STAT 462: Applied Regression Analysis Spring 2014

Class:
MW 11:15 AM - 12:05 PM, 009 Life Sciences Bldg
F 11:15 AM - 12:05 PM, 111 Boucke

Instructor: Xiaoyue (Maggie) Niu
Office 323E Thomas
Phone 863-8127
E-mail xiaoyue@psu.edu
Office Hours M W 1:00-2:30pm or by appointment

TA information: Jingyi Ye
Office 301 Thomas
Email jxy196@psu.edu
Office Hours T 2:30 - 4:00 and Th 2:15 - 3:45

Textbook:
*Applied Linear Regression Models*, 4th edition, by Neter, Nachtsheim, Kutner
(recommended but NOT required)
Or the first half part of *Applied Linear Statistical Models*, 5th edition by Neter, Nachtsheim, Kutner and Li

Course Objective:

STAT 462 is an applied linear regression course that involves "hands on" data analysis. Students enrolling for this course should have taken at least one other Statistics course and should be familiar with the basic fundamentals of statistical testing and estimation. Data analysis is emphasized, so computers will be used frequently during the course. Throughout the course, Minitab for Windows will be used to analyze the data for lecture demonstrations in class and lab activities. Those wishing to install Minitab on their own computers may go to [www.minitab.com/education](http://www.minitab.com/education) for details.

Website:
Class announcements and materials will be regularly posted on ANGEL, so it is recommended that you check the site frequently. Such materials as lecture notes, homework and lab assignments, homework and exam solutions, etc. will be posted.

Homework:
Homework assignments will be given and collected on Friday at the beginning of class. You must show all work on the homework problems to get full credit. Doing the homework promptly and carefully is necessary for learning the material. A reasonable
amount of collaboration is allowed on homework. However, each student must turn in his or her own written work which reflects his or her own individual understanding of the material. **Late home works will not be accepted unless the student has the instructor’s permission in advance.** Note that the lowest homework score will be dropped.

**Labs:**
We will practice the tools learned in previous lectures during lab sessions. Lab activities need to be submitted in the drop box on Angel during the class time. **The lowest grade for lab will be dropped.**

**Exams:**
Two midterm exams will be given. Most questions will be short answer and will include Minitab analysis output. You will be allowed to bring in a simple calculator and one double-sided formula sheet made up by yourself. No early or late exams will be allowed without a legitimate excuse. Tentative dates for the two exams are Feb 24 and April 7.

**Final Project:**
The final project is a team project. The teams will be assigned after the 1st midterm (to allow student drops). The team size ranges from 3-4 members. Students are encouraged to pick their own topics and data to analyze. A back-up dataset will also be provided in case you couldn’t find your own. If you pick your own dataset, please set up an appointment with the instructor before April 1 to get approval. During the last one or two weeks of class, each team will present their project to the rest of the class. Based on the feedbacks given by the instructor and other students, you can revise your analysis and report. The final report is due May 6, 12:00am.

**Evaluation:**

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>20%</td>
</tr>
<tr>
<td>Lab activities</td>
<td>15%</td>
</tr>
<tr>
<td>Midterms (2)</td>
<td>40%</td>
</tr>
<tr>
<td>Final project presentation</td>
<td>5%</td>
</tr>
<tr>
<td>Final project report</td>
<td>20%</td>
</tr>
</tbody>
</table>

**Grading Scale:**
A: 93%; A-: 90%; B+: 87%; B: 83%; B-: 80%; C+: 77%; C: 70%; D: 60%; F: <60%

**Academic Integrity:**
All Penn State and Eberly College of Science policies regarding academic integrity apply to this course. See [http://www.science.psu.edu/academic/Integrity/index.html](http://www.science.psu.edu/academic/Integrity/index.html) for details.
ECOS Code of Mutual Respect and Cooperation:
The Eberly College of Science Code of Mutual Respect and Cooperation:  
http://www.science.psu.edu/climate/Code-of-Mutual-Respect_final.pdf embodies the 
values that we hope our faculty, staff, and students possess and will endorse to make The 
Eberly College of Science a place where every individual feels respected and valued, as 
well as challenged and rewarded.

Disability Policy Notice:
Penn State welcomes students with disabilities into the University's educational 
programs. If you have a disability-related need for reasonable academic adjustments in 
this course, contact the Office for Disability Services (ODS) at 814-863-1807 (V/TTY). 
For further information regarding ODS, please visit the Office for Disability Services 
Web site at http://equity.psu.edu/ods/.
In order to receive consideration for course accommodations, you must contact ODS and 
provide documentation (see the documentation guidelines at 
http://equity.psu.edu/ods/guidelines/documentation-guidelines). If the documentation 
supports the need for academic adjustments, ODS will provide a letter identifying 
appropriate academic adjustments. Please share this letter and discuss the adjustments 
with your instructor as early in the course as possible. You must contact ODS and request 
academic adjustment letters at the beginning of each semester.
Specific Topics Usually Covered

1. Simple Linear Regression Model
   - Model for E(Y), model for distribution of errors
   - Least squares estimation of model for E(Y)
   - Estimation of variance
2. Inferences for Simple Linear Regression Model
   - Inferences concerning the slope (confidence intervals and t-test)
   - Confidence interval estimate of the mean Y at a specific X
   - Prediction interval for a new Y
   - Analysis of Variance partitioning of variation in Y
   - R-squared calculation and interpretation
3. Diagnostic Procedures for Aptness of Model: assessing regression assumptions
   - Residual analyses
   - Plots of residuals versus fits, residuals versus x
   - Tests for normality of residuals
   - Lack of Fit test, Pure Error, Lack of Fit concepts
   - Transformations as solution to problems with the model
4. Multiple Regression Models and Estimation
   - Matrix Notations
   - Hyperplane extension to simple linear model
   - Basic estimation and inference for multiple regression
5. Additional Topics for Multiple Regression Analysis
   - General Linear F test and Sequential SS
   - Effects of a variable controlled for other predictors
   - Sequential SS
   - Partial Correlation
   - Multicollinearity between X variables
   - Effect on standard deviations of coefficients
   - Problems interpreting effects of individual variables
   - Apparent conflicts between overall F test and individual variable t tests
   - Benefits of designed experiments
6. Categorical Predictor Variables
   - Indicator Variables
   - Interpretation of models containing indicator variables
7. Model Comparison and Selection Methods
   - R2, MSE, Cp, and PRESS criteria
   - Stepwise algorithms
8. Logistic Regression: categorical outcome variables
   - Binary outcome: Bernoulli Distribution
   - Interpretation of models: odds ratio
9. Miscellaneous Topics as Time Permits