# Econ 211

Prof. Jeffrey Naecker

Wesleyan University



# Power Calculations



## Type 1 and Type 2 Error

- Type 1 Error: Determining a hypothesis is true when in fact it is false
- Type 2 Error: Determining a hypothesis is false when in fact it is true

*The boy who cried wolf caused the villagers to commit a type 1 and a type 2 error, in that order.*

- Type 1: Believing there is a wolf when in fact there is not one
- Type 2: Believing there is a not wolf when in fact there is one



## Power Calcuations: Building Blocks

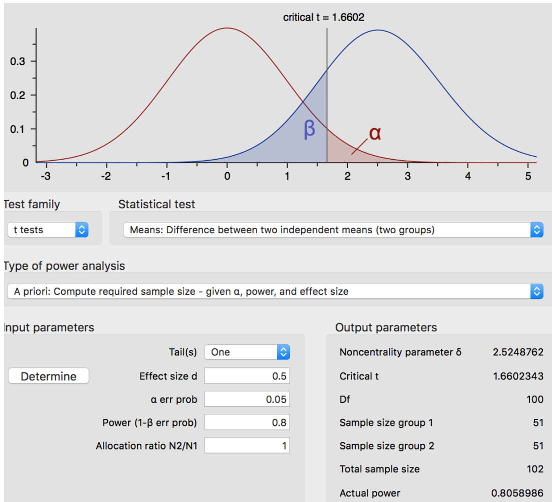
- Significance level $\alpha$ = P(Type 1 error)
  - Typically set this to 0.05
- $\beta$ = P(type 2 error) = 1 - Power
  - Typically set power to 0.80
- Sample size $n$: number of observations (eg subjects) in our study
- Minimum detectable effect $d$: cuttoff in our outcome varaible above which we claim to find an effect (ie we reject the null hypothesis)
  - Effect size typically given in standard deviation units:

$$d = \frac{\text{Mean of treatment group} - \text{Mean of control group}}{\text{Standard Deviation}}$$

- Effect size or sample size can be pulled from existing studies

If we pick any three of these values, the fourth is determined

## Power Calculation: Using gpower App




Source: http://www.gpower.hhu.de/



## Power Calculations: Using pwr Package in R

```
[] library(pwr)
test.results ¡- pwr.t.test(d = 0.5, sig.level = 0.05, power = 0.8, type =
"two.sample", alternative = "greater")
test.results

##
##      Two-sample t test power calculation
##
##               n = 50.1508
##               d = 0.5
##       sig.level = 0.05
##           power = 0.8
##     alternative = greater
##
## NOTE: n is number in *each* group
```



## Plotting Power vs Sample Size

```
[] plot(test.results)
```

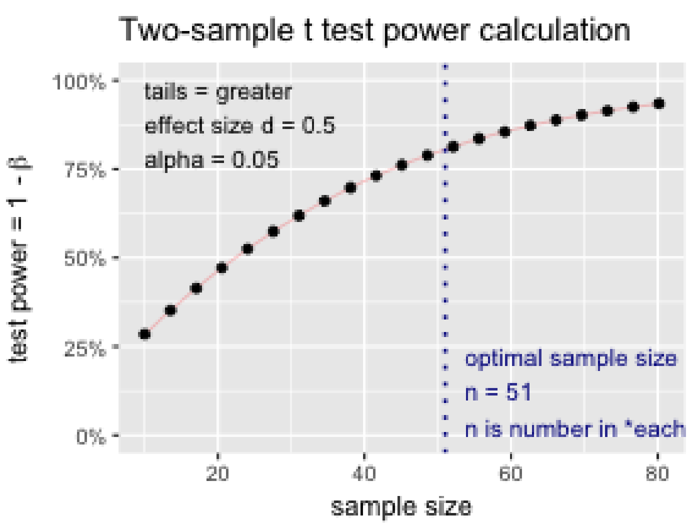



Figure: plot of chunk unnamed-chunk-2



## Zooming Out

- Might be tempted to say that if we set $\alpha = 0.05$, ie only 5% of reported true results will false positives
- However, suppose we run 1000 studies
- Suppose only 10% of studies have a real effect
  - How many real effects?
  - How many no effects?
  - How many true positives?
  - How many false positives?
- False discovery rate $= \frac{\text{False Positives}}{\text{Total Positives}}$
  - FDR in this case?

## False Discovery Rate Visualized

Source: Colquhoun (2014)



## More Tools

Demo by Rafael Charris Dominguez:

- False Discovery Rate: http://shinyapps.org/apps/PPV/
- statcheck: http://statcheck.io/



# Pre-Analysis Plans



## What's a Pre-Analysis Plan?

In a pre-analysis plan (also known as pre-registration), authors **publicly** post

- research hypothesis
- how/where/when data will be collected
- how data will be analyzed

**before** the data is collected

## Why Do We Need Pre-Analysis Plans?

- Reduced file drawer effect
- Reduce publication bias
- Reduce p-hacking
- Reduce HARKing



## Examples of Pre-Analysis Plans

- American Economic Association Registry
  - Good example: https://www.socialscienceregistry.org/trials/2196
  - Only for randomized controlled trials (RCTs) (basically experiments)
- Open Science Pre-Registration
  - Template: https://osf.io/t6m9v/
  - Can include non-experimental studies
- AsPredicted.org
  - Example: https://aspredicted.org/nfj4s.pdf



## Final Project

- PAP for your final projects is posted on Moodle
- Final presentation and paper must cover all topics in PAP
- Presentation: 8-12 minutes
- Paper: 6-10 pages