

PRESIDENTIAL ELECTION POLLS

Lars Perner, Ph.D.
Assistant Professor of Marketing
San Diego State University, Imperial Valley Campus
720 Heber Avenue, Calexico, CA 92231
(760) 768-5614
Internet: lperner@mail.sdsu.edu <http://www.LarsPerner.com>
<http://www.ConsumerPsychologist.com>

September 10, 2004

With public opinion polls showing a tight race again in this year's Presidential election, it might be useful to take a look at how polls are conducted.

A poll conducted by *Newsweek* from September 2-3, 2004 shows 52% for Bush and 42% for Kerry among "likely voters." This poll has a "margin of error" of +/- 3%. It might be noted that this is actually one of the larger differences we have seen for a while. Had we been closer to a 50-50 split, it would have been more difficult to determine who was ahead. Even at 52% vs. 48%, we could not say with much certainty who was actually ahead.

Why such a large margin? The above survey was based on a survey of 1,008 respondents. With so many people, why can't we make more precise determinations?

Computing the margin of error is based on the idea of a "confidence interval" within statistics. If we tossed a coin ten times and got six heads, we would not have much reason to doubt that the coin was unbiased. On the other hand, if we threw 10,000 and got 60% heads, we would worry more.

To determine a confidence interval, we start by determining the "standard deviation" of the percentages for a given sample size. First, we need to make a slight modification since 52% and 42% figures do not add up to 100%--people could vote for a third candidate or still be undecided. Therefore, we will base our calculations on those who are--and those who are not--supporting one of the candidates. For Bush, we have 52% for (p) and 48% "not for" (1-p).

The formula for the standard deviation of a "proportion"--or percentage--is

$$\sqrt{\frac{p(1-p)}{n}}$$

That is, the square root of the quantity p times 1-p over n.

To put in actual numbers:

$$\sqrt{\frac{0.52(0.48)}{1,008}} = 0.0157$$

To be 95% certain—an arbitrarily selected standard favored by many statisticians in the U.S.--that our confidence interval will cover the true proportion, we must go 1.96 standard deviations on either side of the sample proportion. (You would have to look up this number in a table). We thus have

$$1.96(0.0157) = 0.0308.$$

The 95% confidence interval for those supporting Bush is therefore 0.52 +/- 0.0308 or from 48.92% to 55.08%. Repeating the calculations for Kerry, his confidence interval turns out to be 38.95% to 45.05%.

It is interesting to observe that the margin of error is so large precisely because the two candidates are so close to each other. Note that, in the numerator of the formula for the standard deviation, the figures p and $(1-p)$ figure prominently. $0.5(0.5) = 0.25$, which is much larger than the $(0.99)(0.01) = 0.0099$ figures we would have if we were determining the margin of error for a 99% to 1% race. From an intuitive point of view, looking at any one voter, it is difficult to tell how he or she will vote in a 50-50 race. In a 99-1 race, not only is the sample outcome overwhelming, but on top of this, the margin of error will be much less because there is little doubt how each individual voter will behave.

In principle, we could reduce the margin of error proportionately by increasing the sample size. If we had 3,000 respondents instead of 1,008, the margin of error would be only about 1%. Sampling such larger groups would be more expensive, and because of the time sensitive nature of polling data, this increase could also result in delays in finishing the poll.

An important issue in conducting opinion polls is the question of who should be surveyed. The above survey is based on "likely voters." If the sampling base is all eligible--or at least all registered--voters, Kerry's figures appear to be higher. From the point of view of the representation of people in a democracy, that may be a more appropriate figure. However, it is a reality that only those voters who actually turn out will decide the election outcome. Thus, the estimate based on "likely voters" is likely to represent a more accurate prediction of actual election results if the election were held at the time of the election. It should also be noted that despite the best efforts, it is not possible in practice to identify a perfect sampling base of either all eligible voters or those most likely. Some individuals that should have been included will be omitted from any practical survey, and some that are not eligible will be included by mistake. Professional pollsters, however, do the best they can do to minimize such biases.