Arjun Krishnan

Day 06 Pre-class Assignment

Please submit your assignment as a RMarkdown or a Jupyter notebook. Include your code and results/plots for the programming exercises.

1. Generating power curves

Complete what you started in class by working through the script we discussed to generate power curves for the experiment to determined if a coin is biased or not.
The R code is here: https://statgaps2019.slack.com/archives/CPCHCEH37/p1574106006017400

Your tasks are the following:

- Carefully examine and annotate the code by writing detailed comments at each step. Focus your comments on the logical/statistical aspects of each step. In other words, instead of focusing on how things are exactly coded-up (specific choice of R functions or the implementation), comment on what each line/block of code is achieving with respect to the steps in our statistical exercise.
- Run this code again for the following cases:
  - Number of samples equal to 100.
  - Alpha equal to 0.01.
- Make a plot containing these four power curves and write your observations.

2. Sample size calculation

The second part of your assignment involves reading this review article:

> **Power failure: why small sample size undermines the reliability of neuroscience**
> https://www.nature.com/articles/nrn3475
> [PDF uploaded in the #assignments channel.]

Based on your reading, answer the following questions using terms/quantities in this figure https://drive.google.com/open?id=1EbndpgrPNlnuGWmnvKLFp5QHSWjxCNws:

1. How does low power reduce the likelihood that a statistically significant result reflects a true effect?
2. How does low power lead to overestimates of effect size?

3. Regression to the mean

Read the following article, parts of which were used in the in-class lecture.

> **Regression to the mean: what it is and how to deal with it**

> https://academic.oup.com/ije/article/34/1/215/638499
> [You can find a PDF of this article in the #assignment channel.]

Based on your reading, answer the following questions in your own words:

1. Why are group-level statistics more susceptible to RTM compared to individual-level statistics?
2. Describe Figure 3 and the analysis of RTM using graphs.
3. Briefly describes ways to address RTM at the design and analysis stages.