**Week 2 Notes** – **Falsifying Predictions**

1. Falsifying predictions in theory
   1. Null-effect predictions
      1. Sometimes demonstrating absence of an effect is interesting
         1. Is a new, easier intervention **equally effective** as time-intensive intervention?
         2. Theory predicts two things should not differ
   2. What would falsify your hypothesis?
      1. You might think when p > alpha
         1. **Absence of evidence** is not **evidence of absence**
         2. Need appropriate power to detect real differences
      2. A significant result in the **opposite direction?**
         1. If null hypothesis is true, you shouldn’t find significant effects in *either* direction
         2. Not enough to falsify your prediction?
      3. Is **any** effect in the predicted direction support for H1?
         1. D = 10 (10 sd) support for your theory?
            1. This effect is much too large to be in support of your theory
         2. Danziger, Levav, & Avnaim-Pesso, 2011 but see Glockner, 2016
            1. Judges decisions as function of time of day (lunch-effect)

Huge effect size of cohen D ~= 2

Causes question because it much too large to be plausible

* + - 1. Is an effect size of d = 1.0 support for your theory?
         1. Social exclusion
      2. Tiny effects are still support *if we have a directional prediction* but may be meaningless
         1. 90% power for d = 0.001 in a t-test requires 42 million observations in total
  1. You can never prove an effect is exactly 0
  2. **Undead theories (zombies)** Ferguson & Heene, 2012

1. Setting the smallest effect size of interest (SESOI)
   1. Benefits
      1. Power for effects you want, not effects you expect
         1. Design a study to detect effects as large or larger than SESOI that’s useful
         2. Helps design informative studies
      2. Study is falsifiable if data < SESOI
   2. How to determine SESOI
      1. Theoretical predictions
         1. Range of values predicted to specify SESOI
         2. Burriss et al., (2015)
            1. Predictions: increased redness in the face during the fertile phase of ovulatory cycle in women
            2. Data: high resolution, statistically significant increase

However, theory would predict that it should be noticeable by the naked eye (presumably by men)

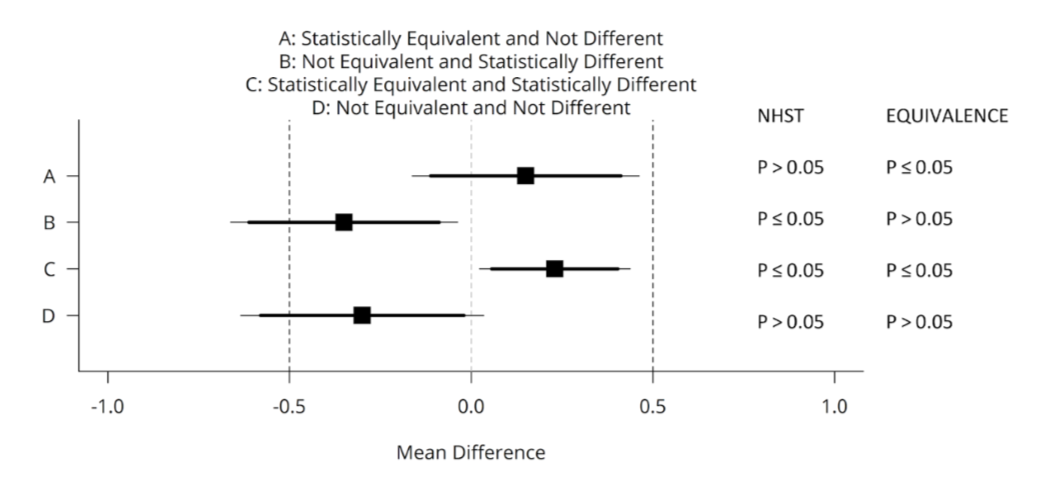
* + - * 1. Result: effect was smaller than predicted by theory!
      1. Anchor-based methods
         1. Self-report “small” difference on a measure
         2. T1 and T2 has your happiness improved?
      2. Norman, Sloan, and Wyrwich (2003)
         1. Surprisingly consistent minimally important difference of cohen’s d = 0.5
    1. Practical significance
       1. Cost-benefit analysis
          1. How large an effect should be to achieve sufficient benefits to outweigh the costs of intervention
          2. Is the effect additive?

Tiny effects can stack (e.g., banking interest)

* + - * 1. Threshold may be required for additive effects
      1. Feasibility
         1. Due to constraints of resources can only study certain size effect sizes

Effect size is a great example

* + - * 1. Sample sizes will limit the lowest effect sizes detectable
  1. Psychological Science Accelerator
     1. 100s of labs collecting data together and pooling resources

1. Falsifying predictions in practice
   1. Overall view
      1. Is the effect smaller than what we predicted/care about?
   2. Equivalence testing
      1. **Reject the presence of a meaningful effect**
      2. Determine **smallest effect size of interest (SESOI)**
         1. Specify equivalence range (close enough to 0 effects)
         2. Test if values outside this range can be statistically rejected
         3. *Reverses the traditional hypothesis test* – *reject hypothesis that effects are larger than what we care about (SESOI)*
      3. Example
         1. Null hypothesis test: compare our value against 0
            1. 95% CI is used in the null hypothesis significance test
         2. Equivalence test: compare our value against the lower bound of SESOI (-0.5) and upper bound (0.5)
            1. 90% CI is used for equivalence test because this is two one-sided tests, best represented by 90% CI
         3. 
            1. A – not sig not meaningful

95% CI overlaps with 0 so null hypothesis test says we cannot reject null (0)

90% CI does not overlap with lower or upper

Conclusion: effect is not statistically different from 0 **and** we can conclude the effect is statistically equivalent (reject any effect large enough to matter)

* + - * 1. B – sig and potentially meaningful

Statistically significant difference

90% CI overlaps with lower – can’t reject effects that are small enough to matter

Statistically different from 0 but not statistically equivalent (might be large enough to matter)

* + - * 1. C – sig but not meaningful

95% CI does not overlap with 0 so reject null

90% CI does not overlap with lower or upper

Conclude: effect is statistically different from 0 but also smaller than anything we care about!!

Statistically significant, but practically insignificant

* + - * 1. D – not sig but potentially meaningful

95% CI overlaps with 0 so cannot reject null

90% CI overlaps with lower!

3Cannot reject hypothesis that the effect is large enough to matter

**Inconclusive result**

* 1. (Bayesian) Estimation
     1. ROPE (Region of practical equivalence) procedure
        1. Uses **posterior distribution** of most plausible values
        2. 95% HDI (highest density interval) compared against ROPE
     2. Philosophically different from equivalence testing
        1. ET: control error rates
        2. Bayes: whole distribution for inference
  2. Bayes factors
     1. Directly test two competing models using **Bayes Factor**
        1. Ratio of two points, indicating likelihood for the null
     2. Priors are an important part of the question you ask and need to be **justified**
     3. **Bishop’s law**
        1. The better designed a study is, the more likely it is to obtain a null result
     4. Design studies to yield informative results both when **H1 is true** as when **H0 is true**