

1 EVSS 624 Syllabus: Biometry

1.1 Overview and Learning Objectives

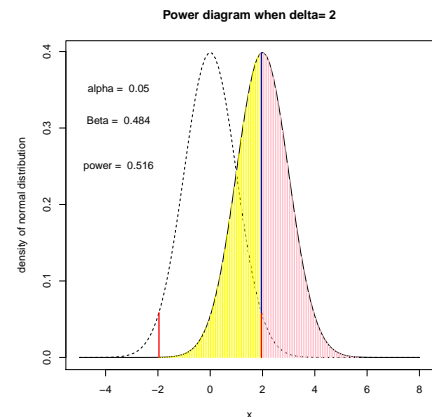
This course is intended as a rigorous introduction into statistical inference for biological data. The course will introduce and cover examples of probability distributions, descriptive statistics, and parametric and non-parametric hypothesis tests. Students should also become versed in automated data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

I am going to distribute data via google drive some this semester. It is in your best interest to obtain a google account so that you can do the same. As part of a reproducible research

¹The username is: bio611 and the password is: fisher

effort, I will also ask you to set up an account with github.org.

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.

If you have your own laptop, I **strongly** encourage you to install the following open-source software:

- R (latest version, currently v3.4.3) <http://cran.r-project.org>. R is an implementation of a statistical programming language.
- RStudio Desktop version (latest version, currently v1.1.383) <http://rstudio.org>². You need both R and R-studio for your platform. Rstudio is merely a front-end (or interface) to the R programming language. You don't technically need it, but your life (and that of your scientific colleagues) will be so much improved if you use it.
- LibreOffice (latest version) libreoffice.org (if you have word and excel, that will be ok, though; basically you need a decent spreadsheet program)

Finally, I know a bit about computer systems in general, but very little about different flavors of windows and macos. It is your responsibility to make your computer talk to the college's wireless network (instructions available in the Grice computer room). Also it's your responsibility to install software yourself. *Your most important research tool is your brain, bar none. However, your computer is an extremely important research tool as well. If you are considering a career in science, you should quickly become versed in how to manage software and files on the computer, not just for this class, but for your training in general.* I'll help if I can, but I am not a computer support technician.

1.6 Labs

Labs will implement concepts discussed in lecture using concrete examples in R. Every Tuesday and Thursday lab sections will meet from 10:30-1:30. 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.

²Later versions of rstudio are very important for some of the coursework

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 your 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 Attendance

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.

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	1/9/2018	tu	Intro; Syllabus; Reproducible research; What is Data?	8	
	1/11/2018	th	What is Data, Computers in Statistics,R		
2	1/16/2018	tu	Descriptive Statistics	3	1
	1/18/2018	th			
3	1/23/2018	tu	Probability Distributions	1, 2	3
	1/25/2018	th			
4	1/30/2018	tu	Hypothesis testing; experimental design	4	2
	2/1/2018	th			
5	2/6/2018	tu	Methods of estimation and model fitting	5	4
	2/8/2018	th	Building a 2-sample MLE		
6	2/13/2018	tu	Problem set 1 available at 5pm		
	2/15/2018	th	Exam Help		
7	2/20/2018	tu			
	2/22/2018	th	ANOVA	6,7,10	7; 6.1
8	2/27/2018	tu	Problem set 1 due at 5pm		
	3/1/2018	th	ANOVA; Exam 1 available at 5pm		
9	3/6/2018	tu	Mixed model ANOVA; Non-parametric analogues	10	6.3
	3/8/2018	th	Regression; non-parametric analogues; Exam 1 Due	9	6.2; 6.4
10	3/13/2018	tu	Multiple Regression; GLMs; MUST HAVE DATASET FOR FINAL PROJECT		
	3/15/2018	th			
11	3/20/2018	tu	Spring Break		
	3/22/2018	th	Spring Break		
12	3/27/2018	tu		11	
	3/29/2018	th	Analysis of Frequencies		8
13	4/3/2018	tu			
	4/5/2018	th	small introduction into multivariate statistics; Problem set II avail at 2am		7
14	4/10/2018	tu			
	4/12/2018	th	Lecture based on popular demand	12	
15	4/17/2018	tu	Presentations (need two days); Exam II available		
	4/19/2018	th	Problem set II due		
Exam	5/1/2018	tu	Exam II due		

3 Lab Schedule

Week	Lab Topic	Daalgard
1	Getting Tech in order	
2	Introduction to Excel, 'R/S+'; import data; Reproducible research	1
3	Description of Data	3
4	Probability distributions	2
5	Determining error rates; hypothesis testing	4
6	OLS; Monte-Carlo; MLEs	
7	Final Project Development	
8	Specifying ANOVA models	
9	nlme and non-parametric tests	6.3
10	Specifying regression models	
12	regression cont	
13	Contingency Table Analyses	
14	Multivariate analyses (clustering, how to implement pca)	
15	presentation spillover	

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.

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 quantitative data numerically
- Describe single and multiple variable 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.
- Conduct and interpret tests of association and independence in categorical data

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](#)