

Little Stat Helper

By Jonathan Simon, Election Defense Alliance

The benefit of statistical sampling lies in the surprisingly strong power of a small part to predict the behavior of a large whole. Although we tend to accept the results of polls and other research based on sampling, most people if they really thought about it would find it quite a head-scratcher that you could predict with great accuracy the preferences of a nation of 300 million--whether in candidates, policies, or favorite kind of cheese--by questioning a mere 3000, or .001%, of them. This is nevertheless the case, providing that certain conditions apply and certain procedures are followed.

To get highly reliable results it is important that the sample be done as randomly as possible. If bias or convenience enter into the sampling process all bets are off and statistical process loses its "crispness," the cut-and-dried rule of simple equations. More on that in a moment, but the key stumbling block I have found for acceptance of statistical sampling is the mind's natural and intuitive protest that larger wholes require larger samples. Once we reach a whole of a certain size--i.e., the size we are dealing with in federal elections--though, larger wholes in fact do not require larger samples, however counterintuitive this may seem. Not only has this been established theoretically, it has also been demonstrated in thousands of experiments. It's just the way it is. High-school statistics, chapter 1.

Now, bearing this basic concept in mind, here are the terms that get thrown around, and that must be understood to have a rational discussion of any statistical protocols and proposals:

Margin of error (MOE) of a sample refers to the range in which we expect to find the discrepancy between the count of the sample and the count of the whole from which the sample is drawn. In most research a +/- x% MOE means that 95% of the time, or 19 out of 20 times, we expect the count of the sample to fall within x% of the count of the whole; that is, the "confidence level" for that MOE is 95%, the standard used in most scientific research. 5% of the time a random sample with an x% MOE will count up more than x% off from the whole.

That's just the way it is. You don't get 100% certainty. But what *is* certain is that if you ran that sample a billion times, the number of times it missed the whole by more than x% would approach 50 million (5%) very closely. That's why computer simulations are so helpful, because you can actually do this and check the results.

A confidence level of 99% (or better), which I've recommended for votecount checking, would tell you to expect a result within the MOE 99 out of 100 times (or better).

The size of the whole numerically has virtually zero impact on the size of the sample needed (once you get above a whole of, say, 50,000; although there is a simple formula, irrelevant here, used for adjusting for such small targets).

As I've said, probably the most difficult and counter-intuitive thing to swallow about sampling is that you don't need to increase your sample size when the size of the target whole jumps from, say, 1,000,000 to 250,000,000; your 30,000 ballots would work about as well as a sample of the whole country as of Rhode Island. Hard to accept but it's true and very elementary statistics.

Given a random sample of a large (numerous) whole, the MOE and Confidence Level as defined above can be calculated quite easily. For a competitive election (60%-40% or closer), the magic formula boils down to: MOE at 95% Confidence = $1/\sqrt{\text{number of ballots sampled}}$ (generally referred to as "N").

So, to plug in a few numbers: If you sample 10,000 ballots, then $1/\sqrt{N} = 1/100 = 1\%$, and you'd say that your MOE is +/-1%. You would expect the sample results to differ from the total tabulated results by less than 1% in 19 out of 20 such elections. If you looked at, say, 1000 such elections, you'd find that the sample/whole difference was less than 1% in just about 950 of them. The more elections you ran, the more exact that 19 out of 20 would become.

If you sampled only 400 ballots, then $1/\sqrt{N} = 1/20 = 5\%$, and you'd say the MOE is $\pm 5\%$, and you'd expect the sample results to differ from the total tabulated results by less than 5% in 19 out of 20 such elections. If you sampled 30,000 ballots then $1/\sqrt{N} = 1/173.2 = 0.58\%$, and you'd say your MOE is $\pm 0.58\%$, and you'd expect the sample results to differ from the total tabulated results by less than 0.58% in 19 out of 20 elections.

Now all those examples presumed a Confidence Level of 95%, the standard for most research. But what that would mean is that if the MOE were used as a trigger for full hand counts or any other relatively drastic check of the results, elections officials would be obliged to proceed to such a step once in every 20 elections or races in the absence of mistabulation, in essence because we set the trigger at an MOE that's only expected to "work" 19 out of 20 times. The Confidence Level that is standard for most research would probably be seen as inadequate for checking on elections.

Fortunately, given a sample size N , a MOE can be easily calculated for any Confidence Level. To find the MOE at a 99% Confidence Level, for example, just take the MOE numbers above and multiply by 1.29: 10,000 ballots would give you a MOE of $\pm 1.29\%$; 400 ballots would give you a MOE of $\pm 6.45\%$; 30,000 ballots would give you a MOE of $\pm 0.75\%$; all at 99% confidence. This would mean that in only one out 100 elections would the difference between sample and whole exceed the MOE in the absence of mistabulation. We believe that one such "false positive" per 100 races would be tolerable to most BOEs (especially since the sample can be run again--that is, resampled--after such a result, rather than proceeding directly to a full hand count).

By the way, the "magic number" of randomly sampled ballots needed for a $\pm 1\%$ MOE at a 99% Confidence Level is about 16,500, as can be checked on a nice website <http://www.raosoft.com/samplesize.html>, very helpful in such calculations.

And, to illustrate an earlier point about the irrelevance of the size of the whole: For a state with 5,000,000 votes, you'd need 16,533 ballots, for a state with 10,000,000 votes, you'd need 16,560, and for a country with 100,000,000 votes, you'd need 16,585. To boost the confidence level to 99.9%, so that you could tell a BOE that they'd have to deal with a "false alarm" only once in 1000 elections, the magic number would be 27,000 ballots.

For a given venue (be it state or Congressional District) of known or predictable size (i.e., number of votes expected to be cast) coming up with sample sizes is child's play, just a question of plugging in a few numbers on a calculating website such as that given above. What's left to tackle is randomness.

There are several factors that can get in the way of randomness in sampling, but in the context of elections they all boil down to convenience or bias. And, given the proper protocol, they all can be avoided. Bias generally crops up when interviews are necessary, as with polling and exit polls. Interviewers may select respondents they "like" rather than say every 7th person to walk through the door; they may frame questions in a leading manner; they may hear what they want to hear in the response and mark it accordingly. Respondents, for their part, may be more likely to participate with an interviewer *they* like, or may give the interviewer answers the respondent thinks the interviewer wants to hear.

All of these possibilities create higher potential for error and are very difficult to quantify. Convenience can take the form of trying to capture your respondents in "bunches," such as at a few precincts (no exit pollster, for example, has the resources to send interviewers to every precinct, so they pick a few precincts carefully based on their likelihood to reflecting the whole), or at a certain time of day, or from the top of a big stack of ballots. Here too error is increased, in a way that is very difficult to quantify, turning statistics from crisp to soggy, straight science to a science-art hybrid.

A well-designed and administered hand-count sampling of ballots avoids all of these pitfalls, and is indeed "crisp" (and in this way very different from exit polling and targeted audits). Since we're counting ballots rather than interviewing, the bias pitfalls are eliminated. Since we propose counting a fixed proportion of the ballots at *every* precinct (rather than counting all or some of the ballots at *selected* precincts, as some have proposed), we avoid the principal convenience pitfall of a "clustered" sample.

All that remains is to insure that the ballots to be sampled at each precinct are, in effect, "shuffled" and a good random sample drawn. This can be achieved by literally shuffling the ballots after they are retrieved from their bin and then selecting ballots from the pile according to a predetermined choosing scale: say, every 15th ballot or every 87th ballot, depending on the overall number needed from the venue. With a modicum of observation and supervision, a random sample can be guaranteed.

Questions will surely arise as to cheating, attempting to rig the handcount sample as well as the machine count. The best answer is in the purpose of the handcount protocol and what cheating would in fact achieve. Since, unlike that of the machine count, the purpose of the handcount is not to get as many votes as possible for your guy, but to match the machine count within the MOE and thereby avoid a full handcount, the incentive for "stuffing" the handcount with extra ballots for your guy wherever possible vanishes.

In fact election officials' goal becomes to do the handcount sample as accurately as possible in order to avoid triggering a full hand count. Even granting that a given official or group of officials knew that the machine count had been rigged to add an extra 5% to "their guy," consider how difficult it would be for them to add the necessary number of handcount ballots to hit that rigged number within the MOE, given that it would have to be done in dozens if not hundreds of precincts in view of both partisan and neutral observers. Effectively impossible.

Now let's turn to a few more concrete numbers. In a Congressional District (CD), a competitive race draws between 200,000 and 250,000 voters. Picking the lower bound, we can achieve a +/-1% MOE at 99% Confidence Level by sampling about 15,000 ballots, or 7.5% of the total cast. Given an average precinct size of 500, that would work out to an average of just about 40 ballots to count at each precinct. Not very labor intensive.

Looking at a medium sized state such as Ohio, with 5,000,000 voters and 11,000 precincts, we'd need about 16,500 ballots, or 0.33% of the total cast. This works out to an average of just 1.5 ballots per precinct. This is so easy that it leads naturally to the idea of a bigger sample, so that the Confidence Level can be improved even further. And indeed it turns out that you can reach a 99.99% Confidence Level (that is, one false alarm in every 10,000 elections!) with a +/-1% MOE by sampling 37,500 ballots or an average of less than *four* (4) per precinct. Woo hoo!

In the state of California a Confidence Level of 99.99% requires those same 37,500 ballots, which boils down to an average of less than *two* ballots per precinct. In such a large venue, therefore, we can do even better: we could go to a MOE of +/- 0.5% at 99.99% confidence by taking 150,000 ballots, or about eight ballots per precinct. Such a MOE would sound the alarm on outcome altering mistabulation of any race decided by greater than one half of one percent. Woo hoo times two!

In conclusion, we propose a uniform, omni-precinct, proportional handcount sampling of ballots be used as the most reliable check mechanism of machine counts, where full hand counting of paper ballots is not yet an acceptable alternative. Such a protocol obeys "crisp" laws of statistics and is highly reliable, with little or no incentive for gaming or practical way to do so.

It can be implemented where paper ballots are in use, whether with opscan systems or, somewhat more problematically, where DREs are fitted with a paper ballot printer. The labor involved at the precinct level is reasonable and within the capacities of virtually every local BOE. The consultant work to generate the parameters is also minimal.

The uniform, omni-precinct, proportional handcount sampling of ballots is a viable and practical protocol that can be rapidly implemented to serve as a check mechanism on computerized recording and tabulation of votes where full hand counts have not been adopted.

Download the UBS paper here:

http://www.electiondefensealliance.org/files/New_UBS_811Update_061707.pdf