

Scientific studies and confounding

Data Science in a Box

datasciencebox.org



Scientific studies



Scientific studies

Observational

- Collect data in a way that does not interfere with how the data arise ("observe")
- Establish associations

Experimental

- Randomly assign subjects to treatments
- Establish causal connections



What type of study is the following, observational or experiment? What does that mean in terms of causal conclusions?

Researchers studying the relationship between exercising and energy levels asked participants in their study how many times a week they exercise and whether they have high or low energy when they wake up in the morning.

Based on responses to the exercise question the researchers grouped people into three categories (no exercise, exercise 1-3 times a week, and exercise more than 3 times a week).

The researchers then compared the proportions of people who said they have high energy in the mornings across the three exercise categories.



What type of study is the following, observational or experiment? What does that mean in terms of causal conclusions?

Researchers studying the relationship between exercising and energy levels randomly assigned participants in their study into three groups: no exercise, exercise 1-3 times a week, and exercise more than 3 times a week.

After one week, participants were asked whether they have high or low energy when they wake up in the morning.

The researchers then compared the proportions of people who said they have high energy in the mornings across the three exercise categories.



Case study: Breakfast cereal keeps girls slim



Girls who ate breakfast of any type had a lower average body mass index (BMI), a common obesity gauge, than those who said they didn't. The index was even lower for girls who said they ate cereal for breakfast, according to findings of the study conducted by the Maryland Medical Research Institute with funding from the National Institutes of Health (NIH) and cereal-maker General Mills. [...]

The results were gleaned from a larger NIH survey of 2,379 girls in California, Ohio, and Maryland who were tracked between the ages of 9 and 19. [...]

As part of the survey, the girls were asked once a year what they had eaten during the previous three days. [...]

Souce: Study: Cereal Keeps Girls Slim, Retrieved Sep 13, 2018.



Explanatory and response variables

- Explanatory variable: Whether the participant ate breakfast or not
- Response variable: BMI of the participant

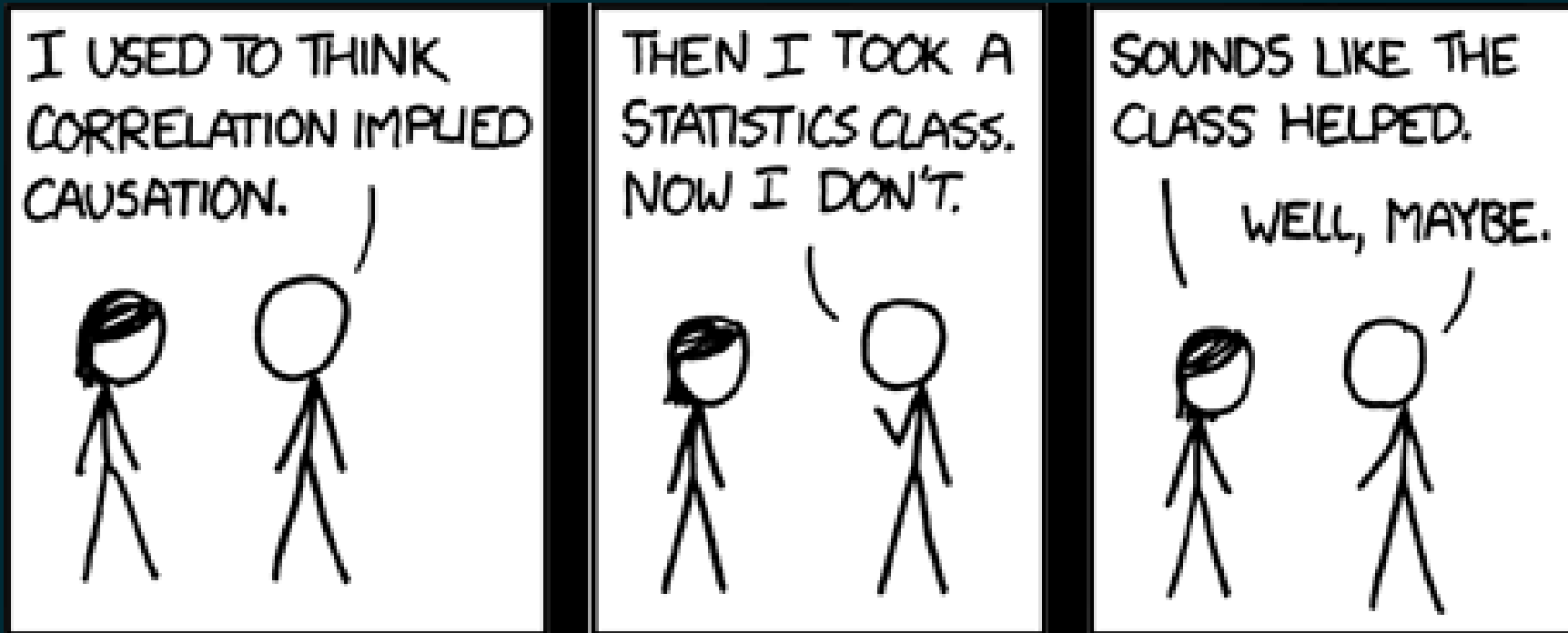


Three possible explanations

1. Eating breakfast causes girls to be slimmer
2. Being slim causes girls to eat breakfast
3. A third variable is responsible for both -- a **confounding** variable: an extraneous variable that affects both the explanatory and the response variable, and that makes it seem like there is a relationship between them



Correlation != causation



Randall Munroe CC BY-NC 2.5 <http://xkcd.com/552/>



Studies and conclusions

	Random assignment	No random assignment	
Random sampling	Causal and generalizable	Not causal, but generalizable	Generalizable
No random sampling	Causal, but not generalizable	Neither causal nor generalizable	Not generalizable
	Causal	Not causal	



Case study: Climate change survey



Survey question

A July 2019 YouGov survey asked 1633 GB and 1333 USA randomly selected adults which of the following statements about the global environment best describes their view:

- The climate is changing and human activity is mainly responsible
- The climate is changing and human activity is partly responsible, together with other factors
- The climate is changing but human activity is not responsible at all
- The climate is not changing



Survey data

	The climate is changing and human activity is mainly responsible	The climate is changing and human activity is partly responsible, together with other factors	The climate is changing but human activity is not responsible at all	The climate is not changing	Don't know	Sum
GB	833	604	49	33	114	1633
US	507	493	120	80	133	1333
Sum	1340	1097	169	113	247	2966

Source: YouGov - International Climate Change Survey



What percent of **all respondents** think the climate is changing and human activity is mainly responsible?

	The climate is changing and human activity is mainly responsible	The climate is changing and human activity is partly responsible, together with other factors	The climate is changing but human activity is not responsible at all	The climate is not changing	Don't know	Sum
GB	833	604	49	33	114	1633
US	507	493	120	80	133	1333
Sum	1340	1097	169	113	247	2966

```
(all <- 1340 / 2966)
```

```
## [1] 0.4517869
```



What percent of **GB respondents** think the climate is changing and human activity is mainly responsible?

	The climate is changing and human activity is mainly responsible	The climate is changing and human activity is partly responsible, together with other factors	The climate is changing but human activity is not responsible at all	The climate is not changing	Don't know	Sum
GB	833	604	49	33	114	1633
US	507	493	120	80	133	1333
Sum	1340	1097	169	113	247	2966

```
(gb <- 833 / 1633)
```

```
## [1] 0.5101041
```



What percent of **US respondents** think the climate is changing and human activity is mainly responsible?

	The climate is changing and human activity is mainly responsible	The climate is changing and human activity is partly responsible, together with other factors	The climate is changing but human activity is not responsible at all	The climate is not changing	Don't know	Sum
GB	833	604	49	33	114	1633
US	507	493	120	80	133	1333
Sum	1340	1097	169	113	247	2966

```
(us <- 507 / 1333)
```

```
## [1] 0.3803451
```



Based on the percentages we calculated, does there appear to be a relationship between country and beliefs about climate change? If yes, could there be another variable that explains this relationship?

```
all
```

```
## [1] 0.4517869
```

```
gb
```

```
## [1] 0.5101041
```

```
us
```

```
## [1] 0.3803451
```



Conditional probability

Notation: $P(A|B)$: Probability of event A given event B

- What is the probability that it will be unseasonably warm tomorrow?
- What is the probability that it will be unseasonably warm tomorrow, given that it was unseasonably warm today?



Independence

- If knowing event A happened tells you something about event B happening, or vice versa, then events A and B are not independent
- If not, they are said to be independent
- $P(A|B) = P(A)$

