



**Syllabus**  
**STAT 340: Data Modeling II**  
**Fall 2022, 4 Credits**

**Description**

Students will learn how to explore and analyze data using R, as well as how to present their findings and analyses clearly. Topics include basic probability models; the central limit theorem; Monte Carlo simulation; one- and two-sample hypothesis testing; Bayesian inference; linear and logistic regression; ANOVA; the bootstrap; random forests and cross-validation. Students will learn how to present their findings in a clear and reproducible manner in a project setting by applying their skills to analyze real-world data sets.

**Prerequisites**

STAT240: Introduction to Data Modeling I and one or more of MATH 217, 221, or 275. Specifically, students should have a broad familiarity with the R programming language and should be comfortable with basic concepts from calculus.

**Instructor**

Keith Levin (kdlevin@wisc.edu); office hours: Wednesdays 12pm-2pm in Medical Sciences Center 6170 or by appointment.

**Teaching Assistants**

- Nursultan Azhimuratov (azhimuratov@wisc.edu); office hours: Mondays 1pm-3pm in Medical Sciences Center 1274
- Alex Hayes (alex.hayes@wisc.edu); office hours: Wednesdays 10am-12pm in Medical Sciences Center 1475
- Shane Huang (shuang457@wisc.edu); office hours: Tuesdays and Thursdays 1pm-2pm in Medical Sciences Center 1274
- Joseph Salzer (jsalzer@wisc.edu); office hours: Tuesdays 10am-11am in Medical Sciences Center 1217C and Tuesdays 5pm-6pm in Medical Sciences Center 1274

**Meetings**

*Lecture:*

- **Section 001:** Tuesday, Thursday 11:00AM-12:15PM in Bardeen 140
- **Section 002:** Tuesday, Thursday 2:30PM-3:45PM in Van Vleck B130

*Discussion:*

Refer to your schedule. Discussion sections are on Wednesday afternoons and Thursday mornings.

**Textbook, Readings & Online Resources**

There is no physical textbook required for this course. I have endeavored to make the course lecture notes as self-contained as possible. Still, to ensure you get a thorough exposure to the material, we will have weekly readings from a variety of sources. We will make frequent reference to *Introduction to Probability and Statistics Using R* by G. Jay Kerns (available online at <http://ipsur.r-forge.r-project.org/book/download/IPSUR.pdf>); *R for Data Science* by Wickham and Golemund (available online at <https://r4ds.had.co.nz/>); and, later in the course, *An Introduction to Statistical Learning with*

*Applications in R* (ISLR) by James, Witten, Hastie and Tibshirani (available online at <https://www.statlearning.com/> and in print from Springer).

Where possible, I will try to give readings from more than one of these resources. You should not feel obligated to read all of these overlapping readings. Rather, the purpose of providing multiple readings is so that you have the option to read from the resource that you find the most useful for your learning style. Other required readings will be made available as we cover relevant material, and supplemental readings will be suggested for those who are interested in learning more.

All class resources will be made available on the course web page, <https://pages.stat.wisc.edu/~kdlevin/teaching/Fall2022/STAT340/> and on canvas. Please contact the instructor if any resources are missing from the webpage. The instructor will make an effort to post lecture notes and demo code a few days ahead of lecture.

### Course Topics

- **Random variables and models.** Basic probability models, conditional probability, Monte Carlo simulation, the strong law of large numbers, central limit theorem.
- **Estimation.** Confidence intervals, point estimates, the bootstrap.
- **Testing.** One- and two-sample hypothesis testing, test statistics, permutation tests, ANOVA.
- **Prediction.** Linear and logistic regression, random forests, cross-validation.
- **Exploratory Data Analysis.** Clustering, unsupervised learning, visualization.

### Learning Outcomes

By the end of this course, you will be able to:

- Use the R programming language to gather, clean and analyze data.
- Understand and apply basic concepts in probability; combine basic probability models to build more complicated ones; and critique models and their assumptions.
- Formulate statistical hypotheses for different kinds of research questions and test those hypotheses using both classical and Monte Carlo methods.
- Understand and apply principles of statistical estimation and prediction, including fitting models and assessing model quality.
- Perform basic exploratory data analysis and present findings visually using `ggplot2`.
- Apply statistical tools to answer research questions using real-world data and present these findings clearly in both spoken and written form to non-experts.

### Grading, Exams, Homeworks & Late Days

Grades will be based on cumulative performance on three exams and a collection of weekly homework assignments. Homeworks will review material from lecture and will be primarily programming-based. Exams will include both programming and short answer questions designed to assess how well students can explain and apply the concepts and methods discussed in lecture. All three exams will be take-home. These will count toward your final grade as follows:

Homeworks:	10%
Exam 1:	25%
Exam 2:	25%
Exam 3:	40%

Note that the exact number of homework assignments is subject to change depending on factors such as lecture cancellations and the speed with which we cover material. The instructor reserves the right to curve scores in the event of skewed class performance. Students may contest their grade on an assignment up to two (2) weeks from the day that an assignment's grades are released, after which grades may not be changed. Homework due dates are strict, and you may turn in work late only with the use of "late days", of which you have five (5) to use over the course of the semester. For each late day you spend, you may extend the

deadline of a homework by up to 24 hours. You may spend multiple late days per homework. Once you have turned in your homework you may not spend more late days to turn in your homework again after the deadline (you may, of course, turn in multiple versions of your homework assignment through Canvas prior to the deadline). Late days will be deducted automatically, and there is no need to notify the instructor that you wish to spend a late day. Homeworks turned in late with no remaining late days to spend will receive a zero. Late days may not be used to change the dates of exams. The purpose of this late day policy is to give you a way to deal with unexpected circumstances (e.g., illness, family emergencies, job interviews) without having to come to the instructor. Of course, if dire circumstances arise (e.g., long-term illness that causes you to miss multiple weeks of lecture), please speak with the instructor as promptly as possible. **Note:** owing to the university grading schedule, you may not use late days to extend any deadline beyond Wednesday, December 21st.

Grades will be assigned based on the scheme outlined below. The instructor reserves the right to relax this grading scheme in the event of skewed class performance, but pledges not to curve grades downward. That is, if you have an AB under this grading scheme, your curved grade will not be worse than an AB.

$\geq 93\%$	A
88% to 93%	AB
83% to 88%	B
78% to 83%	BC
70% to 78%	C
60% to 70%	D
$< 60\%$	F

### Key Dates

First lecture: Thursday, September 8, 2022

Exam 1: week of October 19, 2022

Exam 2: week of November 16, 2022

Last lecture: Wednesday, December 14, 2022

Final Exam (both sections): Thursday, December 22, 2022 at 10:05AM-12:05PM

### Communication and email policies

Questions regarding course content (e.g., clarifications regarding homework assignments, material from lecture, or a technical issue with R) should be posted to the discussion board on Canvas. This ensures that your classmates benefit from the answer to your question.

Questions not regarding course content (e.g., grading disputes or logistical matters such as missed lectures), should be raised via email. Please include the phrase “STAT340” in the subject line of your email and copy your TA on any communication with the instructor. Owing to the size of the course, the instructor and TAs cannot guarantee immediate response to emails. If a response is not received within 48 hours, please send a follow-up message.

Again owing to the size of the course, the instructor and TAs cannot feasibly provide technical support via email. If you have an issue with R (e.g., packages not installing properly or error messages that you cannot decipher), please raise your question at office hours.

### Ethics and class policies

Academic misconduct includes such actions as copying code from the web or from your fellow students, providing code to your fellow students, looking up solutions online, turning in assignments from other classes or previous iterations of this course, and hiring others to complete your work for you. You are welcome to discuss homeworks with your classmates, but the work that you turn in must be yours and yours alone, and you must disclose in your homework the names of those with whom you collaborated. Exams are to be completed in isolation and are to be discussed only with the course instructor and teaching assistants.

From the Office of Student Conduct and Community Standards:

[A]cademic 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.

See <https://conduct.students.wisc.edu/academic-misconduct/> for more information.

Violations of these or other university ethical standards surrounding academic honesty will be met with serious consequences and disciplinary action. At a minimum, cheating on an assignment will result in a 0 for that assignment and the incident will be reported to the appropriate office. At the instructor's discretion, depending on the circumstances, an additional full letter grade may be deducted from the student's final grade in the course.

**COVID-19 Preparation and Policies** The COVID-19 pandemic is, of course, still in progress and is inherently unpredictable and the university is updating its policies accordingly. Please be sure to follow all university guidelines surrounding masks and vaccines. Should the university decide to return to remote instruction, we are prepared to switch course modality accordingly.

### **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 faculty 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. Faculty 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.