IDA+A
Institutional Data Analytics + Assessment

Samples and Response Rates

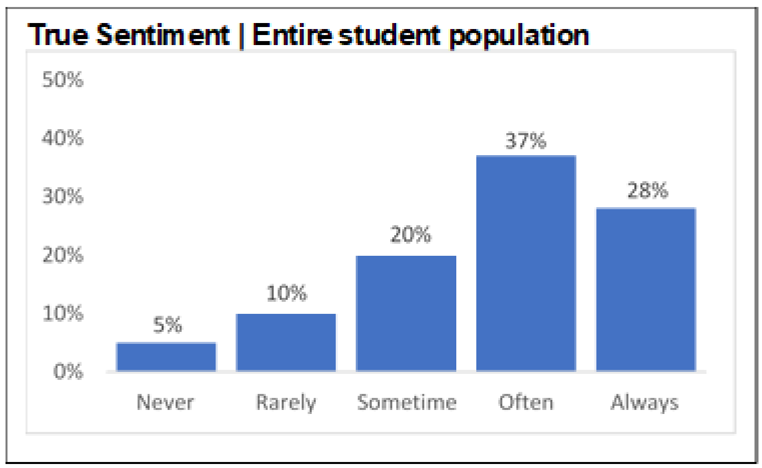
This brief document presents a case for preferring smaller, more targeted survey samples rather than large random samples. The best survey design depends on many factors. Consult with IDA+A (idata@purdue.edu) to develop a plan for your next survey.

To be reliable and useful for research and administrative decision support, a survey must have a "sufficient" number of responses. Most survey researchers strive aggressively for as high a response rate as possible. But response rates are not evidence of a survey's reliability and accuracy. The type of analysis and/or statements you want to be able to make about the population can help you determine how many responses are needed. And the number required is often moderate. Here we present the relationship between response rate and the quality of a sample. It shows that, beyond a moderate number of responses, increasingly larger sample provide little additional benefit.

A random sample does not ensure a representative sample

Researchers often believe that if their survey response is large enough, it will be representative of the whole population. But random, representative sampling methods only ensure that a representative pool of people are invited. When participation is voluntary, as most surveys at Purdue are, the sample is contingent on individual choices to participate. Participation bias is not resolved by incrementally larger numbers. If a survey is promoted in a biased way – even unintentionally – this can make the resulting dataset even more biased.

Suppose a survey is launched with a messaging strategy that produces a 15% response rate campus wide but oversamples one college. This could happen if promotional posters are placed more prominently in that college's building, offers a persuasive incentive, or if that college employs more effective messaging. Survey administrators might decide that the 15% response rate is not adequate and undertake a second wave of messaging. But the same biases will remain in the larger sample. The larger sample will be no more representative or accurate than the smaller one.



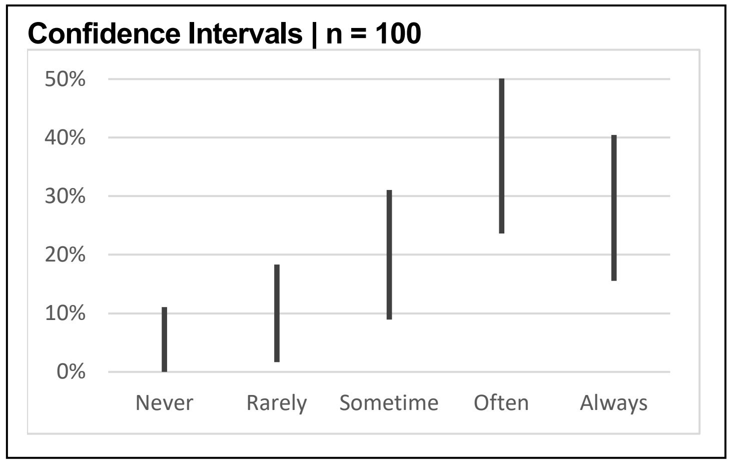
Suppose there is a survey question, to which the true sentiment of the entire student population is as shown at right. The response is roughly "normal" in distribution but very much skewed toward higher values.

If a survey asks a sample of students a question about this hypothetical issue, the sample must be sufficiently large to measure the actual distribution. For a question with only five choices, samples of no more than 30 to 50 would begin to approach the true values. A sample of only n=50 would probably indicate that frequent answers (Often/Always) are chosen more than the infrequent answers (Never/Rarely). But samples that small would retain a great deal of uncertainty and imprecision. Researchers could infer something, but not very much.

INSTITUTIONAL DATA ANALYTICS + ASSESSMENT (IDA+A)
Young Hall
155 S. Grant Streel West Lafayette, IN 47907
idata@purdue.edu

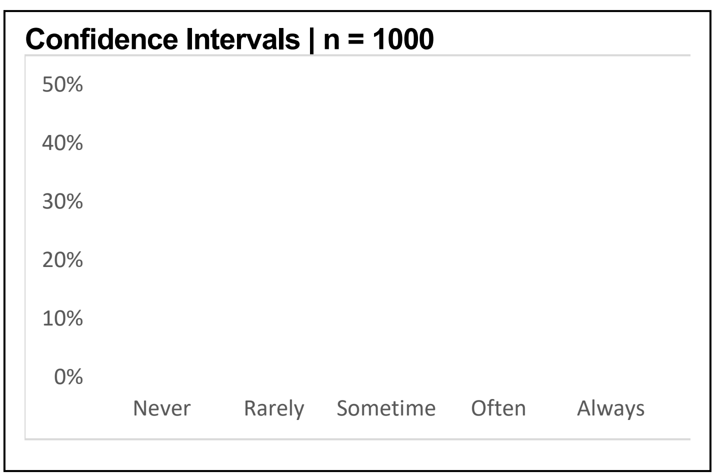
Confidence intervals indicate preciseness

In the chart above the 5 | 10 | 20 | 37 | 28 values are exact values, based on the (hypothetical) student population. A survey sample provides an estimate of that population based on the sample. That estimate is, by nature, less precise than the exact measure. But survey results are often displayed as if they were exact. In a statement released in 2016, the American Statistical Association issued a [Statement on Statistical Significance and p-Values](#) recommending use of confidence intervals in reporting data outcomes as a means of correcting the error. Instead of "Twenty-eight percent of students always encounter the condition or experience," a more apt phrasing would be, "The survey results give us 95% confidence that the share of all students who answer "Always" is 28%(+/-5%)."



The size of that confidence interval is determined by the sample size. The larger the sample, the smaller the range or interval. In the chart at right, the true values (5 | 10 | 20 | 37 | 28) are bracketed within the vertical bars that indicate the confidence interval ranges. We know the true values in this hypothetical example, but in most cases we would not.

The vertical lines depict the range of the confidence intervals for the entire sample (n=100). Those intervals are substantial. The value for 'Never' (which we know to be 5) is stipulated only to be something between zero and 11. The value for "Rarely" is between two and 18. From this, we cannot say if "Never" or "Rarely" is more common. The share saying "Always" lies between 15.5% and 40.5%. If these data were measuring a program that aimed for a 30% "Always" outcome, the sample of 100 could not determine if the program was successful or not.



Even though the data attained in the sample of 100 is accurate and methodologically rigorous, the results may not be precise enough to be useful. A larger sample is needed.

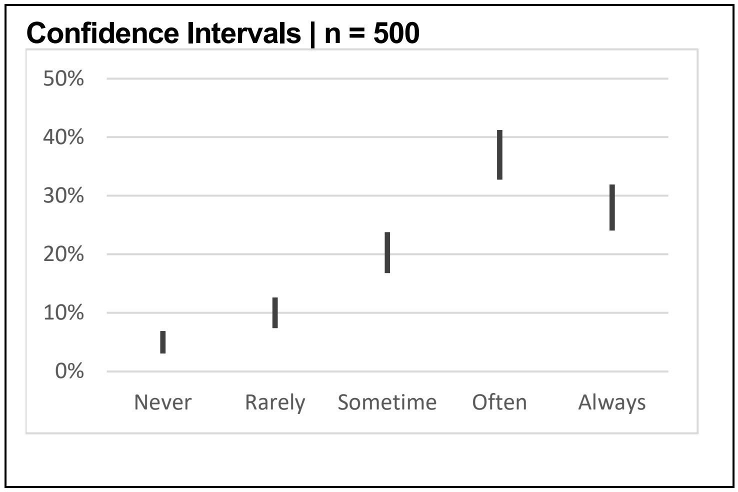
Next consider the same scenario, but with a larger sample (n=1000). The true values are again contained within the ranges. But unlike the sample of 100, the confidence interval ranges are very small. There is no overlap in the ranges, so the distribution is very clear and evident. "Never" is the least common answer. "Often" is the most common answer.

The true answer for "Always" (28%) is tightly bracketed between the low of 25% and the high of 31%.

INSTITUTIONAL DATA ANALYTICS + ASSESSMENT (IDA+A)
Young Hall
155 S. Grant Streel West Lafayette, IN 47907
idata@purdue.edu

The sample of 1000 appears to be both accurate and precise enough to allow research conclusions or administrative decisions. But we cannot yet say whether 1000 is the optimal sample size.

When the sample is cut in half (n=500), the confidence intervals expand as compared to the sample of 1000. But as with the sample of 1000, the ranges do not overlap and the results are clear. We don't know exactly what share of respondents said "Often," but we know it was the largest share and that "Always" is next most common.

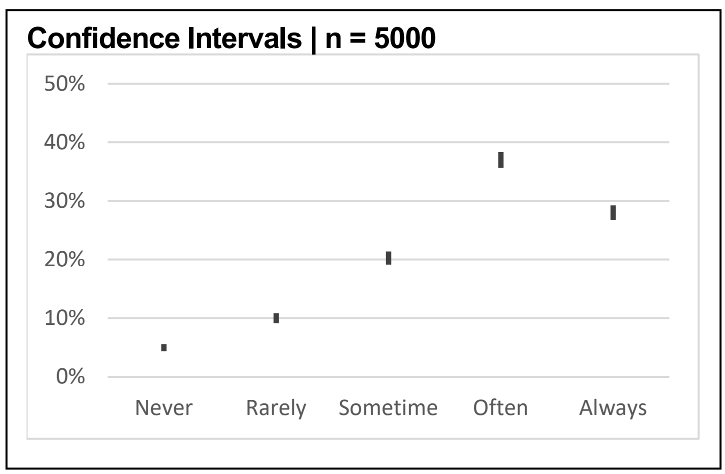


For practical purposes, the sample of 500 might be as good as the sample of 1000. It is not as precise as the larger sample, but for administrative purposes, it supports the same conclusions and decisions as the sample of 1000.

Very large samples can be wasteful

The common practice among researchers is to strive aggressively for the largest possible sample. For a campus the size of Purdue University, a 30% response from 35,000 undergraduates would be a sample of more than 10,500. It is hard to justify, given practical considerations, a need for such a large sample. A sample of 15% or n=5000 achieves great exactness. The vertical bars in the next chart indicates the range of the confidence intervals are almost non-existent.

The "Always" share is clearly something less than 30%. If the "Do-or-Die" threshold for success of the program were 30% "Always," the program could be judged a failure. On the other hand, it might be deemed to be so near to the success threshold that it could be continued with modifications that would improve future performance.



The decision makers could make nearly the same judgement from the sample of 1000 or even the sample of 500. The additional effort necessary to attain the larger sample may not be justified by the additional precision obtained.

Note: To emphasize the effects of various sample sizes on the confidence intervals, which presentation has removed sampling error from consideration. Small samples can incorrectly estimate the value in the population. In practice, samples large enough to sufficiently narrow confidence intervals are also sufficient to eliminate estimation errors.