



THE WORLD BANK



技术路线 第一部分： 因果推断

中国北京
2009

动机

- 医学领域大多数研究问题的本质都是在探讨因果关系
- 例如:
 - 某种药物对某类人群的功效如何?
 - 某种疾病经过某种治疗或政策干预后，其死亡率的降低程度如何?

动机

- 经济学中最富挑战性的实证性问题也涉及因果关系：
 - 学校实行地方分权化之后教学质量是否有改进？
 - 多接受一年教育能否 **带来** 更高的收入？
 - 有条件的现金转移能否 **使** 儿童更加健康？

动机

以下方面内容会引起人们对这些问题的研究兴趣:

- 政策关注
 - 公共项目能否减少贫困?
- 理论考虑
- 决策者面临的问题

对问题的感性认识： 一个假设的例子

- 一个针对孕期妇女的有条件的现金转移项目
 - 她们会定期进行体检
 - 项目集中在营养状况良好的地区
 - 定期开展活动，宣传吸烟和饮酒的风险

- 关键问题：该项目能否影响儿童的出生体重？

有条件的现金转移（CCT）与出生体重

- 假设你的数据来自：一个开展CCT的乡镇的妇女和一个临近的未开展CCT的乡镇的妇女

Y = 平均出生体重（克）				
	开展CCT乡镇的 妇女	临近的未开展CCT 的乡镇的妇女	差异	二次差异
开展CCT之前	?	?		
开展CCT之后	3,250	3,100	150	?

有条件现金转移（CCT）与出生体重

- 假设你的数据来自：实施项目的一个乡镇的妇女和邻近的一个未实施项目的乡镇的妇女

Y = 平均出生体重（克）				
	开展CCT乡镇的 妇女	临近的未开展CCT 的乡镇的妇女	差异	二次差异
开展CCT之前	3,025	2,840	185	
开展CCT之后	3,250	3,100	150	-35

有条件的现金转移（CCT）与出生体重

- 假设有关于随机选出参与本项目的妇女的一些数据

Y = 平均出生体重（克）			
	开展CCT乡镇的妇女	临近的未开展CCT的乡镇的妇女	差异
开展CCT之前	3,025	2,840	185
开展CCT之后	3,250	3,100	150
	开展CCT乡镇的妇女	临近乡镇的妇女随机选择	差异
开展CCT之前	3,028	3,028	0
开展CCT之后	3,250	3,105	145

标准统计分析

- 工具：可能性、其它预测技术
- 目标：从一个分布的样本来推断这个分布的参数
- 应用：借助参数可以：
 - 推断变量之间的关联
 - 估计已发生事件和未发生事件的可能性
 - 根据新证据或新测量方法重新估计事件的概率
- 条件：实验条件要一致

因果分析

- 因果分析比标准统计分析更进一步
- 旨在推断数据生成过程的各个方面
- 借助这些方面，可以推论：
 - **静态**（如在标准统计分析中）**条件下**，事件发生的可能性
 - **和条件变化时**，事件的动态变化情况

因果分析

- “条件改变时事件的动态变化”包括：
 1. 预测干预效应
 2. 预测自发变化引起的效应
 3. 找出该事件的起因

因果关系与相关关系

- 标准统计分析/概率论：
 - 没有“**原因**”这一词汇
 - 我们能说两个事件互相**关联**，或相互依赖
 - 发现一个事件，就能发现另一个事件
- 这对政策制定人而言是不够的
 - 他们需要为政策决策寻找合理依据：假如我们做了XXX，是否就能得到YYY？
 - 因此必须在概率语言中增添**因果关系**这一词汇

鲁宾因果关系模型

- 人群为 U .
 U 的每个单元记为 u .
- 对每一单元 $u \in U$, 研究的变量 Y 均有一个对应值 $Y(u)$,称为:反应变量
- 鲁宾的立场是原因必须能成为假设试验中的处理因素

- 简言之，我们假设处理因素仅有两个原因或干预层面
- 令 D 表示 U 中暴露单元的变量的原因

$$D = \begin{cases} 1 & \text{假设单元 } u \text{ 受到干预} \\ 0 & \text{假设单元 } u \text{ 作为对照} \end{cases}$$

- 在一个设对照组的研究中， D 由试验者构建
- 在一个不设对照组的研究中， D 由不受试验者控制的因素所决定

- 因变量 Y 潜在地受 u 是否受到干预所影响
- 因此我们需要两个因变量：

$Y_1(u)$ 是单元 u 受到干预的结果

$Y_0(u)$ 是单元 u 作为对照的结果

$$D = \begin{cases} 1 & \text{假设单元}u\text{受到干预} \\ 0 & \text{假设单元}u\text{作为对照} \end{cases}$$

$Y_1(u)$ 是单元 u 受到干预的结果

$Y_0(u)$ 是单元 u 作为对照的结果

↓

那么每个单元 u 的结果就是：

$$Y(u) = DY_1(u) + (1 - D)Y_0(u)$$

注：此定义假设对某一单元的干预不影响其他单元的潜在结果

定义：对于每一个单元 u 而言，干预引起的效应为：

$$\delta_u = Y_1(u) - Y_0(u)$$

因果推断的基本问题是：

对于某个 u ，我们只能观察到 $Y_1(u)$ 或 $Y_0(u)$

我们无法同时观察到同一单元 u 的 $Y_1(u)$ 值和 $Y_0(u)$ 值 u
 \Rightarrow 不可能通过 u 本身观察到干预对 u 的效应

问题：我们对于单元 u 没有虚拟条件的证据，

即假如不暴露于处理因素， u 会发生什么

nt.

- 假设无法观察到单个单元 u 的干预效应，我们可以将目标转向确定总体 U （或亚总体）的**平均干预效应**
- 总体 U （或亚组）的平均干预效应（ATE）：

$$TE_u = \delta_u = Y_1(u) - Y_0(u)$$

⇓

$$\begin{aligned}ATE_U &= E_U [Y_1(u) - Y_0(u)] \\ &= E_U [Y_1(u)] - E_U [Y_0(u)] \\ &= \bar{Y}_1 - \bar{Y}_0 \\ &= \bar{\delta} \quad (1)\end{aligned}$$

- 统计学的解决方式是：用可以估计的总体 U 的 **平均**干预效应 t ，来代替无法观察到的一个单元 u 的干预效应 t
- $E_U(Y_1)$ 和 $E_U(Y_0)$ 无法计算，但可以估计
- 多数计量经济学方法试图从估计一致的观察数据中得出

$$E_U(Y_1) = \bar{Y}_1 \quad \text{以及} \quad E_U(Y_0) = \bar{Y}_0$$

因此我们想办法估计:

$$\begin{aligned}ATE_U &= E_U [Y_1(u)] - E_U [Y_0(u)] \\ &= \bar{Y}_1 - \bar{Y}_0\end{aligned}\tag{1}$$

可以考虑对 ATE_U 进行如下简单估计:

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0]\tag{2}$$

- 等式 (1) 适用于总体
- 等式 (2) 是该总体中样本的估计值

推论：若假设

$$[\bar{Y}_1 | D = 1] = [\bar{Y}_1 | D = 0]$$

而且 $[\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$

那么

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0]$$

是

$$\bar{\delta} = \bar{Y}_1 - \bar{Y}_0$$

的一致性估计。

- 因此，用简单估计作真实ATE值一致估计的充分条件是：

$$[\bar{Y}_1 | D = 1] = [\bar{Y}_1 | D = 0]$$

以及

$$[\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$$

- 受到干预的平均结果 \bar{Y}_1 与干预组(D=1)及对照组 (D=0) 的相同
- 作为对照的平均结果 \bar{Y}_0 与干预组(D=1)及对照组 (D=0) 的相同

这些条件在什么情况下能得到满足？

- 干预因素分配 D 与 Y_0 和 Y_1 可能的分布结果无关联
 - 直觉: 以下无关联
 - 某人是否受到干预
 - 某人在干预中的受益程度
- 达到无关联性的最简单方法是随机分配干预因素

阐述该问题的另一种方式

□ 运算可得:

$$\underbrace{\hat{\delta}}_{\text{简单估计}} = \underbrace{\bar{\delta}}_{\text{真实效应}} + \underbrace{\left([\bar{Y}_0 | D = 1] - [\bar{Y}_0 | D = 0] \right)}_{\text{基线差异}}$$
$$+ (1 - \pi) \underbrace{\left(\delta_{\{D=1\}} - \delta_{\{D=0\}} \right)}_{\text{干预因素的异质性}}$$

阐述该问题的另一种方式（用文字描述）

- 观察性研究估计因果效应，需要消除两种来源的偏倚
 - 基线差异（选择性偏倚）
 - 干预因素的异质性
- 目前多数方法仅能处理选择性偏倚

实验组的干预

- ATE 不一定是我们感兴趣的参数
- 通常 *干预组*的平均干预效应才是最受关注的是：

$$\begin{aligned}TOT &= E [Y_1(u) - Y_0(u) | D = 1] \\ &= E [Y_1(u) | D = 1] - E [Y_0(u) | D = 1]\end{aligned}$$

实验组的干预

若要估计 TOT

$$TOT = E [Y_1(u) | D = 1] - E [Y_0(u) | D = 1]$$

假设: TOT 的一致估计为,

$$[\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$$

那么简单估计(2)

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0]$$

“干预组和对照组没有基线差异”

参考文献

- Judea Pearl (2000): Causality: Models, Reasoning and Inference, Cambridge University press. (Book) Chapters 1, 5 and 7.
- Trygve Haavelmo (1944): “The probability approach in econometrics,” *Econometrica* 12, pp. iii-vi+1-115.
- Arthur Goldberger (1972): “Structural Equations Methods in the Social Sciences,” *Econometrica* 40, pp. 979-1002.
- Donald B. Rubin (1974): “Estimating causal effects of treatments in randomized and nonrandomized experiments,” *Journal of Educational Psychology* 66, pp. 688-701.
- Paul W. Holland (1986): “Statistics and Causal Inference,” *Journal of the American Statistical Association* 81, pp. 945-70, with discussion.