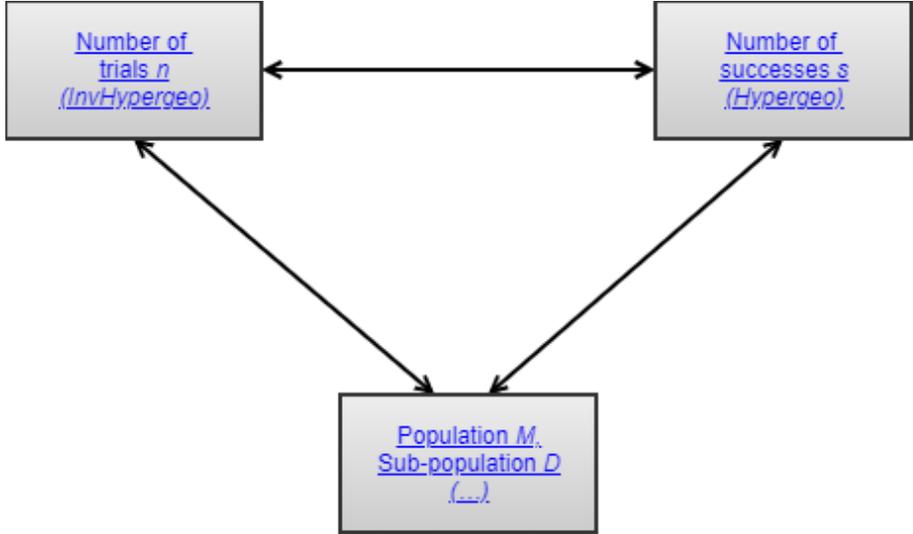


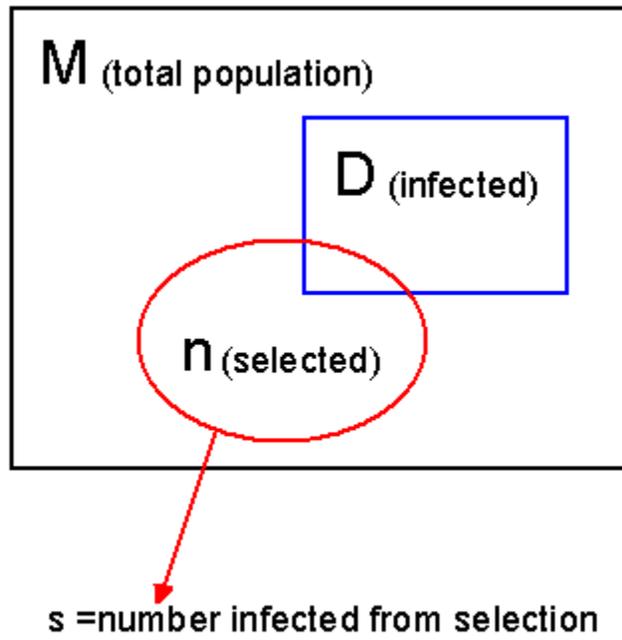
# Introduction - The Hypergeometric Process

## The Hypergeometric Process



## Description

The hypergeometric process occurs when one is sampling randomly without replacement from some population (as opposed to sampling with replacement in the [Binomial Process](#)), and where one is counting the number in that sample that have some particular characteristic. This is a very common type of scenario. For example, population surveys, herd testing, and lotto are all hypergeometric processes. In many situations, the population is very large in comparison to the sample and we can assume that if a sample was put back into the population, the probability is very small that it would be picked again. In that case, each sample would have the same probability of picking an individual with a particular characteristic: in other words this becomes a binomial process. When the population is not very large compared to the sample (a good rule is that the population is less than ten times the size of the sample) we cannot make a [binomial approximation to the hypergeometric](#). This section discusses the distributions associated with the hypergeometric process.



The figure above demonstrates the four parameters of the Hypergeometric process: The population one is sampling from ( $M$ ); the sub-population of interest ( $D$ ), the number being randomly sampled from the population ( $n$ ) and the number ( $s$ ) in that sample that come from  $D$ . We recommend that you draw out a diagram like this when you are faced with a hypergeometric problem to keep that all clear!

1. [Number in a sample with a particular characteristic](#)
2. [Number of samples to get a specific  \$s\$](#)
3. [Number of samples that were taken to have observed a specific  \$s\$](#)
4. [Estimate of population and sub-population sizes](#)

Once you have reviewed the material in this section, you might like to test how much you have learned by taking the self-test quiz:

Take

A quiz on The hypergeometric process: [quiz](#)

Summary of results for the hypergeometric process

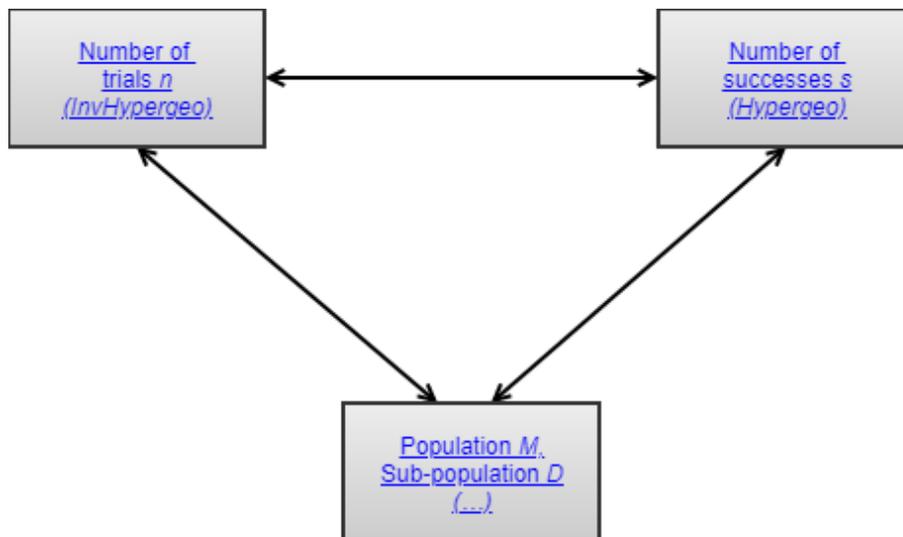
Quantity	Formula	Notes
Number of sub-population in the sample	<i>In Crystal Ball version 5.5-:</i> $s = \text{Hypergeometric}(D/M, n, M)$ <i>In Crystal Ball version 7.0+:</i> $s = \text{Hypergeometric}(D, n, M)$	
Number of samples to observe $s$ from the sub-population	$n = \text{InvHypergeo}(s, D, M)$	
Number of samples there were to have observed $s$ from the sub-population	$n = \text{InvHypergeo}(s, D, M)$	Where the last sample <i>is</i> known to have been from the sub-population

Number of samples $n$ there were before having observed $s$ from the sub-population	$f(n) \propto \frac{n!(M-n)!}{(n-s)!(M-D-n+s)!}$	Where the last sample is <i>not</i> known to have been from the sub-population. This uncertainty distribution needs to be normalized.
Size of sub-population $D$	$f(D) \propto \frac{D!(M-D)!}{(D-s)!(M-D-n+s)!}$	This uncertainty distribution needs to be normalized.
Size of population $M$	$f(M) \propto \frac{(M-D)!(M-n)!}{M!(M-D-n+s)!}$	This uncertainty distribution needs to be normalized.

Useful Excel Functions:

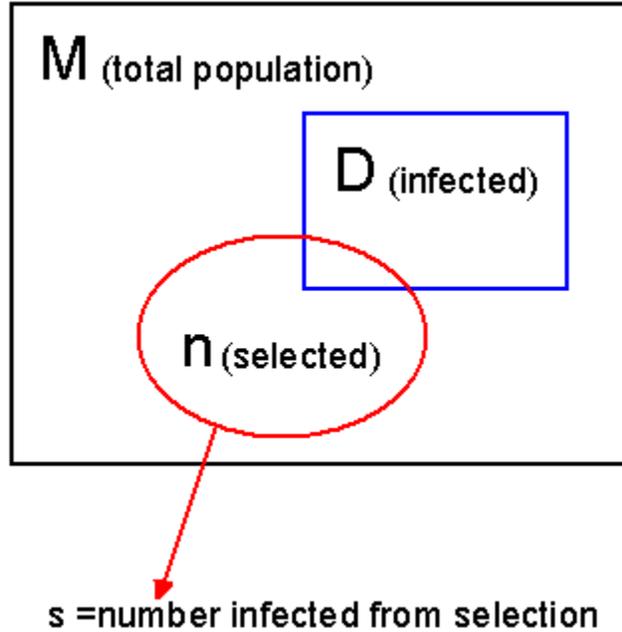
Use	Function	Explanation
Hypergeometric probability	=HYPGEOMDIST(x,n,D,M)	The hypergeometric probability of observing exactly $s$ in a sample of size $n$ that come from sub-population $D$ in a total population $M$ .
Combinations	=COMBIN(n,x)	The <b>binomial coefficient</b> ${}_n C_x = n!/(s!(n-s)!)$
Factorial	=FACT(x)	$x!$

### The Hypergeometric Process



Description

The hypergeometric process occurs when one is sampling randomly without replacement from some population (as opposed to sampling with replacement in the [Binomial Process](#)), and where one is counting the number in that sample that have some particular characteristic. This is a very common type of scenario. For example, population surveys, herd testing, and lotto are all hypergeometric processes. In many situations, the population is very large in comparison to the sample and we can assume that if a sample was put back into the population, the probability is very small that it would be picked again. In that case, each sample would have the same probability of picking an individual with a particular characteristic: in other words this becomes a binomial process. When the population is not very large compared to the sample (a good rule is that the population is less than ten times the size of the sample) we cannot make a [binomial approximation to the hypergeometric](#). This section discusses the distributions associated with the hypergeometric process.



The figure above demonstrates the four parameters of the Hypergeometric process: The population one is sampling from ( $M$ ); the sub-population of interest ( $D$ ), the number being randomly sampled from the population ( $n$ ) and the number ( $s$ ) in that sample that come from  $D$ . We recommend that you draw out a diagram like this when you are faced with a hypergeometric problem to keep that all clear!

1. [Number in a sample with a particular characteristic](#)
2. [Number of samples to get a specific  \$s\$](#)
3. [Number of samples that were taken to have observed a specific  \$s\$](#)
4. [Estimate of population and sub-population sizes](#)

Once you have reviewed the material in this section, you might like to test how much you have learned by taking the self-test quiz:

Take

A quiz on The hypergeometric process: [quiz](#)

Summary of results for the hypergeometric process

Quantity	Formula	Notes
----------	---------	-------

Number of sub-population in the sample	<p><i>In Crystal Ball version 5.5-:</i> <math>s = \text{Hypergeometric}(D/M, n, M)</math></p> <p><i>In Crystal Ball version 7.0+:</i> <math>s = \text{Hypergeometric}(D, n, M)</math></p>	
Number of samples to observe $s$ from the sub-population	$n = \text{InvHypergeo}(s, D, M)$	
Number of samples there were to have observed $s$ from the sub-population	$n = \text{InvHypergeo}(s, D, M)$	Where the last sample <i>is</i> known to have been from the sub-population
Number of samples $n$ there were before having observed $s$ from the sub-population	$f(n) \propto \frac{n!(M-n)!}{(n-s)!(M-D-n+s)!}$	Where the last sample is <i>not</i> known to have been from the sub-population. This uncertainty distribution needs to be normalized.
Size of sub-population $D$	$f(D) \propto \frac{D!(M-D)!}{(D-s)!(M-D-n+s)!}$	This uncertainty distribution needs to be normalized.
Size of population $M$	$f(M) \propto \frac{(M-D)!(M-n)!}{M!(M-D-n+s)!}$	This uncertainty distribution needs to be normalized.

Useful Excel Functions:

Use	Function	Explanation
Hypergeometric probability	<code>=HYPGEOMDIST(x,n,D,M)</code>	The hypergeometric probability of observing exactly $s$ in a sample of size $n$ that come from sub-population $D$ in a total population $M$ .
Combinations	<code>=COMBIN(n,x)</code>	The <b>binomial coefficient</b> ${}_n C_x = n!/(s!(n-s)!)$
Factorial	<code>=FACT(x)</code>	$x!$