

Documentation for Sample Size for a Proportion

Kevin M. Sullivan, PhD, MPH, MHA
cdckms@sph.emory.edu

This module calculates a sample size for a proportion. The data input screen is as follows:

Sample Size for % Frequency in a Population (Random Sample)		
Population size	1000000	If large, leave as one million
Anticipated % frequency(p)	50	Between 0 & 99.99. If unknown, use 50%
Confidence limits as +/- percent of 100	5	Absolute precision %
Design effect (for complex sample surveys=DEFF)	1.0	1.0 for random sample

The four values requested are:

- **Population size** – what is the size of the population to be surveyed? The default value is one million. In general, when sampling from a “large” population, whether there are 100,000 in the population or 100 million, this will not affect the sample size calculation. However, when the population size is small, then this could reduce the sample size.
- **Anticipated % frequency (p)** – provide an educated guess of the percent of the population with the outcome of interest. Of course, if you knew this with any accuracy, you would not need to perform the survey. If you are unsure of the percentage, use of 50% will result in the largest sample size (holding the other three pieces of information being requested the same).
- **Confidence limits as +/- percent of 100** – this question is asking exactly how wide (in absolute terms) you would like the confidence interval to be around your point estimate. Using the default values, with an anticipated frequency of 50% and confidence limits as $\pm 5\%$, for the calculated sample size the confidence interval would be $50\% \pm 5\%$, i.e., (45%, 55%). With an anticipated frequency of 85% with confidence limits of $\pm 10\%$, for the calculated sample size the confidence interval would be (75%, 95%).
- **Design Effect** – if simple random sampling is to be used to select individuals (or whatever the element of analysis), then the design effect (DEFF) should be left as one. If a cluster-type survey is being used in the sampling methodology, the DEFF is frequently larger than one, perhaps in the range of 2 to 10 depending on the outcome being studied. An estimate of the DEFF can usually be found in the literature.

The results of the calculation, using the default values, is shown next:

Sample Size for Frequency in a Population

Population size(for finite population correction factor or fpc)(<i>N</i>):	1000000
Hypothesized % frequency of outcome factor in the population (<i>p</i>):	50%/±5
Confidence limits as % of 100(absolute +/- %)(<i>d</i>):	5%
Design effect (for cluster surveys- <i>DEFF</i>):	1

Sample Size(*n*) for Various Confidence Levels

ConfidenceLevel(%)	Sample Size
95%	384
80%	165
90%	271
97%	471
99%	664
99.9%	1082
99.99%	1512

Equation

$$\text{Sample size } n = \frac{[DEFF + Np(1-p)]}{[(d^2/Z^2_{1-\alpha/2} + (N-1)p*(1-p)]}$$

At the top of the output are the values supplied by the user. In the middle section are sample sizes for various confidence levels, from 80% to 99.99%. In most situations the 95% confidence level is used, therefore in this example, the sample size is 384.

The sample size formula used is as follows:

$$n = deff \times \frac{N\hat{p}\hat{q}}{\frac{d^2}{1.96^2}(N-1) + \hat{p}\hat{q}}$$

where

n = sample size

deff = design effect

N = population size

\hat{p} = the estimated proportion

$\hat{q} = 1 - \hat{p}$

d = desired absolute precision or absolute level of precision

References

Schaeffer RL, Mendenhall W, Ott L. Elementary Survey Sampling, Fourth Edition. Duxbury Press, Belmont, California 1990.