Power Analysis

Ben Kite
KU CRMDA
2015 Summer Methodology Institute

Created by Terrence D. Jorgensen, 2014
Recall Hypothesis Testing?

• Null Hypothesis Significance Testing (NHST) is the most common application in social science
  – Frame research hypothesis as an “alternative” ($H_1$) to a “null” hypothesis ($H_0$) that is given preference
  – Design study to test $H_0$, collect data
    • Reject $H_0$ when data are uncommon if $H_0$ is true
    • If you fail to reject $H_0$, you can’t reject $H_0$ as a plausible explanation for the observed data
Examples of $H_0$

• Effect of wealth on electricity demand is $\beta_1 = 7$
  
  Electricity Demand = $\beta_0 + \beta_1 \text{Wealth} + \epsilon$

  – Estimate from data is $\hat{\beta}_1 = 10$
  
  – Is 10 far enough from 7 for $H_0$ to be rejected?

• Gender difference is $\mu_{\text{Men}} - \mu_{\text{Women}} = \mu_{\text{diff}} = 0$

  – Estimate is $\hat{\mu}_{\text{diff}} = -5$

  – Is the observed difference big enough to convince us that $H_0$ is untenable?
Sampling Distribution

• Estimates vary from sample to sample
• If $H_0$ is true, would it be terribly unusual to observe the data (i.e., the statistic) we observed? ($p$ value)
• Any observations in the critical region beyond the red confidence limits are sufficient evidence to reject $H_0$
What Is Statistical Power?

• The probability of rejecting $H_0$, on the condition that it is FALSE
  – Only makes sense in the context of NHST

• Affected by 4 factors
  – Rejection criterion ($\alpha$ level)
  – Sample size ($N$)
  – Sampling variability ($SD$, $\sigma^2$)
  – Effect size (the degree to which $H_0$ is false)
What If Result Is Not “Significant”?

• $H_0: \beta = 4$  
  From data: $\hat{\beta} = 5$, $SE = 0.6$

• Estimated sampling distribution of $\beta$ is not different enough from $H_0$ to rule it out
  – Does that mean that 4 was the true value of $\beta$?
  – No, it just means we can’t reject $H_0: \beta = 4$

• If $H_0$ were wrong, what could you change to reject it?
  – Collect more data
  – More liberal criterion ($\alpha = .10$)
  – Change $H_0$ to $\beta = 2$. We can reject that!
Motivation Behind Power Analyses

• Important part of research proposals
  – How many cases are required to reject your $H_0$?
  – Funding agencies & dissertation advisors want to make sure they aren’t wasting time & money

• Think backwards
  – Imagine a completed study, with data
  – MUST write down the actual model to be estimated
  – With “made up data” of size $N$, using carefully chosen population parameters, how often is a “significant” effect detected?
  – If not, how large must $N$ be to detect the effect at least as often as a minimum threshold?
Real-Life Research Example

• Researcher collects data on $N = 10$ people to find out whether tobacco causes cancer
  – Statistical procedure says there’s no relationship, so we can’t reject $H_0$ of no relationship
  – Suppose the effect of tobacco on cancer risk is actually present, but we missed it by not collecting enough data

• 80% is a customary threshold for “enough” power
  – We should design experiments so the power $\geq 0.8$
    • Measure variables with little variance; collect large $N$

• Effect must be “large” if it is to be detected with small $N$
  – If effect is “small,” then we increase $N$ to increase chances of finding a “significant” result (i.e., of rejecting $H_0$)
How Much Power Do We Need?

• 80% is a customary threshold for what is considered “enough” power
  – We should design experiments so the power $\geq 0.8$
    • Measure variables with little variance
    • Collect large $N$

• To detect with small $N$, effect must be “large”
  – If effect is “small,” then we increase $N$ to increase chances of finding a “significant” result (i.e., of rejecting $H_0$)
Effect Sizes

• Raw effect sizes are just the parameter estimate minus the null hypothesized value
  – Regression slopes ($\hat{\beta} - \beta_0$)
  – Mean-differences between groups ($\hat{\mu}_{\text{Diff}} - \mu_0$)
  – Often can divide difference by $SE$ for a $t$ statistic

• Let’s look at the R syntax
  – Continuing the example from this morning’s workshop on Monte Carlo Simulation
    • See PowerAnalysis-01.R (or accompanying HTML file)
Effect Sizes

• Effect Size = magnitude of difference between a parameter estimate and its \( H_0 \) value (e.g., \( \hat{\mu} - \mu_0 \))

• APA requires “standardized” effect sizes
  – Seeking a number that is generic across contexts
  – Supposed to represent “practical” significance, but effects in units of SD or proportions are not always intuitive or useful

• Cohen (1988) pioneered the most frequently used criteria for describing effect sizes and estimating power among social scientists
  – Back to R! (see also G*Power)
Monte Carlo Power Analysis

• A Monte Carlo study where:
  – The outcome of interest is statistical power
  – The main manipulated factor is $N$

• Useful because analytical methods only cover simple cases
  – Power = the proportion of samples in a condition for which $H_0$ was rejected

• Can manipulate other factors
  – Effect size, alpha, variability, missing data, etc.
Free Power Analysis Resources

• G*Power ([http://www.gpower.hhu.de/en.html](http://www.gpower.hhu.de/en.html))
  – Linear Models (regression, correlation, t test, ANOVA, ANCOVA, MANOVA, MANCOVA)
  – Some generalized linear models (Poisson or logistic regression)
  – Contingency tables ($\chi^2$, McNemar’s test)
  – Proportion tests
  – The user’s manual on the website is easy to read (lot’s of pictures and easy instructions)
Free Power Analysis Resources

- Multilevel Modeling power analysis software
  - Optimal Design (http://sitemaker.umich.edu/group-based/optimal_design_software)
    - Comprehensive, graphical, like G*Power for MLM
  - PINT (http://www.stats.ox.ac.uk/~snijders/multilevel.htm#progPINT)
    - Uses analytical approximation, 2-level models only
  - MLPowSim (http://www.bristol.ac.uk/cmm/software/mlpowsim/)
    - Makers of MLwiN (among the best MLM software)
    - You input characteristics of your data (summary stats of predictors, sample size at each level) and population parameters, then MLPowSim writes an R script for Monte Carlo simulation-based power analysis
CRMDA Resources

• For SEMs (and more), see KUant Guide #12: Monte Carlo Simulation in Mplus
  – See http://crmda.ku.edu/kuant-guides
  – This is primarily SEM software (not free), but it can also be used for anything that can be framed as a
    • Linear model (t test, ANOVA, regression)
    • Generalized linear model (Poisson or logistic regression)
    • Multilevel / mixed-effects model
  – Just need to know how to write model in Mplus syntax
• Example provided at bottom of today’s R syntax