

## 02随机实验与观测性研究

### 1. 两种方法的比较

	研究思路	优点	局限	其他
随机实验	随机分组-分析数据	不需要控制其他任何混淆变量	存在很多伦理限制	
观测性研究	收集资料-区分被研究者是否接受处理变量处理-分成控制组和实验组-分析数据	数据已经被收集好	潜在的混淆变量无法完全消除	大多数社会科学研究必选方案

### 2. 随机实验

#### 2.1 随机实验的性质与基本类型

- 随机实验是目前为止因果推论最好的方法。
- 随机实验能够保证每个被研究对象进入到控制组和实验组的概率是一样的。

##### 2.1.1 伯努利实验

伯努利实验																
A男	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0	1
B男	0	0	0	1	0	1	0	0	1	1	0	1	1	0	1	1
C女	0	0	1	0	0	0	1	0	1	0	1	1	0	1	1	1
D女	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1	1

- 通过抛硬币的方式决定样本进入实验组或控制组
- 缺点：容易造成样本损失，实验组和控制组人员不均衡。

2.1.2 经典随机实验

经典随机实验																
A男	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0	1
B男	0	0	0	1	0	1	0	0	1	1	0	1	1	0	1	1
C女	0	0	1	0	0	0	1	0	1	0	1	1	0	1	1	1
D女	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1	1

- 经典随机实验需要确定被研究对象里面有多少人去接受处理变量的特定取值

2.1.3 分层随机实验

分层随机实验																
A男	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0	1
B男	0	0	0	1	0	1	0	0	1	1	0	1	1	0	1	1
C女	0	0	1	0	0	0	1	0	1	0	1	1	0	1	1	1
D女	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1	1

- 在实验之前提前控制一些我们想控制的一些混淆因素。例如在男生里面随机分配，同样在女生组内做随机分配。
- 通常来讲，如果有比较强烈的理由认为有一些混淆变量存在，分层随机实验会使用的比较多。

2.2 随机实验在因果推断中的重要作用

- 经验研究的重点就是将影响Y系统性变化的原因找出来。
- 响应变量的两个变化来源：系统变异  $f(D)$  和随机扰动  $e$ 。通常，要求研究结论尽可能归于系统变化而非随机扰动。

$$Y = f(D) + e$$

- 一个前提：样本一定要足够大。被研究对象是**随机分配**到不同的处理变量取值水平的，所以除了处理变量之外，其他所有的可见的不可见的混淆因素在两组之间都会被拉平。
- 随机实验的好处就在于控制组和实验组的被研究对象如果在响应变量Y上有任何变化的话，只能归因于处理变量的取值不同。
- 在现实中，随机实验还存在很多伦理限制。

## 2.3 如何分析随机实验数据

### 2.3.1 T检验

- 零假设：组间均值相等
- 以测试新旧教学项目效果为例，1表示参加新教学项目，0表示参加旧的教学项目

个案	教学项目	考试成绩
A	1	85
B	1	92
C	0	81
D	1	95
E	0	76
F	0	80

- T检验的R语言代码

```
1 y = c(85,92,81,95,76,80)
2 t.test(y~c(1,1,0,1,0,0))
```

- 结果显示  $p = 0.03962$ ，大于  $0.001$ ，说明两组之间存在显著性差异

```
1      Welch Two Sample t-test
2
3 data:  score by c(1, 1, 0, 1, 0, 0)
4 t = -3.5, df = 2.9931, p-value = 0.03962
5 alternative hypothesis: true difference in means between group 0 and group 1
6 95 percent confidence interval:
7  -22.288637 -1.044696
8 sample estimates:
9 mean in group 0 mean in group 1
10      79.00000      90.66667
```

- 实际上估计量的方差由三部分组成，实验组、控制组和协方差。将协方差考虑进去之后，实际上的p值会更小一点。

关心的因果关系 
$$\tau = \frac{\sum_{i=1}^N Y_i(1)}{N} - \frac{\sum_{i=1}^N Y_i(0)}{N} = \overline{Y(1)} - \overline{Y(0)}$$

估计量 
$$\hat{\tau} = \frac{\sum_{i=1}^{N_t} Y_i \times D_i}{N_t} - \frac{\sum_{i=1}^{N_c} Y_i \times (1-D_i)}{N_c}$$

估计量的方差 
$$\text{var}(\hat{\tau}) = \frac{s_T^2}{N_T} + \frac{s_C^2}{N_C} - \frac{s_{TC}^2}{N}$$

- 对于T检验，研究者关心的主要问题是实验组和控制组的均值是不一样的。但均值有时是不可靠的。一种替代方案就是费舍尔精确检验。

### 2.3.2 费舍尔精确检验 (Fisher's Exact Test)

- 强 (strong) 零假设：对于每个个体而言，处理变量没有任何效果 (例如  $Y_i(1) = Y_i(0)$ )
- 

### 2.3.3 总结

1. 随机试验本质上取决于研究者能否把不感兴趣的处理变量**随机分配**到不同的对象上。
2. 实际上，国家层面的随机实验并不多，区域性的政策的实施逐渐使用随机实验。
3. 部分研究问题虽然能做到随机安排，但有时很难找到有社会意义的Y（摇号车牌）。

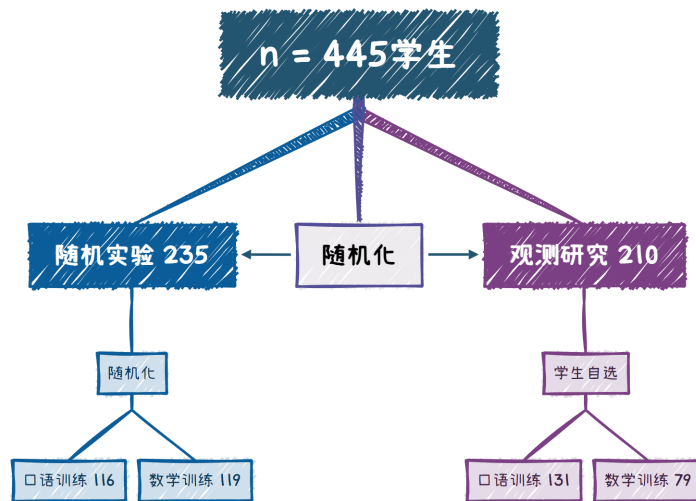
## 3. 观测研究初步：有限混淆变量

### 3.1 回归法

- 传统的回归模型（如OLS，广义线性模型）共同点在于可以将一系列的混淆变量的观测作为控制变量放在模型中，从而起到消除控制变量影响的作用。D表示处理变量，C表示控制变量，e表示随机扰动项。

$$Y = f(D) + f(C) + e$$

- Shadish等人的研究（2008）表明传统的回归方法在满足**一定的前提下**也是能够复制随机实验的结果的。



- **外推(extrapolation)**: 利用别的数据点建立的模型去推广到没有直接数据支撑的部分。
- 外推对于因果推论最大的问题在于，结论有可能不是数据的经验模式，而是混杂了基于模型对数据位置区域的探索。对于因果关系的推论就取决于模型的形式，形式不同，估计也就不同。也就是说我们的结论会很不稳健。
- 在实际研究中，完全可以用回归模型去做因果推断的研究，但要注意控制变量在实验组和控制组之间并不平衡，一旦出现这种问题，回归的结果就有可能不能做因果诠释。

## 3.2 细分法

### 3.2.1 辛普森悖论

- 美国高校两个学院的录取情况汇总，可以发现汇总后的女生的录取率比男生的录取率低，能否说明两个该学校在新生录取方面存在性别歧视呢？

性别	录取	拒收	总数	录取率
男生	209	95	304	<b>68.8%</b>
女生	143	110	253	<b>56.5%</b>

- 如果将录取情况按照学院分开，反而得到相反的结果

Table: 法学院的录取数据

性别	录取	拒收	总数	录取率
男生	8	45	53	15.1%
女生	51	101	152	33.6%

Table: 商学院的录取数据

性别	录取	拒收	总数	录取率
男生	201	50	251	80.1%
女生	92	9	101	91.1%

- 原因在于商学院整体录取率较高，但商学院女生的申请总数与男生相比较少，法学院录取率低，但与男生相比女生的申请总数高于男生，整体就给我们造成了**女生被拒绝得多的假象**。
- **辛普森悖论 (Simpson`s Paradox,1951)** :即在某个条件下的两组数据，分别讨论时都会满足某种趋势，可是一旦合并考虑，却可能导致相反的结论。

### 3.2.2 细分法

- **细分法 (subclassification)** : 传统的回归模型在因果推论上的问题在于研究者无法直接观察到混淆变量在处理变量在不同水平下的平衡情况，一个最直接的方法就是一一查看混淆变量的**平衡性**，然后在混淆变量取值一致的前提下分别考察因果关系，之后汇总（让混淆因素在控制组和实验组之间达到均衡）。
- **维度灾难**：当混淆变量增多时，我们不断对样本进行切割时，针对每个切块，内部能用到的研究对象迅速减少。

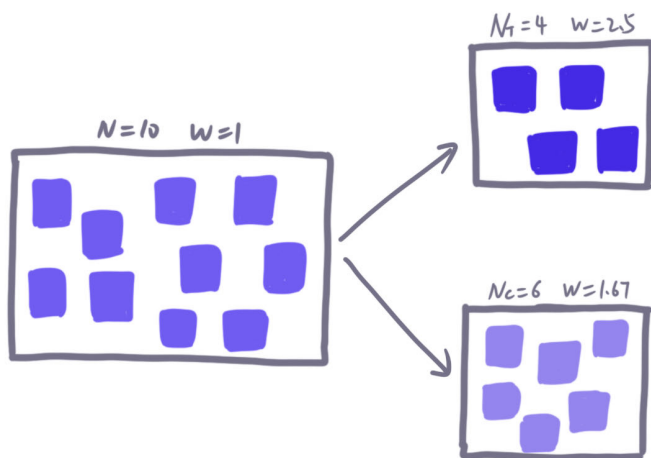
### 3.2.3 总体不均衡和抽样不均衡

- 细分法的工作就是让混淆因素在实验组和控制组之间达到均衡。
- **总体不均衡**：实验组和控制组的人彼此之间有一些系统差异，这种系统差异是数据本身的问题，和获取数据的方式无关。
- **抽样不均衡**：控制变量在总体和样本所占的比例不同。（例如总体中80%男性，而样本中男性占比50%）
- 如果只就因果推论谈均衡的话，是不在乎样本里的比例和总体比例之间的差异的，即不考虑总体是否均衡。

## 3.3 加权法

### 3.3.1 加权法





### 3.3.2 赫维兹-汤普森估计量

- 赫维兹-汤普森估计量 ()

$$Y_{HT} = \frac{1}{n} \times \sum_{i=1}^n \frac{Y_i}{\pi_i}$$

### 3.3.3 加权导致标准误扩大

- 加权通常会导致标准误扩大，因为加权以后相当于认为缩小了样本量，标准误扩大了多少基本取决于权重的大小。
- 通常来讲，权重最好不要超过20。相当于样本量除以20，这实际上是一个很大的样本损失。
- 加权的处理无法处理样本量不够的情形。如果一个模型控制变量有20个的话，就需要保证样本数量在400以上来保证样本量足够大。

### 3.3.4 总结

- 理想的加权方案：研究者既有每一个分析层次的权重，又有总的加权方案
- 如果有限混淆变量在实验组和控制组之间的不平衡性不是很大，那么加权对结果的影响差别也不大。