



- ### ③ 总结

- ### ③ 总结

[illegible]

- 2021 年 10 月 11 日，美国加州大学伯克利分校（UC Berkeley）经济系教授 David Card 因其对劳动经济学的突出贡献、美国麻省理工学院（MIT）经济系教授 Joshua Angrist 和美国斯坦福大学（Stanford）经济系教授 Guido Imbens 因其对因果关系分析的方法论贡献，共同分享了 2021 年诺贝尔经济学奖。
- 诺贝尔奖主页上对三位经济学家的评价是：“他们三人为我们提供了有关劳动力市场的新见解，并且展示了从自然实验可以得到什么样的因果关系。他们的方法已经扩展到其他领域并且彻底改变了实证研究。”

- 潜在结果 (potential outcomes): 每个结果都对应于一种处理或试验 (treatment or manipulation), 遵循“没有处理就没有因果关系 (no causation without manipulation)”的格言 (Rubin, 1975, p238)。这些潜在结果在先验上, 都是可观测的, 也就是说, 如果个体接受了相应的处理, 就可以观察到。但是, 从后验上看, 一旦进行了处理, 最多只能观察到一个潜在的结果。
- 观察多个个体 (multiple units) 的必要性, 以及相关稳定性假设的效用: 事实上, 因果推断的大多数情况下都使用稳定性假设这一概念来处理存在的多个个体。
- 分配机制 (assignment mechanism) 的核心作用: 这对于推断因果效应至关重要, 而且不同的分配机制构成了潜在结果框架的架构模块。

潜在结果 (potential outcomes)

- 例 1: “因为我服用了阿司匹林, 所以我的头痛减弱了。”
- 例 2: “因为她上了大学, 所以她去年找到了一份好工作。”
- 例 3: “她留长发是因为她是女孩。”
- 给定一个个体和一组行为, 我们将每个行为-个体对 (action-unit pair) 与一潜在结果关联起来。我们把这些结果称为潜在结果 (potential outcomes), 最终只有一个会实现, 继而可能被观察到。
- 事后无法观察到其他潜在结果, 因为没有采取相应的行动来实现这些结果。因果效应则涉及对潜在结果的比较, 其中一个已经实现, 而另一个尚未实现, 因此也无法观察到。
- 任何处理必须在观察到任何相关的潜在结果之前进行。

潜在结果 (potential outcomes)

- 一种解释这三种“因果”陈述之间的本质差异的方法是，在采用实际处理/行动之后，考虑与未采用处理相对应的潜在结果的反事实数值。
- 例 1: 所使用的处理方法是“服用阿司匹林”，而反事实的潜在结果是“未服用阿司匹林”下的头痛状态；在这里，考虑与事实相反的结果是明确无疑的。
- 例 2: 反事实：如果她决定不上大学，一年前的工作境况。
- 例 3: 反事实：如果她是男孩而不是女孩，那么她的头发长度。注意没有明确定义行为/处理，因此因果陈述的定义也不明确。
- 如果没有进一步解释干预措施，因果问题就没有明确定义。
- 通过更精确地阐明干预措施（干预措施本可以使另一种潜在结果变为已实现的结果），因果关系的陈述会变得更加明确。

因果效应 (causal effects)

- 例：在一个特定的时间点，我考虑是否服用阿司匹林治疗头痛。也就是说，有两种处理水平，服用阿司匹林和不服用阿司匹林。如果我服用阿司匹林，在一小时后，我的头痛可能会消失，也可能会持续。我们用 $Y(\text{Aspirin})$ 来表示这个结果，可以是“头痛”，也可以是“不头痛”。同样地，如果我不服用阿司匹林，在一小时后，我的头痛可能会持续，也可能不会；我们用 $Y(\text{No Aspirin})$ 表示这一潜在的结果，可以是“头痛”，或者“不头痛”。因此，存在两种潜在结果， $Y(\text{Aspirin})$ 和 $Y(\text{No Aspirin})$ 。
- $Y(\text{Aspirin}) = \text{不头痛}$ ， $Y(\text{No Aspirin}) = \text{头痛}$ ；
- $Y(\text{Aspirin}) = \text{头痛}$ ， $Y(\text{No Aspirin}) = \text{头痛}$ ；
- $Y(\text{Aspirin}) = \text{不头痛}$ ， $Y(\text{No Aspirin}) = \text{不头痛}$ ；
- $Y(\text{Aspirin}) = \text{头痛}$ ， $Y(\text{No Aspirin}) = \text{不头痛}$ ；

因果效应 (causal effects)

Table 1.1. *Example of Potential Outcomes and Causal Effect with One Unit*

Unit	Potential Outcomes		Causal Effect
	$Y(\text{Aspirin})$	$Y(\text{No Aspirin})$	
You	No Headache	Headache	Improvement due to Aspirin

图 1: 示例: 单个个体的潜在结果和因果效应

- 因果效应取决于潜在结果，但不取决于实际观察到的结果。
- 因果效应是比较同一个体，在试验后同一时间的潜在结果。
- 特别地，因果效应并不是比较不同时间的结果，就像在我决定服用或不服用阿司匹林之前和之后比较头痛的状况。
- 因此，“因果推断的基本问题”（Holland,1986, 第 947 页）是最多只可以实现并观测到一种潜在结果的问题。

因果效应 (causal effects)

Table 1.2. Example of Potential Outcomes, Causal Effect, Actual Treatment, and Observed Outcome with One Unit

Unit	Not Observable			Known	
	Potential Outcomes		Causal Effect	Actual Treatment	Observed Outcome
	$Y(\text{Aspirin})$	$Y(\text{No Aspirin})$			
You	No Headache	Headache	Improvement due to Aspirin	Aspirin	No Headache

图 2: 示例: 潜在结果, 因果效应, 实际治疗和可观测的潜在结果

- 因此，即使个体水平的因果效应（比较两种潜在结果）可能会很好地定义，但根据定义我们不能仅从单个已实现的潜在结果中了解其价值。
- 对于因果效应的估计，考虑为什么某些潜在结果实现了，而其它潜在结果却没有实现的原因也很重要，即分配机制。

稳定个体干预值假设 (SUTVA)

- 稳定个体干预值假设 (SUTVA): 任何个体的潜在结果都不随分配给其他个体的处理水平而变化, 而且处理水平对于所有个体都是相同的, 这就导致了不同的潜在结果。
- 挑战 1: 存在着不同个体间相互干预的可能性, 使得接受特定处理的个体潜在因果可能依赖于另一个体接受的处理水平。
- 挑战 2: 因为是多个个体的情境, 在每种处理水平下, 我们必须有多个观测。在接受名义上相同的处理水平时, 不同个体的潜在因果可能因接受处理的不同程度而有所差异。
- 解决方法: 将子样本 (subpopulation) 视为一个个体。

1. *Journal of Management Studies*, 1990, 27, 1, 1-14.

Unit	Unknown			Known	
	Potential Outcomes		Causal Effect	Actual	Observed
	$Y(\text{Aspirin})$	$Y(\text{No Aspirin})$		Treatment W_i	Outcome Y_i^{obs}
You	No Headache	Headache	Improvement due to Aspirin	Aspirin	No Headache
I	No Headache	No Headache	None	No Aspirin	No Headache

1. **Introduction**

- 缺乏的关键信息：每个人如何接受处理/治疗水平的，即分配机制。
- 为了得出有效的因果推断，我们必须考虑为什么有些个体接受一种处理方法而不是另一种。

分配机制 (assignment mechanism)

Table 1.4. Medical Example with Two Treatments, Four Units, and SUTVA: Surgery (S) and Drug Treatment (D)

Unit	Potential Outcomes		Causal Effect
	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
Patient #1	1	7	6
Patient #2	6	5	-1
Patient #3	1	5	4
Patient #4	8	7	-1
Average	4	6	2

图 4: 示例: SUTVA 假设下的四个个体医学案例

- 考虑平均因果效应：手术治疗比药物治疗会延长 2 年寿命。
- 现在假设医生通过专业知识，对这些潜在的结果有足够的了解，因此为每位患者分配对该患者而言更有益的治疗。

分配机制 (assignment mechanism)

Table 1.5. Ideal Medical Practice: Patients Assigned to the Individually Optimal Treatment; Example from Table 1.4

Unit	Treatment	Observed Outcome
i	W_i	y_i^{obs}
Patient #1	1	7
Patient #2	0	6
Patient #3	1	5
Patient #4	0	8

图 5: 示例：四个个体的最优治疗情形

- 观察到的治疗/处理方法和结果：手术治疗的平均观测结果比药物治疗的平均观测结果少 2 年。
- 基于此例，我们不能简单地观察不同治疗方法下潜在结果的实际观测值，而不考虑分配机制就得出结论。

协变量 (covariates)

- 个体特定性质（也称为预处理变量或协变量）：用个体 i 的含有 K 个分量的行向量 X 表示。
- 协变量的关键特征：先验性，不受处理分配机制的影响。
- 协变量通常通过解释因变量/结果中的一些变化使得估计更精确。
- 由于实际原因，研究人员可能更有兴趣探讨处理/治疗对某一子样本（由协变量定义）的平均因果效应。
- 协变量最重要的作用在于它们对分配机制的影响。
- 关于分配机制及其可能不依赖于潜在结果的假设，在某些关于协变量同质的子样本中通常更为合理。

因果估计量 (causal estimands)

- 令 $\mathbb{T}_i =$ 表示个体 i 接受的处理水平的集合。在多数情况下, 这一集合对于所有个体而言都是相同的: $\mathbb{T}_i = \mathbb{T} = \{0, 1\}$.
- 个体因果效应 (unit-level causal effects): 差值 $Y_i(1) - Y_i(0)$ 或者比值 $Y_i(1)/Y_i(0)$.
- 平均因果效应 (average treatment effects, ATE):

$$\tau_{\text{fs}} = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)).$$

因果估计量 (causal estimands)

- 条件平均因果效应 (conditional average treatment effects, CATE):

$$\tau_{\text{fs}}(f) = \frac{1}{N(f)} \sum_{i: X_i=f} (Y_i(1) - Y_i(0))$$

- 其它形式的平均因果效应:

$$\tau_{\text{fs},t} = \frac{1}{N_t} \sum_{i:W_i=1} (Y_i(1) - Y_i(0))$$

$$\tau_{\text{fs, pos}} = \frac{1}{N_{\text{pos}}} \sum_{i: Y_i(0) > 0, Y_i(1) > 0} (Y_i(1) - Y_i(0))$$

样本、总体和大样本

- 在本书的大部分讨论中，我们观测到的协变量、处理水平和实验结果的有限单位集都是我们所感兴趣的集合，我们称之为总体。
- 这一群体是如何选择的，来自何处并不重要。所有结论都取决于这一总体，并且我们并不试图对其他总体进行推论。
- 然而，对于讨论的这一部分来说，仍有必要把我们观测到的集合视为是从一个更大的总体中随机抽取出来的。在这种情况下，我们通常把这一更大的总体视为无穷大。
- 做出这种区分至关重要，我们把观测到的集合视为有限样本 (finite sample, 使用下标 “fs” 表示)，把无限总体视为超总体 (super-population, 使用下标 “sp” 表示)

Lord 悖论

- Lord(1967) 构造了一个假想的案例：一所大学想考察其食堂膳食对学生体重的影响以及影响的性别差异，为此，收集了学生 9 月份入学时的体重，然后次年 6 月份又获得了学生在校一学年后的体重。
- 两个统计学家分别利用这个数据考察了学校食堂对学生体重的影响，但得到了完全不同的结论。
- 第一个统计学家用了初等的方法，计算了男生和女生入学时的平均体重，分别是 150 磅和 130 磅。然后又计算了入学 1 学年后男、女生的平均体重，发现仍然是 150 磅和 130 磅。
- 因而，第一位统计学家认为学生食堂膳食对学生体重没有影响。

1972

总体	某大学 1963-1964 学年全部学生
积极干预	学校食堂膳食
控制干预	?
干预变量	对所有的学生, 均有 $D_i = 1$
G	性别 (1= 男生, 2= 女生)
X	1963 年 9 月入学时的体重
Y	1964 年 6 月放假时的体重

- 食堂膳食对学生体重的影响可以写成 $Y_1 - Y_0$ ，对所有学生的平均影响可以写成：

$$\Delta_i = E[Y_1 - Y_0 \mid G = i], \quad i = 1, 2$$

平均因果影响的性别差异为:

$$\Delta = \Delta_1 - \Delta_2$$

Lord 悖论

- Δ 也可以写成:

$$\Delta = \{E[Y_1 | G = 1] - E[Y_1 | G = 2]\} \\ - \{E[Y_0 | G = 1] - E[Y_0 | G = 2]\}$$

- 第 1 个统计学家根据男女生入学前和放假后平均体重的对比得到没有影响的结论, 他估计的参数为:

$$\delta_i = E[Y_1 - X \mid G = i], \quad i = 1, 2$$

影响的性别差异为:

$$\delta = \delta_1 - \delta_2$$

- 在假设 $Y_0 = X$ 下, δ_i 才会等于平均因果效应参数 Δ_i .

Lord 悖论

- 第 2 个统计学家认为应该控制开学时的体重, 比较相同体重的人放假时体重的变化。因而, 同样初始体重为 X 的个体, 体重的增加为:

$$\delta_i(X) = E[Y_1 - X \mid X, G = i], \quad i = 1, 2$$

增量的性别差异为:

$$\delta(X) = \delta_1(X) - \delta_2(X)$$

- 为了简单起见, Lord 假设条件期望函数均为线性且男女生斜率相同, 即

$$E[Y_1 \mid X, G = i] = a_i + bX, \quad i = 1, 2$$

则 $\delta(X)$ 可以简化为:

$$\delta(X) = a_1 - a_2$$

Lord 悖论

- 第 2 个统计学家将 $\delta(X)$ 解释为 6 月份男生平均体重超过女生平均体重的部分, 这一解释是正确的, 但并不是因果效应的差额。
- 但是, 在 $Y_0 = a + bX$ 假设下, 第 2 个统计学家的估计结果等于因果效应参数:

$$\begin{aligned}\Delta &= E[Y_1 - Y_0 \mid G = 1] - E[Y_1 - Y_0 \mid G = 2] \\ &= E[Y_1 - a - bX \mid G = 1] - E[Y_1 - a - bX \mid G = 2] \\ &= E[Y_1 - bX \mid G = 1] - E[Y_1 - bX \mid G = 2] \\ &= E[E[X_1 \mid X, G = 1] - bX \mid G = 1] \\ &\quad - E[E[Y_1 \mid X, G = 2] - bX \mid G = 2] \\ &= E[a_1 + bX - bX \mid G = 1] - E[a_2 + bX - bX \mid G = 2] \\ &= a_1 - a_2 = \delta(X)\end{aligned}$$

1. *Journal of Management Studies*, 1997, 34, 1, 1-14.

1. *Journal of Management Studies*, 1996, 33, 1, 1-14.

- 对于每个个体，还有一对潜在结果， $Y_i(0)$ 和 $Y_i(1)$ ，表示其在两个治疗状态下的结果值： $Y_i(0)$ 表示在对照治疗下的结果，而 $Y_i(1)$ 表示在主动治疗下的结果。
- 假设潜在结果是一个标量， $\mathbf{Y}(0)$ 和 $\mathbf{Y}(1)$ 表示潜在结果的 N 维列向量，治疗分配的 N 维列向量由 \mathbf{W} 表示，其中元素 $W_i \in \{0, 1\}$ ，如果单元 i 接受了对照治疗，则 $W_i = 0$ ，如果该单元接受了主动治疗，则 $W_i = 1$ 。
- 此外，记 $N_t = \sum_{i=1}^N W_i$ 和 $N_c = \sum_{i=1}^N (1 - W_i)$ 分别为分配给主动治疗和对照治疗的个体数，则有 $N_c + N_t = N$ 。

© 2004 Blackwell Publishing Ltd, *Journal of Internal Medicine* 255: 109–116

1. *Journal of Management Studies*, 1996, 33, 1, 1-14.

倾向性得分 (propensity score)

- 此外, 对具有共同协变量值的子样本集的个体水平分配概率的平均值, 称为“倾向性得分 (propensity score)", x 处的倾向得分是所有满足 $X_i = x$ 的单位的平均治疗分配概率:

$$e(x) = \frac{1}{N(x)} \sum_{i: X_i = x} p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$$

其中 $N(x) = \#\{i = 1, \dots, N \mid X_i = x\}$ 是 $X_i = x$ 的个体数量。

- 约定若 $N(x) = 0$, 则倾向性得分为 0。

分配机制的个体化分配假设 (Individualistic Assignment)

- 个体化分配 (Individualistic Assignment): 限制了个体 i 的治疗分配对其他个体的结果和分配的依赖性。
- 称一个分配机制 $\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ 是个体化的, 若存在函数 $q(\cdot) \in [0, 1]$ 和集合 \mathbb{A} 使得:

$$p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q(X_i, Y_i(0), Y_i(1)), i = 1, \dots, N$$

以及

$$\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = c \cdot \prod_{i=1}^N q(X_i, Y_i(0), Y_i(1))^{W_i} (1 - q(X_i, Y_i(0), Y_i(1)))^{1-W_i}$$

在 $(\mathbf{W}, \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \in \mathbb{A}$ 成立, 且在补集上取值为 0 (其中 c 是使得概率和为 1 的标准化系数)。

分配机制的概率分配假设 (Probabilistic Assignment)

- 概率分配 (Probabilistic Assignment): 要求每个个体被分配到积极治疗 $W_i = 1$ 和对照治疗 $W_i = 0$ 的情形均具有正概率。
- 称一个分配机制 $\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ 是概率分配的, 若对于所有个体 i 被分配到积极治疗的概率都严格地在 0 到 1 之间:

$$0 < p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) < 1, \quad \text{对于所有可能的 } \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)$$

对所有个体 $i = 1, 2, \dots, N$ 均成立。

分配机制的非混淆性假设 (Unconfounded Assignment)

- 非混淆性 (Unconfounded Assignment): 限制了分配机制对潜在结果的依赖性。
- 称一个分配机制 $\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ 是非混淆的, 若该分配机制不依赖于潜在结果:

$$\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}'(0), \mathbf{Y}'(1))$$

对于所有 $\mathbf{W}, \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Y}'(0)$, 和 $\mathbf{Y}'(1)$ 均成立。

实验性研究和观察性研究

- 总体来讲, 可以根据分配机制是否被实验者已知, 将研究分为实验性研究 (Experimental Study) 和观察性研究 (Observational Study)。若分配机制已知, 则称为是实验性研究, 否则称为是观察性研究。
- 对于实验性研究, 随机试验设计被认为是因果推断最可信的基础, 称分配机制是随机试验 (Randomized Experiment), 若该分配机制已知, 且是概率分配的。
- 在此基础上, 称分配机制是经典随机试验 (Classical Randomized Experiment), 若该分配机制是随机试验, 且是个体化分配的, 同时具有非混淆性。

观察性研究

- 对于观察性研究, 研究者可能不知道确切的分配概率, 但是研究者仍然有关于分配机制的大量信息。例如, 在某些情况下, 医疗决策是仅使用患者的病历做出的, 这需要医生的先验知识, 但具体如何可能是未知的。
- 一般来说, 我们将具有未知分配机制的设计称为观察性研究。称观察性研究的分配机制是正规的 (Regular), 若该分配机制同时是个体化的, 概率的, 以及具有不可混淆性。
- 此外, 如果已知正规分配机制的函数形式, 则该分配机制对应于经典随机实验。如果函数形式未知, 则该分配机制对应于具有正规分配机制的观察性研究。

随机化实验的分类

- 伯努利实验 (Bernoulli Trials)
- 完全随机实验 (Completely Randomized Experiments)
- 分层随机实验 (Stratified Randomized Experiments)
- 配对随机实验 (Paired Randomized Experiments)

伯努利实验 (Bernoulli Trials)

- 在伯努利实验中, 可以利用投硬币方式决定每个个体的干预状态。
- 如果实验总体共有 N 个, 则分配向量共有 2^N 种。
- 假设我们的实验总体为 2, 那么, 在伯努利实验中, 共有 4 种可能的分配向量, 分别是:

$$W \in \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

完全随机试验 (Completely Randomized Experiments)

- 考虑实验性研究中一类特殊的分配机制, 若固定数量的受试者被分配接受积极治疗, 则称该分配机制是完全随机试验 (Completely Randomized Experiments)。在完全随机试验中, 分配机制是:

$$\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} \left(\binom{N}{N_t}\right)^{-1} & \text{if } \sum_{i=1}^N W_i = N_t \\ 0 & \text{otherwise} \end{cases}$$

其中

$$\binom{N}{N_t} = \frac{N!}{N_t! (N - N_t)!}$$

- 给定个体数为 N 的群体, 完全随机试验通过固定分配给治疗的单位数 N_t , 从群体中随机抽取 N_t 个单位接受治疗。因此, 每个个体接受治疗的概率 $q = N_t/N$.

[illegible]

-

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \mid \sum_{i: B_i=j}^N W_i = N_t(j), \text{ for } j = 1, 2, \dots, J \right\}$$

且

$$\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} \prod_{j=1}^J \binom{N(j)}{N_t(j)}^{-1} & \text{if } \mathbf{W} \in \mathbb{W}^+ \\ 0 & \text{otherwise} \end{cases}$$

配对随机实验 (Paired Randomized Experiments)

- 有一种特殊的分层随机化实验, 称为配对随机化实验。
- 在这种随机化实验中, 根据某个特征, 将所有个体分成 $N/2$ 组 (N 为偶数), 每组只有 2 个个体, 并且两个个体随机化分配, 1 个在干预组, 1 个在控制组。
-

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \mid \sum_{i: B_i=j}^N W_i = 1, \text{ for } j = 1, 2, \dots, N/2 \right\}$$

且

$$\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 2^{-N/2} & \text{if } \mathbf{W} \in \mathbb{W}^+ \\ 0 & \text{otherwise} \end{cases}$$

四种经典随机实验的分配机制数量

Table 4.1. Number of Possible Values for the Assignment Vector by Design and Sample Size

Type of Experiment and Design	Number of Possible Assignments Cardinality of \mathbb{W}^+	Number of Units (N) in Sample			
		4	8	16	32
Bernoulli trial	2^N	16	256	65,536	4.2×10^9
Completely randomized experiment	$\binom{N}{N/2}$	6	70	12,870	0.6×10^9
Stratified randomized experiment	$\left(\binom{N/2}{N/4}\right)^2$	4	36	4,900	0.2×10^9
Paired randomized experiment	$2^{N/2}$	4	16	256	65,536

图 7: 四种经典随机实验的分配机制数量

举例：序列随机实验 (Sequential Randomized Experiment)

$$\Pr(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}) = \left\{ \begin{array}{ll} 1/2 & \text{if } Y_1(0) > Y_2(1), \text{ and } \mathbf{W} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \\ 1/2 & \text{if } Y_1(1) \geq Y_2(0), \text{ and } \mathbf{W} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \\ 1/2 & \text{if } Y_1(0) \leq Y_2(1), \text{ and } \mathbf{W} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \\ 1/2 & \text{if } Y_1(1) < Y_2(0), \text{ and } \mathbf{W} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \end{array} \right\}$$

举例：序列随机实验 (Sequential Randomized Experiment)

$$p_1(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = p_2(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1/2$$

$$p_3(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 0 & \text{if } Y_1(0) > Y_2(1) \text{ and } Y_1(1) < Y_2(0) \\ 1 & \text{if } Y_1(1) \geq Y_2(0) \text{ and } Y_1(0) \leq Y_2(1) \\ 1/2 & \text{otherwise.} \end{cases}$$

- 1 因果推断的潜在结果框架简介
- 2 完全随机实验 (Completely Randomized Experiments):
R.A.Fisher 与 J. Neyman 之间的分歧
费希尔精确 P 值方法 (R.A.Fisher's Exact P-Values)
奈曼重复抽样方法 (J.Neyman's Repeated Sampling Approach)
- 3 总结

- ① 因果推断的潜在结果框架简介
- ② 完全随机实验 (Completely Randomized Experiments):
R.A.Fisher 与 J. Neyman 之间的分歧
 费希尔精确 P 值方法 (R.A.Fisher's Exact P-Values)
 奈曼重复抽样方法 (J.Neyman's Repeated Sampling Approach)
- ③ 总结

费希尔精确 P 值方法 (R.A.Fisher's Exact P-Values) : 简介

- 给定一个完全随机实验的数据, 费希尔采用了一种“尖锐零假设”来评估积极治疗与对照治疗是否有显著差异。
- 在该零假设下, 对于实验中的每个个体, 两个潜在结果的值是相同的。费希尔开发了计算“P 值”的方法, 称为费希尔精确 P 值 (*FEPs*)。
- 费希尔的方法可以适应除了个体治疗效果均为零的尖锐零假设之外的零假设, 例如假设有一个恒定的可加治疗效果:

$$Y_i(1) = Y_i(0) + C,$$
 或通过取对数, 零假设变为: $Y_i(1)/Y_i(0) = C$ 。

Paul et al. 的蜂蜜实验数据

- 实验介绍：随机实验，评估两种治疗方法对儿童上呼吸道感染相关的夜间咳嗽和睡眠困难的效果。
- 实验对象：接受荞麦蜂蜜 ($N_t=35$) 或没有积极治疗 ($N_c=37$) 的儿童。
- 治疗组：单一剂量的荞麦蜂蜜；对照组：无积极治疗。
- 潜在结果：术后咳嗽频率 (cfa) 和术后咳嗽严重程度 (csa)，均介于 0 (“完全不频繁/严重”) 到 6 (“非常频繁/严重”) 之间。
- 协变量：咳嗽频率前 (cfp) 和咳嗽严重程度前 (cfp)。

Paul et al. 的蜂蜜实验数据

Table 5.1. *Summary Statistics for Observed Honey Data*

Variable	Mean	(S.D.)	Mean Controls	Mean Treated
Cough frequency prior to treatment (cfp)	3.86	(0.92)	3.73	4.00
Cough frequency after treatment (cfa)	2.47	(1.61)	2.81	2.11
Cough severity prior to treatment (csp)	3.99	(1.03)	3.97	4.00
Cough severity after treatment (csa)	2.54	(1.74)	2.86	2.20

图 8: 观测蜂蜜数据的描述性统计

Paul et al. 的蜂蜜实验数据

Table 5.2. Cumulative Distribution Functions for Cough Frequency and Severity after Treatment Assignment for the Honey Study

Value	cfa		csa	
	Controls	Treated	Controls	Treated
0	0.14	0.14	0.16	0.17
1	0.19	0.40	0.22	0.46
2	0.32	0.63	0.35	0.54
3	0.73	0.83	0.59	0.77
4	0.89	0.91	0.86	0.91
5	0.92	0.97	0.95	0.94
6	1.00	1.00	1.00	1.00

图 9: 蜂蜜研究中治疗后咳嗽频率和严重程度的累积分布函数

Paul et al. 的蜂蜜实验数据

Table 5.3. Cough Frequency for the First Six Units from the Honey Study

Unit	Potential Outcomes				
	Cough Frequency (cfa)		Observed Variables		
	$Y_i(0)$	$Y_i(1)$	W_i	X_i (cfp)	Y_i^{obs} (cfa)
1	?	3	1	4	3
2	?	5	1	6	5
3	?	0	1	4	0
4	4	?	0	4	4
5	0	?	0	1	0
6	1	?	0	5	1

图 10: 蜂蜜实验前六个个体的咳嗽频率

Paul et al. 的蜂蜜实验数据

Table 5.4. Cough Frequency for the First Six Units from Honey Study with Missing Potential Outcomes in Parentheses Filled in under the Null Hypothesis of No Effect of the Treatment

Unit	Potential Outcomes					
	Cough Frequency (cfa)		Observed Variables			
	$Y_i(0)$	$Y_i(1)$	Treatment	X_i	Y_i^{obs}	$\text{rank}(Y_i^{\text{obs}})$
1	(3)	3	1	4	3	4
2	(5)	5	1	6	5	6
3	(0)	0	1	4	0	1.5
4	4	(4)	0	4	4	5
5	0	(0)	0	1	0	1.5
6	1	(1)	0	5	1	3

图 11: 在无治疗效果的零假设下 ($H_0: Y_i(0) = Y_i(1)$ 对于 $i = 1, \dots, 6$), 蜂蜜实验中缺失潜在结果的咳嗽频率

1. *Journal of the American Medical Association*, 1997; 278: 1039-1044.

- 费希尔的方法同样需要指定统计量, 用于检验的统计量 T 是 C 个已知的实值函数 $T(\mathbf{W}, \mathbf{Y}^{\text{obs}}, \mathbf{x})$, 其中 \mathbf{W} 是完全随机实验的分配机制, \mathbf{Y}^{obs} 是观测到的潜在结果, 也可被视为是 \mathbf{W} 和潜在结果 $(\mathbf{Y}(0), \mathbf{Y}(1))$ 的函数, \mathbf{X} 是协变量矩阵。
- 统计量 T^{dif} 可以用于评估因果效应的显著性:

$$T^{\text{dif}} = \left| \bar{Y}_{\text{t}}^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}} \right| = \left| \frac{\sum_{i:W_i=1} Y_i^{\text{obs}}}{N_{\text{t}}} - \frac{\sum_{i:W_i=0} Y_i^{\text{obs}}}{N_{\text{c}}} \right|$$

当个体治疗效果被认为是可加的,且数据中异常值较少时,该统计量会发挥更好的效果。

- 当数据中的异常值较多时, 一个具有吸引力的选择是秩统计量。对于所有个体的观测潜在结果 Y_i^{obs} , 可以对每个个体, 计算得到中心化的秩排序:

$$R_i = R_i(Y_1^{\text{obs}}, \dots, Y_N^{\text{obs}}) = \sum_{j=1}^N \mathbf{1}_{Y_j^{\text{obs}} < Y_i^{\text{obs}}} + \frac{1}{2} \left(1 + \sum_{j=1}^N \mathbf{1}_{Y_j^{\text{obs}} = Y_i^{\text{obs}}} \right) - \frac{N+1}{2}.$$

- 给定 N 个个体的秩排序 R_i , 可以构造治疗组和对照组平均秩排序差异的绝对值作为检验统计量:

$$T^{\text{rank}} = |\bar{R}_t - \bar{R}_c| = \left| \frac{\sum_{i:W_i=1} R_i}{N_t} - \frac{\sum_{i:W_i=0} R_i}{N_c} \right|$$

1. *Journal of the American Medical Association*, 2000; 283: 2689-2693.

- 实验者也可以使用霍特林 (Hotelling's) T^2 统计量, 用于衡量治疗组和对照组平均值之间的马氏平方距离。

Paul et al. 的蜂蜜实验数据

Table 5.5. Randomization Distribution for Two Statistics for the Honey Data from Table 5.3

W_1	W_2	W_3	W_4	W_5	W_6	Statistic: Absolute Value of Difference in Average	
						Levels (Y_i)	Ranks (R_i)
0	0	0	1	1	1	-1.00	-0.67
0	0	1	0	1	1	-3.67	-3.00
0	0	1	1	0	1	-1.00	-0.67
0	0	1	1	1	0	-1.67	-1.67
0	1	0	0	1	1	-0.33	0.00
0	1	0	1	0	1	2.33	2.33
0	1	0	1	1	0	1.67	1.33
0	1	1	0	0	1	-0.33	0.00
0	1	1	0	1	0	-1.00	-1.00
0	1	1	1	0	0	1.67	1.33
1	0	0	0	1	1	-1.67	-1.33
1	0	0	1	0	1	1.00	1.00
1	0	0	1	1	0	0.33	0.00
1	0	1	0	0	1	-1.67	-1.33
1	0	1	0	1	0	-2.33	-2.33
1	0	1	1	0	0	0.33	0.00
1	1	0	0	0	1	1.67	1.67
1	1	0	0	1	0	1.00	0.67
1	1	0	1	0	0	3.67	3.00
1	1	1	0	0	0	1.00	0.67

Note: Observed values in boldface (R_i is rank(Y_i)). Data based on cough frequency for first six units from honey study.

- 值得注意的是, 当样本量很大时, 具有正概率的分配机制向量的选择会剧增, 此时采用全部的分配向量来计算某统计量在费希尔尖锐零假设下的分布是困难的。
- 蒙特卡洛模拟可以有效解决该问题, 可以考虑仅针对随机选择的可能分配向量的子集 (采样 K 次), 对于 $k = 1, \dots, K$, 通过计算 $T^{\text{dif},k} = \bar{Y}_{t,k} - \bar{Y}_{c,k}$, 来近似得到零假设成立时的分布, 从而得到近似的 P 值:

$$\hat{p} = \frac{1}{K} \sum_{k=1}^N \mathbf{1}_{T^{\text{dif},k} \geq T^{\text{dif},\text{obs}}}.$$

① 因果推断的潜在结果框架简介

② 完全随机实验 (Completely Randomized Experiments): R.A.Fisher 与 J. Neyman 之间的分歧

费希尔精确 P 值方法 (R.A.Fisher's Exact P-Values)

奈曼重复抽样方法 (J.Neyman's Repeated Sampling Approach)

③ 总结

1. *Journal of the American Medical Association*, 1997; 278: 1039-1044.

- 在费希尔开发这种方法的同一时期,奈曼专注于平均治疗效果的估计和推断方法,他试图找到无偏的点估计量,以及在大样本中的区间估计量。
- 如前所述,他对平均效果的关注不同于费舍尔:即使部分或全部单位水平的治疗效果不同于零,整个人群的平均效果也可能等于零。奈曼的基本问题如下:如果所有单位都接受积极的治疗,平均潜在结果 $Y(1)$ 会是什么?与该结果相比,如果所有单位都接受对照治疗,平均潜在结果 $Y(0)$ 又如何?

奈曼重复抽样方法 (J.Neyman's Repeated Sampling Approach) : 简介

- 在奈曼重复抽样方法中，我们感兴趣的估计量是这些平均值之间的差异 $\tau_{fs} = \bar{Y}(1) - \bar{Y}(0) = \sum_{i=1}^N (Y_i(1) - Y_i(0)) / N$ ，其中下标 fs 表示有限样本 (finite sample)，如前所述，有限样本集中的随机性完全由分配机制 \mathbf{W} 所解释。
- 此外，我们的目标是为因果估计构建一个区间估计量，且该估计量是平均治疗效果估计量的样本方差的无偏估计量。

1. *Journal of the American Medical Association*, 2000; 284: 2689-2694.

- 考虑一个由 N 个个体组成的群体, 对于每个个体, 存在两种潜在结果, 分别对应于控制和治疗下的结果, 即 $Y_i(0)$ 和 $Y_i(1)$ 。潜在结果被视为常量, 唯一的随机性由分配机制 \mathbf{W} 所解释, 其中的第 i 个分量记为 W_i 。在完全随机实验中, 分配机制的概率分布已知。
- 总体的平均治疗效果是:

$$\tau_{\text{fs}} = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = \bar{Y}(1) - \bar{Y}(0)$$

其中 $\bar{Y}(0)$ 和 $\bar{Y}(1)$ 分别是对照组和治疗组的潜在结果向量:

$$\bar{Y}(0) = \frac{1}{N} \sum_{i=1}^N Y_i(0), \quad \text{且} \quad \bar{Y}(1) = \frac{1}{N} \sum_{i=1}^N Y_i(1).$$

- 平均治疗效果的估计量 $\hat{\tau}^{\text{dif}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$ 的样本方差是:

$$\mathbb{V}_W(\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N}$$

其中 S_c^2 和 S_t^2 分别是 $Y_i(0)$ 和 $Y_i(1)$ 样本方差:

$$S_c^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(0) - \bar{Y}(0))^2, \quad S_t^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - \bar{Y}(1))^2$$

以及 S_{tc}^2 是个体单位的治疗效果的样本方差:

$$\begin{aligned} S_{tc}^2 &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - (\bar{Y}(1) - \bar{Y}(0)))^2 \\ &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - \tau_{fs})^2 \end{aligned}$$

奈曼重复抽样方法：另一种表示方法

- 此外, $\hat{\tau}^{dif}$ 的样本方差有另一种有意义的表示: 记 ρ_{tc} 是群体潜在结果 $Y_i(1)$ 和 $Y_i(0)$ 的相关系数, 可以将个体单位的治疗效果的样本方差改写成 ρ_{tc} 的函数:

$$S_{tc}^2 = S_c^2 + S_t^2 - 2 \cdot \rho_{tc} \cdot S_c \cdot S_t$$

其中

$$\rho_{tc} = \frac{1}{(N-1) \cdot S_c \cdot S_t} \sum_{i=1}^N (Y_i(1) - \bar{Y}(1)) \cdot (Y_i(0) - \bar{Y}(0))$$

- 其中, ρ_{tc} 作为相关系数的取值为 $[-1, 1]$, 将以上 S_{tc}^2 的表示代入 $\hat{\tau}^{\text{dif}}$ 的样本方差的表达式可以得到 $\hat{\tau}^{\text{dif}}$ 的样本方差的另一种表示:

$$\mathbb{V}_W(\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}) = \frac{N_t}{N \cdot N_c} \cdot S_c^2 + \frac{N_c}{N \cdot N_t} \cdot S_t^2 + \frac{2}{N} \cdot \rho_{tc} \cdot S_c \cdot S_t$$

奈曼重复抽样方法: 常数治疗效果假设情形

- 当潜在结果 $Y_i(1)$ 和 $Y_i(0)$ 的相关系数 $\rho_{tc} = 1$ 时, 估计量 $\hat{\tau}^{\text{dif}}$ 的样本方差达到最大值:

$$\begin{aligned}\mathbb{V}_W(\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \mid \rho_{tc} = 1) &= \frac{N_t}{N \cdot N_c} \cdot S_c^2 + \frac{N_c}{N \cdot N_t} \cdot S_t^2 + \frac{2}{N} \cdot S_c \cdot S_t \\ &= \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{(S_c - S_t)^2}{N}\end{aligned}$$

- 当治疗效果是常数以及具有可加性时, 即 $Y_i(1) - Y_i(0) = \tau$, 自动满足相关系数 $\rho_{\tau\epsilon} = 1$ 的条件, 此时有:

$$\mathbb{V}^{\text{const}} = \mathbb{V}_W \left(\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \mid \rho_{tc} = 1, S_c^2 = S_t^2 \right) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$$

奈曼重复抽样方法: S_t, S_c 和 S_{tc} 的不可观测性

- 然而, 我们无法从观测样本中获得准确的样本方差 S_t 以及 S_c , 特别是个体单位的治疗效果的样本方差 S_{tc} , 因为我们无法同时观察同一个个体的 $Y_i(1)$ 和 $Y_i(0)$ 。
- 因此, 我们无法直接观察整个人群治疗效果的变化, 导致无法直接估计 S_{tc} 。如前所述, 如果治疗效果是常数且具有可加性, 即 $Y_i(1) - Y_i(0) = \tau_{fs}$, 那么样本方差的这个分量等于零。

- 若个体治疗效果 $Y_i(1) - Y_i(0)$ 是常数, 则估计量 $\hat{\tau}^{\text{dif}}$ 的样本方差的一个无偏估计量是:

其中

以及

$$s_t^2 = \frac{1}{N_t - 1} \sum_{i: W_i = 1} (Y_i(1) - \bar{Y}_t^{\text{obs}})^2 = \frac{1}{N_t - 1} \sum_{i: W_i = 1} (Y_i^{\text{obs}} - \bar{Y}_t^{\text{obs}})^2$$

奈曼重复抽样方法: \hat{V}^{neyman} 估计量的优良性

- 值得注意的是, \hat{V}^{neyman} 的期望至少与平均治疗效果的估计量 $\hat{\tau}^{\text{dif}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$ 的真实样本方差一样大。换言之, \hat{V}^{neyman} 是统计保守的。
- 使用 \hat{V}^{neyman} 作为样本方差的估计量的第二个好处是, 当作为无限超总体的平均治疗效果 τ^{dif} 的样本方差估计量时, 它是一个无偏估计。

奈曼重复抽样方法: 另一种保守估计 $\hat{\mathbb{V}}\rho_{tc}=1$

- 此外, 对于另一个统计保守的样本方差估计量 $\mathbb{V}_W(\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} \mid \rho_{tc} = 1)$, 也可以通过治疗组和对照组的样本方差来实现保守估计:

$$\begin{aligned}\hat{\mathbb{V}}^{\rho_{tc}=1} &= s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N} \\ &= \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t} - \frac{(s_t - s_c)^2}{N}\end{aligned}$$

- 该统计量在治疗组和对照组的样本方差不一致时, 将同样实现保守估计, 且比 \hat{V}^{neyman} 对应的置信区间更紧。

奈曼重复抽样方法: 常数治疗效果假设情形下的估计 \hat{V}^{const}

- 此外, 考虑在治疗效果不变的附加假设下, 即 $Y_i(1) - Y_i(0) = \tau_{fs}$. 这个治疗的无效性假设隐含了, 两个潜在结果的总体方差相等的事实, 即 $S^2 \equiv S_c^2 = S_t^2$. 该共同方差的估计量为:

$$\begin{aligned} s^2 &= \frac{1}{N-2} \cdot (s_c^2 \cdot (N_c - 1) + s_t^2 \cdot (N_t - 1)) \\ &= \frac{1}{N-2} \cdot \left(\sum_{i:W_i=0} (Y_i^{\text{obs}} - \bar{Y}_c^{\text{obs}})^2 + \sum_{i:W_i=1} (Y_i^{\text{obs}} - \bar{Y}_t^{\text{obs}})^2 \right) \end{aligned}$$

- 当治疗效果为常数时, 样本量越大, 统计量 $\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$ 样本方差的估计越精确:

$$\hat{\mathbb{V}}^{\text{const}} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t} \right)$$

Duflo-Hanna-Ryan 教师工资激励计划数据

- 实验介绍：随机实验，在印度农村评估工资激励计划对教师教学质量的影响。
- 实验对象：参与工资激励计划 ($N_t=54$) 或没有参与工资激励计划 ($N_c=53$) 的老师。
- 治疗组：工资结构发生变化，教师的工资与一个月的出勤率挂钩；对照组：无变化。
- 潜在结果：第一个是学校在随机访问期间开放的比例 (open)。第二个结果是完成写作测试 (pctpostwritten) 的学生百分比。第三个是参与写作测试的学生分数均值 (written)。第四个是全部学生的平均写作测试分数 (written- all)。
- 协变量：在实验之前能够完成测试的学生百分比 (pctprewritten)。

Duflo-Hanna-Ryan 教师工资激励计划数据

Table 6.1. Summary Statistics for Duflo-Hanna-Ryan Teacher-Incentive Observed Data

Variable		Control ($N_c = 54$)		Treated ($N_t = 53$)			
		Average	(S.D.)	Average	(S.D.)	Min	Max
Pre-treatment	pctprewritten	0.19	(0.19)	0.16	(0.17)	0.00	0.67
Post-treatment	open	0.58	(0.19)	0.80	(0.13)	0.00	1.00
	pctpostwritten	0.47	(0.19)	0.52	(0.23)	0.05	0.92
	written	0.92	(0.45)	1.09	(0.42)	0.07	2.22
	written_all	0.46	(0.32)	0.60	(0.39)	0.04	1.43

图 13: Duflo-Hanna-Ryan 教师工资激励计划数据的描述性统计

- 若选择 open 作为感兴趣的潜在结果:

$$\hat{\tau}^{\text{dif}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} = 0.80 - 0.58 = 0.22$$

Duflo-Hanna-Ryan 教师工资激励计划数据

- $s_c^2 = 0.19^2$, $s_t^2 = 0.13^2$, 以及 $s^2 = 0.16^2$

$$\hat{V}^{\text{neyman}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t} = 0.0311^2$$

$$\hat{\mathbb{V}}^{\rho_{tc}=1} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N} = 0.0305^2$$

$$\hat{V}^{\text{const}} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t} \right) = 0.0312^2.$$

- 奈曼对于 ATE 的精度感兴趣很大程度由统计保守的置信区间所驱动:

$$\Pr_W \left(C_L \left(\mathbf{Y}^{\text{obs}}, \mathbf{W} \right) \leq \tau \leq C_U \left(\mathbf{Y}^{\text{obs}}, \mathbf{W} \right) \right) \geq 1 - \alpha$$

$$\begin{aligned} \text{CI}_{\text{neyman}}^{1.90}(\tau_{\text{fs}}) &= \left(\hat{\tau}^{\text{dif}} + z_{0.10/2} \cdot \sqrt{\hat{\mathbb{V}}^{\text{neyman}}}, \hat{\tau}^{\text{dif}} + z_{1-0.10/2} \cdot \sqrt{\hat{\mathbb{V}}^{\text{neyman}}} \right) \\ &= (0.2154 - 1.645 \cdot 0.0311, 0.2154 + 1.645 \cdot 0.0311) = (0.1642, 0.2667) \end{aligned}$$

$$\text{CI}_{\text{const}}^{0.90}(\tau_{\text{fs}}) = (0.1640, 0.2668)$$

$$\text{CI}_{\rho_{tc}=1}^{0.90}(\tau_{fs}) = (0.1652, 0.2657)$$

奈曼重复抽样方法：假设检验

$$H_0^{\text{neyman}} : \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = 0, \text{ 以及}$$

$$H_a^{\text{neyman}} : \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) \neq 0$$

$$t = \frac{\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}}{\sqrt{\hat{V}_{\text{neyman}}}} = \frac{0.2154}{0.0311} = 6.9$$

- t-统计量: $2 \cdot (1 - \Phi(6.9)) < 0.001$.
- 结论: 拒绝原假设, 即平均处理效果为零。

Duflo-Hanna-Ryan 教师工资激励计划数据

Table 6.2. Estimates of Components of Variance of Estimator for the Effect of Teacher Incentives on the Proportion of Days that the School is Open; $N_c = 54$, $N_t = 53$, Duflo-Hanna-Ryan Data

Estimated means	\bar{Y}_c^{obs}	0.58
	\bar{Y}_t^{obs}	0.80
	$\hat{\tau}$	0.22
Estimated variance components	s_c^2	0.19 ²
	s_t^2	0.13 ²
	s^2	0.16 ²
Sampling variance estimates	$\hat{V}^{neyman} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$	0.03 ²
	$\hat{V}^{const} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t} \right)$	0.03 ²
	$\hat{V}^{\rho_{tc}=1} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N}$	0.03 ²

图 14: Duflo-Hanna-Ryan 教师激励对学校开放日比例 (open) 影响

Duflo-Hanna-Ryan 教师工资激励计划数据

Table 6.3. Estimates of, and Confidence Intervals for, Average Treatment Effects for Duflo-Hanna-Ryan Teacher-Incentive Data

\widehat{ATE}	$\widehat{(s.e.)}$	95% C.I.
0.22	(0.03)	(0.15,0.28)
0.05	(0.04)	(−0.03,0.13)
0.17	(0.08)	(0.00,0.34)
0.14	(0.07)	(0.00,0.28)

图 15: Duflo-Hanna-Ryan 教师激励数据平均治疗效果估计和置信区间

- 考虑治疗前变量作为伪潜在结果 (以检验随机化):
ATE=-0.03, 95%CI=(-0.10,0.04).

奈曼重复抽样方法 (J.Neyman's Repeated Sampling Approach)

Duflo-Hanna-Ryan 教师工资激励计划数据

Table 6.4. Estimates of, and Confidence Intervals for, Average Treatment Effects for Duflo-Hanna-Ryan Teacher-Incentive Data

Variable	pctpre = 0 (N = 40)			pctprewritten > 0 (N = 67)			Difference		
	$\hat{\tau}$	$\widehat{(s.e.)}$	95% C.I.	$\hat{\tau}$	$\widehat{(s.e.)}$	95% C.I.	EST	$\widehat{(s.e.)}$	95% C.I.
open	0.23	(0.05)	(0.14,0.32)	0.21	(0.04)	(0.13,0.29)	0.02	(0.06)	(-0.10,0.14)
pctpost	-0.004	(0.06)	(-0.16,0.07)	0.11	(0.05)	(0.01,0.21)	-0.15	(0.08)	(-0.31,0.00)
written									
written	0.20	(0.10)	(0.00,0.40)	0.18	(0.10)	(-0.03,0.38)	0.03	(0.15)	(-0.26,0.31)
written	0.04	(0.07)	(-0.10,0.19)	0.22	(0.09)	(0.04,0.40)	-0.18	(0.12)	(-0.41,0.05)
_all									

图 16: Duflo-Hanna-Ryan 教师激励数据平均治疗效果估计和置信区间

- ### ③ 总结

1. *Journal of Management Studies*, 1997, 34, 1, 1-14.

- 关于这两种角度的争论：1935 年，奈曼向英国皇家统计学会提交了一篇论文 “Statistical problems in agricultural experimentation”，费希尔和奈曼在讨论文章时发生了激烈的争执。
- 费希尔关心有限样本情形下，尖锐零假设的假设检验，在零假设成立时，可以精确求出统计量的分布，蕴含了 bootstrap 的思想。
- 奈曼在完全随机实验中关心治疗的平均效果，提出了有限样本中 ATE 的一个无偏估计量。
- 为抽样方差找到一个本身无偏的估计量是不可能的，相反，奈曼证明了这个估计量的抽样方差的标准估计量是正偏的，除非处理效果是常数和可加的，在这种情况下它是无偏的。

Thanks!