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

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

Please watch the video here:

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

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

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

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