STAT GR5703 Statistical Inference and Modeling
Sections 001 Syllabus – Fall 2017

Version: June 1, 2017. Syllabus is subject to change. Make sure you have the latest version. You can always find the most current version on CourseWorks.

Instructor: Dr. Banu Baydil
E-mail: bb2717@columbia.edu
Lectures: Attendance is mandatory. For lecture times and location see
http://www.columbia.edu/cu/bulletin/uwb/sel/STAT_Fall2017.html

Instructor office hours &location: TBD
Course TA, office hours &location: TBD

Required textbooks:
Probability and Statistics, Fourth Edition, M. H. DeGroot and M. J. Schervish
Publisher: Pearson (2012), (ISBN-13) 978-0321500465.
Chapters 7-9, 11, 10 + (time permitting) additional selected topics in statistical modeling. Note: Many suggested problems will be drawn from this text, so the correct edition is required.
+ Material covered in class (which might not necessarily be in the slides posted or in the course textbook).

For students not familiar with elementary statistical inference at the introductory undergraduate level, as a supplement to the above course textbook:
Probability and Statistics for Engineering and the Sciences. 9th Edition, Jay L. Devore
Publisher: Cengage Learning. ISBN: 1305251806. Chapters 5.3-15.
A copy of the 9th edition is made available on Mathematics Library reserve.

Prerequisites: This course assumes that you have a working knowledge of Calculus (single and multi-variable), linear algebra (vectors and matrices), and STAT GU4203/GR5203 or the equivalent.

Course description: In this course, we will systematically cover fundamentals of statistical inference and testing, and give an introduction to statistical modeling. The first half of the course will be focused on inference and testing, covering topics such as maximum likelihood estimates, hypothesis testing, likelihood ratio test, Bayesian inference, etc. The second half of the course will provide introduction to statistical modeling via introductory lectures on linear regression models, generalized linear regression models, nonparametric regression, and statistical computing. Throughout the course, real-data examples will be used in lecture discussion and homework problems. This course lays the foundation, preparing the MA in Data Science students, for other courses in machine learning, data mining and visualization.

Assessment measures:
1) Exams**:** Two exams (including the final) will be administered. The exams will cover material from the beginning of the semester. All material covered will have equal weight in the final exam. Some of the questions in the exams will be chosen from the exercises in the book, in suggested problems and in worksheets. Exams are scored out of 100 points. Midterm exam will count towards 35% and final exam will cover towards 45% of your final grade.

   Exam Dates and location: TBD

   Make-up policy for all the exams: : If you have a valid document such as a doctor's report for missing the midterm exam or an official excuse from the school, your final exam score will also count for the missed midterm exam.

Taking the final exam is mandatory, and failing to do so, might lead to a failing grade.

2) Homework: There will be assigned suggested problems, however they will not be collected. Solutions will be posted periodically.

3) Attendance: **Attendance is required** for the lectures. From time to time, there might be unannounced/announced in-class-work/quizzes, attendance might be taken during the lectures, and you might be asked to complete JiTT's (Just-in Time Teaching activities, each one of which is likely to take around 15-20 minutes) before or during the lectures. The total of all these participation activities will make 20% of your grade. There will be no make-ups or excused absences for missing in-class-work/quizzes/attendance or JiTT's. However, you can miss one lecture due to unexpected circumstances without penalty. Students with documented disabilities should make arrangements with ODS to take their quizzes at ODS after a quiz is administered in class.

4) Final Letter Grade: Catalog ranges will be used. The instructor reserves the right to change the cut-offs for grade distribution based on the overall average of the class.

Use of technology in the course: CourseWorks online course system
Class announcements/e-mails will be made/sent in CourseWorks. You are expected to check CourseWorks course page regularly. A copy of the most recently updated syllabus will be on CourseWorks. Occasionally, there will be other course related handouts posted in CourseWorks. Lecture slides will be posted on Courseworks, however, blackboard lectures will not be posted on Courseworks. You are responsible making sure CourseWorks announcements/e-mails are going to an e-mail you check at least once daily.

Students with disabilities: In order to receive disability accommodations, students must first be registered with Disability Services (DS). More information on the DS registration process is available online at
http://health.columbia.edu/disability-services
Registered students must contact DS to arrange accommodations for this course, including exam accommodations. Students should bring an accommodation letter for signature to the professor for this course to inform the professor of the types of accommodations they will be needing during the course.
Students who have, or think they may have, a disability are invited to contact DS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at disability@columbia.edu.

Academic dishonesty: Cheating in any form is unacceptable. Standard school policies will be enforced in the case any student is caught cheating. In addition, if you get caught cheating during an exam, you get a score of zero from that exam and are strongly encouraged to withdraw from the course. You are encouraged to check The Columbia University Undergraduate Guide to Academic Integrity at
https://www.college.columbia.edu/academics/academicintegrity

Tentative Schedule:

Weeks 1-2	Introduction to inference; prior, posterior, conjugate prior distributions; Bayes, maximum likelihood, method of moments estimators; invariance, consistency
Weeks 1-2	Sufficiency, factorization theorem, improving an estimator
Weeks 3-4	Sampling distributions of estimators, confidence intervals, unbiasedness, Fisher information, efficiency, asymptotic distributions of estimators
Weeks 3-4	Hypothesis testing
Weeks 5-6	Linear statistical models
Weeks 5-6	Nonparametric methods; model selection (time permitting)