Balance Check in RCTs: Hansen & Bowers

Quantitative Workshop

October 18, 2016

My Problem

- I’ve run two experiments in which the unit of randomization is higher than the unit of observation (clustered)
  - ALEKS: Randomly assigned dates to T/C for students who enter a testing center over the summer
  - Louisville: Randomly assigned classrooms to T/C
  - Student level tests/survey results are unit of observation
- Both employ blocking
  - ALEKS: Day of the week (M-F)
  - Louisville: High schools
- What’s the appropriate way to conduct balance analysis (and outcome analysis) for these studies?

Why do we conduct balance checks?

- Bad randomization is possible
  - Experimenter error
  - Bad luck (bigger problem in smaller samples)
- The goal is to ascertain if randomization was successful in generating equivalent groups
- Balance check on observables is a poor test
  - Even if it passes this low bar, it doesn’t mean you have balanced on unobservables
  - Can never prove good randomization
- It’s expected and could consider rerandomization
- Hansen & Bowers: How should we conduct balance checks with complex RCT designs (blocked/clustered)?

Two Ways Not to Do This

1. Ignore clustering
    - Take simple mean difference between treatment and control groups on a covariate and determine p-value of the difference
    - Provides a separate p-value for each baseline variable
    - This is common practice with simple random sampling
    - With clustering however, it provides inaccurate p-values (Table 2 simulations), especially in small samples with small cluster sizes
        - variance in cluster sizes
        - intraclass correlations

Two Ways Not to Do This

2. Logistic regression of treatment on all covariates
   - This is also common for simple random assignment
   - Provides a single omnibus assessment of all covariates instead of separate p-values for a test of each
   - Use likelihood ratio test to compare fit of logit t x_1 x_2…x_k to fit of logit t on a constant (or include blocking variables for blocked study)
   - With small samples, this process overestimates type I error rates by producing p-values that are too small (Table 3/Figure1)
   - Must have many times as many observations as variables

Hansen & Bowers Proposed Solution

- Aggregate covariates up to the cluster level and conduct mean T/C difference test at randomization level with adjustment for cluster sizes and # of clusters
  - This is the d(**z**, **x**) difference in the paper on page 8
- With blocking, include block level weights based on size of block or their proposed optimal weighting (Section 5)
- Can calculate manually or use their R package
- This calculates a p-value on each individual covariate
- Section 4 creates a single test statistic, $d^2(\mathbf{z}; \mathbf{x}_1, \ldots, \mathbf{x}_k)$ for linear combinations of all covariates which can be compared to a χ^2 distribution.

Implications and Questions

- Can make more optimal blocking decisions by minimizing the variance of d(**Z**, **x**)
- We could also estimate the treatment effect this way using d(**z**, **y**)
- How would this method compare to a multilevel modeling approach?