researchdesk

Number of completes

How many completes are required: a beginner guide to estimating the number of completes needed for a statistically significant survey

TL;DR;

Although some surveys are based on "hey, let's get 1000 completes", there are sound statistical rules to calculate the number of completes that a study needs. These are called the "Confidence Interval" and the "Margin of Error". These need to be valid for all the segments you plan to make from your survey data.

The number of completes you need is derived from a formula:

$$N = \frac{0.25}{s} \times \left(\frac{z}{MOE}\right)^2$$

where N is the number of completes, s is the smallest share of the panelists that you want to segment, z is a z-score which is found from a lookup table of the required confidence interval, and the MOE is the margin of error that you are OK with.

Introduction

Usually, the objective of creating a study is to provide an answer to a series of questions. In order to be sure that your survey is delivering data that you can be confident in, you may choose a confidence interval (CI) and acceptable margin of error (MOE).

Confidence interval and margin of error make sense when you fill in the following sentence:

"I am {confidence interval} sure that the numbers from my survey are within +/- {margin of error}"

A third factor is level at which you want to segment your completes into sub-segments. These are typically shown as "base" when showing a chart.

> **Confidence Interval** is typically 95% for a high-quality survey, though 90% confidence intervals are also deemed to be high quality. Lower confidence intervals are used for guidance surveys – surveys that are not specifically meant to deliver actionable results, but provide guidance about how to perform subsequent research. Confidence Intervals above 95% are used where high certainty is required, usually because costly decisions will be based on it.
>
> **Margin of Error** is simpler to understand, as it's the error margin in any numbers. A typical survey uses 3% or 5%, but where accuracy needs to be better, 1% or 2% are used, and where a survey is being used as the basis for further research, or just to give indications, MOEs of more than 5% are sometimes used.
>
> **Smallest Segment Share of Completes** represents the share of completes that is the smallest segment you plan to test. For example, if you are segmenting people into buckets, and you expect these buckets to be distributed as 20%, 25%, 30% and 25%

Number of completes

of the completes, then you need to consider 20% to be the smallest segment share of completes.

Calculating using a Formula

You can use a formula, where **N** is the required number of completes, **z** is the z-score, **MOE** is the margin of error, and **s** is the share of the smallest segment. The z-score is derived from the confidence interval using the table shown:

CI	z-score
80%	1.28
85%	1.44
90%	1.645
95%	1.96
97%	2.17
99%	2.576

$$N = \frac{0.25}{s} \times \left(\frac{z}{MOE}\right)^2$$

or

$$MOE = \frac{z}{2\sqrt{N \times s}}$$

Using Lookup charts

The set of charts allow you to look-up the number of completes required to meet confidence interval, margin of error, and segmentability of your data.

How to use these charts:

Example 1:

Let's assume that you decide that a 90% confidence interval, at 3% margin of error, is acceptable, and the smallest segment you will look at is by gender and age. You have a quota of gender at 50%:50%, and four age groups, each set to 25% of the panel. The smallest segment of age-group X gender is therefore 12.5%.

Look at the 90% Confidence Interval chart. The gray line represents 3%. Looking at 12.5% on the horizontal axis, you can see that 6,000 completes is required.

CI = 90%
z=1.645 (from table)
MOE=3%
s=12.5%

N ~ 6000 (from chart)

Alternatively using the formula:

$$N = \frac{0.25}{s} \times \left(\frac{z}{MOE}\right)^2 = \frac{0.25}{0.125} \times \left(\frac{1.645}{0.03}\right)^2 = 6013$$

Example 2:

You have already conducted a survey, and have 1000 completes. You want to segment males from females, which are 50%, and your business requirements are that the study is at confidence interval of 95%. You want to know the margin of error.

Looking at the 95% confidence interval chart, you can see that 50% segment and 1,000 completes is between the yellow (MOE: 4%) and light blue (MOE: 5%) lines. Your margin of error is around 4.5%.

CI = 95%
z=1.96 (from table)
N=1000
s=50%

MOE ~ 4.5% (from chart)

Number of completes

Alternatively using the formula:

$$MOE = \frac{z}{2\sqrt{N \times s}} = \frac{1.96}{2\sqrt{1000 \times 0.5}} = 0.0438 \cong 4.4\%$$

Example 3:

You have a result from a survey of 2,300 people of which 1000 are male, and 1300 are female. There is a question which says "Do you like video games?", and 71% of the males say yes, and 65% of the females say yes. You want to ask yourself how confidently you can say that males like video games more than females.

In this case, you are asking what is the confidence interval when the margin of error is 3%, because when the margin of error is 3%, then between 68% and 74% of males would definitely answer "yes" even if everyone was asked, and 62% to 68% of females would definitely answer "yes", and so you can definitely say that males like video games more than females.

We need to perform this on the smallest group, because we are certainly more confident about the females than the males, because there were more of them.

N = 2300
s = 1000 / 2300 = 43.5%
MOE = 3%

$$z = 2\sqrt{N \times s} \times MOE = 2\sqrt{2300 \times 0.435} \times 0.03 \approx 1.89$$

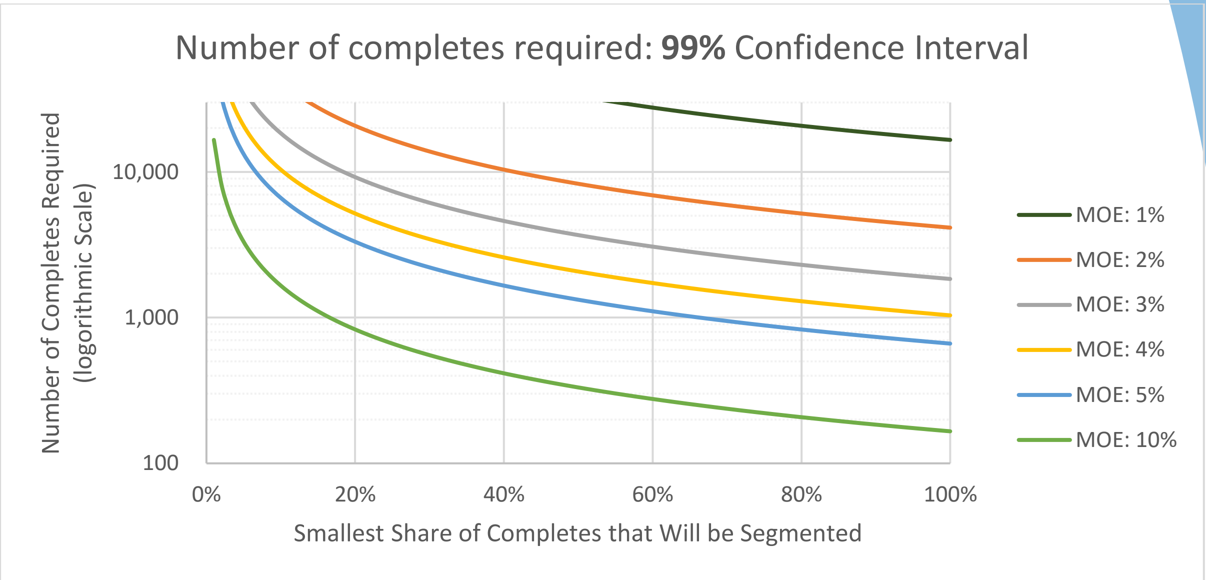
z=1.89 (from formula)

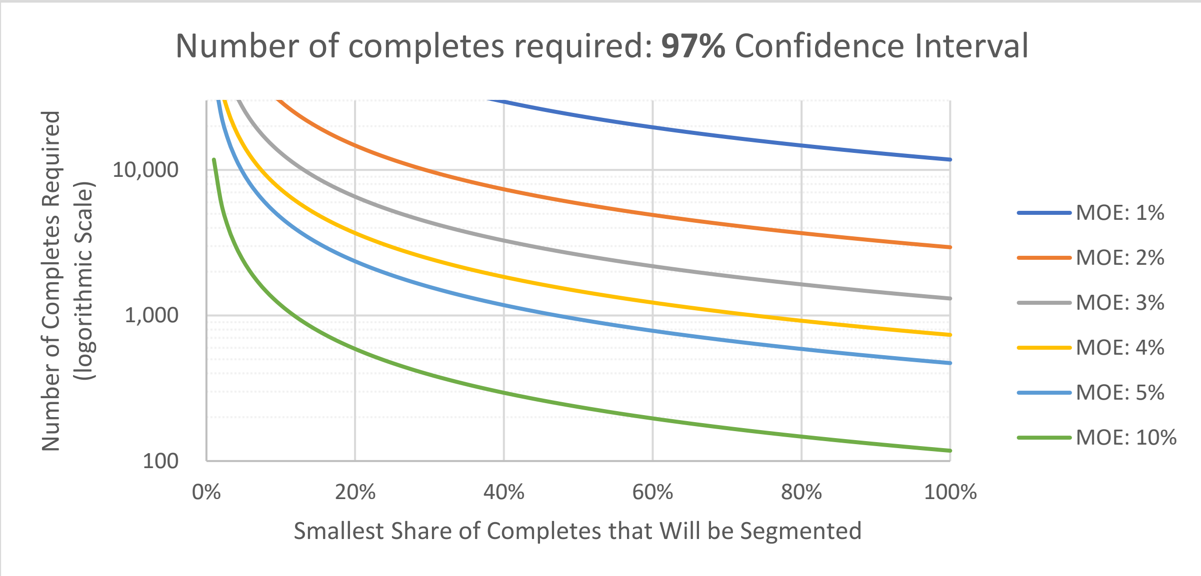
CI ~ 94% (from table)

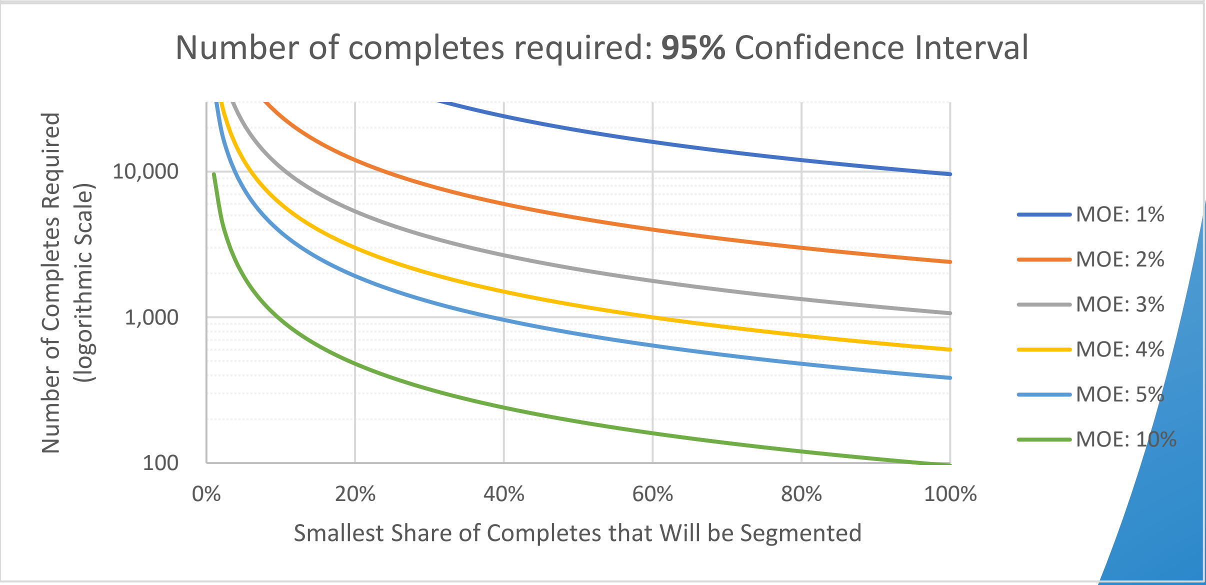
With this data, you can be 94% sure that males like video games more than females.

researchdesk

Number of completes







Number of completes