

1. 四个研究相关问题:

- 研究对象间的关系: **因果关系是什么?**
- 理想条件下的实验
- 识别策略 **FUA: 不能被任何实验回答的研究问题.**
- 推断模式

2. 最可信的研究设计应该使用随机分配.

$$\cdot \text{潜在结果} = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

$$\cdot \text{观察结果 } Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

· 平均效应:

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

$$= \underline{E(Y_{1i} | D_i = 1) - E(Y_{0i} | D_i = 1)} \quad \text{处理平均因果效应}$$

$$+ \underline{E(Y_{0i} | D_i = 1) - E(Y_{0i} | D_i = 0)}$$

选择性偏误

$$\text{其中, } E(Y_{1i} | D_i = 1) - E(Y_{0i} | D_i = 1) = E(Y_{1i} - Y_{0i} | D_i = 1)$$

$$\cdot \text{随机分配: } \begin{cases} E(Y_{0i} | D_i = 1) = E(Y_{0i} | D_i = 0) \\ E(Y_{1i} | D_i = 1) = E(Y_{0i} | D_i = 0) \end{cases}$$

$$\begin{aligned} E(Y_{1i} | D_i = 1) - E(Y_{0i} | D_i = 0) &= E(Y_{1i} - Y_{0i} | D_i = 1) \\ &= E(Y_{1i} - Y_{0i}) \end{aligned}$$

随机分配消除了选择性偏误!

3. 对实验的回归分析:

No. 2.

Date. / /

假设因果效应对所有人都不一样, 即 $Y_{1i} - Y_{0i} = p$ 为常数

$$Y_i = \underbrace{\alpha}_{E(Y_{0i})} + \underbrace{p}_{(Y_{1i}-Y_{0i})} D_i + \underbrace{\eta_i}_{Y_{0i}-E(Y_{0i})}$$

$$\begin{cases} E(Y_i | D_i=1) = \alpha + p + E(\eta_i | D_i=1) \end{cases}$$

$$\begin{cases} E(Y_i | D_i=0) = \alpha + E(\eta_i | D_i=0) \end{cases}$$

$$E(Y_i | D_i=1) - E(Y_i | D_i=0) = p + \underbrace{E(\eta_i | D_i=1) - E(\eta_i | D_i=0)}$$

处理效应 选择性偏差

$$E(\eta_i | D_i=1) - E(\eta_i | D_i=0)$$

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

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

选择性偏差意味着回归残差项 η_i 和回归元 D_i 之间的相关性。

$$Y_i = \alpha + p D_i + \underline{X_i' \gamma} + \eta_i$$

控制变量:

若 X_i 与 D_i 无关, 但对 Y_i 有解释力度, 将这些控制变量包含进回归可以减小残差方差, 从而降低回归的标准误。