Homework #3

Statistical Methods with R, Spring 2019

Due in class on Wednesday February 13, 2019. *Instructions for "theoretical" questions: Answer all of the following questions. The theoretical problems should be neatly numbered, written out, and solved. Please do not turn in messy work. Working in small groups is allowed, but it is important that you make an effort to master the material and hand in your own work;Identical solutions will be considered as a violation of the Student Honor Code.* **Note that you are also required to turn in the computational portion of this assignment. It is on Canvas, under the name:**

`A4570SP19HW03-computational.ipynb.`

ON THE FRONT OF YOUR HOMEWORK CLEARLY PRINT THE FOLLOWING:

- Your full name.
- Your lecture number (either **APPM 4570** or **APPM 5570** or **STAT 4000** or **STAT 5000**).
- Homework number.
- **Points will be deducted if these instructions are not followed.**

Remember that writing style, clarity and completeness of explanations is always important. Justify your answers. (Be sure to place your homework in the correct pile, either "undergraduate" or "graduate".)

Theoretical Questions

1. Answer each of the following.

   (a) The CU Boulder triathlon team has 12 women and 9 men. The team is going to a race and can only enter 5 participants. What is the probability of randomly selecting a race squad of 5 participants with exactly 3 women?

   (b) What is the probability that that at least two people in a room of 45 have the same birthday? Assume that all possible birthdays are equally likely, and ignore leap year. (HINT: Recall that $P(A) + P(A^c) = 1$.)

   (c) Three dice are rolled and their face values are summed up. What is the probability that a sum of 12 appears? What is the probability that a sum of 13 appears? Show all work.

2. What does it mean for one event C to *cause* another event E—for example, smoking (C) to cause cancer (E)? There is a long history in philosophy, statistics, and the sciences of trying to clearly analyze the concept of a cause. One tradition says that causes raise the probability of their effects; we may write this symbolically is

$$P(E|C) > P(E). \tag{1}$$

   (a) Show that equation (1) implies that $P(C|E) > P(C)$. (*Hint:* Use Bayes' Rule.)

   (b) Another way to formulate a probabilistic theory of causation is to say that

$$P(E|C) > P(E|C^c). \tag{2}$$

   Show that equation (1) implies equation (2).

   (c) Let C be the drop in the level of mercury in a barometer and let E be a storm. Briefly describe why this leads to a problem with using equation (1) (or equation (2)) as a theory of causation.

   (d) Let A, C, and E be events. If $P(E|A \cap C) = P(E|C)$, then event C is said to "*screen event A off from event E*". Suppose that $P(E \cap C) > 0$. Show that "screening off" is equivalent to saying that $P(A \cap E|C) = P(A|C)P(E|C)$. What does this latter equation say in terms of independence?

3. Suppose a particular crime is committed in Jerry's apartment. We'd like to know whether Tom is guilty of the crime. We are torn as to whether we think he is guilty: we think it's equally likely that he guilty or not guilty. Suppose that, in similar situations, we know that if a suspect is guilty, 85% of the time their finger prints are found at the scene, and, we know that if a suspect is *not guilty*, 30% of the time their finger prints are found at the scene.

   (a) What is the probability that Tom's finger prints are found at the scene?

   (b) If Tom's finger prints are found at the scene, how likely is it that he is guilty?

4. (**APPM 5570/STAT 5000 Students Only**) The game of Yahtzee is played with five fair dice. The goal is to roll certain 'hands', such as Yahtzee (all five dice showing the same number), full house (three of a kind and two of a kind), small straight (where 4 of the 5 dice are sequential, for example, $\{1, 2, 3, 4, 6\}$) and large straight (all five dice are sequential). In the first round of a player's turn, the player rolls all five dice. Based on the outcome of that roll, the player has a second and third round, where they can then choose to re-roll any subset of the dice to get a desired hand.

   (a) What is the probability of rolling a Yahtzee on the first round?

   (b) Suppose that, on the second round, the dice are $\{2, 3, 4, 6, 6\}$. You decide to re-roll both sixes in the third round. What is the probability that you roll either a small straight or a large straight ?

   (c) What is the probability of rolling a small straight on the first round?