Faculty
Instructor: Dr. Linda Clark
Teaching Assistant: Zhanxiong Xu
(contact via ANGEL email ONLY… emails sent to psu email addresses will NOT be answered)

Office Hours
Dr. Clark Wed 11:15 – 12:15 and Thurs 10:30 – 11:30
Zhanxioing TBA

General Description
Statistics (STAT) 501 is an applied linear regression course that emphasizes data analysis and interpretation. Generally, statistical regression is a collection of methods for determining and using models that explain how a response variable (dependent variable) relates to one or more explanatory variables (predictor variables). A list of specific topics usually covered is given later in this document.

Prerequisites
The prerequisite for this course is STAT 500. Students without a firm understanding of the material in STAT 500 should NOT be in this course. This course will build on this material and knowledge of introductory statistics IS part of the assessment of this course.

Text Requirements
Use either Applied Linear Regression Models (4th edition) by Kutner, Nachtsheim, and Neter (click here for the book’s official site) or Applied Linear Statistical Models (5th edition) by Kutner, Nachtsheim, Neter, and Li. The first 14 chapters of this text are identical to the Applied Linear Regression Models book listed first. This book also includes 16 chapters on analysis of variance and experimental design not covered in this course. Readings will be assigned from the text.

Software Requirements
You will need statistical software to be able to do most assignments.
Minitab: We recommend Minitab version 14 or above. The student version is also okay. The regression capabilities are identical across these versions. If you have access to another statistical program such as SPSS, JMP, R, or SAS, you can use that program however, those programs will NOT be supported in this course. Excel will not be sufficient for this course.

Grading and General Requirements

Eight homework assignments
Regression project
Two mid terms
Final comprehensive exam given during finals week

There will be weekly homework assignments, three “practice” quizzes, and three exams. Homework will count as 45% of the course grade, the practice quizzes will count as 5%, and exams will count as the remaining 50% (each exam counts equally). Grades will be awarded accordingly: [93.0%-100%] = A, [90%-92.9999%) = A-, [87%-89.999%) = B+, [84%-87%) = B, [80%-84%) = B-, [77%-80%) = C+, [70%-77%) = C, [60%-70%) = D.

Discussion Forum: I encourage all class members to post content or class related questions to the discussion forum. I also encourage peer answers to any posted questions. I will respond to any posts within a 24 hour period unless the class is otherwise notified. Please note that homework and exam related questions posted within 24 hours of due dates will NOT be answered. You will need to ask your questions in advance of the due dates for these assignments. (any content questions sent to my ANGEL email will be redirected to the discussion forum. Any emails sent to my psu address will NOT be answered at all)

Lectures: These will take place Monday and Wednesdays of the course. Fridays will be in a computer lab where you will have the opportunity to practice the material. I reserve the right to introduce new material in Friday labs. Attendance is NOT required and please do not email me with notifications that you will not be in class or lab.

Homework: Homework assignments will be “due” by the start of class on Friday. Each assignment will state the due date, so please be aware of what day each assignment is due. NO LATE HOMEWORKS WILL BE ACCEPTED.

Project: A project will be assigned the first week in April. This project will NOT be assigned before this time.

Exams: Exam 1 will be February 14. Exam 2 will be April 4th. The Final Exam will be administered DURING FINALS WEEK. These dates are firm deadlines and you are required to meet them for completion of this course! Please note these dates now and be prepared to meet these deadlines. If you have many obligations and conflicts with the requirements set forth in this course, then I suggest you consider enrolling in this course
during a later semester. As for the content of the exams, the midterm exams will be a blend of conceptual questions and applied questions. The final exam will be cumulative.

Attendance: Attendance in this course is NOT mandatory. If you are going to miss class you do NOT need to notify me. However you will be responsible for getting the lecture material from another student or visiting me in office hours. As I am flexible with lecture and lab attendance I expect the exam dates to be a priority on your schedule and only under extreme circumstances will I allow make up exams.

Academic Integrity: It is encouraged for you to work with your colleagues regarding homework assignments, but the solutions you submit MUST be your own. Furthermore, it is to be understood that no collaboration is to occur regarding the examinations. For any material or ideas obtained from other sources, such as the text or things you see on the web, in the library, etc., a source reference must be given. Direct quotes from any source must be identified as such. This course will abide by the Penn State Academic Integrity Policy.

Accommodations for Students with Disabilities: 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 website. 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 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.

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 feels respected and valued, as well as challenged and rewarded.

Campus Emergency: In case of weather-related delays at the University, this online course will proceed as planned. Your instructor will inform you if there are any extenuating circumstances regarding content or activity due dates in the course due to weather delays. If you are affected by a weather-related emergency, then please contact your instructor at the earliest possible time to make special arrangements.
Topics Covered in STAT 501

1. Simple Linear Regression Model: One Predictor Variable
   - model for $E(Y)$, model for distribution of errors
   - least squares estimation of model for $E(Y)$
   - estimation of variance
   - regression through the origin
   - distinguishing regression from correlation

2. Inferences for Simple Linear Model
   - inferences concerning the slope (confidence intervals and $t$-tests)
   - confidence interval estimate of the mean $Y$ at a specific $X$
   - prediction interval for a new $Y$
   - analysis of variance (ANOVA) partitioning of variation in $Y$
   - calculation and interpretation of $R^2$

3. Diagnostic Procedures for Aptness of Model
   - residual analyses
     - plots of residuals versus fits, residuals versus $x$, residuals versus new $x$
     - tests for normality of residuals
     - lack of fit test, pure error, lack of fit concepts
   - transformations as a solution to problems with the model
   - weighted least squares as a solution for variance problems

4. Matrix Notation for Regression Models
   - $X$ matrix, $\beta$ vector, matrix formula for estimating coefficients
   - linear dependence issues
   - variance-covariance matrix of sample coefficients

5. Multiple Regression Models and Estimation: Multiple Predictor Variables
   - basic estimation and statistical inference within multiple regression
   - interaction terms and the interpretation of interaction
6. General Linear $F$-Test for Testing Hypotheses
   • reduced and full models associated with hypotheses about the model’s coefficients
   • $F$-test for general linear hypotheses

7. Assessing and Interpreting the Effect of a Single Predictor Variable Within a Multiple Regression
   • properly interpreting the $t$-test
   • sequential sums of squares
   • partial correlation between $y$ and an $x$ variable

8. Examining All Possible Regressions to Identify the Potential Models
   • $R^2$, MSE, $C_p$, AIC, BIC, and PRESS criteria
   • stepwise algorithms for identifying models

9. Problems Caused by Correlations (Confounding) Among Predictor Variables
   • inflation effects on standard deviations of coefficients
   • problems in interpreting effects of individual variables
   • variance inflation factors (VIFs)

10. Incorporating Categorical Predictor Variables
    • indicator variables
    • interpretation of models containing indicator variables

11. More Diagnostic Measures and Remedial Measures for Lack of Fit
    • deleted residuals
    • influence statistics - hat matrix, Cook’s D and related measures

12. Time Series Issues: Autocorrelation in Errors and Autoregressive Time Series Models

13. Polynomial Regression Models
    • the hierarchy principle

14. Logistic Regression Models for a Binary Response Variable