INTRODUCTION

The Worst Social Statistic Ever

The dissertation prospectus began by quoting a statistic--a "grabber" meant to capture the reader's attention. (A dissertation prospectus is a lengthy proposal for a research project leading to a Ph.D. degree--the ultimate credential for a would-be scholar.) The Graduate Student who wrote this prospectus* undoubtedly wanted to seem scholarly to the professors who would read it; they would be supervising the proposed research. And what could be more scholarly than a nice, authoritative statistic, quoted from a professional journal in the Student's field?

So the prospectus began with this (carefully footnoted) quotation: "Every year since 1950, the number of American children gunned down has doubled." I had been invited to serve on the Student's dissertation committee. When I read the quotation, I assumed the Student had made an error in copying it. I went to the library and looked up the article the Student had cited. There, in the journal's 1995 volume, was exactly the same sentence: "Every year since 1950, the number of American children gunned down has doubled."

This quotation is my nomination for a dubious distinction: I think it may be the worst--that is, the most inaccurate--social statistic ever.

What makes this statistic so bad? Just for the sake of argument, let's assume that the "number of American children gunned down" in 1950 was one. If the number doubled each year, there must have been two children gunned down in 1951, four in 1952, eight in 1953, and so on. By 1960, the number would have been 1,024. By 1965, it would have been 32,768 (in 1965, the FBI...
identified only 9,960 criminal homicides in the entire country, including adult as well as child victims. In 1970, the number would have passed one million; in 1980, one billion (more than four times the total U.S. population in that year). Only three years later, in 1983, the number of American children gunned down would have been 8.6 billion (about twice the Earth's population at the time). Another milestone would have been passed in 1987, when the number of gunned-down American children (137 billion) would have surpassed the best estimates for the total human population throughout history (110 billion). By 1995, when the article was published, the annual number of victims would have been over 35 trillion—a really big number, of a magnitude you rarely encounter outside economics or astronomy.

Thus my nomination: estimating the number of American child gunshot victims in 1995 at 35 trillion must be as far off—as hilariously, wildly wrong—as a social statistic can be. (If anyone spots a more inaccurate social statistic, I'd love to hear about it.)

Where did the article's Author get this statistic? I wrote the Author, who responded that the statistic came from the Children's Defense Fund (the CDF is a well-known advocacy group for children). The CDF's *The State of America's Children Yearbook--1994* does state: "The number of American children killed each year by guns has doubled since 1950."[1] Note the difference in the wording—the CDF claimed there were twice as many deaths in 1994 as in 1950; the article's Author reworded that claim and created a very different meaning.

It is worth examining the history of this statistic. It began with the CDF noting that child gunshot deaths doubled from 1950 to 1994. This is not quite as dramatic an increase as it might seem. Remember that the U.S. population also rose throughout this period; in fact, it grew about 73 percent—or nearly double. Therefore, we might expect all sorts of things—including the number of child gunshot deaths—to increase, to nearly double just because the population grew. Before we can decide whether twice as many deaths indicates that things are getting worse, we'd have to know more.** The CDF statistic raises other issues as well: Where did the statistic come from? Who counts child gunshot deaths, and how? What do they mean by a "child" (some CDF statistics about violence include everyone under age 25)? What do they mean "killed by guns" (gunshot death statistics often include suicides and accidents, as well as
homicides)? But people rarely ask questions of this sort when they encounter statistics. Most of the time, most people simply accept statistics without question.

Certainly, the article's Author didn't ask many probing, critical questions about the CDF's claim. Impressed by the statistic, the Author repeated it--well, meant to repeat it. Instead, by rewording the CDF's claim, the Author created a mutant statistic, one garbled almost beyond recognition.

But people treat mutant statistics just as they do other statistics--that is, they usually accept even the most implausible claims without question. For example, the Journal Editor who accepted the Author's article for publication did not bother to consider the implications of child victims doubling each year. And people repeat bad statistics: the Graduate Student copied the garbled statistic and inserted it into the dissertation prospectus. Who knows whether still other readers were impressed by the Author's statistic and remembered it or repeated it? The article remains on the shelf in hundreds of libraries, available to anyone who needs a dramatic quote. The lesson should be clear: bad statistics live on; they take on lives of their own.

This is a book about bad statistics, where they come from, and why they won't go away. Some statistics are born bad--they aren't much good from the start, because they are based on nothing more than guesses or dubious data. Other statistics mutate; they become bad after being mangled (as in the case of the Author's creative rewording). Either way, bad statistics are potentially important: they can be used to stir up public outrage or fear; they can distort our understanding of our world; and they can lead us to make poor policy choices.

The notion that we need to watch out for bad statistics isn't new. We've all heard people say, "You can prove anything with statistics."*** My title, Damned Lies and Statistics, comes from a famous aphorism (usually attributed to Mark Twain or Benjamin Disraeli): "There are lies, damned lies, and statistics."2 There is even a useful little book, still in print after more than forty years, called How to Lie with Statistics.[Note 3]

Statistics, then, have a bad reputation. We suspect that statistics may be wrong, that people who use statistics may be "lying"--trying to manipulate us by using numbers to somehow distort the truth. Yet, at the same time, we
need statistics; we depend upon them to summarize and clarify the nature of our complex society. This is particularly true when we talk about social problems. Debates about social problems routinely raise questions that demand statistical answers: Is the problem widespread? How many people--and which people--does it affect? Is it getting worse? What does it cost society? What will it cost to deal with it? Convincing answers to such questions demand evidence, and that usually means numbers, measurements, statistics.

But can't you prove anything with statistics? It depends on what "prove" means. If we want to know, say, how many children are "gunned down" each year, we can't simply guess--pluck a number from thin air: one hundred, one thousand, ten thousand, 35 trillion, whatever. Obviously, there's no reason to consider an arbitrary guess "proof" of anything. However, it might be possible for someone--using records kept by police departments or hospital emergency rooms or coroners--to keep track of children who have been shot; compiling careful, complete records might give us a fairly accurate idea of the number of gunned-down children. If that number seems accurate enough, we might consider it very strong evidence--or proof.

The solution to the problem of bad statistics is not to ignore all statistics, or to assume that every number is false. Some statistics are bad, but others are pretty good, and we need statistics--good statistics--to talk sensibly about social problems. The solution, then, is not to give up on statistics, but to become better judges of the numbers we encounter. We need to think critically about statistics--at least critically enough to suspect that the number of children gunned down hasn't been doubling each year since 1950.

A few years ago, the mathematician John Allen Paulos wrote Innumeracy, a short, readable book about "mathematical illiteracy."[Note 4] Too few people, he argued, are comfortable with basic mathematical principles, and this makes them poor judges of the numbers they encounter. No doubt this is one reason we have so many bad statistics. But there are other reasons, as well.

Social statistics describe society, but they are also products of our social arrangements. The people who bring social statistics to our attention have reasons for doing so; they inevitably want something, just as
reporters and the other media figures who repeat and publicize statistics have their own goals. Statistics are tools, used for particular purposes. Thinking critically about statistics requires understanding their place in society.

While we may be more suspicious of statistics presented by people with whom we disagree--people who favor different political parties or have different beliefs--bad statistics are used to promote all sorts of causes. Bad statistics come from conservatives on the political right and liberals on the left, from wealthy corporations and powerful government agencies, and from advocates of the poor and the powerless. In this book, I have tried to choose examples that show this range: I have selected some bad statistics used to justify causes I support, as well as others offered to promote causes I oppose. I hope that you and everyone else who reads this book will find at least one disconcerting example of a bad statistic presented in behalf of a cause you support. Honesty requires that we recognize our own errors in reasoning, as well as those of our opponents.

This book can help you understand the uses of social statistics and make you better able to judge the statistics you encounter. Understanding this book will not require sophisticated mathematical knowledge. We will be talking about the most basic forms of statistics: percentages, averages, and rates--what statisticians call "descriptive statistics." These are the sorts of statistics typically addressed in the first week or so of an introductory statistics course. (The remainder of that course, like all more advanced courses in statistics, covers "inferential statistics," complex forms of reasoning that we will ignore.) This book can help you evaluate the numbers you hear on the evening news, rather than the statistical tables printed in the American Sociological Review and other scholarly journals. Our goal is to learn to recognize the signs of really bad statistics, so that we won't believe--let alone repeat--claims about the number of murdered children doubling each year.

NOTES

* For reasons that will become obvious, I have decided not to name the Graduate Student, the Author, or the Journal Editor. They made mistakes, but the mistakes they made were, as this book will show, all too common.

** For instance, since only child victims are at issue, a
careful analysis would control for the relative sizes of the child population in the two years. We also ought to have assurances that the methods of counting child gunshot victims did not change over time, and so on.

*** This is a criticism with a long history. In his book *Chartism*, published in 1840, the social critic Thomas Carlyle noted: "A witty statesman said you might prove anything with figures."


2. Twain uses the expression in his autobiography, but refers to it as "the remark attributed to Disraeli."


------

Copyright © 2001 by the Regents of the University of California. Not to be reproduced without written permission of the publisher.