

统计因果推断及应用—2023 年秋季期末作业

截止日期：2023 年 11 月 5 日晚 22: 00

1 Presentation (必选题)

1. 请用课堂所学的方法分析实际数据，提供数据来源、数据整理过程、分析过程、及最终结论。整理成小组报告形式提交，请组长提交，第九周课堂展示。
2. 因果推断方面的论文读书报告。整理成小组读书报告形式提交，此部分内容可以个人 or 小组，第九周课堂展示。
3. 科研交流展示。提交展示 ppt 即可，此部分内容可以个人 or 小组，第九周课堂展示。

注：每个小组控制在 15 分钟以内。

以下作业截止日期：2023 年 11 月 30 日晚 22: 00

2 个人作业 (必选 8 题，多选酌情加分)

1. 叙述因果推断的三个层级。列举你的生活和研究中分别对应三个层级的例子。
2. 叙述 Simpson's paradox 的内容。列举你的生活和研究中可能存在以上悖论的例子。
3. 叙述可识别性的定义。
4. 叙述潜在结果模型中潜在结果的定义，个体因果作用和平均因果作用的定义。
5. 叙述可忽略性的定义和含义。
6. 证明在可忽略性假设下，控制组因果作用可识别。
7. 叙述倾向得分的定义和性质。
8. 叙述估计量的双稳健性质的含义。
9. 验证双稳健估计量的有效性。
10. 叙述平行趋势假设。
11. 叙述不依从问题。
12. 叙述死亡删失问题。

13. 叙述二值处理下的主分层框架。
14. 叙述倾向得分匹配的原理和估计思想。
15. 叙述 $\text{pr}(Y_0 = 0 \mid Z = 1, Y = 1)$ 和 $\text{pr}(Y_1 = 1 \mid Z = 0, Y = 0)$ 的因果含义。
16. 利用单调性假设，即 $Y_1 \leq Y_0$ ，说明 $\text{pr}(Y_0 = 0, Y_1 = 1)$ 和 $\text{pr}(Y_1 = 1 \mid Z = 0, Y = 0)$ 的识别性。
17. 叙述工具变量在单调性假设下的非参识别性。
18. 叙述工具变量线性模型下两阶段最小二乘估计。
19. 叙述工具变量线性模型下基于 control function 的估计方法。
20. 叙述工具变量两阶段最小二乘的有效性。
21. 叙述工具变量两阶段基于 control function 的有效性。
22. 叙述无效工具变量下的一些识别方法。
23. 叙述双重差分和合成控制法的异同。
24. 给定 STUVA，一致性和可忽略性，利用倾向得分估计、回归估计、双稳健估计、和一些机器学习的方法完成平均因果作用的估计，包括数据生成和估计。
25. 我们关心某个二值处理变量 X 对结果变量 Y 的因果作用。给定 STUVA，一致性和可忽略性，及完全观察的协变量 V ，一个参数倾向得分

$$\pi(V; \alpha) = \text{pr}(X = 1 \mid V; \alpha)$$

和一个参数化结果回归

$$m(V; \beta) = \mathbb{E}(Y \mid X = 1, V; \beta)$$

这里 \hat{E} 表示经验均值算子。我们可以用以下三种方法估计潜在结果均值 $\mu = \mathbb{E}(Y_1)$:

• IPW:

$$\begin{aligned} \hat{E} \left[\left\{ \frac{X}{\pi(V; \hat{\alpha})} - 1 \right\} \cdot g(V) \right] &= 0 \\ \hat{E} \left\{ \frac{XY}{\pi(V; \hat{\alpha})} - \hat{\mu}_{\text{ipw}} \right\} &= 0, \end{aligned}$$

• REG:

$$\begin{aligned} \hat{E} \left[X \{Y - m(V; \hat{\beta})\} \cdot h(V) \right] &= 0 \\ \hat{E} \left\{ m(V; \hat{\beta}) - \hat{\mu}_{\text{reg}} \right\} &= 0, \end{aligned}$$

• DR:

$$\begin{aligned} \hat{E} \left[\left\{ \frac{X}{\pi(V; \hat{\alpha})} - 1 \right\} \cdot g(V) \right] &= 0 \\ \hat{E} [X \{Y - m(V; \hat{\beta})\} \cdot h(V)] &= 0 \\ \hat{E} \left[\frac{XY}{\pi(V; \hat{\alpha})} + \left\{ 1 - \frac{X}{\pi(V; \hat{\alpha})} \right\} m(V; \hat{\beta}) - \hat{\mu}_{\text{dr}} \right] &= 0. \end{aligned}$$

作业如下：

- (a) 设计一个模拟试验完成上面三种估计方法，包括数据生成和估计。（参考广义矩估计思想）
- (b) 给定工作模型正确，一些正则性条件，证明所得到的估计量具有相合性和渐进正态性。（参考广义矩估计思想）
- (c) 计算它们的渐进方差，并解释它们如何受到 g 和 h 的选择影响。如果一个或两个工作模型不正确会发生什么？