Final Exam

Instructions

Due date
This exam must be submitted via Blackboard before midnight on Monday, May 11.

Allowed resources
You may use R and any written or internet resources on this test, although you are not allowed to ask for help from anyone except Dr. Clair.

Honor Pledge
The department of mathematics and statistics is requiring all students taking remote exams to sign an honor pledge. In our case, I want you to include this pledge at the beginning of your exam:

The work I have submitted represents my own effort. While working on this exam, I did not communicate in any form with individuals other than the instructor.

Format
For this exam, you are to create an RMarkdown document with your answers to each question. Knit your document. Convert, if needed, to PDF.

Exams that are not knit will lose 10 points.

If you have difficulty knitting your markdown document, contact Dr. Clair as early as possible for help.

Your document should include any R code you needed, the results of running that code, and most importantly a clear statement of the answer to the problem.

Questions
There are 10 questions, worth a total of 100 points.

Put your solutions in order and number each problem clearly, preferably with this syntax:

## Problem 1

I will provide an RMarkdown template that you should use for the exam.
1. In an experiment where you roll two regular six-sided dice, here are some events:
   - $A$: Both of the dice show 5.
   - $B$: The sum of the dice is 5.
   - $C$: At least one of the dice is 5.
(a) What are the exact probabilities $P(A)$, $P(B)$, $P(C)$?
(b) What is the conditional probability $P(A|C)$?

2. Continue with the experiment and events of problem 1. True or false:
   (I) $A$ and $B$ are independent.
   (II) $B$ and $C$ are independent.
   (III) $C$ and $A$ are independent.
   (IV) $A$ and $B$ are disjoint.
   (V) $B$ and $C$ are disjoint.
   (VI) $C$ and $A$ are disjoint.

3. Let $X \sim \text{Unif}(1,2)$ be a uniform random variable on the interval $[1,2]$.
   (a) What is the exact value of the mean of $X$?
   (b) Compute or estimate the standard deviation of $X$.
   (c) Estimate the expected value $E[1/X]$ accurately to two decimal places.

4. In Scrabble, players start with a bag of 100 tiles. Each tile has a letter on it and a point value. Less common letters are worth more points. All 100 tiles are stored in the file `scrabble.csv` on our course website.

   At the start of the game, a player draws 7 tiles from the bag without replacement.
   (a) What is the expected point total of the tiles in a player’s first draw?
   (b) What is the probability that a player’s first 7 tiles contain no vowels (vowels are the letters A, E, I, O, U)?
The next questions use data from the Flint Water Study, available as flint.csv on our course website.

This has information about lead levels for tap water in homes in Flint, Michigan during the 2015 water crisis. Lead levels were taken at first draw (Pb1), after 45 seconds (Pb2), and after 2 minutes (Pb3).

5. (10) The graph shown here shows lead levels for Flint’s eight geographical areas, called “Wards”. Reproduce this graph as well as you can. A couple of points to note:
   - The y axis scale is logarithmic, which you can accomplish with +scale_y_log10()
   - There is no Ward 0 in Flint.
   - The horizontal red line is the EPA “action level” for lead in water, at 15ppb.

![Graph of lead levels across wards in Flint, Michigan](image)

6. (10) Test for a difference in lead levels after first draw (Pb1) and after 45 seconds (Pb2). These variables are quite skew, so be sure to choose an appropriate method.
The next questions use the data frame **Backpack** which is part of the **Stat2Data** library that you will probably need to install.

To access this data, load the library with the usual `library(Stat2Data)` command, and then type the command `data("Backpack")` to expose the data. After this, you can use the **Backpack** variable.

7. (10) (a) What are the five most common values of **Major** in the **Backpack** data?
   (b) Produce a table showing the number of **Female** students in each **Year**.

8. (10) Ignore all students taking less than 10 units. Make a scatterplot showing **BackpackWeight** as a function of **Units**, and show the regression line on your plot. Looking at the plot, do you think that taking more course units leads to a heavier backpack?

9. (10) (a) Is there a difference in **BodyWeight** for Males and Females in the **Backpack** dataset? State the results of your test with a *P*-value.
   (b) Is there a difference in **BackpackWeight** for Males and Females in the **Backpack** dataset? State the results of your test with a *P*-value.
   (c) Is there a difference in **BackpackWeight** between students with **BackProblems** and students without **BackProblems**? State the results of your test with a *P*-value.

10. (10) The data set **Rubber** from the **MASS** library contains information from testing of tire rubber. Harder tires should suffer less loss of material to abrasion, or rubbing.
    Make a linear model to predict abrasion loss (**loss**) from tire hardness (**hard**)
    (a) What is the equation for predicted loss in terms of tire hardness?
    (b) Predict the abrasion loss for a tire with a hardness of 60.

- Did you remember to include your honor pledge?
- Did you knit your Markdown document?