

## Spurious Correlations

© 1996, 1997 by William C. Burns (NOTE: He gave me permission to use this 9/6/13)

The analysis of human resources data typically involves the use of computer databases that were constructed to process transactions. Their purpose normally centers on administration and recordkeeping. Thus the variables that are available for analysis are not necessarily the ones that would be chosen as the ideal set of variables given the purposes of the analysis. A side effect is that in many cases critical analysis variables may be missing. This can lead to "spurious correlations," a common and serious interpretation fallacy. For example, suppose that the critical variable is correlated with race, age, or gender. Thus any other variable that correlates with the critical variable will probably also be correlated with race, age, or gender. These correlations are spurious because their primary cause is the missing critical variable. Nonetheless these spurious correlations are at times used as indicators of discrimination. The purpose of this paper is to illustrate the widespread occurrence of spurious correlations.

My favorite example is to do the following:

1. Get data on all the fires in San Francisco for the last ten years.
2. Correlate the number of fire engines at each fire and the damages in dollars at each fire.

Note the significant relationship between number of fire engines and the amount of damage. Conclude that fire engines cause the damage.

The reason that I like this example is that the conclusion is so absurd. Anyone will quickly recognize that both variables result from and are correlated with the overall size of the fire. However, many spurious correlations do not seem absurd and some seem compelling.

### Other Descriptions and Examples

The spurious-correlation fallacy is not widely recognized by most people. Its occurrence is pervasive, but it is generally unnoticed. Part of the problem is the wide variation in terminology that is used by different authors. The purpose of this paper is to provide a set of examples, which illustrate the various ways that the fallacy can be described and discussed. The goal is to use repetition to help develop a "feel" for the pattern so that recognition becomes easier. Because the examples are highly redundant, once the pattern is clearly understood, reading the entire paper may be unnecessary.

The following excerpts are from a dictionary, two books devoted entirely to mathematical and statistical fallacies, an elementary statistics textbook, and a book on statistical methodology. They all describe the same phenomenon in somewhat different terminology.

#### **Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences**

By W Paul Vogt

(An excellent reference covering procedures, concepts, and issues in both subjects.)

**"Spurious Relation (or Correlation)** (a) A situation in which measures of two or more variables are statistically related (they cover) but are not in fact causally linked—usually because the statistical relation is caused by a third variable. When the effects of the third variable are removed, they are said to have been partialled out. See confound, lurking variable." (Defined below)

"(b) A spurious correlation, as defined in definition a, is sometimes called an "illusory correlation." In that case, "spurious" is then reserved for the special case in which a correlation is not present in the original observations but is produced by the way the data are handled. Compare artifact.

For example, (a) if the students in a psychology class who had long hair got higher scores on the midterm than those who had short hair, there would be a correlation between hair length and test scores. Not many people, however, would believe that there was a causal link and that, for example, students who wished to improve their grades should let their hair grow. The real cause might be gender: that is, women (who usually have longer hair) did better on the test. Or that might be a spurious relationship too. The real cause might be class rank: Seniors did better on the test than sophomores and juniors, and, in this class, the women (who also had longer hair) were mostly seniors, whereas the men (with shorter hair) were mostly sophomores and juniors." (p. 217)

**"Lurking Variable.** A third variable that causes a correlation between two others – sometimes, like the troll under the bridge, an unpleasant surprise when discovered. A lurking variable is a source of a spurious correlation. See also confound. Compare covariate, latent variable, moderator variable.

For example, if researchers found a correlation between individuals' college grades and their income later in life, they might wonder whether doing well in school increased income. It might; but good grades and high income could both be caused by a third (lurking or hidden variable) such as tendency to work hard."

(p.132)

**How to Lie with Statistics**

By Darrell Huff

(A classic.)

"Somebody once went to a good deal of trouble to find out if cigarette smokers make lower college grades than nonsmokers. It turned out that they did. This pleased a good many people and they have been making much of it ever since. The road to good grades, it would appear, lies in giving up smoking; and, to carry the conclusion one reasonable step further, smoking makes dull minds.

This particular study was, I believe, properly done: sample big enough and honestly and carefully chosen, correlation having a high significance, and so on. The fallacy is an ancient one which, however, has a powerful tendency to crop up in statistical material, where it is disguised by a welter of impressive figures. It is the one that says that if B follows A, then A has caused B.

An unwarranted assumption is being made that since smoking and low grades go together, smoking causes low grades. Couldn't it just as well be the other way around? Perhaps low marks drive students not to drink but tobacco. When it comes right down to it, this conclusion is about as likely as the other and just as well supported by the evidence. But it is not nearly so satisfactory to propagandists.

It seems a good deal more probable, however, that neither of these things has produced the other, but both are a product of some third factor. Can it be that the sociable sort of fellow who takes his books less than seriously is also likely to smoke more? Or is there a clue in the fact that somebody once established a correlation between extroversion and low grades – a closer relationship apparently than the one between grades and intelligence? Maybe extroverts smoke more than introverts. The point is that when there are many reasonable explanations you are hardly entitled to pick one that suits your taste and insist on it. But many people do.

To avoid falling for the post hoc fallacy and thus wind up believing many things that are not so, you need to put any statement of relationship through a sharp inspection.

The correlation, that convincingly precise figure that seems to prove that something is because of something, can actually be any of several types." (pp.87 - 89)

...

"Perhaps the trickiest of them all is the very common instance in which neither of the variables has any effect at all on the other, yet there is a real correlation. A good deal of dirty work has been done with this one. The poor grades among cigarette smokers is in this category, as are all too many medical statistics that are quoted without the qualification that although the relationship has been shown to be real, the cause-and-effect nature of it is only a matter of speculation. As an instance of the nonsense or spurious correlation that is a real statistical fact, someone has gleefully pointed to this: There is a close relationship between the salaries of Presbyterian ministers in Massachusetts and the price of rum in Havana.

Which is the cause and which the effect? In other words, are the ministers benefiting from the rum trade or supporting it? All right. That's so farfetched that it is ridiculous at a glance. But watch out for other applications of post hoc logic that differ from this one only in being more subtle. In the case of the ministers and the rum it is easy to see that both figures are growing because of the influence of a third factor: the historic and world-wide rise in the price level of practically everything." (p. 90)

...

"Professor Helen M. Walker has worked out an amusing illustration of the folly in assuming there must be cause and effect whenever two things vary together. In investigating the relationship between age and some physical characteristics of women, begin by measuring the angle of the feet in walking. You will find that the angle tends to be greater among older women. You might first consider whether this indicates that women grow older because they toe out, and you can see immediately that this is ridiculous. So it appears that age increases the angle between the feet, and-most women must come to toe out more as they grow older.

Any such conclusion is probably false and certainly unwarranted. You could only reach it legitimately by studying the same women – or possibly equivalent groups – over a period of time. That would eliminate the factor responsible here. Which is that the older women grew up at a time when a young lady was taught to toe out in walking, while the members of the younger group were learning posture in a day when that was discouraged." (pp. 96 - 97)

...

"Permitting statistical treatment and the hypnotic presence of numbers and decimal points to befog causal relationships is little better than superstition. And it is often more seriously misleading. It is rather like the conviction among the people of the New Hebrides that body lice produce good health. Observation over the centuries had taught them that people in good health usually had lice and sick people very often did not. The observation itself was accurate and sound, as observations made informally over the years surprisingly often are. Not so much can be said for the conclusion to which these primitive people came from their evidence: Lice make a man healthy. Everybody should have them.

As we have already noted, scantier evidence than this— treated in the statistical mill until common sense could no longer penetrate to it—has made many a medical fortune and many a medical article in magazines, including professional ones. More sophisticated observers finally got things straightened out in the New Hebrides. As it turned out, almost everybody in those circles had lice most of the time. It was, you might say, the normal condition of man. When, however, anyone took a fever (quite possibly carried to him by those same lice) and his body became too hot for comfortable habitation, the lice left. There you have cause and effect altogether confusingly distorted, reversed, and intermingled." (pp. 98 - 99)

### **A Mathematician Reads the Newspaper**

By John Allen Poulos

(A follow-up to his bestseller, Innumeracy, which is to numbers what illiteracy is to words. He illustrates widespread innumeracy in newspapers.)

"A more elementary widespread confusion is that between correlation and causation. Studies have shown repeatedly, for example, that children with longer arms reason better than those with shorter arms, but there is no causal connection here. Children with longer arms reason better because they're older! Consider a headline that invites us to infer a causal connection: BOTTLED WATER LINKED TO HEALTHIER BABIES. Without further evidence, this invitation should be refused, since affluent parents are more likely both to drink bottled water and to have healthy children; they have the stability and wherewithal to offer good food, clothing, shelter, and amenities. Families that own cappuccino makers are more likely to have healthy babies for the same reason. Making a practice of questioning correlations when reading about "links" between this practice and that condition is good statistical hygiene." (p. 137)

### **Statistics Second Edition**

By David Freedman, Robert Pisani, Roger Purves, and Ani Adhikari

(This is the textbook used in the introductory statistics course in the Statistics Department at both Stanford University and University of California at Berkeley.)

### **"ASSOCIATION IS NOT CAUSATION"**

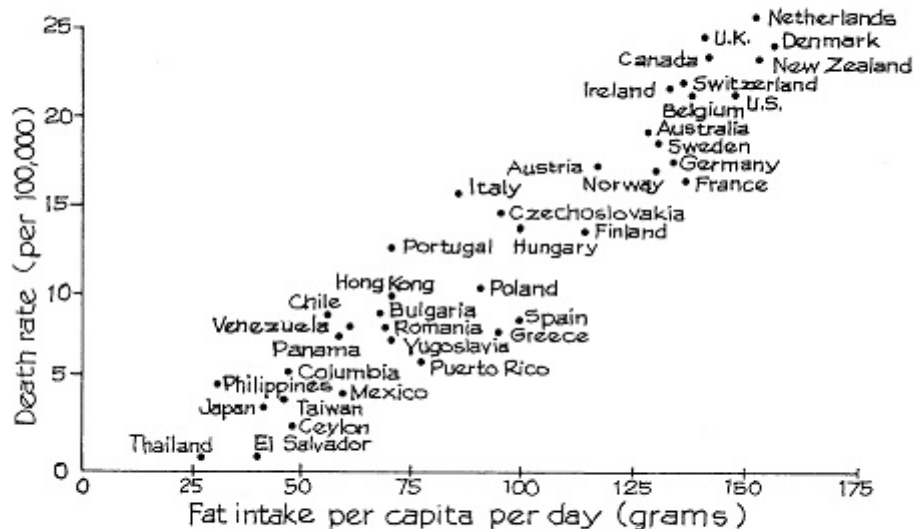
For school children, shoe size is strongly correlated with reading skills. However, learning new words does not make the feet get bigger. Instead, there is a third factor involved – age. As children get older, they learn to read better and they outgrow their shoes. (In the statistical jargon of chapter 2, age is a confounding factor.) In the example, the confounder was easy to spot. Often, this is not so easy. And the arithmetic of the correlation coefficient does not protect you against third factors.

### **Correlation measures association. But association is not the same as causation.**

...

Example 3. Fat in the diet and cancer. In countries where people eat lots of fat like the United States rates of breast cancer and colon cancer are high. See figure 8 (next page). This correlation is often used to argue that fat in the diet causes cancer. How good is the evid

Figure 8. Cancer rates plotted against fat in the diet, for a sample of countries.



Source: K. Carroll, "Experimental evidence of dietary factors and hormone-dependent cancers," *Cancer Research* vol. 35 (1975) p. 3379. Copyright by *Cancer Research*. Reproduced by permission.

ence?

Discussion. If fat in the diet causes cancer, then the points in the diagram should slope up, other things being equal. So the diagram is some evidence for the theory. But the evidence is quite weak, because other things aren't equal. For example, the countries with lots of fat in the diet also have lots of sugar. A plot of colon cancer rates against sugar consumption would look just like figure 8, and nobody thinks that sugar causes colon cancer. As it turns out, fat and sugar are relatively expensive. In rich countries, people can afford to eat fat and sugar rather than starchier grain products. Some aspects of the diet in these countries, or other factors in the life-style, probably do cause certain kinds of cancer and protect against other kinds. So far, epidemiologists can identify only a few of these factors with any real confidence. Fat is not among them." (pp.142 - 144)

### **Statistics as Principled Argument**

By Robert P Abelson

(This is perhaps the best book on the use of Statistics. Abelson is highly respected and widely honored)

"We have seen that the category of methodological artifacts is a broad one. Here we discuss three general categories that come up repeatedly: the influence of third variables; the presence of impurities in the variables; and procedural bias. Cases involving third variables typically apply to correlational studies, procedural bias to experimental studies, and impurities to both types of studies.

#### Third Variables

We go back to basics and begin our discussion by considering an elementary claim from a correlational study that two variables are related as cause and effect. We saw in chapter 1, in our discussion of the purported longevity of conductors, how misleading such claims can be."

(from Chapter 1)

"The longevity datum on famous orchestral conductors (Atlas, 1978) provides a good example. With what should the mean age at their deaths, 73.4 years, be compared? With orchestral players? With nonfamous conductors? With the general public?

All of the conductors studied were men, and almost all of them lived in the United States (though born in Europe). The author used the mean life expectancy of males in the U.S. population as the standard of comparison. This was 68.5 years at the time the study was done, so it appears that the conductors enjoyed about a 5-year extension of life and indeed, the author of the study jumped to the conclusion that

involvement in the activity of conducting causes longer life. Since the study appeared, others have seized upon it and even elaborated reasons for a causal connection (e.g., as health columnist Brody, 1991, wrote, "it is believed that arm exercise plays a role in the longevity of conductors."

However, as Carroll (1979) pointed out in a critique of the study, there is a subtle flaw in life-expectancy comparisons: The calculation of average life expectancy includes infant deaths along with those of adults who survive for many years. Because no infant has ever conducted an orchestra, the data from infant mortalities should be excluded from the comparison standard. Well, then, what about teenagers? They also are much too young to take over a major orchestra, so their deaths should also be excluded from the general average. Carroll argued that an appropriate cutoff age for the comparison group is at least 32 years old, an estimate of the average age of appointment to a first orchestral conducting post. The mean life expectancy among U.S. males who have already reached the age of 32 is 72.0 years, so the relative advantage, if any, of being in the famous conductor category is much smaller than suggested by the previous, flawed comparison."

(end of the passage from Chapter 1) (p. 4)

"Every student in the social sciences is to a greater or lesser degree taught to be reluctant to draw causal conclusions from correlations, but it is surprising how causal implications nonetheless sneak insidiously into interpretations of correlations.

An investigator who takes her correlational results as indicating a causal relationship is subject to a plentiful source of criticisms – the artifact of the third variable. If it be asserted from a significant correlation of A with B that A causes B, the critic can usually rebut forcefully by proposing some variable C as the underlying causal agent for both A and B.

Power Dressing: A Whimsical Example. As a hypothetical example, consider a relationship for high school seniors between the sizes of their wardrobes and scores on the Scholastic Aptitude Test (SAT). Let us imagine that a significant correlation of .40 between these two variables is announced by an investigator, who weaves a tale about the importance of power dressing for success in life.

A critic cries humbug, noting that the relation could easily be explained as an artifact of income differences. Socially advantaged kids have lots of clothes, and by and large do well on standardized tests, whereas disadvantaged kids have fewer clothes, and perform less well on tests.

The critic may make this shot overpowering by reanalyzing the investigator's data (or by analyzing new data), showing that when income differences are partialled out, or when income is held constant, the relationship between wardrobe size and SAT scores disappears. One way to look for such an outcome awkward in practice but conceptually clear is to sort cases on the income variable into class intervals, and then for the cases within each class interval, see whether there is any relationship between the original two variables. If there is very little or none, the third variable can be said to explain the relation between the other two. The investigator will usually be left without any satisfactory rejoinder. Presumably this is what would happen in the hypothetical example of the wardrobes." (pp. 180 - 182)

[Return to WCB&A home page](#)