

Linear Regression Models

Statistics GR5205/GU4205 — Fall 2020

Section 003: MW 1:10pm–2:25pm; HyFlex

Before September 18th, **ALL CLASS WILL BE OFFERED ONLINE ONLY.**

COVID-19 CORONAVIRUS. The policies set forth in this syllabus are subject to change as we try to determine how best to keep you safe from the COVID-19 coronavirus while we provide the education we promised you.

<i>Instructor:</i>	Xiaofei Shi (xs2427@columbia.edu)	<i>Office hours:</i>	Thursday 5:30-7:00 pm
<i>TA:</i>	Arnab Auddy (aa4238@columbia.edu)	<i>Office hours:</i>	Monday 4:00-5:30 pm
	Ari Blau (bsb2144@columbia.edu)	<i>Office hours:</i>	Wednesday 3:00-4:30 pm

Course Description: This is a first course in regression analysis for graduates and undergraduates in Statistics and related majors. Topics covered include: scatterplots and correlation, bivariate and conditional distributions, inference in the simple linear regression model, multivariate regression, ordinary and weighted least squares, categorical predictor variables, transformations, residual analysis, model selection, and logistic regression.

Prerequisite: Statistics, (Calculus based) Probability, Linear Algebra.

Textbook: *Applied Linear Regression Models, fourth edition*; by Kutner, Nachtsheim and Neter. Identical content can be found in the first 14 chapters of *Applied Linear Statistical Models, fifth edition*; by Kutner, Nachtsheim, Neter and Li.

Programming: Coursework will require the use of statistical software. You are free to use any language to finish the homework, including python, Matlab and R.

Exams: There will be two examinations. You must take the final exam at the scheduled time, and there will be no make-up exams.

- The midterm exam will be given on Wednesday, October 28.
- The final exam TBD.

Further details (sections covered, online/on campus, seating assignments) will be announced in class at least two weeks before the exam.

Tentative course outline:

Date	Topics	
Mon 09/07	Labor Day	
Wed 09/09	Introduction to Linear Regression	Online
Mon 09/14	The Simple Linear Regression Model	Online
Wed 09/16	R Tutorial (by Arnab Auddy)	Online
Mon 09/21	MLE and MAP	501 SCH
Wed 09/23	Gaussian SLR: Inference and Prediction	501 SCH
Mon 09/28	Gaussian SLR: Hypothesis Testing	Online
Wed 09/30	Multivariate Linear Regression	501 SCH
Mon 10/05	Weighted Least Squares	501 SCH
Wed 10/07	Inference and Testing for Multivariate Linear Models	Online
Mon 10/12	Delta Method	501 SCH
Wed 10/14	TA Session (by Ari Blau)	Online
Mon 10/19	Beyond Gaussian Models	Online
Wed 10/21	Mixture Models	501 SCH
Mon 10/26	Midterm Review	501 SCH
Wed 10/28	Midterm Exam	Online
Mon 11/02	Election Day; No Class	
Wed 11/04	Categorical Variables and Transformations	
Mon 11/09	Logistic Regression	
Wed 11/11	Kernel Methods	
Mon 11/16	Introduction to Time Series	
Wed 11/18	TBD	
Mon 11/23	TBD	
Wed 11/25	Thanksgiving; No Class	
Mon 11/30	TBD	
Wed 12/02	TBD	
Mon 12/07	TBD	
Wed 12/09	Review Session	
Mon 12/14	Last Day of Classes	
Wed 12/16	Reading and Exam Days	
Mon 12/21	Reading and Exam Days	
Wed 12/23	End of Semester	

Evaluation: Your course grade will be computed as the following weighted average:

- 50% weight for the final exam,
- 25% weight for the midterm exam,
- 20% weight for the homework capped by the exam average,
- 5% weight for class participation.

In particular, at the end of the course you will have a Final Exam Score (100 points possible), a Midterm Exam Score (100 points possible), a Homework Score (100 points possible), and a Writing Assignment Score (10 points possible). I will first compute your Exam Average by the following formula:

$$\text{Exam Average} = \frac{2}{3}\text{Final Exam Score} + \frac{1}{3}\text{Midterm Exam Score}.$$

Then the final grade of this course will be computed by the formula:

$$\begin{aligned} &\text{Course Final Grade} \\ &= 75\% \text{ Exam Average} + 20\% \min \{ \text{Homework Average}, \text{Exam Average} \} + 5\% \text{ Participation} \end{aligned}$$

I will then convert the Course Final Grade to a letter grade. The possible letter grades that will be assigned at this point are A+, A, A-, B+, B, B-, C+, C, C-, D and R.

Homework: Homework will be posted on the Courseworks approximately two weeks in advance of the due date. There will be 6 assignments,

1. Post on Wednesday, September 9, due on Monday, September 21
2. Post on Wednesday, September 23, due on Monday, October 5
3. Post on Wednesday, October 7, due on Monday, October 19
4. Post on Wednesday, October 21, due on Monday, November 9
5. Post on Wednesday, November 11, due on Monday, November 23
6. Post on Wednesday, November 25, due on Monday, December 7

Each homework will have approximately four problems, and two of these will be selected for grading. Solutions to all problems will be posted on Courseworks after the homework are submitted. Each homework is due at midnight on due date. Please submit your homework on Courseworks. You are encouraged to discuss homework problems with your classmates, but **all work submitted must be your own, and you have to include all collaborators**. If multiple students turn in identical solutions, all of them will receive a zero.

You have a busy schedule, with assignments in other courses and, we hope, internship interviews. To give you some latitude with homework submissions, the following policy will be used:

- **Your lowest score will be dropped.** You are advised not to use this as permission not to submit a homework. If you choose not to submit one homework and then cannot submit a second because of illness, a job interview, or a broken scanner, you will have gambled and lost.
- **The homework may be submitted late.** Each homework is graded out of 20 possible points, and there will be a penalty of 1 point for each two hours of lateness. Thus, the last possible submission time is 40 hours after the Monday midnight deadline, which is 4:00p.m. Wednesday. Solutions will be posted Wednesday evening.

Scribing and Monitoring: Students can sign up for scribing and monitoring for lectures, with no more than 4 students signed up for a particular lecture. Scribed notes are due one week after the date of the corresponding lecture.

Policy on Collaboration: You are encouraged to work together on the homework. Discussing the homework problems with one another can be a valuable learning experience. However, it is a violation of the rules on academic integrity to copy another student's solution and submit it as your own. **You should write up your solutions separately, not referring to a common document.** Furthermore, you should not submit any work that you do not fully understand. You should be able to start with a clean sheet of paper and without notes or assistance write out the solution to any homework solution you submit. If you will do that with every homework you submit, the similarity between your solutions and those of other students will not arouse suspicion. More importantly, you will be well prepared for the exams. You are not permitted to use homework solutions for this course from previous years or solutions you find from other sources, including the internet.

You do not need and may not use anything during the exams except writing instruments and a non-programmable calculator. In particular, cell phones must be put away during exams. During the exam, you may not give nor receive assistance from other students.

Violation of these rules will be treated seriously according to school policy. The most common penalty for violations of the homework policy is that you will receive a grade of 0 on that homework and an additional one letter reduction in your course grade. The standard penalty for violations during an exam is failure in the course. Second violations or first violations that are accompanied by lies denying what the evidence shows generally result in suspension or dismissal from the program. Persons providing assistance in violation of these rules are subject to the same penalties as persons receiving assistance.

People caught cheating generally tell a story about being unable to keep up with the material. First of all, remember that you can submit homework late. Secondly, if there is something going on in your life that makes it hard for you to keep up, please let me know. That brings me to the following point.

Taking Care of Yourself. It is easy for me to say and hard for all of us, including me, to do, but taking care of your physical and mental health is essential, especially during the COVID-19 pandemic. Life is a marathon, and you need to pace yourself.

Do your best to maintain a healthy lifestyle by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

If you or anyone you know experiences extreme academic stress, difficult life events, or feelings of anxiety or depression, I strongly encourage you to seek support. Counseling and Psychological Services is here to help 24/7, and everything will be confidential: call 212-854-2878 or visit

<https://health.columbia.edu/content/counseling-and-psychological-services>

In addition, consider reaching out to a friend, faculty or family member you trust for help getting connected to the support. Keep in mind that for serious psychological issues, the first counselor you meet with may not be the right one for you, but this does not mean you should give up on counseling. Keep looking for someone who can help you.

If you or someone you know is feeling suicidal or in danger of self-harm, call immediately, day or night:

Counseling and Psychological Services: 212-854-2878

If the situation is life threatening, call the police:

On campus: Columbia Police: 212-854-2797

Off campus: 911

If you have questions about this advice, your coursework, or anything else about which I might be helpful, please let me know.

Rubric and policies are designed with the experience of Prof. Steven Shreve, the instructor of Stochastic Calculus II, MSCF program at CMU. Materials are based off lectures from Prof. Wayne T. Lee, Prof. Ronald Neath and Prof. Gabriel Young.