

匹配方法 第一节-第四节

朱思颖

2020.11.19

本章框架

- 6.1 匹配方法的直观理解
- 6.2 匹配方法的假设条件
- 6.3 直接匹配法
- 6.4 倾向得分匹配法原理

6.1 匹配方法的直观理解

- 随机分配下，处置组个体和观测组个体平均观测结果的差异反映了平均处置效应
- 药物效果实验
 - 假设个体潜在健康结果可表示为（可观测特征）年龄 Age_i 、（不可观测特征）情绪 e_i 和（处置效应） γ_i 的线性函数：

$$Y_i(0) = a - bAge_i + e_i, \text{ 当 } D_i = 0$$

$$Y_i(1) = a + \gamma_i - bAge_i + e_i, \text{ 当 } D_i = 1$$

- 观测结果为

$$Y_i = D_i \times Y_i(1) + (1 - D_i) \times Y_i(0) = a + \gamma_i D_i - bAge_i + e_i$$

6.1 匹配方法的直观理解

- 年龄和情绪负相关

- $E(e_i | Age_i = 30) = 60\% \times 1 + 40\% \times (-1) = 0.2$

- $E(e_i | Age_i = 50) = 40\% \times 1 + 60\% \times (-1) = -0.2$

- 总体平均处置效应

$$ATE = P(Age_i = 30) \times \gamma(30) + P(Age_i = 50) \times \gamma(50) = 15$$

表：总体人数按年龄和情绪分布情况

	(1)	(2)	(3)	(4)	(5)
		情绪			
		1	-1		
年龄	30	30%	20%	50%	$\gamma(30) = 10$
	50	20%	30%	50%	$\gamma(50) = 20$
		50%	50%		

6.1 匹配方法的直观理解

- 随机抽取2万人，并随机分配药物，产生1万人服药（处置组），1万人不服药（控制组）
 - 随机分配使处置组和控制组个体的特征分布与总体分布相同
 - 两组平均年龄一样 $E(Age_i|D_i = 1) = E(Age_i|D_i = 0) = 40$
 - 两组情绪均值一样 $E(e_i|D_i = 1) = E(e_i|D_i = 0) = 0$

表：随机分配下处置组和控制组人数按年龄和情绪分布情况

		处置组 (D=1)				控制组 (D=0)			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		情绪				情绪			
		1		-1		1		-1	
年龄	30	30%	20%	50%		30	30%	20%	50%
	50	20%	30%	50%		50	20%	30%	50%
		50%		50%		50%		50%	

6.1 匹配方法的直观理解

- 由于处置组和控制组平均年龄和平均情绪没有差异，将二者观测结果平均值相减，得到处置组的平均处置效应

$$\begin{aligned} & E(Y_i | D_i=1) - E(Y_i | D_i=0) \\ &= E(Y_i(1) | D_i=1) - E(Y_i(0) | D_i=0) \\ &= [a + E(\gamma_i | D_i=1) - bE(\text{Age}_i | D_i=1) + E(e_i | D_i=1)] \\ &\quad - [a - bE(\text{Age}_i | D_i=0) + E(e_i | D_i=0)] \\ &= E(\gamma_i | D_i=1) \end{aligned}$$

- 随机分配下三个平均处置效用相同
 - $ATT = E(\gamma_i | D_i=1) = P(\text{Age}_i = 30 | D_i=1) \times \gamma(30) + P(\text{Age}_i = 50 | D_i=1) \times \gamma(50) = 15$
 - $ATT = ATU = ATE$

6.1 匹配方法的直观理解

- 实际研究中大多是个体自选择是否接受处置后观测到的数据，因此处置组和控制组在特征上的分布不同
 - 例子：假设服药的选择受可观测变量年龄的影响，但不受不可观测变量情绪的影响
 - $E(Age_i | D_i = 1) = 46$; $E(Age_i | D_i = 0) = 34$
 - $E(e_i | D_i = 1, Age_i = 30) = E(e_i | D_i = 0, Age_i = 30) = 0.2$
 $E(e_i | D_i = 1, Age_i = 50) = E(e_i | D_i = 0, Age_i = 50) = -0.2$

	处置组 (D=1)				控制组 (D=0)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
	情绪				情绪				
		1	-1			1	-1		
年龄	30	12%	8%	20%	年龄	30	48%	32%	80%
	50	32%	48%	80%		50	8%	12%	20%
		44%	56%			56%	44%		

6.1 匹配方法的直观理解

- 处置组总体情绪均值低于控制组总体情绪均值

$$E(e_i | D_i = 1) = 44\% \times 1 + 56\% \times (-1) = -0.12$$

$$E(e_i | D_i = 0) = 56\% \times 1 + 44\% \times (-1) = 0.12$$

- 原因：年龄较大个体容易选择处置，且其个体情绪均值较低

- 处置组平均观测结果与控制组平均观测结果的差异不仅包括药物治疗效果，还包括两组平均年龄造成的差异以及由于年龄差异造成的情绪均值差异

- $E(Y_i | D_i = 1) - E(Y_i | D_i = 0)$

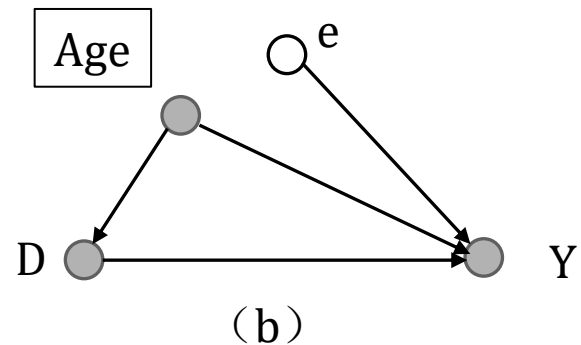
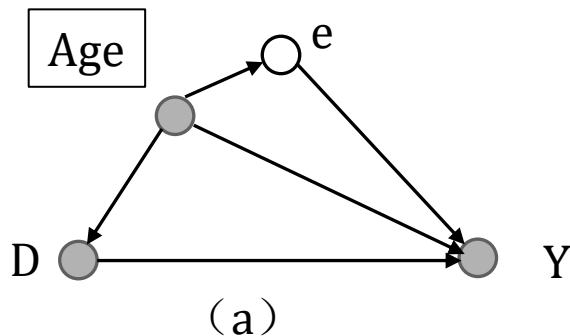
$$= [a + E(\gamma_i | D_i = 1) - bE(\text{Age}_i | D_i = 1) + E(e_i | D_i = 1)] - [a - bE(\text{Age}_i | D_i = 0) + E(e_i | D_i = 0)]$$

$$= E(\gamma_i | D_i = 1) - b[E(\text{Age}_i | D_i = 1) - E(\text{Age}_i | D_i = 0)] + E(e_i | D_i = 1) - E(e_i | D_i = 0)$$

$$= E(\gamma_i | D_i = 1) - b \times 12 - 0.24$$

6.1 匹配方法的直观理解

- 可观测特征 Age 是混淆变量，不可观测特征 e 和 Age 相关并且会影响结果，但并不直接影响选择变量 D
- 基于可观测特征的选择（**selection on observables**）：只根据可观测特征而选择是否接受处置
- 匹配方法通过“控制”可观测变量，即将处置组和控制组的个体按可观测特征匹配，解决基于可观测变量自选择造成的偏差



6.1 匹配方法的直观理解

- 在之前的例子里

- 将控制住和处置组个体根据年龄进行匹配，此时两组的其他特征和处置效应均值无差异

- 对30岁的匹配块，年龄差异造成的影响以及情绪差异被消除

$$E(Y_i | D_i = 1, Age_i = 30) - E(Y_i | D_i = 0, Age_i = 30)$$

$$= E(\gamma_i | D_i = 1, Age_i = 30) = ATT(Age_i = 30)$$

- $ATT(Age_i = 30) = ATU(Age_i = 30) = ATE(Age_i = 30) = \gamma(30) = 10$

- 通过匹配法分别得到30岁和50岁个体的平均处置效应后，进一步计算两组人的平均处置效应

$$ATT = P(Age_i = 30 | D = 1) \times \gamma(30) + P(Age_i = 50 | D = 1) \times \gamma(50) = 20\% \times 10 + 80\% \times 20 = 18$$

6.1 匹配方法的直观理解

■ $ATU = P(Age_i = 30|D = 0) \times \gamma(30) + P(Age_i = 50|D = 0) \times \gamma(50) = 12$; $ATE = P(30) \times \gamma(30) + P(50) \times \gamma(50) = 15$

● 自选择与随机分配下平均处置效应异同：

■ 随机分配下，处置组、控制组和总体个体特征（年龄）分布相同

$$ATE = ATT = ATU$$

■ 自选择下，处置组和控制组个体特征分布不同

$$ATE \neq ATT \neq ATU$$

■ 但两种情况下的 ATE 相等

匹配方法	回归方法	随机分配
控制组和处置组在可观测特征上匹配	观测结果方程里加入可观测变量作为控制变量	
控制可观测特征	控制可观测特征	可观测特征和不可观测特征分布一样

6.2 匹配方法的假设条件

- 匹配方法估计处置效应依赖于一个假设和一个条件

- 条件独立假设（conditional independence assumption）：

给定可观测特征后，潜在结果独立于处置状态

$$\{Y_i(1), Y_i(0)\} \perp D_i \mid X_i$$

- 通俗理解：在可观测变量 X 给定情况下，接受处置与否是随机分配的，不会因为潜在结果的好坏而决定是否接受处置

- 条件独立假设成立，则潜在结果均值条件独立假设成立

$$E(Y_i(0) \mid D_i = 1, \mathbf{X}) = E(Y_i(0) \mid D_i = 0, \mathbf{X}) = E(Y_i(0) \mid \mathbf{X})$$

$$E(Y_i(1) \mid D_i = 1, \mathbf{X}) = E(Y_i(1) \mid D_i = 0, \mathbf{X}) = E(Y_i(1) \mid \mathbf{X})$$

- 如果关心平均处置效应，则只需要潜在结果均值条件独立假设

6.2 匹配方法的假设条件

- 将 $Y_i(1)$, $Y_i(0)$ 表达式分别代入上述潜在结果均值条件独立假设, 则可得到

$$E(e_i | D_i = 1, Age_i) = E(e_i | D_i = 0, Age_i) = E(e_i | Age_i)$$

$$E(\gamma_i | D_i = 1, Age_i) = E(\gamma_i | D_i = 0, Age_i) = E(\gamma_i | Age_i)$$

- 给定相同可观测特征的处置组和控制组, 平均处置效应相同

- 共同支撑域条件 (common support condition) :

给定可观测特征 $X_i = x$, 个体接受处置的概率大于0并小于1

$$0 < P(D_i | X_i = x) < 1$$

- 当条件独立假设和共同支撑域条件满足时, 观测特征为 x 的平均处置效应:

$$E(Y_i | D_i = 1, X_i = x) - E(Y_i | D_i = 0, X_i = x)$$

$$= E(Y_i(1) | D_i = 1, X_i = x) - E(Y_i(0) | D_i = 0, X_i = x) = ATT(x)$$

6.2 匹配方法的假设条件

- 对特征为 x 的个体 $ATT(x) = ATU(x) = ATE(x)$
- 对所有接受处置个体、未接受处置个体和总体而言
 - $ATT = E_{X|D=1}[ATT(X)] = \sum_x ATT(x)P(x|D = 1)$
 - $ATU = E_{X|D=0}[ATU(X)] = \sum_x ATU(x)P(x|D = 0)$
 - $ATE = E_X[ATE(X)] = \sum_x ATE(x)P(x)$

6.3 直接匹配法

- 直接匹配法指根据可观测的特征值直接进行匹配
 - 例子：估计税收优惠政策是否会影响公司研发投入，假设数据满足潜在结果均值条件独立假设

i	D	Assets	R&D
1	0	25	6
2	0	35	8
3	0	35	8
4	0	45	9
5	0	45	9
6	0	120	20
7	1	45	10
8	1	35	8
9	1	40	9
10	1	25	7

6.3 直接匹配法

- 通过匹配法消除在估计税收优惠对研发投入因果关系中企业规模这一混淆变量影响

i	D	Assets	R&D	Match	$E[R\&D(1)]$	$E[R\&D(0)]$	ATT(Assets)	P(Assets D=1)
1	0	25	6					
2	0	35	8					
3	0	35	8					
4	0	45	9					
5	0	45	9					
6	0	120	20					
7	1	45	10	[4][5]	10	9	1	1/3
8	1	35	8	[2][3]	8	8	0	1/3
9	1	40	9					
10	1	25	7	[1]	7	6	1	1/3

6.3 直接匹配法

- $ATT(45) = E(R\&D_i(1) | D_i = 1, Assets_i = 45) - E(R\&D_i(0) | D_i = 0, Assets_i = 45) = 10 - 9 = 1$

同理，可以求得 $ATT(35) = 0$ ； $ATT(25) = 1$

- 用不同规模企业比率计算

- $ATT = \sum_x ATT(x)P(x | D_i = 1)$

$$= ATT(45)P(45 | D_i = 1) + ATT(35)P(35 | D_i = 1) + ATT(25)P(25 | D_i = 1)$$

$$P(25 | D_i = 1) = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0.667$$

- 上式中只使用了潜在结果均值条件独立假设第一条，若第二条也成立，则可进一步算得 ATU 和 ATE

6.4 倾向得分匹配法原理

- “维数的诅咒”
 - 可观测特征维数增加时，在多维上进行直接匹配比较困难，特别的，如果可观测特征还包含连续变量，无法进行直接匹配
- 倾向得分法（Propensity Score Methods , PSM）
 - Rosenbaum and Rubin: “The central role of the propensity score in observational studies for causal effects”
 - 原理：通过函数关系将多维变量 \mathbf{X} 变换为一维的倾向得分 $ps(X_i)$ 之后，再根据倾向得分进行匹配。倾向得分是可观测特征为 $X_i = x$ 的个体接受处置的概率

$$ps(X_i = x) = P(D_i = 1|X_i = x)$$

6.4 倾向得分匹配法原理

- 倾向得分可行原因:

条件独立假设 $\{Y_i(1), Y_i(0)\} \perp D_i \mid X_i$ 等价于

$$\{Y_i(1), Y_i(0)\} \perp D_i \mid ps(X_i)$$

通俗理解: 倾向得分 $ps(X_i)$ 总结了变量 X_i 中包含的所有相关信息

- 如果处置组和控制组的个体有相同的倾向得分, 两组的可观测特征分布就是均衡的

$$X_i \perp D_i \mid ps(X_i)$$

若只关心均值, 则上式可简化为

$$E(X_i \mid D_i = 1, ps(X_i)) = E(X_i \mid D_i = 0, ps(X_i))$$

- 随机分配里, 倾向得分是已知的, 但在观测数据中, 倾向得分未知, 需要进行估计

6.4 倾向得分匹配法原理

- 自选择例子中

- 每个人选择到处置组概率和年龄相关

$$P(D_i = 1 | Age_i) = \theta_0 + \theta_1 Age_i$$

只能根据观测数据估计系数 θ_0 和 θ_1

- 计算给定多维可观测特征 x 的处置效应 $ATT(x)$ 可简化为计算给定一维倾向得分 $ps(x)$ 的处置效应

$$\begin{aligned} ATT(x) &= E(Y_i(1) | D_i = 1, X_i = x) - E(Y_i(0) | D_i = 1, X_i = x) \\ &= E(Y_i(1) | D_i = 1, X_i = x) - E(Y_i(0) | D_i = 0, X_i = x) \\ &= E(Y_i(1) | D_i = 1, ps(X_i = x)) - E(Y_i(0) | D_i = 0, ps(X_i = x)) \end{aligned}$$

- 对所有 $ATT(x)$ 加权平均得到 ATT

$$ATT = E_{ps(x)|D=1}[ATT(x)] = \sum_x ATT(x)P(ps(X_i = x)|D = 1)$$