Correlation Versus Causality

Gray Watson

2004.05.11

A number of years ago, I picked up a magazine in a doctor's office and read how researchers had determined that cars painted yellow got in the fewest accidents. They theorized that this was because other drivers could see yellow the best so were able to avoid them. The question that immediately came to me was "what kind of person drives a yellow car?" Cars with sportier colors (red and black) are purchased by younger kids who drive faster and get into more accidents. Yellow cars may appeal to older drivers, who tend to get in fewer accidents.

This is a good example of confusion between correlation and cause. Correlation is defined by dict.org as:

> a statistical relation between two or more variables such that systematic changes in the value of one variable are accompanied by systematic changes in the other.

In other words, if you are studying two things A and B (car color and accident rate) and you change A then there will be a change in B if they are correlated. There could be a negative correlation in that when A increases, B decreases or it could be a positive correlation where when A increases B also increases. In the car study, the color yellow is negatively correlated with accidents – the more yellow the car, the less accidents it has.

But correlations do not imply that A caused B. You can't go out and repaint all of the cars in Boston and somehow get fewer accidents. It could be that B caused A or rather that C (some other variable that we don't know about) caused both A and B. It could be that C (the age of the person who buys the car) causes both A and B.

Corr-Causation.ctx