STAT 340 (Spring 2021)
Data Science: Introduction to Data Modeling II
University of Wisconsin—Madison

Course Description: Apply statistical methods to learn from data. Topics include inference with Monte Carlo 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 9:30AM-10:45AM	**Room:**	NA
Professor:	Karl Rohe	**Office:**	MSC 1239
Office Hours:	TuTh 10:45-11:15, W 1-2	**E-mail:**	karlrohe@stat.wisc.edu
www:	Canvas and `https://karlrohe.github.io/340-Spring21/`.		

Associate Instructor:	Bi Cheng Wu
E-mail:	bwu62@wisc.edu
Office Hours:	MTu 11-12; WTh 1-2
	`us.bbcollab.com/guest/8a420438709641ee9e059de1a63591ab`
TA:	Yanbo Shen
E-mail:	yshen84@wisc.edu
Office Hours:	9:45-11:45 AM on Mon
	`https://us.bbcollab.com/guest/677f0de311364fb499d4efdd625cf82c`
TA:	Jiaxin Hu
E-mail:	jhu267@wisc.edu
Office Hours:	4:00 - 5:00 PM on Mon
	`https://us.bbcollab.com/guest/f97a1dd867e2477e87cf8baf2844517a`
TA:	Bowen Zhang
E-mail:	bzhang327@wisc.edu
Office Hours:	11-13 on Mon
	`https://us.bbcollab.com/guest/9452c430819941c687d839096105c565`
TA:	Yongjoon Kim
E-mail:	ykim676@wisc.edu
Office Hours:	4 - 6 PM on Mon
	`https://us.bbcollab.com/guest/3028c846179c4917b8045c9cd28084c7`
TA:	Yongfeng Wu
E-mail:	wu364@wisc.edu
Office Hours:	3-4pm on Wed
	`https://us.bbcollab.com/guest/fab83967dc8d420ea4c761e8f4a5ee41`

Section 311:	8:50-9:40 AM, TA Yongjoon Kim	
	https://us.bbcollab.com/guest/5b21047a78af4dec8b8741ec62d32826	
Section 312:	1:20-2:10 PM, TA Yongjoon Kim	
	https://us.bbcollab.com/guest/8eba200aed8c43c4816c8fa91063322c	
Section 313:	5:40-6:30 PM, TA Jiaxin Hu	
	https://us.bbcollab.com/guest/e8e809dbfea84dd683cfd73698284624	
Section 314:	8:50-9:40 AM, TA Yanbo Shen	
	https://us.bbcollab.com/guest/627e0f1100274cad8e79ab39221b15f6	
Section 315:	1:20-2:10 PM, TA Yanbo Shen	
	https://us.bbcollab.com/guest/1596b6b799654d4abf6d667b596741a8	
Section 316:	5:40-6:30 PMTA Yongfeng Wu	
	https://us.bbcollab.com/guest/3db3f07db3234e2bb0ed31296f17345f	
Section 317:	7:45-8:35 AM, TA Yongfeng Wu	
	https://us.bbcollab.com/guest/0195a031ef13466ab4edcac32fd513f1	
Section 318:	9:55-10:45 AM, TA Bowen Zhang	
	https://us.bbcollab.com/guest/b9df07cb226f4a7f97bf621703f15721	
Section 319:	2:25-3:15 PM, TA Bowen Zhang	
	https://us.bbcollab.com/guest/1e81da9e508747bbb0feb717499c6a52	

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 Monte Carlo to preform statistical inference; (2) use linear/logistic regression models; (3) use the bootstrap to generate confidence intervals; (4) use cross-validation to estimate out-of-sample performance; (5) interpret results in the appropriate scientific context; (6) communicate findings using effective writing and data summaries; (7) use the R language and R Markdown for reproducible data analysis.

Text: The first 5 weeks of the course will use lecture notes. The next five 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. Homework should be completed in R Markdown (unless otherwise stated) and delivered to the grader in the fashion discussed in class. There is the possibility 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	Introduction to discussion
2	Random variables	course notes	application problems, R
3	Monte Carlo	course notes	application problems, R
4	Hypothesis testing	course notes	application problems, R
5	Statistical estimation	course notes	application problems, R
6	Linear Models	ISLR ch 3.1, 3.2	application problems, R
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	Visualizations	course notes	application problems, R
14	Project presentations		final review

Project:	There will be a group project.
Exams:	There will be a midterm exam (Thursday March 18) 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 and quizzes, your work is to be entirely your own. Due to the stresses of the pandemic, we will be more lenient in all regards, except on issues of academic integrity and honesty.

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 Laura Juszczyk (`ljuszczyk@stat.wisc.edu`) who is another member of 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`.