CH7 習題演練

陳家威 [1]

November 23, 2022

[1] R10323045@ntu.edu.tw

7-9

題幹

1. 母親吸菸對出生嬰兒會不會有影響？

題幹

1. 母親吸菸對出生嬰兒會不會有影響？
2. 直覺會有影響，但是實證上能不能證明？

題幹

1. 母親吸菸對出生嬰兒會不會有影響？
2. 直覺會有影響，但是實證上能不能證明？
3. 資料集為 bweight_small

A- 基本統計量

- 母親吸菸者，出生嬰兒體重的平均值
- 母親不吸菸者，出生嬰兒體重的平均值
- 使用 t 檢定，檢定兩族群平均體重是否相同。

我們量出來的是什麼？

回歸 $BWEIGHT = \beta_1 + \beta_2 MBSMOKE$，並解釋 β_2 的係數

我們可以將 β_2 解釋為平均處理效果嗎？

β_2 是否可以代表 $E[BWEIGHT_1] - E[BWEIGHT_0]$

讓我們先跳出題目，討論一下上面量出來的係數有沒有意義

分兩群做平均，能代表什麼嗎？

考慮以下思考邏輯：

1. 醫生平均年薪 300 萬
2. 高級社畜平均年薪 200 萬
3. 所以高級社畜去當醫生，年薪可以增加 (?)

分兩群做平均，能代表什麼嗎？

考慮以下思考邏輯：

1. 醫生平均年薪 300 萬
2. 高級社畜平均年薪 200 萬
3. 所以高級社畜去當醫生，年薪可以增加 (?)

請問哪裡怪怪的？

我們到底想量測什麼東西？

「處理效果」- 如果有經過處理（ex: 丟去醫學系），會如何

ATE 如果隨機分配一個人去當醫生，他會比隨機分配他去當高級社畜多賺多少？

我們到底想量測什麼東西？

「處理效果」- 如果有經過處理（ex: 丟去醫學系），會如何

ATE 如果隨機分配一個人去當醫生，他會比隨機分配他去當高級社畜多賺多少？

ATT 如果醫生沒當醫生，那他少賺多少？

我們到底想量測什麼東西？

「處理效果」- 如果有經過處理（ex: 丟去醫學系），會如何

ATE 如果隨機分配一個人去當醫生，他會比隨機分配他去當高級社畜多賺多少？

ATT 如果醫生沒當醫生，那他少賺多少？

ATU 如果把一個當了高級社畜的人抓去當醫生，他會多賺多少？

我們到底想量測什麼東西？

「處理效果」- 如果有經過處理（ex: 丟去醫學系），會如何

ATE 如果隨機分配一個人去當醫生，他會比隨機分配他去當高級社畜多賺多少？

ATT 如果醫生沒當醫生，那他少賺多少？

ATU 如果把一個當了高級社畜的人抓去當醫生，他會多賺多少？

LATE 對一個只差一點點就上醫學系的高級社畜，讓他真的當醫生，可以多賺多少？

我們到底想量測什麼東西？

「處理效果」- 如果有經過處理（ex: 丟去醫學系），會如何

ATE 如果隨機分配一個人去當醫生，他會比隨機分配他去當高級社畜多賺多少？

ATT 如果醫生沒當醫生，那他少賺多少？

ATU 如果把一個當了高級社畜的人抓去當醫生，他會多賺多少？

LATE 對一個只差一點點就上醫學系的高級社畜，讓他真的當醫生，可以多賺多少？

我們到底想量測什麼東西？

「處理效果」- 如果有經過處理（ex: 丟去醫學系），會如何

ATE 如果隨機分配一個人去當醫生，他會比隨機分配他去當高級社畜多賺多少？

ATT 如果醫生沒當醫生，那他少賺多少？

ATU 如果把一個當了高級社畜的人抓去當醫生，他會多賺多少？

LATE 對一個只差一點點就上醫學系的高級社畜，讓他真的當醫生，可以多賺多少？

會需要考慮這麼多「處理效果」的原因，在於「結果通常不是隨機的」，而是「自我選擇的」

潛在結果框架

$$y_i = D_i y_{1i} + (1 - D_i) y_{0i} \tag{1}$$

潛在結果框架

$$y_i = D_i y_{1i} + (1 - D_i) y_{0i} \tag{1}$$

- D_i：個體 i 的選擇（當醫生 $=1$，當社畜 $=0$）
- y_{1i}：個體 i 當醫生的年薪
- y_{0i}：個體 i 當社畜的年薪
- y_i：個體 i 實際的年薪

潛在結果框架

$$y_i = D_i y_{1i} + (1 - D_i) y_{0i} \tag{1}$$

- D_i：個體 i 的選擇（當醫生 $=1$，當社畜 $=0$）
- y_{1i}：個體 i 當醫生的年薪
- y_{0i}：個體 i 當社畜的年薪
- y_i：個體 i 實際的年薪

如果當醫生，只觀察的到 y_{1i}，無發觀察到 y_{0i}，反之亦然。

我們稱另外一個為「反事實結果 (counterfactual result)」。

各種處理効果

ATE	$E[Y_{1i}] - E[Y_{0i}]$
ATT	$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1]$
ATU	$E[Y_{1i} \mid D_i = 0] - E[Y_{0i} \mid D_i = 0]$

各種處理效果

ATE	$E[Y_{1i}] - E[Y_{0i}]$
ATT	$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1]$
ATU	$E[Y_{1i} \mid D_i = 0] - E[Y_{0i} \mid D_i = 0]$

「新聞媒體」的「處理效果」：$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$

當醫生的平均薪資 - 當社畜的平均薪資

事實是，會當高級社畜的人，當初可能就因為技能點不在三類，所以沒選擇當醫生。所以這種新聞媒體的「處理效果」並不能回答「當醫生可以多賺多少（ATT）」

我們需要想辦法找出「選擇當醫生的人，如果不當醫生，他的薪資會是多少」，也就是 $E[Y_{0i} \mid D_i = 1]$（實際上觀測不到！）

什麼時候平均相減有意義？

如果潛在變數跟選擇沒有關聯：$\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i$

$$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] = E[Y_{1i}] - E[Y_{0i}]$$

什麼時候平均相減有意義？

如果潛在變數跟選擇沒有關聯：$\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i$

$$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] = E[Y_{1i}] - E[Y_{0i}]$$

隨機對照試驗 RCT

如果當不當醫生是隨機分配的，則兩族群薪資平均差，就可以看成是 ATE/ATT...。
薪資的平均差異，就代表了當醫生這件事對薪資的增幅。

什麼時候平均相減有意義？

如果潛在變數跟選擇沒有關聯：$\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i$

$$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] = E[Y_{1i}] - E[Y_{0i}]$$

隨機對照試驗 RCT

如果當不當醫生是隨機分配的，則兩族群薪資平均差，就可以看成是 ATE/ATT...。
薪資的平均差異，就代表了當醫生這件事對薪資的增幅。

我們可以放寬一點，假設在控制了某些條件之下，選擇會與潛在變數無關，我們稱之為條件獨立假設

CIA

$\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i \mid X$
給定 X 之下，處理與否 (D=1 or 0) 就與潛在結果無關。

所以一定要做實驗？

隨機分配孕婦抽菸與不抽菸，檢查兩平均（做 a 小題的回歸），即可檢查平均處理效果。

- 實驗不人道，喪盡天良
- 我們看到的資料卻只有孕婦自我選擇過後的結果

所以到底如何檢視 ATE/ ATT？

1. 靠一些自我選擇的模型，例如 Roy Model
2. 靠工具變數 (IV)– 得出 LATE
3. 靠逆傾向分數加權 (inverse propensity weighting) – 在 CIA 下得出 ATE
4. 課本作法– 在 CIA 下得出 ATE

2000（自我選擇），2019（隨機實驗），2021（局部平均處理效果）年諾貝爾經濟學獎。

C - 加入其他變數

- MMARRIED
- MAGE
- PRENTAL1
- FBABY

1. 這些變數是否為重要預設指標？
2. 是否與預期相同？
3. MBSMOKE 的係數估計發生很大的變化嗎？

控制了變數就有 ATE 了嗎？

如果假設我們的選擇（抽不抽菸），在控制了一些變數之後（是否已婚、有無做產檢、是否第一胎...），選擇就變成隨機的決定的話，那我們對選擇的係數，就可以詮釋為（條件）平均處理效果。

這種假設我們稱之為條件獨立假設 (CIA)。

在此題的情況下，有滿強的理由認為就算控制這些變數，抽菸還是自我選擇的結果。

D-Chow Test

抽菸到底會不會有差的另外一種檢定方式 – 分組，看迴歸係數是否一樣

1. 不把抽菸加入回歸，算 SSE_R
2. 只對抽菸的回歸，算 SSE_{U1}
3. 只對不抽菸的回歸，算 SSE_{U0}
4. 計算 $SSE_U = SSE_{U0} + SSE_{U1}$
5. 變數的數量 $= 5$
6. 全部的樣本數 $= 1200$
7. 計算 $F = \left(\frac{SSE_R - SSE_U}{5}\right) / \left(\frac{SSE_U}{1200-5\times 2}\right)$
8. 檢定 F 值

D 的另一種做法

如果 MBSMOKE 沒有影響，那麼加入他們的交乘項，應該係數都會是 0

介紹新指令：`testparm`

控制變數之後的平均處理效果

假設有 CIA，則我們可以算出平均處理效果，也就是在控制這些變數之下，我們分組的平均就有如隨機實驗般。

土法煉鋼做法

$$\begin{aligned}\tau_{ATE} =& (3143.2108 - 3154.0161) + (206.9242 - 96.3475) \times 0.715 \\ &+ (-5.5208 - 4.9272) \times 26.57583 + (8.2981 - 119.8472) \times 0.815 \\ &+ (94.0675 + 83.167) \times 0.440833 \\ =& -222.19\end{aligned}$$

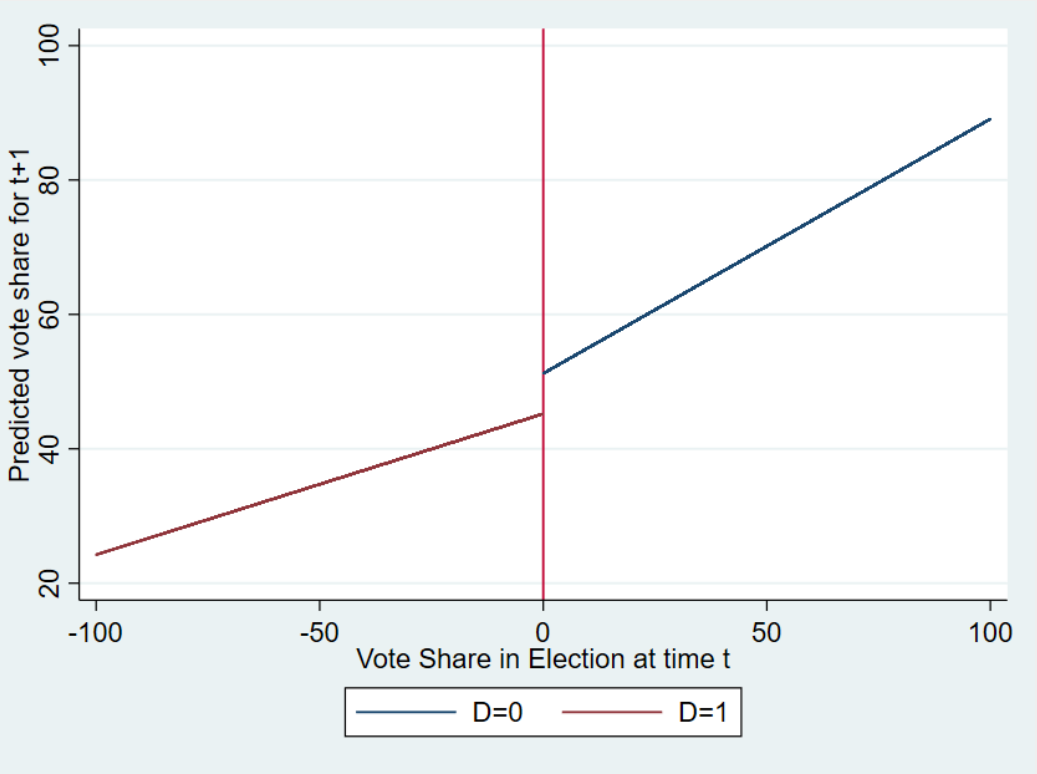
用回歸技巧：$y_i = \alpha + \tau_{ATE} d_i + \beta x_i + \gamma(d_i(x_i - \bar{x})) + e_i$

7-15

題幹

課本翻譯太爛，見 7_15 重新翻譯.pdf

斷點回歸 Regression Discontinuity



$$Y = \alpha_1 + \alpha_2 X + D(\beta_1 + \beta_2 X)$$
$$= reg\ Y\ X\ D\ XD$$