CHAPTER 8
ANALYSIS

財金四　王棋瑩　**Chi-Ying Wang**
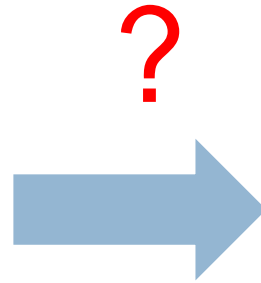工管四　王元翰　**Yuan-Han Wang**

Agenda

1. Prior to analysis
2. Potential problems
3. Pre-analysis plan (PAP)

Prior to analysis

Prior to analysis

Consider an experiment:

Prior to analysis

What should be done?	How?
Obvious errors	Feasible range, built-in skip patterns, consistency checks
Outliers	Check sensitivity
Attrition rates	Collect more data
Plotting and describing the data	Summary statistics

Prior to analysis

- $Y_i = \alpha + \beta * T_i + \gamma * X_i + \varepsilon_i$
  1. Y: the increase in grades
  2. α: mean in the comparison group
  3. T: dummy variable, β: treatment effect
  4. X: other variables (age, sex…)
  5. ε: the part of outcome that can't be explained by the above variables

Potential problems

Randomize at the group level

Spillovers

Randomize at group level

- Control
- City
- Rigid requirement of attendance rate

- Comparison
- Country
- Encourage students to attend freely
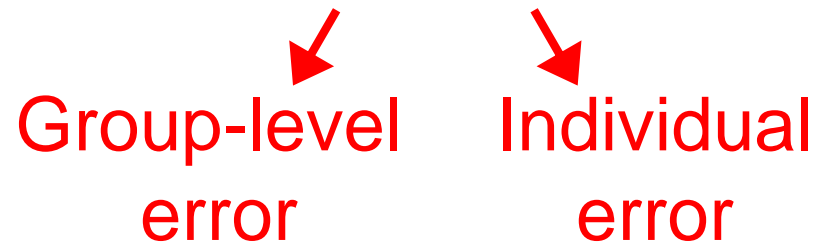
Randomize at group level

- Individual-level randomization:

$$Y_i = \alpha + \beta * T_i + \gamma * X_i + \varepsilon_i$$

- Group-level randomization:

$$Y_{ij} = \alpha + \beta * T_i + \gamma * X_i + v_j + \omega_{ij}$$

Group-level error   Individual error

What is spillover?

- When the effect on some people produces a secondary effect on other people, we say there are spillovers.

Spillovers within treatment

- Negative or positive spillovers within treatment.

- Violate the assumption that data are independent to each other → Biased and inefficient estimation.

Pre-analysis Plans

Data mining

- If two groups of people are compared on many different characteristics, it's very likely that at least one will turn out to be significant simply by chance. – thus we may look for the result we want in the data until we find it.

What is PAP ?

□ Pre-analysis plan describes, ahead of time, how the data will be analyzed

Why and when do we need one ?

□ Why we need PAP ?

Ans: Gain credibility by avoiding the risk of being accused of data mining.

Why and when do we need one ?

- When? There are three situations..

- When there are many alternative ways to measure outcomes.

EX. Is poor educational level harmful for economic growth?

Why and when do we need one ?

- When we want to test the effectiveness of a program on different subgroups.

EX. Is the academic performance between men and women in NTU different from each other?

Why and when do we need one ?

- When there are alternative ways to specify estimating equation.

EX. Running the regression in levels or logs, OLS, logit….

At what point should PAP be written?

- The greatest protection against accusations of data mining comes when a PAP is written before an evaluation is started, before baseline data are collected, before the intervention has been implemented.

We are NOT Sherlock Holmes

Drawbacks of PAP

- Writing a PAP very early means that we have to ignore lots of information that is acquired throughout the process.
- May left out an important determinant.
- The right way to perform our analysis will often be dependent on the main findings

Drawbacks of PAP

- It's possible to write an PAP that is conditional: If I find result X, I do this; If I find result Y, I do that….

But it is hard since in reality it is a HUGE tree

Conclusion

- We runs a tension between the benefits of the credibility that comes from tying one's hands versus the benefit of flexibility to respond to unforeseen events and results.

Conclusion

☐ PAP is still relatively new in economics, lots of things still remain debatable.

Conclusion

Conclusion

- Prior to analysis
- Potential problems of analysis
- PAP