# INTERNAL ASSESSMENT

## Modelling Microwave Popcorn

Exopixel. *Stock Photo: Frame Made Of Popcorn*. Web. 13 Nov. 2016. <https://www.dreamstime.com/stock-photo-frame-made-popcorn-over-white-background-image45909144>.

## Introduction and Rationale

The applications of mathematics in real-life have always fascinated me. As an aspiring engineer, I am often motivated to discover how I can practically apply the theory that I learn in the classroom to real-life. I became highly enthusiastic recently when I stumbled across something that not only I could use to demystify the elusive uses of maths in the world, but was also highly personally relevant: popcorn.

My affinity to popcorn manifests whether I am watching a movie or simply looking for a light snack. Too many a time have I also taken a break from futile attempts on a challenging maths question to make popcorn, simply to take some time to consider the problem from a different perspective. Yes – popcorn is inspirational.

Being such an important part of my daily life, I have concocted and perfected a method to cook popcorn kernels on the stove. Being quite the popcorn connoisseur, I have always insisted that fresh, homemade popcorn beats any pre-packaged, or microwave popcorn. This changed after last summer when I entered $6^{th}$ year. I was thrilled that the microwaves that for so long had been kept from us at school became at our disposal. A few weeks in, the packaged popcorn in the vending machines simply wasn't cutting it for me – I wanted, fresh, hot, popcorn, even if it wasn't of ideal stovetop quality.

I tried to take full advantage of the new opportunity to cook my favourite healthy snack in the microwave, but very quickly, a problem arose. The transition from stove to microwave was not without tribulation, as it seemed that no matter for what length of time I microwaved it for, the popcorn would always either be undercooked (largely unpopped) or overcooked (mostly burnt). Since popcorn is my snack of choice when doing maths, I thought it would be appropriate to strengthen the relationship between my favourite healthy snack and my favourite school subject by using maths to solve this enigma.

And since cooking food in the microwave only has one variable, time, I simply needed only find a mathematical model that I could use to work out for exactly how long I should microwave my popcorn. A consideration in this exploration is that my peers and classmates don't seem to share my enthusiasm, despite my willingness to share, for popcorn, and do not appreciate the lengthy time that it takes to cook. I don't want to eat a bowl full of unpopped kernels, but at the same time, I don't want to keep my fellow microwave regulars waiting in line. I could use maths to make the perfect maize treat in the least amount of time.

In other words, reflecting the engineer that I am ambitious to become, I would use a theoretical model in context and with real-life considerations to conclude the optimal duration for which to microwave popcorn.

## Aim and Approach

As established in the introduction, the aim of this exploration will be to determine the length of time to keep popcorn in the microwave by tailoring and interpreting a mathematical model. This model will relate the percentage of popcorn popped to time under the typical conditions under which I make popcorn – the exact bowl, number of kernels, brand of popcorn, and microwave.

Because I wanted the model to correspond to the conditions under which I make popcorn, I of course needed to collect first hand data through experiment. There is no secondary data available, and certainly no data from an experiment using the exact same specific conditions.

I decided to approach the problem by seeing if relationship between the total amount of popcorn that had popped and the time elapsed described as an equation. I realised that I would first need to create a plot for the percentage of popped kernels vs. time.

My first instinct was to record the exact time of each pop, and then plot the total number of pops at each of these instances in time. I soon realised however that the number of popped kernels is discrete but time is continuous so the time would need to be accurate to the number of decimal places the time measuring device I used, and that the data extraction would be highly laborious.

Instead, I decided to take the approach of dividing the duration for which the popcorn was in the microwave into discrete intervals and counting the number of pops in each interval. I could then plot a cumulative distribution graph and the ogive would represent the percentage of popped popcorn vs. time.

Intuition suggested that the number of popped kernels would increase over time and that the number of popped kernels could not exceed the number of kernels that were placed into the microwave, i.e. the proportion of popped kernels could not go over 100%. From personal experience and observation, it seemed that this pops occurred very quickly at first, then slowed down until very few pops are heard. I also noticed that kernels only begin to pop a short time after the microwave is turned on. These indicated an increasing curve with a horizontal asymptote, decreasing slope, and horizontal offset – this sounded to me like an exponential relationship. If this were the case I expected the base to be $e$ as this is often the case in a relationship seen in the real world.

The expected observed relationship was $p = a + be^{ct}$, where p is the percentage of the kernels popped, $e$ is Euler's number, t is time elapsed, and a,b, and c are constants (transformation factors) that either dilate or translate the graph.

To test my hypothesis, I conducted the experiment. I chose to use 100 kernels exactly in the experiment because this number reflected my typical serving amount. It would also mean that the percentage of popped kernels would be equal to the actual number of popped kernels, and that the percentages would only be whole numbers.

## Data Collection and Results

Exactly 100 kernels were placed into the bowl that I use to make popcorn and the accompanying lid was placed on top. The popcorn was microwaved and measurements were taken for 5 minutes (300 seconds) starting from when the microwave was turned on. This reflected just over the amount of time after which the popcorn would start to burn, and the people queuing to use the microwaves would begin to become exceedingly disgruntled. A sound recorder was used to record the pops.

The recording was then analysed manually and the total number of pops in each 10 second interval, starting at 0 seconds, was counted. This sized class width was chosen as it was deemed sensible: it divided the total amount of time (300 seconds) evenly into 30 groups so that each interval could be of the same width, was large enough to count manually, and was small enough to give a good representation of the data with minimal loss of fidelity, allowing for easy interpretation.

Each class, except for the first one, excluded the lower class-boundary but included the upper class-boundary for time, e.g. the second boundary was $10\text{s} < \text{time} \leq 20\text{s}$, or $(10, 20]$. If a pop occurred exactly on a boundary, it would be counted towards the class where the pop occurs on the upper boundary.

The data obtained is summarised on the following which shows the time interval, frequency of pops, and cumulative frequency of pops (calculated by adding the frequency of the class interval to the cumulative frequency for the previous class interval). Since the number of kernels is 100, the frequency is also equal to the relative frequency (percentage), and the cumulative frequency is equal to the relative cumulative frequency (percentage).

| Time (s) | Number (or percentage) of pops | Cumulative number (or percentage) of pops | Time (s) | Number (or percentage) of pops | Cumulative number (or percentage) of pops |
|---|---|---|---|---|---|
| 0-10 | 0 | 0 | 150-160 | 5 | 42 |
| 10-20 | 1 | 1 | 160-170 | 6 | 48 |
| 20-30 | 0 | 1 | 170-180 | 5 | 53 |
| 30-40 | 1 | 2 | 180-190 | 6 | 59 |
| 40-50 | 1 | 3 | 190-200 | 5 | 64 |
| 50-60 | 1 | 4 | 200-210 | 6 | 70 |
| 60-70 | 2 | 6 | 210-220 | 4 | 74 |
| 70-80 | 3 | 9 | 220-230 | 4 | 78 |
| 80-90 | 3 | 12 | 230-240 | 2 | 80 |
| 90-100 | 3 | 15 | 240-250 | 1 | 81 |
| 100-110 | 4 | 19 | 250-260 | 1 | 82 |
| 110-120 | 4 | 23 | 260-270 | 0 | 82 |
| 120-130 | 4 | 27 | 270-280 | 1 | 83 |
| 130-140 | 5 | 32 | 280-290 | 0 | 83 |
| 140-150 | 5 | 37 | 290-300 | 0 | 83 |

**Table 1** (Relative) cumulative frequency distribution table of the times at which the kernels popped

From these data, I could then plot a histogram:

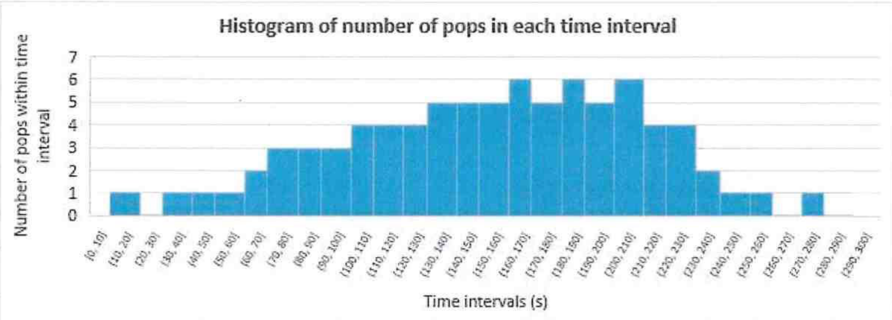


**Fig. 1** Histogram of the amount of time taken for each of the 83 kernels to pop

A cumulative frequency graph could also be plotted on a scatter-graph. By convention, the cumulative frequency for an interval is represented by the point at the upper boundary of that interval.

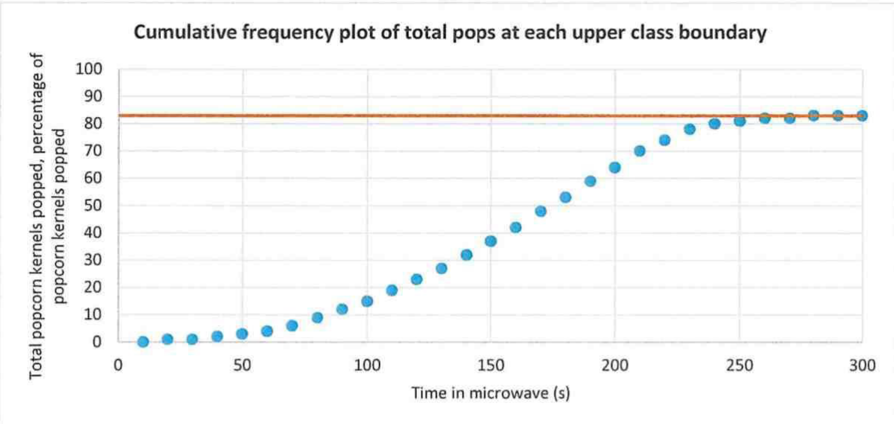


**Fig. 2** Cumulative frequency plot of the number pops (blue) with a horizontal asymptote of y = 83 (orange)

A horizontal asymptote of 83 kernels (83%) is also shown for clarity. The significance of this was that the microwave was kept on for a minute after data collection (to check the asymptote, but there were no new pops after that time. This is strictly not an asymptote however, as the graph not only approaches, but can reach 83(%). It is often the case that no matter for how long some of the kernels are heated, they will not pop. In this case, such kernels account for 17% of the total number kernels. In an ideal case where every kernel can pop, the asymptote would be 100(%).

## Modelling and Mathematical Manipulation of Results

***Exponential model***

From the hypothetical exponential relationship: $p = a + be^{ct}$, we can define a function of time, E, where $E(t) = a + be^{ct}$ and equals to both the number and percentage of kernels popped. When graphed, t, the independent variable goes on the x-axis, and the dependent variable E(t) goes on the y-axis.

From observing the plotted graph, a few assumptions can be made, presuming that the ogive/line of best fit of the relationship between time elapsed and number/percentage of kernels popped follows the hypothesised exponential relationship.

**1. The exponent, ct, is negative:** as previously asserted, the number of popped kernels cannot exceed the number of kernels placed in the bowl: 100. The presence of an asymptote indicates that the y value must approach a certain number as $x \to \infty$. This can only occur if the power of the exponent is negative, i.e. $ct < 0$. This effectively reflects the graph of $e^x$ in the y-axis.

**2. The vertical stretch factor, b, is negative:** this can be inferred from the fact that y approaches a number as x increases. This means that $b < 0$. This effectively reflects, in the x-axis, the graph achieved from the first assumption.

**3. The horizontal asymptote is 83:** without a vertical translation factor, the curve would approach 0 as $x \to \infty$. This would indicate that the vertical shift factor would be 83 to ascertain a horizontal asymptote of $E(t) = 83$. This is not strictly an asymptote, however, as the plot does reach this value for E(t), but for the purposes of modelling, this can be condoned.

**4. When t = 0, the number of pops also equals 0**: The Y-intercept and X-intercept are both (0,0); the graph goes through the origin.

Assumption 4 states that when y = 0, x = 0. From assumptions 1-3, it can be conjectured that function P will be such that $E(t) = be^{ct} + 83$ to model a relationship with such attributes, where b is the vertical stretch factor prior to the translation, and c is the horizontal stretch factor.

To find the constants b and h, figures from the data can be substituted into the equation. The fourth-to-last data point, (270,82), was used along with the origin. The last 3 data points cannot be used because the number of pops equals the value of the asymptote; no solutions can be found by using these points as substitutions.

| | |
|---|---|
| $0 = be^0 + 83$ *(0,0)*<br>$\mathbf{b = -83}$ | $82 = be^{c270} + 83$ *(270,82)*<br>$be^{270c} = -1$ |

$$-83e^{270c} = -1 \quad \textit{substituting b into } be^{270c} = -1$$
$$e^{270c} = \frac{1}{83}$$
$$\ln(e^{270c}) = \ln\left(\frac{1}{83}\right)$$
$$270c(\ln e) = \ln 1 - \ln 83 \quad \textit{log rules}$$
$$270c = 0 - \ln 83$$
$$\mathbf{c = -\frac{\ln 83}{270}}$$

Substituting the values for b and h into the function E results in $E(t) = -83e^{-\frac{\ln 83}{270}t} + 83$. When graphed, the function produces the following curve:

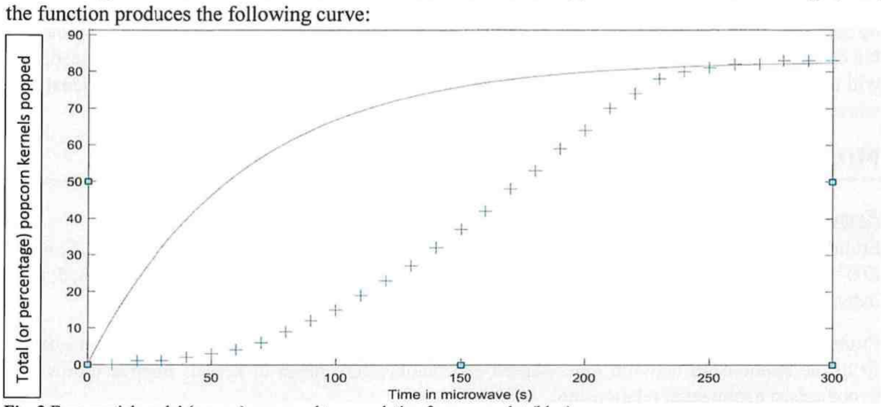


**Fig. 3** Exponential model (orange) compared to cumulative frequency plot (blue)

Without doing any mathematical analysis on the curve produced from the model, it can be seen by eye that the curve deviates significantly from the data points – particularly the centre-portion where the curve shows values much higher than the data points.

Though I only mainly needed to look at the right-hand-side portion of the graph to determine the amount of time to keep the popcorn in the microwave, I decided that the model was not sufficient to model the relationship. I want to use the model for any time shorter than ~220, the data will deviate significantly (more than 10%) from the actual data.

After looking at the plot and failed attempted for some time, I noticed that the ogive of the scatterplot resembled a cubic curve, so I tried to find an equation of such a model to fit the data…

***Cubic Model***

For a cubic model, I suspected that the leading coefficient was a negative number, and that the shape of the plotted popcorn data approximately corresponded to the portion of the function shown in the blue rectangle (right). From this, a few characteristics could be derived for a function C.

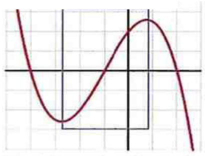

**Fig. 4** General cubic function with negative leading coefficient, image from Wikipedia

**1. When t = 0, y = 0. The curve passes through origin**

**2. There is a double root and local minimum at the origin (0,0):**

A cubic equation must have 3 complex roots, and since complex roots come in conjugate pairs, the function in question must either have 3 real roots or 1 real and 2 imaginary roots. In this case, the right portion of the graph will tend to negative infinity as the x value, t, increases and tends to infinity. Since the maximum turning point occurs where the y value is greater than 0, the right portion of the graph will cross the x axis.

It would also seem that the origin, (0,0), is the minimum turning point of the function, as the plot seems to have a derivative/slope of approximately 0 at this point. If this is the case, it would mean that $(x + 0)$ is a root, and has a multiplicity of two as it is the repeated root. This means that $(x + 0)(x + 0)$, i.e. $x^2$ is a factor

**3: The local maximum is (300, 83):**

$C'(t) = 0$ at the local maximum, (300, 83), and at the local minimum, (0,0) - the turning points have derivatives (slopes) of 0.

C'(0) = 0 and C'(300) = 0, so (t – 0) and (t – 300) are factors of C'(t).

C'(t) = q(t)(t – 300) where q is a constant factor of C'(t), resulting in $C'(t) = qt^2 - q300t$

| | |
|---|---|
| 1. Integrating this expression:<br>$q\int t^2 - 300t\ dt$<br>$q(\frac{1}{3}t^2 - 150t^2 + c)$<br>$\mathbf{C(t) = q(\frac{1}{3}t^3 - 150t^2 + c)}$ | 2. Then substituting in the values for (0,0)<br>$C(0) = q(\frac{1}{3}(0)^3 - 150(0)^2 + c) = 0$<br>**c = 0** |
| 3. Substituting c = 0 into the function<br>$\mathbf{C(t) = q(\frac{1}{3}x^3 - 150x^2)}$ | 4. Then substituting the values for (300, 83)<br>$C(300) = q(\frac{1}{3}(300)^3 - 150(300)^2) = 83$<br>$q(9000000 - 13500000) = 83$<br>$q(-4500000) = 83$<br>$\mathbf{q = -\frac{83}{4500000}}$ |
| Resulting in the final function: $\mathbf{C(t) = -\frac{83}{4500000}(\frac{1}{3}x^3 - 150x^2)}$ | |

Plotting this in Matlab gives the following result in relation to the original data:

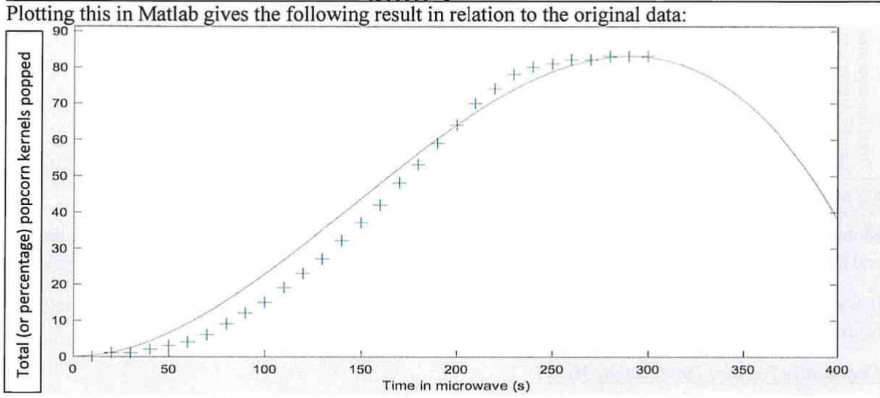


**Fig. 5** Cubic model (orange) compared to cumulative frequency plot (blue)

This model seemed to be much closer to the data points, especially for the first few and last few data points, despite some discrepancy in the left portion of the graph where the model is consistently higher than the data, and in the right portion where the model is lower.

There was a problem with this model however. There is a max turning point instead of an asymptote at (300, 83) which means that the modelled curve starts to decrease after 300 seconds. This would mean that when extrapolating to account for a longer time more time, the model would be incorrect.

Furthermore, though prevalent in digital applications, cubic relationships are quite rare in nature. Considering this, I noticed that the data points seemed to resemble a curve that is commonly observed in nature: a sine function.

***Sine Model***

The model would be in the form: S(t) = a sin (b(t + c)) + d

The plot has a range of 83 (83-0), so it has an amplitude of 41.5 (83/2), which becomes the vertical stretch factor: a = 41.5

After being vertically stretched, the graph must be translated up by the value of the amplitude, 41.5, so that the minima of the sine graph occur at y = 0: d = 41.5

The period of the function dictates the horizontal stretch factor. The period of the function is equal to $\frac{2\pi}{b}$. Half of the period is 300, so the period is 600.

$600 = \frac{2\pi}{b}$ → $b = \frac{\pi}{300}$

The horizontal phase shift (translation factor) is 1/4 of the period, and to the right, so c = 150.
This gives the function $S(t) = 41.5 \sin\left(\frac{\pi}{300}(t - 150)\right) + 41.5$
When graphed, it gives the following curve in relation to the data:

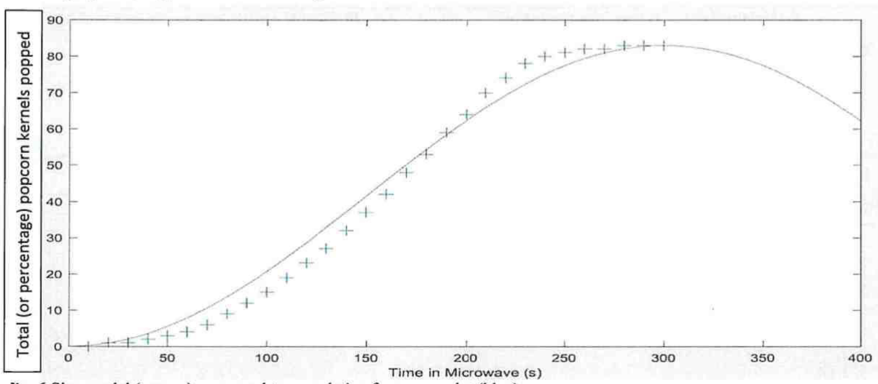


Fig. 6 Sine model (orange) compared to cumulative frequency plot (blue)

More likely to be the actual underlying relationship, but was further from the data then the cubic model, and had the same problem that the curve decreased after t = 300 and so the model cannot be extrapolated.

It was then that it hit me. Could it be that the number of kernels that popped in each time interval followed a normal distribution?

***<u>Cumulative Normal Distribution Model</u>***

Recalling the chapter on probability distributions, I noticed that the plot very closely resembled the ogive of a cumulative distribution function for a normal distribution. Despite some skew on the graph, I felt that this was likely the actual relationship because of the nature of popping popcorn and because the data fit many of the characteristics a normal distribution – for example, the true x-variable, time, is continuous just as it is in the normal distribution. The histogram plotted also resembles a bell curve that is approximately symmetrical about a mean value, indicating a normal distribution.

The expression for a normal cdf (cumulative distribution function) was not on our course, but it was known that it could be found by integrating the normal equation (the cdf of a value is equal to the area under the pdf (probability density function) between zero and that value). After some research, I came across the error function which describes a related expression: $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$

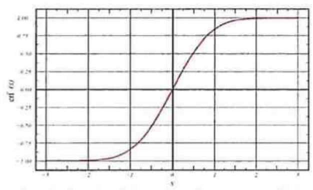

Fig. 7 Graph of the error function, erf (x), image from Wikipedia

The error function has a sigmoid (s) shape, just as in the cdf of the normal distribution, and approximately resembled the shape of the data plot.

Its relationship to the standardised cdf of a normal distribution is $F(z) = a\left[1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right]$, where F(z) is the curve of the cdf, a is the vertical scale factor, and $z = \frac{x - \mu}{\sigma}$, the standardized normal variable, where x is a specific x value, μ is the mean, and σ is the standard deviation in a normal distribution ($X \sim N(\mu, \sigma^2)$).

With this knowledge, it meant that I could simply find the mean, μ, and standard deviation, σ, of the data and find z in terms of x and set it as the input for the error function to hopefully produce a curve to model the plot. μ would be the horizontal translation factor, and σ would be the horizontal stretch factor.

| Time | Midpoint $m$ | Frequency $f(m)$ | $m_i \times f(m_i)$ | $(m_i - \mu)^2$ | $(m_i - \mu)^2 \times f(m_i)$ |
|---|---|---|---|---|---|
| 0-10 | 5 | 0 | 0 | 22212 | 0 |
| 10-20 | 15 | 1 | 15 | 19331 | 19331 |
| 20-30 | 25 | 0 | 0 | 16650 | 0 |
| 30-40 | 35 | 1 | 35 | 14170 | 14170 |
| 40-50 | 45 | 1 | 45 | 11889 | 11889 |
| 50-60 | 55 | 1 | 55 | 9808 | 9808 |
| 60-70 | 65 | 2 | 130 | 7927 | 15855 |
| 70-80 | 75 | 3 | 225 | 6247 | 18740 |
| 80-90 | 85 | 3 | 255 | 4766 | 14298 |
| 90-100 | 95 | 3 | 285 | 3485 | 10456 |
| 100-110 | 105 | 4 | 420 | 2405 | 9618 |
| 110-120 | 115 | 4 | 460 | 1524 | 6095 |
| 120-130 | 125 | 4 | 500 | 843 | 3372 |
| 130-140 | 135 | 5 | 675 | 362 | 1812 |
| 140-150 | 145 | 5 | 725 | 82 | 408 |
| 150-160 | 155 | 5 | 775 | 1 | 5 |
| 160-170 | 165 | 6 | 990 | 120 | 721 |
| 170-180 | 175 | 5 | 875 | 439 | 2197 |
| 180-190 | 185 | 6 | 1110 | 959 | 5753 |
| 190-200 | 195 | 5 | 975 | 1678 | 8390 |
| 200-210 | 205 | 6 | 1230 | 2597 | 15584 |
| 210-220 | 215 | 4 | 860 | 3717 | 14866 |
| 220-230 | 225 | 4 | 900 | 5036 | 20143 |
| 230-240 | 235 | 2 | 470 | 6555 | 13110 |
| 240-250 | 245 | 1 | 245 | 8274 | 8274 |
| 250-260 | 255 | 1 | 255 | 10194 | 10194 |
| 260-270 | 265 | 0 | 0 | 12313 | 0 |
| 270-280 | 275 | 1 | 275 | 14632 | 14632 |
| 280-290 | 285 | 0 | 0 | 17152 | 0 |
| 290-300 | 295 | 0 | 0 | 19871 | 0 |
| **Totals** | | $\Sigma f(m_i) = 83$ | $\Sigma m_i \cdot f(mi) = 12785$ | $\Sigma (m_i - \mu)^2 \cdot f(mi) = 249722.9$ | |
| | | **Mean (μ)** | $\frac{12785}{83} = 154$ | **Variance ($\sigma^2$)** | $\frac{249722.9}{83} = 3083$ |
| | | | | **Standard Deviation (σ)** | $\sqrt{3083} = 55.5$ |

**Table 2** Calculation of mean and standard deviation

Accuracy of calculations kept to a large multiple decimal places, calculated values shown to 1 (for final calculated values) or 0 (for non-final calculated values) decimal places.

Substituting in the resulting values for μ and σ into $z = \frac{x - \mu}{\sigma}$ yields $z = \frac{x - 154}{55.5}$

Which when substituted into F gives $F(z) = a\left[1 + \text{erf}\left(\frac{x - 154}{55.5\sqrt{2}}\right)\right]$

Since the original graph of erf(z) has a range 2 and the plot has a range of 83, the vertical scale factor, a, will be equal to $\frac{83}{2}$, yielding the result of $F(t) = \frac{83}{2}\left[1 + \text{erf}\left(\frac{t - 154}{55.5\sqrt{2}}\right)\right]$.

This gives the graph:

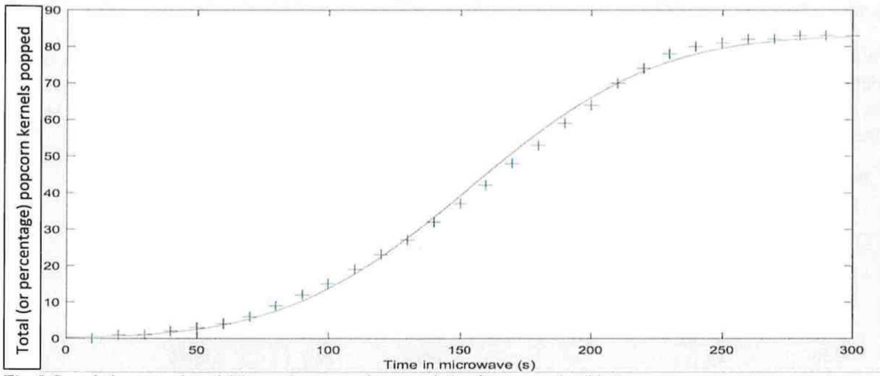


**Fig. 8** Cumulative normal model (orange) compared to cumulative frequency plot (blue)

This graph seemed to fulfil all the expectations and requirements for a model for the amount of popcorn popped over a certain amount of time and could be used as a model for the relationship. In addition to having the correct asymptote at y = 83, the curve is very close to the data points, with only some small discrepancy in the right portion where the data is first lower, then higher than the curve. A possible explanation for this slight skew is that the kernels are still heating up to popping temperature prior to approximately 120 seconds. Nevertheless, the normal cdf curve seems like a good model for the ogive.

## Analysis and Conclusion

To compare the goodness of fit of each of the models, the regression analysis method of looking at residuals was chosen. This would allow how far each value predicted by the model is from the observed value for each time to be seen. The residual, r is defined as: (observed value – predicted value). To find overall how far the data is from the model, the residuals cannot simply be added because they have different signs. To mitigate against this, the r value was first squared and the values for $r^2$ were added to find a value which we will call R. The calculations are summarised on the tables below. Figures are shown correct to 2 decimal places.

| Data | 1. Exponential model | | | 2. Cubic model | | | 3. Sine model | | | 4. Cumulative Normal model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Value | r | $r^2$ | Value | r | $r^2$ | Value | r | $r^2$ | Value | r | $r^2$ |
| **0** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | -0.23 | 0.05 |
| **1** | 12.53 | -11.53 | 132.95 | 0.27 | 0.73 | 0.53 | 0.23 | 0.77 | 0.60 | 0.39 | 0.61 | 0.37 |
| **1** | 23.17 | -22.17 | 491.48 | 1.06 | -0.06 | 0.00 | 0.91 | 0.09 | 0.01 | 0.65 | 0.35 | 0.12 |
| **2** | 32.20 | -30.20 | 912.16 | 2.32 | -0.32 | 0.10 | 2.03 | -0.03 | 0.00 | 1.06 | 0.94 | 0.89 |
| **3** | 39.87 | -36.87 | 1359.46 | 4.03 | -1.03 | 1.07 | 3.59 | -0.59 | 0.35 | 1.66 | 1.34 | 1.80 |
| **4** | 46.38 | -42.38 | 1796.24 | 6.15 | -2.15 | 4.61 | 5.56 | -1.56 | 2.43 | 2.53 | 1.47 | 2.16 |
| **6** | 51.91 | -45.91 | 2107.76 | 8.63 | -2.63 | 6.93 | 7.93 | -1.93 | 3.71 | 3.75 | 2.25 | 5.07 |
| **9** | 56.60 | -47.60 | 2266.13 | 11.45 | -2.45 | 5.99 | 10.66 | -1.66 | 2.75 | 5.40 | 3.60 | 12.95 |
| **12** | 60.59 | -48.59 | 2360.89 | 14.56 | -2.56 | 6.55 | 13.73 | -1.73 | 3.00 | 7.57 | 4.43 | 19.62 |
| **15** | 63.97 | -48.97 | 2398.29 | 17.93 | -2.93 | 8.57 | 17.11 | -2.11 | 4.44 | 10.33 | 4.67 | 21.84 |
| **19** | 66.84 | -47.84 | 2289.14 | 21.52 | -2.52 | 6.34 | 20.75 | -1.75 | 3.06 | 13.72 | 5.28 | 27.89 |
| **23** | 69.28 | -46.28 | 2142.20 | 25.29 | -2.29 | 5.26 | 24.62 | -1.62 | 2.63 | 17.76 | 5.24 | 27.48 |
| **27** | 71.35 | -44.35 | 1967.33 | 29.22 | -2.22 | 4.91 | 28.68 | -1.68 | 2.81 | 22.42 | 4.58 | 21.02 |
| **32** | 73.11 | -41.11 | 1690.25 | 33.25 | -1.25 | 1.56 | 32.87 | -0.87 | 0.76 | 27.62 | 4.38 | 19.23 |
| **37** | 74.61 | -37.61 | 1414.17 | 37.36 | -0.36 | 0.13 | 37.16 | -0.16 | 0.03 | 33.24 | 3.76 | 14.17 |
| **42** | 75.87 | -33.87 | 1147.36 | 41.50 | 0.50 | 0.25 | 41.50 | 0.50 | 0.25 | 39.12 | 2.88 | 8.32 |
| **48** | 76.95 | -28.95 | 838.03 | 45.64 | 2.36 | 5.55 | 45.84 | 2.16 | 4.67 | 45.07 | 2.93 | 8.57 |
| **53** | 77.86 | -24.86 | 618.13 | 49.75 | 3.25 | 10.56 | 50.13 | 2.87 | 8.25 | 50.92 | 2.08 | 4.35 |
| **59** | 78.64 | -19.64 | 385.65 | 53.78 | 5.22 | 27.21 | 54.32 | 4.68 | 21.86 | 56.46 | 2.54 | 6.44 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **64** | 79.30 | -15.30 | 233.98 | 57.71 | 6.29 | 39.61 | 58.38 | 5.62 | 31.59 | 61.56 | 2.44 | 5.94 |
| **70** | 79.86 | -9.86 | 97.13 | 61.48 | 8.52 | 72.57 | 62.25 | 7.75 | 60.06 | 66.10 | 3.90 | 15.20 |
| **74** | 80.33 | -6.33 | 40.07 | 65.07 | 8.93 | 79.71 | 65.89 | 8.11 | 65.72 | 70.01 | 3.99 | 15.91 |
| **78** | 80.73 | -2.73 | 7.47 | 68.44 | 9.56 | 91.37 | 69.27 | 8.73 | 76.23 | 73.27 | 4.73 | 22.34 |
| **80** | 81.08 | -1.08 | 1.16 | 71.55 | 8.45 | 71.37 | 72.34 | 7.66 | 58.67 | 75.91 | 4.09 | 16.74 |
| **81** | 81.37 | -0.37 | 0.13 | 74.37 | 6.63 | 43.98 | 75.07 | 5.93 | 35.12 | 77.97 | 3.03 | 9.19 |
| **82** | 81.61 | 0.39 | 0.15 | 76.85 | 5.15 | 26.50 | 77.44 | 4.56 | 20.79 | 79.53 | 2.47 | 6.11 |
| **82** | 81.82 | 0.18 | 0.03 | 78.97 | 3.03 | 9.20 | 79.41 | 2.59 | 6.70 | 80.67 | 1.33 | 1.77 |
| **83** | 82.00 | 1.00 | 1.00 | 80.68 | 2.32 | 5.40 | 80.97 | 2.03 | 4.13 | 81.48 | 1.52 | 2.31 |
| **83** | 82.15 | 0.85 | 0.72 | 81.94 | 1.06 | 1.12 | 82.09 | 0.91 | 0.82 | 82.04 | 0.96 | 0.93 |
| **83** | 82.28 | 0.72 | 0.52 | 82.73 | 0.27 | 0.07 | 82.77 | 0.23 | 0.05 | 82.41 | 0.59 | 0.35 |
| | | | **26699.99** | | | **537.03** | | | **421.48** | | | **299.12** |

**Table 3** Calculation of R for each model

The R values decrease from the exponential model to the cumulative normal model, suggesting that each model was an improvement over the last in terms of goodness of fit. The cumulative normal model has a R value of 299.12, a relatively low value, suggesting that this is comparatively the best model.

To investigate further whether the data followed a normal distribution, the normal curve could be found and then compared to the histogram of the observed data. The equation of the normal-shaped curve could be found by using the data to assume points on the curve to form multiple simultaneous equations from the general formula for a normal curve and then solve for the constants of the formula, but another method of simply substituting figures into the formula for a normal distribution curve found on the HL Data Booklet was deemed to be more efficient: $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

$X \sim N(154, 3083) \rightarrow \left(\frac{1}{55.5\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-154}{55.5}\right)^2}\right)q$, where q is the vertical stretch factor.

To approximate q, the approximate value of the pdf function was found by finding the value of the function for each value of t. The maximum was found to be (150, 0.004045). In the histogram in Fig. 1, the highest number of pops in an interval seems to be 6. 6 ÷ 0.004045 = q = 1483.33, giving the expression $\left(\frac{1}{55.5\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{t-154}{55.5}\right)^2}\right)1483.33$.

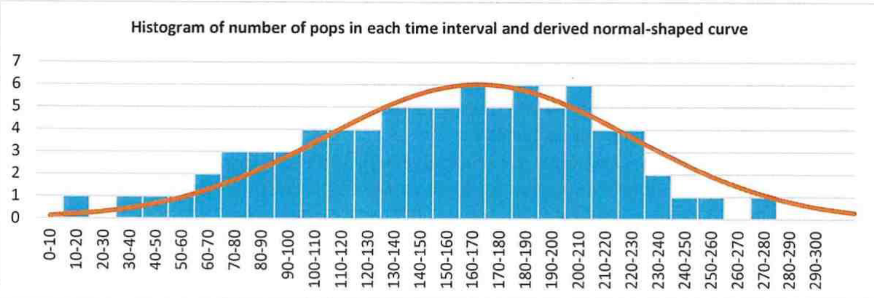


**Fig. 9** Histogram of data compared to normal curve of model

By eye, it seems that despite some negative skew, the calculated curve models the histogram quite well which also supports the conjecture that the data does follow a normal distribution. For a more reliable indication however, a quantitative test would have to be used. After some research, various normality tests were found. These are used to determine whether a data set is well modelled by a normal distribution. It just so happened that one such test was to simply compare the histogram to the normal curve. Another test however was to compare the data to the empirical (68–95–99.7) rule.

- 68% of the data should lie between one standard deviation either side of the mean:
  Between $\mu - \sigma \approx 100$ and $\mu + \sigma \approx 210$, there were 55 pops. 55 ÷ 83 ≈ 0.66 = 66%.

- 95% should lie between two standard deviations either side of the mean:
  Between $\mu - 2\sigma \approx 40$ and $\mu + 2\sigma \approx 270$, there were 80 pops. $80 \div 83 \approx 0.96 = 96\%$.
- 99.7% should lie between three standard deviations either side of the mean:
  Between $\mu - 3\sigma \approx -10 \rightarrow 0$ and $\mu + 2\sigma \approx 320 \rightarrow 300$, there were 83 pops. $83 \div 83 \approx 1 = 100\%$.

The data very closely follows the empirical rule which is strong evidence that the popcorn followed a normal distribution. Despite being a completely different model to what I expected ("$P = a + be^{ct}$"), the model could still be used to fulfil the original aim of the investigation: to determine for how long I should keep my popcorn in the microwave.

From Fig. 8 (page 9), the optimal time can be decided qualitatively from three main factors:

1. Where the rate of popping begins to decrease – this corresponds to the point of inflection on the cdf (where the curve changes curvature from concave up to concave down)
   - The point of inflection occurs at approximately t = 170s
2. The actual rate of popping - indicated by the slope of the tangent to a point on the graph
   - The slope decreases to a particularly low value (approximately 0.2) at t = 240s
3. The percentage popped at each time - shown on the y-axis
   - Approximately 80 out of the 100 (80%) of the kernels are popped by t = 240s

Taking these three aspects into account, it seems most reasonable that the lowest acceptable time is approximately 240 seconds (4 minutes) – it is at this point where more than 4/5 of the kernels have popped and the rate of popping begins to slow drastically. In other words, after this point, additional time spent waiting on the popcorn will yield continually diminishing returns of popped kernels as well as anger other queueing microwave-goers.

I am also satisfied that this investigation has fulfilled personal aim of seeing how maths can be used in real-life. In this exploration, I have directly seen and participated in the use of integrals, continuous normal probability distributions, statistics, logs, and Euler's number. This, for me, has indeed helped to demystify the elusive uses of maths in our daily lives.

## Evaluation and Extensions

- The histogram cannot be compared directly to the vertically stretched normal curve. This is because the class width is not infinitely small (i.e. continuous) due to the nature of the way the data was collected. In a future experiment, my original proposed method of recording the exact time of each pop could be used to produce a series of continuous data.
- The R-squared test cannot reveal any biases in the data. This could be mitigated against by examining the residual plots for each model.
- Many of the comparisons are made on a qualitative basis (by visual inspection and comparison of patterns and trends). If the investigation were to be extended, a variety of complex analysis methods could be used.
- The law of large numbers states that the more data is collected, the closer the data will be to its true expected value. This means that the experiment could be repeated with more kernels to allow more resolution and precision in the data and lead to a more accurate model.
- Though not done in this investigation, the experiment could be repeated for different brands of popcorn, different powered microwaves, different ambient temperatures etc. and constants (stretch and shift factors) could be further derived to modify the expression for these different parameters to make the model more versatile and applicable in a wider variety of conditions.
- The concept of a skewed normal distribution was encountered during the investigation, though it was too advanced to include in this investigation. Over the course of my further study of maths, perhaps I could revisit this and produce an even more accurate model considering the skewing of data.
- Other optimisation techniques involving calculus could be used to determine the optimal time.

## BIBLIOGRAPHY


Abramowitz, Milton and Irene Ann Stegun. *Handbook Of Mathematical Functions With Formulas, Graphs, And Mathematical Tables*. 1st ed. New York: Dover Publications, 1970. Print.

"Error Function". *Mathworks*. Web. 22 Nov. 2016. <https://uk.mathworks.com/help/matlab/ref/erf.html>.

"Normality Test". *Originlab.com*. Web. 16 Dec. 2016. <http://www.originlab.com/doc/Origin-Help/Normality-Test>.

"The Law of Large Numbers". Math.uah.edu. Web. 17 Dec. 2016. <http://www.math.uah.edu/stat/sample/LLN.html>.

Wazir, Ibrahim. Mathematics. 1st ed. Harlow: Pearson Education, 2012. Print.

Weisstein, Eric. "Erf". *MathWorld--A Wolfram Web Resource*. Web. 21 Nov. 2016. <http://mathworld.wolfram.com/Erf.html>.

Weisstein, Eric. "Normal Distribution". *MathWorld--A Wolfram Web Resource*. Web. 12 Nov. 2016. <http://mathworld.wolfram.com/NormalDistribution.html>.

Weisstein, Eric. "Sigmoid Function". *MathWorld--A Wolfram Web Resource*. Web. 22 Nov. 2016. <http://mathworld.wolfram.com/SigmoidFunction.html>.