Carlo Sanabria
Mathematics SL Coursework

Population Trends in China

The aim of this investigation is to find out more about different functions that best model the population of China from 1950 to 1995.

The controlled variables would be the place where the population was measured (China) and the amount of years between each reading (5 years every time the population is given). The dependant variable is the population in China and the independent variable would be the year at which the population was recorded.

The parameters leading to the amount of Chinese people involve the family planning policy, this mean that from 1970 onwards, urban families would only be allowed to have one child. However the majority of the population would be exempt from the rule because they are either rural families, ethnic minorities or if either parent was a single child he or she would be allowed to have more than one child[1]. Fertility rate in China was 5 babies per woman until 1970, the new laws were implemented and there was a sharp reduction to 3 babies per woman in 1980. In 2008 the fertility rate was less than 2. The implementation of the restrictions helped the Chinese government to reduce an estimate of 400 million births since the policy was implemented. After the policy was implemented people in China looked for ways to only give birth to boys because they would be able to sustain the family, on the other hand girls wouldn't. This idealism affected the population trends, the ratio became for every 100 girls born 119 boys were born.

Birth rates in China are currently at 14 babies born for every 1000 and the death rates are currently at less than 21 babies dead for every 1000. These numbers have been improving year over year, this means that there is less infant mortality rate and births have decreased. Again, the major thing affecting these numbers is that the government has restricted the number of babies a family can have. Doing so has made the health care system be more effective which leads to more prevention of diseases and death.

The investigation consists in finding a model that fits as close as possible the data given. This can be useful because the model can determinate how, if the parameters stay the same, the population numbers will be like in the future. However as parameters do change the models that are found will not be very precise to what the future could be, the past that is given can be modeled but that doesn't mean it will be similar to it.

The data given shows, as a starting point the year 1950, it then continues for another 45 years of data which has been given in multiples of 5.

---

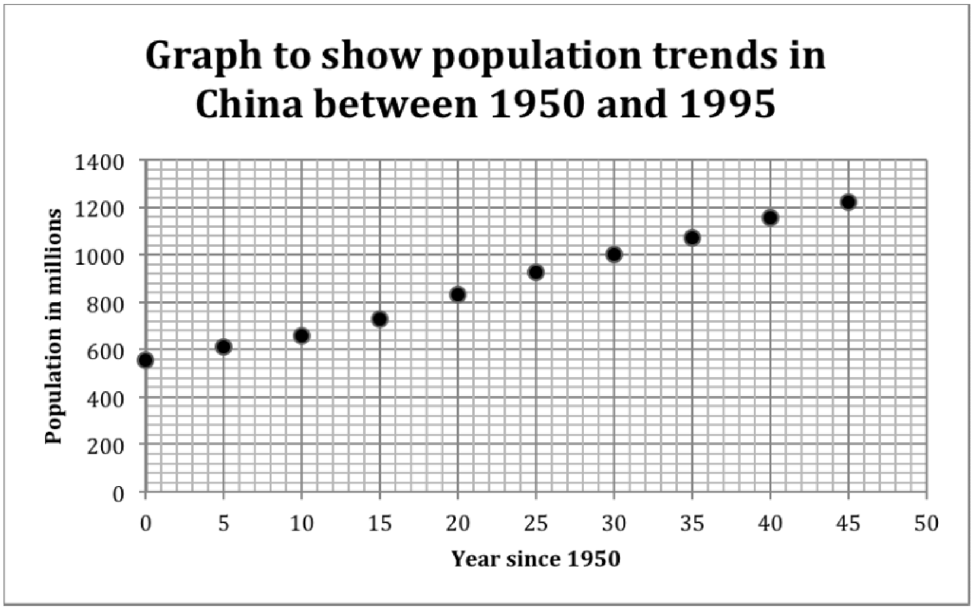
[1] Xiaofeng, Guan, "Most people free to have more child", In Chinadaily.com, 07/11/07, <http://www2.chinadaily.com.cn/china/2007-07/11/content_5432238.htm> (Accessed 26/09/10)

Year	Population in millions
1950	554.8
1955	609.0
1960	657.5
1965	729.2
1970	830.7
1975	927.8
1980	998.9
1985	1070.0
1990	1155.3
1995	1220.5

To make the investigation easier the data can be represented as where is the year given in the table. This will make it easier to manipulate the data and will give a graph starting at zero.

Year	Population in millions
0	554.8
5	609.0
10	657.5
15	729.2
20	830.7
25	927.8
30	998.9
35	1070.0
40	1155.3
45	1220.5

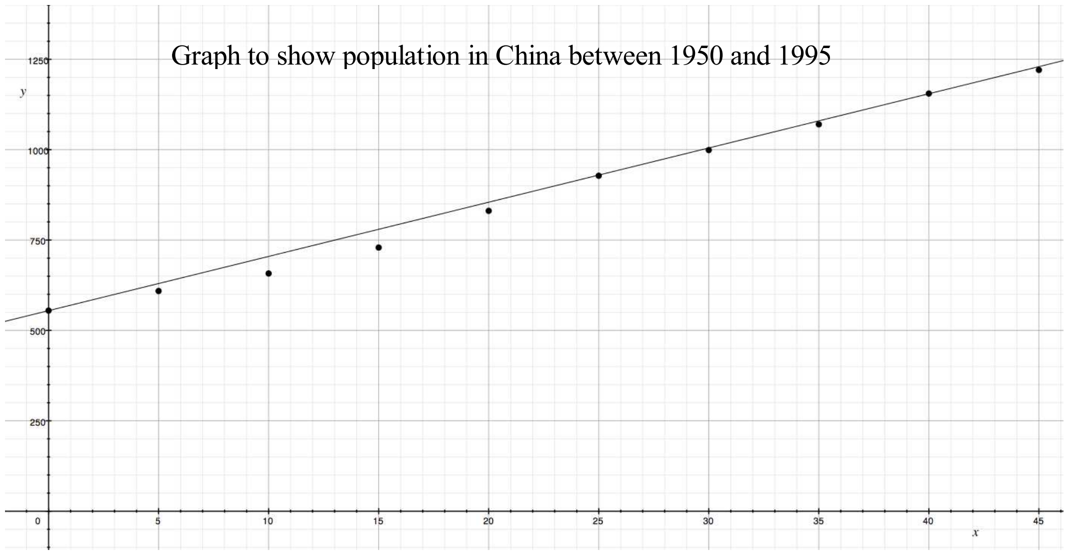
The graph for this data would be:



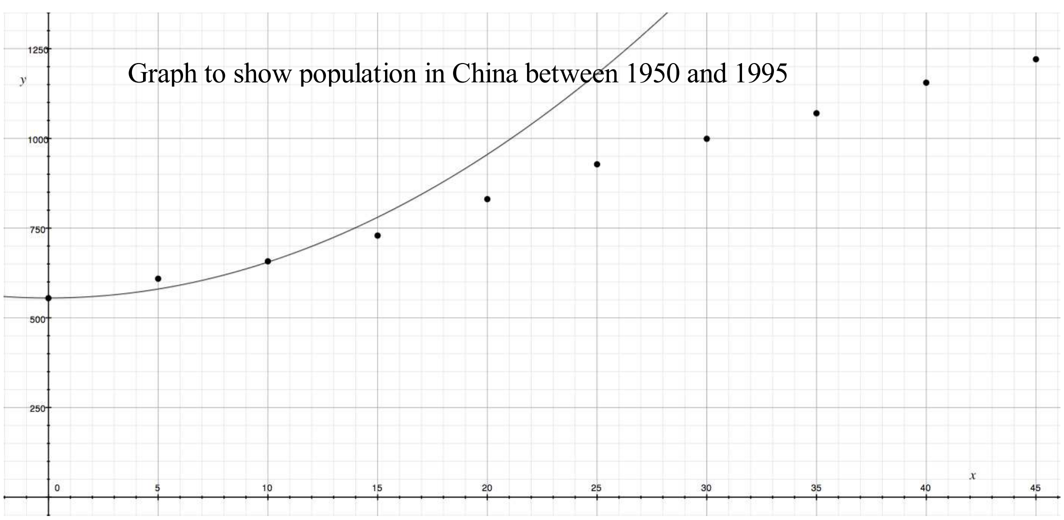
The graph above shows the trend that the population in China had, each of the axis is labled with it's repsective meaning but they are actually represented on the model by having for the years since 1950 and for the population in millions.

The models I will be researching will be linear lines, polynomial (2nd and 3rd degrees) and exponential curves. I will also use trigonometry to find if a or graph can be similar to the data points. The least likely model to fit the data will be a logarithmic curve, however it will also be investigated.

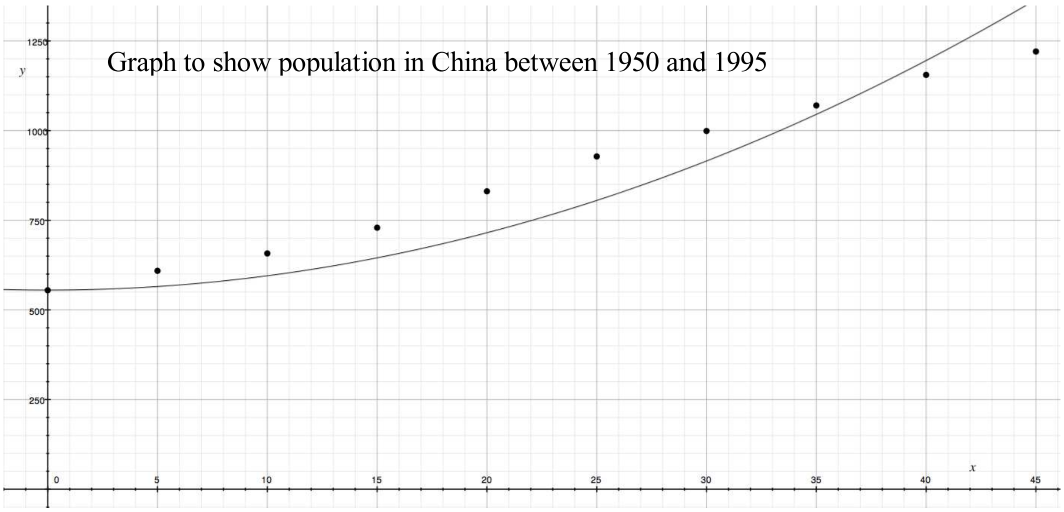
I predict that a linear graph won't suit the data points because the points are not linearly set. The or graphs might be able to represent the data points although further on in time the curve will drop meaning that the prediction is having a decrease in the population. A polynomial of 2nd degree will likely be similar in the way they behave as the trigonometric curves. They will drop down and in real life it wouldn't apply because China's population won't decrease at the rate the graphs for a quadratic, or would show. An exponential graph would only be able to represent the initial 20 years closely because further on it would increase at an even faster rate than that of the actual population.. The logarithmic curve isn't likely to fit the data because it has a "L" like shape meaning that the initial years would be very far away from the model. The model I will develop will depend on how well it seems to fit into the data point set.



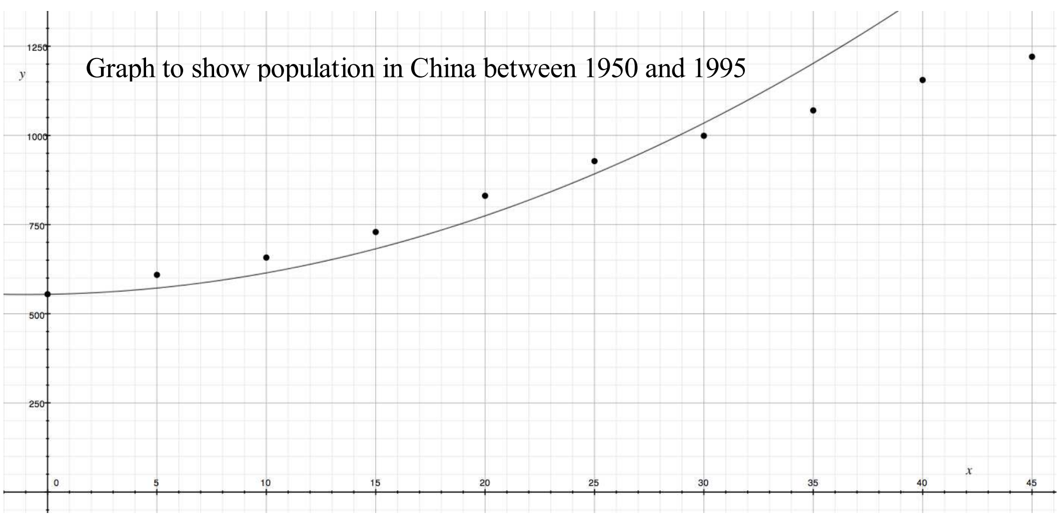
This model is , this is a linear equation, many points (represented by circles) are not close to the line and therefore the model is not likely to represent what happens next or what happened before because population doesn't normally increase constantly, this is the case for population in China. The number which represents is the only point which is exactly on the line, the year 1950 is represented as being the one on the and at this point in time there was a population in China of 554.8 million people. This is actually the closest a linear equation can ever get to the data therefore it can't be the model.



The first trial at getting a model from a quadratic curve wasn't at all successful. The current equation is , there are too many points of the data that means that there is no way that this can be the model, however the graph can still be manipulated to find a similar model. Again the in the equation is relevant because it was the population at the year where the data given starts.



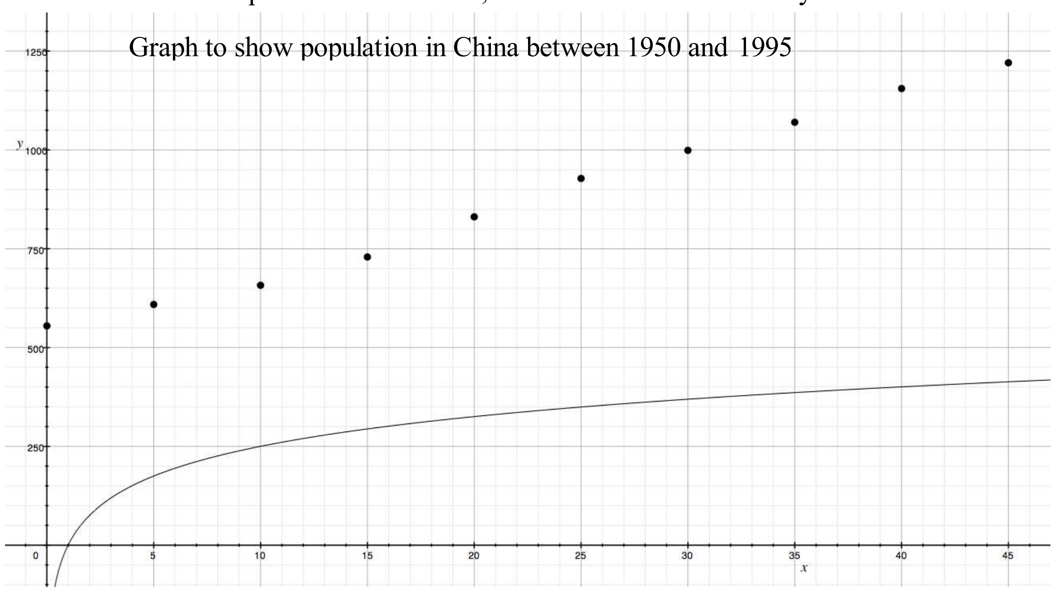
The curve for this model is represented by , it is the closest a quadratic curve can get before adding an with 1 as its power. Although any number of 's can be added or subtracted this won't change the way a quadratic curve is and therefore keeping on with this type of model wouldn't make much sense since the data has a "N" like shape. To prove that the quadratic curve won't change the following model is .



The trigonometric graphs can be modified to look like the data points, however as the years pass in the graph the curve for both and the curve will go down. In a way it can be somewhat realistic because if any of the two curves is extended by a lot it will mean that there is a certain point in time where people wont live in China for a particular reason which cant be predicted.

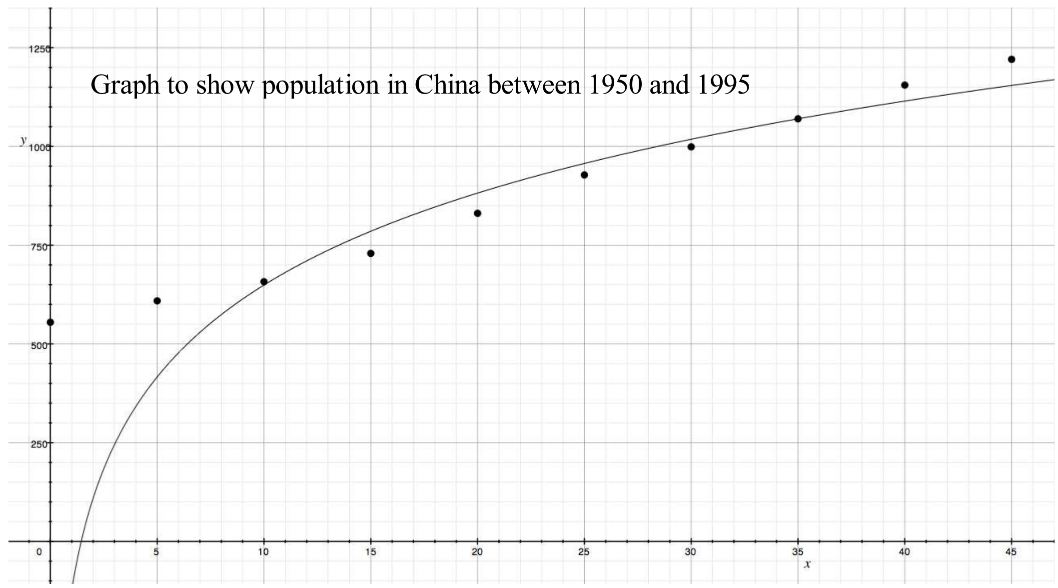
The following model is a logarithmic model, the normal curve is very curved and isn't at all what the data looks like in a graph. It starts increasing disproportionally, it then curves and starts to decrease the increase per year of population. It then stays at the same point which although somewhat realistic a population of a country wouldn't stay the same, it would either decrease (the case in some developed countries) or increase (the case in developing countries). This would be very similar to what the behavior of would be; the only difference would be that the curve of would take longer to stay at the same amount of population.

This model is of equation , the number 250 is there only to increase the



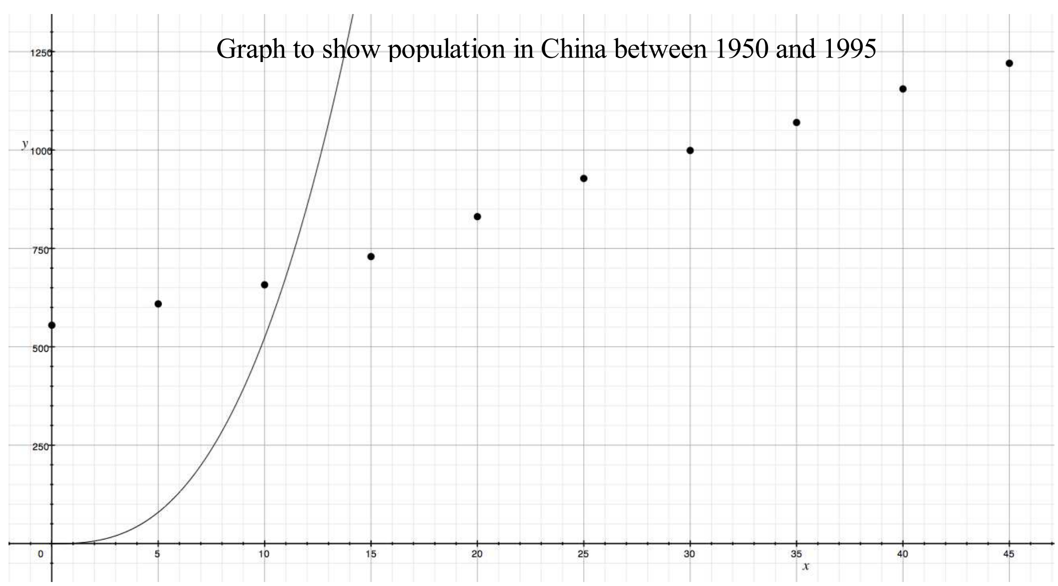
Graph to show population in China between 1950 and 1995

size of the curve and for it to be visible to such a scale. The actual model isn't very similar to the data and it can't be changed by a lot in terms of the shape that the curve has.

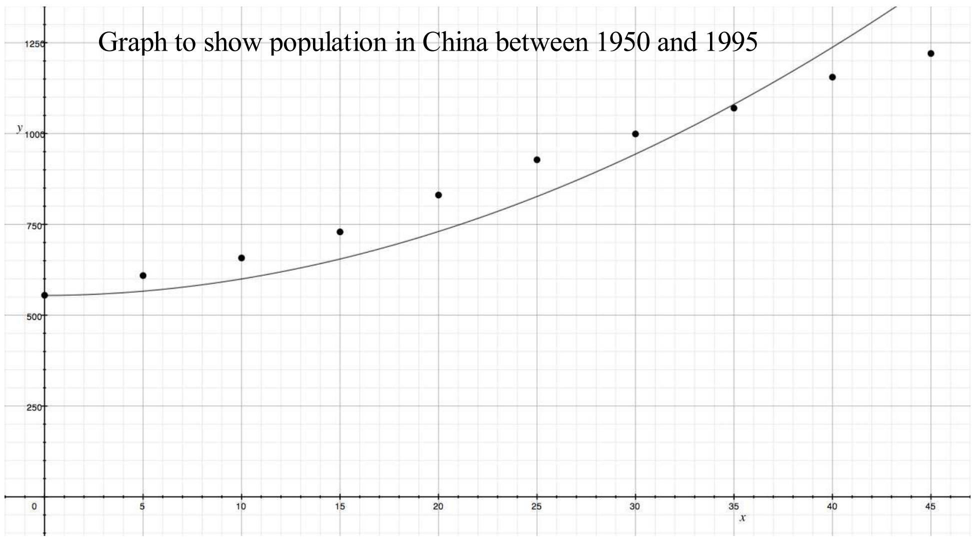


This model of equation is very similar to the data from the year 1960 to the year 1990, the other years are very far away to the curve and it can't represent what happened before 1960 because there wasn't such an increase in population in 10 years and the curve will never reach the year 1950 which means that there wasn't any population in China in the year 1950. This model isn't useful because it can't correctly represent the past and the future that it represents is not as realistic.

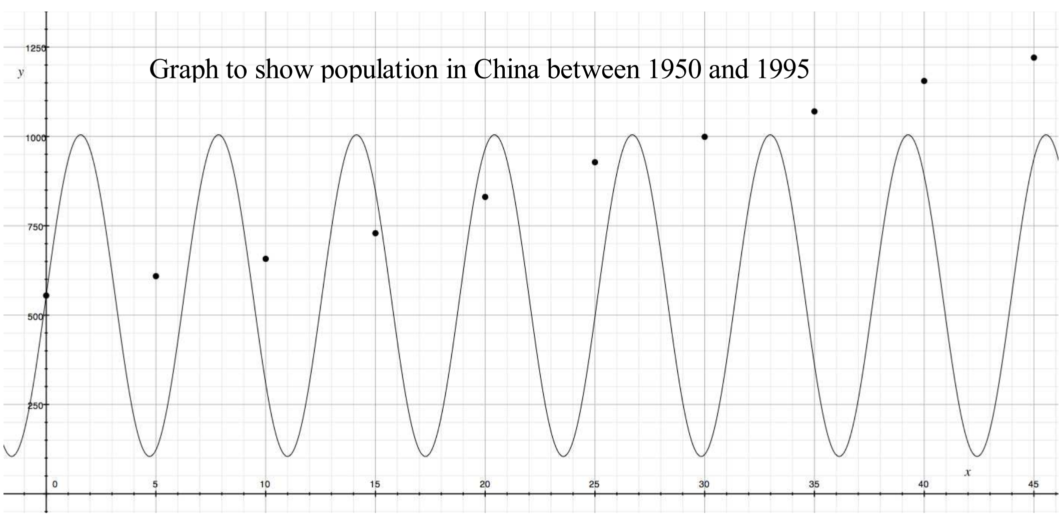
The following model is of the form to show an exponential graph, it can be relevant because that is the way the entire world's population is increasing. The Chinese government has restriction to control population; this means that the model would most likely not work in this situation because of the parameters set by their government.



The previous model is , it isn't in any way similar to the data and although manipulation can be done the shape of the curve resembles that of logarithmic curves because of its curvature.

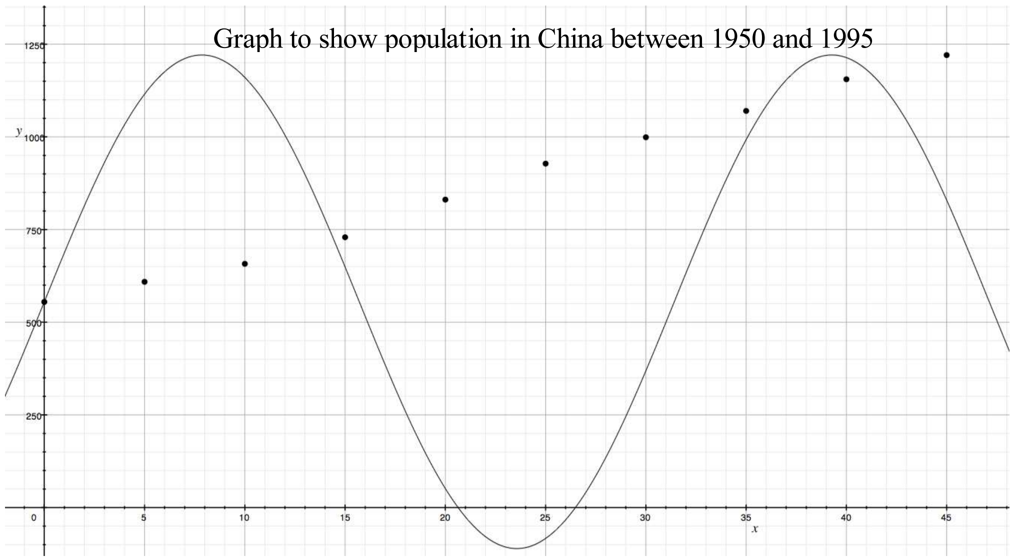


This model has an equation of and is very similar to the quadratic models. This means that as a model the exponential curve isn't suited for on certain country such as China although an exponent curve does represent the world's

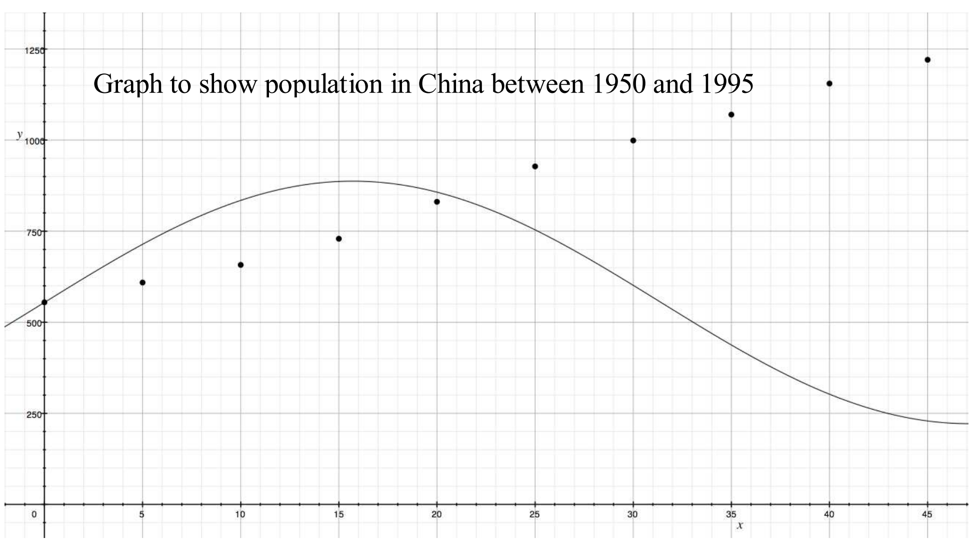


population growth.

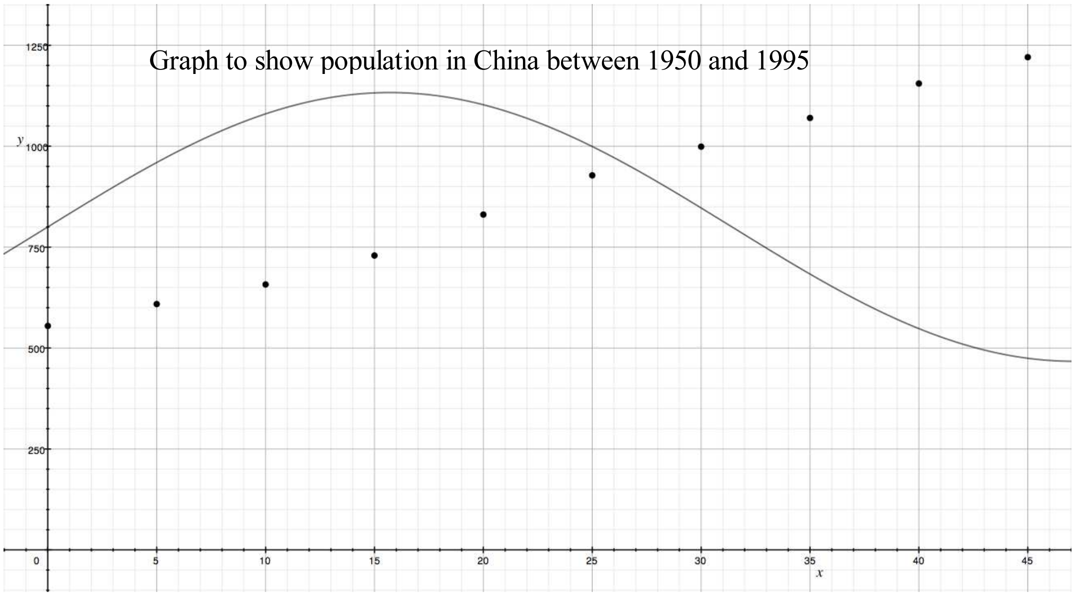
This is a curve and it doesn't resemble the points at all, this will be changed as much as possible to find a way to make the curve fit the data points. The equation for the current graph is which allows for much manipulation. This is the model I am going to develop as it closely associates itself to the data, not when is multiplied by one but by less than one.



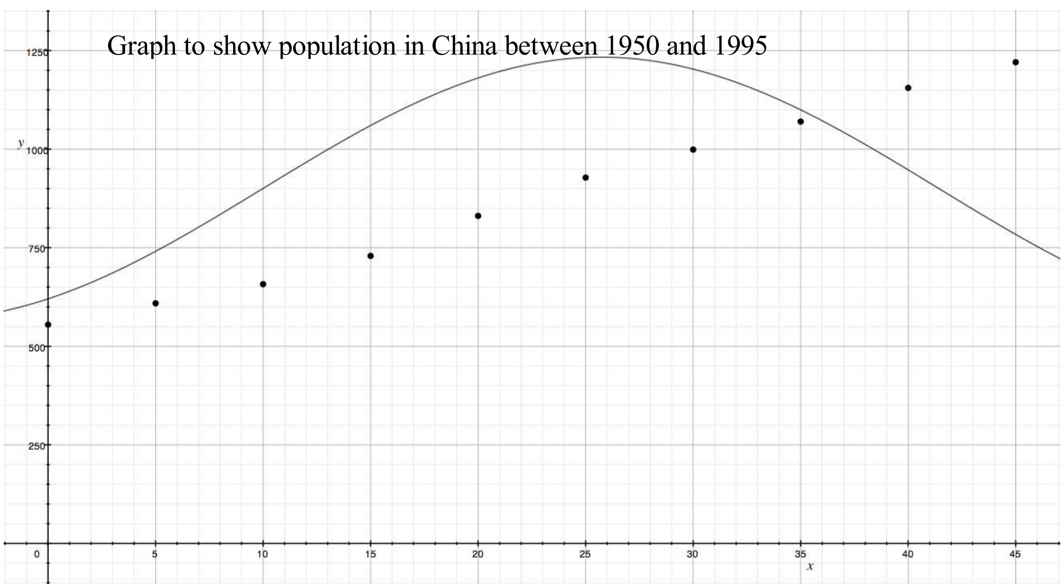
This curve is and there are two relevant numbers to what the investigation gives in terms of numbers, the in the equation is again the point at which the curve intersects the or the population axis. The number 665.7 comes from subtracting the population in the initial given year (1950) from the population in the final year, this is the amount of people which were born between those 45 years and to increase the amplitude of the wave it is put before . The amount of times the is multiplied changes the frequency and therefore a smaller frequency is needed. Currently it doesn't look very much alike the data. The amplitude of the curve is too big and should become much smaller, this means that the change in population doesn't really have an effect on the function.



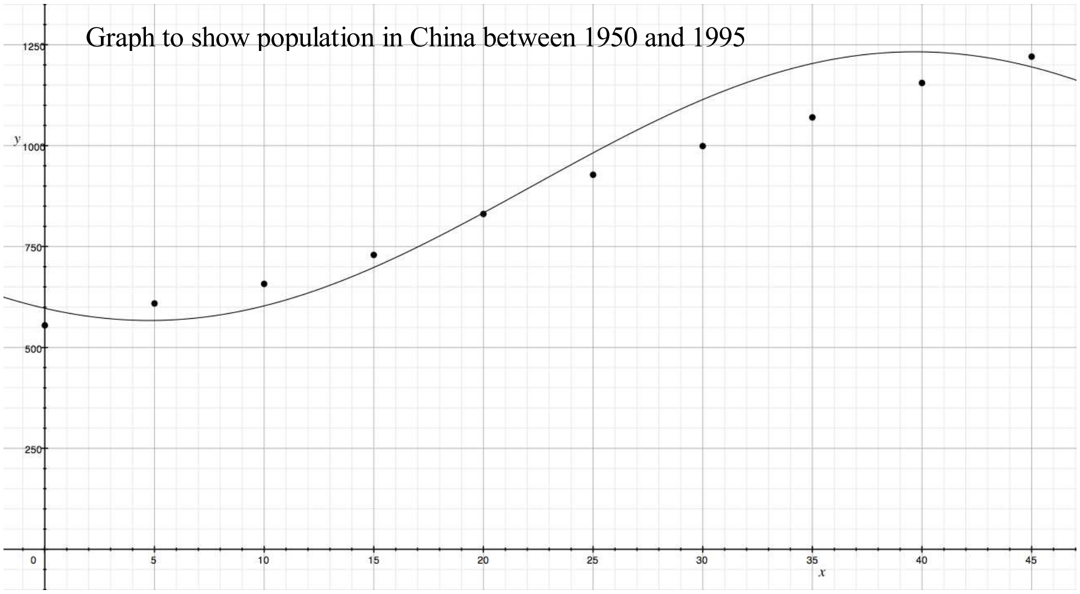
The equation for this curve is and the changes come from having very big amplitude and a lot of frequency. I halved both the factors that affect the amplitude and the frequency to 332.85 and 0.1 respectively.



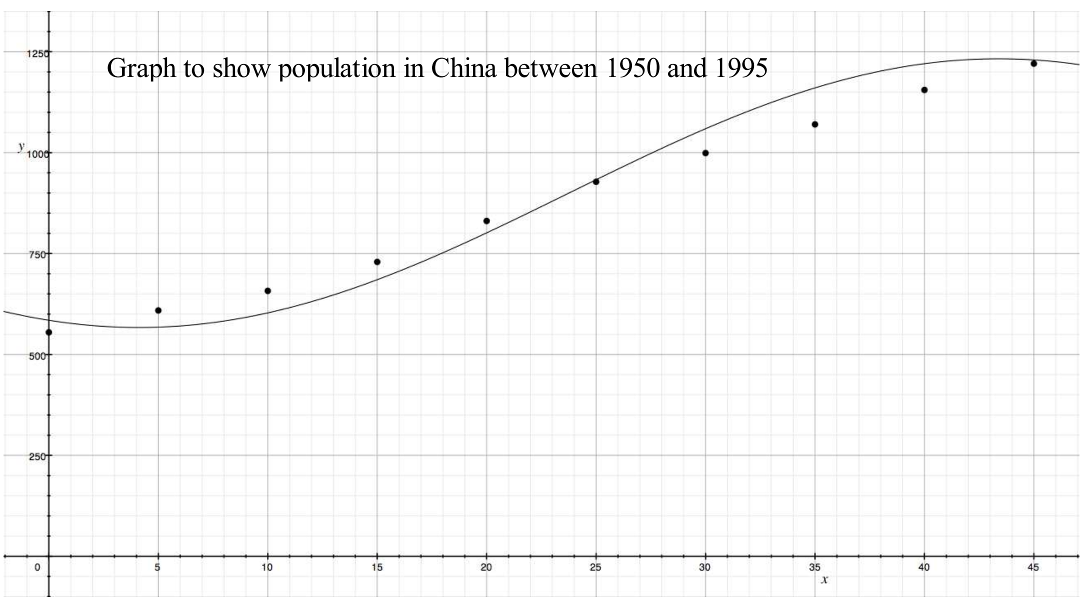
The sine curve starts to assimilate the points but it requires to be shifted towards the data points. The current equation is , from the previous graph I learnt that the curve had to be shifted upwards and to do so the had to be increased.



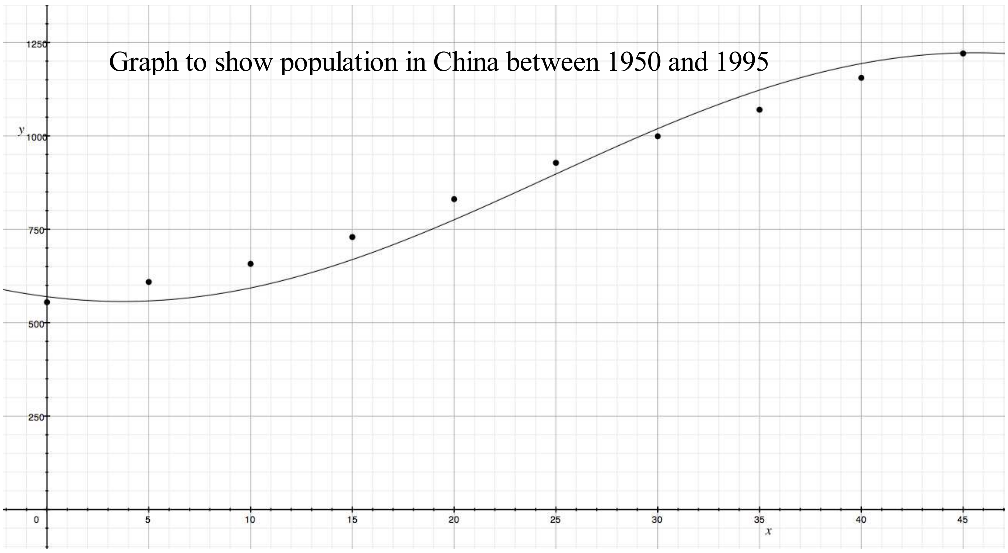
The shift of the curve was caused by having an equation with the parameters where is yet another constant which will create a shift in the sine curve. The formula for this curve is in which the major changes include the addition of to the equation as being -1 and the change in the constant, the curve wasn't near the point in the year 1995 which meant that another shift in the curve upwards had to be done.



The focus of this investigation is to get as close as possible a curve that will represent the data given. Therefore is a much better equation because the change in the multiplier of has to decrease because the amplitude at 0.1 was too big, it still is because it doesn't fit very closely. Doubling the constant has approached the cure to the 1995 point that I believe is crucial to having a "curve of best fit".

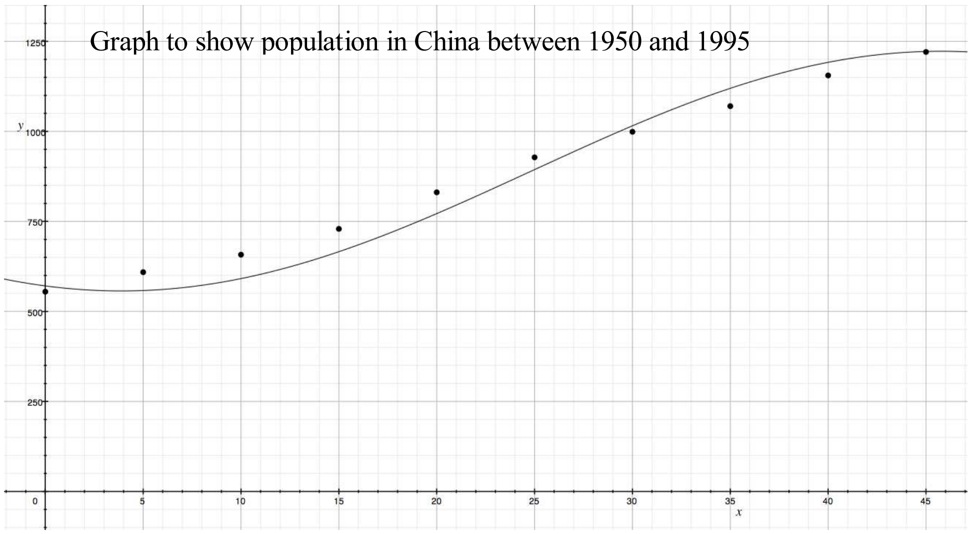


The curve has reached the 1995 population, however the curve of equation can be approximated to the data; mainly the data between the years 1955 to 1965 and 1985 to 1990 are somewhat further away from the curve drawn.



This curve resembles the data very closely; the equation is and if modified there will be more data points further away from the curve even though some will be closer. The first year's population is now much closer and the data between the years I had mentioned on the previous attempt are now closer, mainly the data between the years 1985 to 1990. The main changes are that most constants had to be decreased (except for), by doing so the line's amplitude and frequency decreased even more making it closer to the data points. Increasing constant meant that the curve shifted towards the right but only by the 0.05 that were added.

I decided to show what the curve would look like and it is very similar in terms of numbers in the equation and the curve itself. I decided to only show the closest result I could get because there are many similarities between the two trigonometry such as being the derivative of , both being involved in finding the hypotenuse of a right-angled triangle and finding sides or angles in non-right angled triangles.



As predicted the closest curve to the data is very similar to the curve above. The equation is and most numbers are very similar to the sine equation, the number that isn't is 3 digits bigger than that of the sine curve. The reason for showing this is I wanting to investigate both trigonometric curves; because the cosine curve is too similar to the sine curve I find no value in producing a model with .

The best curve to develop is the sine curve; I believe that by being shaped the way it is the points can be closest than to what other models were. The curve has one important factor to being similar to the data set because the population difference between 1950 and 1995 divided by 2 is the constant . This is more remarkable than for a straight line to cross exactly at the point where the data starts in the (554.8).

The biggest limitations that the sine curve model has are that it isn't correctly representing the past nor the future; the past has been represented by the curve with a negative gradient meaning that before there were more people in China, the future is represented as a deceleration of increase in population, at the year 1995 population starts to stay the same. The sine curve will go towards the , this means that in a near future to 1995 the population will decrease (China's population is to present day still growing). Another limitation is, as shown briefly before, how the curve can't reach the points between 1955 and 1965 very closely. This can't be improved because if it is done the curve will become further apart from other points.

The model where , and are constants can be developed by continuing the research with the base of . This means that is equal to as it did in the research for a curve which is close to the data of population between 1950 and 1995.

The first value of would be 0, this can be then substituted to get

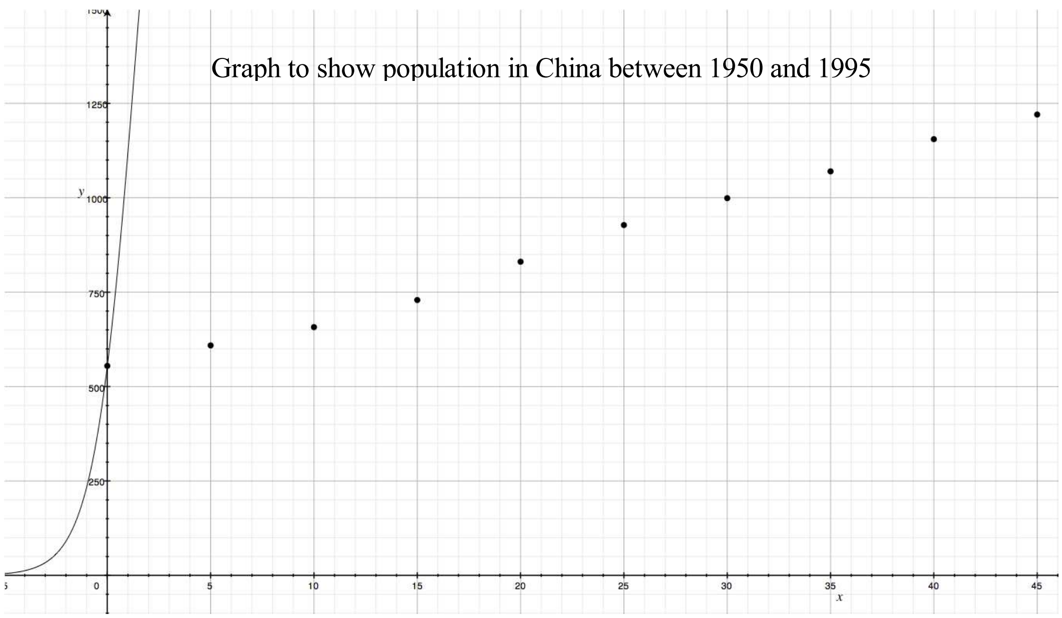
this then simplifies to (the is the same as).

The constants which are left when are and , this means that as best as possible.

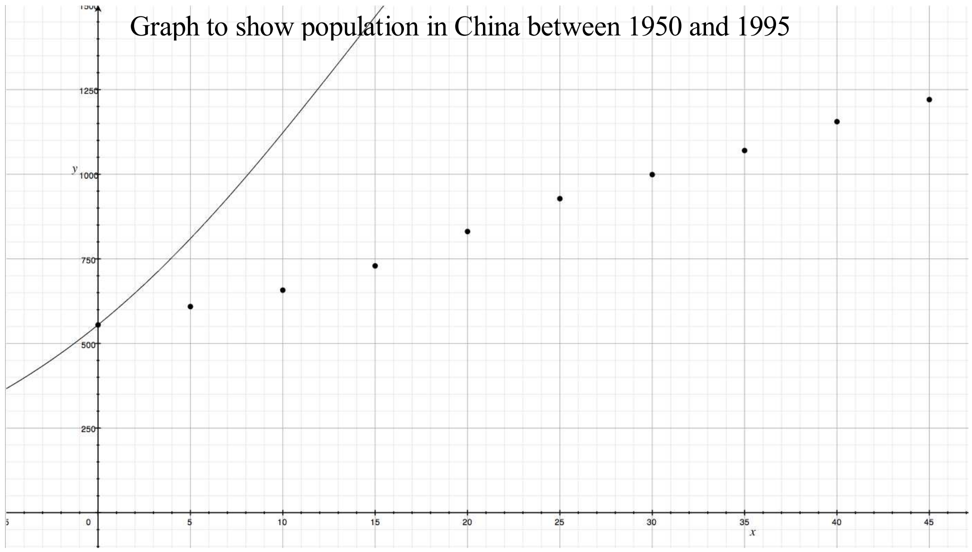
This being true means that and should be estimates.

If is equal to 2774 and is equal 4 then the equation results in 554.8, this was done by trial and error. These constants can be tested to see if they are useful when time has a value that isn't zero.

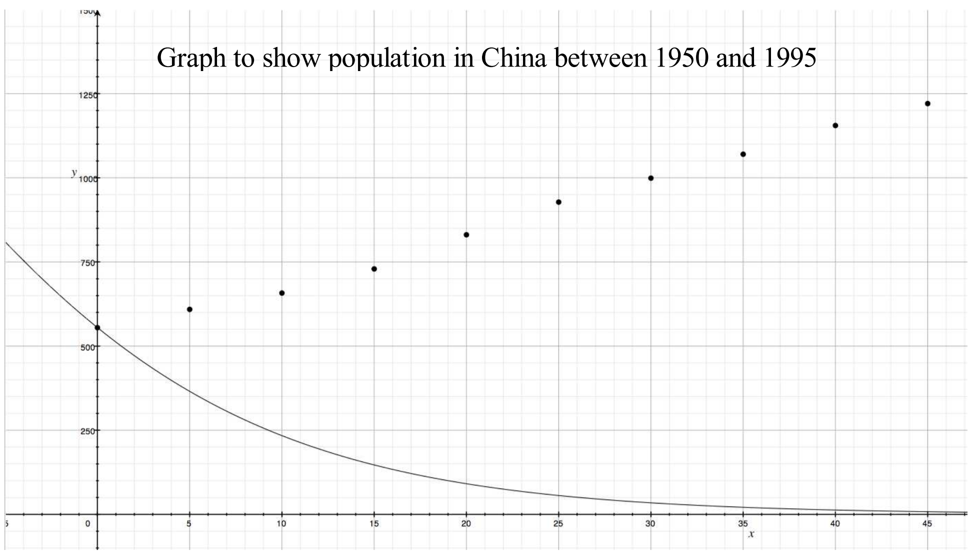
This equation will be a way to find out what the value of is, this way it will be possible to graph the equation. The time can be considered as so it can be graphed, as stated at the start of this investigation, the represent the year since 1950. The therefore represent time which means that .



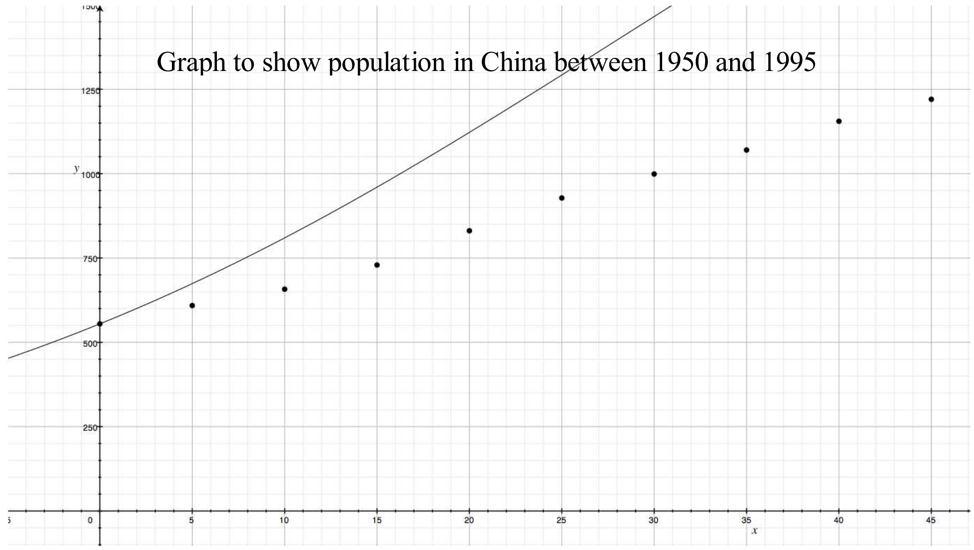
The value given to is 1, as it can be seen from the plotted graph the steepness of the curve is too steep. To be able to get a curve that assimilates the data pints it is necessary to decrease the gradient's value.



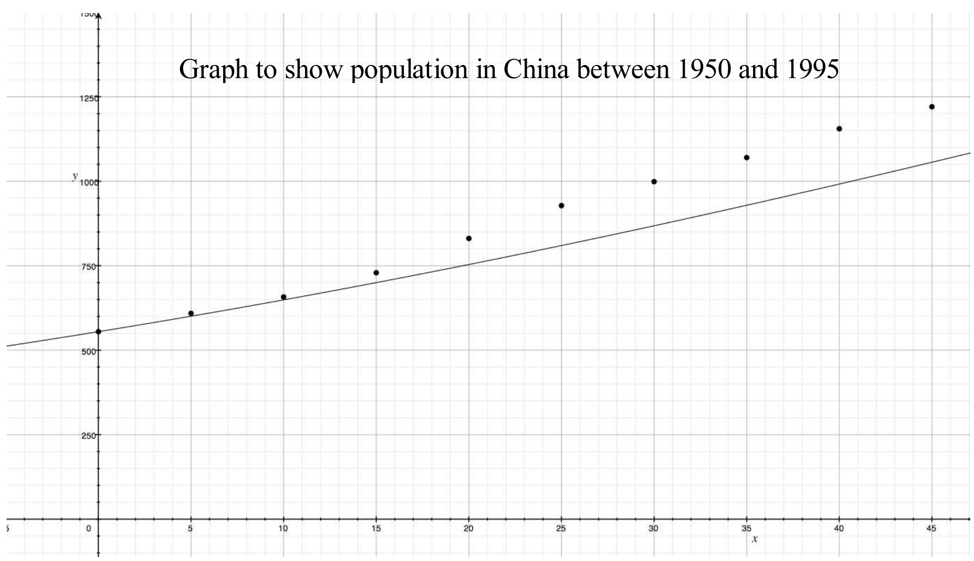
As predicted, decreasing the value of meant that the curve's steepness would decrease as well. Next I will try to have be of a negative value, this will negate the effect that the formula given by the researcher had on the value of the exponent.



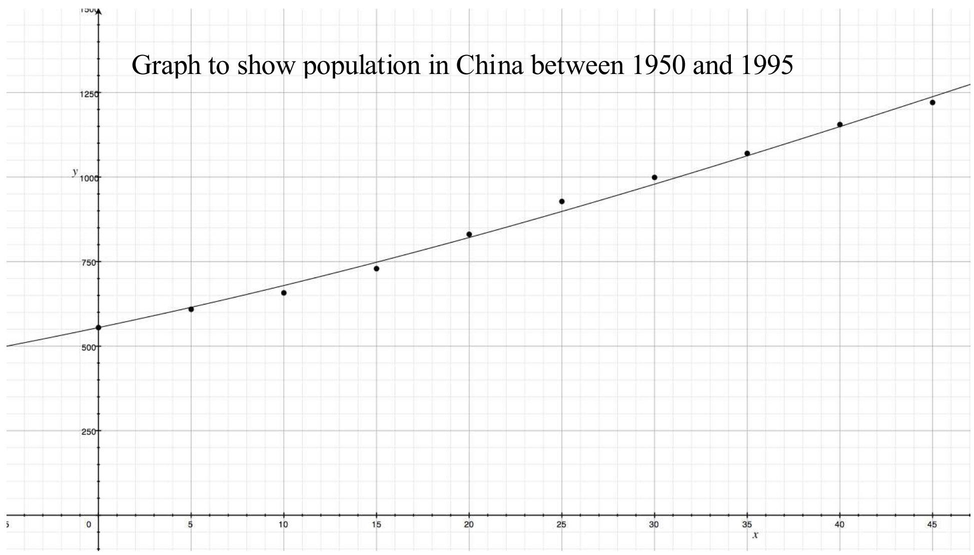
This time the value of is -0.1, as it can be seen changing the value of to a negative value will only turn the curve around. This means that for the curve to be similar to the data points . To follow this rule I will half the previous value of from 0.1 to 0.05.



In this model the is 0.05, this has made it become closer to the data points, it seems that it is rotating with the point 554.8 as its center. I believe this because every graph I have drawn has had the line cross through the point . This means that the population in the year 1950 is very important to the investigation because it is the first piece of data that is given.



This model has being equal to two; the curve has flattened a lot and is now past the data points. This means that the value of has to be more to enable the curve to b as close as possible to the points.



From graphing the model and trying to find a value for I found out that the closest the line can get to the points is when is equal to 0.026 and even when it is that close to the data it still has some gaps between the points and the line itself. The final equation of this model is or as it is represented by the researcher,

.

The researcher's model with the constants I put in is a good way to represent the population in China between the years 1950 and 1995. In most cases the points are very close to the graphed line. The model shows how the population increased correctly although it should be more curved order to be able to represent what actually happened. The model shows a reasonable past where population was less than in 1950. However the model isn't accelerating the population growth, this may or may not be true for the future but the population growth is almost never goes at a constant rate for a long period of time.

The model I developed which was has a relatively good way to present the data between the years 1950 and 1995, however the future, as I already mentioned, is not very realistic. If the function is seen with a bigger window then my model will start to decrease the population in a very short period of time. This is unlikely to happen, therefore the model is not very accurate when it has to deal with a year or more before nor a year or more after what the data given is.

The model given by the researcher and with my parameters of equation

has a very slight curvature that isn't very much alike what the real situation of other countries has been like. According to the research made in order to introduce the investigation the many restrictions imposed on the population has meant that this model is a good one but only for China because there is a way to keep an order in the population growth. Other countries' graphs would look more like a curve because the population in certain countries is going up very fast whilst in others it is either flat because it is approaching a decline or it is declining already. This particular concept of controlling the population isn't enrolled anywhere else and therefore I believe that the researcher's model for a middle term future is good enough. This model has a problem with the long-term future, by the year 2110 population will stay the same in China and will never stop being of the same amount of people. This is unrealistic which means that for a long-term future the model doesn't work properly.

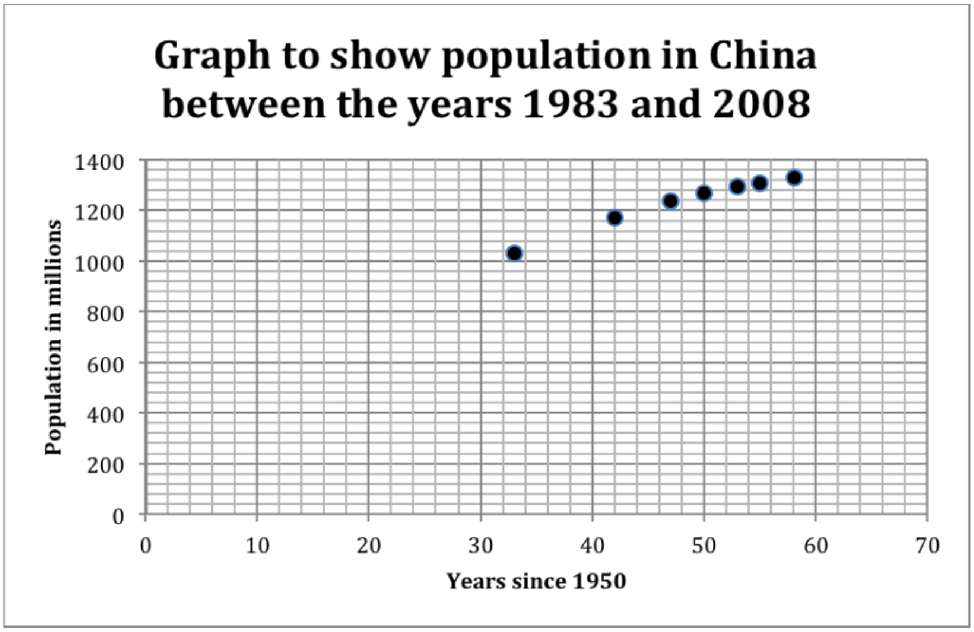
A new set of data is given

Year	Population in millions
1983	1030.1
1992	1171.7
1997	1236.3
2000	1267.4
2003	1292.3
2005	1307.6
2008	1327.7

To make the data easier to manipulate I will apply the same rule .

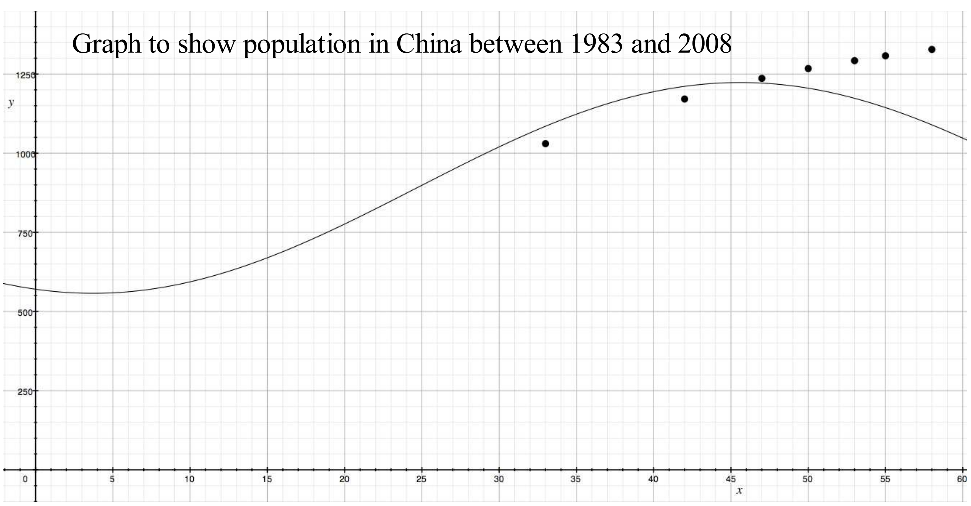
Years since 1950	Population in millions
33	1030.1
42	1171.7
47	1236.3
50	1267.4
53	1292.3
55	1307.6
58	1327.7

The data is easier to manipulate because the years are not the same size and this allows for better formulas and better graphing.

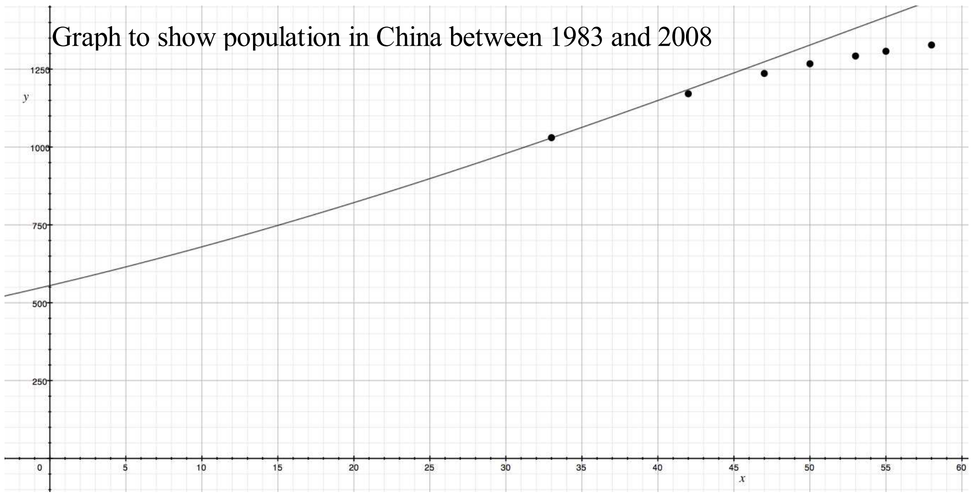


The graph above represents the data given for the years 1983 to 2008, each of the axes is labeled with years being time or more specifically the and the population is the .

Firstly I will put my model to the test to see if the new data is inside a range of my developed model .

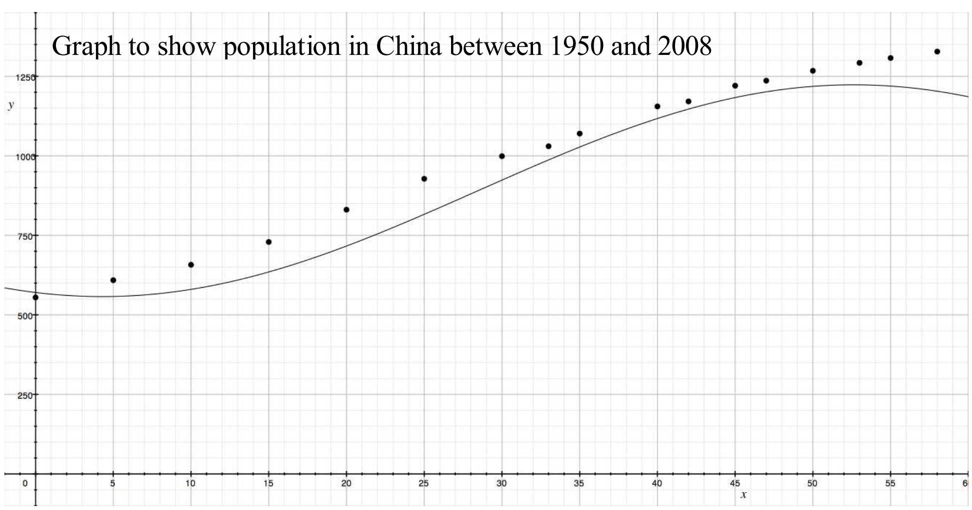


My model isn't similar to what most of the data is like, the initial there points which are close to a date of the previous set of data are in a good range from my curve, however the points from the year 2000 onwards are very distant and aren't a prediction of what could happen according to my model. In order to fit this data the frequency must be decreased even more and the amplitude increased so the curve reaches the population in the year 2008.



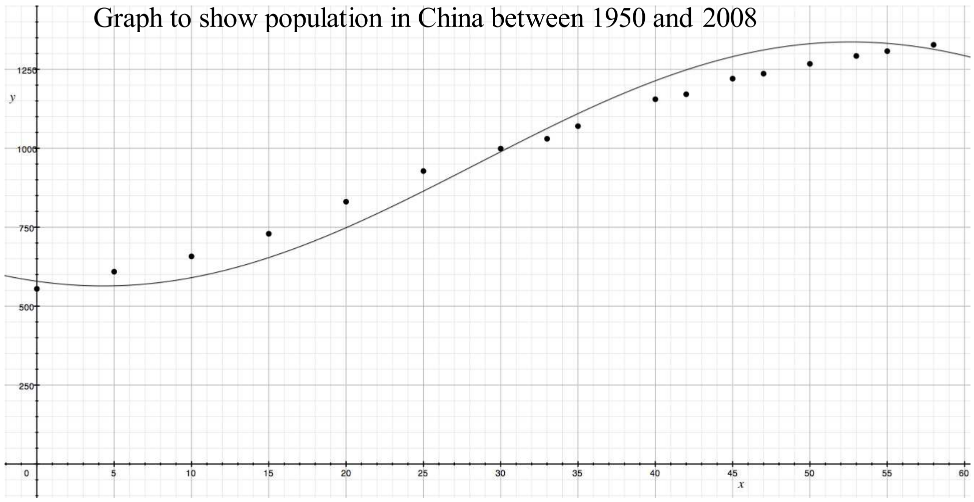
This model of formula isn't very accurate either; the points that are in a good range are as well years 1983 to 1997. Since the year 2000 onwards the model predicted that the population wouldn't decelerate its rate of increase, however it did and it is starting to flatten in a near future to 1995. In order to have the years 2000 to 2008 be closer to the model, the model would need to become a little curved such as a quadratic. This would enable the line to reach the actual population at that certain time.

I believe that my model would be easier to modify because my model did predict a flattening of the population growth rate in a near future, this means that the only thing that has to be done to my model is decrease the frequency of the sin curve to include the data from 1983 to 2008.

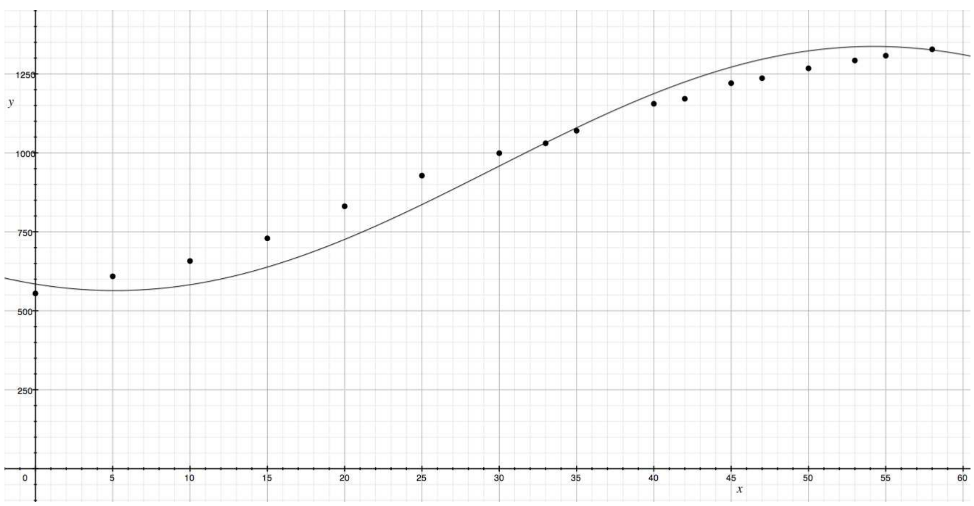


The frequency has been reduced from what I had in my original model by 0.01 units; the equation is now and still requires being closer to the population in the year 2008. This can be achieved by decreasing the frequency again and by increasing the amplitude.

Graph to show population in China between 1950 and 2008



The equation for this curve is , the two changes are increasing the amplitude by using the same method of the population in the first year given minus the population in the last year given, this divided by two which gives 386.45, the other change was the shift towards the 2008 population data point which needed an increase to 950 in order to be at a similar level to that point. I still need to decrease the frequency so that the maximum number of points are as close as possible to the curve.



This curve of best fit has the equation , the big change in the amplitude was enough to get the curve to reach the population data of the year 2008 as well most of the other data points. The change was a decrease from -1.85 to -1.9 and this also brought the curve closer to the data points after 1985. The change in the frequency of the curve wasn't very significant but was useful in order to get closer to the first data point in the year 1950. This is the closest this curve can get because if an attempt is made to get closer to the data points between 1960 and 1985 then the new data points from 2000 to 2008 would be further apart from the curve.

I believe that the limitations that this investigation has are very limited to China, the policies in place in China to control population growth are very strict and because that doesn't happen in other parts of the world a trend line similar to china's (not in numbers but in shape) might be difficult to encounter. A prediction of China's population in the future is more difficult to predict after the second set of data was revealed. I believe this because is the population were to keep growing at a certain rate then that would be what is expected from a normal country, China is definitely the exception and mainly because of the factors affecting population growth. This has happened because there was too may people and therefore an intervention must take place.

The Chinese government probably used a model of their own , they must have found out that if the growth of population went at the same rate as it was going at that time they wouldn't be able to control and govern that many people. I believe that is why the trend has a sudden drop in the increase of the population, the year 1970 was when the policies started to come into effect and 10 years later the population growth decelerated which made the anomalous curve of population in China between all those years.