Arjun Krishnan

Day03 Assignment

Due before class Wed, Nov 18

> Download and work with R or Jupyter notebook and submit your completed notebook via Slack.

Folks, unlike previous years, I have written the code already. So, this assignment should not take more than **30–40 min** *max*. So, please submit this before next class. However, if you need more time, do not hesitate to send me a message.

The R or Jupyter notebook – whichever one you choose – contains all the details in the `Assignments` directory in the class website: https://github.com/krishnanlab/teaching/tree/master/2020-fall_statgaps/Assignments

- `Day-03_Assignment.Rmd`
- `Day-03_Assignment.ipnb`

Overview of the assignment

Submit the R or Jupyter Notebook back to me via Slack with your comments/annotations on the code and the results, along with your interpretation of the results and answers the questions at the end of each part.

The assignment contains **three** parts.

1. **Calculating p-values using a permutation test.**
   a. This first part of the assignment is identical to the exercise we did in-class yesterday (Nov 16).
   b. Your tasks for this part 1 are to complete what you started in class:
      i. Add comments/annotations on the code and the results, along with your interpretation of the results.
      ii. Answers a couple of questions at the end.
2. **Effect of effect_size, variance, and sample_size on p-values.**
   a. Like we discussed in class, apart from the null hypothesis (which need not always be equivalent to random chance), the p-value of a statistical test depends on effect size, variance with each group, and sample size.
   b. To explore how these factors influence the p-value, I have written the below to simulate data for two groups multiple times (just like the exercise we did in class), each time varying the

`effect_size`, `std_deviation`, and `sample_size`, and calculating the p-value using a T-test.

c. Your tasks are the following:
    i. Carefully examine and annotate the code by writing detailed comments at each step.
    ii. Examine the figure that is being produced and write a short paragraph about you observations on how these three quantities influence the p-value.

3. **P-hacking**
    a. P-hacking is the practice of collecting or selecting data or statistical analyses until nonsignificant results become significant.
    b. You tasks:
        i. Carefully examine and annotate the code by writing detailed comments at each step.
        ii. Describe your thoughts on what this coding exercise has to do with p-hacking.