**Math SL Type II Portfolio**

Gold Medal Heights

Bea Zimmermann

11/15/11

The Olympic Games are an international event where athletes from over 200 countries around the globe come together to compete in a variety of winter and summer sporting events. The ancient Olympic Games began in Olympia, Greece and were held from 8th century BC to 4th century AD. The first modern Olympic Games was held in 1896, and the Olympic Games have been a mark of historical events and conflicts since their beginning; World Wars led to the cancellation of the 1916, 1940, and 1944 Games, and boycotts during the Cold War very nearly cancelled the 1980 and 1984 Games.[[1]](#footnote-1)

Men’s high jump is a very popular sport in the summer Olympic Games, where the goal is to jump over a bar of a certain height without knocking it down; the gold medalist is the athlete who can jump over the tallest bar height. The aim of this paper is to consider the winning heights for men’s high jump in several Olympic Games and analyzing the trends and fluctuations. To do this, we will create several functions and use them to analyze the data.

**The table below** shows the height (in centimeters) achieved by the gold medalists at various Olympic Games.

**Plotting the Data:**

To better understand the data, our first step will be to plot the data into the calculator and then graph it. That way, we can better determine what kind of equation to create.

The graph of the points from our table (Figure 1) is shown below in a screen-shot from the calculator.

|  |  |
| --- | --- |
| Year | Height (cm) |
| 1932 | 197 |
| 1936 | 203 |
| 1948 | 198 |
| 1952 | 204 |
| 1956 | 212 |
| 1960 | 216Height (cm) |
| 1964 | 218 |
| 1968 | 224 |
| 1972 | 223 |
| 1976 | 225Figure 2 |
| 1980 | 236Year |

 Figure 1

The parameters are limited to positive numbers because it is not logical to have a negative year or negative winning height. The variables are defined as labeled on Figure 2.

Some possible constraints of the task at hand include the fact that we have no data from before 1896 and the data we have is somewhat sporadic from varying years, so we are constrained by the data we have been given. In addition, the Olympic Games were not held in 1940 and 1944 due to World War II, and in some years there were fewer participants because of boycotts of the games.

Now that the points have been plotted onto a graph, the next step is to create and equation by hand that precisely fits the data by use of matrices. This will help to eventually predict trends in the data.

**Creating an Equation Analytically:**

In order to analytically (by hand) create an equation that will fit this data, one must first decide which type of function best fits the data points. I have chosen a **cubic function** because its standard curved shape seems as though it will fit the plotted points well.

The standard cubic formula is:

$$ax^{3}+bx^{2}+cx+d$$



The next step is to choose four points which we will plug into this function. As Figure 1 shows above, our data is in terms of years. When creating an equation by hand, 1932, for example would be a very big number to plug into the above standard equation, and most likely the calculator could not compute 19323 very easily. So in order to make this somewhat simpler, we will subtract 1900 from our List1 (L1) values in the table and use the new L3 points for the X values instead, as shown in Figure 4.

Figure 4

I have chosen the following points:

(32, 197), (52, 204), (64, 218), (80, 236)

Next, **plug the points into to the X and Y values** of the standard cubic formula, as shown below.

$$197=(32^{3})a+(32^{2})b+\left(32\right)c+d$$

Continue plugging in values and the resulting system of equations looks like this:

$$197=32768a+1024b+32c+d$$

$$204=140608a+2704b+52c+d$$

$$218=262144a+4096b+64c+d$$

$$236=512000a+6400b+80c+d$$

Our goal is to find out the values for the variables *a, b, c, d* that will plug into the standard cubic equation above.

Matrices are ideal for solving systems of equations with multiple variables. A matrix equation will allow us to find all of the variables at once.

To do that, plug the equations into two matrices, one for the values on the left side of the equal sign, and one for the values on the right side of the equal sign.

$$\left[\begin{matrix}32768&1024&32&1\\140608&2704&52&1\\262144&4096&64&1\\512000&6400&80&1\end{matrix}\right]×\left[\begin{matrix}a\\b\\c\\d\end{matrix}\right]=\left[\begin{matrix}197\\204\\218\\236\end{matrix}\right]$$

**A B C**

To find the values of *a, b, c, d*, we must solve for matrix B, so we must solve this matrix equation in the following way.

$$\left[B\right]=\left[\begin{matrix}A\end{matrix}\right]^{-1}×\left[C\right]$$

Once that formula is applied to the matrix equations above, one ends up with this:

$$\left[\begin{matrix}a\\b\\c\\d\end{matrix}\right]=\left[\begin{matrix}32768&1024&32&1\\140608&2704&52&1\\262144&4096&64&1\\512000&6400&80&1\end{matrix}\right]^{-1}× \left[\begin{matrix}197\\204\\218\\236\end{matrix}\right]$$

Use a scientific or graphing calculator to solve the right-hand side of the equation:

$$\left[\begin{matrix}a\\b\\c\\d\end{matrix}\right]=\left[\begin{matrix}-5.62686×10^{-4}\\0.108798\\-5.755059\\288.19048\end{matrix}\right]$$

After the calculations, plug each corresponding value for a, b, c, d into the original standard cubic equation formula to create our cubic equation that fits this data.

**Cubic Equation:**

**A B C**

**A B C**

$$y=-0.00056x^{3}+0.10879x^{2}-5.75506x+288.19048$$

Here is **the graph of the equation** we have created. It fits the data well; nearly as well as the automatic calculator regression would fit. In any case, it is accurate enough for creation by hand.



*Years*

*Height of jump (cm)*

Figure 5

To check our first equation, we can use the calculator. The calculator can automatically create the most accurate regression equation possible, with a few simple steps.

However, in analyzing the equation we have created above, **it is necessary to see if the equation logically fits the pattern of the data**. If the equation fits the data within the domain, it is different than if the equation logically fits what the data would likely do over a long period of time.



The cubic equation increases exponentially and then sharply declines after creating an exponential curve, as shown in Figure 6, a larger view of our equation as a whole. It is not logical that the winning high jump heights would sharply decline after a certain time period, or after reaching a certain height

*Years*

*Height of jump (cm)*

Figure 6

Instead, we would expect that the winning heights would level off after reaching a certain height, because so far, and most likely for the next few decades, the human body is only capable of jumping a certain height. That height may vary, increasing or decreasing as the years progress, but we would expect that graph of the winning heights would remain somewhat stable and level.

We are looking for an equation that allows us to predict trends and to analyze the data over long periods of time. Being restricted to a domain, like this cubic equation is, prevents us from doing that. We must conclude that our equation will not fulfill the purposes of what we are trying to accomplish with this particular information.

The better option would be a **logistic** equation, which does level-off after curving upwards. However, creating a logistic function analytically is beyond the scope of this paper, so we will utilize a graphing calculator to create this equation.

**Creating the Logistic Regression Equation with a Calculator:**

In order to create this equation, the y-values in our data table must not be more than about 10. If so, the calculator cites a domain error and does not create the regression. So it is necessary to create a new table, where the y-values have been subtracted as much as possible.



Figure 7

Figure 8

Starting with our original data table, we will create a new list of x-values and y-values by subtracting. As you can see in Figure 7, I subtracted 1930 from List 1 or L1 to create L3. In Figure 8, I subtracted L2 by 190 to create L4. We will be working with L3 as the x-values and L4 as the y-values.

We will now require a new window for this set of data. The points have been subtracted by the same amount in each column, so they remain spatially the same.

Here are the new points, plotted on a graph:



*Years*

*Height of jump (cm)*

Figure 9

The next step is to use the calculator to create a logistic equation for our data using L3 and L4.

Here is a snapshot of the regression equation information after the calculator created it:

Figure 10

**Logistic Regression Equation:**

$$y=85.350986/(1+12.384766e^{-.051178x})$$

Here are both of the graphs of the **cubic** (Figure 11) and **logistic** (Figure 12)



*Years*

*Height of jump (cm)*

*Years*

*Height of jump (cm)*

Figure 12

Figure 11

The logistic regression fits the data very well. Now it is necessary to **compare** it to our previous cubic equation and discuss the **merits and shortcomings of both functions in describing and displaying the trends in the data.**

While the cubic function seems to fit the data much more precisely than the logistic, it makes much more sense that the data would level off after a certain height, because it is illogical that suddenly the winning height would drop down to almost zero and into the negative numbers by the next Olympic Games. It is much more logical that the winning heights would level-off after reaching a certain maximum, as displayed in the graph of the logistic regression in Figure 13.

Figure 13

*Years*

*Height of jump (cm)*

We must also consider that it is optimized only within a certain domain. Outside of the domain of (32, 197) and (80, 36), the cubic function would not make sense in relation to the trends of the data

However, both of the models have their limitations. As Figure 13 shows, the winning height would increase exponentially, to a height of about 271 inches, for a couple of years before leveling off. This is not ideal, considering that we would not expect the winning height to increase by more than a few inches over the course of the next few decades. However, it is still more logical considering the data at hand and more logical considering our goal of evaluating trends.

**Further Evaluation of the Equations:**

The Olympic Games were not held in 1940 and 1944 due to World War II, and so to further compare the two equations, we will use both functions to estimate what the winning heights would have been in 1940 and 1944.

To solve for this data with the cubic function, we will plug in the year 1940 for X. We must first take 1940 and subtract 1900 to get 40, like we did in order to create the function (refer to figure 4). We will then plug in 40 for X and solve for Y. We will repeat the same process for the year 1944.

Cubic function:

$$y=-0.00056\left(40\right)^{3}+0.10879\left(40\right)^{2}-5.75506(40)+288.19048$$

$$y=196.21 cm$$

$$y=-0.00056\left(44\right)^{3}+0.10879\left(44\right)^{2}-5.75506(44)+288.19048$$

$$y=197.88 cm$$

To solve for the data with the logistic function, we must subtract 1930 from our x-value of 1940 to get 10, again like we did when creating the logistic function (refer to figure 7). When we get the answer for y, we must also add 190 back to it, because we subtracted 190 from the y-values when creating the equation (refer to figure 8). We will repeat that process for the year 1944 as well.

Logistic function:

$$y=85.350986/(1+12.384766e^{-.051178(10)})$$

$$y=200.13 cm$$

$$y=85.350986/(1+12.384766e^{-.051178(14)})$$

$$y=202.11 cm$$

The answers for both functions make sense within the context of the problem at hand, considering that the winning heights for those years would be expected to be between 190 and 210. Both functions predicted values for the missing years that could very likely have been the winning heights. As the table (figure 1) shows, the winning heights did not increase by steady increments, instead, some years the heights increased and the following years they might have decreased by a few centimeters. So while the values for the logistic function are slightly higher, they would be as logical of predictions as the cubic values.

|  |  |  |
| --- | --- | --- |
|  | Cubic function | Logistic function |
| Winning height 1940 (cm) | 196.21 | 200.13 |
| Winning height 1944 (cm) | 197.88 | 202.11 |

Figure 14

A more demonstrative evaluation of the equations would be to predict values for years that are outside of the domain of the data that was given. Both equations can so accurately predict the values for 1940 and 1944 because those values fall within the domain of the given points, and both equations are very similar and extremely accurate within those domains. So now, we will predict the heights for the years 1984 and 2016 in order to further evaluate the functions.

For explanation of this following process, refer back to the explanations on page 6 under “Further Evaluation of the Equations.”

Cubic function:

$$y=-0.00056\left(84\right)^{3}+0.10879\left(84\right)^{2}-5.75506(84)+288.19048$$

$$y=240.47 cm$$

$$y=-0.00056\left(116\right)^{3}+0.10879\left(116\right)^{2}-5.75506(116)+288.19048$$

$$y=210.38 cm$$

|  |  |  |
| --- | --- | --- |
|  | Cubic function | Logistic function |
| Height 1984 (cm) | 240.47 | 237.92 |
| Height 2016 (cm) | 210.38 | 264.09 |

Logistic function:

$$y=85.350986/(1+12.384766e^{-.051178(54)})$$

$$y=237.92 cm$$

$$y=85.350986/(1+12.384766e^{-.051178(86)})$$

$$y=264.09 cm$$

Figure 15

|  |  |
| --- | --- |
| Year | Height |
| 1896 | 190 |
| 1904 | 180 |
| 1908 | 191 |
| 1912 | 193 |
| 1920 | 193 |
| 1928 | 194 |
| 1984 | 235 |
| 1988 | 238 |
| 1992 | 234 |
| 1996 | 239 |
| 2000 | 235 |
| 2004 | 236 |
| 2008 | 236 |

The values found and displayed in Figure 15 demonstrate how well the two equations we created are able to predict future values and trends. The cubic function predicts a quite reasonable height for 1984, but the height for 2016 is more than 30 cm lower than the height for 1984, which is not reasonable, considering that athletes competing in high jump would have decreased their best heights by more than 30 cm over the span of 8 Olympic Games. If anything, they would most likely keep improving one or two centimeters every four years. As the additional data from additional years not already in our first table show us, the actual height in 1984 was 235 cm.

The logistic function predicts a reasonable height for 1984 as well; however its prediction for the height in 2016 is quite a bit higher than would be possible. The function predicts that by 2016 the winning height will be 264 cm, and as the additional data in Figure 16 shows, the winning height as late as 2008 was only 236 cm and trends show the height increasing or decreasing by no more than 5 cm from 1984 onwards. But while the height for 2016 given by the logistic is somewhat unreasonable, it is more logical to conclude that the height would increase rather than decrease drastically.

Figure 16

Having shown that the logistic model is the more reasonable of the two, we will now use the data given in Figure 16 in addition to the data from Figure 1 and evaluate how well this function fits that new data.

Shown below is the graph of the logistic equation and all of the available data, including the years before and after our original domain of years:

As the graph shows, the logistic model fits the data very well in the earlier years of the high jump heights. From about 1896 to 1988, the data fits quite well. But the logistic function continues to increase steadily until about 2040, when it finally levels off. But the actual data starts to level off around 2008, and that is where the logistic function ceases to be quite as accurate a predictor of future heights.

*Years*

*Height of jump (cm)*



Figure 7

*The graph above has several negative values on the graph that are in reality not. In order to graph the function with the additional points, because we have to subtract 1930 and 190 when creating the function, the points are shifted downwards, causing the earlier years of the data to appear as though they are negative. The tables (figures 1 and 16) are correct in their values for each year.*

**Discussion of Overall Trends in the Data:**

The overall trend from 1896 to 2008 is one that increases gradually, with a few outliers. After 2008, it begins to level-off, with fluctuating only a few centimeters while remaining around 235 cm.

In years where conflicting global events caused athlete attendance to be lower, it is likely that the winning height would be shorter. When high jump first became an event in the Olympics, men started at a baseline for the best possible height, because at first there was less intensive training and there was not a winning height to gage their performance upon. Afterwards, a standard was set and athletes continued to improve and train harder to achieve a higher and higher jump. As the sport developed, athletes and coaches developed new methods of jumping over the bar, for example the straddle technique instead of approaching the bar head on or with a scissor technique.

With the implementation of higher, softer landing mats allowed Dick Fosbury to develop the Fosbury Flop which won him the gold in 1986, and the flop, where one approaches the bar from the side and jumps with a twisting flopping motion, has become standard high jumping practice.[[2]](#footnote-2) New, more rigorous training methods were devised and put into practice, and equipment innovations helped athletes to improve their jumps.[[3]](#footnote-3) Specialized footwear with thicker soles and spikes has been designed to improve jumping technique. All of these factors account for the exponential increase in heights after about 1952. Steroids have been known to be used in track and field events, which could potentially account for the spike in heights from 1976 to 1980.

The logistic model we created clearly does not fit the data as well as it should. The function continues to increase exponentially despite the earlier leveling off of the actual data. Predictions for future heights would be much higher than realistically possible. The logistic would be in need of modifications in order to fit the data that continues past 1980. If we used the new data in our data table and created a new logistic regression, the new equation would better predict winning heights because it would take into account when the real-life data begins to level off. Sometimes it is necessary to create new equations when new data is presented, because the more data is available, the more accurate the equation will be. Then, it is possible to predict future trends more accurately, which can be vital when considering more serious topics like climate change or volcano eruptions or global population and water usage.

I would proceed to create a new logistic function that includes all of the available data; however the range of data makes it very difficult to shift the data so that the calculator can process the new logistic function, as it does not work with negative values or values that are too large.

1. <http://en.wikipedia.org/wiki/High_jump> [↑](#footnote-ref-1)
2. <http://www.benchallenger.com/high-jump/history-of-the-high-jump> [↑](#footnote-ref-2)
3. <http://en.wikipedia.org/wiki/High_jump> [↑](#footnote-ref-3)