

I stand ready to help students with disabilities succeed in this course and at the College of Charleston in general. We do have an office of disability services and I do encourage students to reach out to that office. I will also provide accommodations in this course while protecting privacy, if you reach out to me directly.

Academic honesty

Don't cheat. If you have to ask yourself if you are skirting too close to academic honesty rules, you should engage in serious self-evaluation. You are a graduate student because you have signaled that you consider yourself a professional scientist. Cheating will not help you reach your professional goals–one of the most important characteristics you each bring to your career is integrity. If you lose your integrity you will not truly succeed over the long-term in science. It's hard enough without having the community doubt your results, analyses, and writing.

To reiterate information from the assignments section above: I will ask you to work collaboratively on Problem Sets this semester. You must, however, *hand in your own paper that you wrote* and should only rely on your colleagues for approaches to problems, not

for the details of implementation and reporting. You have to work alone on Exams. Any resource other than other students can be used to help with exams, however.

I will prosecute students who do not work independently on exams or who copy text and code rather that write it themselves on problem sets.

2 Lecture Schedule

I reserve the right to deviate from the lecture topic schedule. I've included relevant chapters from Gotelli and Ellison (2004) and Dalgaard (2002), but remember these are suggested, not required, texts.

Week	Date	Day	Topic	Gotelli.Ellison..1st.ed.	Daalgard
1.00	2020-08-25	tu	Intro; Syllabus; Reproducible research; What is Data?	8	NULL
	2020-08-27	th	What is Data, Computers in Statistics,R	8	NULL
2.00	2020-09-01	tu	Descriptive Statistics	3	1; 3
	2020-09-03	th		NULL	NULL
3.00	2020-09-08	tu	Probability Distributions	1514851200	2
	2020-09-10	th		NULL	NULL
4.00	2020-09-15	tu	Hypothesis testing; experimental design	4	4
	2020-09-17	th		NULL	NULL
5.00	2020-09-22	tu	Methods of estimation and model fitting	5	NULL
	2020-09-24	th		NULL	NULL
6.00	2020-09-29	tu	Building a 2-sample MLE	NULL	NULL
	2020-10-01	th	Problem set 1 available at 5pm	NULL	NULL
7.00	2020-10-06	tu	ANOVA	6,7,10	7; 6.1
	2020-10-08	th	Problem set 1 due at 5pm	NULL	NULL
8.00	2020-10-13	tu		NULL	NULL
	2020-10-15	th		NULL	NULL
9.00	2020-10-20	tu	ANOVA; Exam 1 available at 5pm	NULL	NULL
	2020-10-22	th	Mixed model ANOVA; Non-parametric analogues	10	6.3
10.00	2020-10-27	tu	Regression; non-parametric analogues; Exam 1 Due	9	6.2; 6.4
	2020-10-29	th	Multiple Regression; GLMs; MUST HAVE DATASET FOR FINAL PROJECT	NULL	NULL
	2020-11-03	tu	Election Day–GO VOTE!	NULL	NULL
11.00	2020-11-05	th	Regression, maximum-likelihood approaches	NULL	NULL
12.00	2020-11-10	tu	Analysis of Frequencies	11	NULL
	2020-11-12	th		NULL	NULL
13.00	2020-11-17	tu	small introduction into multivariate statistics; Problem set II avail at 2:27am	12	8
	2020-11-19	th	Presentations (need two days); Exam II available	NULL	7
14.00	2020-11-24	tu		NULL	NULL
	2020-11-26	th	THANKSGIVING	NULL	NULL
15.00	2020-12-01	tu		NULL	NULL
16.00	2020-12-03	th	Lecture based on popular demand; Problem set II due (last day of class)	NULL	NULL
	2020-12-08	tu	Exam II due; 5 pm	NULL	NULL

3 Lab Schedule

Week	Lab.Topic	Daalgard
1.00	Getting Tech in order (everyone welcome)	
2.00	Intro to 'R'; import data; Reproducible research; Description of Data	1; 3
3.00	Probability distributions	2
4.00	Determining error rates; hypothesis testing	4
5.00	OLS; Monte-Carlo; MLEs	
6.00	Final Project Development	
7.00	Specifying ANOVA models	7; 6.1
9.00	nlme and non-parametric tests	
10.00	Specifying regression models	6.2; 6.4
11.00	Likelihood and regression	
12.00	Contingency Table Analyses	
13.00	Multivariate analyses (clustering, how to implement pca)	8
	Final presentation spillover, problem set questions	7

Stuff I am required by the College of Charleston to include in a syllabus, but that I think is a waste of your time

If I don't include them, however, I can be fired:
https://www.insidehighered.com/news/2016/08/08/should-professor-lose-his-job-over-ref

Pre- and Co- requisites

If you are a graduate student at the College of Charleston, you can take this class. That is not to say that you would not benefit from undergraduate courses in statistics and mathematics.

There are no co-requisites.

Storm Plan

If classes are cancelled for a storm, we will readjust the syllabus. This happens most Falls. However, because the course is set up online throughout the semester, I believe that cancellation may last for a shorter period of time than would occur with in-person classes.

I plan to hold to our normal schedule unless the college officially closes. So in other words, we will (as a group) employ common sense to decide how to adjust to storm closures, incluing taking advantage of scheduled storm days.

OAKS

We will use many online learning, research, and communication tools this semester. OAKs will not really be an important player in this collection of tools. I am required to upload

a syllabus to OAKs and I will do so. Other content will be available through the course web-page, slack, and our RStudio server (see technology guidelines).

Learning Outcomes

By the end of the semester students will be able to

- Calculate goodness of fit statistics for discrete biologically-derived datasets.
- Describe single and multiple variable environmental quantitative data numerically
- Describe single and multiple variable environmental quantitative data graphically
- Conduct and interpret 2-way analysis of variance.
- Conduct and interpret single variable linear regression using least-squares and maximum likelihood.
- Conduct and interpret two-sample tests using least-squares and maximum likelihood.

It is not clear to me that if you do these things that you will actually understand statistics, but you will have committed facts to memory for a while. At the same time, a bureaucrat somewhere will be somehow relieved. And feeding bureaucracy is an essential skill that we all need to master. In my opinion, it would be a pedagogical disaster if we all believe that the outcomes of any course can be boiled down to a short list of bullet points.

The most important point (one unfortunately, *not* required by the College of Charleston)

I hope that over the semester, you will develop a less task-oriented and more nuanced view of statistical inference than emphasized in the previous section. That's when you start to really love the topic. Lots of students have done this in the past, *you can do it too*. I'm looking forward to the semester and helping you along that path.

References

Dalgaard, P. (2002). *Introductory Statistics with R*, Academic Press, Inc Springer. book goes with R package ISwR. 1, 4

Gotelli, N. and Ellison, A. (2004). *A Primer of Ecological Statistics*, Sinauer. 1, 4

Sokal, R. R. and Rohlf, F. J. (1995). *Biometry. The principles and practice of statistics in biological research*, third edn, W. H. Freeman, San Francisco,. 1