Semester Project

Jake Vissicchio
Jake.Vissicchio1@Marist.edu

December 14, 2021

1 Results

Case 1: Represents a pool where there were no infected samples.
Case 2: Represents when there is one or more infected samples within one half of the pool.
Case 3: Represents where there are infected samples within both halves of the pool.

Expected Results:
Case (1): $125 \times 0.8500 = 106.25$ instances requiring 107 tests
Case (2): $125 \times 0.1496 = 18.70$ instances requiring 131 tests
Case (3): $125 \times 0.0004 = 0.05$ round up to 1 instance requiring 11 tests

The table below represents the simulated results using different total population sizes...

Pooled Testing Results Using Different Population Sizes				
Population:	Total Tests:	Case 1 Total:	Case 2 Total:	Case 3 Total:
1,000	243	106	18	1
10,000	2,410	1,064	175	11
100,000	23,960	10,642	1,780	78
1M	240,028	106,326	17,928	746

As shown, the results are close to the expected results. Also, each time we add a 0 to the population we often see each case amount go up by a new digit which fits.

2 Binomial Distribution & Hypergeometric Distribution

2.1 Binomial Distribution

Definition:
Binomial Distribution is a discrete probability distribution of the possible number of successful outcomes in a given number of trials in each of which there is the same probability of success.

Formula:

$$P_x = \binom{n}{x} p^x q^{n-x}$$

P = binomial probability
x = number of times for a specific outcome within n trials
$\binom{n}{x}$ = number of combinations
p = probability of success on a single trial
q = probability of failure on a single trial
n = number of trials

2.2 Hypergeometric Distribution

Definition:
Hypergeometric Distribution is a discrete probability distribution that describes the probability of success given a number of trials, without replacement.

Formula:

$$P_x = \frac{{}_{s_p}C_{s_s} * {}_{n_p - s_p}C_{n_s - s_s}}{{}_{n_p}C_{n_s}}$$

P = Hypergeometric probability
s_p = number of successes in the population
s_s = number of successes in the sample
n_p = size of the population
n_s = size of the sample

2.3 Difference

For binomial distribution, the probability is the same for every trial because it is WITH replacement. For Hypergeometric Distribution, the probability changes each trial because it is WITHOUT replacement. This will

cause each trial after one another to be different.
The difference between the probabilities are very slim so it is usually okay to ignore for most applications.
It will be best to use Hypergeometric Distribution when the population is small since that is when we will see a difference due to its dependent nature.

3 How Can I Improve My Simulation?

3.1 Adding False Positives/Negatives

Perhaps the adding of false positives and negatives will add another layer of realism within my simulation. While false positives are not too much too worry about besides taking up more tests, false negatives can be extremely dangerous since it allows an infected individual to do things as if they were not infected.

3.2 Vicinity Testing

Perhaps the adding of vicinity testing will also add realism within my simulation. By vicinity testing, I mean having the pools be grouped together by people that live close to one another. This will add many more Case 3s to the results as those that are infected are likely to infect those around them.

3.3 Missing Testee

Perhaps adding a checker to see if an individual missed their test can prove to be beneficial for my simulation. At Marist, there were many times when students would skip their testing in favor of doing their Semester-Projects and would have to be reminded to go as soon as possible after or else there be punishment. This can be quite bad because they can be positive without knowing if they skipped their testing.