

Chapter Five

STATISTICAL EVIDENCE

[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

1. Explaining the Numbers

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. Suppose I am struck by the remarkably high percentage of phone calls I received last night from worthy causes asking for money. I might generalize to a hypothesis about what is going on outside my home.

e_1 . Three phone calls last night; all from worthy causes.

=====

t_0 . Unusually high percentage of charitable solicitations last night.

Just a moment's reflection, however, yields a rival explanation that I

hope strikes you as every bit as good.

- t₁. It's just a coincidence that all of my phone calls last night came from charitable organizations.

On the other hand, 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.

- e₁. Of the 771 men in the 40-49 year age group, 29% showed some signs of coronary heart disease.

- e₂. Of the 954 women in the 40-49 year age group, only 14% showed signs of coronary heart disease.

- =====
- t₀. Coronary heart disease appears much more often in men than in women.

In this latter argument, as in the case of the phone calls, 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. Statistics can be seen as the 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.

2. Explaining the Correlations

If we are confident in our explanatory inference from the raw numerical data to some genuine connection or correlation, the interesting explanatory work has just begun. A Mississippi study established a remarkably strong correlation between hearing loss and being incarcerated in the state penitentiary -- almost fifty percent of the prisoners showed some hearing loss. The researchers hypothesized that this data provided strong evidence that hearing loss causally contributed

to a life of crime. The argument is a classic inference to the best explanation.

e₁. Forty-eight percent of the prisoners show some hearing loss.

t₀. Hearing loss [partially] causes a life of crime.

Given the inference that this forty-eight percent figure is not simply a fluke or coincidence -- for the modern social scientist, that it is statistically significant -- we must now determine whether t₀ is the *best* explanation of the correlation. We must compare it to some rival explanations. Perhaps, for example, the life of the career criminal is particularly hard on one's ears. All the necessary time on the firing range might be causally responsible for hearing loss.

t₁. A life of crime [partially] causes hearing loss.

I take it that such a rival is initially very implausible in this particular case. Much more challenging would be a rival account that brings in some third factor -- a common cause -- that causally explains the hearing loss, and independently explains the life of crime; child abuse comes immediately to mind.

t₂. Being an abused child often causes significant hearing loss (the actual mechanism might be either physiological, or psychological), independently, coming from a background of abuse causes antisocial behavior and often leads to a life of crime.

In our first inference when we were attempting to explain an observed quantitative relationship modern statistical theory was of great help. We might have, for instance, simply entered the numbers into a computer and the program simply told us whether our result was statistically significant (that a genuine correlation was the best explanation) or not (that the null hypothesis [the coincidence hypothesis] was the best). When it comes to sorting out explanations of genuine correlations the task is more difficult.

Returning to the Framingham study, gender seems to play some role

in coronary heart disease. But what is the actual causal mechanism? I can think of a number of possible explanations.

- t₁. Male and female hormones have different effects in the human circulatory system.
- t₂. "Traditional" male jobs [remember the study went back to the 1950s] are more stressful, causing a higher rate of coronary heart disease.
- t₃. Males are socialized to prefer a diet that is much higher in fat and salt.

It may well be that we will need to go out and uncover some additional data -- conduct some new studies or experiments -- before we are in a position to confidently commit ourselves to any particular explanation of an established correlation as being the best. At other times, however, will be confident about the relative explanatory plausibility as things stand.

3. *Some Very Disturbing Data*

Most of my students seem to believe that the problems of racial and sexual discrimination have been pretty much solved. Sure, we all know blatant racists or sexists, but at a national level the problem is much less severe than even twenty years ago. Many Americans believe that affirmative action is actually a worse social problem than the racism and sexism it was designed to combat. I think my students are dead wrong about this, and here's some statistical evidence to support that claim.

Labor market statistics suggest that there are great disparities between blacks and whites, both in terms of the characteristics they take into the labor market and in terms of the benefits derived from the labor market, as the following table illustrates.

RACIAL DIFFERENCES IN INCOME, POVERTY, AND EDUCATION 1991

RACE	% WITH 4 YRS COLLEGE	MEDIAN INCOME	POVERTY RATE	MEDIAN WKLY PAY

White	22.2%	\$36,915	10.7%	\$446
Black	11.5%	\$21,423	31.9%	\$348

In addition to the above racial differences in poverty rates, income and earnings differentials, and education levels, black employment tends to be concentrated in low-paying, low-status jobs and under-represented in high-paying, high-profile careers. Thus 30% of nursing aides are black, 29% of domestic servants are black, 25% of vehicle washers are black, and 21% of janitors are black, while only 0.7% of geologists are black), 1.5% of dentists are black, 2.1% of architects are black, and 2.6% of lawyers are black. Thus many employment opportunities for blacks appear to be concentrated in the secondary labor market, with low-paying, high turnover jobs, while careers in the primary labor market, jobs that are high-paying with advancement potential, are limited in their availability to blacks.

There are, of course, many other data that support the contention that blacks and whites differ in characteristics that significantly affect their economic success. Blacks are more likely to be high school dropouts and thus have a lower high school graduation rate; black children are 2.8 times more likely to be poor than white children; and blacks are twice as likely to be in the bottom 20% of the income distribution than whites. Given this wealth of data, how do we explain it?

As sad as things are with respect to the wealth and race, things are even worse with respect to health. In 1915, the Department of Labor, Children's Bureau issued a report on infant mortality which demonstrated the coincidence between infant mortality rates and socioeconomic factors, including low earnings and poor housing. Infant mortality is defined as death within the first 364 days after birth, and while infant mortality rates have declined over the last eighty years, the differential between whites and blacks remains virtually unchanged. Consider the data in the table below.

***Selected Pregnancy Outcome Rates By Race:
United States, 1970 and 1985***

	1970			1985		
	WHITE	BLACK	RATIO	WHITE	BLACK	RATIO
TOTAL MORTALITY*	29.8	55.1	1.85	16.2	30.5	1.89
FETAL MORTALITY*	12.3	23.2	1.89	7.0	29.3	1.81
INFANT MORTALITY*	17.8	32.6	1.84	9.3	18.2	1.95

Overall, the black infant mortality rate is about twice the white infant mortality rate. More specifically, black infants are 2.5 times as likely as whites infants to die of pneumonia and influenza; over twice as likely to die of maternal complications of pregnancy, interuterine hypoxia, and birth asphyxia; and 3.5 times as likely to die as a result of short gestation and unspecified low birth weight. Low birth weight infants are 5 times more likely to die during the first year of life than a normal birth weight infant, with black infants being twice as likely to be born with low birth weights. Although the factors leading to low birth weight are not completely understood, it is generally believed that inadequate prenatal care is a primary determinant and that socioeconomic status limits prenatal care, affects nutrition during pregnancy, increases stress, and increases the likelihood of harmful behavior during pregnancy, such as smoking and drug use. The weight of evidence seems to argue that poverty (and therefore race) strongly influence the adequacy of prenatal care, which is a determinant of low birth weight, which in turn is a leading cause of infant mortality.

If we schematize all of this data in the familiar form, we get an awfully powerful *prime facie* case that something is seriously wrong.

- e₁. There is a strong correlation between wealth and poverty, and race.
- e₂. There is a strong correlation between health -- infant mortality, morbidity rates, etc. -- and race.

-
-
- t₀. Blacks are at a distinct economic and health disadvantage, *because* of their race.

Many of you will be extremely uncomfortable with this diagnosis of the data; you will seek rival explanations. I fear that some will be tempted further down the racist path and suggest that these racial differentials can be explained in terms of worth, ability, and natural physiological differences. I almost hate to put it into words, but one rival explanation to be considered is the following.

- t₁. Whites are smarter and more-hard working, hence their superior economic success. Whites are also physiologically better off, hence the improved health statistics.

A slightly less offensive rival explanation is very popular in contemporary economics. According this view the individual is seen as a "rational actor." Each person makes individual decisions for him or herself, including decisions to "invest" in one's future occupation through training and education, and one's future health through the consumption of medical services and the adoption of a healthy lifestyle and diet. The model of "human capital" sees some of these decisions as rational -- blacks tend to have shorter work lifetimes and earn lower wages, hence there would be less return from investing in job training, for instance.

- t₂. Whites have more economic success, and a brighter health picture, because of differing decisions between blacks and whites regarding investments in human capital.

I suspect that the human capital model does get at something interesting. If I knew that I would only earn eighty cents on the dollar compared to a white co-worker, I might be less inclined to invest in my professional career. But notice that this explanation presupposes a relevant racial difference in our earning potentials. Furthermore, the human capital model can be carried to implausible extremes, as when economists suggest that pregnant women in the ghetto make "rational" investment decisions in the health of their unborn children.

Thus, I stand by the diagnosis of racism as the best explanation of at least part of the racial differentials in wealth and health. For a nation founded on the principle that "all men are created equal," and a nation

whose Constitution guarantees "equal protection of the laws," this is an incredibly damning account.

4. *Statistics and the Death Penalty*

I will be arguing in Chapter Seven that capital punishment is unconstitutional. Now, before we even begin that discussion I have to concede that our Supreme Court does not agree with me about any of this. And, of course, a majority of my fellow citizens, and I suppose a majority of my readers, feel quite strongly that we should have the death penalty in this country. Nevertheless, I will be defending the following interpretation of constitutional language and previous Supreme Court precedent.

- t₀. Capital punishment may be exercised by Congress or the State, if and only if, (i) there is due process of law, (ii) capital punishment is neither cruel nor unusual, and (iii) capital punishment is administered in a non-discriminatory manner that constitutes equal protection of the laws.

I want now to begin my case against the death penalty by arguing that the third necessary condition for permissible executions is 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 statute 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.

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

- e₁. 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 than 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.

- e₂. 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, however, that the judgments, both arbitrary and prejudicial, of other legal officials are even more problematic. I take it to be obvious that the best explanation of the Baldus data is the following.

- t₀. The race of the murder victim causally influences the decision whether to seek the death penalty.

What's the Best Explanation?

As I hope by this point you are all already thinking, the crucial question is whether t₀ 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.

- t₁. 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.

- t₂. 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.

- t₃. 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 t₀.