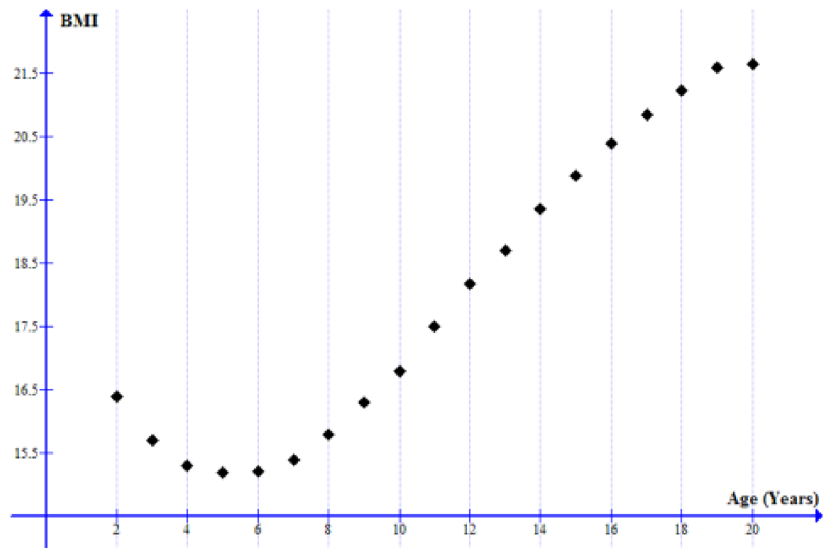
Calculus 1 IB - Body Mass Index – SL Type II

Introduction

This is an investigation of body mass index (BMI) and age of females between 2 and 20 years old to find a model that can represent given data and a method to find such a model.

Section 1 – Given Data



The variables used in the given graph are age in years and BMI. The age is independent and BMI is dependent.

The general trend of the BMI tends to rise as the age of subjects increase. This generalization is not consistent from age 2 to 6. In this range, the median BMI of females drop to a minimum and then rises for the remainder of the given graph.

The scale of this graph is adequate for the given data. The data can be fully represented with no vertical or horizontal overflow. Additionally, in this scale, the contours of the graph can easily be analyzed.

Section 2 – Finding a Model

There are a few functions that could model the behavior of this graph. A polynomial whose degree is higher than 2 could conceivably match closely with the given data. Another function may be a periodic, implemented by $sine$ or $cosine$. A periodic function might model the data if it was to continue indefinitely and a polynomial with a high degree might describe the given data's curve precisely if electronically computed.

Because it is difficult to find a polynomial that can correctly model the given data by hand, a model using $sine$ will be found instead.

The equation, $\boldsymbol{a\sin(bx+c)+d}$, gives the parameters that must be found. The easiest parameter to find for the given graph is doubtlessly a.

$$21.65 - 15.20 = amplitude = 6.45$$

Using the given data table, the range of the data table can be found by subtracting the lowest BMI from the highest BMI, 21.65 at 20 years old and 15.20 at 5 years old. This finds the amplitude.

$$\frac{amplitude}{2} = \frac{6.45}{2} = 3.225$$

The amplitude can then be divided by two to find the a parameter. The a parameter must also be negative in order to flip the $sine$ curve over. The d parameter can be found next by using the amplitude found previously.

$$15.20 + 3.225 = 18.425$$

The lowest BMI plus half the amplitude can be used to find the d parameter, 18.425. This leaves two parameters, $b\ and\ c$.

$$-3.225\sin(x) + 18.425$$

The current equation must be horizontally shifted to right. The current period is not modified, it is 2π.

$$maximum\ of\ 1\ cycle \approx 4.71$$

$$mimimum\ of\ 1\ cycle \approx 1.57$$

$$maximum - minimum = 4.71 - 1.57 = 3.14 \approx \pi$$

The minimum and maximum of the current model can help determine the b parameter. Each cycle of the current model is π, which is closely mirrored by finding approximate values for the maximum and minimum.

$$-3.225\sin\left(\left(\frac{\pi}{15}\right)x\right) + 18.425$$

According to the given data, a single cycle of the body mass index of a female is close to 15 years. The current model can be stretched to meet that requirement by multiplying x by $\frac{\pi}{15}$. The current model still does not match the given graph adequately. To align the current model, the c parameter must be found.

$$given\ point = (11, 17.50)$$

$$current\ model, f(11) \approx 17.1133$$

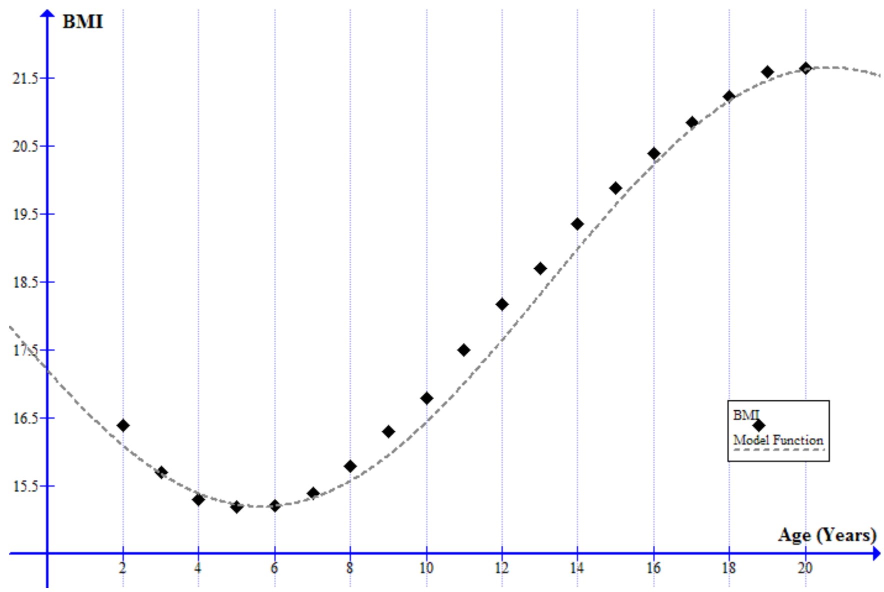
$$horizontal\ shift = 17.50 - 17.1133 = .3867$$

By taking an arbitrary point in the given data table using the x-value as input for the current model, a difference is found. This this difference is the c parameter, the horizontal shift. The model now is sufficiently aligned with the given graph.

$$-3.225\sin\left(\left(\frac{\pi}{15}\right)x + .3867\right) + 18.425$$

Section 3 – Revising the Model

Comparing the original data and graph to the new model is necessary at this stage.



The model created in **Section 2** appears to the right as a dashed gray line. The diamond points are the original data points given.

There are points that fall on the model's curve, such as $(20,21.65)$ and $(4,15.30)$. While those points fit well, there are indeed points well off the curve such as $(12,18.18)$ and $(14,19.36)$. It seems that a larger horizontal shift to the left would correct this discrepancy, but fixing it may cause the given points that *do fit the model's curve* to become unaligned.

The data table to the right contains the independent variable, age, in years and the two versions of the dependent variable, BMI in given and created model form. The final column in the table is the difference between each value from the given data and the value found from the created model.

The difference column shows how close the created model is to the given data. The largest difference is age 12, **.5267**. The smallest difference is age 4, **-.0915.** The closer the difference is to zero, the better fit the model is. The average of the differences is **.1927**. That's reasonably close to zero.

Age (Years)	BMI (Given)	BMI (Created Model)	Given - Created (Difference)
2	16.4	16.099	0.301
3	15.7	15.6854	0.0146
4	15.3	15.3915	-0.0915
5	15.2	15.2302	-0.0302
6	15.21	15.2085	0.0015
7	15.4	15.3274	0.0726
8	15.8	15.5816	0.2184
9	16.3	15.9602	0.3398
10	16.8	16.4464	0.3536
11	17.5	17.0192	0.4808
12	18.18	17.6533	0.5267
13	18.7	18.3212	0.3788
14	19.36	18.9937	0.3663
15	19.88	19.6413	0.2387
16	20.4	20.2357	0.1643
17	20.85	20.751	0.099
18	21.22	21.1646	0.0554
19	21.6	21.4585	0.1415
20	21.65	21.6198	0.0302

Refining the current model is possible at this point. The average difference, .1927 can be added to the current d parameter, .3867 to result in .5794. This establishes a minimum and maximum shift needed to align the current model to the given data. Taking the average of these two numbers results in **.48305**. This is the new c parameter value.

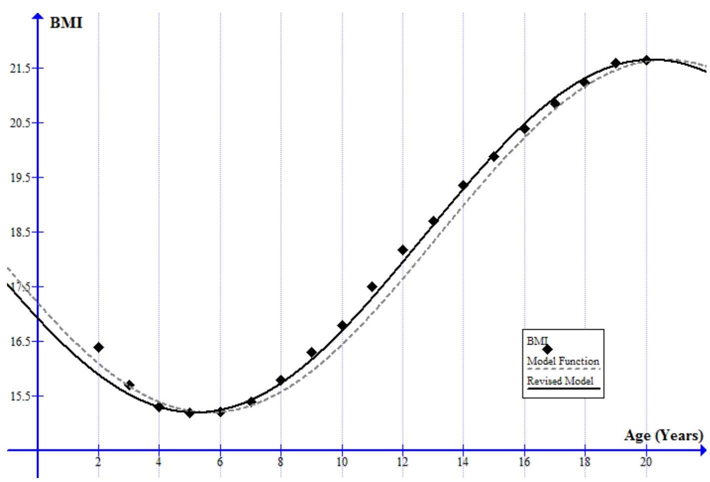
$$avg.difference = .1927$$

$$minimum = .3867$$

$$maximum = .3867 + .1927 = .5794$$

$$\frac{(minimum + maximum)}{2} = .48305 = d$$

$$\boldsymbol{revised\ model = -3.225 \sin\left(\left(\frac{\pi}{15}\right)x + .48305\right) + 18.425}$$



This revised model, the solid curve to the left, appears to fit in line with the given data better than the previous model. The mid-section of the graph (10-18) is generally closer to their given counterparts than it was in the previous model in the same range.

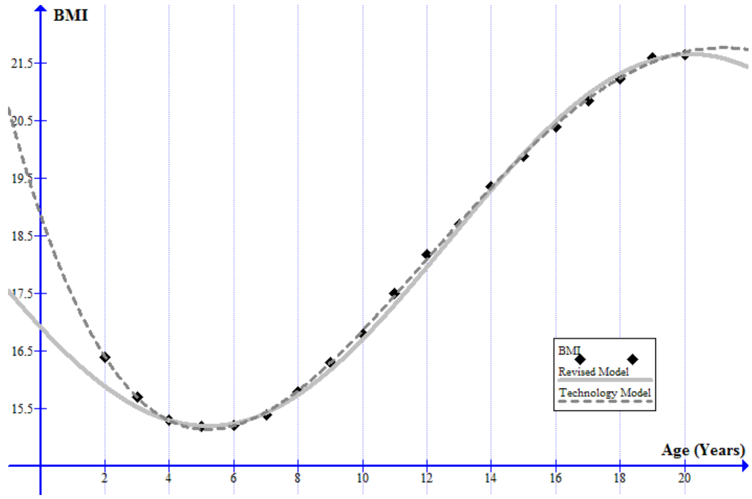
The average difference between the given data and the revised model is closer to zero than the previous model, **.06127**. This revised model is indeed a better fit for the given data.

Section 4 – Finding a Model with Technology

Another model for the given data might be a polynomial. A polynomial with a degree higher than 2 could model the given data well. Using technology, a fourth degree polynomial regression, an equation can be found. This equation is troublesome to repeat, so it shall simply be called the *technology model*.

$$(9.7256554 \times 10^{-5})x^4 - (0.0083538185)x^3 + (0.21673969)x^2 - (1.6300002)x + 18.863421$$

The revised model that was created by hand is represented by the solid line; the dashed line represents the technology model. The diamonds are for reference to the given data.



The end behavior of the technology model is out of scope when looking at a graph at this scale; however, it is good to note that the behavior would not match the reality of BMI; a 90 year old would have a BMI close to 2000, which is impossible. The two models appear visually to be close to the given data in the given scale. The technology model might be slightly more in line with the given data, however. Notice that a baby that is just born, which would be on the y-axis, is close to a BMI of 18.6. That is probably unreasonable for a newborn baby.

Looking at a data table with the given data and technology model differences also shows how closely the technology model mirrors the given data.

Age (Years)	BMI (Given)	Technology Model	Given - Tech. Model Difference
2	16.4	16.4051	-0.0051
3	15.7	15.7064	-0.0064
4	15.3	15.3015	-0.0015
5	15.2	15.1485	0.0515
6	15.21	15.2077	0.0023
7	15.4	15.4418	-0.0418
8	15.8	15.816	-0.016
9	16.3	16.2975	0.0025
10	16.8	16.8561	-0.0561
11	17.5	17.4639	0.0361
12	18.18	18.0952	0.0848
13	18.7	18.7268	-0.0268
14	19.36	19.3377	0.0223
15	19.88	19.9093	-0.0293
16	20.4	20.4253	-0.0253
17	20.85	20.8718	-0.0218
18	21.22	21.2372	-0.0172
19	21.6	21.5122	0.0878
20	21.65	21.6898	-0.0398

The differences are very close to 0. Many of these are too far left, so they are negative values. Positive values would indicate that they are too far to the right. The average difference in this case is extremely small, much smaller than the average difference for the revised model. The exact value of the average is **0.0000105263157887947**. It is agreeable that this average difference exceeds the level similarity to the given data than either model created by hand.

Section 5 – Testing the Models

The two models at this point are the revised model created by hand and also the technology model. They both have their uses. For instance, the revised model is periodic, so it could theoretically be used to find the body mass index of someone older than 20 and still be somewhat relevant. The technology model would be used in cases where extreme accuracy to the given data is required, as mentioned in **Section 4**, the level of similarity between the given data and the technology model is unrivaled.

To find the BMI at birth and other values that exceed 20, the technology model cannot be used since it is unreasonable for a baby to have an 18.6 BMI and a 90-year old to have 2000 BMI. Therefore, the revised model will be used.

i) BMI of a newborn - By using the revised model function, the BMI at birth can be found by setting the x parameter to 0.

$$-3.225 \sin\left(\left(\frac{\pi}{15}\right)(0)+.48305\right)+18.425 \approx 16.9270$$

The BMI at birth is 16.9270. This is reasonable, because babies tend to have little height and quite a bit of weight.

ii) BMI of a 17 Year Old - Again, using the revised model function, set the x parameter to 17.

$$-3.225 \sin\left(\left(\frac{\pi}{15}\right)(17)+.48305\right)+18.425 \approx 20.9551$$

The BMI of a female age 17 is 20.9551. This is reasonable according to the given data; a 17 year old female should have about 20.85 BMI.

iii) BMI of a 30 Year Old - This is where both the revised model and the technology model fail slightly, the inability to accurately predicting behavior outside of the original scale. The technology model does not function at all when the age is 30, so the revised model must be used.

$$-3.225 \sin\left(\left(\frac{\pi}{15}\right)(30)+.48305\right)+18.425 \approx 16.9270$$

The BMI of a 30 year old female probably not 16.9270. It is reasonable to assume there is fluctuation in life, where the BMI of a person can rise and drop over time, but it does not seem likely that a 30 year old is 16.9270 when a 10 year old female would have the same BMI too. At the very least, the revised model gives a value within a range suitable for BMI, compared to the technology model.

iv) BMI of a 40 Year Old - This is similar to *iii*. The revised model cannot account for an age outside of the original scale, but it can provide at the very least a number that is suitable for a BMI.

$$-3.225 \sin\left(\left(\frac{\pi}{15}\right)(40)+.48305\right)+18.425 \approx 16.7006$$

v) BMI of a 90 Year Old - The BMI of a 90 year old woman should be lower than a woman in their 50's because as people age, they get slightly shorter and usually thin out. According to the revised model, a 90 year old woman has the same BMI as that of a 30 year old woman. This is possible but again unlikely and hard to predict without data going into the range of higher ages.

$$-3.225 \sin\left(\left(\frac{\pi}{15}\right)(90)+.48305\right)+18.425 \approx 16.9270$$

Section 6 – Applying the Model to Other Situations

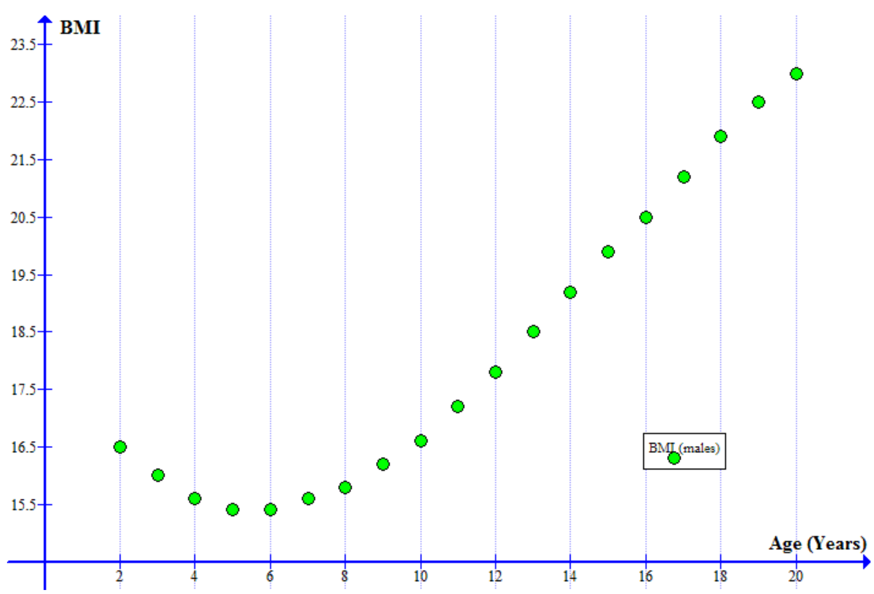
The median of all data on the males' median BMI graph is a curve marked as 50, that is, the 50th percentile. The 50th percentile represents half of the data and half of the data below the average. A given graph of males' BMI is used find the subsequent data table.

Age (Years)	BMI (male)	BMI (female)
2	16.5	16.4
3	16	15.7
4	15.6	15.3
5	15.4	15.2
6	15.4	15.21
7	15.6	15.4
8	15.8	15.8
9	16.2	16.3
10	16.6	16.8
11	17.2	17.5
12	17.8	18.18
13	18.5	18.7
14	19.2	19.36
15	19.9	19.88
16	20.5	20.4
17	21.2	20.85
18	21.9	21.22
19	22.5	21.6
20	23	21.65

This data table contains the original given data for females and additionally the male data found by *only looking* at a graph for male BMI, so these values are approximations of the eye.

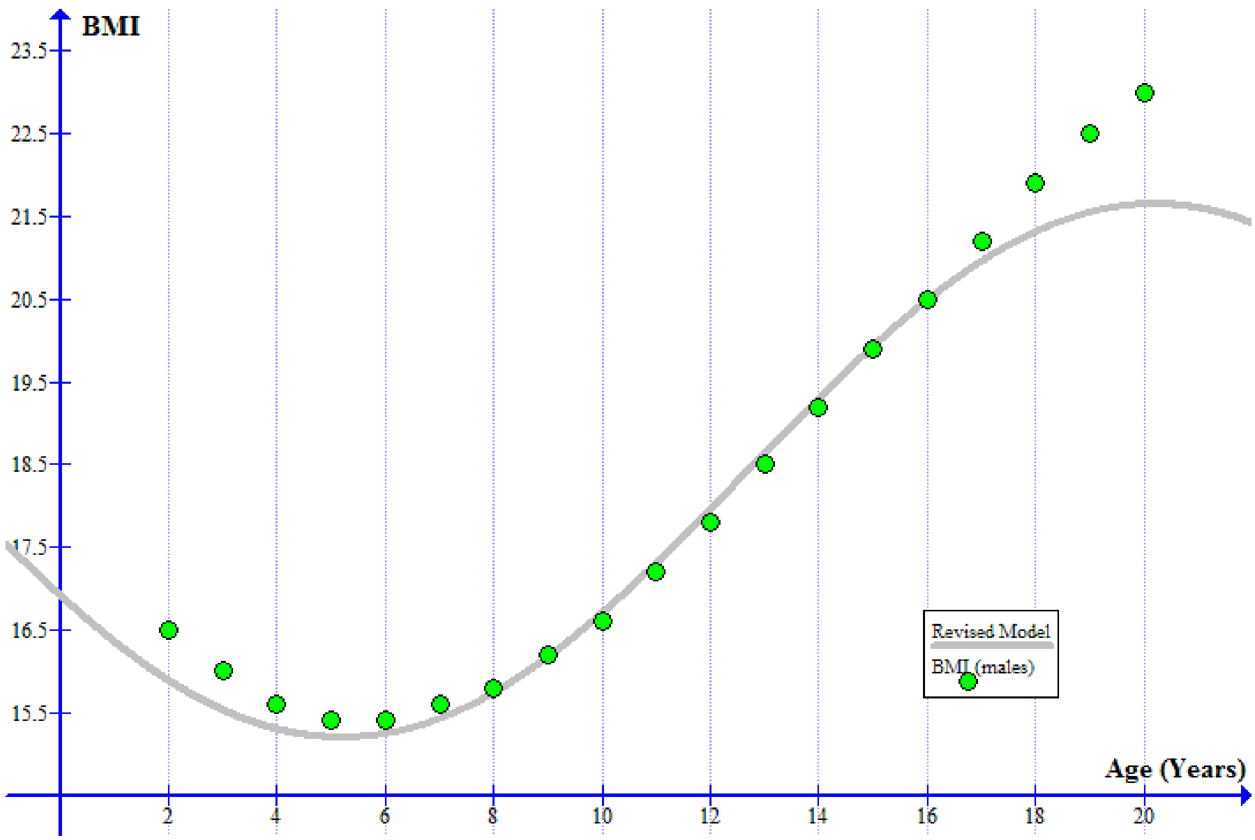
The graph to the right represents the male data extracted from the male's BMI graph.

There are no parameters concerning this graph, the variables involved are the same as the female data, the age and the BMI. Age is independent and BMI is dependent.



Section 7 – Comparing Related Data to the Model

Since the male data is now available, it can be compared to the model function found for the female given data.



The graph above contains the revised model found previously and also the points that represent the median BMI of males from age 2 to 20.

The revised model that was found analyzing female data fits some of the data for males. Notice the behavior of the male-points from age 2 to 6: the points deviate increasingly as x approaches 0. Also, notice the male-points from age 17 to 20: the points continue at an almost constant ascent while the revised model levels off and then drops. From age 7 to 16, the female based revised model fits the males' data generally well.

The female based revised model could be revised once more for the male data. A different amplitude exists for the male data, so a different a parameter would be needed and thus a different d parameter. The period remains the same for both sets of data, so the b and c parameters may remain the same.

Section 9 – Scope and Limitations of the Model

There are limitations for the revised model.

One such limitation is that is a $sine$ based model. This means that the values of BMI will be return eventually, after one full period. This both good and bad; it allows the model to give an idea of a BMI outside of ages 2 to 20, but it is also inaccurate for ages outside of 2 to 20.

Another limitation of the revised model is that it is comprised of values that are medians. The given BMI values of females age 2 to 20 are medians and this limits the variation in the data. This limited variation means that someone could easily be off the curve created by the revised model.

Because the revised model was found by hand, the model cannot be absolutely accurate even in the scale that is functions well in. This is proven to be correct by the average difference found in the revised model and the technology model. This is a limitation; models created by hand are difficult to be perfected.

Finally, a major problem with the revised model, which is also a limitation, is that it included negative values for BMI. This is absolutely impossible as a person cannot have a negative height or weight. The model could be constrained by a domain to prevent negative BMI values from appearing, if necessary.

Section 10 – Conclusion

A function that decently models the given female data has been found, it has been revised by analyzing a numeric representation. That model has been compared against a model found by using technology and it fared better in situations where a computed model simply could not. The model has been compared against similar data to see how well it matches. Finally, the scope and limitations of the model were found and explained. The graph below is a composite of graphs that have appeared previously; it is comprised of the given female data, the found male data, and the revised and technology models.