Effect Size & Statistical Power

When too much or too little can be bad things…

Statistical power

- Represents the power of a test to correctly reject the null hypothesis
- Varies with each test that you run
- Sample size is very important in driving power

Effect Size

- "A measure that allows one to judge the relative importance of a difference or relationship by reporting the size of the difference" relative to the sample size (error).

Chi-square:

Roughly 0.1 = small, 0.3 = medium, 0.5 = large effect size

Student's t:

Roughly 0.2 = small, 0.5 = medium, 0.8 = large effect size

Effect Size

- A small effect size with large samples means power may be too high
  - Differences may be significant, but not important
- A large effect size small samples means that power may be too low
  - Differences may not be significant, but important

For Student's *t*

- One sample Cohen's *d*

$$d = \frac{|t|}{\sqrt{n}}$$

- Two sample Cohen's *d*

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{pooled}} \quad \text{or} \quad d = \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

For Chi-Square

- For Goodness of Fit, Cohen's *w*

$$w = \sqrt{\frac{X^2}{n}}$$

- For Test of Independence

$$\phi = \sqrt{\frac{X^2}{n}}$$

Summary

- If $p < \alpha$ and the samples are large, check for Effect Size
  - If effect size is LOW, the difference may not mean much
    - Test is too powerful
- If $p > \alpha$ and the sample is small, check for Effect Size
  - If effect size is HIGH, there may be a meaningful difference that was not detected
    - Test is not powerful enough
- The easiest way to manipulate power is by adjusting sample size