Problem Set 3

Name:

Due: Tues, Oct 6

Rubric

Points	Criteria
5	The solution is correct and well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no nontrivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).

Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.

```
# library packages for data manipulation
library(tidyverse)
```

1. Suppose I flip a fair coin and roll a fair die.

   a) Write out the sample space for this random process.

   ```
   # you can use code if you want!
   ```

   b) Write out the probability distribution for this random process.

   ```
   # you can use code if you want!
   ```

   c) Are the outcome of the coin flip and die roll independent events? Why or why not?

   d) Let A denote rolling an even number and B denote landing on heads. Are A and B mutually exclusive? Why or why not?

   e) What is $P(A \cap B)$? Show your work.

   f) What is $P(A \cup B)$? Show your work.

2. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 30% of students miss exactly one day of school, 12% miss two days, and 36% miss three or more days due to sickness.

   a) What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year? Show your work.

   b) What is the probability that a student chosen at random misses at most one day? Show your work.

   c) What is the probability that a student chosen at random misses at least one day? Show your work.

d) If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.

e) If a parent has two kids at a DeKalb County elementary school, what is the probability that both kids will miss some school, i.e. at least one day? Note any assumption you make.

f) If you made an assumption in part d or e, do you think it was reasonable?

3. Suppose you read an article that states that at Montana State University, 36% of students major in engineering, 12% major in mathematics, and 52% major in something else. The article then claims that 20%, 28%, and 22% of students graduate with honors if majoring in engineering, mathematics, and something else, respectively. For this question, assume that students can only have one major at MSU.

   a) Do the three major categories defined in the prompt partition the students at MSU? Why or why not?

   b) What is the probability that a student selected at random from Montana State graduates with honors? Show your work.

4. It is estimated that 40% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 98% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 7%. If an email is detected as spam, what is the probability that it is in fact a non-spam email?

5. It is estimated that 10% of whitefish in the Yellowstone river have a condition known as *Proliferative kidney disease*, or PKD. Suppose there exists a test to detect the presence of PKD with a sensitivity of .95 and a specificity of .97. Suppose **two** whitefish are selected at random from the Yellowstone river, and **both** test positive for PKD. What is the probability that neither fish actually has PKD? You may assume that the fish are independent of one another.