University of Florida
College of Public Health & Health Professions Syllabus
PHC 6068: Biostatistical Computing (3 credit hours)
Fall, 2018
Delivery Format: on-campus/online
Course Website and E-Learning

Instructor Name:	Zhiguang Huo
Room Number:	CTRB, Room 5235
Phone Number:	(352)-294-5929
Email Address:	zhuo@ufl.edu
Lecture time:	Mon 12:50 – 14:45 Wed 13:55 – 14:45
Teaching Assistants:	Hanz Gao gaohanzhi@ufl.edu
Office Hours:	HZ: Monday 3-5pm, CTRB, Room 5230 GH: Thursday 4-5pm, CTRB, Room 5254-C08

Preferred Course Communications: Lecture notes will be distributed on https://caleb-huo.github.io/teaching/2018FALL/biostatisticalComputing.html and assignments will be distributed through the canvas system at http://elearning.ufl.edu/.

Prerequisites

- PHC 6092: Biostatistical Theory,
- PHC 6050: Statistical Methods for Health Sciences Research I,
- PHC 6051: Biostatistical Methods II,
- Or permission of the instructor.

PURPOSE AND OUTCOME

Course Overview

This course is intended to develop your ability to perform statistical computing. The course will focus primarily on the R programming language using the RStudio interface, both of which are free and open-source software programs. The course will cover programming topics (vectorization, data input and output, data manipulation, GitHub and building R packages), statistical and computational methods (visualization, optimization, simulation, resampling, classification, clustering, regression and modern statistical methods such as LASSO and ElasticNet), and direct integration and dynamic reporting using LaTeX and R through programs such as knitr. Additionally, this course will include the use of high-performance computing resources at the University of Florida such as HiPerGator.

Relation to Program Outcomes

Students will develop the knowledge and skills to translate ideas and methods into workable software and interface with diverse data structures and objects, and write functions to implement statistical methods. These computational skills are essential for applied biostatistics.

Course Objectives and/or Goals

Upon successfully completing this course, students should be able to:

1. Convert an algorithm into a workable program and write functions that others can use and understand.
2. Smooth and visualize data, including ggplot2 functions for multi-panel displays.
3. Construct a simulation study and use it to evaluate the size and power of a statistical test or method.
4. Use resampling techniques such as the bootstrap and cross-validation to assess model fit and compare competing models.
5. Implement computational methods for optimization (e.g., Newton-Raphson), numerical integration (e.g., Monte Carlo integration), classification (e.g., LDA, SVM, tree-based methods, random forests), and regression (e.g., LASSO).
6. Build R packages and upload to the GitHub.

DESCRIPTION OF COURSE CONTENT

Topical Outline/Course Schedule

Week	Date(s)	Topic(s)	Assignments
1		Overview of statistical computing	
2		Basics of R programming Writing functions in R	HW 1
3		Vectorized calculations	
4		Using HiPerGator (Guest Lecture) R graphics (basic)	HW 2
5		R graphics (advanced) Data manipulation (Tidyverse)	
6		R packages Rcpp	Final project
7		GitHub Matrix operation	Midterm
8		Mid-term (Take home) Simulating random variables	HW3
9		Dimension reduction Clustering method	
10		Optimization Decision Tree	HW 4
11		Linear models Lasso and Ridge regression	
12		Generalized linear models Mixed models	HW 5
13		Bootstrap and Permutation	
14		Simulation	
15		Monte Carlo methods Cross validation	HW 6
16			Final exam
17		Final exam (Take home)	

Course Materials and Technology
There is no required text. You are also referred to the following texts as follow-up resources.

1. An Introduction to R. (From R official website). https://cran.r-project.org/doc/manuals/R-intro.html
2. Google's R Style Guide. https://google.github.io/styleguide/Rguide.xml
3. Hadley Wickham (2014). *Advanced R*. Taylor & Francis. http://adv-r.had.co.nz

The primary mechanism for communication in this course, other than class meetings, is conducted through the course website https://caleb-huo.github.io/teaching/2018FALL/biostatisticalComputing.html to post lecture notes and Canvas system https://ufl.instructure.com/ to deliver in class labs, assignments, final exams and grades. It is imperative that students familiarize themselves with Canvas, check Canvas frequently for possible announcements, and make sure that their e-mail account in Canvas is correct and active.

For technical support for this class, please contact the UF Help Desk at:

- Learning-support@ufl.edu
- (352) 392-HELP - select option 2
- https://lss.at.ufl.edu/help.shtml

ACADEMIC REQUIREMENTS AND GRADING

Final project:

The goal of the final project is to develop an R package that will be useful to other statisticians and R users. Students can either form groups (at most 3 people) themselves or work individually. Readable R documentations are necessary. Students will be encouraged to use higher level knowledge from the class to the R package. (E.g. Rcpp, vignette). Details instructions about the final project will be released on Wed Sep 26. The final R package and a short report of how to use the R package is due on Mon Dec 10 (11:59 pm).

Exam Policy

Both the midterm exam and the final exam will be "take home exam". You must work alone. You may only ask clarification questions from the instructor; you may not ask for hints to the questions since these are exam, not a homework assignment. Exams will be delivered in R Markdown format and must be submitted in R Markdown format. We will not grade exams in other formats. All the exams must be submitted electronically to the instructor and/or the teaching assistant of the class. **No hard copy is required.** Your responses must be supported by both textual explanations and the code you generate to produce your result.

Homework

All the homework problems will require R programming involving various statistical computational topics outlined before. Students will be required to use their own computers as well as HiPerGator in order to complete the assignments. All homework will be delivered in R Markdown format and must be submitted in R Markdown format. We will not grade homework in other formats. All the homework assignments must be submitted electronically to the instructor and/or the teaching assistant of the class. **No hard copy is required.** Your responses must be supported by both textual explanations and the code you generate to produce your result. You are allowed to discuss with your classmates, but you need to write your own code. The topics and dates of the homework assignments are provided in the previous table and the submission schedules and the credit distributions are provided below.

Late Homework policy:

Full credit will be given for assignments turned in on the due date (by 11:59pm). 80% credit for one day late. Assignments turned in the next school day after the due date will have a maximum possible credit of 80%. 50% credit for two days late. Assignment turned in two school days after the due date will have a maximum credit of 50%. NO credit is given after two days late. If you are out sick, no deduction will be taken if you inform me before the homework is due that you are ill. Please stay home and do not get other people sick. Just turn in your homework as soon as you can. If you are going to miss school on the day the homework is due (going out of town, religious holiday, etc.) please turn your homework in early. Each student is granted for two-day grace periods, which can be applied to any homework (not exam or final project). For all 6 homework, the total grace periods cannot exceed 2 days. Each time at least one day of grace period will be used. The instructor and the TA will check the last submission date and determine if you use the grace periods.

Grading

Requirement	Due date	Points or % of final grade (% must sum to 100%)
Homework 1	09/12/2018	5%
Homework 2	09/26/2018	5%
Mid-term exam	10/10/2018	20%
Homework 3	10/24/2018	5%
Homework 4	11/07/2018	5%
Homework 5	11/26/2018	5%

Homework 6	12/05/2018	5%
Final project	12/10/2018	20%
Final exam	12/14/2018	30%

Point system used (i.e., how do course points translate into letter grades).

Points earned	94-100%	90-93%	87-89%	83-86%	80-82%	77-79%	73-76%	70-72%	67-69%	63-66%	60-62%	< 60%
Letter Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	E

Please be aware that a C- is not an acceptable grade for graduate students. A grade of C counts toward a graduate degree only if an equal number of credits in courses numbered 5000 or higher have been earned with an A. In addition, the Bachelor of Health Science Program does not use C- grades.

You must include the letter grade to grade point conversion table below. Letter grade to grade point conversions are fixed by UF and cannot be changed.

Letter Grade	**A**	**A-**	**B+**	**B**	**B-**	**C+**	**C**	**C-**	**D+**	**D**	**D-**	**E**	**WF**	**I**	**NG**	**S-U**
Grade Points	4.0	3.67	3.33	3.0	2.67	2.33	2.0	1.67	1.33	1.0	0.67	0.0	0.0	0.0	0.0	0.0

For greater detail on the meaning of letter grades and university policies related to them, see the Registrar's Grade Policy regulations at:
http://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx

Policy Related to Make up Exams or Other Work
Please see the instructor as early as possible regarding a possible absence during an exam. Make-up exams due to an excused absence will be handled on an individual basis.

Please note: Any requests for make-ups due to technical issues MUST be accompanied by the ticket number received from LSS when the problem was reported to them. The ticket number will document the time and date of the problem. You MUST e-mail me within 24 hours of the technical difficulty if you wish to request a make-up.

All faculty are bound by the UF policy for excused absences. For information regarding the UF Attendance Policy see the Registrar website for additional details:
https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx

STUDENT EXPECTATIONS, ROLES, AND OPPORTUNITIES FOR INPUT

Expectations Regarding Course Behavior
Students are expected to spend an average at least 2-1/2 hours per week per credit hour on the course exclusive of class time. This time includes but is not limited to reading, research, preparations for class, team or group meetings (electronic or otherwise), and course deliverables.

Communication Guidelines
The preferred methods of communication for the course are messages in Canvas (see Course Materials above) or e-mail.

Academic Integrity

Students are expected to act in accordance with the University of Florida policy on academic integrity. As a student at the University of Florida, you have committed yourself to uphold the Honor Code, which includes the following pledge:

"**We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity**."

You are expected to exhibit behavior consistent with this commitment to the UF academic community, and on all work submitted for credit at the University of Florida, the following pledge is either required or implied:

"On my honor, I have neither given nor received unauthorized aid in doing this assignment."

It is your individual responsibility to know and comply with all university policies and procedures regarding academic integrity and the Student Honor Code. Violations of the Honor Code at the University of Florida will not be tolerated. Violations will be reported to the Dean of Students Office for consideration of disciplinary action. For additional information regarding Academic Integrity, please see Student Conduct and Honor Code or the Graduate Student Website for additional details:
https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/
http://gradschool.ufl.edu/students/introduction.html

Please remember cheating, lying, misrepresentation, or plagiarism in any form is unacceptable and inexcusable behavior.

Online Faculty Course Evaluation Process
Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at https://evaluations.ufl.edu. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at https://evaluations.ufl.edu/results/.

SUPPORT SERVICES

Accommodations for Students with Disabilities
If you require classroom accommodation because of a disability, you must register with the Dean of Students Office http://www.dso.ufl.edu within the first week of class. The Dean of Students Office will provide documentation of accommodations to you, which you then give to me as the instructor of the course to receive accommodations. Please make sure you provide this letter to me by the end of the second week of the course. The College is committed to providing reasonable accommodations to assist students in their coursework.

Counseling and Student Health
Students sometimes experience stress from academic expectations and/or personal and interpersonal issues that may interfere with their academic performance. If you find yourself facing issues that have the potential to or are already negatively affecting your coursework, you are encouraged to talk with an instructor and/or seek help through University resources available to you.

- The Counseling and Wellness Center 352-392-1575 offers a variety of support services such as psychological assessment and intervention and assistance for math and test anxiety. Visit their web site for more information: http://www.counseling.ufl.edu. On line and in person assistance is available.
- You Matter We Care website: http://www.umatter.ufl.edu/. If you are feeling overwhelmed or stressed, you can reach out for help through the You Matter We Care website, which is staffed by Dean of Students and Counseling Center personnel.
- The Student Health Care Center at Shands is a satellite clinic of the main Student Health Care Center located on Fletcher Drive on campus. Student Health at Shands offers a variety of clinical services. The clinic is located on the second floor of the Dental Tower in the Health Science Center. For more information, contact the clinic at 392-0627 or check out the web site at: https://shcc.ufl.edu/
- Crisis intervention is always available 24/7 from:
  Alachua County Crisis Center
  (352) 264-6789

http://www.alachuacounty.us/DEPTS/CSS/CRISISCENTER/Pages/CrisisCenter.aspx

Do not wait until you reach a crisis to come in and talk with us. We have helped many students through stressful situations impacting their academic performance. You are not alone so do not be afraid to ask for assistance.