Gold Medal Heights

In this task I will develop a function that best fit the data points in the graph, which will be plotted based on the table below showing the different gold medal heights.

Year	1932	1936	1948	1952	1956	1960	1964	1968	1972	1976	1980
Height (cm)	197	203	198	204	212	216	218	224	223	225	236

Number of years start from 1932 X-axis	Height (cm) Y-axis
0	197
4	203
16	198
20	204
24	212
28	216
32	218
36	224
40	223
44	225
50	236

Note: There are no data of 1940 and 1944

For all the graphs in this table, the y- axis will be represent the height in cm and the x-axis will be the year when the height was obtained.

The graph below shows the relationship between the years and the heights obtained between the years of 1932 and 1980:
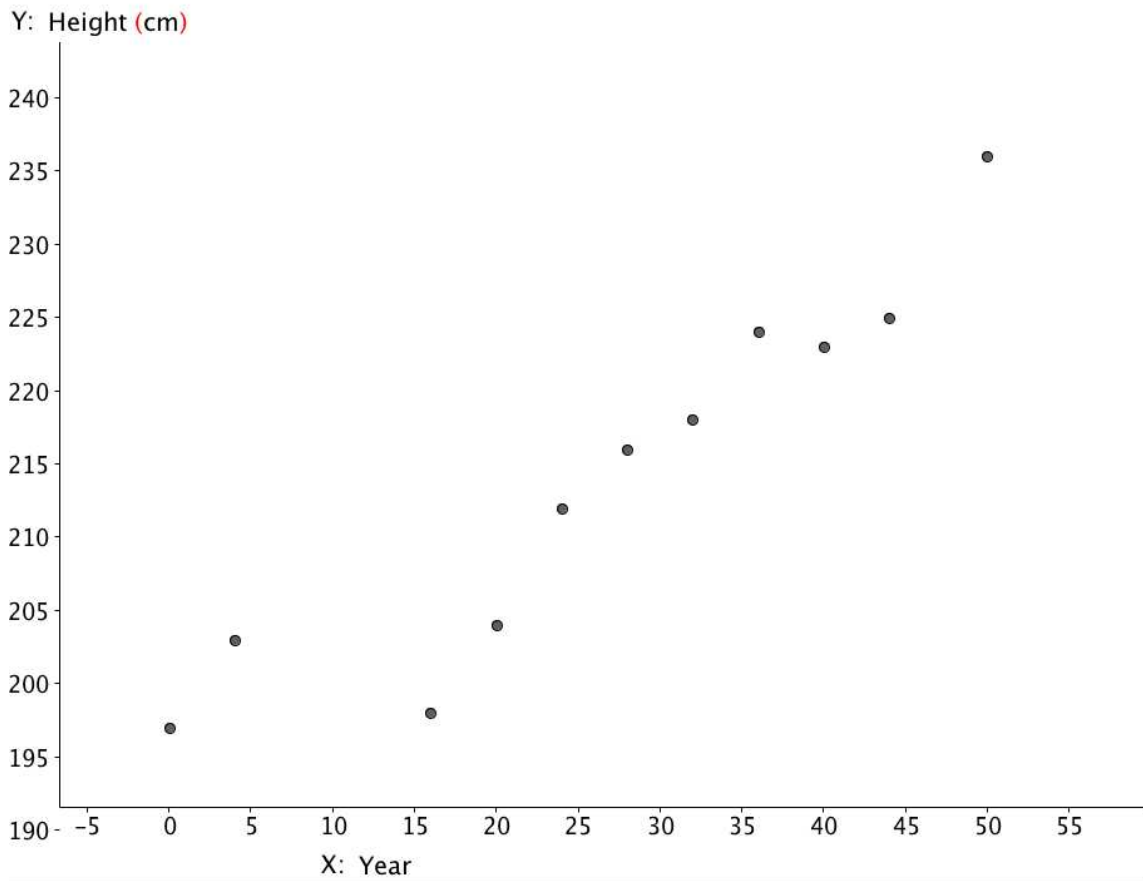
The further explanation for the missing data in 1940 and 1944 was due to World War II. Although the data does not tell us the reason why the height increased in 1932 and 1936 and drop down abruptly in 1948, we can assume that the World War II had affected athletics health critically.

With this graph and the numbers given in the table I am able to develop a model of function that fits the data points in my graph by using Geogebra software.

I chose linear function to model it, because the graph shows a general increase on the heights from 1948-1980 so in my opinion the linear will be the best function for it.

After sketching the new model function using the original data using Geogebra. The gradient calculated by the software is represented by :

y= 0.7456x+194.2535

Y: Height (cm)

X: Year

Regression Model

Linear $y = 0.7456x + 194.2535$
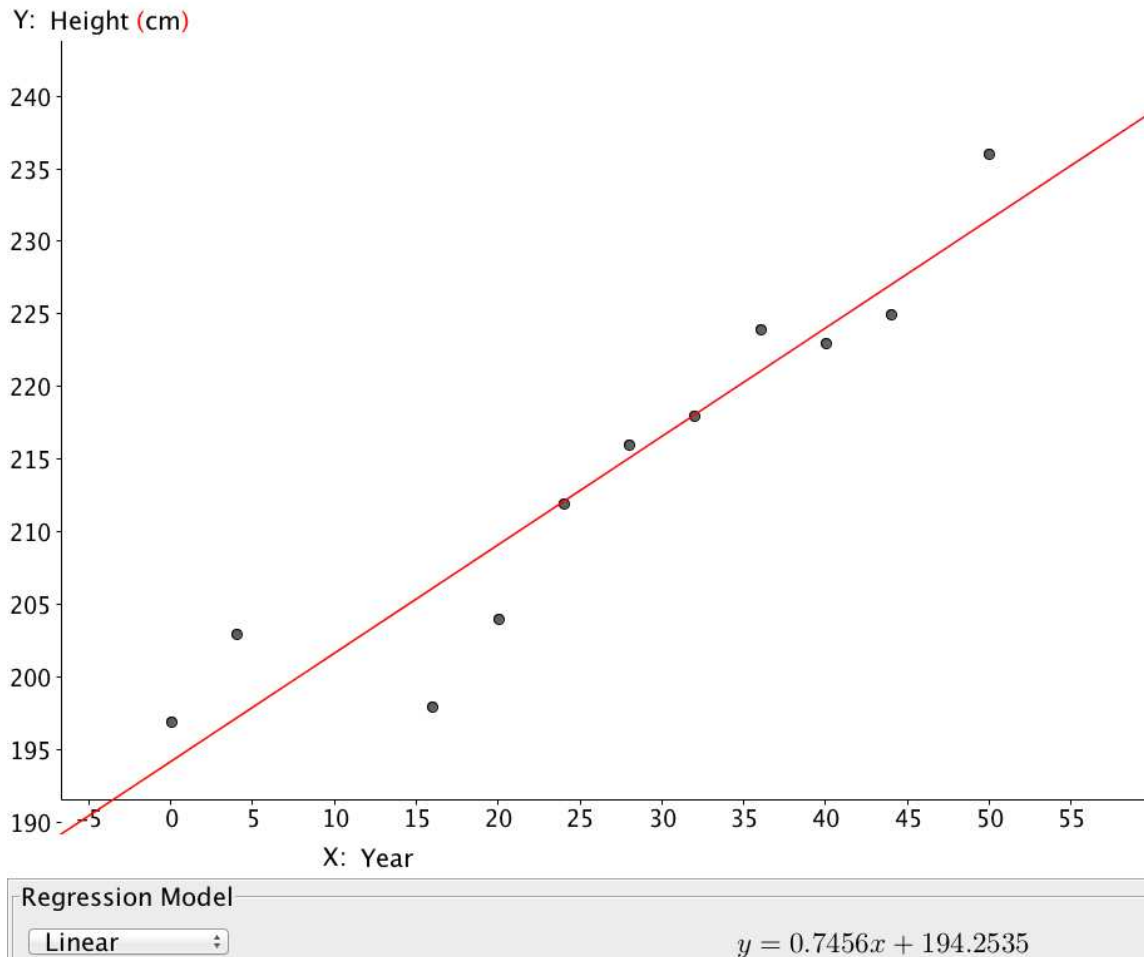
Testing the software linear equation:
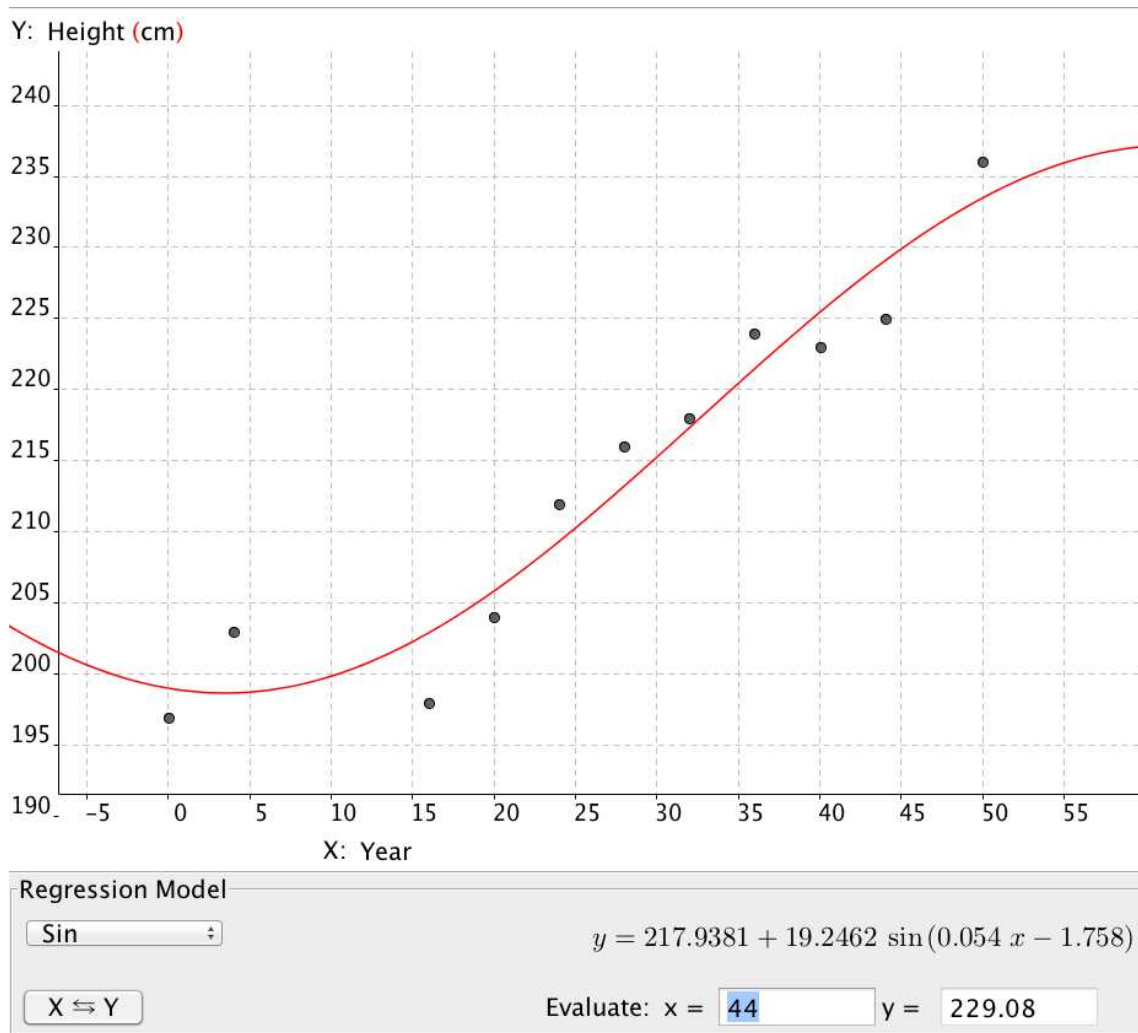
y= 0.7456x+194.2535

Let x = 44, we expect y= 225.

y= 0.7456 (44)+ 194.2535
y= 227.0599

There is a bit different from the actual data.
According to software, the R^2 = 0.8927, R= 0.94482802668

Subsequently, I decided to test with the Sine equation using Geogebra to find out a better line to fit with the data given.

Regression Model

Sin

$$y = 217.9381 + 19.2462 \sin(0.054\,x - 1.758)$$

X ⇆ Y   Evaluate: x = 44   y = 229.08

The software gave the equation of y= 217.9381 + 19.2462 sin (0.054x- 1.758).
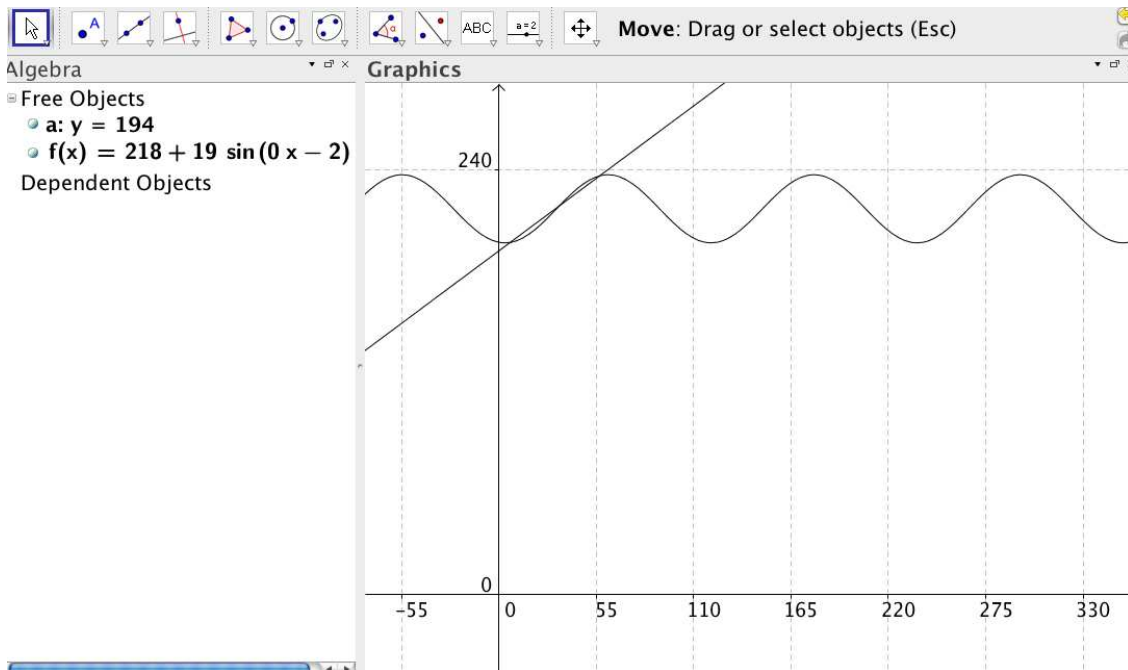And R^2 = 0.9361
R= 0.967522609555

Conclusion:

After sketching the linear graph and sine graph to investigate the best-fit model, I can see that the sine model fit the data from 1932- 1980 better than the linear function. The evidence is given by the value of R,

Sine function R =0.967522609555
Linear function R= 0.94482802668

Sine R is closer to the value of 1.

Furthermore, after sketching two functions on the new set of axes, I realize there is limitation in linear function, which, the x value will increase as y value increases. It is clearly impossible in the year, lets say 2050, the athletic will be able to perform a high jump of 5 meter despise gravitational force. On the other hand, the sine curve seems to be more reasonable to predict the jump height in the future. Even though we do not know whether the height increases or decreases but the sine function still keep the data in range of human ability.

Therefore in the next step, I decide to use sine function to predict the height in 1940 and 1944 if had the game been held:

1932-1940 = 8 years

y= 217.9381 + 19.2462 sin (0.054x- 1.758)
x= 8
$\Rightarrow$ y = 199.27 cm

1932-1944= 12 years

y= 217.9381 + 19.2462 sin (0.054x- 1.758)
x= 12
=> y = 200.7 cm

The answer would be mathematical reasonable increasing if we assume the data follows sine equation. The data are still in human range that the athletic can perform, however it is hard to say that they are 100% correct since we only rely on technology.

I also decided to use the function to predict the result in 1984 and 2016

1932 – 1984 = 52 years

y= 217.9381 + 19.2462 sin (0.054x- 1.758)
y= 234.63 cm

1932- 2016 = 84 years
y= 217.9381 + 19.2462 sin (0.054x- 1.758)
y= 224.8 cm
The answer are mathematical correct. However, in reality we do not know whether the data will truly describe the result since we only the data up to 1980.

Data table from 1896 – 2008:

Numbers of year	Actual height/cm
0	190
8	180
12	191
16	193
24	193
32	194
36	197
40	203
52	198
56	204
60	212
64	216
68	218
72	224
76	223
80	225
84	236
88	235
92	238
96	234
100	239
104	235
108	236
112	236

The graph below represents the data from 1896 to 2008 :

n	24
MeanX	61.6667
MeanY	214.5833
Sx	34.0826
Sy	19.2939
r	0.9638
ρ	0.9643
Sxx	26717.3333
Syy	8561.8333
Sxy	14576.6667
R²	0.9615
SSE	329.6886

Y: Height

240 — 220 — 200 — 180

0   20   40   60   80   100   120
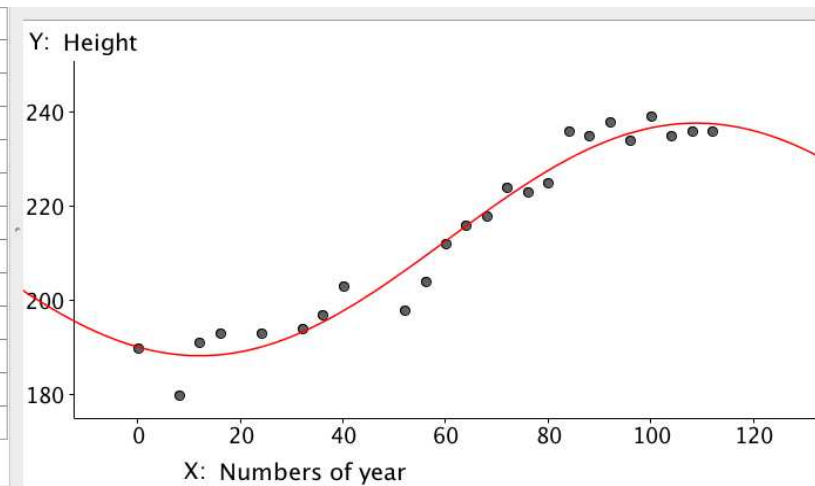
X: Numbers of year

The sine model I used fit even better when I obtain additional data. The value of R= 0.980561063881, which is closer to R=1. However, it only went through some data points.

Numbers of year	Height calculated by sine function
0	190.1
8	227.5
12	236.1
16	210.9
24	201.8
32	233.9
36	206.1
40	188.3
52	231.0
56	201.6
60	188.8
64	211.1
68	236.1
72	227.3
76	197.6
80	190.2
84	216.0
223.1	237.3
194.1	223.1
192.5	194.1
220.7	192.5

237.6	220.7
218.6	237.6

Different between the actual data and data calculated by using

y= 212.9337 + 24.6777 sin (0.0324x- 1.9601)

y= 212.9337 + 24.6777 sin (0.0324x- 1.9601)	Actual height	Difference
190.1	190	0.1
227.5	180	47.5
236.1	191	45.1
210.9	193	17.9
201.8	193	8.8
233.9	194	39.9
206.1	197	9.1
188.3	203	-14.7
231.0	198	33.0
201.6	204	-2.4
188.8	212	-23.2
211.1	216	-4.9
236.1	218	18.1
227.3	224	3.3
197.6	223	-25.4
190.2	225	-34.8
216.0	236	-20.0
237.3	235	2.3
223.1	238	-14.9
194.1	234	-39.9
192.5	239	-46.5
220.7	235	-14.3
237.6	236	1.6

218.6	236	-17.4

After comparing the data between the actual provided height and the new best-fit model when we have addition data, it seems to be likely describing the nature of the data. However, the function does not fit in most of the data. Moreover, a better function should be investigated to make sure it goes through all the data point. However, due to the limit knowledge at the moment when I only know about linear, exponential, sine, logistic functions, it prevents me from going further to develop a more advance function which fit all the data points. Although the two functions have their R-values near to 1, especially sine function, it does not go through all the data point.