105

Mar. 13

Mathematics SL • Type II Portfolio Piece

INTERNATIONAL BACCALAUREATE

Body Mass Index

Body Mass Index is a measure of one's body fat. It is calculated by taking one's weight in kg and dividing it by the square of one's height in meters.

$$BMI = \frac{weight(kg)}{height(m)^2}$$

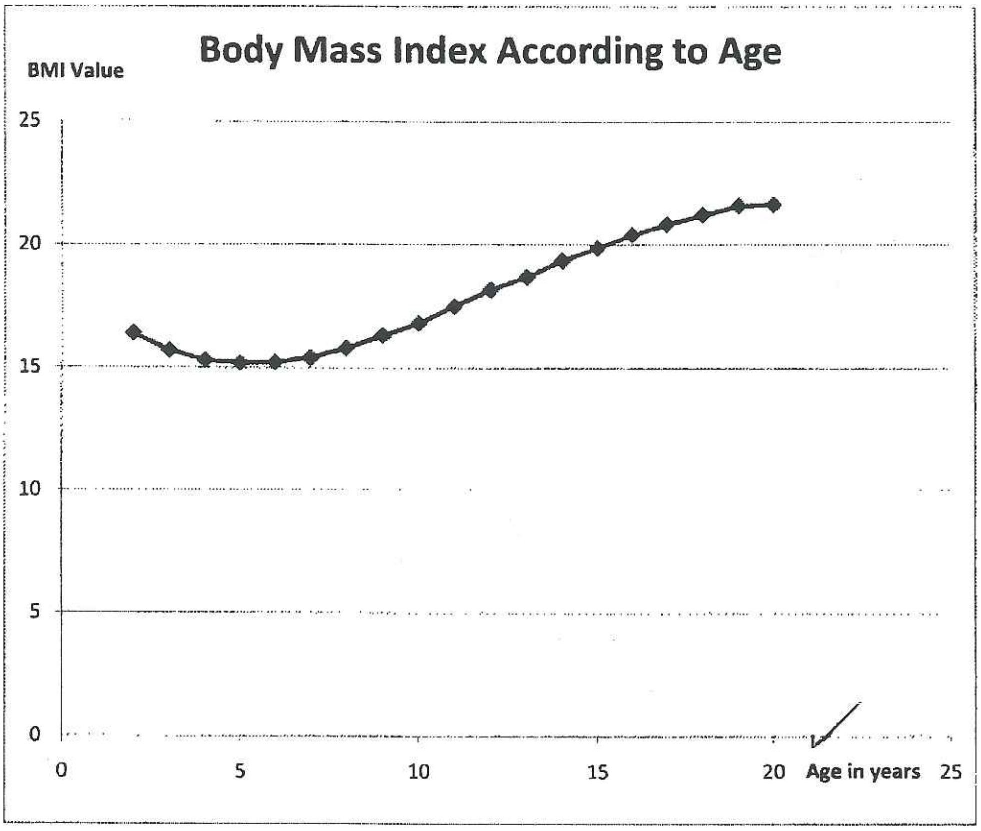
Portfolio Outline: GOOD SUMMARY!

- ✓ First I will plot the graph of the table using MS Excel and subs sequentially define all variables used and state the parameters employed.
- ✓ Secondly, I will manually (by hand) determine a quadratic function equation which models to behavior of the graph and justify why I chose this function.
- ✓ Afterwards, I will draw a graph with both the original line (from the table) and the new line discovered from the quadratic formula in order to juxtapose both lines and look for differences. I will also use a table and correlation coefficient values to display disparities from the original line.
- ✓ Next, I will refine the model using a cubic function (determined using the GDC) and draw its line on the graph with the original and show that it is more precise than the quadratic.
- ✓ Then I will display an amalgamation of graphs for different functions which could model the graphs such as quartic and discuss them all.
- ✓ Now I will use the cubic model I created in step 4 and apply it to attempt to find the BMI for a 30 year old woman.
- ✓ A new model will be created with the use of a cubic function and a quadratic function to more precisely model the age of a 30 year old female.
- ✓ This new model will then be compared with data from Guatemalan women.

The following is a table denoting the median BMI for females of different ages in the USA in the year 2000 obtained from the Center for Disease Control website.

2	16.4
3	15.7
4	15.3
5	15.2
6	15.21
7	15.4
8	15.8
9	16.3
10	16.8
11	17.5
12	18.18
13	18.7
14	19.36
15	19.88
16	20.4
17	20.84
18	21.22
19	21.6
20	21.65

The table has henceforth been illustrated in the graph below:



As we can see, by observing the form and shape of the curve, it is possible to denote that between the ages of 2 and 6, the BMI for young girls decreases. However, from ages 7 to 18, the BMI continues to increase progressively. Finally, at the age of 19, the graph shows a slight stagnating nature of the BMI of women, suggesting that the BMI index is beginning to stagnate. GOOD OBSERVATIONS

The variables used in this graph and table are as follows:

- Independent variable: "What do I change?"
  - The age of the women is changed.
- Dependent variable: "What do I observe?"
  - We then observe the change in the BMI value.
- Controlled variable: "What do I keep the same?"
  - The rate of change in years is always by 1 year.

C1

Mar. 13

the parameters:

- Definition:
  - A parameter can be defined as a constant in the equation of the curve that can be varied to yield a family of similar curves.
  - Therefore, according to this definition of parameter, the following could be parameters to the curve:
    - Amplitude
    - Period
    - Phase shift
    - Vertical shift
- However, the following could also be parameters to the curve:
    - The domain: the age of the subjects being from 2 years to 20 years old
    - The fact that only women are being analyzed.
    - The fact that these women are only from the United States of America.
    - The fact that these values are only medians, therefore do not denote distribution.
    - And the fact that these values apply only for the year 2000.

C2

Now we will manually create an equation for this graph using a quadratic formula. A quadratic type of equation was chosen because it provides a certain type of precision due to the exponents to the second power whilst also being relatively simple to create without the use of a GDC or other technological methods (by hand, essentially).

In order to create this equation, three points from the graph will be used. These points will be directly taken from the data table in order to preserve accuracy and create an equation in the form $y = ax^2 + bx + c$.

- At (4 ; 15.3) → $15.3 = a(4^2) + 4b + c$
  $15.3 = 16a + 4b + 1c$
- At (10 ; 16.80) → $16.8 = a(10^2) + 10b + c$
  $16.8 = 100a + 10b + 1c$
- At (16 ; 20.40) → $20.4 = a(16^2) + 16b + c$
  $20.4 = 256a + 16b + 1c$

It is henceforth possible to solve this tri-system of equations using matrix algebra:

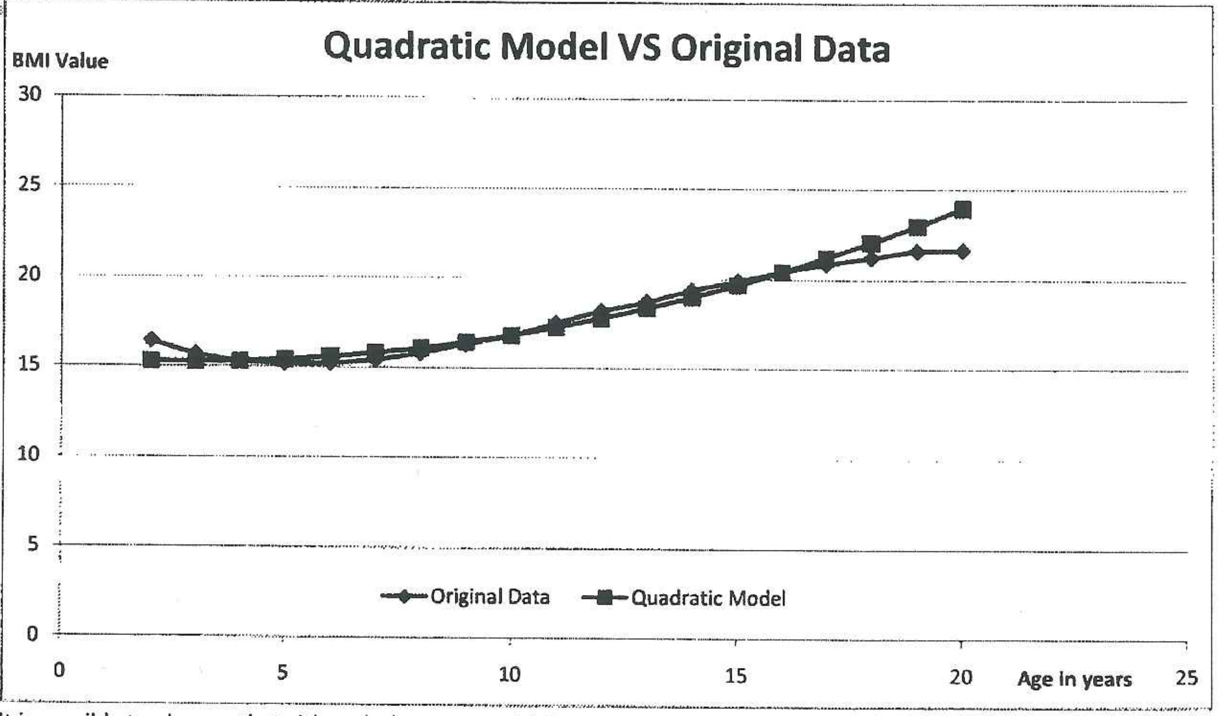
$$\begin{bmatrix} 16 & 4 & 1 \\ 100 & 10 & 1 \\ 256 & 16 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix}\begin{bmatrix} 15.3 \\ 16.8 \\ 20.4 \end{bmatrix}\end{bmatrix} \quad \text{With } \begin{bmatrix} 16 & 4 & 1 \\ 100 & 10 & 1 \\ 256 & 16 & 1 \end{bmatrix} = A \quad \begin{bmatrix} a \\ b \\ c \end{bmatrix} = X \quad \begin{bmatrix} 15.3 \\ 16.8 \\ 20.4 \end{bmatrix} = B$$

The laws of matrix algebra state that: $AX = B \therefore A^{-1}AX = A^{-1}B \therefore IX = A^{-1}B \therefore X = A^{-1}B$ (due to the identity matrix).

Therefore: $X = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \approx \begin{bmatrix} 0.0291 \\ -0.1583 \\ 15.4666 \end{bmatrix}$ → Using these values for a, b and c ; a graph using $y=ax^2+bx+c$ will be created using Excel modeling the results found through this manual quadratic equation.

We will now insert this model into the graph of the original data using Microsoft Excel and observe:

C3



It is possible to observe that although the new model does follow the original date trend until a certain point, it then becomes un realistic and only regains the trend at the last point.

2	15.26667	16.4	1.133334
3	15.25417	15.7	0.445834
4	15.3	15.3	6.65E-07
5	15.40417	15.2	0.204166
6	15.56667	15.21	0.356666
7	15.7875	15.4	0.387499
8	16.06667	15.8	0.266666
9	16.40417	16.3	0.104166
10	16.8	16.8	6.6E-07
11	17.25417	17.5	0.245834
12	17.76667	18.18	0.413334
13	18.3375	18.7	0.362501
14	18.96667	19.36	0.393334
15	19.65417	19.88	0.225834
16	20.4	20.4	6.53E-07
17	21.20417	20.84	0.364166
18	22.06667	21.22	0.846666
19	22.9875	21.6	1.387499
20	23.96667	21.65	2.316666
			Sum = 9.62

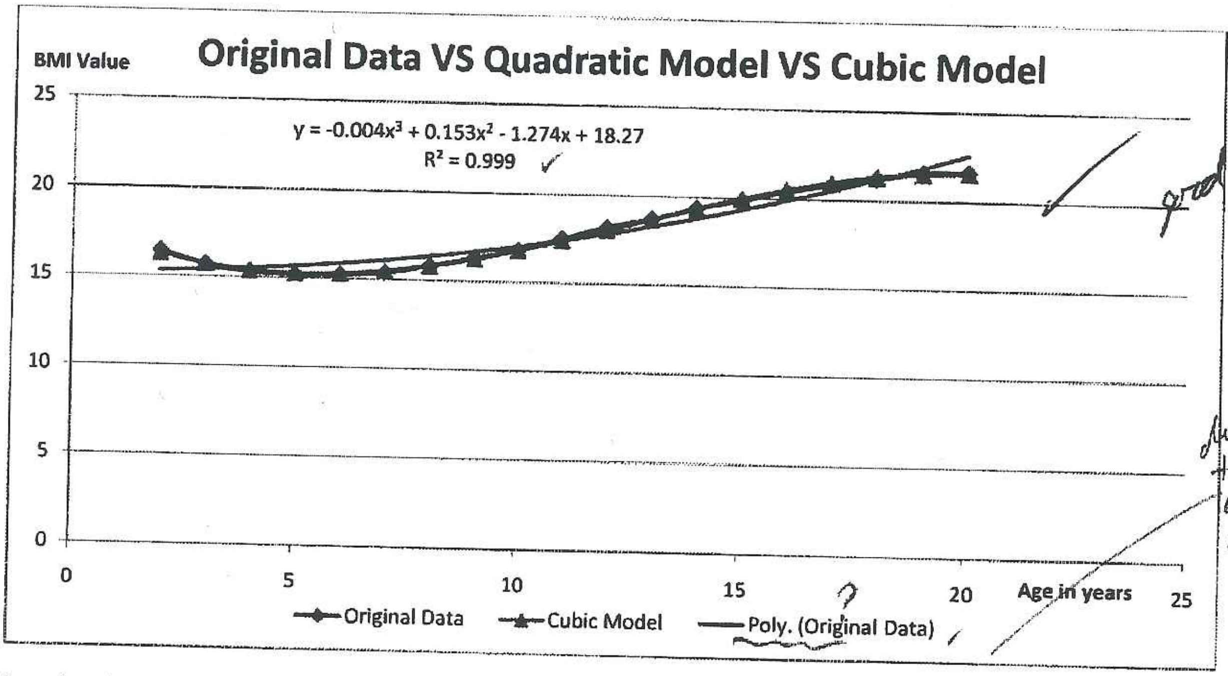
C4

Moreover, the above table (previous page) denotes the difference in values between the original data and that obtained through the quadratic equation found. The Sum value of 9.62 is a strong testament to the un-reliability of this method of modeling. MAYBE DIVIDE BY THE # OF VALUES & GET AN AVERAGE ERROR ? (HOWEVER, COMPARED BELOW)

A new model will therefore be devised and analyzed. This new model will be an equation of cubic form. I have chosen to use a cubic equation as the higher power value will lead to a more accurate result. The values for a, b, c and d will be determined using the "Cubic Regression" function of the GDC. WELL DONE

The following values were obtained: for $y = ax^3 + bx^2 + cx + d$ $\quad a = -0.0040745301 \quad c = -1.27573621$ ✓
$b = 0.1535646101 \quad d = 18.27287926$

Below is a table comparing both models created along with the original data using Microsoft Excel:



It can be observed, by looking at the graph, that the Cubic model does do a much better job than the Quadratic model at mimicking the original data. The new model (in green) almost entirely hides the original data line, meaning that a high data correlation is present, rendering the model very accurate. ✓

Moreover, a table of differences, similar to that created on page 4 has been created for this model. It states that the sum of all the differences between the original data and that obtained through the cubic function (all absolute values) is approximately 1.139 – a much smaller number than that of the quadratic model. Thus, it can be said that my original expectations concerning the Cubic equation were correct, as the cubic function is much more accurate than the quadratic. GOOD COMPARISON!

However, the correlation coefficient for this function is approximately 0.9991 – a value which is not to far way from 1 (the perfect correlation); therefore suggesting that a perfect model is perhaps close by. ✓

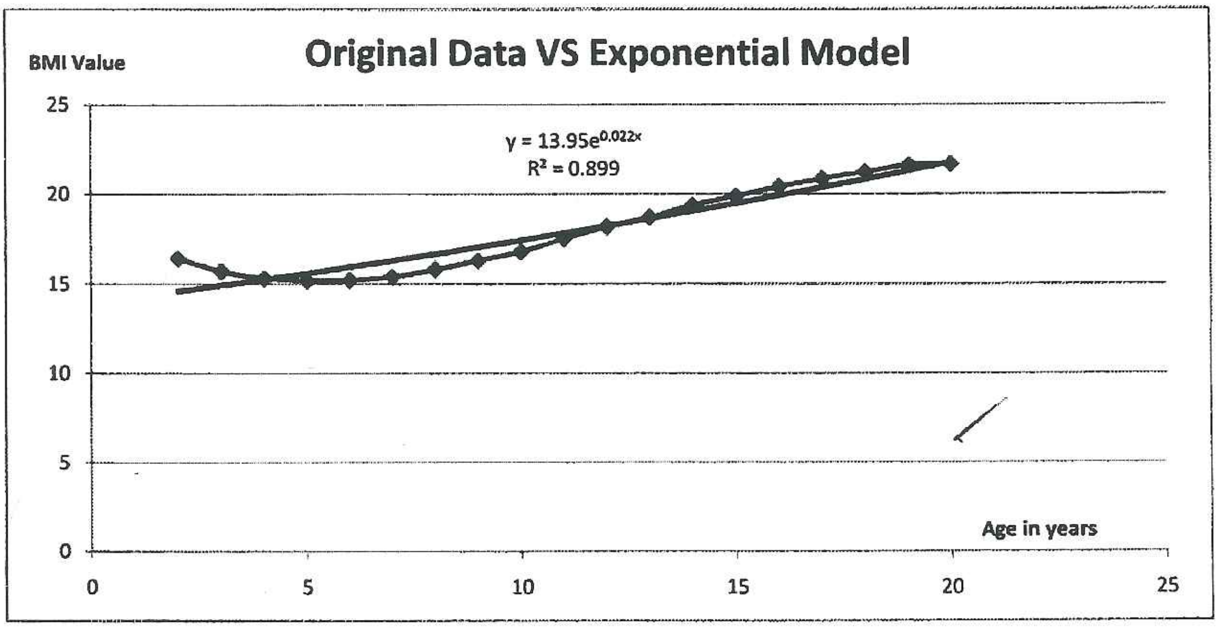
R^2 is not the correlation coefficient – This is the coefficient of determination

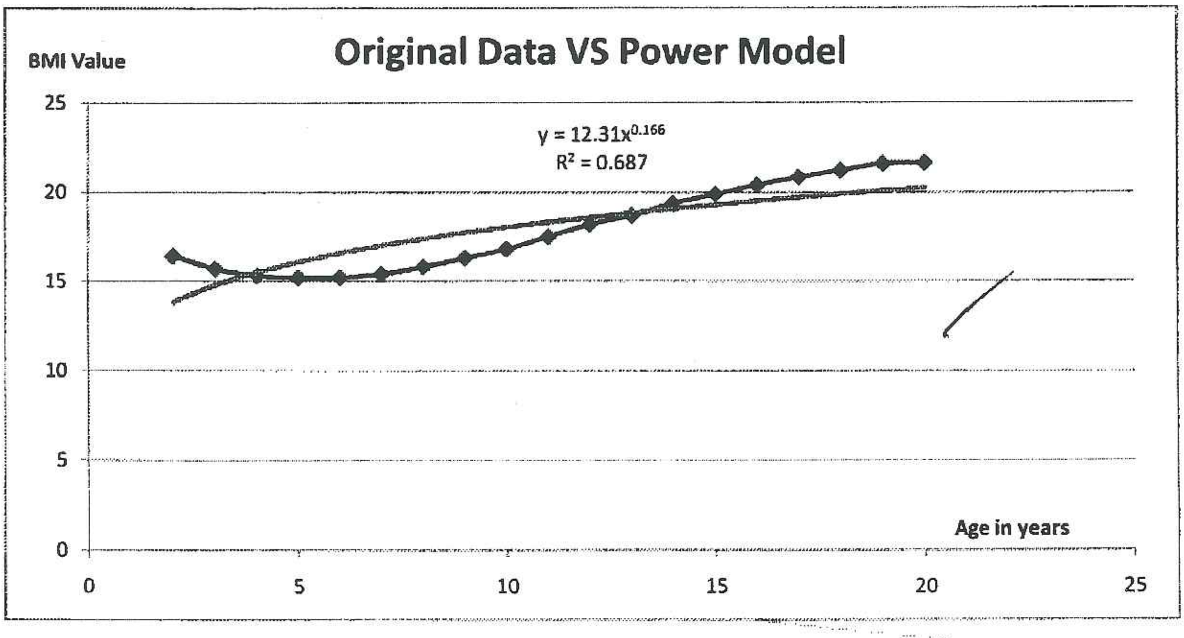
...er to be assured that the best fit model has indeed been discovered in the cubic function, we will simulate other types of models available.

These other models will be: All of these models will be determined using Microsoft Excel.

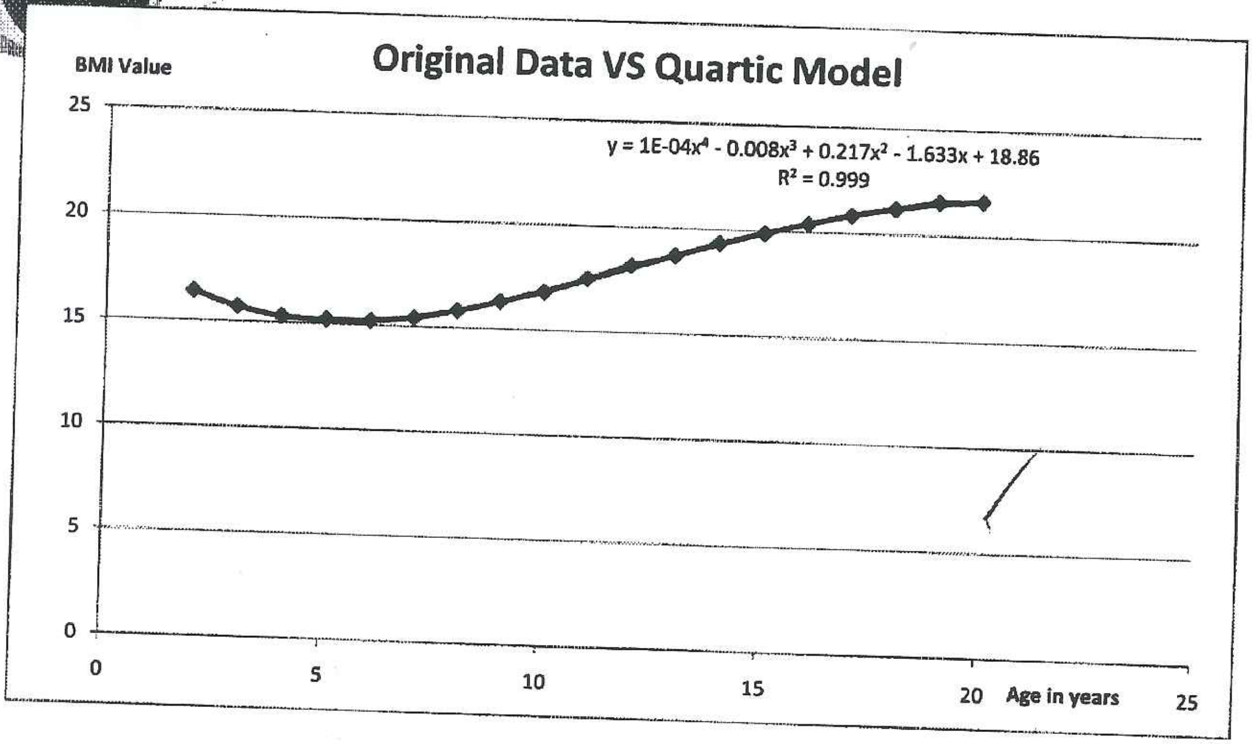
- An exponential regression.
- A power regression.
- A quartic regression.
- A logarithmic regression.

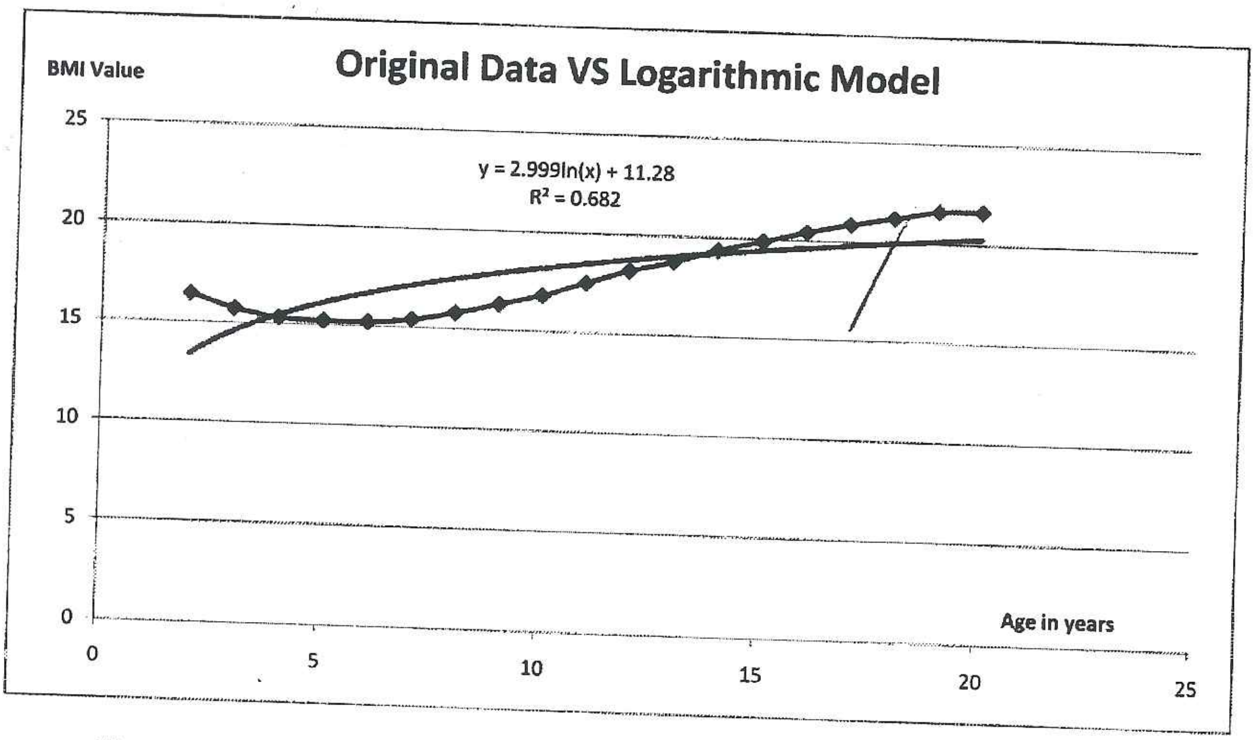
why?





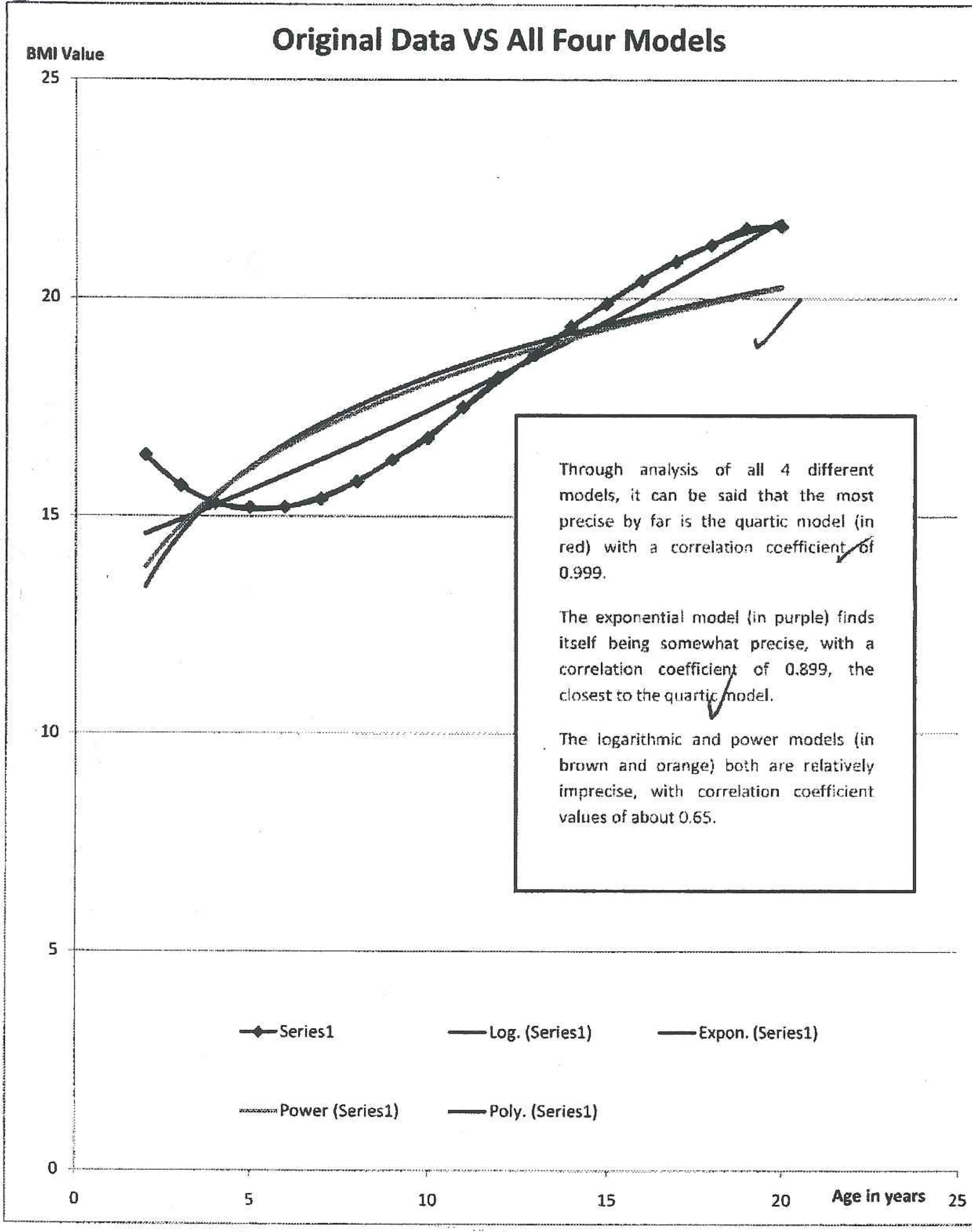
111


Original Data VS Quartic Model
BMI Value
y = 1E-04x⁴ - 0.008x³ + 0.217x² - 1.633x + 18.86
R² = 0.999
25
20
15
10
5
0
0
5
10
15
20
25
Age in years


Original Data VS Logarithmic Model
BMI Value
y = 2.999ln(x) + 11.28
R² = 0.682
25
20
15
10
5
0
0
5
10
15
20
25
Age in years

GOOD VISUAL SUMMARY

compare all four of the above models against the original data:

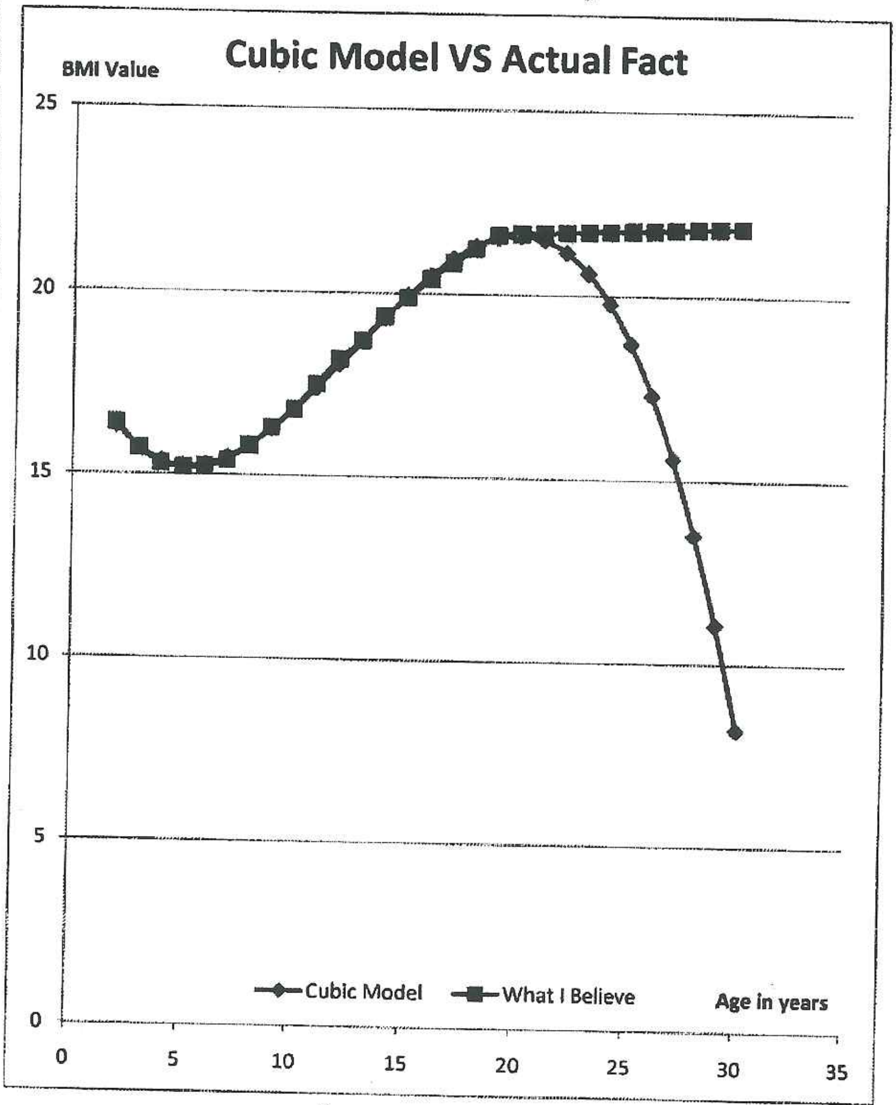


Mar. 13

[illegible] consideration, I have decided that the model I will move forwards with is the cubic model, since it provides the required level of precision (correlation coefficient of approximately 0.999) and does not require complex calculations such as the quartic model, which itself does not provide any additional visible precision.

I will therefore use the cubic model to attempt to estimate the BMI of a 30 year old woman in the USA. The following prediction can be seen in the graph and table below:

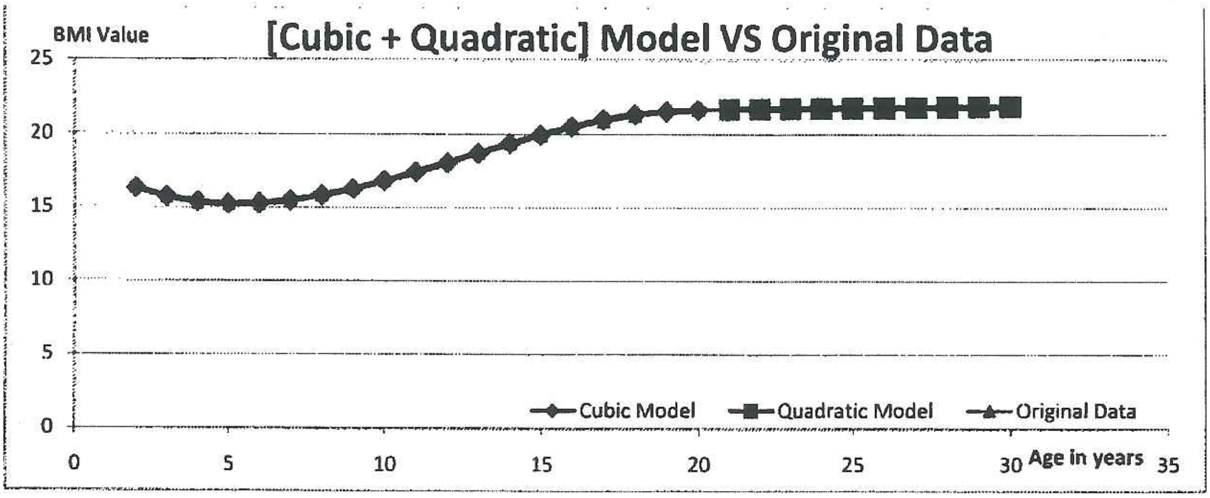
2	16.30307	16.4
3	15.71774	15.7
4	15.3662	15.3
5	15.224	15.2
6	15.26669	15.21
7	15.46983	15.4
8	15.80897	15.8
9	16.25965	16.3
10	16.79745	16.8
11	17.3979	17.5
12	18.03656	18.18
13	18.68899	18.7
14	19.33073	19.36
15	19.93733	19.88
16	20.48436	20.4
17	20.94737	20.84
18	21.3019	21.22
19	21.52351	21.6
20	21.58776	21.65
21	21.47019	21.67
22	21.14636	21.69
23	20.59182	21.71
24	19.78212	21.73
25	18.69282	21.75
26	17.29947	21.77
27	15.57763	21.79
28	13.50283	21.81
29	11.05065	21.83
30	8.196629	21.85



The above graph depicts the cubic model up to 30 years of age in blue, contrasted with what actually occurs. The cubic model states that the BMI of a 30 year old woman is the same of an unborn child (not even in the table). In the USA, for women in 2000, it is highly unlikely that their BMI will decrease so exponentially after 20 years of age. This therefore makes the cubic model only relevant for the domain 2 to 20 years of age. A new graph will therefore have to be created in order to sustain the model for what actually occurs. (Note: the BMI values up to 30 years old are accurate, obtained from the Mass General Hospital for Children website.)

GREAT DISPLAY OF MODEL LIMITATION

[illegible]ubic function does not model the actual data of an American woman in 2000 up to 30 years of age, a new model must thus be created that will fit the model clearly. This model will be a merging of a cubic model for the domain 2 to 20 years of age and a quadratic model for the domain 21 to 30 years of age. The GDC will be used to find both the cubic and the quadratic equations for this new graph. ← WHAT IS THE QUADRATIC MODEL?



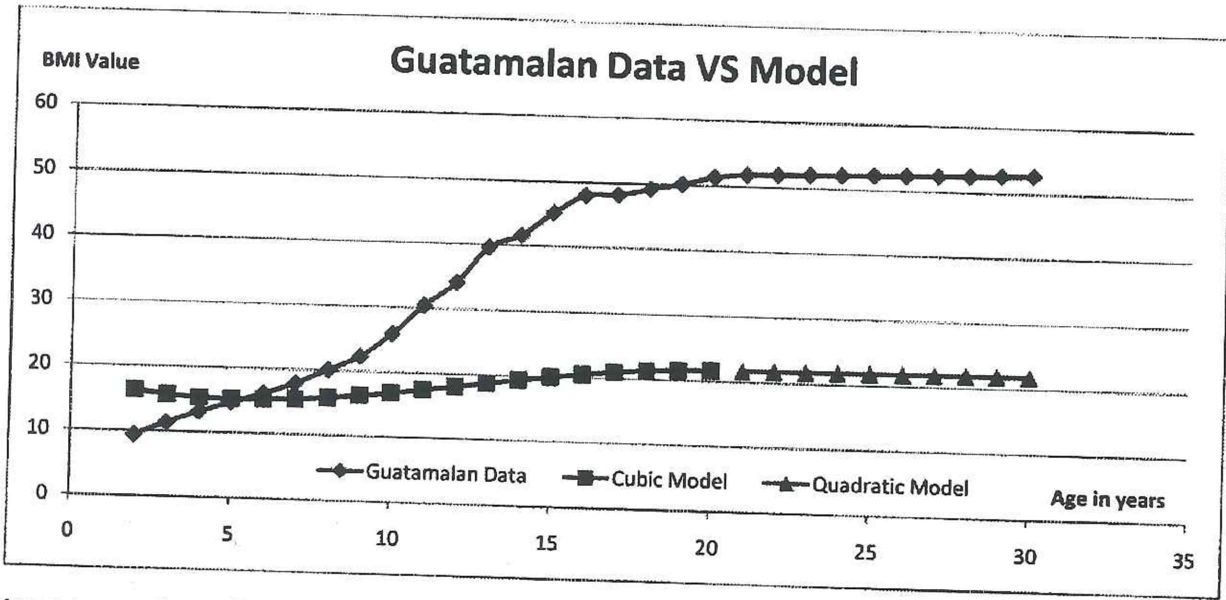
It can therefore be observed that this new model fits the original data quite precisely. The sum of all absolute differences being only 1.131, this model merging is indeed very precise.

2	16.30306904		16.4	0.09693096
3	15.71773981		15.7	0.017739808
4	15.36619826		15.3	0.066198255
5	15.2239972		15.2	0.0239972
6	15.26668946		15.21	0.056689462
7	15.46982786		15.4	0.069827861
8	15.80896522		15.8	0.008965215
9	16.25965435		16.3	0.040345655
10	16.79744807		16.8	0.00255193
11	17.39789921		17.5	0.102100791
12	18.03656058		18.18	0.143439418
13	18.68898501		18.7	0.011014993
14	19.33072531		19.36	0.029274695
15	19.9373343		19.88	0.057334295
16	20.4843648		20.4	0.084364796
17	20.94736963		20.84	0.107369628
18	21.30190161		21.22	0.081901609
19	21.52351356		21.6	0.07648644
20	21.5877583		21.65	0.0622417
21		21.67	21.67	0
22		21.69	21.69	0
23		21.71	21.71	0
24		21.73	21.73	0
25		21.75	21.75	0
26		21.77	21.77	0
27		21.79	21.79	0
28		21.81	21.81	0
29		21.83	21.83	0
30		21.85	21.85	Sum = 1.131

115

Mar. 13

...will now be applied to BMI data for females from Guatemala. (note, the values found were incomplete, therefore, averages were made as there were gaps in between the ages in years – the data may be inaccurate but it will test the model regardless of authenticity.)



e5

As we can see, the model which modeled the data for the American women so successfully does not model women from Guatemala (although there is some inaccurate data). This could be due to the fact that there is a different diet in Guatemala, that women's body structures are different from that in the USA or simply that the data obtained online is completely false.

The changes required in making this model fit the Guatemalan data would be to change the cubic model for it so be more fluctuating and to increase the shift in height of the quadratic model. do it

In all, although the [Quadratic+Cubic] model does successful model the BMI for women in the USA, there are some limitations that I will now evaluate:

- This model is only for women in the USA, in the year 2000.
- The model only works between the domain of 2 to 30 years.
- The model is based on average values only.
- Therefore, the model should not be applied to other countries' data as it will be inaccurate.

In the end, nevertheless, I believe that this experimental piece was a success, since a good model was found, although not replicating and modeling the data from many other countries precisely.

My grades:	Group grades:	large group
A - 2	A - 2	A - 2
B - 2 3	B - 3	B - 3
C - 5 → applied to guatamala	C - 5	C - 5
D - 4	D - 4	D - 4-5
E - 3	E - 3	E - 3
F - 1 2	F - 1	F - 1

17 - 19

see actual pg 96/ comments pg 94