ASSIGNMENT #4

EPsy 8252

This assignment covers probability simulation. The questions are adapted from Exercise #1 (Section 7.6), and Exercise #4 (Section 7.6) of Gelman & Hill (2007). Please submit your responses to each of the questions below in a printed document. Only provide your responses to the question asked. You do not need to include any R syntax and output unless it is specifically required in the question.

Any graphics you include should be resized so that they do not take up more room than necessary and all should have an appropriate caption. Any equations should be appropriately typeset within the document.There are 10 points possible for the assignment (each question is worth one point unless otherwise noted).

Adapted from Exercise #1

Suppose that a basketball player has a 60% chance of making a shot, and he keeps taking shots until he misses two in a row. Also, assume his shots are independent (so that each shot has a 60% probability of success, no matter what happened in before). Your goal is to write an R program that will simulate this process and determine how many shots the player takes before missing two in a row.

This is not an easy problem. Below, I will help you think through this process in stages to help you think about how we might write a simulation program. The following syntax produces a simulation of 100 shots for the player. The `set.seed()` function sets the seed for the random generator so that we can reproduce our results.

```
set.seed(25)
b = rbinom(100, 1, 0.60)
b
```

```
##   [1] 1 0 1 0 1 0 0 1 1 1 1 1 0 1 0 1 1 0 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 0
##  [36] 1 0 1 1 1 0 1 1 0 1 1 0 0 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 1 0 0 1
##  [71] 0 1 1 0 1 1 1 0 1 1 0 1 1 1 0 0 1 1 0 1 1 1 0 0 1 1 0 0 1 1
```

1. How many shots did the player take before missing two in a row.
2. Use the `sum()` function to compute the sum of the first two elements of the object `b`. Provide your syntax and output.
3. Now, write some syntax to generalize computing the sum of any two consecutive elements in the object `b`. (Hint: Use indexing (square brackets).) Provide your syntax and output.
4. Use the following pseudocode to write the syntax to actually compute how many shots were attempted before the player missed two shots in a row. Provide your syntax and output based on running the program on the object `b`.

initialize element number $i = 1$;
while *the sum of elements* i *and* $i+1$ *is not zero* **do**
  $i = i + 1$;
end
print $i + 1$

Algorithm 1: Pseudocode

5. Put your program in a `for()` loop and carry out 1000 trials of the simulation, each time recording the number of attempted shots before the player misses two shots in a row. Plot the distribution of these 1000 results. (Note: I do not need your syntax, only the plot of the results from the simulation.)

6. Compute the mean and standard deviation for the distribution of 1000 results.

7. Use your simulation results to collect both the number of attempted shots and the proportion of those shots that were successes (Note, you may need to change your program to do this.). Present these results for 1000 trials in a scatterplot.

Adapted from Exercise #4

Download the contents of the folder `beauty` from `http://www.stat.columbia.edu/~gelman/arm/examples/`. This folder contains data from (Hamermesh & Parker, 2005) on student evaluations of instructors' beauty and teaching quality for several courses at the University of Texas. The teaching evaluations were conducted at the end of the semester, and the beauty judgments were made later, by six students who had not attended the classes and were not aware of the course evaluations.

8. Fit a regression model that includes the predictors of beauty, age, sex, and whether the instructor is a native English speaker to explain variation in course ratings. Write the regression equation.

9. Now consider two instructors,

   - **Instructor A:** A 50-year old woman who is a native English speaker and has a beauty score of -1
   - **Instructor B:** A 60-year old man who is a native English speaker and has a beauty score of -0.5

   Simulate 1000 random predictions of the course evaluation rating for these two instructors. In your simulation, be sure to account for both sampling uncertainty and prediction uncertainty. Create a plot of the differences in course evaluations for the two instructors.

10. Based on your simulation results, what is the probability that Instructor A will have a higher rating than Instructor B?