Statistics GR5205-005/D05
Linear Regression Models
Fall 2019

Instructor
Gabriel Young
Email: gjy2107@columbia.edu
Office: Room 614, 6th floor of Watson Hall, 612 West 115th Street

Teaching Assistants

- **In-class:** Reed Palmer
  Email: wrp2110@columbia.edu
- **Online:** Casey Bradshaw
  Email: cb3431@columbia.edu

Office Hours

- Gabriel Young:
  - **TBA**
- Reed Palmer:
  - On campus students: **TBA**, SSW 10th floor lounge.
  - **The in-class TA will not be available for online students**
- Casey Bradshaw :
  - Online students: **TBA**, Zoom.
  - **The online TA will not be available for in-class students**

Learning Outcomes
STAT GR5205 is a first course in regression analysis for beginning graduate students in Statistics and related fields. The course covers the following topics: review of bivariate probability distributions, correlation, estimation and testing for linear, polynomial, and multiple regression models; indicator variables; analysis of residuals; selection of variables introduction to logistic regression & poisson regression, projection approach to linear models. Every topic is implemented in `R`. Students are required to perform simulations, exploratory analysis and inferential procedures using `R`.

Class Time and Location
STAT GR5205 is a 3 credit hour course. The face-to-face students meet MW 8:40am - 9:55am, 903 School of Social Work.

Prerequisites for GR5205
STAT GR5203 and GR5204 or the equivalent.

Text/Supplies

- *Applied Linear Statistical Models*, fourth edition; by Michael H. Kutner, Christopher J. Nachtscheim, and John Neter. Note: Identical content can be found in the first 14 chapters of *Applied Linear Statistical Models*, fifth edition; by Michael H. Kutnerm, Christopher J. Nachtscheim, and John Neter.
- Statistical Software `R`

Course Structure
The class follows a traditional lecture environment. New topics are presented in each class. In class examples and other relevant course materials will be posted on Canvas. Students should print class examples/notes out before each lecture. The solutions are covered during the scheduled contact hours.

Attendance
Students should ideally attend each class meeting. Even though attendance is not required, I will frequently give examples or hints during lecture that may show up on assessments.

Grading

Homework	25%	
Project	10%	+EC%
Midterm	30%	
Final	35%	

Homework
Homework assignments will be posted on Canvas at least one week in advance of the due date. Some homework problems are selected from the text and some are not. There will be 5-7 assignments given during the semester.

You are encouraged to discuss homework problems with your classmates, but all work submitted must be your own. If multiple students turn in identical solutions, all of them will receive a zero.

Not every problem of every homework set will be graded, and thus a perfect score does not necessarily mean that every problem was done correctly. Students are responsible for checking their own work against the solutions posted on the Canvas.

Homework/Project Submission (In-class)
Hard copies are required for submission. Turn the assignment in for full credit no later than the posted due date. Late homework assignments will occasionally be accepted with significant penalties.

Homework/Project Submission (Online)
Completed homework assignments must be uploaded as a .pdf extension on Canvas. Turn the assignment in for full credit no later than the posted due date. Late homework assignments will occasionally be accepted with significant penalties.

Project
The project will be an open ended case study constrained to a linear regression setting. This will assess the following topics:

1. Exploratory data analysis
2. Variable/model selection
3. Diagnostics
4. Model validation
5. Hypothesis testing

The instructor will provide the dataset used for this case study. Students who present high quality work on the project will receive extra credit. The amount of extra credit will be determined during final grade calculations.

Midterm
There will be one midterm exam. The midterm exam is scheduled for

- 10/16/2019

Final
The final exam is tentatively scheduled for

- **TBA**

You must take the final at the scheduled time.

Exams
You will have 75 minutes to complete the midterm exam, and 150 minutes to complete the final. Allowable materials will be discussed one week before the exam.

Exam absences
Make-up exams will not be given routinely. If you have a legitimate conflict with an exam date, it is incumbent upon you to make arrangements with the instructor to take the exam. An exam missed due to a documented illness or other unforeseeable (and documented) extraordinary circumstances must be made up before the test papers are returned to the class.

Test-Run Exam
For the online students only, please make sure you attend class on 10/02/2019 at 9:40am-10:00am, so that you may test out your camera, scanner and printer. If you do not attend class live on 10/02/2019 at 9:40am-10:00am and your web cam or other technologies do not work during your exam, you will not receive additional time to make up for any time lost. The *test-run exam* also counts as 10% of the Midterm grade **(online students only)**. During the test run you are required to:

1. Print out the document TestRun.pdf (posted under assignments).
2. Fill out a few questions.
3. Scan the completed document.
4. Upload the scanned document in Canvas (similar to homework).

Zoom pre-exam check

1. Prior to the Test-Run Exam, you must also complete the pre-exam check.
2. Follow the instructions presented in the assignments section.

We will not help trouble shoot during class for those who have technology difficulties. Rather, we ask that anyone who identifies an issue (or if we identify one and share with you), you must contact sps-help@columbia.edu to resolve the issues before the exam. Again, it is your responsibility to attend class next week to ensure your web cam, printer and scanner works properly before the exam.

Grading Scale

93 or more	A
90 to 92	A-
87 to 89	B+
83 to 86	B
80 to 82	B-
77 to 79	C+
70 to 76	C
60 to 69	D
Below 59	F

The above grading scale is guaranteed. However, final grades are often curved up based on the performance of your current class. If the class as a whole does extremely well, then do not expect a curve. If you are worried about earning a letter grade A, then work for at least a 93%. The letter grade $A+$ will be given to the top few students in the class (assuming they earned a final grade of at least 93%).

Academic Honesty
The university expressly prohibits academic dishonesty such as cheating, plagiarism, etc. It provides for a number of rather unpleasant consequences for students who are caught in violation of its academic honesty policies. Any suspected cheating on examinations will be referred to the Dean?s Discipline process, possibly resulting in course failure or College dismissal.

Tentative course outline

Date	Topic	Section in text
09/04/2019	Models, statistical models, review of conditional probability distributions	1.1, 1.2, A.1-A.9
09/09/2019	The simple linear regression model, least squares estimation	2.9, 1.3, 1.6
09/11/2019	Mean square error, coefficient of determination, maximum likelihood estimation	1.7, 2.9, notes
09/16/2019	Inference for simple linear regression	2.1-2.3
09/18/2019	Comparing models: analysis of variance & general F-statistic	2.7, 2.8
09/23/2019	Indicator variable for simple linear regression, prediction, extrapolation	notes, 2.4, 2.5
09/25/2019	Correlation, normal correlation model	2.11
09/30/2019	Simultaneous inference	2.6
10/02/2019	Introduction to multiple regression, primer on matrices **(Test Run Exam)**	notes, CH 5
10/07/2019	Multiple regression model, estimation of parameters	6.1-6.4
10/09/2019	Inference on slope parameters, multiple coefficient of determination	6.6, 6.5
10/14/2019	Comparing models: analysis of variance & general F-statistic	2.8
10/16/2019	**Midterm**	
10/21/2019	Prediction for multiple regression, regression diagnostics	6.9, 3.2
10/23/2019	Homogeneity of variance, F lack of fit test, variable transformations	3.3, 3.7, 3.9
10/28/2019	Non-linear response surface, indicator variables for multiple regression	8.1, 8.3
10/30/2019	Interaction, sum of squares type I, II, III	8.2, 8.5, 7.1-7.4
11/04/2019	No class	
11/06/2019	Multicollinearity	7.8
11/11/2019	Model selection	9.1, 9.2, 9.3
11/13/2019	Model selection, model validation	9.4, 9.5
11/18/2019	Influential observations	10.2-10.4
11/20/2019	Ethics in statistical modeling	notes
11/25/2019	Introduction to logistic regression	14.1-14.3
11/27/2019	No class	
12/02/2019	Multiple logistic regression	14.4
12/04/2019	Introduction to general linear models, poisson regression	14.14, 14.13
12/09/2019	The projection approach to regression	notes
12/11/2019	Study week	
12/16/2019	Finals week	
12/18/2019	Finals week	