Lecture 7: Correlation

Madeleine Udell
Stanford University

Upcoming deadlines

This Thursday 11:59pm (4/27): HW 3

This Friday (4/28): Project proposal
[Required project meetings happen this week]

Next Monday (5/1): Quiz 1
[In class]

Demo

https://colab.research.google.com/github/stanford-mse-125/demos/blob/main/correlation.ipynb

The standard deviation line

1. Goes through the point of averages.

2. Climbs [or falls] at the rate of one vertical SD for each horizontal SD.

Correlation

Measure of association between two variables

Quantifies the dispersion of the points around the SD line.
Ranges from -1 to 1.

Definition: the correlation is the average of the products of the variables, when both are measured in standard units.

Correlation properties

Correlation is

- Scale invariant
- A measure of linear dependence
- Sensitive to outliers

Association is not causation

Examples from *Statistics* by Freedman et al.

For school children, shoe size is strongly correlated with reading skills. Does learning new words make your feet grow?

Association is not causation

Examples from *Statistics* by Freedman et al.

For school children, shoe size is strongly correlated with reading skills. Does learning new words make your feet grow?

Age is a confounding factor!

Association is not causation

Examples from *Statistics* by Freedman et al.

During the Great Depression of 1929-1933, better-educated people tended to have shorter spells of unemployment.

Does education protect you against unemployment?

Association is not causation

Examples from *Statistics* by Freedman et al.

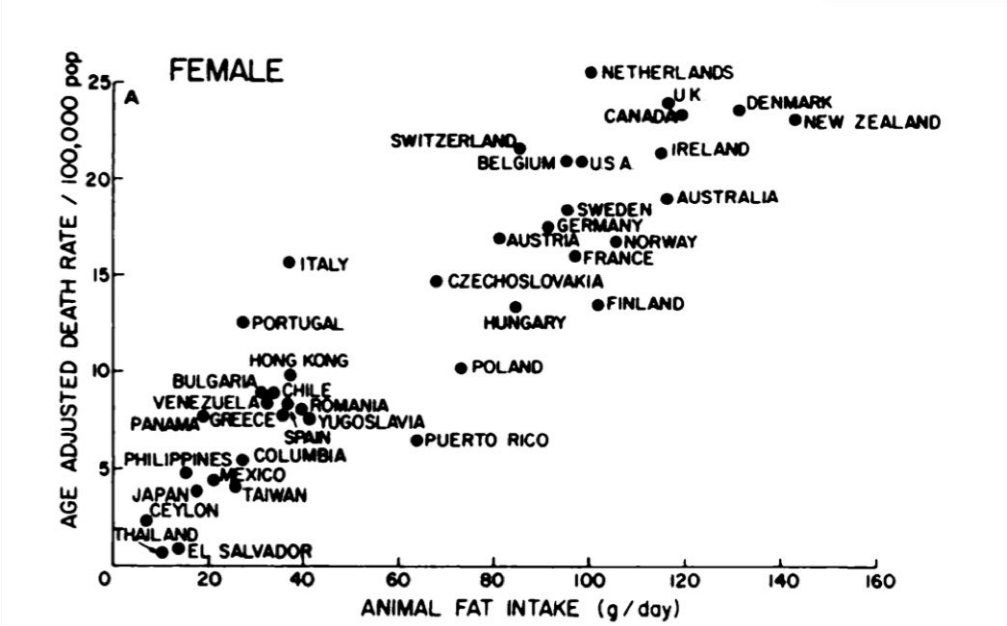
During the Great Depression of 1929-1933, better-educated people tended to have shorter spells of unemployment.

Does education protect you against unemployment?

Age is again a confounding factor. Employers tended to prefer younger job-seekers, and younger people were better educated.

Association is not causation

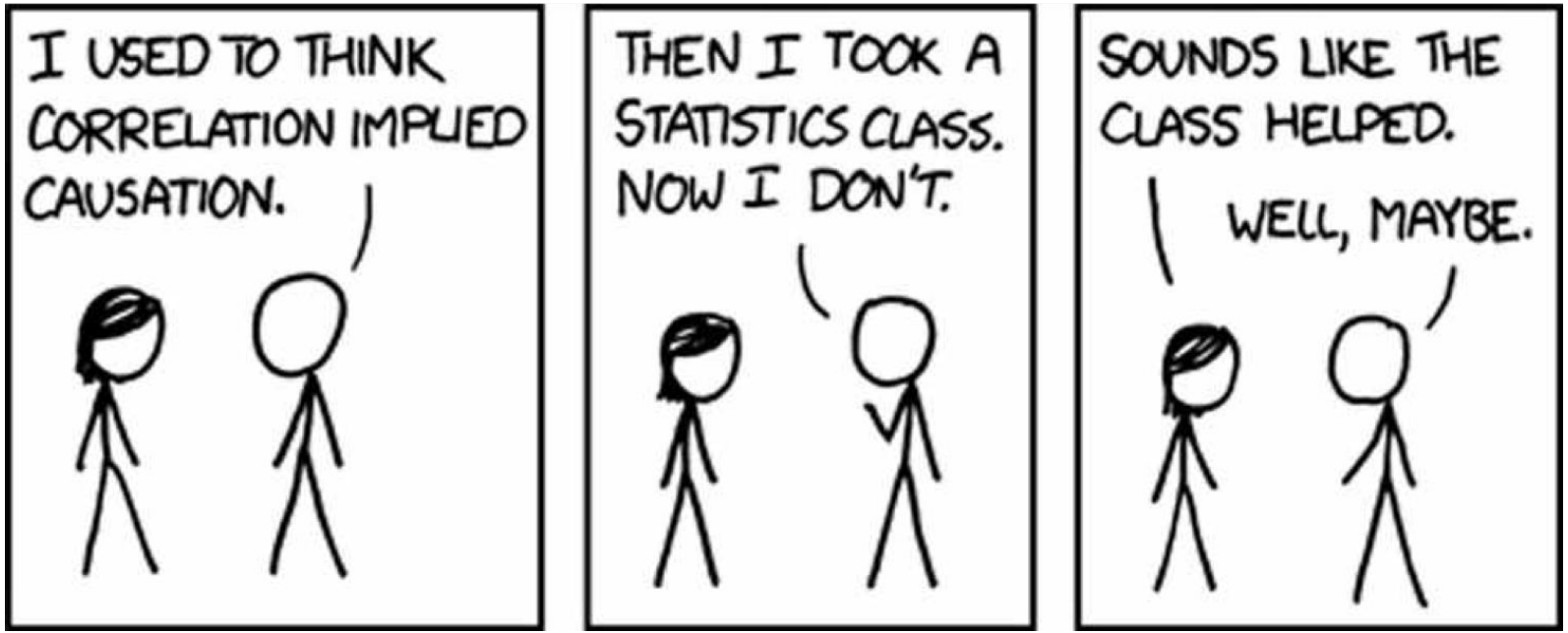
Fat in the diet and breast cancer



Association is not causation

Fat in the diet and breast cancer

Fat is relatively expensive so high fat intake occurs primarily in rich countries. Rich countries differ in a lot of ways from poorer ones.


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.