

**STATISTICS 505**  
**Applied Multivariate Statistical Analysis**  
**Fall 2009**  
**Course Syllabus**

- PREREQUISITES:** STAT 500-501 (applied statistics courses)  
or equivalent and matrix algebra  
or permission of the instructor.
- CREDITS:** 3
- FORMAT:** Lecture
- HOURS:** MF, 2:30 - 3:20 pm, 117 Thomas  
W, 112 Boucke
- INSTRUCTOR:** Debashis Ghosh  
Dept. of Statistics  
Address: 514A Wartik Lab  
Phone: 863-9886  
Email: ghoshd@psu.edu
- OFFICE HOURS:** M, 4-5 pm; Th, 1-2 pm  
514A Wartik  
or by appointment
- CLASS WEBPAGE:** available through ANGEL; check  
regularly for updated announcements
- HOMEWORK:** Given out weekly Wednesdays, due the following Wednesday,  
with exceptions for holidays. Late homework assignments  
will not be accepted, unless the student obtains the permission  
of the instructor. You may work on the problem sets in groups;  
however, you must write up the assignment on your own.  
Plagiarism on any assignments will be dealt with as per  
the policy of the Eberly College of Science on Academic Integrity  
(modification of Senate Policy 49-20), available at  
  
<http://www.science.psu.edu/academic/Integrity/Policy.htm>
- TEXT (Recommended):** Johnson, R.A., and Wichern, D.W. (2007).  
Applied Multivariate Statistical Analysis. 6th ed.  
Prentice Hall, New York.

**EVALUATION:** Homework assignments: 20%  
Exam 1, October 2 (in class, 2:30 - 3:20 pm) : 25%  
Exam 2, November 13 (in class, 2:30 - 3:20 pm): 25%  
Final Exam (date TBA): 30%

**COMPUTER:** Most of the examples we talk about in class have been implemented in the software package R (<http://www.r-project.org/>), as well as in the SAS package. For the homework assignments, I will ask you to analyze datasets. It will be very important to have R available on either a work or home computer. If you have never used R, the class webpage has available links for R (available in the Resources section – Course Resources, Other Tools). If you have other software packages available from your work or research that can implement the proposed methods, then that will be fine as well.

### Learning Objectives

Upon entering this course, you are expected to have completed six credit hours in applied statistical methods as well as have some knowledge of matrix algebra. You should know how to fit basic descriptive statistics to datasets and to have an understanding of hypothesis testing and confidence intervals, either through coursework or practical experience.

After completing the course, the student can ordinarily expect to be able to:

1. Identify and understand the structure of multivariate data and be able to phrase the appropriate scientific questions in terms of parameters of interest.
2. Understand the various assumptions needed for the various methodologies covered in the class as well as their implementation.
3. Implement analyses of these methods in a statistical software package.
4. Read the scientific literature and comprehend the use (and misuse) of multivariate analysis methodologies reported by study authors.

### ACADEMIC INTEGRITY:

The conduct of a student registered or taking this course should be consistent with that

of a professional person. Courtesy, honesty and respect should be shown by students toward faculty members, guest lecturers, administrative support staff and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions and striving to help them achieve maximum benefits from their experience in the course.

Student academic misconduct refers to behavior that may include plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials (including library materials), and aiding and abetting the perpetration of such acts. The preparation of reports, papers, and examinations, assigned on an individual basis, must represent each student's own effort. Reference sources should be indicated clearly. The use of assistance from other students or aids of any kind during a written examination, except when the use of aids such as electronic devices, books or notes has been approved by an instructor, is a violation of the standard of academic conduct. All violations of academic integrity will be assessed and dealt with based on the policy of the Eberly College of Science on Academic Integrity, available for reading at

<http://www.science.psu.edu/academic/Integrity/Policy.htm>

## **Course Topics**

1. Review of basic matrix operations
2. Numerical and graphical summaries of multivariate data
3. Multivariate normal distribution
4. Simultaneous estimation of means for multivariate data
5. Comparison of two or more mean vectors
6. Principal Components
7. Factor Analysis
8. Discrimination and Classification
9. Cluster Analysis and Data Mining Methods
10. (If time permits) Canonical Correlation: Methods for analyzing the linear relationships between a vector of dependent variables and a vector of explanatory variables