

# 《传播统计学》

## 社会科学因果推断介绍

教师：林志良

邮箱：[linzhl@nfu.edu.cn](mailto:linzhl@nfu.edu.cn)

个人网站：[www.zhilianglin.com](http://www.zhilianglin.com)



# 目录

- 相关与因果
- 反事实与因果
- 随机对照实验
- 自然实验与准实验
- 回归分析

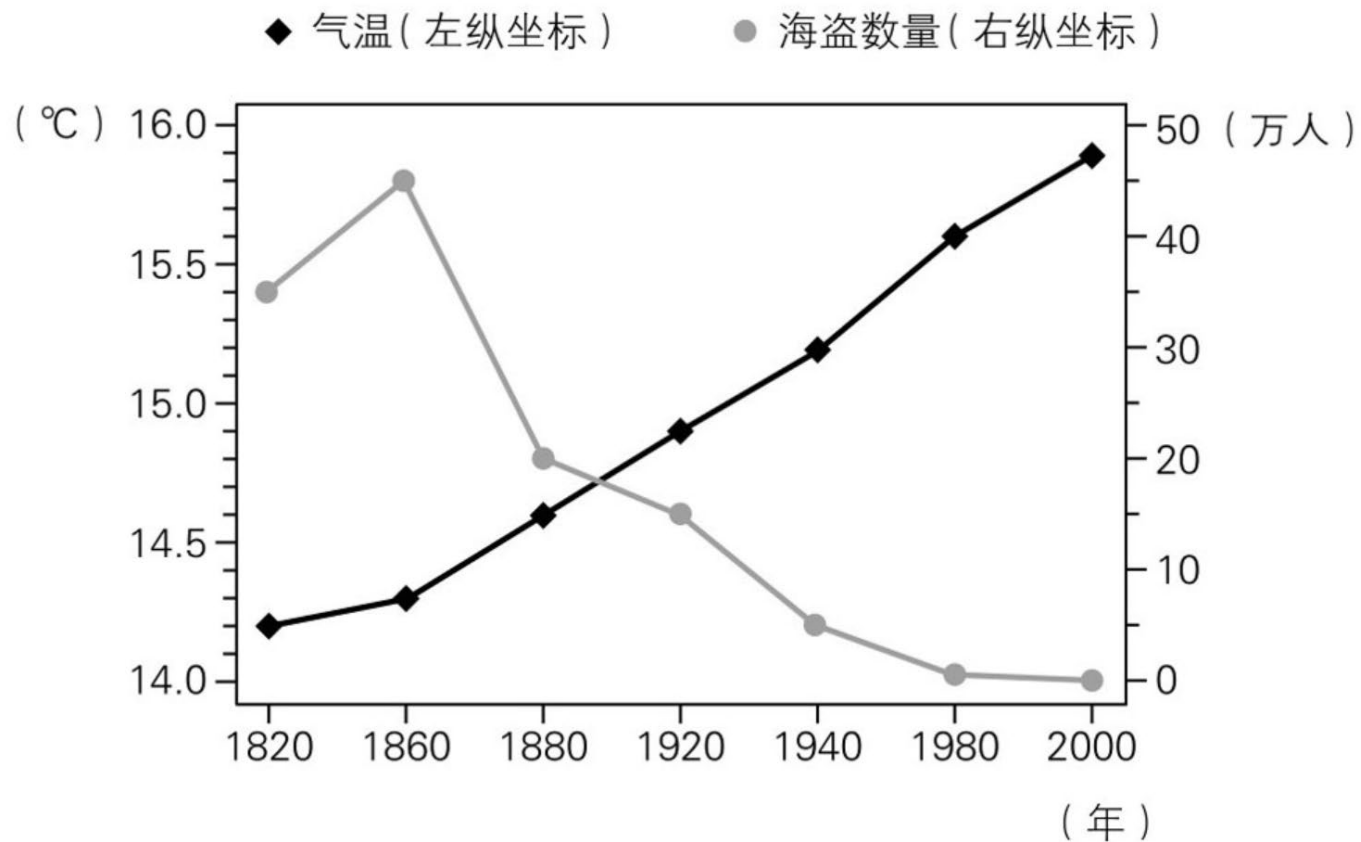
# 相关与因果

## 因果关系成立的三个条件

- 相关关系（“相关不等于因果”）
- 先因果后果
- 排除其它因素

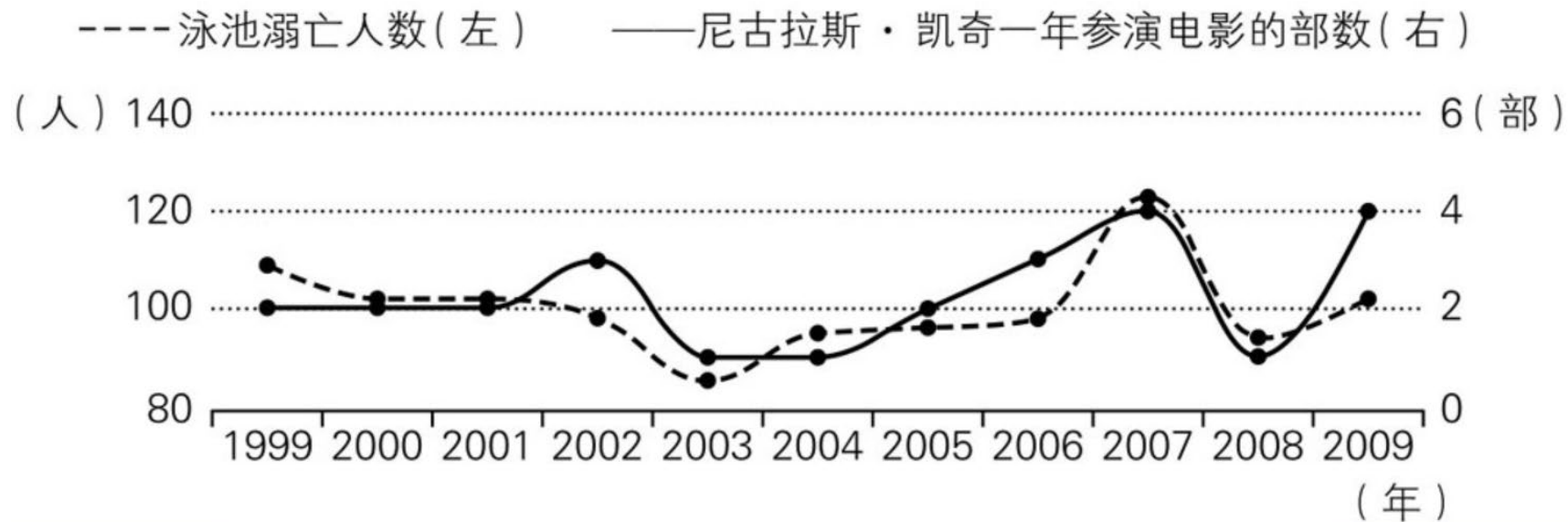
# 相关与因果

## 1、“伪相关”（巧合）

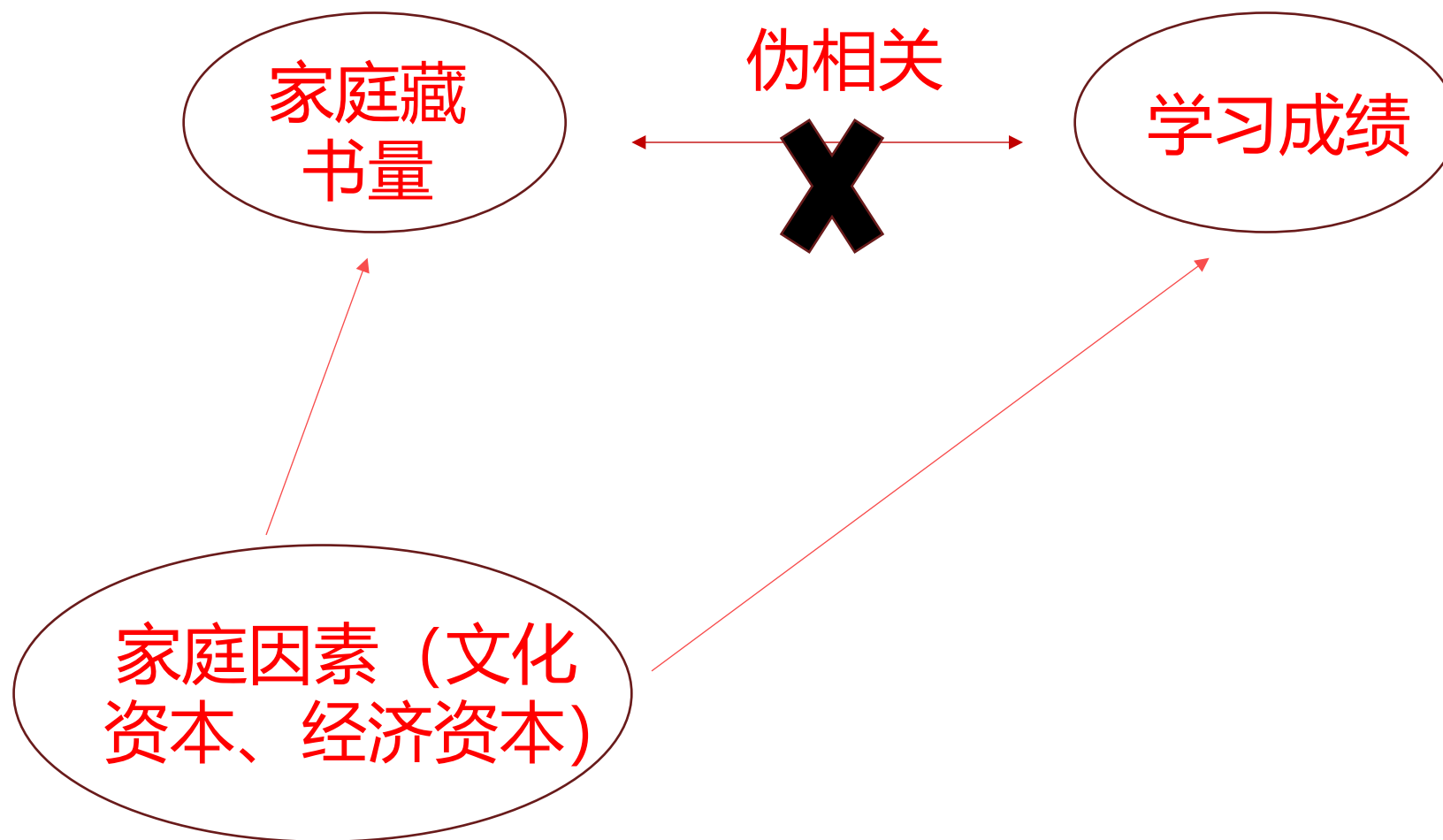


# 相关与因果

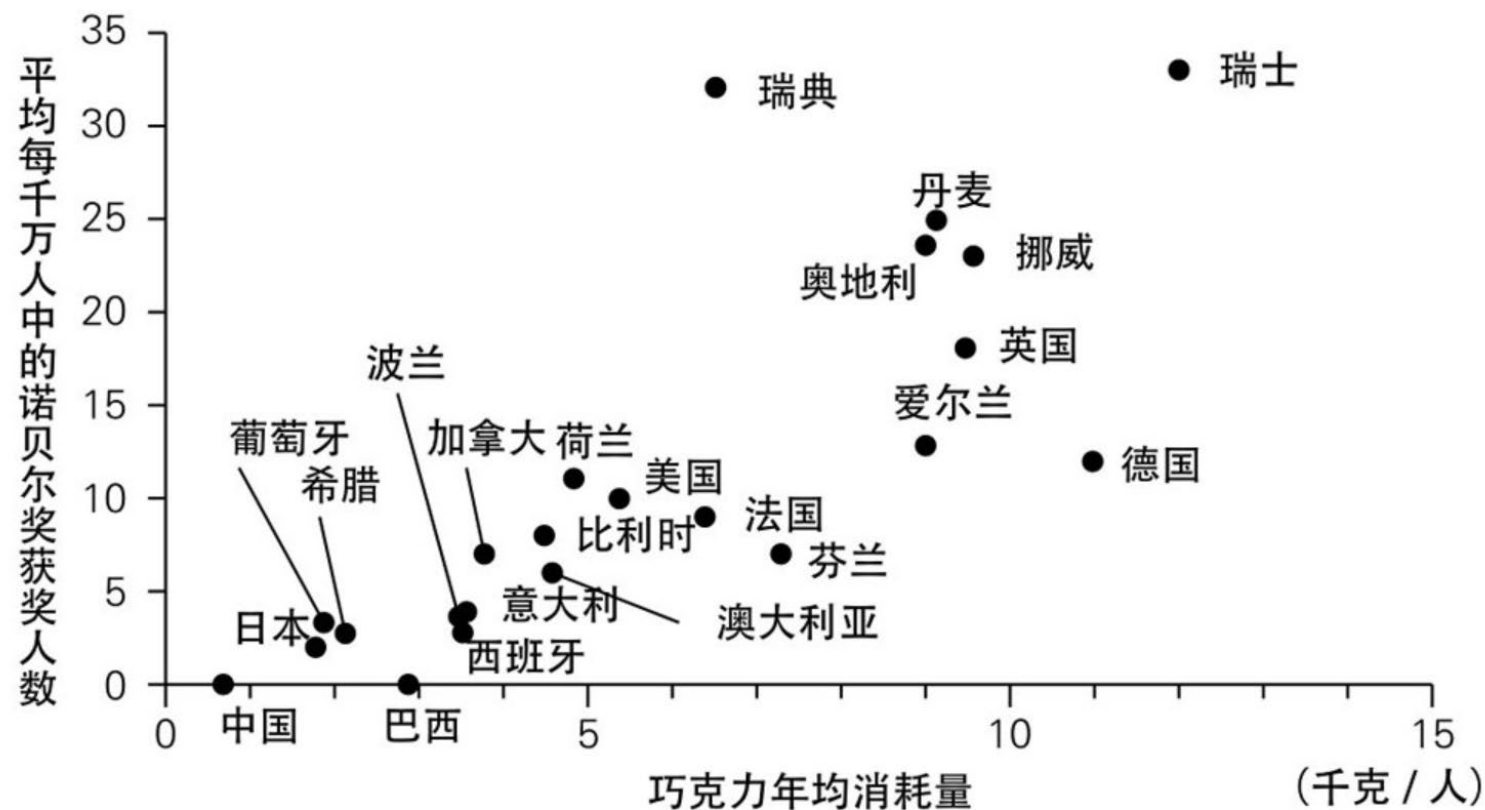
## 1、“伪相关”（巧合）



## 2、第三变量



## 2、第三变量



巧克力消耗量与诺贝尔奖获奖人数的关系 (2012年数据)

# 反事实与因果

她老家甘肃张掖，少年丧父，因而比一般的孩子性格更加坚韧，更有自强的意志。

——《金榜题名之后》序言（熊庆年）





## 反事实与因果

- 假设一位把孩子送入名牌大学的母亲在书中写道：  
“**我从来不让孩子看电视。**” 那么读了这本书的很多人便会认为孩子学习能力高是因为家长没有让孩子看电视。

## 反事实与因果

要想得到因果关系也许需要 “重启人生” / “多重宇宙”



## 反事实与因果

如果一个人吃了某道菜之后死了，那么，只有当他同时又没有吃这道菜之后活着，这两种状态都为我们所观察到时，我们才能说，这道菜是他死去的原因。

——英国哲学家约翰·密尔

# 反事实与因果

## 证明因果关系需要 “反事实”

- **反事实**是指对过去未曾发生的事实所做的假设，例如 “如果当时没有……，那么……”。
- 我们将现实中**实际发生**的事称为 “**事实**”，所以将**设想的与现实完全相反的情况**称为 “**反事实**”。

## 反事实与因果

### 投了广告，销量增加，这是广告的功劳吗？

- 假设你在一家经营全国连锁珠宝店的企业负责广告品牌业务，正在考虑通过投放广告来提高销售额。
- 所幸聘请网红制作的媒体广告十分具有吸引力，投放时期也定在12月下旬，刚好配合元旦促销活动。结果如你所料，广告刊登后，客流量激增，销售额和去年同期相比增长了30%。
- **这是广告的功劳吗？**

## 反事实与因果

**投了广告，销量增加，这是广告的功劳吗？**

- 说不定 “即使不投放广告，销售额也会增长” 。

# 反事实与因果

需要有反事实才能证明因果关系

**事实：**投放广告            销售额：150万

**反事实：**不投放广告            销售额：100万

**因果效应：**广告带来50万的收益增长

## 反事实与因果

**但现实是我们没有反事实.....**

- **“因果推理中的根本问题”**：现实中我们可以观测事实，但无法观测到反事实。



## 反事实与因果

### 没有反事实，怎么办？

- 用最贴切的值来替换无法观测到的“反事实的结果”

## 反事实与因果

### 没有反事实，怎么办？

- 假设你所在的珠宝连锁店在全国四个地区设有店铺。为配合元旦商战，正打算投放广告。考虑到预算有限，最后决定在4个地区当中，为地区1和地区2的店铺投放广告。



# 反事实与因果

## 没有反事实，怎么办？

- 观测到反事实就能明确因果效应——虚构的珠宝店销售额

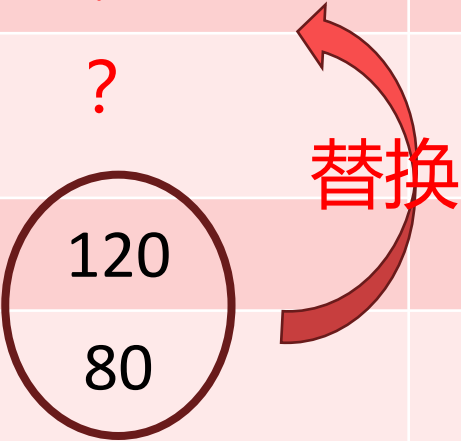
	地区	投广告后的销 售额 [A] （万元）	未投广告后的销 售额[B] （万元）	因果效应 [A-B] （万元）
有广告	1	130	90	40
	2	170	110	60
无广告	3	160	120	40
	4	140	80	60
	平均值	150	100	50

# 反事实与因果

## 没有反事实，怎么办？

- 现实中无法观测反事实——实际的珠宝店销售额

	地区	投广告后的销 售额 [A] （万元）	未投广告后的销 售额[B] （万元）	因果效应 [A-B] （万元）
有广告	1	130	?	?
	2	170	?	?
无广告	3	?	120	?
	4	?	80	?
	平均值	150	100	?



# 反事实与因果

## 没有反事实，怎么办？

- 用“最贴切的值”替换反事实

有广告组的事实 (平均销售额)	无广告组的反事实 (平均销售额)
150万	?



有广告组的事实 (平均销售额)	无广告组的反事实 (平均销售额)	无广告组的事实 (平均销售额)
150万	100万	100万



替换

# 反事实与因果

## 没有反事实，怎么办？

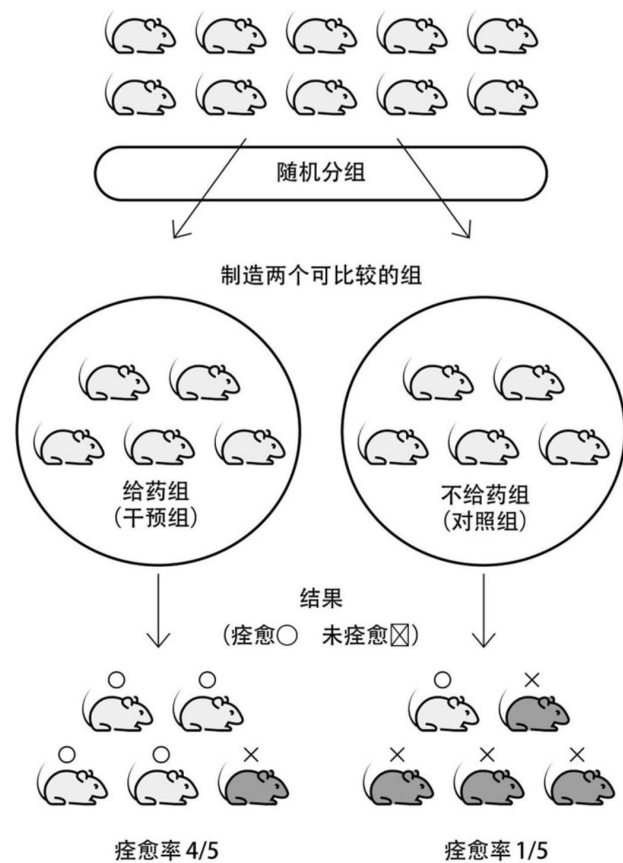
- 上述思路成立需要一个条件：有广告组和无广告组必须是“**可比较**”的。
- 西方谚语：**苹果和橘子是不可以比较的。**

## 小结

- 明确因果关系的方法有很多种，但是这些方法本质上的共同目标都是“制造可比较的组，用最贴切的值替换**反事实**”。
- 在社会科学领域，要探索因果关系，最可靠的方法是用**随机对照实验**，其次是**自然实验与准实验**，再次是**回归分析**。

# 随机对照实验

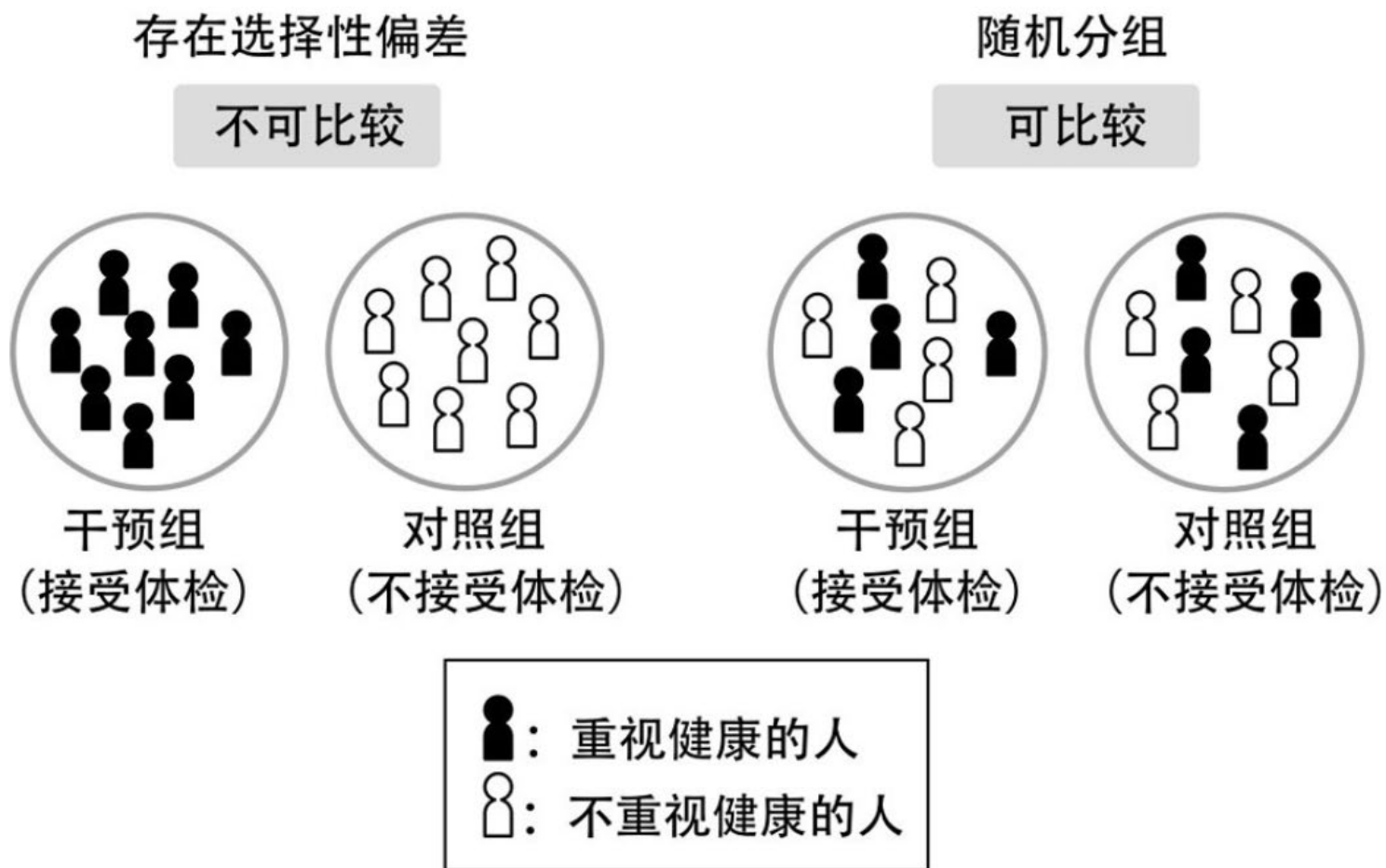
## 用随机对照实验方法验证药物疗效



随机分组制造反事实



## 随机分组的必要性



## 随机对照实验

### 实验法经典案例：柠檬与坏血病

- 18世纪之前，远洋航行是一件高风险事业，水手们饱受坏血病的困扰。
- 18世纪中叶，英国皇家海军的随船医生詹姆斯·林德（James Lind）无意中发现了柠檬可以治疗这种疾病，于是，他就在他所在的船上做了一次临床试验。

## 随机对照实验

### 实验法经典案例：柠檬与坏血病

- 他随机将海员分为两组：
  - 一组的食物供应一如平常（**对照组**）；
  - 另一组则提供柠檬等新鲜水果（**干预组**）。

## 随机对照实验

### 随机对照实验案例：就业市场存在种族歧视吗？

- 制作一些假简历投递给招聘雇员的公司。
- 这个实验主要的差别就在姓名上，是姓名把两份简历分成两组。
- 白人：Emily Walsh, Greg Baker
- 黑人：Lakisha Washington, Jamal Jones
- 白人收到回复概率：1/10
- 黑人收到回复概率：1/15

# 自然实验与准实验

## 自然实验

- 自然实验利用研究对象人群由于法律制度变更、自然灾害等 **“外生冲击”** 的影响而自然分成受影响组（干预组）和不受影响组（对照组）的现象，来验证因果关系。

# 自然实验与准实验

## 自然实验的案例：经济贡献和女孩生存率

- 1980年代，家庭联产承包责任制推广至全国。该制度允许农户充分根据采购价格的变化情况，决定是**种植粮食还是其他经济作物**。
- 有的地区更多地种植像**茶叶**（对女性收入提高有利）
- 有的地区更多地种植像**果树**（对男性收入提高有利）
- 有的地区和以前一样主要**种植粮食**
- 我们依此可以分析**经济因素**和**女孩存活率**的关系。

# 自然实验与准实验

## 准实验

- 有时候我们想探讨的问题没法通过实验的办法来解决。
  - 读大学 VS 不读大学，对收入的影响有多大？
  - 婚前同居有助于婚姻稳定吗？
- 这时候我们只能去使用观测数据，而没办法使用实验数据。

# 自然实验与准实验

## 准实验

- 准实验是**对实验的“模仿”**，即通过观测数据和统计学方法，打造出与实施随机对照试验相类似的状态。
- “准实验”方法包括**匹配法、断点回归**、工具变量、双重差分等。



- 你所在的珠宝店目前共有100家店铺。其中**30家投放了广告，70家未投放广告**。从未投放广告的70家店铺中，**选出和投放广告的20家店铺情况相似的20家店铺**，将二者进行比较不就好了吗？这便是匹配法的思路。
- 所谓**匹配法**，即从对照组中选出和干预组非常接近的配对，使两组成为可比较组的方法。

如果干预组与实验组差异太大，则两者不可比较

	有广告 (30家店铺/干预组)	无广告 (70家店铺/干预组)
店长年龄	50岁	30岁
地区人口	300万人	500万人
地区人均收入	4万元	6万元
销售额	100万	140万

需要使用匹配法使得两组是“可比的”

	有广告 (20家店铺/干预组)	无广告 (20家店铺/干预组)
店长年龄	45岁	44岁
地区人口	400万人	410万人
地区人均收入	4.6万元	4.5万元
销售额	120万	100万

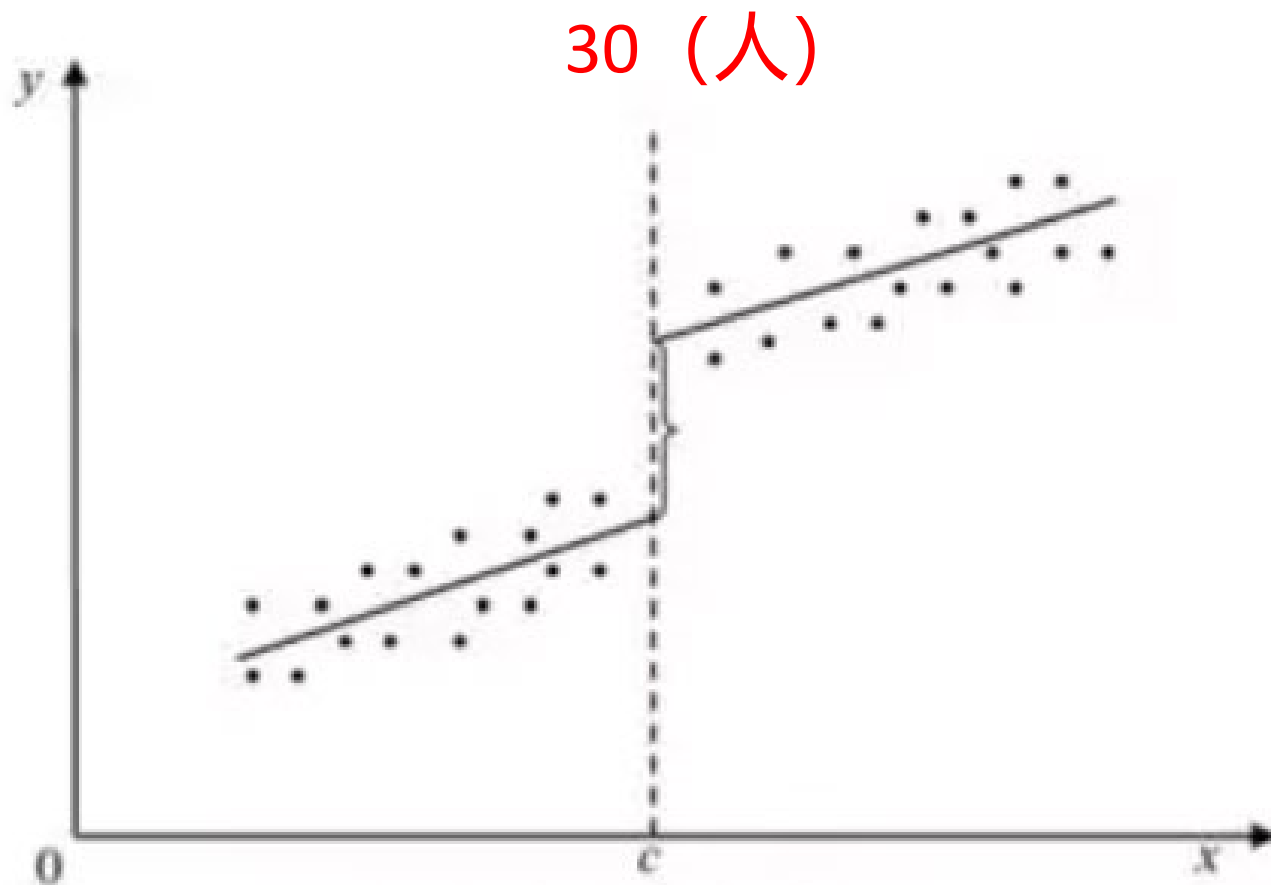
20万的因果效应

### 匹配法其它案例：读名校会增加未来收入吗？

- 上名校和上普通高校，未来收入会有所不同？
- 假设有两个人，他们正在申请美国的大学。他们都参加了美国的高考，成绩相近，都申请哈佛大学和马萨诸塞大学，而且同时被录取。但是由于各种原因，其中一个人选择读哈佛，另外一个人选择读马萨诸塞大学。
- 去比较这两人未来的收入，我们就可以知道上名校和普通高校对收入的影响。

- 就职于珠宝店的你还在为如何论证广告的效果而一筹莫展。这时，下属向你提出了一个方案：**只有员工人数30人以上（含30人）的大型店铺才在年底元旦商战期间投放广告。**
- 对此，你产生了一个想法：按照这个方案，员工人数29人的店铺不投放广告，而30人的店铺投放广告。不过，**29人的店铺和30人的店铺除了是否投放广告之外，应该没有其他特别显著的差异。**

销售额



店铺人数

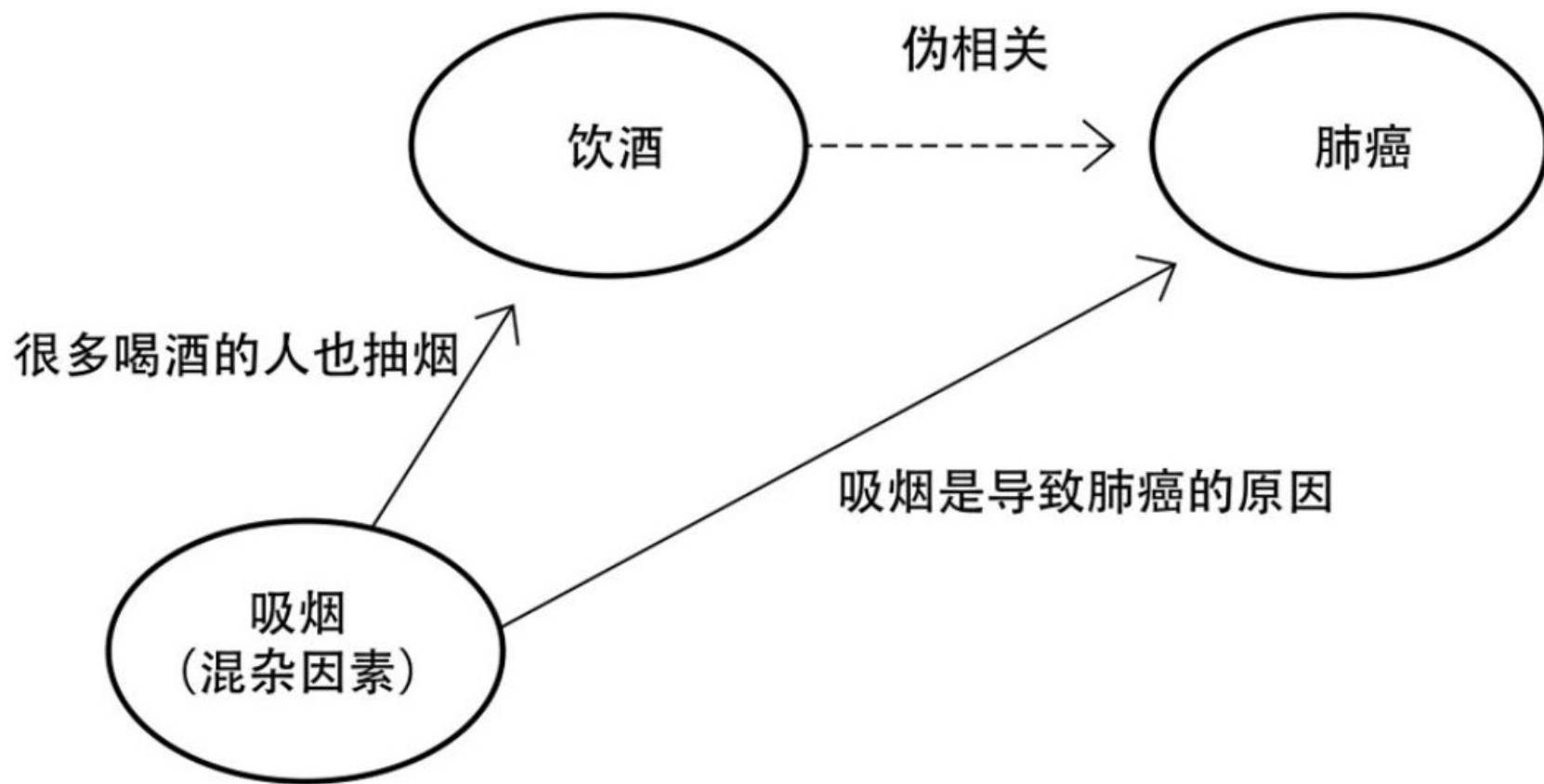
断点回归示意图

[https://blog.csdn.net/Claire\\_chen\\_jia](https://blog.csdn.net/Claire_chen_jia)

### 断点回归其它案例：上重点高中对高考成绩的帮助

- 想知道**上重点高中**和**上普通高中**的学生高考成绩差异有多大？
- 直接做比较是不公平的，也不可能去做随机分组实验。
- 假如重点高中录取分数线是既定的，那么**擦线上重点高中**的和**差一点上线**的学生应该能力是差不多的，他们的高考分数是可比较的。

# 回归分析



- 抽烟水平一样，饮酒程度不一样，得肺癌的风险一样吗？



## 回归分析的作用

- （多元）回归分析可以**排除混杂因素**的影响，并在此基础上评估原因与结果的关系。假如我们掌握所有混杂因素的数据（不过这种情况极为罕见），用多元回归分析也完全能够证明因果关系。

## 参考资料

- 《原因与结果的经济学》（[日] 中室牧子/ [日] 津川友介）
- 《大侦探经济学：现代经济学中因果推断革命》（李井奎）
- [经济学与大侦探 | 如果经济学家连股都不炒，那他们都在干什么呢 | 李井奎 一席第864位讲者\\_哔哩哔哩\\_bilibili](#)



谢谢！