Correlation, causation confused

False conclusions drawn from statistical disparities

THOMAS SOWELL

Syndicated columnist

One of the first things they teach in introductory statistics is that correlation is not causation. It is also one of the first things forgotten. People have for centuries been mistaking correlation to mean causation. The causation they have seen has been based on their own preconceptions, seemingly supported by empirical evidence.

Centuries ago, Galileo invented the barometer. But since he used water rather than mercury in the tube, his barometer was much larger than the later mercury barometers, because the atmosphere supports a much taller column of water than of mercury.

Galileo's barometer went all the way up through the roof of his house and came outside. In order to be able to tell how high the water was on a given day, Galileo floated a wooden figure of a red devil on the water. That way he could keep track of whether the atmospheric pressure was going up or down.

Galileo's neighbors began to notice that the red devil came up out of the house on bright sunny days and went back down inside on rainy days. Having no idea about how atmospheric pressure changes with the weather, the neighbors attributed this correlation to sinister goings-on with the devil.

What really set them off was that the red devil kept behaving this way even when there was nobody home. He just kept coming out on sunny days, and going back inside on rainy days, on his own. That did it! They broke into Galileo's house and destroyed the barometer.

Centuries later, when the pioneers were settling the American northwest, one group of them brought a measles epidemic that struck whites and Indians alike. Dr. Marcus Whitman, for whom Whitman College would later be named, treated patients of both races. The whites recovered and the Indians died.

Indians noticed the correlation between race and outcome, and concluded that Dr. Whitman was doing something to cause the Indians to die while the whites lived. They burned down his outpost and killed him and many other whites — an action that set off retaliatory attacks by whites against the Indians.

Today we realize that Indians throughout the Western Hemisphere had no biological resistance to European diseases and were wiped out by many of these diseases that were not nearly as devastating to people of European ancestry. In other words, now we know the causation, but the Indians saw only the correlation.

Even well into the 20th century, American national policy was based on correlations that supported preconceptions — and which were therefore accepted as causation.

After Japan's shocking attack on Pearl Harbor, in the midst of "peace" negotiations with the United States, someone noticed that many Japanese Americans were living clustered in the vicinity of military installations.

The correlation was far greater than could be accounted for by random chance. The conclusion was that they were living there in order to be able to sabotage American military forces for the benefit of Japan. Such reasoning played a role in the decision to intern 100,000 people during the war.

It turned out that the Japanese Americans had chosen their locations because the land was cheap — and the military later came in and built installations there for the same reason.

Today, many national policies and even Supreme Court decisions are based on statistical correlations that seem to fit the prevailing preconceptions of our time.

If women are "under-represented" in certain industries or occupations, that is automatically regarded as evidence of discrimination. If the proportion of youngsters suspended or expelled from school because of bad conduct differs from one racial or ethnic group to another, charges of discrimination are sure to follow.

This kind of confusion of correlation with causation is so widespread that you might think an even representation of people was some kind of norm, so that any deviations from it are suspicious. In reality, large statistical disparities exist between all sorts of groups in countries around the world.

At the running of this year's Boston marathon, the first five runners across the finish line were all from Kenya.

All the baseball players who have stolen a hundred bases in a season are black.

For years now, most of the people laying cable in Sydney, Australia, have been Irish. All the billionaires in Thailand and Indonesia are of Chinese ancestry. Four-fifths of the doughnut shops in California are owned by people of Cambodian ancestry.

It would be no feat to fill this whole page with statistical disparities that have nothing to do with discrimination. What would be a real feat would be to get people to realize that correlation is not causation — especially when the numbers fit their preconceptions.

Sowell is a fellow with the Hoover Institution at Stanford University. His column is distributed byt Creators Syndicate, Inc.