第五章 处置效应 第五节-第八节

汇报人：陈远致

2020.11.5

Overview

- 5.5 随机分配
- 5.6 控制可观测特征
- 5.7 回归方法和处置效应
- 5.8 随机分配实例：“田纳西学生/教师比例对学生学业成就影响”实验

5.5随机分配

5.5.1 随机分配的两种理解

- 理解一：**潜在结果独立性假设**

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

- 理解二：可观测特征、不可观测特征和处置效应完全独立于是否接收处置
  - 若潜在结果可以表示为可观测特征X_i、不可观测特征e_i和处置效应γ_i的函数

$Y_i(0) = a + bX_i + \mathrm{e_i}, D_i = 0$

$Y_i(1) = a + \gamma_i + bX_i + \mathrm{e_i}, D_i = 1$

$$(X_i, e_i, \gamma_i) \perp D_i$$

- 通俗理解：将总体随机分为处置组和控制组，个体的特征在总体、处置组、控制组均一致

5.5.2 潜在结果独立假设包含的两个“独立”(1)

- 独立性一：未接受处置潜在结果独立于处置变量

$$\{Y_i(0)\} \perp D_i$$

进一步，未接受处置潜在结果的均值也和D_i不相关

$$\mathbb{E}[Y_i(0) \mid \mathrm{D}_i = 0] = \mathbb{E}[Y_i(0) \mid \mathrm{D}_i = 1] = \mathbb{E}[Y_i(0)]$$

化简为：$\mathbb{E}\,[Y_i(0) \mid \mathrm{D}_i] = \mathbb{E}[Y_i(0)]$

该条件称为平均未处置潜在结果独立于处置变量条件（即 $\mathrm{T0} = \mathrm{C0}$）

- 通俗理解：可以用未处置组的观测结果C0来衡量不可观测的反事实结果T0，此时接受处置组的平均处置效应ATT无偏

$$T1 - C0 = \underbrace{(\mathrm{T1} - \mathrm{T0})}_{\mathrm{ATT}} + \underbrace{(\mathrm{T0} - \mathrm{C0})}_{\mathrm{ATT}\text{的偏差} = 0} = ATT$$

5.5.2 潜在结果独立假设包含的两个“独立”(2)

- 独立性二：接受处置潜在结果独立于处置状态

$$\{Y_i(1)\} \perp D_i$$

同理，有 $\mathbb{E}\,[Y_i(1) \mid \mathrm{D}_i] = \mathbb{E}[Y_i(1)]$

该条件称为平均接受处置潜在结果独立于处置变量条件（即 $\mathrm{T1} = \mathrm{C1}$）

- 通俗理解：可以用处置组的观测结果T1来衡量不可观测的反事实结果C1，此时未接受处置组的平均处置效应ATU无偏

$$T1\ - C0 = \underbrace{(\mathrm{C1} - \mathrm{C0})}_{\mathrm{ATU}} + \underbrace{(\mathrm{T1} - \mathrm{C1})}_{\mathrm{ATU}\text{的偏差}=\ 0} = ATU$$

5.5.3 小结

- 随机分配的假设下，有**均值独立性假设**

$$\mathbb{E}\left[Y_i(0) \mid \mathrm{D}_i\right] = \mathbb{E}[Y_i(0)],\quad \mathrm{T0} = \mathrm{C0}$$

$$\mathbb{E}\left[Y_i(1) \mid \mathrm{D}_i\right] = \mathbb{E}[Y_i(1)],\quad \mathrm{T1} = \mathrm{C1}$$

- ATT = ATU = ATE 平均处置效应没有偏差
- 随机分配可以使得处置组和控制组（未处置）的处置效应没有区别，均可以用已知的T1 和 C0去估计，看起来完美解决了因果推断的根本难点，但是现实中社会科学无法保证完全随机和对照处理，随机分配的思想指导我们进一步研究

5.6控制可观测特征

5.6.1消除选择偏差——一个例子

- 药物效果实验
  - 对处置组和控制组的年龄进行分类，控制年龄以消除不同年龄段潜在健康状况的差异同一个年龄段，处置组和控制组可以看成随机分配，满足前一节的独立性假设

表5.6 年龄为30岁的个体

潜在结果		处置情况	观测结果
如果处置	如果未处置		
T1(30) $=\mathbb{E}[Y_i(1)\mid D_i=1,X_i=30]$	T0(30) $=\mathbb{E}[Y_i(0)\mid D_i=1,X_i=30]$	D = 1	T1(30) $=\mathbb{E}[Y_i(1)\mid D_i=1,X_i=30]$
C1(30) $=\mathbb{E}[Y_i(1)\mid D_i=0,X_i=30]$	C0(30) $=\mathbb{E}[Y_i(0)\mid D_i=0,X_i=30]$	D = 0	C0(30) $=\mathbb{E}[Y_i(1)\mid D_i=0,X_i=30]$

  - ATT (30) = ATU(30) = ATE (30) = T1(30) — C0(30)
  - ATT (40) = ATU(40) = ATE (40) = T1(40) — C0(40)
  - ATT = P(30 | D = 1) × ATT(30) + = P(40 | D = 1) × ATT(40)

5.6.1消除选择偏差——一般化(1)

- 对于给定可观测特征 $X_i = x$ 的处置组和控制组

表5.6 年龄为30岁的个体

潜在结果		处置情况	观测结果
如果处置	如果未处置		
$\mathrm{T1}(x) = \mathbb{E}[Y_i(1) \mid D_i = 1, X_i = x]$	$\mathrm{T0}(x) = \mathbb{E}[Y_i(0) \mid D_i = 1, X_i = x]$	$\mathrm{D} = 1$	$\mathrm{T1}(x) = \mathbb{E}[Y_i(1) \mid D_i = 1, X_i = x]$
$\mathrm{C1}(x) = \mathbb{E}[Y_i(1) \mid D_i = 0, X_i = x]$	$\mathrm{C0}(x) = \mathbb{E}[Y_i(0) \mid D_i = 0, X_i = x]$	$\mathrm{D} = 0$	$\mathrm{C0}(x) = \mathbb{E}[Y_i(0) \mid D_i = 0, X_i = x]$

- $ATT(x) = T1(x) - C0(x)$
- $ATT = \sum_x P(x \mid D = 1) \times \mathrm{ATT}(x)$

5.6.1消除选择偏差——一般化(2)

- 则有，$\text{ATE} = \mathbb{E}_{\text{x}}[\text{ATE}(\text{X})] = \sum_x ATE\,(x)$

  该假设称为：**条件均值独立假设** CMI

  $$\mathbb{E}\,[Y_i(0) \mid D_i = 1, X_i = x] = \mathbb{E}[Y_i(0) \mid D_i = 0, X_i = x] = \mathbb{E}[Y_i(0) = x]$$

  $$\mathbb{E}\,[Y_i(1) \mid D_i = 1, X_i = x] = \mathbb{E}[Y_i(1) \mid D_i = 0, X_i = x] = \mathbb{E}[Y_i(0) = x]$$

- 满足CMI 最直接的方式是条件随机分配，例如给定30岁个体，随机抽取一些人服药、一些人不服药

- CMI只关心平均处置效应，更强的假设是**条件独立假设CIA**

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

5.6.2 小结

- 控制可观测变量方法的问题：
  - X只能是可观测的变量，个体还有不可观测的特征
  - 条件独立假设或均值条件独立假设太强

5.7 回归方法和处置效应

从潜在结果和处置效用角度理解回归方法得到的回归函数要反映因果关系的条件

5.7.1 处置效用和回归函数系数的关系(1)

- 由第二节知，

$$Y_i = Y_i(0) + [Y_i(1) + Y_i(0)] \times \mathrm{D}_i$$

改写为： $\mathrm{Y_i} = \underbrace{\mathbb{E}[\mathrm{Y_i}(0)]}_{a} + \underbrace{[Y_i(1) - Y_i(0)] \times \mathrm{D}_i}_{\gamma} + \underbrace{Y_i(0) - \mathbb{E}[Y_i(0)]}_{u_i}$

$$= a + \gamma \times D_i + u_i$$

系数：a：所有个体未处置潜在结果均值

γ：处置效应（假设所有个体处置效应相同）

u_i：第i个个体未处置时的潜在结果和所有个体未处置时的平均潜在结果之差

5.7.1 处置效用和回归函数系数的关系(2)

- 将Y_i对D_i进行回归：

$$\mathbb{E}(Y_i \mid D_i) = a + \gamma \times D_i + \mathbb{E}(u_i \mid \mathrm{D_i})$$

- 要使回归得到的条件期望函数D_i的系数等于处置效应γ

$$\mathbb{E}(u_i \mid D_i) = \phi_i$$

干扰项必须满足均值独立于处置变量，即u_i的均值不随D_i而改变

- 将 $u_i = Y_i(0) - \mathbb{E}[Y_i(0)]$ 代入上式：

$$\mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_i(0) \mid D_i = 0]$$

意味着，干扰项均值独立等价于未处置潜在结果均值独立

5.7.1 处置效用和回归函数系数的关系(3)

- 满足干扰项均值独立的数值例子

（1）	（2）	（3）	（4）	（5）	（6）
个体	$Y_i(1)$	$Y_i(0)$	$u_i = Y_i(0) - \mathbb{E}[Y_i(0)]$	D_i	Y_i
1	30	20	5	1	30
2	20	10	-5	1	20
3	35	25	10	0	25
4	15	5	-10	0	5
	$\mathbb{E}[Y_i(1)] =$ 25	$\mathbb{E}[Y_i(0)] =$ 15			

$$\mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_i(0) \mid D_i = 0] = 15$$

$$\mathbb{E}(u_i \mid D_i) = \phi_i = 0$$

5.7.1 处置效用和回归函数系数的关系(4)

- Stata： reg Y D

得出：$\mathbb{E}(Y_i \mid D_i) = 15 + 10 \times D_i$

回顾：$\mathrm{Y_i} = \underbrace{\mathbb{E}[\mathrm{Y_i}(0)]}_{a} + \underbrace{[Y_i(1) - Y_i(0)] \times \mathrm{D}_i}_{\gamma} + \underbrace{Y_i(0) - \mathbb{E}[Y_i(0)]}_{u_i}$

5.7.1 处置效用和回归函数系数的关系(5)

- 不满足干扰项均值独立的数值例子

（1）	（2）	（3）	（4）	（5）	（6）
个体	$Y_i(1)$	$Y_i(0)$	$u_i = Y_i(0) - \mathbb{E}[Y_i(0)]$	D_i	Y_i
1	30	20	10	1	30
2	20	10	0	1	20
3	16	6	-4	0	6
4	14	4	-6	0	4
	$\mathbb{E}[Y_i(1)] = 20$	$\mathbb{E}[Y_i(0)] = 10$			

$$\mathbb{E}[Y_i(0) \mid D_i = 1] = 15 \;；\; \mathbb{E}[Y_i(0) \mid D_i = 0] = 5$$

$$\mathbb{E}(u_i \mid D_i = 1) = 5 \;；\; \mathbb{E}(u_i \mid D_i = 0) = -5$$

5.7.1 处置效用和回归函数系数的关系(6)

- Stata： reg Y D

得出：$\mathbb{E}(Y_i \mid D_i) = 5 + 20 \times D_i$

错误的平均处理效应

回顾： $\mathrm{Y_i} = \underbrace{\mathbb{E}[\mathrm{Y_i}(0)]}_{a} + \underbrace{[Y_i(1) - Y_i(0)] \times \mathrm{D}_i}_{\gamma} + \underbrace{Y_i(0) - \mathbb{E}[Y_i(0)]}_{u_i}$

事实上，由OLS得到的D_i的系数 $= \mathbb{E}(Y_i \mid D_i = 1) - \mathbb{E}(Y_i \mid D_i = 0)$

$=(a + \gamma \times D_i + \mathbb{E}[u_i \mid D_i = 1]) - (a + \mathbb{E}[u_i \mid D_i = 0])$

$= \gamma + \mathbb{E}[u_i \mid D_i = 1] - \mathbb{E}[u_i \mid D_i = 0]$

$= 10 + 5 + 5 = 20$

5.7.2回归方法和控制变量(1)

- $\mathbb{E}[Y_i(0) \mid D_i = 1] \neq \mathbb{E}[Y_i(0) \mid D_i = 0]$时，考虑造成差异的原因是个体的潜在未处置结果是个体未处置时可观测特征和不可观测特征的线性函数

$$Y_i(0) = \alpha + \beta X_i + e_i$$

代入方程: $\mathrm{Y_i} = \underbrace{\mathbb{E}[\mathrm{Y_i}(0)]}_{\alpha} + \underbrace{[Y_i(1) - Y_i(0)] \times \mathrm{D}_i}_{\gamma} + \underbrace{Y_i(0) - \mathbb{E}[Y_i(0)]}_{u_i}$

得：$Y_i = \alpha + \gamma D_i + \beta X_i + e_i$（观测结果、处置状态、观测特征、不可观测特征的关系）

将Y_i对D_i、X_i归回：$\mathbb{E}(\mathrm{Y_i} \mid \mathrm{D_i}, \mathrm{X_i}) = \alpha + \gamma D_i + \beta X_i + \mathbb{E}[e_i \mid \mathrm{D_i}, \mathrm{X_i}]$

5.7.2回归方法和控制变量(2)

- 同之前思路一样，若要使条件期望函数的D_i 的系数等于 γ ，需要干扰项e_i的条件均值与处置变量无关

  即$\mathbb{E}[e_i \mid \mathrm{D_i}, \mathrm{X_i}] = \mathbb{E}[e_i \mid X_i]$

  也即 $\mathbb{E}[e_i \mid \mathrm{D_i} = 1, \mathrm{X_i}] = \mathbb{E}[e_i \mid \mathrm{D_i} = 0, \mathrm{X_i}]$

  该条件称为干扰项条件均值独立于处置变量

  将 $e_i = \alpha + \beta X_i - Y_i(0)$代入上式：

  即$\mathbb{E}[\alpha + \beta X_i - Y_i(0) \mid \mathrm{D_i} = 1, \mathrm{X_i}] = \mathbb{E}[\alpha + \beta X_i - Y_i(0) \mid \mathrm{D_i} = 0, \mathrm{X_i}]$

  $$\mathbb{E}[Y_i(0) \mid \mathrm{D_i} = 1, \mathrm{X_i}] = \mathbb{E}[Y_i(0) \mid \mathrm{D_i} = 0, \mathrm{X_i}]$$

  这意味着，干扰项均值条件独立和平均未处置潜在结果条件独立是等价的

- 这个条件使得回归方法可以通过使用控制变量X 来达到估计处置变量D的因果效应系数γ 的目的

5.7.3 控制变量(1)

- 正式定义控制变量 X_i：

  能够让干扰项e_i条件均值独立于处置变量D_i的变量，即给定了控制变量后，干扰项与处置变量无关

控制了X

$$Y_i = \alpha + \gamma D_i + \beta X_i + e_i$$

D_i 和 e_i 不相关

5.7.3 控制变量(2)

- 给定控制变量 X_i ，当D_i发生变化时，e_i的均值不发生变化，解释变量Y_i的均值变化就可以完全归因于D_i的变化，从而识别出处置效应
- 只是保证了D_i的系数是γ ，不能保证控制变量X_i 的系数β无偏估计
- 进一步将解释变量区分为：处置变量和控制变量

5.8 随机分配实例：“田纳西学生/教师比例对学生学业成就影响”实验

5.8.1实验的理论基础(1)

- 为了振兴经济，提高教育质量，应对入学人数减少，在选民的强烈要求下，20世纪八十年代美国田纳西州进行了迄今最具权威性的小班化教育改革实验STAR计划。它耗资1200万，历时4年，实验学生达12,000名，被誉为美国历史上最伟大的教育实验。STAR计划及后续研究表明，小班化教育有利于学生学业成就的提高，尤其对处境不利学生更具优势，并且这种优势具有累积性
- 学校：79所
- 学生：随机分配到三组班级：普通班（22-25）；加强普通班（22-25，加辅导老师）；小班（13-17）
- 处置变量：班级类型
- 潜在结果：学习成绩

5.8.1实验的理论基础(2)

$$\begin{cases} Y_i(regular), & D_i = regular \\ Y_i(regularaid), & D_i = regularaid \\ Y_i(small), & D_i = small \end{cases}$$

- 实验目的：研究小班和加强普通班相对于普通班对学习成绩的影响
- 小班和加强普通班的平均处置效应：

$$ATE(small) = \mathbb{E}[Y_i(small)] - \mathbb{E}[Y_i(regular)]$$
$$ATE(regularaid) = \mathbb{E}[Y_i(regularaid)] - \mathbb{E}[Y_i(regular)]$$

5.8.1实验的理论基础(3)

- 给定学校g，学生的潜在成绩和班级类型是独立的，即CIA条件独立假设

$$\{Y_i(small), Y_i(regularaid), Y_i(regular) \perp D_i \mid school_i\}$$

  对于同一个学校，不同班级的学生没有区别

- $ATE(small) = \sum_{\mathrm{i}} p_i \, ATE(small \mid school_i)$

  p_i为每个学校学生占所有学校学生的比例

5.8.1实验的理论基础(4)

- 使用回归方法，有如下线性关系：

  $$Y_{gi}=\alpha+\gamma_1\, small_{gi}+\gamma_2 regularaid_{gi}+\beta school_g+e_{gi}$$

  Y_{gi}：学校g学生i的成绩

  $small_{gi}=1$：学生上小班，否则为0

  $regularaid_{gi}=1$：学生上加强普通班，否则为0

  α：普通班的效应

  β：学校固定效应

  - 条件随机分配满足干扰项条件均值独立：

    $$\mathbb{E}(e_{gi}\,|\,small_{gi}, regularaid_{gi}, regular_{gi}, school_g)=\mathbb{E}(e_{gi}\,|school_g)$$

  - 基于以上分析，该回归方程系数反映了因果关系

5.8.2回归分析(1)

- 截取了STAR实验中一年级学生的数据
- Step1： 检测是否随机分配？

  分别看学生的性别、年龄、人种在不同班级类型中是否有区别

  Stata命令：

  table glclasstype ,c(mean female mean whiteasian mean age1985 mean glclasssize) cell(18)

  结果：除了小班平均人数少于加强普通班和普通班，其他均无差异

5.8.2回归分析(2)

- 没有控制学校固定效应的模型：

$$Y_{gi} = \alpha + \gamma_1\ small_{gi} + \gamma_2 regularaid_{gi} + \mathrm{e_{gi}}$$

- Stata命令：

```
gen small = (glclasstype == 1)
gen regular = (glclasstype == 2)
gen regularaid = (glclasstype == 3)
reg score small regularaid
```

- 结果：

score	Coeff.	Std.Err.
small	29.78802	2.8076
regularaid	11.93308	2.6862
_cons	1039.393	1.8361

5.8.2回归分析(3)

- 查看平均处置效应ATE
- Stata命令：

table glclasstype ， c(mean score)

glclasstype	Mean(score)
1	1069.181
2	1039.393 PS：普通班，常数项
3	1051.326

5.8.2回归分析(4)

- 加入学校固定效应的模型：

$$Y_{gi} = \alpha + \gamma_1\ small_{gi} + \gamma_2 regularaid_{gi} + \beta school_g + e_{gi}$$

- Stata 命令：

xi： reg score small regularaid i.glschid

- 结果：

score	Coef.
small	29.005
regularaid	7.217
_Iglschid_123056	55.609
_Iglschid_128076	43.541
_Iglschid_128079	47.238

对这两项系数影响不太大

5.8.2回归分析(5)

- 还可以加入对学习成绩有解释力度的其他控制变量，如性别，年龄，人种，加入这些解释变量有助于降低干扰项方差，提高估计精度

固定效应	Root MSE	Adj R—square
不加参数	90.512	0.0171
固定学校	79.17	0.2480
固定性别，年龄，人种	78.084	0.2685

- PS :虽然女生的平均成绩高于男生，但性别在最开始随机分配，故不影响不同班级类型对学习成绩的处置效应的估计

5.8.2回归分析(6)

- 最后，需要考虑集群方差

  Stata命令：

  xi reg score small regularaid female whiteasian age1985 i.glschid, cluster(glschid)

  结果：标准误差变大，t值下降，说明存在显著的集群方差