STAT 340 (Spring 2020)
Introduction to Data Modeling II
University of Wisconsin—Madison

Course Description: Apply statistical methods to learn from data. Topics include one- and two-sample inference; an introduction to Bayesian inference and associated probability theory; linear and logistic regression models; the bootstrap; and cross-validation. Use an integrated statistical computing environment to explore and analyze data, develop models, make inferences, and communicate results in a reproducible manner through a project-oriented approach to learning.

Requisite:	MATH 217 or 221 or 275 and STAT 240	**Designations:**	Intermediate level, 4 credits Natural Science
Classes:	TuTh 11:00AM - 12:15PM	**Room:**	BIOCHEM 1120
Professor:	Karl Rohe	**Office:**	MSC 1239
Office Hours:	Th 12:15-2:15	**E-mail:**	karlrohe@stat.wisc.edu
www:	All course materials will be here www.stat.wisc.edu/~karlrohe/wdm340/main.html.		

TA: Nick Montalbano
E-mail: npmontalbano@wisc.edu
Office Hours 1: 11-12 on Mon in MSC 1210
Office Hours 2: 8-9 AM on Tuesdays in MSC 1217C

Section 311 meets in 1153 Mechanical Engineering Building.	T 3:30-4:20 PM.
Section 312 meets in 114 Van Hise Hall.	W 8:50-9:40 AM.
Section 313 meets in 114 Van Hise Hall.	W 12:05-12:55 PM.

Grader: David White
E-mail: dmwhite5@wisc.edu
Office Hours: 10-11am on Wed in MSC B315.

Course format: This class meets for two lecture periods and one 50-minute discussion section period each week over the semester and carries the expectation that students will work on course learning activities (reading, writing, problem sets, studying, etc) for about 3 hours out of the classroom for every class period.

Learning Outcomes: Students will: (1) use the bootstrap and permutation test statistical inference in one- and two-sample settings; (2) use Bayesian inference for some one- and two-sample settings; (3) use linear/logistic regression models; (4) use the bootstrap to generate confidence intervals; (5) use cross-validation to estimate out-of-sample performance; (6) interpret results in the appropriate scientific context; (7) master probability concepts associated with these models and methods; (8) communicate findings using effective writing; (9) use the R language and R Markdown for reproducible data analysis.

Text: We will use "Introduction to Statistical Learning with R", by James, Witten, Hastie, and Tibshirani, referred to as ISLR. You may find ISLR at goo.gl/Pi5zs8. Print versions are also available.

Software: R cran.r-project.org and R Studio www.rstudio.com are required software.

Homeworks: Assignments should be neat and well organized. Unless the question requires mathematical derivations, homework should be completed in R Markdown and delivered to the grader in the fashion discussed in class. There is the possibility of a small number of quizzes and they would count as homeworks.

Computing:	We will make extensive use of the software R. It is assumed that you are familiar with the basics of R from STAT 240. The penultimate section within each chapter of ISLR is devoted to the R code necessary to perform the analyses described in the text. Discussion section will often focus on these steps.		

Topics: The following table gives a rough sketch of the course.

Week	Content	Reading	Discussion Emphasis
1	Course Overview	course notes	does not meet
2	One and two sample inference	course notes	application problems, R
3	Prediction	ISLR ch 2	application problems, R
4	Probability and Random Variables	course notes	simulation and concept problems
5	Expectation and variance	course notes	simulation compared to derivation
6	Linear Models	ISLR ch 3.1, 3.2	simulation and application problems
7	Random Forest	ISLR ch 8.2.2	application problems, R
8	Catch-up and Midterm	ISLR ch 3.1	applications problems, R
9	Logistic regression	ISLR ch 4	application problems, R
10	Initiate projects!	course notes	thesis statements
11	Model checking	course notes	application problems, R
12	Cross-validation and bootstrap	ISLR ch 5	application problems, R
13	Project presentations		final review
14	Project presentations		final review

Project: There will be a group project.

Exams: There will be a midterm exam (Thursday March 12) and a final exam.

Grading: Homework is worth a total 300 points. The Midterm is worth 200 points. The final project is worth 400 points. The final exam is worth 100 points. Over 930 is an A. Over 900 is at least an AB. Over 850 at least a B. Over 800 at least a BC. Over 750 is at least a C. Over 700 is at least a D.

Honesty: You are permitted, in fact encouraged, to talk to other students, your teaching assistant, or me about homework. However, you may not present other people's work as your own. If you work with other students solving problems, make sure that you write up your own solution independently. It is not acceptable for one student to write a solution for another student to copy. – On exams, your work is to be entirely your own.

Public data: I have not fully explored these sites. Use at your own risk. Certainly, there are many others. If you find one you like, please let me know. In particular, this list is heavily focused on "social science" type data sets. It is lacking in other areas. Other popular data types come from sports or is based upon zip code aggregates.
https://www.data.gov,
https://opr.princeton.edu/archive/,
http://www.cpc.unc.edu/projects/addhealth,
https://data.cityofmadison.com,
https://mediacloud.org,
http://www.pewresearch.org/data/download-datasets/
http://www.ropercenter.uconn.edu/data_access.html

Academic integrity and data ethics:

By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin-Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of disciplinary action include, but is not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension, or expulsion. For detailed information, please see https://conduct.students.wisc.edu/academic-misconduct/.

The members of the faculty of the Department of Statistics at UW-Madison uphold the highest ethical standards of teaching, data, and research. They expect their students to uphold the same standards of ethical conduct. Standards of ethical conduct in data analysis and data privacy are detailed on the ASA website, and include:

- Use methodology and data that are relevant and appropriate; without favoritism or prejudice; and in a manner intended to produce valid, interpretable, and reproducible results.
- Be candid about any known or suspected limitations, defects, or biases in the data that may affect the integrity or reliability of the analysis. Obviously, never modify or falsify data.
- Protect the privacy and confidentiality of research subjects and data concerning them, whether obtained from the subjects directly, other persons, or existing records.

By registering for this course, you are implicitly agreeing to conduct yourself with the utmost integrity throughout the semester.

Diversity and inclusion:

Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.

The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background–people who as students, faculty, and staff serve Wisconsin and the world. https://diversity.wisc.edu/

Accommodations for students with disabilities:

The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform me of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. I will work either directly with the student or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA.

Complaints:

If you have a complaint about a TA or course instructor, you should feel free to discuss the matter directly with the TA or instructor. If the complaint is about the TA and you do not feel comfortable discussing it with him or her, you should discuss it with the course instructor. Complaints about mistakes in grading should be resolved with the instructor in the great majority of cases. If the complaint is about the instructor (other than ordinary grading questions) and you do not feel comfortable discussing it with him or her, contact the Director of Undergraduate Studies, Professor Cecile Ane, cecile.ane@wisc.edu.

If your complaint concerns sexual harassment, please see campus resources listed at https://compliance.wisc.edu/titleix/resources/. In particular, there are a number of options to speak to someone confidentially.

If you have concerns about climate or bias in this class, or if you wish to report an incident of bias or hate that has occurred in any statistics class, you may contact the Chair of the Statistics Department Climate & Diversity Committee, Professor Karl Rohe (karlrohe@stat.wisc.edu). If you would prefer someone who is not the instructor for this course, you may contact Rick Chappell (chappell@biostat.wisc.edu) who is also on the Statistics Department Climate & Diversity Committee. You may also use the University's bias incident reporting system, which you can reach at https://doso.students.wisc.edu/services/bias-reporting-process/.

University level rules, rights, and responsibilities for students:

See: https://guide.wisc.edu/undergraduate/#rulesrightsandresponsibilitiestext.