

Special Topics: Biostatistics (in R)
PCB 4454, CRN 10501

Version 1.1
Updated: 1 January 2018

Elective for biology majors
(may be accepted for mathematics, marine science, and environmental science).

Quantification, including statistical analyses, is one of the unifying characteristics of science, and therefore biology. To fully understand a study, we must be able to evaluate data objectively. In science, we accomplish this typically by quantifying our data, regardless of whether those data are continuous or categorical; normal or non-normal; complete or incomplete in nature. We then evaluate potential patterns in the data using statistical inference. In many ways, the explosion of science, including biology, over the past 100 years is directly related to the advent of modern statistical techniques and the associated inference. As much as any other piece of technology, the development of statistics has allowed us to understand the structure of life, the universe, and everything. Yes; as we all know, the statistical answer to the world is 42 (read Douglas Adams!!!!). Remember – modern science began a little over 500 years ago. However, modern statistics are only 100 some odd years old. I suspect that we would all trust the drugs, medicine, and therapy of the past 100 years over the treatments that people received in the 1500's. Statistics has helped revolutionize science.

In this introductory biostatistical course, we will explore the use of statistical methodologies in designing, analyzing, interpreting, and presenting biological experiments and observations. We will cover descriptive statistics, elements of experimental design, probability, hypothesis testing and statistical inference, analysis of variance, correlation, regression techniques, non-parametric statistical methods, permutation procedures, and other approaches (time permitting). Throughout the course the application of statistical techniques within a biological context will be emphasized, using data from laboratory and field studies; we will focus on student generate data where appropriate. Most importantly, you will learn how to understand statistical procedures instead of memorize formulae and derive equations. This is not a traditional mathematics course, where formula derivations are stressed. Instead, we will explore the application of statistical ideas and methods to the design and interpretation of biological experiments and comparative data.

Ultimately, this class will help you avoid statistical illiteracy!!!!!!!!!!

Instructor: Dr. Charles (billY) Gunnels
Edwards Hall 214b
Email: cgunnels@fgcu.edu
Phone: 590-7210
Office Hours: Wednesday 12:30 – 2 pm or by appointment
I will likely need to change office hours on a weekly basis. Updates will be announced on Canvas.

Credits: 3

Prerequisites: (BSC 1010C with a minimum grade of C or (BSC 1010 with a minimum grade of C and BSC 1010L with a minimum grade of C) and (BSC 1011C with a minimum grade of C or (BSC 1011 with a minimum grade of C and BSC 1011L with a minimum grade of C) and (STA 2023 with a minimum grade of C or STA 2037 with a minimum grade of C)

Class Time: T & Th 1:30 – 2:45 pm in Reed Hall 146
Cumulative Final Exam: Due by Tuesday, May 2nd at 02:45pm

Sometimes required but always encouraged.

You are adults; I expect you to make the responsible decisions. I expect you to attend all classes because each day provides you the opportunity to learn. If you must miss a class, make sure that you keep up with the material. I will not be able to reteach the class in office hours.

Required Programs and Texts:

R (<http://cran.r-project.org/>)

This is by far the best and most powerful statistical software that is available. It is an open-source statistical language, with 1000's of packages. Most importantly, it is FREE. There is a learn curve – a steep, but manageable, learning curve. It will be well worth the journey.

We will use R (and the many associated packages) to analyze data, which we can then interpret.

While much of what we will need comes from the Base package, I will draw from many packages (free add-ons) throughout the semester, including:

- psych
- ggplot2
- vcd
- coin
- lmerPerm
- circular

Excel

I know; I know. Excel is not a favorite for many people. In fact, many people who are quantitatively inclined and many people who are quantitatively disinclined hate Excel equally. However, this is a great program to organize and manipulate data. You can write simple codes; you can create simple graphs; and most importantly, you can organize complex data sets easily, quickly, and accurately.

We will use Excel to organize and manipulate data. We will then import our data sets into R.

Dalgaard, Peter. 2008 *Introductory Statistics with R*. Springer.

This is as good as it gets for an introductory stats book that uses R. R is such a powerful and wonderful platform that we need to use an appropriate text. I like this book for three major reasons. One, the book does a great job showing you coding examples that let your work in R; two, the book draws most of its examples from biological and sociological data sets; and three, the book combines parametric and non-parametric approaches into the same section, which makes understanding the relationship between both approaches much easier.

We will use this book for the organization, coding examples, and problem sets.

Motulsky, Harvey. 2013. *Intuitive biostatistics: a nonmathematical guide to statistical thinking*. Oxford University Press.

This is the best book that I have ever read about statistics. The book focuses on statistical logic and concepts that underlie statistical thinking. It is simply brilliant. Some will love it immediately, while others will come to enjoy it in future years.

We will use this book to drive our discussions, lectures, and conversations about statistics.

Additional Sources

Quick-R: Accessing the Power of R (<http://www.statmethods.net>)

This is without question the best website that I have found to learn and practice with R. This is my go-to location when I want to figure out a new procedure. The website does not have everything, and some instructions are incredibly curt. However, it is exceptional.

Kabacoff, Robert I. 2015. *R in Action, Second Edition, Data analysis and graphics with R*, Manning Publication, ISBN 9781617291388

This is my statistical dictionary. In time, this will become your go-to reference. And, yes, it is written by the creator of Quick-R: Accessing the Power of R.

Google

You may have heard of this new fangled search engine – Google. It does a great job compiling all of the ideas and solutions that people generate throughout the world. All you need to do is type in a question in the search engine, punctuation (and spelling) not required. At this point, Google will scour the

“internet”, providing you all of the times that such a query has been posted on the net. I use this every time that I work with R.

You

Always make sure to save your coding notes. You will forget your solutions. Your notes will provide you with the reminders that you need to work efficiently and successfully.

Student Learning Outcomes:

- 1) Skills
 - a. Critical Thinking
 - b. Mathematical/ Statistical Literacy
 - c. Quantitative Communication
 - d. Written Communication
 - e. Professionalism
 - f. Scientific Process
- 2) Content
 - a. Excel Data Manipulation
 - b. R
 - c. Calculate summary statistics
 - d. Calculate common probability distributions and apply those calculations to solve problems based on biological studies
 - e. Randomly allocate experimental units to treatments and apply this technique to solve problems based on biological studies
 - f. Calculate the distribution of observations about the mean based on the assumption of normality and apply those calculations to solve problems based on biological data
 - g. Calculate the distribution of sample means about the mean and apply those calculations to solve problems based on biological data
 - h. Design simple biological experiments
 - i. Compare two means (from paired and unpaired data) using both parametric and non-parametric methods and use those methods to test hypotheses
 - j. Analyze categorical data to test both goodness-of-fit and contingency hypotheses
 - k. Compare more than two means using analysis of variance methods and use those methods to test hypotheses derived from both single-factor and two-factor experimental designs
 - l. Calculate least-squares regression lines and apply those calculations to solve problems based on biological studies.

Canvas: Important material is provided on Canvas. Please consult Canvas for up-to-date information, deadlines and additional information for lectures. Emails should be sent through Canvas.

Evaluation: Your grade in the course is based on two equal parts: laboratory and lecture.

- 1) Two Exams (40%).
 - a) There will two lectures exams. See schedule for dates and times.
- 2) Final Exam (30%)
 - a) The final will be comprehensive. There will be a disproportionate focus on material that comes from the final few weeks of the course.
- 3) Homework (20%)
 - a) All homework assignments will be due at the beginning of the specified class, unless stated otherwise. 1 letter grade will be deducted for late assignments. Late assignments can be turned in for up to one week after the assignment's due date.
- 4) Participation (10%)
 - a) Here are the behaviors that count:
 - i) Asking questions
 - ii) Answering questions
 - iii) Making comments
 - (1) (Extra recognition for comments that relate to material in the text, and for sharing relevant experiences)
 - b) Here are the value-added behaviors—the ones the put your contributions over the top:
 - i) Responding to something another student says (including answering a question asked by a student)
 - ii) Constructively disagreeing with something in the text or said in class by me or another student

- c) I will post my impression sometime between the 1/2 to 2/3 mark in the semester.
- d) Finally, while attendance is not mandatory. Attendance does influence your participation score indirectly; you cannot participate unless you are in class.

Please Note: Lectures and materials provided for this course are the property of the instructor and may not be used for any commercial purposes. In addition, material may not be posted on the web or any social media.

Feel free to make audio (or video) recordings of class for your personal use.

Honesty: It is expected that you will show ethical behavior concerning all of your class work. Students are expected to do their own work, use their own words and reference all outside sources. Failure to uphold the standards of academic honesty will result in the appropriate disciplinary action.

Academic Behavior Standards and Academic Dishonesty: All students are expected to demonstrate honesty in their academic pursuits. The university policies regarding issues of honesty can be found in the FGCU Student Guidebook under the Student Code of Conduct and Policies and Procedures sections. All students are expected to study this document, which outlines their responsibilities and consequences for violations of the policy. The FGCU Student Guidebook is available online at <http://studentservices.fgcu.edu/judicialaffairs/new.html>

Disability Statement: Florida Gulf Coast University, in accordance with the Americans with Disabilities Act and the university's guiding principles, will provide classroom and academic accommodations to students with documented disabilities. If you need to request an accommodation in this class due to a disability, or you suspect that your academic performances is affected by a disability, please see me or contact the Office of Adaptive Services. The Office of Adaptive Services is located in Howard Hall, room 137. The phone number is 590-7956 or TTY 590-7930.

Center for Academic Achievement Syllabus Statement: The Center for Academic Achievement (CAA) offers academic support services for any FGCU student. The services are at no extra charge to students and include: peer tutoring, Supplemental Instruction, Student Success Workshops, and individualized academic coaching. If you would like to participate in or learn more about these services, please visit the CAA in Library 103. You may also email the CAA at caa@fgcu.edu or call at (239) 590-7906. The CAA website is <http://www.fgcu.edu/caa>.

Letters of Recommendation: Students seeking a letter of recommendation should provide a signed letter waiving their rights to keep their academic record in this class private. This waiver allows faculty members to discuss your record; grades; work ethic; as well as your profession, personal, and intellectual ability. I will not be able to write an honest recommendation without this waiver.

Student Observance of Religious Holidays: All students at Florida Gulf Coast University have a right to expect that the University will reasonably accommodate their religious observances, practices, and beliefs. Students, upon prior notification to their instructors, shall be excused from class or other scheduled academic activity to observe a religious holy day of their faith. Students shall be permitted a reasonable amount of time to make up the material or activities covered in their absence. Students shall not be penalized due to absence from class or other scheduled academic activity because of religious observances. Where practicable, major examinations, major assignments, and University ceremonies will not be scheduled on a major religious holy day. A student who is to be excused from class for a religious observance is not required to provide a second party certification of the reason for the absence.

Tentative Schedule: All aspects of the schedule may be modified. This class is designed to maximize your learning. I will make changes that enhance your learning. Please check Canvas and class notes for any changes.

Required readings and homework will be assigned each week.

Week	Date	Special Announcements	Topics	Dalgaard Reading	Motulsky Readings
1	Jan 9		Introduction (Gunnels returns from Trinidad and Tobago)		
1	Jan 11		Statistical (Non)Intuition		1
2	Jan 16		Populations, Sampling, and Research		3
2	Jan 18		Excel Manipulation		
3	Jan 23		Excel Manipulation		
3	Jan 25		Introduction to R	1 & Appendix A	
4	Jan 30		Introduction to R	1 & Appendix A	
4	Feb 1		Descriptive Statistics and Variation	4	8, 4, 9 & 12
5	Feb 6		Probabilities & Distribution	3	2, 10, 11
5	Feb 8		P-Value		15, 16, 18 19 & 20
6	Feb 13		P-Value		15, 16, 18 19 & 20
6	Feb 15		The 3 Evils Normality, Outliers & Transformations		24, 25
7	Feb 20	Exam 1			
7	Feb 22		One Sample Tests & Two Sample Tests	5	22, 23, 30, 31, 41
8	Feb 27		One Sample Tests & Two Sample Tests	5	22, 23, 30, 31, 41
8	Mar 1		One Sample Tests & Two Sample Tests	5	22, 23, 30, 31, 41
9	Mar 6	Spring Break			
9	Mar 8	Spring Break			
10	Mar 13		One Sample Tests & Two Sample Tests	5	22, 23, 30, 31, 41
10	Mar 15		K-Sample Tests	7	39, 41
11	Mar 20		K-Sample Tests	7	39, 41
11	Mar 22		Two-Way Tests	7	39, 41
12	Mar 27		Two-Way tests	7	39, 41
12	Mar 29		Count Data Tests	8	27, 28
13	Apr 3		Regression & Correlation	6	32, 33, 35
13	Apr 5		In-Class Data Analysis Exercise		
14	Apr 10	Exam 2			
14	Apr 12		PCA/ Factor Analyses		
15	Apr 17		Circular Statistics		
15	Apr 19		Logistic Regression		
16	Apr 24		Survival Curve		
15	Apr 26		Onward and Upward		
16	May 1	Final Exam (The in-class exam will be taken starting at 12:30 on Tuesday in Reed Hall 146)			