

Mathematics Portfolio

Topic: Populations trend in China

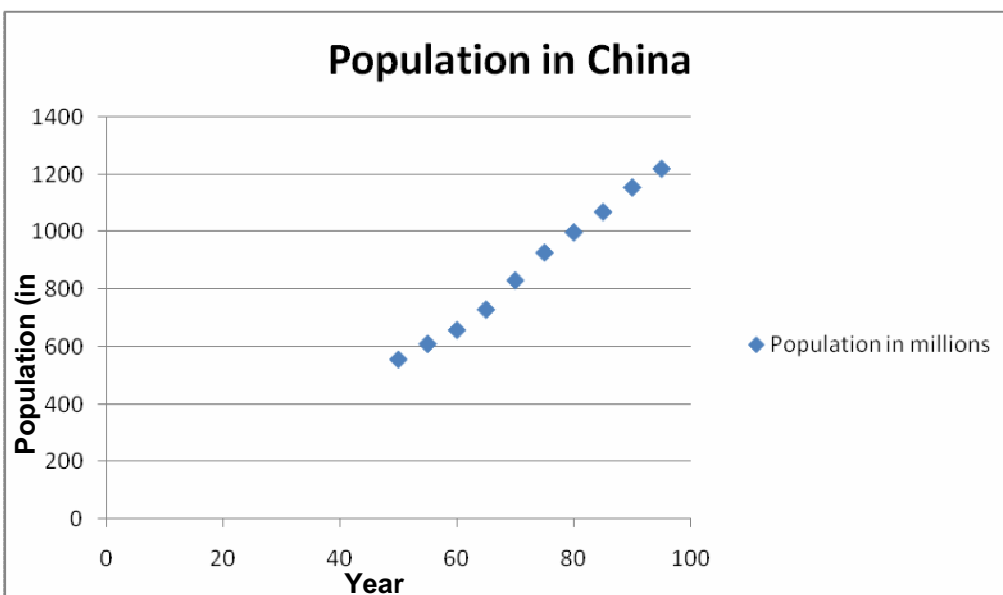
Date: 28.02.2012

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

In this portfolio, we should investigate and use mathematical functions and equations that would portray the model in the best way.

The relevant variables in this investigation are the population in millions in different years. The parameter is the initial population growth.

I have plotted the points given in the table above in the graph shown below.



**the year points of the picture above are replaced from 1950 to 50, 1960 to 60, etc. [plotted with Microsoft Excel 2007]*

We can clearly see from the graph above that the obvious trend occurring is the rise of the population over a period of time and we can see it is increasing gradually, so we could possibly present the model through a linear equation like: $Y = aX + b$ where a are the variables whilst b are the parameters.

Now I will try to develop the model using a linear function.

In order to start, we will firstly need to find the slope. And to find the slope, we divide the difference between the y-values with the x-values.

$$\text{Slope} = m = \frac{\Delta y}{\Delta x} = \frac{(1155.3 - 609.0)}{(90 - 55)} = \frac{546.3}{35} \approx 15.6$$

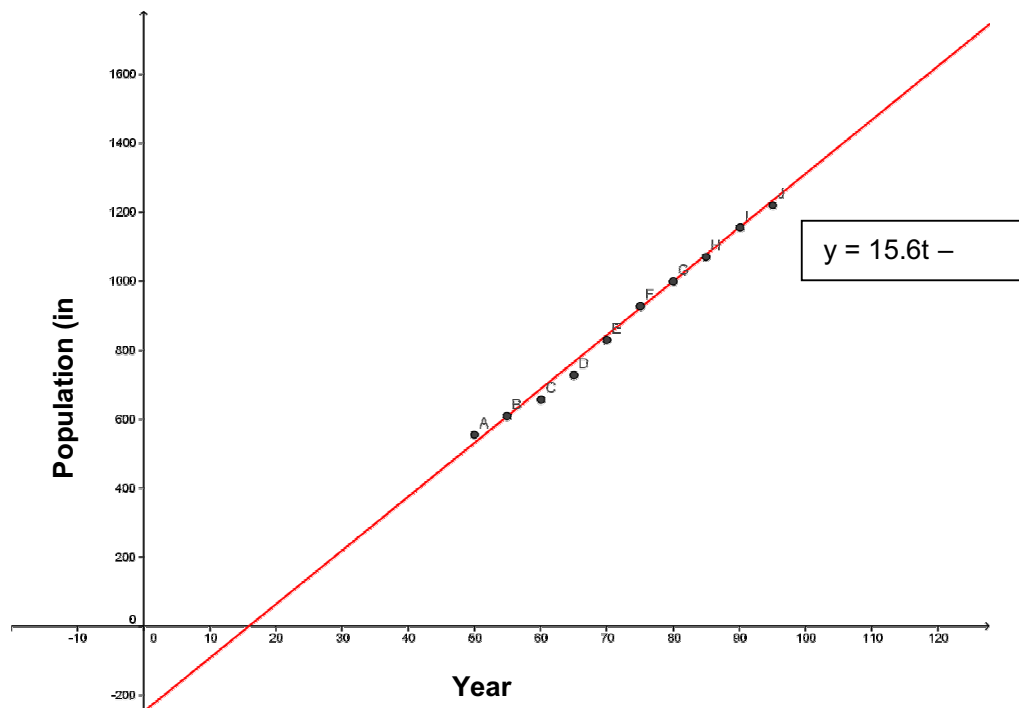
After the slope, we find the y-intercept, y_1 .

$$y - y_1 = m * (x - x_1)$$

$$y - 1155.3 = 15.6 * (x - 90)$$

$$y = 15.6x - 248.7$$

After we find this formula, we plot it into the graph, and we get the line that approximates the original points we got from the data. Below is the graph:



*graph above shows the linear function we found, plotted with the actual points. **[plotted with GeoGebra⁴]**

We can clearly see that the approach we took is close, but not accurate enough because the line does not pass through all the points, but merely touches them.

So by using a linear function we can try and be accurate with the points but it is not reliable when it comes to plotting population growth [or decay] because at one point we can see that the values become negative which is inaccurate and impossible for a population statistics. The linear function does not give us very accurate results about the long run as well, and it is most suitable just for portraying present data, because with this trend, it will show that the population will grow continuously and constantly up to infinity.

Points	Value of the table (a)	Value of the linear curve (b)	Difference between the values (b-a)	Systematic error percentage ((b-a)/b)*100
50	554.8	531.3	-23.5	-4.4

55	609.0	609.3	0.3	0.05
60	657.5	687.3	29.8	4.3
65	729.2	765.3	36.1	4.7
70	830.7	843.3	12.6	1.5
75	927.8	921.3	6.5	-0.7
80	998.9	999.3	0.4	0.04
85	1070.0	1077.3	7.3	0.07
90	1155.3	1155.3	0.0	0.0
95	1220.5	1233.3	12.8	1.04

This model does represent our data quite accurately but as I discussed above, its limitations keep it from being an appropriate model for population growth. The average systematic error percentage is **1.7 %**.

Now we take in account the formula that the researcher suggested which is:

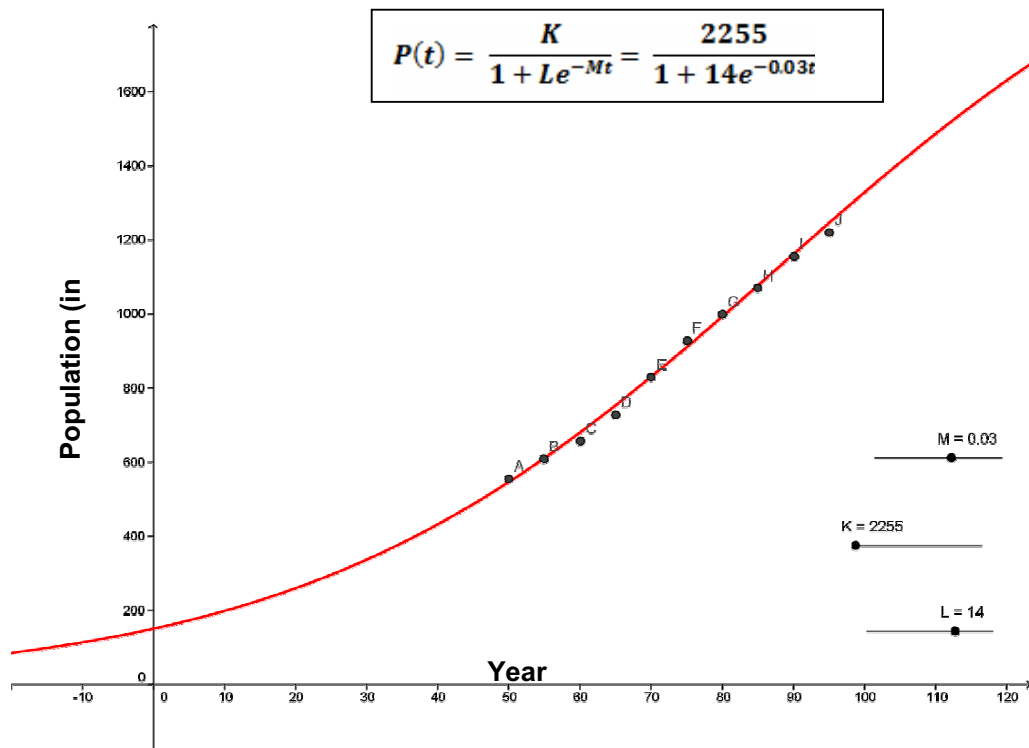
$$P(t) = \frac{K}{1 + Le^{-Mt}}, \text{ where } K, L \text{ and } M \text{ are parameters.}$$

With the use of GeoGebra⁴ we found the approximate values for K, L and M that best fit this model.

K=2255

L=14

M=0.03



**year is shown as 50 – 1950, 70 – 1970 etc, whilst 100 represents 2000. [plotted with GeoGebra⁴]*

We can clearly see that this function fits the model better, there are more improvements than the linear function model and it represents the data more realistically and its approach is more detailed instead of the much broader one of the previous model.

Points	Value of the table (a)	Value of the logistic curve (b)	Difference between the values (b-a)	Systematic error percentage ((b-a)/b)*100
50	554.8	546.8	-8	-1.5
55	609.0	611.3	2.3	0.4
60	657.5	680.4	22.9	3.4
65	729.2	753.7	24.5	3.3
70	830.7	830.6	0.1	0.01
75	927.8	910.9	-26.9	-2.9
80	998.9	993.4	-5.5	-0.6
85	1070.0	1077.3	7.3	0.7
90	1155.3	1161.8	6.5	0.6
95	1220.5	1245.9	25.4	2.1

When we look at this table, we can see that the systematic error percentage is lower than the one of the linear model. The average systematic error percentage is **1.5 %**.

By taking these two approaches, we are able to see the difference, evaluate and find the proper function that most likely suited the model I was given, and got the points presented with quite an accurate precision.

One of the limitations that were noticed in the first linear model function was that although it approached the points close, it wasn't accurate enough, and it does not represent the population statistics as we would imagine, because of its tendency to portray results just in the short run. If we see at the beginning of the 1st picture of the linear function above, we can notice that the line is decreasing and eventually going towards the negative set of numbers. We know that this is not possible when it comes to the count and growth of population, so it stands out as the biggest weakness of the model I presented. Also, in the future, the line tends to grow constantly, just as it is expected from a linear function, but that would mean that the population is growing to infinity at a constant rate, and does not account for any changes that might happen, so the second biggest weakness and limitation of the first model is its ability to predict results and data.

If we take this model as the one that should be considered when looking the growth of population of China, we would find out and expect a very fast growth to several millions of people in the next couple of years, and believe that the population in the past at one point had a negative value (which is impossible to state).

With the second function we are able to come to an even more accurate model that resembles closely the original data points given, which helps us to view the model with a function that is able to cover some of the weaknesses of the first model. We can see that it does start quite realistic, rising at the present moment, with a continuous but not constant rate of growth which does seem more reliable than the constant one of the linear model.

Its limitations could be its lack of complete accuracy, or just a bigger accuracy, its possible predicaments, even if they do come very close to the data, they still do not account any possible changes to the way the population grows, but the data would be correct if the population continues to grow under the same circumstances and at the same rate as the present time in the model.

If we take this model as the one that shows us the growth of the population in China, we are left with a slightly more clear and precise model that would give us more or less close to accurate answers

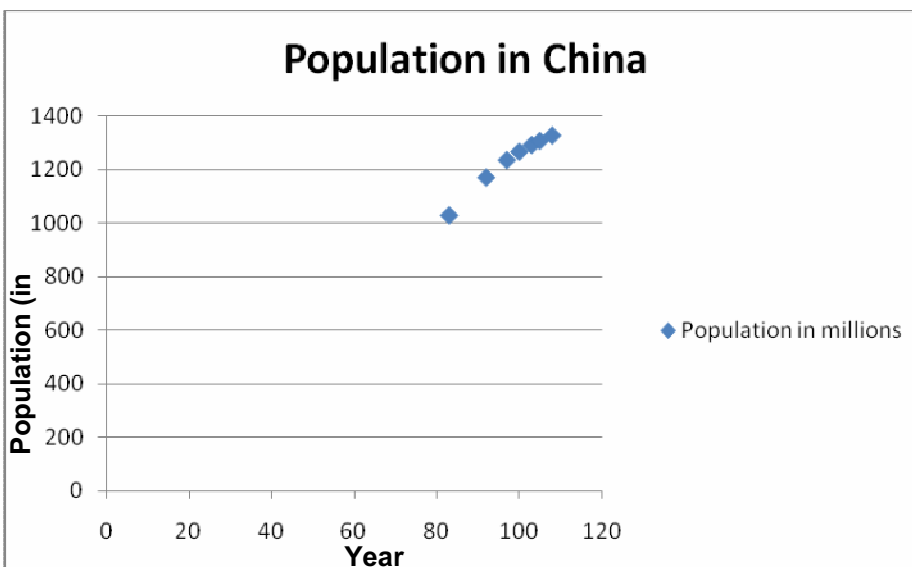
as to how the population was in the past, how it increased, the present time and the continuity of the growth of the population after the present

So the choice as to the more appropriate model for this particular situation would definitely be the second, logistic model because of the bigger accuracy and better predicaments and past data presentation than the linear one.

Now we take a look at the data published by the International Monetary Fund from 2008.

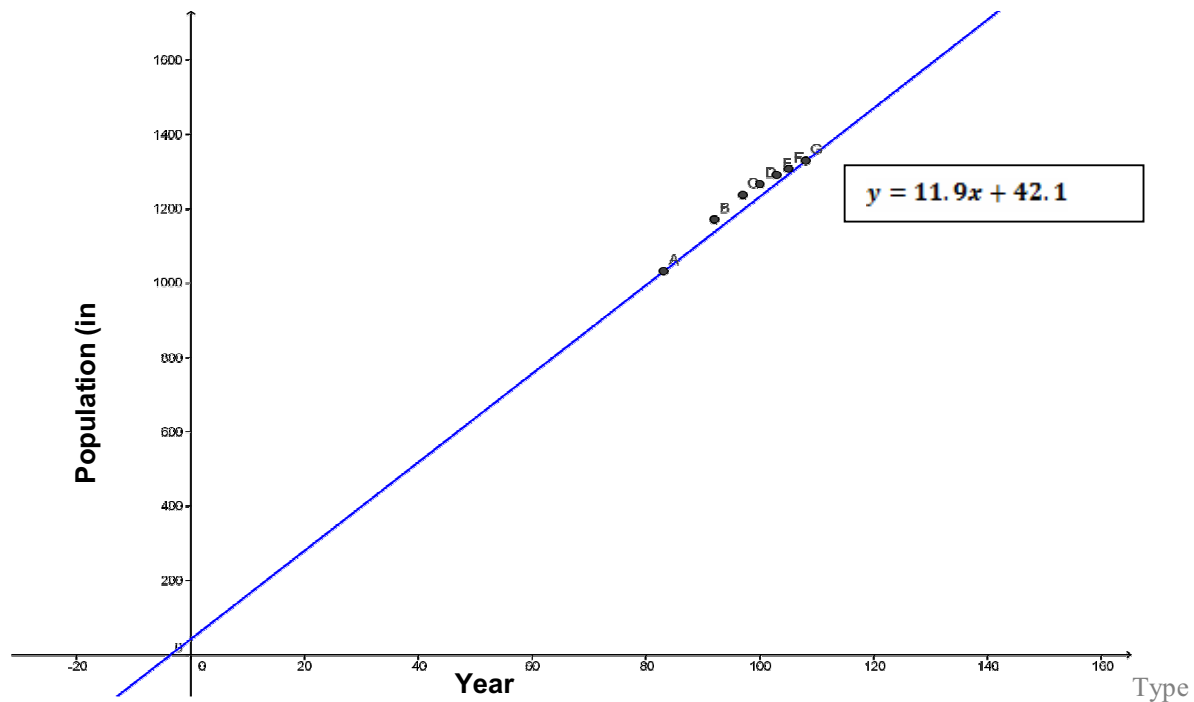
Year	1983	1992	1997	2000	2003	2005	2008
Population in Millions	1030.1	1171.7	1236.3	1267.4	1292.3	1307.6	1327.7

We can see that the data above closely resembles the first data.



**the scattered points show us a growth in population again and the values for the years are counted as 80 for 1980, 90 for 1990 and 100 for 2000. [plotted with Microsoft Excel 2007]*

As we did with the first given data on the sheet, we will start firstly with the linear equation and model. We find the slope and y-intercept and we get the following graph



equation here.

**[plotted on GeoGebra⁴]*

We can see that the linear model did not work out as accurate as with the first model for the first table of data. The possibilities for that are because this has some newer data points that don't increase with the same intensity as before. The population is still growing but not as fast as before, and this is one of the models where we can notice why it is not wise to use a linear function model to present a population growth.

Points	Value of the table (a)	Value of the linear model (b)	Difference between the values (b-a)	Systematic error percentage ((b-a)/b)*100
83	1030.1	1029.8	-0.3	-0.03
92	1171.7	1136.9	-34.8	-3.1
97	1236.3	1196.4	-39.9	-3.3
100	1267.4	1232.1	-35.3	-2.9
103	1292.3	1267.8	-24.5	-1.9
105	1307.6	1291.6	-16	-1.9
108	1327.7	1327.3	-0.4	-0.03

With the table above we can see how the error percentage has increased significantly from the model based on the first data we had. The average systematic error percentage is **1.88%**.

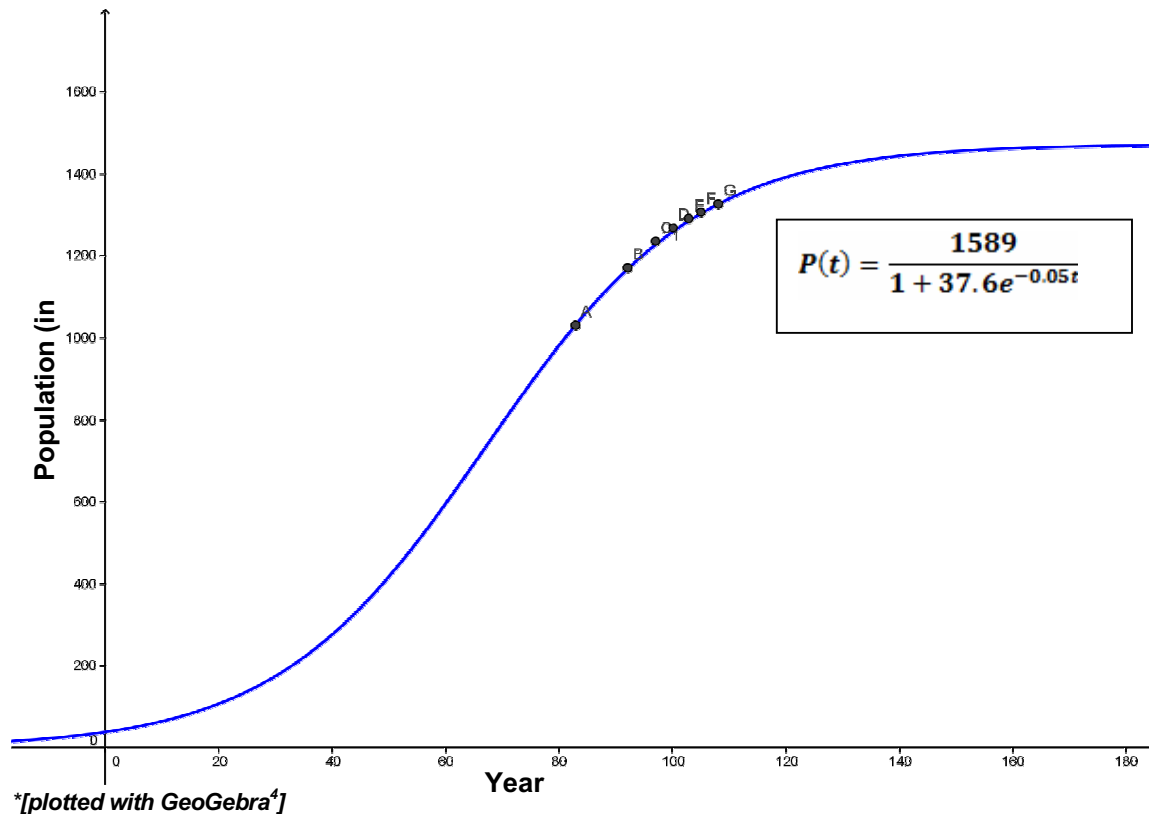
Now for this data I will try to model a presentation through a logistic function, or more precisely, a model based on the logistic function we were given.

The values are: Type equation here.

K = 1589

L = 37.6

M = 0.05



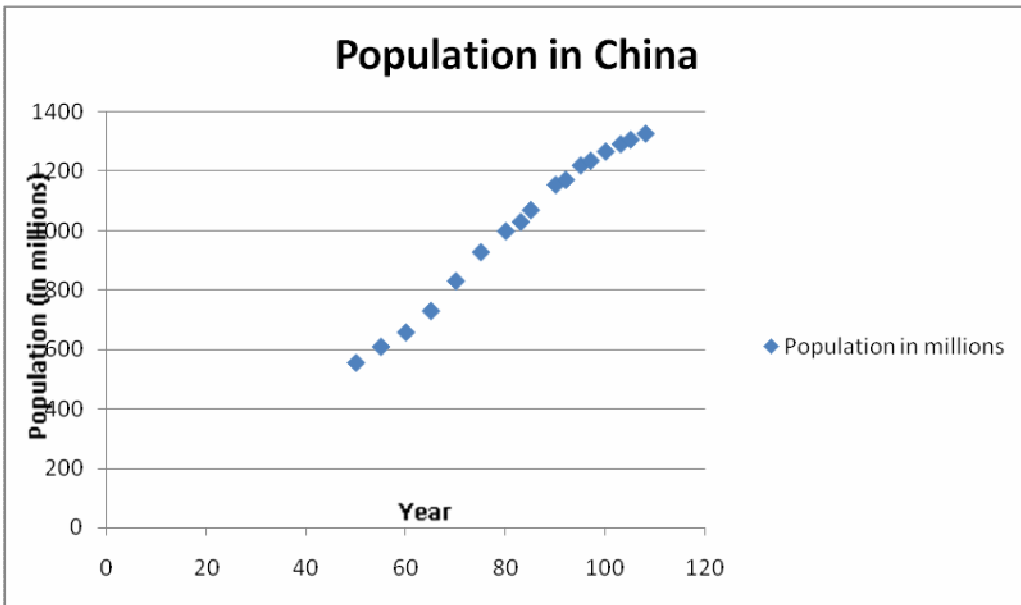
What we are able to interpret with the graph above is that the formula fits the model quite good and accurately.

Points	Value of the table (a)	Value of the logistic model (b)	Difference between values (b-a)	Systematic error percentage ((b-a)/b)*100
83	1030.1	997.7	-32.4	-3.2
92	1171.7	1153.2	-18.5	-1.6
97	1236.3	1227.6	-8.7	-0.7
100	1267.4	1267.8	0.4	0.03
103	1292.3	1304.5	12.2	0.9
105	1307.6	1327.1	19.5	1.5
108	1327.7	1358.3	30.6	2.2

The average systematic error percentage is **1.4 %** which by itself is quite good for our model, and it is represented close to its original data.

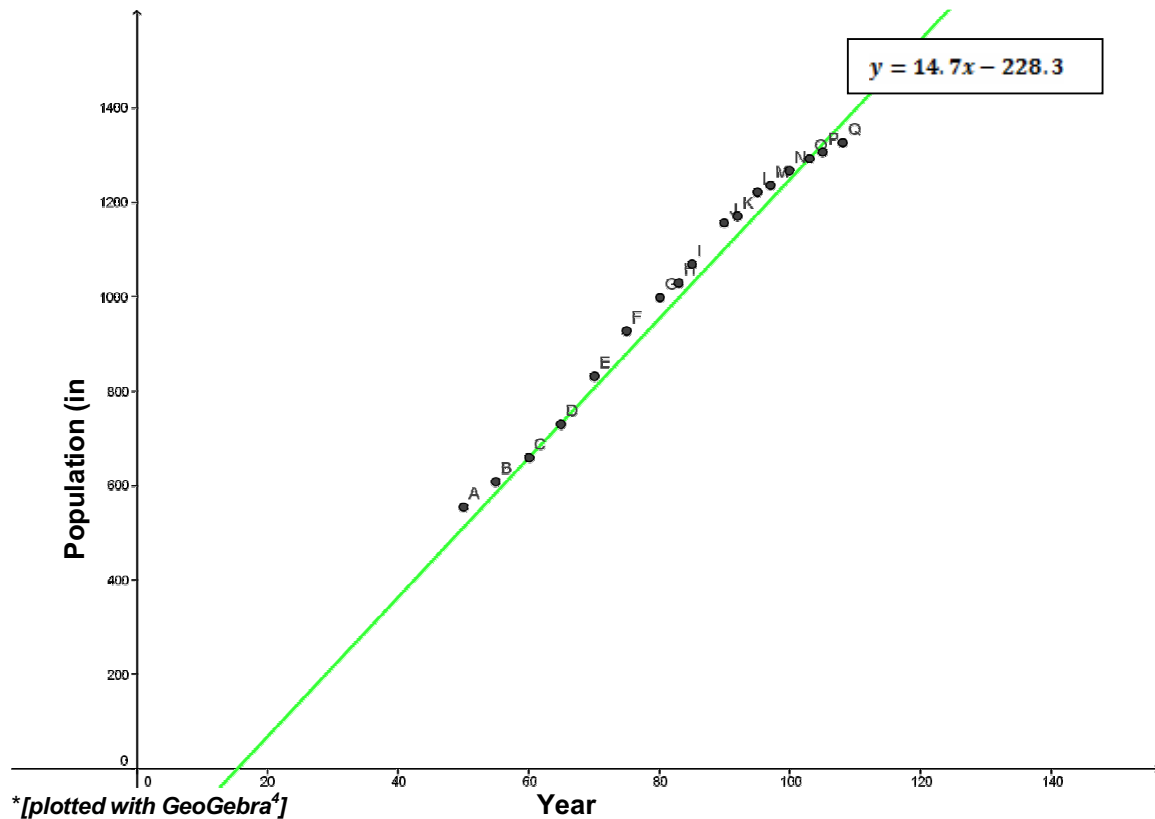
And again, as expected, the logistic model gives us a better presentation of the data given by the IMF about the population growth, because just as it was with the first data for the population, the linear function gave approximate results that make sense just at the present moment, and are quite inaccurate when we discuss about the long run of the population growth. The logistic model on the other hand, gives a more realistic approach as to the population growth and it has better accuracy than the linear model.

If we combine both data and tables we got we will get the following linear and logistic models and tables for the data:



**[plotted with Microsoft Excel 2007]*

Linear model:



**[plotted with GeoGebra⁴]*

Logistic model:

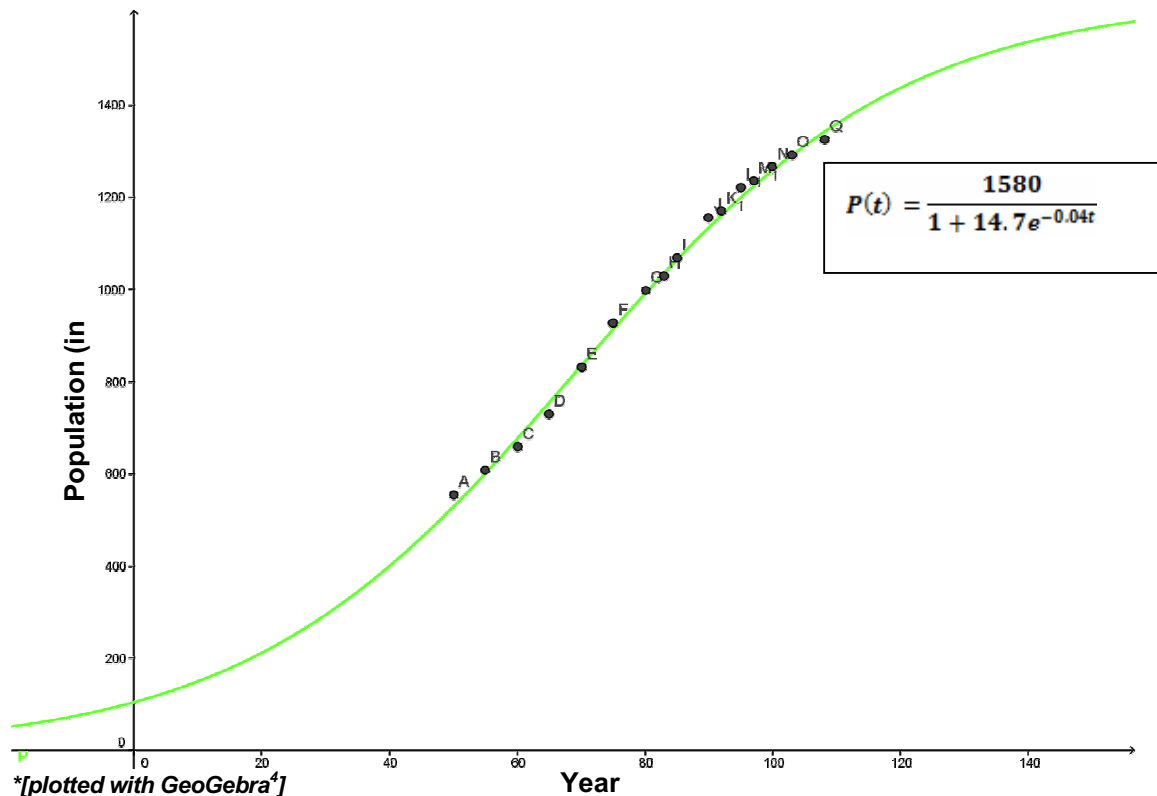


Table:

Points	Value of the table (a)	Value of the logistic model (b)	Difference between the values (b-a)	Systematic error percentage ((b-a)/b)*100
50	554.8	528.5	-26.3	-4.9
55	609.0	601.1	-7.9	-1.3
60	657.5	677.1	19.6	2.9
65	729.2	755.3	26.1	3.4
70	830.7	834.2	3.6	0.4
75	927.8	912.3	-15.5	-1.7
80	998.9	987.9	-10.9	-1.1
83	1030.1	1031.7	1.6	0.1
85	1070.0	1059.9	-10.1	-0.9
90	1155.3	1127.2	-28.1	-2.5
92	1171.7	1152.6	-19.1	-1.6
95	1220.5	1188.9	-31.5	-2.6
97	1236.3	1212.1	-24.2	-2
100	1267.4	1244.8	-22.6	-1.8
103	1292.3	1275.4	-16.9	-1.3
105	1307.6	1294.6	-12.9	-1.002
108	1327.7	1321.6-26.3	-6.1	-0.5

The average systematic error percentage is **1.7%** and considering the range of the data, this is very close to being considered as accurate data, and the precision is high.

So I can conclude that the model that I developed fits the data rather good and accurately, the linear models having their limitations in the prediction and long run data, altogether with the past impossibilities they present whilst the data for the present moment seems as though it is presented correctly. The logistic models,

adjusted to the data that was given to us, best fit the overall model and presentation of the population in China because it models the data with high precision (or rather with low imprecision), and it takes into account any possible long run changes to the population, and can be described as the bestfit model to represent the data in this case.