Chapter Eight

CORRELATIONS AND CAUSES

[W]e are not concerned with the routine random differences in outcomes based on luck. Rather the concept of discrimination encompasses only those differences that are so systematic that they do not cancel each other out within large groups

Ehrenberg & Smith

EXPLAINING THE NUMBERS

As we saw in Chapter Five, statistical reasoning in the social and natural sciences can easily be reconstructed as a related pair of inferences to the best explanation. In the first inference the explanatory question focuses on some quantitative relationship. Consider the extensive medical data that was uncovered over several decades in the famous Framingham study. Medical researchers were surprised to discover that twenty-nine percent of the men in the forty to forty-nine year range suffered from coronary heart disease while only fourteen percent of the women in the same age range suffered from the disease. This tells us something potentially very important about gender and heart disease.

e_1. Of the 771 men in the 40-49 year age group, 29% showed some signs of coronary heart disease.

e_2. Of the 954 women in the 40-49 year age group, only 14% showed signs of coronary heart disease.

=============================

t_0. Coronary heart disease appears much more often in men than in women.

The rival explanation that must be ruled out is the so-called null hypothesis. We want to know if the numbers indicate some real world tendency, or whether they are simply a coincidence. Much of statistics can be seen as the sophisticated application of mathematical theory, particularly the probability calculus, to developing reliable techniques for distinguishing between the null hypothesis as the better explanation of the numerical data, or some genuine correlation as the better explanation.

EXPLAINING THE CORRELATIONS

Consider the following data that played such a prominent role in Al Gore's, *An Inconvenient Truth.*
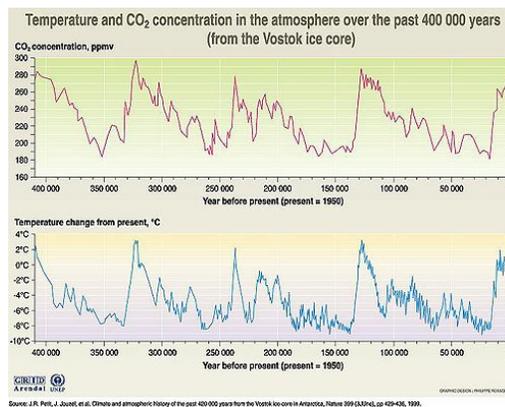
Mr. Gore used this data as evidence that CO_2 concentrations cause global temperature variations.

> e_1. There is a strong correlation between CO_2 levels and the Earth's average temperature.
> ============================
> t_0. CO_2 concentrations cause global temperature variations.

Given that the correlation is real, and not simply a fluke or coincidence -- for the modern social scientist, that it is statistically significant -- we must now determine whether t_0 is the best explanation of the correlation. We must compare it to some rival explanations. Perhaps, as some skeptics have claimed, the direction of causation is reversed.

> t_1. Global temperature variations cause varying CO_2 concentrations

This rival is probably a better account of the historical data, because many believe that we see the changes in temperatures before we see changes in CO_2 level. In addition, before the advent of the industrial revolution, it was hard to see what else could initiate such large scale changes in the CO_2 concentrations.

> It is likely that the temperature variations ... drove the CO_2 variations, not the reverse. That might have occurred, for example, when warmer temperatures increased the rate of bacterial breakdown of plant material, releasing CO_2 to the atmosphere as it warmed. This historical relationship does not, however, refute the modern relationship of human additions of CO2 to the atmosphere driving increases in temperature.[1]

Why, you may ask doesn't the "reverse cause" rival, t_1, refute the anthropogenic hypothesis?

> [One] potential explanation for the observed warming of the Earth is human activity. There are several reasons to think that this can account for some portion of the observed warming. We know that human activities have been increasing the concentration of CO_2 and other greenhouse gases in the atmosphere for at least the past century or two. Measurements show the concentration of CO_2 has increased about 30 percent over that time ..., while other greenhouse gases have increased by similar or larger amounts. Basic physics provides strong theoretical reasons to believe that such an increase in greenhouse gases should warm the Earth.[2]

It now seems likely that the best explanation of the correlation is that the causal relationship between CO_2 and global warming actually points in both directions and that increased CO_2 concentrations cause increased temperatures, and that simultaneously, increased temperatures cause increases in CO_2 concentrations. We probably have a kind of feedback loop.

> t_2. Increased CO_2 concentrations cause increased temperatures, while, increased temperatures cause increases in CO_2 concentrations.

A little mantra that you often hear in science and statistics courses is that "you cannot infer a cause from a correlation." But actually you often can. You must remind yourself, however, that there are many causal possibilities. When **A** and **B** are genuinely correlated **A** may be causing **B**, **B** may be causing **A**, some third "common cause," **C** may be causing both **A** and **B**. And as we have just seen, in cases of feedback **A** may be causing **B**, while at the very same time **B** is causing **A**. You will probably need more data than the simple correlation to sort out which is causal possibility is operative.

STATISTICS AND THE DEATH PENALTY

You will remember that I had started to climb on a soapbox about capital punishment at the end of the last chapter. I claimed that the best constructive interpretation of our nation's abstract intentions with respect to the death penalty was the following.

t_0. The death penalty must be imposed fairly, and with reasonable consistency, or not at all.

I want now to continue with my case against the death penalty by arguing that both fairness and reasonable consistency are demonstrably absent. My argument to this effect will depend on the analysis of statistical evidence.

I take it that legal historians would agree with me that capital punishment has, in the past, been applied in a manner that was clearly discriminatory. We would like to think, however, that we have made some progress in the area of racial justice. That is why the following data is so disappointing.

Professor Baldus examined over twenty-four hundred homicide cases in the state of Georgia during the period between 1974 and 1979. The dates are significant because the Georgia murder stature had been rewritten after *Furman v. Georgia* in order that death sentences not be administered in a "random and capricious manner." Here's a brief summary of what Professor Baldus discovered.

Race Killer/Victim	Death Sentence	Percentage
Black/White	50 of 223	22%
White/White	58 of 748	8%
Black/Black	18 of 1443	1%
White/Black	2 of 60	3%
Total by Victim		
White	108 of 981	11%
Black	20 of 1503	1%

The original Baldus study controlled for over two hundred non-racial variables such as the defendant's record and the severity of the crime. When all of this data was considered, the study concluded that murderers of white victims were 4.3 times as likely to receive the death penalty. Justice Brennan expressed this correlation in characteristically vivid language.

> At some point in this case, Warren McCleskey doubtless asked his lawyer whether the jury was likely to sentence him to die. A candid reply to this question would have to tell McCleskey that few of the details of the crime or of McCleskey's past criminal record were more important than the fact that his victim was white. Furthermore, counsel would feel bound to tell McCleskey that defendants charged with killing white victims in Georgia are 4.3 times as likely to be sentenced to die as defendants charged with kill blacks.[3]

I have discussed the McCleskey case with hundreds of students in the last several years. Many simply refuse to accept the following data.

e_1. When controlled for over two hundred non-racial variables such as the defendant's record and the severity of the crime, the Baldus study concluded that murderers of white victims were 4.3 times as likely to receive the death penalty.

It is, of course, true that life in the ghetto is different than life in the suburbs, and that the black culture is different that the white culture. The shocking figure that over four times as many murderers of whites receive the death penalty takes all of that into account. I know some of you will continue to believe that "statistics always lie." But the very same techniques that tell us that cigarette smoking causes cancer, or that so-and-so will win next month's election, tell us that the connection between race and the death penalty in Georgia is for real. Thus, the question before us is producing an explanation of why this correlation holds.

There is no big mystery about the reason for this disparity. The original study contained the crucial data.

e_2. District attorneys ask for a capital sentence in seventy percent of the cases involving a black defendant and a white victim. When the victim is black and the defendant is white, however, a mere nineteen percent are even prosecuted as capital cases.

Clearly the discretion of local law enforcement officials is the point at which racial attitudes enter the criminal process. Apparently, black on black murders are considered even less important, since the Baldus study showed that the death penalty was requested in only nineteen percent of these cases.

In *Furman* the Supreme Court was concerned with the arbitrary and capricious actions of trial judges, and particularly juries. It appears now, however, that the judgments, both arbitrary and prejudicial, of other legal officials are even more problematic. An obvious explanation of the Baldus data is the following.

t$_0$. The race of the murder victim causally influences the decision whether to seek the death penalty.

Given my earlier interpretation of the Eighth Amendment, this account of the murder statistics in Georgia seem to demand that the Supreme Court declare capital punishment, at least in the state of Georgia, to be unconstitutional.

WHAT'S THE BEST EXPLANATION?

As I hope by this point you are all already thinking, the crucial question is whether t$_0$ is really the best explanation. I, personally, cannot see anyway in this could be a case of "reverse causation." Thus, I reject any possibility that the following needs to be considered at all.

t$_1$. The decision to seek the death penalty causes the race of the victim.

I, also, find it pretty hard to explain the Baldus study results as simply a "statistical fluke." Such things are always possible, but modern statistical analysis guarantees us that they are exceedingly unlikely. Consequently, the following is also very low on my plausibility ranking.

t$_2$. It's just a coincidence that victims' race "correlated" with capital sentences in Georgia.

The only serious competitor I can imagine, therefore, is that there is some unnoticed "common cause" that is independently responsible for both the race of the homicide victims, and the fact that their murderers received the sentences they did. The Baldus team tried to think of some of the possible factors in their original evaluation of the data. That's what they were up to when they performed the

statistical tests that "controlled for over two hundred non-racial variables." Even when they did this, it turned out that murderers of white victims were 4.3 times as likely to receive death sentences. Maybe something else is responsible for the correlation, but we have yet to see what it is. Hence, I am willing to take the following seriously as a potential rival explanation.

> t_3. Some unidentified non-racial factor is responsible for the correlation of victim race and death sentences.

Since we have yet to even think of what this non-racial factor might be, I admit it's possibility, but rank it significantly lower than the racial explanation in t_0.

---

[1] **Dessler and Parson, p. 59.**

[2] *Ibid*, **p. 73.**

[3] *McClesky v. Kemp*, **Justice Brennan dissenting.**