

# 因果推断入门

## Introduction to Causal Inference

杨点溢

Department of Government  
London School of Economics and Political Science

February 23, 2024



# 本节课内容

1 回顾相关性 Correlation

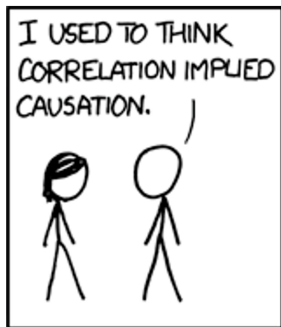
2 回顾因果关系

3 一些别的概念和问题



# 经典老番

我曾经以为相关性意味着因果性



在上过一节统计课后，我现在不这么认为了



听起来这节课挺管用的



- 我们说相关性不意味着因果性，不过相关性是什么？



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)
- 但现实生活中大家往往忽视有变化才构成相关性
  - 很多“专家”也犯这个低级错误



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)
- 但现实生活中大家往往忽视有变化才构成相关性
  - 很多“专家”也犯这个低级错误
- 例 1：美国中学生辍学（讲座里提到的例子）



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)
- 但现实生活中大家往往忽视有变化才构成相关性
  - 很多“专家”也犯这个低级错误
- 例 1：美国中学生辍学（讲座里提到的例子）
- 例 2：孩子被教坏





# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)
- 但现实生活中大家往往忽视有变化才构成相关性
  - 很多“专家”也犯这个低级错误
- 例 1：美国中学生辍学（讲座里提到的例子）
- 例 2：孩子被教坏
  - 美国的 PMRC 组织声称摇滚乐的“不好”歌词导致了青少年自杀和犯罪的高发
- 中国的比较案例：游戏，喜羊羊？



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)
- 但现实生活中大家往往忽视有变化才构成相关性
  - 很多“专家”也犯这个低级错误
- 例 1：美国中学生辍学（讲座里提到的例子）
- 例 2：孩子被教坏
  - 美国的 PMRC 组织声称摇滚乐的“不好”歌词导致了青少年自杀和犯罪的高发
    - 依据：1997-2000 年内的 37 起学校恶性案件中，过半数袭击者通过电影、游戏、图书和其他媒介接触暴力。
  - 中国的比较案例：游戏，喜羊羊？



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)
- 但现实生活中大家往往忽视有变化才构成相关性
  - 很多“专家”也犯这个低级错误
- 例 1：美国中学生辍学（讲座里提到的例子）
- 例 2：孩子被教坏
  - 美国的 PMRC 组织声称摇滚乐的“不好”歌词导致了青少年自杀和犯罪的高发
    - 依据：1997-2000 年内的 37 起学校恶性案件中，过半数袭击者通过电影、游戏、图书和其他媒介接触暴力。
  - 中国的比较案例：游戏，喜羊羊？
  - 但是这构成“变化”吗？（更不用说因果关系）



# 重温相关性

- 相关性：A 变高时伴随着 B 变高/低 (Higher A is accompanied with higher/lower B)
  - 重点：变化 (Variation)
- 但现实生活中大家往往忽视有变化才构成相关性
  - 很多“专家”也犯这个低级错误
- 例 1：美国中学生辍学（讲座里提到的例子）
- 例 2：孩子被教坏
  - 美国的 PMRC 组织声称摇滚乐的“不好”歌词导致了青少年自杀和犯罪的高发
    - 依据：1997-2000 年内的 37 起学校恶性案件中，过半数袭击者通过电影、游戏、图书和其他媒介接触暴力。
  - 中国的比较案例：游戏，喜羊羊？
  - 但是这构成“变化”吗？（更不用说因果关系）
    - 需要对比接触这些媒介和不接触这些媒介的少年哪波**犯罪率**更高



# 测试一下对相关性的掌握

- (失败) 后的反思对吗？



# 测试一下对相关性的掌握

- (失败) 后的反思对吗？

- Post-Mortem Analysis 事后分析
- 案例：乐队演出 (Bueno de Mesquita and Fowler., 2021)

	Do well	Do poorly	Total
Extra rehearsals	?	80	?
Take it easy	?	8	?
Total	?	88	?



# 我们必须确保自己知道事情全貌 (Bueno de Mesquita and Fowler., 2021)

To do a better job, you could look at the history of all the band competitions you've participated in, whether you performed well or poorly. Now you have variation in both variables and can fill in all the data, as shown in Table 4.5.

	Do well	Do poorly	Total
Extra rehearsals	300	80	380
Take it easy	12	8	20
Total	312	88	400



# 我们必须确保自己知道事情全貌 (Bueno de Mesquita and Fowler., 2021)

To do a better job, you could look at the history of all the band competitions you've participated in, whether you performed well or poorly. Now you have variation in both variables and can fill in all the data, as shown in Table 4.5.

	Do well	Do poorly	Total
Extra rehearsals	300	80	380
Take it easy	12	8	20
Total	312	88	400

- 总结成功人士的经验对吗？





# 我们必须确保自己知道事情全貌 (Bueno de Mesquita and Fowler., 2021)

To do a better job, you could look at the history of all the band competitions you've participated in, whether you performed well or poorly. Now you have variation in both variables and can fill in all the data, as shown in Table 4.5.

	Do well	Do poorly	Total
Extra rehearsals	300	80	380
Take it easy	12	8	20
Total	312	88	400

- 总结成功人士的经验对吗？
  - 成功人士不见得知道失败者的经验



# 本节课内容

1 回顾相关性 Correlation

2 回顾因果关系

3 一些别的概念和问题



# 因果关系 Causation

- "A causal effect is a change in some feature of the world that would result from a change to other feature of the world."  
(Bueno de Mesquita and Fowler., 2021)
  - 一些情况的变化导致另外一些情况的变化
- 因果关系是对显示和**反事实**(counterfactuals) 的比较
  - 反事实：一个某特征不同的假想世界



# 一些符号 Notation

- 标注对象/单位：下表  $i$  指代第  $i$  个对象/单位， $1 \leq i \leq N$ ， $N$  为对象总数（样本量）。
- 处理 Treatment: 对实验对象的干预 A manipulation or intervention administered to subjects.
- 定义处理：变量  $d_i$  代表第  $i$  个对象有没有在实验组 (treatment group)
  - $d_i = 1$  代表第  $i$  个对象属于实验组 (treatment group)
  - $d_i = 0$  代表第  $i$  个对象属于对照组 (control group)



# 潜在结果 Potential Outcome

- 潜在结果：如果第  $i$  个对象（没有）收到了处理，它的结果  $Y$  会是多少
  - $Y$  是我们感兴趣的因变量 (dependent variable)
- 潜在结果可以用  $Y_i(d)$  来表示， $d$  代表有没有受到处理。
  - $Y_i(1)$  是第  $i$  个对象受处理的潜在结果
  - $Y_i(0)$  是第  $i$  个对象没有受处理的潜在结果
- 我们假设潜在结果是固定的——哪怕它不一定可被观测 (not necessarily observable)。
  - 稳定性假设 (SUTVA)



# 稳定性假设 (SUTVA)

## Stable Unit Treatment Value Assumption

- AKA no interference/Individualized Treatment Reponse
- 单位 ( $i$ ) 的潜在结果不受到其他单位 ( $i$ ) 的处理情况 (treatment assignment) 影响。
- 处理没有溢出效应 (no spillover)

举例：补课对成绩的影响

- SUTVA 要求小丽补课不会影响小明的成绩
- 小丽补课可以带动小明补课 → 不违反 SUTVA



## 例子：补课对成绩有什么影响？

在 SUTVA 满足的条件下，我们可以画出以下潜在模型表格：

subject $i$	$Y_i(0)$	$Y_i(1)$
	Test score if not tutored	tutored
1	3	4.5
2	5	5.5
3	5	4.5
4	4.5	5
5	4	5.5
6	6	6



# 计算影响

补课对成绩有什么影响？

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$
	Test score if not tutored	tutored	Treatment effect
1	3	4.5	1.5
2	5	5.5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0
Average	4.58	5.08	0.5





# 潜在结果和观察到的结果 Potential and Observed Outcomes

- 因果推断的根本问题 (The Fundamental Problem) 指出, 对于任意对象, 我们只能观察到  $Y_i(0)$  或者  $Y_i(1)$ , 而不能同时观察到两者。
- 因此, 对于单一对象的处理效应 ( $\tau_i$ ) 是**不可被观察的**(unobservable)。
- 被观察到的结果  $Y_i$  和两种潜在结果的关系可以通过**切换等式**(switching equation)表示:
  - $Y_i = d_i Y_i(1) + (1 - d_i) Y_i(0)$



# 回到计算补课的影响

现实中我们不能观察到所有的潜在结果，而只能观察到一种。

- 因此个体处理影响 ( $\tau$ ) 是不可计算的。

subject $i$	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ tutored	$\tau_i$ Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?
Average	4.5	5	0.5

我们只能**估计**平均影响。



# 被估量、估计量和估计值

## 被估量 (Estimand)

- 通常是不可观测的，是我们估计的目标
- 例：平均值、相关性、影响

## 估计量 (Estimator)

- 是我们用来估计的工具
- 例：样本均值、线性回归、DID
- 是本系列的重点

## 估计值 (estimate)

- 是我们使用估计量和样本数据之后，得到的数字结果
- 我们想办法让估计值接近被估量
- 通常在符号表达上加一个小帽子
  - E.g.  $\hat{\beta}$



# 平均处理效应 (ATE)

Average Treatment Effect (ATE) 的计算方法:

$$ATE = E(\tau_i) = E(Y_i(1) - Y_i(0)) = E(Y_i(1)) - E(Y_i(0))$$

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$ Treatment effect
	Test score if not tutored	tutored	
1	3	4.5	1.5
2	5	5.5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0
Average	4.58	5.08	0.5

这个例子中  $ATE = 5.08 - 4.58 = 0.5$ .



## 其他被估量 (estimands)

处理组平均处理效应 (ATT):

- Average treatment effect on the treated

$$\tau_{ATT} = E[Y_i(1) - Y_i(0) \mid D_i = 1]$$

- 通常比 ATE 更容易估计

控制组/对照组平均处理效应 (ATC):

- Average treatment effect on the controls

$$\tau_{ATC} = E[Y_i(1) - Y_i(0) \mid D_i = 0]$$

- 匹配法 (Matching) 可以估计 ATC

条件平均处理效应 (CATE):

- Conditional average treatment effect

$$\tau_{CATE}(X) = E[Y_i(1) - Y_i(0) \mid X_i = x]$$

- 用于衡量异质性影响 - 比如, 对于女性的平均处理效应



# 算一算 ATT/ATC

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$	$D_i$
	Test score if not tutored	Test score if tutored		
1	3	4.5	1.5	1
2	5	5.5	0	1
3	5	4.5	-0.5	1
4	4.5	5	0.5	0
5	4	5.5	1.5	0
6	6	6	0	0



# 算一算 ATT/ATC

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$	$D_i$
	Test score if not tutored	Test score if tutored		
1	3	4.5	1.5	1
2	5	5.5	0	1
3	5	4.5	-0.5	1
4	4.5	5	0.5	0
5	4	5.5	1.5	0
6	6	6	0	0

- $ATT = \frac{1}{3}$



# 算一算 ATT/ATC

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$	$D_i$
	Test score if not tutored	Test score if tutored		
1	3	4.5	1.5	1
2	5	5.5	0	1
3	5	4.5	-0.5	1
4	4.5	5	0.5	0
5	4	5.5	1.5	0
6	6	6	0	0

- $ATT = \frac{1}{3}$
- $ATC = \frac{2}{3}$





# 算一算 ATT/ATC

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$	$D_i$
	Test score if not tutored	Test score if tutored		
1	3	4.5	1.5	1
2	5	5.5	0	1
3	5	4.5	-0.5	1
4	4.5	5	0.5	0
5	4	5.5	1.5	0
6	6	6	0	0

- $ATT = \frac{1}{3}$
- $ATC = \frac{2}{3}$
- 接下来我们从被估量（看不到的）转换到估计量（看得到的）



# 平均值比较 (Diff-in-means/naive estimator)

我们可以通过计算我们观测到的两组平均值的差来估计ATE。

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$
	Test score if not tutored	tutored	Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?
Average	4.5	5	0.5

$\widehat{ATE} = 5 - 4.5 = 0.5$  —— 在这个例子中我们运气不错



# 运气不好

操作不对/运气不好的情况下，我们的结果就会出现偏差 (bias)/误差 (Error)

subject $i$	$Y_i(0)$	$Y_i(1)$	$\tau_i$
	Test score if not tutored	tutored	Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	?	5.5	?
6	6	?	?
Average	5.17	4.83	-0.34

在实际影响为正的情况下，我们甚至测出了负影响！

- 为什么会有这种偏差/误差呢



# 偏差 Bias vs 误差 Error

- 偏差 (Bias) 是系统性的 - 如果我们重复我们的研究, 或者增加样本数量, 理论上我们得到的结果的平均值不一定会接近真实的结果。
- 误差 (Error) 是随机的 - 通过增加样本数量和重复我们的研究可以减少误差。



# 选择偏差（系统性）

$$\underbrace{E[y_i|d_i = 1] - E[y_i|d_i = 0]}_{\text{观察到的平均之差 (naive comparison)}} = E[y_i(1)|d_i = 1] - E[y_i(0)|d_i = 0] =$$

观察到的平均之差 (naive comparison)

$$\underbrace{E[y_i(1)|d_i = 1] - E[y_i(0)|d_i = 1]}_{\text{Average Treatment Effect on the Treated (ATT)}} + \underbrace{E[y_i(0)|d_i = 1] - E[y_i(0)|d_i = 0]}_{\text{选择偏差 (Selection Bias)}}$$

Average Treatment Effect on the Treated (ATT)

选择偏差 (Selection Bias)

- 如果选择偏差足够大且和处理效应 (ATT) 是**同一方向**，那么处理效应的大小 (magnitude) 可能被**高估** (overestimated)。
- 如果选择偏差足够大且和 ATT 是**相反方向**，那么处理效应的大小可能被**低估** (attenuation bias)。



# 什么情况下没有系统性选择偏差呢？

处理分组 (Treatment Status) 应该**独立 (independent)**于潜在结果 (potential outcomes) 和其他背景特征 background attributes(X)

$$Y_i(0), Y_i(1), X \perp D_i$$

那么如何才能做到呢？



# 什么情况下没有系统性选择偏差呢？

处理分组 (Treatment Status) 应该**独立 (independent)**于潜在结果 (potential outcomes) 和其他背景特征 background attributes(X)

$$Y_i(0), Y_i(1), X \perp D_i$$

那么如何才能做到呢？

- **随机分配** (Random Assignment)



# 什么情况下没有系统性选择偏差呢？

处理分组 (Treatment Status) 应该**独立 (independent)**于潜在结果 (potential outcomes) 和其他背景特征 background attributes(X)

$$Y_i(0), Y_i(1), X \perp D_i$$

那么如何才能做到呢？

- **随机分配** (Random Assignment)
- 让样本中的每一个对象接受处理的机会相同





# 什么情况下没有系统性选择偏差呢？

处理分组 (Treatment Status) 应该**独立 (independent)**于潜在结果 (potential outcomes) 和其他背景特征 background attributes(X)

$$Y_i(0), Y_i(1), X \perp D_i$$

那么如何才能做到呢？

- **随机分配** (Random Assignment)
- 让样本中的每一个对象接受处理的机会相同
  - 我们需要设定这个几率和随机分配的方式



# 什么情况下没有系统性选择偏差呢？

处理分组 (Treatment Status) 应该**独立 (independent)**于潜在结果 (potential outcomes) 和其他背景特征 background attributes(X)

$$Y_i(0), Y_i(1), X \perp D_i$$

那么如何才能做到呢？

- **随机分配** (Random Assignment)
- 让样本中的每一个对象接受处理的机会相同
  - 我们需要设定这个几率和随机分配的方式
  - 例：扔硬币置随机数



# 本节课内容

1 回顾相关性 Correlation

2 回顾因果关系

3 一些别的概念和问题



# 不可观察的特征

- 我们已经讨论了潜在结果的不可观察性
- 但是对象的一些别的特征可能也是无法观察的
- 在实验中，我们经常需要保证实验组和对照组很“像”
  - 来减少选择偏差
  - 一些可以被观察的特征（年龄，性别）很好控制
  - 但是有一些特征无法控制
    - 生活习惯，身体健康（医保），对风险的态度
    - 即使可观察特征相同，不可观察特征可能会有区别
    - 比如说医保投保，两个相像的人中健康条件更差的可能更愿意投保



# 人口 (population) 与样本 (sample)

- 我们往往用样本 (Samples) 来推断人口 (Populations) 特征
  - 选举民调
  - 调查 (Survey)
- 估计量 (estimator): 通过样本和算法来推断人口的特征
  - 简单地估计量: 平均数 (mean), OLS
  - 较复杂的估计量: DiD, SC, RDD, IV
- 如何判断一个估计量 (estimator) 的好坏呢?



# 偏差 (bias)、一致性 (Consistency) 与效率 (Efficiency)

- 符号:  $\mu$  为人口参数 parameter (特征),  $\hat{\mu}$  为估计量 estimator
- 无偏性:  $E(\hat{\mu}) - \mu = 0$ 
  - 如果进行很多次测量, 测量的期望值 (平均值) 是  $\mu$
  - 不取决于样本大小
- 一致性:  $\hat{\mu} \xrightarrow{p} \mu$ 
  - 样本量 (sample size) 越大, 越接近真实值
  - 有些估计量有偏差 (biased), 但是一致 (e.g. IV)
- 效率: 对比两个估计量  $\hat{\mu}$  和  $\tilde{\mu}$ , 如果  $\text{var}(\hat{\mu}) < \text{var}(\tilde{\mu})$  则  $\hat{\mu}$  更有效率
  - 比如说 SDID 估计量比 DID 和 SC 的估计量更有效率
  - 特征: 相同情况下更小的置信区间 Confidence Interval



# References I

Bueno de Mesquita, E., & Fowler., A. (2021). *Thinking clearly with data: A guide to quantitative reasoning and analysis*. Princeton University Press.  
<https://books.google.com/books?id=aHYvEAAAQBAJ>

