Book Review

Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists

Reviewed by Lynn Arthur Steen

Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists
Joel Best
The University of California Press, 2001
*196 pages, ISBN: 0-520-21978-3, $19.95*

In the late 1980s the press was filled with stories about an epidemic of crack babies, addicted at birth and afflicted with irreparable mental handicaps, that would create insurmountable problems for our medical, educational, and social systems. Years later it turned out that the crack baby problem, albeit serious, was only about one-tenth that which experts had predicted.

Similar hyperbole abounds in the media and professional literature: Reports that serial killers are responsible for as many as 4,000 homicides a year turned out to be exaggerated by a factor of ten. Reports that anorexia leads to 150,000 deaths a year were exaggerated by a factor of 20. And reports that white males would soon make up only 15 percent of U.S. workers turned out to be wrong by a factor of three. (Actually, the 15 percent refers to "net additions", not to the work force itself or even to new entrants to the work force.)

That many published "facts" are wrong (and often by an order of magnitude) would not surprise those who were raised on Darrell Huff's classic

---

Lynn Arthur Steen is professor of mathematics at St. Olaf College. His email address is steen@stolaf.edu.
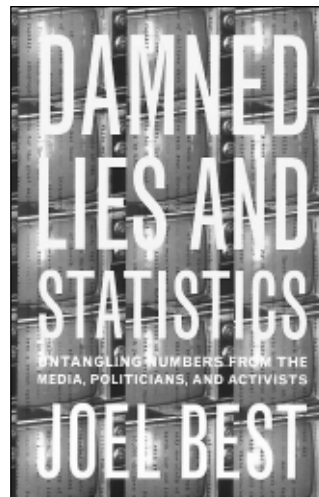
How to Lie with Statistics (Huff, 1954). The association of lies with statistics—carried on by the title of Joel Best's book—goes back at least to nineteenth-century British Prime Minister Benjamin Disraeli's reputed characterization of three kinds of lies in political life: "lies, damned lies, and statistics." Today this association remains as strong as in earlier centuries: Huff's book is still in print fifty years after its first publication, and Disraeli's "lies" quotation is, according to the popular BBC quiz show *Quote...Unquote*, the most quoted remark in the British media (Rees, 2002).

Educated citizens are rightly skeptical of statistics, especially when used by politicians, advertisers, and other advocates to promote particular causes. Mathematicians often have an additional reason to be suspicious: professional caution about the applicability of statistical inference, knowing that reality rarely conforms to assumptions on which these inferential models are constructed. Even when deployed in the most neutral and professional manner, the inexactitude of

statistical inference contrasts sharply (and for many, negatively) with the deductive certainty that is the hallmark of mathematical reasoning.

The reality is, however, that statistics have become the "facts" on which modern society is built. "After all, facts are facts," noted Leonard Henry Courtney, the British economist and politician in a speech on proportional representation at Saratoga Springs in August 1895. "Although we may quote one to another with a chuckle the words of the Wise Statesman, 'Lies—damn lies—and statistics,' still there are some easy figures the simplest must understand and the astutest cannot wriggle out of" (cited in Baines, 1896, p. 87). [It was this speech, supposedly, that led Mark Twain to attribute the "lies" quotation to Disraeli (Twain, 1924) and that has led many others to attribute the quotation to Twain.]

Much as we now take them for granted, "facts" were not always facts—at least not as we know them today, and certainly not described in quantitative terms. For the first several millennia of recorded history, most humans lived in a qualitative rather than quantitative milieu. Hours were not one-twenty-fourth of a day but times for monastic prayers; feet were not twelve inches but an anatomical comparison. Yet as early as the thirteenth century people began learning the value of imposing standardized measures (of length, of time, of money) through such innovations as mechanical clocks, perspective drawing, and double-entry bookkeeping (Crosby, 1997). Gradually numbers lost their ancient metaphysical meanings and became simply quantities devoid of qualities, which made them useful as tools for measuring just as the value of measuring things became apparent and the means to measure became available.

The increasing prominence of numbers (used to measure specifics) and arithmetic (used to aggregate individual numbers) enabled rudimentary statistics to mediate the transition from the Aristotelian tradition of facts as universals awaiting recognition to the modern scientific understanding of facts as particulars—specific, empirical, and individualistic. In the seventeenth and eighteenth century facts became "nuggets of experience detached from theory," and numbers came to epitomize (modern) facts, because they began to be seen as "preinterpretive or even noninterpretive" at the same time as they became "the bedrock of systematic knowledge" (Poovey, 1998).

In the early nineteenth century, when revolutions threatened social stability, the reporting of societal data (births, marriages, and deaths) and economic measures (agriculture, manufacturing, shipping) offered welcome hints of underlying social order. Decennial censuses became conventional, and counting things—populations, incomes, properties, jobs, crops—became a common and accepted political activity. Thus was statistics born as the "science of the state", but not without considerable contention. For example, the 1840 census, conducted on the eve of the Civil War and just one year after the founding of the American Statistical Association, generated intense political debate because an apparent gradient of black insanity rising from south to north seemed to support slave owners' arguments that slaves could not survive freedom (Cohen, 1982). Many years later this was discovered to be a statistical artifact that magnified certain routine errors of enumeration, but at the time emotions were too intense and statistical understanding too meager for anyone to see this.

The desire to comprehend society through quantitative facts lent the new field of (social) statistics considerable influence. Statistics offered an effective means of creating new universals by making separate facts "hold together". It appeared to make real such social abstractions as fertility, wealth, unemployment, and inflation. By focusing on objects purged of the "unlimited abundance of the tangible manifestations of individual cases," statistics helped objectify the social world (Desrosières, 1998). Slowly, numbers came to be believed *simply because they were numbers.*

Contrary to widespread belief, the drive for quantitative rigor during the last two centuries has been due not so much to increased demands of natural science as to social pressure for objectivity in political, economic, and social affairs (Porter, 1995). Indeed, the increased propensity (oftentimes demand) to offer numbers in support of arguments of all kinds gives at least the appearance of objectivity. Our contemporary drive for objective data—what we now think of as "facts"—is principally a recent cultural phenomenon. Witness the headlines featuring refinements about the death toll at the World Trade Center: surely the individual deaths mean a great deal, but does the precise total to four significant digits have any newsworthy meaning? This drive to count everything is so strong that numbers used to certify facts often take on what Porter calls "totemic" significance, the antithesis of the "lies" reputation that worried Disraeli, Twain, Huff, and now Best.

So how is it that so many widely disseminated facts are not facts at all? Followers of the lies-and-statistics school of thought would probably attribute common howlers either to deliberate misrepresentation or to innumerate reporting. This is the implicit message of Huff's masterpiece: it is so easy to lie with statistics that (almost) everyone does it. Joel Best's brief monograph, written for journalists and lay persons, makes a different and more subtle argument: Statistics not only measure but also create ("objectify") social issues, and the dynamics of objectifying social structures is inherently skewed in the direction of bad statistics.

Here's why, according to Best, a professor of sociology and criminal justice at the University of Delaware:

People who are concerned about some new social issue (e.g., child abductions) seek to justify their concern using the contemporary standard of objective fact, statistics. However, precisely because their concern is new, no one will have collected accurate, systematic data on the problem they are worried about. What little data there may be on child abductions are byproducts of other work, for example, police reports gathered in varied jurisdictions, subject to disparate definitions and uneven standards, or journalistic accounts of cases that appear specially newsworthy.

Faced with the lack of sound data in a political environment that demands numbers for legitimacy, those who are concerned about the new issue of child abductions choose from whatever numbers are available those that will draw greatest attention to their cause. Big numbers will justify their concern; small numbers will not. The media play along, because big numbers make more compelling news. Even experts favor big numbers, because it makes their work seem more important and justifies research grants. Small differences do not produce publishable results; big ones do. (This scenario assumes the best of intentions by all parties. Best's argument is not about the relatively few who fabricate or deliberately misrepresent data but about the way data flows through the hands of those who are trying to be fair and honest.)

As the problem (of child abductions) takes on public significance, better data becomes available and is reported in various sources. Inevitably, secondary sources introduce misinterpretations which make the statistics seem more dramatic. These "mutant statistics" get repeated precisely because they are compelling. Accidental transformations that make a statistic seem less dramatic are likely to be forgotten. Few can escape the Darwinian pressure that confers survival value on dramatic numbers.

Whether motivated by sincerity or opportunity, whether honestly enraged about some social ill or merely hired to advocate a new position in the press or the courts, those who speak for new causes prefer large numbers to make the problem seem serious and the need urgent. To achieve this end, advocates typically favor inclusive rather than well-focused definitions. (Does an abduction of a child by a divorced parent count the same as by a stranger? What about runaways?) Then there is the ever-present unaccounted "dark figure", which, like the dark matter in the universe, we know must exist but which we cannot see or measure. For every one abduction that gets officially counted, there may be two (or ten) that do not.

Once in circulation, mutant statistics are difficult to retract, especially when the number is large and dramatic: drama ensures repetition, while public innumeracy inhibits critical thinking, even if the number is wrong by an order of magnitude (Paulos, 1988; Dewdney, 1993). Three centuries ago Samuel Johnson quipped that "round numbers are always false." Now Joel Best explains why and updates Gresham's Law: Bad statistics drive out good ones.

According to Best's analysis, social problems are constructed through the activity of people who identify, name, describe, measure, and promote their significance. A widely recognized name turns a condition we take for granted into something we consider troubling and worth measuring. Most commonplace statistics (e.g., the consumer price index, minority unemployment, breast cancer rates) evolved in this manner: only after someone (or some organization) agitated about their importance did systematic measurement begin. In other words, (social) statistics are a product of social activity, not just a representation of society. All statistics are social constructs: they are how we make the world meaningful. Yet too often we treat socially constructed numbers as nuggets of indisputable truth.

Mathematicians will recognize that the "statistics" under discussion here is the plural of the word *statistic* (meaning "numerical fact"), not the singular "science of data" that is the subject of high school and college courses in statistics. Most popular attempts at demystifying statistics are about the latter; Best's monograph is almost entirely concerned with the former. Best eschews standard topics such as probability, polls, correlation, and regression in favor of case studies ranging from AIDS and alcoholism to traffic fatalities and victims of crime. As Best amply illustrates, every such statistic is the result of human choices and thus is as much the product as the reflection of social reality.

Damned Lies and Statistics thrives on relevance: nearly all its examples are about important contemporary issues where competing claims about statistics have shaped policy debates in Congress and state legislatures. Recognizing the statistical and mathematical illiteracy of his intended audience, the author rarely discusses any mathematical idea more complicated than a percentage. Instead, Best uses a wide variety of examples to illustrate the manifold ways in which bad statistics can so easily be created: bad guesses, deceptive definitions, confusing questions, biased samples, inadequate measurement, overgeneralization, incomparable comparisons (different times, places, or social groups), public innumeracy, and more.

The aim of Best's book is similar to Cynthia Crossen's *Tainted Truth: The Manipulation of Fact*

in America (Crossen, 1994). But Best's book differs in two important respects: First, it is half as long as Crossen's, thus more accessible to a broad audience. More important, whereas Crossen relentlessly documents distortions related to research for hire (where the research design produces results favored by sponsors of the research), Best focuses on natural and mostly innocent forces that distort data. In many respects, Best's analysis is the more alarming, since it illustrates how "mutant facts" can infiltrate where one least suspects.

Despite its brevity—170 small pages—*Damned Lies and Statistics* is a slow read. It is somewhat repetitive and lacks both the humor that made Huff's monograph a classic and the passion of Crossen's treatment. Readers will gain a good deal of caution but will not learn much about how one should create good statistics nor how to transform bad ones into better ones. Best says next to nothing about inferential statistics, the subject of virtually every introductory course on the subject. There are very few good examples in the book: it is not a handbook about creating good statistics.

Best's goal is different from these other treatments and in some ways more scholarly. He sets out to document a sobering social theory of facts in our number-crazed age, namely, that statistics—the plural word, not the singular—are primarily social products (not social measures) subject to inherent forces that skew numbers in the direction of the large and the dramatic. As such, statistics must be approached first, not with the tools of Minitab or SPSS, but with the skepticism of a good investigative reporter: who created them, why were they created, what was their intended purpose, and how accurate might they be?

One might wonder why a monograph devoted to data that is virtually devoid of standard inferential or quantitative analysis should stir up so much interest in higher education—a lengthy excerpt was printed in the *Chronicle of Higher Education* (Best, 2002)—or be reviewed in the *Notices*. I suspect that one reason is the recent widespread recognition that numeracy is a failing of our educational system parallel to, but different from, the system's well-known weaknesses in mathematics (e.g., Steen, 2001; Madison, 2002).

But another reason, noted by Best, is a poorly recognized paradox of our educational priorities: the quantitative devices subject to the deceptions that Best analyzes depend on only the simplest of mathematics—averages, percentages, rates—yet our educational strategies focus primarily on more advanced aspects of statistics (and mathematics), overlooking many sources of corrupt data simply because the underlying mathematics appears too simple to worry about. To help students deal with the deluge of quantitative facts, we need to find some way to reinforce and extend their sophistication in using the most elementary aspects of mathematics to think about data while at the same time continuing to advance their knowledge and skills in more advanced mathematics. Without such skills for thinking about data, students will be left to the mercy of numbers as totems, forever thinking of statistics "as facts we discover not as numbers we create" (Best, p. 160).

References

[1] J. A. Baines, Parliamentary representation in England illustrated by the elections of 1892 and 1895, *J. Royal Stat. Soc.* **59** (March 1896), 38–124.

[2] Joel Best, Telling the truth about damned lies and statistics, *The Chronicle Review* (May 4, 2001), B7–B9.

[3] Patricia Cohen, *A Calculating People: The Spread of Numeracy in Early America*, Univ. of Chicago Press, Chicago, IL, 1982, 1999.

[4] Alfred W. Crosby, *The Measure of Reality: Quantification and Western Society, 1250–1800*, Cambridge University Press, Cambridge, UK, 1997.

[5] Cynthia Crossen, *Tainted Truth: The Manipulation of Fact in America*, Simon & Schuster, New York, 1994.

[6] Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*, Harvard University Press, Cambridge, MA, 1998.

[7] A. K. Dewdney, *200% of Nothing: An Eye-Opening Tour through the Twists and Turns of Math Abuse and Innumeracy*, John Wiley & Sons, New York, 1993.

[8] Darrell Huff, *How to Lie with Statistics*, W. W. Norton, New York, 1954.

[9] Bernard L. Madison, Educating for numeracy: A challenging responsibility, *Notices Amer. Math. Soc.* **49** (February 2002), 181.

[10] John Allen Paulos, *Innumeracy*, Hill and Wang, New York, 1988.

[11] Mary Poovey, *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*, University of Chicago Press, Chicago, IL 1998.

[12] Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton University Press, Princeton, NJ, 1995.

[13] Nigel Rees, *Quote…Unquote*, BBC Radio 4, URL `http://www1c.btwebworld.com/quote-unquote/p0000149.htm`.

[14] Lynn Arthur Steen, *Mathematics and Democracy: The Case for Quantitative Literacy*, National Council on Education and the Disciplines, Princeton, NJ, 2001.

[15] Mark Twain, *Mark Twain's Autobiography*, Harper & Bros., New York, 1924.