Project Title: Pandemic Flu Spread

1. ABSTRACT

Our objective was to apply material learned in class to simulate a flu pandemic in an elementary school classroom. The classroom consisted of 21 school children. On Day 1, 20 of the children were healthy and one child, Tommy, was sick. Tommy arrived at school every day with the potential to infect other children while he was infectious for the first three days. The probability that he infected another student is $p = 0.02$ and all children and days are independent (i.i.d. Bern(p) trials). Our project simulated the number of children who were sick on subsequent days after Tommy introduced the flu to his classmates.

We were able to simulate the flu pandemic by applying our knowledge learned in class and by using SimPy with an initial unsuccessful attempt in ARENA. This document will cover our theoretical approach, simulation and outputs as well as the challenges we faced during the project. The results of the simulation revealed that the pandemic lasted in the classroom anywhere from 4 to 17 days.

2. BACKGROUND & DESCRIPTION

The flu simulation project was an application-based problem and required programming. We implemented the project in Python using the SimPy package after an initial approach in ARENA. As a sense check, we verified our results using a basic Excel model. Our project answers the following three questions: (1) What is the distribution of kids infected on Day 1? (2) What is the expected number of infected kids on Day 1? (3) What is the expected number of infected kids on Day 2? We then simulated the flu epidemic in ARENA and Python in order to validate or reject our initial manual calculations.

3. MAIN FINDINGS

3.1 ASSUMPTIONS

The following assumptions were established when calculating statistical values and simulating our model:

1. Once a child has been infected, he or she cannot be re-infected after recovery
2. All children who are infected are still coming to school

3.2 THEORETICAL CALCULATIONS

Before beginning the simulation, the team worked together to answer the following theoretical and statistical questions.

A. DISTRIBUTION OF KIDS INFECTED ON DAY 1

The prompt of the problem indicated that each child was an independent and identically distributed Bernoulli distribution. Thus, the distribution of kids infected on Day 1 was a Binomial distribution with $p = 0.02$.

B. EXPECTED NUMBER OF INFECTED KIDS ON DAY 1

From part a, we know that this problem was Binomially distributed with each child being an independent and identically distributed trial with $p = 0.02$.

We calculated the expected number of infected kids as below assuming that we needed a perfect world scenario to initialize our simulation:

$$E[X] = np = (20)(.02) = 0.4$$

C. EXPECTED NUMBER OF INFECTED KIDS ON DAY 2

Our first approach was to recalculate the probability of infection since the initial condition from day one will ultimately change over time as more students get infected. The calculation for day 2 was attempted as below.

[Equation]

The conditional probability gave a 100% response which led to a re-work of our initial strategy as it would be impossible that there is 100% chance of infecting another child if one is already infected. As a result, our new expected probability was to calculate the number of infected students multiplied by the original p-value. Consider a scenario where Tommy infects 1 child on Day 1:

$$p = 2 \text{ infected kids} * .02 = .04$$

However, we know that there are many possible outcomes of Tommy infecting other children on Day 1. For example, if Tommy infects 0 children on day 1, the expected value on Day 2 could be:

$$E[X] = np = (20)(.02) = 0.4$$

If Tommy infects 2 children on Day 1, the expected value on Day 2 could be:

$$E[X] = np = (19)(.04) = 0.76$$

If Tommy infects 10 children on Day 1, the expected value on Day 2 could be:

$$E[X] = np = (10)(.2) = 2$$

As seen from these examples, as the number of students infected increased on Day 1, the number of expected children infected on Day 2 also increased. As more children are infected on Day 1, the more likely the rest of the class was to encountering an infectious kid, increasing their chances of becoming sick.

3.3 SIMULATION FINDINGS

To increase our opportunity to learn, the team simulated the flu pandemic in ARENA and SimPy. After simulating, an Excel spreadsheet was created to verify and validate that our understanding of the underlying statistics matched what the simulations were outputting. The below sections showcase the findings from the ARENA, SimPy, and Excel simulation methods.

When simulating, the following important components were considered:

1. The number of days a child is contagious for. This needed to be an attribute of each individual child with some calculation that added up to 3 days when he/she was infected and to indicate that he/she had recovered.
2. The number of days of the total simulation to determine when exactly the pandemic had ended
3. The number of children infected with the virus to indicate if all or any had “survived” the pandemic without being infected or not.
4. A condition stating that children cannot be re-infected as this was assumed based on the information provided.

3.3.1 ARENA APPROACH

The ARENA model began with a handful of students (CREATE modules) entering the classroom and a “Tommy” which will be handled separately due to his initial infection. The following considerations required a significant amount of trial and error in ARENA as described below:

I. How to keep track of days as the simulation cycles through. This allowed us to know when the “pandemic” had officially ended. This was attempted by having a clone of the entities to keep track of overall days that can indicate the end of the overall pandemic.

II. How to keep an internal tracker for the number of days a child is infected and indicate that he or she is no longer contagious. Similar to the overall counter, this was tracked by adding a value for each day until the 3-day limit was reached to signal the recovery from the illness.

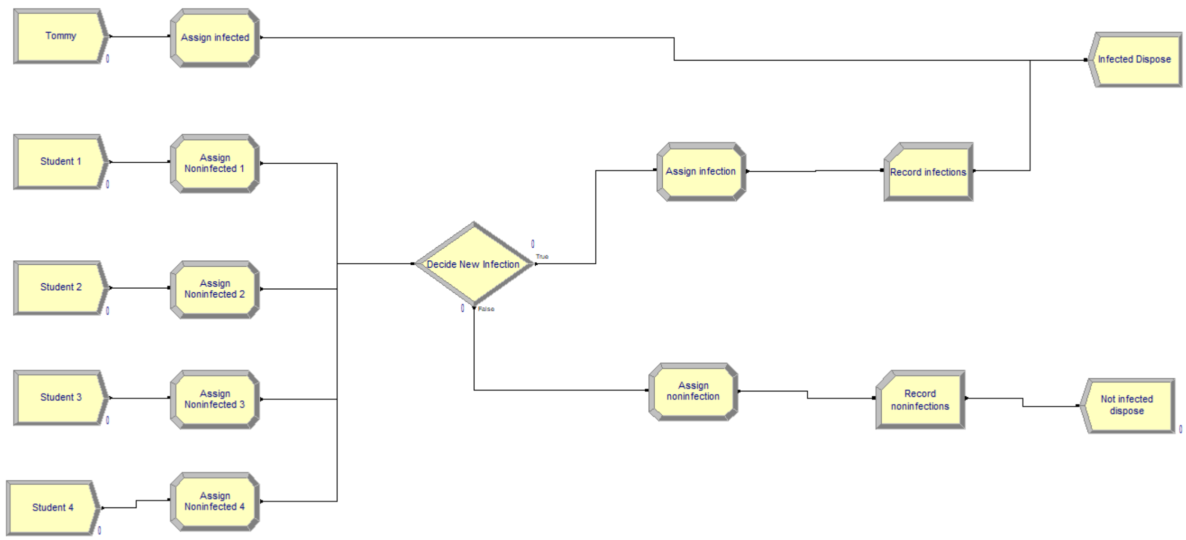

Figure 1: ARENA Simulation Model

The attempt in ARENA as shown in the figure above was proving difficult to implement due to the difficulties of the logic in keeping track of the number of days that had passed in addition to the limitations of the student, free version. The run time errors resulted in a final decision to use SimPy for the simulation.

3.3.2 SIMPY APPROACH

We implemented the simulation in Python using SimPy to see if we could use the planned logic for ARENA for improved results. We began by creating an environment with 20 healthy children, 1 infected, and 0 recovered. The program documents the number of children infected on each simulated day. A *for* loop was utilized to generate a random number between 0 and 1. If a given number was less than the calculated *p*-value, a child became infected. Each infected child was contagious for three days. The histogram in Figure 2: SimPy Histogram shows the number of children that were currently sick on a given day. With the seed chosen, there were no more infected children on the 15th day, and the pandemic ended.

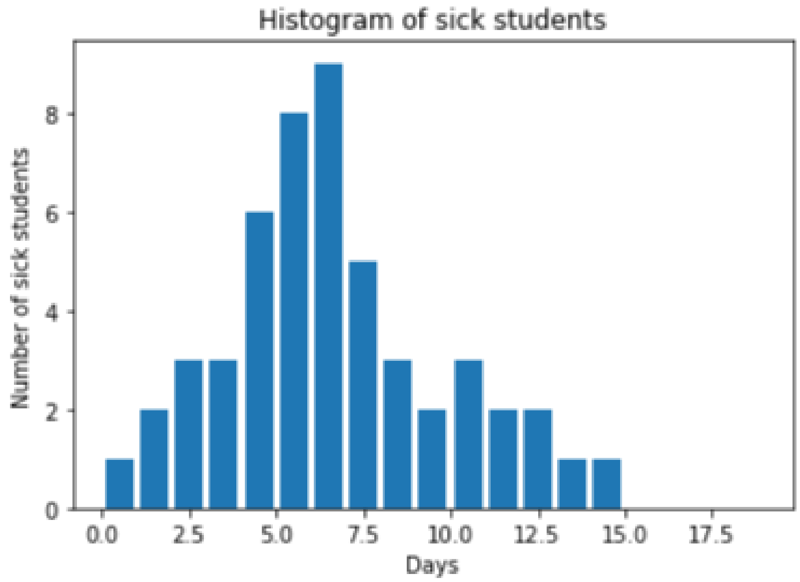

Figure 2: SimPy Histogram

With other random seeds, we saw the pandemic last anywhere from 12 to 17 days. This simulation was run for 100,000 iterations. A histogram of the length of the simulations can be found in Figure 3: Simulation Length Histogram: 100,000 Iterations.

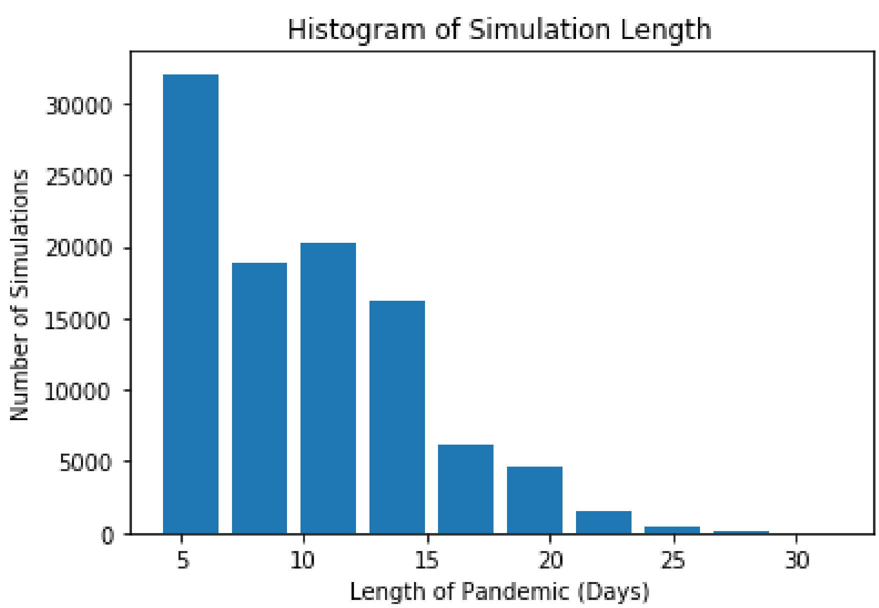

Figure 3: Simulation Length Histogram: 100,000 Iterations

Similarly, the simulation was run for 10,000 iterations. A histogram of the length of the simulations can be found in Figure 4: Simulation Length Histogram: 10,000 Iterations.

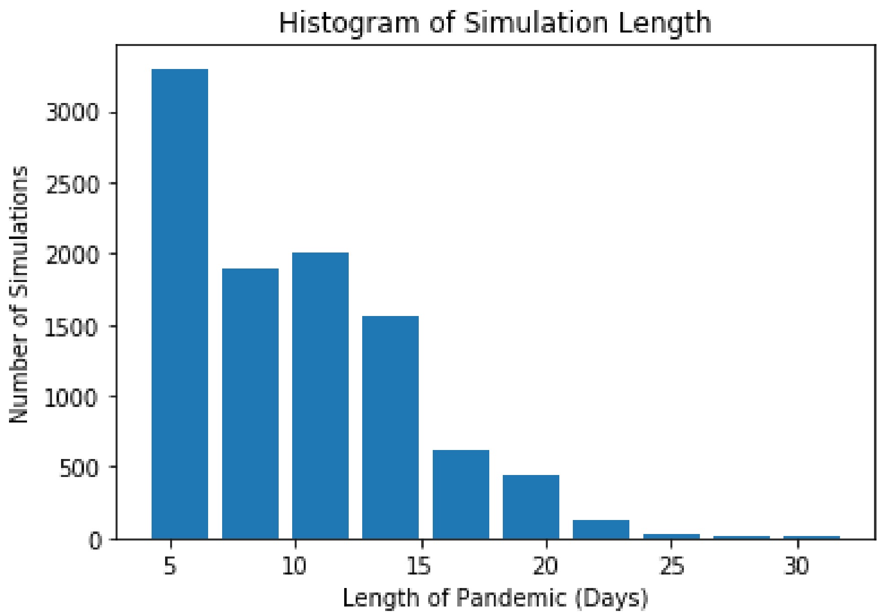

Figure 4: Simulation Length Histogram: 10,000 Iterations

Many typical SimPy simulations imitated long queues and breakdowns, where the simulation records the time and if an event occurs, an interruption is triggered. However, with the flu pandemic simulation, we had to move children from the "healthy" category to the "infected" category, as well as ensure the same child did not get reinfected, and recorded the number each time in order to graph the histogram, which was challenging for us.

3.3.3. EXCEL

Lastly, an Excel spreadsheet was created as a method of validating our simulation methods. The Excel spreadsheet served as a theoretical backbone and no random numbers were used in the Excel spreadsheet. Figure 5: Excel Spreadsheet below shows the Excel spreadsheet utilized by the team.

Day	# Infectous	# Pre-infected	# Total recovered	# Newly recovered	New chance of Infection	# Newly Infected	Total Kids in Class	P value of infection	Newly infected + newly recovered (sense check)
1	1	20	0	0	0.02	1	21	0.02	1
2	2	19	0	0	0.04	1	21	0.02	1
3	3	18	0	0	0.06	2	21	0.02	2
4	4	16	1	1	0.08	2	21	0.02	3
5	5	14	2	1	0.1	2	21	0.02	4
6	6	12	3	1	0.12	2	21	0.02	5
7	6	10	5	2	0.12	2	21	0.02	7
8	6	8	7	2	0.12	1	21	0.02	8
9	5	7	9	2	0.1	1	21	0.02	10
10	4	6	11	2	0.08	1	21	0.02	12
11	3	5	13	2	0.06	1	21	0.02	14
12	3	4	14	1	0.06	1	21	0.02	15
13	3	3	15	1	0.06	1	21	0.02	16
14	3	2	16	1	0.06	1	21	0.02	17
15	3	1	17	1	0.06	1	21	0.02	18
16	3	0	18	1	0.06	0	21	0.02	18
17	2	0	19	1	0.04	0	21	0.02	19
18	1	0	20	1	0.02	0	21	0.02	20
19	0	0	21	1	0	0	21	0.02	21
20	0	0	21	0	0	0	21	0.02	21
21	0	0	21	0	0	0	21	0.02	21
22	0	0	21	0	0	0	21	0.02	21

Number of total kids
21

Figure 5: Excel Spreadsheet

3.4 REAL WORLD APPLICATION

This project has numerous applications to real world problems, more recently related to the Coronavirus pandemic. With the ability to forecast the potential length of the pandemic with and without precautionary methods in place to prevent virus spread, this type of simulation may be one of many effective ways to demonstrate the impact on the length of the pandemic. The complexities of simulating a real-world health-related event brings its own challenges including the uncertainty of recovery as well as the supporting research, which may or may not exist regarding the infection rate and survivability.

4. CONCLUSION

Our approach for this project was to simulate the scenario in ARENA at first. This proved to be difficult given the limitations of the student version of the software and led to our attempt in SimPy with a supporting Excel approach for a basic sense check. This sense check allowed us to confirm the findings from a simulation in order to ensure consistency for this "experiment".

This project was a fascinating application of real-world simulation, especially in a topic so relevant to today's post- or perhaps still mid-pandemic world. We learned the nuances of the SimPy software as well as the challenges of simulating such a real-world topic within software such as ARENA. No matter the simulation software that is being used, the key to success is the importance of "GIGO" and real-world context. Our original attempts for determining the probability of infection provided some odd results when using the conditional probability function. The resulting 100% infection rate was clearly incorrect and enabled us to redirect our approach to determining the new probability rate for infection.

It is also important to note the variability in the results of the simulation. Depending on the number of runs, we saw a different impact to the length of time the “pandemic” would last in the class. This variability is natural in simulation scenarios, and we tried to run 100,000 runs in order to get the best overall results. However, as a result, it is also crucial to understand the variability in the number of days that the pandemic could potentially last. This large number of runs resulted in a particularly large range of days as well.

For future work, it may be interesting to observe impacts to the model if children can become re-infected which is the case in real life. In addition, it would also be interesting to provide a variable recovery rate as it was simply assumed to be 3 days when in fact, it very much depends on the child and their immunity levels and care routine etc. In addition, while the number of students was assumed to be constant throughout the simulation, another interesting approach could consider that a certain number or percentage of students could be absent either before or after getting infected which further could impact the way the pandemic spreads in the classroom.

Overall, this was a very interesting project with real world applications and provided us with a unique challenge for learning simulation techniques.