STAT 530, Applied Multivariate Statistics and Data Mining – Fall 2018

Instructor:
David Hitchcock, associate professor of statistics
209A LeConte College
Phone: 777-5346
Email: `hitchcock@stat.sc.edu`
Course Web Page: `http://people.stat.sc.edu/hitchcock/stat530.html`
                    (Also accessible via Blackboard)

Classes:
Meeting Times: Mon-Wed 2:20-3:35 pm, Wardlaw 116
OR via distance through Adobe Connect/Ensemble or online streaming video

Office Hours: Monday, Wednesday, Friday, 10:50 am - 11:50 am, Tuesday 1:00-2:00 pm, and please feel free to make appointments to see me at other times.

Textbooks:
An R and S-PLUS Companion to Multivariate Analysis (2005), by Brian Everitt.
An Introduction to Statistical Learning with Applications in R (2013), by James, Witten, Hastie, and Tibshirani, available as a free download at
`http://www-bcf.usc.edu/~gareth/ISL/index.html`

Prerequisite: A grade of C or higher in STAT 515 or equivalent. *For this purpose*, courses equivalent to STAT 515 include PSYC 228 or 709; EDRM 710; STAT 205, 509, 512, 700, or 704; MGSC 291, 391 or 692; BIOS 700; ECON 436. See me regarding other questions about prerequisites.

Course Outline: Much of Chapters 1 – 8 of the Everitt textbook, and Chapters 4, 5, 8, 9 of the James et al. textbook. Topics covered include: summary statistics for multivariate data, multivariate data visualization, principal components analysis, exploratory factor analysis, multidimensional scaling and correspondence analysis, cluster analysis, classification and supervised learning, tree-based methods, support vector machines, cross-validation, and (time-permitting) MANOVA, multivariate regression.

Learning Outcomes:
Upon successful completion of this course, students should be able to:
- Determine which multivariate methods are appropriate for a given situation
- Understand the basic logic behind each method's construction
- Verify whether the assumptions needed to implement the methods are satisfied
- Analyze a data set using the methods in an appropriate software package
- Interpret the output for each of the methods

Class Lectures:
You may attend the lectures live on Mondays and Wednesdays in Wardlaw 116, or you may watch them live online via Adobe Connect, or after the fact by viewing the lectures that are posted on the Blackboard STAT 530 course page (click the Course Lectures Playlist link on the left side of the page). Information about how to access online lectures has been emailed to you.

Homework:
Homework problems will be assigned periodically. Each student's homework must be done independently. You may ask each other informal questions about the homework, but everyone is to do his/her own work. If homework is found to be copied, all students involved will receive a 0. Of course, you may always ask me questions about the homework. [To be clearer, students can ask each other informal ORAL questions about homework, but **cannot look at or copy each other's homework papers**. All submitted homework must be their own work.] Note that for the take-home tests, you MAY NOT talk to each other about the problems at all.

Graduate Students: The university requires that 500-level classes be more rigorous for graduate students than for undergraduates. Therefore, any students enrolling in the course for graduate credit will be asked to do some extra problem(s) on some homework assignments.

Disabilities: Any student with a documented disability should contact the Student Disability Resource Center at 777-6142 to make arrangements for appropriate accommodations.

Exams:
There will be one midterm take-home exam (due date to be determined, but it will definitely be in October) and a take-home final exam due Monday, Dec. 10, by 4:00 p.m. The exams may be turned in to me personally, turned in to my office in 209A LeConte, or sent by e-mail in a pre-approved format. **You are not allowed to receive help from *anyone except me* on the exams.** For example, you may not talk to other students about the exam problems, and you may not look at other students' exams. Violations of this policy may result in a 0 on the exam, an F for the course, and/or punishment by the USC Office of Academic Integrity.

Grading:
The course grade will be based on homework average (30%), the midterm exam (35%), and the final exam (35%). The overall course average will result in the following grades: 90-100 = A, 87-89 = B+, 80-86 = B, 77-79 = C+, 70-76 = C, 67-69 = D+, 60-66 = D, 59 and below = F.

Computing:
Use of a computer is required for the analysis of multivariate data. The examples in class will be done using R. No previous knowledge of R is assumed. Everyone is encouraged to download a free copy of R (see the course page for downloading instructions). All necessary analyses will be able to be done in R, at least --- you may use other software if it accomplishes the appropriate task and if you know how to use it.