36-401 MODERN REGRESSION
Fall 2011

Instructor: Rebecca Nugent
Baker Hall 232C
(412) 268-7830
rnugent@stat.cmu.edu
http://www.stat.cmu.edu/∼rnugent
Office Hours: Wed 1:30-2:30pm, Baker 232C

TAs: Zachary Kurtz, Sonia Todorova, Anne-Sophie Charest, Cong Lu
zkurtz, sktodoro, acharest, congl@stat.cmu.edu
Office Hours: ZK Thurs 10:30-11:30am, ST Wed 4-5pm, AC Tues 1-2pm
all in FMS 320

Class Meetings: Tuesdays and Thursdays 3:00-4:20pm, MM 103

Website: http://www.cmu.edu/blackboard
http://www.stat.cmu.edu/∼rnugent

Prerequisites: 36-226 or 36-626

Textbook: *Applied Linear Regression Models* by M. H. Kutner, C. J. Nachtsheim and J. Neter, McGraw-Hill (Fourth Edition, 2004)

General Course Plan: This course is an introduction to the real world of statistics and data analysis. We will explore real data sets, examine various models for the data, assess the validity of their assumptions, and determine which conclusions we can make (if any). Data analysis is a bit of an art; there may be several valid approaches. We will strongly emphasize the importance of critical thinking about the data and the question of interest. Our overall goal is to use a basic set of modeling tools to explore and analyze data and to present the results in a scientific report.

The course will begin with a review and discussion of exploratory methods, informal techniques for summarizing and viewing data. We then consider simple linear regression, a model that uses only one predictor. After briefly reviewing some linear algebra, we turn to multiple linear regression, a model that uses multiple variables to predict the response of interest. For all models, we will examine the underlying assumptions. More specifically, do the data support the assumptions? Do they contradict them? What are the consequences for inference? Finally, we will explore extra topics such as nonlinear regression or regression with time-dependent data.

Course Objectives:

1. Demonstrate how/when to use exploratory data analysis tools (e.g., graphical displays)
2. Develop model-building skills including evaluation of assumptions and interpretation of model-fitting results for linear regression models
3. Learn and apply the basic mathematical theory underlying linear regression models
4. Develop written and verbal communication skills for discussing conclusions and limitations of statistical evidence; present data analysis appropriately in a scientific report
5. Effectively use R, a widely-used statistical package, in data analysis

Course Work: Your grade in this course will be determined by homework assignments, in-class theory exams, and applied data analysis exams.

- Weekly homework assignments are due at the beginning of class (3pm), usually Thursdays (ten minute grace period; 15 point penalty 3:10-4:20; zero points after 4:20pm). Assignments should be submitted on paper; electronic versions will not be accepted (exceptions may be made depending on circumstances; instructor permission required).

  Homework Format: stapled; name on front page, initials on remaining pages; questions should be answered in order; ALL answers should be clearly marked and labeled; *just circling answers on R output is not acceptable; answers should be written up in the context of the problem.* Graphs should be as close to the corresponding problem as possible. Deviating from this format may result in loss of points on homework.

  Please see the TA or instructor during office hours for help with homework problems. Questions posed by email must be sent at least 24 hours before the time an assignment is due in order to guarantee a response.

- The theory examinations will focus primarily on your understanding of the mathematical material that is covered in the course. Both exams are scheduled in-class (see syllabus) and are required.

- The applied data analysis exams will correspond to the current modeling material. You will have one week with a data set; your analysis and results will be turned in as a structured report. We will cover the specifics in detail later in the course. There are three data analysis exams; the final exam will be during finals week and will be cumulative. Only two data analysis exams are required. Details are below in the grading scheme.

Grading policy: You are encouraged to discuss homework problems with your fellow students, however the work you submit must be your own. Acknowledge any help received on your assignments. Copied assignments will receive no credit. **Late assignments will not be accepted. Please come talk to me if there are difficulties; problems/conflicts must be discussed IN ADVANCE.** Your lowest homework grade will be dropped.

You have one week from the day an assignment, exam, etc is handed back in class to bring any grading issues, comments, complaints, etc to the attention of the instructor. Please note that if you are absent the day something is handed back, this deadline will not be extended unless arrangements have been made in advance with the instructor.

Final grades will be computed with the following weights (two options):

Homework	.25		
Theory Exam 1	.15		
Theory Exam 2	.20		
Data Analysis Exam 1	.15	Data Analysis Exam 2	.20
Data Analysis Exam 2 or Final	.25	Data Analysis Exam Final	.20

Final letter grades will be determined as usual: [90,100] = A, [80,89] = B, [70,79] = C, [60,69] = D, [$<$ 60] = R. Grades may be curved at the instructor's discretion (effort, improvement, etc).

Computing: The statistical computing package we will use in this course is R. R is available on many campus computers, and you may download a free version from `www.r-project.org`. You may also use the nearly-identical (but not free) program called S+, available on all campus computers. You can obtain a free temporary version from `myandrew`. This version is good for 1 year; you can keep renewing the license as long as you are a CMU student.

R References: manuals available on R website;
`http://www.stat.cmu.edu/~rnugent/teaching/introR`
Introductory Statistics with R, Peter Dalgaard; Springer-Verlag
Modern Applied Statistics with S-Plus Venables, Ripley; Springer

Laptop Policy: Students are expected to be participating in class; any laptop use during class should pertain directly to the class. Instructor reserves the right to not allow laptop use during class. When the class has a guest speaker, laptops must be turned off and put away.

Cellphones/Pagers, etc: All cellphones, pagers, beepers, and anything else that makes noise should either be turned off or silenced during class.

Communication: Assignments and class information will be posted on Blackboard. Help with using blackboard is available at www.cmu.edu/blackboard/help/.

Email: Sending email to your professor or teaching assistants should be treated as professional communication. Emails should have an appropriate greeting and ending; students should refrain from using any kind of "shortcuts", abbreviations, acronyms, slang, etc. in the email text. Emails not meeting these standards may not be answered.

Academic Integrity: All students are expected to comply with the CMU policy on academic integrity. This policy is online at www.studentaffairs.cmu.edu/acad_integ/acad_int.html

Cheating, copying, etc will not be tolerated; please ask if you unsure of whether or not your actions are complying with assignment/exam instructions. Always ask if you are unsure; always default to acknowledging any help received.

Video/Audiotaping: No student may record or tape any classroom activity without the express written consent of the professor. If a student believes that he/she is disabled and needs to record or tape classroom activities, he/she should contact the Office of Equal Opportunity Services, Disability Resources to request an appropriate accommodation.

Disability Services: If you have a disability and need special accomodations in this class, please contact the professor. You may also want to contact the Disability Resources office at 8-2013.

TENTATIVE SCHEDULE *subject to change*
(Do the readings before the class where they are listed)

Date	**Topic**	**Reading**	**Due**
Tu 8/30	Introduction and Review; Random Variables	Appendix A	
Th 9/1	Motivating Regression; Data Collection	1.1 – 1.5	Intro Survey
Tu 9/6	Statistical Models; Simple Linear Regression	1.1 – 1.5	
Th 9/8	EDA; Intro to R **Meet in Computer Labs**		HW 1
Tu 9/13	Estimation: Method of Least Squares	1.6 – 1.7	
Th 9/15	Estimation: Normal Model (MLE)	1.8	HW 2
Tu 9/20	SLR: Inference $\hat{B}$'s	2.1 – 2.3	
Th 9/22	SLR: Inference $\hat{Y}$, $\hat{Y}_{new}$	2.4 – 2.6	HW 3
Tu 9/27	ANOVA; F-tests; Correlation	2.7 – 2.9	
Th 9/29	Diagnostics; Transformations, Sp. Topics	3.1 – 3.3, 3.8 – 3.9, 4.1 – 4.5	HW 4
Tu 10/4	Midterm Review		
Th 10/6	**Theory Midterm 1; Data Analysis Exam 1 out**		
Tu 10/11	Linear Algebra; SLR in matrix form	Ch. 5	
Th 10/13	Multivariate Linear Regression; EDA	6.1 – 6.4	**Data Exam 1**
Tu 10/18	Categorical Variables; Polynomial Terms	8.1-8.7	
Th 10/20	Diagnostics, Transformations	6.8	HW 5
Tu 10/25	MLR Inference	6.6 - 6.8	
Th 10/27	Multicollinearity; Added Variables	7.1 - 7.3, 10.1-10.5	HW 6
Tu 11/1	Extra SS Inference; Partial F-tests	7.3 - 7.4	
Th 11/3	Interaction Terms; Model Selection	8.1 – 8.2, 9.1 - 9.4	HW 7
Tu 11/8	Model Selection; Influential Points	9.1 – 9.4, 10.1 - 10.5	
Th 11/10	Midterm Review		HW 8
Tu 11/15	**Theory Midterm 2; Data Analysis Exam 2 out**		
Th 11/17	DA Review		
Tu 11/22	TBA/Special Topics		**Data Exam 2**
Th 11/24	Thanksgiving Break; no class		
Tu 11/29	Special Topics	TBA	
Th 12/1	Special Topics	TBA	
Tu 12/6	Special Topics	TBA	
Th 12/8	Data Analysis in Practice; Regression Contest		HW 9
	Data Analysis Final out		

Fri 12/16 **DA Final (5pm)**