Lab 4

BSS Stat 20

2022-07-18

NOTICE

If you have any collaborators, please write a sentence before Question 1 which acknowledges them. In addition, make sure that the sentence they therefore are also required to write acknowledges you.

For a template you can follow, see the course syllabus. This notice will be pasted at the beginning of every lab assignment.

All code MUST be shown. No credit will be given for correct answers without supporting code.

Questions

Question 1 - Confidence Intervals

Answer this question using a separate RMarkdown file knitted to a pdf and submit it under "Lab 4 Question 1" on Gradescope.

Parts a-f are to be done *without* looking at the `ppk` data frame or using code.

part a

What was the goal(s) of the Chancellor's office in commissioning this survey?

part b

To which population does the Chancellor's office wish to generalize their findings?

part c

Describe two parameters that the Chancellor's office is trying to estimate using the survey data.

part d

What was the initial sample size of students selected to take the survey? What was the final sample size that actually did? What was the response rate?

part e

Describe one scenario wherein this response rate would lead to results that they could not generalize from the sample to the population. Be specific.

part f

Sketch a plot of what the data might look like that is generated by Question 18 and 21. These two questions together show the change before and after participants were given basic information on the People's Park project.

Now, let's take a look at the `ppk` data set in R. The `ppk` data set represents a subset of questions that were asked in the questionnaire and have had random noise added to them. The results, in aggregate, share similar statistical properties to the raw data, but a given row no longer reflects an individual student's response completely.

part g

Create the visualization of **part f** using the questionnaire data. Add a title and axis labels. Then, describe the distribution in words.

part h

Create a new column called `support_before` that takes the response data from question 18 and returns `TRUE` for answers of `"Very strongly support"`, `"Strongly support"`, and `"Somewhat support"` and `FALSE` otherwise. What proportion of the survey participants in each class (freshman, sophomore, etc) supported the People's Park Project before being presented with the information on the bottom of page 14?

part i

Create a new column called `change_in_support` that measures the change in support from question 18 to 21. What is the average change in support of the survey participants in each class (freshman, sophomore, etc) for the People's Park Project after reading the information?

part j

Create a 95% confidence interval for the overall proportion of students that support the People's Park Project without having been exposed to the information on page 14. Interpret the interval in the context of the problem.

part k

Create a 95% confidence interval for the overall change in support the Project before and after being exposed to the information on page 14. Does the interval contain 0? What are the implications of that for those working in the Chancellor's Office on the People's Park Project?

Question 2 - Testing

Answer this question directly on this lab sheet and submit it under "Lab 4 Question 2" on Gradescope.

Is Yawning Contagious? An experiment conducted by MythBusters tested if a person can be subconsciously influenced into yawning if another person near them yawns. Our goal is to see if the results below they obtained are convincing evidence that yawning is contagious.

```
##        yawned
## seeded  0  1
##      0 12  4
##      1 24 10
```

First, we have some background questions which are important to answer.

part a

What are the explanatory and response variables?

part b

What was the proportion of yawners in the seeded group?

part c

What was the proportion of yawners in the unseeded group?

part d

If there were *no association* between yawning and the proximity of another yawner, what would you expect the difference to be between these two proportions?

part e

Let X be a random variable corresponding to the number of people in the unseeded group that yawned. What are the possible values that X can take?

part f

In terms of X, what is an example of a result x that would demonstrate a *strong* association between yawning and being exposed to a seed?

Now, to determine whether the results the Mythbusters obtained were "special," we will perform a simulation. What kind of data would be observed if there was no association between yawning and seeding; that is, if the only variation was caused by the process of randomly assigning subjects to the two conditions?

part g

Perform a simulation of the Mythbusters data and in doing so, fill out the below table. The simulation steps are as follows:

1. Create a deck of cards, 36 of which represent subjects who did not yawn, 14 of which represent subjects who yawned.
2. Shuffle the deck of cards to simulate the process of randomly assignment to the two conditions: being exposed to a yawn (seeded) and not being exposed (unseeded).
3. Deal them into two decks of size 16 and 34, representing the 1/3 of the subjects that were assigned to the unseeded group and the 2/3 assigned to the seeded group.
4. Count the number of yawners that happened to end up in the unseeded group just by chance and record your count below as x_1.
5. Repeat steps 1-4 five more times.

x_1	x_2	x_3	x_4	x_5	x_6

part h

Now, each lab group will contribute to a dotplot on the chalkboard of realizations of X. Reconstruct the dotplot below.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

part i

How many yawners would you expect to find in the unseeded group just by chance?

part j

What value of X did the MythBusters actually observe?

part k

Is this data convincing evidence that yawning is contagious? Why or why not?