Introduction to Econometrics

Lecture 2 : Causal Inference in Social Science

Zhaopeng Qu

Sep. 20th, 2018

Outlines

Review the Previous Lecture

Causal Inference in Social Science

- Causality is our main goal in the studies of empirical social science
- The existence of Selection Bias makes social science more difficult than science.
- Randomly Assignment or RCT can eliminate the Selection Bias.
- Although experimental method is a powerful tool for economists, every project can not be carried on by it.

Program Evaluation Econometrics

- Question: How to do empirical research scientifically when we can not do experiments?
  - It means that we always have selection bias in our data, or in term of "endogeneity".
- Answer: Build a reasonable counterfactual world by naturally occurring data to find a proper control group is the core of econometrical methods.
  - It is the main reason why modern econometrics exists and develops.

"The modern menu of econometric methods can seem confusing, even to an experienced number cruncher. Luckily, not everything on the menu is equally valuable or important. Some of the more exotic items are needlessly complex and may even be harmful. On the plus side, the core methods of applied econometrics remain largely unchanged, while the interpreta-

Furious Seven Weapons（七种武器）

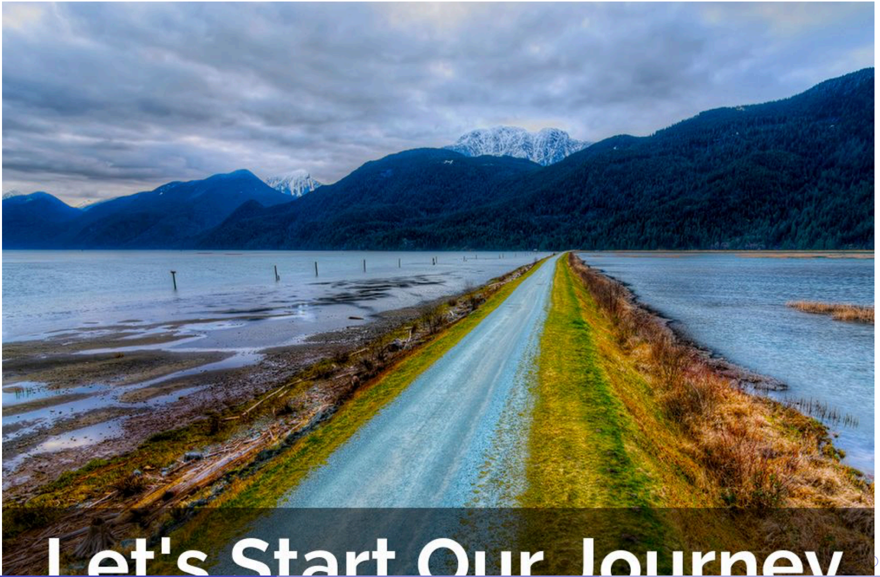
- Build a reasonable counterfactual world or find a proper control group is the core of econometrical methods.

1. **Random Controlled Trail(RCT)**
2. **Ordinary Least Squre Regression(OLS)(最小二乘回归)**
3. **Matching and Propensity Score（匹配与倾向得分）**
4. **Decomposition（分解）**
5. **Instrumental Variable（工具变量）**
6. **Regression Discontinuity（断点回归）**
7. **Difference in Differences （双差分或倍差法)**

Furious Seven Weapons（七种武器）

- These **Furious Seven** are the most basic and popular methods in applied econometrics and so powerful that
  - even if you just master one, you may finish your empirical paper and get a good score.
  - if you master several ones, you could have a higher chance to publish your term paper.
  - If you master all of them, you might to teach applied econometrics class just as what I am doing now.
- We will introduce the essentials of these methods in the class as many as possible. Let's start our journey together.

An Amazing Journey



Overview: Furious Seven

Selection on Observables

- In contrast to RCT, in observational studies, researchers cannot control the assignment of treatment into a treatment group versus a control group.
- To make two groups comparable, we need to keep treatment and control group "other thing equal" in observed characteristics and unobserved characteristics.
  - OLS
  - Matching
- Assume the source of selection bias is only from the difference in observed characteristics.

Identification Assumption for Regression

Conditional Independence Assumption(CIA)

$$(Y_{0i}, Y_{1i}) \perp D_i \mid C_i$$

also called **selection on observables**.

- Thus the causal inference can be obtained

$$E[Y_i|C_i, D_i = 1] - E[Y_i|C_i, D_i = 0] = E[Y_{1i} - Y_{0i}|C_i]$$

- CIA asserts that conditional on observable characteristics C_i, potential outcomes, Y_i are independent of assigned treatment, D_i.
  - In other words, observed covariates C_i can *fully* explain the difference in potential outcome between treatment and control groups.
  - After controlling for value of covariates C_i, the assignment of units to treatment is "**as good as random**".
  - The probability of receiving treatment is same for the individuals with

OLS Regression

- Then, we get causal effect of treatment by comparing the outcomes of treatment and control groups with the same observed characteristics

$$Y_i = \alpha + \beta D_i + \gamma C_i + \epsilon_i$$

- One regression always have ***dependent variables*** and ***independent ones***. Generally, we have one ***dependent*** variable Y , the outcome of the relationship we care about.
- And several independent variables, but we would like to pay more attention to only one among them. So specifically we call it ***independent*** or **treatment** variable D (sometimes it can be multiple, X_j)
- Other variables in the right hand, we call them ***control variables*** C, which we would like to explicitly *hold fixed* when studying the effect of Y on D .
- The concept of ceteris paribusmeans "*other things equal*"

Matching and Propensity Score (匹配与倾向得分法)

- Matching is another way to eliminate selection bias by controlling observable covariates.
- Idea: directly recreate a control group based on the treatment group (preprocess the data)
  - Based on treatment unit' s covariate values, match each unit in the treatment group to an untreated unit
  - use the information C as much as possible for both treatment and control groups.

Matching and Propensity Score

- An example: How much wage premium does the worker get by receiving a job training program.
- For selection bias, we can't get the right causal effect(ATT) via naive comparison, thus workers joined the program are not similar to the workers didn't join it.
- We could use information in other covariates (C_i) to find the corresponding samples in untreated group as the sample of **control group**

Matching and Propensity Score

ID	C_i	D_i	Y_i
1	1	0	10
2	1	1	15
3	1	1	20
4	2	0	25
5	2	1	30
6	2	0	30
7	3	1	25
8	3	0	35
9	4	0	50
10	5	1	55

$\Longrightarrow$

ID_1	C_i	Y_{1i}	ID_0	Y_{0i}	$Difference$
2	1	15	1	10	5
3	1	20	1	10	20
5	2	30	$4,6$	27.5	2.5
7	3	25	8	35	-10
10	5	55	NA	NA	NA

- Thus $ATT = \frac{(5+20+2.5-10)}{4} = 6.875.$

Matching and Propensity Score

- So we use a certain method to calculate the information of the worker such as gender,age, education, pre-income, occupation, industry, etc. to identify a "pair" of similar samples from primary control group and treatment group respectively.
- Then we use the new sample to make a comparison between the pairs.
- In an ideal control group, all people are as same as those in treatment group on one or three main variabels, such as gender, age and education. As adding more variables into the process of matching, we can't find exact match more impossible.
- Matching may become **unfeasible** with many covariates since it is difficult to find good matches *in large dimensions*.

Propensity Score Matching

- Propensity Score Matching: Rosenbaum and Rubin(1983)
- Instead of matching over k dimensions, the method of **propensity score matching** (PSM) allows the matching problem to be reduced to a single dimension.
- The propensity score is defined as the treatment probability conditional on a set of observed variables C_i

$$p(C_i) = E[D_i \mid C_i] = Pr(D_i = 1 \mid C_i)$$

- Intuitively, propensity score $P(C_i)$ summarized all information of a set of covariates C_i into a single value.Then, we can just control $p(C_i)$ to eliminate selection bias.

Propensity Score Matching

Propensity Score Theorem

Rosenbaum and Rubin(1983)
Suppose the CIA holds, thus $(Y_{1i}, Y_{0i}) \perp D_i \mid C_i$, then
$(Y_{1i}, Y_{0i}) \perp D_i \mid p(C_i)$.

- If potential outcomes are independent of treatment status conditional on a set of covariates X_i,then potential outcomes are independent of treatment status D_i conditional on a scalar function of covariates, the propensity score $p(C_i)$.
- Then we can estimate causal effect of treatment by comparing the outcomes of treatment and control groups with the same propensity score $p(C_i)$.
- Basic idea：把所有的信息，通过回归方程归结为一个值：倾向分数，然后只对这一个值进行排序。Rosenbaum and Rubin(1983) 证明这样一个值排序是对所有变量排序的最佳近似

Matching v.s.OLS

- **Similarity**
  - Both matching and regression relies on CIA (selection on observables).
- **Difference:** Why we still need matching?
  - Matching does not impose any restriction on empirical specification (e.g. linear function) when estimating treatment effect. It don't need estimate specific parameters in the model.
  - Regression does not account for the **common support** issue.

Some Examples

- Dejaha and Wahba(2002) "Job Training Program", "Propensity Score-Matching Methods for Non-experimental Causal Studies", *The Review of Economics and Statistics,* Vol. 84, No. 1 (Feb., 2002).
- Angrist Joshua(1998) "Veteran Return", "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants",*Econometrica*, Vol. 66, No. 2 (Mar., 1998).
- 盛丹（2013），"外资进入是否提高了劳动者的讨价还价能力"，《世界经济》，2013 年第 10 期。
- 陈玉宇和吴玉立 (2008)，"信息化对劳动力市场的影响：个人电脑使用回报率的估计，《经济学（季刊)》，2008 年第 4 期。

OLS formula: the Returns to Schooling

- We would like to estimate the effect of **schooling on wage**, thus

$$y_i = \alpha + \beta D_i + u_i$$

  $y - logwage; x - schooling; u - error\,term$ and α is a constant. The main hypothesis is $E(u_i \mid D_i) = 0$ which means that D_i is randomly assigned.
- So β is the parameter we are interested in, which indicates how much the wage will increase in term of percentage with one more schooling year.
- Taking the covariance in both sides

$$Cov(y_i, D_i) = Cov(\alpha, D_i) + \beta Cov(D_i, D_i) + Cov(D_i, u_i)$$

- Then OLS estimate is

$$\beta_{ols} = \frac{Cov(y_i, D_i)}{Var(D_i)}$$

Omitted Variable Bias(OVB)

- One of the most common and important problem in applied econometrics: we cannot control for everything that we would like to
- **Omitted Ability Bias**: We would like to estimate the effect of schooling on wage, thus

$$y_i = \alpha + \beta D_i + \gamma A_i + u_i$$

where A_i represents unobservables, such as ability.

- So β is the parameter we are interested in, which indicates how much the wage will increase in term of percentage with one more schooling year.
- But we could **NOT** observe the ability, so we have regress

$$y_i = \alpha + \beta D_i + e_i$$

where

Omitted Variable Bias(OVB)

- Review OLS estimate is

$$\beta_{ols} = \frac{Cov(y_i, D_i)}{Var(D_i)}$$

- Then

$$\begin{aligned}
\beta_{OVB} &= \frac{Cov(\alpha + \beta D_i + \gamma A_i + u_i, D_i)}{Var(D_i)} \\
&= \frac{Cov(\alpha, D_i) + \beta Cov(D_i, D_i) + \gamma Cov(A_i, D_i) + Cov(u_i, D}{Var(D_i)} \\
&= \beta_{ols} + \gamma \frac{Cov(A_i, D_i)}{Var(D_i)}
\end{aligned}$$

- Thus if $\gamma \neq 0$ **and** $Cov(A_i, D_i) \neq 0$, the OLS estimate under OVB is biased.
- To obtain causal effect (eliminate OVB), we need a variation in D_i is

Instrument Variable

- The Instrument Variable (IV) is an exogenous sources of variation that drives the treatment X_i but unrelated to other confounding factors A_i that affect outcome Y_i.
- Intuitively, IV breaks variation of the treatment D_i into **two** parts
    - A part that might be correlated with other unobservable confounding factors, such as A_i
    - A part that is not(correlated with....).
- We use the variation of **the second part in D** to estimate causal effect of the treatment.

Instrument Variable

- IV is a signature technique in the econometrics toolkit. The earliest application involved attempts to estimate demand and supply curve for product.
- **Question:** How to find the supply or demand curves?
- **Difficulty:** We can only observe intersections of supply and demand, yielding pairs.
- **Solution:** Wright(1928) use variables that appear in one equation to shift this equation and trace out the other.
- The variables that do the shifting came to be known as **Instrumental Variables**.
- The most important contemporary use of IV is to solve the problem of omitted variable bias and other endogenous problems.

Identification Assumption

- Assume we have another variable Z_i satisfy the following Three Key Assumptions
  - Z should be **correlated** with the casual variable of interest, Di (endogenous variable)

$$Cov(D, Z) \neq 0$$

  - Z must be **uncorrelated** with any other determinants of the dependent variable.

$$Cov(e, Z) = 0$$

No direct or indirect effect of the instrument on the outcome **unless through the treatment**

Identification Assumption

- Independent assumption (Instrument exogeneity):
  - An instrument variable should be as good as **randomly assigned.**
  - It should not be correlated with other factors that might affect treatments and outcomes.

$$\begin{array}{ccccc} Z & \longrightarrow & D & \longmapsto & Y \\ & & \uparrow & \nearrow & \\ & & u & & \end{array}$$

- IV initiates **a causal chain**: the instrument Z affects D, which in turn affects Y

Where to find an IV?

- **Generally Speaking**
  - **"可遇不可求"**

Where to find an IV?

- Institutional background
  - Angrist(1990)-**draft lottery**: Vietnam veterans were randomly designated based on birth day used to estimate the wage impact of a shorter work experience.
  - Acemoglu, Johnson, and Robinson(2001): the dead rate of some diseases in some areas to estimate **the impact of institutions to economic growth**.
- Natural conditions(geography, weather)
  - Miguel et al(2004)-Variation of Rainfall: Estimating the impact of **economic conditions on the likelihood of civil conflict** is difficult because of endogeneity and omitted variable bias. They use rainfall variation as an instrumental variable for economic growth in 41 African countries during 1981–1999.
- Economic theory and logics
  - Angrist & Evans(1998): have **same sex or different sex children** used to estimate the impact of an additional birth on **women labor supply**.

Regression Discontinuity Design(断点回归)

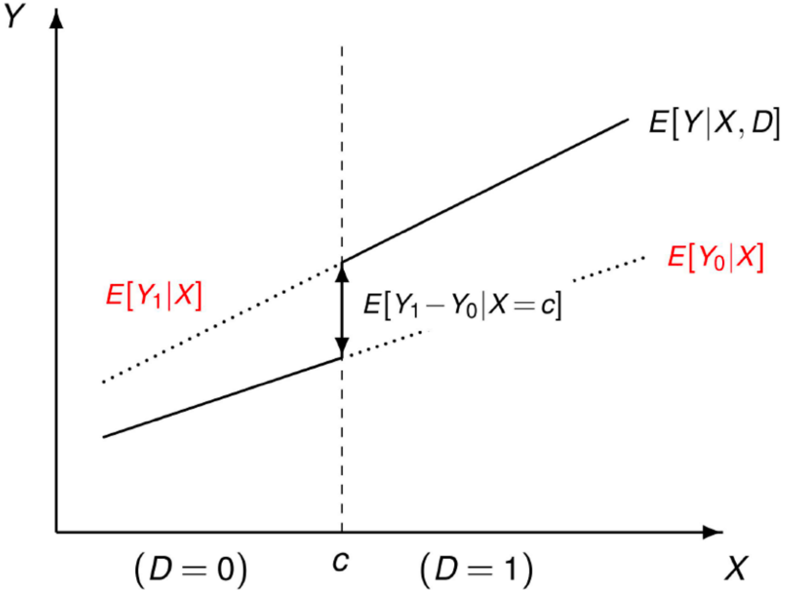
- Instead of controlling treatment assignment process, if researchers have detailed institutional knowledge of treatment assignment process. Then we could use this information to create an "experiment".
- Some rules are arbitrary and generate a **discontinuity** in treatment assignment.
- The treatment assignment is determined based on whether a unit exceeds some **threshold** on a variable (assignment variable, running variable or forcing variable)
- Assume other factors do not change abruptly at threshold.(continued assumption)
- Then any change in outcome of interest can be attributed to the assigned treatment

Regression Discontinuity Design(断点回归)

- Context: Individual receives treatment only if observed covariate C crosses known threshold C_0.
- Logic: compare individuals just above and below threshold.
- 举例：上大学影响收入。
- 但是是否能上大学是由分数线和分数共同决定的。
  - 刚过分数线的人上了大学。
  - 差一点够分数线的人没有上大学。
- 可以认为分数上的这一很小的不同，肯定不能说明二者的能力存在巨大的差异，所以这两部分人可以近似的看成是同质的。进而两者进入劳动力市场后的工资差异，应该全部归结于两人是否“上大学”而带来的。

Sharp RDD and Fuzzy RDD

- In general, depending on enforcement of treatment assignment, RDD can be categorized into two types
  1. **Sharp RDD**: nobody below the cutoff gets the "treatment" , everybody above the cutoff gets it.
     - Everyone follows treatment assignment rule (all are compliers).
     - Local randomized experiment with perfect compliance around cutoff.
  2. **Fuzzy RDD**: the probability of getting the treatment jumps discontinuously at the cutoff (NOT jump from 0 to 1)
     - Not everyone follows treatment assignment rule.
     - Local randomized experiment with partial compliance around cutoff.
     - Using initial assignment as an instrument for actual treatment.


Y
$E[Y|X,D]$
$E[Y_0|X]$
$E[Y_1|X]$
$E[Y_1 - Y_0|X = c]$
$(D = 0)$
c
$(D = 1)$
X

Regression Discontinuity(断点回归)

- Yuyu Chen(陈玉宇)，Avraham Ebenstein, Michael Greenstone and Hongbin Li（李宏彬）(2013)，"Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy"，PNAS
    - 主题：空气污染与健康
    - 结论：长期暴露于污染空气中，总悬浮颗粒物（TSP）每method上升 100 微克/立方米，死亡率上升 14%，平均预期寿命将缩短 3 年，且死亡率的上升几乎都是由心肺疾病导致的。
    - 按照北方地区总悬浮颗粒物的水平，这意味着中国北方 5 亿居民因严重的空气污染平均每人失去 5.5 年寿命。
    - 方法：自然实验 +RD：1949 年中华人民共和国成立后，按照淮河分界，中国南北方采取不同的供暖模式。

Regression Discontinuity: Chen et al(2013)

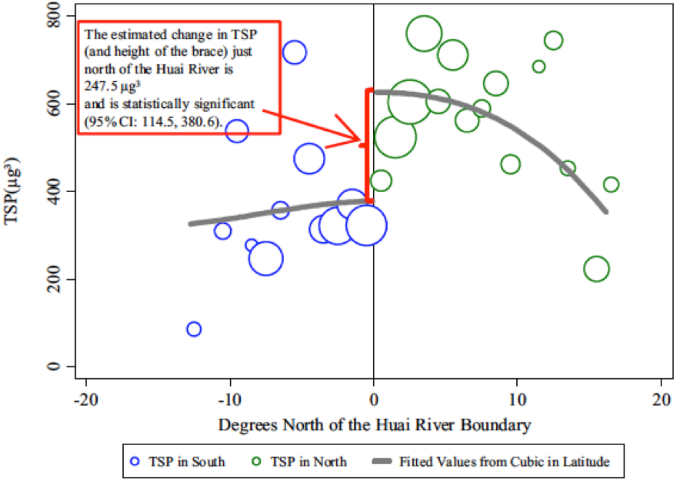


Fig. 2. Each observation (circle) is generated by averaging TSPs across the Disease Surveillance Point locations within a 1° latitude range, weighted by the

Regression Discontinuity: Chen et al(2013)

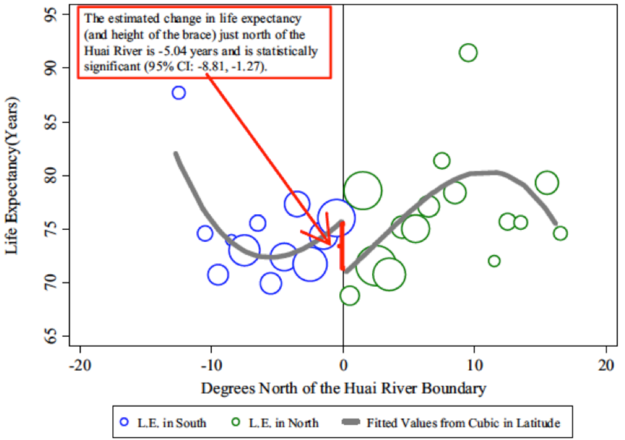


Figure: Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

Difference in Differences（双差分）

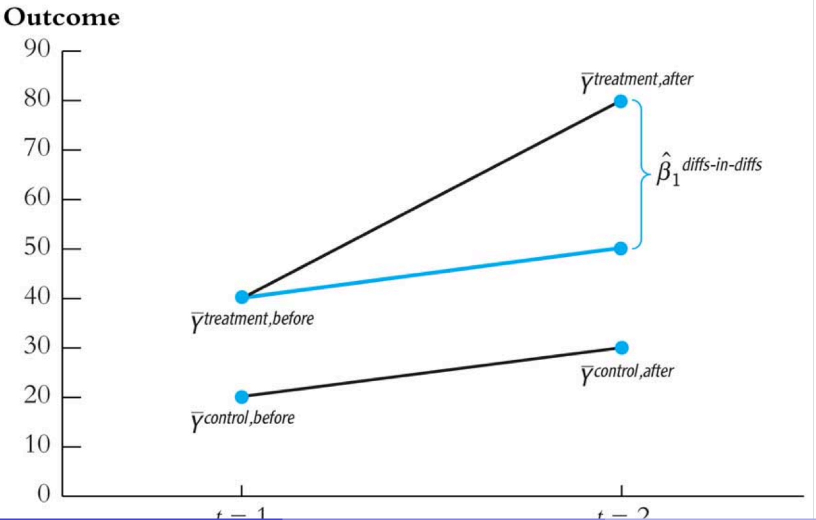
- In the absence of random treatment assignment, we cannot estimate causal effect by comparing outcomes of treatment and control groups.
- Since the outcomes of treatment and control groups are likely to be differ for many reasons, even though they receive the same treatment.
- If we can observe **group-level** outcomes **several times** (at least before and after treatment)
- Assume treatment and control outcomes move in parallel in the absence of treatment.
- Then, we can use diverged trend in outcome of treatment group to represent causal effect of treatment.

Difference in Differences（双差分）

- A typical experimental design requires a causal studies to do as follow
  - Randomly assignment of treatment to divide the population into a "treatment" group and a "control" group.
  - Collecting the data at the time of both pre-treatment and post-treatment
  - then comparing them to get the difference for the outcome
- Actually it is the DID estimator here.

Difference in Differences （双差分）

Figure:



Difference in Differences （双差分）

- Thus the control group that does not receive the treatment but experiences some or all the other influences that affect the treatment group.
- The outcome of treatment group
  - Before treatment: $E(Y_{0i}|D_i = 1)$
  - After treatment : $E(Y_{1i}|D_i = 1)$
- The outcome of control group
  - Before treatment: $E(Y_{0i}|D_i = 0)$
  - After treatment: $E(Y_{0i}|D_i = 1)$
- The casual effect interested is

$$ATT = E[(Y_{1i} - Y_{0i})]|D_i = 1] - E[(Y_{1i} - Y_{0i})]|D_i = 0]$$

Identification Assumption for DID

Common Trend Assumption

The treatment group and control group would have exhibited the same trend in the absence of the treatment.

- The average treatment effect on treated (ATT) can be identified by difference in trend of outcome between treatment and control groups.

Some Examples

- Minimum Wage Story by Card David & Alan Krueger(1994), “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania” , American Economic Review, Vol.84(4).
- “the Great Chinese Famine” in 1959-1961 by Yuyu Chen & Zhou, Li-An(2007). "The long-term health and economic consequences of the 1959-1961 famine in China," Journal of Health Economics, Elsevier, Vol. 26(4), pp 659-681, July.
- 周黎安和陈烨（2005）“中国农村税费改革的政策效果: 基于双重差分模型的估计”，《经济研究》，第 8 期。
- 孙文凯、白重恩和谢沛初（2011）“户籍制度改革对中国农村劳动力流动的影响”，《经济研究》，第 1 期。

An Simple Intuition: Mean Comparison

- **Common Idea**: match similar units, and produce a mean comparison
  - OLS gives conditional mean comparison.
  - Matching make a weighted conditional mean comparison.
  - DID compares difference in mean across locations.
  - RD compares means around the cutoff.
  - IV compares means of instrumented and non-instrumented.
- **Goal**: give a believable and reliable mean comparison.