PEOPLE COUNT: THE SOCIAL CONSTRUCTION OF STATISTICS

Joel Best, University of Delaware
322 Smith Hall, Newark, Delaware, 19716-2580, USA

Abstract
Social statistics often play a key role in discussions of social issues. Estimates for the extent of social problems and other similar figures tend to be treated as objective, factual, indisputable evidence. This assumption ignores the social processes involved in the creation and dissemination of these numbers – processes that are often glossed over in statistics courses. Statistical education ought to foster critical thinking about the social construction of statistics.

Keywords: social construction, social statistics, statistical education

All statistics are products of social activity–the process sociologists call social construction. While this might seem painfully obvious, this point is almost completely ignored when we think about--and particularly when we teach–statistics. Statistics instruction tends to focus on matters of calculation: that is, students are taught a measure's underlying logic, the formula used to compute the measure and/or the software commands used to extract it from the computer, and some guidelines for interpreting the numbers that result from these computations; once this sequence is completed, the class proceeds to the next measure. These are complicated lessons: few students have an intuitive grasp of any but the simplest statistics; and instruction tends to focus on trying to clarify the computational complexities.

By and large, this instructional process ignores the sorts of basic statistics that appear in newspaper articles and on news broadcasts, and that shape the public's understanding of social issues. These figures are simple descriptive statistics–percentages, averages, ratios, rates and the like. They pose few interesting mathematical problems, and most statistics instructors deal with them in their initial lectures, if at all.

Unfortunately our culture assigns statistics "facticity," that is they are presumed to be facts. We tend to envision statistics as little nuggets of truth that we uncover, much as rock collectors find stones. After all, we think, a statistic is a number, and numbers seem solid, factual, proof that somebody must have actually counted something. But that's the point: somebody had to do the counting. We'd do better to think of statistics as jewels: jewels must be selected, cut, polished, and placed in settings so that they can be viewed from particular angles. In much the same way, people create statistics; they choose what to count, how to go about counting, and which of the resulting numbers they will share with others. Numbers do not exist independent of people; understanding numbers requires knowing who counted what, and why.

This is what is meant by saying that statistics are socially constructed. Sociologists use social construction to refer to the process by which people assign meaning to the world. This term means different things in different disciplines, but when sociologists use it, they do not imply that something is false or fanciful. All statistics, from the best to the worst, are socially constructed.

An example: In April, 2001, an article about bullying appeared in the Journal of the American Medical Association (JAMA) indicating that 30 percent of students in the sixth through tenth grades had moderate or frequent involvement in bullying. This article received widespread coverage in various news media. JAMA sends out press releases about articles in its current issue that its editors hope will prove newsworthy. JAMA's visibility encourages submissions from top researchers by offering opportunities to bring their work to the notice of, not just fellow professionals, but the larger public. And researchers who successfully place their work in prestigious journals in turn please their funders–the government agencies or private foundations that supply the grants needed to pay for large-scale research; knowing that their grants led to highly visible publications confirms that the funders' money was spent wisely.

It is important to appreciate that this is an extremely competitive process. Newspapers are flooded with press releases, and must choose which ones will be carried in the increasingly limited space they have available for news. Would-be authors submit many more manuscripts to JAMA than it can publish, and the journal's editors must not only choose among them, but also must decide which articles merit press releases. And, of course, funding agencies have to winnow through many grant applications to select those worthy of support.

Now, let's try a little thought experiment. Remember that bullying affected about 30 percent of students. Imagine that the study had discovered that bullying was one-third as common, that it affected 10 percent of students. The finding now seems less

impressive, doesn't it? We can suspect that reports featuring a 10 percent figure would have resulted in less news coverage, that JAMA's editors might have been less likely to circulate a press release or even publish those results, that the authors might have been less likely to submit their papers to such a highly-selective journal, and that the funding sources also would have been less impressed with the reception given the published results. In other words, we can suspect that everyone in the process–in the media, at JAMA, the researchers, and the funders–might prefer the bigger percentage.

I am not suggesting that anything fraudulent is involved in this process, nor am I criticizing the quality of this particular research article (which featured a very large, well-drawn sample and careful analysis). Still, the study's authors had to make a set of methodological choices.

In this study, students were asked whether they bullied others or were themselves bullied, and they were asked how often this bullying occurred. The key finding–that 30 percent of youths are involved in bullying--depended upon two choices. First, the authors combined all students who reported bullying others and all students who reported being bullied into students who were "involved" in bullying. Second, they classified bullying that occurred at least once a week as "frequent" and bullying that occurred "sometimes" as moderate. The product of these choices was that 30 percent of students had moderate or frequent involvement in bullying. Other choices would have produced other results–the percentage of students who reported being victims of frequent bullying, for example, has about 8 percent. Both figures–30 percent and 8 percent–appear in the report, but it was the larger number that was featured in the article's abstract, the press release, and the resulting media coverage.

The point is not that there is something misleading about this research. Nor is it atypical. Two months later, JAMA publicized an article showing that 20 percent of youths who used the Internet frequently reported unwanted sexual solicitations; two months after that, JAMA promoted a finding that 20 percent of female high-school students had experienced physical or sexual abuse during a date. A close reading of those articles reveals the same sorts of methodological choices shaping the results.

Statistical instruction tends to whisk past any consideration of how real-life statistics come into being. All statistics are products of choices and compromises that inevitably shape, limit, and distort the outcome. Statistics instructors tend to dismiss this as a melodramatic irrelevance. A surprisingly number of book titles draw a distinction between statistics and lies: *How to Lie with Statistics*, *How to Tell the Liars from the Statisticians*, *How Numbers Lie*, even my own *Damned Lies and Statistics*. It is as though statistics are pure, unless they unfortunately become contaminated by biased people with their bad motives.

But all statistics are produced by people who have reasons for choosing to count particular things in particular ways. Statistics instruction needs to address this social process. It needs to concern itself with matter of construction–as well as calculation.

REFERENCES

Best. Joel (2001). *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*. University of California Press.

Hooke, Robert (1983). *How to Tell the Liars from the Statisticians.* Marcel Dekker.

Huff, Darrell (1954). *How to Lie with Statistics*. W.W. Norton & Company.

Mitchell, Kimberly J., David Finkelhor and Janis Wolak (2001). *Risk Factors for and Impact of Online Sexual Solicitation of Youth.* JAMA 2001;285:3011-3014

Nansel, Tonja R., Mary Overpeck, Ramani S. Pilla, W. June Ruan, Bruce Simons-Morton and Peter Scheidt (2001). *Bullying Behaviors Among US Youth: Prevalence and Association With Psychosocial Adjustment.* JAMA. 2001;285:2094-2100.

Runyon, Richard P. (1981). *How Numbers Lie*. Lewis Publishing.

Silverman, Jay G.; Anita Raj, Lorelei A. Mucci and Jeanne E. Hathaway (2001). *Dating Violence Against Adolescent Girls and Associated Substance Use, Unhealthy Weight Control, Sexual Risk Behavior, Pregnancy, and Suicidality*. JAMA. 2001;286:572-579.