Modelling the effect of temperature on the cooking time of simple syrup

Chantal Zhang¹

¹Affiliation not available

January 4, 2024
Modelling the effect of temperature on the cooking time of simple syrup

Chantal Zhang

January 2023
1 Abstract

This research paper delves into the meticulous modeling of temperature changes during the preparation of Italian meringue buttercream. The primary focus is on controlling the cooking time for a crucial step in the process, the creation of simple syrup. The aim is to ensure the syrup reaches precisely 235°F to attain the desired buttercream consistency without crystallization issues. The study investigates various mathematical methods to model the temperature-time relationship based on collected experimental data points. The analysis involves rigorous mathematical computations to determine the parameters of the best-fitting functions. The study ultimately presents detailed insights into the complexities of this culinary process and provides a comprehensive exploration of mathematical methodologies for function approximation in similar scenarios.
2 Introduction

I have always loved baking. It is one of my favorite hobbies, and it has become a household tradition that I bake a birthday cake for everyone’s birthday. There are many different flavors and preferences, but I always frost the cakes with Italian meringue buttercream. One of the steps in making the buttercream is creating a simple syrup that must be boiled until it reaches exactly 235 F. Going even five degrees over this temperature can result in a dense buttercream with chunks of sugar. Going five degrees below this temperature can lead to similar issues. As such, it is imperative that the cooking process be carefully controlled. Unfortunately, I have a very short attention span. More than that, the thermometer I own at home is an infrared thermometer, and only measures the temperature if I press a button. This means that I have to test the temperature of my simple sporadically. When the simple syrup reaches 235 F, it is very easy for short seconds to make a large difference in the temperature.

While it is possible for me to buy a new thermometer, there is another way to solve my issue. I can model the temperature of the simple syrup with respect to time, and use this function to determine exactly how much time must elapse before I can take the simple syrup off the stove.

To do so, I will need to conduct an experiment in order to take measurements of the temperature of the simple syrup. This is so that I can have discrete points from which I can find an equation of best fit. The equation of best fit will be found through function transformations on a base function.

To begin, the recipe I use to make the simple syrup in Italian meringue buttercream - courtesy of Preppy Kitchen - calls for \( \frac{3}{4} \) cups of water and 1 cup of white sugar. However, I’ve found that this tends to make too much frosting. As such, I typically half that amount. Thus, for the experiment, I will use \( \frac{1}{6} \) cups of water and \( \frac{1}{2} \) cups of sugar.

2.1 Aim of the Exploration

To create a function which will allow me to solve for the amount of time I have to wait before my simple syrup reaches 235°F.

3 Data Collection

From past experience, I know it takes about eight minutes for the simple syrup to come to temperature. As such, I will take temperature measurements fifteen seconds in order to have at least fifteen points to create my scatter plot.

I used an infrared thermometer in order to take measurements of the temperature. It measured the temperature with a precision of one decimal place. However, when the temperature exceeded 200°F, it only reported whole number values. As such, all my data has been recorded as an integer in order to keep everything at the same level of precision. My experiment set up is shown below.

\(^1\text{A syrup made by boiling water and sugar together}\)
As I began the experiment, I noticed an issue. I originally conducted this experiment with the heat on high. Due to the intense heat, it took less than four minutes for the simple syrup to come to temperature. With the incredibly short time span, I only managed to take six temperature measurements, which while a satisfactory amount for a science lab report, I felt did not suit my purposes since I wanted more points to indicate an overall trend. There were two ways to solve this problem: I could take temperature measurements at shorter time intervals, or I could turn down the heat. While it would be possible to record the measurements at shorter time intervals, this would make it much more difficult for me to write down the values in time. Thirty seconds was already very difficult for me to stay on-pace; anything less and the accuracy and precision of the data - I felt - would need to be called into question. With that, I made the decision to turn the heat of the stove down to low. This meant the simple syrup took more time to come to temperature, and so I have sixteen data points, instead of six.

Table 1. Raw data acquired from experiment

<table>
<thead>
<tr>
<th>Time (seconds)</th>
<th>Temperature (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>73</td>
</tr>
<tr>
<td>30</td>
<td>107</td>
</tr>
<tr>
<td>60</td>
<td>129</td>
</tr>
<tr>
<td>90</td>
<td>142</td>
</tr>
<tr>
<td>120</td>
<td>148</td>
</tr>
<tr>
<td>150</td>
<td>164</td>
</tr>
<tr>
<td>180</td>
<td>175</td>
</tr>
<tr>
<td>210</td>
<td>194</td>
</tr>
<tr>
<td>240</td>
<td>198</td>
</tr>
<tr>
<td>270</td>
<td>205</td>
</tr>
<tr>
<td>300</td>
<td>213</td>
</tr>
<tr>
<td>330</td>
<td>218</td>
</tr>
<tr>
<td>360</td>
<td>221</td>
</tr>
<tr>
<td>390</td>
<td>229</td>
</tr>
<tr>
<td>420</td>
<td>230</td>
</tr>
<tr>
<td>450</td>
<td>237</td>
</tr>
</tbody>
</table>
4 Modelling the simple syrup

Now that the raw data has been acquired, I can begin applying function transformations to various functions to determine which best represents my data.

Note that thenceforth, \( x \) will refer to the time elapsed during the simple syrup experiment, as measured in seconds. \( f(x) \) will refer to the temperature of the simple syrup at time \( x \), as measured in \(^\circ\text{F}\).

![Figure 2. Scatter plot of the raw data, showing the relationship between the temperature of the simple syrup in degrees Fahrenheit and the time elapsed in seconds](image)

Looking at the graph, my mind thinks of a square root function right away. This is because of the concave downward increase of the function that does not seem to asymptotic behavior as the time approaches infinity. While there are other possible functions I could base my line of best off of - such as a logarithmic function - I believe a square root function gives me enough flexibility to manipulate the parameters to my liking.

Any square root function can be written in its general form as:

\[
 f(x) = a\sqrt{x - h} + k \tag{1}
\]

where \( a, h, \) and \( k \) are all real numbers, representing function transformations. Those values can be manipulated such that the resulting square root function closely resembles the overall trend of the data points collected in the experiment.

Before the calculations begin, a note regarding precision should be addressed. All values presented will be rounded to three decimal places in order to maintain a consistent level of precision and neatness within the calculations. Any exemptions will be noted.

4.1 Method 1

I chose three random points: (0, 73), (210, 194), and (420, 230). They are evenly enough spaced across my data, and include the point (0, 73), which must be a point on the line of best fit.

Solving for the values of \( a, h, \) and \( k \) proceeded as follows:

Subbing the coordinate values into equation 1:
\[ 73 = a\sqrt{0 - h} + k \]  
(a)
\[ 194 = a\sqrt{210 - h} + k \]  
(b)
\[ 230 = a\sqrt{420 - h} + k \]  
(c)

Subtracting equation a from equation b:

\[ 194 - 73 = a\sqrt{210 - h} + k - (a\sqrt{-h} + k) \]
\[ 121 = a(\sqrt{210 - h} - \sqrt{-h}) \]

Rearranging to solve for an expression for a:

\[ \frac{121}{\sqrt{210 - h} - \sqrt{-h}} = a \]  
(d)

Next, subtracting equation b from c:

\[ 230 - 194 = a\sqrt{420 - h} + k - (a\sqrt{210 - h} + k) \]
\[ 36 = a(\sqrt{420 - h} - \sqrt{210 - h}) \]

Rearranging to solve for an expression for a:

\[ \frac{36}{\sqrt{420 - h} - \sqrt{210 - h}} = a \]  
(e)

Setting equation d equal to equation e:

\[ \frac{121}{\sqrt{210 - h} - \sqrt{-h}} = \frac{36}{\sqrt{420 - h} - \sqrt{210 - h}} \]
\[ 121(\sqrt{420 - h} - \sqrt{210 - h}) = 36(\sqrt{210 - h} - \sqrt{-h}) \]  
(f)

Solving for h:
\[
121\sqrt{420 - h} + 36\sqrt{-h} = 157\sqrt{210 - h} \\
\left(121\sqrt{420 - h} + 36\sqrt{-h}\right)^2 = \left(157\sqrt{210 - h}\right)^2 \\
2 \cdot 121 \cdot 36 \sqrt{(-h)(420 - h)} = 157^2(210 - h) - 121^2(420 - h) + 36^2(h) \\
2 \cdot 121 \cdot 36 \sqrt{(-h)(420 - h)} = 157^2(210) - 157^2(h) + 121^2(h) - 121^2(420) + 36^2(h) \\
2 \cdot 121 \cdot 36 \sqrt{(-h)(420 - h)} = (121^2 - 157^2 + 36^2)(h) + 157^2(210) - 121^2(420) \\
\left[2 \cdot 121 \cdot 36 \sqrt{(-h)(420 - h)}\right]^2 = [(121^2 - 157^2 + 36^2)(h) + 157^2(210) - 121^2(420)]^2 \\
(2 \cdot 121 \cdot 36)^2(h)(h - 420) = (121^2 - 157^2 + 36^2)^2(h^2) + (157^2(210) - 121^2(420))^2 + \\
2 \cdot (121^2 - 157^2 + 36^2)(h) \cdot (157^2(210) - 121^2(420)) \\
(2 \cdot 121 \cdot 36)^2h^2 - (2 \cdot 121 \cdot 36)^2 \cdot 420(h) = (121^2 - 157^2 + 36^2)^2(h^2) + \\
(121^2 - 157^2 + 36^2) \cdot (157^2(210) - 121^2(420))(h) + \\
(157^2(210) - 121^2(420))^2 \\
0 = -75907656 \cdot h^2 + 48829888800 \cdot h - 946592784900 \\
\]

Using the quadratic formula to solve for \(h\):

\[
h = \frac{-48829888800 \pm \sqrt{48829888800^2 - 4 \cdot (-75907656)(-946592784900)}}{2 \cdot (-75907656)}
\]

\(h_1 = 662.11424758177\), \(h_2 = -18.834091493171\)

Note that extra decimal places were kept for precision purposes.

This was originally a square root equation. In squaring both sides, I have obtained extraneous root(s). From equation a, it is evident that \(h < 0\), or else a negative square root would be obtained. Thus, the first root must be discarded. The value of \(h\) is \(-18.834091493171\).

However, checking to see if that value of \(h\) aligns with the equations yields the following conundrum.

Subbing the value of \(h\) into equation f gives:

\[
\text{LS} \\
= \frac{121}{\sqrt{210 - h} - \sqrt{-h}} \\
= \frac{121}{\sqrt{210 + 18.83} - \sqrt{18.83}} \\
= 11.224
\]
RS

\[
\begin{align*}
36 &= \sqrt{420 - h} - \sqrt{210 - h} \\
\frac{36}{\sqrt{420 + 18.83} - \sqrt{210 + 18.83}} &= 6.183
\end{align*}
\]

As can be seen, the right side does not equal the left side; this value of \( h \) is also an extraneous root. There is only one possible explanation for this: this system of equations has no real solutions. But how can I prove this? I thought for some time and came up with an idea. I can graph equation \( d \) and equation \( e \) separately, with \( a \) as my independent variable and \( h \) as my dependent variable. If they intersect, I have a solution to my system of equations. If they do not, then there exists no real solutions.

Looking at the graph, how do I know there can be no solution? Looking at the above figure, as the \( h \) value of each function approaches negative infinity, the \( a \) value of function \( d \) increases more rapidly than equation \( e \). This means that the difference between the \( a \) values is increasing; there can be no intersection as the \( h \) value approaches negative infinity. Furthermore, there can be no intersection as the \( h \) value of the functions approach positive infinity. This is because equation \( d \) starts at \( h = 0 \) and extends to the left; if there is no intersection at \( h = 0 \), there will never be an intersection between the two functions. Thus, it is impossible for equation \( d \) and \( e \) to intersect. In other words, it is impossible for me to derive an equation for the square root function using a system of three equations.

But how do I know there’s no solution for any combination of three points? The truth is, I don’t. However, the three points I chose were arbitrary. There is a large chance that a new set of three points would also yield the same results. Furthermore, any \( h \) values I might solve for using any other set of points would need to be similar to the one I’ve solved for in this question. This is because that \( h \) value would’ve pertained to the line of best fit, and all of my data points seem to follow an overall trend. Thus, through this deduction I can assume that there is no possible way for me to solve a system of three equations for this set of data.

4.2 Method 2

There seems to be some kind of limit I hit when I try and solve a system of three equations. But what if I simplified my problem to two equations instead?
According to the general form of a square root function:

\[ f(x) = a\sqrt{x - h} + k \]

I can "assume" one value of the parameters, leaving me to solve for the value of the other two parameters. The simplest manner that immediately came to mind was the following.

What if I set \( h = 0 \)?

Looking at the data points, the temperature has an initial value of 73 °F. Thus, when \( x = 0, f(x) = 73 \). This is the only point that I know for certain must align with the line of best fit. This is because the initial temperature of the simple syrup is guaranteed. This is only one guaranteed point, from which I have to solve for two variables - mathematically impossible. However, if I set \( h = 0 \), then I can deduce a value for \( k \). As can be seen with subbing \( x = 0 \) into equation 1:

\[
\begin{align*}
73 &= a\sqrt{0 - 0} + k \\
73 &= k
\end{align*}
\]

Thus, \( k = 73 \). This makes sense, since this is equivalent to vertically translating the graph of \( f(x) = \sqrt{x} \) 73 °F upwards. Graphing the equation \( f(x) = \sqrt{x} + 73 \) gives:

![Scatterplot of raw data with first attempt of line of best fit](image)

This very obviously does not fit the data points. However, a value for \( a \) - the compression factor - has not been determined. It can be solved for, using any point from the data and substituting that value into \( f(x) \).

Using the arbitrarily chosen point (30, 107), solving for \( a \) proceeds as follows:

\[
\begin{align*}
f(x) &= a\sqrt{x} + 73 \\
107 &= a\sqrt{30} + 73 \\
\frac{107 - 73}{\sqrt{30}} &= a \\
6.208 &= a
\end{align*}
\]
Using this value of \( a \) gives a function of

\[
f(x) = 6.208\sqrt{x} + 73
\]

Graphing that function with the data point shows:

![Graph showing function](image)

Figure 5. Second attempt at graphing my line of best fit. Points on graph are raw data; line is second attempt at line of best fit

While this graph does look significantly better than my precious attempt, it still can be improved to better suit the data points. There was a flaw with my method of calculating my \( a \) value. In using only one point, I only take into account that point’s position, and not the entire trend. Thus, in order to account for all sixteen of my data points, I will repeat the calculations for my \( a \) value another fourteen times, once for every data point not yet used. This yielded the data in Table 2.

<table>
<thead>
<tr>
<th>Time (seconds)</th>
<th>Temperature (°F)</th>
<th>( a ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>129</td>
<td>7.230</td>
</tr>
<tr>
<td>90</td>
<td>142</td>
<td>7.273</td>
</tr>
<tr>
<td>120</td>
<td>148</td>
<td>6.847</td>
</tr>
<tr>
<td>150</td>
<td>164</td>
<td>7.430</td>
</tr>
<tr>
<td>180</td>
<td>175</td>
<td>7.603</td>
</tr>
<tr>
<td>210</td>
<td>194</td>
<td>8.350</td>
</tr>
<tr>
<td>240</td>
<td>198</td>
<td>8.069</td>
</tr>
<tr>
<td>270</td>
<td>205</td>
<td>8.033</td>
</tr>
<tr>
<td>300</td>
<td>213</td>
<td>8.083</td>
</tr>
<tr>
<td>330</td>
<td>218</td>
<td>7.982</td>
</tr>
<tr>
<td>360</td>
<td>221</td>
<td>7.800</td>
</tr>
<tr>
<td>390</td>
<td>229</td>
<td>7.899</td>
</tr>
<tr>
<td>420</td>
<td>230</td>
<td>7.661</td>
</tr>
<tr>
<td>450</td>
<td>237</td>
<td>7.731</td>
</tr>
</tbody>
</table>

The average of those values gives us an \( a \) value of 7.971. This means our function is now

\[
f(x) = 7.971\sqrt{x} + 73 \quad (f_1)
\]

Graphing the function yields:
This very obviously matches my data far better than the previous two graphs. It passes through approximately half the points, and for the ones it misses, it comes very close. However, there are some issues. For values of time $x < 200$, the graph tends to over predict the temperature. There are several possible reasons for this, but the most glaring one is the fact that I assumed my function would not need to be horizontally translated from its original form as stated in equation 1.

I’ve already shown that solving for all three parameters at a time doesn’t work. But what if I assume the value of a different parameter?

### 4.3 Method 3

It goes without saying that in a square root function, $a$ cannot equal 0. If it does, the square root function becomes a linear function. As such, to investigate an alternate line of best fit, I can set $k$ equal to 0.

Looking once more at equation 1, if I set $k = 0$, the function becomes:

$$f(x) = a\sqrt{x} - h$$  \hspace{1cm} (2)

Once more, I need two equations to solve for two unknowns. One of the points I use has to be $(0, 73)$. The other point I choose arbitrarily to be $(240, 198)$. This creates a set of equations:

$$73 = a\sqrt{0} - h$$
$$198 = a\sqrt{240} - h$$

Rearranging to solve for $a$ in each equation:

$$a = \frac{73}{\sqrt{-h}}$$ \hspace{1cm} (g)
$$a = \frac{198}{\sqrt{240} - h}$$ \hspace{1cm} (h)

Setting equation $g$ equal to equation $h$ and solving for the value of $h$: 
\[ 73 \sqrt{-h} = 198 \sqrt{240 - h} \]
\[ 73\sqrt{240 - h} = 198\sqrt{-h} \]
\[ 73^2(240 - h) = 198^2(-h) \]
\[ 73^2 \cdot 240 = h(73^2 - 198^2) \]
\[ \frac{73^2 \cdot 240}{73^2 - 198^2} = h \]
\[ h = -37.755 \]

Subbing the value of \( h \) into equation g to solve for \( a \):

\[ a = \frac{73}{\sqrt{-(-37.755)}} \]
\[ = 11.880 \]

Thus, my final function is

\[ f(x) = 11.880\sqrt{x + 37.755} \] (4)

Graphing shows:

Figure 7. Graph of equation 4 compared to the data points

This graph produces a line distinct to past ones in the sense that there are suddenly very obvious limitations to what values \( x \) - time - can have. Equation 4 is only reasonable when \( x \geq 0 \).

Interestingly, this line of best fit tends to under predict the value of the first few data points, while largely over predicting the last five data points. Once more, I can take the average of multiple calculated values of \( a \) and \( h \) in order to produce a more representative line of best fit. In order to maximize this process, I will calculate three more values of \( a \) and \( h \) using three sets of points. One of the data points in each set, once more, must be \((0, 73)\). As for the other data points, two will be taken from two points that the current line of best fit tends to over predict, and one point where the line tends to under predict its value. This is because there are more points which are over predicted that under predicted. By accounting for these points, I can aim to rectify as many of the discrepancies in my graph as possible.

The sets of points I’ve chosen are \((60, 129)\), \((360, 221)\), and \((420, 230)\). The results of the calculations are shown in Table 5.
Table 3. Results of calculating new values of $a$ and $h$

<table>
<thead>
<tr>
<th>Set of points $(x_1, f(x_1)), (x_2, f(x_2))$</th>
<th>$h$-value</th>
<th>$a$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0, 73), (60, 129)$</td>
<td>-28.256</td>
<td>13.733</td>
</tr>
<tr>
<td>$(0, 73), (360, 221)$</td>
<td>-44.090</td>
<td>10.994</td>
</tr>
<tr>
<td>$(0, 73), (420, 230)$</td>
<td>-47.049</td>
<td>10.643</td>
</tr>
</tbody>
</table>

Taking the average of the four $a$ and $h$ values:

$$a_{\text{avg}} = \frac{13.733 + 10.994 + 10.643 + 11.880}{4} = 11.813$$

$$h_{\text{avg}} = \frac{-28.256 + (-44.090) + (-47.049) + (-37.550)}{4} = -39.236$$

Thus, my final function is:

$$f(x) = 11.813\sqrt{x} + 39.236, \ x \geq 0 \quad (f_2)$$

Graphing yields:

This line of best fit looks incredibly similar to the one in figure 6: the temperature at times $x > 200$ are over predicted, and the temperature at time $x < 100$ are underpredicted, almost to the same degree. But why does this graph look so similar, despite my attempts to rectify this issue? In choosing three points that were either below or above the line of best fit, their effect on the $a$ and $h$ values ultimately cancelled out. This is a sign that no matter how many points I use to try and modify the line of best fit, it will ultimately circle back to a function with an incredibly similar shape as that of $f_2(x)$. Thus, the equation of my line of best fit is $f_2(x) = 11.813\sqrt{x} + 39.236, \ x \geq 0$

5 Determining the best function

I have two functions, and while from a visual standpoint arguments can be made as to which function better represents the data, ultimately I will need a unarguable, quantitative manner to determine which function better represents my data points.
In SL Math, I was taught to use Pearson’s Correlation Coefficient in order to evaluate the reliability through which a line of best fit represents a set of data points. However, the \( r \) value is best used for a linear relationship (Pennsylvania State University, 2018). There is an alternative: the \( R^2 \) value. It is a coefficient used to measure the relative error between a line of best fit and a set of data points. In other words, it is an objective indication as to how well a line of best fit represents a set of data points (Pennsylvania State University, 2018).

Its formula is as follows (Pennsylvania State University, 2018):

\[
R^2 = 1 - \frac{\text{sum squared regression}}{\text{total sum of squares}}
\]

The sum of the squared regression is essentially the distance from the data point to the line of best fit, squared. The total sum of squares is the sum of the differences between the y-values and the average y-value, squared. Knowing this, the formula can be written to:

\[
R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}
\]

Where:

- \( y_i \) is the i-th y-coordinate in the data set
- \( \hat{y}_i \) is the i-th y-coordinate as predicted by the line of best fit
- \( \bar{y} \) is the average y-coordinate value of the data points

Note that in this case the y-coordinate values refer to \( f(x) \), or the temperature of the simple syrup in °F.

The table below illustrates the calculations for the \( R^2 \) value of equation \( f_1 \).
Table 4. Calculations for $R^2$ value of $f_1$

<table>
<thead>
<tr>
<th>$y_i$</th>
<th>$y_i$</th>
<th>$y_i - y_i$</th>
<th>$(y_i - y_i)^2$</th>
<th>$y_i - \bar{y}$</th>
<th>$(y_i - \tilde{y})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>73</td>
<td>0</td>
<td>0</td>
<td>-107.188</td>
<td>11489.160</td>
</tr>
<tr>
<td>107</td>
<td>116.659</td>
<td>-9.659</td>
<td>93.300</td>
<td>-73.188</td>
<td>5356.410</td>
</tr>
<tr>
<td>129</td>
<td>134.743</td>
<td>-7.743</td>
<td>32.983</td>
<td>-51.188</td>
<td>2620.160</td>
</tr>
<tr>
<td>142</td>
<td>148.620</td>
<td>-6.620</td>
<td>43.818</td>
<td>-38.188</td>
<td>1458.290</td>
</tr>
<tr>
<td>148</td>
<td>160.318</td>
<td>-12.318</td>
<td>151.731</td>
<td>-32.188</td>
<td>1036.035</td>
</tr>
<tr>
<td>164</td>
<td>170.624</td>
<td>-6.624</td>
<td>43.883</td>
<td>-16.188</td>
<td>262.035</td>
</tr>
<tr>
<td>175</td>
<td>179.942</td>
<td>-4.942</td>
<td>24.425</td>
<td>-5.188</td>
<td>26.910</td>
</tr>
<tr>
<td>194</td>
<td>188.511</td>
<td>5.489</td>
<td>30.132</td>
<td>13.813</td>
<td>190.785</td>
</tr>
<tr>
<td>198</td>
<td>196.486</td>
<td>1.514</td>
<td>2.292</td>
<td>17.813</td>
<td>317.285</td>
</tr>
<tr>
<td>205</td>
<td>203.977</td>
<td>1.023</td>
<td>1.047</td>
<td>24.813</td>
<td>615.660</td>
</tr>
<tr>
<td>213</td>
<td>211.062</td>
<td>1.938</td>
<td>3.757</td>
<td>32.813</td>
<td>1076.660</td>
</tr>
<tr>
<td>218</td>
<td>217.800</td>
<td>0.200</td>
<td>0.040</td>
<td>37.813</td>
<td>1429.785</td>
</tr>
<tr>
<td>221</td>
<td>224.239</td>
<td>-3.24</td>
<td>10.492</td>
<td>40.813</td>
<td>1665.660</td>
</tr>
<tr>
<td>229</td>
<td>230.415</td>
<td>-1.415</td>
<td>2.001</td>
<td>48.813</td>
<td>2382.660</td>
</tr>
<tr>
<td>230</td>
<td>236.357</td>
<td>-6.357</td>
<td>40.410</td>
<td>49.813</td>
<td>2481.285</td>
</tr>
<tr>
<td>237</td>
<td>242.090</td>
<td>-5.090</td>
<td>25.913</td>
<td>56.813</td>
<td>3227.660</td>
</tr>
</tbody>
</table>

$\bar{y} = \frac{73 + 107 + \ldots + 237}{16} = 180.188$

From the table:

$\sum(y_i - \hat{y}_i)^2 = 506.219$

$\sum(y_i - y)^2 = 35636.4375$

Subbing these values into the formula for $R^2$ gives:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{506.219}{35636.438} = 0.986$$

Thus, $R^2 = 0.986$

Repeating this process for the $R^2$ value of $f_2$ gives:
Table 5. Calculations for $R^2$ value of $f_2$

<table>
<thead>
<tr>
<th>$y_i$</th>
<th>$\hat{y}_i$</th>
<th>$y_i - \hat{y}_i$</th>
<th>$(y_i - \hat{y}_i)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>73.996</td>
<td>-0.996</td>
<td>0.993</td>
</tr>
<tr>
<td>107</td>
<td>98.294</td>
<td>8.706</td>
<td>75.795</td>
</tr>
<tr>
<td>129</td>
<td>117.678</td>
<td>11.322</td>
<td>128.188</td>
</tr>
<tr>
<td>142</td>
<td>134.293</td>
<td>7.707</td>
<td>59.398</td>
</tr>
<tr>
<td>148</td>
<td>149.067</td>
<td>-1.067</td>
<td>1.139</td>
</tr>
<tr>
<td>164</td>
<td>162.504</td>
<td>1.496</td>
<td>2.239</td>
</tr>
<tr>
<td>175</td>
<td>174.911</td>
<td>0.089</td>
<td>0.008</td>
</tr>
<tr>
<td>194</td>
<td>186.495</td>
<td>7.505</td>
<td>56.326</td>
</tr>
<tr>
<td>198</td>
<td>197.400</td>
<td>0.600</td>
<td>0.360</td>
</tr>
<tr>
<td>205</td>
<td>207.733</td>
<td>-2.733</td>
<td>7.470</td>
</tr>
<tr>
<td>213</td>
<td>217.577</td>
<td>-4.577</td>
<td>20.949</td>
</tr>
<tr>
<td>218</td>
<td>226.993</td>
<td>-8.993</td>
<td>80.875</td>
</tr>
<tr>
<td>221</td>
<td>236.035</td>
<td>-15.035</td>
<td>226.052</td>
</tr>
<tr>
<td>229</td>
<td>244.742</td>
<td>-15.742</td>
<td>247.811</td>
</tr>
<tr>
<td>230</td>
<td>253.151</td>
<td>-23.151</td>
<td>535.969</td>
</tr>
<tr>
<td>237</td>
<td>261.288</td>
<td>-24.288</td>
<td>589.907</td>
</tr>
</tbody>
</table>

Note that the calculations for $y_i - \bar{y}$ and $(y_i - \bar{y})^2$ were omitted from this table because their values are the same as found in Table 4.

From the table, $\Sigma(y_i - \hat{y}_i)^2 = 2033.479$. It was previously calculated that $\Sigma(y_i - \bar{y})^2 = 35636.438$.

Thus, the value of $R^2$ for equation $f_2$ is:

$$
R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} = 1 - \frac{2033.479}{35636.438} = 0.943
$$

6 Conclusion

6.1 Conclusion

The higher the $R^2$ value, the more representative the line of best fit, with respect to the data points (Pennsylvania State University, 2018). Ultimately, depending on the context of a situation, the threshold for an "acceptable" $R^2$ value may change (Pennsylvania State University, 2018). However, it is agreed that the higher the $R^2$ value, the more representative the line of best fit.

The $R^2$ value for $f_1$ was the largest at $R^2 = 0.986$ - but even still it was only by 0.043. This suggests that both functions are extremely close in their reliability. This makes sense as the calculated values for the parameters were averaged to increase the accuracy with respect to the data points.

Thus, since the $R^2$ value of $f_1$ was the greatest, this means that I should use that equation to help calculate when my simple syrup will reach its desired temperature.

Returning to the original aim of the exploration, the amount of time I should wait before my simple syrup reaches 235 °F is:
\[ f_1(x) = 7.971\sqrt{x} + 73 \]
\[ 235 = 7.971\sqrt{x} + 73 \]
\[ x = 413.052 \]

Thus, I should wait approximately 413 seconds before my simple syrup can be used.

6.2 Limitations and Areas of Improvement

By choosing to only look at the possibilities of a model using the square root function, I limited myself to a very specific range of possibilities. While this decision ultimately helped create a streamlined structure for my exploration, it also meant that I discounted possibilities along the way.

Furthermore, although I worked around my initial problem of being unable to solve a system of three equations, I did not fully explore the possibilities. What might happen if both \( h \) and \( k \) are set to 0? What might happen if \( a \) is set to 1? Could a more accurate function be presented? While I did want to explore both those avenues and other combinations, I ultimately had to exclude them from my exploration in order to stay within the page limit. In the future, perhaps in a more flexible environment, it would be interesting to investigate how fixing multiple parameters might affect the square root function.

Additionally, my value for time is only usable under very specific conditions. I used a copper pot to boil the sugar, with the heat on low and the lid taken off. It is an ineffective way of boiling anything and has compromised the applicability of this exploration’s results. In spite of that, I was still able to create a function that models the amount of time it takes before my simple syrup will reach my desired temperature. For that, I consider this exploration a great success. In celebration of completing this paper and hopefully attaining a passing grade, I hope to use my newfound results and frost a cake.
7 Works Cited

1. 2.5 - the coefficient of determination, R-squared. 2.5 - The Coefficient of Determination, r-squared — STAT 462. (2018). Retrieved February 4, 2023, from https://online.stat.psu.edu/stat462/node/95/
