**SAMPLE SIZE**

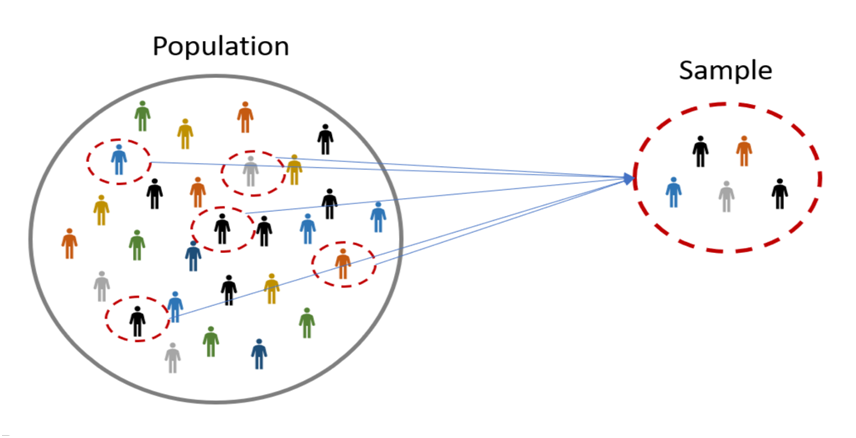
**Introduction**

When you survey a large population of respondents, you’re interested in the entire group, but it’s not realistically possible to get answers or results from absolutely everyone. So you take a random sample of individuals which represents the population as a whole.

A sample is a subset or a smaller, manageable group of individuals, items, or data points that is selected from a larger population or dataset.

**What is Sample Size?**

Sample size is the number of observations or individuals included in a study or experiment. It is the number of individuals, items, or data points selected from a larger population to represent it statistically.



**Example:**

Imagine you work at a chocolate factory, and your job is to ensure the quality of chocolate bars as they come off the production line. To do this, you cannot taste all the chocolates in a batch; instead, you take a random sample of chocolate bars from each batch.

Suppose you took a sample of 10 chocolate bars from each batch of 100 bars for quality control; here, 10 is the sample size.

[The Sample Size Explained in One Minute: From Definition to Examples and Research Tips](https://www.youtube.com/watch?v=Uyd_Fk9cDjA)

**So why does sample size matter?**

When conducting a survey or study involving a large population, it's often impossible to collect responses or data from every single individual within that population.

The size of the sample is very important for getting accurate, statistically significant results and running your study successfully.

* If your sample is too small, you may include a disproportionate number of individuals which are outliers and anomalies. These skew the results and you don’t get a fair picture of the whole population.
* If the sample is too big, the whole study becomes complex, expensive and time-consuming to run, and although the results are more accurate, the benefits don’t outweigh the costs.

[Lesson 5 The importance of sample size](https://www.youtube.com/watch?v=_qO_qtLN6Nw) (Google Data Analytics)

**What is sample size determination?**

Sample size determination is the process of choosing the right number of observations or people from a larger group to use in a sample. The primary objective of this process is to strike a balance between having a sample size that is sufficiently large to yield statistically valid and accurate findings while also keeping it manageable and cost-effective.

The choice of an optimal sample size depends on various factors, including the size of the overall population, the desired level of precision for estimating population parameters, the desired level of confidence in the study's outcomes, the expected variability within the population, and the available resources in terms of time and budget. Statistical methods are commonly employed to calculate the required sample size for specific research inquiries and study designs.

Determining the appropriate sample size is of paramount importance as it ensures the credibility and reliability of research results and conclusions.

**Factors Influencing Sample size:**

**1.** **Population Size**

A population is the entire group that you want to draw conclusions about. It is from the population that a sample is selected, using probability or non-probability samples.

The population size may be known (such as the total number of employees in a company), or unknown (such as the number of pet keepers in a country)

**2.** **Confidence Interval (Margin of Error)**

Confidence interval tells you how confident you can be that the results from a study reflect what you would expect to find if it were possible to survey the entire population being studied. The confidence interval is usually a plus or minus (±) figure.

For example, if your confidence interval is 5 and 60% percent of your sample picks an answer, you can be confident that if you had asked the entire population, between 55% (60-5) and 65% (60+5) would have picked that answer.

**3.** **Confidence Level**

The confidence level refers to the percentage of probability, or certainty that the confidence interval would contain the true population parameter when you draw a random sample many times

For example, a 99% confidence level means that should you repeat an experiment or survey over and over again, 99 percent of the time, your results will match the results you get from a population.

The larger your sample size, the more confident you can be that their answers truly reflect the population. In other words, the larger your sample for a given confidence level, the smaller your confidence interval.

**4.** **Standard Deviation**

Standard deviation measures a data set’s distribution from its mean. In calculating the sample size, the standard deviation is useful in estimating how much the responses you will receive vary from each other and from the mean number.

**How to calculate Sample size?**

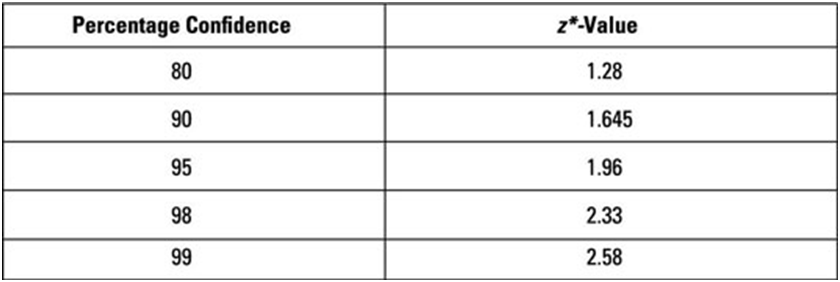
1. Determine the population size (if known).

2. Determine the confidence interval.

3. Determine the confidence level.

4. Determine the standard deviation (a standard deviation of 0.5 is a safe choice where the figure is unknown)

5. Convert the confidence level into a Z-Score. This table shows the z-scores for the most common confidence levels:



6. Put these figures into the sample size formula to get your sample size.

**Method we follow** – (Based on Roscoe’s (1975) guidelines)

* Sample size larger than 30 and less than 500 is appropriate for most research studies.
* A minimum sample size of 30 for each subset.
* In multivariate research sample size should be preferably 10 times or more as the number of variables in the study.
* For simple experimental research with tight control (qualitative research), a sample size of 10-20 is adequate.
* 95% level of confidence (5% level of significance) is the accepted level in business research
* if we know N and SD n = [z2 \* p(1-p)] / e2 / 1 + [z2 \* p(1-p)] / e2 \* N]
* if SD is unknown n=[z2\* p(1-p)]/e2]
* Slovin's formula n=N/[1+Ne^2]

**Examples:**

Suppose you choose to work with a 95% confidence level, a standard deviation of 0.5(estimated population proportion), and a confidence interval (margin of error) of ± 5%, you just need to substitute the values in the formula:

**Case 1**: Finite Population (N and SD are known)

N= 500 and SD=0.5

Sample size equation is,

n = [z2 \* p(1-p)] / e2 / 1 + [z2 \* p(1-p)] / e2 \* N]

=[ (1.96)^2\*(0.5)(0.5)/(0.05)^2]/ 1 +[ (1.96)^2\*(0.5)(0.5)/(0.05)^2\*N]

=217

**Case 2**: Infinite population

Sample size

n = [z2\* p(1-p)/e2]

= [(1.96)^2\*(0.5)(0.5)/(0.05)^2]

=384

**Case 3**: Solvin’s equation

Sample size

n = N/[1+Ne^2]

= 500/[1+(500\*(0.05)^2)]

= 222

For more examples, and to give it a try yourself, here's the link to the sheet :[Formula Sample Size - Operations](https://docs.google.com/spreadsheets/d/1zy4O241i6pGBLMPFr7uVxQH-8LKi15NzmBvbL3DpJ14/edit#gid=0)