Modelling COVID-19 Using the Logistic Function

Contents

Introduction ……………………………………………………………………

While pondering the disappointing reality that IB exams would most likely be held at the end of my senior year of high school, I reflected upon the early stages of the COVID-19 pandemic to contemplate how the virus spread over the years. During early 2020, the pandemic was regarded by many as a distant problem, an epidemic that was not as pressing as other events occurring globally. Following suit with this mindset, I lived ignorantly, thinking nothing of the incoming complications. However, after case numbers started climbing exponentially, my daily life was drastically impacted. This sudden change shifted my attitude towards learning more about the pandemic. Although I lacked a proper understanding of the mechanisms of a virus spreading, I knew one important thing. Surely, case numbers could not exponentially increase forever, right?

Recently, after learning about probability in class, I wanted to discover how deeply this concept could be applied to real life situations. By knowing the probability of an event occurring and the number of trials taking place, I could determine the expected number of successful events in simple actions such as rolling a die or flipping a coin. However, I wanted to further extend my knowledge by using probability as a basis to explore patterns in the spreading of COVID-19. Unfortunately, the simple probability models learned in class were not complex enough to accurately model the case counts of a virus due to the existence of countless external factors. Luckily, a more complex model, the logistic model, can be derived using concepts learned in class and used to model the growth of a virus in a population. Thus, the aim of this investigation is to determine whether the logistic model accurately fits COVID-19 spreading patterns through the use of existing data and other basic models.

Background ……………………………………………………………………………

Before starting this investigation, an understanding of the logistic model must be established.

Exponential model explained

In a simple probability model, the expectation of the number of new COVID-19 infections from one person can be represented by the relationship:

$$\Delta N = np$$

ΔN is the expected number of new infections in a day, n is the number of people exposed to the infected individual, and p is the probability of an exposed person getting infected. However, this equation only accounts for 1 person being infected. Due to the nature of viruses, the more people are infected, the faster they will spread. Thus, to account for a larger population, the equation can be altered:

$$\Delta N_x = npN_x$$

In this new equation, N_x represents the number of people already infected on a given day and x represents the day number. Evidently, the equation accounts for N interactions between infected individuals and other people rather than just 1 interaction. It can be seen that the number of existing cases (N_x), is an important factor in the amount of new cases per day (ΔN_x). From this, the number of total cases in the subsequent day (N_{x+1}) can be found to be:

$$N_{x+1} = npN_x + N_x$$

Factoring out N_x gives: $N_{x+1} = N_x(1 + np)$

To visualise this relationship, an example scenario can be used.

An infectious disease of fatigue has spread in IB Land. Assuming 100 people in IB Land are infected initially, the number of interactions per person is 5 people, and the probability of infection is 10%, what is the case count after a week?

$$n = 5, p = 0.1, \text{ and } N_0 = 100.$$

$$\Delta N_1 = (5)(0.1)(100) = 50$$

Therefore, the number of new cases between day 1 and day 2 is 50.

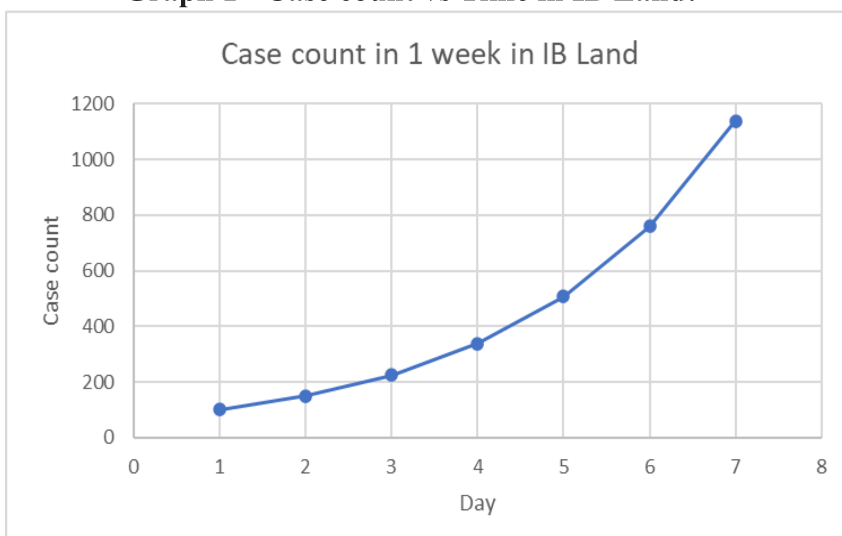
Finding the number of cases in the next day:

$$N_2 = (100)[(1 + (5)(0.1)] = 150$$

Table 1 - Cases for 1 week in IB Land:

Day	Case Count
1	100
2	150
3	225
4	338
5	506
6	759
7	1139

Graph 1 - Case count vs Time in IB Land:



As seen in **Table 1**, the number of cases from the previous day is multiplied by a constant factor of $(1 + np)$, in this case, by 1.5, for each subsequent day. The exponential function in **Graph 1** can be expressed using the following equation:

$$N_u = N_0(1 + np)^u$$, where *u* represents the number of days after the first.

In the early stages of a pandemic, the spread may look similar to **Graph 1**. Due to changes in factors such as lockdown measures and vaccinations (factors which decrease *n* and *p*), the case numbers will still increase exponentially, just at a lower rate.

Deriving the logistic model

However, the major flaw with this exponential model is that it does not account for population limits. Assuming this model, the pandemic will eventually reach an infinitely widespread state. However, if the pandemic were to ever reach such a state in real life, an individual would interact with people who are already infected, not increasing the case count. Therefore, in order to account for a limit of infections, the logistic differential equation can be used:

$$rate\ of\ change\ of\ infection = \frac{dN}{dt} = pN(1 - \frac{N}{K})$$

t = time

p = probability of infection

N = number of cases at time *t*

K = total population size

Just by looking at the equation, it is evident that as *N* approaches 0, $\frac{dN}{dt} = pN$. This is identical to the exponential model explained above. However, as *N* approaches *K*, $(1 - \frac{N}{K})$ approaches a value of 0, and thus $\frac{dN}{dt} = 0$. This means that eventually, the case count will reach a plateau. In order to derive a function of infections in terms of time, $P(t)$, the logistic differential equation must be solved.

Solving the logistic differential equation:

$\frac{dN}{dt} = pN(1 - \frac{N}{K})$ can be rearranged as such:

$$\frac{dN}{dt} = pN - \frac{pN(N)}{K}$$

$$\frac{dN}{dt} = \frac{pNK}{K} - \frac{pN(N)}{K}$$

$$\frac{dN}{dt} = \frac{pNK - pN(N)}{K}$$

$$\frac{dN}{dt} = \frac{pN(K-N)}{K}$$

This is a separable differential equation that can be expressed in the following way:

$$\frac{dN}{N(K-N)} = \frac{p}{K}\, dt$$

$$\frac{K}{N(K-N)} dN = p\, dt$$

$\frac{K}{N(K-N)}$ can be simplified using partial fractions:

$$\frac{K}{N(K-N)} = \frac{A1}{N} + \frac{A2}{K-N}$$

$$K = A1(K - N) + A2(N)$$

To solve for $A1$ and $A2$, look at the equation. $A1(-N)$ must cancel out with $A2(N)$, so $A1$ must equal $A2$. However, since $A1$ is also multiplied by K, and the left side is K, there is 1 solution:

$$A1 = 1,\ A2 = 1$$

$$K = 1(K - N) + 1(N)$$

$$K = K - N + N$$

$$K = K \checkmark$$

Therefore, $\frac{K}{N(K-N)} = \frac{1}{N} + \frac{1}{K-N}$

Integrating both sides: $\int(\frac{1}{N} + \frac{1}{K-N})dN = \int p\,dt$

$$\int \frac{1}{N} dN + \int \frac{1}{K-N} dN = pt + C$$

$$ln|N| - ln|K - N| = pt + C$$

Multiplying both sides by (-1): $-(ln|N| - ln|K - N|) = -(pt + C)$

$$ln|K - N| - ln|N| = -pt - C$$

$$ln\left|\frac{K-N}{N}\right| = -pt - C$$

Raising e to the power of both sides: $e^{ln\left|\frac{K-N}{N}\right|} = e^{-pt-C}$

$$\left|\frac{K-N}{N}\right| = e^{-pt-C}$$

$\frac{K-N}{N} = Ae^{-pt}$, where $A = \pm e^{-C}$

Rearranging the equation: $K = NAe^{-pt} + N$

$$K = N(Ae^{-pt} + 1)$$

$$N = \frac{K}{Ae^{-pt}+1}$$

To solve for new A, substitute $t = 0$ and $N = N_0$ into $\frac{K-N}{N} = Ae^{-pt}$

$$\frac{K-N_0}{N_0} = A(1)$$

$$A = \frac{K-N_0}{N_0}$$

FINALLY, the resulting equation is:

$N(t) = \frac{K}{Ae^{-pt}+1}$, where $A = \frac{K-N_0}{N_0}$

This formula is the logistic function, where the number of infections is expressed as a function of time. N is the number of cases, t is time, K is the maximum population, p is the probability of infection, and N_0 is the initial number of cases. It will be the main function used to attempt to fit COVID-19 spreading patterns.

COVID-19 case count data collection ....

In order to accurately model COVID-19 using existing data, the infection rates after the Omicron variant was introduced will be used. As a result of the beginning of a new variant, there is a significant spike in case counts. Thus, if data from before Omicron were to be used, there would be an irregularity in the trend at the time when Omicron began. Shown below in **Table 2** below are the case numbers from the first Omicron date until March 19th 2022 in Canada.

Table 2 - Canada COVID-19 cases since the arrival of the Omicron variant:

Date (dd-mm-yyyy)	Daily case count	Cumulative case count since the beginning of the pandemic	Cumulative change in cases beginning from 29-11-2021
29-11-2021	4920	1797072	0
30-11-2021	2409	1799481	2409
01-12-2021	3252	1802733	5661
02-12-2021	3274	1806007	8935
03-12-2021	3564	1809571	12499
04-12-2021	1997	1811568	14496
05-12-2021	1803	1813371	16299
06-12-2021	5986	1819357	22285
07-12-2021	2984	1822341	25269
08-12-2021	3801	1826142	29070
09-12-2021	4193	1830335	33263
10-12-2021	3097	1833432	36360

11-12-2021	4707	1838139	41067
12-12-2021	2453	1840592	43520
13-12-2021	8034	1848626	51554
14-12-2021	4499	1853125	56053
15-12-2021	6253	1859378	62306
16-12-2021	7611	1866989	69917
17-12-2021	9265	1876254	79182
18-12-2021	5445	1881699	84627
19-12-2021	6026	1887725	90653
20-12-2021	19350	1907075	110003
21-12-2021	12252	1919327	122255
22-12-2021	15684	1935011	137939
23-12-2021	22620	1957631	160559
24-12-2021	26234	1983865	186793
25-12-2021	16082	1999947	202875
26-12-2021	13666	2013613	216541
27-12-2021	25073	2038686	241614
28-12-2021	23768	2062454	265382
29-12-2021	54725	2117179	320107
30-12-2021	41698	2158877	361805
31-12-2021	46363	2205240	408168
01-01-2022	24350	2229590	432518
02-01-2022	23817	2253407	456335
03-01-2022	56181	2309588	512516
04-01-2022	45527	2355115	558043
05-01-2022	40111	2395226	598154
06-01-2022	42069	2437295	640223
07-01-2022	44405	2481700	684628
08-01-2022	23225	2504925	707853
09-01-2022	21536	2526461	729389
10-01-2022	72364	2598825	801753
11-01-2022	29361	2628186	831114
12-01-2022	35685	2663871	866799
13-01-2022	31605	2695476	898404
14-01-2022	40605	2736081	939009
15-01-2022	12934	2749015	951943
16-01-2022	11563	2760578	963506

17-01-2022	54742	2815320	1018248
18-01-2022	20328	2835648	1038576
19-01-2022	24040	2859688	1062616
20-01-2022	22890	2882578	1085506
21-01-2022	22675	2905253	1108181
22-01-2022	10731	2915984	1118912
23-01-2022	10007	2925991	1128919
24-01-2022	24058	2950049	1152977
25-01-2022	24347	2974396	1177324
26-01-2022	19186	2993582	1196510
27-01-2022	18105	3011687	1214615
28-01-2022	17712	3029399	1232327
29-01-2022	7467	3036866	1239794
30-01-2022	6278	3043144	1246072
31-01-2022	24751	3067895	1270823
01-02-2022	11271	3079166	1282094
02-02-2022	15492	3094658	1297586
03-02-2022	13937	3108595	1311523
04-02-2022	13963	3122558	1325486
05-02-2022	5576	3128134	1331062
06-02-2022	4579	3132713	1335641
07-02-2022	18804	3151517	1354445
08-02-2022	9068	3160585	1363513
09-02-2022	11510	3172095	1375023
10-02-2022	10158	3182253	1385181
11-02-2022	8084	3190337	1393265
12-02-2022	6704	3197041	1399969
13-02-2022	5777	3202818	1405746
14-02-2022	14083	3216901	1419829
15-02-2022	6305	3223206	1426134
16-02-2022	8326	3231532	1434460
17-02-2022	7239	3238771	1441699
18-02-2022	7436	3246207	1449135
19-02-2022	2850	3249057	1451985
20-02-2022	3585	3252642	1455570
21-02-2022	2700	3255342	1458270
22-02-2022	9889	3265231	1468159

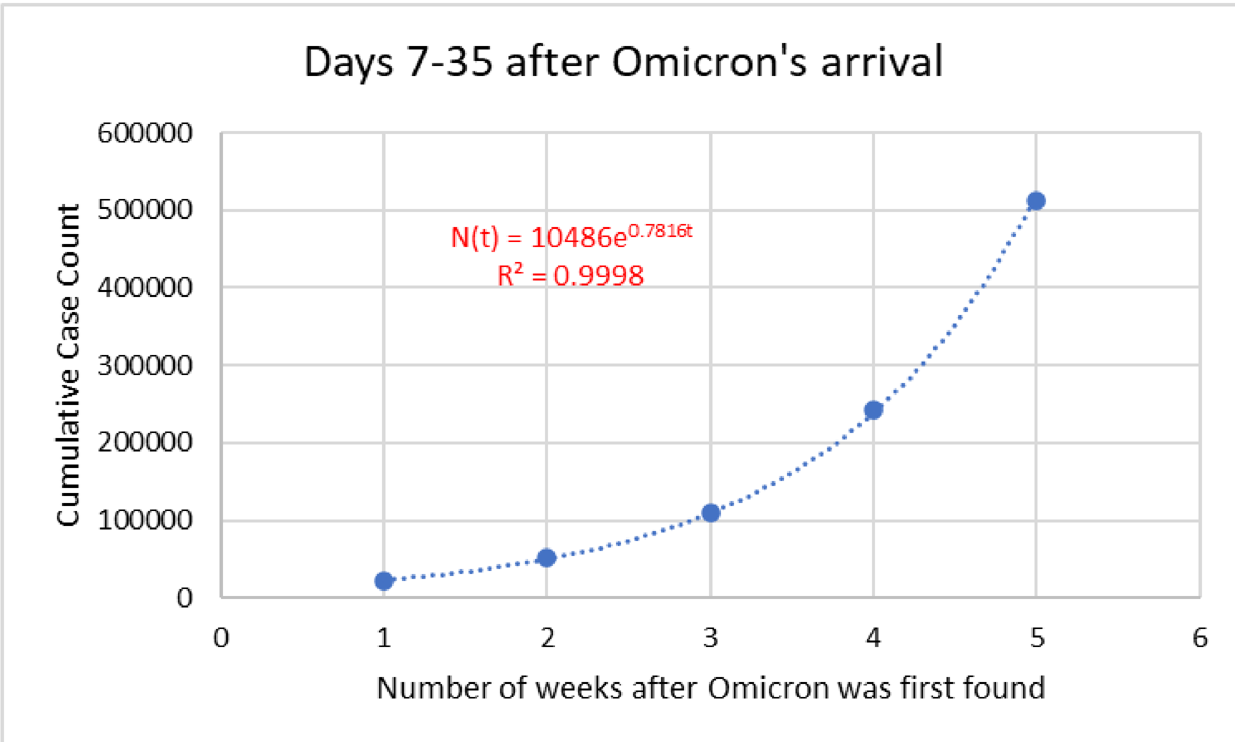
23-02-2022	10243	3275474	1478402
24-02-2022	5837	3281311	1484239
25-02-2022	7924	3289235	1492163
26-02-2022	1017	3290252	1493180
27-02-2022	4159	3294411	1497339
28-02-2022	11139	3305550	1508478
01-03-2022	3321	3308871	1511799
02-03-2022	8087	3316958	1519886
03-03-2022	5179	3322137	1525065
04-03-2022	8922	3331059	1533987
05-03-2022	1767	3332826	1535754
06-03-2022	2321	3335147	1538075
07-03-2022	9433	3344580	1547508
08-03-2022	5540	3350120	1553048
09-03-2022	6968	3357088	1560016
10-03-2022	5910	3362998	1565926
11-03-2022	7292	3370290	1573218
12-03-2022	2099	3372389	1575317
13-03-2022	2479	3374868	1577796
14-03-2022	8295	3383163	1586091
15-03-2022	4582	3387745	1590673
16-03-2022	6509	3394254	1597182
17-03-2022	6629	3400883	1603811
18-03-2022	7088	3407971	1610899
19-03-2022	1624	3409595	1612523

Data analysis ……………………………………………………………………

Plotting initial case counts

The data points in the fourth column of **Table 2** can be plotted. Because of a short incubation period of the virus, the data will be plotted starting 7 days after the Omicron variant was found in Ontario. **Graph 2** below illustrates 4 weeks of plotted data from days 7 to 35 in intervals of 7 days. Daily data is not preferable, as irregular situations often cause outliers in case counts. Weekly data more accurately portrays the general trend in spreading of the virus.

Graph 2 - Cumulative change in COVID-19 case counts in Canada after Omicron Variant



As seen in **Graph 2**, the initial 1-5 weeks of COVID-19 data can be modelled extremely closely using an exponential curve with the equation:

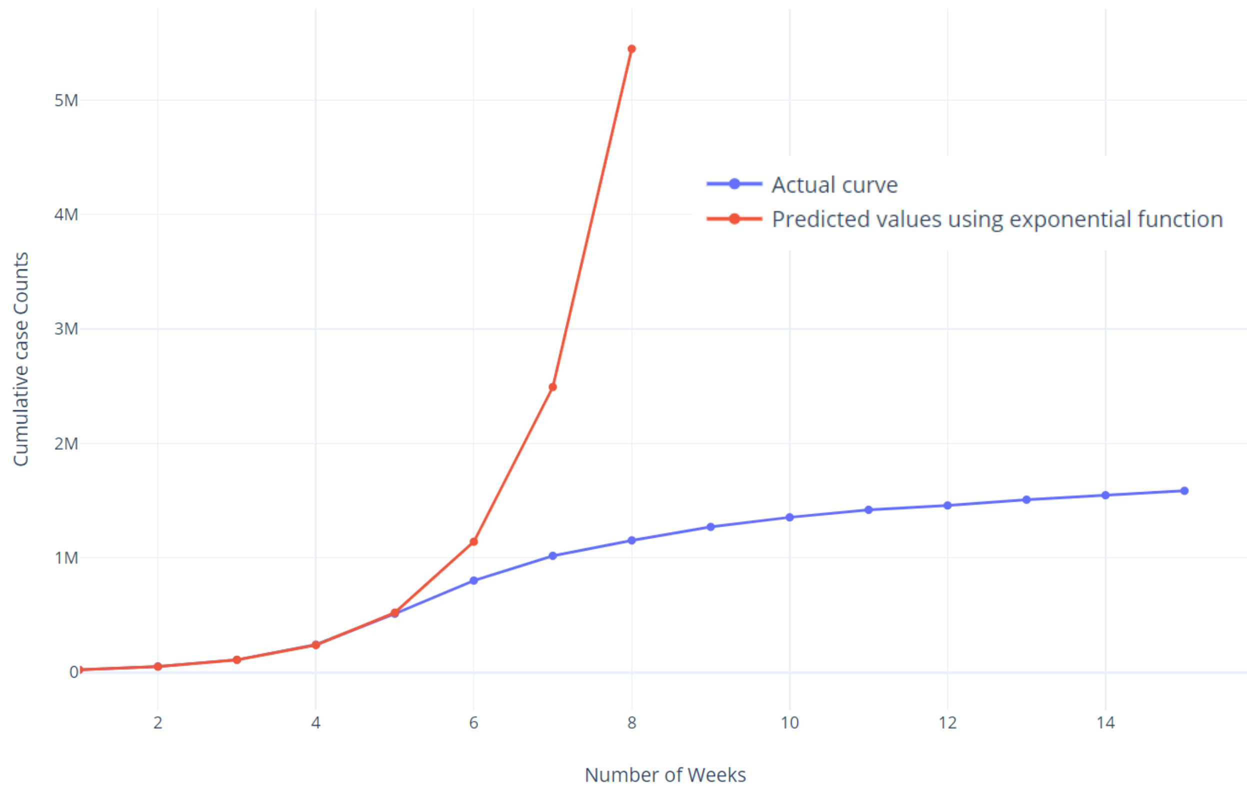
$$N(t) = 10486e^{0.7816t}$$

The coefficient of determination (R squared) is approximately 0.9998, signifying a nearly 100% accurate relationship between the exponential function and data points. However, as mentioned previously, this exponential model can only be used in the early stages of a virus spread.

Exponential model predictions and flaws

Using the exponential equation from **Graph 2**, more data points can be predicted by substituting additional values of t. Shown below in **Graph 3** is the exponential function extended to 8 weeks, compared with the actual values from **Table 2**.

Graph 3: Predicted values from exponential function vs. Actual values



Clearly, the exponential model becomes inaccurate as time passes. This is to say that once the virus starts spreading at a slower rate, a new model must be adapted. Therefore, the logistic function derived earlier can now be employed.

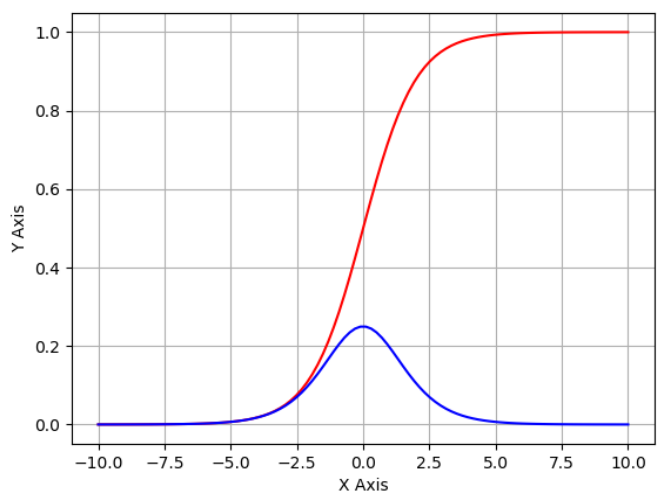
Determining parameters

In order to visualise the logistic function using data, a few parameters must first be defined; most notably, the inflection point. The essential difference between the exponential model and the logistic model is the inflection point, which can be used to deduce other variables in the logistic function. A simple model is shown below:

Inflection point

Graph 4 - Basic logistic function given by

$f(x) = \frac{1}{1+e^{-x}}$ **and its derivative:**



Graph 4 shows a basic logistic function (in red), where the inflection point takes place at (0, 0.5). When considering the slope of the tangent at each point (in blue), the maximum also occurs when x = 0. Therefore, the inflection point takes place when the rate of change transfers from increasing to decreasing.

How can this be quantified using data? To determine the inflection point without an existing function to find the derivative of, the growth factor method is used. The growth factor can be calculated by $\frac{\text{number of new cases in 1 day}}{\text{number of new cases in previous day}}$.

In the context of COVID-19 case counts, the significance of the inflection point is as follows:

$t <$ inflection point	Before the inflection point, the number of new cases per day increases exponentially.

	Growth factor is > 1
t = inflection point	At the inflection point, the number of new cases per day is equal to the previous day. Growth factor = 1
t > inflection point	Once the inflection point is passed, the number of new cases per day decreases. Growth factor < 1

The data in **Table 2** can be used to calculate the growth factor using the formula derived above.

Once again, the data points will be taken in intervals of 7.

Table 3 - Growth factor for each week:

Week number	New cases (not cumulative)	Growth factor
1	22285	N/A
2	29269	1.313395
3	58449	1.996959
4	131611	2.251724
5	270902	2.058354
6	289237	1.067681
7	216495	0.748504
8	134729	0.622319
9	117846	0.874689
10	83622	0.709587
11	65384	0.7819
12	38441	0.587927

13	50208	1.306105
14	39030	0.777366
15	38583	0.988547

From **Table 3**, the growth factor is consistently above 1 from weeks 2 to 6. It then drops consistently below 1 after week 6 (There are outliers such as Week 13. However, the consistent trend is < 1). Using the data table, the inflection point can be estimated to be at approximately Week 6, as it is the closest value to 1.

<u>Maximum Number of Infections (K):</u>

In the logistic model, the maximum amount of infections is double that at the inflection point. Since the case count at the end of Week 6 was 2598825 cases, the maximum number of infections (K) is double this number.

$$K = 2598825 \times 2 = 5197650\, cases$$

However, the data points begin at 1797072. Since a logistic function is symmetrical about the inflection point, this original value must be subtracted from K.

$$K = 3400578\, cases$$

Logistic fitting …………………………………………………………………

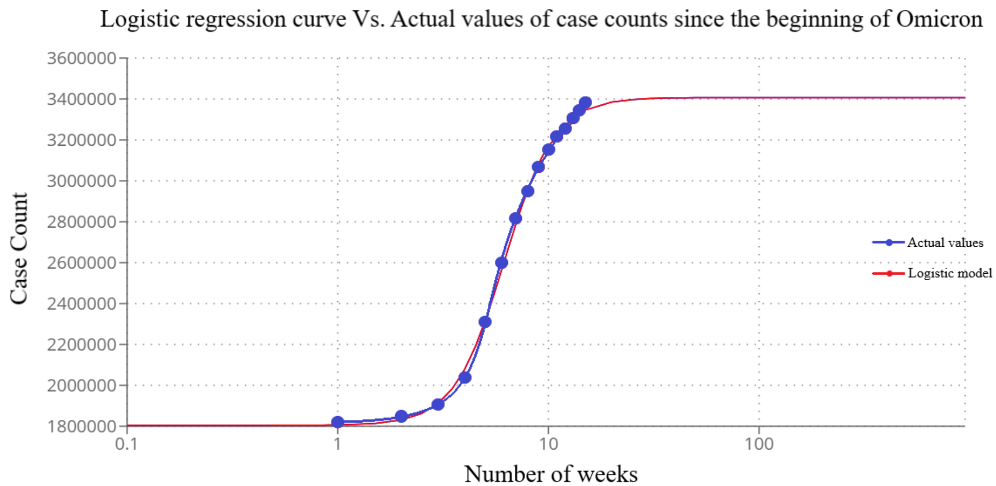
Ideal logistics variables

Referring back to the logistic equation, $N(t) = \frac{K}{Ae^{-pt}+1}$ where $A = \frac{K-N_0}{N_0}$, there are many variables that are necessary to define in order to graph the function. The *K* value is already known and the *A* value can be calculated easily. However, there is no reliable and consistent way to determine e^{-p}. Evident from the presented data, the probability of an individual being infected is sporadic. If one value of *p* were to be used, the trend would not correspond to each data point. If an average value of *p* were to be used, the result would be similar. Thus, the most accurate way to graph the logistic function is by using the two parameters that were previously solved for; namely, *K* and inflection point.

Graphing predictions

The logistic function can be fit using existing data, where the point of inflection takes place around Week 6. Because the *K* value of the function is 3400578, there is a horizontal asymptote at $y = 3400578$. To fit the curve, the initial case count $N_0 = 1797072$ must be the lower bound of the range of the function. These aspects are shown in **Graph 4** below.

Graph 4 - Fitting logistic curve to actual data



The red curve in **Graph 4** has the equation in the form of a slightly modified form of the logistic function that accounts for the initial case count, the maximum case count, and the slope of the function at the inflection point.

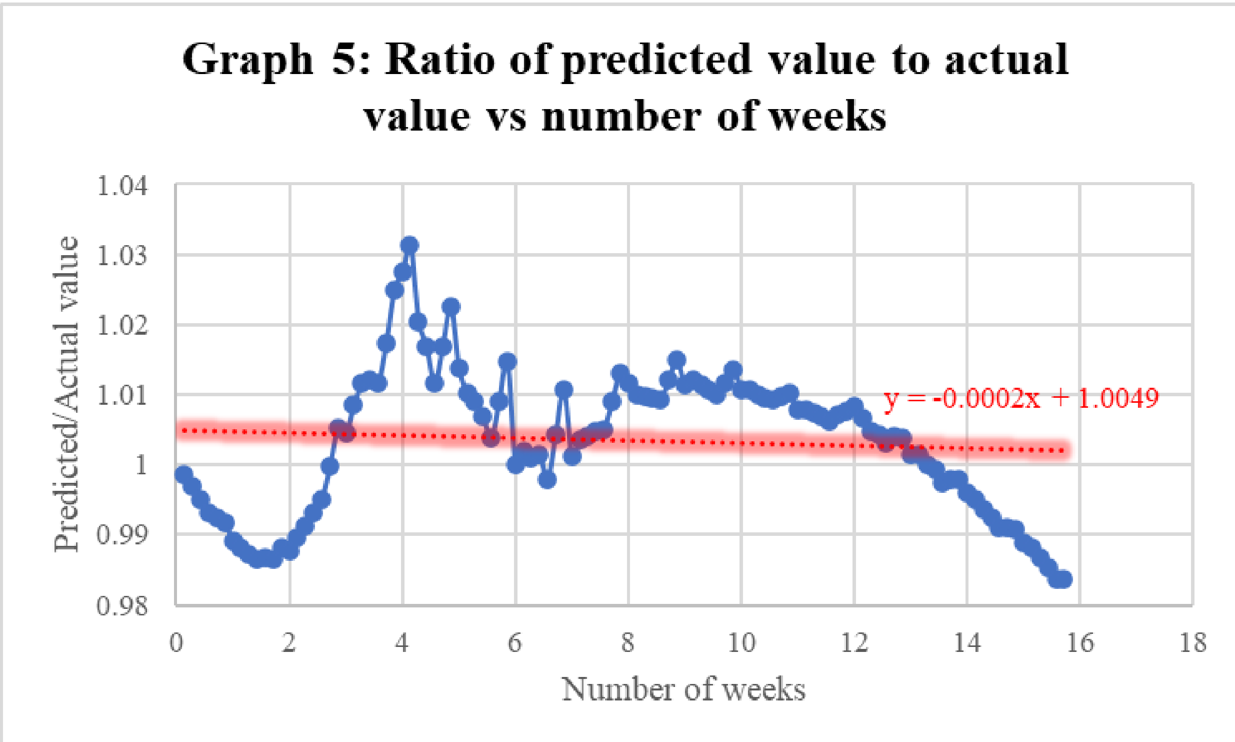
$$N(t) = 1797072 + \frac{1603506}{1+(\frac{t}{6})^{-3.6479}}.$$

<u>Accuracy of predictions</u>

Visually, the logistic function is relatively close to the actual values of case counts. The trends followed are the same and it does not deviate by a noticeable amount. To determine how close the graphs actually are, the actual daily case numbers will be compared to the function at the same *t* values. For example, at *t* = 1 (1 week), the actual case count was 1819357. Substituting *t* = 1 into $N(t) = 1797072 + \frac{1603506}{1+(\frac{t}{6})^{-3.6479}}$, $N(1) = 1799394$. Then, the predicted value is divided by the actual value to find the percent ratio. Shown below is the sample calculation for *t* = 1 in table form.

Actual value	Predicted value	Predicted / Actual value
1819357	1799394	98.90%

After performing this calculation for each day in the data set, **Graph 5** is produced.



From **Graph 5**, the percent ratio ranges from approximately 98% to 103%. The predicted value is higher than the actual value at approximately Weeks 3 - 13. The predicted value is lower than the actual value at approximately Weeks 0 - 3 and 13 - 16. The trendline has a miniscule slope of -0.0002, and stays between 101% and 100%, showing an overall balance in data points.

Conclusion ……………………………………………………………………….

Interpretation of results

Based on the results, the logistic model is a solid method to model the spread of COVID-19. The aim of this investigation, which was to determine the accuracy of the logistic function in COVID-19 spreading patterns by using existing data and other basic models, was satisfied. Case count data was used to create an exponential trend, which was eventually advanced to the logistic model. This logistic model was incredibly consistent with the actual values, with an average percentage error of approximately $\pm\ 2\%$ for the selected time period. However, as seen in **Graph 5**, the model tended to overcompensate when there was a steeper slope and undercompensate when there was a flatter slope. Thus, as the infection number approaches the maximum, there tends to be more cases than predicted. Nevertheless, the logistic function can still be used to make predictions about future case counts. For example, I am curious about the number of COVID-19 cases during the upcoming exam season, which will be approximately 31 weeks after the first Omicron case. Substituting $t = 31$ into the logistic curve's equation, a prediction of 3,396,576 cumulative cases is calculated. Knowing that the theoretical maximum number of infections is 3,400,578, this prediction is reasonable assuming there is no new variant of the virus. Referencing **Graph 5**, $N(31)$ takes place where the curve is remarkably flat, indicating that the daily case counts during this time will not be high. Thus, with this model, I can plan my future accordingly, knowing that exams will likely occur.

Limitations and extensions

However close to the true values this model may be, there still exists many limitations. The most notable limitation is the rising of new variants. As seen in the beginning of this investigation, data specifically after Omicron was found in Canada was taken. At any point in time, if a new variant were to occur, this model would fail. For example, if this model were used at the beginning of the pandemic, the Delta variant would disrupt the trend. To work around this, if the growth factor is steadily below 1 and suddenly begins to rise again, a new function should be created. Additionally, the model does not account for deaths or exposure to previously infected individuals. These changes in infection rate can greatly shift the pandemic spread. Similarly, changing policies such as quarantine measures in a region affects the shape of the curve. The model lacks the complexity to account for these factors, but can be improved by introducing more variables into the growth rate, rather than it just being a constant number. Finally, the data used was based on Canada's case count. However, different provinces have different lockdown procedures and laws, which causes inconsistencies in data collected. Thus, if available, it would have been more preferable to collect data by province or region, rather than the entire country as a whole. However, for this investigation, the data collected was sufficient in determining an accurate trend.

With the seemingly chaotic changes in the daily situation of the pandemic, the logistic model allowed me to tame the sporadic data into a predictable trend.

References

3.4. the logistic equation 3.4.1. the ... - nu math sites. (n.d.). Retrieved March 23, 2022, from https://sites.math.northwestern.edu/~mlerma/courses/math214-2-03f/notes/c2-logist.pdf

4 parameter logistics. AAT Bioquest. (n.d.). Retrieved March 22, 2022, from https://www.aatbio.com/tools/four-parameter-logistic-4pl-curve-regression-online-calculator

COVID-19 Canada tracker. COVID. (n.d.). Retrieved March 22, 2022, from https://covid19tracker.ca/province.html

COVID-19 Canada. COVID. (n.d.). Retrieved March 22, 2022, from https://resources-covid19canada.hub.arcgis.com/

DP Maths: Applications & interpretation: Coronavirus & modelling. Subscription websites for IB teachers & their classes. (n.d.). Retrieved March 22, 2022, from https://www.thinkib.net/mathapplications/page/32838/coronavirus-modelling

Feldman, D. P. (n.d.). *Chaos and fractals*. Oxford Scholarship Online. Retrieved March 22, 2022, from https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199566433.001.0001/acprof-9780199566433

Khan Academy. (n.d.). *Logistic equations (part 1) | differential equations (video)*. Khan Academy. Retrieved March 22, 2022, from https://www.khanacademy.org/math/ap-calculus-bc/bc-differential-equations-new/bc-7-9/v/solving-logistic-differential-equation-part-1

Libretexts. (2021, November 3). *8.4: The logistic equation*. Mathematics LibreTexts. Retrieved March 22, 2022, from https://math.libretexts.org/Bookshelves/Calculus/Book%3A_Calculus_(OpenStax)/08%3A_Introduction_to_Differential_Equations/8.4%3A_The_Logistic_Equation

Logistic growth model for covid-19. Wolfram Cloud. (n.d.). Retrieved March 22, 2022, from https://www.wolframcloud.com/obj/covid-19/Published/Logistic-Growth-Model-for-COVID-19.nb