Name:

uID:

Homework 5: Joint Probability, Independence, Covariance, and Correlation

Instructions: Write your answers directly on this pdf (via an editor, iPad, or pen/pencil). The answers should be in the specified place. Students will be responsible for loading their assignments to GradeScope, and **ensuring the answers align with the GradeScope template**.

The assignment should be uploaded by 11:50pm on the date it is due. There is some slack built into this deadline on GradeScope. Assignments will be marked late if GradeScope marks them late.

If the answers are too hard to read you will lose points (entire questions may be given 0).

Please make sure your name appears at the top of the page.

You may discuss the concepts with your classmates, but write up the answers entirely on your own. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

1. Let $X_1, ..., X_{50}$ be independent $U(2, 6)$ distributed random variables. Suppose that $\overline{X}$ is define as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   (a) Calculate $E[\overline{X}]$.

(b) Calculate $Var[\overline{X}]$.

(c) Assume n is very large, compute the approximate $\Pr(\overline{X} \geq 4.5)$ (hint, central limit theorem, you can look up the result for the function you need on line).

2. The US National Soccer Team (known as "football" much of the rest of the world) schedules a home-and-home series with the German team, where each team plays one game in their home country. A win is worth 3 points, a loss worth 0 points, and a tie 1 point. The probability mass function for these variables in the two-game seres is given in the following probability table.

		Home		
		0 (= loss)	1 (= tie)	3 (= win)
Away	0 (= loss)	0.1	0.08	0.05
	1 (= tie)	0.07	0.2	0.15
	3 (= win)	0.05	0.1	0.2

Answer the following questions, showing and explaining your work:

(a) What is the probability that the USA does not win at least one game?

(b) What is the marginal distribution for the USA's points at Home?

(c) Is the event that USA loses at home independent of the event that they tie away?

(d) Compute the covariance and correlation coefficient for Away and Home.

(e) Is the result from Home independent from the result from Away?

3. In this problem you are going to analyze the built-in R data set `iris`. First, extract the petal information from the Setosa species and save it to a vector `x` using this command:

   ```
   x = iris$Petal.Width[iris$Species == "setosa"]
   ```

   Answer the following: **Write down the R code you use to compute parts (a), (b), (c), (d) - it should be very short.**

   (a) Produce the following plots for `x`: (1) A histogram, (2) a box-plot, and (3) an empirical cdf. (Remember to place your plots in the space provided below. Also remember that items that won't fit in the space provided should be included at the END of the assignment.)

(b) Run an experiment where you choose $n = 10$ measurements at random from this data (small set of samples). Run this experiment 20 times and show the histogram.

(c) Run the same experiment as above with $n = 20$. Comment on the difference between histograms.

Each of the flowers in this example have four features: sepal length, sepal width, petal length, and petal width. Extract all four of these features (e.g. as vectors) and compute the following (show code examples):

(a) The covariance and correlation coefficient between sepal length and petal length. Are they strongly correlated? Does the result make sense?

(b) The covariance and correlation coefficient between petal length and petal width. Are they strongly correlated? Does the result make sense?