Part VI

因果机制：处理后变量

post-treatment variable

Chapter 26 主分层

Principal Stratification

Frangakis & Rubin (2002 Biometrics)

$$Z \longrightarrow M \longrightarrow Y$$

$$Z \longrightarrow Y \longrightarrow M$$

$$\begin{tikzcd} Z \arrow[r] & Y \arrow[r] & M^* \arrow[d] \\ & & M \end{tikzcd}$$

不依赖于因果图

RCT: $\Pr(Y \mid Z=1, M=m)$ v.s. $\Pr(Y \mid Z=0, M=m)$

$M(1) \neq M(0)$

$\Pr(Y(1) \mid Z=1, M(1)=m)$ $\Pr(Y(0) \mid Z=0, M(0)=m)$

Frangakis & Rubin (2002): 给定 M 的潜在结果

$\Pr(Y(1) \mid M(1), M(0))$ v.s. $\Pr(Y(0) \mid M(1), M(0))$

主分层

好处: 不依赖因果机制

e.g. 死亡删失

机制变量
$$\begin{cases} \underset{}{Z} \longrightarrow \underset{\text{死亡}}{M} \longrightarrow \underset{\text{生存质量}}{Y} \\ Z \longrightarrow Y \longrightarrow M \end{cases}$$

e.g. 替代指标

作用($Z \to M$) 预测 作用($Z \to Y$)

坏处: 无法解释因果机制

Chapter 27 中介分析

mediation analysis

$$
\begin{tikzcd}
Z \arrow[r] \arrow[rr, bend right] & M \arrow[r] & Y
\end{tikzcd}
$$

目标：分解 $Z \to Y$ 作用

$\parallel$

通过M间接作用 + 不通过M直接作用

e.g. 27.1

$$
\begin{tikzcd}
\text{15q25.1} \arrow[r] \arrow[rr, bend right] & \text{吸烟} \arrow[r] & \text{肺癌}
\end{tikzcd}
$$

以前：混杂因素

现在：处理变量

e.g. 27.2

$$
\begin{tikzcd}
\text{贫困社区} \arrow[r] \arrow[rr, bend right] & \text{学区、同学质量} \arrow[r] & \text{吸毒}
\end{tikzcd}
$$

问题非常普遍：如何回答？

$$
\begin{tikzcd}
& X \arrow[dl] \arrow[d, "\beta_2"] \arrow[dr, "\theta_4"] & \\
Z \arrow[r, "\beta_1"] \arrow[rr, bend right, "\theta_1"'] & M \arrow[r, "\theta_2"] & Y
\end{tikzcd}
$$

看出是 Baron-Kenny 方法

线性模型

$$
\begin{cases}
E(M \mid Z, X) = \beta_0 + \beta_1 Z + \beta_2' X \\
E(Y \mid Z, M, X) = \theta_0 + \theta_1 Z + \theta_2 M + \theta_4' X
\end{cases}
$$

直接作用： θ_1

间接作用： $\beta_1 \theta_2$

看山不是山

问：线性回归到底在估计什么？

如何定义直接和间接作用？

Robins & Greenland (1992) 引入了

嵌套之潜在结果

nested potential outcomes

① Z 可以被改变

$M(z),\ Y(z),\ z=0,1$

② Z 和 M 都可以被改变

因子实验 factorial experiment

$Y(z, m),\ z=0,1$

$m \in \mathcal{M}$

③ nested potential outcomes

$$Y(z,\ M(z'))$$

z, z' 可不同

也可相同

e.g. $Y(1,\ M(0))$

$Y(0,\ M(1))$

Assumption 27.1 (复合性)

$$Y(z,\ M(z)) = Y(z)$$

有时候，我们定义 $Y(z)$ 为 $Y(z, M(z))$

困难在于 $z \neq z'$: $Y(z, M(z'))$

如此：定义直接和间接作用

$$\tau = E\left(Y(1) - Y(0)\right) \quad \text{总作用}$$

natural/pure

$$NDE = E\left\{Y(1, M(0)) - \overbrace{Y(0, M(0))}^{=Y(0)}\right\} \quad \text{直接作用}$$

$$NIE = E\left\{\overbrace{Y(1, M(1))}^{=Y(1)} - Y(1, M(0))\right\} \quad \text{间接作用}$$

定义保证：$\tau = NDE + NIE$

对这套的批评：

— Rubin: $Y(z, M(z'))$

$z \neq z'$

这个量不存在于任何一个实验之中

只有数据本身无法证伪关于这个量的任何表述

Frangakis & Rubin (2002) 批评

$Y(1, M(0))$, $Y(0, M(1))$

为 a priori counterfactuals

先验反事实

— 其他人也批评

Popper 理论：科学 = 可证伪性

Science as Falsification

The following excerpt was originally published in Conjectures and Refutations *(1963).*

by Karl R. Popper

(1902 – 1994)

like simulation studies in statistics papers

1. It is easy to obtain confirmations, or verifications, for nearly every theory — if we look for confirmations.

2. Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory — an event which would have refuted the theory.

3. Every "good" scientific theory is a prohibition: it forbids certain things to

Sir Karl Popper "Science as Falsification," 1963 http://www.stephenjaygould.org/ctrl/popper_falsific

rules out some results

happen. The more a theory forbids, the better it is.

4. A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice.

5. Every genuine *test* of a theory is an attempt to falsify it, or to refute it. Testability is falsifiability; but there are degrees of testability: some theories are more testable, more exposed to refutation, than others; they take, as it were, greater risks.

6. Confirming evidence should not count *except when it is the result of a genuine test of the theory*; and this means that it can be presented as a serious but unsuccessful attempt to falsify the theory. (I now speak in such cases of "corroborating evidence.")

7. Some genuinely testable theories, when found to be false, are still upheld by their admirers — for example by introducing *ad hoc* some auxiliary assumption, or by reinterpreting the theory *ad hoc* in such a way that it escapes refutation. Such a procedure is always possible, but it rescues the theory from refutation only at the price of destroying, or at least lowering, its scientific status. (I later described such a rescuing operation as a "*conventionalist twist*" or a "*conventionalist stratagem.*")

Key

One can sum up all this by saying that the criterion of *the scientific status of a theory* is its *falsifiability, or refutability, or testability.*

中介分析公式 mediation formula

Pearl (2001)

假定

$Z \perp\!\!\!\perp Y(z,m) \mid X$

Z随机化

$(Z, M) \perp\!\!\!\perp Y(z,m) \mid X$

$M \perp\!\!\!\perp Y(z,m) \mid X, Z$

M随机化

$Z \perp\!\!\!\perp M(z) \mid X$

Z随机化

$Y(z,m) \perp\!\!\!\perp M(z') \mid X$

Cross-world independence

定理 27.1 在以上假定下

$E(Y(z, M(z'))$

$$= \iint E(Y \mid Z=z, M=m, X=x) \cdot \Pr(M=m \mid Z=z', X=x) \cdot f(x) \, dm \, dx$$

一个更简单的公式

$$E(Y) = \iiint E(Y \mid Z=z, M=m, X=x) \cdot \Pr(M=m \mid Z=z', X=x) \cdot \cancel{\Pr(Z=z \mid X=x)} \cdot f(x) \, \cancel{dz} \, dm \, dx$$

固定 $M(z')$

固定 z

$\Rightarrow$ 也可编到中介分析公式

证明　假设变量都是离散的

$$E\big(Y(z, M(z'))\big)$$

全概

$$= \sum_x \underbrace{E\big(Y(z, M(z')) \mid x\big)}_{\text{全概}} \Pr(x)$$

$$\sum_m E\big(Y(z, \cancel{M(z')}\ m) \mid \cancel{M(z') = m},\ x\big) \underbrace{\Pr(M(z') = m \mid x)}_{=\ \Pr(M=m \mid Z=z', x)}$$

Consistency

Cross-world independence

Z 随机化

$$= \sum_m \underbrace{E\big(Y(z,m) \mid x\big)}_{=\ E(Y \mid Z=z,\ M=m,\ x)} \Pr(M=m \mid Z=z', x)$$

$(Z, M) \perp\!\!\!\perp Y(z,m) \mid X$

$$= \sum_m E(Y \mid Z=z, M=m, x) \Pr(M=m \mid Z=z', x)$$

$$\Rightarrow NDE = \sum_x \sum_m \Big(E(Y \mid Z=1, M=m, X=x) - E(Y \mid Z=0, M=m, X=x) \Big) \cdot \Pr(M=m \mid Z=0, X=x) \cdot \Pr(X=x)$$

nonparametric identification formula

$$NIE = \sum_x \sum_m E(Y \mid Z=1, M=m, X=x) \cdot \Big[\Pr(M=m \mid Z=1, X=x) - \Pr(M=m \mid Z=0, X=x) \Big] \cdot \Pr(X=x)$$

看山依然是山

回到线性模型

$$\begin{cases} E(M \mid Z, X) = \beta_0 + \beta_1 Z + \beta_2' X \\ E(Y \mid Z, M, X) = \theta_0 + \theta_1 Z + \theta_2 M + \theta_4' X \end{cases}$$

$$\begin{tikzcd} & X \arrow[dl] \arrow[d] \arrow[dr] & \\ Z \arrow[r, "\beta_1"] \arrow[rr, bend right, "\theta_1"'] & M \arrow[r, "\theta_2"] & Y \end{tikzcd}$$

Tyler VanderWeele 论文

θ_3 为 $Z \cdot M$ 系数

HW 27.5

$$\Rightarrow NDE = \sum_x \sum_m \theta_1 \Pr(M=m \mid Z=0, X=x) \Pr(X=x)$$

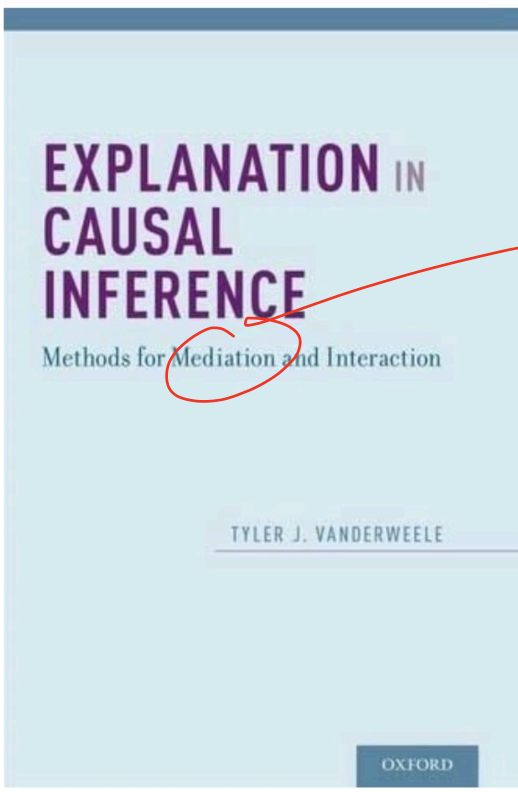
$$= \theta_1$$

$$NIE = \sum_x \sum_m \left(\cancel{\theta_0} + \cancel{\theta_1} + \theta_2 m + \cancel{\theta_4' x}\right) \cdot \left(\Pr(M=m \mid Z=1, X=x) - \Pr(M=m \mid Z=0, X=x)\right) \cdot \Pr(x)$$

$$= \theta_2 \sum_x \left(E(M \mid Z=1, X=x) - E(M \mid Z=0, X=x) \right) \Pr(x)$$

$$= \theta_2 \sum_x \beta_1 \Pr(x)$$

$$= \theta_2 \beta_1$$



中介分析

一切微妙结果

Cross-world independence:

$$Y(z,m) \perp\!\!\!\perp M(z') \mid X$$

对一切 z, z', m 都成立

无法被证实或证伪

即使有 Z 和 M 联合随机化，

也不能保证

e.g. 27.6

Cross-world independence $\Longleftarrow$ 成立

$$\begin{cases} Z = \mathbb{1}\left(f_Z(X, \varepsilon_Z) \geq 0\right) \\ M(z) = \mathbb{1}\left(f_M(X, z, \varepsilon_M^{\cancel{(z)}}) \geq 0\right) \\ Y(z,m) = f_Y\left(X, z, m, \varepsilon_Y^{\cancel{(z,m)}}\right) \\ \varepsilon_Z \perp\!\!\!\perp \varepsilon_M \perp\!\!\!\perp \varepsilon_Y \end{cases}$$

如何扔掉这个假定，还能做中介分析？

修正定义： $M(z')$

$$\Downarrow$$

$$Y\left(z,\ \Pr(M(z') \mid X)\right)$$

stochastic intervention

随机干预

$$\overset{\text{定义}}{=} \sum_m Y(z, m) \Pr\left(M(z') = m \mid X\right)$$

HW 27.8

chapter 28 CDE

Controlled direct effect

控制的直接作用

$$\begin{tikzcd} Z \arrow[r, "//"] \arrow[rr, bend right] & M = m \arrow[r] & Y \end{tikzcd}$$

无法定义
间接作用

$$CDE(m) = E\{Y(1,m) - Y(0,m)\}$$

$Y(z,m)$ 看成 (z,M) 联合干预下的潜在结果（因子实验）

为了简化，Z, m 二值

假定 2.8.1 $\quad \underbrace{(Z, M)}_{2\times 2\ \text{处理}} \perp\!\!\!\perp \underbrace{Y(Z, m)}_{\text{四个潜在结果}} \mid X$

二值之结果 $\Longrightarrow$ 拓展至四值 (Z, M)

自己阅读