

Advanced Methods in Impact Assessment Workshop

Day 2: Constructing the Appropriate Counterfactual: Randomized Control Trials & Matching

Today we will analyze the results from an RCT in which the counterfactual for the treatment group is a randomly selected control group. We will also construct control groups using matching techniques and compare the results to those from the RCT. For these exercises we will use data from a real RCT on the effects of marketing in encouraging households to purchase index insurance. This RCT was conducted in conjunction with ICRISAT, again in India.

First, some background on the RCT. The data comes from a paper by Shawn Cole, Xavier Giné, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery titled “Barriers to Household Risk Management: Evidence from India.” It was published in the *American Economic Journal: Applied Economics* in 2013. The RCT was “designed to estimate the slope of the demand curve for rainfall insurance and to determine the sensitivity of demand to a range of non-price factors including trust, liquidity constraints, and framing effects.” In brief, the researchers randomly selected 700 households from a sample of 1,047 households to receive treatment, where treatment is information about weather insurance. Note that the 1,047 households were selected using stratified random sampling from census data on over 7,000 households in Andhra Pradesh. Of the 700 households only 660 made it into the treatment group because 40 households could not be located. They then consider the number of households that purchase insurance as their outcome.

1. *Experimental Design:* What do you think of this experimental design? How does the treatment inform their research question? Are there limitations?
2. *Sample Design:* Are those in the treatment group representative? Representative of what? Should we be concerned about the missing 40 households?

Not only did the authors randomly assign treatment, they randomized over the content of the treatment. First, they randomly assigned whether an ICRISAT insurance educator received an endorsement from the local BASIX LSA. BASIX is a microfinance institution with an extensive rural network of local agents known as Livelihood Services Agents (LSAs). Two-thirds of villages were endorsement-eligible. Within these villages, the LSA endorsed the educator for half the visits. The goal of this treatment was to measure how trust in the insurance provider influences demand. Second, they offered a random amount of cash compensation for the household’s time, between ₹25 and ₹100. The goal of this treatment was to test the sensitivity of insurance demand to liquidity constraints. Third, they randomized whether the household received additional insurance education. The goal of this treatment was to measure if knowledge about the value of the insurance influences demand. See the Cole paper for more details of the experimental design.

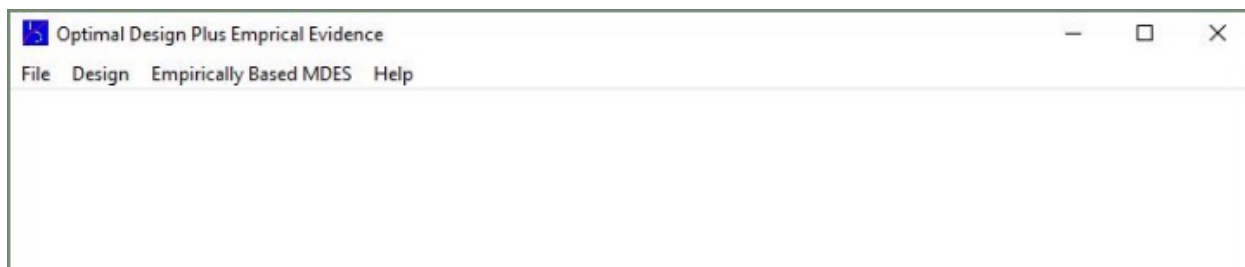
There are three objectives for today’s exercises:

1. Determine the minimum detectable effect size and the minimum sample size for conducting an RCT.
2. Conduct analysis of RCT data to determine if the sample was balanced and to measure the ATE.
3. Construct alternative control groups to the randomized control group using a variety of matching techniques.

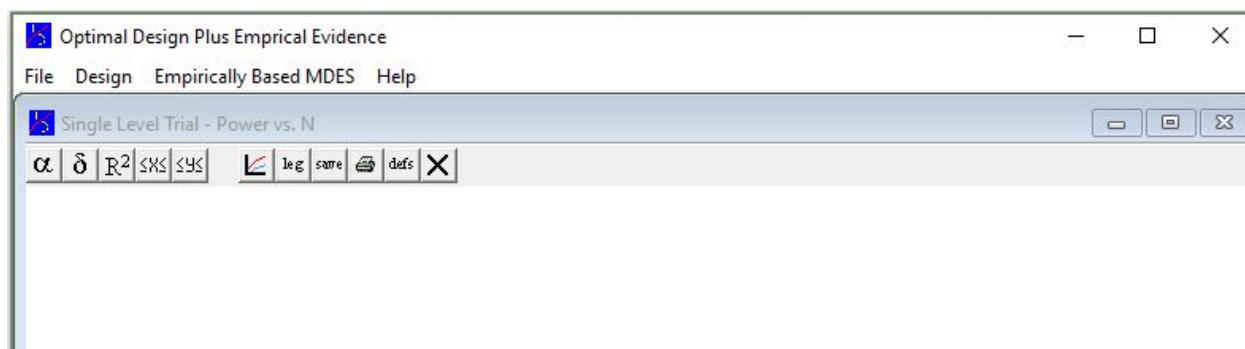
Analysis of Data from a Randomized Control Trial

First, we’ll go through the steps to calculate the minimum detectable effect size given a sample and the minimum sample size to achieve a detectable effect. While **Stata** has had the ability to do power calculations since version 12, I find the set-up unintuitive and limited. Thus, we will use the software Optimal Design (OD) to conduct power calculations.

Open up OD and you should see a fairly basic window with drop down menus as below:



We will start by determining the necessary sample size for a single level trial. This approach is appropriate when you want to test the effect of some intervention, measuring the difference in your variable of interest for treatment and control. Select the Design → Person Randomized Trials → Single Level Trial → Power vs. Total Number of People. This will open up a new window similar to below:



Here we have several buttons as options.

- α – Sets the significance level, or Type I error rate. By default, it is set to 0.05.
- δ – Sets the minimum detectable effect size measured in standard deviations. By default, it is set to 0.20.
- R^2 – Sets the estimated proportion of explained variance in our regression. The default is zero.
- $\leq x \leq$ – Sets the range of the x-axis (the number of subjects).
- $\leq y \leq$ – Sets the range of the y-axis (the power).

The remaining buttons allow you to edit and save the graph.

First, select the δ button and set $\text{delta}(1) = 1.000$, $\text{delta}(2) = 0.500$, and $\text{delta}(3) = 0.250$. These are equivalent to a one-standard deviation change, a half-standard deviation change, and a quarter-standard deviation change. Once you click “OK” the OD software will immediately draw a graph of the relationship between power and sample size for the level of confidence and effect size that we set. Using your mouse to click along the trajectory allows you to determine specific sample size for a given power. You may need to adjust the x-axis to read off the numbers.

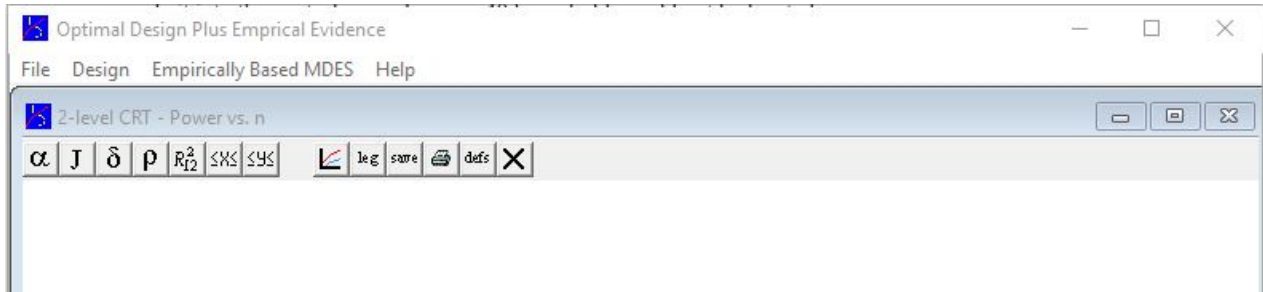
3. What are the necessary sample sizes for detecting a one, half, and quarter standard deviation change in our variable of interest?
4. What is the new sample sized needed to detect a half-standard deviation change if we can include covariates in the regression which explain 20% of the total variance (i.e., and $R^2 = 0.64$)?

Next, we'll consider the necessary sample size for a clustered RCT with individual or household level outcomes. Cluster randomization is much more common than person randomization, especially in agricultural economics. Our interventions are often at the village level to help avoid spillover effects, with everyone in the village assigned to the same group. We can either look at village level outcomes or person level outcomes, depending on the intervention. In the case of the Cole et al. (2013) study,

Stata Version

randomization was at the household level, while some sub-treatments were randomized at the village level. Outcomes (purchase of insurance) was measured at the household level.

Select the Design → Clustered Randomized Trials with person-level outcomes → Clustered Randomized Trials → Treatment at Level 2 → Power vs. Cluster Size. Now we have two new buttons.



- J – Sets the number of clusters, requires a minimum of 2.
- ρ – Sets the intra-class correlation (ICC), or the ratio of the variability between clusters to the total variability.

Set $J = 37$, which is the number of villages in Cole et al. (2013). Set $\delta(1) = 0.500$ and $\delta(2) = 0.250$. Finally, set $\rho(1) = 0.000$ and $\rho(2) = 0.200$.

5. If there is no intra-class correlation, how many total subjects are needed to achieve a MDE of a half-standard deviation with a power of 0.80?
6. How many total subjects are needed if the ICC is 0.20 with an MDE of a half-standard deviation and a power of 0.80?

Now that you have a sense of how to conduct power calculations, we can return to **Stata** and begin to analyze the RCT data. To get started, load into **Stata** the data set `Index_Insurance_data.dta`. Open a `.log` file called `data_exercise_2.smcl` and write your RCT and matching code in a `.do` file so you can reference it later. Note that the variable which measures if a household was in treatment or control is `d_visit`.

7. Examine the data to see if randomization worked. Compare the means of `wealth_indexi` and `sexheadi` between treatment and control using the `ttest`. Based on the results, do you believe those who did not receive the treatment are an appropriate control group for those who received the treatment?
8. Conduct a `ttest` to compare the mean number of insurance purchases by households who received a visit with the mean insurance purchases of the control households. What does the test indicate? Is this estimate the intention to treat effect, the effect of treatment on the treated, or the average treatment effect? Explain.
9. Run a regression including only the `d_visit` variable and if the household purchased insurance as the outcome. What is the result? Under what circumstances is this regression unbiased?
10. Now add household control variables `DK_basixi` `wealth_indexi` `logpce_new_wi` `riskavl_jul06i` `norm_exp_rainMay06` `lcultirrpcti` `mean_payouts` `buy_ins04i` `ins_otheri` `buy_newi` `group_addi` `sched_ct` `muslim` `sexheadi` `lage_headi` `lhsizei` `d_highedu` `ins_skilli` to the regression you just did. How does this change the result and why?

In this study, households were randomly selected to receive a visit to discuss insurance. But, 2/3 of villages were selected to receive an additional treatment – that of an endorsement by a trusted individual. In those “endorsement eligible” villages, half the households were randomly selected to receive the endorsement.

11. How might this randomization structure change our results? Regress the outcome variable on `d_visit`, `endors_LSA`, and the interaction of these two variables. Run the same regression but add the control variables. Did the endorsement have any effect on insurance purchases?

Propensity Score Matching

When we match we first want to check if the balancing property is satisfied. To do this, we will need the `pscore` command. You can find it by typing `findit pscore` into your command window. When the results come up, click on the newest update for the `pscore` package: `st0026_2`.

```

search for pscore

-----
Keywords: pscore
Search: (1) official help files, FAQs, Examples, SJs, and STBs
       (2) Web resources from stata and from other users


-----
Search of official help files, FAQs, Examples, SJs, and STBs

SJ-8-4 st0136_1 . . . Erratum and discussion of propensity-score reweighting
      . . . . . A. Nichols
      Q4/08 SJ 8(4):532--539 (no commands)
      discusses propensity-score reweighting

SJ-7-4 st0136 . . . . . Causal inference with observational data
      . . . . . A. Nichols
      Q4/07 SJ 7(4):507--541 (no commands)
      discusses problems with inferring causal relationships
      from nonexperimental data and describes four broad classes
      of methods designed to allow estimation of and inference
      about causal parameters

SJ-5-3 st0026_2 . . . . . Software update for pscore suite
      . . . . . S. O. Becker and A. Ichino
      (help attk, attnd, attnw, attr, atts, and pscore if installed)
      Q3/05 SJ 5(3):470
      bug fix for pscore

SJ-4-2 st0026_1 . . . . . Software update for pscore suite
      . . . . . S. O. Becker and A. Ichino
      (help attk, attnd, attnw, attr, atts, and pscore if installed)
      Q2/04 SJ 4(2):224
      pscore now allows for bootstrap options in Stata 8 and for
      identification of matched pairs in nearest neighbor matching
  
```



Then install the package, by clicking install


```

-----
package st0026_2 from http://www.stata-journal.com/software/sj5-3
-----

TITLE
  SJ5-3 st0026_2. Estimation of average treatment effects

DESCRIPTION/AUTHOR(S)
  Estimation of average treatment effects
  by Sascha O. Becker, University of Munich
  Andrea Ichino, EUI
  Support: so.b@gmx.net, andrea.ichino@iue.it
  After installation, type help attk, attnd, attnw,
  attr, atts, and pscore

INSTALLATION FILES
  st0026_2/attk.ado
  st0026_2/attk.hlp
  st0026_2/attnd.ado
  st0026_2/attnd.hlp
  (click here to install)
  
```



Once the package is installed, carry out the `pscore` command. In the first stage, use the `logit` option. Calculate the propensity score as follows, specifying that **Stata** should restrict the analysis of the balancing property to all treated plus those controls in the region of common support (`comsup`).

```

pscore d_visit DK_basixi wealth_indexi logpce_new_wi riskavl_jul06i
norm_exp_rainMay06 lcultirrpcti mean_payouts buy_ins04i bua_newi group_addi
sched_ct muslim sexheadi lage_headi lhssizei d_highedu ins_skilli, pscore
(pscore1) logit comsup
  
```

12. What does it mean to say that the sample is balanced? Is your sample balanced? Are any of the control households outside the region of common support?

Stata Version

Now carry out the second stage of the propensity-score-matching estimator by conducting t-tests. Do this using the commands that start with `att`. Calculate the ATT first using nearest neighbor matching (command is `attnd`), then radius matching (`attr`), then kernel matching (`attk`—be sure to do bootstrap standard errors, using 100 repetitions). The code for the t-test, with bootstrapping is:

```
attk ins_lev d_visit DK_basixi wealth_indexi logpce_new_wi riskav1_jul06i  
norm_exp_rainMay06 lcultirrpcti mean_payouts buy_ins04i bua_newi group_addi  
sched_ct muslim sexheadi lage_headi lhssizei d_highedu ins_skilli,  
pscore(pscore1) comsup bootstrap reps(100)
```

13. What are your results for the three t-tests? Are they similar?
14. How do the different ATT values compare to your results to the simple t-test from Question 7?
15. Compare these outcomes to those calculated earlier in the regression in Question 9. Does the PSM reproduce the outcomes of the randomization?

Remember that some villages were selected to receive an additional treatment and others did not.

16. Should you do the propensity score match within villages that had the treatment option or by comparing households that received the endorsement treatment with comparable households in the same village but that did not receive the endorsement treatment? (The **Stata** programs you just ran do not distinguish between households within villages with programs and households within villages without programs. They find the nearest neighbor regardless of village). Explain.