What is Sampling?

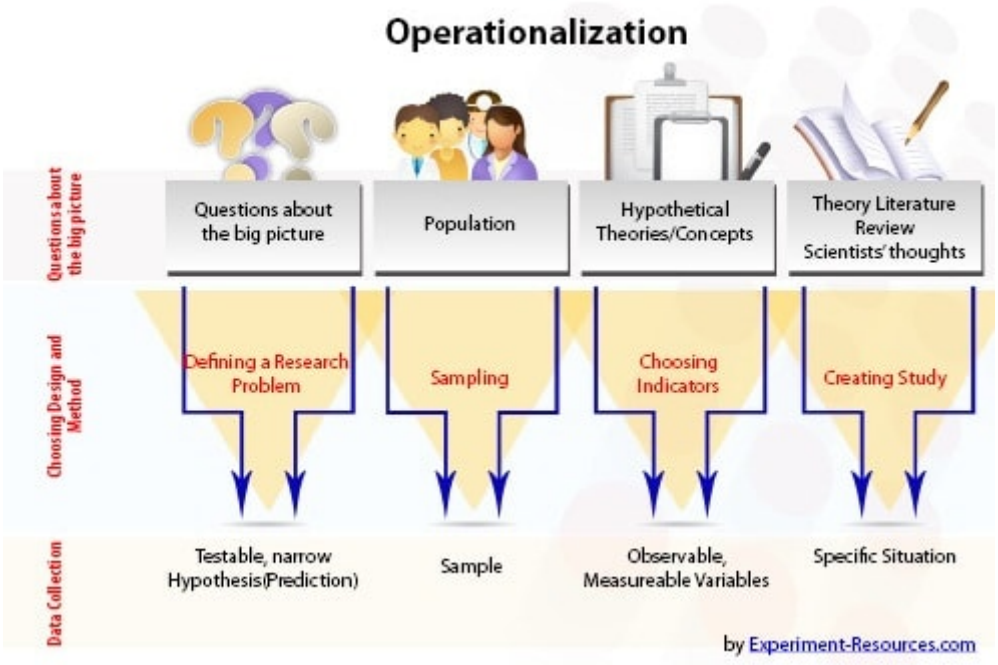
Martyn Shuttleworth160.9K reads

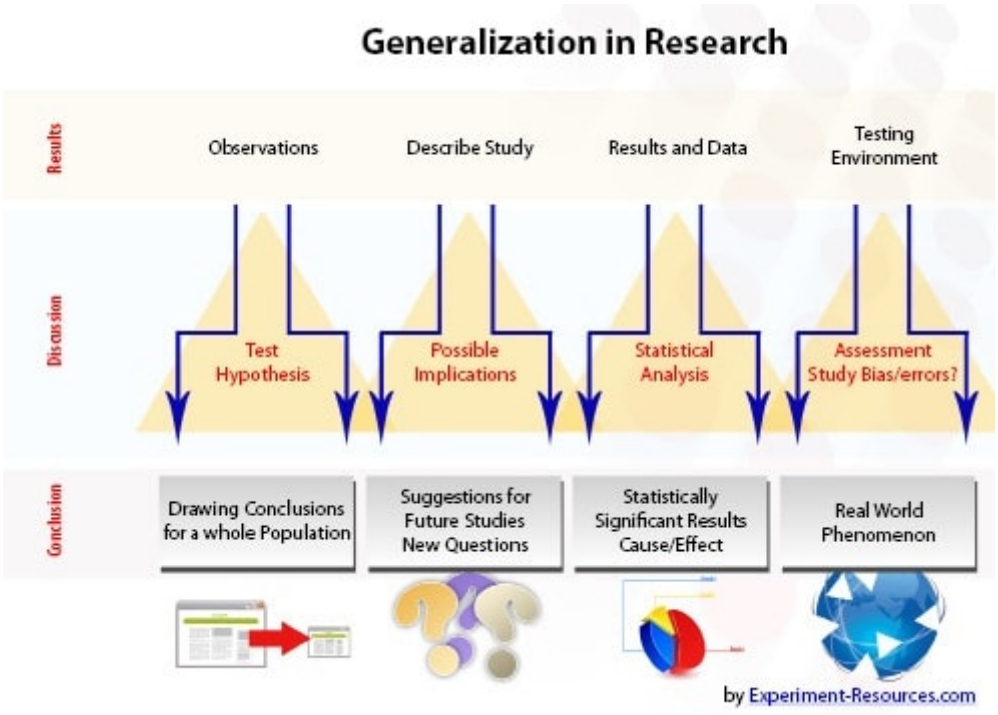
In many experiments, sampling an entire population as part of a research experiment is impossible, due to the time, expense and sheer number of subjects.

What is Sampling? Imagine, for example, an experiment to test the effects of a new education technique on schoolchildren. It would be impossible to select the entire school age population of a country, divide them into groups and perform research.

A research group sampling the diversity of flowers in the African savannah could not count every single flower, because it would take many years.

This is where statistical sampling comes in, the idea of trying to take a representative section of the population, perform the experiment and extrapolate it back to the population as a whole.





In the education example, the research group could test all of the schools in a city, or select one school in a few different cities. Of course, the process is not that easy, and the researchers must use a battery of statistical techniques, and a good research design, to ensure that this subset is as representative as possible.

Failure to take into account all of the various experimental biases and errors that can creep into an experiment, if the sample group is chosen poorly, will inevitably lead to invalid results.

The basic question that a researcher should be asking when selecting a sample group is:

> How many subjects will I need to complete a viable study, and how will I select them?

The Advantages of Sampling

- It involves a smaller amount of subjects, which reduces investment in time and money.
- Sampling [1] can actually be more accurate than studying an entire population, because it affords researchers a lot more control over the subjects. Large studies can bury interesting correlations [2] amongst the ‘noise.'
- Statistical manipulations are much easier with smaller data sets, and it is easier to avoid human error when inputting and analyzing the data.

The Disadvantages of Sampling

- There is room for potential bias in the selection of suitable subjects for the research. This may be because the researcher selects subjects that are more likely to give the desired results, or that the subjects tend to select themselves.

  For example, if an opinion poll company canvasses opinion by phoning people between 9am and 5pm, they are going to miss most people who are out working, totally invalidating their results. These are called determining factors, and also include poor experiment design [3], confounding variables [4] and human error [5].

- Sampling requires a knowledge of statistics [6], and the entire design of the experiment [3] depends upon the exact sampling method required.

Selecting Sample Groups and Extrapolating Results

When sampling, a researcher has two distinct choices:

1. Ideally, they will take a representative sample of the whole population and use randomization techniques [7] to establish sample groups and controls [8].
2. In many cases, this is not always possible, and the make-up of the groups has to be assigned.

For example, a study that needs to ask for volunteers is never representative of a population. In such cases, the researcher needs to be aware that they cannot extrapolate [9] the findings to represent an entire population.

A study into heart disease that only looks at middle aged men, between 40 and 60, will say very little about heart disease in women or younger men, although it can always be a basis for future research involving other groups.

However robust the research design [10], there is always an inherent inaccuracy with any sample-based experiment, due to chance fluctuations and natural variety. Most statistical tests take this into account, and this is why results are judged to a significance level [11], or given a margin of error [12].

Sampling is an essential part of most research, and researchers must know how to choose sample groups that are as free from bias as possible, and also be aware of the extent to which they can extrapolate their results back to the general population.

Links
[1] https://explorable.com/population-sampling
[2] https://explorable.com/statistical-correlation
[3] https://explorable.com/design-of-experiment
[4] https://explorable.com/confounding-variables
[5] https://explorable.com/systematic-error
[6] https://explorable.com/statistics-tutorial
[7] https://explorable.com/random-sampling-error
[8] https://explorable.com/scientific-control-group
[9] https://explorable.com/what-is-generalization
[10] https://explorable.com/research-designs
[11] https://explorable.com/significance-test
[12] https://explorable.com/statistics-margin-of-error