STAT 462: Applied Regression Analysis
Summer 2017

Instructor: Muzi Zhang
Email: muz135@psu.edu
TA: Rui Wan
Email: rxw89@psu.edu
Office/Hours: By appointment. Appointments should be made by contacting the instructor though email or Canvas.

Class Schedule: MWF 11:10am-12:25pm Willard Bldg 103 (Lecture)
TuTh 11:10am-12:25pm IST Bldg 203 (Lab)
Attendance to lecture and lab is required.

Text: *Linear Models with R*, Faraway, Second Edition. (Strongly recommended, will be used for homework and references on application part of the course)
Applied Linear Regression Models, 5th ed., by Neter, Nachtsheim, Kutner, Wasserman (Old textbook, will be used for theory part of the course)

Description: Statistics 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.

Prerequisite: Any intro stat course (200 or higher). This prerequisite is taken very seriously! Any student that is not comfortable with the introductory statistics may struggle with this course. Such students should consider taking STAT 460, which also covers regression and uses less mathematics.

Computer Usage: This course will utilize R for all computational needs. R is free and can be able to download on your computer. Through this link: https://www.r-project.org/, you can download R software. Through this link: https://www.rstudio.com/, you can download R-studio, which is an interface for R. You have to install R first in order to use R-studio. R-studio is recommended to use.

Announcements: Lecture notes, lab activities, assignments, and all due dates will be posted on *Canvas*, available at www.canvas.psu.edu or from the PSU homepage. Students are expected to check this regularly for updates.

Homework: 30%
There will be (tentatively) 5 homework assignments through the semester. The homework will be posted on Canvas on Friday and due the following Friday. Each student need to turn in their homework on Canvas on Friday before the end of the day. No late homework will be accepted unless permitted in advance.

Exams: 40%
There will be two exams, and no final exam. A simple calculator and a one-sided note sheet may be used. Tables will be provided as needed. The tentative dates of the exam will be Fridays in week 3 and week 5. *Conflicts on exam dates must be resolved in advance.*

Quiz: 5%
Quizzes will be given randomly in lecture class. They will be graded as for attendance.
The lowest score of quizzes can be dropped. Quizzes **cannot** be made up for any reason.

Lab Activities: 5%
There will be 10-15mins of introduction or illustration of R at the beginning of lab.
On Tuesday lab, lab assignments will be assigned at the beginning and need to be finished and submitted on Canvas by the end of lab. The lowest lab score can be dropped. Lab assignments cannot be made up and not accepted late.
On Thursday lab, it will be reserved for students to work on the homework with the instructor present. No actual lab assignments will be given, just homework problems.

Final Project: 20%
This class will have a final project. The project consists of putting together all of the tools learned throughout the class. Students will have to analyze a particular data set and put together a report summarizing the analysis. More details will follow.

Letter Grades: Semester grades are assigned according to this scale. The final score may be curved.

93 – 99%	A	77 – 79%	C+
90 – 92%	A-	70 – 76%	C
87 – 89%	B+	60 – 69%	D
83 – 86%	B	0 – 59%	F
80 – 82%	B-		

Academic Integrity:
All Penn State policies regarding ethics and honorable behavior apply to this course. Academic integrity is the pursuit of scholarly activity free from fraud and deception and is an educational objective of this institution. All University policies regarding academic integrity apply to this course. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating of information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used without informing the instructor, or tampering with the academic work of other students. Any instances of academic dishonesty WILL be pursued under the University and Eberly College of Science regulations concerning academic integrity.

ECOS Code of Mutual Respect and Cooperation:
The Eberly College of Science Code of Mutual Respect and Cooperation 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 feel respected and valued, as well as challenged and rewarded.

Disability Services:
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. In order to receive consideration for course accommodations, you must contact ODS and provide documentation (see the documentation guidelines). If the documentation supports the need for academic adjustments, ODS will provide a letter identifying appropriate academic adjustments. Please provide the letter and discuss any adjustments with me as early in the course as possible. You must contact ODS and request academic adjustment letters at the beginning of each semester.

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 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
- R-squared calculation and interpretation

3. Diagnostic procedures for aptness of model

- 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. Matrix Notation and Literacy

- X matrix, vector, y vector
- $(X'X)^{-1}$ X'Y estimates coefficient vector
- Variance- Covariance matrix

5. Multiple Regression Models and Estimation

- Hyperplane extension to simple linear model
- Interaction models
- Basic estimation and inference for multiple regression

6. General Linear F test and Sequential SS

- Reduced and Full models
- F test for general linear hypotheses
- Effects of a variable controlled for other predictors
    - Sequential SS
    - Partial correlation

7. 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

8. Polynomial Regression Models

9. Categorical Predictor Variables

- Indicator Variables
- Interpretation of models containing indicator variables
- Piecewise regression

10. More Diagnostic Measures and Remedial Measures for Lack of Fit

- Variance Inflation Factors
- Deleted Residuals
- Influence statistics - Hat matrix, Cook's D and related measures

11. Examining All Possible Regressions

- R^2, MSE , C_p, and PRESS criteria
- Stepwise algorithms

12. Additional Topics as Time Permits

- Estimating the regression model if residuals have autocorrelation.
- Logistic regression models