

PSY 392 – Data Mining
Homework 5
Due Friday 8 November at 5pm

Purpose: to develop your understanding about p-hacking (online app:
<https://shinyapps.org/apps/p-hacker/>)

Use the online app with the True effect (Cohen's D) set to zero, which means the data are going to be generated from distributions where there is no effect. Thus, any “significant” finding is a false positive (Type 1 Error).



Assignment:

Part 1 – Settings for initial data collection

Begin by thinking of a research area you find interesting and consider what the experimental and control groups would be in an experiment in this realm of research.

- A. Enter the name for your experimental group. Explain why this group is considered the experimental group. (5 points)
- B. Enter the name for your control group. Explain why this group is considered the control group. (5 points)
- C. Consider the number of dependent variables that interest you. **Come up with at least three DVs** and provide a brief explanation for each one, how they might be measured, and how they relate to your construct of interest. Note, I am not expecting you to be an expert or review journal articles to find measures people actually use. (10 points)

Part 2 – p-hack!

Modify as many options as you want (except for the True effect slider, which should be set to 0).

- A. Provide a detailed description of the steps you had to take to find a significant effect. Demonstrate to me that you put some effort into this process and spent time thinking about what the parameters you are adjusting do to the data. (this should be at least half a page). (15 points)
- B. Take screen shots of (10 points)
 - a. “Settings for initial data collection:”
 - b. “Now: p-hack”
 - c. “Tests for each DV”
 - d. Plots of any significant findings

Part 3 – tell your story

- A. Using your variables defined in Part 1, explain the pattern of the results obtained in a coherent manner. This should be longer than half a page and no more than 2 pages. Demonstrate to me that you are able to interpret the results table in the context of your chosen example. (30 points)

Part 4 – What’s wrong with P-hacking?

- A. Reflect on the process you described in 2A. What is wrong with this approach to analyzing data? What should you do differently when exploring future data sets. Demonstrate to me that you have thought about p-hacking in a critical way and how this exercise will make you a better data scientist in the future. (25 points)