

Let's learn about sampling theory!

Sampling Theory is a study of relationships existing between a population and the samples drawn from the population. For instance, the population may be the entire population of the United States, while the sample may be 1000 people you polled for a survey.

There are a ton of important questions we want to ask pertaining to sampling theory, such as whether the observed difference between two samples are due to chance variation or if they are statistically significant. These are sometimes known as *tests of significance*. We'll talk about them later in another iPython notebook.

## Random Samples and Numbers

Samples must be chosen so that they are representative of a population. One method of doing this is by selecting using *random sampling*. This method means that every member of the population has an equal chance of being selected. There's lots of ways to create a "random" sampling method, it can be as simple as picking names from a hat, or using pseudorandom number generators such as a Mersenne Twister (the default PRNG for python).

Learn more about it at: <https://docs.python.org/2/library/random.html>  
(<https://docs.python.org/2/library/random.html>)

## Sampling with and without Replacement

Drawing samples without replacement from a finite source of samples is considered a finite sampling. Drawing samples from a finite source with replacement can be considered infinite. Such as tossing a coin N number of times, while there are only two possible outcomes, there is an infinite possibility of any series of heads or tails as N goes to infinity.

## Sampling Distribution of Means

If you draw N samples from a finite population of  $N_p$  samples where  $N_p > N$  then the mean and the std are denoted as the following:

The mean:

$$\mu_x = \mu$$

The standard deviation:

$$\sigma_x = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$$

This means that we can relate our mean and std from our sample to the total sample population.

## Sampling Distribution of Proportions

If the population is infinite and the probability of success is denoted as  $p$  and probability of failure is  $q=1-p$  (just like in a binomial distribution, like a coin flip). The sampling distribution of proportions then has a mean and standard deviation notes as the following:

The mean:

$$\mu_p = p$$

The standard deviation:

$$\sigma_p = \sqrt{\frac{pq}{N}} = \sqrt{\frac{p(1-p)}{N}}$$

## Sampling Distribution of Differences and Sums

If you are given two sample populations  $N_1$  and  $N_2$  both of which are normally distributed, we can obtain the mean and standard deviation of the sampling distribution of differences or sums of the statistics. If  $S_1$  is a statistics of  $N_1$  and  $S_2$  a statistic of  $N_2$  we can obtain the mean and standard deviation as follows:

For the *differences of the statistics* :

The mean is:

$$\mu_{S_1-S_2} = \mu_{S_1} - \mu_{S_2}$$

The standard deviation is:

$$\sigma_{S_1-S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

For the *sum of the statistics* :

The mean is:

$$\mu_{S_1+S_2} = \mu_{S_1} + \mu_{S_2}$$

The standard deviation is:

$$\sigma_{S_1+S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

Notice that the standard deviations are summed for both cases, but the means are not!

This is just a quick intro to the topic of sampling theory, check out more here:

1.) [http://en.wikipedia.org/wiki/Sampling\\_distribution](http://en.wikipedia.org/wiki/Sampling_distribution)  
([http://en.wikipedia.org/wiki/Sampling\\_distribution](http://en.wikipedia.org/wiki/Sampling_distribution))

