Stat 705: Data Analysis II

Spring 2019

- Lecture: 9:40–10:55AM, Wednesday/Friday LeConte 201A.

- Instructor: Dr. Yen-Yi Ho
  Office: LeConte College 209G
  E-mail: hoyen@stat.sc.edu.
  Office hours: Wednesday, Friday 11a-12pm, and by appointment.
  Class website: http://people.stat.sc.edu/hoyen/Stat705/Stat705.html.
  Homework and announcements will be posted on the class website and
  it will be updated regularly.

- Text:

  - (KNN) Kutner MH, Nachtsheim CJ, Neter J, and Li W. (2005) Applied Linear Statistical Models. 5th edition, McGroww-Hill/Irwin.

  - (RB) Rosner B. (2010) Fundamentals of Biostatistics. 7th, Duxbury.

  - (HL) Hosmer DV and Lemeshow S. (2014) Applied Logistic Regression. 3th, Wiley and Son.

  - (AG) Agresti A. (2013) Categorical Data Analysis. 3th, Wiley and Son.

- Prerequisite: Prerequisites are successful completion of STAT 704 and STAT 712.

- Course description: Continuation of STAT 704. We will cover categorical and count data analysis in the first part. In the second part of this course, we will study some more advanced topics such as: random/mixed effects models, repeated measures, multiple comparison and false discovery rate.

- Learning Objective: Upon successfully completing this course, students will be able to:

  - Formulate a scientific question about the relationship of a continuous (or discrete) response variable Y and predictor variables X in terms of the appropriate linear (or logistic) regression model.

– Interpret the meaning of regression coefficients in scientific terms as if for a substantive journal.

– Develop graphical and/or tabular displays of the data to display the evidence relevant to model fitting.

– Estimate the regression coefficients using statistical packages such as R or SAS or Stata.

– Propose appropriate statistical models and analysis strategies for the datasets at hand.

– Write a methods and results section for a substantive journal, interpret the results and answer specific scientific questions with analysis results in a way that can be understood by substantive experts.

- Homework: Approximately 5 homework assignments will be announced in class and posted on the course web page. Use the homework template posted on the class website to write homework. There will be a take-home data analysis final project, due the week after the last lecture.

  Each student's homework must be done independently. You may discuss the homework, compare answers, et cetera, but *all submitted homework must be your own work.* **Do not copy any part of another student's work or computer code.**

- Exams: There will be two in-class exams. The exams will help prepare students for the qualifying examination at the end of their first year. Exams may not normally be made up, except in extreme circumstances, for which written documentation of excuse (doctor's note, funeral notice, etc.) is required. If you suspect you may miss an exam day, it is important to contact me well in advance of the test date.

- **Students may not use previous student's homeworks or previous solutions.**

- Grading: The course grade will be based on homework (40%), the final take-home data analysis project (10%), and the two exams (25% each).

- Attendance: Though attendance is never a problem with graduate students, I would like to note that my policy corresponds to the policy stated in the student handbook: If you miss more than 10% of your

classes (> 3 classes), the teacher may choose an appropriate penalty. I will deduct 2% from your final class average for each additional day that you miss after the third absence; excused and unexcused absences both count as absences.

- Computing: Some problems in this course involve significant computations, and for these, we will continue to use the software package R. We will also incorporate analysis using SAS in this course. All code used in class will be available on the course webpage; extensive coding examples are available on previous instructors' websites (Prof. Hitchcock) as well. R is an extremely useful statistical programming language that has become widely used in recent years.

- Disabilities: If you qualify for accommodations because of a disability, please submit a letter to me from the Office of Student Disability Services in a timely manner so that your needs can be addressed. The Office of Student Disability Services determines accommodations based on documented disabilities. Contact: 777-6142, LeConte 112A.

Week	Date	Topic	Homework Due	Readings
1	1/16	Syllabus, Test for Binomial Proportions		RB Ch7, RB Ch10
2	1/23	Contingency Tables, Fisher Exact Test,		RB Ch13
		χ^2 Test, Confounding/effect modification	Homework 1 (1/28)	
3	1/30	Introduction to Logistic Regression		HL Ch1
4	2/6	Statistical Inference of Logistic Regression	Homework 2 (2/11)	HL Ch2, Ch3
5	2/13	Classification using Logistic Regression		HL Ch5
6	2/20	Conditional Logistic Regression		HL Ch7, AG Ch10
7	2/27	Conditional Logistic Regression (Cont.)	Homework 3 (2/25)	
8	3/6	Midterm		
9	3/13	Spring Break: No Class		
10	3/20	Log-linear Regression for Count Data	Homework 4 (3/25)	AG Ch8
11	3/27	Log-linear Regression for Count Data		
12	4/3	Random/mixed effects models	Homework 5 (4/8)	KNN Ch 25
13	4/10	Repeated measures		KNN Ch27, HL Ch7, AG Ch12
14	4/17	Multiple Comparisons, FDR		
15	4/24	Project Presentations		
16	5/6	Final Exam (at 9AM)		

Some Important Questions

1. **Where can I find weekly schedule?** Please refer to the class website for important information such as lecture notes, weekly schedule, syllabus. The class website will be updated frequently.

2. **Can I hand in the homework late?** No. Late homework will lose 30% of total points per day, unless arrangements have been made with the instructor for an extension. Homework will not be accepted after the time at which graded homework are returned.

3. **Must I use R?** No. Use whatever you would like. You will be required to do simulation and programming, for which software like R are particularly well suited. The SAS software is also widely used by statisticians, so is STATA. Programs such as SPSS, Minitab, and Excel, which can be great for many analyses, are insufficient for the need of this class.

4. **What should I use to edit programs?** You should use an editor that has syntax highlighting, and automatic indenting. So, don't se Microsoft Word or Notepad to edit programs. In RStuido, you can open a new R script document. Or in R, under the File tab, open a new script by clicking on "New Document". Or if you're a hacker type, use emacs. SAS comes with its own editor that has auto-ident and syntax hightlighting.

5. **I have an event with dates that conflict with the final exam. Can I take the exam early?** No. I probably won't have the exam written to be taken early.