

**PHC 6937 Introduction to Statistical Learning with Applications in Health Sciences
(3 credits)**

Semester: Fall 2022

Delivery Format: On-Campus

Instructor Name: Qing Lu

CTRB 5238

E-mail: lucienq@ufl.edu

Telephone: 352-294-5928

Class Meets: Mondays (9:35am-11:30am) and Wednesdays (10:40am-11:30am)
HPNP G105

Office hours: Monday and Wednesday (11:30am-12:30pm) HPNP G105

Preferred Course Communications: e-mail (lucienq@ufl.edu)

Prerequisites: PHC 6068 Introduction to Biostatistical Computing or the permission of the instructor.

Purpose and Outcome

Course Overview This course is a core course in the Health Data Science concentration of the MS program in Biostatistics, which gives a brief introduction to commonly used methods (e.g., trees and deep learning) in statistical/machine learning and their applications in health science and other disciplines. These statistical/machine learning methods have been increasingly used in the fields of data science and artificial intelligence (AI) to address various problems we encounter in public health and other disciplines. Through this course, students will also learn how to use statistical/machine learning packages in the R software to analyze data (e.g., data from public health science research).

Relation to Program Outcomes The knowledge and skills learned from this course can enhance students' ability in data analysis and professional advancement, especially using statistical/machine learning tools in data analysis. Successful completion of this course fulfills one of the core course requirements for the Health Data Science concentration and one of the elective requirements for the Biostatistical Methods and Practice Concentration in the MS program in Biostatistics.

Course Objectives and/or Goals

- Comprehend key concepts in statistical/machine learning, such as regularization and bias–variance tradeoff
- Compare and use different methods, such as basis expansion, kernel smoothing, trees, support vector machines, and neural networks
- Learn to use statistical/machine packages in R statistical software
- Use statistical/machine learning methods and the relevant software to solve research problems in the public health sciences and other disciplines

Instructional Methods

This course is an in-person course. Lectures will be given Mondays (9:35am-11:30am) and Wednesday (10:40am-11:30am). We will focus on major concepts and R codes in Mondays' lecture and use Wednesday to work on concept questions and R data analysis problems.

Description of Course Content

Topical Outline/Course Schedule

The course will cover major concepts and a variety of methods in the field of statistical/machine learning.

Week	Date(s)	Topic(s)	Reading
1	8/22-8/24	Introduction to Statistical Learning	Chapter 2 from ISLR
2	8/29-8/31	Linear Regression	Chapter 3 from ISLR
3	9/5-9/7	Linear Regression (Labor Day)	Labor Day
4	9/12-9/14	Classification	Chapter 4 from ISLR
5	9/19-9/21	Resampling Methods	Chapter 5 from ISLR
6	9/26-9/28	Linear Model Selection and Regularization	Chapter 6 from ISLR
7	10/3-10/5	Moving Beyond Linearity	Chapter 7 from ISLR
8	10/10-10/12	Kernel Smoothing	Chapter 6 from ESL
9	10/17-10/19	Tree-Based Methods	Chapter 8 from ISLR
10	10/24-10/26	Random forests and Ensemble learning	Chapter 8 from ISLR
11	10/31-11/2	Midterm Exam	Midterm Exam
12	11/7-11/9	Support Vector Machines	Chapter 9 from ISLR
13	11/14-11/16	Deep Learning	Chapter 10 from ISLR
14	11/21-11/23	Unsupervised Learning (Thanksgiving holiday)	Chapter 12 from ISLR
15	11/28-11/30	Multiple Testing	Chapter 13 from ISLR
16	12/5-12/7	Final Projects	Final Projects

Course Materials and Technology

The course is developed based on two textbooks, both of which are recommended and are freely available online

[1. James, G., Witten, D., Hastie, T., & Tibshirani, R. \(2021\). An Introduction to Statistical Learning with Applications in R \(2nd ed. 2021.\). Springer New York, NY. \(main textbook\)](#)

[2. Hastie, T., Tibshirani, R., Friedman, J. \(2009\). The Elements of Statistical Learning \(2nd ed. 2009\). Springer New York, NY.](#)

Statistical Software:

We will mainly use R in this course. R is free, and you can download R from <http://www.r-project.org/>.

Rstudio is a recommended interface for the R software. It is also free and can be downloaded from

<http://www.rstudio.org>. R packages related to this course can be found under <http://cran.rproject.org/web/packages/ISLR/index.html>

Diversity, Equity, & Inclusion:

Whenever possible, we will practice the statistical/machine methods with applications/datasets focusing on minority and vulnerable populations.

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>

Additional Academic Resources

[Career Connections Center](#): Reitz Union Suite 1300, 352-392-1601. Career assistance and counseling services.

[Library Support](#): Various ways to receive assistance with respect to using the libraries or finding resources.

[Teaching Center](#): Broward Hall, 352-392-2010 or to make an appointment 352- 392-6420.

General study skills and tutoring.

[Writing Studio](#): 2215 Turlington Hall, 352-846-1138. Help brainstorming, formatting, and writing papers.

[Student Complaints On-Campus](#): [Visit the Student Honor Code and Student Conduct Code webpage for more information.](#)

[On-Line Students Complaints](#): [View the Distance Learning Student Complaint Process.](#)

Academic Requirements and Grading

Grading The course will be evaluated based on Exercises (e.g., R data analysis problems) (30%), a midterm exam (30%), and a final project (30%). Professionalism (e.g., presence and being involved in group discussions) will account for 10% of the final grade.

Exercises Each week, students will work with instructor and group members on concept and data analysis problems related to materials learned in that week. Discussion on problems is encouraged, but plagiarism is prohibited. Each student should turn in his/her answers via Canvas. Full credit will be given for assignments turned in on the due date (by 11:59pm). Reduced credit (i.e., 20% per day late) will be given for assignments turned in after the due date. Students whom missed class needs to come up with a plan to make up the assignments. Otherwise, students will be given reduced credit.

Midterm exam There will be one midterm exam. The midterm exam will be a take-home exam with both statistical learning methods and R data analysis questions. Students will have one week to complete the exam and submit the neatly typed exams in Word or LaTeX online in Canvas or by email to the instructor. Students are not allowed to work together on the midterm exam. Questions about the midterm exam should be directed to the instructor as early as possible, at least 24 hours before the exam is due. Full credit will be given for assignments turned in via Canvas on the due date (by 11:59pm). Reduced credit (i.e., 20% per day late) will be given for assignments turned in after the due date.

Final Project For the final project, each student/group will work on a project from Kaggle (<https://www.kaggle.com>), preferable one of the competition projects, and present the project based on 1) the analysis of data from the project, or 2) comparing existing methods using the data from the project. The grade is given based on the quality of the project (e.g., an innovative way of using an existing method or comprehensively comparing several methods), and the presentation of the project (i.e., delivering an easy-to-follow and informative talk). Additional 10 bonus points will be given to students/groups who achieve a high rank in the competition (i.e., achieve the top 10% among all contributors). Reduced credit (i.e., 20% per day late) will be given for assignments turned in after the due date.

**University of Florida
College of Public Health & Health Professions Syllabus**

Professionalism Full points (i.e., 10 points) will be given to students who attend the class on time, and reduced points will be given to students who are late for classes (20% point reduction for each class) or miss classes (50% point reduction for each class). Cell phones should be silenced, and laptops should be turned off during class unless needed. Students are also encouraged to be actively engaged in group discussion, participate in solving assigned R data analysis problems, and ask questions during the class.

Requirement	Due date	Points or % of final grade (% must sum to 100%)
R data analysis problems	The end of the week	30%
Midterm Exam	11/07	30%
Final project	12/5-12/7	30%
Attendance	8/24-12/7	10%

The numerical final score will be converted to the letter grades according to the following scale and cutoffs:

Percentage Earned	Letter Grade
93-100	A
90-92	A-
87-89	B+
83-86	B
80-82	B-
77-79	C+
73-76	C
70-72	C-
67-69	D+
63-66	D
60-62	D-
Below 60	E

Please be aware that a C- is not an acceptable grade for graduate students. The GPA for graduate students must be 3.0 based on 5000 level courses and above to graduate. A grade of C counts toward a graduate degree only if based on credits in courses numbered 5000 or higher that have been earned with a B+ or higher.

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

More information on UF grading policy may be found at:

<http://gradcatalog.ufl.edu/content.php?catoid=10&navoid=2020#grades>

Exam Policy

Policy Related to Make up Exams or Other Work

Please note: Any requests for make-ups due to technical issues MUST be accompanied by the UF Computing help desk (<http://helpdesk.ufl.edu/>) correspondence. You MUST e-mail me within 24 hours of the technical difficulty if you wish to request a make-up.

Policy Related to Required Class Attendance

Excused absences must be consistent with university policies in the Graduate Catalog (<https://catalog.ufl.edu/graduate/regulations/#text>). Additional information can be found here: <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 show up for class prepared and on time. Cell phones are to be silenced during class unless there is an emergency, in which case please inform the instructor.

Communication Guidelines

The preferred methods of communication for the course are messages in e-learning 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.

Recording Within the Course:

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A “class lecture” is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To “publish” means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring

services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

Online Faculty Course Evaluation Process

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.

Policy Related to Guests Attending Class

Only registered students are permitted to attend class. However, we recognize that students who are caretakers may face occasional unexpected challenges creating attendance barriers. Therefore, by exception, a department chair or his or her designee (e.g., instructors) may grant a student permission to bring a guest(s) for a total of two class sessions per semester. This is two sessions total across all courses. No further extensions will be granted. Please note that guests are **not** permitted to attend either cadaver or wet labs. Students are responsible for course material regardless of attendance. For additional information, please review the Classroom Guests of Students policy in its entirety. Link to full policy: <http://facstaff.php.ufl.edu/services/resourceguide/getstarted.htm>

SUPPORT SERVICES

Accommodations for Students with Disabilities

If you require classroom accommodation because of a disability, it is strongly recommended you register with the Dean of Students Office <http://www.dso.ufl.edu> within the first week of class or as soon as you believe you might be eligible for accommodations. The Dean of Students Office will provide documentation of accommodations to you, which you must then give to me as the instructor of the course to receive accommodations. Please do this as soon as possible after you receive the letter. Students with disabilities should follow this procedure as early as possible in the semester. 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.
- **U 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>
- **University Police Department:** [Visit UF Police Department website](#) or call 352-392-1111 (or 9-1-1 for emergencies).
- **UF Health Shands Emergency Room / Trauma Center:** For immediate medical care call 352-733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville, FL 32608; [Visit the UF Health Emergency Room and Trauma Center website.](#)

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.

Inclusive Learning Environment

Public health and health professions are based on the belief in human dignity and on respect for the individual. As we share our personal beliefs inside or outside of the classroom, it is always with the understanding that we value and respect diversity of background, experience, and opinion, where every individual feels valued. We believe in, and promote, openness and tolerance of differences in ethnicity and culture, and we respect differing personal, spiritual, religious and political values. We further believe that celebrating such diversity enriches the quality of the educational experiences we provide our students and enhances our own personal and professional relationships. We embrace The University of Florida's Non-Discrimination Policy, which reads, "The University shall actively promote equal opportunity policies and practices conforming to laws against discrimination. The University is committed to non-discrimination with respect to race, creed, color, religion, age, disability, sex, sexual orientation, gender identity and expression, marital status, national origin, political opinions or affiliations, genetic information and veteran status as protected under the Vietnam Era Veterans' Readjustment Assistance Act." If you have questions or concerns about your rights and responsibilities for inclusive learning environment, please see your instructor or refer to the Office of Multicultural & Diversity Affairs website: www.multicultural.ufl.edu