DEPARTMENT:	BIOS		
COURSE NUMBER:	726	SECTION NUMBER:	1
CREDIT HOURS:	2	SEMESTER: Fall	

COURSE TITLE: Applied Multivariate Analysis

CLASS HOURS AND LOCATION: W 3-4:50

INSTRUCTOR NAME: Mary E. Kelley, PhD

INSTRUCTOR CONTACT INFORMATION

EMAIL: mekelle@emory.edu

PHONE: 404-712-0804

SCHOOL ADDRESS OR MAILBOX LOCATION:

OFFICE HOURS: by appt

Teaching Assistant(s): None

COURSE DESCRIPTION

The course covers multivariate analysis methods including multivariate regression and ANOVA, data reduction through the use of principal components or factor analysis, multivariate classification and clustering methods. Where applicable, nonparametric and/or categorical adaptations of traditional multivariate normal analyses will also be presented.   Links of traditional methods to the concepts of supervised and unsupervised learning will be reviewed. Prerequisites include BIOS 510, 511 or equivalent.

MPH/MSPH FOUNDATIONAL COMPETENCIES:

CONCENTRATION COMPETENCIES:

- Conduct complex statistical analyses for a broad range of applications

COURSE LEARNING OBJECTIVES:

- to understand the uses of multivariate techniques in the analysis of research data;
- The purpose of this course is to give both masters and doctoral level students experience with multivariate techniques in the analysis of research data.

EVALUATION

100% of the grade will be based on homework assignments. There will be 4 assignments, with equal weight. Late homework will result in a grade of 0 unless there are serious extenuating circumstances.

COURSE STRUCTURE

The course will comprise of lectures on the material, interspersed with specific analysis examples from various software platforms. The examples will be mainly in SAS and Stata (factor analysis) with additional examples in R. It is noted that the focus is on the interpretation of the analysis and not the software used as software use is not the topic of the class. It is expected that assigned readings are done prior to the lecture in order for the material to be covered successfully. Following each major section of material (see course outline), a homework assignment using the methods reviewed will be given, and the performance on these assignments will demonstrate the analysis competency.

Required textbooks:
- Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, New Jersey, 2007. ISBN: 978-0-13-187715-3 (JW)
- James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. Springer, New York, 2015 (JWHT)

Recommended access to:
- Khattree R, Naik DN. Applied Multivariate Statistics with SAS Software, 2nd edition. SAS Institute Inc, Cary NC, 1999. ISBN: 1-58025-357-1 (KN1)

Supplemental readings:
- Bartholomew DJ, Steele F, Moustaki I, Galbraith JI. The Analysis and Interpretation of Multivariate Data for Social Scientists. CRC Press, Boca Raton, 2002 (BSMG)
- Zhang H, Singer BH. Recursive Partitioning and Applications, 2nd edition. Springer, New York, 2010 (ZS).
- Kaufman L, Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, Hoboken, NJ, 2005 (KR)

COURSE POLICIES

I do not allow the use of cell phones, tablets or laptops in class, as they interfere with the learning objective. All materials will be provided and notes should be taken when necessary.

RSPH POLICIES

Accessibility and Accommodations

Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact the Office of Accessibility Services (OAS). It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed.

Students who registered with OAS and have a letter outlining their academic accommodations are strongly encouraged to coordinate a meeting time with me to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible.

Contact Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html

Honor Code

You are bound by Emory University's Student Honor and Conduct Code. RSPH requires that all material submitted by a student fulfilling his or her academic course of study must be the original work of the student. Violations of academic honor include any action by a student indicating dishonesty or a lack of integrity in academic ethics. *Academic dishonesty refers to cheating, plagiarizing, assisting other students without*

authorization, lying, tampering, or stealing in performing any academic work, and will not be tolerated under any circumstances.

The RSPH Honor Code states: "Plagiarism is the act of presenting as one's own work the expression, words, or ideas of another person whether published or unpublished (including the work of another student). A writer's work should be regarded as his/her own property."
(http://www.sph.emory.edu/cms/current_students/enrollment_services/honor_code.html)

COURSE CALENDAR

Dates are subject to change given class flow

All HWs are due 2 weeks after date of assignment

8/29	Introduction
	HW 1 assigned due 9/12
9/5	Multivariate regression and MANOVA
9/12, 9/19	Repeated measures as multivariate
	HW2 assigned due 10/3
9/26,10/3	Principal component analysis
10/10	Factor analysis MVN
10/17	Ordinal and binary FA
10/24	Path analysis and SEM
	HW3 assigned due 11/7
10/31,11/7	Discriminant analysis and classification
11/14	Clustering for interval scaled variables hierarchical and non-hierarchical(1)
11/21	Clustering for categorical and mixed data types
11/28	Model based clustering (mixture models)
	HW 4 assigned due 12/12
12/5	Mixture models (2)

COURSE OUTLINE

1) Introduction to multivariate data
   a) Data structure and the multivariate problem (JW 1.3)
   b) Measures of distance (JW 1.5)
   c) Multivariate estimation (JW, 4.1-4,4, 5.2,5.3,6.1-6.3,6.6, KN1 chpt 1)
   d) Matrix Algebra review (length and topics depend on class composition) (JW, chpt
      HW1
2) Multivariate regression and MANOVA (JW 6.4 – 6.10, 7.7; KN1, chpts 3-4)
3) Repeated measures as a multivariate problem (JW 6.1-6.2, 6.9, KN1 chpt 5,6)
   HW2
4) Data reduction and latent variable analysis
   a) Principal components analysis (PCA), partial least squares (PLS), and singular
      value decomposition (SVD) (JW Chpt 8; BMSG Chpt 5) (JWHT, 6.3, 10.2,10.4)
   b) Factor analysis (FA)
      i) MVN based inference (JW chpt 9; BSMG chpt 7)
      ii) Ordinal and binary FA (BSMG chpt 8, 9)
   c) Path analysis, confirmatory factor analysis (CFA) and structural equation
      modeling (SEM) (BSMG chpt 11)
      HW3
5) Classification (supervised learning)
   a) Linear Discriminant Analysis (JW chpt 11) (JWHT chpt 4)
   b) Nonparametric classification (kernel, KNN, CART) (JW 11.8, JWHT chpt 8; ZS
      chpt 2,4)
6) Clustering (unsupervised learning) (JW chpt 12; KN2 chpt 6, JWHT 10.3, 10.5)
   a) Interval scaled variables
      i) hierarchical (MVN, agglomerative)
      ii) non-hierarchical (k-means)
   b) Clustering methods for categorical and mixed variable types (KR)
   c) Model based methods (mixture models, JW 12.5)
      i) Latent class analysis
      ii) Latent profile analysis
      iii) Latent trajectory analysis
      HW4