Family Studies Center Methods Workshop
Statistical Power Analysis

Adam Davey

Temple University

November 14, 2014

Overview

- Understand the role of statistical power analysis in family studies research
- Introduce concept of statistical power
- Develop intuitions about factors affecting statistical power
- Learn applications of power analysis when sample size is fixed

Motivating Problem

- Research is *difficult*, *time-consuming*, and *expensive* to conduct
- *Before* we conduct a study, we want to be assured that we have a reasonable change of finding an effect if, in fact, one exists
- We must recruit *sufficient numbers* of subjects into our study
- We must *also* consider efforts (and potential for risk) of study participants
- Nearly all studies entail at least some *risk* for participants (even after data are collected!)
- We must not recruit *too many* research subjects into our study

Motivating Problem

- When number of potential subjects is *limited*, need to *identify design* that gives us the best chance of answering our question
- When number of subjects is *fixed* in advance, need to know *how big an effect* we can detect in our data with desired probability

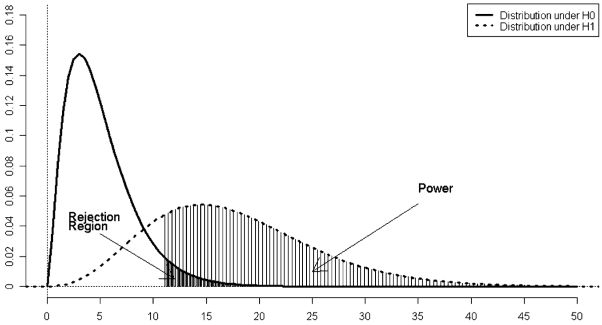
Ways to be Wrong in Hypothesis Testing

	True State of Affairs	
Decision	H_0 True	H_0 False
Accept H_0	Correct$(1-\alpha)$	β
Reject H_0	α	Correct$(1-\beta)$

Ways to be Wrong in Hypothesis Testing

	True State of Affairs	
Decision	No Effect	Effect
No Effect	Correct$(1-\alpha)$	Type II Error
Effect	Type I Error	Correct$(1-\beta)$

Central and Non-Central Distributions



Central and Non-Central Distributions

- *Central* distributions apply when the null hypothesis (H_0) is *true*
- They are *standardized*
- *Non-Central* distributions apply when H_0 is *false*
- They are *not standardized*
- Non-Centrality Parameter (NCP, λ) reflects *degree* to which (H_0) is false
- Non-centrality parameter can affect both *location* and *shape* of distribution.

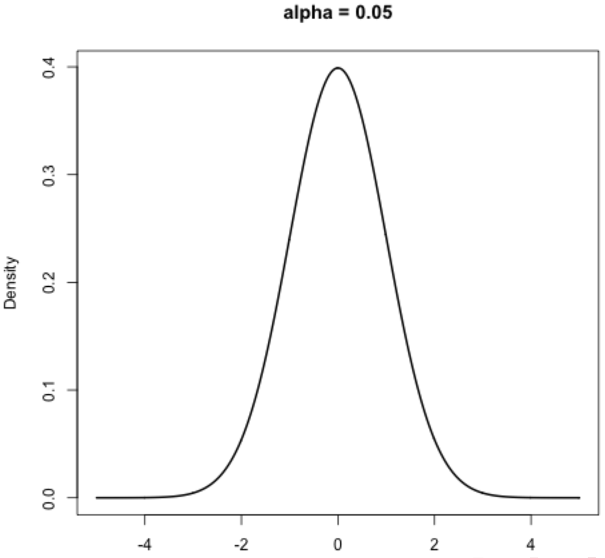
Central and Non-Central Distributions

- Central χ^2 distribution with *df* degrees of freedom can be generated by squaring and summing *df* different random normal variates with means of 0 and variances of 1
- Non-central χ^2 distribution with *df* degrees of freedom and $NCP = \lambda$ can be generated by squaring and summing *df* different random normal variates with means of $\mu = \sqrt{\frac{\lambda}{df}}$

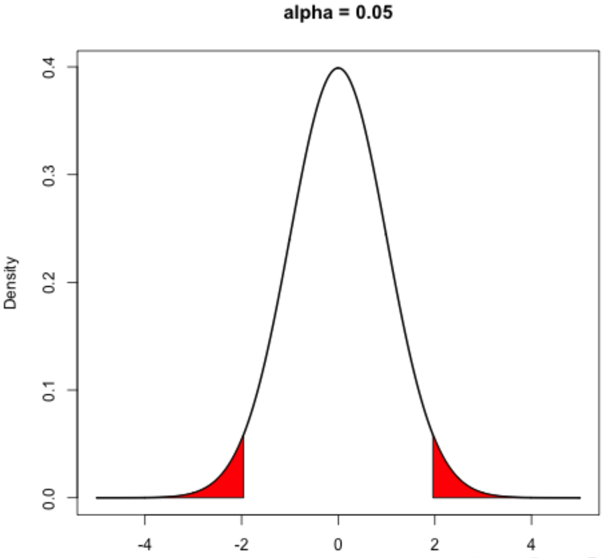
- Four variables are important for power analysis
- α
- Power, $(1 - \beta)$
- N
- Effect Size, (ES, λ)
- Knowing any 3, solve for fourth
- Two other factors include choice of H_0 and $Pr(H_0$ is false)

- Conventions
- $\alpha = .05$
- Power $\geq .80$
- *N* (Some applications, may define minimum acceptable standards or heuristics for overall sample size, distinct from power conventions)
- Effect Size
- Power analyses are invaluable *a priori*, not so useful *a posteriori* (http://www.stat.uiowa.edu/files/stat/techrep/tr378.pdf)

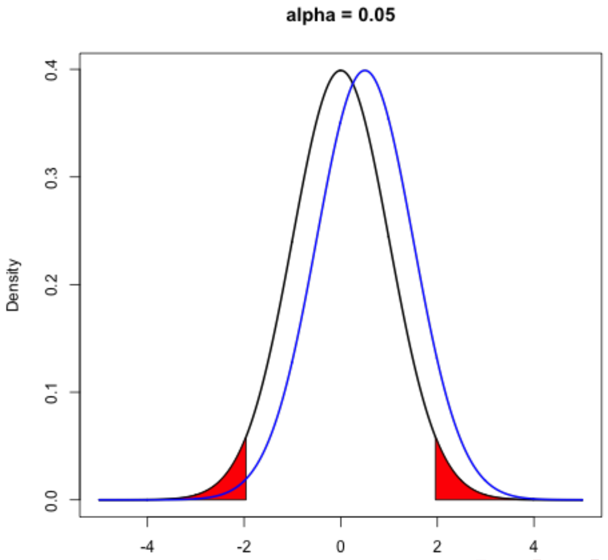
Visualizing Statistical Power



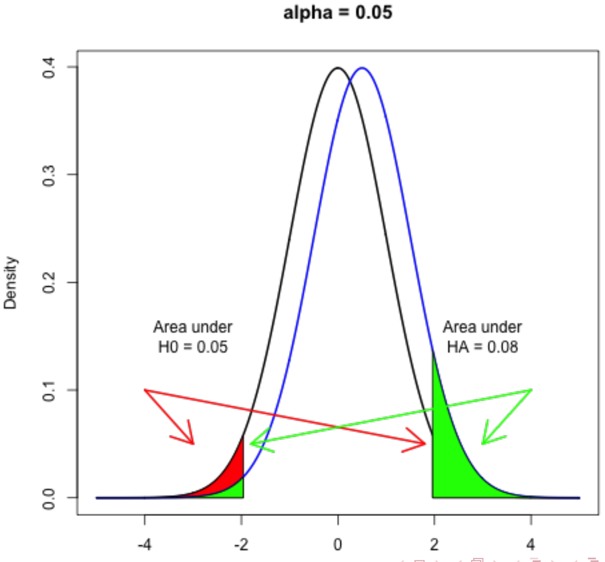
Visualizing Statistical Power



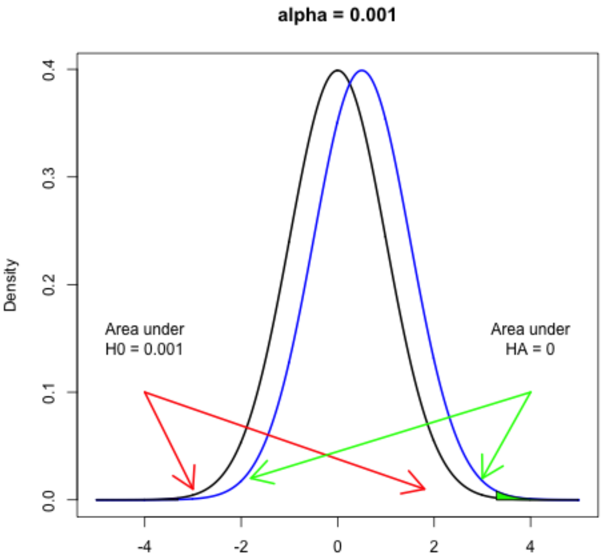
Visualizing Statistical Power



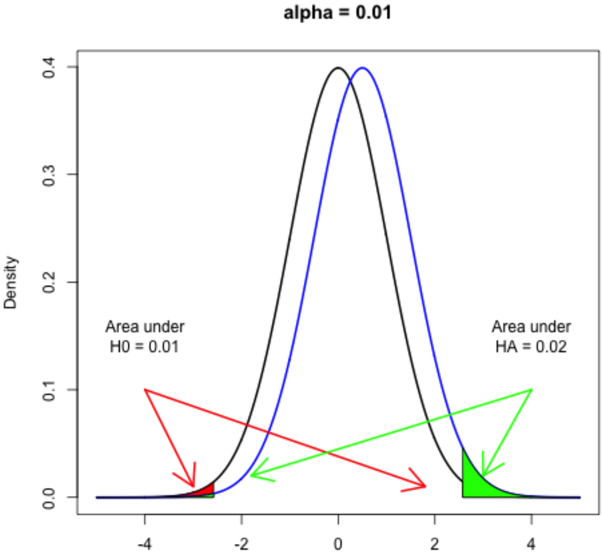
Visualizing Statistical Power



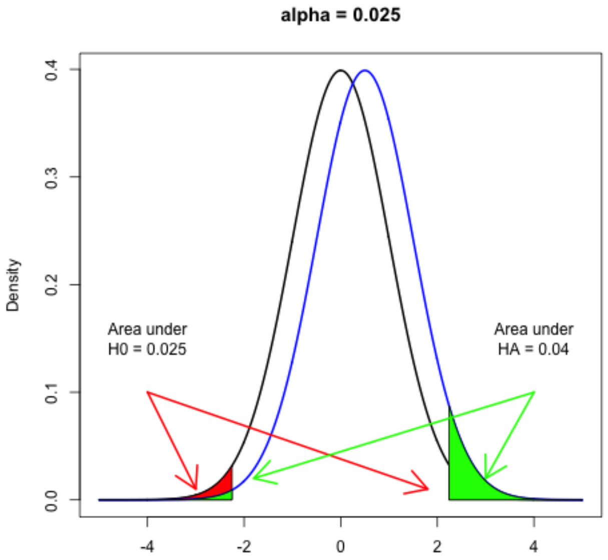
Effects of Increasing Alpha



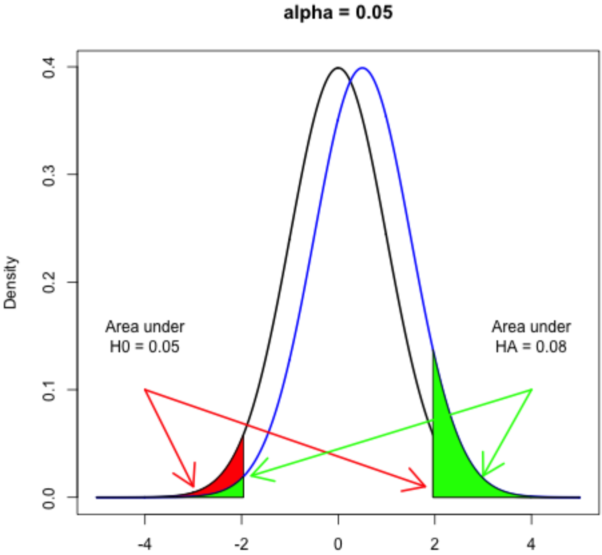
Effects of Increasing Alpha



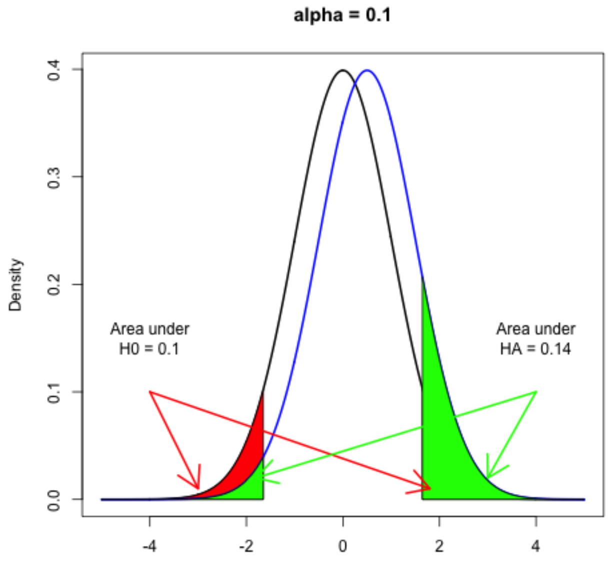
Effects of Increasing Alpha



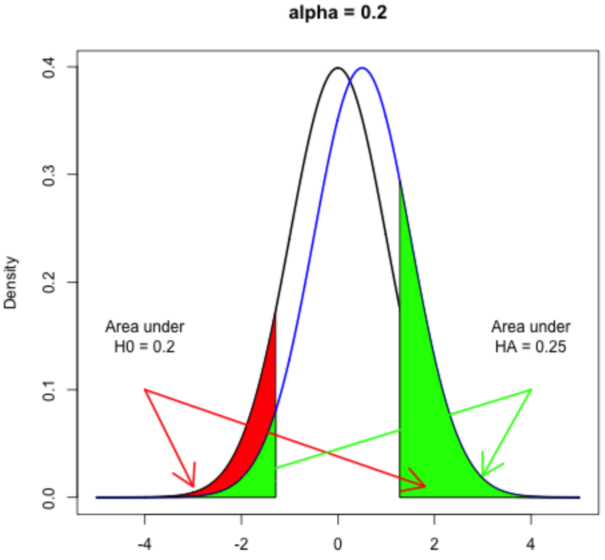
Effects of Increasing Alpha



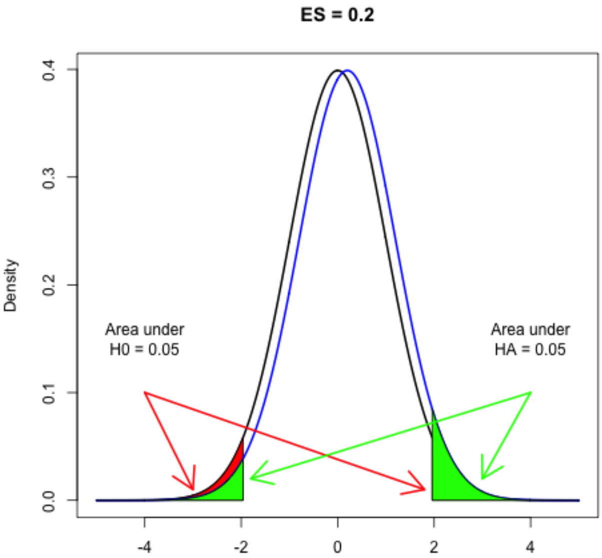
Effects of Increasing Alpha



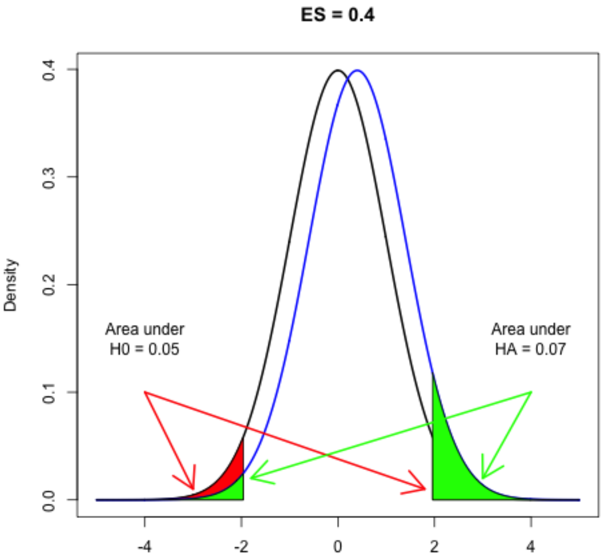
Effects of Increasing Alpha



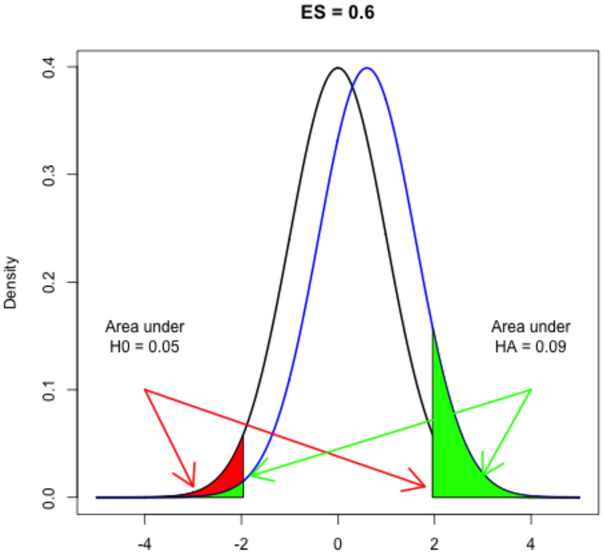
Effects of Increasing Effect Size



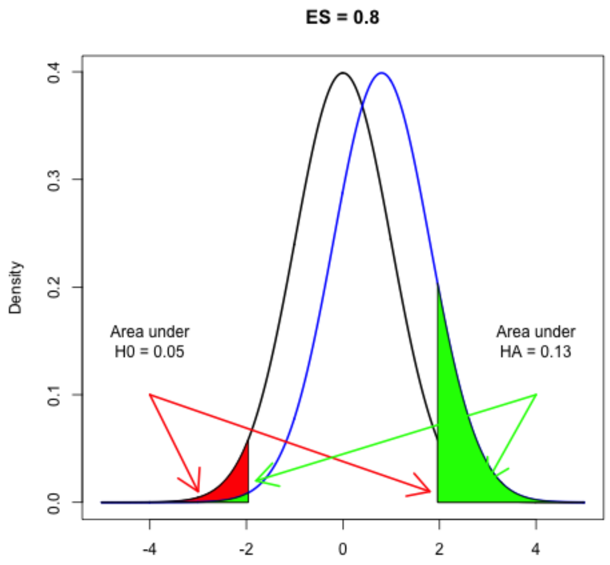
Effects of Increasing Effect Size



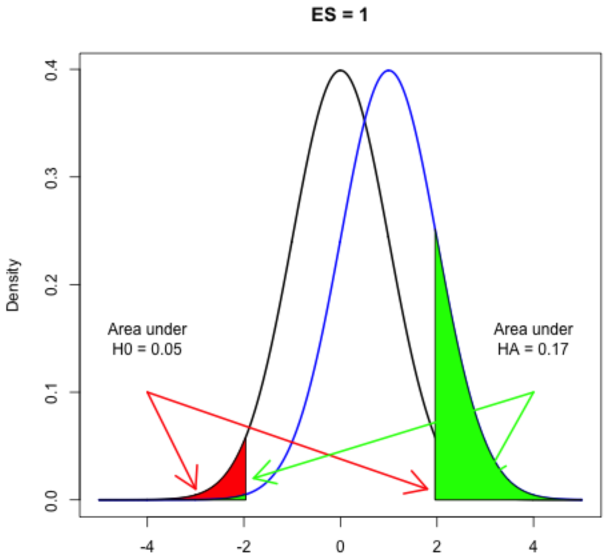
Effects of Increasing Effect Size



Effects of Increasing Effect Size



Effects of Increasing Effect Size

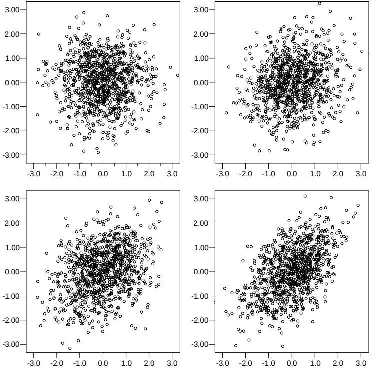


Statistical Power Analysis

Adam Davey

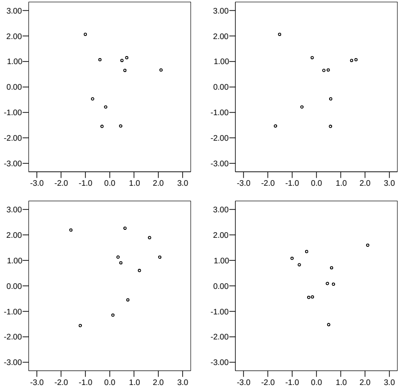
Discerning Patterns: Large N

- Clockwise: (None, Small, Large, Moderate, $N = 1000$)

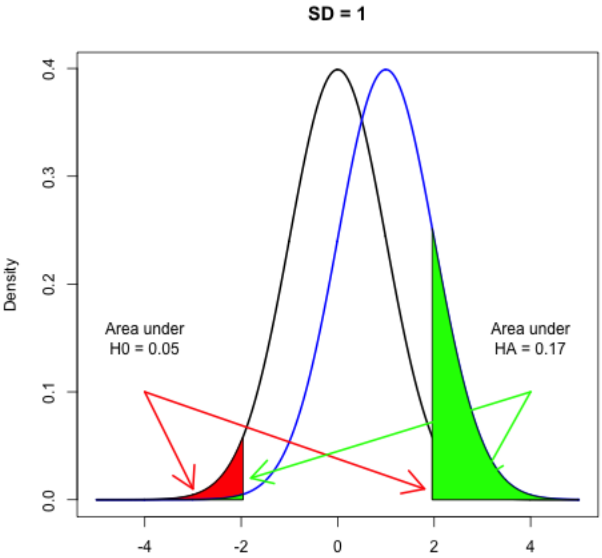


Discerning Patterns: Small N

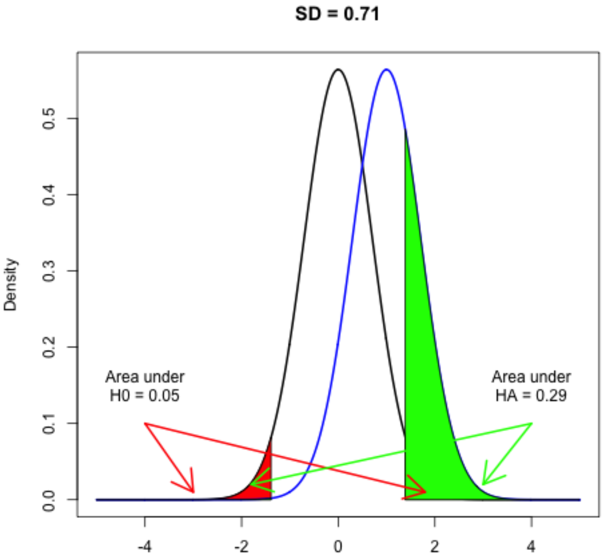
- Clockwise: (None, Small, Large, Moderate, $N = 10$)



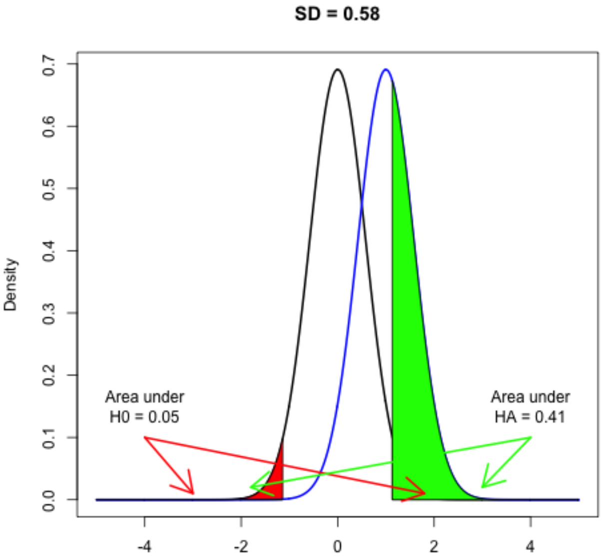
Effects of Increasing N or Decreasing Variance/SE



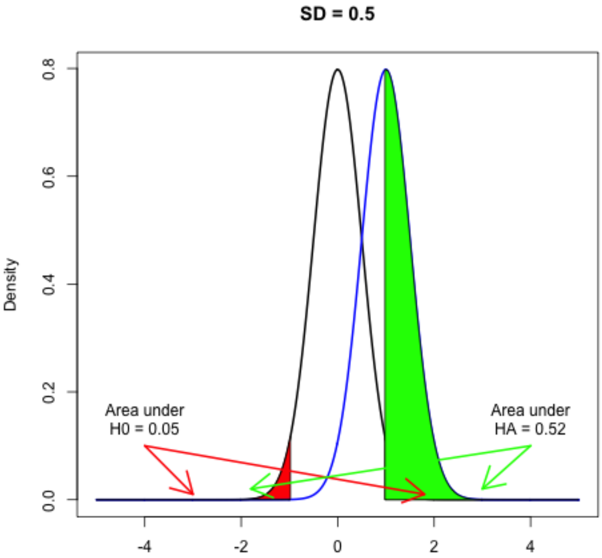
Effects of Increasing N or Decreasing Variance/SE



Effects of Increasing N or Decreasing Variance/SE



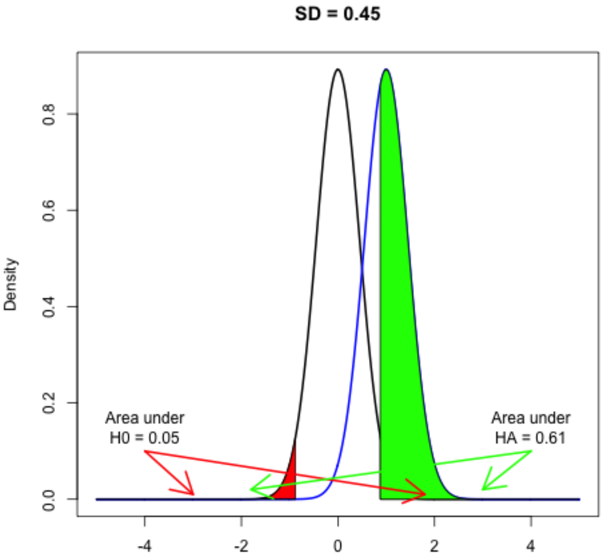
Effects of Increasing N or Decreasing Variance/SE



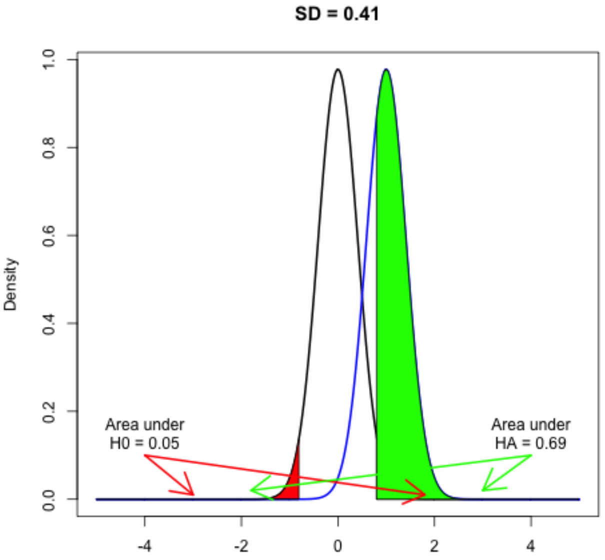
Statistical Power Analysis

Adam Davey

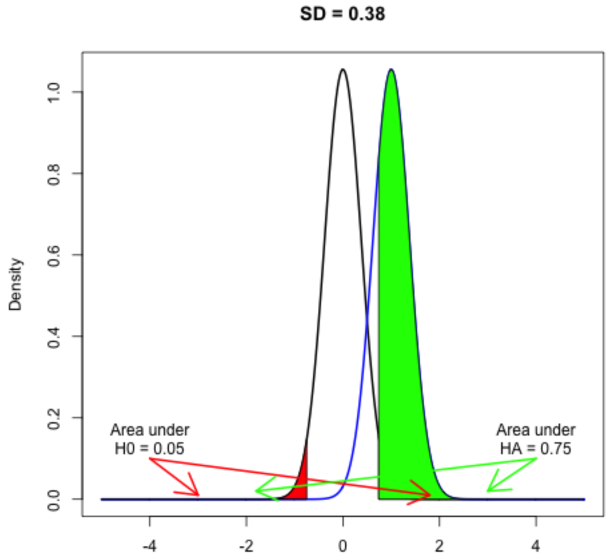
Effects of Increasing N or Decreasing Variance/SE



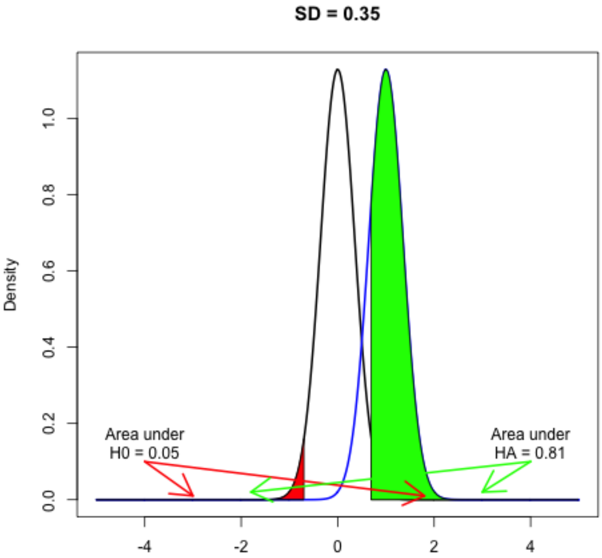
Effects of Increasing N or Decreasing Variance/SE



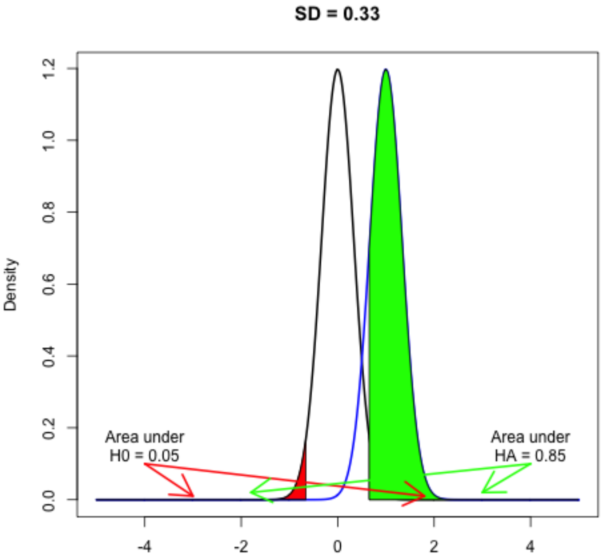
Effects of Increasing N or Decreasing Variance/SE



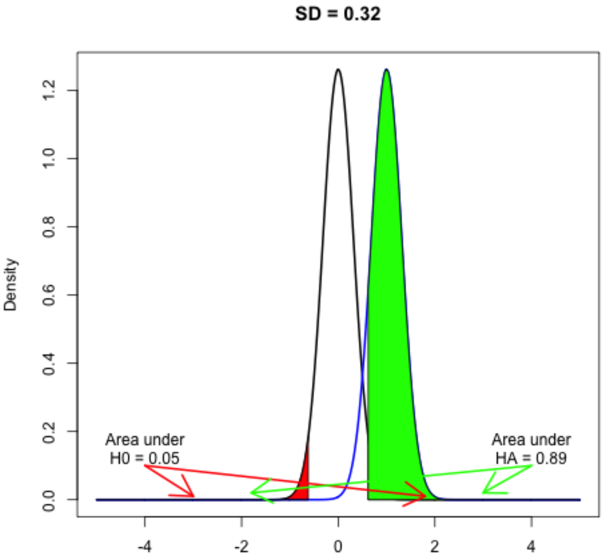
Effects of Increasing N or Decreasing Variance/SE



Effects of Increasing N or Decreasing Variance/SE



Effects of Increasing *N* or Decreasing Variance/SE



Summing Up

- Power of hypothesis test with significance level α is probability we reject null hypothesis when the alternative is true
- Power is probability that data gathered will be sufficient to reject null hypothesis when it is false
- Power is of *critical importance*

Summing Up

- Uses of power
- *A priori:* When designing study, select a sample size large enough to detect and effect of magnitude you believe is meaningful
- *A posteriori:* When test finds no significant difference/association, was there enough power to detect effect of meaningful magnitude?
- (Too little, too late. Can still be used to properly power next study.)
- See http://www.ats.ucla.edu/stat/seminars/Intro_power/ for more.

Statistical Power of a Test

- Significance testing is a balancing act
  - Chance α of making Type I error
  - Chance β of making Type II error
  - Reducing α increases β, and thus reduces the power of a test. It might be tempting to emphasize greater power (the more the better)
- With "too much power" statistical significance may be clinically inconsequential
- A Type II error is not definitive since a failure to reject the null hypothesis does not imply that the null hypothesis is correct
- Since H_0 is either always true or false, we are only in danger of making one kind of error or the other (but we have no idea which one)

Factors Affecting Power

- Size of effect an important factor in determining power. Higher probability of detecting larger effects
- More conservative significance levels (lower α) yield lower power. Less power with $\alpha = .01$ than with $\alpha = .05$.
- Increasing the sample size decreases the spread of the *sampling distribution* and increases power, but there is a trade-off between gain in power and the time/expense of testing a larger sample
- Larger variance $\left(\sigma^2\right)$ implies a larger spread of the sampling distribution, $(\sigma/\sqrt{N})$. The larger the variance, the lower the power.
- Variance is partly a property of the population, but can be reduced through careful study design.

Power with Fixed Sample Size

- Many times, N is fixed, either by resource constraints or with secondary data analysis
- In this context, power analysis serves a different function
- Minimum detectable effect (MDE)
- What is the smallest effect size I can detect with power $= (1 - \beta)$, sample size $= N$, and alpha $= \alpha$?
- (Stata Users: `db power`)

Power with Fixed Sample Size

- Accuracy in parameter estimation (AIPE: http://www.ats.ucla.edu/stat/stata/dae/aipe.htm)
- Bracketing effect sizes (half-width, w). For sample size N, find range that give p% chance that the estimated interval will be $\leq 2 \times w$
- The AIPE paradigm is a framework for managing width of confidence interval, independent of effect size)
- (Stata Users: `findit aipe`)

Statistical Power Analysis
Adam Davey
Overview
Motivating Problem
Hypothesis Testing
Noncentral Distributions
Factors
Intuitions
α
ES
N
Variance
Summary
Resources

Additional Resources for Power (Books)

- (Too) Simple
  Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155.
  http://classes.deonandan.com/hss4303/2010/cohen%201992%20sample%20size.pdf
- Just Sufficient
  Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
  http://www.lrdc.pitt.edu/schneider/P2465/Readings/Cohen,%201988%20(Statistical%20Power,%20273-406).pdf

Additional Resources for Power (Books)

- More Contemporary
  Murphy, K. R., Myors, B., & Wolach, A. H. (2009). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Routledge.
- Extensions
  Davey, A. (2009). *Statistical power analysis with missing data: A structural equation modeling approach*. Routledge.
  Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599-620. http://www.statmodel.com/bmuthen/ED231e/RelatedArticles/Article_097.pdf

Additional Resources for Power (Software)

- G*Power
  http://www.gpower.hhu.de/en.html
- R Package `pwr`
  http://www.statmethods.net/stats/power.html
  http://cran.r-project.org/web/packages/pwr/pwr.pdf
- R Package `powerMediation`
  http://cran.r-project.org/web/packages/powerMediation/powerMediation.pdf