Margins of Error vs Bayesian Credibility Intervals and their Use in Online Survey Research

James Wyatt

April,9 2017

Overview

The shift from phone-based interviews to self-administered (opt-in panel) online surveys has presented the polling industry with a wealth of new opportunities and challenges.[1]

One of those key challenges is determining how to report the level of confidence in online survey results. In the traditional sense, using probability sampling methods, this was done using margins of error.[2]

To calculate a margin of error for a particular survey one must know the exact probability of participation, or "inclusion probability," for each individual in the sampling frame. Furthermore, to use margins of error, It's also necessary to assume that non-respondents are random, which we know is not the case in online surveys.

Since this inclusion probability is unknown for online survey respondents from most survey panels, traditonal "margins of error," cannot be used to accurately relay measurement error.

In traditional random sampling approaches to survey research, each individual in the survey population has a non-zero chance of being selected to take the survey. This fundamental property of sampling theory allows researchers to make generalizable claims of larger populations just by interviewing a few respondents. For example, many national surveys that are tracking the 2016 presidential election have sample sizes that vary between 800 - 1,500 respondents.[3]

Margins of Error (Somewhat) Explained

Margin's of error are based on the frequentist approach to statistics that assumes, with large enough repeated samples, that the true estimate will fall within a given range (given a particular confidence interval, i.e 95%). These margins of error can effectively be used to describe how reliable survey results are.

For example, let's say we have the results of a hypothetical national survey of 1,000 randomly selected registered voters in the United States with a sampling error of plus or minus 3 percentage points. The results of the Clinton/Trump head-to-head match up is 50% for Clinton and 47% for Trump with 3% undecided.

To most political pundits, Clinton is *beating* Trump in the head-to-head match up, but this may not actually be the case. In fact, Clinton's true level support can be anywhere within a range of +3 or -3 of 50%. This is to say that if we repeated this survey with the same methods and sample size 100 times then 95% of the time Clinton's average level of support will fall between 47-53%… and 5% of the time it will fall outside of this range). This is the same for Trump's level of support, which could be as high as 50% or as low as 44%. Given this, the true state of the race is a statistical tie - we really don't have a strong case, based on this example poll, to reject the null hypothesis that Clinton and Trump are *NOT* equal in terms of support.

[1] Pew research on survey responses and the mode of interview: http://www.pewresearch.org/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/

[2] Margins of Error: https://en.wikipedia.org/wiki/Margin_of_error

[3] Real Clear Politics (RCP): http://www.realclearpolitics.com/epolls/2016/president/us/general_election_trump_vs_clinton-5491.html

Quickly Calculate the MoE - Simple Random Sample

For a random sample of adults in the United States, the margin of error, MoE (or maximum margin of error), is a function of the sample size of the survey n. The margin of error can be calculated for the varying degrees of confidence using the following quick formulas to derive the maximum margin of error:

At 99% confidence, the $MoE = \frac{1.29}{\sqrt{n}}$

At 95% confidence, the $MoE = \frac{.98}{\sqrt{n}}$

At 90% confidence, the $MoE = \frac{.82}{\sqrt{n}}$

MoE Example 1:

- Sample size: random sample of 1,000 registered voters
- Confidence level: 95% confident

$$MoE = \frac{.98}{\sqrt{n}} = \frac{.98}{\sqrt{1000}} = \pm 3.1\%$$

In contrast, the general formula to calculate the MoE is:

$$MoE = z * \sqrt{\frac{p(1-p)}{n}}$$

where p is the sample proportion, n is the sample size, and z is the level of confidence desired in the results.

MoE Example 2

- Sample size: random sample of 1,000 registered voters
- Confidence level: 95% confident (z score of 1.96)[4]
- 53% Support Hillary Clinton for President and 47% do not

$$MoE = 1.96 * \sqrt{\frac{(.53)(.47)}{1,000}} = \pm 3.1\%$$

Notice how in both cases the margin of error is the same (after rounding, of course). It should be noted that the differences between both methods *decrease* as the sample sizes *increase*. Try it for your self using smaller n samples, so for a sample of 50 or 100.

So why can't we do the same MoE calculations for Online Surveys?

Well, it comes down again to the conditions that surround sampling theory. If we're trying to make broader claims about registered voters in the United States on, say, a political head-to-head match up, we would need to assume that every registered voter (theoretically) has the opportunity to participate in the survey and those *not* participating in the survey (non-respondents) are completely random. The fact is, online surveys do not meet these two essential fundamental aspects of frequentist sampling methods that allow us to use margins of error.

[4] Check out a table of Z-scores for the corresponding levels of confidence: https://www.ltcconline.net/greenl/courses/201/estimation/smallConfLevelTable.htm

For online, opt-in panel surveys we know:

- There's a digital divide in the United States that skews towards older and ethnic/racial minorities not having access to the Internet - this makes it harder to define the inclusion probability, which in many times is zero for vast portions of the registered voter population
- Most online samples are opt-in, they are non-probability based, and rely mainly on filling quotas of known census quantities
- There could potentially be fundamental differences between individuals that take online surveys and individuals that do not. As a result, online samples may not accurately reflect the broader population it intends to measure [5]

So we can't use MoE, then what *can* we use?

Bayesian Credibility Intervals do the trick.

Anyone doing online surveys can calculate Bayesian Credibility Interval to assess confidence around survey questions using a beta-binomial conjugate model: $x|\pi \sim bin(n, \pi)$ where $\pi \sim beta(\alpha, \beta)$. The intervals of the proportion are calculated from the largest possible credibility interval for the observed samples using the naive priors: $\alpha = 1$ and $\beta = 1$, where $\sum_{i=1}^{n} y_i$ is the observed the number "Yes's," $n - \sum_{i=1}^{n} y_i$ is the observed number of "No's," and n is the sample size. The credibility interval for any particular question in the survey can be calculated from the posterior beta distribution with the following parameters:

$$\alpha + \sum_{i=1}^{n} y_i, \beta + n - \sum_{i=1}^{n} y_i$$

An Applied Example

An online survey of 2,000 registered voters were asked:

Q1 Have you ever voted either in person or by mail in your precinct or election district?

- Yes - 60%
- No - 30%
- Don't Know - 10%

The following is the R-code used to calculate the Bayesian Credibility Interval (BCI) for the above survey question:

```
library(emdbook)
priora <- 1                 # prior hyper parameter - alpha
priorb <- 1                 # prior hyper parameter - beta
p <- 0.50                   # probability of yes/no (Not including a dk/no option)
n <- 2000                   # sample size
total_support <- n * p      # number of yes's
total_oppose <-  n * (1 - p)  # number of no's

#calcuate 95% credibility interval based on posterior distribution
x <- tcredint("beta",list(shape1=priora+total_support,
              shape2=priorb+total_oppose),verbose=TRUE)

cred_interval <- (x[2] - x[1]) / 2
cred_interval
```

[5] Evaluating Non-Probability Survey Panels: PEW http://www.pewresearch.org/2016/05/02/evaluating-online-nonprobability-surveys/

The credibility interval for the example question above is: $\pm 2.2\%$

In the Bayesian sense, we collected some data (ran a survey), where we assume the data is binomially distributed (for each question). In truth, we have some sense of the probability of a success which we characterize through a probability distribution (beta-dist). This is what is known as *a prior*. Based on the range of successes we obtained in repeated survey samples, we assumed the largest possible credibility interval. After we collect the data, we then calculate the probability of getting different values based on the data collected, which we call the "posteriori probability distribution".

The uncertainty in the results, or range of values on the posteriori, includes 95% of the probability that the results will fall within this range, ergo the "95% credibility interval." That is, the probability that *YES* (answer option in the example above) lies between a range of $\pm 2.2\%$ from the total proportion of those that chose option *YES* is 95%.

Another way to say this is: there is 95% chance that those empirical means lie in the credible interval, defined using quantiles of the posterior distribution.

The credibility t interval can be expressed as such:

$$57.8\% \leq t \leq 62.2\%$$

Credible intervals are comparable in certain aspects to confidence intervals in frequentist statistics.

It's important to note that the t-credibility interval can be approximated:

$$\frac{1}{\sqrt{n}}$$

Conclusion

Bayesian Credibility Intervals provide a valid methodological approach to reporting survey findings that is intuitive and useful for users. For researchers conducting on line research, especially using opt-in panels, BCI's provide a way to calculate how much confidence one can have in the survey results.