Stat 343: Applied Linear Statistical Methods — Autumn 2022 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: Joonsuk Kang, `joonsukkang@uchicago.edu`
  Wei Kuang, `weikuang@uchicago.edu`
  Yonghoon Lee, `yhoony31@uchicago.edu`
- The main course page is on Canvas and you can find all slides, assignments, etc there. Homework is due on Wednesdays, and 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 the videos from 2020 will be available through Canvas for any student that needs to miss class due to quarantining or any other reason. **Please do not share these recordings with anyone outside the class—they are provided for your personal use only.**
- The weekly OH schedule is:
  - Monday 7–8pm: TA OH (zoom) (starting week 2)
  - Tuesday 12:30–1:30pm: TA OH (Jones 226) (starting week 2)
  - Tuesday 5-6pm: TA OH (Jones 226) (starting week 2)
  - Friday 10-11:30am: instructor OH (zoom)
- Exam dates: **Thursday Oct 27 (evening)** for the midterm, **Monday Dec 5 (evening)** for the final. Time & location will be announced soon. Please contact the instructor ASAP if you have any time conflicts with the scheduled exams. There will also be a final data analysis due **Friday Dec 9 at noon**.

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 or clarifications on the grading for HW or exams, please use the regrade request feature on Gradescope.
- For administrative questions such as enrollment, prerequisites, accommodations, makeup times for exams, etc (i.e., anything that is not related to the course material)—please contact the instructor by email.

Handing in assignments

- Assignments are due at the start of class on Wednesdays.
- 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.
- 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). We cannot make exceptions to the late penalty.
- Assignments are submitted and graded via Gradescope (which can be accessed from the Canvas course page). Please tag the pages for each problem.
- 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 on Gradescope (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 will be **Thursday Oct 27, in the evening (5:30pm or later)** (exact time & location announced later).
- The final consists of a written exam and a data analysis assignment. The written portion of the final exam will be **Monday Dec 5, in the evening (5:30pm or later)** (exact time & location announced later). The data analysis will be due on **Friday Dec 9, 12pm (noon)**.
- For the midterm and the written final exam, no collaboration or discussion is permitted. For the data analysis, no collaboration or discussion is permitted, but students are permitted to work either alone or in a pair. You may use your textbook, notes, etc for all exams.
- Please contact the instructor ASAP if you have any time conflicts with the scheduled exams.

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`

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.
SDS contact info — Phone: (773) 702-6000, Email: `disabilities@uchicago.edu`

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.

Schedule (may change as needed)

Week	Dates	Topics	HW (due Weds)
1	Wed Sep 28	Lecture 1a: intro to regression Lecture 1b: simple linear regression & inference for OLS	—
2	Mon Oct 3 1	Lecture 2a: intro to multiple linear regression Lecture 2b: intro to multiple linear regression—data example	
	Wed Oct 5	Lecture 3a: multivariate normal distribution Lecture 3b: inference for multiple linear regression (part 1)	PSet 1 due
3	Mon Oct 10	Lecture 4a: inference for multiple linear regression (part 2) Lecture 4b: diagnostics & outliers—example	
	Wed Oct 12	Lecture 5a: diagnostics & outliers Lecture 5b: clusters—example	PSet 2 due
4	Mon Oct 17	Lecture 6a: bootstrap Lecture 6b: bootstrap—example	
	Wed Oct 19	Lecture 7a: model selection Lecture 7b: model selection—simulations	PSet 3 due
5	Mon Oct 24	Lecture 8a: robust regression Lecture 8b: robust regression—simulation	
	Wed Oct 26	Lecture 9a: weighted least squares Lecture 9b: weighted least squares—example	—
	Thu Oct 27	**Midterm (time TBD — will be 5:30pm or later)**	
6	Mon Oct 31	Lecture 10a: iteratively reweighted least squares Lecture 10b: transformations	
	Wed Nov 2	Lecture 11a: ridge regression Lecture 11b: ridge regression—examples	PSet 4 due
7	Mon Nov 7	Lecture 12a: Lasso & sparse regression Lecture 12b: Lasso & sparse regression—examples	
	Wed Nov 9	Lecture 13a: missing data Lecture 13b: missing data—example	PSet 5 due
8	Mon Nov 14	Lecture 14a: categorical covariates & ANOVA—part 1 Lecture 14b: categorical covariates & ANOVA—part 2	
	Wed Nov 16	Lecture 15a: factor models & pairwise comparisons Lecture 15b: factor models & pairwise comparisons—simulation	PSet 6 due
	Nov 21–25	Thanksgiving break	
9	Mon Nov 28	Lecture 16a: factorial design Lecture 16b: experiment design & blocking—part 1	
	Wed Nov 30	Lecture 17a: experiment design & blocking—part 2 Lecture 17b: review	PSet 7 due
	Mon Dec 5	**Final—written exam (time TBD — will be 5:30pm or later)**	
	Fri Dec 9	**Final—data analysis (due at 12:00pm)**	