第三讲 因果推断的基本逻辑

张开平 助理教授

2021年3月8日

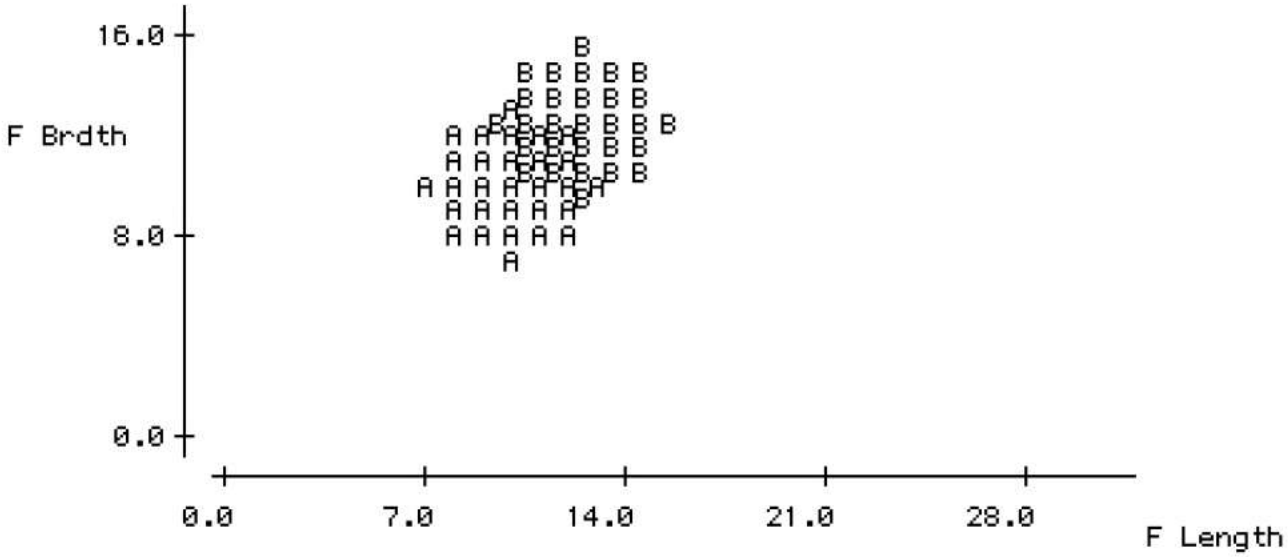


FIGURE 2 A=F Brdth vs.F Length B=M Brdth vs.M Length

The "data" of Figure 1 combined on the same set of axes; here the correlation r = 0.4.

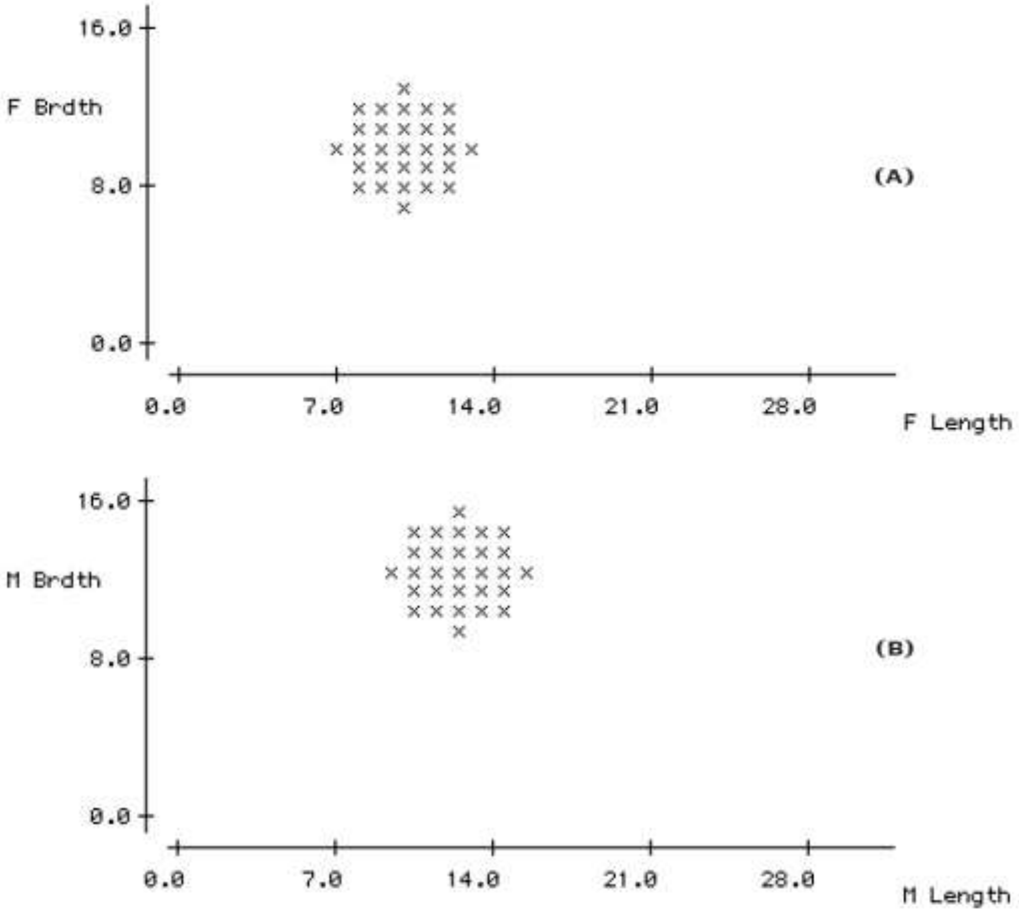

16.0
F Brdth
8.0
0.0
0.0
7.0
14.0
21.0
28.0
F Length
(A)
16.0
M Brdth
8.0
0.0
0.0
7.0
14.0
21.0
28.0
M Length
(B)

FIGURE 1

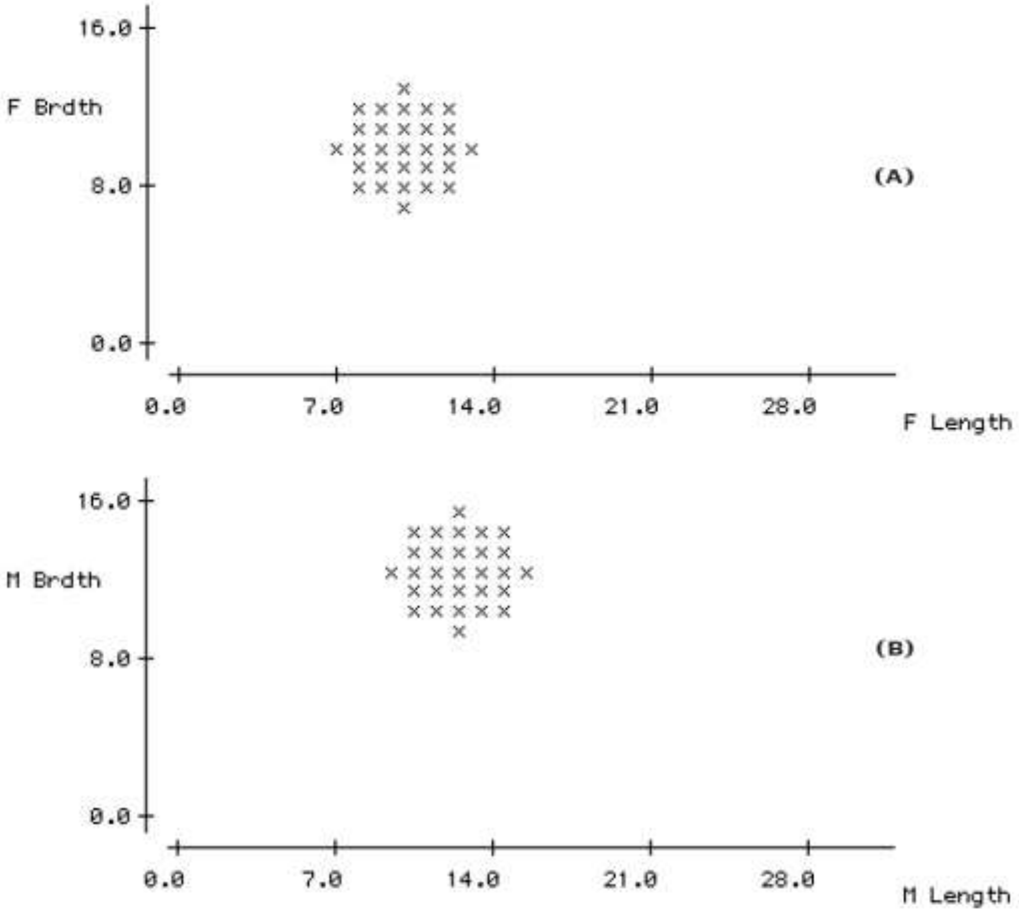


FIGURE 1

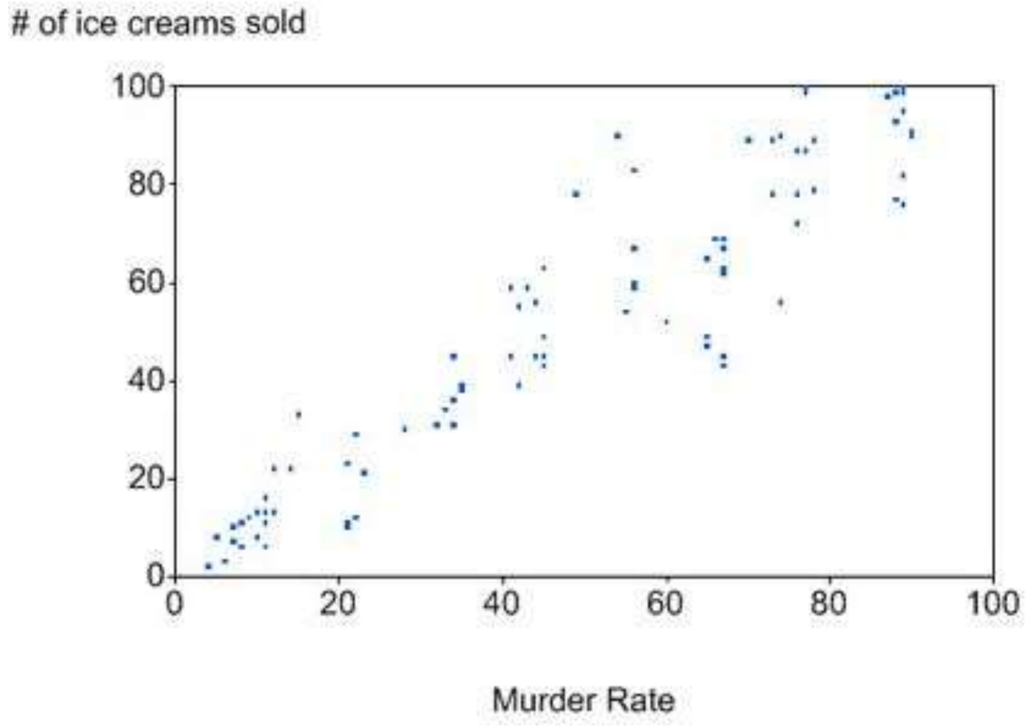
*Stylized "data" depicting the possible relationship between the breadth and length of female skulls (*top panel, *1A) and male skulls (*lower panel, *1B), given sex. In each case the correlation r = 0, but when the data are combined (Figure 2), the correlation is positive.*

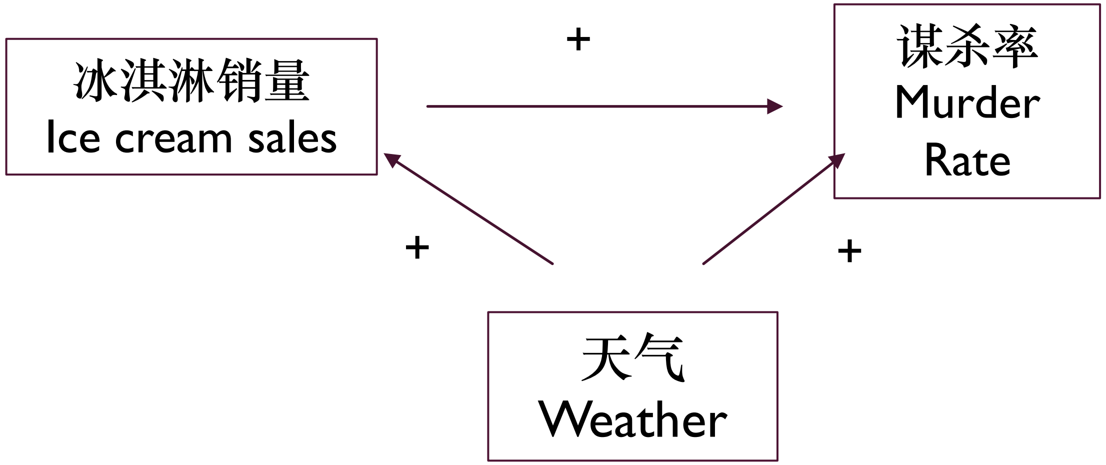
SPURIOUS CORRELATION 虚假相关

- 两个事实上毫不相关的变量却具有显著的相关系数
- 由数学家和生物统计学家Karl Pearson首先提出 (Pearson 1897; Pearson & Bramley Moore, 1899)
  - 当两个变量看似相关时，有可能由于他们都与第三个变量存在相关关系；而当第三个变量(confounding variable)被遗漏时，这两个变量就呈现出虚假相关关系
  - 在时间序列数据常见

Pearson, Karl. "Mathematical contributions to the theory of evolution—on a form of spurious correlation which may arise when indices are used in the measurement of organs." *Proceedings of the royal society of London* 60.359-367 (1897): 489-498.

Spurious Correlation





Confounding
variable

Another Example of Spurious Correlation: Redskins Rule

Year	Electoral vote result	Washington (Score)	Opponent (Score)	Washington Win or Lose?	Incumbent Party Win or Lose?	Rule upheld?	Popular vote winner
1932	Roosevelt defeats Hoover 472–59	Braves 19[Notes 1]	Staten Island Stapletons 6	win	lose	no	Roosevelt
1936	Roosevelt defeats Landon 523–8	Redskins 13[Notes 1]	Chicago Cardinals 10	win	win	yes	Roosevelt
1940	Roosevelt defeats Willkie 449–82	Redskins 37	Pittsburgh Steelers 10	win	win	yes	Roosevelt
1944	Roosevelt defeats Dewey 432–99	Redskins 14	Cleveland Rams 10	win	win	yes	Roosevelt
1948	Truman defeats Dewey & Thurmond 303–189–39	Redskins 59	Boston Yanks 21	win	win	yes	Truman
1952	Eisenhower defeats Stevenson 442–89	Redskins 23	Pittsburgh Steelers 24	lose	lose	yes	Eisenhower
1956	Eisenhower defeats Stevenson 457–73	Redskins 20	Cleveland Browns 9	win	win	yes	Eisenhower
1960	Kennedy defeats Nixon 303–219	Redskins 10	Cleveland Browns 31	lose	lose	yes	Kennedy
1964	Johnson defeats Goldwater 486–52	Redskins 27	Chicago Bears 20	win	win	yes	Johnson
1968	Nixon defeats Humphrey & Wallace 301–191–46	Redskins 10	New York Giants 13	lose	lose	yes	Nixon
1972	Nixon defeats McGovern 520–17	Redskins 24	Dallas Cowboys 20	win	win	yes	Nixon
1976	Carter defeats Ford 297–240	Redskins 7	Dallas Cowboys 20	lose	lose	yes	Carter
1980	Reagan defeats Carter 489–49	Redskins 14	Minnesota Vikings 39	lose	lose	yes	Reagan
1984	Reagan defeats Mondale 525–13	Redskins 27	Atlanta Falcons 14	win	win	yes	Reagan
1988	G. H. W. Bush defeats Dukakis 426–111	Redskins 27	New Orleans Saints 24	win	win	yes	G. H. W. Bush
1992	B. Clinton defeats G. H. W. Bush 370–168	Redskins 7	New York Giants 24	lose	lose	yes	B. Clinton
1996	B. Clinton defeats Dole 379–159	Redskins 31	Indianapolis Colts 16	win	win	yes	B. Clinton
2000	G. W. Bush defeats Gore 271–266	Redskins 21	Tennessee Titans 27	lose	lose	yes	Gore
2004	G. W. Bush defeats Kerry 286–251	Redskins 14	Green Bay Packers 28	lose	win	no[Notes 2]	G. W. Bush
2008	Obama defeats McCain 365–173	Redskins 6	Pittsburgh Steelers 23	lose	lose	yes	Obama
2012	Obama defeats Romney 332–206	Redskins 13	Carolina Panthers 21	lose	win	no	Obama
2016	Trump defeats H. Clinton 304–227	Redskins 27	Philadelphia Eagles 20	win	lose	no	H. Clinton
2020	Biden defeats Trump 306–232	Washington 25	Dallas Cowboys 3	win	lose	no[Notes 2]	Biden

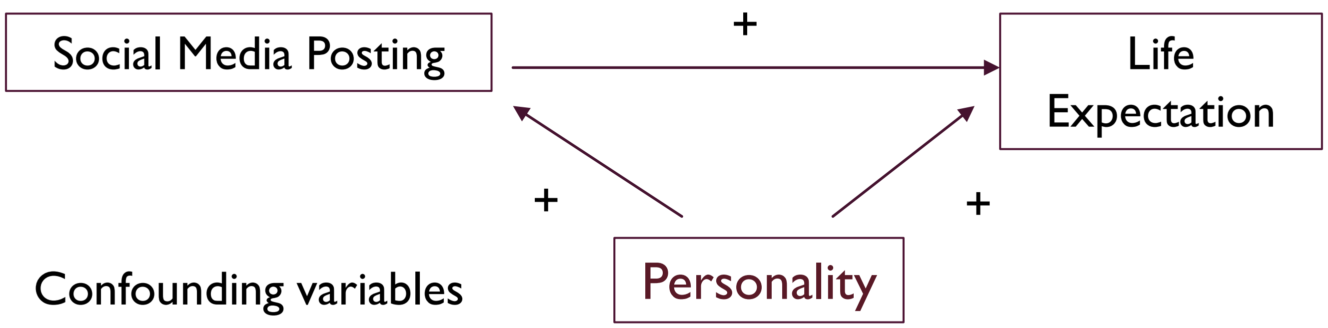
“转发：看来我得多发朋友圈”

国际顶刊《PNAS》：爱发朋友圈的人，更容易长寿

2019年11月15日　生活数据

近几十年来，智能手机和网络的普及率越来越高，与此同时，“朋友圈”应运而生。
在这个朋友圈里，有人十分活跃，而也有些人是“国家级潜水运动员”，那么，在朋友圈的活跃与否对我们有什么影响呢?
2016年10月31日，著名的美国科学院院刊《PNAS》在线发表的一篇研究表明，在控制了其他影响因素后，朋友圈的活跃度与死亡率呈现负相关，也就是说，**在社交网络上越活跃的人，死亡的风险越低!**

https://www.pnas.org/content/pnas/113/46/12980.full.pdf



Confounding variables

SOCIAL SCIENTISTS ASK

- 如果x会怎样? (What-if question)
  - 精准扶贫政策能够消除贫困吗？ (policy implications)
  - 接受研究生教育能够增加收入吗？(individual decisions)
  - 民主转型促进经济发展吗？ （theoretical concern)

Simpson's Paradox

Treatment	Mortality	Mild	Severe
A	16% (240/1500)	15% (210/1400)	30% (30/100)
B	19% (105/550)	10% (5/50)	20% (100/500)
	E(Y|T)	E(Y|T,C=0)	E(Y|T,C=1)

1400/1500 * 0.15 + 100/1500 * 0.3 = 0.16
50/550 * 0.1 + 500/550 * 0.2 = 0.19

辛普森悖论：不同组出现的趋势在整合数据后呈现相反的趋势或原有趋势消失。

Simpson's Paradox

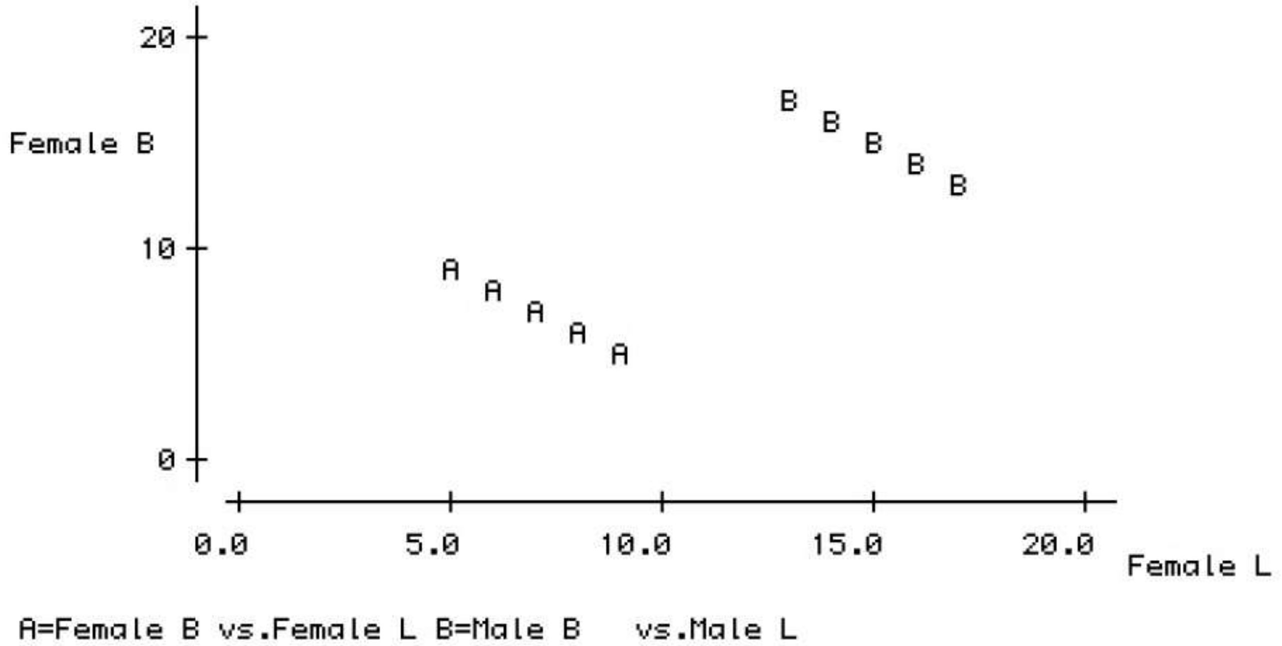


FIGURE 3

A more extreme example, where separately for male and female the correlations are $r = -1$, and together the correlation is $r = 0.8$.

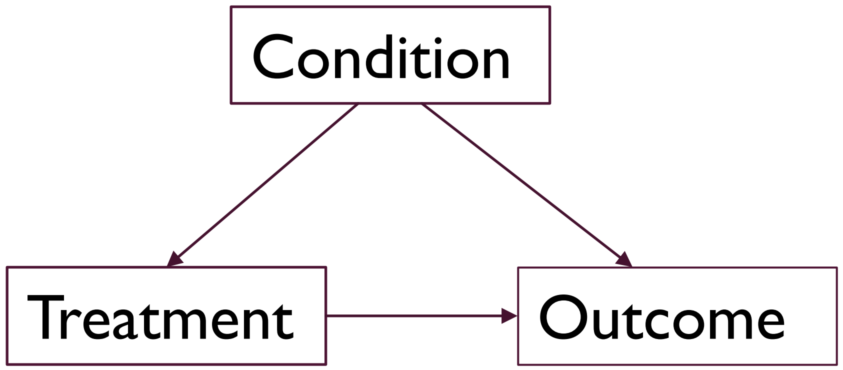
中药是冠状病毒的有效疗法吗?

媒体：吃中药的新冠肺炎患者治愈率高，吃西药的患者治愈率低，所以中药更有效。

问题：吃中药的患者都是轻症患者，吃西药的患者是重症患者。而轻症患者多数能够自愈。

轻症患者 $\xrightarrow{\text{方舱隔离+中药}}$ 治愈率高

重症患者 $\xrightarrow{\text{西药}}$ 治愈率低



This is not a causal graph!

IS CORRELATION GOOD ENOUGH?

- Association: 1 unit increase in X is associated with Z amount of increase in Y, holding control variables constant.

	Sample Size	Mean GDP (Billion RMB)	Std. Error
实施扶贫政策的地区	422	0.21	0.014
其他地区	1292	3.67	0.003

扶贫政策造成了贫困?

	Sample Size	Mean GDP (Billion RMB)	Std. Error
实施扶贫政策的地区	422	0.21	0.014
其他地区	1292	3.67	0.003

扶贫

贫困

反向因果

肥胖能够传染吗?

HOME \ CULTURE \ HEALTH

Blame your location: Obesity spreads through communities like a contagious disease

By Laura Kelly - The Washington Times - Wednesday, January 24, 2018

Obesity is deadly, epidemic and ... contagious?

A study in the Journal of the American Medical Association Pediatrics shows that being around obese people for extended periods makes it more likely that you will become obese.

The study also suggests that people who are surrounded by a community and environment that promote healthy living will mirror that lifestyle.

Follow Us

SIGN UP FOR OUR DAILY NEWSLETTERS

Submit

FRONT PAGE PODCAST

February 21, 2020

RECOMMENDED

Quiz: How much do you know about U.S. presidents?

Nuclear arms race among U.S., Russia, China could heat up as Obama era treaty nears expiration

Quiz: Can you match the romantic movie quote with the film?

Changing definition of 'sex' since ERA was written poses thorny legal issues

https://jamanetwork.com/journals/jamapediatrics/article-abstract/2668504?redirect=true

肥胖传染吗?

- 有肥胖的朋友使你更容易肥胖吗?
  - 相似的人更容易成为朋友
  - 肥胖的人更可能是经济上紧张，街区充满了麦当劳而没有昂贵的健康食物，更没有钱健身，更爱吃某些高热量食物

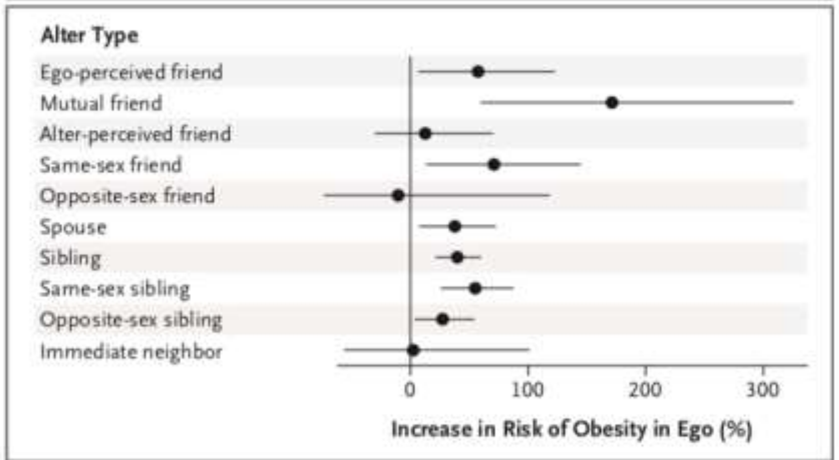


Figure 4. Probability That an Ego Will Become Obese According to the Type of Relationship with an Alter Who May Become Obese in Several Subgroups of the Social Network of the Framingham Heart Study.

The closeness of friendship is relevant to the spread of obesity. Persons in closer, mutual friendships have more of an effect on each other than persons in other types of friendships. The dependent variable in each model is the obesity of the ego. Independent variables include a time-lagged measurement of the ego's obesity; the obesity of the alter; a time-lagged measurement of the alter's obesity; the ego's age, sex, and level of education; and indicator variables (fixed effects) for each examination. Full models and equations are available in the Supplementary Appendix. Mean effect sizes

反向因果问题 RESERVE CAUSALITY PROBLEM (SELECTION BIAS)

- 住院使人身体更糟糕？
- 抽烟导致抑郁？
- 观看竞选广告能够使人们更容易参与投票？
- 男性养的狗侵略性更强？

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	2.79	0.014
No Hospital	90049	2.07	0.003


POLITICS PLAN YOUR VACCINE COVID-19 U.S. NEWS OPINION BUSINESS WORLD TECH

Dogs Walked by Men More Aggressive

Dogs being walked by men are four times more likely to threaten and bite other dogs and dogs on a leash are more likely to act aggressively than dogs off the leash.

By Jennifer Viegas

Sponsored Stories

CORRELATION IS NOT CAUSATION!

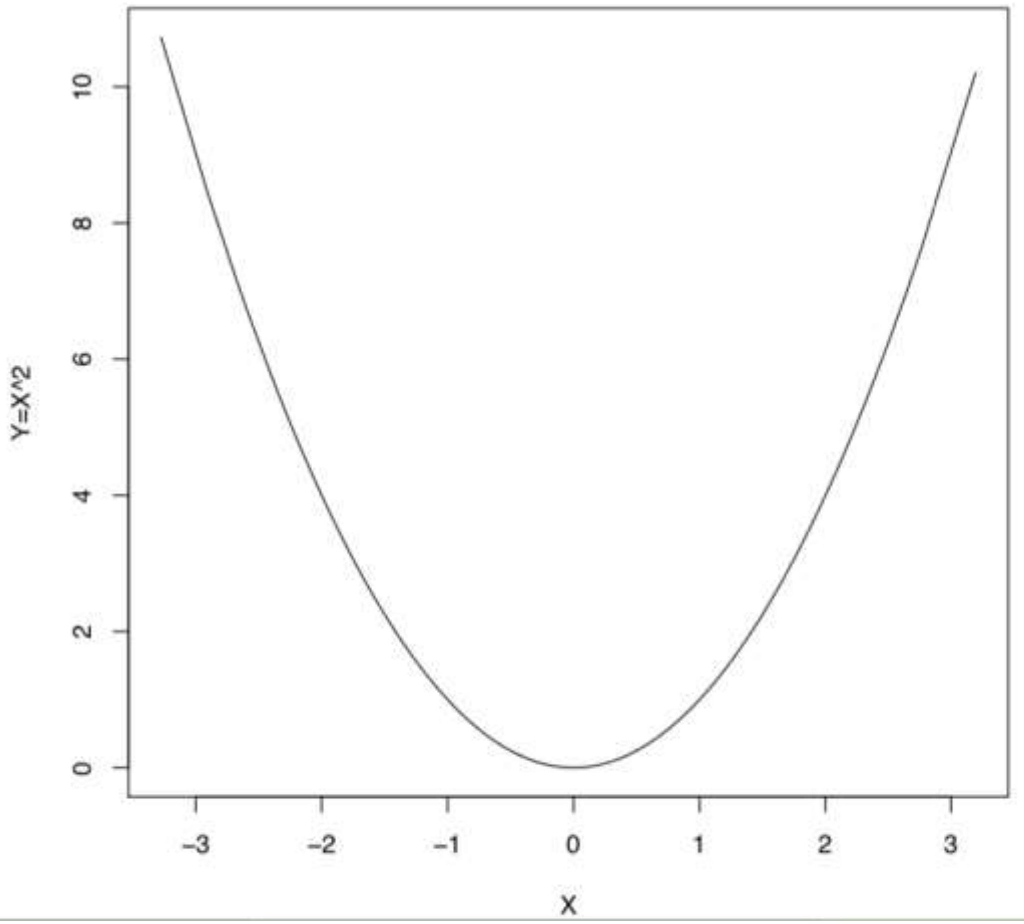
- OLS回归最关键的假设“残差与因变量不存在相关关系”经常并不满足，在政策领域尤其显著。
  - 遗漏变量 omitted variable bias
  - 内生性（反向因果）endogeneity (reverse causal relation)

- 政治学传统上采用田野调查、案例研究、问卷调查、统计分析等方法，然而由观察性数据(observational data) 得出的变量之间普遍存在自我选择、遗漏变量以及反向因果等问题，成为实现因果推断难以逾越的鸿沟。

- 基于观察性数据借助回归分析进行所谓的“因果推论”在本质上仅仅只能发现相关性关系，但“相关”并不等于“因果” 。

Correlation $\neq$ Causation

It's extremely hard to make causal inference with observational data!



X和Y可能存在因果关系，但并不具备相关性。

存在相关性不代表存在因果关系；反之，存在因果关系也不一定具有相关性。

二、因果推断的框架：鲁宾因果模型

因果推断：从自然哲学思辨到“反事实”模型

- 休谟：“事物后跟着另一个事物，所有类似于第一个对象的事物后面也会跟着类似第二个对象的事物。换言之，若第一个事物不发生，那么后面事物（即第二个对象）也不存在。”
  - 因果事件在时空上有毗连性（contiguity）；时间的顺序性（succession），即原因发生在结果之前；原因与结果之间存在必然联系（necessary connection），即总是不可分离、相伴而生
- 密尔逻辑：求同法、求异法、求同求异并用法、剩余法以及共变法

问题：相关社会事件一旦发生“事实”就只有一个，无法对同一时空条件下的原因与事实重现。

因果推断：从自然哲学思辨到“反事实”模型

- 二十世纪七十年代，哲学家刘易斯(David Lewis)转变了“如果原因不发生，那么结果也不会发生”的休谟主义思维，将其发展为反事实分析(counterfactual analysis of causation)，使因果推论从哲学思辨转变为可操作的分析框架。

  - 界定两个事物的因果关系无须同时观测到原因X与结果Y的（相继）发生；相反，如果在一个各方面都相似的情境下，X没有发生，Y就没有发生，那么便可界定二者间的因果关联。

问题：在现实中寻找一个各方面特征都相似且X没有发生的“反事实”存在并不容易实现，且X没有发生常伴随着其他因素的变化。因而，很难界定若X没有发生，结果有何不同。

实验

- 实验法通过人为的干预来构建可比较的情境，认为“没有操纵就没有因果（no causation without manipulation）”（Holland 1986）。
- 莫顿：“当研究者通过主动操纵数据的生成过程的相关因素,从而介入到数据的生成过程（data generate process，DGP）的便是实验。”（Morton & Williams 2010)

潜在结果模型 POTENTIAL OUTCOME MODEL

- 由著名统计学家唐纳德·鲁宾提出，又称“Rubin Causal Model”, 是目前社会科学领域因果推论最重要的理论模型。

Donald Rubin

POTENTIAL OUTCOME MODEL （RUBIN CAUSAL MODEL)

- 每个个体(unit)有两种潜在的可能结果

$$potential\ outcome = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}.$$

- 因果关系即某个体接受了实验干预与其没有接受实验干预的潜在结果的差异

$$\delta = Y_{1i} - Y_{0i}$$

Potential Outcome Model（Rubin Causal model)

- 平均干预效应

$$Average\ Treatment\ Effect(ATE) = E(Y_{i1})\ - E(Y_{i0})$$

- 干预者平均干预效应

$$Average\ Treatment\ Effect\ on\ the\ Treated(ATT) = E(Y_{i1}|D_i = 1)\ - E(Y_{io}|D_i = 1)$$

- ATE = ATT吗？

干预效应对于主动寻求干预的人可能更大，ATE不等与ATT。

- 接受大学教育能够增加多少收入？

我们能够观测到的：

	收入	收入
高中学历	3000	
大学学历		8000

练习

- 问题：接受大学教育能够增加多少收入？

假设我们知道：

	收入	收入
高中学历	3000	5000
大学学历	4000	8000

那么ATE是多少？ ATT是多少？

练习

- 问题：接受大学教育能够增加多少收入？

	平均收入	平均收入	Difference
高中学历	3000	5000	2000
大学学历	4000	8000	4000

Naïve Comparison = 5000　ATE = 3000　ATT = 4000

8000 - 4000

$$\underbrace{E\left[\mathrm{Y}_i|\mathrm{D}_i=1\right]-E\left[\mathrm{Y}_i|\mathrm{D}_i=0\right]}_{\text{Observed difference in average health}} = \underbrace{E\left[\mathrm{Y}_{1i}|\mathrm{D}_i=1\right]-E\left[\mathrm{Y}_{0i}|\mathrm{D}_i=1\right]}_{\text{average treatment effect on the treated}}$$

$$+\underbrace{E\left[\mathrm{Y}_{0i}|\mathrm{D}_i=1\right]-E\left[\mathrm{Y}_{0i}|\mathrm{D}_i=0\right]}_{\text{selection bias}}$$

8000 - 3000

+ 4000 - 3000

练习 2

Treatment	Yo: Outcome had they not receive treatment	Y1: Outcome had they receive treatment	Observed Outcome
0	6	5.5	6
1	7	6.5	6.5
0	7	6	7
1	9	8	8
1	8.5	8	8
0	7.5	7	7.5
1	10	9	9
0	8	7	8

练习 2

Treatment	Observed Outcome
0	6
1	6.5
0	7
1	8
1	8
0	7.5
1	9
0	8

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{\text{Observed difference in average health}} = \underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]}_{\text{average treatment effect on the treated}}$$

$$+\underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{selection bias}}$$

然而

- 贺兰：“进行因果推论面临的根本性问题便是：我们无法同时观察到同一个单元接受干预与不接受干预的两种状态。因而无法观察到干预对于受试单元的影响。”

Holland，P．*Statistical and Causal Inference*．Journal of the American Statistical Association, 1986,81(396) : 945-960.

随机实验解决该根本问题

- 由于实验干预是随机分配的，对照组与实验组之间在可观测和非可观测的特点上均没有系统性差异，因而具有相同的期望。

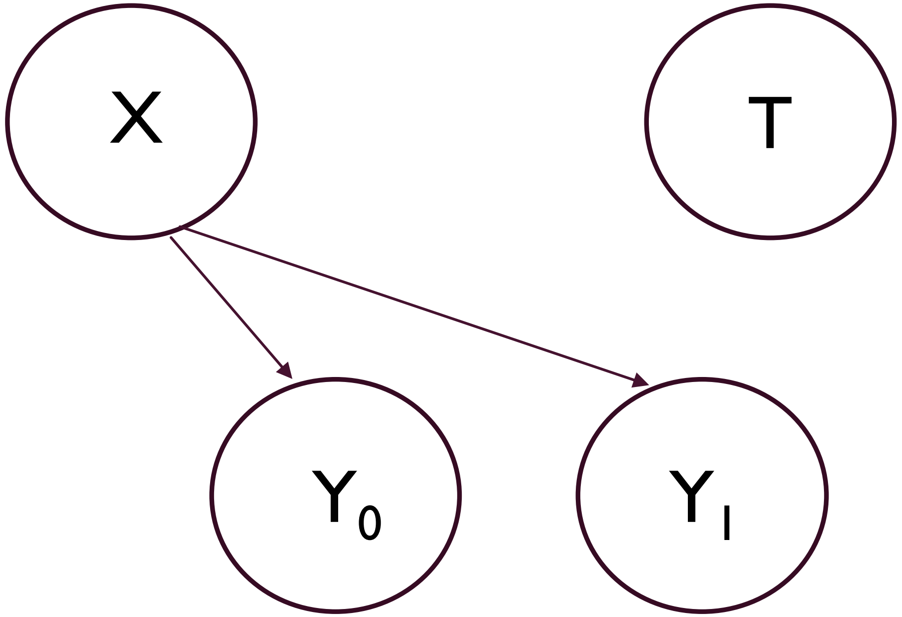
$$E(Y_{i1}|D_i = 1) = E(Y_{i1}|D_i = 0) \qquad E(Y_{i0}|D_i = 0) = E(Y_{i0}|D_i = 1)$$

- 对照组和实验组可以作为彼此的“反事实”存在。平均干预效应（average treatment effect）便是对照组与实验组的期望之差，即可由样本均值之差估算而来

$$Average\ Treatment\ Effect(ATE) = E(Y_{i1}|D_i = 1) - E(Y_{io}|D_i = 0)$$

假设：干预分配机制的可忽略性（IGNORABILITY）

Gender, age,
SES, weight,
diet, etc.

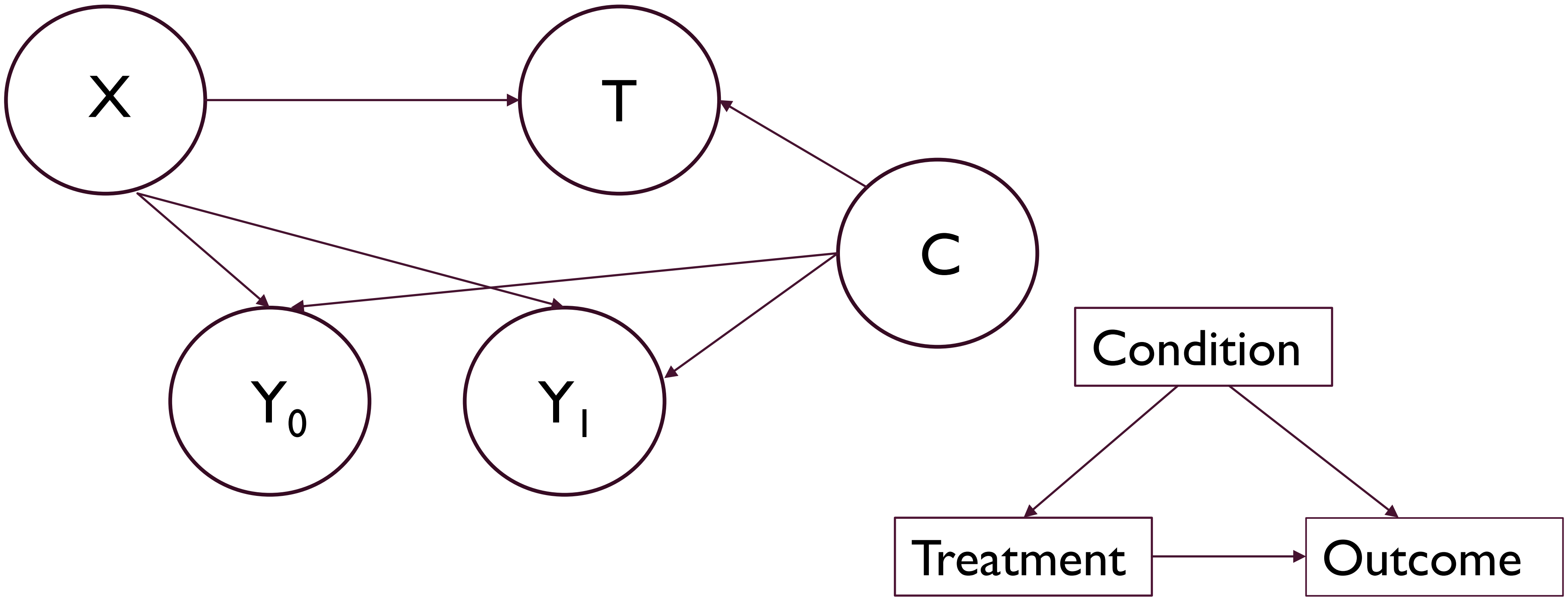


$$(\mathrm{Y}_0, \mathrm{Y}_1 \perp T \mid X)$$

- Test for balance to check covariate balance across experimental conditions

违背假设（NO IGNORABILITY）

Gender, age,
SES, weight,
diet, etc.



假设条件 ASSUMPTIONS

1. 排他性限制 Exclusion Restriction
   - Assignment to the treatment group only affects outcomes insofar as subjects receive the treatment

霍桑效应

假设条件 ASSUMPTIONS

1. 排他性限制 Exclusion Restriction
   - Assignment to the treatment group only affects outcomes insofar as subjects receive the treatment
   - 每个单元的结果只与干预（发生或不发生）有关，与其他原因无关。
   - 无安慰剂效应 (No placebo effect)、实验者效应 (experimenter effect)
   - 解决方法：安慰剂组，双盲实验

假设条件 ASSUMPTIONS

2. **稳定的单元处理值假设** Stable Unit Treatment Value Assumption (SUTVA)

- the treatment of *i* only affects outcomes of *i*;
- 对某个单元的干预只影响该单元的效果
- 无溢出效应(No spillover effect)
- 解决方法：在不同组变化干预强度
- 违背假设条件产生估计偏差（biased estimation)

假设条件 ASSUMPTIONS

- 单向性 Monotonicity
  - 实验组接受干预的概率不小于对照组接受干预的概率
    - 只有实验组才能接触到某个干预
- 随机分配的非零效果 Nonzero causal effect of assignment on treatment
  - 至少一些被随机分配到实验组的人实际接受了干预

NON-COMPLIANCE 不服从

- 人们并不总是服从研究者的安排，导致实际接受干预的人与随机分配应该接受干预的人不同
  - 例如，实验组的人没有联系到(non-response)；对照组的人通过其他方式接受了干预（新药）
  - 实地实验面临的重要挑战

Experiment Assignment （$\boldsymbol{Z_i \neq D_i}$）

实际接受干预情况	实验组 （$Z_i = 1$）	对照组（$Z_i = 0$）
Treated （$T_i = 1$）	$D_1 = 1$	$D_0 = 1$
Not treated （$T_i = 0$）	$D_1 = 0$	$D_0 = 0$

Non-compliers

Non-compliers

Compliers

NON-COMPLIANCE

- Intent-to-treat Effect(ITT) 试图干预效果
  - 可以用来评估一个政策项目
  - 较低的ITT说明ATE(平均干预效果）很低，或者实验的实施干预水平较差。

$$\text{ITT} = E(Y_{i1}|Z_i = 1) - E(Y_{io}|Z_i = 0)$$

Angrist, Imbens, & Rubin (1996)

LOCAL AVERAGE TREATMENT EFFECT (LATE)

- Local Average Treatment Effect (LATE) 实际接受干预者平均干预效应
  - Average treatment effect for the compliers if and only if it is assigned to the treatment condition
- 当存在不服从时，对于干预效应的估计只限于服从者人群，推广至一般人群需要额外假设，即服从者与一般人群的干预效应相似。
  - When non-complier = 0, LATE = ATT

ITT

$$LATE = \frac{E(Y_1 - Y_0)}{p(D_1|T_i = 1)}$$

总结

- 随机分配机制剔除了不可观测的相关因素的干扰，从而解决了“因果推论的根本难题”。
- 将随机实验与潜在结果有机结合，为实验探寻因果关系提供了重要的理论指导。
- 非实验方法的因果推断共同原理是在“潜在结果模型”的指导下寻找可比较的对照组
  - 匹配法 Matching
  - 中断时间序列分析 Interrupted Time Series
  - 断点回归 Regression Discontinuity