Center for Effective Global Action
University of California, Berkeley

Module 2.2: Randomized Assignment

Contents

List of Figures

1. INTRODUCTION

In the previous module we covered causal inference and counterfactual analysis, which are two key concepts used to conduct a rigorous impact evaluation. We also described selection bias and omitted variable bias and how randomization mitigates/eliminates this problem. In the next modules we will cover the various methods typically used when conducting an impact evaluation, including experimental and quasi-experimental methodologies.

This module will focus on randomized assignment. We will spend time making sure you understand how to actually analyze the impacts of a program, where to look, and what to look for. We will explain the various methods that one can use to find the impact, comparing when to use one over the other. Finally, we will walk through more advanced program designs where we may need to stratify on some existing variable (e.g. gender, age, occupation) or cluster at a higher level than the individual (e.g. school, market, village). These ways get at a more precise estimate of the impact, but are also relatively trickier to set up and implement.

At the end of this module, you should be able to:

- ✓ Determine how successful the randomization was (or wasn't)
- ✓ Conduct basic data analysis of an RCT
- ✓ Understand how and when to implement a stratified or clustered randomization

2. ANALYSIS FOR EVALUATING IMPACTS

2.1 Regression Analysis to Quantify Impacts

We can write the observed outcome Y for an individual I as,

$$Y_i = T_i . Y_{i_{trt}} + (1 - T_i) . Y_{i_{ctr}}$$

where T_i = 1 if the individual i is assigned to a treatment group and T_i = 0 if the individual is assigned to the control group, and Y_i is that individual's observed outcome. The *trt* and *ctr* are used to clarify that the individual can be either in treatment or control groups in "real life" analysis. Rearranging the terms

$$Y_i = Y_{i_{ctr}} + \left(Y_{i_{trt}} - Y_{i_{ctr}}\right) . T_i \text{ , and}$$

$$Y_i = E[Y_i]_{ctr} + \left(Y_{i_{trt}} - Y_{i_{ctr}}\right) . T_i + \left(Y_{i_{ctr}} - E[Y_i]_{ctr}\right)$$

which in usual linearity regression notation, assuming linearity, can be represented as,

$$Y_i = \beta_0 + \beta_1 . T_i + \varepsilon_i .$$

Based on the above regression model, we can estimate the conditional outcome with and without the treatment T and then estimate the causal effect as follows,

$$\boldsymbol{T_i = 1}: \qquad E[Y_i|T = 1] = \beta_0 + \beta_1 + \varepsilon_i$$

$$\boldsymbol{T_i = 0}: \qquad E[Y_i|T = 0] = \beta_0 + \varepsilon_i$$

$$E[Y_i|T = 1] - E[Y_i|T = 0] = \beta_0 + \beta_1 + \varepsilon_i - (\beta_0 + \varepsilon_i) = \beta_1.$$

Therefore, coefficient β_1 quantifies the impact as group mean difference in outcomes between the treatment and control group. Remember, estimates of β_1 obtained in this way are unbiased only if selection bias is 0.

Exercise: Open `PanelPROGRESA_97_99year.dta`. This dataset is a repeated cross-section of different waves of the ENCEL survey for March and October 1998 and November 1999. It also includes the baseline data collected in 1997. We have used this or part of this dataset previous modules as well. Also refer to `Module2.2 Learning Guide.do` file.

Let's assume that we expect PROGRESA (`D_HH`) to change the income of the household (`IncomeLabHH`) in 1999. Specify a regression model in STATA as discussed above and restrict the data to be used from year 1999 (variable `year`). What is the impact of the intervention on income levels? Is it statistically significant? How do you interpret the coefficient?

Answer Key: Specify regression model as, `reg IncomeLab_HH D_HH if year == 1999`. The STATA output is given in Figure 1. We find that PROGRESA participation by the household (D_HH) did not change the household income levels in 1999 statistically significantly at α = 0.1. Remember, these causal inferences are based on the assumption that the treatment was effectivelyrandomized by the study organizers.

```
. reg IncomeLab_HH D_HH if year == 1999
```

Source	SS	df	MS	Number of obs	=	18,370
				F(1, 18368)	=	0.16
Model	38331591.5	1	38331591.5	Prob > F	=	0.6908
Residual	4.4509e+12	18,368	242320693	R-squared	=	0.0000
				Adj R-squared	=	-0.0000
Total	4.4510e+12	18,369	242309588	Root MSE	=	15567

IncomeLab_HH	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
D_HH	94.88663	238.5733	0.40	0.691	-372.7393	562.5125
_cons	1898.628	144.1233	13.17	0.000	1616.133	2181.123

Figure 1. STATA output for OLS regression model to evaluate impacts

2.2 T-test Based Analysis

We have amply practiced t-test to detect difference between the two groups; for example, whether household assets value is different between highly-educated versus poorly-educated household heads (Module 1.3).

Here, we can extend t-test analysis to comparing the two groups which differ by their treatment assignment (T_i). As before, the null hypothesis is: the outcome of treated individuals is the same as it would have been had those individuals not been treated. The alternative hypothesis can be specified as a two-sided or one-sided (one is larger/smaller than other) comparison.

Remember, we are making an assumption that the individuals receiving the treatment are exactly like those not receiving the treatment because of the "missing data" problem. In other words, randomization and independence implies that there is no selection bias.

Exercise: Conduct a two-sided t-test to compare if household income is different by the treatment assignment. Are the results same as those from the regression analysis?

Answer Key: we can conduct a t-test in STATA as `ttest IncomeLab_HH if year == 1999, by(D_HH)`. We find that the magnitude (group mean difference) and significance of the causal effect is precisely the same as that in Figure 1 above. Indeed, OLS just performed a t-test for us.

```
. ttest IncomeLab_HH if year == 1999, by(D_HH)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	11,666	1898.628	83.78446	9049.497	1734.396	2062.859
1	6,704	1993.514	278.9093	22836.52	1446.764	2540.265
combined	18,370	1933.256	114.8499	15566.3	1708.139	2158.373
diff		-94.88663	238.5733		-562.5125	372.7393

```
    diff = mean(0) - mean(1)                                      t =  -0.3977
Ho: diff = 0                     degrees of freedom =    18368

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3454         Pr(|T| > |t|) = 0.6908          Pr(T > t) = 0.6546
```

Figure 2. t-test output to evaluate impact of PROGRESA on household income

2.3 A Decision to Make: Regression Models or T-test?

The above exercise demonstrates that causal effects are the same whether you use regression analysis or t-test. However, both of these analyses assumed that selection bias was zero. Was this the case? Whether randomized or non-randomized design is used, it is possible that the comparison groups were imbalanced at the baseline. What if the randomization is stratified by geography, for instance (we will discuss randomization strategies in a later module)? If that was the case, we would have to account for how we modeled the individuals in our analysis. What if treatments were

conducted at the village level, but we believe that individuals within a village are influenced by each other and share common facilities, so that there is some "correlation" among their behaviors? Then we have to estimate the standard errors clustered as the village level (refer to Module 1.3 to learn more). Also, we will later learn more robust methods, such as difference-in-difference, which provide additional robustness relative to regression analysis in some cases.

In general, impact evaluations often face problems that have to be "controlled for" or adjusted for in the analysis.

Regression methods give us the tools to do such adjustment which a simple t-test cannot. T-tests (or chi-squared tests if the outcome is categorical) alone are valid only if our groups are properly randomized, sample size is adequately large to achieve baseline balance in the two groups, data collection is unbiased, and you can reasonably adhere to the standard assumptions of t-test.

Exercise: As a demonstration, let's assume that poverty status of a household (`pov_HH`) is a confounder and we should control for it in the regression analysis. We also want to cluster the standard errors at the village level. We can most easily accomplish this using regression models in STATA as: `reg IncomeLab_HH D_HH pov_HH if year == 1999, cluster(villid)`. Notice that the effect size (magnitude of the coefficient) and the standard error are both changed markedly. This analysis provides additional evidence against the hypothesis that the PROGRESA program increased local income levels. However, this analysis is still very basic; we will see in a later module that PROGRESA actually did have significant impacts, though we have been unable to isolate them empirically.

```
. regress IncomeLab_HH D_HH pov_HH if year == 1999, cluster( villid )

Linear regression                               Number of obs     =     17,942
                                                F(2, 498)         =       5.34
                                                Prob > F          =     0.0051
                                                R-squared         =     0.0001
                                                Root MSE          =      15749

                                    (Std. Err. adjusted for 499 clusters in villid)
------------------------------------------------------------------------------
             |               Robust
IncomeLab_HH |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        D_HH |     229.035   324.9309     0.70   0.481    -409.3695    867.4394
      pov_HH |  -439.9662   138.2007    -3.18   0.002    -711.4945   -168.4379
       _cons |   2204.446   75.01622    29.39   0.000     2057.058    2351.833
------------------------------------------------------------------------------

.
```

Figure 3. OLS regression output with clustered standard error and controlling for poverty

3. RANDOMIZATION IN PRACTICE

Typically, there are important experimental steps to take before you randomize the treatment group. These steps include: (1) based on several assumptions, you estimate the sample size you would need to detect the impact you are interested in (covered in later modules); and (2) you select

a population which you can include in your experiment. This can be a purposive selection in discussion with intervention implementing agencies, a representative sample of the target population and anything in between. Here, we discuss how to randomly assign the people into the treatment and control groups. We will cover the quasi-experimental selection of control groups later.

3.1 Simple Randomization

Under a simple randomization framework each individual, household, or any other "unit of analysis" has equal chance of being selected in treatment or control groups; that is, a 0.5 probability of being part of each group. You can imagine flipping a coin and assigning the individual to the treatment group if the coin turns heads and to the control group if it is tails. You can conduct such a randomized allocation through STATA as demonstrated below.

- ✓ Open `PanelPROGRESA_97_99year.dta` if you have not already
- ✓ Generate a random variable which takes value of 0 or 1 as:

```
gen random_T = 0+int((1-0+1)*runiform()) if year == 1997
```

- ✓ You must perform randomization before the program or at the baseline (whether or not you conduct a baseline survey). However, in this example, we have data for 1998 and 1999 already so we restricted the randomization to 1997 observations only.

3.2 Stratified or Block Randomization

In some experiments it is practical or theoretically proper to randomize at some unit of grouping (called blocks or strata). For example, you may want to control for regional administrative, ecological, or political factors and maximize "exchangeability" by randomizing at an administrative district or block level. It is possible that you have selected your study participants in a 2-stage sample and now you want to randomize the treatment in the same manner. Or you are concerned with attrition (individuals leaving your sample before you conduct follow-up analysis) so that at the end of the study you will be faced with "un-exchangeable" groups. In this case, you can: (a) break the study sample in random groups of sizes (for example) 2, 4, and 6 villages each. Within each group, you can randomize the treatment – half to treatment and half to controls. This is done so that even if you lose controls in a "group" you can discard that group of 2, 4 or 6 villages but still be assured of exchangeability in the remaining groups. In short, there are strong reasons for stratifying the randomization. It can be conducted in STATA as follows:

- ✓ Suppose we want to stratify by federal administrative regions (`geopolid`)
- ✓ We will repeat what we did in 3.1 but just stratified as, `bysort geopolid : gen random_strata_T = 0+int((1-0+1)*runiform()) if year == 1997`
- ✓ Note, anytime you use a random generator (e.g. `runiform` function above), it is best to `set seed` to a fixed value so that you can always reproduce the results. `runiform()` creates a uniform distribution which is basically random number generation.

3.3 Clustered Randomization

This is a practical and popular strategy often employed in the development sector. Most development interventions or programs are not targeted at individuals but at some cluster: a village, an office department, or some other group of people. All individuals within the targeted cluster can be the intended customers/beneficiaries/target of the intervention, or there can be a "selection criteria" or eligibility criteria within the cluster, or there may be random selection of participants within the selected cluster for the program (the latter is termed "block randomization"). In STATA you can do cluster randomization as follows.

- ✓ Create an identifier that flags unique records at a cluster level. For example, suppose we want to randomize the villages then we should identify one unique observation for each village. We can do this as `egen uniqvill = tag(villid) if year == 1997`
- ✓ Now, randomly assign half of the villages to treatment group as follows, `gen random_cluster_T = 0+int((1-0+1)*runiform()) if year == 1997 & uniqvill == 1`
- ✓ Note, you can always combine cluster and stratified randomization as, `bysort geopolid : gen random_strata_cluster_T = 0+int((1-0+1)*runiform()) if year == 1997 & uniqvill == 1`

3.4 Testing the Success of Randomization

We discussed how the randomization assumption is based on large sample size to randomize the treatment so that we can assume that all measured and unmeasured confounders are distributed equally in the comparison groups. However, in reality we have to assess whether randomization was "successful" in achieving balance. The way to test for "success of randomization" is by evaluating the balance in measured variables between the treatment and control groups at the baseline, which should be the first table in any kind of report. Consider the following:

- ✓ We have to use baseline data because some of the measured variables can be affected by the intervention, and they can change differentially in the treatment and control group, so that comparing them after the commencement of a treatment might be biased

- ✓ Just because we find statistically significant difference at the baseline does not mean that the groups are imbalanced. For example, continuous measurements of age (years) and income (US Dollars) will likely be statistically different between the two groups even if the sample size of each is very large, because large sample size allows us to detect even very small differences in continuous variables. Therefore, we should assess only whether the difference is economically, biologically and logically large, not necessarily relying only on statistical significance.

- ✓ The converse of above is also true. Just because the difference is statistically insignificant does not mean that the groups are well balanced, if the differences have large magnitudes.

- ✓ It is best to compare the treatment and control groups on all available measurements for balance at the baseline so that you can be reasonably confident that the comparison groups are balanced, at least on observables. Best practice is to select these comparison variables "before" you randomize and faithfully check the group mean difference after the randomization at baseline. Plus, we can always add this covariate into our regression afterwards to check for any potential omitted variable bias!

For the dataset example we have been following so far, let's demonstrate how to check for the balance.

- ✓ Download and install the STATA command `ttable2` is you haven't already done so.
- ✓ Run the following command to check the balance for a few selected variables at the baseline. Note, `D` indicates whether the village was randomized to treatment group or not. `ttable2 IncomeLab_HH famsize eduhead sexhead agehead pov_HH if year ==1997, by(D)`
- ✓ Figure 4 is the output that shows how well the groups are balanced. We find that all the factors are balanced very well, but some of the differences are statistically significant.
- ✓ Note, you can use the `regress` or `ttest` commands also for each one of these variables separately and get the same results. STATA offers you several options for most kinds of analysis, and it is up to you which one to use.
- ✓ In the case of stratified or block randomization, you should evaluate the balance between each block or strata. You can do so by using the `by` or `bysort` options in most STATA commands.

```
. ttable2 IncomeLab_HH famsize eduhead sexhead agehead pov_HH if year ==1997, by(D)
```

Variables	G1(Control)	Mean1	G2(Treated)	Mean2	MeanDiff
IncomeLa~H	7995	1353.695	12333	1299.585	54.110**
famsize	9221	5.257	14856	5.196	0.061*
eduhead	9178	2.768	14774	2.783	-0.014
sexhead	9218	0.888	14852	0.890	-0.002
agehead	9202	47.056	14810	46.616	0.441**
pov_HH	9221	0.508	14856	0.528	-0.020***

Figure 4. Testing for baseline balance

4. BIBLIOGRAPHY/FURTHER READINGS

1. Duflo, Esther; Rachel Glannester and Micheal Kremer (2008). "Using Randomization in Economic Development Research: A Toolkit," *Handbook of Development Economics*, Vol. 4, Elservier Science.

2. Gerber, Alan S., and Donald P. Green. "Field experiments: Design, analysis, and interpretation." WW Norton, 2012.
3. Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel MJ Vermeersch. "Impact evaluation in practice." World Bank Publications, 2011.