ISYE 6644 Summer 2023 – Final Project

Pandemic Flu Spread

Shiv Chhabra, Kartikaye Gomber, Adarshkumar Gouda and Melirose Liwag

Group: 123

ABSTRACT

Living through the pandemic over the last few years taught us to keep a distance from others in fear of spreading the virus. Any individual is equally likely to get infected no matter how healthy and is more susceptible if not immune. This project will simulate the pandemic on a much smaller scale while answering the question of whether one infected person is enough to spread a hypothetical virus and how long the epidemic would last. Additionally, we will test the importance of vaccination (immunity to the virus) and its role in stopping the epidemic in a shorter amount of time. This project will be exploring 3 different methods: Chain Binomial, Bernoulli Trials, and the SIR Differential method.

1 Project Overview

1.1 Background

The project will explore a hypothetical scenario of a classroom of 31 children, and supposing Tommy to be our patient zero. To mirror real-life, each of the 30 children is equally likely to get infected by Tommy and is equally as likely to be susceptible to the flu. To start, we will assume that Tommy comes to school every day whether he is sick or not and will be infectious for 3 days starting on Day 1. With this, we will also assume that Tommy will be the only one infectious on Day 1 and the rest of the 30 children are healthy. Additionally, Tommy has a 0.02 probability of infecting other children on each of the 3 days he is infectious. Any kid infected by Tommy from Day 1 onwards will also be infectious for 3 days starting the day after infection. Furthermore, we will assume that each day and kid in the classroom are all independent and identical with a Bernoulli distribution (Bern(p)).

1.2 Approach

We will be generating a random sample of 30 children after each replication on the ARENA simulation program since we are dealing with a hypothetical scenario. Each child will be created at the start of each day and each of the 31 children (including Tommy) will be monitored per replication to test for whether each child is infected, recovered from infection,

or immune (as we will explore later). To simplify the problem, we will also assume that any child who recovers from the infection (after the 3 days of being infectious are up) is no longer susceptible. This is to enable the simulation to run smoother and will enable us to run at least 100,000 replications to produce reasonable analysis.

Another approach we will be exploring is using a regular group of Bernoulli trials. Since we are dealing with 30 children with Bernoulli trials each, the distribution will result in a Binomial distribution of $n = 30$ and $p = 0.02$. Furthermore, the SIR Differential method is also explored in this project where people are divided into Susceptible, Infectious, and Recovered. These variables are then derived to estimate values for the formula as described later in this project.

2 Methodology

2.1 Reed-Frost Model

We can generalize the probability of infection to:

$$I_{t+1} \sim Bin\left(S_t, Q_t\right)$$

This is also known as the Reed-Frost model, belongs to the family of "Chain Binomial" models (quantpie, 2020) and is widely used in pandemic-like cases such as the scenario in this project. Figure 1(b) shows the probability of an individual not getting infected given that there is contact with 5 infected individuals.

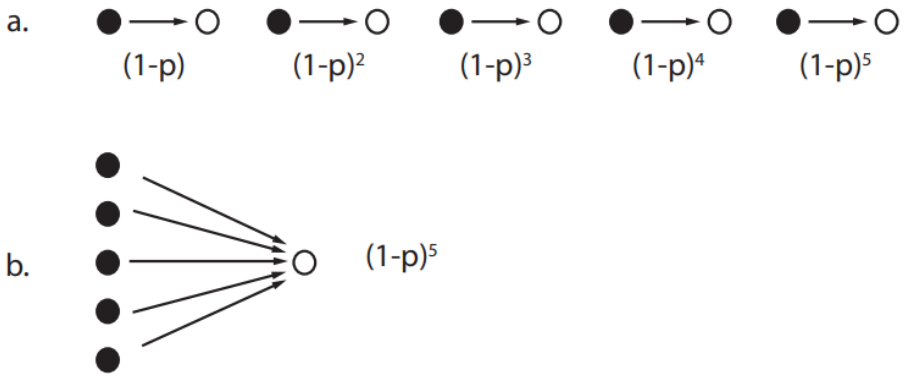

Figure 1: a) The escape probability with five consecutive contacts. b) The escape probability with five simultaneous independent contacts, as in the Reed-Frost model. In both cases, the probability is 1 - (1 – p)^5 (Halloran et al., 2009)

The Reed-Frost model has 3 entity types:

- Infectives: These are the individuals who have been infected and can further spread the disease amongst the susceptible portion of the population. The number of Infectives at time t is given by I_t.
- Susceptibles: This is the portion of the population that is susceptible to infection. The number of children susceptible at time t is given by S_t .
- Recovered: These are the individuals that have recovered from infection. The number of recovered individuals at time t is given by R_t. An important assumption of this model is that individuals that have recovered are no longer considered as susceptible to infection.

As mentioned earlier in Figure 1(b), the probability of an individual not getting infected given I_t infectives is:

$$Q_t = (1-p)^{I_t}$$

Thus, the potential number of infectives at time t+1 will be binomially distributed:

$$Bin \sim \left(S_t,\ 1-Q_t\right)$$

Furthermore, the transition probability of i_{t+1} infectives at time t+1 given $I_t$and $S_t$is

$$Pr\left(I_{t+1}=i_{t+1} \mid S_t, I_t\right) = \begin{pmatrix} S_t \\ i_{t+1} \end{pmatrix} \left(1-Q_t\right)^{i_{t+1}} \left(Q_t\right)^{S_t - i_{t+1}}$$

This approach was used to build the ARENA simulation and the results of the simulation will be compared to these theoretical expected values.

2.2 Bernoulli Trials

In this approach, the event a student is infected by another student is defined by a Bernoulli trial. The number of students that could get infected by any other infected student is a Binomial distribution since we have iid Bernoulli trials with 30 susceptible students with a probability of infection of 0.02. In this model, any infected student on any given day can infect other susceptible in the same day. For example:

```
Day 1: Student_18 was infected by Student_31!!
Day 1: Student_30 was infected by Student_18!!
Day 1: Student_3 was infected by Student_18!!
Day 1: Student_14 was infected by Student_3!!
Day 1: Student_13 was infected by Student_14!!
Day 1: Student_11 was infected by Student_13!!
Day 1: Student_8 was infected by Student_13!!
```

The distribution that Tommy infects other students on Day 1 is given by PMF:

$$\Pr(X=k) = \binom{n}{k} p^k q^{n-k}$$

Where n = 30 and p = 0.02, k = 1, 2, 3, …, n.

This approach is similar to the Reed-Frost model however; it will be using a regular Binomial distribution instead of a chain binomial as mentioned before.

3 Results

3.1 Reed-Frost Model

3.1.1 Distribution of the Data

To start, we needed to calculate the distribution and expected number of children that Tommy would infect on Day 1. Since each day and child is independent and identically distributed with the Bernoulli distribution, the distribution for the entire day would result in the Binomial distribution with 30 individual Bernoulli trials to count for each of the 30 children in the classroom. Let I_i be the number of children infected on day i and so we have the following distribution for Day 1.

$$I_1 \sim Binomial(30,0.02)$$

Suppose that being in contact with an infected individual is counted as a 'success,' the following is the Probability Mass Function (PMF) considering that the probability of infection upon contact is *p* = 0.02

$$f(x) = \begin{cases} p & , x = 1 \\ 1 - p & , x = 0 \end{cases}$$

The deterministic limit of the Reed-Forest Model is found by replacing the random variables (I *and* S) with their expectations:

$$I_{t+1} = S_t(1 - q^{I_t})$$

$$S_{t+1} = S_t(q^{I_t})$$

3.1.2 Expected Number of Infected on day 1

Following the PMF described before, we can also let $q = 1 - p$ be the probability that a child does not get infected and thus, let $Q_{i+1} = q^{I_i}$ be the probability that a child will not get sick on the following day given the number of infected children on Day i. Additionally, we can also calculate the probability of infection with: $1 - q^{I_i}$. We can also let S_i be the number of

those susceptible on Day i . With this, we can calculate the expected number of infected on Day 1 expanding from the Binomial distribution to be

$$E[I_{i+1}] = S_i(1 - q^{I_i})$$

$$E[I_1] = S_0(1 - q^{I_0})$$

$$= 30(1 - (0.98)^1) = 30(0.02)$$

$$= 0.6 \approx 1.0$$

Since it is not realistic to have a fraction of a person, we applied a rounding function to the above expected number. Although, the rounding function will not be applied in the later simulation and results of the expected value calculations the following sections. The rounding in this section is just to apply to real-world scenarios as we consider a small sample size of 30.

3.1.3 Expected Number of Infected on Day 2

Following the formula mentioned above, we can easily calculate the expected number of infected will be on Day 2 being 1.15. Since we cannot have a fraction of a child becoming infected, we will use the rounding function to round up the expected number of infected on Day 1 to 1. Using the deterministic model, we will get:

$$S_1 = S_0(q^{I_0}) = 30(0.98^1) = 29.4 \approx 29$$

$$I_1 = I_0 + Tommy = 2$$

$$E[I_2] = S_1(1 - q^{I_1})$$

$$E[I_2] = 29(1 - .98^2)$$

$$E[I_2] = 1.15 \approx 1$$

As mentioned above, the expected value of the number of kids infected on day 2 is 1.15. However, this was rounded down to 1 to easily apply to a real-world scenario.

The total number of kids infected on days 1 and 2 (excluding Tommy) is:

$$E[I_1] + E[I_2] = 1 + 1 = 2$$

3.2 ARENA Simulation

The Reed Frost Model was implemented in Arena to simulate the spread of the pandemic in the classroom.

3.2.1 Setup

A large value of 100 days was taken as the outer limit of the epidemic's time period. Accordingly, a fake customer was used to change the days from one to the next and to record the number of *Infectives*, *Susceptibles* and *Recovered* at the end of each day.

The 30 students were generated using a *CREATE* module. Using an *ASSIGN* module, each student was assigned with attributes to monitor whether they were sick as well as the number of days they were so. Furthermore, an attribute was also initialized to monitor when they recovered if they fell ill. Next, a *DECIDE* module was used to determine whether a student fell ill using the probability calculated via the Reed Frost equations mentioned earlier:

$$U(0,1) < \left(1 - q^{I_{t-1}}\right)$$

If the student was determined to be sick as per the previous *DECIDE* module, their corresponding 'sick' attribute was set as 1. Furthermore, the counter for the number of sick students was increased by 1 depending on the value of this attribute. After this, the student was delayed for a day and the counters for the number of sick and recovered students were updated depending upon their attributes.

The terminating conditions for the simulation were as follows:

- The number of infected kids becomes 0
- The number of susceptible kids becomes 0

3.2.2 Results and Analysis

The ARENA simulation was run with 100,000 replications and an average of the results taken as the observed number of infections per day. As mentioned before, the simulation is simplified by assuming that any infected child will not be counted as susceptible after the 3 days of being infectious. With this, we observed that the number of infected individuals starts to decrease starting on Day 20 on average. Results of the number of infected children per day including Tommy are shown in Table 1.

Day	Average Number of Infected	Day	Average Number of Infected
1	1.60130	14	3.92207
2	2.52319	15	3.03874
3	3.88184	16	2.31456
4	4.78218	17	1.73656
5	6.20982	18	1.29064
6	7.55702	19	0.94720
7	8.53547	20	0.69236
8	8.85743	21	0.50687
9	8.77751	22	0.36848
10	8.18189	23	0.26935

11	7.20970	24	0.19611
12	6.06803	25	0.14202
13	4.95784	26	0.10328

Table 1: Average number of Infected children per day after 100,000 replications using ARENA

We also calculated the expected number of infections using the Reed-Frost model. Including Tommy, the expected number of infections is shown in Table 2.

Day	Expected Number of Infected	Day	Expected Number of Infected
1	1.6	14	3.857
2	2.535	15	2.951
3	3.876	16	2.196
4	4.317	17	1.601
5	5.459	18	1.148
6	6.475	19	0.813
7	7.143	20	0.571
8	7.702	21	0.397
9	7.818	22	0.275
10	7.49	23	0.189
11	6.833	24	0.129
12	5.904	25	0.088
13	4.868	26	0.059

Table 2: Calculated Expected number of infections per day using the Reed-Frost Model

It is interesting to see that the epidemic starts to disappear as early as Day 20 on average. This makes sense as the number of susceptible children should start to decrease after the 4th Day following those children that Tommy first infected. It is important to note that all of these expected number values are calculated by hand and may observe some truncating between days and may be rounded for easier calculation. Just from looking at Table 1 and Table 2, it is clear to see that the results of our simulation approach the expected number after 100,000 replications.

3.2.2.1 Histograms and Plots

The below histogram shows the duration of the pandemic for all simulation runs. The expected pandemic duration value is about 17 days as can be observed in the histogram. One interesting point of note is that there are around 9,364 runs that end on days 4 and 5 because the number of infected kids becomes 0. A majority of these runs (8,816) ended on day 4 because no other children were infected by Tommy, indicating that the relatively low chance of infection (0.02) combined with the small sample size (30) allowed for an 8% probability that the pandemic would never even start. The remaining 548 runs that ended on day 5 did so because there were no kids infected on days 2, 3 and 4

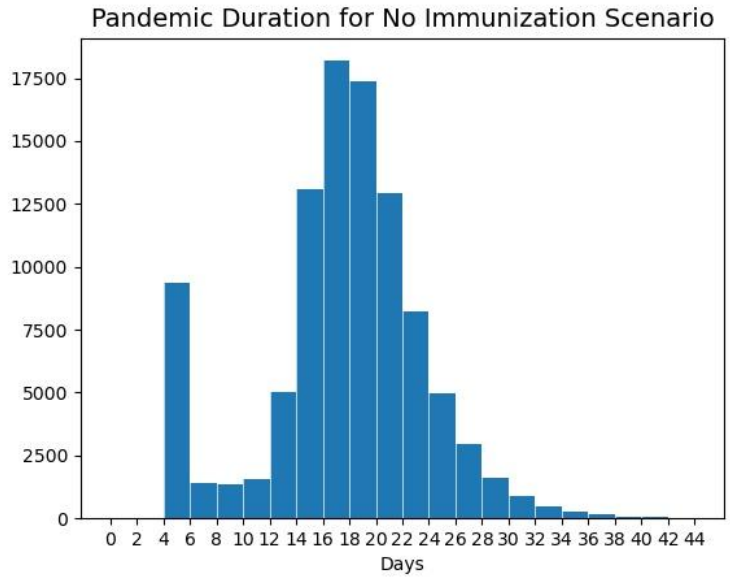

Figure 2(a): Histogram showing the duration of the pandemic for all simulation runs for the 'No Immunization' scenario

The below violin plot shows the spread of the pandemic duration for all 100,000 runs. The sample variance was calculated to be 37.08. Assuming that this data across replications is iid, by the Central Limit theorem the mean pandemic duration will be normally distributed with a standard error of:

$$\sigma^2/n = 37.08/100000 = 0.00037$$

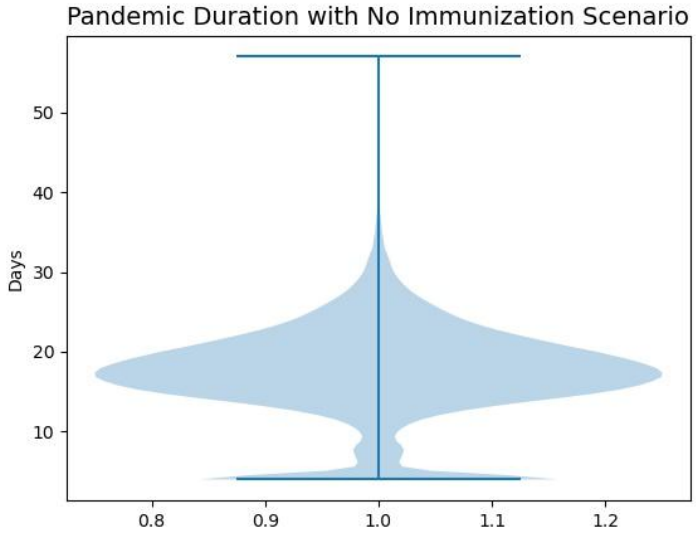

Figure 2(b): Violin plot showing the spread of the duration of the pandemic for all simulation runs for the 'No Immunization' scenario

3.2.2.2 Percentage Error per Day

To validate whether the results from our simulation are reasonable, we will calculate the percentage error of infection per day. This will tell us how far the results of the simulation were from what we expect to see theoretically. The percentage error is shown in Table 3 for each day taking results from previous tables above.

Day	Percentage Error	Day	Percentage Error
1	0.08	14	1.69
2	0.47	15	2.97
3	0.15	16	5.40
4	10.78	17	8.47
5	13.75	18	12.43
6	16.71	19	16.51
7	19.49	20	21.25
8	15.00	21	27.68
9	12.27	22	33.99
10	9.24	23	42.51
11	5.51	24	52.02
12	2.78	25	61.39
13	1.85	26	75.05

Table 3: Percentage error of the infected number of children from the ARENA simulation built.

From Table 3 above, the percentage of error increases following the days in which we expect the epidemic to end. This may be due to the values for both observed and expected values being so small that a difference in them seems large. The equation below is used to calculate the percentage error between the observed value and expected values:

$$\epsilon_i = \frac{(O_i - E_i)}{E_i}$$

Where ϵ_i denotes the error per day through all 100,000 replications. Without taking those last few days into account, it looks like our ARENA simulation produces results very close to what we expect which makes the model reliable for statistical analysis.

3.2.2.3 Confidence Intervals

Following the observations gathered for each replication, we can set a confidence interval for how many infections we can expect to see for each day. The confidence interval can be calculated like so:

$$\overline{X_n} \pm t_{n-1,\frac{1-\alpha}{2}}\sqrt{\frac{S_n^2}{n}}$$

Using an alpha value of $\alpha = 0.05$ we can get a 95% confidence interval for Day 1 and 2 to be [1.5965,1.6061] and [2.5139,2.5325] respectively. With this, we can be 95% confident that the number of infected on Day 1 and Day 2 will fall into these intervals should we run the simulation many other times.

4 Immunization

Now, suppose that all 30 children (excluding Tommy) has a 0.50 probability of being vaccinated and in turn immune to the virus spreading in the classroom. This case should decrease the number of days that the epidemic lasts and should be shorter than the prior scenario. This is because the number of children susceptible to the virus should be less than 30 (and at most half of the classroom) and thus, the epidemic should end in a shorter timeframe.

4.1 ARENA Simulation

4.1.1 Setup

In this simulation, the setup is similar to the previous simulation but with a chance of immunization. Each of the 30 children generated will go through a *DECIDE* module in a '2-way by chance' to receive a 1 if immune and a 0 if not. We will also assume that any child who is immune will not be susceptible to the virus and will thus not get the virus even if they are in contact with any children who are infected. In addition to this assumption, we decided that any children who do have the immune attribute will be disposed in that replication as these children will not be affected by any modules in the simulation.

4.1.2 Results and Analysis

Similar as before, this simulation was run with 100,000 replications and simplified to have infected children be marked as *Recovered* and not susceptible after 3 days. Results are the average number of infected from all 100,000 replications and is shown in Table 3.

Day	Average Number of Infected	Day	Average Number of Infected
1	1.30018	14	0.51130
2	1.67882	15	0.39223
3	2.14521	16	0.29733
4	1.70745	17	0.22550
5	1.81226	18	0.16693
6	1.83404	19	0.12350
7	1.74181	20	0.08983
8	1.50885	21	0.06405

9	1.36263	22	0.04512
10	1.18150	23	0.03152
11	0.98799	24	0.02155
12	0.79984	25	0.01474
13	0.64747	26	0.01031

Table 4: Average number of infected children per day after 100,000 replications using ARENA with a 50% chance of immunization.

Since each child has a 50% chance of being immune, the expected number of susceptible children will decrease from 30 to 15 at the start of each replication. The formula for the start of each replication (Day 1) changes to:

$$E[I_1] = 15(1 - (0.98)^1)$$

$$= 0.3$$

4.1.2.1 Histograms and Plots

The histogram below shows the duration of the pandemic for all simulation runs when 50% of the students are immunized. It is assumed that vaccination provides perfect immunity i.e. the probability of vaccinated students catching the disease is 0%. The expected pandemic duration value is about 10 days as can be observed in the histogram. Note that in around 30% of the cases the simulation terminated at day 4 because there were no infected individuals

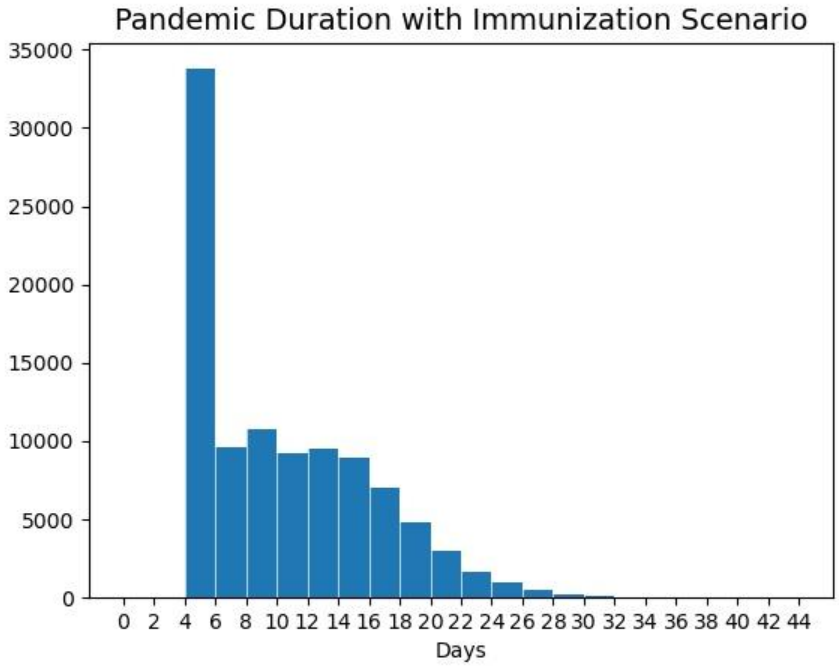

Figure 3(a): Histogram showing the duration of the pandemic for all simulation runs when half the students are immunized

Similar to the previous scenario, the below violin plot shows the spread of the pandemic duration. The variance of the pandemic duration is 32.5. Per the central limit theorem, the mean pandemic duration will be normally distributed with a variance of 0.00032:

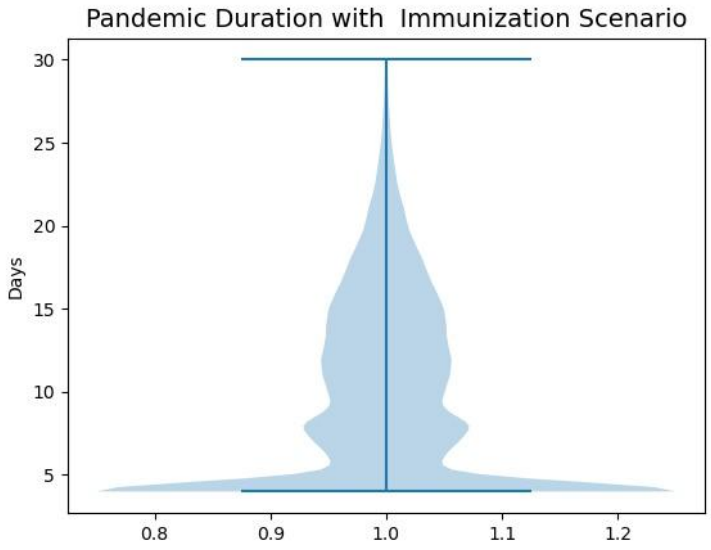

Figure 3(b): Violin plot showing the spread of the duration of the pandemic for all simulation runs when half the population is immunized

4.1.2.2 Confidence Intervals

Similar to before, we will be using the same formula to calculate the confidence intervals for both Day 1 and Day 2 of our model with immunization. Using an alpha value of $\alpha = 0.05$ we get a 95% confidence interval for Day 1 and Day 2 being [1.2968,1.3035] and [1.6731,1.6845] respectively. As mentioned before, this will make us 95% confident that the expected number of infected persons on Day 1 and Day 2 will fall into these intervals should we run the simulation more.

4.2 Without Immunization VS Immunization

With our results from before, it is clear to see that the length of the epidemic is expected to end earlier with immunization than without. Despite the chance that Tommy will never infect anyone in Days 1, 2, or 3, it is less likely for an epidemic to last longer than 20 days with immune individuals than with a group who do not have immune individuals. Additionally, the probability of Tommy failing to infect others increases as the sample of susceptible individuals decreases. It is also interesting to note that the number of replications that terminate on Day 4 shows that there is a possibility for a pandemic to end should the sample

set not come into contact with those who are infected (in this case, it is our patient Zero: Tommy).

5 Bernoulli Trials Method

Following the Bernoulli Trials Method results in the following Binomial Distribution on day one is shown in Figure 4 below. The method uses the scenario in which no child has immunization to the virus.

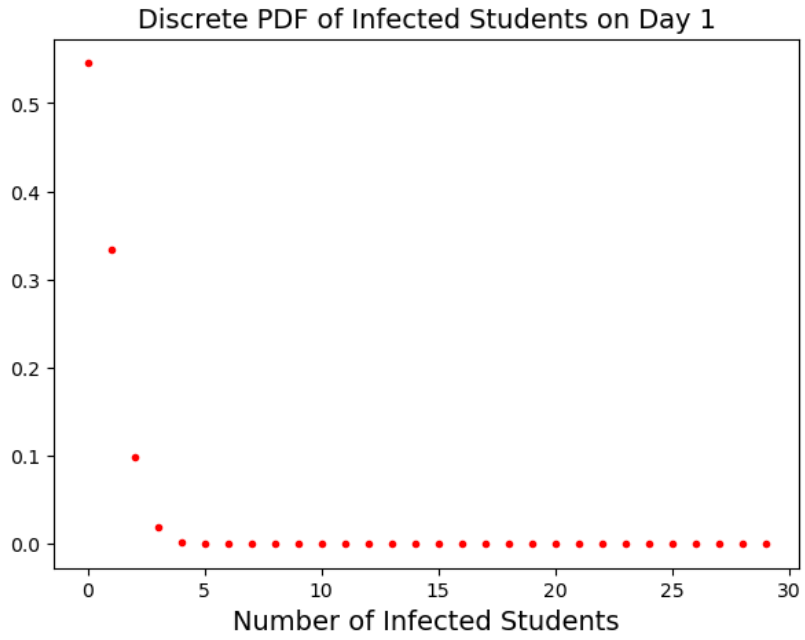

Figure 4: Number of Infected students on Day 1 using Bernoulli Trials

The plot above implies that Tommy can only infect at the max 1 student on Day 1. However, there is higher probability that Tommy would not infect anyone on Day 1. This is similar to the results we found in our simulation where a portion of replications ended around Day 4. Following the PMF of a Binomial distribution, the expected number of infected students can be calculated like so:

$$E[X] = n * p = 30 * 0.02 = 0.6$$

Additionally, to simulate the Bernoulli Trials, Student, Infected Student, Recovered Student and Simulate Classes, the method was constructed in Python. The infected students are generated based on binomial Random Variates and the infected students were picked from the pool of susceptible students by random sampling using the binomial RVs to simulate "infection". A single simulation run results in following plot:

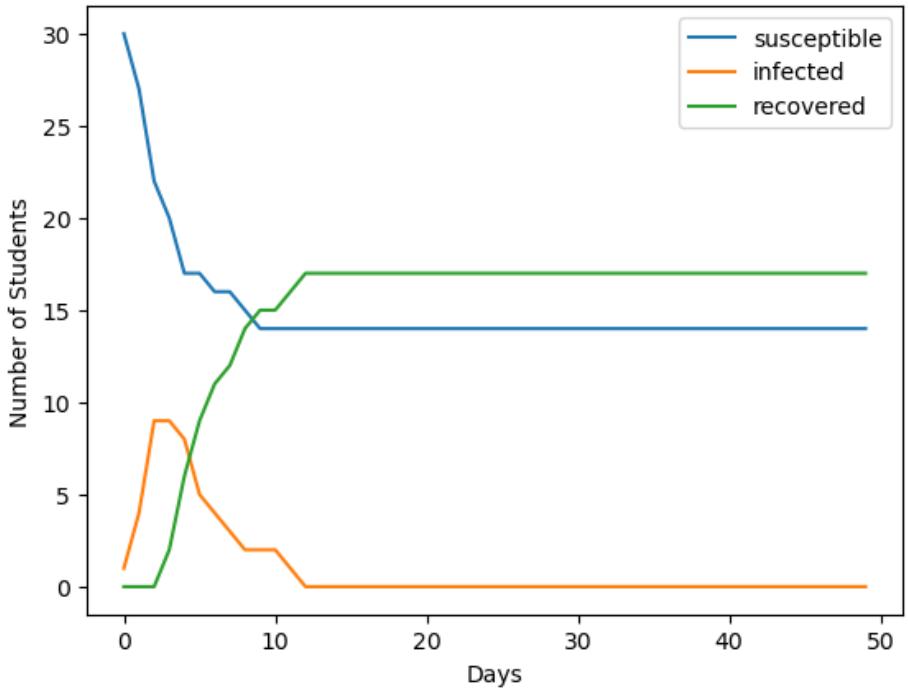

Figure 5: Number of Infected children from a single simulation run from the Bernoulli Trials method

The corresponding daily log is shown and in this run, the simulation ended on Day 12 where the number of infected reached zero as seen in the below:

day	susceptible	infected	recovered
0	30	1	0
1	27	4	0
2	22	9	0
3	20	9	2
4	17	8	6
5	17	5	9
6	16	4	11
7	16	3	12
8	15	2	14
9	14	2	15
10	14	2	15
11	14	1	16
12	14	0	17

Table 5: Daily log for one simulation. Epidemic ends when there are 0 infected left

If we run this simulation for 10,000 trials and get the mean of susceptible (S), infected (I) and recovered (R) for each day, we get the expected values of S I and R for any given day. The corresponding plot is shown below:

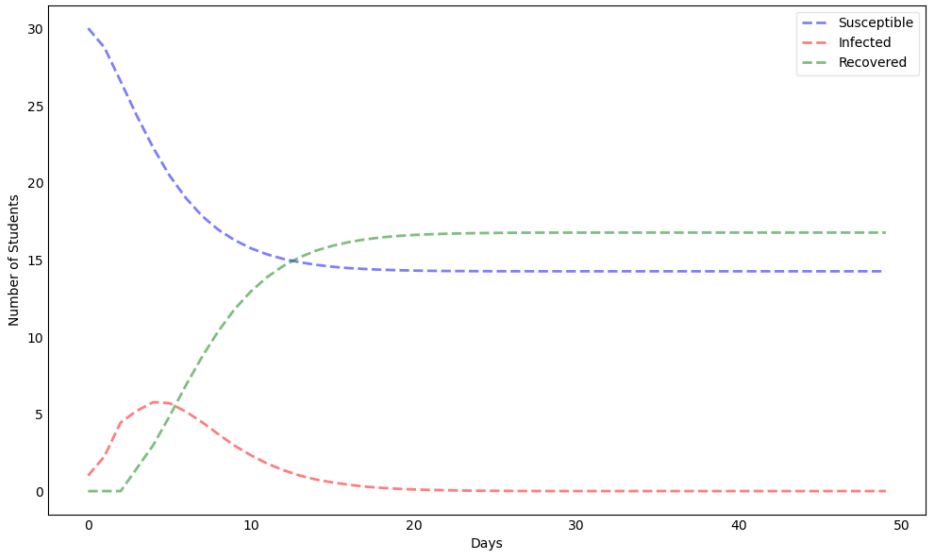

Figure 6: Simulation with 10,000 trials with values of Susceptible, Infected, and Recovered per day

The observed values for 50 days are tabulated below:

Days	Susceptible	Infected	Recovered	Days	Susceptible	Infected	Recovered
0	30.00	1.00	0.00	**25**	14.21	0.02	16.78
1	28.72	2.28	0.00	**26**	14.21	0.01	16.78
2	26.50	4.50	0.00	**27**	14.20	0.01	16.79
3	24.24	5.27	1.49	**28**	14.20	0.00	16.79
4	22.20	5.79	3.02	**29**	14.20	0.00	16.79
5	20.39	5.71	4.90	**30**	14.20	0.00	16.80
6	18.91	5.18	6.90	**31**	14.20	0.00	16.80
7	17.75	4.48	8.77	**32**	14.20	0.00	16.80
8	16.84	3.71	10.45	**33**	14.20	0.00	16.80
9	16.15	2.98	11.87	**34**	14.20	0.00	16.80
10	15.64	2.32	13.04	**35**	14.20	0.00	16.80
11	15.25	1.78	13.97	**36**	14.20	0.00	16.80
12	14.96	1.34	14.70	**37**	14.20	0.00	16.80
13	14.74	1.01	15.25	**38**	14.20	0.00	16.80
14	14.60	0.74	15.67	**39**	14.20	0.00	16.80
15	14.49	0.53	15.99	**40**	14.20	0.00	16.80
16	14.40	0.38	16.22	**41**	14.20	0.00	16.80
17	14.35	0.27	16.38	**42**	14.20	0.00	16.80
18	14.30	0.19	16.50	**43**	14.20	0.00	16.80

19	14.27	0.14	16.59	**44**	14.20	0.00	16.80
20	14.25	0.10	16.65	**45**	14.20	0.00	16.80
21	14.23	0.07	16.69	**46**	14.20	0.00	16.80
22	14.22	0.05	16.72	**47**	14.20	0.00	16.80
23	14.22	0.04	16.75	**48**	14.20	0.00	16.80
24	14.21	0.03	16.76	**49**	14.20	0.00	16.80

Table 6: SIR values from SIR Differential Equation model

From Table 6 above, we can see that the epidemic is expected to end around Day 16 using this model which is much earlier than our simulation than before.

6 The SIR Model – A Differential Equation Model

Infectious disease dynamics in a population can be described using the SIR model. People are divided into three categories: Susceptible, Infectious, and Recovered (with immunity). At its most basic level, SIR describes the number (or proportion) of people in each compartment at each point in time using equations. Flow diagrams depicting the SIR model often show the three states (S, I, and R) and the direction of flow between them.

Figure 7: Flow diagram depicting the SIR states (Bachman et al. 2023)

Whenever a contagious person comes into close contact with a susceptible person, he/she will infect them. The contagious person will come in close contact with an average of β (contact rate) people each day; if they are susceptible (i.e., not immune), they will get infected. On day t, if $S_t$susceptible people are in close contact with our carrier, the expected number of infected people is $\frac{\beta S_t}{N}$, where $\frac{S_t}{N}$ is the probability of close contact with a susceptible person. The mean recovery rate is given by γ . And $\frac{1}{\gamma}$ is the mean period of time during which as infected individual (carrier) can pass it on. The differential equations were first developed by Kermack and McKendrick and are as follows:

$$\frac{dS}{dt} = -\frac{\beta SI}{N},$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I,$$

$$\frac{dR}{dt} = \gamma I$$

The transition rate is calculated as below:

$$\beta = P(transmission) \times No.of\ Susceptible$$

$$\beta = 0.02 \times 30 = 0.6$$

And recovery rate is calculated as:

$$\gamma = \frac{1}{days\ to\ recover} = \frac{1}{3} = 0.33$$

6.1 Results

The SIR derivatives for a set of ordinary differential equations which are integrated over t to arrive at the S, I and R values. The integration of these equations is implemented in python using Scipy.integrate.odeint library over 100 days. The results of said method is depicted in Figure 8 below and is using the scenario in which no children has immunization to the virus.

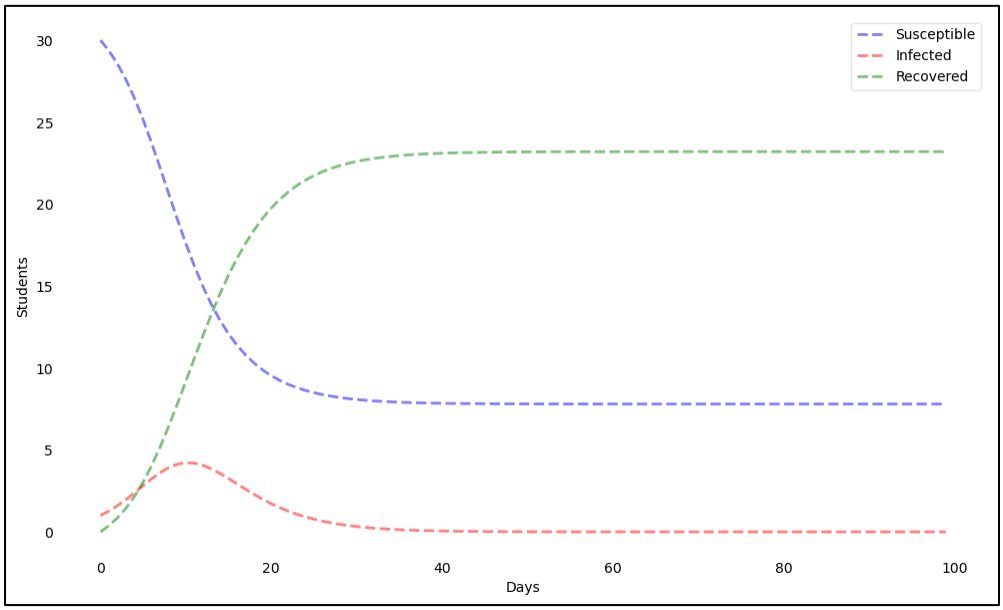

Figure 8: Plot for distribution with transition rate = 0.4 and recovery time of 3 days

Additionally, the numerical results of the method is reported in Table 7 for value of Susceptible, Infected, and Recovered.

Days	Susceptible	Infected	Recovered	Days	Susceptible	Infected	Recovered
0	30.00	1.00	0.00	**25**	8.49	0.77	21.74
1	29.35	1.27	0.38	**26**	8.37	0.65	21.98
2	28.55	1.60	0.85	**27**	8.28	0.55	22.17
3	27.58	1.97	1.45	**28**	8.20	0.46	22.34
4	26.44	2.38	2.17	**29**	8.13	0.39	22.48
5	25.15	2.81	3.04	**30**	8.08	0.32	22.60
6	23.72	3.24	4.05	**31**	8.03	0.27	22.70
7	22.19	3.62	5.19	**32**	7.99	0.23	22.78
8	20.63	3.92	6.45	**33**	7.96	0.19	22.85
9	19.08	4.13	7.79	**34**	7.93	0.16	22.91
10	17.60	4.22	9.19	**35**	7.91	0.13	22.96
11	16.22	4.19	10.59	**36**	7.89	0.11	23.00
12	14.97	4.06	11.97	**37**	7.88	0.09	23.03
13	13.87	3.84	13.29	**38**	7.86	0.08	23.06
14	12.91	3.57	14.52	**39**	7.85	0.06	23.08
15	12.08	3.26	15.66	**40**	7.84	0.05	23.10
16	11.38	2.93	16.69	**41**	7.84	0.04	23.12
17	10.79	2.60	17.61	**42**	7.83	0.04	23.13
18	10.29	2.28	18.43	**43**	7.83	0.03	23.14
19	9.87	1.99	19.14	**44**	7.82	0.03	23.15
20	9.53	1.72	19.75	**45**	7.82	0.02	23.16
21	9.24	1.48	20.29	**46**	7.81	0.02	23.17
22	9.00	1.26	20.74	**47**	7.81	0.01	23.17
23	8.80	1.07	21.13	**48**	7.81	0.01	23.18
24	8.63	0.91	21.46	**49**	7.81	0.01	23.18

Table 7: SIR values from SIR Differential Equation model

7 Model Comparison

In comparing Table 3, 4 and 5, it can be noted that the SIR differential equation model, Bernoulli Trial Model and the Chain Binomial Model produce considerably different results. Although, the Bernoulli Trials method produces a result quite close to the expected values of the Chain binomial method. This should make sense since the Chain Binomial method is just an extension of a group of Bernoulli trials. Additionally, we can observe the expected end of the epidemic without immunization as derived from each method as follows:

	Chain Binomial	Bernoulli Trials	SIR Differential
End of Epidemic (Day Interval)	[22,27]	[16,22]	[28,40]

Table 8: Model intervals of expected end of the epidemic from the 3 models explored

Taking the decimal nature of the results into account, we will determine the upper and lower bounds of the expected end of the epidemic by rounding the number of infected (i.e. $0.1 < x \leq 0.5$). Overall, following the Bernoulli Trials method will result in the pandemic ending

a little earlier than the other 2 methods. Interestingly, the SIR differential model results in the expected end to happen on the upper bound of the Chain Binomial as does the Chain Binomial to the upper bound of the Bernoulli Trials method.

8 Conclusion

In this small-scale pandemic simulation, we found that an epidemic will last for about 20 days on average given that patient Zero (Tommy) has a probability of infecting others by Bern(0.02). This is also assuming that each infected person is infectious for 3 days, are independent but identically distributed, and that any infected persons will be counted as "Recovered" after the third day and is not susceptible anymore. Furthermore, we found that although the probability for each child is independent, each day is dependent on the outcome of the day before. This led us to pursue the Reed-Frost method to build the basis of our simulation. Additionally, we explored other methods such as the Bernoulli Trials method and the SIR Differential method.

Vaccination reduced the average duration of the pandemic to 10 days versus 20 days for the base case. This was because the susceptible population was reduced by half due to immunization. This shows the importance of immunization, and that increased immunization will decrease the length of a pandemic by a wide margin.

Furthermore, in 35% of the runs from the 'Immunized' scenario, the pandemic didn't spread at all from the original infected (Tommy) versus only 8% in the base case. However, it is important to note that the small population sample (30 students) size had a large role to play in these results as it limited the number of potential contacts/chances for the infection to spread. It is highly likely that the pandemic would have spread to more people from Tommy given a larger sample size, even with immunization. This was observed in real world scenarios during the Covid epidemic during which high (~75%) vaccination rates in certain European countries weren't enough to end the pandemic even when the vaccinations had a high efficacy rate against symptomatic infection. The large population allowed the virus to spread even though the probability of getting infected was dramatically reduced through vaccination.

8.1 Future Work

The above scenario may be observed should we have counted "recovered" children as susceptible again after 3 days of being infected. As mentioned before, this assumption was put in place to simplify our simulation but it still produces reasonable results. Should we have more time to run our simulation, we would consider this case as well and run it through 100,000 replications to get results closer to a real-world scenario. Additionally, we can expand our work and build more models than what we have currently. Some additional analysis and model comparison can be made should we create more models can also be further explored in the future.

9 References

Bachmann, R., Topa, G., & Klaauw, W. van der. (2023). Handbook of Economic Expectations. Academic Press.

Halloran, M. E., Longini, I. M., & Struchiner, C. J. (2009). *Design and Analysis of Vaccine Studies*.

quantipie. (2020, August 8). *Chain Binomial Method for COVID-19 Spread* [Video]. YouTube. https://www.youtube.com/watch?v=J-3xLLG6yOI&ab_channel=quantpie