IBDP Mathematics: Analysis and Approaches (HL)

Internal Assessment

Modeling Seed Germination

Candidate code: kbx008

Exam session: November 2022

Introduction

Since I was a child, I have always been intrigued by how mathematics is applied in everyday life. More specifically, I have been fascinated by mathematical modeling because it is useful and effective to make predictions and grasp the true nature of problems. Hence, I thought that I wanted to explore mathematical modeling and I wondered whether there is anything which I can apply it to.

When I was conducting experiments about plants' germination and growth for my biology internal assessment, I wondered if I could apply mathematical modeling to plants. This is because I had a lot of interests in them and I wanted to deeply understand them. I especially wanted to apply mathematical modeling to seed germination because how it proceeds was unclear to me and I thought that mathematical modeling would greatly help me deepen my understanding about it.

In this investigation, I want to develop a mathematical model which describes the relationship between time elapsed and the total number of seeds which have germinated. The main aim of this investigation is to create a suitable model of seed germination in order to deepen my comprehension about how seed germination proceeds.

Data collection and results

To create the model, I decided to gather data by conducting an experiment specifically for this IA (there has been no duplication of work). I chose to plant seeds of Kaiware Daikon (Raphanus sativus L.) because I had researched about them in my biology IA and I was familiar with them. I planted 100 seeds so that I can convert the number to percentage. I thought this was beneficial because I can apply the model which I create to other cases where the number of seeds is different. In the experiment, I used 5 petri dishes. I planted 20 seeds and used 25 ml of water in each petri dish. Since germination of seeds of Kaiware Daikon usually takes about 2 days, I decided to use my smartphone and record with it. In order to record, an app which takes pictures at regular intervals was used. I planted seeds for 60 hours at about 20 ℃ and pictures were taken every 2 hours with my smartphone. In total, 31 pictures were taken from time $t = 0$ (hours) to time $t = 60$. The reason why I set the interval as 2 hours is because I thought it was small enough to create an accurate model. After

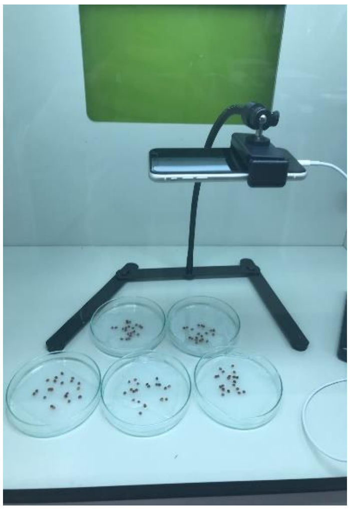


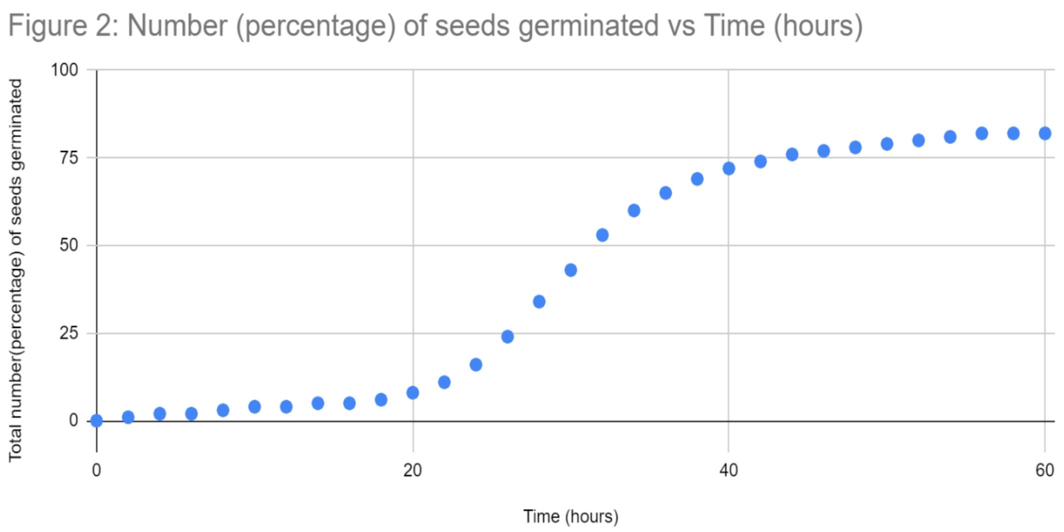
Fig 1: A picture showing how I recorded seed germination

finishing planting them, I counted manually the number of seeds which have germinated at each time seeing the pictures taken. The data is shown in the table below.

Time (hours)	Number (percentage) of seeds germinated	Time (hours)	Number (percentage) of seeds which have germinated	Time (hours)	Number (percentage) of seeds which have germinated
0	0	22	11	42	74
2	1	24	16	44	76
4	2	26	24	46	77
6	2	28	34	48	78
8	3	30	44	50	79
10	4	32	53	52	80
12	4	34	60	54	81
14	5	36	65	56	82
16	5	38	69	58	82
18	6	40	72	60	82
20	8				

Table 1: The number (percentage) of seeds which have germinated at each time

Next, I am going to plot the scatter graph using google spreadsheet.



From the scatter graph, it can be seen that the number (percentage) of seeds germinated increased rapidly starting after about time $t = 20$. I hypothesize that exponential function with the exponential growth will be a suitable model because similar to exponential function, the figure increased slowly at the beginning but increased rapidly as time get bigger.

Modeling

<u>1.Exponential model</u>

In order to assess the hypothesis which I made above, I am going to create an exponential model. Natural base e is going to be used as the base of the function mainly because an exponential function with the base e is easier to differentiate as a derivative of a function $y = e^x$ can simply given by $\frac{dy}{dx} = e^x$. I thought this was advantageous because I might examine the derivative to analyze the exponential model. A function of time, f is defined by $f(t) = ae^{bt} + c$ where a, b and c are constants. $f(t)$ shows the number (percentage) of seeds which have germinated at time t (hours). Since time elapsed cannot be negative, the input t has to be greater than or equal to 0. Therefore, the domain of the model is D: $[0, \infty)$. All of the models which are created in this exploration are going to have the same domain. On the graph, t is going to be plotted on horizontal axis and $f(t)$ is going to be plotted on vertical axis. A few predictions about the function can be made.

<u>a and b are both positive:</u>

The derivative of the function $f(t) = ae^{bt} + c$ is $f'(t) = abe^{bt}$. This is simply obtained by letting $u = bt$ and using chain rule. In order for the exponential function to have the exponential growth which I expect, the derivative of it must be always positive and increase as t get bigger for $t \geq 0$. As for the sings of the two constants, there are following four cases.

① $a > 0$ and $b > 0$ ② $a > 0$ and $b < 0$ ③ $a < 0$ and $b > 0$ ④ $a < 0$ and $b < 0$

The derivative of the function can be always positive and increase as t get bigger for $t \geq 0$ only with case ①. Thus, in order for the exponential function to have the exponential growth which I expect, a and b must be positive.

<u>$f(0) = 0$:</u>

Since I started planting at $t = 0$ and none were germinated at the beginning, f has to be equal to 0 at $t = 0$. Thus, $0 = ae^0 + c$ and $c = -a$.

Since I could find $c = -a$, the function can be expressed with two unknown constants a and b. To find the two unknown constants, two coordinates from the collected data can be used to solve a pair of simultaneous equations.

Two coordinates $(24,16)$ and $(32,53)$ were chosen since it seemed exponential model which has exponential growth would fit well especially at the middle of the data.

By substituting the coordinate pair (24,16) into the function f, the following equation can be made.

$$16 = ae^{24b} - a$$

Dividing both sides by a gives

$$\frac{16}{a} = e^{24b} - 1$$

$$\therefore \ \frac{16}{a} + 1 = e^{24b}$$

$$\therefore \ 24b = \ln\left(\frac{16}{a} + 1\right)$$

$$\therefore \ b = \frac{1}{24}\ln\left(\frac{16}{a} + 1\right) \cdot\cdot\cdot\cdot ①$$

Similarly, substituting the coordinate pair (32, 53) into the function gives

$$b = \frac{1}{32}\ln\left(\frac{53}{a} + 1\right) \cdot\cdot\cdot\cdot ②$$

By equating ① and ②, we can make the following new equation.

$$\frac{1}{24}\ln\left(\frac{16}{a} + 1\right) = \frac{1}{32}\ln\left(\frac{53}{a} + 1\right)$$

By using GDC (graphic display calculator), we can find $a \cong 0.48245$

Thus, $c \cong -0.48245$

By substituting the value of a into equation ①, we can obtain the following equation.

$$b = \frac{1}{24}\ln\left(\frac{16}{0.48245} + 1\right)$$

Since $\left(\frac{16}{0.48245}+1\right)$ is greater than 1, $\ln\left(\frac{16}{0.48245}+1\right)$ is positive. Thus b is positive. As both a and b are positive, the exponential model which is created has the exponential growth.

Since the two constants were obtained, we gain the exponential function

$f(t) = 0.48245e^{\frac{t}{24}\ln\left(\frac{16}{0.48245}+1\right)} - 0.48245$. The graph of this function is shown in figure 3.

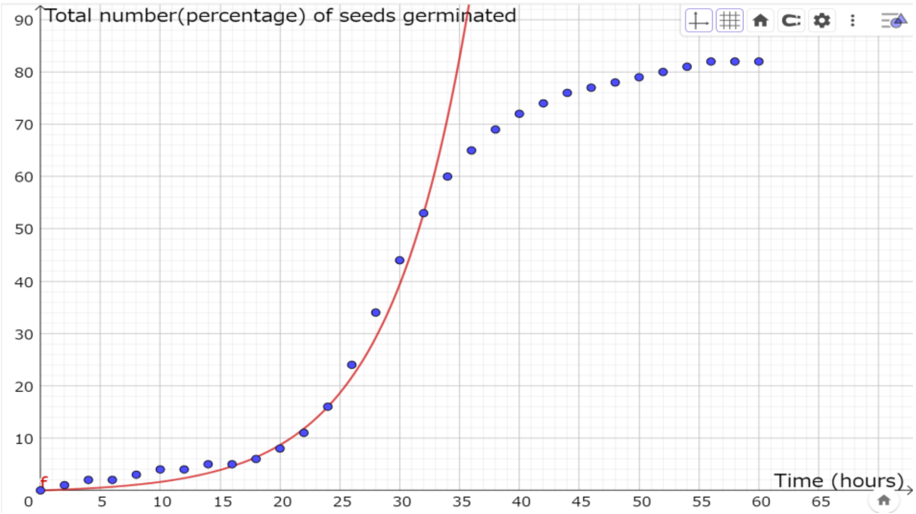


Fig 3: Exponential model compared to the original data

From the plotted graph, it can be seen that although the exponential model fits the data well at first, it does not fit well after about $t = 34$. The data only has exponential growth until the middle. The rate of increase of the collected data decreased after $t = 34$ and the figure reached 82 at the end. 82 represents the germination rate in the conducted experiment. The value of the exponential function which I made get bigger and bigger as t get bigger. This never happen with seed germination because the number(percentage) of seeds in the experiment can never exceed 100. This is another reason why the exponential function was inappropriate. Although 82 is not technically a horizontal asymptote in the collected data as some data has the value of 82, I am going to strive to create a model with a horizontal asymptote of 82 for the sake of convenience.

In order to find a better model, I looked for a function in which the figure increases like exponentially until the middle but converge to a specific value at the end. I realized that the graph of the inverse tangent function has the two properties and the graph looked similar to the scatter graph of the collected data.

2.Inverse tangent model

The inverse tangent function $y=\arctan x$ has horizontal asymptotes at $y=\frac{\pi}{2}$ and $y=-\frac{\pi}{2}$.

The rate of change of the function is given by $\frac{dy}{dx}=\frac{1}{x^2+1}$. As x gets closer to 0 from left, the rate of change increases. However, as x passes 0 and get bigger, the rate of change decreases, creating the s-shaped graph. A function of time, g is defined by $g(t)=A\arctan B(t-C)+D$, where A, B, C and D are constants.

Since I want the graph to fit the scatter graph of the collected data, I will try to make the function have horizontal asymptotes of y=0 and y=82.
As the difference between two horizontal asymptotes of $y=\arctan x$ is π, $g(t)$, which is going to have horizontal asymptotes of $y=0$ and $y=82$ has to be vertically stretched by $\frac{82}{\pi}$. Therefore $A=\frac{82}{\pi}$. In order to have the horizontal asymptotes of $y=82$ and $y=0$, the graph also has to be translated upward by 41. Hence, $D=41$.

To find the two unknown constants, B and C, two coordinates from the collected data can be used to solve pair of simultaneous equations. A few predictions about the constants, B and C, can be made.

<u>B is positive:</u>
To create a function which has the s-shape graph, B has to be positive since A is positive.

<u>C is positive and between 24 and 32:</u>
For the inverse tangent function, the constant C is equal to the value of t at which the instantaneous rate of change is maximum. The derivative of the function $g(t)=\frac{82}{\pi}\arctan B(t-C)+41$ is $g'(t)=\frac{82B}{\pi((B(t-C))^2+1)}$. This is obtained by letting $u=B(t-C)$ and using chain rule. Since the numerator of the derivative must be positive as B must be positive, the smallest value of the denominator makes the value of the derivative the biggest. The value of denominator becomes smallest when $(B(t-C))^2=0$. Therefore, when $t=C$, the instantaneous rate of change is maximum. Since the constant C is equal to the value of t at which the instantaneous rate of change is maximum, C has to be positive and between 24 and 32 because the value of the collected data increased significantly between $t=24$ and $t=32$.

Two coordinates (26, 24) and (54, 81) were chosen so that the model would fit well not only at the beginning but also at the end.

By substituting the coordinate pair (26,24) into the function g, the following equation can be made.

$$24 = \frac{82}{\pi} arctanB(26 - C) + 41$$

$$\therefore \ \frac{82}{\pi}\arctan(26B - BC) = -17$$

$$\therefore \ \arctan(26B - BC) = -\frac{17\pi}{82}$$

$$\therefore \ \tan\left(-\frac{17\pi}{82}\right) = 26B - BC$$

$$\therefore \ BC = 26B - \tan\left(-\frac{17\pi}{82}\right) \cdots ③$$

Similarly, by substituting the coordinate pair (54, 81) into the function, the following equation can be obtained.

$$BC = 54B - \tan\left(\frac{20\pi}{41}\right) \cdots ④$$

By equating ③ and ④, we can make the following new equation.

$$26B - \tan\left(-\frac{17\pi}{82}\right) = 54B - \tan\left(\frac{20\pi}{41}\right)$$

$\therefore \ B \cong 0.959$ (Using GDC)

$\therefore \ BC = 54 \times 0.959 - \tan\left(\frac{20\pi}{41}\right)$ (Using equation④)

$$\therefore \ BC \cong 25.695$$

Dividing both sides by B gives $C \cong 26.795$

This means that the instantaneous rate of change of the inverse tangent function is maximum at $t = 26.795$. As I predicted, C is between 24 and 32.

Since the two constants were obtained, we get the inverse tangent function,

$g(t) = \frac{82}{\pi}\arctan\big(0.959(t - 26.795)\big) + 41$. The graph of this function is shown in figure 4.

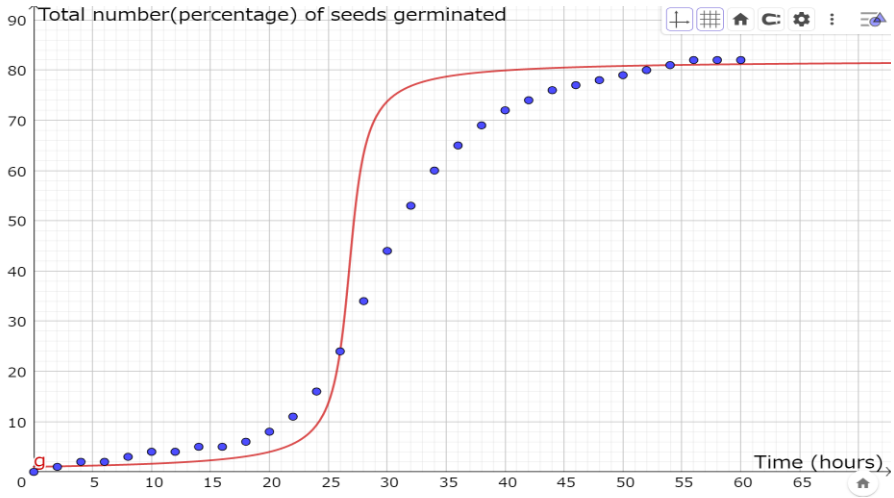


Fig 4: Inverse tangent model compared to the collected data

From the plotted graph, it can be seen that the model does not fit the original data well at the middle. The inverse tangent model increases steeply at the middle compared to the collected data. The steepness of the inverse tangent model is determined by the constant B because the function is horizontally stretched with scale factor $\frac{1}{B}$. If the value of B is large, the maximum rate of change of the function become larger and the graph become steeper at the middle. If the value of B is small, the maximum rate of change of the function become smaller and the graph become less steep at the middle. I realized that the value of B is dependent on the two coordinates which I use to solve pair of simultaneous equations. Therefore, I thought that I could improve the inverse tangent model by choosing different coordinate pair.

3.Modified inverse tangent model

Since the previous inverse tangent model seems to differ greatly from the collected data especially at $t = 22$ and $t = 36$, I am going to use a coordinate pair (22, 11) and (36, 65) to solve pair of simultaneous equations and create the modified inverse tangent function, h which is defined by $h(t) = \frac{82}{\pi}\arctan B(t - C) + 41$, where B and C are constants. I suspect that the modified model fit the original data well compared to the previous inverse tangent model because it passes the two coordinates (22, 11) and (36, 65) which are the points the previous inverse tangent model did not fit well.

By substituting the coordinate pair (22,11) into the function, following equation can be obtained.

$$11 = \frac{82}{\pi} arctan B(22 - C) + 41 \cdots ⑤$$

And by substituting the coordinate pair (36, 65) into the function, following equation can be obtained.

$$65 = \frac{82}{\pi} arctan B(36 - C) + 41 \cdots ⑥$$

Solving ⑤ and ⑥ simultaneously as I did on page 7, following values are obtained.

$$B \cong 0.253, \ C \cong 30.816$$

Since the value of B is much smaller than that of the previous inverse tangent model, the graph of the modified inverse tangent function is less steep at the middle. Also, the instantaneous rate of change of the modified inverse tangent model is maximum at $t = 30.816$.

Since the two constants were obtained, we get the modified Inverse tangent function, $g(t) = \frac{82}{\pi} \arctan\big(0.253(t - 30.816)\big) + 41$. The graph is shown below.

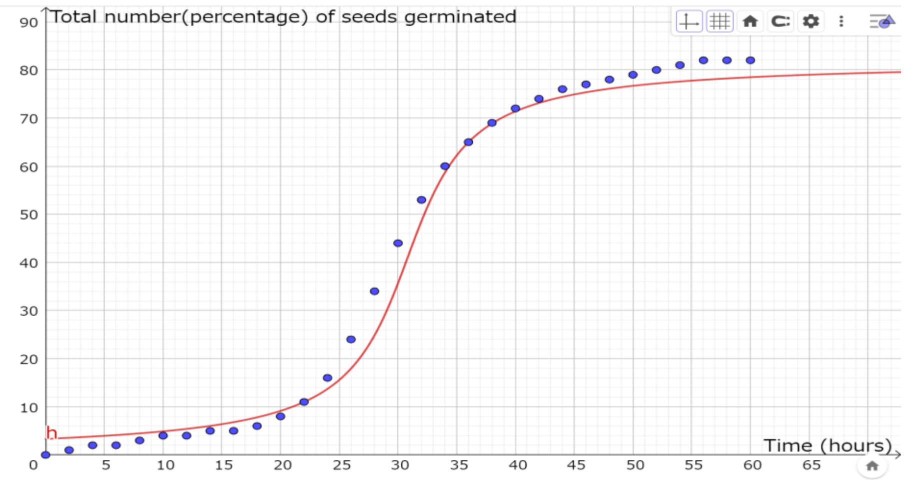


Fig 5: Modified inverse tangent model compared to the collected data

From figure 5, it can be seen that the modified model fits the collected data much better than the previous inverse tangent model. By changing coordinate pair used, I could make graph less steep at the middle. However, although the function fits the original data relatively well, I think that the function is not accurate enough because there are still many small discrepancies.

When I was looking at my math textbook, I realized that the graph of function obtained from the logistic equation is also s-shaped and looked similar to the scatter graph of the collected data.

4. Logistic model

Logistic function can be obtained by solving the following differential equation:

$\frac{dy}{dx} = ky\left(1 - \frac{y}{B}\right)$, where k and B are constants. Although the function obtained from the differential equation increases exponentially at first, the value of $\frac{dy}{dx}$ approaches 0 as the value of y approaches the value of B because the value of $\left(1 - \frac{y}{B}\right)$ approaches 0. Therefore, the constant B in the differential equation is the value of horizontal asymptote of the function obtained from the differential equation. Thus $B = 82$. Substituting $B = 82$ gives the following differential equation.

$$\frac{dy}{dx} = ky\left(1 - \frac{y}{82}\right)$$

$$\therefore\ \frac{dy}{dx} = ky\left(\frac{82-y}{82}\right)$$

Dividing both sides by $y\left(\frac{82-y}{82}\right)$ gives

$$\frac{1}{y}\left(\frac{82}{82-y}\right)\frac{dy}{dx} = k$$

Integrating both sides with respect to x gives

$$\int \frac{1}{y}\left(\frac{82}{82-y}\right)\frac{dy}{dx}dx = \int kdx$$

$$\therefore\ \int \frac{1}{y}\left(\frac{82}{82-y}\right)dy = kx + c$$

Partial fraction decomposition gives

$$\frac{1}{y}\left(\frac{82}{82-y}\right)=\frac{1}{y}+\frac{1}{82-y}$$

$$\therefore\ \int\left(\frac{1}{y}+\frac{1}{82-y}\right)dy=kx+c$$

$$\therefore\ \ln|y|-ln|82-y|=kx+c$$

$$\therefore\ \ln\left|\frac{y}{82-y}\right|=kx+c$$

$$\therefore\ e^{kx+c}=\left|\frac{y}{82-y}\right|$$

$$\therefore\ \pm e^{kx+c}=\frac{y}{82-y}$$

$$\therefore\ \frac{1}{\pm e^{kx+c}}=\frac{82-y}{y}$$

$$\therefore\ \frac{1}{e^{kx}\times\pm e^{c}}=\frac{82}{y}-1$$

$$\therefore\ Ae^{-kx}=\frac{82}{y}-1\ \text{ (letting } \pm e^{-c}=A)$$

$$\therefore\ Ae^{-kx}+1=\frac{82}{y}$$

By multiplying both sides by $\frac{y}{Ae^{-kx}+1}$, we can get the following logistic function[1].

$$y=\frac{82}{Ae^{-kx}+1}$$

[1] In order to solve the logistic differential equation, I referred to my math textbook.
Haese, M., Humphries, M., Sangwin, C., & Vo, N. (2019). *Mathematics Analysis and Approaches HL 2*. Marleston, South Australia: Haese Mathematics.

A function of time, j is defined by $j(t) = \frac{82}{Ae^{-kt}+1}$, where A and k are constants.

To find the two unknown constants, A and k, two coordinates from the collected data can be used to solve pair of simultaneous equations. A few predictions about the constants, A and k, can be made.

A is positive:
A must be positive because $j(t)$ become positive if A is positive. Since the number(percentage) of seeds is always greater than or equal to 0, $j(t)$ must be positive and thus A has to be positive.

k is positive:

The derivative of the logistic function, $y' = ky\left(1 - \frac{y}{82}\right)$ can be rearranged as follow.

$$y' = \left(ky - \frac{ky^2}{82}\right)$$

$$\therefore\ y' = -\frac{k}{82}(y^2 - 82y)$$

$$\therefore\ y' = -\frac{k}{82}(y-41)^2 + \frac{41^2k}{81}$$

The derivative $y' = -\frac{k}{82}(y-41)^2 + \frac{41^2k}{81}$ is a quadratic function with input y $(0 < y < 82)$.

The graph of $y' = -\frac{k}{82}(y-41)^2 + \frac{41^2k}{81}$ with input y $(0 < y < 82)$ has to open downwards so that the logistic function obtained from the differential equation has the s-shaped graph. This means that the coefficient $-\frac{k}{82}$ of y^2 has to be negative. Therefore, k must be positive. The graphs of the derivative for $k > 0$ and $k < 0$ are shown on the next page.

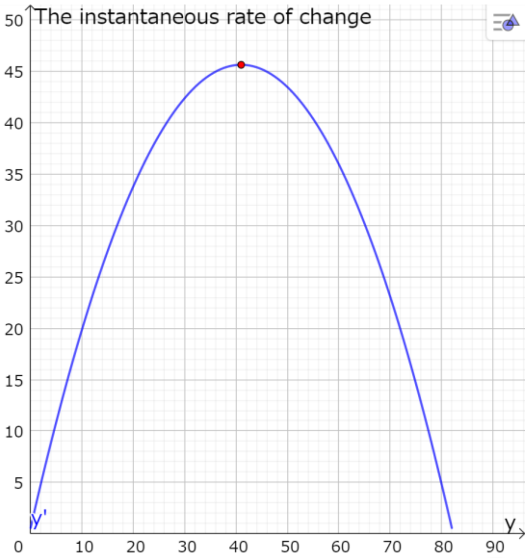


Fig 6: The graph of the derivative when $k > 0$

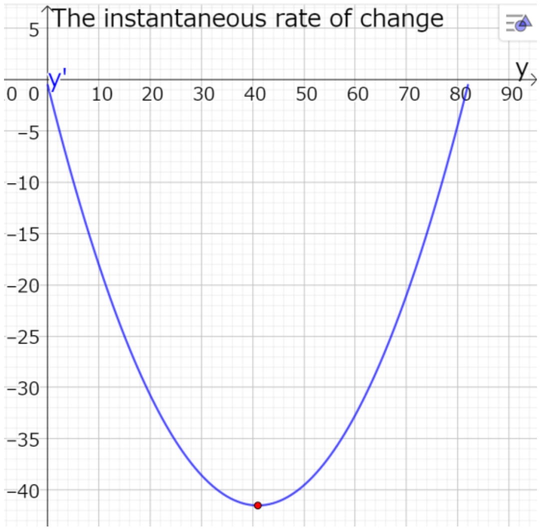


Fig 7: The graph of the derivative when $k < 0$

The red points on the graphs in figure 6 and 7 are representing the vertex $\left(41, \frac{41^2k}{81}\right)$ of the derivative (quadratic function). Since k must be positive, the graph of the derivative should look like the graph in figure 6. Thus, the maximum instantaneous rate of change must be $\frac{41^2k}{81}$ and dependent on the value of k. If k is small, the maximum instantaneous rate of change become small. On the other hand, if k is big, the maximum instantaneous rate of change become big.

Two coordinates (26, 24) and (54, 81) were chosen so that the model would fit well not only at the beginning but also at the end (same as the first inverse tangent function).

By substituting the coordinate pair (26,24) into the function, following equation can be obtained.

$$24 = \frac{82}{Ae^{-26k} + 1}$$

Multiplying both sides by $Ae^{-26k} + 1$ gives

$$24\left(Ae^{-26k} + 1\right) = 82$$

Dividing both side by 24 gives $\qquad Ae^{-26k} + 1 = \frac{41}{12}$

$$\therefore \ Ae^{-26k} = \frac{29}{12}$$

$$\therefore \ e^{-26k} = \frac{29}{12A}$$

$$\therefore \ -26k = \ln\left(\frac{29}{12A}\right)$$

$$\therefore \ k = -\frac{1}{26}\ln\left(\frac{29}{12A}\right) \cdots\cdots ⑤$$

Similarly, by substituting the coordinate pair (54, 81) into the function, the following equation can be obtained.

$$k = -\frac{1}{54}\ln\left(\frac{1}{81A}\right) \cdots\cdots ⑥$$

By equating ⑤ and ⑥, we can make a following new equation.

$$-\frac{1}{26}\ln\left(\frac{29}{12A}\right) = -\frac{1}{54}\ln\left(\frac{1}{81A}\right)$$

$$\therefore \ A \cong 324.5081 \ \text{(Using GDC)}$$

$$\therefore \ k = -\frac{1}{54}\ln\left(\frac{1}{81\times 328.2}\right) \ \text{(using ⑥)}$$

$$\therefore \ k \cong 0.18846$$

Therefore, we get the logistic function $j(t) = \frac{82}{324.5081^{-0.18846t}+1}$. The graph of this function is shown on the next page.

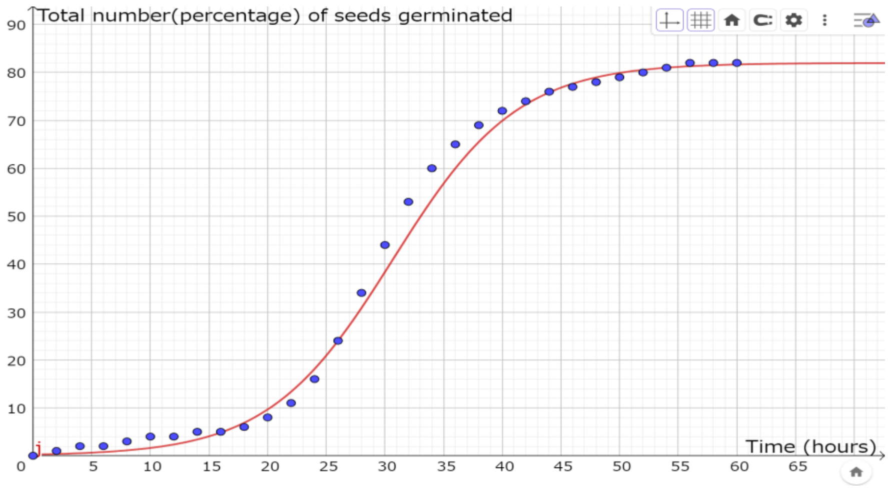


Fig 8: Logistic model compared to the original data

From figure 8, it can be seen that the model fits the original data fits well especially at the beginning and the end. Although there are some discrepancies at the middle, the graph seems to have less discrepancies compared to the modified inverse tangent model. In addition to fitting the original data well, the logistic model has a horizontal asymptote of $y = 82$. Thus, this model seems to be valid to describe seed germination because the number (percentage) of seeds germinated is not likely to increase from 82 after 60 hours.

Analysis

In order to know which model fits the original data best, I am going to find the residual sum of squares (RSS) for each model. I thought that finding which model is the best using mathematical method was crucial because both of the last two models seem to fit the original data really well and mathematical method can tell us which is the best objectively. Residual is the difference between the observed value and the value from the model[2]. If we add up the residuals, the negative residuals can offset the positive residuals. Thus, they have to be squared. RSS is the sum of all the squared residuals and it can be defined as follow[3]:

$$RSS = \sum_{i=1}^{n} (\hat{y}i - yi)^2$$

[2] 一色政彦. (2021, November 22). 残差平方和（RSS：Residual Sum of Squares）／［損失関数］二乗和誤差（SSE：Sum of Squared Error）とは？. 残差平方和（RSS：Residual Sum of Squares）／［損失関数］二乗和誤差（SSE：Sum of Squared Error）とは？：AI・機械学習の用語辞典 - @IT (itmedia.co.jp).

[3] 一色政彦. (2021, November 22). 残差平方和（RSS：Residual Sum of Squares）／［損失関数］二乗和誤差（SSE：Sum of Squared Error）とは？. 残差平方和（RSS：Residual Sum of Squares）／［損失関数］二乗和誤差（SSE：Sum of Squared Error）とは？：AI・機械学習の用語辞典 - @IT (itmedia.co.jp).

In the equation, n represents the number of observed values, $\hat{y}$ represents the observed values and y represents the values from the model.

I calculated the RSS for each model using google spreadsheet. The table below shows the result.

Data	Exponential model		Arctangent model		Modified Arctangent		Logistic model	
	$\hat{y}i - yi$	$(\hat{y}i - yi)^2$	$\hat{y}i - yi$	$(\hat{y}i - yi)^2$	$\hat{y}i - yi$	$(\hat{y}i - yi)^2$	$\hat{y}i - yi$	$(\hat{y}i - yi)^2$
①	0.00	0.00	-1.02	1.03	-3.33	11.11	-0.25	0.06
②	0.83	0.70	-0.10	0.01	-2.56	6.54	0.63	0.40
③	1.60	2.67	0.81	0.65	-1.82	3.31	1.47	2.15
④	1.31	1.73	0.69	0.48	-2.12	4.51	1.22	1.50
⑤	1.92	3.67	1.55	2.41	-1.48	2.18	1.87	3.51
⑥	2.38	5.67	2.38	5.67	-0.90	0.81	2.37	5.61
⑦	1.66	2.76	2.16	4.68	-1.40	1.97	1.64	2.70
⑧	1.70	2.88	2.88	8.28	-1.03	1.05	1.61	2.59
⑨	0.40	0.16	2.49	6.18	-1.80	3.26	0.15	0.02
⑩	-0.34	0.11	2.92	8.52	-1.81	3.27	-0.88	0.78
⑪	-0.67	0.45	4.03	16.20	-1.14	1.31	-1.66	2.76
⑫	-0.80	0.64	5.41	29.28	0.00	0.00	-2.36	5.59
⑬	0.00	0.00	6.68	44.62	2.28	5.21	-2.13	4.54
⑭	2.36	5.57	0.00	0.00	6.06	36.75	0.00	0.00
⑮	4.79	22.96	-29.38	863.23	9.16	83.86	3.15	9.90
⑯	3.63	13.21	-30.79	948.01	7.31	53.49	4.63	21.44
⑰	0.00	0.00	-23.84	568.31	4.40	19.39	6.93	48.08
⑱	-11.30	127.65	-18.25	333.01	1.30	1.69	6.58	43.33
⑲	-30.86	952.19	-14.06	197.56	0.00	0.00	5.02	25.17
⑳	-59.82	3578.32	-10.58	111.89	0.13	0.02	3.49	12.21
㉑	-101.06	10212.77	-7.94	63.09	0.61	0.37	2.08	4.31
㉒	-158.43	25101.11	-6.21	38.60	0.87	0.75	0.69	0.47
㉓	-236.12	55753.98	-4.42	19.54	1.60	2.57	0.16	0.03
㉔	-342.08	117016.84	-3.58	12.85	1.65	2.71	-0.67	0.45
㉕	-484.63	234861.39	-2.72	7.38	1.90	3.61	-0.98	0.96
㉖	-676.29	457362.48	-1.83	3.34	2.30	5.31	-0.90	0.82
㉗	-933.86	872101.18	-0.92	0.85	2.81	7.92	-0.55	0.30
㉘	-1279.91	1638169.61	0.00	0.00	3.41	11.61	0.00	0.00
㉙	-1744.70	3043965.18	0.93	0.87	4.06	16.51	0.69	0.47
㉚	-2369.85	5616166.32	0.87	0.76	3.77	14.20	0.47	0.22
㉛	-3208.88	10296922.4	0.82	0.67	3.51	12.35	0.33	0.11
RSS		**22372354.6**		**3297.99**		**317.65**		**200.51**

Table 2: Calculation of RSS

The value of RSS decreases from exponential model to the logistic model. The logistic model has the smallest RSS value (200.51) of the four models, meaning that the logistic model fits the original data best and it is the most suitable model of the four to describe seed germination.

Although the RSS value of logistic model was smaller than that of the modified inverse tangent model, they looked almost identical to me. I thought that I could find the difference between them and the reason why logistic model is better by finding and comparing the derivative of the two models. I believe that looking at derivative of the two models enable me to analyze models from different perspective.

The derivative of both models can be obtained by using chain rule.

For the modified Inverse tangent function $h(t) = \frac{82}{\pi}\arctan\left(0.253(t-30.816)\right) + 41$,

letting $u = \left(0.253(t-30.816)\right)$ gives

$$h(u) = \frac{82}{\pi}\arctan u + 41$$

$$\therefore\ h'(t) = \left(\frac{82}{\pi} \times \frac{1}{u^2+1}\right) \times u' \text{ (Using chain rule)}$$

$$\therefore\ h'(t) = \frac{82}{\pi} \times \frac{1}{(0.253(t-30.816))^2+1} \times 0.253$$

For the logistic function $j(t) = \frac{82}{324.5081e^{-0.18846t}+1}$, letting $v = 324.5081e^{-0.18846t} + 1$ gives

$$j(v) = 82v^{-1}.$$

$$\therefore\ j'(t) = \left(82 \times -1 \times \frac{1}{v^2}\right) \times v' \text{ (Using chain rule)}$$

$$\therefore\ j'(t) = -\frac{82}{(324.5081e^{-0.18846t}+1)^2} \times 324.5081 \times -0.18846e^{-0.18846t}$$

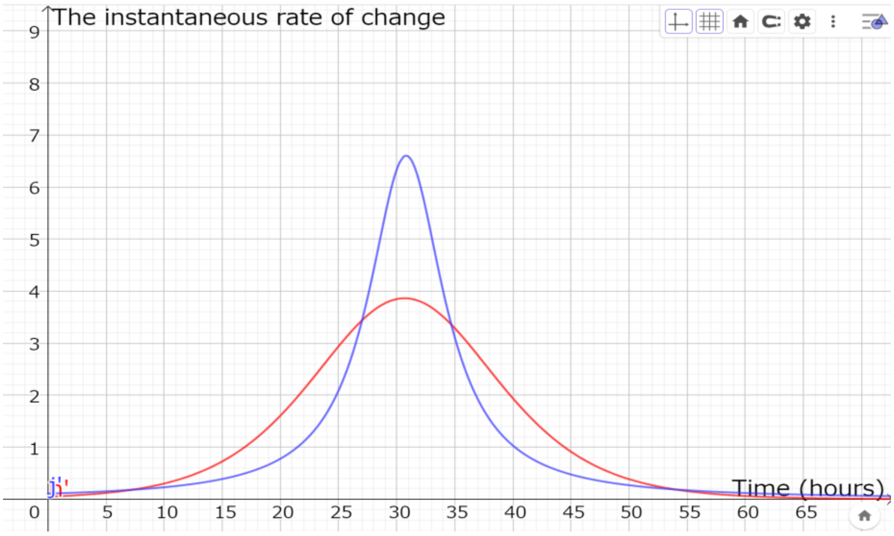


Fig 9: The graph of h' (blue) and the graph of j' (red)

From figure 9, it can be seen that the instantaneous rate of change of both models become maximum at around $t = 31$. However, the maximum value of instantaneous rate of change differs greatly. h' has the maximum value of around 6.6, whereas j' has the maximum value of around 3.8. I collected data with two-hour interval and the value of the collected data increases by 10 at most in 2 hours, meaning that the maximum average rate of change is 5. This suggests that the maximum value of around 3.8 is a little bit small. Because of this, the logistic function has lower values than the collected data at the middle, where the value of collected data increases most significantly. Also, the maximum value of around 6.6 is not appropriate as this is a little bit large. Because of this, the modified inverse tangent is relatively steep at the middle.

In addition, the instantaneous rate of change of the modified inverse model changes relatively drastically, whereas that of the logistic model changes comparatively slowly. For example, although the instantaneous rate of change of the modified inverse model has larger maximum value, it decreases drastically after reaching the maximum. Because of that, the modified inverse tangent model has smaller values than the collected data and does not fit the collected data well compared to the logistic model at the end.

Conclusion

To conclude, it was found that the logistic model $j(t) = \frac{82}{324.5081e^{-0.18846t}+1}$ is the most suitable model for seed germination of Kaiware Daikon (Raphanus sativus L.) of the four models which I made. I could successfully fulfill my aim of creating a suitable model for seed germination by trying different types of models and comparing them. The fact that the logistic model describes the germination well tells us that the number of seeds germinated increase exponentially at first but eventually level off to reach a maximum value. Through the investigation, I could deepen not only my understanding of mathematics by using concepts from a variety of areas but also my understanding of how seed germination proceeds.

Evaluation and extensions

Through the investigation, I realized that the fitness of models to the observed data can be highly dependent on which coordinates I use to substitute into the model. For example, the first inverse tangent model was improved greatly by choosing a different coordinate pair. This implies that the logistic model which I created might be able to be more accurate if I choose different coordinates to use for substitution. Therefore, to create more accurate model, I should have tried a few sets of coordinates to use for substitution. In addition, the number of observed data can be increased to create more accurate model.

The biggest limitation of the logistic model is that it can only be used to describe how seed germination proceeds at the same condition as the experiment which I conducted. This is because how seed germination proceeds is likely to change depending on the condition of germination. I am really curious about how changing the temperature affect the rate of change of the model.

For further research, I would like to explore more about how seed germination is distributed. From figure 9, I realized that the logistic model, which describes the seed germination best, has symmetric derivative. This indicates that a few seeds germinate at the beginning and the end and many seeds germinate at the middle. I thought this implied the possibility of seed germination being normally distributed. If seed germination is normally distributed, it indicates that the relationship between time elapsed and number of seeds germinated can be described better with the cumulative distribution function than logistic function. Thus, exploring how seed germination is distributed contributes to creating more accurate model of seed germination.

Additionally, I would like to explore model of seed germination for different plants. Since how seed germination proceeds differ greatly between plants species, I believe that it will be interesting to find models of seed germination for other plants. By doing that, interesting patterns or connections might be discovered.

Bibliography

Haese, M., Humphries, M., Sangwin, C., & Vo, N. (2019). *Mathematics Analysis and Approaches HL 2*. Marleston, South Australia: Haese Mathematics.

一色政彦. (2021, November 22). 残差平方和（RSS：Residual Sum of Squares）／［損失関数］二乗和誤差（SSE：Sum of Squared Error）とは？.

https://atmarkit.itmedia.co.jp/ait/articles/2111/22/news011.html