Context for Interpreting SurveyUSA Research Conducted Using Non-Probability Samples

\- - -

The Evolution from
"Margin of Sampling Error"
to
"Credibility Interval"

For 25 years, SurveyUSA research has been conducted using *probability* samples, where every member of the respondent population has a known, non-zero chance of being included. While there are many sources of error that occur in any opinion survey --- even a survey that aspires to the highest methodological standards --- one source of error that journalists and other consumers of poll data have become accustomed to seeing when research results are displayed is a "margin of sampling error." A "margin of sampling error" tells the consumer of a poll that if every other aspect of that particular survey is perfect and zero incremental error exists, then the same survey, repeated multiple times, would produce measurements that did not vary from the estimates in the particular survey by more than a certain number of percentage points in either direction.

By way of illustration:

In a world of door-to-door interviewing, if a researcher knows that in a particular geography there are exactly 1 million dwellings and each dwelling has 1 and only 1 person in residence, and if the researcher assumes (however unlikely) that no one is homeless and every person in the particular geography lives in one of the 1 million dwellings, then a research project that randomly chooses to knock on every 1,000th door, assuming everyone is home and everyone opens the door when they hear the knock, will produce 1,000 completed interviews from a randomly chosen, probability sample. If a survey today of those 1,000 respondents finds that 20% are bald (or left-handed, or Roman Catholic), then an identically executed survey (of a different 1,000 dwellings) re-conducted in 1 week, 1 month or 1 year, should produce estimates that do not vary one way or the other by more than the researcher's stated margin of sampling error. If, for this particular survey, the researcher calculates the stated margin of sampling error to be 3 percentage points, then, in theory, the next 100 times any researcher completes the identical survey (each time selecting at random a different 1,000 dwellings for inclusion) the number of bald individuals should fall between 17% and 23% --- in 95 of those next 100 surveys.

In today's world, individuals do not open their front door to a stranger, do not answer their ringing home telephone (if they even have a home telephone), and perceive much of their US mail to be junk.

Opinion research companies have fewer and fewer ways to conduct scientific research in theory and almost no way to draw a probability sample in practice.

The good news, as is often the case, is: just as one door closes, another opens. Over the past 25 years, more and more Americans have found ways to access the Internet. The single-purpose cell-phones of the 1990s have given way to the multi-purpose smartphones of the 2000s. Many Americans are walking around with a 2-way transponder in their purse or pocket.

For just a moment, imagine a world in which every American is issued exactly one (and only one) smart-phone, and each individual is assigned a unique smart-phone number that he/she owns his/her entire life; and that phone number never changes; and that smart-phone is the only device from which it is possible for that individual to access the Internet; and further imagine that there is a centrally (and benevolently) maintained database of all known smart-phone numbers; and that individuals always attend promptly to their smart-phones the moment they vibrate, then: it would be possible in such a world for research companies to draw a random, probability sample from the central database of all known smart-phone numbers, and, it would be possible for the researcher to calculate a margin of sampling error for an internet survey.

Alas: that is not today's world. In today's world, there is no central database of smart-phone numbers; some individuals have several smart-phones and others have none. Some individuals have multiple desktop computers, laptop computers and tablets that allow them to access the internet from many different locations using many different devices, and others have no such devices. Some individuals have a half-dozen email addresses and others have none.

So?

So that means that although research companies have many different ways of inviting individuals to complete opinion surveys electronically, using the Internet, samples drawn from internet populations --- however randomly the respondents may be chosen --- are non-probability samples.

For the foreseeable future, until an entirely new communications paradigm is invented, SurveyUSA and other opinion research companies see non-probability samples, for many studies, as their best option --- not because non-probability samples are ideal, preferred or perfect, but because they are now superior to older, dated methodologies which cost more and yield much less than they used to.

In a world of non-probability samples, SurveyUSA, other research companies, statisticians and research trade associations need to express to journalists and other consumers of poll data, how much of a bracket to place around a given survey finding, where such bracket isolates the error that might be attributable to sampling alone.

Adding additional complexity, most internet research today is conducted of pre-recruited individuals, so that if SurveyUSA, right this very minute, needs to complete 1,000 interviews in geography "x," SurveyUSA does not have to start from scratch, but rather has access to a so-called "panel" of willing cooperators who have previously provided their U.S. mailing address, which permits SurveyUSA to know in advance whether a potential respondent does or does not reside in geography "x."

A Credibility Interval is SurveyUSA's best estimate of an interval around a measured percentage within which the true percentage, if all eligible respondents were to be interviewed, would have a 95% chance of falling. A Credibility Interval uses Bayesian analysis to update a prior probability distribution (which, in the absence of other information, falls uniformly between 0% and 100%) based on the survey results. The posterior distribution is represented by an interval bounded by the 2.5% and 97.5% percentile points.

For sufficiently large sample sizes and in the absence of prior data, the Credibility Interval will be similar to the 2-standard-deviation confidence interval that would be obtained from a probability sample, after estimating an effective sample size based on the respondent weights using a formula developed by Leslie Kish (square of sum of weights over sum of squares of weights). When the weighted percentage of an answer choice is p and the effective sample size is N, this confidence interval is given by p +/- 2sqrt(p(1-p)/N). The maximum width occurs where p=1/2 and the interval is given by p +/- 1/sqrt(N) (p +/- 10% for n=100, p +/- 5% for n=400, p +/- 4% for n=625, etc.).

Bayesian Credibility Intervals represent uncertainty as a subjective probability estimate, which should not be interpreted as a frequency. For a 95% Credibility level, there is no collection of alternative research outcomes which would fall within the interval 19 times out of 20 following a bell-curve distribution.

Rather a Credibility Interval represents a "***degree of belief***" in a proposition.

Theoretically, if one gives a proposition 95% Bayesian Credibility, one ought to be willing to bet against that belief when the odds exceed 19 to 1. Bayes's Theorem indicates the correct way of using evidence to transform probability estimates. Previous knowledge is represented by the "prior probability distribution" (which, in the absence of previous evidence, is chosen as the distribution that maximizes mathematical entropy, which is the uniform distribution on (0,1) for variables with the units of probability such as answer choice percentages).

The Credibility Interval works according to the following rule:

$$P(H|E) = P(H) \times \frac{P(E|H)}{P(E)}$$

In this formula, P(H) is the prior probability of a hypothesis H, P(H|E) is the posterior probability of H given the evidence E, P(E) is the probability of E, without knowing anything about H, and P(E|H) is the probability of E given the hypothesis H, also known as the "likelihood" of the evidence. The ratio is greater than 1 if the hypothesis being true makes the evidence more likely than the hypothesis being false, and in that case the subjective probability of H is increased when E is seen.

For example: if half of adults are men, 10% of men are over 6 feet tall, and 2% of women are over 6 feet tall, then knowing that an adult is over 6 feet tall changes the subjective estimate of the person's gender being male from 50% to 83.3% (the probability that an adult of unknown gender is over 6 feet tall is 6%, so the fraction has numerator 10% and denominator 6%, which when multiplied by 50% gives 83.3%).

By way of illustration: If SurveyUSA calculates a Credibility Interval of 3.4 percentage points for a given measurement of 60%, SurveyUSA might graphically display the interval between 56.6% and 63.4% as the range of credible outcomes. But in practice, SurveyUSA will adopt the industry shorthand and record, in this illustration, the Credibility Interval with a text notation next to the measurement that says "Plus or minus 3.4 percentage points."

When SurveyUSA displays alongside the results of a particular survey a Credibility Interval that accounts for sampling error, this does not mean that sampling error is the only source of error in that survey, nor necessarily even the largest source of error. Other sources of error can be introduced, in practice, in the way questions are worded and ordered, the way in which text and images are displayed on the respondent's electronic device, which days and how many total days interviews are completed, the literacy of the respondent, the difficulty of translating the questionnaire into all possible languages and dialects, and the fact that those who volunteer to be included in online research may differ in some unknowable and therefore uncorrectable way from the population as a whole. Non-sampling errors cannot be routinely quantified, and often cannot be quantified at all.

These statements conform to the principals of disclosure as recommended by the National Council on Public Polls (NCPP). Questions about SurveyUSA research can be addressed to: editor@surveyusa.com.