Homework 1: Tidy Data and Simulation Studies

Introduction to Thinking Statistically

1. Suppose a book gives information on 72 different hikes that one can take in San Luis Obispo county. For each of the 72 hikes, the book reports the distance of the hike (in miles), the anticipated hiking time (in minutes), the elevation gain (in feet), the rating (easy, medium, or hard), and the region of the county in which the hike can be found (North County, Morro Bay, etc., for a total of eight regions).

a. What are the observational units here?

b. For each of the following, specify whether or not it is a legitimate variable for the observational units you specified in part a.

- The longest hike in the book

- The number of easy hikes in the book

- Whether or not the hike is easy.

- Whether or not the hike is in the Morro Bay region

- The proportion of hikes with an elevation gain of more than 500 feet

    c. Complete the top 6 rows of a possible data set in which the information could be stored. Label the columns. How many rows would the full data set contain?

```
head(hike_data)
```

_____

_____ _____ _____ _____ _____ _____

2. Consider a forced-choice procedure known as the "2 in 6 Test" which can be used to evaluate a patient claiming memory loss. For five seconds, a researcher presents the patient with a card displaying four letters, and the patient is instructed to remember the letters. After a delay of a few minutes, the subject is shown a second card which shows the same four letters plus two distractor letters that were not on the original card. The patient is then asked to recall any one of the letters that were on the original card.
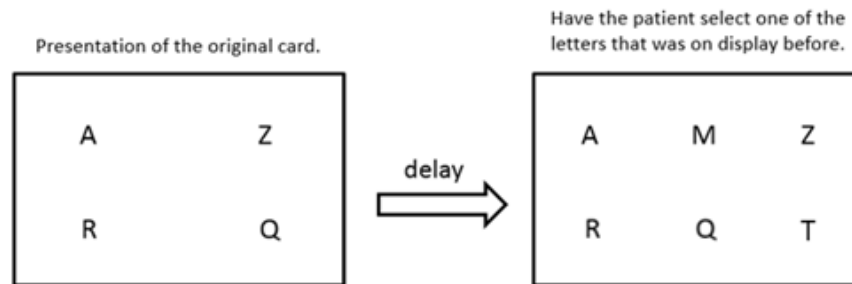
This process is repeated a total of 36 times, and it is noted whether the subject answers correctly each time. Suppose that in one study, a patient who is suspected of faking their memory loss gave a correct answer in 21 of the 36 trials.

    a. What is the expected value of the number of correct answers a subject would give if they were truly suffering from memory loss and were guessing when presented with the second card in each of the 36 trials? Answer this question with a single value, not a range of values. *Careful!*

b. Sketch a stacked bar chart to display the results from the patient suspected of faking their memory loss (our observed data) and next to it, sketch a stacked bar chart to display what we would expect to see if the patient truely was suffering from memory loss.

c. Now, set up a simulation study that would help us understand what outcomes (for the number of correct answers) are expected by chance if the patient is really guessing. Conduct 100 runs of the simulation using the `Online Simulation Applet` for `One Proportion Inference` (http://www.rossmanchance.com/applets/2021/oneprop/OneProp.htm).

- What are the two possible outcomes on each of the trials?

- What is the probability that the patient selects one of the letters that were on the original card under the assumption that the patient has memory loss? Change your `Probability of heads` accordingly.

- How many trials were conducted in this study? Keep this value in mind when setting the `Number of tosses` value.

  Copy/paste/sketch your simulation setup below.

d. Carry out a simulation study using the applet to determine what outcomes (for the number of correct answers) are expected by chance if the patient is really guessing. In the end, you should have a graph showing the results of 100 runs of the simulation. Copy and paste (or sketch) this dot plot below.

e. Why do we need to repeat the set of 36 "coin tosses" 100 times? Why not just look at one set of "36 coin tosses"?

f. Clearly describe what each dot on the dot plot summarizing your simulated results represents.

g. Recall that in this study, the patient who is suspected of faking their memory loss gave a correct answer in 21 of the 36 trials. Based on the results of your simulation study, does this outcome provide statistical evidence that this patient is faking their memory loss by answering wrong on purpose? Explain your reasoning, and be sure to use the results of the simulation study in your explanation.

h. Suppose that this patient had answered 27 out of 36 trials correctly. Based on the results of your simulation study, does this outcome provide statistical evidence that this patient is faking their memory loss by answering wrong on purpose? Explain your reasoning, and be sure to use the results of the simulation study in your explanation.

i. Can we use the results from this patient and simulation study to make claims about whether or not their sibling has memory loss? Why or why not?

> ◉ Assignment credit
>
> 1. Complete the homework quiz on Canvas (2 attempts; keeps average score).
> 2. Upload completed assignment (pdf, take scan on phone, etc.) to the assignment portal on Canvas to be checked for intellectual engagement.