

1.2: Population and sample

Let us research the simple case: which of two ice-creams is more popular? It would be relatively easy to gather all information if all these ice-creams sold in one shop. However, the situation is usually different and there are many different sellers which are really hard to control. In situation like that, the best choice is **sampling**. We cannot control everybody but we can control somebody. Sampling is also cheaper, more robust to errors and gives us free hands to perform more data collection and analyses. However, when we receive the information from sampling, another problem will become apparent—how representative are these results? Is it possible to estimate the small piece of sampled information to the whole big **population** (this is not a biological term) of ice-cream data? Statistics (mathematical statistics, including the theory of sampling) could answer this question.

It is interesting that sampling could be more precise than the total investigation. Not only because it is hard to control all variety of cases, and some data will be inevitably mistaken. There are many situations when the smaller size of sample allows to obtain more detailed information. For example, in XIX century many Russian peasants did not remember their age, and all age-related total census data was rounded to tens. However, in this case selective but more thorough sampling (using documents and cross-questioning) could produce better result.

And philosophically, full investigation is impossible. Even most complete research is a subset, sample of something bigger.

This page titled [1.2: Population and sample](#) is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by [Alexey Shipunov](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.