

**Instructor:** Dr. Amanda S. Hering

E-mail: [ahering@mines.edu](mailto:ahering@mines.edu)

Phone: 303.384.2462

Office: 235 Chauvenet Hall

Office Hours: Tuesday 9-10am, Wednesday 2-3pm, Friday 12-1pm

**Prereq:** Undergraduates: MATH 335 (Introduction to Mathematical Statistics) or MATH 323 (Probability and Statistics for Engineers).

Graduate Students: MATH 530 (Applied Statistical Methods I)

In particular, **it assumed that you have an understanding of and a working knowledge of linear algebra**, so having had a course like MATH 332 (Linear Algebra) or MATH 500 (Linear Vector Spaces) will be very beneficial.

**Course Schedule:** MWF 10-10:50 am, Location Green Center 265

**Web Page:** The username and password for the website will be given in class.

<http://inside.mines.edu/~ahering/math437/multivariate.html>

**Course Description:** The body of methodology used when the data includes simultaneous measurements on many variables is called *multivariate analysis*. Common scientific objectives for which multivariate methods naturally lend themselves are the following:

1. *Data reduction or structural simplification.* The phenomenon being studied is represented as simply as possible without sacrificing valuable information, hopefully making interpretation easier.
2. *Sorting and grouping.* Groups of “similar” objects or variables are created based on measured characteristics. Alternatively, rules for classifying objects into well-defined groups may be required.
3. *Investigation of dependence among variables.* The nature of the relationship among variables is of interest. Are all the variables mutually independent or are one or more variables dependent on the others? If so, how?
4. *Prediction.* Relationships between variables must be determined for the purpose of predicting the values of one or more variables on the basis of observations on the other variables.
5. *Hypothesis construction and testing.* Specific statistical hypotheses, formulated in terms of the parameters of multivariate populations, are tested. This may be done to validate assumptions or to reinforce prior convictions.

This is primarily designed to be an applied course. Matrix calculus will not be required; however, some algebra is required to understand the conceptual development of different methods. Multivariate techniques are applied to a wide array of disciplines, such as business, health, medicine, economics, education, ecology, geology, sociology, energy, atmospheric sciences, law enforcement, and defense. In many cases, a multivariate method is used as one component of another overarching application, so they appear frequently in the literature.

Many academic disciplines use software designed to meet their specific needs, but in this course, students will learn how to use a free programming language (called **R**, which is closely related to **S-PLUS**) to develop the tools needed to properly analyze multivariate data. Learning to program in **R** gives students the power to double-check their discipline-specific software, and it also gives them the flexibility to analyze data with methods that are not automatically provided by their software.

**Course Objectives:** By the end of the course, the student should be able to:

1. Understand the major types of multivariate statistical analyses and how they differ from each other.
2. Apply the concepts of multivariate statistics to real datasets.
3. Use the **R** software (or other software package of your choosing) to perform multivariate analysis of real data sets.

**Course Outline:** The following is a list of topics that will be covered in this course:

- Review of basic vector and matrix algebra
- Random vectors
- Multivariate distributions—normal, Wishart, Hotelling's-T, Skew-T, Skew-Normal
- Estimation and testing of multivariate distribution parameters
- Principal component analysis (PCA)
- Canonical correlation analysis (CCA)
- Discriminant and cluster analysis
- Multidimensional scaling (MDS)
- Tentative topics: MANOVA (comparison of multiple multivariate means), multivariate linear regression, factor analysis, and facet theory.

**Textbook: (Required)** Johnson, R. A. and Wichern, D. W. (2007) *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River, NJ: Prentice Hall.

**Course Work:** Your grade for the course will be based on the following (relative weights given in percentage):

- **Homework Assignments (30%):** Bi-weekly homework assignments will be given throughout the semester. Assignments will be collected at the START of class on the due date. Late assignments will not be accepted.
- **Exam (30%):** There will be one in-class closed book exam tentatively scheduled for Wednesday, April 4th.
- **Project (40%):** A project will be due at the end of the semester, and a poster session to present the work will be held in place of the class' 2-hour final exam. More details will be given as the semester progresses. Late projects will not be accepted.

The following letter grades are guaranteed:

A	B	C	D	F
100-90%	89-80%	79-70%	69-60%	59-0%

**Other References:** This is a list of other commonly used texts in multivariate statistics.

Anderson, T. W. (1984) *An Introduction to Multivariate Statistical Analysis*. Wiley.

\*Borg, I. and Groenen, P. J. F. ( ) *Modern Multidimensional Scaling*. New York, NY: Springer.

Chatfield, C. and Collins, A. J. (1980) *Introduction to Multivariate Analysis*. Cambridge, UK: University Press.

\*Izenman, A. J. (2008) *Modern Multivariate Statistical Techniques*. New York, NY: Springer.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979) *Multivariate Analysis*. London, UK: Academic Press.

Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*. Wiley.

\*Denotes texts available through CSM's library access to Springer via <http://www.springerlink.com>.

**Notes:** A few more things...

- Check the website frequently for updates.
- I would like to know about any particular academic difficulties or personal problems that are affecting a student's performance.