

Lying with Statistics: The Case of Campaign Contributions and Iraq Reconstruction Contracts

By Jack Glaser
Assistant Professor
Goldman School of Public Policy
UC Berkeley

This article appeared in the March 2004 *PolicyMatters* (Vol. 1, Issue 1, pp. 55-57)
(<http://www.policy-matters.org/>)

In his November 11, 2003 column in the New York Times, David Brooks provided a nice illustration of how misleading statistics can be. Brooks cited an analysis reported by Daniel Drezner on Slate.com (Nov. 3, 2003) to conclude that "there is no statistically significant correlation" between the size of campaign contributions to President Bush and the size of contracts awarded to companies working in Iraq. Drezner, asserting that the Center for Public Integrity, whose dataset he used, "has no evidence to support its allegations" of political favoritism, reports a correlation of .192 (small to medium in size by scientific standards) with a sample size of 70. When Brooks and Drezner say the correlation is "not statistically significant," they mean that it does not meet a conventional standard used to determine if a result should be relied upon.

Usually, we use a five percent chance of a false positive result as our cutoff, somewhat arbitrarily, to call something "statistically significant." This vaunted " $p < .05$ " basically means that, based on the size of the effect and the size of the sample from which it was calculated, there is less than a .05 probability that the observed relation (correlation, difference in averages or

percentages, etc.) one has obtained is merely due to chance, perhaps poor sampling, as opposed to reflecting a real state of affairs in the population in which one is interested. There seems to be something about a one-in-twenty chance that people are comfortable with.

The significance level for Drezner's .192 correlation would be .056¹ (meaning a 5.6% chance that the relation observed in the sample does not reflect a real one in the population). But it is irresponsible to utterly dismiss a finding that comes that close. There is no magical difference between a .05 probability and a .056 probability! Would you disregard a result with a p-value of .051, but take your .049 to the bank?

Policy analysts should be especially wary of falling prey to .05 demagoguery. Our samples are often small, and smaller samples have higher p-values. It is our job to make accurate assessments and projections, and significance testing is useful in giving us a sense of the confidence we can have in our results, but it should not

¹ This is a "one-tailed" p-value, which is appropriate because there is a clear, *a priori* directional claim (that the correlation is greater than zero) that is being tested.

necessarily lead us to reject useful information based on inflexible adherence to an arbitrary standard. On the other side of the spectrum, many samples are so large that even trivial effects are “statistically significant,” but they may not be meaningful. Rigid use of the $p < .05$ criterion to determine if something is worth reporting can prove misleading under these conditions as well.

The eminent psychologist and statistician, Jacob Cohen, in fact, published a forceful essay challenging the orthodoxy of the .05 criterion. He titled the paper, with tongue firmly planted in cheek, “The Earth is Round ($p < .05$).” As a result of efforts by Cohen and other respected statisticians, social scientists are moving away from an over-reliance on p-values, focusing increasingly on the actual size of the effects in question, whether or not they replicate, and other approaches.

Brooks’s second-hand report skipped over the correlation coefficient, so those who don’t read Slate didn’t even have a chance, unless they went snooping, to judge the effect size for themselves or see just how not statistically significant it was. This further illustrates the pitfalls of judging results by the dichotomous standard of whether the p-value is greater than or less than .05. Once an effect gets tagged “not significant” it loses all nuance.

Having said all that, the point is somewhat moot. Huh? Why? Because in Drezner's analysis he was not really working with a "sample" but rather with the data from essentially the whole "population" (or very close to it) of contractors working in Iraq. Remember, the point of significance testing – of calculating that p-value – is to generate an estimate of how likely it is that the result observed in the sample is representative of

the population. So with population data it is meaningless to engage in this kind of significance testing. The correlation is what it is. The one in question, .192, is not huge (indeed there must be many other factors, such as appropriateness of capabilities and negotiating skills, that predict contract size) but it's clearly greater than zero.

People who question that there is a *quid pro quo* in Iraq reconstruction contracting would be well advised to conduct an analysis that takes into account the size of the companies in question. Bigger companies are more likely to give bigger contributions *and* get bigger contracts (although General Electric is a gargantuan company that made mammoth campaign contributions but got relatively small contracts, perhaps merely due to a mismatch of capabilities and needs, so that may undermine such an attempt).

There must be many factors that contribute to whether or not a company is awarded a contract and the size of the contract. Careful consideration of such factors might indeed explain away any observed correlation between campaign contribution and contract size. But until then the most reasonable interpretation of Drezner’s result as reported in Slate (and misused in The New York Times and elsewhere) is simply that there is a relation between campaign contributions and contract size, and it should not be so readily dismissed by statistical sleight (or Slate?) of hand.