

Correlation and Causation



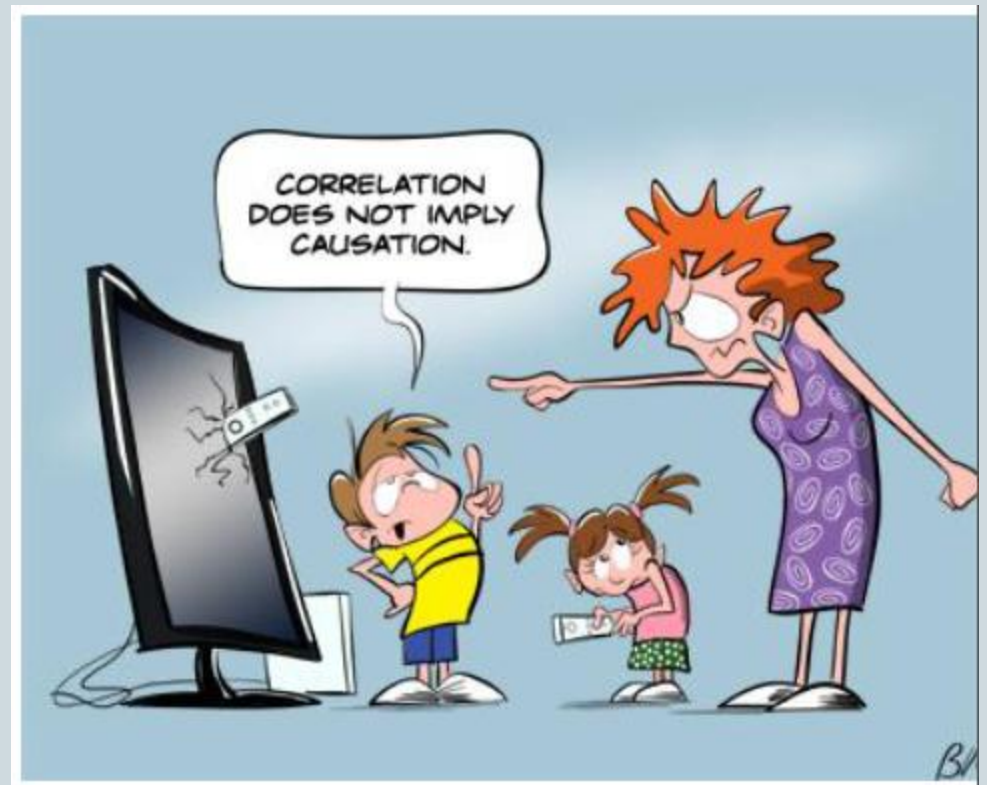
1.1.6



- Explain that the existence of a correlation does not establish that there is a causal relationship between two variables

Correlation does not imply causation

- A phrase used to emphasize that correlation between two variables does not automatically imply that one causes the other



Correlation



- Correlation will be positive or negative
 - Direct relationship exists, correlation equals causation
 - No direct relationship exists, correlation does not equate to causation.

Correlation



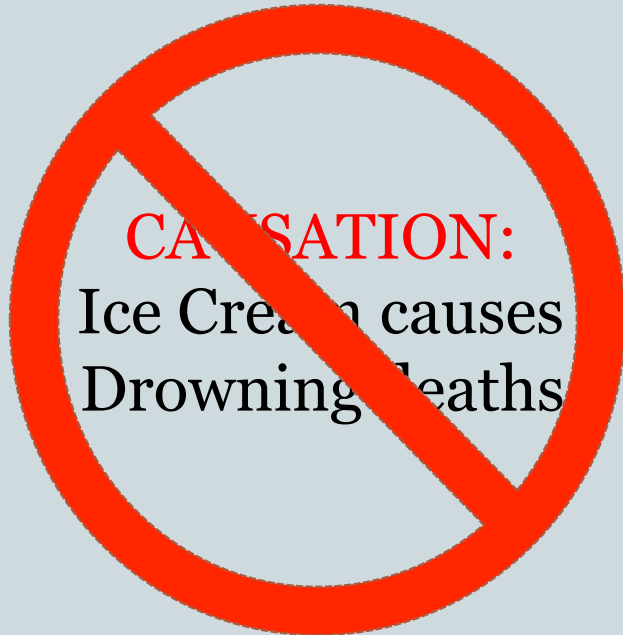
- The more fireman fighting a fire, the bigger the fire is going to be



Correlation



- As ice cream sales increase, the rate of drowning deaths increases sharply



Correlation



- Since the 1950s, both the atmospheric CO² level and crime levels have increased sharply

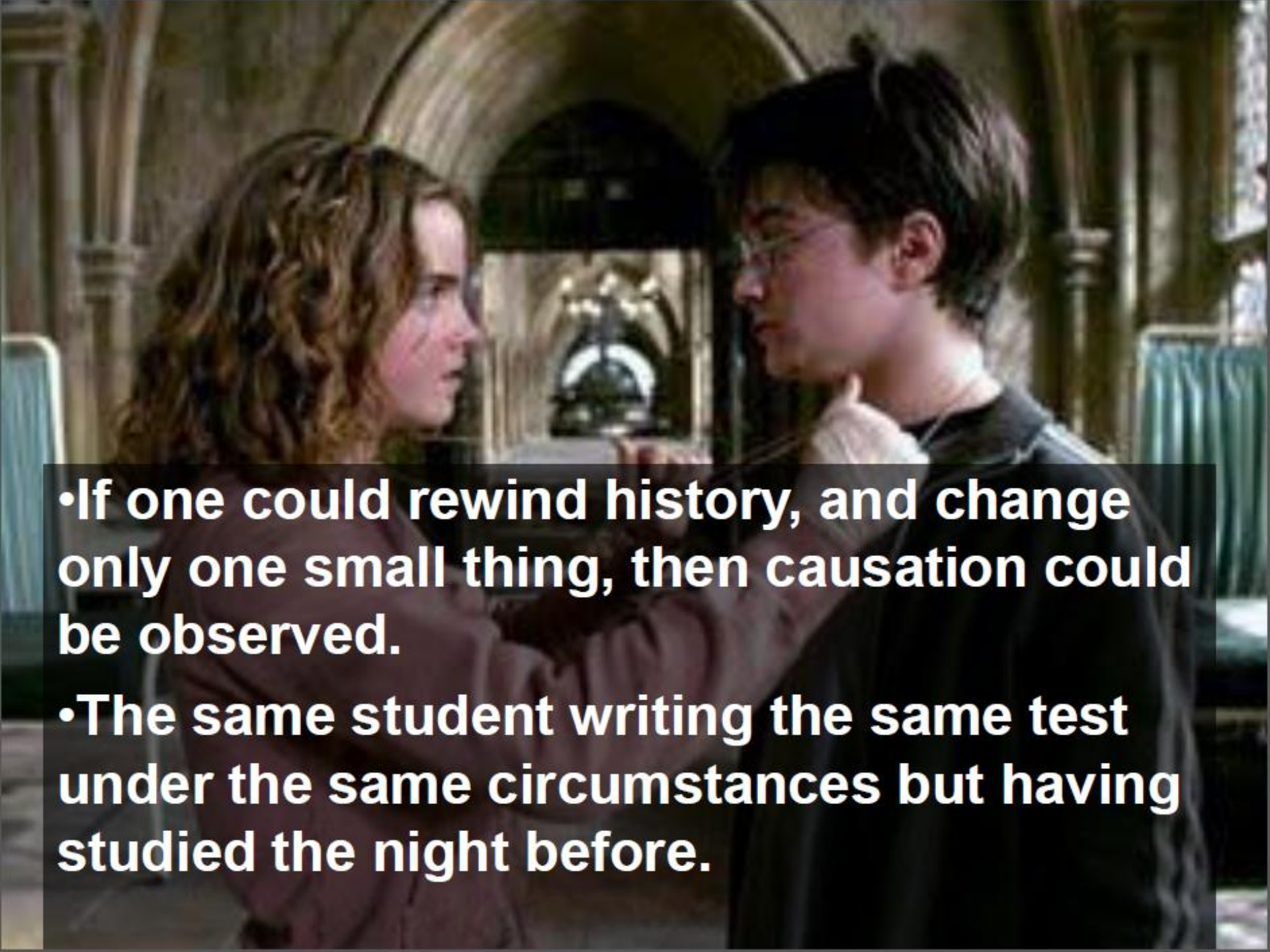


Determining Causation

- Suppose that a student performed poorly on a test and guesses that the cause was his not studying



**How could he
prove this?**

- 
- A young woman with long, curly brown hair and a young man with dark hair and glasses are standing in a hallway, facing each other and talking. The woman is on the left, wearing a dark purple sweater. The man is on the right, wearing a dark jacket over a white shirt. The background shows a hallway with arched doorways and a green curtain.
- If one could rewind history, and change only one small thing, then causation could be observed.
 - The same student writing the same test under the same circumstances but having studied the night before.

A major goal of scientific experiments is to control variables as best as possible.

- One could run an experiment on identical twins who were known to consistently get the same grades on their tests
- One twin is sent to study for six hours while the other is sent to the amusement park



- If their test scores suddenly diverged by a large degree, this would be strong evidence that studying had a causal effect on the test scores

- Correlation between studying and test scores would almost certainly imply causation



I USED TO THINK
CORRELATION IMPLIED
CAUSATION.

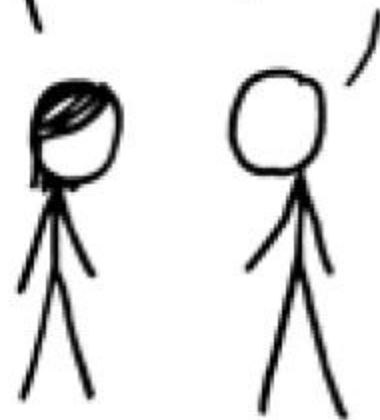


THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



SOUNDS LIKE THE
CLASS HELPED.

WELL, MAYBE.



HEADLINE:

“Diet of fish can prevent teen violence”



- Participants were a group of 3-year-olds given an “enriched diet, exercise, and cognitive stimulation.” They were compared to a control group who did not go through this same program
- By age 23 they were 64% less likely than a control group of children not on the program to have criminal records.
- Assume, of course, that the enriched diet included fish.
- Note, also, that the media article does not mention what the other kids ate or did.

**Does the data support the headline?
What are some “third variable” explanations?
How could you reword the headline?**

HEADLINE:

“Higher beer prices cut gonorrhoea rates”

- The research suggests "that raising the price of a six-pack of beer by 20 cents cut gonorrhoea rates by almost 9%
- Researchers considered gonorrhoea rates from 1981 to 1995 among teens and young adults in states that raised the legal drinking age or increased the state beer tax
- “Of the 36 beer tax increases that were reviewed, gonorrhoea rates declined among teens aged 15 to 19 in 24 instances. Among young adults aged 20 to 24, they declined in 267 instances”
- Important side note” 1981 is also when the CDC recognized AIDS and HIV; condoms protect against both HIV and gonorrhoea

**Does the data support the headline?
What are some “third variable” explanations?
How could you reword the headline?**

HEADLINE:

“Luckiest people born in summer”



- Online public survey (40,000 people)
- Those born in May were most likely to consider themselves lucky; those born in October had most negative views in their life
- People who took part in the survey gave their birthdates and rated the degree to which they saw themselves as lucky or unlucky.
- The poll found there was a summer-winter divide between people born from March to August and those born from September to February.
- 50% of those born in May considered themselves lucky; 43% of those born in October
- It isn't clear when the survey took place (i.e. What month)

**Does the data support the headline?
What are some “third variable” explanations?
How could you reword the headline?**