

Spring 2019

CSC 21700 Probability and Statistics for Computer Science

Evan Agovino
CUNY City College

NYC Tech-in-Residence Corps

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: https://academicworks.cuny.edu/cc_oers

 Part of the [Computer Sciences Commons](#)

Recommended Citation

Agovino, Evan and NYC Tech-in-Residence Corps, "CSC 21700 Probability and Statistics for Computer Science" (2019). *CUNY Academic Works*.
https://academicworks.cuny.edu/cc_oers/129

This Syllabus is brought to you for free and open access by the City College of New York at CUNY Academic Works. It has been accepted for inclusion in Open Educational Resources by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

CSC 217 - Probability and Statistics for Computer Science

Course Title	Probability and Statistics for Computer Science
Time	Wednesday, 6:30 PM - 9:00 PM
Location	NA 6311
Credits & Hours	3 Credits, 3 Hours
Instructor	Evan Agovino
Email	evan.agovino.ccnyc@gmail.com
Office	Varies
Office Hours	Varies, by appointment

Description

This course is an introduction to the practical tools of probability and statistics, including but not limited to descriptive statistics, probability theory, knowledge of discrete and continuous distributions, random variables and estimation, hypothesis testing and regression. The course takes a computational and applied approach to these topics. Though we will cover the mathematical theory behind our topics, students' output will involve applying said theory to real-world problems and datasets.

The course will be presented entirely in **Python** to mimic the workflow and tools used by professional Data Analysts and Data Scientists. No prior knowledge of Python is required. The majority of the class will use a few core packages, including **Numpy** for number simulation and statistical learning, **Pandas** for data exploration and cleaning, and **Matplotlib** for data visualization. In-class assignments, homework, and projects will be submitted via **Jupyter Notebook** files.

Class time will be split between lectures and hands-on group work, with occasional quizzes, announced and unannounced, to check for understanding. In-class participation is essential to the course as a means of understanding and applying the concepts covered in lecture.

Pre-Requisites

Math 20100 with minimum C grade, CSC 10300, CSC 10400.

Course Objectives

By the end of the course, students should be proficient at:

1. **Single Variable Explorations:** Examine a single variable, understand its underlying distribution, and choose the appropriate summary statistics for it.
2. **Pair-Wise Exploration:** Identify possible relationships between variables and compute correlations and linear fits.
3. **Estimation and Hypothesis Testing:** Understand the following three questions when reporting statistical results: 1) How big is the effect? 2) How much variability should we expect if we run the same measurement again? 3) Is it possible that the apparent effect is due to chance?
4. **Visualization:** Use data visualization as a tool for examining data and communicating results

Grading

	Weight
Group Project	25%
Midterm Exam	25%
Final Exam	25%
Homework/Quizzes	15%
Participation	10%

Group Project: Students will work on a small group project throughout the second half of the course that they will be expected to present to the class at the end of the semester. Projects will be graded based on a demonstration of core principles taught in class, and effectiveness of communication in their presentation. Details of the project will be shared later in the semester.

Exams/Quizzes: The midterm and final exams will focus on the core concepts covered in the class and will mimic the style of questions frequently asked in interviews for data-related roles.

In-class quizzes, both announced and unannounced, will occasionally be administered to check for understanding.

Homework: Students will be given weekly homework assignments in Python to check for comprehension of the material. Homework will be graded on a 5-point scale for completeness and effort. Homework will be due at the beginning of class every week. **Any homework submitted after 6:30 PM on the Wednesday it is due will be scored as 0/5.** Exceptions will be granted only as mandated by CUNY policy.

Participation: Students are expected to attend class and be active participants in discussion.

Texts and Materials

Texts

Students will be given weekly reading assignments from the following texts:

- *Introduction to Probability and Statistics for Engineers and Scientists*, Sheldon M. Ross, Third Edition.
- *Think Stats: Exploratory Data Analysis in Python*, Allen B. Downey, Second Edition. Available [here](#)

Additional Readings

Students will be given additional readings throughout the semester related to the material on a given week. Additional readings will be shared in the Github repo as they are added.

Github

All class materials, including the syllabus, readings, assignments and more can be accessed at the Github repository [here](#).

Binder

Students will use **Binder** as a means of running Python via Jupyter Notebook. Binder is a cloud-based executable environment that lets anyone interact with a Jupyter notebook through an internet browser (and no registration required!). The Binder link is below and may be periodically updated throughout the semester.

launch binder

Slack

Students will receive an invite to join a Slack channel for the class, which they are required to join. While email and/or Blackboard (TBD) may be used as a channel for some administrative updates, updates may also be sent via the Slack channel, which students are responsible for monitoring. Students are also encouraged to use Slack to communicate with each other to work on class materials, replicating the workflow used by professionals.

Tentative Schedule: Spring 2019

Week	Date	Topic
1	January 30	Introduction
2	February 6	Descriptive Statistics
3	February 13	Basic Probability I
4	February 20	Basic Probability II
5	February 27	Random Variables and Distributions
6	March 6	The Normal Distribution and Central Limit Theorem
7	March 13	Midterm
8	March 20	Estimation and Confidence Intervals
9	March 27	Hypothesis Testing
10	April 3	Relationships Between Variables
11	April 10	Regression
12	April 24	Regression II
13	May 1	TBD: Bayesian Modeling
14	May 8	Project Presentation
15	May 22*	Final Exam

There will be no class on Wednesday, April 17 due to Spring Recess and no class on

Wednesday, May 15 due to Reading Day.

*Final Exam date subject to change

CUNY Policy on Academic Integrity

[The CUNY Policy on Academic Integrity](#). The policy, as adopted by the Board, is available to all students. Academic dishonesty is prohibited in the City University of New York and is punishable by penalties, including failing grades, suspension, and expulsion.