

1.3: How to obtain the data

There are two main principles of sampling: replication and randomization.

Replication suggests that the same effect will be researched several times. This idea derived from the cornerstone math “big numbers” postulate which in simple words is “the more, the better”. When you count replicates, remember that they must be independent. For example, if you research how light influences the plant growth and use five growing chambers, each with ten plants, then number of replicates is five, not fifty. This is because plants within each chamber are not independent as they all grow in the same environment but we research differences between environments. Five chambers are replicates whereas fifty plants are *pseudoreplicates*.

Repeated measurements is another complication. For example, in a study of short-term visual memory ten volunteers were planned to look on the same specific object multiple times. The problem here is that people may remember the object and recall it faster towards the end of a sequence. As a result, these multiple times are not replicates, they are repeated measurements which could tell something about learning but not about memory itself. There are only ten true replicates.

Another important question is how many replicates should be collected. There is the immense amount of publications about it, but in essence, there are two answers: (a) as many as possible and (b) 30. Second answer looks a bit funny but this rule of thumb is the result of many years of experience. Typically, samples which size is less than 30, considered to be a small. Nevertheless, even minuscule samples could be useful, and there are methods of data analysis which work with five and even with three replicates. There are also special methods (*power analysis*) which allow to estimate how many objects to collect (we will give one example due course).

Randomization tells among other that every object should have the equal chances to go into the sample. Quite frequently, researchers think that data was randomized while it was not actually collected in the random way.

For example, how to select the sample of 100 trees in the big forest? If we try simply to walk and select trees which somehow attracted the attention, this sample will not be random because these trees are somehow deviated and this is why we spotted them. Since one of the best ways of randomization is to *introduce the order which is knowingly absent in nature* (or at least not related with the study question), the reliable method is, for example, to use a detailed map of the forest, select two random coordinates, and find the tree which is closest to the selected point. However, trees are not growing homogeneously, some of them (like spruces) tend to grow together whereas others (like oaks) prefer to stay apart. With the method described above, spruces will have a better chance to come into sample so that breaks the rule of randomization. We might employ the second method and make a transect through the forest using rope, then select all trees touched with it, and then select, saying, every fifth tree to make a total of hundred.

Is the last (second) method appropriate? How to improve it?

Now you know enough to answer another question:

Once upon a time, there was an experiment with a goal to research the effect of different chemical poisons to weevils. Weevils were held in jars, chemicals were put on fragments of filter paper. Researcher opened the jar, then picked up the weevil which first came out of jar, put it on the filter paper and waited until weevil died. Then researcher changed chemical, and start the second run of experiment in the same dish, and so on. But for some unknown reason, the first chemical used was always the strongest (weevils died very fast). Why? How to organize this experiment better?

This page titled [1.3: How to obtain the data](#) 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.