Formulae for determining sample size

Contents

1 Introduction......................................................................................................................................................1

2 Estimating the sample size based on a proportion.............................................................................................1

3 Estimating the sample size based on a mean......................................................................................................2

1 Introduction

This supplement expands the discussion of sample size in Chapter 9 by providing additional details on using formulae to determine sample size. It should be read in conjunction with that chapter.

A number of formulae are available for working out sample size. We will introduce two of them here to highlight some of the issues involved in calculating sample size for a quantitative project.

2 Estimating the sample size based on a proportion

To calculate the sample size based on the sample required to estimate a proportion with an approximate 95% confidence level\(^1\), you can use the following formula:

\[ n_r = \frac{4pq}{d^2} \]

Where \( n_r \) = required sample size, \( p \) = proportion of the population having the characteristic, \( q \) = 1-\( p \) and \( d \) = the degree of precision. The proportion of the population (\( p \)) may be known from prior research or other sources; if it is unknown use \( p = 0.5 \) which assumes maximum heterogeneity (i.e. a 50/50 split). The degree of precision (\( d \)) is the margin of error that is acceptable. Setting \( d = 0.02 \), for example, would give a margin of error of plus or minus 2%. We apply this formula in the example in the worked example below.

\(^1\) More accurately the formula is: \( n = \frac{(1.96)^2pq}{d^2} \)
2.1 Worked example

You are investigating the use of mobile phones for online banking and want to estimate what proportion of the population uses their phones in this way at an approximate 95% confidence level. Since no data are available on the proportion currently using their mobile phones you take the worst case scenario and set \( p = 0.5 \) (and therefore \( q = 1-0.5 = 0.5 \)). As this is a preliminary study you are prepared to accept a margin of error of ±5% so you set \( d = 0.05 \). To determine the minimum sample size you then apply the formula:

\[
\begin{align*}
  n &= \frac{4pq}{d^2} = \frac{4 \times 0.5 \times 0.5}{0.05^2} = \frac{1}{0.0025} = 400
\end{align*}
\]

So your minimum sample size would be 400.

Note that accepting a higher margin of error (making \( d \) bigger) will reduce the sample size. For a 10% margin of error in this scenario (i.e. \( d = 0.1 \)), the sample size would reduce to 100. Similarly if the population were more homogenous the sample size would also reduce. In this example, for instance, if we know that only 10% of the population used their phones for mobile banking then \( p = 0.1, q = 0.9 \) and the required sample size would drop to 144, assuming a margin of error of 5%.

3 Estimating the sample size based on a mean

The second formula applies when estimating the arithmetic mean (average) of a particular variable for a population. Suppose, for example, that you wanted to know the average employee satisfaction level in your organisation.

To calculate the sample size based on the sample required to estimate a population mean with an approximate 95% confidence level\(^2\), you can use the following formula:

\[
  n_r = \frac{4\sigma^2}{d^2}
\]

\(^2\) More accurately the formula is: \( n = \frac{(1.96)^2\sigma^2}{d^2} \)
Where \( n_r \) = required sample size, \( \sigma \) (the Greek letter sigma) = the population standard deviation, a measure of the variation in the population (see Chapter 13) and \( d \) = the degree of precision required by the researcher. A drawback with this formula is the need to know the population standard deviation. This may be known from prior research; if a good estimate is unavailable the formula will not be reliable. We apply this formula in the example in the box below.

### 3.1.1 Worked example

You are investigating the average (mean) level of employee satisfaction and want to know the required sample size. You decide on a 95% confidence level. Prior studies have reported a standard deviation (\( \sigma \)) of 1.5 so you decide to use the same figure in your estimate. Satisfaction will be measured on a 7-point scale and you set a margin of error of ±0.25 units. To determine the minimum sample size you then apply the formula:

\[
\frac{n_r}{d^2} = \frac{4 \times 1.5^2}{0.25^2} = \frac{9}{0.0625} = 144
\]

So your minimum sample size would be 144.

Note that as with the proportion formula, accepting a higher margin of error will decrease the sample size. (You can see this from the formula: as \( d \) gets bigger, \( n_r \) gets smaller.) The sample size will also decrease if the population is more homogenous (i.e. the standard deviation is smaller).

### 3.2 Finite population correction factor

When the sample represents a significant (e.g. over 5%) proportion of the population, a finite population correction factor can be applied. This will reduce the sample size required. The formula for this is:

\[
n_a = \frac{n_r}{1 + \frac{(n_r - 1)}{N}}
\]

Where \( n_a \) = the adjusted sample size, \( n_r \) = the original required sample size and \( N \) = population size. The formula is applied in the worked example below:
3.3 Worked example

Having calculated your sample size \((n_r = 144)\) for the employee satisfaction survey in the previous example, you decide to apply a finite population correction factor because the total number of employees is only 650 \((N = 650)\). To determine the adjusted sample size you apply the following formula.

\[
n_a = \frac{n_r}{1 + \frac{(n_r - 1)}{N}} = \frac{144}{1 + \frac{144 - 1}{650}} = \frac{144}{1.22} = 118
\]

So your adjusted minimum sample size would be 118.