Statistics 2221: Advanced Applied Multivariate Analysis

Spring, 2024

Instructor: Jason Fine
Room 1822 Posvar Hall
jaf367@pitt.edu

TA/Grader: Fan Yang
1825 Posvar Hall
ffy1@pitt.edu

Lectures: Regularly scheduled class times are Tuesday/Thursday, 9:30 to 10:45 am, in Cathedral of Learning 216. Lecture notes will be made available to students in Canvas. Handouts from texts will also be provided.

Office Hours: Tue/Thu after class, 10:45-12. Fri, 11-12 am, and by appointment.

Overview: Important statistical methods and rlevant theory for analyzing continuous multivariate data are introduced. The first half of the courses examines traditional and fundamental topics in some depth, and the second half of the course surveys modern topics. There is no formal prerequisite for this course. However, it is assumed that students have a working knowledge of probability, statistics, and matrix algebra, as well as applied experience with statistical methodology, including descriptive statistics, hypothesis testing and confidence intervals, regression modelling, including linear and generalized linear models. An ability to conduct data analyses in statistical software will be essential (R, SAS, Stata).

Material to (Potentially) be Covered (readings from texts in parentheses)

1. Multivaraite Data Exploration (HS, Ch 1, Izenman, Ch 4)
2. Matrix Algebra Review (HS, Ch 2, Izenman Sec 3.2)
3. Multivariate Normal Distribution (HS, Ch 4, 5, Izenman Sec 3.3)
4. Wishart Distribution (HS Ch 4,5, Izenman Sec 3.4)
5. Inference for Multivariate Normal (HS, Ch 6,7, Izenman Sec 3.5)
6. Linear Dimension Reduction: Principle Compoments (PCA) and Canonical Correlations (CCA) (HS Ch 9, 10, Izenman Ch 7)
7. Classification (HS, Ch 13, Izenman Ch 8)
8. Latent Variable Models (HS, Ch 11, Izenman Sec 15.4)
9. Clustering (HS, Ch 12, Izenman Ch 12)
10. Multidimensional Scaling (HS, Ch 16, Izenman Ch 13)
11. Classification and Regression Trees (CART) (HS Sec 19.5, Izenman Ch 9)

Computing: There is no designated software for this course. You may use the software that makes the most sense for you. Many pharmaceutical companies use SAS for compliance with FDA regulations. Academic intuitions as well as labs often use R and python. Corporations often use MATLAB, Stata, Minitab, S, etc. because of the relatively high reliability despite the cost. However, it is expected that the student immerse herself with use of at least one software. Reference manuals, for example SAS documentation (Ravindra and Naik, 2019, 2020) may prove invaluable for data analysis exercises and projects.

Texts : Hardle and Simar (2012) and Izenman (2008) are primary resources, with other books listed below being useful. You can access the books electronically or via hard copy from the Pitt library system, or download the books as .pdf files or buy hard copies directly from the publisher or off Amazon

1. Applied multivariate statistical analysis, W. Hardle and L. Simar. Springer, 2012. (elementary)

2, Modern multivariate statistical techniques: regression classification and manifoold learning. A. Izenman, 2008. (advanced applied)

3, Applied multivariate statistical analysis, R.A. Johnson and D. Wichern. Pearson Prentice Hall, 2007. (classic elementary)

4, Elements of statistical learning: data mining inference and prediction, T. Hastie, R. Tibshirani, J. Friedman. Springer 2009. (modern applied)

5, An introduction to applied multivariate analysis with R. B. Everitt and T. Hothorn. New York: Springer, 2011 (computing)

Course Requirements and Grading: Regular homework will be assigned, including both theoretical and applied examples (55%). There will be a mid term exam (25%), with the data to be determined. There will be a final project (20%), with details to provided later in the semester. Lecture attendance and participation is encouraged (5%). Homework assignments will be graded by the TA and returned to the students in a timely manner. Any questions regarding grading on the HWs should first be directed to the TA. The midterm and final will be evaluated by the course instructor.