集智俱乐部
www.swarma.org

因果推断的潜在结果框架在观察性研究的应用

李昊轩

北京大学，大数据科学研究中心

2021.11.14

实验性研究：线性回归

- 回归方法可以有效应用于实验和观察性研究来估计因果效应。在因果推断中，回归方法的自变量包括接受治疗的指标变量 W_i，通常还包括一些治疗前变量 X_i 。
- 回归方程的参数通过最小二乘法估算，我们主要关注治疗指标的系数。本方法中考虑的回归模型使用超总体视角。
- 在条件均值的模型中，我们主要关注平均治疗效果，是统计模型的一个参数，即可以被视为统计模型参数的推断问题。
- 在当前完全随机实验的背景下，我们主要关注对于平均治疗效果的最小二乘估计量的大样本无偏性，而非线性模型的有效性，即模型是否提供条件均值的准确描述。

线性回归：不纳入协变量情形

- 首先不考虑将协变量 X_i 纳入回归模型，此时线性模型的解释变量只有个体接受治疗的指标 W_i，建立观察结果 Y_i^{obs} 的模型如下：

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + \varepsilon_i$$

- 其中 ε_i 是无法观测的残差项，τ 的最小二乘估计基于 α 和 τ 的残差平方和最小化：

$$\left(\hat{\tau}^{\text{ols}}, \hat{\alpha}^{\text{ols}}\right) = \arg\min_{\tau,\alpha} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i\right)^2$$

线性回归：不纳入协变量情形

- 解得：

$$\hat{\tau}^{\text{ols}} = \frac{\sum_{i=1}^{N}(W_i - \bar{W}) \cdot \left(Y_i^{\text{obs}} - \bar{Y}^{\text{obs}}\right)}{\sum_{i=1}^{N}(W_i - \bar{W})^2}, \text{ 以及} \hat{\alpha}^{\text{ols}} = \bar{Y}^{\text{obs}} - \hat{\tau}^{\text{ols}} \cdot \bar{W}$$

其中

$$\bar{Y}^{\text{obs}} = \frac{1}{N}\sum_{i=1}^{N} Y_i^{\text{obs}} \quad \text{以及} \quad \bar{W} = \frac{1}{N}\sum_{i=1}^{N} W_i = \frac{N_t}{N}$$

- 此时，平均治疗效果的最小二乘估计 $\hat{\tau}^{\text{ols}}$ 与有限样本情形下的自然估计量 $\hat{\tau}^{\text{dif}}$ 相同：

$$\hat{\tau}^{\text{ols}} = \bar{Y}_{\text{t}}^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}} = \hat{\tau}^{\text{dif}}$$

其中 $\bar{Y}_{\text{t}}^{\text{obs}} = \sum_{i:W_i=1} Y_i^{\text{obs}}/N_{\text{t}}$ 和 $\bar{Y}_{\text{c}}^{\text{obs}} = \sum_{i:W_i=0} Y_i^{\text{obs}}/N_{\text{c}}$ 分别是治疗组和对照组的观测潜在结果均值。

线性回归：不纳入协变量情形

- 除了点估计以外，我们希望能够估计最小二乘估计 $\hat{\tau}^{\text{ols}}$ 的样本方差以及置信区间。
- 首先考虑治疗组和对照组的方差同质化假设：$\sigma^2_{Y|W} = \sigma^2_c = \sigma^2_t$，此时最小二乘回归方法的残差项的方差估计为：

$$\hat{\sigma}^2_{Y|W} = \frac{1}{N-2}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 = \frac{1}{N-2}\sum_{i=1}^{N}\left(Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}}\right)^2$$

其中残差的点估计是 $\hat{\varepsilon}_i = Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}}$，观测结果 $\hat{Y}_i^{\text{obs}}$ 的预测值是：

$$\hat{Y}_i^{\text{obs}} = \begin{cases} \hat{\alpha}^{\text{ols}} & \text{if } W_i = 0 \\ \hat{\alpha}^{\text{ols}} + \hat{\tau}^{\text{ols}} & \text{if } W_i = 1 \end{cases}$$

线性回归：不纳入协变量情形

- 最小二乘估计的方差估计为：

$$\hat{\sigma}^2_{Y|W} = \frac{1}{N-2}\left(\sum_{i:W_i=0}\left(Y_i^{\text{obs}} - \bar{Y}_c^{\text{obs}}\right)^2 + \sum_{i:W_i=1}\left(Y_i^{\text{obs}} - \bar{Y}_t^{\text{obs}}\right)^2\right)$$

该结果与同质化假设的 s^2 相同，即治疗组和对照组的同质化样本方差。

- 因此，在治疗组和对照组的方差同质化假设下，最小二乘估计 $\hat{\tau}^{\text{ols}}$ 的样本方差估计为：

$$\hat{\mathbb{V}}^{\text{homosk}} = \frac{\hat{\sigma}^2_{Y|W}}{\sum_{i=1}^{N}(W_i - \bar{W})^2} = s^2 \cdot \left(\frac{1}{N_{\text{c}}} + \frac{1}{N_{\text{t}}}\right)$$

该结果与 $\hat{\mathbb{V}}^{\text{const}}$ 相同，由于常数治疗效果的假设隐含了线性模型中的方差同质化假设。

线性回归：不纳入协变量情形

- 此外，考虑该估计的相合性，设 p 是样本量趋于无穷时个体的倾向性得分 e_i，则随着样本量的增加，归一化的样本方差估计量依概率收敛于真实样本方差：

$$N \cdot \hat{\mathbb{V}}^{\text{homosk}} \xrightarrow{p} \frac{\sigma^2_{Y|W}}{p \cdot (1-p)}$$

线性回归：不纳入协变量情形

- 接下来考虑异质性方差时，最小二乘估计的样本方差的一个稳健估计：

$$\hat{\mathbb{V}}^{\text{hetero}} = \frac{\sum_{i=1}^{N} \hat{\varepsilon}_i^2 \cdot (W_i - \bar{W})^2}{\left(\sum_{i=1}^{N} (W_i - \bar{W})^2\right)^2}$$

其中

$$s_c^2 = \frac{1}{N_{\text{c}} - 1} \sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \bar{Y}_c^{\text{obs}}\right)^2, \; s_t^2 = \frac{1}{N_{\text{t}} - 1} \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \bar{Y}_{\text{t}}^{\text{obs}}\right)^2$$

线性回归：不纳入协变量情形

- 因此，在治疗组和对照组的方差异质化假设下，最小二乘估计 $\hat{\tau}^{\text{ols}}$ 的样本方差估计为：

$$\hat{\mathbb{V}}^{\text{hetero}} = \frac{s_c^2}{N_{\text{c}}} + \frac{s_t^2}{N_{\text{t}}}$$

- 该结果与 $\hat{V}^{\text{neyman}}$ 相同，因此，在不引入协变量时，回归方法的样本方差与奈曼重复抽样方法得到的样本方差估计大致相同。

线性回归：纳入协变量情形

- 若在线性模型中考虑协变量，观测结果的回归模型为：

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + X_i\beta + \varepsilon_i$$

其中 X_i 是个体 i 的协变量向量，使用最小二乘法来估计回归系数：

$$\left(\hat{\tau}^{\text{ols}}, \hat{\alpha}^{\text{ols}}, \hat{\beta}^{\text{ols}}\right) = \arg\min_{\tau,\alpha,\beta} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta\right)^2$$

- 类似地，考虑超总体视角，用上标 $*$ 表示超总体中的最小二乘结果：

$$\left(\alpha^*, \tau^*, \beta^*\right) = \arg\min_{\alpha,\beta,\tau} \mathbb{E}\left[\left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta\right)^2\right]$$

线性回归：纳入协变量情形

- 在线性模型中考虑协变量时：首先，有限样本中的最小二乘回归 $\hat{\tau}^{\text{ols}}$ 是超总体的治疗效果 τ_{sp} 的无偏估计，且超总体在期望意义下的最小二乘回归 τ^* 与超总体的治疗效果 τ_{sp} 等价；此外，有限样本中的最小二乘回归 $\hat{\tau}^{\text{ols}}$ 依分布收敛到超总体的治疗效果 τ_{sp}
- 考虑在超总体中进行有限抽样，在该有限样本集上进行的完全随机试验，则有：$\tau^* = \tau_{\text{sp}}$ 且

$$\sqrt{N}\cdot\left(\hat{\tau}^{\text{ols}} - \tau_{\text{sp}}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}\left[(W_i - p)^2 \cdot \left(Y_i^{\text{obs}} - \alpha^* - \tau_{\text{sp}} \cdot W_i - X_i\beta^*\right)^2\right]}{p^2 \cdot (1-p)^2}\right)$$

线性回归：纳入协变量情形

- 基于以上的观点，无论模型的线性假设是否正确，回归方法都具有无偏性和相合性。
- 然而，当协变量能够从一定程度上解释潜在因果时，会导致 $\mathbb{V}_{\text{sp}}\left(Y_i(w) \mid X_i = x\right) = \sigma^2_{YW,X}$ 显著小于未纳入协变量考虑的条件方差 $\sigma^2_{Y|W}$，此时纳入协变量的线性模型会使治疗效果的最小二乘估计更加稳健。

线性回归：纳入协变量情形

- 此时，平均治疗效果的样本方差可以使用最小二乘方法进行估计，在方差异质性和同质性假设下，样本方差分别为：

$$\hat{V}_{sp}^{\text{hetero}} = \frac{1}{N\left(N-1-\dim\left(X_i\right)\right)} \cdot \frac{\sum_{i=1}^{N}\left(W_i-\bar{W}\right)^2 \cdot \left(Y_i^{\text{obs}} - \hat{\alpha}^{\text{ols}} - \hat{\tau}^{\text{ols}} - X_i\hat{\beta}^{\text{ols}}\right)^2}{\left(\bar{W}\cdot\left(1-\bar{W}\right)\right)^2}$$

以及

$$\hat{V}_{\text{sp}}^{\text{homo}} = \frac{1}{N\left(N-1-\dim\left(X_i\right)\right)} \cdot \frac{\sum_{i=1}^{N}\left(Y_i^{\text{obs}} - \hat{\alpha}^{\text{ols}} - \hat{\tau}^{\text{ols}} - X_i\hat{\beta}^{\text{ols}}\right)^2}{\bar{W}\cdot\left(1-\bar{W}\right)}$$

线性回归：含交互作用情形

- 在考虑将协变量纳入线性模型中时，如果我们预期协变量和结果之间的关联因治疗状态而异，可以考虑协变量与治疗指标的交互作用。含交互作用的模型可以进一步提高估计器的精度，同时提升了错误指定模型的鲁棒性，此时回归模型如下：

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + X_i\beta + W_i \cdot (X_i - \bar{X})\gamma + \varepsilon_i$$

- 设 $\hat{\alpha}^{\text{ols}}$ ，$\hat{\tau}^{\text{ols}}$ ，$\hat{\beta}^{\text{ols}}$ 和 $\hat{\gamma}^{\text{ols}}$ 是以上模型的最小二乘回归结果：

$$\left(\hat{\tau}^{\text{ols}}, \hat{\alpha}^{\text{ols}}, \hat{\beta}^{\text{ols}}, \hat{\gamma}^{\text{ols}}\right) = \arg\min_{\tau,\alpha,\beta,\gamma} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta - W_i \cdot (X_i - \bar{X})\gamma\right)^2$$

线性回归：含交互作用情形

- 此外，设 $\alpha^*, \tau^*, \beta^*$，和 γ^* 是超总体视角下对应的回归结果：

$$\left(\alpha^*, \tau^*, \beta^*, \gamma^*\right) = \arg\min_{\alpha,\beta,\tau,\gamma} \mathbb{E}_{\text{sp}}\left[\left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta - W_i \cdot (X_i - \mu_X)\gamma\right)^2\right]$$

  则我们有类似的结论，证实该模型的无偏性和相合性。
- 考虑在超总体中进行有限抽样，在该有限样本集上进行的完全随机试验，则有：$\tau^* = \tau_{\text{sp}}$ 且

$$\sqrt{N} \cdot \left(\hat{\tau}^{\text{ols}} - \tau_{\text{sp}}\right) \xrightarrow{d} \mathcal{N}$$

$$\left(0, \frac{\mathbb{E}_{\text{sp}}\left[(W_i - p)^2 \cdot \left(Y_i^{\text{obs}} - \alpha^* - \tau_{\text{sp}} \cdot W_i - X_i\beta^* - W_i \cdot (X_i - \mu_X)\gamma^*\right)^2\right]}{p^2 \cdot (1-p)^2}\right)$$

检验因果效应：平均效果为零

- 此外，对于治疗效果相关的回归系数，需要进行假设检验来判断治疗效果的显著性。若以上模型被正确指定，则在超总体观点下：

$$\mathbb{E}_{\text{sp}}\left[Y_i^{\text{obs}} \mid X_i = x, W_i = w\right] = \alpha + \tau \cdot w + x\beta + w \cdot (x - \mu_X)\gamma'$$

- 首先考虑在所有协变量相同的条件下，平均治疗效果为零的假设检验：

$$H_0 : \mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0) \mid X_i = x\right] = 0, \quad \forall x$$

以及备择假设：

$$H_a : \mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0) \mid X_i = x\right] \neq 0, \text{对于某些} x$$

检验因果效应：平均效果为零

- 此时，若零假设 $\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x] = 0$ 对所有 x 均成立，则超总体中的回归系数 τ_{sp} 和 γ^* 均为零；反之则不成立，即使回归系数 τ_{sp} 和 γ^* 均为零，零假设仍有可能被违背，即对于某些 x，$\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x]$ 不等于 0 。
- 考虑回归系数 $\left(\hat{\tau}^{\text{ols}}, \hat{\gamma}^{\text{ols}}\right)$ 的协方差矩阵：

$$\mathbb{V}_{\tau,\gamma} = \begin{pmatrix} \mathbb{V}_\tau & \mathbb{C}_{\tau,\gamma} \\ \mathbb{C}_{\tau,\gamma}^T & \mathbb{V}_\gamma \end{pmatrix}$$

可以构造如下的统计量：

$$Q_{\text{zero}} = \begin{pmatrix} \hat{\tau}^{\text{ols}} \\ \hat{\gamma}^{\text{ols}} \end{pmatrix}^T \hat{\mathbb{V}}_{\tau,\gamma}^{-1} \begin{pmatrix} \hat{\tau}^{\text{ols}} \\ \hat{\gamma}^{\text{ols}} \end{pmatrix}$$

检验因果效应：平均效果为常数

- 同理，考虑对治疗效果为常数的另一种假设检验：

$$H_0' : \mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0) \mid X_i = x\right] = \tau_{\text{sp}}, \quad \forall x$$

以及备择假设：
H_a'：$\exists x_0, x_1$，使得
$\mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0) \mid X_i = x_0\right] \neq \mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0) \mid X_i = x_1\right]$.

- 可以构造如下的统计量：

$$Q_{\text{const}} = \left(\hat{\gamma}^{\text{ols}}\right)^T \hat{\mathbb{V}}_{\gamma}^{-1} \hat{\gamma}^{\text{ols}}$$

检验因果效应：渐近统计量

- 考虑在超总体中进行有限抽样，在该有限样本集上进行的完全随机试验。
- 若存在常数 τ，使得对于所有个体均有 $Y_i(1) - Y_i(0) = \tau$，则有：$\gamma^* = 0$ 且 $Q_{\text{const}} \xrightarrow{d} X\left(\dim\left(X_i\right)\right)$
- 若对于所有个体均有 $Y_i(1) - Y_i(0) = 0$，则有：$Q_{\text{zero}} \xrightarrow{d} X\left(\dim\left(X_i\right) + 1\right)$

LRC-CPPT 胆固醇数据

- 实验介绍：随机实验，研究消胆胺对胆固醇水平的影响
- 实验对象：$N_t = 165$ 接受消胆胺药物和 $N_c = 172$ 接受安慰剂
- 协变量：chol1，在对全部 337 个体宣传关于低胆固醇饮食益处的信息之前的胆固醇水平；chol2，是在这一建议之后，但在随机分配给消胆胺或安慰剂之前。
- cholp = 0.25· chol1 +0.75· chol2
- 潜在结果：主要结果 cholf，是随机分组后胆固醇读数的平均值，次要结果 comp，是依从性度量，即个体实际服用的名义上指定剂量的胆胺或安慰剂的百分比。

LRC-CPPT 胆固醇数据

Table 7.1. ***Summary Statistics for PRC-CPPT Cholesterol Data***

	Variable	Control (N_c =172)		Treatment (N_t =165)			
		Average	Sample (S.D.)	Average	Sample (S.D.)	Min	Max
Pre-treatment	chol1	297.1	(23.1)	297.0	(20.4)	247.0	442.0
	chol2	289.2	(24.1)	287.4	(21.4)	224.0	435.0
	cholp	291.2	(23.2)	289.9	(20.4)	233.0	436.8
Post-treatment	cholf	282.7	(24.9)	256.5	(26.2)	167.0	427.0
	chold	−8.5	(10.8)	−33.4	(21.3)	−113.3	29.5
	comp	74.5	(21.0)	59.9	(24.4)	0	101.0

图 1: PRC-CPPT 胆固醇数据的描述性估计

LRC-CPPT 胆固醇数据

Table 7.2. ***Regression Estimates for Average Treatment Effects for the PRC-CPPT Cholesterol Data from Table 7.1***

Covariates	Effect of Assignment to Treatment on Post-Cholesterol Level		Effect of Assignment to Treatment on Compliance	
	Est	$\widehat{(\text{s. e.})}$	Est	$\widehat{(\text{s. e.})}$
No covariates	−26.22	(3.93)	−14.64	(3.51)
`cholp`	−25.01	(2.60)	−14.68	(3.51)
`chol1, chol2`	−25.02	(2.59)	−14.95	(3.50)
`chol1, chol2`, interacted with W	−25.04	(2.56)	−14.94	(3.49)

图 2: PRC-CPPT 胆固醇数据平均治疗效果的回归估计

LRC-CPPT 胆固醇数据

Table 7.3. ***Regression Estimates for Average Treatment Effects on Post-Cholesterol Levels for the PRC-CPPT Cholesterol Data from Table 7.1***

Covariates	Model for Levels		Model for Logs	
	Est	$\widehat{(\text{s.e.})}$	Est	$\widehat{(\text{s.e.})}$
Assignment	−25.04	(2.56)	−0.098	(0.010)
Intercept	−3.28	(12.05)	−0.133	(0.233)
chol1	0.98	(0.04)	−0.133	(0.233)
chol2-chol1	0.61	(0.08)	0.602	(0.073)
chol1 × Assignment	−0.22	(0.09)	−0.154	(0.107)
(chol2-chol1) × Assignment	0.07	(0.14)	0.184	(0.159)
R-squared	0.63		0.57	

图 3: PRC-CPPT 胆固醇数据中平均治疗效果的回归系数

LRC-CPPT 胆固醇数据

Table 7.4. ***P-Values for Tests for Constant and Zero Treatment Effects, Using*** `chol1` ***and*** `chol2-chol1` ***as Covariates for the PRC-CPPT Cholesterol Data from Table 7.1***

		Post-Cholesterol Level	Compliance
Zero treatment effect	$\mathcal{X}^2(3)$ approximation	<0.001	<0.001
	Fisher exact p-value	<0.001	0.001
Constant treatment effect	$\mathcal{X}^2(2)$ approximation	0.029	0.270

图 4: 常数和零因果效应假设检验的 P 值

实验性研究：基于模型的推断方法

- 考虑基于模型的推断方法，即使是在有限样本集中，我们依然将潜在结果视为随机变量，所以它们的任何函数也将是随机变量。
- 为所有潜在结果建立一个随机模型，它通常依赖于一些未知的参数，通过贝叶斯方法，使用假设模型来估算给定观察数据的缺失潜在结果，并使用这些结果来对感兴趣的统计量进行推断。
- 在某种程度上，所有因果推断的方法都可以被视为插补方法，因为任何因果效应的估计都依赖于缺失的潜在结果，所以基于模型的推断方法的本质是对这些缺失的潜在结果的估计。
- 一般来讲，与费希尔的精确 P 值方法、奈曼的重复抽样方法或是回归方法相比，基于模型的方法非常灵活。

实验性研究：基于模型的推断方法

- 例如，该方法可以有效适应各种对因果效应估计的统计量 $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}, \mathbf{W})$，通过对缺失的潜在因果的估计，实验者可以推断中感兴趣的统计量分布。
- 考虑在没有纳入协变量的完全随机实验中基于贝叶斯模型的推理方法，该方法的主要目的是为缺失的潜在结果建立一个模型 $f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$，由此可以导出感兴趣的估计 $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$ 的分布，也可以用观察到的和缺失的潜在结果来表示估计该估计 $\tau = \tau\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$.

基于模型的推断方法：输入潜在结果分布

- 基于模型的推断方法的第一个输入是潜在结果 $(\mathbf{Y}(0), \mathbf{Y}(1))$ 的联合分布 $f(\mathbf{Y}(0), \mathbf{Y}(1) \mid \theta)$ 。由德 · 费内蒂定理 (de Finetti's theorem)，在矩阵 $(\mathbf{Y}(0), \mathbf{Y}(1))$ 的行可交换的条件下，可以不失一般性，将 $(\mathbf{Y}(0), \mathbf{Y}(1))$ 的联合分布建模为独立同分布的个体分布的乘积：

$$f(\mathbf{Y}(0), \mathbf{Y}(1)) = \int \prod_{i=1}^{N} f\left(Y_i(0), Y_i(1) \mid \theta\right) \cdot p(\theta) d\theta$$

其中 θ 是一个未知的有限维参数，$\theta \in \Theta$，且 $p(\theta)$ 是先验的边际分布。

基于模型的推断方法：输入先验参数

- 基于模型的推断方法的另一个输入则是参数 θ 的先验分布 $p(\theta)$.
- 这里常常将治疗组和对照组的均值考虑服从正态分布，而方差服从逆卡方分布（inverse chi-square distribution）。

Lalonde NSW 职业培训数据

- 实验介绍：随机实验，评估职业培训计划对收入的影响
- 实验对象：在劳动力市场上处于实质性劣势的男性，大多数人的劳动力市场历史非常糟糕，很少有长期就业的情况。
- 协变量：年龄（age）、受教育年限（education）、是否已婚（married）、他们是否高中辍学（nodegree）和种族（black）。两个衡量培训前收入的指标：第一个是 1975 年的收入（earn'75），第二个是培训前十三到二十四个月的收入（earn'74）。1975 年的零收入指标（earn'75=0），以及随机接受或不接受培训前 13 至 24 个月的零收入指标（earn'74=0）。
- 潜在结果：项目后劳动力市场指标，1978 年的收入（earn'78）。

Lalonde NSW 职业培训数据

Table 8.1. ***Summary Statistics: National Supported Work (NSW) Program Data***

Covariate	Mean	(S.D.)	Average Controls ($N_c = 260$)	Average Treated ($N_t = 185$)
age	25.37	(7.10)	25.05	25.82
education	10.20	(1.79)	10.09	10.35
married	0.17	(0.37)	0.15	0.19
nodegree	0.78	(0.41)	0.83	0.71
black	0.83	(0.37)	0.83	0.84
earn'74	2.10	(5.36)	2.11	2.10
earn'74=0	0.73	(0.44)	0.75	0.71
earn'75	1.38	(3.15)	1.27	1.53
earn'75=0	0.65	(0.48)	0.68	0.60
earn'78	5.30	(6.63)	4.56	6.35
earn'78=0	0.31	(0.46)	0.35	0.24

图 5: Lalonde NSW 职业培训数据的描述性估计

Lalonde NSW 职业培训数据

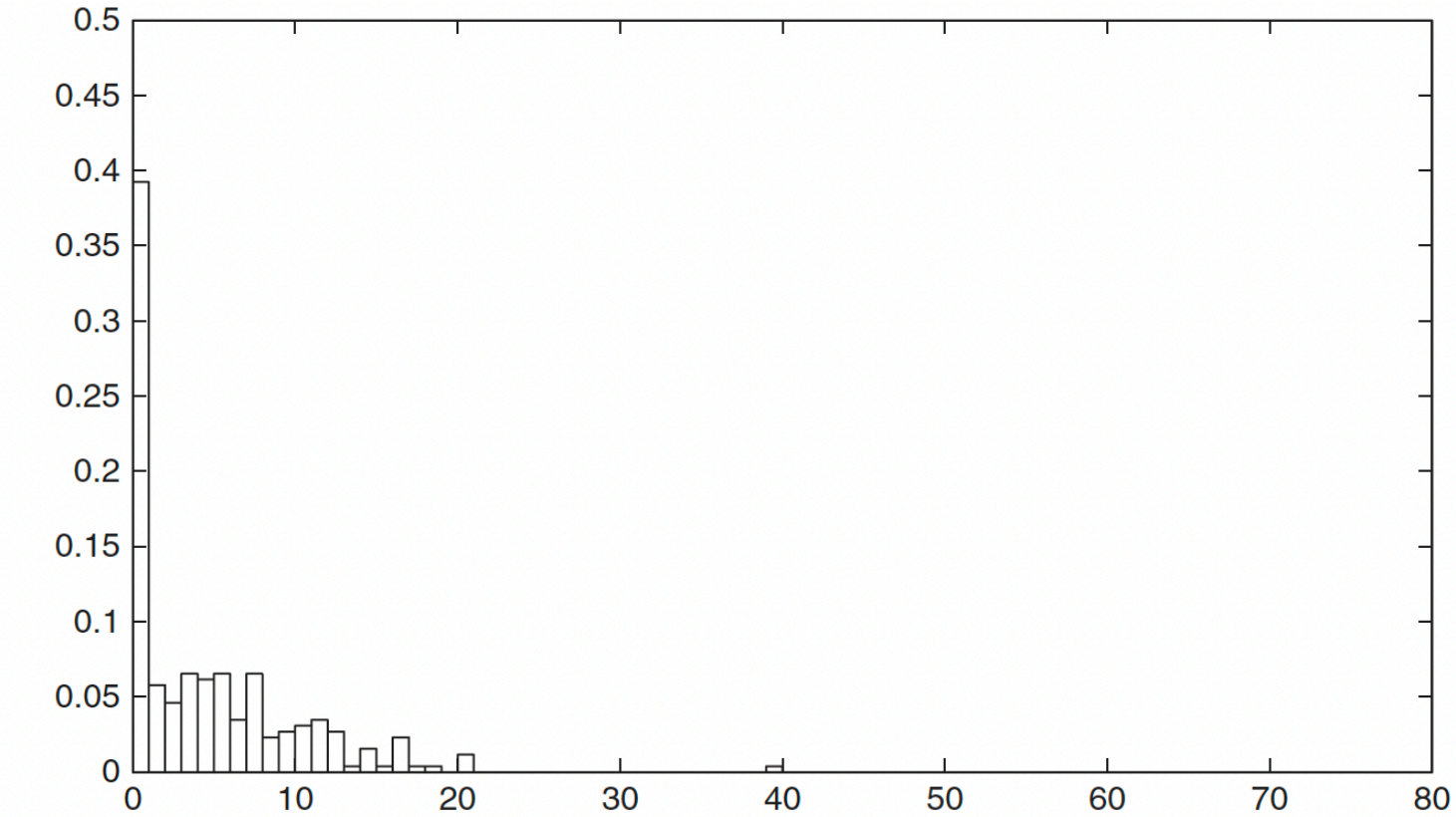

Figure 8.1. Histogram of earnings for control group – NSW job-training data

图 6: 对照组个体 1978 年收入直方图

Lalonde NSW 职业培训数据

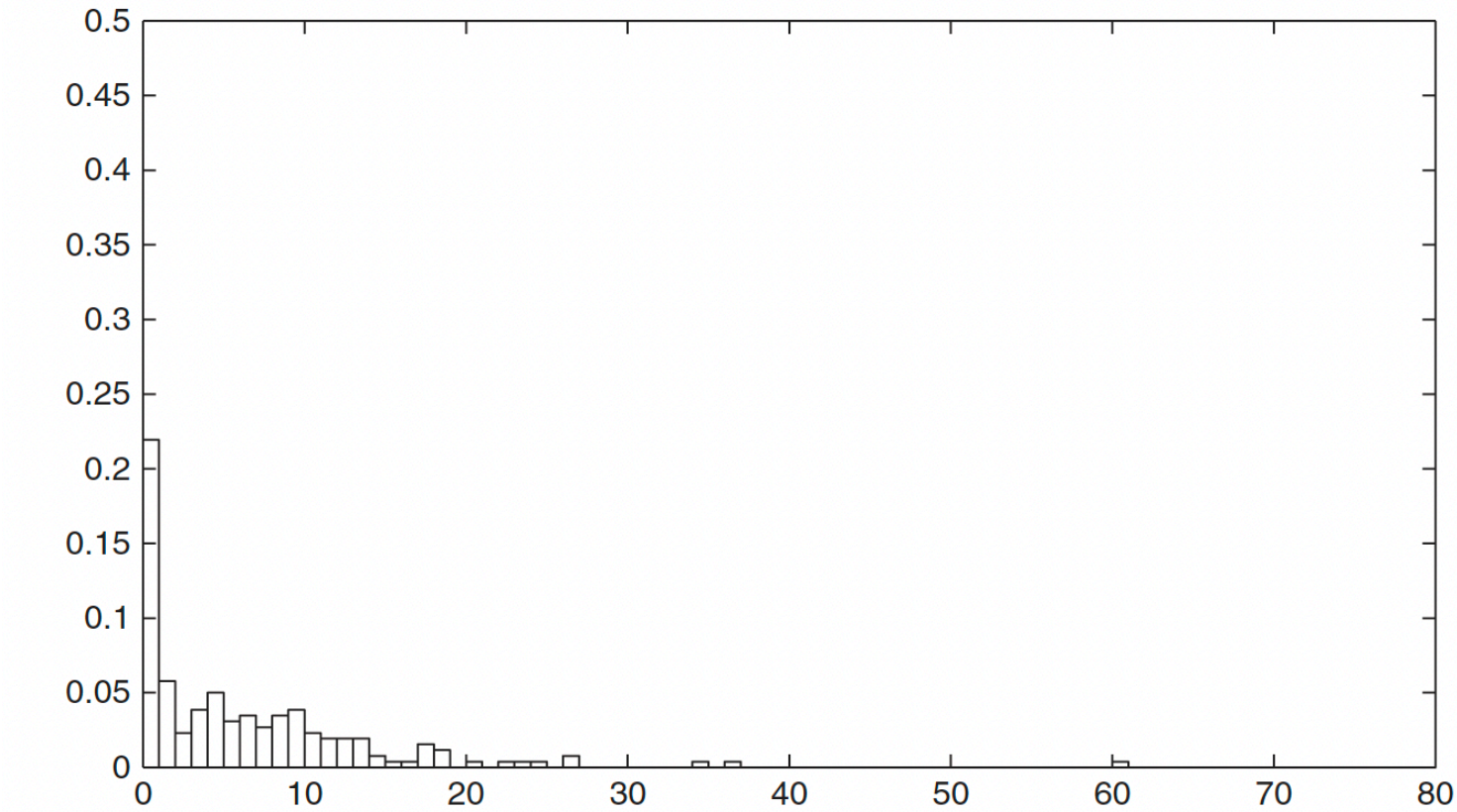

Figure 8.2. Histogram of earnings for trainee group – NSW job-training data

图 7: 治疗组个体 1978 年收入直方图

Lalonde NSW 职业培训数据：前六个样本分析

Table 8.2. ***First Six Observations from NSW Program Data***

Unit	Potential Outcomes		Treatment	Observed Outcome
	$Y_i(0)$	$Y_i(1)$	W_i	Y_i^{obs}
1	0	?	0	0
2	?	9.9	1	9.9
3	12.4	?	0	12.4
4	?	3.6	1	3.6
5	0	?	0	0
6	?	24.9	1	24.9

图 8: NSW 职业培训数据的前六个样本

Lalonde NSW 职业培训数据：前六个样本分析

- $$\tau_{\text{fs}} = \tau(\mathbf{Y}(0), \mathbf{Y}(1)) = \frac{1}{6} \cdot \sum_{i=1}^{6} (Y_i(1) - Y_i(0))$$

- $$Y_i(0) = \begin{cases} Y_i^{\text{mis}} & \text{if } W_i = 1, \\ Y_i^{\text{obs}} & \text{if } W_i = 0, \end{cases} \quad \text{and} \quad Y_i(1) = \begin{cases} Y_i^{\text{mis}} & \text{if } W_i = 0 \\ Y_i^{\text{obs}} & \text{if } W_i = 1 \end{cases}$$

- $$\begin{aligned} \tau_{\text{fs}} &= \tilde{\tau}\left(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}}, \mathbf{W}\right) \\ &= \frac{1}{6} \cdot \sum_{i}^{N} \left(\left(W_i \cdot Y_i^{\text{obs}} + (1 - W_i) \cdot Y_i^{\text{mis}}\right) - \left((1 - W_i) \cdot Y_i^{\text{obs}} + W_i \cdot Y_i^{\text{mis}}\right)\right) \\ &= \frac{1}{6} \cdot \sum_{i=1}^{N} \left((2 \cdot W_i - 1) \cdot \left(Y_i^{\text{obs}} - Y_i^{\text{mis}}\right)\right) \end{aligned}$$

- $$\hat{\tau} = \tilde{\tau}\left(\mathbf{Y}^{\text{obs}}, \hat{\mathbf{Y}}^{\text{mis}}, \mathbf{W}\right) = \frac{1}{6} \cdot \sum_{i=1}^{N} \left((2 \cdot W_i - 1) \cdot \left(Y_i^{\text{obs}} - \hat{Y}_i^{\text{mis}}\right)\right)$$

Lalonde NSW 职业培训数据：前六个样本分析：均值填补法

Table 8.3. ***The Average Treatment Effect Using Imputation of Average Observed Outcome Values within Treatment and Control Groups for the NSW Program Data***

Unit	Potential Outcomes		Treatment	Observed Outcome
	$Y_i(0)$	$Y_i(1)$	W_i	Y_i^{obs}
1	0	(12.8)	0	0
2	(4.13)	9.9	1	9.9
3	12.4	(12.8)	0	12.4
4	(4.13)	3.6	1	3.6
5	0	(12.8)	0	0
6	(4.13)	24.9	1	24.9
Average	4.13	12.8		
Diff (ATE):	8.67			

图 9: NSW 职业培训数据的前六个样本分析：均值填补法

Lalonde NSW 职业培训数据：前六个样本分析：抽样法

Table 8.4. ***The Average Treatment Effect Using Imputed Draws from the Empirical Distributions within Treatment and Control Groups for the First Six Units from the NSW Program Data***

Unit	Potential Outcomes		Treatment	Observed Outcome
	$Y_i(0)$	$Y_i(1)$	W_i	Y_i^{obs}
Panel A: First draw				
1	0	(3.6)	0	0
2	(12.4)	9.9	1	9.9
3	12.4	(9.9)	0	12.4
4	(12.4)	3.6	1	3.6
5	0	(9.9)	0	0
6	(0)	24.9	1	24.9
Average	6.2	10.3		
Diff (ATE):	4.1			
Panel B: Second draw				
1	0	(9.9)	0	0
2	(0)	9.9	1	9.9
3	12.4	(24.9)	0	12.4
4	(0)	3.6	1	3.6
5	0	(3.6)	0	0
6	(0)	24.9	1	24.9
Average	2.1	12.8		
Diff (ATE):	10.7			

基于模型的推断方法：Step 1：推导 $f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta\right)$

- 将给定潜在结果下分配机制的条件分布 $\Pr(\mathbf{W} \mid \mathbf{Y}(0), \mathbf{Y}(1))$ 和给定参数 θ 下潜在结果的条件分布 $f(\mathbf{Y}(0), \mathbf{Y}(1) \mid \theta)$ 结合，得到给定参数 θ 下 $(\mathbf{W}, \mathbf{Y}(0), \mathbf{Y}(1))$ 的联合分布：

$$f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} \mid \theta) = \Pr(\mathbf{W} \mid \mathbf{Y}(0), \mathbf{Y}(1), \theta) \cdot f(\mathbf{Y}(0), \mathbf{Y}(1) \mid \theta)$$

- 推导给定参数 θ 时 $f(\mathbf{Y}(0), \mathbf{Y}(1) \mid \mathbf{W}, \theta)$ 的条件分布：

$$f(\mathbf{Y}(0), \mathbf{Y}(1) \mid \mathbf{W}, \theta) = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} \mid \theta)}{\Pr(\mathbf{W} \mid \theta)} = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} \mid \theta)}{\int f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} \mid \theta) \mathrm{d}\mathbf{Y}(0) \mathrm{d}\mathbf{Y}(1)}$$

基于模型的推断方法：Step 1：推导 $f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta\right)$

- 完全随机实验的假设使得分配机制 $\mathbf{W}$ 与潜在结果 $(\mathbf{Y}(0), \mathbf{Y}(1))$ 无关，因此该条件分布与只给定参数 θ 下的条件分布等价：

$$f(\mathbf{Y}(0), \mathbf{Y}(1) \mid \mathbf{W}, \theta) = f(\mathbf{Y}(0), \mathbf{Y}(1) \mid \theta)$$

- 事实上，对于所有正规随机实验，上式均成立。然后考虑将给定分配机制 $\mathbf{W}$ 和参数 θ 下潜在结果 $(\mathbf{Y}(0), \mathbf{Y}(1))$ 的分布，转换为给定观测潜在结果 $\mathbf{Y}^{\text{obs}}$ 、分配机制 $\mathbf{W}$ 和参数 θ 下的缺失潜在结果 $\mathbf{Y}^{\text{mis}}$ 的分布，由于：

$$Y_i^{\text{obs}} = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1, \end{cases} \qquad Y_i^{\text{mis}} = \begin{cases} Y_i(0) & \text{if } W_i = 1 \\ Y_i(1) & \text{if } W_i = 0 \end{cases}$$

基于模型的推断方法：Step 1：推导 $f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta\right)$

- 从而 $\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}\right)$ 可以看成是 $(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$ 的变换，即 $\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}\right) = g(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$ 。使用该变换，可以得到给定 $\mathbf{W}$ 和 θ 下的 $\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}\right)$ 的联合分布：

$$f\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} \mid \mathbf{W}, \theta\right)$$

- 从而

$$f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta\right) = \frac{f\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} \mid \mathbf{W}, \theta\right)}{f\left(\mathbf{Y}^{\text{obs}} \mid \mathbf{W}, \theta\right)} = \frac{f\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} \mid \mathbf{W}, \theta\right)}{\int_{\mathbf{y}^{\text{mis}}} f\left(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} \mid \mathbf{W}, \theta\right) \mathrm{d}\mathbf{y}^{\text{mis}}}$$

由此得到了给定观测潜在结果 $\mathbf{Y}^{\text{obs}}$ 、分配机制 $\mathbf{W}$ 和参数 θ 下的缺失潜在结果 $\mathbf{Y}^{\text{mis}}$ 的后验分布。

基于模型的推断方法：Step 2：推导参数 θ 的后验分布 $p\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$

- 将参数 θ 的先验分布 $p(\theta)$ 与给定参数 θ 下观测结果的分布结合得到参数 θ 的后验分布 $p\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$。考虑第一步中推导的 $\theta, f\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} \mid \mathbf{W}, \theta\right)$ 关于缺失潜在结果的积分，可以得到给定参数 θ 下观测结果的边际分布，即似然函数：

$$\mathcal{L}\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right) \equiv f\left(\mathbf{Y}^{\text{obs}}, \mathbf{W} \mid \theta\right) = \int_{\mathbf{y}^{\text{mis}}} f\left(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W} \mid \theta\right) \mathrm{d}\mathbf{y}^{\text{mis}}$$

基于模型的推断方法：Step 2：推导参数 θ 的后验分布 $p\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$

- 将似然函数与参数的先验分布 $p(\theta)$ 结合可以得到参数的后验分布：

$$p\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right) = \frac{p(\theta) \cdot \mathcal{L}\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)}{f\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right)}$$

- 其中 $f\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$ 是通过对参数 θ 积分得到的 $(\mathbf{Y}, \mathbf{W})$ 的边际分布：

$$f\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right) = \int_{\theta} p(\theta) \cdot \mathcal{L}\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right) \mathrm{d}\theta$$

基于模型的推断方法：Step 3：推导缺失潜在结果的后验分布 $f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$

- 将给定 $\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta\right)$ 的条件下 $\mathbf{Y}^{\text{mis}}$ 的分布与参数 θ 的后验分布结合，可以得到给定 $\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$ 的条件下，$\left(\mathbf{Y}^{\text{mis}}, \theta\right)$ 的分布：

$$f\left(\mathbf{Y}^{\text{mis}}, \theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right) = f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta\right) \cdot p\left(\theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$$

- 接下来对 θ 可以得到给定 $\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$ 的条件下，$\mathbf{Y}^{\text{mis}}$ 的分布：

$$f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right) = \int_{\theta} f\left(\mathbf{Y}^{\text{mis}}, \theta \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right) \mathrm{d}\theta$$

从而得到给定观测潜在结果的条件下，缺失潜在结果的后验分布。

基于模型的推断方法：Step 4：推导估计量的后验分布 $f\left(\tau \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$

- 最后，使用给定观测潜在结果的条件下，缺失潜在结果的后验分布 $f\left(\mathbf{Y}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$ 以及观测数据 $\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$ 来得到给定观测结果的条件下，我们感兴趣的估计量的分布。
- 估计量的一般形式为 $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$，可以将 τ 重写为观测潜在结果、确实潜在结果和治疗指标的函数 $\tilde{\tau}\left(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$。
- 与给定 $\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$ 的条件下 $\mathbf{Y}^{\text{mis}}$ 的分布相结合，可以得到给定 $\left(\mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$ 的条件下 τ 的后验分布 $f\left(\tau \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$。

分层随机试验（Stratified Randomized Experiments）

- 分层随机试验（Stratified Randomized Experiments）是完全随机试验的直接推广，设层数为 $J, N(j), N_c(j)$ 和 $N_t(j)$ 分别是层 j 内的总样本数，对照组样本数和治疗组样本数。对 $j = 1, \ldots, J$，设 $G_i \in \{1, \ldots, J\}$ 是个体 i 所在的层，记 $B_i(j) = \mathbf{1}_{G_i=j}$ 是示性函数，若个体 i 在层 j 则取值为 1，否则为 0。
- 在层 j 内有 $\begin{pmatrix} N(j) \\ N_t(j) \end{pmatrix}$ 种可能的分配机制，因此分配机制为：

$$\Pr(\mathbf{W} \mid \mathbf{S}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{j=1}^{J} \begin{pmatrix} N(j) \\ N_t(j) \end{pmatrix}^{-1} \quad \text{for } \mathbf{W} \in \mathbb{W}^+$$

其中

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \mid \sum_{i=1}^{N} B_i(j) \cdot W_i = N_t(j) \text{ 对于} j = 1, \ldots, J \right\}$$

分层随机试验:Fisher 精确 P 值方法

- 总体来讲，在分层随机实验中，估计总体因果效应的思想是先估计每个层 j 内的因果效应，然后进行加权得到总体的因果效应估计:

$$T^{\text{dif},\lambda} = \left| \sum_{j=1}^{J} \lambda(j) \cdot \left(\bar{Y}_{\text{t}}^{\text{obs}}(j) - \bar{Y}_{\text{c}}^{\text{obs}}(j) \right) \right|$$

- 第一个自然的想法是使用每个层内的人数进行加权:

$$\lambda(j) = \frac{N(j)}{N}, \text{ 从而} T^{\text{dif},\lambda_{\text{RSS}}} = \left| \sum_{j=1}^{J} \frac{N(j)}{N} \cdot \left(\bar{Y}_{\text{t}}^{\text{obs}}(j) - \bar{Y}_{\text{c}}^{\text{obs}}(j) \right) \right|$$

分层随机试验:Fisher 精确 P 值方法

- 第二个权重的选择依赖于每个层 j 内，在治疗组与对照组遵从同方差假设时，可以先估计出每个层 j 的治疗效果的方差，然后使用方差的倒数进行加权:

$$\lambda^{\text{opt}}(j) = \frac{N(j) \cdot \frac{N_{\text{t}}(j)}{N(j)} \cdot \frac{N_{\text{c}}(j)}{N(j)}}{\sum_{k=1}^{J} N(k) \cdot \frac{N_{\text{t}}(k)}{N(k)} \cdot \frac{N_{\text{c}}(k)}{N(k)}}$$

- 从而

$$T^{\text{dif},\lambda_{\text{opt}}} = \left| \frac{1}{\sum_{j=1}^{J} N(j) \cdot \frac{N_{\text{t}}(j)}{N(j)} \cdot \frac{N_{\text{c}}(j)}{N(j)}} \sum_{j=1}^{J} N(j) \cdot \frac{N_{\text{t}}(j)}{N(j)} \cdot \frac{N_{\text{c}}(j)}{N(j)} \cdot \left(\bar{Y}_{\text{t}}^{\text{obs}}(j) - \bar{Y}_{\text{c}}^{\text{obs}}(j) \right) \right|$$

分层随机试验:Fisher 精确 P 值方法

- 该权重的选择通常会导致一个更鲁棒的估计，从一定程度上解决了由于层内接受治疗和对照的样本不平衡导致的估计不稳健。
- 例如两个层的人数相同，其中一个层的倾向得到接近 0.5，而另一个层的倾向得分接近 1：若使用每个层内的人数进行加权，则将会赋予这两个层相等的权重，但是显然倾向得分接近 1 的层将导致层内效果估计的不稳健。

分层随机试验:Fisher 精确 P 值方法

- 此外，还可以使用个体 i 在其所在层内的秩排序 R_i^{strat} 进行加权:

$$R_i^{\text{strat}} = \sum_{i':G_{i'}=G_i} \mathbf{1}_{Y_{i'}^{\text{obs}} < Y_i^{\text{obs}}} + \frac{1}{2}\left(1 + \sum_{i':G_{i'}=G_i} \mathbf{1}_{Y_{i'}^{\text{obs}} = Y_i^{\text{obs}}}\right) - \frac{N(G_i)+1}{2}$$

- 对于接受治疗和对照的个体，分别计算该秩排序的差异，得到统计量:

$$T^{\text{rank, stratum}} = \left|\bar{R}_t^{\text{strat}} - \bar{R}_c^{\text{strat}}\right|$$

其中，$\bar{R}_t^{\text{strat}}$ 和 $\bar{R}_c^{\text{strat}}$ 分别是治疗组和对照组的层内标准化秩排序的平均值。

分层随机试验:Neyman 重复抽样法

- 由于不同层内个体的治疗分配是独立的，因此奈曼重复抽样方法在分层随机实验中的推广是自然的。考虑有两个层，分别是男性（male）和女性（female）。对于女性所在的层，一个自然的对于层内治疗效果 $\tau_{\text{fs}}(f)$ 的无偏估计量是：

$$\hat{\tau}^{\text{dif}}(f) = \bar{Y}_{\text{t}}^{\text{obs}}(f) - \bar{Y}_{\text{c}}^{\text{obs}}(f) = \frac{1}{N_{\text{t}}(f)} \sum_{i:G_i=f} W_i \cdot Y_i^{\text{obs}} - \frac{1}{N_{\text{c}}(f)} \sum_{i:G_i=f} (1 - W_i) \cdot Y_i^{\text{obs}}$$

- 在分配机制的随机化意义下，该统计量的方差为：

$$\mathbb{V}_W\left(\hat{\tau}^{\text{dif}}(f)\right) = \frac{S_c^2(f)}{N_{\text{c}}(f)} + \frac{S_t^2(f)}{N_{\text{t}}(f)} - \frac{S_{ct}^2(f)}{N(f)}$$

对于男性所在的层，有类似的结论。

分层随机试验:Neyman 重复抽样法

- 接下来考虑总体的平均治疗效果:

$$\tau_{\text{fs}} = \frac{N(f)}{N(f)+N(m)} \cdot \tau_{\text{fs}}(f) + \frac{N(m)}{N(f)+N(m)} \cdot \tau_{\text{fs}}(m) = \frac{1}{N} \sum_{i=1}^{N} (Y_i(1) - Y_i(0))$$

- 由于 $\hat{\tau}^{\text{dif}}(f)$ 和 $\hat{\tau}^{\text{dif}}(m)$ 的无偏性，我们得到了对于总体平均治疗效果 τ_{fs} 的一个无偏估计量:

$$\hat{\tau}^{\text{strat}} = \frac{N(f)}{N(f)+N(m)} \cdot \hat{\tau}^{\text{dif}}(f) + \frac{N(m)}{N(f)+N(m)} \cdot \hat{\tau}^{\text{dif}}(m)$$

分层随机试验:Neyman 重复抽样法

- 对于该统计量，由于层之间分配机制的独立性，因此:

$$\begin{aligned}\mathbb{V}_W\left(\hat{\tau}^{\text{strat}}\right) &= \left(\frac{N(f)}{N(f)+N(m)}\right)^2 \cdot \mathbb{V}_W\left(\hat{\tau}_f\right) + \left(\frac{N(m)}{N_f+N(m)}\right)^2 \cdot \mathbb{V}_W\left(\hat{\tau}_m\right) \\ &= \left(\frac{N(f)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{S_c(f)^2}{N_{\text{c}}(f)} + \frac{S_t^2(f)}{N_{\text{t}}(f)} - \frac{S_{ct}^2(f)}{N(f)}\right) \\ &\quad + \left(\frac{N(m)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{S_c(m)^2}{N_{\text{c}}(m)} + \frac{S_t^2(m)}{N_{\text{t}}(m)} - \frac{S_{ct}^2(m)}{N(m)}\right)\end{aligned}$$

分层随机试验:Neyman 重复抽样法

- 如前所述，我们没有直接的方法来估计个体潜在结果差异的方差。然而可以通过忽略这一项来得到总体平均治疗效果的估计量的采样方差的一个保守估计。
- 通过估计两个层内的采样方差来获得采样方差的估计上界:

$$\hat{\mathbb{V}}^{\text{neyman}} = \left(\frac{N(f)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{s_c^2(f)}{N_{\text{c}}(f)} + \frac{s_t^2(f)}{N_{\text{t}}(f)}\right) + \left(\frac{N(m)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{s_c^2(m)}{N_{\text{c}}(m)} + \frac{s_t^2(m)}{N_{\text{t}}(m)}\right)$$

分层随机试验: 线性回归

- 在超总体的视角，我们考虑分层随机实验的回归方法。假设层数为 J，每个层内均有无限数量的个体。在分层随机实验的回归方法中，自变量除了处理指标外，还包括层指标:

$$Y_i^{\text{obs}} = \tau \cdot W_i + \sum_{j=1}^{J} \beta(j) \cdot B_i(j) + \varepsilon_i$$

- 考虑的 τ 最小二乘估计:

$$\left(\hat{\tau}^{\text{ols}}, \hat{\beta}^{\text{ols}}\right) = \arg\min_{\tau,\beta} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \tau \cdot W_i + \sum_{j=1}^{J} \beta(j) \cdot B_i(j) \right)^2$$

- 类似地，在超总体观点的期望意义下:

$$\left(\tau^*, \beta^*\right) = \arg\min_{\tau,\beta} \mathbb{E}\left[\left(Y_i^{\text{obs}} - \tau \cdot W_i + \sum_{j=1}^{J} \beta(j) \cdot B_i(j) \right)^2 \right]$$

分层随机试验: 线性回归

- 在分层随机实验的回归分析中，最小二乘估计 $\hat{\tau}^{\text{ols}}$ 不一定是超总体平均治疗效果 τ_{sp} 的相合估计。
- 然而，分层随机实验回归的最小二乘估计 $\hat{\tau}^{\text{ols}}$ 是层内平均效应 $\tau_{\text{sp}}(j)$ 的加权平均值的相合估计，其中每个层的权重记为 $\omega(j)$：

$$\omega(j) = q(j)\cdot e(j)\cdot(1-e(j)), \quad \text{以及} \quad \tau_\omega = \sum_{j=1}^{J} \omega(j)\cdot\tau_{\text{sp}}(j) \Big/ \left(\sum_{j=1}^{J} \omega(j)\right)$$

- 其中 $q(j) = N(j)/N, e(j) = N_{\text{t}}(j)/N(j)$ 以及 $\tau_{\text{sp}}(j) = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid B_i(j) = 1\right]$。此时 $\hat{\tau}^{\text{ols}}$ 是 τ_ω 的相合估计。

分层随机试验: 线性回归

- 考虑在超总体中进行有限抽样，在该有限样本集上进行的分层随机试验，则有 $\tau^* = \tau_\omega$ 且

$$\sqrt{N} \cdot \left(\hat{\tau}^{\text{ols}} - \tau_\omega \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}\left[\left(W_i - \sum_{j=1}^{J} q(j) \cdot B_i(j) \right)^2 \cdot \left(Y_i^{\text{obs}} - \tau^* \cdot W_i - \sum_{j=1}^{J} \beta_j^* \cdot B_i(j) \right)^2 \right]}{\left(\sum_{j=1}^{J} q(j) \cdot e(j) \cdot (1 - e(j)) \right)^2} \right)$$

分层随机试验: 线性回归

- $Y_i^{\text{obs}} = \tau \cdot W_i \cdot \frac{B_i(j)}{N(j)/N} + \sum_{j=1}^{J} \beta(j) \cdot B_i(j) + \sum_{j=1}^{J-1} \gamma(j) \cdot W_i \cdot \left(B_i(j) - B_i(J) \cdot \frac{N(j)}{N(J)} \right) + \varepsilon_i$
- $\tau^* = \tau_{\text{sp}}$
- $\sqrt{N} \cdot \left(\hat{\tau}^{\text{ols,inter}} - \tau_{\text{sp}} \right) \xrightarrow{d} \mathcal{N}\left(0, \sum_{j=1}^{J} q(j)^2 \cdot \left(\frac{\sigma_c^2(j)}{(1-e(j)) \cdot q(j)} + \frac{\sigma_t^2(j)}{e(j) \cdot q(j)} \right) \right)$
- 一般而言，$\hat{\tau}^{\text{ols,inter}}$ 的样本方差大于 $\hat{\tau}^{\text{ols}}$.

分层随机试验: 基于模型的推断方法

- 在分层随机实验中，也可以考虑基于模型的推断方法。分为两种情形考虑。
- 若层数很少，而每个层有很多个体: 则可以将参数先验地设置为独立。
- 不失一般性，层 j 的潜在结果的条件联合分布为:

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \mid B_i(j), \theta \sim \mathcal{N}\left(\begin{pmatrix} \mu_c(j) \\ \mu_t(j) \end{pmatrix}, \begin{pmatrix} \sigma_c^2(j) & 0 \\ 0 & \sigma_t^2(j) \end{pmatrix} \right)$$

- 其中 $(\mu_c(j), \mu_t(j))$ 和 $\left(\sigma_c^2(j), \sigma_t^2(j)\right)$ 分别是层 j 的潜在结果服从的正态分布的期望和方差。
- 此时参数向量是
  $\theta = \left(\mu_c(j), \mu_t(j), \sigma_c^2(j), \sigma_t^2(j), w = 0, 1, j = 1, \ldots, J\right)$ 。
- 一般假设 $\mu_c(j)$ 和 $\mu_t(j)$ 的先验服从正态分布，而 $\sigma_c^2(j)$ 和 $\sigma_t^2(j)$ 的先验服从逆卡方分布。

分层随机试验: 基于模型的推断方法

- 若层数很多，而每个层个体数量较少时：则可以通过施加限制条件 (如等式约束)，将所有层 j 的潜在结果的方差 σ_{c}^2 和 σ_t^2 限制为跨层相同：

$$\begin{pmatrix} \mu_{\mathrm{c}}(1) \\ \mu_{\mathrm{c}}(2) \\ \vdots \\ \mu_{\mathrm{c}}(J) \\ \mu_{\mathrm{t}}(1) \\ \mu_{\mathrm{t}}(2) \\ \vdots \\ \mu_{\mathrm{t}}(J) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \gamma_{\mathrm{c}} \\ \gamma_{\mathrm{c}} \\ \vdots \\ \gamma_{\mathrm{c}} \\ \gamma_{\mathrm{t}} \\ \gamma_{\mathrm{t}} \\ \vdots \\ \gamma \end{pmatrix}, \begin{pmatrix} \eta_{\mathrm{c}}^2 & 0 & \cdots & 0 & \rho\sigma_c\sigma_t & 0 & \cdots & 0 \\ 0 & \eta_{\mathrm{c}}^2 & & \vdots & 0 & \rho\sigma_c\sigma_t & & \vdots \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & \cdots & & \eta_{\mathrm{c}}^2 & 0 & \cdots & & \rho\sigma_c\sigma_t \\ \rho\sigma_c\sigma_t & 0 & \cdots & 0 & \eta_{\mathrm{t}}^2 & 0 & \cdots & 0 \\ 0 & \rho\sigma_c\sigma_t & & \vdots & 0 & \eta_{\mathrm{t}}^2 & & \vdots \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & \cdots & & \rho\sigma_c\sigma_t & 0 & \cdots & & \eta_{\mathrm{t}}^2 \end{pmatrix} \right)$$

- 此时，参数向量为 $\theta = \left(\sigma_c^2, \sigma_t^2, \gamma_c, \gamma_t, \eta_c^2, \eta_t^2\right)$ 。

配对随机实验（Pairwise Randomized Experiments）

- 配对随机实验（Pairwise Randomized Experiments）是分层随机实验的一个特殊情况。其中每个层正好包含两个个体，随机选择一个分配给治疗组，另一个则分配给对照组。
- 然而，每一个阶层中治疗组和对照组均只有一个个体的事实意味着我们无法使用奈曼抽样方差的估计量，由于奈曼对平均因果效应的方差估计量要求存在至少两个单位分配给每个层内的治疗组与对照组。
- 此外，配对随机实验中每个层具有相同比例的处理单元，这允许我们对称地分析层内估计，即平均治疗效果的自然估计量将对每一层以等权重进行加权。

配对随机实验（Pairwise Randomized Experiments）

- 假设群体中有 N 个个体，则配对随机实验的层数为 $J = N/2$，每层中均有一个个体被分配接受治疗，另一个个体接受对照，即 $N_t(j) = N_c(j) = 1, N(j) = 2$ 对于所有 $j = 1, \ldots, J$。设 G_i 个体 i 所属层的示性函数，则 $G_i \in \{1, \ldots, N/2\}$ 。由于该示性函数不受治疗指标影响，故可认为是治疗前指标（即协变量）。
- 配对随机实验的每一层中，分配机制均有 $\begin{pmatrix} N(j) \\ N_t(j) \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2$ 种可能的情形，因此总体的分配机制 $\mathbf{W}$ 是：

$$p(\mathbf{W} \mid \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{j=1}^{N/2} \begin{pmatrix} N(j) \\ N_t(j) \end{pmatrix}^{-1} = \prod_{j=1}^{N/2} \frac{1}{2} = 2^{-N/2}, \text{ 对于} \mathbf{W} \in \mathbb{W}^+$$

其中

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

配对随机实验（Pairwise Randomized Experiments）

- 在每一层中，随机将一个个体标记为 A，另一个体标记为 B 。对于所有层 $j = 1, \ldots, N/2$，设 $\left(Y_{j,A}(0), Y_{j,A}(1)\right)$ 和 $\left(Y_{j,B}(0), Y_{j,B}(1)\right)$ 分别为个体 A 和个体 B 的潜在结果。在层 j 内，设 $W_{j,A}$ 和 $W_{j,B}$ 分别为个体 A 和个体 B 的治疗指标，显然有 $W_{j,A} = 1 - W_{j,B}$，且 $\Pr\left(W_{j,A} = 1 \mid \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\right) = 1/2$.
- $$Y_{j,A}^{\text{obs}} = \begin{cases} Y_{j,A}(0) & \text{if } W_{j,A} = 0, \\ Y_{j,A}(1) & \text{if } W_{j,A} = 1, \end{cases} \quad \text{以及} \quad Y_{j,B}^{\text{obs}} = \begin{cases} Y_{j,B}(0) & \text{if } W_{j,A} = 1 \\ Y_{j,B}(1) & \text{if } W_{ji,A} = 0 \end{cases}$$
- 则层 j 内的平均治疗效果 $\tau_{\text{pair}}(j)$ 为：

$$\tau_{\text{pair}}(j) = \frac{1}{2} \sum_{i:G_i=j} \left(Y_i(1) - Y_i(0)\right) = \frac{1}{2} \left(\left(Y_{j,A}(1) - Y_{j,A}(0)\right) + \left(Y_{j,B}(1) - Y_{j,B}(0)\right)\right)$$

配对随机实验（Pairwise Randomized Experiments）

- 总体的平均治疗效果为：

$$\tau_{\text{fs}} = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) = \frac{2}{N}\sum_{j=1}^{N/2}\tau_{\text{pair}}(j)$$

- 此外，记层 j 内的观测结果和缺失结果分别为：

$$Y_{j,c}^{\text{obs}} = \begin{cases} Y_{j,A}^{\text{obs}} & \text{if } W_{i,A} = 0, \\ Y_{j,B}^{\text{obs}} & \text{if } W_{i,A} = 1, \end{cases} \quad \text{以及} \quad Y_{j,t}^{\text{obs}} = \begin{cases} Y_{j,B}^{\text{obs}} & \text{if } W_{i,A} = 0 \\ Y_{j,A}^{\text{obs}} & \text{if } W_{i,A} = 1 \end{cases}$$

配对随机实验：Fisher 精确 P 值方法

- 在配对随机实验中，通过费希尔的尖锐零假设，可以评估因果效应的显著性，其中零假设仍然为个体无治疗效果：

$$H_0 : Y_i(0) = Y_i(1), \text{ 对于所有} i = 1, \ldots, N$$

- 在已知分配机制时，假设零假设 H_0 成立时，对于任何固定的统计量，可以导出其随机化分布，从而计算对应的 P 值（FEP）。对于统计量，一个自然的选择是每对受试者治疗结果和对照结果之间的平均差异：

$$\begin{aligned} T^{\text{dif}} &= \left| \frac{1}{J} \sum_{j=1}^{J} \left(Y_{j,t}^{\text{obs}} - Y_{j,c}^{\text{obs}} \right) \right| \\ &= \left| \frac{1}{J} \sum_{j=1}^{J} \left(W_{j,A} \cdot \left(Y_{j,A}^{\text{obs}} - Y_{j,B}^{\text{obs}} \right) + \left(1 - W_{j,A} \right) \cdot \left(Y_{j,B}^{\text{obs}} - Y_{j,A}^{\text{obs}} \right) \right) \right| \end{aligned}$$

配对随机实验：Neyman 重复抽样法

- 接下来考虑奈曼重复抽样方法的统计量:

$$\hat{V}^{\text{neyman}}\left(\bar{Y}_{\text{t}}^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}}\right) = \frac{s_c^2}{N_{\text{c}}} + \frac{s_t^2}{N_{\text{t}}}$$

- $$s_c^2 = \frac{1}{N_{\text{c}} - 1} \sum_{i:W_i=0} \left(Y_i(0) - \bar{Y}_{\text{c}}^{\text{obs}}\right)^2 = \frac{1}{N_{\text{c}} - 1} \sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}}\right)^2$$

- $$s_t^2 = \frac{1}{N_{\text{t}} - 1} \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \bar{Y}_{\text{t}}^{\text{obs}}\right)^2$$

配对随机实验：Neyman 重复抽样法

- 与费舍尔的精确 P 值方法不同，由于每层只有一个个体接受治疗和对照，因此无法直接得到 s_t^2 和 s_c^2 的估计，从而奈曼的重复抽样方法无法直接应用于分层随机试验。
- 为了解决样本方差无法估计的问题，可以假设层内及层间的治疗效果均为相同的常数且具有可加性，从而层内的样本方差为：

$$\mathbb{V}_W\left(\hat{\tau}^{\text{pair}}(j)\right) = 2 \cdot S^2(j), \quad \text{其中} S^2(j) = S_c^2(j) = S_t^2(j)$$

配对随机实验：Neyman 重复抽样法

- 因此总体的样本方差为：

$$\mathbb{V}_W\left(\hat{\tau}^{\text{dif}}\right) = \frac{1}{(N/2)^2}\sum_{j=1}^{N/2}\left(S_c^2(j) + S_t^2(j) - \frac{S_{ct}^2(j)}{2}\right) = \frac{4}{N}\cdot S^2$$

- 对该统计量可以通过计算对水平治疗效果估计的样本方差来估计：

$$\hat{V}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right) = \frac{4}{N\cdot(N-2)}\cdot\sum_{j=1}^{N/2}\left(\hat{\tau}^{\text{pair}}(j) - \hat{\tau}^{\text{dif}}\right)^2$$

- 值得注意的是，如果治疗效果存在异质性，那么该样本方差估计量是向上偏置的，相应的置信区间在通常的统计学意义上是保守的。

配对随机实验：Neyman 重复抽样法

- 假设有 J 对个体，从每对中随机分配一个个体接受治疗，另一个个体接受对照，则 $\hat{\tau}^{\text{dif}}$ 是 τ_{fs} 的无偏估计，且 $\hat{\tau}^{\text{dif}}$ 的真实样本方差为：

$$\mathbb{V}_W\left(\hat{\tau}^{\text{dif}}\right) = \frac{1}{N^2}\sum_{j=1}^{N/2}\left(Y_{j,A}(0) + Y_{j,A}(1) - \left(Y_{j,B}(0) + Y_{j,B}(1)\right)\right)^2$$

- 该样本方差的估计量为：

$$\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right) = \frac{4}{N\cdot(N-2)}\cdot\sum_{j=1}^{N/2}\left(\hat{\tau}^{\text{pair}}(j) - \hat{\tau}^{\text{dif}}\right)^2$$

$$\mathbb{E}\left[\hat{\mathbb{V}}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right)\right] = \mathbb{V}_W\left(\hat{\tau}^{\text{dif}}\right) + \frac{4}{N\cdot(N-2)}\cdot\sum_{j=1}^{N/2}\left(\tau_{\text{pair}}(j) - \tau\right)^2$$

- 可以看出，若层内及层间的治疗效果均为相同的常数且具有可加性，则 $\hat{V}^{\text{pair}}\left(\hat{\tau}^{\text{dif}}\right)$ 是 $\mathbb{V}_W\left(\hat{\tau}^{\text{dif}}\right)$ 的无偏估计量。

因果推断在观察性研究中的非混淆性

- 在观察性研究中，治疗分配的概率是一个未知的函数。然而，实验者常常保持分配机制的非混淆性假设，即在给定协变量的条件下，分配不依赖于潜在结果。此外，我们假设分配机制是个体化的和概率的。
- 对于非混淆性假设，一个很好的理解是协变量包含了充分的信息，即在两个个体具有同样的充分信息 (协变量) 的条件下，可以将总体的分配机制视为分层随机实验。
- 本质上，基于非混淆性假设，实验者可以利用协变量的匹配方法来推断个体的缺失潜在结果。
- 然而，由于协变量通常在样本中呈现出许多不同的值，且协变量在许多研究中是高维的，因此可能会有相当多的个体无法精确匹配协变量。
- 在这种情况下，我们可以比较治疗前变量值“相似”但不完全相同的治疗组和对照组的结果，该比较的合理性基于平滑度和建模假设，以及关于一个协变量与另一个协变量之间差异权衡的决策。

因果推断在观察性研究中的非混淆性

- 在超总体的视角中，同样可以考虑非混淆性的假设。
- 若超总体中的分配机制具有非混淆性，则：

$$(Y_i(0) \mid W_i = 1, X_i) \sim (Y_i(0) \mid W_i = 0, X_i), \quad \text{for } i = 1, \ldots, N$$

且

$$(Y_i(1) \mid W_i = 0, X_i) \sim (Y_i(1) \mid W_i = 1, X_i), \quad \text{for } i = 1, \ldots, N$$

其中表示服从相同的分布。

- 在超总体中的非混淆性假设使得我们可以通过调整观察到的协变量的差异来消除治疗组和对照组之间比较的所有偏差。
- 虽然在理论上可行，但在实践中，高维的协变量将很难实现精确的匹配。

因果推断在观察性研究中的非混淆性

- 与充分统计量的想法类似，引入平衡分数从而找到低维的协变量函数，且平衡分数的相等足以消除与治疗前变量差异相关的偏差。
- 因此，在给定平衡分数的情况下，接受积极治疗的概率就不依赖于协变量。
- 称协变量的函数 $b(x)$ 为均衡得分，若

$$W_i \perp\!\!\!\perp X_i \mid b(X_i)$$

- 显然，均衡得分不唯一：首先协变量 X_i 本身就是一个均衡得分，此外任何一个均衡得分的一对一函数也是一个均衡得分。
- 这里我们最感兴趣的是低维均衡得分。其中，一个标量均衡得分是倾向性得分 $e(X_i)$。

因果推断在观察性研究中的非混淆性

- 倾向性得分是均衡得分，即 $W_i \perp\!\!\!\perp X_i \mid e(X_i)$ 或 $\Pr(W_i = 1 \mid X_i, e(X_i)) = \Pr(W_i = 1 \mid e(X_i))$。
- 证明：

$$\Pr(W_i = 1 \mid X_i, e(X_i)) = \Pr(W_i = 1 \mid X_i) = e(X_i)$$

其中第一个等式成立，由于倾向性得分是协变量 X_i 的函数，第二个等式由倾向性得分的定义可证。由重期望公式：

$$\Pr(W_i = 1 \mid e(X_i)) = \mathbb{E}[W_i \mid e(X_i)] = \mathbb{E}[\mathbb{E}[W_i \mid X_i, e(X_i)] \mid e(X_i)]$$

$$= \mathrm{E}[e(X_i) \mid e(X_i)] = e(X_i)$$

因果推断在观察性研究中的非混淆性

- 在给定个体的均衡得分 $b(X_i)$ 的条件下，个体的治疗指标 W_i 与潜在结果独立。
- 若分配机制具有非混淆性，则在给定倾向性得分时，分配机制满足：

$$W_i \perp\!\!\!\perp Y_i(0), Y_i(1) \mid b(X_i).$$

- 证明：我们证明

$$\Pr_W(W_i = 1 \mid Y_i(0), Y_i(1), b(X_i)) = \Pr_W(W_i = 1 \mid b(X_i))$$

首先，由重期望公式：

$$\begin{aligned}\Pr_W(W_i = 1 \mid Y_i(0), Y_i(1), b(X_i)) &= \mathbb{E}_W[W_i \mid Y_i(0), Y_i(1), b(X_i)] \\ &= \mathbb{E}[\mathbb{E}_W[W_i \mid Y_i(0), Y_i(1), X_i, b(X_i)] \mid Y_i(0), Y_i(1), b(X_i)]\end{aligned}$$

由非混淆性，内层的期望等价于 $\mathbb{E}[W_i \mid X_i, b(X_i)]$，由均衡得分的定义，上述表达等价于 $\mathbb{E}[W_i \mid b(X_i)]$ 。因此

$$\mathbb{E}[\mathbb{E}_W[W_i \mid b(X_i)] \mid Y_i(0), Y_i(1), b(X_i)] = \mathbb{E}[W_i \mid b(X_i)] = \Pr(W_i = 1 \mid b(X_i))$$

因果推断在观察性研究中的非混淆性

- 倾向性得分是最粗的均衡得分，即倾向性得分是每个均衡得分的函数。
- 证明：若倾向性得分不行写成均衡得分的函数，则存在不同的 x 和 x'，使得 $b(x) = b\left(x'\right)$，同时 $e(x) \neq e\left(x'\right)$ 。
- $\Pr\left(W_i = 1 \mid X_i = x\right) = e(x) \neq e\left(x'\right) = \Pr\left(W_i = 1 \mid X_i = x'\right)$.
- 因此 W_i 和 X_i 在给定 $b\left(X_i\right) = b(x)$ 的条件下不独立，而这与均衡得分的定义相矛盾。

倾向性得分估计

- 在实证研究中，协变量的数量通常相对于个体的数量很大。因此，将所有协变量都包含在倾向性得分的估计模型中并不是一个好的选择。
- 此外，对于一些影响作用较大的协变量，仅仅包含它们的线性项可能会导致模型欠拟合，这里我们利用逐步回归的思想来选择协变量和二次项（包括平方项和交互项），包含在倾向性得分的预测模型中。
- 考虑使用逻辑回归模型来估计倾向性得分，其中接受治疗的对数几率（log odds ratio）是协变量和二次项（包括平方项和交互项）的线性函数，其中系数未知，利用最大似然估计来估计系数。利用逐步回归的思想，我们接下来讨论协变量的选择。

倾向性得分估计 Step 1：基础协变量的选择

- 在第一步中，我们在实验的先验性决策下选择 K_{B} 个基础协变量。
- 记个体的协变量为 X_i，维数为 K 。基础协变量的选择包括两类协变量，均依赖于先验性信息。
- 一类是对分配机制有解释作用的协变量，另一类是对潜在结果高度相关的协变量。
- 如果研究人员对协变量无先验性信息，可以设置 $K_{\mathrm{B}} = 0$ 。

倾向性得分估计 Step 2：增加协变量的线性组合项

- 在第二步中，我们选择一些其余的协变量，逐步引入倾向性得分的估计模型，现有 $K - K_{\mathrm{B}}$ 个协变量未被引入模型。
- 每次考虑将一个剩余协变量迭代地纳入逻辑回归模型，设某时刻已经选择了 $\tilde{K}_{\mathrm{L}}$ 个协变量，作为回归模型中的线性组合项，其中包括维数为 K_B 的基础协变量。
- 对每个剩余的 $K - \tilde{K}_{\mathrm{L}}$ 个协变量，分别将其纳入已有的逻辑回归模型，并计算似然比统计量（likelihood ratio statistic）来评估新纳入的协变量在回归模型中系数为 0 的零假设。
- 若所有协变量被分别纳入原回归模型中，似然比统计量均低于一个预先设置的临界值 C_{L}，则不纳入剩余的协变量，此时模型中只包括原模型的 $\tilde{K}_{\mathrm{L}}$ 个协变量的线性组合。

倾向性得分估计 Step 2：增加协变量的线性组合项

- 若至少有一个协变量被纳入后的似然比统计量超过了预先设置的临界值 C_{L}，则将似然比统计量最大的一个协变量纳入模型中。
- 目前我们有 $\tilde{K}_{\mathrm{L}}+1$ 个协变量被纳入模型，接下来检验剩余的 $K-\tilde{K}_{\mathrm{L}}-1$ 个协变量的似然比统计量，来判断其是否应被纳入现有模型。
- 我们迭代上述过程，直到剩余协变量的似然比统计量均小于 C_{L}，记迭代最终模型的协变量维数为 K_{L} 。

倾向性得分估计 Step 3：引入二次项和交互项

- 在第三步中，我们考虑哪些二次项和交互作用应包含在倾向性得分的模型中。
- 假设在第二$\boxed{\text{F}}$之后，我们已经将 $K_{\text{L}} \leq K$ 个协变量的线性组合纳入了回归模型，现考虑将这些协变量的 $K_{\text{L}} \cdot (K_{\text{L}} + 1)/2$ 个二次项和交互作用纳入回归模型。
- 应该注意，该方法只考虑了第二$\boxed{\text{F}}$选择的 $K_{\text{L}} \leq K$ 个协变量的高阶项。
- 与第二步的思想类似，设某时刻已经在 $K_{\text{L}} \cdot (K_{\text{L}} + 1)/2$ 个二次项和交互作用中选择了 $\tilde{K}_{\text{Q}}$ 个，接下来对剩余的项分别纳入现有模型，得到 $K_{\text{L}} \cdot (K_{\text{L}} + 1)/2 - \tilde{K}_{\text{Q}}$ 个逻辑回归模型，并分别计算似然比统计量 (likelihood ratio statistic) 来评估新纳入的项在回归模型中系数为 0 的零假设。

倾向性得分估计 Step 3：引入二次项和交互项

- 若至少有一个项被纳入后的似然比统计量超过了预先设置的临界值 C_{Q}，则将似然比统计量最大的一个协变量纳入模型中。
- 接下来检验剩余的 $K_{\mathrm{L}} \cdot (K_{\mathrm{L}}+1)/2 - \tilde{K}_{\mathrm{Q}} - 1$ 个项的似然比统计量，来判断其是否应被纳入现有模型。
- 我们迭代上述过程，直到剩余项的似然比统计量均小于 C_Q。

The Reinisch et al. 巴比妥暴露数据

Table 13.1. ***Summary Statistics Reinisch Data Set***

Label	Variable Description	Controls (N_c =7198)		Treated (N_t =745)		t-Stat
		Mean	(S.D.)	Mean	(S.D.)	Difference
sex	Sex of child (female is 0)	0.51	(0.50)	0.50	(0.50)	−0.3
antih	Exposure to antihistamine	0.10	(0.30)	0.17	(0.37)	4.5
hormone	Exposure to hormone treatment	0.01	(0.10)	0.03	(0.16)	2.5
chemo	Exposure to chemotherapy agents	0.08	(0.27)	0.11	(0.32)	2.5
cage	Calendar time of birth	−0.00	(1.01)	0.03	(0.97)	0.7
cigar	Mother smoked cigarettes	0.54	(0.50)	0.48	(0.50)	−3.0
lgest	Length of gestation (10 ordered categories)	5.24	(1.16)	5.23	(0.98)	−0.3
lmotage	Log of mother's age	−0.04	(0.99)	0.48	(0.99)	13.8
lpbc415	First pregnancy complication index	0.00	(0.99)	0.05	(1.04)	1.2
lpbc420	Second pregnancy complication index	−0.12	(0.96)	1.17	(0.56)	55.2
motht	Mother's height	3.77	(0.78)	3.79	(0.80)	0.7
motwt	Mother's weight	3.91	(1.20)	4.01	(1.22)	2.0
mbirth	Multiple births	0.03	(0.17)	0.02	(0.14)	−1.9
psydrug	Exposure to psychotherapy drugs	0.07	(0.25)	0.21	(0.41)	9.1
respir	Respiratory illness	0.03	(0.18)	0.04	(0.19)	0.7
ses	Socioeconomic status (10 ordered categories)	−0.03	(0.99)	0.25	(1.05)	7.0
sib	If sibling equal to 1, otherwise 0	0.55	(0.50)	0.52	(0.50)	−1.6

The Reinisch et al. 巴比妥暴露数据

Table 13.2. ***Estimated Parameters of Propensity Score: Baseline Case; Barbituate Data***

Variable	EST	$\widehat{(\text{s. e.})}$	t-Stat
Intercept	−2.38	(0.06)	−41.0
sex	−0.01	(0.08)	−0.2
lmotage	0.48	(0.04)	11.7
ses	0.10	(0.04)	2.6

The Reinisch et al. 巴比妥暴露数据

Table 13.3. ***Estimated Parameters of Propensity Score: Baseline Case with*** `lpbc420` ***Added; Barbituate Data***

Variable	EST	$\widehat{(\text{s.e.})}$	t-Stat
Intercept	−3.71	(0.10)	−36.3
`sex`	0.07	(0.09)	0.8
`lmotage`	0.22	(0.05)	4.7
`ses`	0.15	(0.05)	3.3
`lpbc420`	2.11	(0.08)	27.2
LR statistic	1308.0		

The Reinisch et al. 巴比妥暴露数据

Table 13.4. ***Likelihood Ratio Statistics for Sequential Selection of Covariates to Enter Linearly; Barbituate Data***

Covariate	Step →										
sex	–	–	–	–	–	–	–	–	–	–	–
antih	17.5	0.5	1.6	1.3	2.1	1.8	1.6	1.6	1.7	1.3	–
hormone	3.9	0.3	0.7	0.7	0.4	0.8	0.7	0.7	0.7	0.8	0.9
chemo	10.0	36.6	41.9	–	–	–	–	–	–	–	–
cage	0.8	5.8	6.4	7.2	7.6	7.9	–	–	–	–	–
cigar	4.3	2.3	3.5	3.7	3.0	2.1	2.1	1.7	2.1	–	–
lgest	0.4	11.1	5.0	6.4	7.3	5.5	5.6	–	–	–	–
lmotage	–	–	–	–	–	–	–	–	–	–	–
lpbc415	0.6	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.0
lpbc420	1308.0	–	–	–	–	–	–	–	–	–	–
motht	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
motwt	6.1	1.5	0.6	1.2	2.5	2.7	2.4	3.4	–	–	–
mbirth	4.6	66.1	–	–	–	–	–	–	–	–	–
psydrug	93.1	29.8	38.9	46.8	–	–	–	–	–	–	–
respir	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ses	–	–	–	–	–	–	–	–	–	–	–
sib	21.0	13.8	12.5	15.0	15.7	–	–	–	–	–	–

The Reinisch et al. 巴比妥暴露数据

Table 13.5. ***Estimated Parameters of Propensity Score: Baseline Case with*** `lpbc420` ***and*** `mbirth` ***Added; Barbituate Data***

Variable	EST	(s. e.)	t-Stat
Intercept	−3.73	(0.10)	−35.9
`sex`	0.08	(0.09)	0.9
`lmotage`	0.21	(0.05)	4.5
`ses`	0.16	(0.05)	3.4
`lpbc420`	2.21	(0.08)	27.5
`mbirth`	−1.96	(0.30)	−6.6
LR statistic	66.1		

The Reinisch et al. 巴比妥暴露数据

Table 13.6. *Estimated Parameters of Propensity Score: Final Specification; Barbituate Data*

Variable	EST	$\widehat{(\text{s.e.})}$	t-Stat
Intercept	−5.67	(0.23)	−24.4
Linear terms			
sex	0.12	(0.09)	1.3
lmotage	0.52	(0.11)	4.7
ses	0.06	(0.09)	0.6
lpbc420	2.37	(0.36)	6.6
mbirth	−2.11	(0.36)	−5.9
chemo	−3.51	(0.67)	−5.2
psydrug	−3.37	(0.55)	−6.1
sib	−0.24	(0.22)	−1.1
cage	−0.56	(0.26)	−2.2
lgest	0.57	(0.23)	2.5
motwt	0.49	(0.17)	2.9
cigar	−0.15	(0.10)	−1.5
antih	0.17	(0.13)	1.3
Second-order terms			
lpbc420 × sib	0.60	(0.19)	3.1
motwt× motwt	−0.10	(0.02)	−4.5
lpbc420 × psydrug	1.88	(0.39)	4.8
ses× sib	−0.22	(0.10)	−2.2
cage× antih	−0.39	(0.14)	−2.8
lpbc420× chemo	1.97	(0.49)	4.0
lpbc420 × lpbc420	−0.46	(0.14)	−3.3
cage × lgest	0.15	(0.05)	3.0
lmotage × lpbc420	−0.24	(0.10)	−2.5
mbirth× cage	−0.88	(0.39)	−2.3
lgest × lgest	−0.04	(0.02)	−2.0
ses× cigar	0.20	(0.09)	2.2
lpbc420× motwt	0.15	(0.07)	2.0
chemo × psydrug	−0.93	(0.46)	−2.0
lmotage× ses	0.10	(0.05)	1.9
cage × cage	−0.10	(0.05)	−1.8
mbirth × chemo	−∞	(0.00)	−∞

The Reinisch et al. 巴比妥暴露数据

Table 13.7. ***Determination of the Number of Blocks and Their Boundaries; Barbituate Data***

Step	Block	Lower Bound	Upper Bound	Width	# Controls	# Treated	t-Stat
1	1	0.00	0.94	0.94	4462	742	36.3
2	**1**	0.00	0.06	0.06	2540	61	3.2
	2	0.06	0.94	0.88	1922	681	23.7
3	**1**	0.00	0.02	0.01	1280	20	2.2
	2	0.02	0.06	0.05	1260	41	0.5
	3	0.06	0.20	0.14	1163	138	3.9
	4	0.20	0.94	0.74	759	543	10.9
4	**1**	0.00	0.01	0.00	644	6	−0.0
	2	0.01	0.02	0.01	636	14	1.7
	3	0.02	0.06	0.05	1260	41	0.5
	4	0.06	0.11	0.05	604	46	−0.3
	5	0.11	0.20	0.09	559	92	1.0
	6	0.20	0.37	0.17	458	192	1.2
	7	0.37	0.94	0.57	301	351	5.6
5	1	0.00	0.01	0.00	644	6	−0.0
	2	0.01	0.02	0.01	636	14	1.7
	3	0.02	0.06	0.05	1260	41	0.5
	4	0.06	0.11	0.05	604	46	−0.3
	5	0.11	0.20	0.09	559	92	1.0
	6	0.20	0.37	0.17	458	192	1.2
	7	0.37	0.50	0.13	181	144	2.5
	8	0.50	0.94	0.44	120	207	2.3
6	1	0.00	0.01	0.00	644	6	−0.0
	2	0.01	0.02	0.01	636	14	1.7
	3	0.02	0.06	0.05	1260	41	0.5
	4	0.06	0.11	0.05	604	46	−0.3
	5	0.11	0.20	0.09	559	92	1.0
	6	0.20	0.37	0.17	458	192	1.2
	7	0.37	0.42	0.05	101	61	0.3
	8	0.42	0.50	0.08	80	83	0.7
	9	0.50	0.61	0.11	73	90	0.8
	10	0.61	0.94	0.34	47	117	−0.3

Note: Boldface block numbers indicate blocks that were split at this step.

总结

- 实验性研究：线性回归：估计量的相合性不依赖于模型的正确假设。
- 实验性研究：基于模型的推断方法：为所有潜在结果建立一个随机模型，通过贝叶斯方法，使用假设模型来估算给定观察数据的缺失潜在结果。
- 分层随机试验（Stratified Randomized Experiments）：线性回归模型的如何设定，使得估计量具有相合性？
- 配对随机实验（Pairwise Randomized Experiments）：如何解决层内治疗组与对照组样本方差不可估？
- 因果推断在观察性研究中的非混淆性：均衡得分与倾向性得分。
- 倾向性得分估计：基于逐步回归的思想。

Thanks!