Syllabus for S&DS 107: Introduction to Statistics

Instructor:

Xiaoqian (Dana) Yang xiaoqian.yang@yale.edu

Teaching Fellows:

Tal Sarig tal.sarig@yale.edu

Dylan O'Connell dylan.oconnell@yale.edu

Time:

M - F 10:00 - 11:15, July 3rd - August 4th.

No classes on July 4th.

Location:

60 Sachem St (Watson Center), Room A51.

Office hours:

M - F 13:00 - 15:00.

24 Hillhouse Ave (Department of Statistics and Data Science), Room 107.

You will need:

- High school level algebra. It would help to have some knowledge of linear algebra or programming. But it is not required.
- A laptop. Please bring it to class so you can participate in in-class activities.
- Textbook: *Stats: Data and Models* by De Veaux, Velleman and Bock, 4th Edition. Available on Amazon and Yale bookstore for rent or purchase. Since the summer course moves at a fast pace, please make sure you have the textbook no later than July 6th.

Grades will be assigned:

- 30%: Homework.
- 30%: Quizzes. We will have daily reading assignments. There will be quizzes in class regarding the material covered in lectures and in the corresponding chapters in the textbook.
- 10%: Participation. This is not a class where I talk the whole time. We will have activities that needs your participation. If you are not able to attend a class, please let me know the previous day by midnight.

- 30%: Final Exam.

Academic honesty:

It is very hard to fail this course, unless the honor code is broken. For homework assignments you are allowed to collaborate with each other, or look things up online. But please remember to declare/cite/acknowledge help in any form. Failing to do so, no credit will be given to that homework.

Tentative Schedule:

Date	Topic	Supplementary reading: chapter number in the textbook
July 3rd	Sampling: sample and population; univariate data-sets.	10, 11.
July 4th	Federal holiday. No classes.	
July 5th	Intro to R; histograms and boxplots.	2, 3, 4.
July 6th	Sample spaces and events; probability and distributions; (population and sample) mean and variance.	13(partial), 14, 15
July 7th	Discrete distributions; random variables; independence; Bernoulli and binomial distribution.	16(partial)
July 10th	Law of large numbers.	13
July 11th	Continuous distributions; the normal distribution. Approximation of the binomials to normals.	16.
July 12th	Normal approximation for histograms; central limit theorem.	17.
July 13th	Quantiles; confidence intervals; confidence interval based on a normal distribution.	5.
July 14th	Confidence interval for a population proportion (review of CLT).	18.
July 17th	Confidence interval for the proportion in geometric distributions.	
July 18th	Review with exercise problems	

July 19th	Testing for a population proportion with normal approximation. Formal discussion of hypothesis testing; p-values; type-I and type-II error; power.	19, 21.
July 20th	The Student's t-distribution. Testing for a population mean with unknown variance (one-sample t-tests). Paired data: confidence interval and t-test for difference of means.	22.
July 21st	Midterm exam.	
July 24th	Two-sample t-tests (equal variance).	22.
July 25th	Joint probability distributions; covariance and correlation; scatter plots.	6.
July 26th	Simple linear regression: least squares optimization; residuals.	7.
July 27th	Simple linear regression: estimation and prediction; transformation of variables.	9.
July 28th	R^2; analysis of variance; geometric understanding of linear regression.	26.
July 31st	Bayesian statistics: the philosophy and Bayes' rule.	End of Ch. 14.
Aug 1st	Review by case study.	
Aug 2nd	Review by exam-like problems.	
Aug 3rd	Final exam.	
Aug 4th	Multiple linear regression; basic contrasts.	28.