**Sampling**

In statistics, quality assurance, and survey methodology, sampling is the selection of a subset (a statistical sample) of individuals from within a statistical population to estimate characteristics of the whole population. ... Results from probability theory and statistical theory are employed to guide the practice.



Samples are parts of a population. For example, you might have a list of information on 100 people (your “sample”) out of 10,000 people (the “population”). You can use that list to make some assumptions about the entire population’s behavior.

However, it’s not that simple. When you do stats, your sample size has to be ideal—not too large or too small. Then once you’ve decided on a sample size, you must use a sound technique to collect the sample from the population:

Probability Sampling uses randomization to select sample members. You know the probability of each potential member’s inclusion in the sample. For example, 1/100. However, it isn’t necessary for the odds to be equal. Some members might have a 1/100 chance of being chosen, others might have 1/50.

Non-probability sampling uses non-random techniques (i.e. the judgment of the researcher). You can’t calculate the odds of any particular item, person or thing being included in your sample.

**Common Types**
The most common techniques you’ll likely meet in elementary statistics or AP statistics include taking a sample with and without replacement. Specific techniques include:

* Bernoulli samples have independent Bernoulli trials on population elements. The trials decide whether the element becomes part of the sample. All population elements have an equal chance of being included in each choice of a single sample. The sample sizes in Bernoulli samples follow a binomial distribution. Poisson samples (less common): An independent Bernoulli trial decides if each population element makes it to the sample.
* Cluster sampes divide the population into groups (clusters). Then a random sample is chosen from the clusters. It’s used when researchers don’t know the individuals in a population but do know the population subsets or groups.
* In systematic sampling, you select sample elements from an ordered frame. A sampling frame is just a list of participants that you want to get a sample from. For example, in the equal-probability method, choose an element from a list and then choose every kth element using the equation k = N\n. Small “n” denotes the sample size and capital “N” equals the size of the population.
* SRS : Select items completely randomly, so that each element has the same probability of being chosen as any other element. Each subset of elements has the same probability of being chosen as any other subset of k elements.
* In stratified sampling, sample each subpopulation independently. First, divide the population into homogeneous (very similar) subgroups before getting the sample. Each population member only belongs to one group. Then apply simple random or a systematic method within each group to choose the sample. Stratified Randomization: a sub-type of stratified used in clinical trials. First, divide patients into strata, then randomize with permuted block randomization.

**Primary data**

An advantage of using primary data is that researchers are collecting information for the specific purposes of their study. In essence, the questions the researchers ask are tailored to elicit the data that will help them with their study. Researchers collect the data themselves, using surveys, interviews and direct observations.

**Secondary data**

There are several types of secondary data. They can include information from the national population census and other government information collected by Statistics Canada. One type of secondary data that’s used increasingly is administrative data. This term refers to data that is collected routinely as part of the day-to-day operations of an organization, institution or agency. There are any number of examples: motor vehicle registrations, hospital intake and discharge records, workers’ compensation claims records, and more.