

The Olympic games were created thousands of years ago with the purpose to determine who could push himself the hardest to jump higher or run faster. High jump has always been part of the revived Olympics and has been carried out ever since the 1896's Olympics.

The achievements of athletes almost put itself forward for analysis, as the data is already precisely recorded and therefore it is rather easy to make predictions. This data is useful for trainer or scouts when they have to estimate whether an athlete is worth their while or not.

As in all of the other programs a steady improvement of the athletes' performances have been recorded as the games advanced and almost all skilled sportsmen were taken under supervision of trainers and were made professionals.

The recorded heights of the gold medal winners starting from the 1932's Olympics until 1980 are as follow:

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

**Table 1.0** Data table for the height of the gold medallists in high jump from 1932 until 1980.

Using these raw values to plot a graph without any manipulation results in a picture which is both hard to read and not well suited to convey the wanted information.

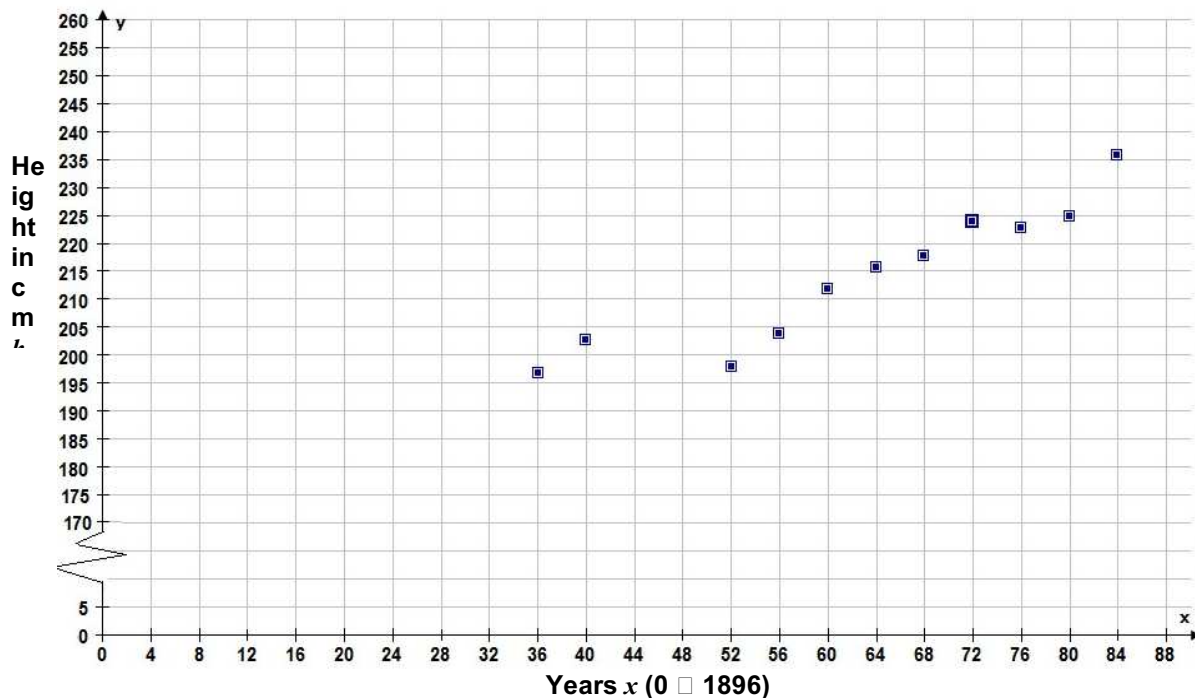
If the x-Axis was to correspond with the years starting from year zero, the graph would be unnecessarily broad. The data of **Table 1.0** has to be processed in order to provide a readable graph. As well no years before 1896 are relevant for this task as the first modern Olympic games were held in that year.

Therefore zero on the x-Axis will represent the year 1896 on the graph and the new data is listed in the following table.

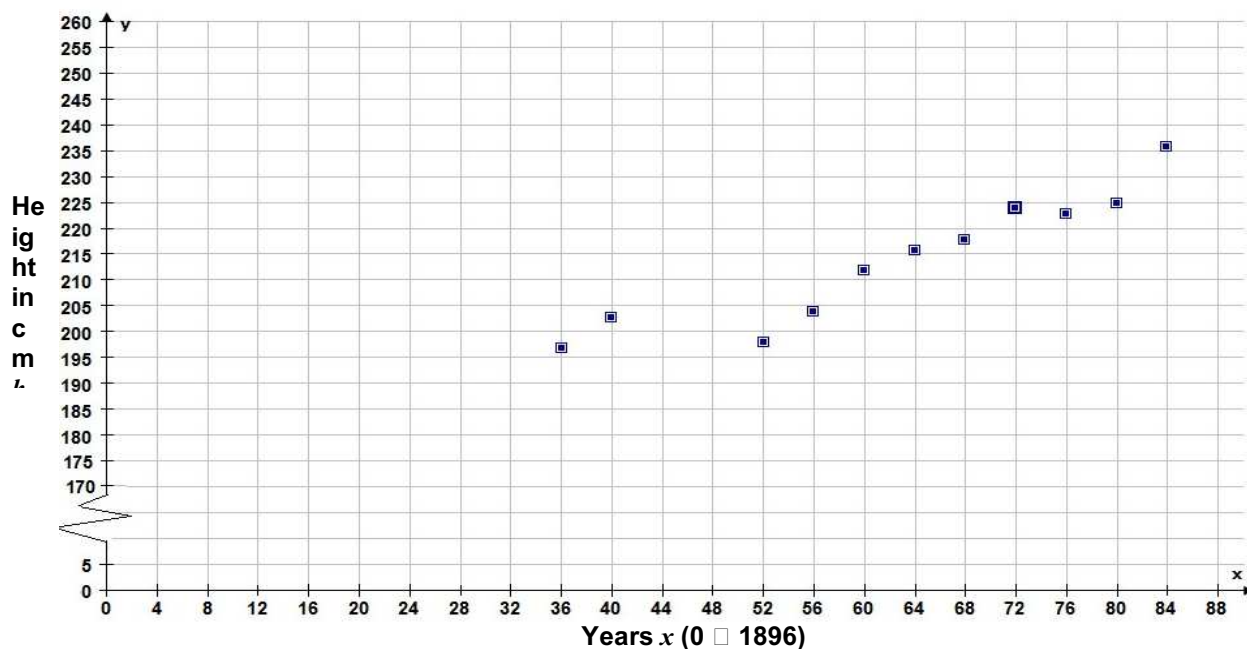
Year with year 1896 as year 0	36	40	52	56	60	64	68	72	76	80	84
Height in cm	197	203	198	204	212	216	218	224	223	225	236

**Table 1.1** Processed data from **Table 1.0** for the height of the gold medallist in high jump from 1932 until 1980 with 1896 as year 0

Plotted on a graph the data from **Table 1.1** results in this graph.



**Figure 1.0** Graph from **Table 1.1** with years  $x$  versus Height in centimetres  $h$



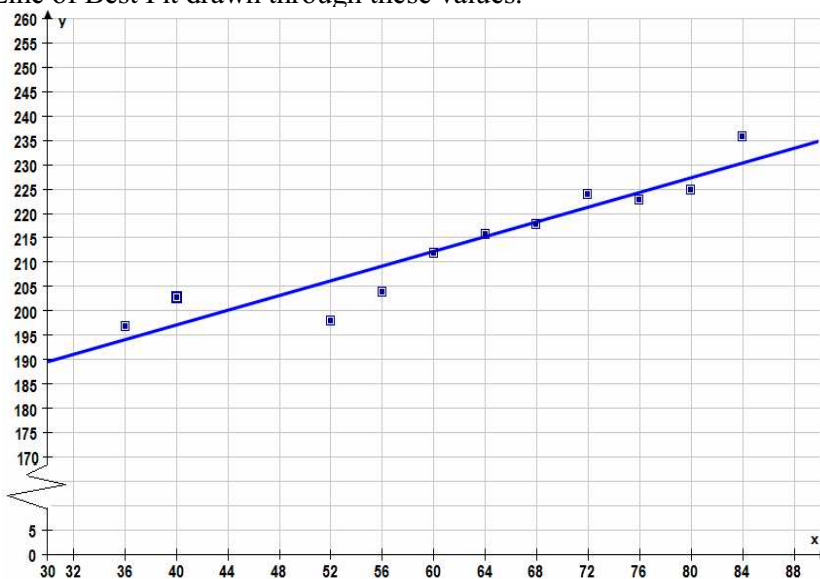
**Figure 1.1** Graph from **Table 1.1** with years  $x$  versus Height in centimetres  $h$  (Identical to **Figure 1.0** inserted here again to make reading more convenient)

On the x-Axis the years are listed and the letter  $x$  will be assigned to this variable, and the y-Axis indicates the Height of the athlete's in centimetres and will be referred to as  $h$ .

In addition the values of  $h$  are restricted to positive values, since the a jump cannot go any lower than its base level. To provide a reasonably sized and proportioned graph the y-values starting with 10 to 170 were skipped, the (SQUISHY LIGHTENING THINGGY) represents a summarization of these values.

The Olympics were not held in 1940 and 1944 as the whole world was caught up in the Second World War. As one can see the break had a negative impact on the athlete's performance as the best achieved height fell by almost five centimetres in comparison to the previous games.

This disrupts the general trend of improvement as one can clearly deduce from the following graph which uses the values of **Table 1.1**, however in comparison to **Figure 1.0** and **Figure 1.1** has a generated Line of Best Fit drawn through these values.



The position of the x-Axis had been shifted to lay focus on the relevant area. Instead of zero being the origin, the value of 30 (1926) is used as it provided a fitting window.

Using a continuous line leads to several problems since it implies that there are valid values for years where no Olympics took place e.g. 39 (1935). Therefore the Domain for the graph must be defined so that only reasonable x-Values are permissible.

The data from **Table 1.1** states that the Olympics occurred every four years starting from with a starting year defined as 36 (1932). **Years  $x$  (0 = 1896)**

**Figure 1.2** Graph based on **Table 1.1** years  $x$  versus Height in centimetres  $h$  with a line of best fit.

Therefore a Domain can be stated as  $0 \leq x \leq 1996$ , with the restriction that  $x$  and  $h$  are integers.

Also as previously mentioned the Rage must be limited to  $0 \leq h \leq 353$  since a jump cannot reach its maximum height at a point lower than the starting point.

A linear Graph will not be suitable as one can see from **Figure 1.2** as there is a rather steep increase of height as the years progress. If one was to follow the graph for some period of time one would get entirely unreal results. A linear graph is also bound to intersect the x-Axis at some point and from there on continue on decreasing. It is also very stiff and unsuited to depict the **only slight improvements that happen**.

To prove this and give an example of the unrealistic results the already generated line of best fit from **Figure 1.2** was used and the x-Axis extended to evaluate.

If one takes a rough estimate at the highest point of the graph one would get a result such of (250, 353).

Which would imply a height of 3.50 meters in 2146.

At first sight this might not seem to be an unachievable value, however the height of one storey of an average house is about 3.00 meters.

So to get a proper picture of the height one should look outside of the 2<sup>nd</sup> floor of a building and then image a person almost jumping through that window from the ground.

Windows are normally placed at a height of 0.70 meters above the floor. This gives one a feeling at how unrealistic the predicted height is and how unsuited therefore a line of linear nature is.

Any cubic function can be discarded as well. The eventual gradient would prove to be even steeper than the one of the linear graph. (Add other things)

A sinus function might seem a reasonable solution as one can see fluctuations that produce an wave like form, however the general trend towards a greater height would not be depicted when using a sine function and the anomalies do not reoccur regularly. The sin or cosine function for that matter are unsuited, as they are only able to depict discrepancies that occur with a regular pattern.

The function of a logarithm seems to provide the most suitable choice. It shows a slow steady slope, that decreases as the x-values grow.

The general Logarithmic function is expressed as  $h = b + k \log(x - c)$  and includes 4 parameters  $b$ ,  $c$ ,  $d$  and  $k$ . Firstly the focus will solely rest on  $b$ , the parameter who controls the vertical stretch and  $k$  which represents to vertical translation. Restricting the parameters until later to  $b$  and  $k$ , makes it easier to determine the value of these and both of them are sufficient to model a suitable graph for the given

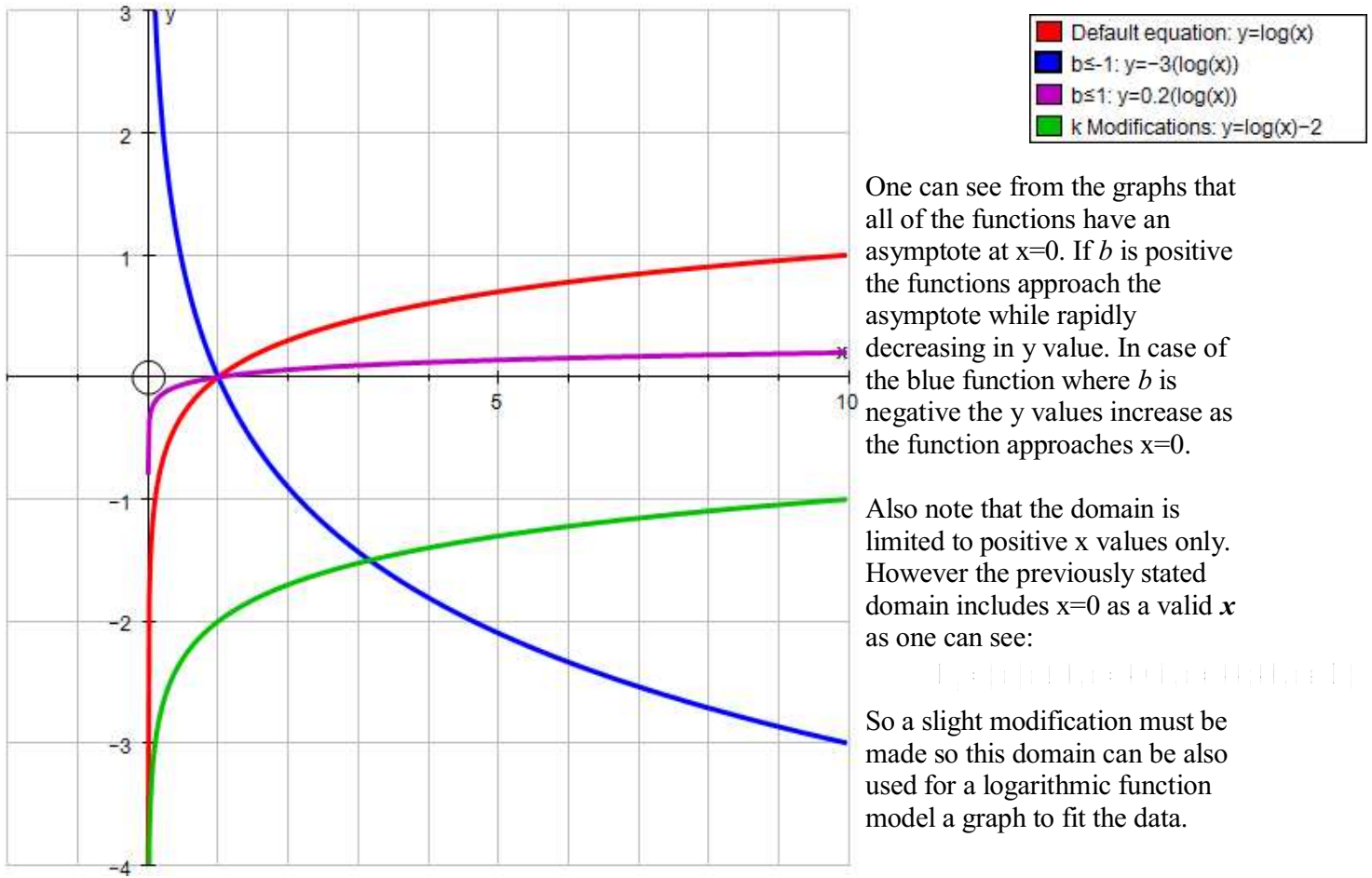
data. So the new equation we are dealing with goes as follow .

As the values for  $x$  increase expression of  $\log(x)$  increases as well. If  $\log(x)$  increases furthermore, resulting in a vertical expansion, however if  $k$  is negative the graph is vertically compressed.

Years  $x$  ( $0 \leq x \leq 1996$ )

**Figure 1.3** Graph based on **Table 1.1** years  $x$  versus Height in centimetres  $h$  with a line of best fit.  $x$  extended to the value of 250

A negative  $b$  causes a reflection along the x-axis, even if  $b$  is now negative a vertical compression is found when and a vertical expansion occurs if  $b$  is positive. The parameter  $k$  determines the y value when  $x=1$ , as the expression of  $\log(1)=0$ , so by default, meaning  $k$  equals zero, the y value at  $x=1$  is 0, the effect of the  $k$  can be expressed by stating  $(1,k)$ . As whenever the x-value is 1 the y value equals  $k$ . This all seems very abstract and in order to give a proper understanding here are some graphs that represent modified graphs of  $y = \log(x)$ .



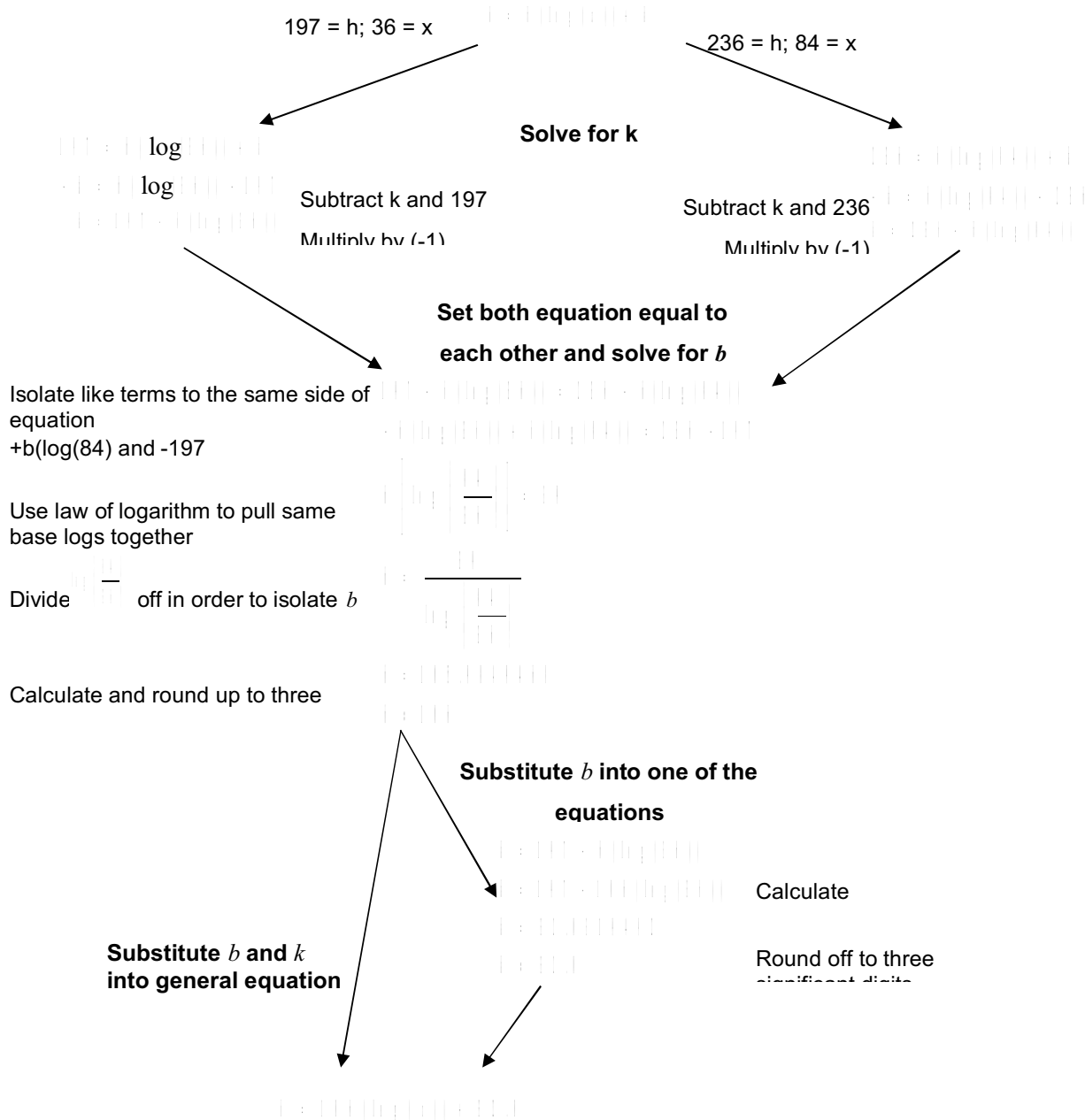
**Figure 2.0** Graph used to explain the effects of parameter  $b$  and  $k$  on logarithmic functions. See legend for equations.

The new Domain must be stated as  $x > 0$ . Zero is now a non-permissible value and the restriction that  $x$  must be a multiple of 4 has also been revoked so a proper graph can be created.

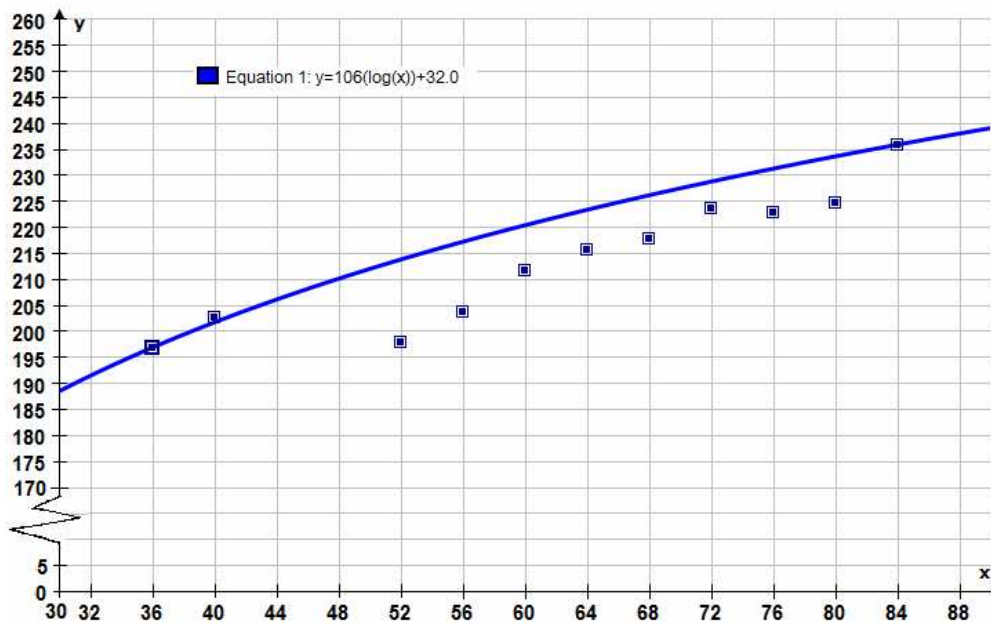
The Range however stays  $y \in \mathbb{R}$ .

So now in order to find a suitable logarithmic function with the format of  $y = b \log(x) + k$  to fit the data from **Table 1.1** the system of equations is used. Only the parameters of  $b$  and  $k$  were used, since they give one enough room to model a function and the system of equations is limited to two unknowns only.

See next page for the acquisition of  $b$  and  $k$ .



One should not at this point that these points were not randomly chosen. They both represent anomalies and therefore do not fit the general pattern, however their effect on the curve of the graph must also be taken into consideration. The graph attained by this calculation does fit these points very well as the graph on the following page proves.

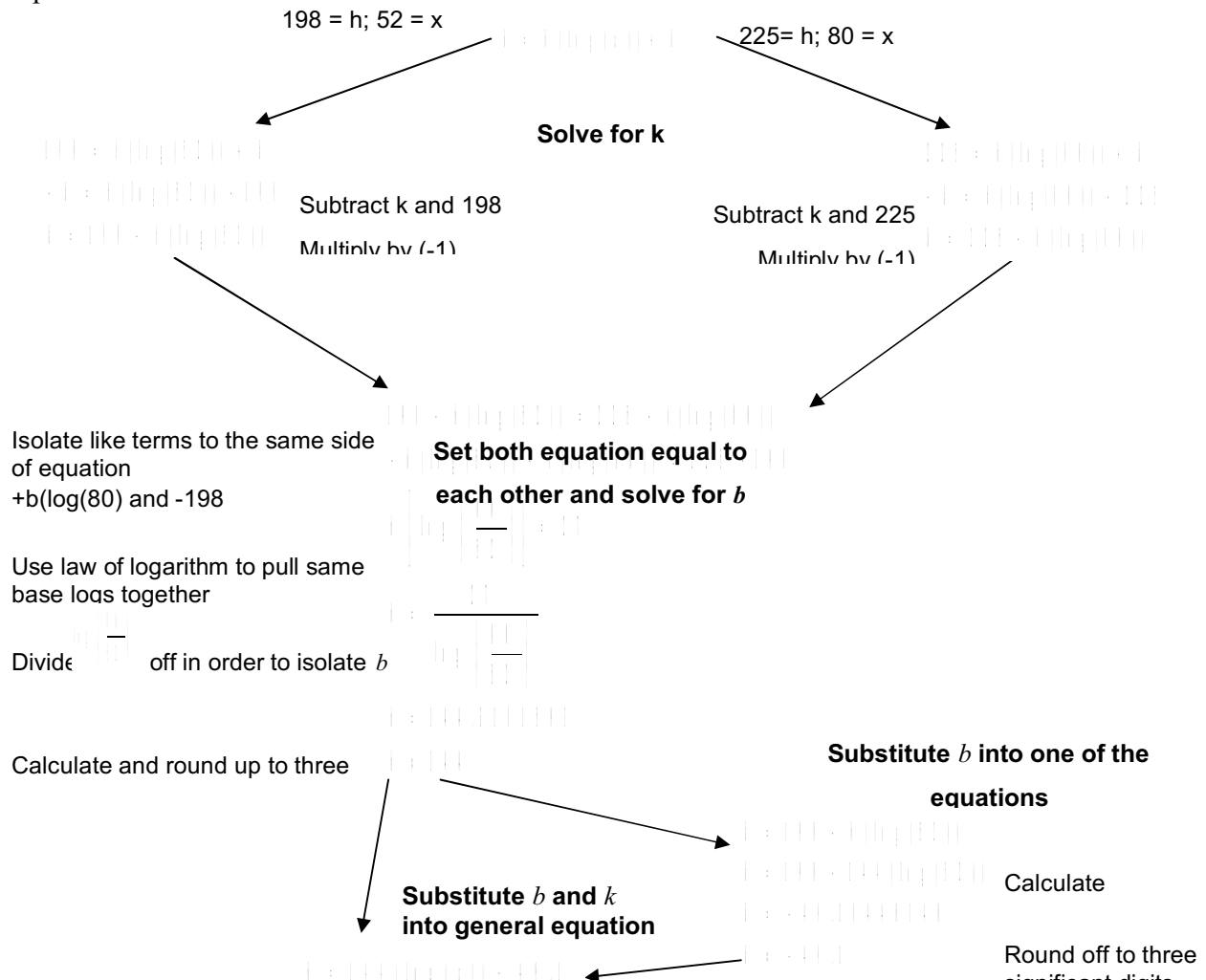


The axis' have again been shifted to provide a better visualization of the graph. The x-axis begins at 30(1926). One can see that the graph fits really well with (36, 197), (40, 203) and (84, 236). However the other points are not really included on the graph. Therefore it is fit to create a graph which suits the other points and then average the value of the parameters in order to find the best fitting graph.

**Figure 3.0** Graph with data points from **Table 1.1** years  $x$  versus Height in centimetres  $h$  with logarithmic Equation 1

Years  $x$  (0

One can see from **Figure 3.0** that the points from (52, 198) to (80, 225) follow a rather regular pattern and so following the same steps as before using the system of equations one can attain a suitable graph for these points.

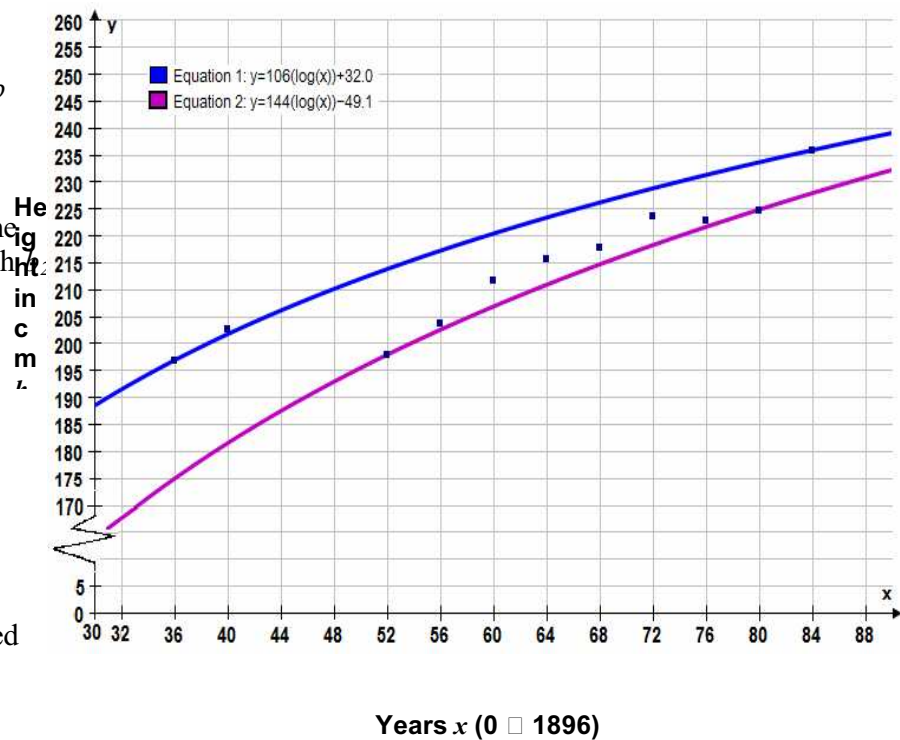




Now using the just attained new equation  $y = 125(\log(x)) - 8.55$  the functions graph will added to the graph already present **Figure 3.0's** draw a comparison between the two and prove that an averaged function would provide the best fit.

A graph that represents the average function can be calculated by averaging the  $b$  and  $k$  of Equation 1 and Equation 2. To distinguish between the two function's parameters let  $b_1$  and  $k_1$  be the parameters of Equation 1 with  $b_2$  and  $k_2$  belonging to Equation 2.

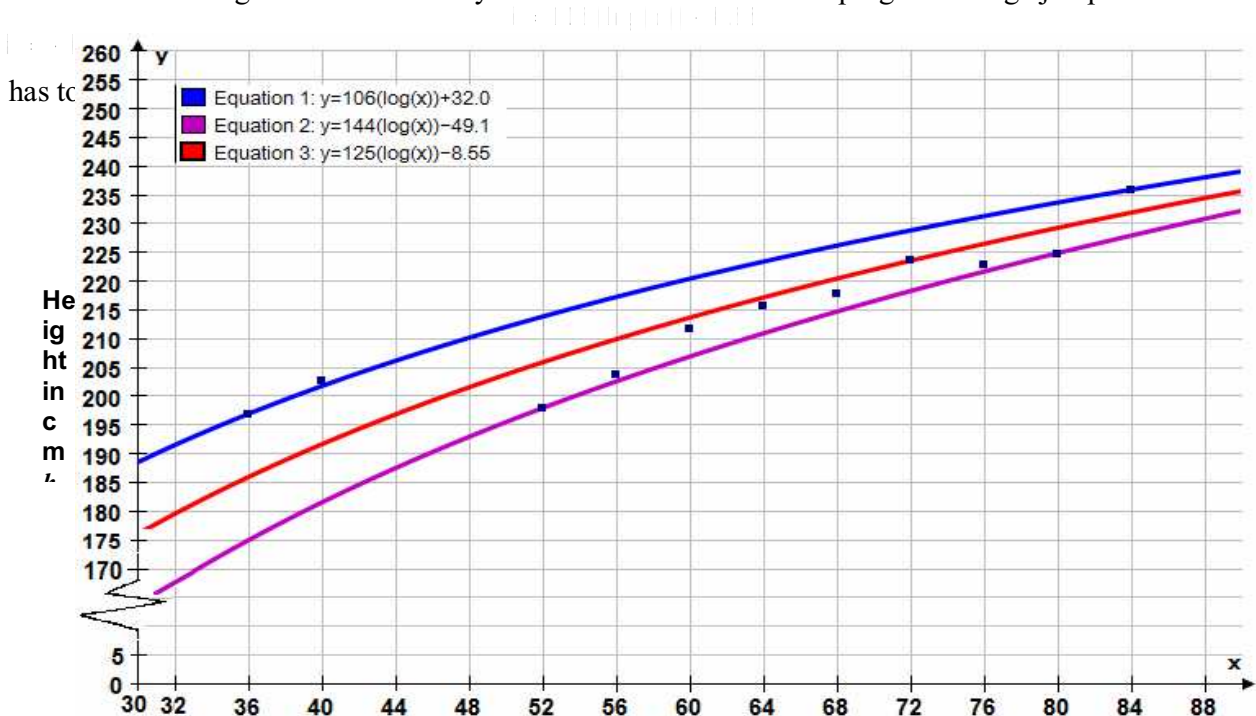
$b$  now represents the averaged value.



Forming a new equation using the newly acquainted values one gets the following function

**Figure 3.1** Graph with data points from **Table 1.1** years  $x$  versus Height in centimetres  $h$  Graphs of logarithmic Equation 1 and logarithmic Equation 2

To test if the averaged function is really better suited to visualise the progress of high jump athletes one



Years  $x$  (0  $\square$  1896)

**Figure 3.2** Graph with data points from **Table 1.1** years  $x$  versus Height in centimetres  $h$ . Graphs of logarithmic Equation 1 and logarithmic Equation 2 and Equation 3 the average from the first and second.

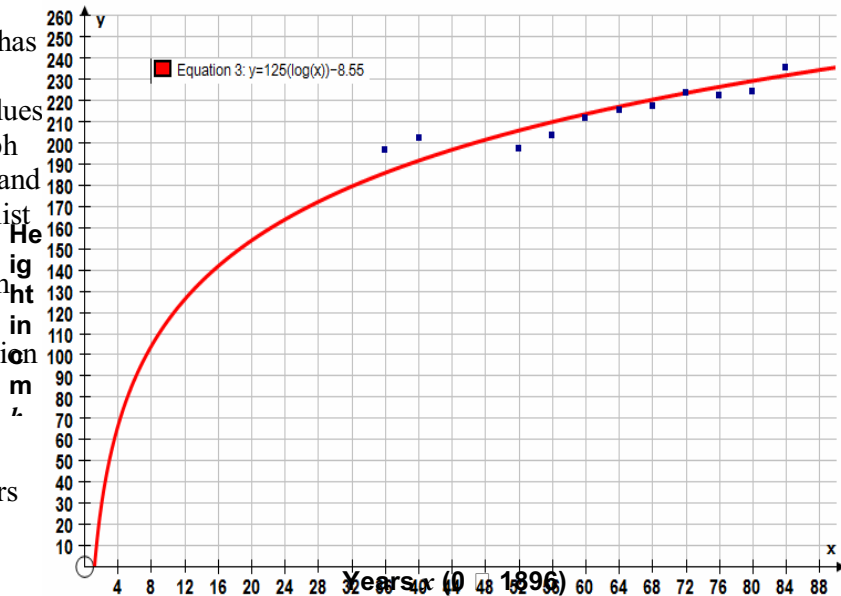
Note that even though the graph only directly passes explicitly thorough one point it does not differ greatly from any points on the graph.

After finding a fitting logarithmic function, the limitations of this model function need to be explained and analysed. The axis' chosen from **Figure 3.0** through to **Figure 3.2** do only offer a limited view on the function and one of the main problematic with logarithmic functions lie in their asymptote along  $x=0$ . To understand the problem properly it is suited to move the  $x$ -axis back to the origin.

As one can see this graph as well has an asymptote at  $x=0$ . This proves problematic, since no negative values are permissable, however the graph clearly enters the fourth quadrant and the likelihood that the gold medallist height of the 1908's (8) Olympics

Looking back at the general equation which was already introduced as

one can see that for the here used model function only the parameters of  $b$  and  $k$  were used.



**Figure 3.3** Graph with data points from **Table 1.1** years  $x$  versus Height in centimetres  $h$  with Equation 3.

If one had added  $d$  for example, the parameter that controls the horizontal translation one could have avoided facing the problem of rapidly decreasing height as the function approaches  $x=0$ , because assigning  $d$  a value lesser than 0 would have resulted in a shift of the graph and the asymptote by the value of  $d$  towards the left.

If the asymptote moves further into the positive values.

The last parameter  $c$  controls the horizontal compression or expansion. Causes a horizontal compression, effectively flattening the graph. On the other hand if the graph is expanded by a factor of  $-$ , this expression is also valid for the factor of the compression.

A negative value of  $c$  causes a reflection along the  $y$ -axis.

Up until now fits the give data from **Table 1.1** well enough. So no refinements are necessary, yet.

However there needs to be a more suitable equation without the problems of asymptotes logarithmic functions propose.

Using technology a line of best fit of exponential nature was generated. The default equation of exponential function is however the parameters of the logarithmic function were already ladled so they have to be renamed in order to avoid confusion.  $b$  will be substituted by  $l$ , so the new equation is

There is a special restriction regarding  $l$  also called the base. The base has to be greater than zero, so a valid function can be produced. In addition if the gradient is negative, meaning the  $y$  values are decreasing as the  $x$  values increase. The opposite happens if then the graph grows.

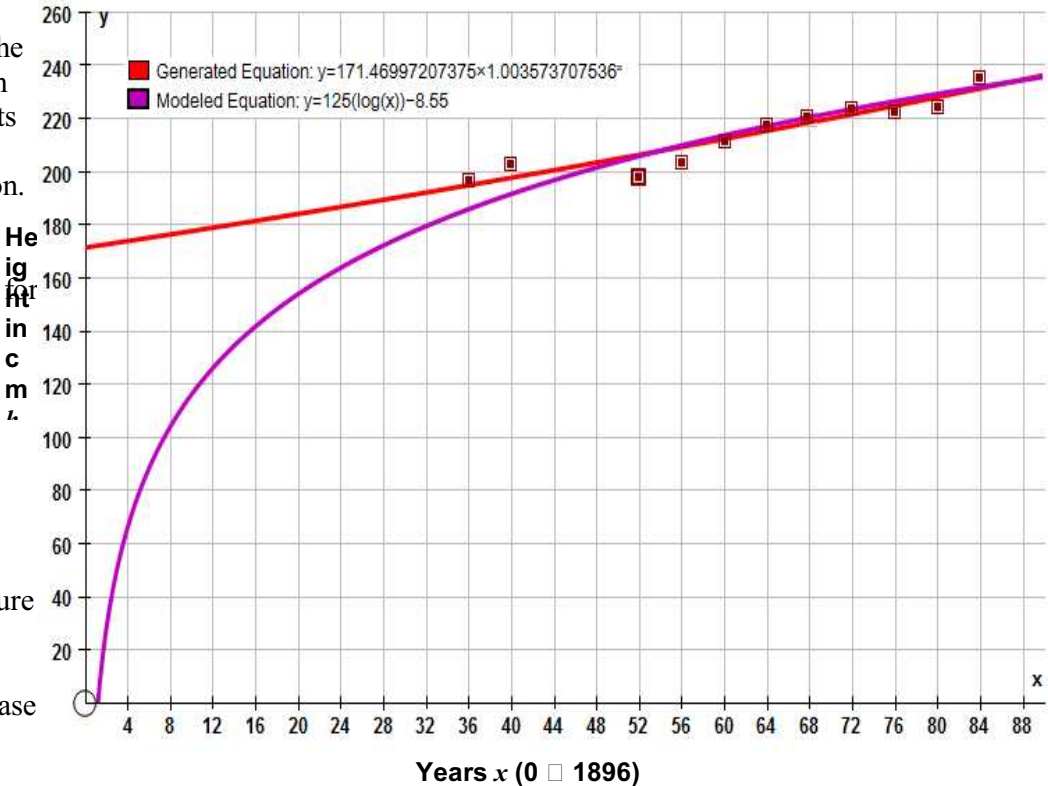
$A$  represents the  $y$ -intercept this is easily explained as the graph has to intercept the  $y$ -axis if  $x=0$ . So if zero is substituted for  $x$ : the expression will always equal 1 regardless of the value of  $l$ .



The exact equation that was generated by the software was  $y = 171.46997207375 \times 1.003573707536^x$ . Usually the rule of only three significant digits applies, however in this case  $1$  would end up to be rounded off to  $1.00$  which would result in a flat line at  $y = 171.46999\dots$ . Therefore it is necessary in this case to keep the decimal places as inconvenient as it is, the equation as seen is extremely sensible and should be kept as accurate as possible.

To receive an accurate understanding how well the modelled and generated function compare a graph is necessary.

It is obvious that the generated equation suits the data points very well and also fits the data later on. There is no asymptote which generates an error in 1896(0). As well the values are far more realistic as the  $x$ -values decrease. However the modelled function might prove more accurate in the future as the logistic functions gradient continues to decrease whereas the exponential function's slope will increase.



**Figure 4.0** Graph with data points from **Table 1.1** years  $x$  versus Height in centimetres  $h$  with Equation 3.

Using the two equations one is able to estimate what the winning height could have been if the Olympics happened in 1940 and 1944.

	Equation 3 ( $h=125(\log(x))-8.55$ )	Generated equation ( $h=171.46997207375 \times 1.003573707536^x$ )
Calculate $h$ of 1940 by replacing $x$ with 44		

Calculate $h$ of 1944 by replacing $x$ with 48	$1 \div 114699 \div 20 \div 3 \div 1003 \div 3 \div 0 \div 36^{48}$ $203 \div 141413$ $203$
--	---

The results of the Generated equation seems more reasonable as, for example in 1940 the drop from 1936's 203cm to 200cm is more realistic than the almost 6cm drop Equation 3's result states. However no definite answer can be given as there is no great difference between the results to eliminate the possibility of one or the other.

If the graph is well fitted to the data points one is also able to predict or retrace possible heights that athletes should have achieved in the past.

However one cannot wholly trust the results common sense is needed in evaluating the likely hood of the height.

Starting with prediction of the gold medallist height in the Olympics after this year's the corresponding  $x$ -value has to be calculated.

$$1 \div 114699 \div 20 \div 3 \div 1003 \div 3 \div 0 \div 36^{48}$$

$$203 \div 141413$$

$$203$$

Then following the same steps as above, a prediction for 2016 can be made.

$$1 \div 114699 \div 20 \div 3 \div 1003 \div 3 \div 0 \div 36^{48}$$

$$203 \div 141413$$

$$203$$

The height of 251cm seems reasonable looking at it purely mathematically. Within a timespan of 48 years (1932 to 1980) the height increased by 39 centimetres. So it following the same trend an increase by 15 centimetres over 36 years (1980 to 2016) should be achievable.

However if one adds common sense into the equation, it becomes firstly clear that a prediction is very hard to make since there are a lot of outside factors.

For example the environmental factors are not ideal or because of the lack of fit athletes the height might decrease in comparison to the precious years.

As well we have seen that there is no definite pattern the results follow. Though there is a general trend that suggest the height will continue to grow one also has to take into account that human bodies can only sustain so much strain.

Considering that the strain increases with the height achieving a greater height will become harder and harder and at some point, a maximum height should be reached.

However a prediction that is not so far off the data which was used to model the function might not be that much off.

$$1 \div 114699 \div 20 \div 3 \div 1003 \div 3 \div 0 \div 36^{48}$$

$$203 \div 141413$$

$$203$$

A height of 235cm seems reasonable prediction but again there are many factors which can cause a random glitch, meaning an exceptional high or low height as before mentioned.

These values should not be used as definite predictions but more to give someone an idea of what the potential height could be.

Year	1896	1904	1908	1912	1920	1928	1932	1936	1948
------	------	------	------	------	------	------	------	------	------

Height in cm	190	180	191	193	193	194	197	203	198
--------------	-----	-----	-----	-----	-----	-----	-----	-----	-----

<b>Year</b>	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992
<b>Height in cm</b>	204	212	216	218	224	223	225	236	235	238	234

Year	1996	2000	2004	2008
Height in cm	234	235	236	236

**Table 2.0** Data table for the height of the gold medallists in high jump from 1896 until 2008.

In order to use the modelled graph the years have to be modified so 1896 equals 0 on the graph, so the same modification as occurred on *Table 1.0* to *Table 1.1* must be made.

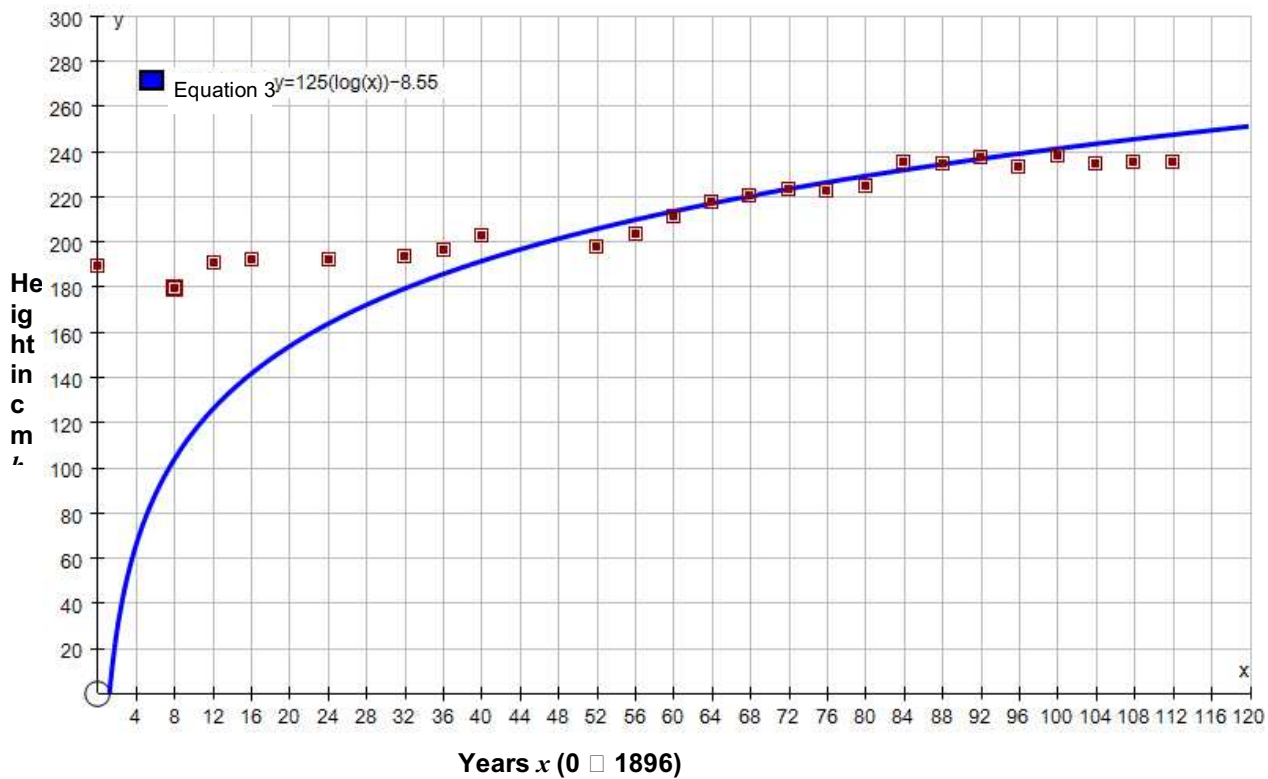
Year with year 1896 as 0	0	8	12	16	24	32	36	40	52
Height in cm	190	180	191	193	193	194	197	203	198

Year with year 1896 as 0	56	60	64	68	72	76	80	84	88	92	96
<b>Height in cm</b>	204	212	216	218	224	223	225	236	235	238	234

Year with year 1896 as 0	100	104	108	112
Height in cm	234	235	236	236

**Table 2.1** Data table for the height of the gold medallists in high jump from 1896 until 2008, with year 1896 as year 0.

In order to be able to rate how well the graph is fitted to depict the additional data the new data from *Table 2.1* is taken and plotted with the modelled equation giving following graph.



**Figure 4.0** Graph with data points from **Table 2.1** years  $x$  versus Height in centimetres  $h$  with Equation 3.

Note that the graph is quite well fitted starting from 1928 (32) until 2008 (112). However then the graph falls more rapidly as it approaches  $x=0$  and deviates greatly from the data points.

However the graph can be modified to better fit the data points.

The general equation  $y = a \log(x) + b$  gives one more opportunity to modify the function further to meet one's needs. Until now only  $b$  and  $k$  have been used.

In order to attain a better fit graph the suitable values for  $c$  and  $d$  will be determined.

One of the main problems as one can see from **Figure 4.0** and as repetitively mentioned is the fact that

the graph does not give a valid value for  $x=0$ .

To change that a  $d$  has to be defined as it shifts the asymptote.

Using the point  $(0, 190)$  and substituting for  $x$  and  $h$  respectively into  $y = 125(\log(x) + 0.55)$  is bound to give one a new graph with a valid  $y$ -value for  $x$ .

$y = 125(\log(x) + 0.55)$

$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$

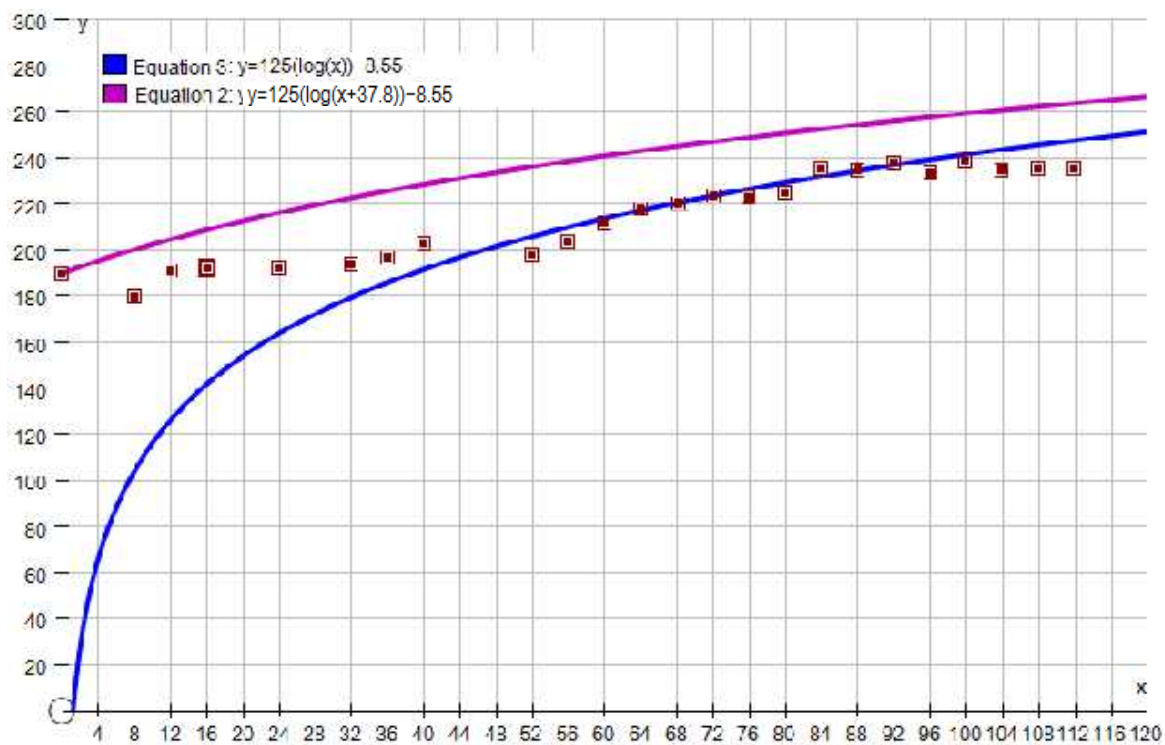
$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$

$y = 125(\log(x + 37.8)) - 8.55$



The new equation now is  $y = 125(\log(x + 37.8)) - 8.55$ .

To check if there has been any improvement the new graph is added to **Figure 4.0** in order to compare it with the original equation.

**Figure 4.1** Graph with data points from **Table 2.1** years  $x$  versus Height in centimetres  $h$  with Equation 3 and the newly attained Equation here Equation 2.

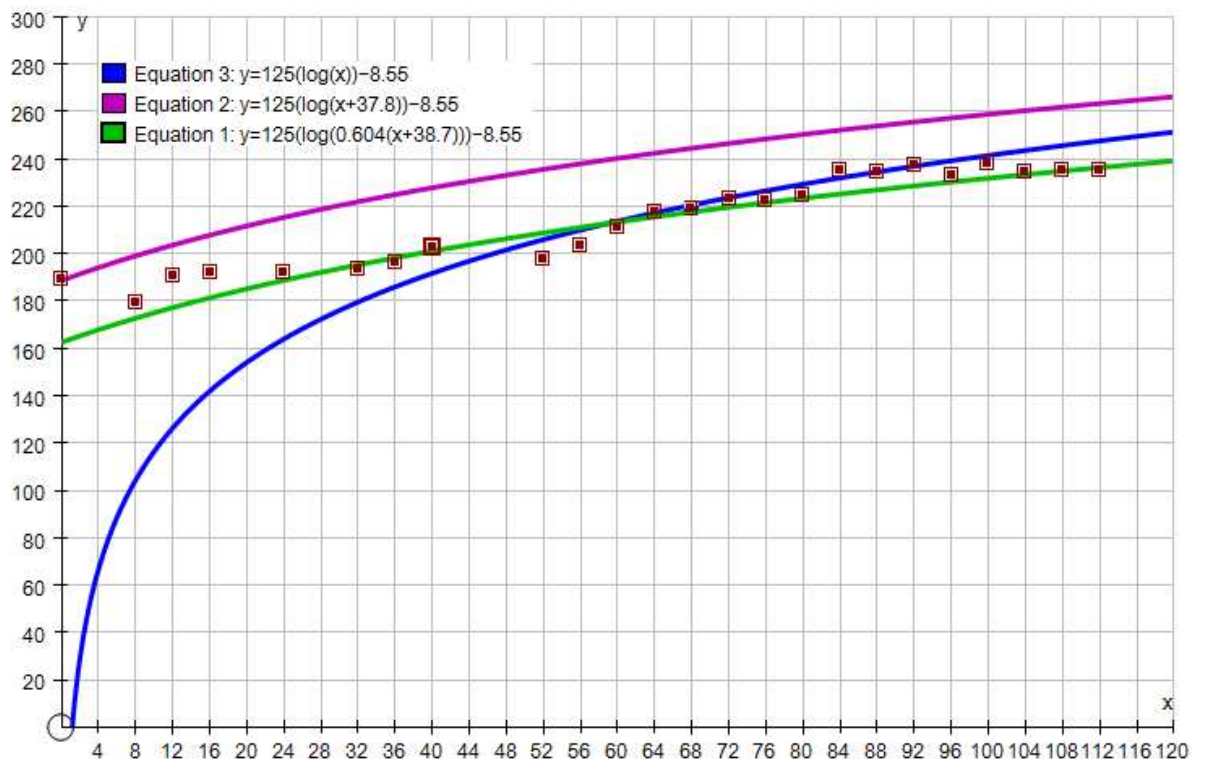
It becomes clear that even though the problem of the asymptote has been solved that the new graph misses the other data points by a great deal. This can however be easily corrected as the parameter  $c$

which has not been used yet, can flatten the equation and so moves it towards the data points again.

In order to get the best possible value of  $c$  the last point (112, 236) is used as it is the last point and in order with the general trend. The formulae of the function will be expanded by adding  $c$  and its value will be calculated by substituting 112 as  $x$  value and 236 as  $h$  value.

$$\begin{aligned}
 &125 \times \log \left( \frac{112 + 37.8}{0.604(112 + 38.7)} \right) - 8.55 \\
 &= \frac{125 \times \log \left( \frac{150.8}{4.14288} \right) - 8.55}{125} \\
 &= \frac{125 \times \log(36.4002) - 8.55}{125} \\
 &= \frac{125 \times 1.5613 - 8.55}{125} \\
 &= \frac{195.1625 - 8.55}{125} \\
 &= \frac{186.6125}{125} \\
 &= 1.4929
 \end{aligned}$$

And again in order to confirm if the addition of a parameter improved the graph the function will be added to Figure 4.1 to compare the graphs with each other.



The new equation here represented by the green graph is the best

Height in



suited one as it does not have a visible asymptote nor differs greatly from the data points. The trend line passes the data points with the general shape of a slight positive slope into a plateau.

The general trend of the data points show an over all increase of height as the years progress.

However there are a few random fluctuations throughout the graph.

The first notable swing can be seen from year 1896 to 1908. During this period there were foreshadows of the Great War, where the nations main concern were not athletic competition, but preparations for the on coming war. The First World is also responsible for the missing value at  $x=20$  which represents the year 1916, as with the Olympics there is also the ancient tradition that no games can be held if there is war.

The same explanation is there for the missing values of the years 1940 (44) and 1944 (48) as the Second World War was raging.

No value is available for the 1900 Olympics which were actually held in Paris. However they were part of the Parisian World Exhibition, the same goes for the 1924 Olympics also held in Paris, however they were not part of a World Exhibition.

One can see a pattern that whenever a year's value is missing the next years height drops below the height of the previous years.

The graph starting from 1960 (64) seems to be reaching a plateau however this trend is broken in 1968 (72). This is the same year as the Fosbury Flop was introduced and started to become the prominent high jump technique.

Starting from that year again a steady increase can be seen until again a plateau is reached 1988 (92). From that point on the height mostly stayed the same.

In this portfolio for the display of equations the software MathType Light (Version 6.7a) was used.

The graphs were displayed using Autograph (Version 3.3.0010) and the same software was also used to generate the Lines of best Fit.

The data tables were taken from Microsoft Excel Starter (Version 14.0.5128.5000).