

Body Mass Index Math Coursework

By Warren Hui



Body Mass Index Maths Coursework

by Warren Hui

Body Mass Index Math Coursework

By Warren Hui

Introduction

This piece of coursework will be based on data of the median BMI for females of different ages in the US in the year 2000. It will be a modeling task and I will model the data using different functions.

BMI or Body Mass Index is a measurement of someone's height in relation to their weight; this standardized ratio is usually used to determine someone's health. BMI can be calculated by dividing your weight (in kilograms) by the square of your height (in meters).

1. Using technology, plot the data points on a graph.
Define all variables used and state any parameters clearly

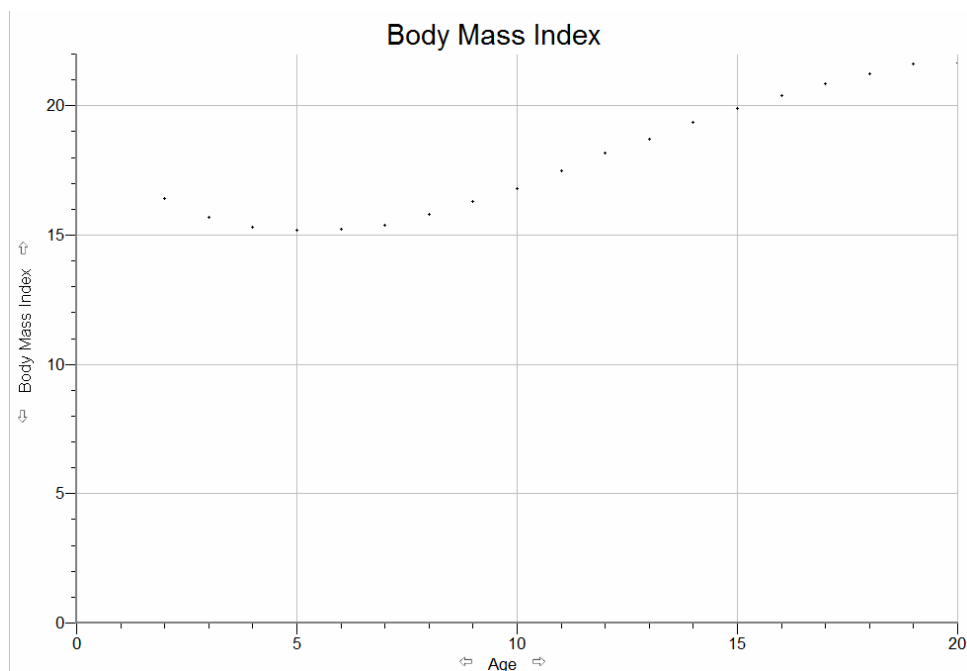
	Data Set	
	Age	y
1	2	16.40
2	3	15.70
3	4	15.30
4	5	15.20
5	6	15.21
6	7	15.40
7	8	15.80
8	9	16.30
9	10	16.80
10	11	17.50
11	12	18.18
12	13	18.70
13	14	19.36
14	15	19.88
15	16	20.40
16	17	20.85
17	18	21.22
18	19	21.60
19	20	21.65

The first thing I did was input the data into the x column. This is because age is the constant. Then I named the column 'Age' to make sure the x axis has a title.

The graph is plotted but all the points are connected and this makes the data become inaccurate, so I right click and chose 'Graph Options' and uncheck the Connect Points box. The dependent variable in this graph is the BMI which is dependent on the Age (independent variable). The parameter a constant in the equation of a curve that can be varied to yield a family of similar curves. The black dots show the median BMI in respect to the age.

Body Mass Index Maths Coursework

by Warren Hui



2. What type of function models the behavior of the graph? Explain why you chose this function. Create an equation (a model) that fits the graph.

I chose the sine function because the curve would fit the data provided. The curve I see shows a great resemblance to a sine graph and after stretching and shifting a sine function, it will match the graph. I did not choose cosine because there is no regression for cosine. Also there are lots of similarities between sine and cosine simple because they have the same function translation.

$$Y = A \sin(Bx + c) + d$$

A= To find the Amplitude (A) I must find the length between the maximum to the middle of the graph.

From looking at the data I can tell that the Maximum value is 21.65

The minimum value is 15.20

Middle point of graph is calculated by $(\text{Maximum} + \text{Minimum}) \div 2$

Body Mass Index Maths Coursework

by Warren Hui

$$\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right)$$

$$(21.65 + 15.20) \div 2 \\ = 18.425$$

Then I need to find the amplitude (length between max point and mid point)
To do this I minus the middle point from the maximum point

$$21.65 - 18.425 = 3.225$$

$$A = 3.225$$

B= the stretch of the graph, which is $\sin x = 2\pi$. However, in the data provided, the period is incomplete so it becomes $\sin x = 2\pi b$, where I have to find b.

From looking at the graph, I assume the graph is half the period, so I'll calculate the value of half the period which is the maximum (20) – minimum (5). To get the full period, I'll need to multiply the half period by 2.

$$20 - 5 = 15$$

$$15 \times 2 = 30$$

b being the period, I'll substitute 30 into the equation of $\sin x = 2\pi b$.

$$2\pi \div 30 = 0.2094$$

$$B = 0.2094$$

C= The phase shift of the function is $C \div B$. A positive phase shift will mean the graph moves to the right while a negative phase shift means the graph moves to the left.

The maximum points of $y = \sin x$ are at $2n\pi + \pi/2$. So I'll try to locate that point on the data. The minimum points are at $2n\pi + 3\pi/2$ and the x intercepts are $n\pi$.

First, we need to find the difference between the maximum (20) and minimum (5) values, $20 - 5 = 15$

Body Mass Index Maths Coursework

by Warren Hui

$$15 \div 2 = 7.5$$

The half of the difference is 7.5, then we add to the minimum (5)

$$7.5 + 5 = 12.5$$

12.5 is the period end in a normalized sine curve. The whole period is 30 in a complete sine curve, so the offset is $30 - 12.5$ which is 17.5.

We then need to multiply 17.5 by the stretch (B) to normalize the curve, which gives us C.

$$20 - 5 = 15$$

$$15 \div 2 = 7.5$$

$$12.5 + 5 = 17.5$$

$$17.5 \times 0.2094 = 3.664$$

$$\mathbf{C = 3.664}$$

D = the amount of vertical displacement which shows how the function graph moves up by using the midpoint formula

Middle point of graph is calculated by $(\text{Maximum} + \text{Minimum}) \div 2$

$$\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right)$$

$$(21.65 + 15.20) \div 2$$

$$\mathbf{D = 18.425}$$

Next to check the accuracy between my calculated values and the actual values of the graph modeled from the data, I will plot both my model and the data's model onto a graph.

RMSE or the root means square error is a measure of the difference between the values predicted by a model (the graph in this case) and the actual values from the thing being modeled. RMSE is a good indicator of the accuracy and the individual differences between two models.

Body Mass Index Maths Coursework

by Warren Hui

To check the values I got I used a Regression Curve fit to check.

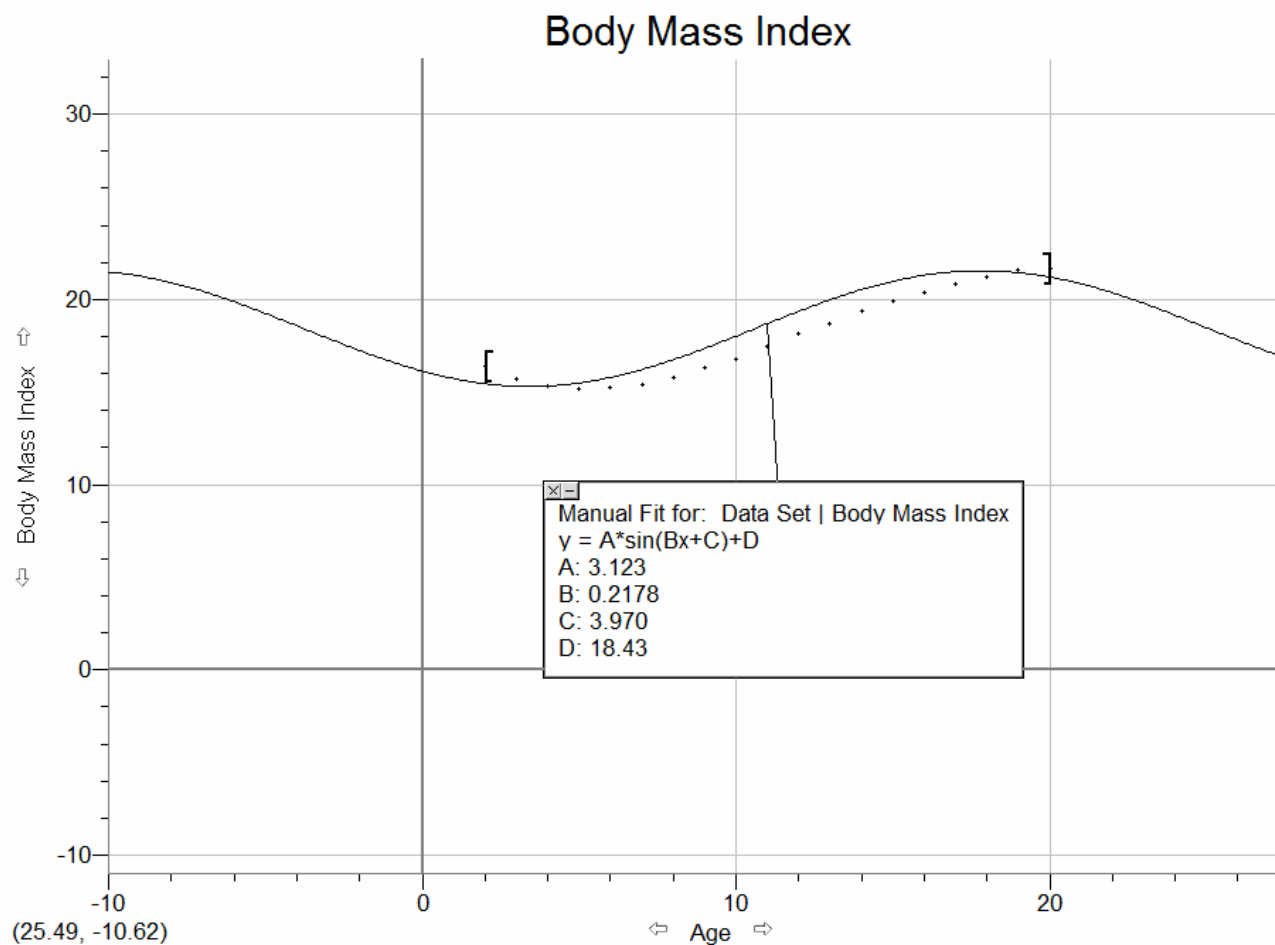
$$A=3.123$$

$$B=0.2178$$

$$C= 3.970$$

$$D= 18.43$$

$$RMSE= 0.970936$$



The Regression Curve fit graph with the original graph from data

After I looked at the Regression Curve RMSE, I refined my model to

$$A=3.129$$

$$B=0.184$$

$$C=3.96$$

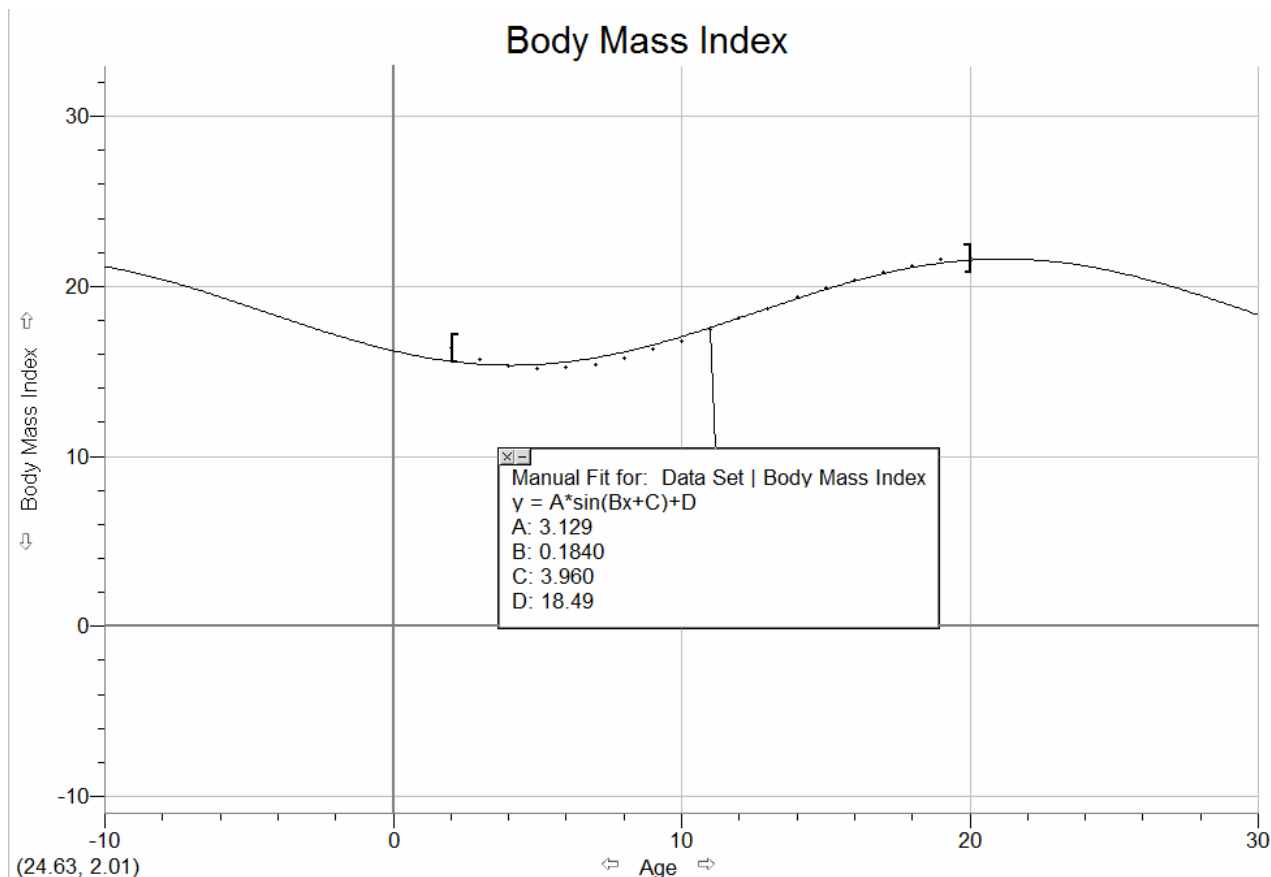
Body Mass Index Maths Coursework

by Warren Hui

D=18.49

By tweaking certain decimal placed to reduce the RMSE.

Now the RMSE to the original graph is 0.305792



The tweaked Regression Curve Fit graph and the original graph from data

- On a new set of axes, draw your model function and the original graph. Comment on any differences. Refine your model if necessary

Next I use Curve Fit to check the RMSE of the values I worked out

RMSE: 0.188044

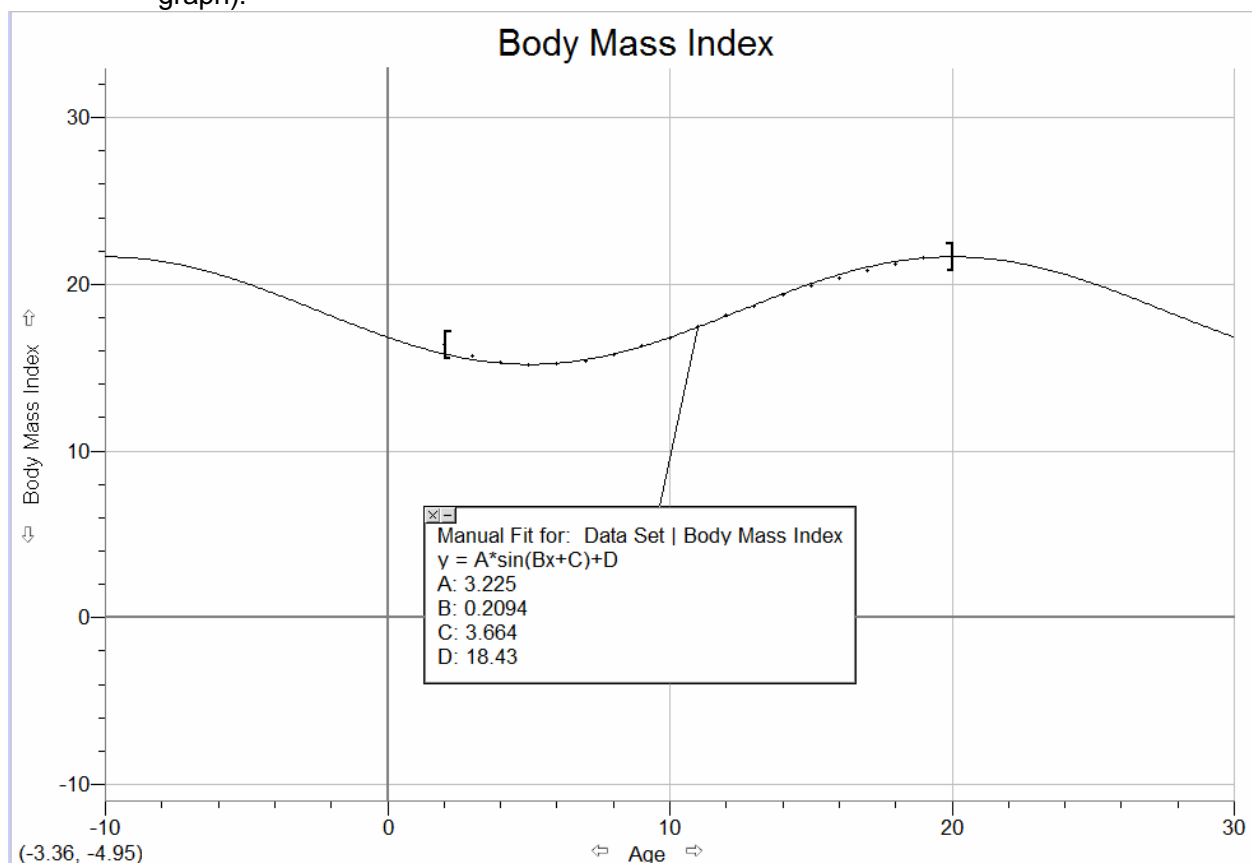
Since the RMSE is so low, it shows that my model function is extremely close to the original model. However there are still minor discrepancies due to

Body Mass Index Maths Coursework

by Warren Hui

manual calculation and the rounding off decimal points.

Even though I tweaked the Regression Curve Fit graph, the values I work out has a lower RMSE than 0.305792 (of the tweaked Regression Curve Fit graph).



The graph modeled from the values I worked out and the graph modeled from original data

4. Use technology to find another function that models the data. On a new set of axes, draw your model functions and the function you found using technology. Comment on any differences.

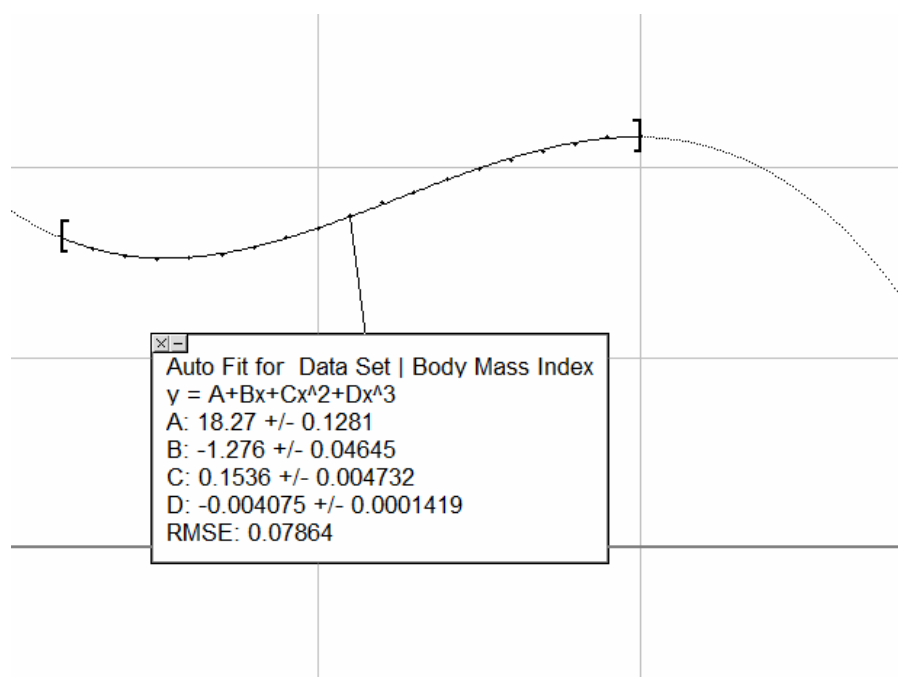
Before I chose another function that models the data I had two equations in mind. Either the Cubic, Quartic or Quintic. In order to see which equation models my graph more accurately I would look at the RMSE

Cubic graph

The RMSE is 0.0786418. This is already very accurate.

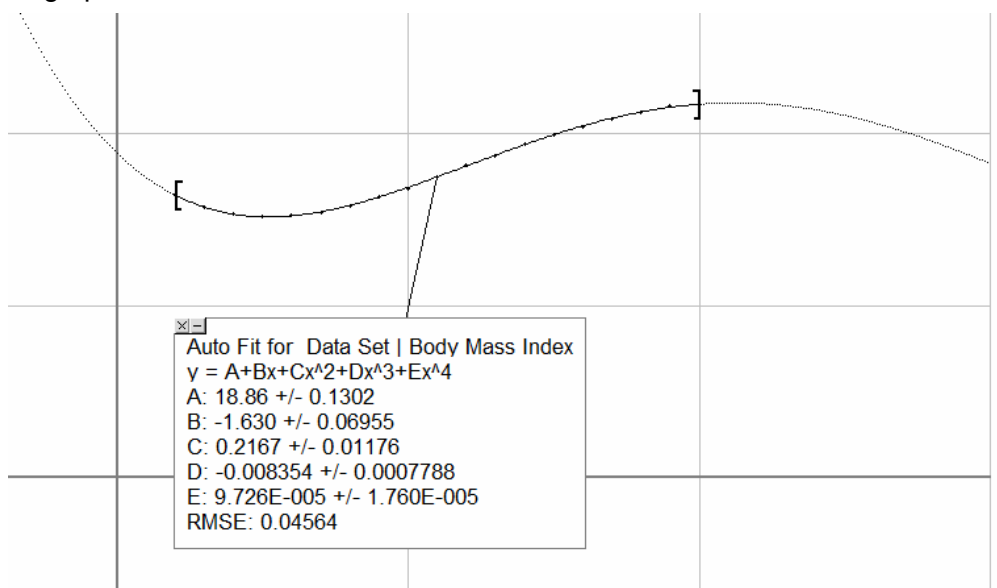
Body Mass Index Maths Coursework

by Warren Hui



Quartic graph

The RMSE is 0.04564, this is very accurate and smaller then the Cubic graph's RMSE.



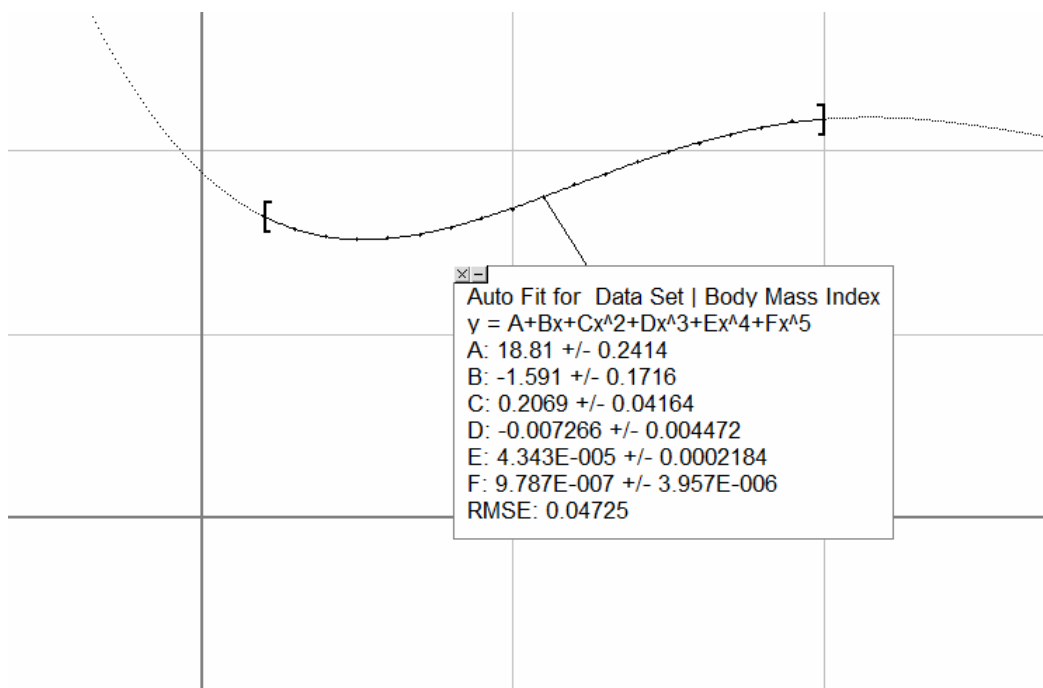
Quintric

RMSE: 0.04725, this number is bigger then the Quartic graph but smaller then

Body Mass Index Maths Coursework

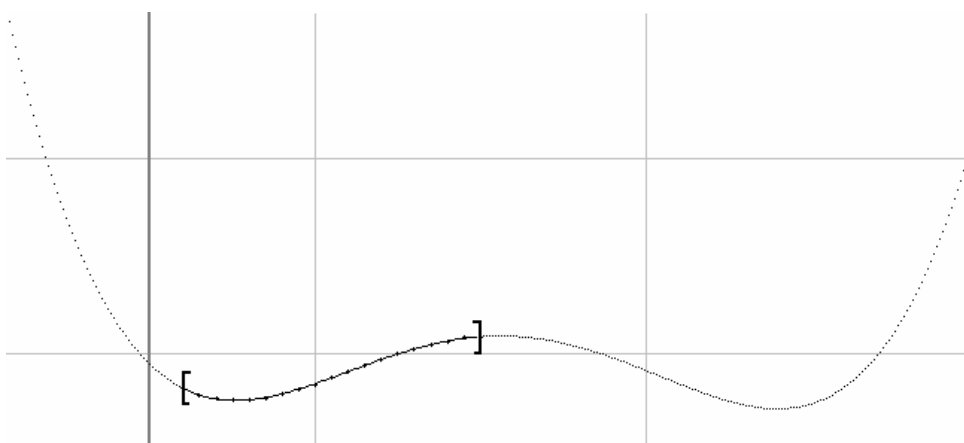
by Warren Hui

the Cubic graph, so I chose to use Quartic



Quartic is the most accurate graph, therefore I will use it for the next question. Some possible limitations in using the Quartic function is that it is only a best fit for age 2-20. If new data was presented, or we used this function to try and predict data outside of the data provided there may be errors. This is because the quartic function is only best fit for the current data provided, therefore we cannot solely rely on the quartic function for predicting values outside of the data provided.

If we zoom out, we will see the quartic graph dropping a lot after age 30. This proves this Quadratic may be the current best fit, but it's unreliable if we look outside of our current age range.

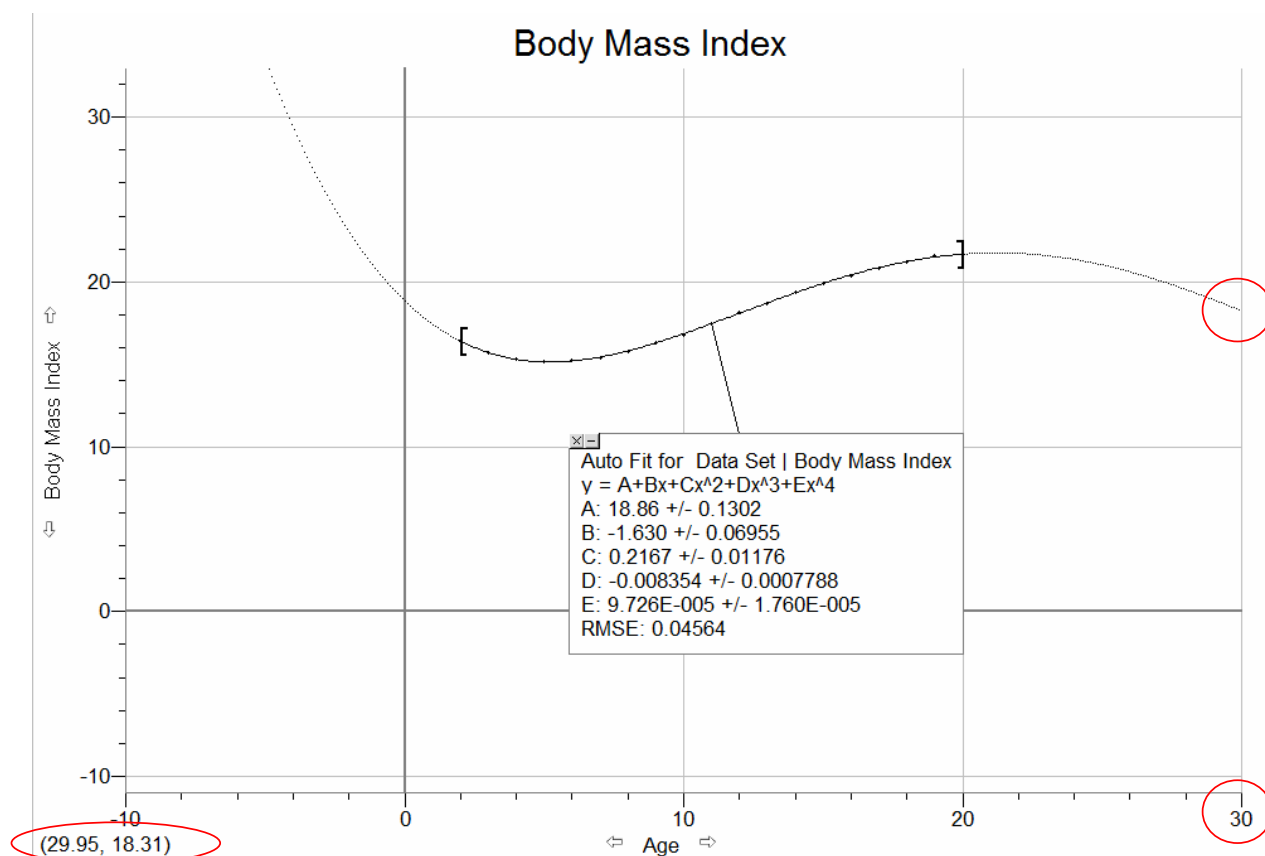


Body Mass Index Maths Coursework

by Warren Hui

5. Use your model to estimate the BMI of a 30-year-old woman in the US. Discuss the reasonableness of your answer

From the previous question I discovered that the Quartic is the most accurate compared to the graph modeled from the original data, so we should start by using the Quartic function



Quartic graph compared with origin

The co-ordinates are (29.95, 18.31)

Which means the BMI estimate for 30 year old women are around 18.38

This is unreasonable because this BMI of 30 year old women is much lower then the BMI of an age 20 or 25 year old woman.

30 year old women BMI: 18.38

20-25 year old women BMI: 20-21 (we can tell this by placing cursor on the line

Body Mass Index Maths Coursework

by Warren Hui

from the age of 20-30)

The BMI of these 30 year old women resembles that of an 11-12 year old girl.

30 year old women BMI: 18.38

11 year old girl BMI: 18.18

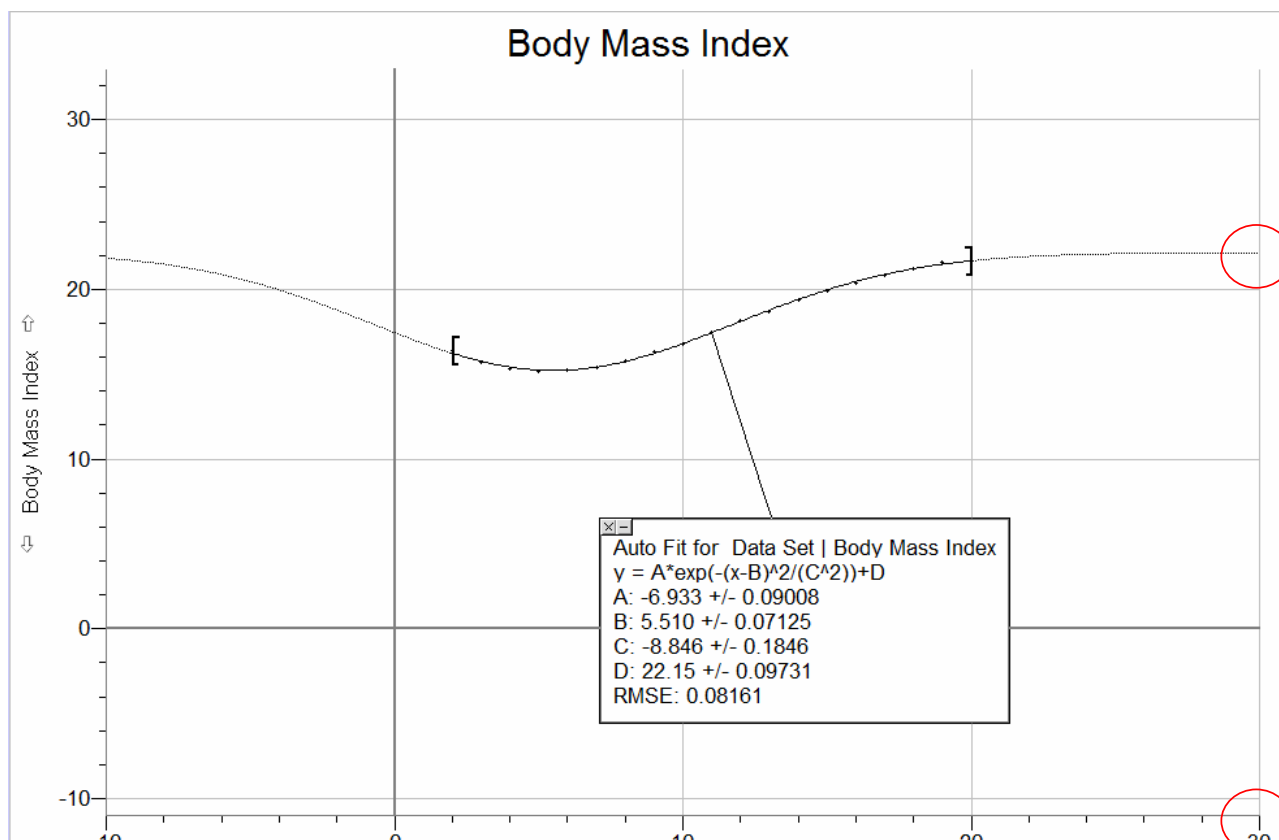
12 year old girl BMI: 18.70

If the BMI is calculated through taking the weight and dividing it by the square of the height, either a 30 year old women's weight suddenly decrease a lot or the height increase a lot during age 20-30 (which is impossible due to fact that girls finish growth and reach maturity around 14).

If the quartic graph does not fit, then I will look at other graphs which can show a realistic BMI for 30 year old women

After looking at the sine, cubic, quartic, quintic graphs I noticed there is a deep decline in the BMI from age 20-30 so I had to look for a new function.

From clicking the equations available in the Curve Fit window I found that the



Body Mass Index Maths Coursework

by Warren Hui

Gaussian function shows a realistic trend from age 20-30.

From the Gaussian graph we can tell a 30 year old women's BMI is 22.21, this is slightly higher than the 20 year olds BMI which is 21.65. It's reasonable to assume that a 30 year old women's BMI will be slightly higher than a 20 year old's BMI.

So I decided that the Gaussian graph shows a reasonable estimate for a 30 year old's BMI.

6. Use the Internet to find BMI data for females from another country. Does your model also fit this data? If not, what changes would you need to make? Discuss any limitations to your model

I found a graph of Bahraini girl and their BMI and age. I will now place this data into a graph form and compare it to the sine function.

Source: http://www.nature.com/ijo/journal/v24/n9/fig_tab/0801371t2.html

Table 2 Sample size per age group, means and standard deviations (s.d.) for six measurements of body fat of Bahraini girls

Age (y)	Sample size	Sum of skinfold (mm)		Body mass index (kg/m ²)		Mid arm muscle circumference (cm)		Percentage body fat		Fat weight (kg)		Lean body weight (kg)	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
6-6.9	23	29.8	11.5	15.4	2.2	12.9	0.9	13.8	2.6	2.9	1.0	17.5	2.2
7-7.9	81	27.2	12.5	15.2	2.2	12.7	1.2	13.5	3.1	3.0	1.5	18.2	3.1
8-8.9	83	30.2	14.9	15.6	2.4	13.6	2.0	14.0	3.6	3.5	1.8	20.7	4.1
9-9.9	72	28.3	13.1	15.5	3.1	13.6	1.5	13.6	3.1	3.6	2.4	21.7	4.1
10-10.9	81	36.1	21.4	16.8	3.7	14.7	1.9	15.4	4.8	5.1	3.5	25.5	5.1
11-11.9	50	36.7	19.2	17.5	3.7	15.7	2.5	15.4	4.4	5.9	4.3	29.2	7.1
12-12.9	57	42.0	19.2	18.5	3.8	16.4	1.8	16.4	4.5	6.9	3.6	33.3	7.1
13-13.9	47	54.1	17.1	18.2	4.3	17.3	2.8	17.0	4.1	7.8	3.9	36.2	7.1
14-14.9	70	53.8	24.6	21.1	4.8	17.8	2.2	19.2	5.7	10.2	6.4	39.8	7.1
15-15.9	58	56.9	24.5	21.7	4.3	18.0	2.4	19.8	5.4	10.8	5.4	41.3	6.1
16-16.9	43	50.7	19.0	21.1	4.6	18.1	2.3	18.6	4.5	10.2	5.5	41.9	8.1
17-17.9	56	59.3	22.3	22.5	5.0	18.4	3.1	20.8	5.1	11.9	6.0	42.7	7.1
18-18.9	49	56.7	21.3	21.9	4.0	18.1	2.7	19.8	4.9	10.7	5.2	41.4	6.1

To be able to compare the data of this table with my sine graph I will need to work out the sine function.

$$y = A \sin(Bx + C) + D$$

Body Mass Index Maths Coursework

by Warren Hui

D: Change vertical shift

Maximum = 22.5 (age 17-17.9)

Minimum = 15.2 (age 7-7.9)

$$(22.5+15.2)\div 2 = 18.85$$

$$D = 18.85$$

A: Change in amplitude

$$22.5 - 18.85 = 3.65$$

$$A = 3.65$$

B: Change horizontal scaling

$$\text{Maximum: } (17+17.9)\div 2 = 17.45$$

$$\text{Minimum: } (7+7.9)\div 2 = 7.45$$

$$17.45 - 7.45 = 10$$

$$10 \times 2 = 20$$

$$2\pi \div 20 = 0.3142$$

$$B = 0.3142$$

C: Change in horizontal shift

$$\text{Maximum: } 17.45$$

$$\text{Minimum: } 7.45$$

$$17.45 - 7.45 = 10$$

$$10 \div 2 = 5$$

$$5 + 5 = 10$$

$$10 \times 0.3142 = 3.142$$

$$C = 3.142$$

Finally

$$A = 3.65$$

$$B = 0.3142$$

$$C = 3.142$$

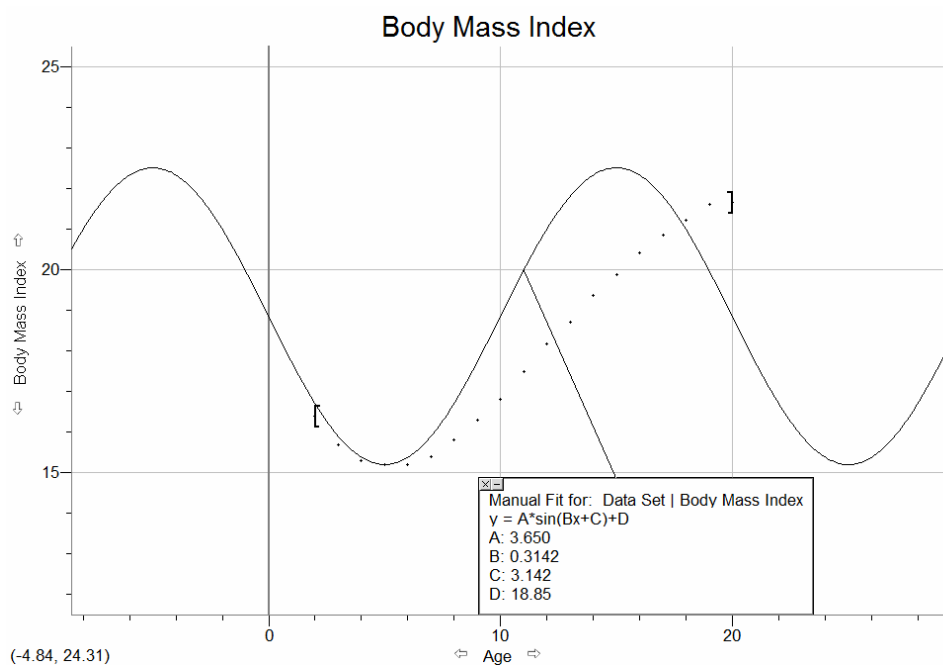
$$D = 18.85$$

$$y = 3.65 \sin(0.3142x + 3.142) + 18.85$$

RME to original sine function: 2.03322

Body Mass Index Maths Coursework

by Warren Hui



There are many limitations to the data of the BMI of the Bahrani girls. The first one being the age is a range which really decreases the accuracy of my manual calculation. As I am forced to take the average between the lower bound and upper bound of an age range. Second the age only goes up to 18-18.9 and not 20 like the other data.

There could also be an anomaly in the data presented this is because the maximum point is 22.5 for an age 17-17.9 while all the other girls are around 21.1-21.9 from age 14-18.9. This could explain the height difference of the graph for the Bahrani girls.

However apart from this anomaly the general shape of the curve B greatly matches the curve of graph A especially from age 2-6 when both graphs are touching.

Changes I will make will be to not take into account the anomaly 22.5 and see what happens.

D: Change vertical shift

Body Mass Index Maths Coursework

by Warren Hui

Maximum = 22.5 (age 17-17.9) if I don't use 22.5 I'll use the second highest number 21.9

Minimum = 15.2 (age 7-7.9)

$$(21.9+15.2)/2 = 18.55$$

$$D = 18.55$$

A: Change in amplitude

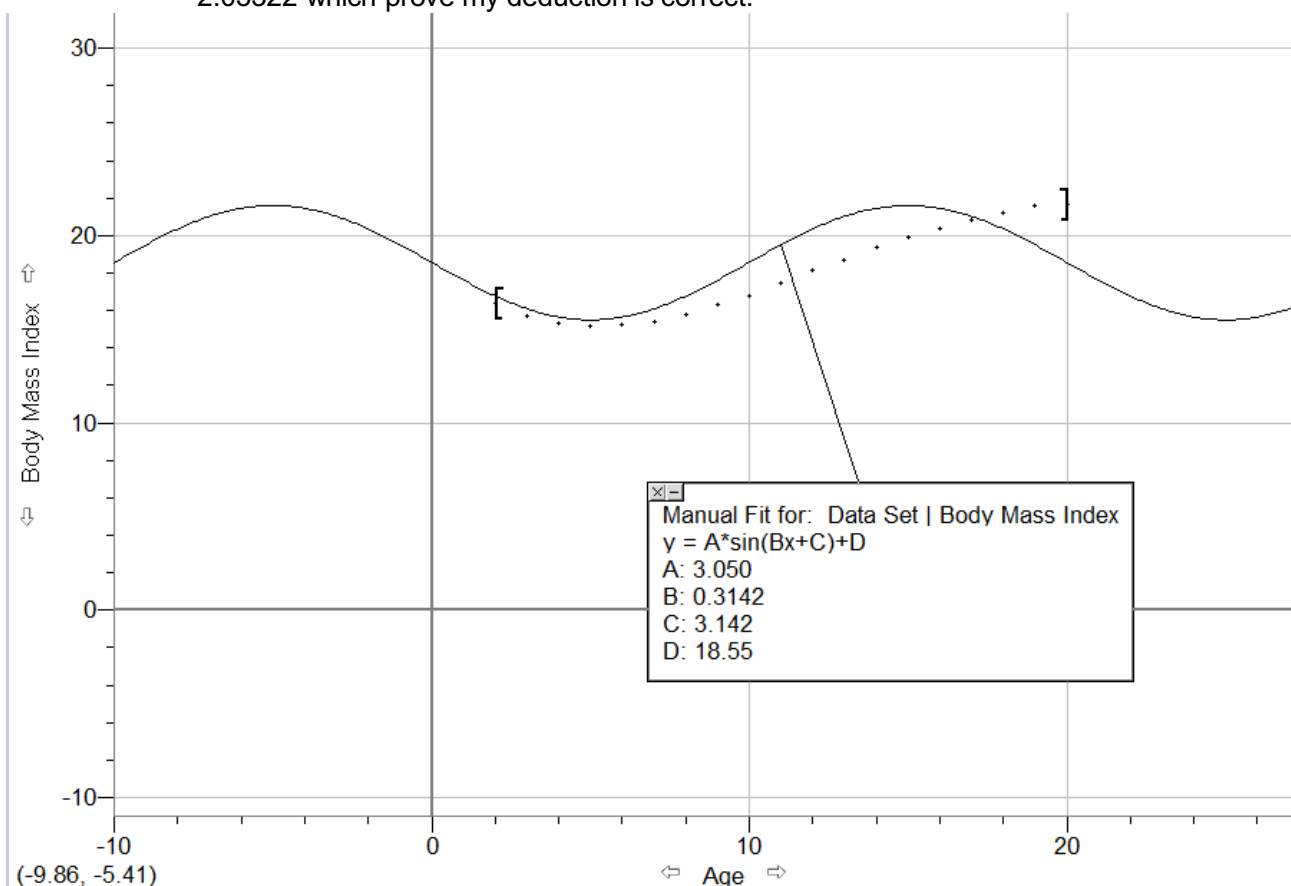
$$21.9 - 18.85 = 3.05$$

$$A = 3.05$$

This changes the function to:

$$y = 3.05 \sin(0.3142x + 3.142) + 18.55$$

The RMSE becomes 1.717 which is smaller than the original Bahrani graph's 2.03322 which proves my deduction is correct.



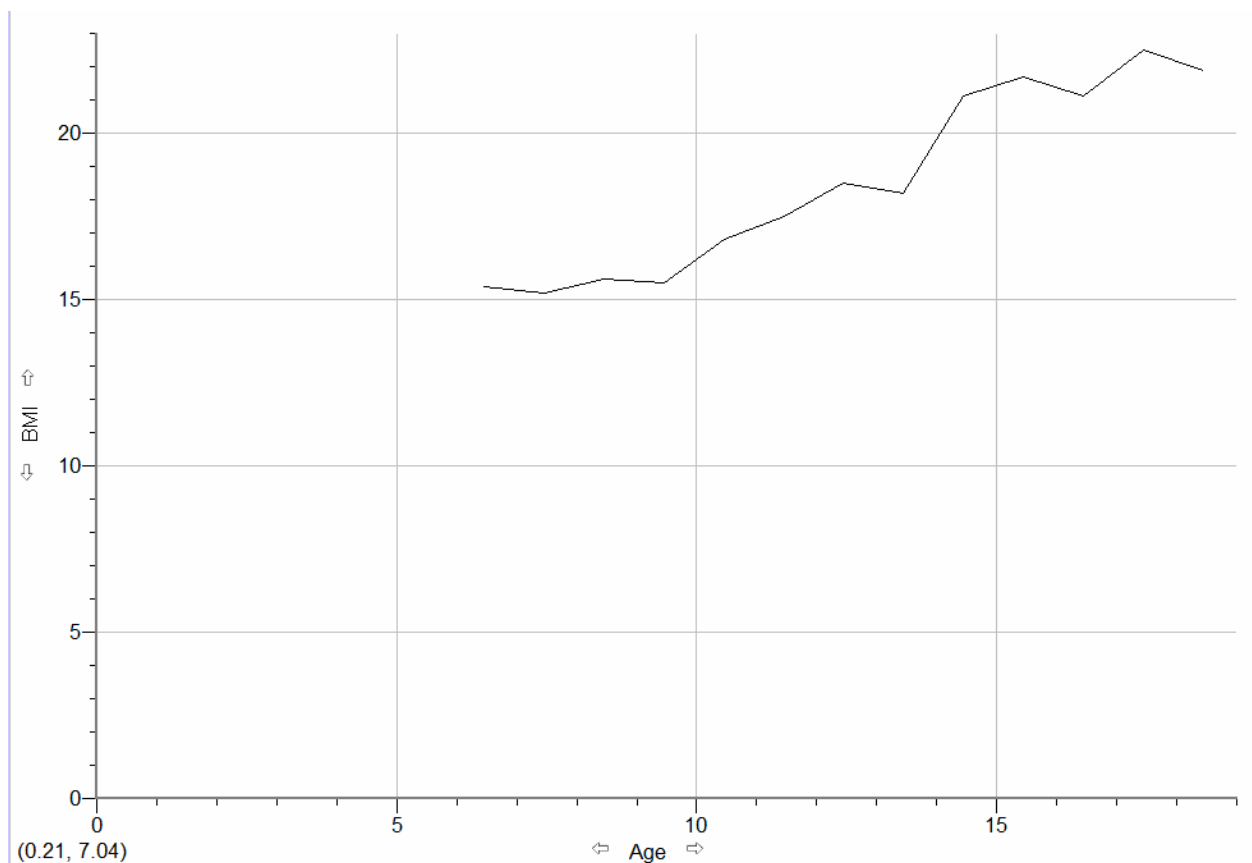
Still the sine graph produced from the data is not close enough and I am not satisfied with the results, so I will look at other functions and try to find one that

Body Mass Index Maths Coursework

by Warren Hui

fits more.

To do this I first have to model the data into a graph. As you can see, it is not a perfect curve. This is due to a few limitations. Firstly the ages are in ranges (which mean I have to add the two ages and divided by 2 to produce one value for the age). Also the BMI is a mean, which means it only shows the average which can result in this graph. In order to compare this with the graph made from the original BMI data, I must first find a function to model this graph.

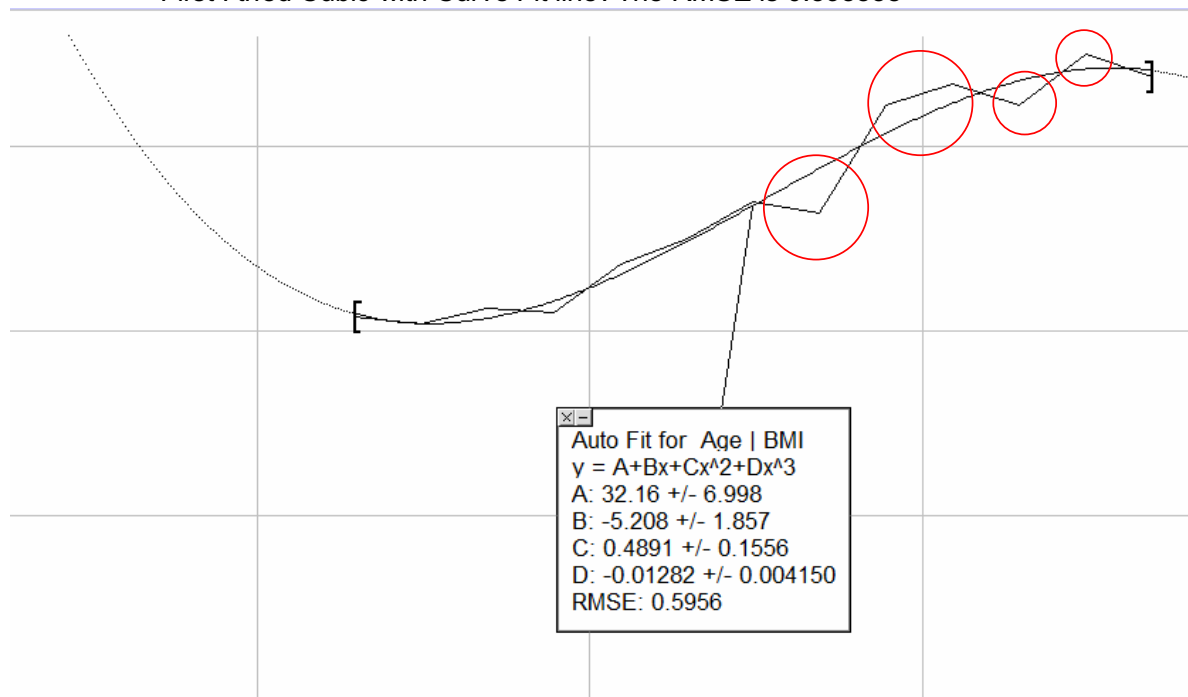


I will try Cubic, Quintic, Quartic and Gaussian because these are the previous functions I have used. All I have to do is click Curve Fit and the computer automatically creates a graph that best fits using the equation I have selected.

Body Mass Index Maths Coursework

by Warren Hui

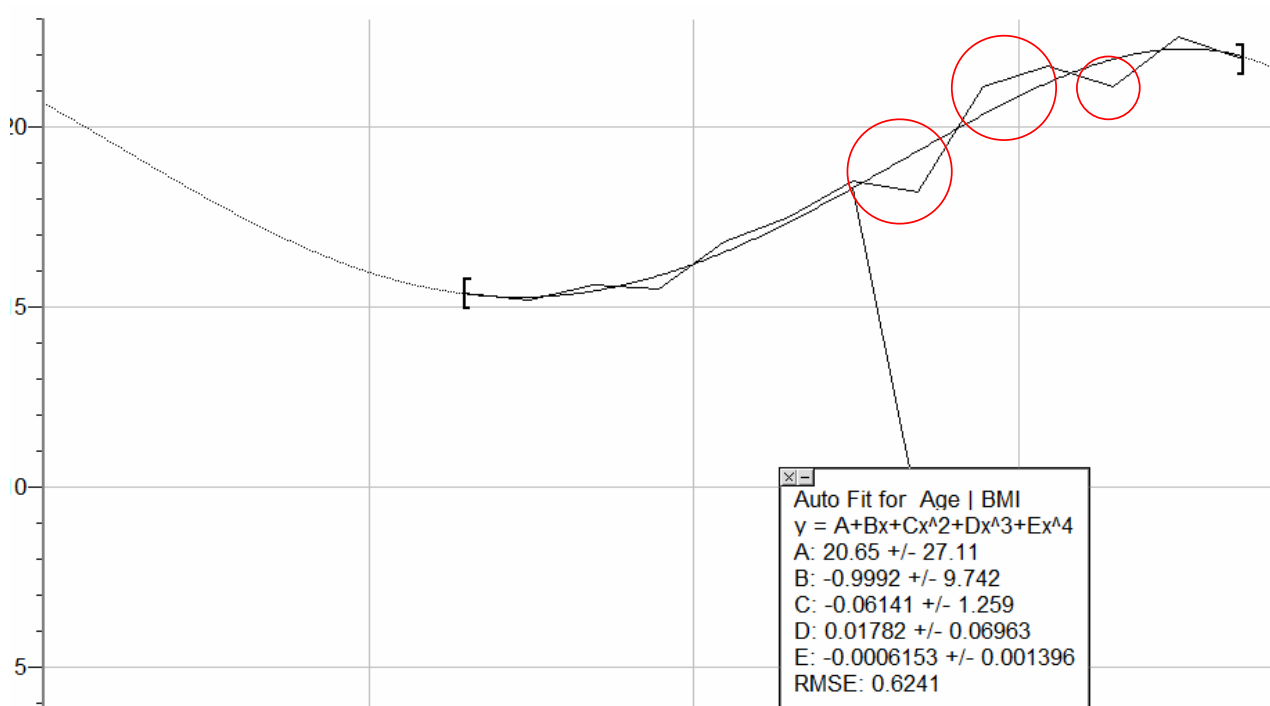
First I tried Cubic with Curve Fit line. The RMSE is 0.595558



Red circles show the differences between Bahrani gir I graph and Curve Fit line

Quartic graph with Curve Fit line

RMSE: 0.624147

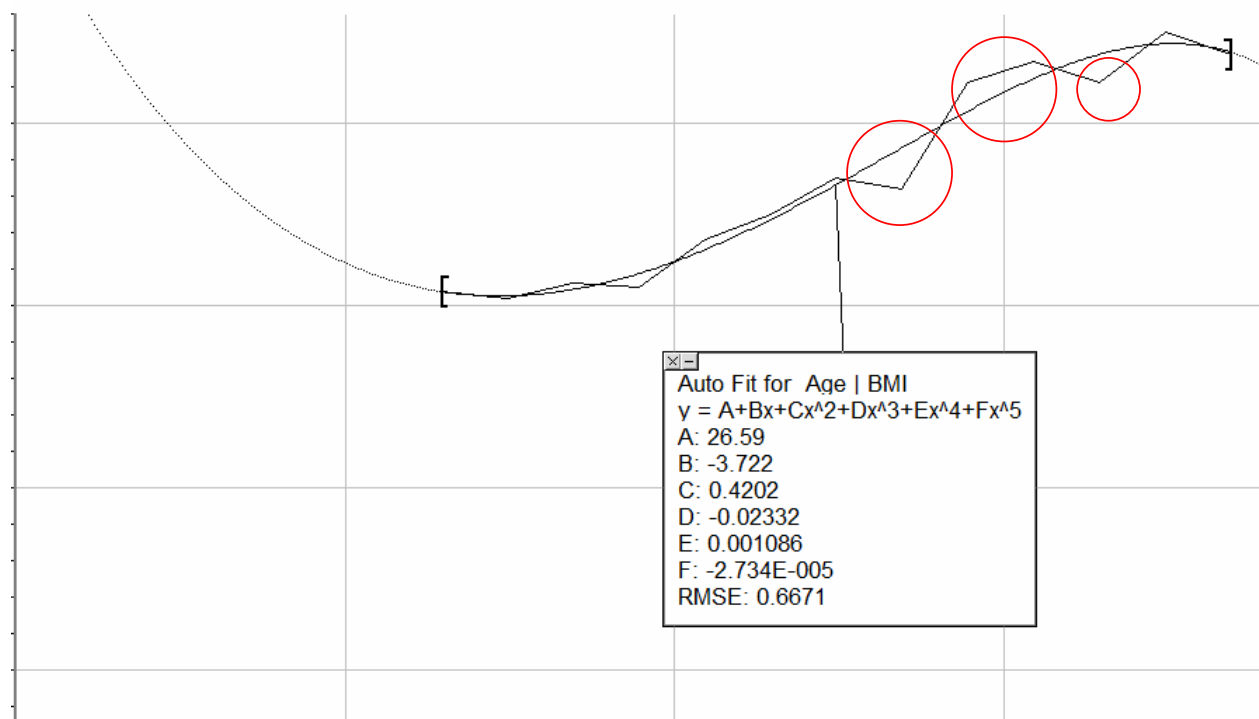


Body Mass Index Maths Coursework

by Warren Hui

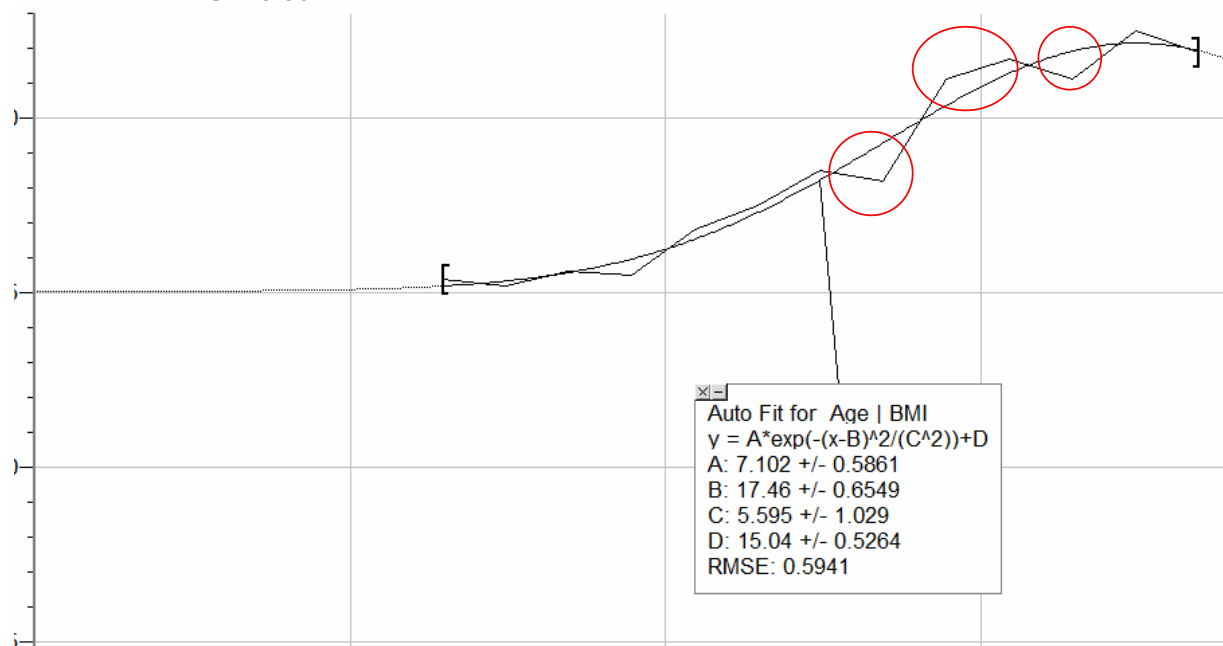
Quintic with Curve Fit line

RMSE: 0.667095



Gaussian with Curve Fit Line

RMSE: 0.594122



Body Mass Index Maths Coursework

by Warren Hui

After looking at all these equations, we can now tell indeed all the functions that fit the original BMI data fits the Bahraini women data as well.

However I also noticed that after 20 or 30, the graphs for the Bahraini girls started to drop significantly. This shows that the equations are not reasonable; the reason for the graph dropping may be insufficient data. Or because the data is the average.

Even with this weakness in the data, I chose to use the Gaussian graph of the Bahraini women to compare with the original data graph because the RMSE for the Gaussian graph is the lowest.

Gaussian graph values:

A: 7.102

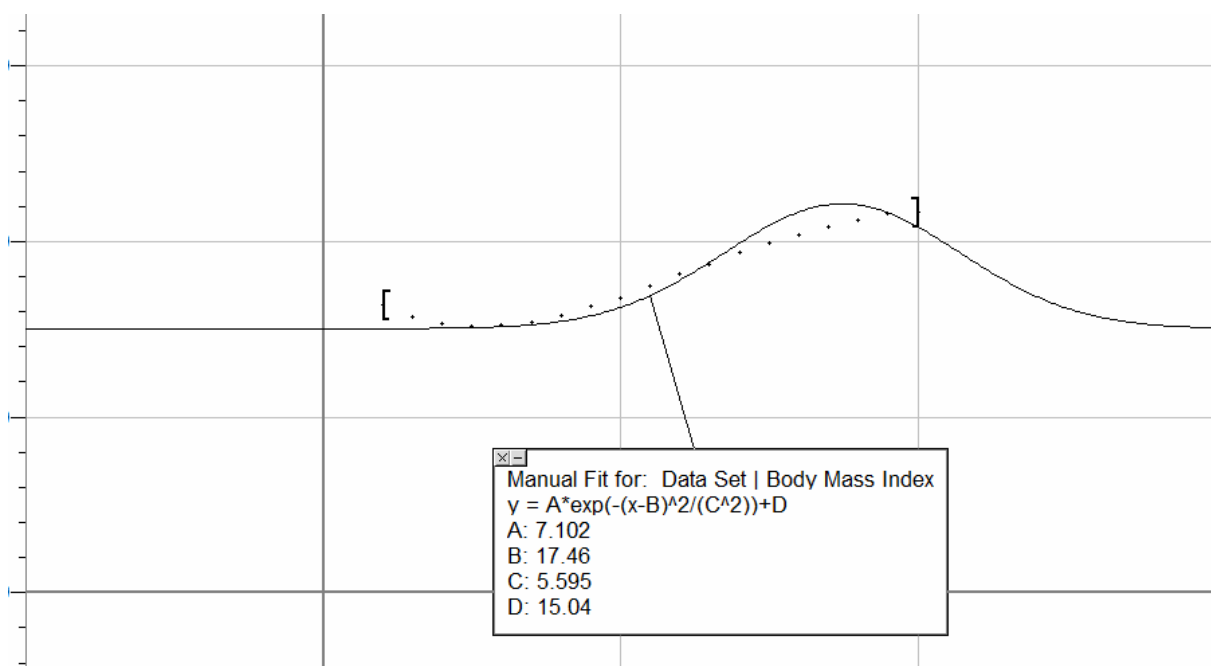
B: 17.46

C: 5.595

D: 15.04

The Gaussian graph compared with the original data graph

RMSE: 0.795802



Body Mass Index Maths Coursework

by Warren Hui

Conclusion

From looking at all these functions modeled from the Bahraini girl's graph, we can tell it fits with the original BMI data for women.

After completing the task, I have found a flaw using the Gaussian function. Since it is named the 'bell shape' graph, after age 30 it drops back down to a straight line. This is extremely unreasonable, as women's BMI will not drop significantly after age 20 to 15.1, which is equivalent to the BMI of a 3 year old. However due to the limited data provided for the Bahraini girl, and I've tried every single function, there is no solution to that issue. However this can be solved by gaining more data on Bahraini women from age 19-30.

However from accuracy point of view a RMSE of 0.795802 when comparing the Gaussian graph and the original BMI of women shows the similarities. Based on my findings, the function that fits best is Gaussian, even though when it drops down to a constant of 15.1 it is still more reasonable than the Quartic graph which drops eventually to a 14.3. Dropping to 15.1 on the Gaussian graph and staying at that is reasonable if the BMI was slightly higher, this is because after age 30, a women's BMI should not fluctuate much, if at all.

