**STAT 501 – Course Syllabus – Fall 2015**

#### STAT 501 Regression Methods (4)

Analysis of research data through simple and multiple regression and correlation; polynomial models; indicator variables; step-wise, piece-wise, and logistic regression.

#### Instructor – Srabashi Basu

Feel free to contact me any time through Angel using the "Communicate" tool. I will try to respond as quickly as possible, but remember that there is considerable time difference between India and USA.

#### Teaching Assistant – Yuji Samizo

Email: yzs138@psu.edu.

#### Prerequisites

Stat 500 or equivalent; matrix algebra. Students enrolling for this course should have taken at least one other statistics course and should be conversant with the basic fundamentals of statistical testing and estimation. They also should have a rudimentary knowledge of matrices.

#### 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 is given later in this document.

This course is cohort-based, which means that there is an established start and end date, and that you will interact with other students and the instructor throughout the course. The course consists of the textbook, software (Minitab version 17), a course website in Angel, and a Drupal website containing supplemental notes.

#### Required Course Text

#### This course requires the textbook *Applied Linear Regression Models* (4th edition) by Kutner, Nachtsheim, and Neter. There is a website for the book: <https://netfiles.umn.edu/users/nacht001/www/nachtsheim/>. The extended version of this book, *Applied Linear Statistical Models* (5th edition) by Kutner, Nachtsheim, Neter, and Li will also do. The first 14 chapters of the second text are identical to the first text. The second text also includes 16 chapters on analysis of variance and experimental design not covered in this course (but covered in STAT 502). This text is considered to be one of the bibles of applied statistics, so it probably will have value to you beyond this course. Supplemental notes covering the course material are available on the Drupal website.

#### Statistical Software

You will need statistical software to be able to do the homework assignments and exams.

* **Minitab 17**: The student version is okay. If you have access to another statistical program such as SPSS, JMP, R, or SAS, feel free to use that program. However, Excel will not be sufficient. I will be able to help only with Minitab. You may use the free Web version of Minitab at <https://webapps.psu.edu/>.
* **Purchasing the Student Version**: Some students may already have the Minitab 17 student version as it is required for STAT 500. Others may wish to do an Internet search to find a good purchase price for the Minitab student version.
* **Leasing the Full Version**: The full version of Minitab can be leased at <http://estore.onthehub.com/WebStore/ProductsByMajorVersionList.aspx>.

#### Grading and General Requirements

* **Grades**: There will be twelve weekly homework assignments and three exams. Homework will count as 40% of the course grade and exams will count as the remaining 60% (each exam counts equally). Grades will be awarded according to the following: [94%-100%] = A, [90%-94%) = A-, [87%-90%) = B+, [83%-87%) = B, [80%-83%) = B-, [77%-80%) = C+, [70%-77%) = C, [60%-70%) = D.
* **Lectures**: The lectures for the course have been combined into what we call the **Lessons**. You are expected to read over the listed sections from the textbook as well as the online material on the Drupal website. These readings will be crucial to help focus your studies. **Most students will spend, on average, about 8 hours working through each lesson. You may need more (or less) time depending on your prior experience in statistics**.
* **Homeworks**: Each homework will have the due date written clearly on the course website in Angel and homework assignments will be due by **11:00 p.m. (Eastern Time - USA)**. You *may* be given a day as grace period for homeworks if you need it (you must contact myself and the grader ahead of time if you need this grace period). I understand that some of you will have work/personal obligations that might make it necessary to use this grace period; however, please do not abuse this as it is meant to be used as a last resort. I also ask that you please try to submit your homework answers by the due date so that I may post the solutions as soon as possible for the rest of the students. **Late homework for which no grace period has been pre-arranged cannot be accepted for credit.**
* **Exams**: Unlike the homework assignments, there will be NO grace period for the exams regarding the due date. 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! As for the content, the exams will be a blend of conceptual and applied questions. The two **midterm exams** must be completed within a 3-hour time period at a time of your choice between the assigned and due dates. The **final exam** is a **proctored exam**. The proctor's role is ensuring the academic integrity of the exam process on behalf of Penn State. You will need to secure a proctor in order to take exams in your courses. *A proctor will not automatically be assigned to you*; rather, you must make the necessary contacts to secure a professional who will serve in this capacity. Further details are available from the course website on Angel.
* **Assignment and Exam Due Dates**: Theses are given in the Course Schedule available in Angel.
* **Discussion/Message Board**: The online sections of STAT 501 follow closely with the in-class sections. Students in the online sections will have essentially the same course activities and requirements as students in the in-class sections. For online students, there are also message boards that allow you to ask and answer questions (thus gaining a sense of participation in a learning community). I highly encourage you to answer questions posted on the message boards. If a fellow student correctly answers a question that you post, then I will not post a response unless I wish to add something else. Otherwise, I will respond to your posting as soon as possible (usually within 24 hours, but I will let you know if I will be unable to access the course for any longer period of time). The term **Netiquette** refers to the etiquette guidelines for electronic communications, such as e-mail and bulletin board postings. Netiquette covers not only rules to maintain civility in discussions, but also special guidelines unique to the electronic nature of forum messages. Please review Virginia Shea's "[The Core Rules of Netiquette](http://www.albion.com/netiquette/corerules.html)" for general guidelines that should be followed when communicating in this course.

#### 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. 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. All exam answers must be your own, and you must not provide any assistance to other students during exams. For more information on academic integrity, see [Penn State's Academic Integrity Policy](http://science.psu.edu/current-students/Integrity/Policy.html).

#### Accommodations For Students with Disabilities

#### Penn State welcomes students with disabilities into the University's educational programs. Every Penn State campus has an office for students with disabilities. The Office for Disability Services (ODS) Web site provides contact information for every Penn State campus: <http://equity.psu.edu/ods/dcl>. For further information, please visit the Office for Disability Services Web site: <http://equity.psu.edu/ods>. In order to receive consideration for reasonable accommodations, you must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: <http://equity.psu.edu/ods/guidelines>. If the documentation supports your request for reasonable accommodations, your campus’s disability services office will provide you with an accommodation letter.  Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. You must follow this process for every semester that you request accommodations.

#### Course Topics

1. Simple Linear Regression (SLR)

* SLR model and least squares estimation
* Mean squared error
* R-squared
* Correlation

1. SLR Model Evaluation

* Confidence intervals for the intercept & slope
* Hypothesis tests for the intercept & slope
* ANOVA
* Lack of fit test for SLR

1. SLR Estimation & Prediction

* Confidence interval for the mean response
* Prediction interval for a new response

1. SLR Model Assumptions

* SLR assumptions/conditions
* Residuals vs fits
* Residuals vs predictor
* Residuals vs order
* Normal probability plot

1. Multiple Linear Regression (MLR)

* MLR model
* Matrix formulation of the MLR model

1. MLR Model Evaluation

* General linear test
* Partial R-squared
* Lack of fit test for MLR

1. MLR Estimation, Prediction & Model Assumptions
   * Confidence interval for the mean response
   * Prediction interval for a new response
   * MLR assumptions/conditions
   * Assessing the assumptions graphically
   * Testing the assumptions
2. Categorical Predictors

* Coding qualitative/categorical predictors
* Additive effects
* Interaction effects (quantitative by indicator)
* Piecewise linear models

1. Data Transformations

* Transforming predictors
* Transforming the response
* Polynomial regression
* Interaction models (quantitative by quantitative)
* Box-Cox transformations

1. Model Building

* Overfitting and variable selection
* Cross-validation
* PRESS, Mallow’s Cp, information criteria
* Subset selection
* Stepwise selection

1. Outliers & Influential Points

* Standardized residuals
* Leverage
* Cook’s distance

1. Multicollinearity & Other Regression Pitfalls

* Multicollinearity
* Detecting and responding to multicollinearity
* Extrapolation
* Other pitfalls

1. Weighted Least Squares & Robust Regression

* Weighted least squares
* Robust regression
* Resistant regression
* Regression depth

1. Time Series & Autocorrelation

* Time series
* Identifying and responding to autocorrelation
* Advanced methods

1. Nonlinear Regression & Generalized Linear Models

* Nonlinear least squares
* Logistic regression
* Poisson regression
* GLMs