

SYLLABUS FALL 2007
BINF 5354 Post-Genomic Analysis & BINF 5113 Seminar
Multivariate Data Analysis and Probabilistic Modeling
T 9:00 - 11:20am, R 9:00am - 12:20pm Bell Hall 143

INSTRUCTORS: Naijun Sha and Ming-Ying Leung

OFFICE: Bell Hall 203 (Sha) and 225 (Leung)

PHONE: 747-6844 (Sha) and 747-6836 (Leung)

E-MAIL: nsha@utep.edu; mleung@utep.edu

WEBPAGE: <http://www.math.utep.edu/Faculty/naijun/sha.html>
<http://www.math.utep.edu/Faculty/mleung/index.html>

OFFICE HOURS: MW: 1:00-2:00pm (Sha) and T: 11:30am-12:30pm (Leung) or by appointment

DEADLINE for Automatic W: Nov. 2

Text: *Applied Multivariate Statistical Analysis, 5th edition* by R. A. Johnson and D. W. Wichern

Course Description: Statistical analysis of a multivariate response. Topics covered: Descriptive multivariate statistics, multivariate normal distribution, multivariate multiple linear regression, principal component analysis, classification and clustering analysis. Applications with the use of statistical packages will be considered. Prerequisite: STAT 5380, STAT 5385 or equivalent, or consent of instructor.

Probabilistic modeling for nucleic and amino acid sequences. Topics covered: Markov chains and Hidden Markov Models (HMM). Probabilistic approaches to sequence alignment, phylogeny, and RNA structure analysis. Prerequisite: Instructor approval.

Assignment and Attendance: Homework and data analysis projects will be assigned throughout the semester. You may do the exercises with your partner, but each student must answer the questions individually. ZERO grades will be got for those whose solutions are the exact copies of someone else. Class attendance is required and helpful to decide borderline grades.

NO LATE HOMEWORKS WILL BE ACCEPTED!

Grading: Homework	30%
Midterm Exam	30%
Final/Project	40%

The final grade is based on a scale of $90 - 100 = A$, $80 - 89 = B$, $70 - 79 = C$, $60 - 69 = D$, below $60 = F$. Attendance and class participation will be used to help decide borderline grades.

References:

1. *An Introduction to Multivariate Statistical Analysis* by T. W. Anderson. Wiley.
2. *Methods of Multivariate Analysis* by Alvin C. Rencher. Wiley.
3. *Multivariate Data Analysis* by J. F. Hair, R. L. Tatham, R. E. Anderson and W. Black.
4. *Methods of Microarray Data Analysis* eds by J. Showmaker and S. Lin. Wiley.
5. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids* by R. Durbin, S. Eddy, A. Krogh and G. Mitchison.