

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 data 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 Gine, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery title “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 to receive treatment from 1,047 households. 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 control group because 40 households could not be located.

1. What do you think of this experimental 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 Rs 25 and Rs 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

To get started, load into Stata the data set `Index_Ins_data.dta`. Open a `.log` file and write your RCT and matching code in a `.do` file so you can reference it later.

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. Since Stata 12 the process for doing power calculations has changed. These instructions outline how to do power calculations in Stata 12. If you have Stata 13 or 14 there is now an easier way to do power calculations. In these newer versions, to conduct power calculations select “Power and sample size” on the “Statistics” dropdown menu. There is good

documentation now on how to do power calculations by just typing `help power`. Alternatively, you can use the `power` command or the `sampsi` command as below.

For those without Stata 13 or later you will have to use the `sampsi` command. `sampsi` estimates require sample size or power of tests for studies comparing two groups. `sampsi` can be used when comparing means or proportions for simple studies where only one measurement of the outcome is planned and for comparing mean summary statistics for more complex studies where repeated measurements of the outcome on each experimental unit are planned.

We will start by determining the power and sample size for a one-sample test of mean. This is appropriate when you want to test the effect of some intervention, measuring the difference in your variable of interest for each subject before and after the intervention. This uses the `onesample` option for `sampsi`. For this we'll assume that our invention is some a visit to teach prospective buyers about insurance. Our outcome of interest is whether or not people ultimately bought the insurance. Using data from the VDSA we find that 30% of households have bought index insurance in the recent past. So, the null hypothesis is that 30% of our population, even without our treatment, would buy index insurance.

First we'll consider what sample size we'll need for a particular level of power. We'll want to reject the null with a level of power of 0.80 and a significance level of 0.05. Further, we have hypothesis some expected difference in the probability of purchasing the insurance, which we'll assume is 0.1 – we assume our intervention will increase the purchase of insurance from 30% of the population to 40% of the population, a fairly large increase. The code to run this is shown below:

```
sampsi 0.3 0.4, power(0.8) a(0.05) onesample
```

or for those of you with Sata 13

```
power oneproportion 0.3 0.4, power(0.8) a(0.05)
```

Next, we'll consider what power we would need for a particular sample size. We'll assume the same information as before, but instead of power, this time we have information about sample size. So, while the mean difference is 0.1 and the significance level of 0.05, this time we'll have a number of observations of 1000. The code to run this is shown below:

```
sampsi 0.3 0.4, n(1000) a(0.05) onesample
```

Now, let's determine the power and sample size for a two-sample test of equality of means. This is appropriate when you want to test the impact of some intervention on two different groups of people. We'll assume the same information as before, testing some intervention to encourage people to buy insurance. Our null hypothesis is again that the mean of individual differences for insurance purchase will be zero.

Now, we'll consider what the sample size needed is for a particular level of power. This is the same procedure, with the same information, as for the one-sided test – except that it is now a two-sided test. The code to run this is shown below:

```
sampsi 0.3 0.4, power(0.8) a(0.05)
```

Finally, we'll consider the resulting power for a particular sample size. Again, this uses the same information as above, but only for a two-sided test. The code to run this is shown below:

```
sampsi 0.3 0.4, n(1000) a(0.05)
```

Now that you have a sense of how to conduct power calculations, we can begin to analyze the RCT data.

2. Examine the data to see if randomization worked by comparing means using the `tabstat`. Based on the table of means for the treatment and control group that you generated do you believe those who did not receive the treatment are an appropriate control group for those who received the treatment?
3. Conduct a `ttest` to compare the mean number of insurance purchases by households who received a visit with that of control. 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
4. 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 a valid one to run?
5. Now add household control variables `DK_basixi` `wealth_indexi` `logpce_new_wi` `riskavl_jul06i` `norm_exp_rainMay06` `lcultirrpcti` `mean_payouts` `buy_ins04i` `ins_otheri` `bua_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.

6. 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. Under what circumstances is this regression a valid one to run? What else might we need to account for in the regression?

In addition to the visit treatment and the endorsement treatment the study included two more treatments randomly assigned to households. One was a random amount of cash compensation (`d_highreward`). The other was an additional amount of education about the value of the product (`ins_edu`).

7. Regress our outcome variable on each one of these treatments individually including household controls variables and village level effects. Are the coefficients on the treatment the ATE? How should we go about testing the overall effect of all of these various treatments?

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 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 a function of DK_basixi wealth_indexi logpce_new_wi riskavl_jul06i norm_exp_rainMay06 lcultirrptci mean_payouts buy_ins04i ins_otheri bua_newi group_addi sched_ct muslim sexheadi lage_headi lhsizei d_highedu ins_skilli.

Make sure to specify the common support option using comsup?

8. What does it mean to say that the sample is balanced? Is your sample balanced?
9. Use the detail option to see what variables in which blocks are balanced and which are not balanced. How should you address this situation?

It is often useful to graph the results of the propensity score calculation. Generate histograms for both the treated and non-treated groups. Use the options bin(50) color(bluishgray) fintensity(inten70) for the treatment group and the options bin(50) fcolor(none) for the control group.

10. What do you learn from the histogram of the propensity score by treatment status? Use the same command but now generate histograms for `wealth_indexi`. Also graph `ins_skilli`. And finally graph `ins_otheri`. What do you learn from each of these graphical comparisons?

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).

11. What are your results for the three t-tests? Are they similar?
12. What happens if you change the radius in the `attr` command from the default of 0.1 to 0.5? Why does this happen?
13. What happens if you change the kernel method used from its default of Gaussian to Epanechnikov?
14. Summarize your results and compare to your results to the simple t-test from Question 3. What is the effect of a visit from an insurance educator on a household's probability of purchasing insurance?
15. Remember that some villages were selected to receive an additional treatment and others did not. 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.
16. Compare these outcomes to those calculated earlier in the regression of `ins_lev` on `d_visit`, the full set of household control variables, and the village level effects. Does the PSM reproduce the outcomes of the randomization?

As a final exercise today we want to calculate Rosenbaum bounds. Rosenbaum bounds estimate the degree to which unobservables might affect both selection into treatment and outcome, biasing the estimate of treatment. Specifically, they produce an odds ratio (I') of how much could an unobserved variable bias outcome by affecting selection. To do this we need to install `Rbounds`. Then, re-run the above PSM matching using `psmatch2`, which should have been installed when you installed `pscore`.

```
bootstrap r(att): psmatch2 d_visit DK_basixi wealth_indexi
logpce_new_wi riskavl_jul06i norm_exp_rainMay06 lcultirrpti
mean_payouts buy_ins04i ins_otheri bua_newi group_addi sched_ct
muslim sexheadi lage_headi lhhsizei d_highedu ins_skilli, logit
common out(ins_lev)

generate att
gen delta = ins_lev - _ins_lev if d_visit==1

estimate bounds
rbounds delta, gamma (1 (1) 3)
```

17. What do these estimates tell you about the probability that treatment is affected by unobservables?