Calculus-based Introduction to Statistics
S1201

Time and Location
MTWR 10:45-12:20, 227 Seeley W. Mudd Building

Instructor
Hane Lee (they/them)
Email: hl3333@columbia.edu
Office hours: Monday 13:00-13:45 (via Zoom), Wednesday 13:00-13:45 (in-person, at SSW 10th floor lounge)

TA
Wribhu Banik (he/him)
Email: wb2387@columbia.edu
Office hours: TBA

Objectives
This course provides a comprehensive introduction to statistics, the science of data. It is designed for students who desire a strong grounding in statistical concepts with a greater degree of mathematical rigor than in STAT UN1101 (see the prerequisites at the next paragraph). After this course, students are expected to be able to understand real-life phenomena from a statistical perspective. Also, they are expected to be capable of taking more advanced statistics courses to explore these concepts and methods more deeply. Topics include basic probability, random variables, parameter estimation, hypothesis testing, and linear regression.

Prerequisites
High-school level algebra and counting, a solid understanding of univariate calculus (such as sigma symbol, differentiation, and integration). Some familiarity with multivariate calculus (such as partial derivatives, double integration) is recommended but not required. Before we use calculus results, I will briefly review them in class. A self-test for the mathematical background will be uploaded in Courseworks (HW0).

References
- Devore, J. L. (2011). Probability and Statistics for Engineering and the Sciences. Cengage learning. (9th edition) [Main text, for reading. Some homework questions will be taken from here.]
- Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2012). Open Intro statistics. Boston, MA,USA. OpenIntro. [Optional, for additional reading. Available online for free.]

Grading
- Homework (20%)
  There will be 5 (or 6) equally weighted weekly homeworks, to be turned in through Courseworks. The overall homework score will be based on the best 4 (or 5) homeworks, excluding the lowest score. Late submissions will NOT be accepted, unless you are under exceptional circumstances. Please email the TA in that case with documentation (and cc me).
  Homeworks can be discussed, but must be written up individually. Flat out copying is

unacceptable.
- Midterm (40%)
in-class, tentatively on June 6.
- Final (40%)
in-class, tentatively on June 27.
- The final grade is subject to a curve, where the median is guaranteed a B or above and the upper quartile (Q3) is guaranteed an A- or above. The letter grade A+ will be given to outliers. Make-up exams will not be given except for well-documented emergencies.

Tentative Course Outline

- Week 1: introduction to data, descriptive statistics, data visualization (Ch 1), basic probability (Ch 2)

- Week 2: discrete random variables (Ch 3), continuous random variables (Ch 4)

- Week 3: joint probability distributions (Ch 5), point estimation (Ch 6), midterm review

- Week 4: confidence intervals (Ch 7), large sample theory

- Week 5: hypothesis testing (one sample mean, two sample mean; Ch 8, 9), testing contingency tables (if time permits; Ch 14)

- Week 6: linear regression (if time permits; Ch 12), final review