## STAT GR5701 Probability and Statistics for Data Science
## Sections 001 Syllabus – Fall 2017

Version: September 7, 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 Imperative: Use the subject STAT GR5701 in all your e-mails, otherwise delays in response/no response is possible. (E-mail is monitored Monday through Friday 7:30am - 8:30pm and is answered on a first come first serve basis.)

E-mails with questions whose answers can be found in the syllabus or in the sent CourseWorks announcements will not be answered individually. If multiple students ask the same question, an announcement through Courseworks will be sent in place of individual responses.

**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:** TTh 8:55pm-9:30pm, course lecture room.

**Course TA, office hours &location:**
Pratyay Datta, pd2511@columbia.edu
ThF 10:30am-12:30pm, 1019 SSW (School of Social Work)
SSW is on Amsterdam Avenue, between 121st and 122nd streets.

**Required textbooks:**
Probability and Statistics, Fourth Edition, M. H. DeGroot and M. J. Schervish
Publisher: Pearson (2012), (ISBN-13) 978-0321500465.
Chapters 1-9, 11 (time permitting).
Note: Many suggested problems will be drawn from this text, so the correct edition is required.
You are not responsible from the diamond or star marked sections, topics in the book that are not covered in class.
+ Material covered in class (additional examples, topics etc. which might not necessarily be in the course textbook or in the slides posted).

For students not familiar with elementary probability or 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 1-9.
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).

**Course description:** Objective: Statistics is the language in which data is analyzed and interpreted, and thus any serious data scientist must have a firm understanding of the mathematical principles of probability and statistics. Assiduous students of this course will build this critical foundation.

Details: This course is a self-contained introduction to probability and statistics with a focus on data science. The topics covered include fundamentals of probability theory and statistical inference, including: probabilistic models, random variables, useful distributions, expectations, the law of large numbers, the central limit theorem, point and

©Dr. Banu Baydil, 2017

confidence interval estimation, maximum likelihood methods, hypothesis tests, and linear regression (time permitting).

**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: Midterm, Thursday, November 2, 2015, in class.
   Final, 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) Class Participation: Students will receive 1 point extra credit for every problem they solve on the board up to a total of 3 points during the lectures. Class participation extra credit will be taken into account if the student's grade is at the cutoff line between two letter grades.

4) 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 10% 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.

5) Projects: You will be required to work on a Probability and/or Statistics related project in groups during the semester. The project will contribute towards enhancing your knowledge of applying Probability and/or Statistics to real world problems and strengthen your understanding of important concepts. You will be working in small groups and each group will hand in a joint report on their project and will give a short presentation to the class. The project will make 10% of your final grade. Finalized project reports are due last lecture of the semester. More information on the project will be posted later in the semester.

6) 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 are based on a set of lecture slides initially prepared by Prof. Cunnigham for earlier offerings of the course. Lecture slides will be posted on Courseworks, however, blackboard lectures will not be posted on Courseworks. You should NOT base your study of the material solely on the lecture slides but read the corresponding chapters/sections from the book while the slides are being covered. You are responsible making sure CourseWorks announcements/e-mails are going to an e-mail you check at least once daily.

©Dr. Banu Baydil, 2017

**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:**

| | |
|---|---|
| Week 1 | Introduction, sample spaces, probability axioms, counting (combinatorics) (Chapter 1) |
| Week 2 | Conditional probability, independence of events, Bayes Rule, random variables (Chapters 2-3) |
| Week 3 | Random variables, common random variables, expectation (mean), variance (Chapters 3, 4, 5) |
| Week 4 | Multivariate (joint) distributions (discrete and continuous), conditional distributions, independence of random variables (Chapters 3, 4, 5) |
| Week 5 | Covariance, correlation (Chapters 3, 4, 5) |
| Week 6 | Common random variables (Chapter 5) |
| Week 7 | Law of large numbers, central limit theorem (Chapter 6) |
| Week 8 | Point Estimation (Chapter 7) |
| Week 9 | Sampling distributions, Interval Estimation (Chapter 8) |
| Week 10 | Interval Estimation (Chapter 8) |
| Week 11 | Hypothesis Testing (Chapter 9) |
| Week 12 | Hypothesis Testing (Chapter 9) |
| Week 13 | Hypothesis Testing, Linear Regression (Chapters 9, 11) |
| Week 14 | Linear Regression, Student Presentations (Chapter 11) |
| Week 15 | Student Presentations |

©Dr. Banu Baydil, 2017