Regression and Analysis of Variance (EDUC 767-001)
Fall 2019

Instructor: Dr. Wendy Chan
Email: wechan@upenn.edu
Phone: 215-573-2038
Office: GSE Room 342
Office Hours: By appointment

Course Logistics: This class will meet once per week on Wednesday from 2:00 - 4:30 PM in GSE 203.

Course Description: This course covers concepts related to general linear models and will primarily focus on topics such as correlation, regression and analysis-of-variance. We will compute examples of many of these models to understand their applications in practice. The prerequisite for this course is EDUC 667 or an equivalent graduate-level introduction to statistics course.

Course and Teaching Philosophy: This course is designed to provide you with opportunities to improve your understanding of quantitative methods and your ability to conduct statistical analysis. You should finish this course with an enhanced ability to think critically about how data is generated and how researchers and practitioners apply inferential methods to reach conclusions. In statistics, you will often be asked to explain and communicate concepts and research findings in a way that is easily interpretable by others.

As with any course, *you* are responsible for making sure that you understand the material. If you feel that you need help, please do not hesitate to ask me. Do not let yourself fall behind because concepts tend to build upon one another. This class is an open learning environment so please ask questions if there is any confusion. Your questions may help others who are confused and will also help me gauge student understanding. There will be a few times where I will decide that we need to move on and ask that you follow up outside of class, but it is never bad to ask a question. If you take responsibility for making sure that you understand the material and seek help when you don't, I will take responsibility for being there to help you.

Textbooks, Software, and Other Course Material:

- Faraway, J. (2005). *Linear Models with R*. Boca Raton, FL: Chapman & Hall/CRC.
- Hertzog, C., & Rovine, M. (1985). Repeated measures analysis of variance in developmental research: Selected issues. *Child Development, 56*, 787 – 809.
- Kleinbaum, D., Kupper, L., Nizam, A., & Rosenberg, E. (2014). *Applied Regression Analysis and Other Multivariable Methods*. (3rd edition). Boston, MA: CENGAGE Learning.
- Kutner, M., Nachtscheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th edition). New York, NY: McGraw-Hill.
- Rovine, M., & von Eye, A. (1991). *Applied Computational Statistics in Longitudinal Research.* Boston: Academic Press
- We will use R and R Studio in this class, which are free and available for download at `https://www.r-project.org/` and `https://www.rstudio.com/`, respectively. During class, I will periodically use the R Studio Cloud workspace (`https://rstudio.cloud/`) to demonstrate and

teach R code. You are welcome to use another software program, but please be aware that I will not be able to troubleshoot issues related to any other software program other than R. However, many software programs, including R, have useful help forums online. If you run into computing issues, I encourage you to try searching for solutions online first as another person may have already solved it.

Course Grading:

- Problem Sets 20%
- Midterm Exam 40%
- Final Exam 40%

Problem Sets: There will be **4** problem sets. The problem sets are designed to reinforce the material from lectures and readings as well as provide opportunities to apply the concepts using data. Please ensure that you allocate enough time for each problem set so that you can try the problems and return to them at a later point, if needed. For problems involving R, be sure to summarize and interpret the findings. Please **do not** include unedited R code or output unless you are asked to do so. For problems that explicitly ask you to use R, solutions based on software other than R will **not** be accepted.

Problem sets will be posted on *Canvas*. Each assignment must be turned in at the beginning of each class. Late assignments will **not** be accepted. If you have any questions, concerns, or comments about a problem set, please inform me *in writing* (by e-mail) and *prior* to any due date.

Collaboration Policy: You are free to work together on the problem sets. In fact, you are *encouraged* to work in groups. Working in groups is an excellent way to learn, motivate each other, and reinforce understanding of key concepts. However, the completed problem sets that are turned in must be individual work and must be written *in your own words*. Frequently, students (and professors!) think that they understand something, but when they sit down to try it themselves, they realize they do not. Writing out your assignment in your own words is how you will know that you really understand.

Midterm/Final Exam: There will be an in-class midterm and final exam. You will need a calculator (scientific or graphing) for both exams. You are allowed to bring an 8.5 x 11 inch cheat sheet for both exams, which can include formulas, definitions, and diagrams to help you solve the problems. You must take the exam in the scheduled time slot except in the event of truly extenuating circumstances, such as illness, illness in the family, an important and immovable court date or comparable event. In the event that any problem occurs, please contact me as soon as you learn of it.

The midterm exam is tentatively scheduled during our class meeting on **Wednesday, October 30, 2019**. The final exam, which will be cumulative, is tentatively scheduled for **Wednesday, December 18, 2019** during class time.

Other Information:

Academic Honesty: Please consult the GSE Student Handbook on the following webpage for details on expected student conduct: `http://www.gse.upenn.edu/policies/academicintegrity`. Please be sure to read the material in this document. Plagiarism or cheating of any kind will be dealt with according to University policy, which can be found at: `http://www.upenn.edu/academicintegrity`.

Use of cell phones and laptops: Please do not use your cell phone in class. Of course, you may set your phone to vibrate and leave the class to answer emergency calls or send emergency texts. Please do not use your laptop for anything other than taking notes in class and using R.

Topic	Reading Assignment
Introduction: The General Linear Model Review of some descriptive statistics	Kleinbaum, Chapters 1-4, Faraway (2005), Chapter 1, Neter (2005), Chapter 1, pp. 2 – 9
Simple straight-line regression	Kleinbaum, Chapter 5 Neter (2005), Chapter 1, pp. 9 – 27
Correlation and its relation to straight line regression	Kleinbaum, Chapter 6 Neter, Chapter 2, 5
Multiple regression: regression with two or more predictor variables	Kleinbaum, Chapters 7 – 9 Faraway, Chapter 2, 3 Neter, Chapter 6
The correlation matrix	Kleinbaum, Chapter 10
Interactions and product terms in regression	Kleinbaum, Chapter 11 Neter, Chapter 8
Describing and graphing the data; residuals	Kleinbaum, Chapter 14 Faraway, Chapters 4, 5 Neter, Chapter 3
Polynomial and nonlinear regression	Kleinbaum, Chapter 15 Faraway, Chapters 7, 10 Neter, Chapter 8
Hierarchical and stepwise regression R^2 and F	Kleinbaum, Chapter 16 Faraway, Chapter 8 Neter, Chapter 9
One-way ANOVA	Kleinbaum, Chapter 17 Faraway, Chapter 14 Neter, Chapter 16
Two-way and higher-order ANOVA	Kleinbaum, Chapters 19, 20 Faraway, Chapter 15
Repeated measures ANOVA	Rovine (1991), Chapter 2 Hertzog & Rovine (1985), Chapter 19
ANCOVA	Kleinbaum, Chapter 13 Faraway, Chapter 13, 22
Missing data	Faraway, Chapter 12