Data 100, Spring 2025

Homework 6 Statistics

Due Date: Thursday, April 10th, 11:59 PM PT

Submission Instructions

You must submit this assignment to Gradescope by the on-time deadline, **Thursday, April 10th, 11:59 PM PT**. Please read the syllabus for **the Slip Day policy**. No late submissions beyond the Slip Day policy will be accepted unless additional accommodations have been arranged prior. While course staff is happy to help you if you encounter difficulties with submission, we may not be able to respond to last-minute requests for assistance (TAs need to sleep, after all!). **We strongly encourage you to plan to submit your work to Gradescope several hours before the stated deadline.** This way, you will have ample time to contact staff for submission support.

There are two parts to this assignment listed on Gradescope:

- **Homework 6 Coding**: Submit your Jupyter Notebook zip file for Homework 6, which can be generated and downloaded from DataHub using the `grader.export()` cell provided.
- **Homework 6 Coding Written**: Gradescope will automatically submit the PDF from the zip file submitted earlier. You do not need to submit anything to this assignment yourself, but **you are responsible** for checking that it is submitted properly.
- **Homework 6 Statistics**: Submit a PDF to Gradescope that contains all your answers to all questions in Homework 6 (Statistics).

To receive credit on this assignment, **you must submit both your coding and written portions to their respective Gradescope portals**.

You can answer the below Homework 6 Statistics questions in one of many ways:

1. Type your answers. We recommend LaTeX, the math typesetting language. Overleaf is a great tool to type in LaTeX.
2. Download this PDF, print it out, and write directly on these pages. If you have a tablet, you may save this PDF and write directly on it.
3. Write your answers on a blank sheet of physical or digital paper. Note: If you write your answers on physical paper, use a scanning application (e.g., CamScanner, Apple Notes) to generate a PDF.

Important: When submitting Homework 6 (Statistics) on Gradescope, you **must tag pages to each question correctly** (it prompts you to do this after submitting your work). This significantly streamlines the grading process for our readers. Failure to do this may result in a score of 0 for untagged questions.

You are responsible for ensuring your submission follows our requirements and that the automatic submission for Homework 6 Coding Written answers went through properly. We will not be granting regrade requests nor extensions to submissions that don't follow instructions. If you encounter any difficulties with submission, please don't hesitate to contact staff before the deadline.

Collaboration Policy

Data science is a collaborative activity. While you may talk with others about the homework, we ask that you write your solutions individually. If you discuss the assignments with others, please include their names below.

Homework 6 Manually Graded Questions

- This is not a question. This is a reminder to check and make sure that your Homework 6 Coding Written (manually graded questions) have been automatically submitted as a PDF by Gradescope when the HW 6 Coding assignment is submitted.

From Bernoulli to Geometric Variables

1. (9 points) Let $X_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, where $p \in [0, 1]$.

   For this problem, you may find it helpful to review the fundamental rules of probability from Chapter 9.5 of the DATA C8 Textbook.

   (a) (1 point) Compute $P(X_1 = 1)$.

   (b) (1 point) Compute $P(X_1 = 1, X_2 = 1)$.

   (c) (1 point) Compute $P(X_1 = 1, X_2 = 1, X_3 = 0)$.

   (d) (2 points) Compute $P(X_1 = 1, X_2 = 1, \ldots, X_n = 1, X_{n+1} = 0)$. Express your answer in terms of n and p.

   (e) (2 points) Suppose you flip a coin repeatedly until you observe a single tail. For example, you might have a streak of two heads before you see a tail (H,H,T). A geometric RV Y represents the **number of heads** you observe in a series of coin flips until you observe the **first** tail.

(i) What are the possible values of Y?

Hint: As a starting point, you might think about the (H,H,T) scenario presented above, where $P(Y = 2)$.

(ii) Suppose a particular coin has $P(\text{Heads}) = p$. Derive $P(Y = y)$ in terms of p and y.

Hint: You might start by computing $P(Y = 2)$ *for the (H,H,T) scenario.*

(f) (2 points) The U.S. stock market has a positive return on 53% of trading days. For this problem, assume the positive or negative sign of daily stock market returns are i.i.d. Bernoulli RVs with $p = 0.53$.

What is the probability of observing a consecutive positive gain streak of five or more trading days, starting from the first trading day after this HW is due? For example, if the first trading day after the due date has a positive return, but the following day has a negative return, we would have a streak length of one day.

Note: For full credit, your answer cannot include an infinite sum. But, you can leave your answer as an unsimplified algebraic expression. For example, $(0.3)^3(0.5)^4 + 10$