Sample Size for Specified Margin of Error

There are two formulas we will work with, and both require that you (or someone) specify first the confidence level **C** and the margin of error **m** required.

Generally, the bigger you make the sample size the smaller will be the confidence interval for a given confidence level. A small confidence interval is good, but expensive if it requires an unreasonably huge sample size to produce. Expensive is bad. We want to estimate—**and we should emphasize that these are just estimates**—how large the sample size $\boldsymbol{n}$ must be to produce a confidence interval of specified width (or smaller.)

In any actual experiment, the margin of error is determined by the data and is not in your control and must be reported to be whatever it turns out to be! The estimate below tells you how big your sample size should be if you hope the margin of error will turn out to be near $\boldsymbol{m}$.

Our goal is to take a sample and produce an interval $\hat{\boldsymbol{p}} \pm \boldsymbol{m}$ or $\overline{\boldsymbol{x}} \pm \boldsymbol{m}$ that will capture $\boldsymbol{\mu}$ or $\boldsymbol{p}$ with probability **C**.

The first formula below is for a confidence interval for the **population proportion** of a value of a categorical variable, and the second is for a confidence interval for the **population mean** of a numerical variable.

The formulas for the $\boldsymbol{n}$ which would be expected to do this are

$$\boldsymbol{n} \geq \left(\frac{\boldsymbol{z^*}}{\boldsymbol{m}}\right)^{\boldsymbol{2}} \boldsymbol{p^*(1-p^*)} \qquad \text{or} \qquad \boldsymbol{n} \geq \left(\frac{\boldsymbol{z^* s}}{\boldsymbol{m}}\right)^{\boldsymbol{2}}.$$

$\boldsymbol{z^*}$ is the z-score of the confidence level $\boldsymbol{C}$.

$\boldsymbol{p^*}$ is your guess about the population proportion. That could come from a pilot study or other information about the background population. It should be close to the $\hat{\boldsymbol{p}}$ you end up getting, but of course **there is no way to guarantee that.** The "safe" choice is to choose $\boldsymbol{p^*} = 0.5$ but that might result in an $\boldsymbol{n}$ considerably larger than you needed.

$\boldsymbol{s}$ is an estimate of the standard deviation $\boldsymbol{\sigma}$ of the background population (which you ordinarily will not know.) Sometimes it comes from a pilot study, sometimes from knowledge of other similar distributions. You hope that $\boldsymbol{z^* s}$ will end up being at least as big as $\boldsymbol{t^* S_x}$ where $\boldsymbol{t^*}$ is the t-score of confidence level $\boldsymbol{C}$ and $\boldsymbol{S_x}$ is the sample standard deviation of the sample you end up picking. **But there is no way to guarantee that it will be!** Pick the number $\boldsymbol{s}$ to be a bit larger than your estimate of the population standard deviation $\boldsymbol{\sigma}$ and hope for the best.

Example: Population Proportion

You have a huge field of flowers and want to estimate how many blue ones are in the field. Someone told you that in the past it was found that the proportion of blue flowers was about 0.2 (that is, 20% blue flowers.)

You want a 95% confidence interval for these flowers to be 0.02 wide so the margin of error you require is half that, $\mathbf{m} = 0.01$.

To be reasonably sure that you will accomplish this, your random sample size should be *at least*

$$\left(\boldsymbol{\frac{z^*}{m}}\right)^{\mathbf{2}} \boldsymbol{p^*(1-p^*)} = \left(\frac{2}{0.01}\right)^2 (0.2)\,(0.8) = 6400.$$

Notice that if you were willing to settle for a 90% confidence interval and the wider target margin of error of 0.02 we get

$$\left(\boldsymbol{\frac{z^*}{m}}\right)^{\mathbf{2}} \boldsymbol{p^*(1-p^*)} = \left(\frac{1.645}{0.02}\right)^2 (0.2)\,(0.8) \approx 1083$$

and a sample of this size is much more manageable.

If you start out with no idea of the fraction of blue flowers you would choose $\boldsymbol{p^*} = 0.5$ and would choose a sample size at least

$$\left(\boldsymbol{\frac{z^*}{m}}\right)^{\mathbf{2}} \boldsymbol{p^*(1-p^*)} = \left(\frac{1.645}{0.02}\right)^2 (0.5)\,(0.5) \approx 1691.$$

Example: Population Mean

You want to create a 95% confidence interval for the average weight of a species of fish and need to know how big the random sample must be if you would like to have that confidence interval no more than 0.5 kilograms wide so the margin of error you want is 0.25 kilograms.

A pilot study has been done and based on that and other information you estimate that the standard deviation of weight in this species is 1.5 kilograms.

To accomplish your margin-of-error goal you should take a random sample of size at least

$$\left(\boldsymbol{\frac{z^* s}{m}}\right)^{\mathbf{2}} = \left(\frac{(2)(1.5)}{0.25}\right)^2 = 144.$$

If you want a 99% confidence interval to be only 0.5 kilograms wide you would need a sample size at least

$$\left(\boldsymbol{\frac{z^* s}{m}}\right)^{\mathbf{2}} = \left(\frac{(2.576)(1.5)}{0.25}\right)^2 \approx 239.$$