Statistics W2024 – Applied Linear Regression Analysis
Fall 2011

Class location and time: 703 Hamilton Hall, TR 10:35am-11:50am

Textbook: Chatterjee and Hadi, Regression Analysis by Example, 4th Edition

Course website: http://www.stat.columbia.edu/~martin/W2024/W2024.html

Lecturer: Martin Lindquist
Office: SSW 1031 (1255 Amsterdam Ave., 10th floor)
Telephone: (212) 851-2148
Email: martin@stat.columbia.edu

TA: Yunxiao Chen
Email: yc2710@columbia.edu

Prerequisites: STAT W1001, W1111 or W1211

Course description: The course goals are to develop critical thinking for statistical modeling in scientific and policy contexts and to develop useful data analysis skills based on the use of linear statistical models. Topics to be covered include: simple linear regression, multiple regression, non-linear regression and logistic regression models; Random and mixed effects models; and penalized regression. Implementation of statistical methods in the R statistical software will be covered for all topics. Examples will form an important part of the lectures. Problem sets will focus on conducting data analyses and reporting the results.

Calculators and computing: Each student needs a calculator for doing the homework and exams. The course will also use the statistical software package R which will be needed to solve homework problem.

Office Hours: The instructor’s and TA’s office hours can be found on the class webpage. It is important that you attend office hours if you have problems with the homework or the course material.

Homework: There will be weekly homework assignments, due one week after they are assigned. Late homework will not be accepted. The lowest homework score will be dropped from the calculation of your final grade.

Exams: There will be one in-class midterm and a final exam. Both exams will contain a take-home portion. The in-class portion of the midterm is scheduled for October 27. The Final Exam will cover material from the entire course. No make-up exams will be given.
Grading: Grades will be decided using the following weights:

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>30%</td>
</tr>
<tr>
<td>Midterm Exam</td>
<td>25%</td>
</tr>
<tr>
<td>Final</td>
<td>45%</td>
</tr>
</tbody>
</table>

Lecture Outline: This lecture outline is tentative and subject to change as the semester progresses.

**Week 1:** September 6 – 8

**Material:** Course motivation; Introduction to linear regression analysis; Review of statistical inference: sampling distributions, tests and confidence intervals; computer packages

**Reading:** 1.1 – 1.5

**Week 2:** September 13 – 15

**Material:** Two-sample t-tests, one-way ANOVA, multiple comparison procedures

**Reading:** Handouts

**Week 3:** September 20 - 22

**Material:** Simple linear regression, tests and confidence intervals for slope and intercept, prediction, model assessment.

**Reading:** 2.1 - 2.8

**Week 4:** September 27 – 29

**Material:** Correlation models; R^2 and the ANOVA table

**Reading:** 2.9 – 2.10, Handouts

**Week 5:** October 4 - 6

**Material:** Multiple regression and inferential tools for multiple regression

**Reading:** 3.1 – 3.11

**Week 6:** October 11 - 13

**Material:** Regression Diagnosis

**Reading:** 4.1 – 4.14
Week 7: October 18 - 20

Material: Dummy variables, two-way ANOVA, ANCOVA
Reading: 5.1 – 5.7, Handouts

Week 8: October 25 - 27

Material: Review, Midterm Exam in class on October 27.

Week 9: November 1 - 3

Material: Regression with transformed variables, polynomial regression
Reading: 6.1 – 6.9

Week 10: November 10

Material: Weighted least squares, serial correlation
Reading: 7.1-7.5, 8.1-8.10

Week 11: November 15 - 17

Material: Multicollinearity; bias variance tradeoff, penalized regression
Reading: 9.1-9.9, 10.1 – 10.8

Week 12: November 22

Material: Variable selection; non-linear regression
Reading: 11.1 – 11.15, Handouts

Week 13: November 29 - December 1

Material: Logistic regression, Poisson Regression
Reading: 12.1 – 12.8

Week 14: December 6 - 8

Material: Generalized linear models; Review
Reading: 13.1 – 13.6