14.74 Recitation 3

Causality, Dummy Variables, Diff in Diff

Maheshwor Shrestha

September 27, 2013

Agenda

- A framework for causality
- Regression with dummy variables
- Difference in Difference estimator
- Statistics review

Causality

- What is the causal impact of a policy (or a treatment)?
- Think about a parallel universe with everything identical to this world. In one of the worlds, provide the treatment to the individual and in the other world, do nothing.
  - The only difference between these two worlds is that this individual got the treatment. *Ceteris paribus*
  - After some time, compare the outcome of the individual with his/her twin in the parallel world. The difference in outcome is the causal effect of the treatment on the individual
  - The outcome for the twin in the other world that did not get the treatment is the *counterfactual* for the person in this world that got the treatment
- Since we never observe the other world, we never observe the counterfactual

What we like to know

- What is the average effect of a policy (or treatment) on the entire population?
  - This is the average treatment effect (ATE)
  - What is the effect of deworming on the population of schoolchildren

$$E\left[\text{outcome if treated} - \text{counterfactual outcome}\right]$$

- What is the average effect of a policy (or treatment) on the population affected by the policy?
  - This is the average treatment effect on the treated (TOT)
  - What is the effect of providing income support to the poor?

$$E\left[\begin{array}{r}\text{outcome if treated for the treated group}\\ -\text{counterfactual outcome for the treated group}\end{array}\right]$$

The problem

- In real world, we usually have a group affected by the policy (treatment group, T) and a group not affected by the policy (control group, C)
  - Schools were built in certain parts of the country and not in others
- Since we don't observe the counterfactual, we can only take the difference in outcome between the two groups

  $$E\left[\text{realized outcome for } T - \text{realized outcome for } C\right]$$

  is this what we want?
- The difference equals

  $$\begin{aligned} & E\left[\text{realized outcome for } T - \text{counterfactual outcome for } T\right] \\ +& E\left[\text{counterfactual outcome for } T - \text{realized outcome for } C\right] \end{aligned}$$

The problem...

$$E\left[\text{realized outcome for } T - \text{counterfactual outcome for } T\right]$$
$$+E\left[\text{counterfactual outcome for } T - \text{realized outcome for } C\right]$$

- The first expression is what we want. The treatment effect on the treated (TOT)
- The second expression is called the **selection bias**.
  - It is the difference in outcomes between the T and C if neither were treated
- When the selection bias is zero, the average difference between the treatment and control group gives us what we want.
- The exercise of establishing causal effect of a treatment is about arguing (via assumptions, program design and econometric techniques) that the selection bias is zero (or negligible). That the T and C are comparable

A Solution

- Suppose in a population, we could randomly choose some to get the treatment and some to not get the treatment. That is, T and C are randomly assigned. What is the advantage?
- Since treatment was randomly assigned, group C now provides a good proxy for the counterfactual. That is

$$E\left[\text{counterfactual outcome for } T\right] = E\left[\text{realized outcome for } C\right]$$

  so a comparison of mean outcome between the T and C gives the causal TOT (and the ATE)!
- Randomized control trials establish causality will minimal set of assumptions.
  - Issues?

Selection Bias Example

- Government conducts regular immunization camps in places lagging in immunization rates. You are measuring child health outcomes. In what ways could the places be different? What is the bias?
  - Lagging areas have bad infrastructures and lack supplies
  - Lagging areas have other diseases that constantly affect children
  - Places which didn't get the program are angry at the government and refuse to use government services
- Deworming pills randomly given to some children (seems ideal). But when measuring outcomes (adult wages) there were lots of children that were not found. What is the bias if
  - the more educated moved to the city for better jobs and are harder to find
  - worms killed some children (who are the ones likely to die?)
  - the really poor people can't find job in their village and move to the city

Selection Bias Example

- The US eradicated malaria in the 50s after the introduction of DDT spray. You want to examine the effect of malaria (eradication) on future wages.
  - You compare wages of cohorts born after the eradication program with wages of cohorts unaffected by the eradication program. Problems?
  - You compare wages of the same cohort in places that had malaria, with those that didn't have malaria. Problems?
    - What places had malaria and what places didn't. Were they comparable?
- What if you compare the difference in wages between the older and younger cohorts, in places with and without malaria?
  - What does the selection bias term reduce to?
- This approach is called the difference-in-difference.

Regression recap

- Linear model:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

  where

$$\beta = \frac{Cov(y,x)}{Var(x)} = \frac{E\left[(y - E[y])(x - E[x])\right]}{E\left[(x - E[x])^2\right]}$$

$$\alpha = E[y] - \beta \cdot E[x]$$

- And we have

$$E[y_i | x_i] = \alpha + \beta \cdot x_i$$

  and $E[\varepsilon_i | x_i] = 0$
- Today, restrict to cases where $x_i \in \{0, 1\}$
  - x is a dummy variable (indicator variable) that only takes values 1 or 0

When $x_i = 0$

- From our regression coefficients, when $x_i = 0$

$$\alpha = E\left[y|x_i = 0\right] - \beta \cdot E\left[x|x_i = 0\right] = E\left[y|x_i = 0\right]$$

  - That is, the expected value (mean) of y when $x_i = 0$
- When $x_i = 0$, the model reduces to

$$y_i = \alpha + \varepsilon_i$$

where the coefficient α represents the mean of y when $x = 0$. Taking conditional expectations,

$$E\left[y_i|x_i = 0\right] = E\left[\alpha|x_i = 0\right] + E\left[\varepsilon_i|x_i = 0\right] = \alpha$$

When $x_i = 1$

- When $x_i = 1$, the model becomes

$$y_i = \alpha + \beta + \varepsilon_i$$

- Taking expectations

$$E\left[y_i | x_i = 1\right] = \alpha + \beta + E\left[\varepsilon_i | x_i\right] = \alpha + \beta$$

- Combining

$$\begin{aligned} \beta &= (\alpha + \beta) - \alpha \\ &= E\left[y_i | x_i = 1\right] - E\left[y_i | x_i = 0\right] \end{aligned}$$

  - That is, the coefficient β is the difference in mean outcome between the two groups ($x_i = 1$ and $x_i = 0$)

More detailed derivation

- Using $E[y \cdot x] = E[y|x=1]E[x] + 0$ and $E[x^2] = E[x]$

$$\begin{aligned}
\beta &= \frac{E[(y - E[y])(x - E[x])]}{E\left[(x - E[x])^2\right]} \\
&= \frac{E[yx - yE[x] - xE[y] + E[x]E[y]]}{E[x^2] - E[x]^2} \\
&= \frac{E[y|x=1]E[x] - E[y]E[x] - E[x]E[y] + E[x]E[y]}{E[x] - E[x]^2} \\
&= \frac{E[x](E[y|x=1] - E[y])}{E[x](1 - E[x])} \\
&= \frac{(E[y|x=1] - E[y])}{(1 - E[x])}
\end{aligned}$$

More detailed derivation (cont...)

- Using $E[y] = E[y|x=1]E[x] + E[y|x=0](1 - E[x])$

$$\begin{aligned}
\beta =& \frac{E[y|x=1] - E[y]}{1 - E[x]} \\
=& \frac{E[y|x=1] - (E[y|x=1]E[x] + E[y|x=0](1 - E[x]))}{1 - E[x]} \\
=& \frac{E[y|x=1](1 - E[x]) - E[y|x=0](1 - E[x])}{1 - E[x]} \\
=& E[y|x=1] - E[y|x=0]
\end{aligned}$$

Interpretation in experimental context

- The coefficient β represents the mean difference between two groups
- In case of a social experiment (think of providing deworming pills), we often denote the treatment group with $T_i = 1$ and the control group by $T_i = 0$ and the model is

$$y_i = \alpha + \beta T_i + \varepsilon_i$$

- Hence, the regression coefficient β is the difference in mean between the treatment group and the control group.
- When does β represent the causal effect of the treatment?

Two dummy variables

- Now, imagine there are two dummy variables (T_i, A_i). Think of T_i as the indicator for whether individual i was treated (affected by policy) and A_i as an indicator of whether the outcome was measured after the policy (treatment) was enacted. There will be four groups:
  1. $(T_i = 0, A_i = 0)$: Control group before the policy
  2. $(T_i = 0, A_i = 1)$: Control group after the policy
  3. $(T_i = 1, A_i = 0)$: Treatment group before the policy
  4. $(T_i = 1, A_i = 1)$: Treatment group after the policy
- The difference-in-difference (DD) regression is

$$y_i = \alpha + \beta A_i + \gamma T_i + \delta (A_i \times T_i) + \varepsilon_i$$

Difference in Difference mechanics

$$y_i = \alpha + \beta A_i + \gamma T_i + \delta \left(A_i \times T_i \right) + \varepsilon_i$$

- What are the mean outcomes of the four groups?

	Before $(A_i = 0)$	After $(A_i = 1)$
Control $(T_i = 0)$	$\mathcal{A}$	$\mathcal{B}$
Treatment $(T_i = 1)$	$\mathcal{C}$	$\mathcal{D}$

$$\begin{aligned}
\mathcal{A} &= \alpha \\
\mathcal{B} &= \alpha + \beta \\
\mathcal{C} &= \alpha + \gamma \\
\mathcal{D} &= \alpha + \beta + \gamma + \delta
\end{aligned}$$

DD in a grid

$$y_i = \alpha + \beta A_i + \gamma T_i + \delta \left(A_i \times T_i\right) + \varepsilon_i$$

	Before $(A_i = 0)$	After $(A_i = 1)$	Diff (A-B)
Control $(T_i = 0)$	α	$\alpha + \beta$	β
Treatment $(T_i = 1)$	$\alpha + \gamma$	$\alpha + \beta + \gamma + \delta$	$\beta + \delta$
Diff (T-C)	γ	$\gamma + \delta$	δ

- What do each of the coefficients represent?
  - α: Mean of the Control, Before the policy
  - β: Mean difference in the control before and after the policy
  - γ: Mean difference before policy between treatment and control
  - δ: Difference in difference

$$\begin{aligned}\delta &= (\mathcal{B} - \mathcal{A}) - (\mathcal{D} - \mathcal{C}) \\ &= (\mathcal{D} - \mathcal{B}) - (\mathcal{C} - \mathcal{A})\end{aligned}$$

DD Assumptions

- δ is the difference in means between the treatment and control group of the differences before and after

$$\delta = E\left[\text{Gain in outcome for } T\right] - E\left[\text{Gain in outcome for } C\right]$$

  - When is it causal?

- Can rewrite δ in a manner similar to our causal framework

$$\begin{aligned}\delta =& E\left[\text{Gain in outcome for } T - \text{counterfactual gain for } T\right] \\ &+ E\left[\text{counterfactual gain for } T - \text{Gain in outcome for } C\right]\end{aligned}$$

- δ is causal when the second term, the selection bias, is zero
  - The gain in outcome for the control group mimics the counterfactual gain in outcome for the treatment group
    - That is, the control group and the treatment group are on the same trend
  - This assumption is called the **parallel trend** assumption

DD Assumptions

- When the parallel trend assumption holds, the DD estimator δ gives the average treatment effect for the treated group (TOT)
- When we also assume homogeneous treatment effect: that is, the treatment effect in the treated and control group are the same, then δ is the average treatment effect on the population (ATE).

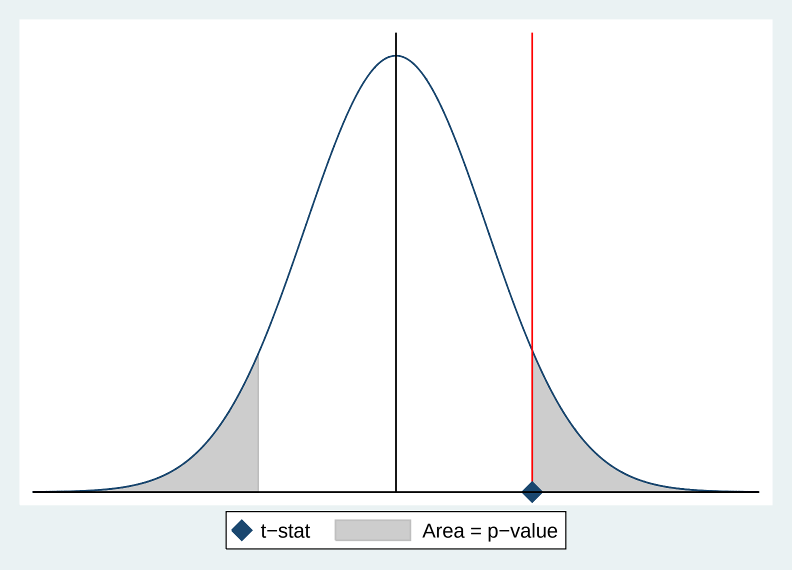
Parallel Trends

- How to make a convincing argument for parallel trends:
    - Diagram: Plot the outcome for treatment and control group for as many time periods as you can before and after the treatment. If the graphs actually look parallel before the treatment, then you are in good shape
    - Placebo test for timing: Do the regression with a fake treatment time. For example, there should be no DD effect between time 0 and -1.
    - Placebo test for outcome: Do the same regression with an outcome that is unlikely to be affected by the policy. If you find an effect, then people will think that something else must be going on besides just the policy
    - Covariate balance: Look at other covariates (district characteristics) and show that the treatment and control group are not very different. If they are, theoretically, you can fix it by including those covariates in your regression. But people will be skeptical about whether other unobserved variables being systematically different between the treatment and control groups, and that is driving the result.

Stats Review: t-stat and p-value

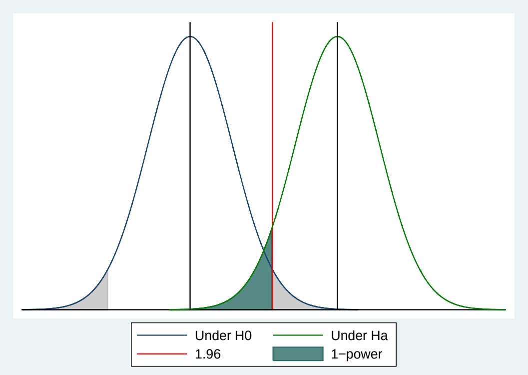
- **t-stat**: $t = \frac{\hat{\beta} - \beta(=0)}{\text{se}(\hat{\beta})}$
  - tells us how many standard deviations far is our estimate from the null (zero)
  - the further it is (higher values), the less likely that the null is true (reject null)
- **p-value**: If we were to repeat the experiment several times, what is the probability of getting (t-)statistic that is worse for the null than the current value
  - Area outside the t-stat in a t-distribution
  - Smaller values => reject the null

Stats Review: t-stat and p-value



Power of a test

- Suppose that the null is false. That is, $\beta \neq 0$, then we would like to reject our null $H_0 : \beta = 0$
- **Type-II error**: Under a test of given significance, α, what is the probability that we fail to reject the null.
- **Power**: 1-(type-II error rate). Denoted by β



Power calculation in experiment design

- When we are designing an experiment, we want to choose the number of observations (sample size of experiment) so that we have enough power (80%) to falsely reject the null of no effect when there is an effect of some magnitude (effect size). This depends upon
  - The effect size of the treatment. Larger we expect the effect to be, smaller sample size is required
  - Variance in outcome. Larger variance in outcome measure, need larger sample size
  - Power, β:. The more power we want, need larger sample sizes
- If you are randomizing at the level of group. For example schools instead of students, then required sample size also depends on
  - The intra-cluster correlation: If students within a school look very similar (high intra-cluster correlation) then need more schools/students