

随机控制实验（RCT）及其实例应用

2019.06.09

王裕、刘娟伶

一、随机控制实验

随机控制实验(RCT)是因果推断的常用方法。RCT 最初应用于医学领域，主要用于对医疗干预效果的评估。实验者会被随机分组，在实验者不知情的情况下，一部分人服用真实药物而另一部分人则服用类似的安慰剂，这样可以排除所有的人为主观因素，达到真实的评估效果。后来 RCT 才逐渐应用于经济学研究。

下面先详细说明随机控制实验背后的统计理论知识。

（1）处理效果与样本产生的选择性偏差以及随机控制实验

如果我们有一大批吃阿司匹林的人群，和另外一群不吃的人群，我们想计算吃药和不吃药的人头疼的样本均值。一个自然的对处理效果(treatment effect)的估计量如下。基于样本均值的性质，我们知道这是一个很好的估计量，让我们看一下这个估计量的含义是什么：

观察到的偏差

$$\begin{aligned} &= E[Y_i^{obs} | W_i = 1] - E[Y_i^{obs} | W_i = 0] \\ &= E[Y_i(1) | W_i = 1] - E[Y_i(0) | W_i = 0] \\ &= E[Y_i(1) | W_i = 1] + (-E[Y_i(0) | W_i = 1] + E[Y_i(0) | W_i = 1]) - E[Y_i(0) | W_i = 0] \\ &= (E[Y_i(1) | W_i = 1] - E[Y_i(0) | W_i = 1]) + (E[Y_i(0) | W_i = 1] - E[Y_i(0) | W_i = 0]) \\ &= \text{处理效果(treatment effect)} + \text{选择性偏差(selection bias)} \end{aligned}$$

其中 $(E[Y_i(1) | W_i = 1] - E[Y_i(0) | W_i = 1])$ 是处理效果(treatment effect)，

$(E[Y_i(0) | W_i = 1] - E[Y_i(0) | W_i = 0])$ 是选择效应(selection effect)，即选择性偏差。

比如：

头疼更厉害的人更可能选择吃药，那么选择吃药与不吃药的两组人，在都不吃药（不处理）的情况下，对比两组的头疼程度，选择吃药组的人会有更强的头疼程度，即对比值不为 0（更可能是大于 0）。此时就有了选择性偏差，两组样本均值的差就不再是对处理效果的无偏估计了。

上大学的与不上大学的人在很多方面是不同的。

招聘者可能会说两个应聘者除了种族不同，还有其它的不同点。

什么情况下这个（选择性）偏差会消失呢？要么创造一种情形让选择性偏差消失，或者在现实世界找到一个我们能够相信的情形，再做一个合适的比较。也就是我们可以进行随机控制实验，随机化解决选择问题。

随机控制实验是说，在一个完全随机化的实验中，从样本中随机抽取 N_t 单元作为处理组，随机抽取 N_c 单元作为控制组，那么分配的概率不依赖潜在结果，处理组与控制组在都不进行任何处理的情况下观察值相等，进而：

观察到的偏差

$$\begin{aligned} &= E[Y_i^{obs} | W_i = 1] - E[Y_i^{obs} | W_i = 0] \\ &= E[Y_i(1) | W_i = 1] - E[Y_i(0) | W_i = 0] \\ &= E[Y_i(1) | W_i = 1] + (-E[Y_i(0) | W_i = 1] + E[Y_i(0) | W_i = 1]) - E[Y_i(0) | W_i = 0] \\ &= (E[Y_i(1) | W_i = 1] + E[Y_i(0) | W_i = 1]) + (E[Y_i(0) | W_i = 1] - E[Y_i(0) | W_i = 0]) \\ &= (E[Y_i(1) | W_i = 1] + E[Y_i(0) | W_i = 1]) + 0 \\ &= E[Y_i(1) - Y_i(0) | W_i = 1] \\ &= E[Y_i(1) - Y_i(0)] \\ &= \text{处理效果} \end{aligned}$$

（2）随机控制实验的类型及应用

随机控制试验(RCT)一般有以下几种类型：

1) 简单随机(Completely randomized)；

2) 分层随机(Stratified randomization)：基于一些协变量 X 创建分块，随机化在每个分块内进行；比如，药品疗效实验可以采用年龄分层；

成对随机(Pairwise randomization)：创建一个个配对，随机化在每个配对内进行；比如，两个非常亲密的人作为一对，然后随机地把其中一个放到处理组，另一个放到控制组；

3) 群随机(Clustered randomization)：单元不是单独的个体，而是一组个体(例如：班级)。

无论那种随机化，只要让选择性偏差消失，结果都是有效的。为什么呢？因为可以用样本均值估计期望。

什么时候我们会更倾向于其中某一种呢？回到前面我们提到的例子，想想针对下面这些问题哪种 RCT 更适合：

1) 阿司匹林案例：

这是经典的医药 RCT，可以进行简单随机化实验，在个体级别进行随机化，也可以按年龄、种族、性别等分层。

2) 上大学案例：

我们不能随机分配让人们“上大学”：一些人可能无论如何都不想上大学，并且一些人会无论如何都要去上大学，不会同意进入控制组。这不是一个可以实际实施的随机化实验。

我们能做什么？有一个和巴菲特基金会一起做的一个研究。巴菲特基金会为内布拉斯加州的一些人提供资助，但他们不能为每个申请人都提供资助，所以他们在同等合格的候选人中随机选择并给予奖学金。奖学金是一种鼓励，并且保证了一些人比另外一些人更“可能”上大学。这个奖学金计划本身可以被看作一个简单随机化实验。但研究的不是“上大学”的效果，而是“鼓励上大学”的效果，这个被称之为一个“鼓励设计”。因为一些获得奖学金的人可能没有去上大学，另一些没有获得奖学金的人仍然去了大学。

3) 种族和工作案例：

我们需要从重新定义问题开始。看看问题的意思具体是在说：非裔美国人在总体上存在劣势？或者他们因为给定的特征受到雇主的潜在歧视？

如果是这种情况，我们可以在不改变其它信息的情况下操纵人们对种族的认知：

试镜研究(Audit studies)：通过派遣除了种族不同之外具有相同特征的一对演员，来操纵种族认知；试镜研究是一种在经济学、社会学、政治科学与心理学中被用到的一种真实环境试验(field experiment)。参与试验的人是受过训练的雇员，被配对成除了被研究的(可能会被歧视的)特征之外所有特征都一样的组。

简历研究：通过发送除了名字不同（听起来像或不像一个非裔美国人）其他信息一模一样的简历。这是一个很有名的歧视研究，由 Marianne Bertrand 和 Sendhil Mullainathan 一起完成，同样的简历，名字听起来像非裔美国人的候选人只收到相对 1/2 的面试机会，同时该研究做了教育分层分析，在高等教育分层中，差别更大。同时这是一个成对的随机试验，把一对除了名字其它完全相同的简历发给同一个雇主。

二、RCT 的实例应用

下面结合一个具体实例来说明平均处理效果的估计与标准差的计算，使用置信区间与基于正态分布的假设检验这一传统 RCT 的数据分析。

(1) 平均处理效果

我们知道，在 RCT 条件下，就是平均处理效果，怎样才能做到：

- 1) 找到一个好的估计量 (estimator)
- 2) 获得这个估计量的标准误差 (standard error) 估计值 (estimate)
- 3) 检验它是不是零 (估计量的实际实验计算结果，即平均处理效果)

假设我们有一个简单随机实验，有 N_t 个被处理单元，有 N_c 个控制单元，对兴趣对象

的合理估计量可能是什么呢？

(2) 估计处理效果和他们的标准差

两组样本均值的差值是一个对处理效果的无偏估计：

$$\hat{T} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs} - \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{obs} = \overline{Y_t^{obs}} - \overline{Y_c^{obs}}$$

两组统计独立变量差值的方差是他们方差之和，因此这个估计量的方差是：

$$V(\hat{T}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$$

使用基于样本数据计算的值替换上式中的分子作为方差的估计：

$$S_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} (Y_i(0) - \overline{Y_c^{obs}})^2$$
$$S_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} (Y_i(1) - \overline{Y_t^{obs}})^2$$

(3) 置信区间

回顾关于置信区间的定义：我们要找到关于随机样本的函数 **A** 和 **B**，以满足以下条件：

$$P(A(X_1 \dots X_N) < \theta < B(X_1 \dots X_N)) > 1-\alpha$$

我们知道两组样本均值的差值与估计标准差的比值会符合一个 **t** 分布，所以：

- 1) 在小样本情况下从自由度为 **Nt+Nc-1** 的 **t**-分布表取 α 对应的关键值
- 2) 在大样本的情况下，使用近似正态，从标准正态分布表取关键值，例如 $\alpha=0.1$ 的关键值 1.645， $\alpha=0.05$ 的关键值是 1.96

(4) 假设检验

让我们从一个标准假设开始（零与非零）

$$H_0 : \frac{1}{N} \sum_{i=1}^N Y(1) - Y(0) = 0$$
$$H_a : \frac{1}{N} \sum_{i=1}^N Y(1) - Y(0) \neq 0$$

比较自然能想到的检验统计量为：

$$t = \frac{\overline{Y}_t^{obs} - \overline{Y}_c^{obs}}{\sqrt{\hat{V}}}$$

服从一个自由度为 $N-1$ 的 t 分布，或当 N 足够大时的正态分布，相关的双边 p 值为： $2 * (1 - \Phi(t))$ 。

（5）例子：俄勒冈州医保实验

医保成为近几届美国总统大选的一个焦点，特朗普说：“没有人知道医保会这么复杂”，奥巴马推动通过了奥巴马医保法案—平价医保法案（Affordable Care Act, ACA），而奥巴马的继任特朗普致力于废除 ACA。

那么 ACA 到底起了怎样的作用，如何研究它的因果效应呢？

幸运的是有一个独特的实验可以揭示这个法案的效应。美国俄勒冈州想要在 ACA 之前扩大医疗补助计划---俄勒冈州健康计划（OHP），但没有足够的钱去全面覆盖：他们决定采用抽签方式。这样就满足了 RCT 的要求。

Amy Finkelstein 带了一组研究人员实施了这个研究，跟踪抽到签的人和没抽到签的人的健康情况和收入情况。数据的统计结果见下表：

TABLE 1.6

Does winning the Oregon Health Plan lottery in 1996 affect health and financial health? (1996-2000)

OHP effects on health indicators and financial health

Outcome	Oregon		Portland area	
	Control mean (1)	Treatment effect (2)	Control mean (3)	Treatment effect (4)
A. Health indicators				
Health is good	.548	.039 (.008)		
Physical health index			45.5	.29 (.21)
Mental health index			44.4	.47 (.24)
Cholesterol			204	.53 (.69)
Systolic blood pressure (mm Hg)			119	-.13 (.30)
B. Financial health				
Medical expenditures >30% of income			.055	-.011 (.005)
Any medical debt?			.568	-.032 (.010)
Sample size	23,741		12,229	

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on health indicators and financial health. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.

表中报告了抽到俄勒冈州健康计划（OHP）的人在健康指数和财务健康方面的效果。奇数列显示的是控制组的均值，偶数列显示的是关于中签者哑变量的回归系数，括号里的是（回归系数的）标准误差。

让我们基于这个表来计算一下，俄勒冈州健康计划在健康指数和财务健康方面的效果。

1) 表的使用

波特兰的实验组的平均健康指数是多少？

实验组均值=控制组均值+处理效果=45.5+0.29= 45.79

2) 置信区间

让我们计算一下拥有保险在“health is good”变量上“效果”的 95%置信区间，表中 0.039 是处理效果的估计，0.008 是该估计的标准误差，所以置信区间为：

$$(0.039-0.008*1.96, 0.039+0.008*1.96) \rightarrow (0.02332, 0.05468)$$

3) 假设检验

在 5%水平上是否拒绝零假设：抽到和没抽到保险的人在“health is good”这项指标上没有差别？

拒绝零假设，因为 $0.039/0.008=4.875>1.96=q_{0.025}$ （使用标准正态分布双边 5%水平的关键值）

在 10%水平上是否拒绝零假设：在波特兰抽到和没抽到保险的人在身体健康指数上没有差别？

不拒绝零假设， $0.29/0.21=1.381 \leq 1.645=q_{0.05}$ （使用标准正态分布双边 10%水平的关键值）

在 10%水平上是否拒绝零假设：在波特兰胆固醇水平降低了？

不拒绝零假设， $0.53/0.69=0.768 \leq 1.645=q_{0.05}$ （使用标准正态分布双边 10%水平的关键值）

参考文献：

http://www.sohu.com/a/314491288_617676

http://www.sohu.com/a/315841264_617676?spm=smpc.author.fd-d.12.1560050685246mFBhuih

<http://3g.163.com/dy/article/EFA8D8TJ0518F074.html>