Part IV　Chapter 16 — 20

本书有趣的部分

讨论观察性研究中的复杂性

$$\left\{\begin{array}{l} Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X \\ 0 < e(X) < 1 \end{array}\right\}$$

$\hat{\tau}^{reg}$, $\hat{\tau}^{ipw}$

$\hat{\tau}^{dr}$, $\hat{\tau}^{mbc}$

过强的假设

Part IV
- 有未观测的混杂
- 无完全重叠

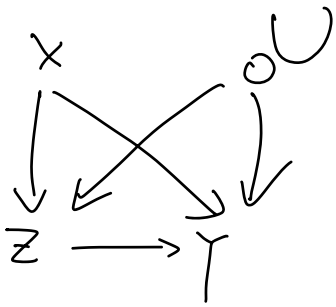
Chapter 16　有未观测的混杂

无混杂 $\Leftrightarrow$

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ Z \arrow[rr] & & Y \end{tikzcd}$$

Z和Y共同原因有被观察

common causes



unmeasured confounding

未观测的混杂

流行病学

epidemiology

$Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$ 不成立

$Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X, U$

如果你永远都担心 U 的存在

那么你可能成为 David Hume.

本质上 $Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$ 不可

验证或证伪

可观测

$$\Longleftrightarrow \begin{cases} \Pr(Y(1) \mid Z=1, X) = \Pr(Y(1) \mid Z=0, X) \\ \Pr(Y(0) \mid Z=1, X) = \Pr(Y(0) \mid Z=0, X) \end{cases}$$

哲学术语 反事实 counterfactual

可以做什么？

16.2 Assessing ignorability

评估可忽略性

~~testing~~ ~~verifying~~

negative control
阴性对照

negative outcome
阴性结果

negative exposure
阴性暴露

<u>negative outcome</u>

Y 关心结果

ACE $\tau_{Z \to Y}$

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ Z \arrow[rr] & & Y \end{tikzcd}$$

?

对可忽略性无信心. 怎么办?

目的：增强信心.

Y^n : 阴性结果

$\tau_{Z\to Y^n} = 0$

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] \arrow[rr] & & Y^n \\ Z \arrow[rr] & & Y & \end{tikzcd}$$

且 Y^n 与 Y 心理学因素相同

估计 $\hat{\tau}_{Z\to Y^n}$ $= 0$ 提供一定支持

$\neq 0$ 糟糕

难点：如何找 Y^n ?

<u>e.g.</u> Cornfield et al (1959)

$$\begin{tikzcd} & \triangle \arrow[dl] \arrow[dr] & \\ \text{smoking} \arrow[rr] & & \text{lung cancer} \end{tikzcd}$$

Y^n = 车祸

e.g. Imbens and Rubin (2015年)

job training program ⟶ wage

遗漏 Y^n = 上一期 wage

但是一般地，上一期 wage 是 X

的一个分量.

e.g. Jackson et al (2006)

流感疫苗 ⟶ 流感相关病

流感疫苗　　秋天 Y　　冬天 Y

time

假设 $\tau_{Z\to 积天Y} < \tau_{Z\to 考天Y}$

但是 $\hat{\tau}_{Z\to 积天Y} > \hat{\tau}_{Z\to 考天Y}$

（假设可忽略性）

统计学的尴尬：负面结论容易，

正面结论难.

<u>阳性最密</u>

$$\begin{tikzcd} Z^n & & X \arrow[ll] \arrow[d] \arrow[dr] & \\ & & Z \arrow[r] & Y \end{tikzcd}$$

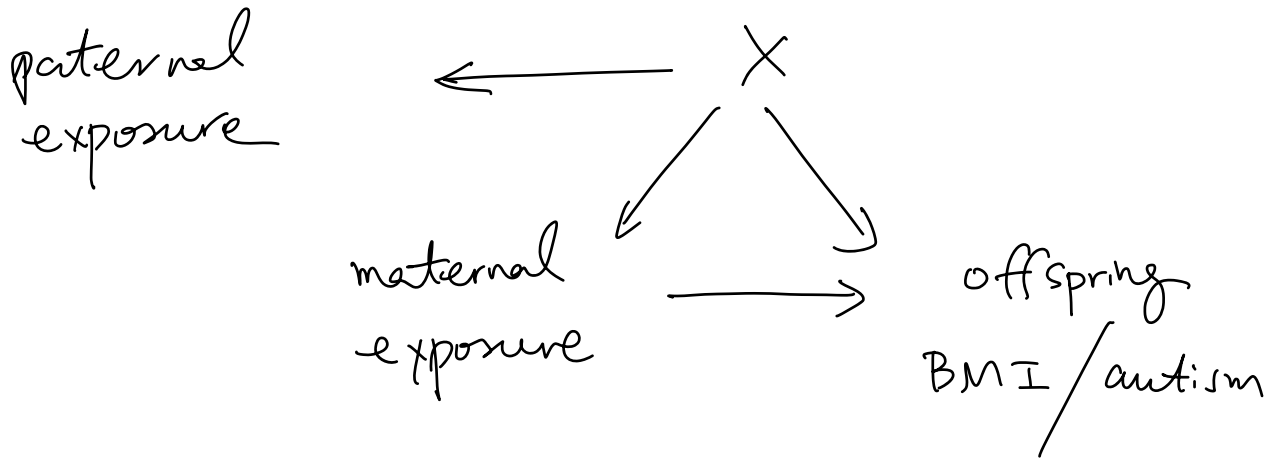
$\tau_{Z^n\to Y} = 0$

Z^n 与 Z 有相同混杂因素

估计 $\hat{\tau}_{Z^n\to Y}$ $=0$ OK

$\neq 0$ 精准的确定

e.g. Sanderson et al (2017)



假设 $\tau_{Z^n \to Y} < \tau_{Z \to Y}$

用概率论证伪.

falsify : Karl Popper

《猜想与反驳》

科学 versus 玄学/形而上

science versus metaphysics

猜测：$Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$

何时满足？

如何选取 X 使之满足？

Rosenbaum + Rubin：X 包含所有

会影响

$e(X) = \Pr(Z=1 \mid X)$

与 Z 相关之

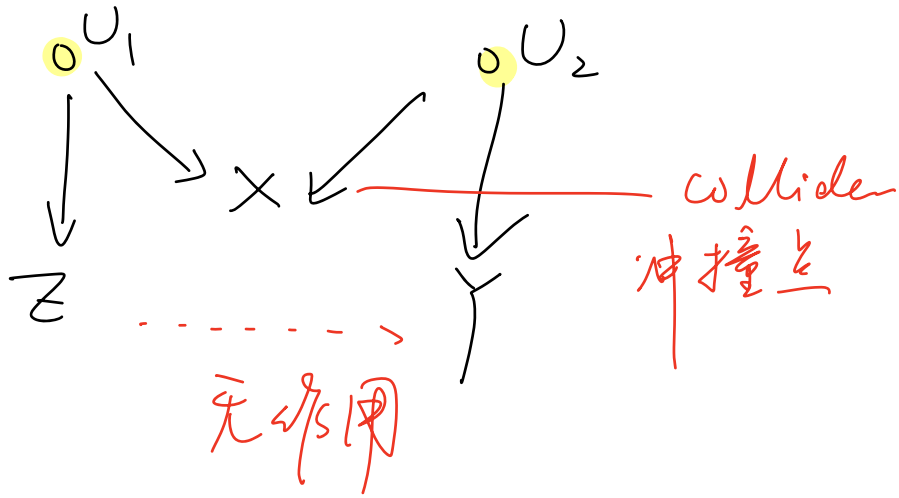
全部 X

Pearl 的反例

① M-bias

未观测

观测



不调整 X: 0 = 真正因果作用

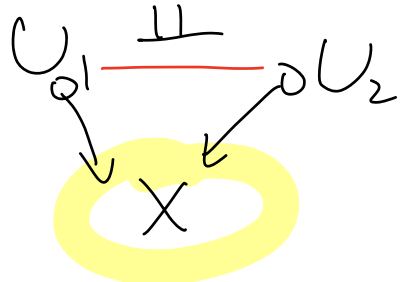
调整 X: $\mathrm{lm}(Y \sim Z + X)$

$\Rightarrow \mathrm{coef}(Z) \neq 0$

直觉: $\underbrace{Z \leftarrow U_1 \rightarrow X \leftarrow U_2 \rightarrow Y}_{\text{不活跃的路径}}$

$\Rightarrow Z, Y$ 不相关

给定 X:



$\Rightarrow$ Z与Y相关 | X

讲义：线性模型下的计算

② Z-bias

工具变量

part V

工具变量

$$\begin{tikzcd} & & U \arrow[dl, "b"'] \arrow[dr, "c"] & \\ X \arrow[rr, "a"'] \arrow[rrr, bend right, "\times"] & & Z \arrow[rr, "\tau"'] & & Y \end{tikzcd}$$

线性模型下计算

Ding, VanderWeele, Robins (2017 Biometrika)

$$\begin{cases} Z = aX + bU + \varepsilon_Z \\ Y_{(Z)} = \tau Z + cU + \varepsilon_Y \end{cases}$$

细节见讲义

$$\tau_{unadj} = lm(Y \sim Z) : coef(Z)$$

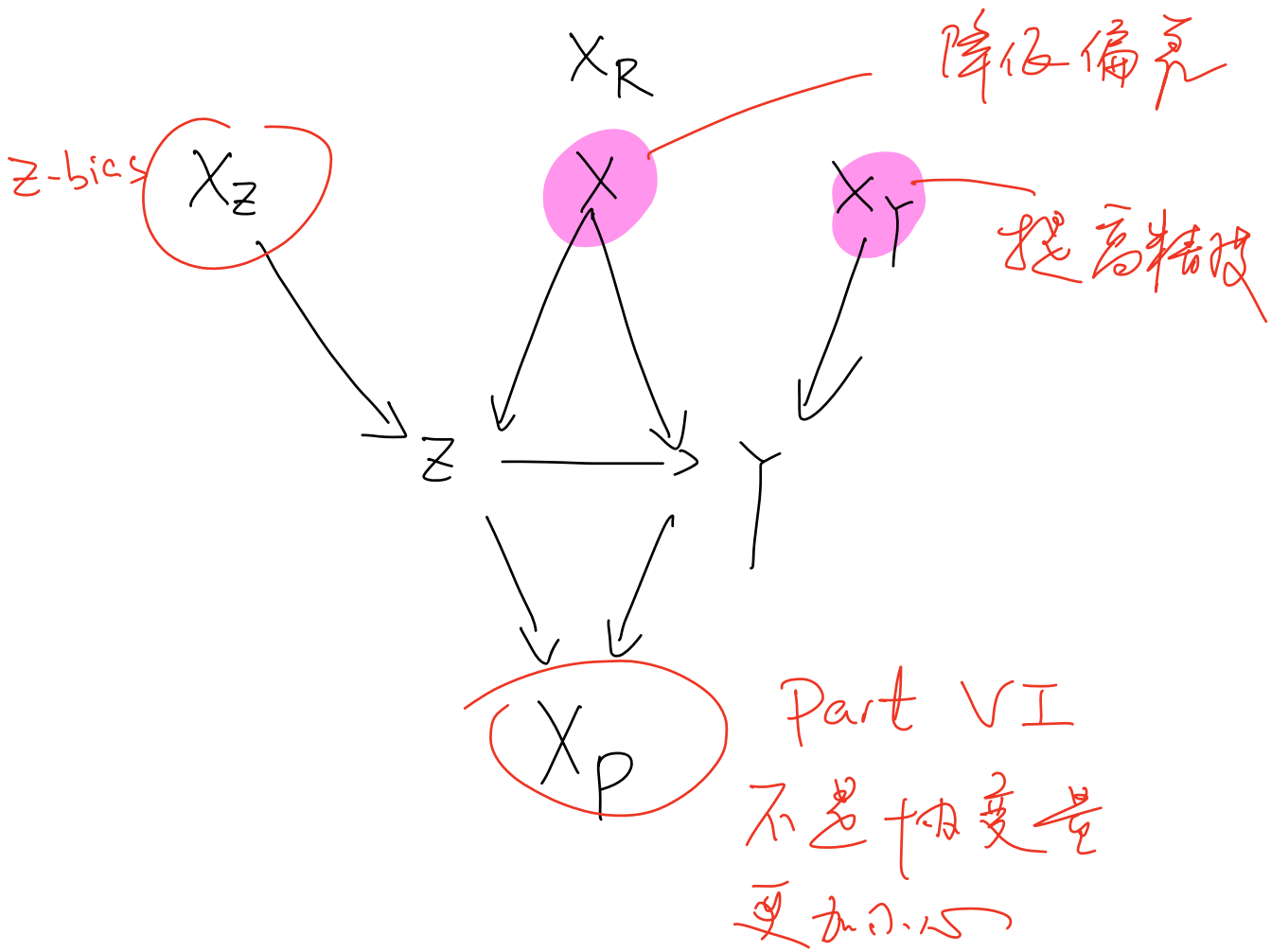
$$= \tau + \frac{cb}{a^2 + b^2 + 1}$$

$$\tau_{adj} = lm(Y \sim Z + X) : coef(Z)$$

$$= \tau + \frac{cb}{b^2 + 1}$$

结论：过度调整 X 可能有害

尤其是那些只影响处理的 X


X_R
降低偏差
Z-bias
X_Z
X
X_Y
提高精度
Z
Y
X_P
Part VI
不是协变量
更加小心

Chapter 17 E-value

p-value

Fisher FRT

随机实验：强零假设 多大 拒绝 被接受

观测性研究

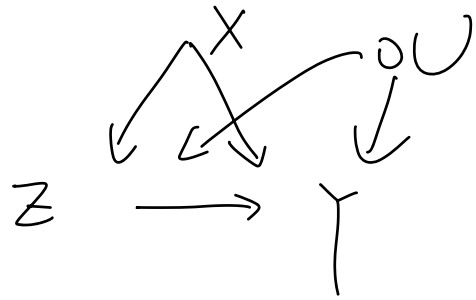
$$\hat{\theta} \pm \frac{\hat{\sigma}}{\sqrt{n}}$$

p-value 不够用

E : evidence for causation

Ding & Van der Weele : 2016 Epidemiology

VanderWeele & Ding : 2017 Annals of Internal Medicine



$Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X, U$

U 不可观测

Simpson's Paradox: U 会使可以完全改变你的结论

也是 Fisher 的论述:

Doll & Hill (1950): 吸烟 → 肺癌

RR ≅ 9

Fisher (1957): 未观测的基因

吸烟 肺癌

无因果

Fisher 并不完全错（被遗忘的发现了这个基因）

但是 Fisher 太极端：

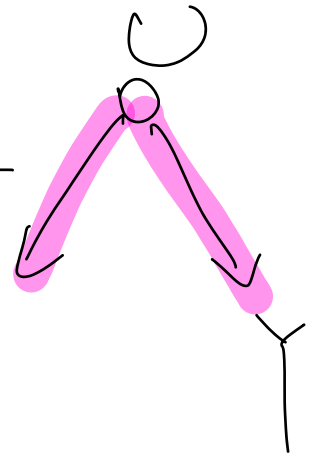
吸烟和肺癌之相关 $\boxed{\text{完全}}$ 由这个（些）基因导致

Cornfield et al (1959)：

类似反证法：反驳 Fisher

$\Rightarrow$ 有用果之成分

给定 X



$Z \perp\!\!\!\perp Y \mid (X, U)$

Cornfield et al: 如果Fisher好，

那么 Cornfield inequality

$$\underbrace{RR_{ZU}}_{\text{不可观测的}} \geq \underbrace{RR_{ZY}}_{\cong q}$$

如果我们没有因素U这样的

那么 $Z \longrightarrow Y$ 有因果作用.

<u>定理 17.1</u> （更广的 Cornfield 不等式）

如果 $Z \perp\!\!\!\perp Y \mid (X, U)$

那么 $$RR^{obs}_{ZY|X} \leq \frac{\alpha\beta}{\alpha+\beta-1}$$

其中 $\alpha = RR_{ZU|X}$

$\beta = RR_{UY|X}$

关于 α、β 对称

① $RR^{obs}_{ZY|X}$ 上界与 α、β 有关

② Simpson's Paradox 反面：混杂因素要足够强才能产生悖论

③
$$\begin{cases} RR_{ZU|X} \geq RR^{obs}_{ZY|X} \\ RR_{UY|X} \geq RR^{obs}_{ZY|X} \end{cases}$$

$$\Rightarrow \min\left(RR_{ZU|X},\ RR_{UY|X}\right) \geq RR^{obs}_{ZY|X}$$

④

$$\max\left(RR_{ZU|X},\ RR_{UY|X}\right)$$

$$\geq RR^{obs}_{ZY|X} + \sqrt{RR^{obs}_{ZY|X}\left(RR^{obs}_{ZY|X} - 1\right)}$$

定义为 E-value

⑤ Bradford Hill 的因果准则 (1965)

— strength (E-value 越大 一个支持)

— consistency 一致性

— specificity 特异性

有<因 negative outcome

来得即这多

— temporality

— biological gradient

— plausibility

— coherence

— experiment (Part II 实验流行)

— analogy