SYLLABUS: Spring, 2014
STATISTICS 689: Large-Scale Inference & Covariance Estimation

INSTRUCTOR: Mohsen Pourahmadi
TIME & PLACE: MW 5-6:15 pm, Blocker 411 (Subject to change)
OFFICE HOURS: MW 4-5 pm or by Appointment
OFFICE: 439 Blocker
PH. & FAX NOs: 979-845-3164, 979-845-3144
E-MAIL ADDRESS: pourahm@stat.tamu.edu

TEXTS:
1. (Required) Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction, B. Efron, 2010, IMS Monographs.
2. (Recommended) High-Dimensional Covariance Estimation, M. Pourahmadi, 2013, Wiley.

PREREQUISITE: STAT 612 and/or STAT 613.

PLAN OF THE COURSE:
STAT 689 is a course for graduate students in statistics and other fields who seek an advanced background in and deeper understanding of the theory and methods of high-dimensional data analysis. According to Efron, author of the text for the course, "we live in a new age for statistical inference, where modern scientific technology such as microarrays and fMRI machines routinely produce thousands and sometimes millions of parallel data sets, each with its own estimation or testing problem. Doing thousands of problems at once involves more than repeated applications of classical methods".

The modern data matrices are relatively large with correlated rows and columns. **Various ways of modeling and assessing the impacts of dependence/correlation on inference is a major goal of the course.** The classical least square estimation and multiple comparisons are known to fail in the high-dimensional framework, however, we discuss how to revive these by taking advantage of *sparsity* in the data/model and introduce the ideas of penalized likelihood/least squares and false discovery rates (FDR). Mildly Bayesian ideas and techniques will be used to motivate and develop applicable frameworks for estimation, testing and prediction.

The **course project** is extremely helpful in creating a balance between the theory and applications to biomedical sciences or areas of interest to students in the course. In the first two weeks, students will develop the frame for a project and identify the relevant research papers and a dataset to study, analyze and present at various times in the course, see item 2 below for more details. There is a wealth of genuine datasets and R programs in the first text.

GRADE POLICY:

1. **Exams**: There is a midterm and no final. The midterm will constitute 40% of the course grade.

2. **Course Project**: Will involve a significant amount of data analysis, reading recent research papers, computational effort and discussion. There will be bi-weekly project progress reports and presentations in class by students. The project starts by choosing a suitable area, papers and a dataset, the first written report could be a page long describing the data and application area, and the first oral presentation will be about five minutes long describing the data, etc. Subsequent project presentations may take 15 and 30 minutes. The project is worth 40% of your grade. The project reports should be organized and typed following the format of a research article in statistics or your area of applications. They should contain your names, have a title, abstract, objectives,..., references. The quality of writing and presentation in class will contribute greatly to the grade for this part of the course.

3. **Homework**: Will be assigned regularly and posted on DoStat (Reference and Registration codes are : DS- and), it will contribute 20% to your grade. The quality of writing and logical presentation of the arguments leading to a result, not just the correct answer, will contribute greatly to the grade for this part of the course. You may consult with other students about the homework, but always write up your solutions by yourself. You should never just copy from another person. **Do not include R programs and computer printouts in your HW and project reports, unless asked to do so.**

4. The final course grade will be based on the standard scale where a total of 90 to 100 percent will be an A, 80 to 89 percent will be a B, etc.

5. Attendance and **classroom participation** are encouraged and will be rewarded, they are integral parts of the learning process .

6. ACADEMIC INTEGRITY STATEMENT: "An Aggie does not lie, cheat, or steal or tolerate those who do." The Aggie Honor Council Rules and Procedures are available at http://www.tamu.edu/aggiehonor.

7. STATEMENT ON PLAGIARISM: As commonly defined, plagiarism consists of passing off as one's own ideas, words, writing, etc., which belong to another. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turn it in as your own, even if you should have the permission of that person. Plagiarism is one of the worst academic sins, for the plagiarist destroys the trust among colleagues without which research cannot be safely communicated. If you have any questions regarding plagiarism, please consult the latest issue of the Texas A&M University Student Rules, under the section "Scholastic Dishonesty."

8. STATEMENT ON DISABILITIES: The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation for their disabilities. If you believe you have a disability requiring an accommodation, please contact the Office of Disability Services in Room B118 of Cain Hall. The phone number is 845-1637.