Subclassification and Matching

ECON 258 DATA ANALYTICS

PREPARED BY ARDINA HASANBASRI

Outline

1) Review Potential Outcome Model

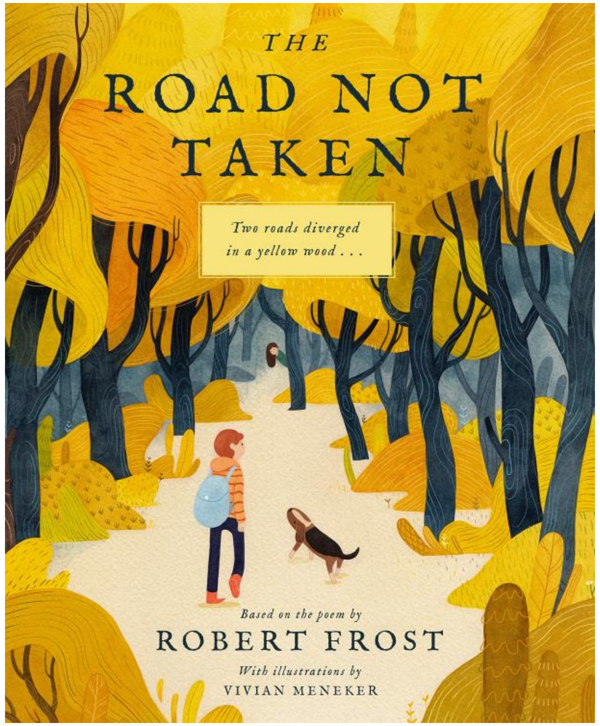
2) Some Theory! Subclassification and Matching Procedures to get ATE

3) Propensity Score Matching

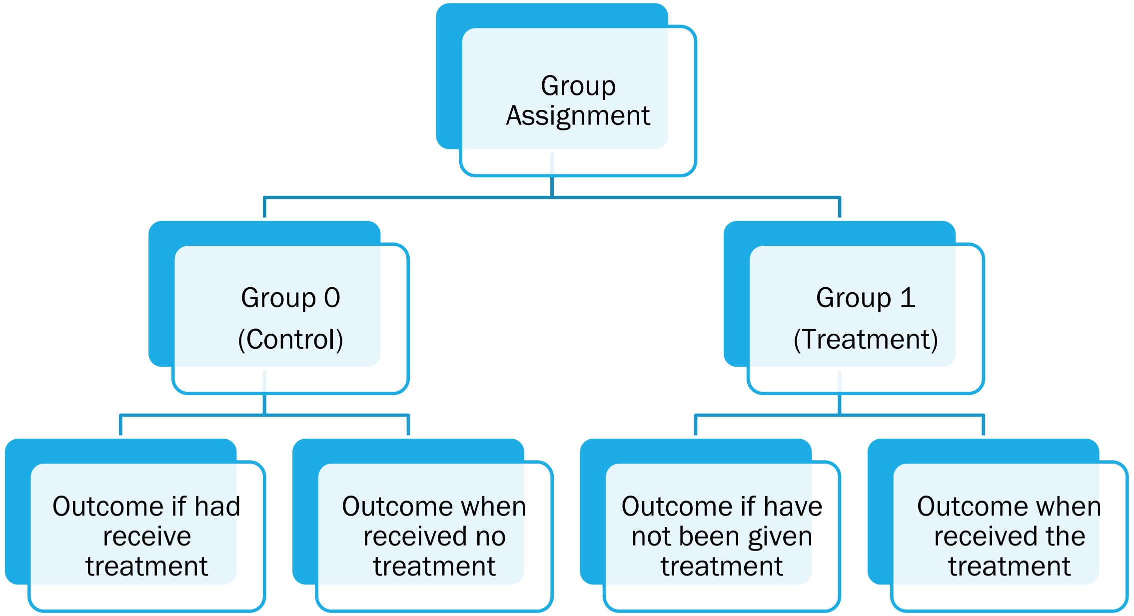
- Motivation and Possible Issues

4) Coding Exercise + Study Worksheet: Propensity Score Matching

Review:
1.1 Group Assignment VS Outcomes



<u>THE PROBLEM:</u> what happens in the road not taken?



Review:
1.2 Potential Outcome Model

$$Y_i^1 = \beta + \delta_i + u_i$$
$$Y_i^0 = \beta + u_i$$

Recall, δ_i is the effect received from treatment D_i

Therefore:

$$Y_i = \beta + \delta_i D_i + u_i$$

Review:
1.3 Two Causes of Bias in OLS Regression

1. Selection into treatment
2. Heterogeneous effect between treated and untreated group

$$
\begin{aligned}
E[Y^1|D=1] - E[Y^0|D=0] \quad = \quad & ATE \\
& + E[Y^0|D=1] - E[Y^0|D=0] \\
& + (1-\pi)(ATT - ATU) \qquad (77)
\end{aligned}
$$

Example:

What is the effect of an app that restricts social media (treatment) on student's study time (outcome)?

- selection: students who would benefit from the app are the ones who download it
- heterogenous effect: even without selection, students benefit differently from the app

Review:
1.4 Assumptions to Recover Causal Effect

1) Assume there is no heterogenous effect $\delta_i = \delta$

   The last term will then disappear.

2) Assume or make sure you have **independence.**

$$(Y^1, Y^0) \perp\!\!\!\perp D$$

Treatment is assigned to individuals **INDEPENDENT** of potential outcome.

Goal is to get rid of the third term.

$$E[Y^1|D = 1] - E[Y^1|D = 0] = 0$$
$$E[Y^0|D = 1] - E[Y^0|D = 0] = 0$$

2.0 Intro to Subclassification and Matching

What happens if you do not have experimental data? (focus on selection bias for now)

Experimental data creates a treatment and control group that independent of treatment.

$$(Y^1, Y^0) \perp\!\!\!\perp D$$

The idea of matching: one way to achieve independence is by conditioning on X

$$(Y^1, Y^0) \perp\!\!\!\perp D | X$$

2.1 Smoking and Mortality (Cochran 1968)

Table: Mortality rates by smoking type and country

Smoking group	Canada	British	US
Non-smokers	20.2	11.3	13.5
Cigarettes	20.5	14.1	13.5
Cigars/pipes	35.5	20.7	17.4

It seems that cigars/pipes are worse than cigarettes.

Is this true?

2.2 Independence Assumption

Average treatment effect depends on independence assumption being TRUE

$$E[Y^1|Cigarette] = E[Y^1|Pipe] = E[Y^1|Cigar]$$
$$E[Y^0|Cigarette] = E[Y^0|Pipe] = E[Y^0|Cigar]$$

Is this true?

2.3 Checking for Balance

Look at the mean age of different groups.

Smoking group	Canada	British	US
Non-smokers	54.9	49.1	57.0
Cigarettes	50.5	49.8	53.2
Cigars/pipes	65.9	55.7	59.7

2.4 Subclassification Example

	Death rates	Number of	
	Cigarette-smokers	Cigarette-smokers	Pipe/cigar-smokers
Age 20-40	20	65	10
Age 41-70	40	25	25
Age ≥71	60	10	65
Total		100	100

If cigarette smokers have the same X (age distribution with the pipe/cigar group):

Cigarette smokers' mortality using pipe smokers age $y_0 = 20 * \frac{10}{100} + 40 * \frac{25}{100} + 60 * \frac{65}{100} = 51$

Cigarette smokers' actual mortality $y_1 = 20 * \frac{65}{100} + 40 * \frac{25}{100} + 60 * \frac{10}{100} = 29$

Compare this 51 (adjusted for age) with the mortality rate of the pipe/cigar smokers (both look at same age group).

Unadjusted Mortality Rates

Smoking group	Canada	British	US
Non-smokers	20.2	11.3	13.5
Cigarettes	20.5	14.1	13.5
Cigars/pipes	35.5	20.7	17.4

Adjusted Mortality Rates for Specific Age Group

Smoking group	Canada	UK	US
Non-smokers	20.2	11.3	13.5
Cigarettes	29.5	14.8	21.2
Cigars/pipes	19.8	11.0	13.7

2.6 Assumptions for ATE

1. $(Y^1, Y^0) \perp\!\!\!\perp D | X$ (conditional independence)
2. $0 < Pr(D = 1|X) < 1$ with probability one (common support)

These two assumptions yield the following identity

$$
\begin{aligned}
E[Y^1 - Y^0|X] &= E[Y^1 - Y^0|X, D = 1] \\
&= E[Y^1|X, D = 1] - E[Y^0|X, D = 0] \\
&= E[Y|X, D = 1] - E[Y|X, D = 0]
\end{aligned}
$$

2.7 Common Support and Curse of Dimensionality

	Death rates	Number of	
	Cigarette-smokers	Cigarette-smokers	Pipe/cigar-smokers
Age 20-40	20	65	10
Age 41-70	40	25	25
Age $\geq$71	60	10	65
Total		100	100

Observe that in every column there exist some observation.
What happens when there isn't?

3.1 Propensity Score Matching

Conditional on the <u>PROBABILITY</u> of getting into treatment.

Intuition:

- Instead of conditioning on so many variables, conditioning on the probability of receiving a treatment instead.

- Run a logit or probit or linear probability model and get the probability for each person of going into treatment.

Create a model that predicts the probability to treatment:

$$P(D_i) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \epsilon_i$$

3.2 Example: Propensity Score Matching

The effect of an unconditional cash transfer (UCT) program:

- Indonesia rolled out a UCT across the country.
- Two groups: one where program is already being rolled out versus another where was not yet.
- The government provided the criteria on a person's eligibility.
- Can create a probability score matching for the two groups, and then match.

A Propensity Score Matching Exercise

Q: Does government training programs improve wage outcomes?

> Download the complete R-Code for this lecture.

> Download the worksheet to help you take notes while we go through the code together.

4.1 On the Job Training!

Question: Is job training effective for increasing the wages of disadvantages works who enter the labor market?

Experiment data is great! (Lalonde 1986)

 - Mid-70s National Work Demonstration (NSW)

 - Result: Experimental data show treatment group earn approx. $800 - $900 dollars more earnings after 3 years.

Can we compare this results with non-experimental data?

- Use clean PSID data and CPS data.

Today is all about analyzing results and understanding intuition of propensity score matching!

Question 1) Load the Lalonde data in R. Run a simple regression of annual earning of individuals in 1978 (re78) on whether they received the training program or not (treat).

Interpret the coefficient on "treat" from the regression results.

Question 2) Graph the density plot and barplot of the treatment and control group by age, and then by race.

- Before conducting any type of "policy" or "treatment" analysis, we first check if the sample is balanced.
- Balanced means that the characteristics of people in the control look similar to the ones in the treatment group.

Is the sample between treatment and control group balanced? Document your observation.

Question 3) Now, let's conduct the matching procedure. The main input for the "matchit" command is a model that determines the propensity score (the likelihood someone will be in treatment).

- When you have a dummy (0,1) variable as the dependent variable Y in your regression, how do you interpret a coefficient? For example, if the Y is treatment (receiving training).

What would be the interpretation of coefficient in front of age if that coefficient is xxx?

- Given you know the answer above, you then know that what the algorithm is doing is calculating the probability of attending training based on the individual characteristics. Once everyone is assigned a probability, the algorithm finds individuals with similar characteristics and assigned them into treatment and control group.

From the m.out output, how many from the sample were "matched" and put into control and treatment group?

Question 4) Graph the propensity score.

What do you notice about the distribution of propensity scores before and after matching?

What does the graphs mean?

Question 5) m.data gives us the sample of matched control and treatment group based on the propensity score. Check again the balance between this created control and treatment group.

How is the balance now?

Question 6) Graph the propensity score.

Now using the new matched sample, what is the coefficient of treatment on annual earnings in 1978? Interpret the results in your own words.

Question 7)

m.data gives us the sample of matched control and treatment group based on the propensity score. Check again the balance between this created control and treatment group.

Does running a simple versus more complicated regression make a difference on the result using the new propensity score matching data?

Reference

Material taken from chapter Potential Outcomes Causal Model.

“Causal Inference: The Mixtape (V.1.7)” by Scott Cunningham.

Propensity Score Matching worksheet material taken from:

Dayal (2015) “An Introduction to R for Quantitative Economics”