Determining sample size

Introduction

When you are doing research, for example, amongst all the participants in a support programme, or amongst association members, or perhaps amongst the population of small businesses, the first question most people ask is about the sample size required. There are a number of factors that influence the decision and, if you a quick review on the web, a myriad of formulae for calculating it.

The factors

Level of precision

The 'level of precision', sometimes called the 'confidence interval' or 'sampling error', is the range in which the population's actual value is estimated to lie. If you find that 60 per cent of the SMEs in a sample have adopted a specific practice with a precision rate of ±5 per cent, then you can conclude that the actual number of SMEs in the total population that have adopted the practice lies in the range 55-65 per cent.

Confidence level

The 'confidence level' indicates how likely it is that the actual answer falls within the range confidence interval. So, for a confidence level of 95 per cent, 95 samples out of 100 will have the true population value within the confidence interval. There is always some risk that your sample does not represent the true population value. Occasionally, it might be necessary to use a 99 per cent confidence level, but most researchers use 95 per cent.

If you are interested, 95 per cent is the proportion of the population covered by (near enough) ±2 standard deviations from the mean in a normal distribution.

The wider you can accept for the confidence interval, then the more confident you can be that the real answer lies within the range.

Degree of variability

The 'degree of variability' describes the distribution of attributes in the population. The more heterogeneous a population, the larger the sample size required to obtain a given level of precision. The more homogeneous a population, the smaller the sample size required. You should note that a 50/50 split on a specific attribute or response indicates maximum variability in the population, whereas a 90/10 split means that 90 per cent of the population share an attribute, so the sample is less variable. If you don't know what level of variability to expect, then assume that it is 50 per cent. This may mean that you use a larger sample size than was really needed, but that is better than using a sample size that is too small, and then having no confidence in the results.

Budget

Inevitably, cost will also play a part in deciding the sample size, or more precisely, the in determining the confidence you can place in your results.

Determining sample size

Options for determining sample size

Assuming that you do not intend just to pick a figure out of thin air – which appears to be the technique favour by at least some researchers – there are essentially three options open to you:

Census

One option is to undertake a census, that is, to survey every member of the population. For small populations this may well be the only way to guarantee a degree of accuracy. It eliminates all sampling error and provides data on the whole population. However, cost considerations make this impractical once populations exceed a few hundred.

Using formulae

If you want different confidence intervals, or have different degrees of variability, then you may find it easiest to use a formula to calculate the sample size.

For populations that are large, and particularly for population where the total population is unknown, this formula will tell you the sample size required.

$$n_0 = \frac{Z^2 p(1-p)}{e^2}$$

Z depends on the degree of confidence that you want. (Actually, it relates to the area under a normal distribution curve, but you don't need to worry about that!) For a confidence level of 95 per cent, Z=1.96; for 90 per cent, Z= 1.64; and for 99 per cent, Z=2.58. p is the degree of variability, expressed as a decimal; if you don't know this, then use 0.5. e is the level of precision, expressed as a decimal.

Imagine that you need to survey the total population of SMEs to discover how many have a loan from a bank. You are happy with a confidence level of 95 per cent, a precision rate of ±5 per cent and a degree of variability of 50 per cent.

$$n_0 = \frac{Z^2 p(1-p)}{e^2} = \frac{1.96^2 x 0.5 x 0.5}{0.05^2} = 385$$

The answer is 384.16, but you obviously cannot interview a fraction of a person so you need to round upwards.

If you know that the total population is small, then your sample can be smaller. You can adjust it by using this equation, where n is the new sample size and N is the size of the population. Suppose that you are running an association of SMEs and have 1,000 members and want to answer the same question.

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} = \frac{385}{1 + \frac{(385-1)}{1000}} = 279$$

You need to do the earlier calculation to discover the sample size for a large population and then you can apply the 'finite population correction'.

Helpfully, however, there is a quicker way of calculating sample sizes for known populations. This is shown below. Again, imagine that you want to survey your population of 1,000 members and are still looking for 95 per confidence, 5 per cent precision.

$$n = \frac{N}{1 + Ne^2} = \frac{1000}{1 + 1000(0.05)^2} = 286$$

This gives a slightly different answer to that using the finite population correction, but it is close enough.

Using published tables

Possibly the quickest option is to find a table of sample sizes, say on the internet. These will typically give sample sizes for different populations and with different levels of precision, confidence levels and degrees of variability.

If you are happy with a 95 per cent confidence level, 5 per cent precision and 50 per cent degree of variability, then you can choose the sample size from the table shown below.

Population	250	500	750	1,000	2,500	5,000	7,500	10,000	25,000	100,000
Sample	154	223	261	286	345	371	380	385	394	399

Whether you use tables or a formula to calculate the sample size, it is important to remember that the sample refers to the number of respondents, and not to the number of people invited to participate in a survey. You may find, for example, that some of the people for whom you have contact details cannot, in fact, be reached; and others may simply not respond. So you will want to plan your work in such a way that you end up with enough responses.

These sample sizes also assume a truly random sample is used. If you need to use a stratified sampling system, perhaps to reflect differences in gender or age or geographic distribution, then you might need a larger sample size.