


THEME 2 : FROM RANDOMIZATION TO RANDOMNESS

- CHOCOLATES & NOBEL PRIZES
- EXPERIMENTAL & OBSERVATIONAL DATA
- CONFOUNDER
- DEATH PENALTY & RACIAL BIAS

- RECALL : ONE OF GOALS OF LEARNING FROM DATA
 - ↳ CAUSALITY
(INTERVENTION ANALYSIS)
- THE GOLD STANDARD OF CAUSAL INFERENCE IS RANDOMIZED CONTROL TRIAL (RCT)
 - ↳ EXPERIMENT

AoS

Event	Percentage in 10,267 people allocated placebo	Percentage in 10,269 people allocated statin	% (relative) risk reduction in those allocated statins
Heart attack	11.8	8.7	27%
Stroke	5.7	4.3	25%
Death from any cause	14.7	12.9	13%

Table 4.1

The outcomes at five years in the Heart Protection Study, according to treatments allocated to patients. The absolute reduction in the risk of a heart attack was $11.8 - 8.7 = 3.1\%$. So out of 1,000 people taking a statin, around 31 heart attacks were prevented – this means that around 30 people had to take a statin for five years to prevent one heart attack.

Second, since this is a statistical or stochastic world, we need to intervene more than once in order to amass evidence.

And that leads us naturally to a delicate topic: conducting medical experiments on large numbers of people. Few of us might relish the idea of being experimented on, especially when life and death are concerned. Which makes it all the more remarkable that thousands of people have been willing to be part of huge studies in which neither they nor their doctor knew which treatment they would end up getting.

Do statins reduce heart attacks and strokes?

Every day I take a little white pill – a statin – because I have been told it lowers cholesterol and so reduces the risk of heart attacks and strokes. But what is its effect on me personally? I am almost certain that it causes my low-density cholesterol (LDL) to drop, since I was told it reduced soon after I started taking the tablets. This drop in LDL

is a direct, essentially deterministic effect that I can assume is caused by the statin.

But I will never know if this daily ritual does me any good in the long run; it depends on which of my many possible future lives actually occurs. If I never have a heart attack or a stroke, I will have no idea whether I would have never had one even if I had not taken the tablets, and all this pill-popping for years was a waste of time. If I do have a heart attack or a stroke, I will not know if this event was delayed by taking the statin. All I can ever know is that, on average, it benefits a large group of people like me, and this knowledge is based on large clinical trials.

The purpose of a clinical trial is to carry out a ‘fair test’ that properly determines causation and estimates the average effect of a new medical treatment, without introducing biases that could give us the wrong idea of its effectiveness. A proper medical trial should ideally obey the following principles:

1. *Controls:* If we want to investigate the effect of

AoS

learn . .

The main recent innovation in randomized experimentation concerns ‘A/B’ testing in web design, in which users are (unknowingly) directed to alternative layouts for web pages, and measurements made of time spent on pages, click-throughs to advertisements, and so on. A series of A/B tests can rapidly lead to an optimized design, and the huge sample sizes mean that even small, but still potentially profitable, effects can be reliably detected. This has meant an entirely new community has had to learn about trial design, including the perils of making multiple comparisons that we will come to in Chapter 10.

← TECH
INDUSTRY

SCIENCE

NASA's Study of Astronaut Twins Creates a Portrait of What a Year in Space Does to the Human Body

Wide-ranging research compares astronaut Scott Kelly to his earthbound twin brother, Mark



Maddie Burakoff

April 11, 2019

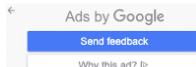


Identical twin astronauts, Scott and Mark Kelly, are subjects of NASA's Twins Study. Scott (right) spent a year in space while Mark (left) stayed on Earth as a control subject. NASA

Astronaut's gene expression no longer same as his identical twin, NASA finds

By Susan Scutti, CNN

Updated 6:10 PM EDT, Thu March 15, 2018



Space travel altered Kelly's chromosomes

00:57 - Source: CNN

vodafone 14:45

4G



bbc.co.uk/ideas



BBC



IDEAS

Browse

My Ideas



Is a vegan diet healthier than eating meat and dairy?

7:06 | 613.4K VIEWS

Identical twins Hugo and Ross Turner went on a 12-week diet with a twist: Hugo turned vegan, Ross continued to eat meat. What happened?



- RANDOMIZATION
 - COIN
 - RANDOM NUMBERS

Previous Examples

- ↳ STATION
- ↳ NASA
- ↳ BBC



EXPERIMENTAL DATA

Let's now talk about ORGANIZATIONAL DATA

This cautious attitude has a long pedigree. When Karl Pearson's newly developed correlation coefficient was being discussed in the journal *Nature* in 1900, a commentator warned that 'correlation does not imply causation'. In the succeeding century this phrase has been a mantra repeatedly uttered by statisticians when confronted by claims based on simply observing that two things tend to vary together. There is even a website that automatically generates idiotic associations, such as the delightful correlation of 0.96 between the annual per-capita consumption of mozzarella cheese in the US between 2000 and 2009, and the number of civil engineering doctorates awarded in each of those years.²

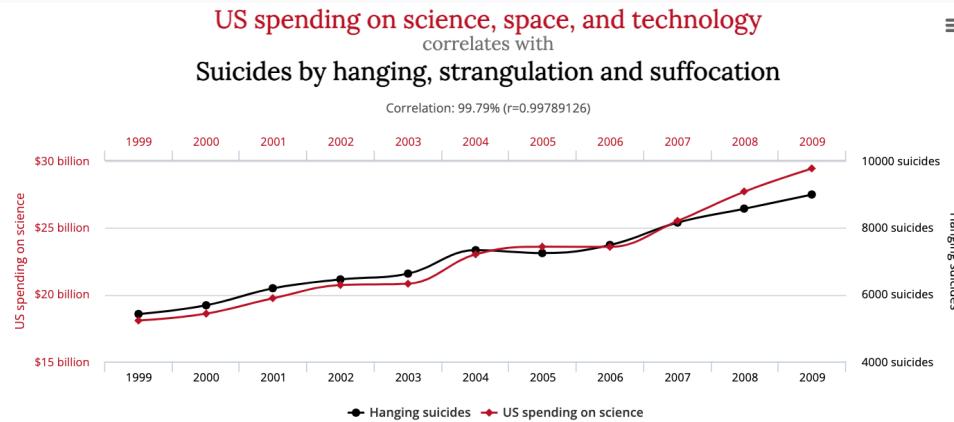
Spurious correlations

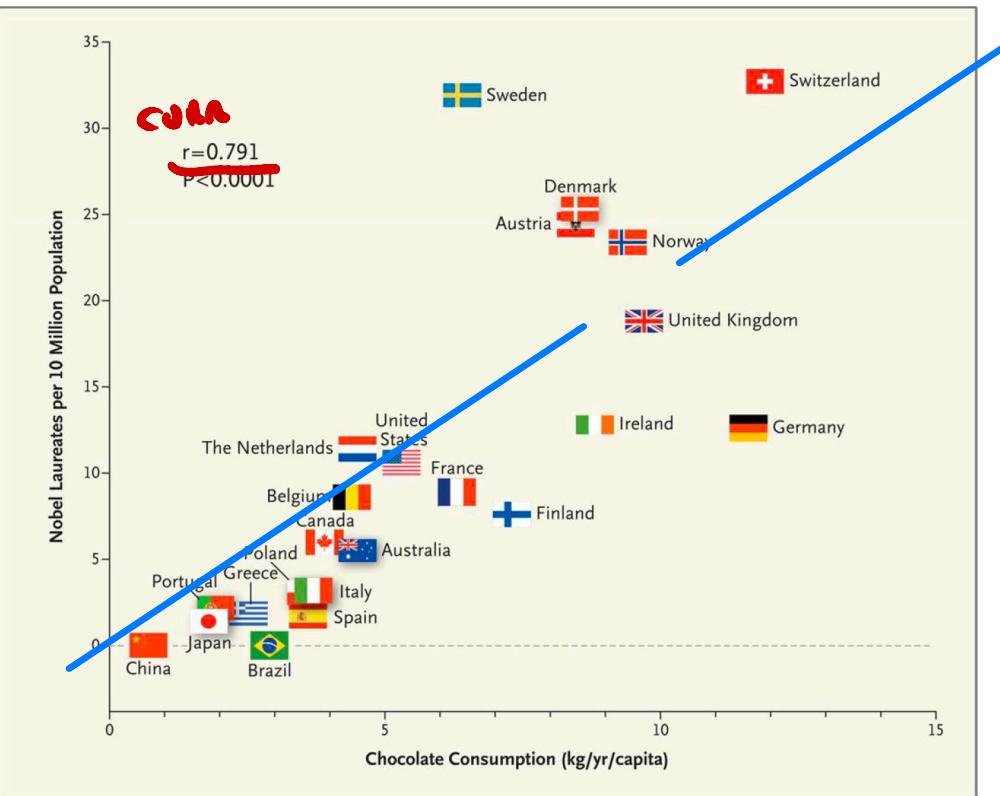


Now a ridiculous book!

- Spurious charts
- Fascinating factoids
- Commentary in the footnotes

[Amazon](#) | [Barnes & Noble](#) | [Indie Bound](#)





F. H. Messerli: Chocolate Consumption, Cognitive Function, and Nobel Laureates, N Engl J Med 2012

A LITTLE MATH ON CAUSALITY

X : "TREATMENT"

Y : OUTCOME

Z : CONFFOUNDERS

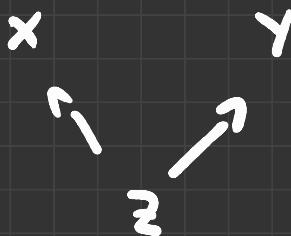
MATHEMATICAL LANGUAGE FOR CAUSALITY

$X \rightarrow Y$

X CAUSES Y

CAUSAL GRAPHS

↳ BAYESIAN NETWORK



Z CONFOUNDER

LUNG CANCER CONTROVERSY

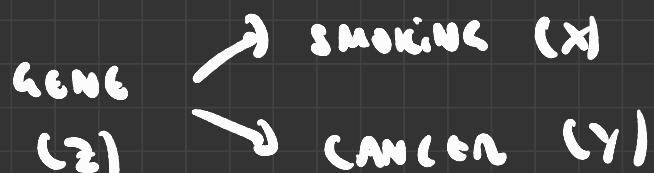
en.wikipedia.org/wiki/Ronald_Fisher

Gmail YouTube Maps

During this time he also worked on mouse chromosome mapping; breeding the mice in laboratories in his own house.^[57]

Fisher publicly spoke out against the 1950 study showing that smoking tobacco causes lung cancer, arguing that correlation does not imply causation.^{[58][59][60][61][62][63]} To quote his biographers Yates and Mather, "It has been suggested that the fact that Fisher was employed as consultant by the tobacco firms in this controversy casts doubt on the value of his arguments. This is to misjudge the man. He was not above accepting financial reward for his labours, but the reason for his interest was undoubtedly his dislike and mistrust of puritanical tendencies of all kinds; and perhaps also the personal solace he had always found in tobacco."^[5] Others, however, have suggested that his analysis was biased by professional conflicts and his own love of smoking.^[64]

FISHER'S POINT



SIMPSON'S PARADOX

AoS

	Women			Men		
	Applied	Accepted	%	Applied	Accepted	%
Computer Science	26	7	27%	228	58	25%
Economics	240	63	26%	512	112	22%
Engineering	164	52	32%	972	252	26%
Medicine	416	99	24%	578	140	24%
Veterinary Medicine	338	53	16%	180	22	12%
TOTAL	1,184	274	23%	2,470	584	24%

Table 4.2

Illustration of Simpson's Paradox using admission data for Cambridge in 1996. Overall, the acceptance rate was higher for men. But in each subject the acceptance rate was higher for women.

X : GENDER

Y : ADMISSION
DECISION

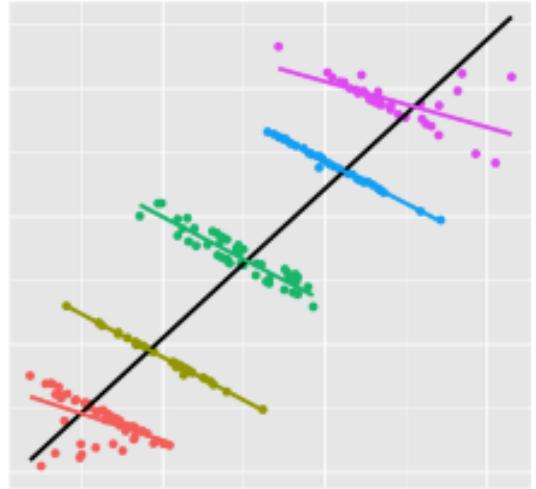
Z : DEGREE
APPLIED FOR

FLORIDA
STATE IN
US

Defendant's race	DP	No DP	Total	% DP
Caucasian	53	430	483	11%
African American	15	176	191	7.9%
Total	68	606	674	

Victim's race	Defendant's race	DP	No DP	Total	% DP
Caucasian	Caucasian	53	414	467	11.3%
Caucasian	African American	11	37	48	22.9%
African American	Caucasian	0	16	16	0%
African American	African American	4	139	143	2.8%
Total		68	606	674	

X : RACE of
Defendant
Y : DP
Z : RACE of
victim



A FEW CONSIDERATIONS FROM ANALYTICAL RCT

i. SIGNAL & NOISE

INFER TO WHAT EXTENT
THE OBSERVED DIFF.
ARE DUE TO TREATMENT
(SIGNAL) OR DUE
RANDOM VARIATION
(NOISE)

Event	Percentage in 10,267 people allocated placebo	Percentage in 10,269 people allocated statin	% (relative) risk reduction in those allocated statins
Heart attack	11.8	8.7	27%
Stroke	5.7	4.3	25%
Death from any cause	14.7	12.9	13%

Table 4.1

The outcomes at five years in the Heart Protection Study, according to treatments allocated to patients. The absolute reduction in the risk of a heart attack was $11.8 - 8.7 = 3.1\%$. So out of 1,000 people taking a statin, around 31 heart attacks were prevented – this means that around 30 people had to take a statin for five years to prevent one heart attack.

ii . UNCERTAINTY QUANTIFICATION

↳ RANGE OF PLAUSIBLE TREATMENT EFFECTS ?
(CONFIDENCE INTERVAL)

iii DESIGN OF EXPERIMENT

↳ REINFORCEMENT LEARNING

IV.

SUBGROUP ANALYSIS

↳ PERSONALIZED MEDICINE

V.

DEALING WITH PEOPLE

↳ Dropout

↳ REAUDITION