This portfolio is an investigation into how the median Body Mass Index of a girl will change as she ages. Body Mass Index (BMI) is a comparison between a person's height (in meters) and weight (in kilograms) in order to determine whether one is overweight or underweight based on their height. The goal of this portfolio is to prove of disprove how BMI as a function of Age (years) for girls living in the USA in 2000 can be modeled using one or more mathematical equations. This data can be used for parents wanting to predict the change in BMI for their daughters or to compare their daughter's BMI with the median BMI in the USA.

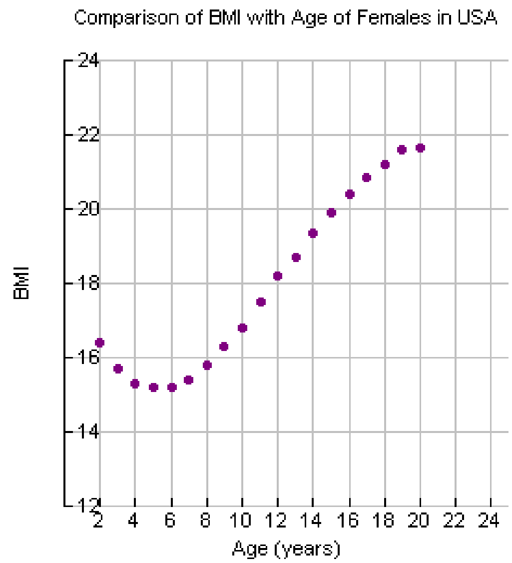
The equation used to measure BMI is:

$$BMI = \frac{kg}{m^2}$$

The chart below shows the median BMI for girls of different ages in the United States in the year 2000:

Age (years)	Median BMI
2	16.40
3	15.70
4	15.30
5	15.20
6	15.21
7	15.40
8	15.80
9	16.30
10	16.80
11	17.50
12	18.18
13	18.70
14	19.36
15	19.88
16	20.40
17	20.85
18	21.22
19	21.60
20	21.65

The values "height (m)" and "weight (kg)" can be discarded when trying to analyze this data because it is not consistent that a girl will have a fixed weight to height or vice versa when they are a certain age. For example, the median BMI for a 10 year old girl is 16.80. With only the data provided, one cannot isolate either the height variable or the weight variable using the formula for calculating BMI.
Using the computer program TI InterActive!™ to graph a scatterplot using the data, the resultant graph is as follows:



Window settings : $x:[0,25,2]$ and $y:[12,24,2]$

One of the parameters of this data are that all x and y values are positive, because one cannot have a negative value for one's weight, height or Age. The scope of this question can be extended to ages beyond the maximum value given by the data, and could conceivably not end until the maximum possible lifespan of a human female (unknown). If this data was a record of the extremes rather than the median BMI of a population, the y values could lower than the window setting of 12 for this graph, as well as much higher than the window setting of 24.

If a curve was fitted to this data, the domain should be: $\{x|2 \le x \le 20, x \in \Re\}$, those values being the lowest and highest ages given in the table of data.

The lowest y-value on the scatter plot is the data point at 5 years (15.20 BMI); the highest y-value on the scatter plot is the data point at 20 years (21.65 BMI).

Using those values, the range of a curve fitted to this data should be:

$\{y|15.2 \le y \le 21.65, y \in \Re\}$

Judging from the distribution of data, the pattern created looks very much like that of a sinusoidal function. Based on this assumption, Finding the values of *a, b, c* and *d* in the function $y = a\sin[b(x-c)] + d$ should result in a graph that fits the data points. Assuming that the point 21.65 is the maximum of the sinusoidal curve and using 15.20 as the minimum, one can find the value for *d* using the formula:

$$d = \frac{Max + Min}{2}$$

d = 18.425

d = 18.4

The line of symmetry in a sinusoidal function is equal to its vertical displacement (*d)* value. Knowing that a sine function begins by curving up from the line of symmetry, the horizontal phase shift *(c)* value can be approximated by looking at what the x value is where the y value is ≈ 18.425 and the graph is curving up. From looking at the graph:

c = 12.0

Using the same assumptions for the maximum value, one can find the value for *a* by using the formula: $a = \frac{Max - Min}{2}$

a = 3.225

a = 3.23

Assuming that the maximum value is when Age = 20 and using Age = 5 as the minimum, one can find the period of the function knowing that the difference between the x value at y max and the x value at y min is ½ of a period. The period of the function is therefore 30.

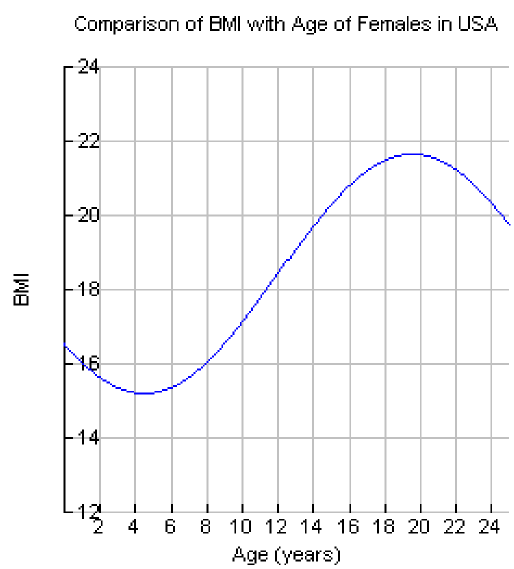
One can find the value for *b* using the formula: $period = \frac{2\pi}{b}$

b = $\frac{\pi}{15}$

Using the approximated *a, b, c* and *d* values, the resultant equation is:

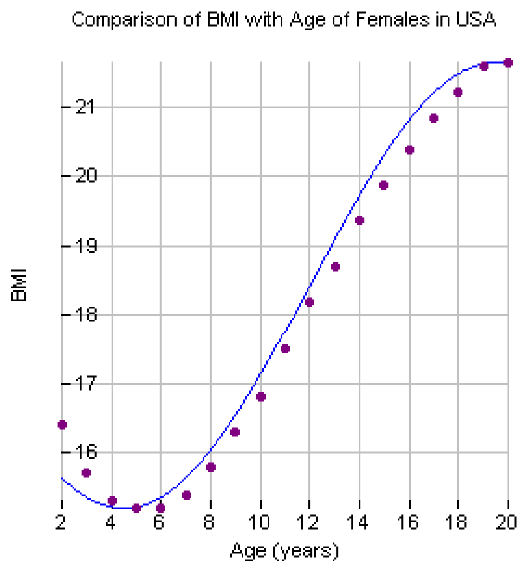
$$y = 3.225sin[\frac{\pi}{15}(x - 12)] + 18.425$$

Using TI InterActive!™ to graph that function, the resultant graph is as follows:



Window settings : $x : [0, 25, 2]$ and $y : [12, 24, 2]$

Overlaid with the original data, and restricting the graph of the function to the same domain and range as the domain and range on the scatterplot, the graph of both the data and the function together is as follows:



Window settings : $x : [2, 20, 2]$ and $y : [15.2, 21.65, 2]$

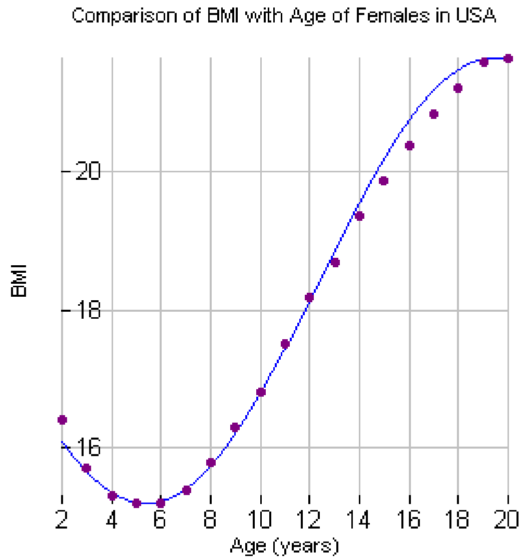
Comparing the two graphs, one can see that where BMI = 18.4 in the function is around 0.400 years off from the corresponding point on the scatterplot where BMI = 18.4. That can be fixed by adjusting the horizontal phase shift value of *c* to 12.4 to compensate.

There is also a noticeable discrepancy between the function graph and the scatter plot in the domain $\{x|2 \leq x \leq 4, x \in \Re\}$ and $\{x|16 \leq x \leq 18, x \in \Re\}$. A slight stretch about the y axis would fit the curve closer to the data points in those areas, therefore the *b* value should be changed from π/15 to π/14. The number π/14 was chosen because reducing the denominator's value by 1 will increase the b value by a slight amount, and an increased b value will shorten period and compress the graph about the y axis, resulting in the values within domain $x|2 \leq x \leq 4$ to be closer to that of the data points.

Making these changes, the revised formula for the sinusoidal function is as follows:

$$y = 3.225sin(\frac{\pi}{14}(x - 12.4)) + 18.425$$

Using TI InterActive!™ to graph that function, overlaid with the original data, and restricting the graph of the function to the same domain and range as the domain and range on the scatterplot, the graph is as follows:



Window settings : $x : [2, 20, 2]$ and $y : [15.2, 21.65, 2]$

The refined equation fits the data much more closely at the beginning, yet around the Age = 13 mark it starts to deviate again. Looking at this graph, one can come to the conclusion that it will take more than one mathematical function to graph the entirety of the data.

If one were to relate the concept of human growth with the measurable value of BMI, then there is a problem in assuming the graph follows a sinusoidal pattern when extrapolating data from the graph beyond 20 years of age. Based on common knowledge, girls are usually finished growing at ≈ 20 years of age – yet one can see that had the domain not been restricted in the sinusoidal function, the BMI would curve back down and back up for women ages 20 to 38 (the difference of 20 and 38 being ½ of the period). Because that data would be wrong based on the common knowledge of growth stopping after ≈ 20 years of age, one can conclude that a sinusoidal function will not match the data in the domain: $\{x|20 \leq x, x \in R\}$.

Using the assumption that girls are usually finished growing at ≈ 20 years of age, one can think of human growth as an instance of *bounded exponential growth*, where a girl grows rapidly as a child before slowing down and eventually stopping her growth once they have reach adult size. With this frame of mind, the mathematical function that best models growth towards a fixed capacity would be a Logistic Regression function.

Because a logistic regression function models data which whose deviance is getting smaller which each point, it will be best used to model data from the point in which the rate of growth is no longer growing. That can be estimated as the *c* value or the horizontal phase shift because in a regular sin function, 0 to $\frac{period}{4}$ is when the rate of increase starts decreasing until the max point. *c* in the revised sinusoidal function equation is 12.4. Therefore, the logistic function should replace the sinusoidal function in predicting or extrapolating data in the domain:

$\{x|12.4 \leq x, x \in R\}$.

A logistic function has the general equation of:

$$y = \frac{c}{1 + ae^{-bx}}$$

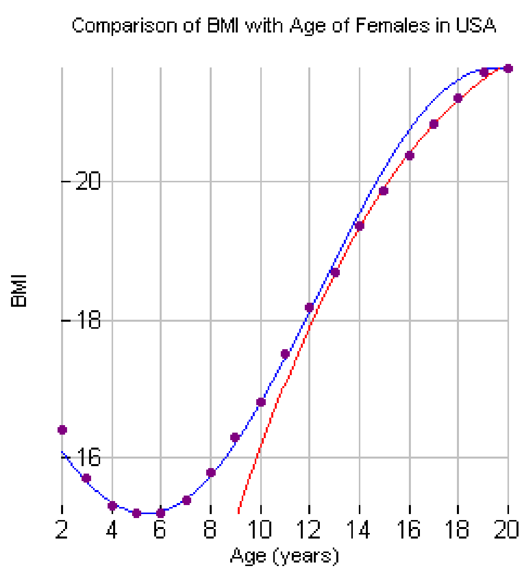
Using the "calculate logistic function" function on TI InterActive!™ , using the data restricted to 13 to 20 years of age only (13 is rounded up from the starting point of 12.4), one gets the following values:

a = 3.36
b = 0.208
c = 22.9

Using those values, the resultant function is:

$$y = \frac{22.87767825}{1 + (3.361268603)e^{-0.208177202x}}$$

Using TI InterActive!™ to graph that function, overlaid with the sinusoidal function in the same graph, the result is as follows:

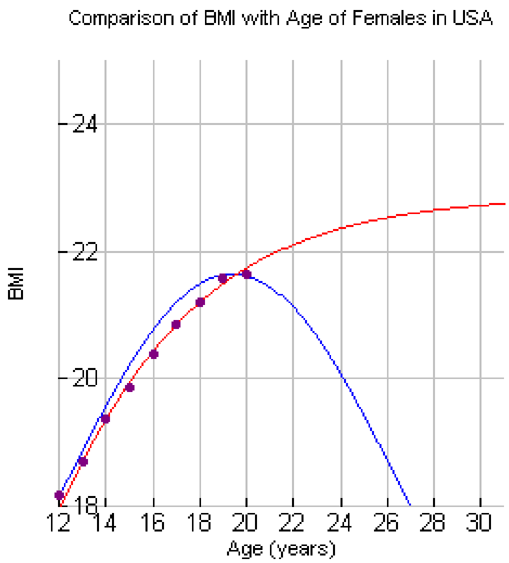


Window settings : $x : [2, 20, 2]$ and $y : [15.2, 21.65, 2]$

*

The logistic function clearly fits the data more closely than the sinusoidal function for data in the domain $\{x|12.4 \leq x, x \in R\}$.

When the functions are allowed to extrapolate beyond the data points provided, one can see the great difference between the two:



Window settings : $x:[12,31,2]$ and $y:[18,25,2]$

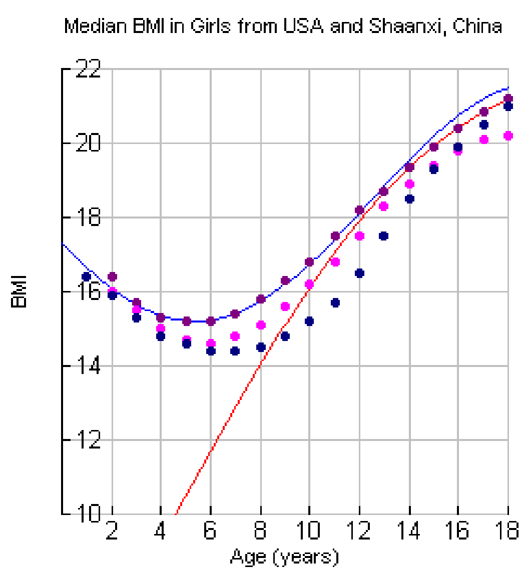
As expected, the sinusoidal function curves down, increasing at the rate it is decreasing from its max point at Age = 20. The logistic function on the other hand, does not decrease after Age = 20. Rather, it levels off with an asymptote of 22.9 This is much more reasonable as a model of a girl's BMI as a function of Age because of how unlikely it is that one's weight to height ratio will have such drastic changes every 18 years (one period of the sinusoidal function graph).

Using the logistic regression function, the BMI of a 30-year-old woman in the US in 2000 would be 22.7. Using the same assumption that human growth is related to the measurable value of BMI, that value is off because a human stops growing at ≈ 20 years of age and that value is more than a full unit higher than the BMI value at age 20 (21.65) based on the data. However, because BMI is not unerringly related to the concept of human growth into adult size, and that it is rather a measurement that compares one's height with one weight; it is possible that one could keep gaining weight to a certain point after reaching adult height. This assumption would be that one reaches adult height and adult weight at different times, causing the BMI value of height to remain constant after 20 years of age but the value of weight to keep on increasing until a separate plateau – adult weight, is reached. Therefore, the BMI value would not remain constant after ≈ 20 years of age and it will keep on increasing until both adult height and weight is reached. Under this assumption, the value found is very reasonable. Also under this assumption, one can conclude based on this data that the median BMI of a woman who has reached both adult height and adult weight would be a value very close to the asymptote of the function – 22.9.

Data taken from an American source should not be used as a generalization for all girls in the world. Many external factors such as diet, health, quality of life and standard of living will affect one's BMI. To contrast the data provided for American girls, the following is a collection of data from the Department of Health Statistics for the People's Republic of China for the province of Shaanxi in 2005.

Age (y)	America						Shaanxi											
	boys			girls			urban boys			urban girls			rural boys			rural girls		
	95th	50th	5th	95th	50th	5th	95th	50th	5th	95th	50th	5th	95th	50th	5th	95th	50th	5th
1	19.9	17.2	14.6	19.3	16.6	14.7	19.7	16.8	14.1	19.6	16.4	13.8	19.8	17.0	14.4	19.2	16.4	13.9
2	19.0	16.5	14.4	18.7	16.0	14.3	19.0	16.4	14.0	18.9	16.0	13.7	18.8	16.3	14.0	18.4	15.9	13.7
3	18.4	16.0	14.0	18.3	15.6	13.9	18.4	15.9	13.8	18.2	15.5	13.4	17.9	15.6	13.5	17.6	15.3	13.2
4	18.1	15.8	13.8	18.2	15.4	13.6	17.9	15.4	13.4	17.5	15.0	13.0	17.4	15.2	13.3	17.1	14.8	12.8
5	18.0	15.5	13.7	18.3	15.3	13.5	17.5	15.0	13.0	17.1	14.7	12.7	16.9	14.8	12.9	16.9	14.6	12.6
6	18.1	15.4	13.6	18.8	15.3	13.3	17.2	14.7	12.6	17.1	14.6	12.6	16.7	14.6	12.7	16.7	14.4	12.4
7	18.9	15.5	13.6	19.7	15.5	13.4	17.7	14.7	12.5	17.6	14.8	12.7	16.8	14.7	12.8	16.7	14.4	12.4
8	19.7	15.7	13.7	21.0	16.0	13.6	18.5	15.0	12.6	18.2	15.1	12.8	17.1	14.9	13.0	17.0	14.5	12.5
9	20.9	16.0	14.0	22.7	16.6	14.0	19.6	15.5	12.8	19.4	15.6	13.0	17.6	15.2	13.2	17.5	14.8	12.6
10	22.2	16.6	14.2	24.2	17.1	14.3	20.7	16.1	13.2	21.0	16.2	13.3	18.3	15.5	13.5	18.3	15.2	12.8
11	23.5	17.2	14.6	25.7	17.8	14.6	22.2	16.6	13.5	22.5	16.8	13.6	18.9	15.8	13.7	19.1	15.7	13.1
12	24.8	17.8	15.1	26.8	18.3	15.0	23.3	17.0	13.9	23.9	17.5	14.1	19.5	16.1	14.0	20.3	16.5	13.6
13	25.8	18.4	15.6	27.9	18.9	15.4	24.1	17.4	14.2	25.1	18.3	14.8	19.9	16.5	14.3	21.7	17.5	14.3
14	26.8	19.1	16.1	28.6	19.4	15.7	24.6	18.0	14.7	25.5	18.9	15.3	20.6	17.1	14.8	22.9	18.5	15.1
15	27.7	19.7	16.6	29.4	19.9	16.1	25.2	18.6	15.2	25.4	19.4	15.8	21.3	17.9	15.5	23.6	19.3	15.8
16	28.4	20.5	17.2	30.0	20.2	16.4	25.9	19.2	15.6	25.2	19.8	16.3	22.3	18.7	16.1	24.0	19.9	16.4
17	29.0	21.2	17.7	30.5	20.7	16.9	26.4	19.7	15.9	25.1	20.1	16.6	23.0	19.2	16.4	24.4	20.5	17.0
18	29.7	21.9	18.3	31.0	21.1	17.2	26.8	20.2	16.2	24.7	20.2	16.7	23.8	19.8	16.7	24.5	21.0	17.5

Using the data from the 50th percentile (the median value of the data collected) and overlaid with the graph of my sinusoidal function, logistic regression function and original data points, the graph is as follows:



Purple –Girls in USA, 2000

Blue – Girls in rural Shaanxi, 2005

Pink – Girls in urban Shaanxi, 2005

Window settings : $x : [0, 18, 2]$ and $y : [10, 22, 2]$

The Department of Heath Statistics have also fitted a curve to the data using the median BMI of urban and rural girls as well as urban and rural boys:

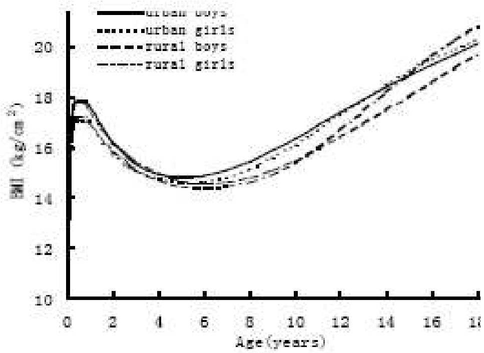


Fig. 3 Median (M curve) of BMI in Shaanxi children

Window settings : $x : [0, 18, 2]$ and $y : [10, 22, 2]$

One can see from this comparison that while the BMI is generally lower across the board, the data patterns for urban girls in Shaanxi and girls in the USA are extremely similar. The model of having the data up to Age = 12.4 plotted as a sinusoidal function and the data from 12.4 onwards as a logistic regression function also fits the data for urban girls in Shaanxi, after slight changes to values of vertical displacement and horizontal phase shift. The majority of the people in USA live in urban conditions, so it is understandable that given similar living conditions the data for urban girls in Shaanxi follows the same pattern. As for the BMI being lower overall, that is likely because of a lower standard of living in Shaanxi compared USA, causing a difference in weight that is contributed to a steady diet.

Where the model fails is for the data of rural girls in Shaanxi. A sinusoidal function does not fit the data for rural girls Ages 1 - 12.4 at all. This can be concluded from the values before and after the min point of 14.4 BMI. A sinusoidal function will have symmetrical data before and after the min point, but the data of rural girls in Shaanxi has the data after the min point growing exponentially while the data before the min point decreasing in almost a linear fashion. Therefore, the sinusoidal model cannot be adjusted to fit the data unless the sinusoidal function is abandoned completely for linear function with a negative slope followed by an exponential growth function. However, the second half of model, plotted using a logistic function, is still likely to fit the data after some changes to the variables. This is due to the theory behind why a logistic function is perfect in fitting data for “growth towards a fixed capacity”. Unless rural girls in Shaanxi never reach an “adult height” or “adult weight” value, it is certain that their BMI will also stop growing after a certain point. Had there been more data collected by the Department for Heath Statistics – data for women above the age of 18 in rural Shaanxi, this assumption could be confirmed.

In conclusion, this study has shown that for girls living in urban conditions, BMI can be modeled by the sinusoidal and logistic mathematical functions. Hopefully thanks to the work done in this report, parents will be able to track whether their children are above or below the median BMI for their age, and predict how much their BMI is likely to change in the following year using the functions derived in the report. They will also be able to compare the BMI of their daughters with that of children in rural and urban China, able to judge whether their daughter is over or under the median statistics. This portfolio has also shown how the difference in lifestyle and living conditions will change the pattern of BMI as a function of Age. Hopefully this report can serve as a starting point for others researching change in BMI over time for people of both genders and across a much larger age span.

Appendix

"Body mass index." Wikipedia, the free encyclopedia. 16 Mar. 2009. 16 Mar. 2009 <http://en.wikipedia.org/wiki/Body_mass_index>.

"Logistic regression." Wikipedia, the free encyclopedia. 16 Mar. 2009. 16 Mar. 2009 <http://en.wikipedia.org/wiki/Logistic_regression>.

Mueller, William. "Logistic Functions." Exploring Precalculus. 16 Mar. 2009 <http://www.wmueller.com/precalculus/families/1_80.html>.

Shang, Lei, Yong-yong Xu, Xun Jiang, and Ru-lan Hou. "Body Mass Index Reference Curves for Children Aged 0-18 Years in Shaanxi, China." Department of Health Statistics 1.1 (2005). 12 Mar. 2009 <http://www.ijbs.org/User/ContentFullText.aspx?VolumeNO=1&StartPage=57&Type=pdf>.

"Sigmoid function." Wikipedia, the free encyclopedia. 16 Mar. 2009. 16 Mar. 2009 <http://en.wikipedia.org/wiki/Sigmoid_curve>.

Appleby, A., Letal, R., & Ranieri, G. (2007). Pure Math 30 Workbook. Calgary: Absolute Value Publications

Technology Used:

TI InterActive!™ Free Trial by Texas Instruments
TI 84-Plus Silver Edition Graphing Calculator
Microsoft Office Excel 2003
Microsoft Office Word 2003

*Extra:
Graph of Sinusoidal function in Blue followed by Logistic Regression function in Red, with the extraneous values removed.