

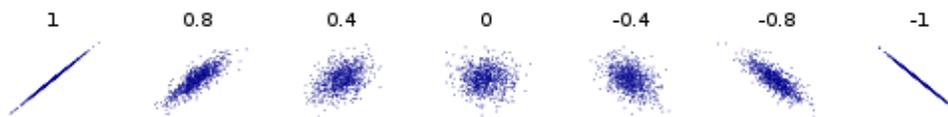
LECTURE 07: CORRELATION AND CAUSATION

I. Of Correlation and Causation

- a. The scatter diagram graphically illustrates if two variables are correlated, or that if one value changes the other will change in a predictable way.
 - i. *Positively correlated* means the values change in the same direction, such as “time studied” and “grade earned.”
 - ii. *Negatively correlated* means the values change in the opposite direction, such as “time partying” and “grade earned.”
- b. Correlation is actually a spectrum. Sometimes, there is weak correlation and other times the correlation is strong. Consider the three dot plots below. Each represents positive correlation but note that the data sets are very different:



- c. We can sum up those differences with the *correlation coefficient*—a single number which captures the strength and kind of the correlation.
 - i. A positive value indicates positive correlation and a negative value indicates negative correlation.
 - ii. The closer the value is to 0, the weaker the correlation.
 - iii. The value cannot be greater than 1 or less than -1.



- d. Of course, *correlation does not mean causation*. Just because it looks like two variables run together doesn't mean they do. Two other things could be going on:
 - i. *Reverse causation*—when the dependent and independent are confused (Greater CO₂ emissions cause people to earn more?)
 - ii. *Confounding variable*—variable which correlates with both independent and dependent variables (Does a greater portion of agricultural workers lower infant mortality?)

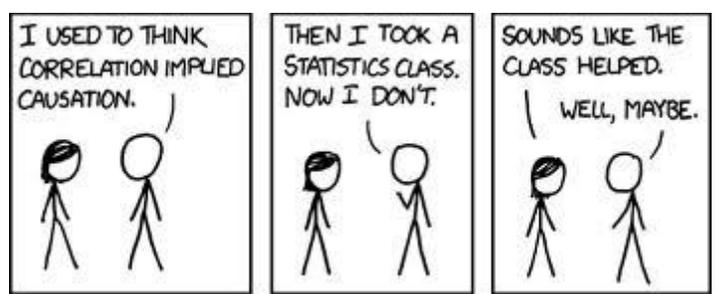
e. At the same time, *correlation is evidence for causation*. The internet is full of people, upon seeing a strongly correlated pair of variables, dismissing any possibility that the two are connected because it doesn't mean causation. But it should give you reason to pause. And if you can find an explanation *why* one would cause another, you're in good standing.



i. Hedge

fund manager and blogger James Altucher dismisses evidence of higher earning potential thanks to a college degree by invoking the tired mantra.¹ Of course, there are good reasons to think college causes higher earnings such as credentials, signaling, and skill building.

f. You need a narrative—some sort of reason—why one thing can cause another. In the comic on the left, it makes sense the male character knows “correlation doesn't mean causation” because the statistics course would emphasize such thinking. If he learned, since the class, that North Korea is an oppressive dictatorship which puts disgruntled citizens into death camps, then that's probably a coincidence. North Korea politics aren't covered in (most) statistic courses.



¹ <http://www.jamesaltucher.com/2011/01/10-more-reasons-why-parents-should-not-send-their-kids-to-college/>