Stat 343: Applied Linear Statistical Methods — Autumn 2021 Syllabus

Course description This course introduces the theory, methods, and applications of fitting and interpreting multiple regression models. Topics include the examination of residuals, the transformation of data, strategies and criteria for the selection of a regression equation, nonlinear models, biases due to excluded variables and measurement error, and the use and interpretation of computer package regression programs. The theoretical basis of the methods, the relation to linear algebra, and the effects of violations of assumptions are studied. Techniques discussed are illustrated by examples involving both physical and social sciences data.

Course info

- Instructor: Rina Barber, `rina@uchicago.edu`
- TAs: Andrew Goldstein, `andrewgoldstein@uchicago.edu`
  Joonsuk Kang, `joonsukkang@uchicago.edu`
  Subhodh Kotekal, `skotekal@uchicago.edu`
- The main course page is on Canvas and you can find all slides, assignments, etc there. Homework will be handed in and graded on Gradescope. We will also use Ed Discussion for announcements and Q&A.
- This course will be held in person, but last year's videos will be available through Canvas for any student that needs to miss class due to quarantining or any other reason.
- All office hours will be on Zoom (links available through Canvas). The weekly OH schedule is:
  - Monday 10:30am–12:00pm: instructor OH (starting week 2)
  - Tuesday 7:00–8:30pm: TA OH
  - Wednesday 3:00–4:30pm: TA OH
- COVID-19 policies: per UChicago policy, students and instructors must wear masks in classrooms. Always stay home if you feel sick or test positive for COVID-19. Last year's recorded lectures will be posted on Canvas to accommodate students who are not able to attend class.

Contacting us

- We will aim to reply to all questions within 24 hours on weekdays (response time will be slower on weekends).
- For any questions about the material or for general questions about the HW, please post a public question on Ed (you can choose to post anonymously).
- For specific questions about your work on the HW (i.e., questions that cannot be posted publicly because it would reveal too much of the solution), please ask us via a private post on Ed.
- For any regrade requests on HW or exams, please use the regrade request feature on Gradescope.
- For other questions such as enrollment, prerequisites, accommodations, makeup times for exams, etc, please contact the instructor by email.

Textbook & resources The textbook for this course is:

- *Linear Models with R*, Faraway, 2nd edition.

Important note—if you own the 1st edition of this book, you can use this as reference for the material, but the exercises are often different. We will not give credit for HW problems handed in using the wrong edition.

Additional suggested resources:

- *Applied Linear Regression*, Weisberg
- *Elements of Statistical Learning*, Hastie, Tibshirani, & Friedman
  Available for free online at `https://web.stanford.edu/~hastie/ElemStatLearn/`
- *Mathematics for Machine Learning*, Thomas
  Available for free online at `https://gwthomas.github.io/docs/math4ml.pdf`

Computing The problem sets will often involve working with simulations or with data. All computing for this course should be done with R (preferably using RStudio). The TAs can provide support as needed for computing questions.

Handing in assignments

- Assignments are due at the start of class on Thursdays.
- At the end of the quarter, the lowest HW grade (or one missing grade) will be dropped. We cannot excuse any missed HWs beyond the one that is dropped.
- To give additional flexibility due to COVID and remote learning, late HWs will be accepted with a penalty of 4% per hour (late time is rounded up, i.e., one minute late counts as one hour late).
- Assignments are submitted and graded via Gradescope (which can be accessed from the Canvas course page).
  - For each problem, Gradescope will prompt you to tag the pages containing your answer to that problem. Be sure to tag all the pages that contain any part of your answer—for example if for problem 1, your written explanation is on page 1 + half of page 2, and your code is in an appendix and appears on page 6, you will need to tag all three of these pages for problem 1.
  - It's fine to have multiple problems on the same page (be sure to tag that page for all problems it contains).
  - If you have separate PDFs for the theory problems vs the code problems, you will need to merge them into a single file to submit (it is fine if this means the problems are then out of order in the final file).
- For the code components of homework or exams, we recommend using R Markdown (via R Studio) to produce a single file that weaves together your code, plots/output, and written explanations/comments.
- If you are having trouble uploading to the website and run out of time, please email your work to the instructor or TA before the time HW is due as proof of completion. The time of your email will count as the time of your HW submission. We do not accept the time stamp of the file on your computer as proof of completion.

Exams

- The midterm exam should be taken during class time on Thursday 10/28. If you need a different exam time (e.g., due to a time zone difference), please arrange this early in the quarter with the instructor.
- The final consists of a written exam and a data analysis assignment. The written exam will be given a specific time slot assigned by the registrar during Finals week, and should be taken at that time. If you need a different exam time (e.g., due to a time zone difference), please arrange this early in the quarter with the instructor. The data analysis will be assigned over several days (but can be completed in a few hours). The dates&times will be announced later in the quarter.
- We will decide whether exams are remote or in person, later on in the quarter, depending on current conditions.

Grading The final grade will be determined by homework plus midterm and final exams, in these proportions:

- Problem sets: 30% (with lowest HW grade or one missing HW will be dropped)
- Midterm exam: 30%
- Final: 40% (30% exam, 10% data analysis)

Collaboration guidelines & plagiarism policy For problem sets, students are free to discuss the problems and collaborate on strategies for solving the problems, but all writing, code, etc, should be done completely on your own. For example, working out a solution as a group, then transferring it to your own page, is not acceptable. No collaboration or discussion of any kind is allowed on the exams.

Any copied material (from websites, published materials, or another students' work) that is handed in without attribution will be considered to be plagiarism and will be reported to the appropriate university department. Feel free to reach out to the instructor or TAs if you have any questions about what is appropriate for collaboration or online resource use.

Please consult the student manual on university policies and regulations that make it clear that the University will not tolerate cheating and plagiarism: `https://studentmanual.uchicago.edu`

Recording and Deletion Policies The Recording and Deletion Policies for the current academic year can be found in the Student Manual under Petitions, Audio & Video Recording on Campus.

- Do not record, share, or disseminate any course sessions, videos, transcripts, audio, or chats.
- Do not share links for the course to those not currently enrolled.
- Any Zoom cloud recordings will be automatically deleted 90 days after the completion of the recording.

By attending course sessions, students acknowledge that:

- They will not: (i) record, share, or disseminate University of Chicago course sessions, videos, transcripts, audio, or chats; (ii) retain such materials after the end of the course; or (iii) use such materials for any purpose other than in connection with participation in the course.
- They will not share links to University of Chicago course sessions with any persons not authorized to be in the course session. Sharing course materials with persons authorized to be in the relevant course is permitted. Syllabi, handouts, slides, and other documents may be shared at the discretion of the instructor.
- Course recordings, content, and materials may be covered by copyrights held by the University, the instructor, or third parties. Any unauthorized use of such recordings or course materials may violate such copyrights.
- Any violation of this policy will be referred to the Area Dean of Students.

Special Accommodations The University of Chicago is committed to ensuring equitable access to our academic programs and services. Students with disabilities who have been approved for the use of academic accommodations by Student Disability Services (SDS) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by SDS for using accommodations. Timely notifications are required in order to ensure that your accommodations can be implemented. Please contact the instructor to discuss your access needs in this class after you have completed the SDS procedures for requesting accommodations.
Phone: (773) 702-6000, Email: `disabilities@uchicago.edu`

Schedule (may change as needed)

Week	Dates	Topics	Due (Thursday)
1	Tue Sep 28	Lecture 1a: intro to regression Lecture 1b: simple linear regression & inference for OLS	
	Thu Sep 30 1	Lecture 2a: intro to multiple linear regression Lecture 2b: intro to multiple linear regression—data example	—
2	Tue Oct 5	Lecture 3a: multivariate normal distribution Lecture 3b: inference for multiple linear regression (part 1)	
	Thu Oct 7	Lecture 4a: inference for multiple linear regression (part 2) Lecture 4b: diagnostics & outliers—example	PSet 1 due
3	Tue Oct 12	Lecture 5a: diagnostics & outliers Lecture 5b: clusters—example	
	Thu Oct 14	Lecture 6a: bootstrap Lecture 6b: bootstrap—example	PSet 2 due
4	Tue Oct 19	Lecture 7a: model selection Lecture 7b: model selection—simulations	
	Thu Oct 21	Lecture 8a: robust regression Lecture 8b: robust regression—simulation	PSet 3 due
5	Tue Oct 26	Lecture 9a: weighted least squares Lecture 9b: weighted least squares—example	
	Thu Oct 28	Midterm (Taken during class time, except by prior arrangement)	—
6	Tue Nov 2	Lecture 10a: iteratively reweighted least squares Lecture 10b: transformations	
	Thu Nov 4	Lecture 11a: ridge regression Lecture 11b: ridge regression—examples	PSet 4 due
7	Tue Nov 9	Lecture 12a: Lasso & sparse regression Lecture 12b: Lasso & sparse regression—examples	
	Thu Nov 11	Lecture 13a: missing data Lecture 13b: missing data—example	PSet 5 due
8	Tue Nov 16	Lecture 14a: categorical covariates & ANOVA—part 1 Lecture 14b: categorical covariates & ANOVA—part 2	
	Thu Nov 18	Lecture 15a: factor models & pairwise comparisons Lecture 15b: factor models & pairwise comparisons—simulation	PSet 6 due
	Tue/Thu Nov 23/25	Thanksgiving break	
9	Tue Nov 30	Lecture 16a: factorial design Lecture 16b: experiment design & blocking—part 1	
	Thu Dec 2	Lecture 17a: experiment design & blocking—part 2 Lecture 17b: review	PSet 7 due
	Tue–Fri Dec 7–10	Final exam / data analysis exam (Dates TBD)	