

STA 306: Applied Multivariate Analysis

I. Basic Course Information

Applied Multivariate Analysis is an upper-level statistics option, both for Data Science and Statistics majors and Statistics minors. The course meets twice weekly for 80-minute periods. The course is suitable for students from many disciplines. Prerequisites: STA 215 or STA 216 or (ECO 105 and (MAT 125 or MAT 127)) or (STA145 and MAT 127).

II. Learning Goals

The American Statistical Association's guidelines for undergraduate programs in statistical science state that such programs should "emphasize concepts and approaches for working with complex data and provide experiences in designing studies and analyzing non-textbook data." More specifically, they recommend that programs should provide statistical topics that include random sampling, stratification in sample surveys, data exploration in observational studies, and a variety of formal inference procedures. They also recommend that programs should require familiarity with a standard statistical software package.

This course covers aspects of statistical inference in cases where there are multiple variables. In contrast to Regression Analysis, the methods in multivariate analysis focus on areas of unsupervised data exploration and classification. It provides an introduction to methods such as clustering and principle component analysis for exploration and linear discriminant analysis for classification. The rationale underlying each test will be conveyed and an industry relevant statistical software package will be used as decided by the individual instructor.

While topics may evolve with the field, on completion students will typically have covered

- (i) Matrices, eigenvalues and eigenvectors
- (ii) Principal components analysis
- (iii) Factor analysis
- (iv) Discriminant analysis (two and three group analyses)
- (v) Cluster analysis, using a variety of metrics and clustering algorithms
- (vi) Multi-dimensional scaling (classical and ordinal)
- (vii) Correspondence analysis
- (viii) Canonical correlation analysis
- (ix) Tests of significance for multivariate data

III. Student Assessment

Students will receive regular feedback on their work through the assignment of homework, quizzes, student presentations and examinations. Through this feedback, students will be able to see and correct their misunderstandings and improve their performance. Student performance on these assessment instruments and the performance of students in their statistics options will be used to assess the success of Applied Multivariate Analysis in achieving its learning goals and its contribution to the fulfillment of the Data Science and Statistics program goals. Peer reviews and student evaluations will also be used to evaluate the course.

IV. Learning Activities

Learning activities will consist of a combination of lectures, discussions, student presentations, and use of an industry statistical software package to analyze data. The specific choice will depend upon the individual instructor. Outside of class, students are expected to do a significant amount of individual and group work to achieve the learning goals, including multiple lengthy problem sets that analyze real world data and involve significant use of computer analysis. Students will receive detailed feedback in order to improve their analysis skills and application of analysis methods. Students are encouraged to work together to more fully develop a deep understanding of multivariate statistics. The ability to do many tasks on laptop computers, including working remotely on the ELSA HPC cluster, will permit students to work as a group in the library in addition to in the School of Science computer laboratories.

Approved 4-9-25