# ST437/537: Applied Multivariate and Longitudinal Data Analysis

Spring 2018

**Course:**

*Meeting time/place*: Tu and Th 10:15AM - 11:30AM, 1216 SAS Hall
*Webpage*: https://st537s18.wordpress.ncsu.edu/home/

**Instructor:** Ana-Maria Staicu
*Office*: 5220 SAS Hall
*Email:* astaicu@ncsu.edu
*Webpage*: http://www4.stat.ncsu.edu/~staicu/
*Office Hours:* Tuesday, 11:30-12:30 and 1:10 - 2:10PM

**Teaching Assistant:** Rui Zhu
Office: 1101 SAS Hall
*Email*: rzhu4@ncsu.edu
*Office Hours*: Monday 1:30-3PM and Wednesday 1:30-3PM

**Prerequisites:**

ST 422, ST 430 (Intro to Regression Analysis)
ST 512 or ST 514 or ST 515 or ST 517

**Textbooks**

1. Applied Multivariate Statistics with R by Daniel Zelterman. (2015) Springer (free download through the university library)
2. Applied Longitudinal Analysis by Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004) . New York: Wiley

*Optional:* Applied Multivariate Statistical Analysis. by Johnson, R. and Wichern, D. Pearson Prentice Hall

**Computing:** The primary computing language will be R, which is freely available at http://www.r-project.org/.

**Grading:** Your final (letter) grade is a conversion from a numerical score (0-100%) which is calculated based on the following scheme:

40% Homework
20% Midterm 1 (covering the part on multivariate statistics, Sections 3-5)
20% Midterm 2 (covering the part on longitudinal analysis, Sections 6-8)
20% Project (group project to be presented at the end of the course)

**Midterms:** are held during the regular meeting time/location. The midterms are tentatively scheduled for March 1 and April 19. The midterm exams will be open book, open notes (no computer access), and may include a take-home portion.
**Project:** Proposal due on April 3. Presentation as poster will be on April 26. Any conflicts with the scheduled exam dates must be submitted in writing to the instructor well in advance. Unexcused missing exams, or inadequate notice of missing an exam will result in a grade of 0 for the exam.

**Homework:** There will be roughly weekly homework assignments. Problems and due dates will be posted on the course webpage. Unexcused late homework will be discounted by 50%.

**Course objectives:**

1. Distinguish between multivariate data and longitudinal data and adequately suggest methods for analysis
2. Compare multivariate techniques and longitudinal methods and identify their advantages/limitations
3. Use computer code ( R ) to carry out analysis of multivariate and longitudinal data
4. Correctly interpret the computer output in these situations

**Policy on Academic Integrity:** The University policy on academic integrity is spelled out in Appendix L of the NCSU Code of Student Conduct. For a more details see the NCSU Office of Student Conduct website http://www.ncsu.edu/student_conduct/. For this course group work on homework is encouraged. However, copying someone else's work and calling them your own is plagiarism, so the work you turn in should be your own.

**Students with Disabilities:** Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with Disability Services for Students (DSS), 1900 Student Health Center, CB 7509, 515-7653.

**Lecture Topics** (Tentative schedule/plan of lecture topics)

1. **Introduction**
    a. What are multivariate/longitudinal data? What to expect from this course.
    b. Basic algebra [individual study, HWa]
    c. Univariate Normal distribution in R. Maximum likelihood estimation (MLE).
    d. Common univariate distributions [HWb]
    e. Bivariate Normal and multivariate Normal distributions in R. Their properties (see ch 7.1 book)

2. **Graphical displays of data in R**
    a. Univariate data
    b. Bivariate data
    c. Multivariate data [HW]

Part I: Multivariate Data Analysis

3. **Inference**
    a. Multivariate normal density. Likelihood function. Maximum likelihood estimation.
    b. Inference about a single mean vector (Hoteling's T2, Likelihood ratio test, simultaneous confidence intervals)
    c. Inference about two or more mean vectors (paired comparison, repeated measured design, MANOVA, profile analysis) [HW]
    d. Inference beyond the mean vectors. Tests for the covariance matrix [covariance pattern/totally un-structured]. Tests for multivariate normality [time permitted]

4. **Factor methods [relationship between variables]**
    a. Principal component analysis[HW]
    b. Factor analysis [HW]

5. **Classification and clustering**
    a. Discrimination and classification [HW]
    b. Clustering [time permitted?]

Part II: Longitudinal Data Analysis

6. **General Linear model**

a. Parametric mean model [polynomial mean, spline mean, smooth mean where penalties are used to shrink the parameters]
b. Parametric models for covariance [known covariance pattern]
c. Estimation and inference by maximum likelihood [what is it new from multivariate].
d. Restricted maximum likelihood estimation [HW]

**7. Linear mixed model**

a. Growth curves
b. Random coefficient model. Estimation of mean and covariance parameters
c. Linear mixed effects model. Estimation, inference, and prediction[HW]

**8. Generalized linear model**

a. Marginal models (generalized estimating equations). Estimation and inference of regression parameters
b. Generalized linear mixed models. Estimation and inference[HW]