Power Analysis

Ben Kite and Terrance Jorgensen

KU CRMDA

2017 Stats Camp

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 $\beta_1 = 7$

  $$\text{Electricity Demand} = \beta_0 + \beta_1 \text{Wealth} + \epsilon$$

  – 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?

What Is Statistical Power?

- The probability of rejecting H_0, on the condition that it is FALSE (1 – Type II error)
  - Only makes sense in the context of NHST
  - Conduct before data collection (avoid *post hoc*)
- Affected by 4 factors
  - Rejection criterion (α level)
  - Sample size (*N*)
  - Sampling variability (*SD*, σ^2)
  - Effect size (the degree to which H_0 is false)

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 (Type II error)
- 80% is a customary threshold for "enough" power
  - We should design experiments so the power ≥ 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)

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 ($\hat{\mu}_{\mathrm{Diff}} - \mu_0$)
  - 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

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)
  - Linear Models (regression, correlation, *t* test, ANOVA, ANCOVA, MANOVA, MANCOVA)
  - Some generalized linear models (Poisson or logistic regression)
  - Contingency tables (χ^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

- WebPower (http://webpower.psychstat.org/wiki/)
  - Correlation, regression
  - Proportion/Mean differences
  - Mediation
  - Multilevel and Longitudinal modeling
  - Structural equation modeling
  - Fairly new, may have bugs

Free Power Analysis Resources

- Multilevel Modeling power analysis software
  - 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 M*plus*
  - 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 M*plus* syntax