Lab07: The dance of the p-values and p-hacking

name and student ID

Today's date

- Due date: July 31, 10:00pm (make sure to provide enough time for Gradescope submission to be uploaded).
- Each answer must be on a separate page. Don't delete or add any `\newpage` tags.

Remember: This homework does not involve autograder checking! The submission process is different than usual

- Submission process: Once finished, export your PDF and upload directly to Gradescope. Make sure to not remove or add any `\newpage` tags.

Helpful hints:

- Every function you need to use was taught during lecture! So you may need to revisit the lecture code to help you along by opening the relevant files on Datahub. Alternatively, you may wish to view the code in the condensed PDFs posted on the course website. Good luck!
- Knit your file early and often to minimize knitting errors! If you copy and paste code for the slides, you are bound to get an error that is hard to diagnose. Typing out the code is the way to smooth knitting! We recommend knitting your file each time after you write a few sentences/add a new code chunk, so you can detect the source of knitting errors more easily. This will save you and the GSIs from frustration! **You must knit correctly before submitting.**
- If your code runs off the page of the knitted PDF then you will LOSE POINTS! To avoid this, have a look at your knitted PDF and ensure all the code fits in the file (you can easily view it on Gradescope via the provided link after submitting). If it doesn't look right, go back to your .Rmd file and add spaces (new lines) using the return or enter key so that the code runs onto the next line.
- Useful mathematical notation in markdown:

$$\mu$$

$$\sigma$$

Overview: In today's lab you will read and watch articles from the internet about differeces in p values and confidence intervals. You will also review information about p-hacking and data dredging to bring you up-to-speed on some of the language used to talk about bad scientific practice around the misuse of p-values.

To do:

- Watch this 11-min Youtube video on P-hacking: https://www.youtube.com/watch?v=Gx0fAjNHb1M

- Read this wikipedia article on data dredging: https://en.wikipedia.org/wiki/Data_dredging
- Read this Vox article about the Cornell food researcher: https://www.vox.com/science-and-health/2018/9/19/17879102/brian-wansink-cornell-food-brand-lab-retractions-jama
- Read this two-page ASA brief on statistical significance and p-values: https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf

In your own words:

- What is p-hacking?

[YOUR ANSWER HERE]

- What is data dredging?

[YOUR ANSWER HERE]

- One of these sources provides an example of p-hacking in epidemiology related to cancer clusters. Describe in your own words what the problem is.

[YOUR ANSWER HERE]

- What are three practices noted in one of the articles to reduce p-hacking? Name each one and describe them in 1-2 sentences.

[YOUR ANSWER HERE]

- One of the sources give a correction method for calculating p-value when you are going to conduct multiple tests. What is the name of the method? Write down the correction using an equation.

[YOUR ANSWER HERE]

Please watch the video here:

https://www.youtube.com/watch?reload=9&v=5OL1RqHrZQ8

Q1. Which p-value is mentioned as leading to "Elation"

[YOUR ANSWER HERE]

Q2. How big was the "true" difference in the imaginary experiment described?

[YOUR ANSWER HERE]

Q3. Which measure gave a better estimate of the variability in results over multiple simulated studies?

[YOUR ANSWER HERE]