

# Model Function of Women's 1500m World Record Improvement over Time

Annie Allmark  
*University of South Florida*

Advisors:

Arcadii Grinshpan, Mathematics and Statistics  
Waren Bye, USF Athletics Department, Track and Field Head Coach

Problem Suggested By: Annie Allmark

Follow this and additional works at: <https://scholarcommons.usf.edu/ujmm>

 Part of the [Mathematics Commons](#)

UJMM is an open access journal, free to authors and readers, and relies on your support:

[Donate Now](#)

## Recommended Citation

Allmark, Annie (2018) "Model Function of Women's 1500m World Record Improvement over Time," *Undergraduate Journal of Mathematical Modeling: One + Two*: Vol. 8: Iss. 2, Article 3.

DOI: <http://doi.org/10.5038/2326-3652.8.2.4890>

Available at: <https://scholarcommons.usf.edu/ujmm/vol8/iss2/3>

---

# Model Function of Women's 1500m World Record Improvement over Time

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

## Abstract

We give an example of simple modeling of the known sport results that can be used for athletes' self-improvement and estimation of future achievements.

This project compares the women's 1500-meter world record times to the time elapsed between when they were run. The function of time which describes this comparison is found through graphing the data and interpreting the graphs. Then the obtained model function is compared to the real time data. The conclusions drawn from the result include that the calculated function of time lacks in accuracy as time elapsed increases, but the model could be used to estimate the future world records.

## Keywords

track and field, running, exponential modeling, line of best fit

## PROBLEM STATEMENT

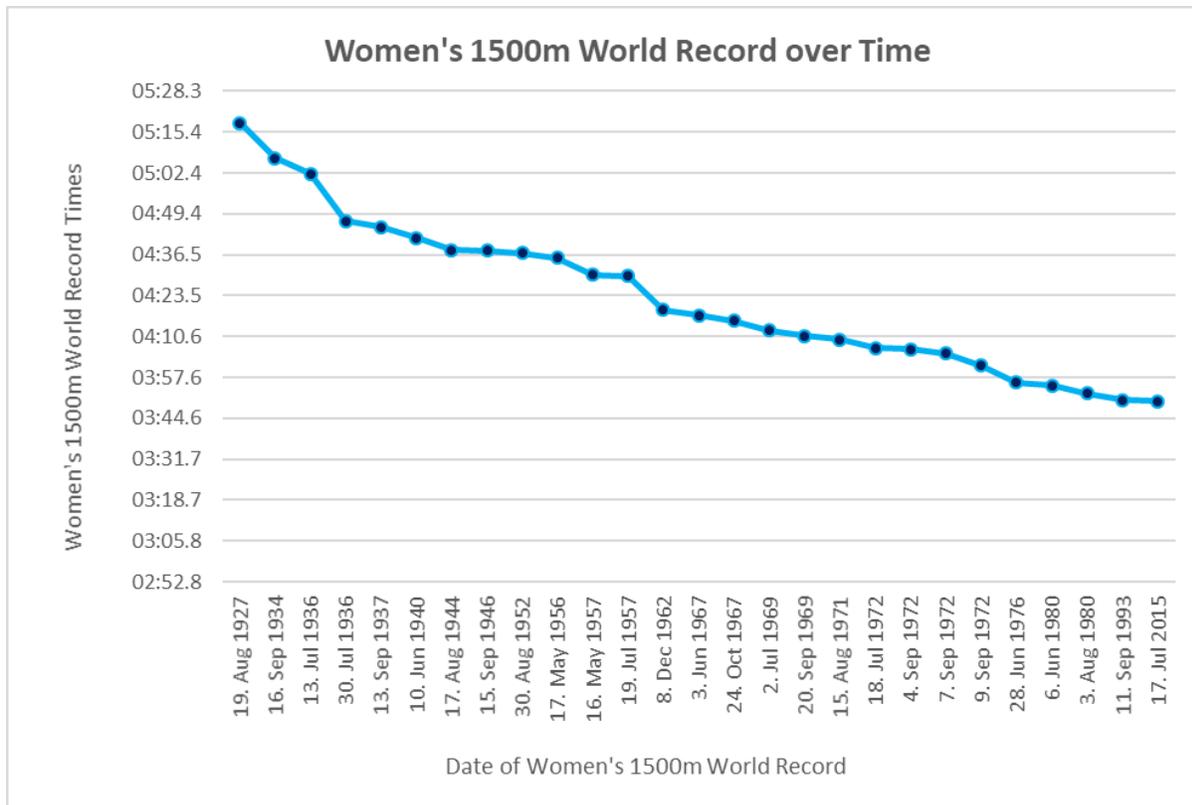
The problem investigated by the author in this paper is to find a function of time that models the women's 1500-meter world record over time. The obtained function is then compared to the real time data. The discussed example of simple modeling of the known sport results can be used for athletes' self-improvement and estimation of future achievements in various sports.

## MOTIVATION

I am a track runner at the University of South Florida, and my specialty event in high school was the 1600-meter run. Now, that I am at a new level, this distance is changed to the 1500 meter, for the Outdoor season of Track and Field. Since I have never run this distance I think it would be interesting to learn a little bit more about it, especially about its improvements over time. This way I can gauge what a good 1500-meter time is, therefore I can set an achievable yet far-reaching goal for myself. I want to do this by finding a function that describes the improvement of the women's 1500m world record over time. I can indirectly use this to measure my own progress of the 1500m times. Obviously I am not a world class athlete and therefore cannot run these times, but I can use it as a ratio of my improvements.

## MATHEMATICAL DESCRIPTION AND SOLUTION APPROACH

The first step is to graph the data to see if there is any correlation between the women's 1500-meter world records, the dependent variable, and the dates when they had been run, the independent variable.



It is easy to notice a downward sloping trend; however, it is not possible to immediately determine a more specific shape that the data might fit. Let us convert the time variable into months, and the time records into seconds. Here is an example of calculation:

*July 1972 – August 1927*

$$1972 - 1927 = 45 \times 12 = 540 - 1 = 539$$

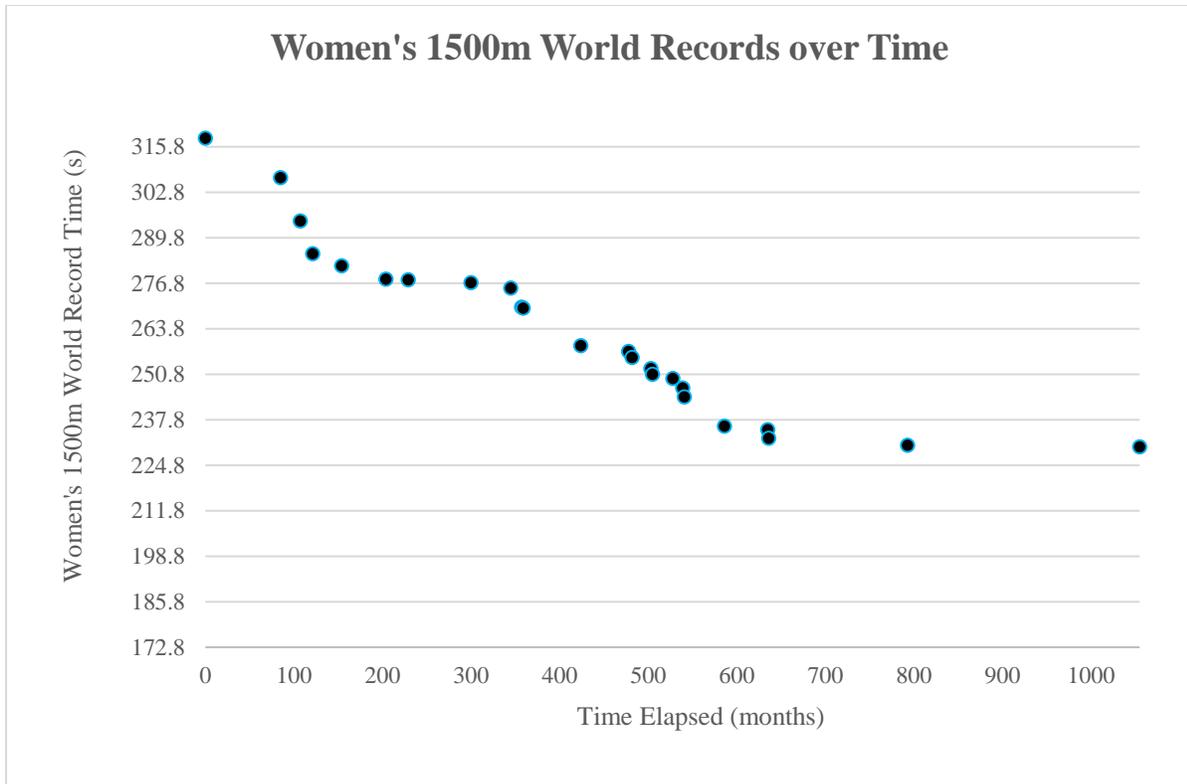
$$4:06.9 \rightarrow \text{seconds}$$

$$4\text{min} \times 60 = 240\text{s} + 6.9\text{s} = 246.9\text{s}$$

Let us also combine and take the average of world records that are run within the same month, as the following example illustrates:

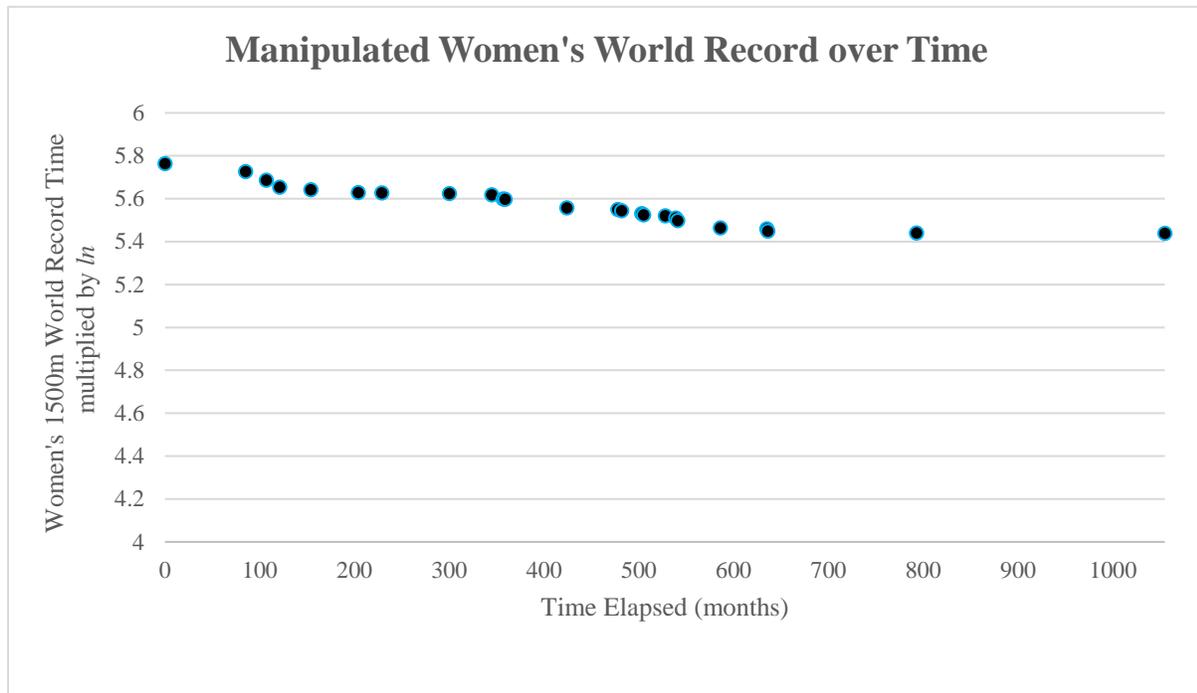
$$\frac{246.5\text{s} + 245.1\text{s} + 241.4\text{s}}{3} = 244.3\text{s}$$

With these new data points we can consider a new graph which produces a clearer pattern.



It roughly resembles the lower part of a graph of an exponential curve that is reflected across the  $y$ -axis, i.e. the data points can be approximately described by a function of the following type:

$f(t) = ae^{-bt}$  ( $a, b > 0$ ). After performing a transformation on the dependent variable we can find the best fit curve with semi-logarithmic coordinates.



We use the standard approach to find the line of best fit (see, e.g., Biocalculus, Section 11.3). It is determined by finding a line  $y(x) = b_0 + b_1x$  that has the least total sum of squared deviations from the data points on the scatter plot:  $(x_k, y_k)$ ,  $k = 1, \dots, n$ . In fact, we need to minimize the following function of two variables:  $S(b_0, b_1) = \sum_{k=1}^n (y_k - b_0 - b_1x_k)^2$ . To do it we use partial differentiation with respect to  $b_0$  and  $b_1$  separately to find the critical points of  $S$ :

$$\frac{\partial S}{\partial b_0} = -2 \sum (y - b_0 - b_1x) \quad (1)$$

and

$$\frac{\partial S}{\partial b_1} = -2 \sum x(y - b_0 - b_1x). \quad (2)$$

Partial derivatives (1) and (2) are both set equal to 0. We have:  $0 = \sum y - \sum b_0 - \sum b_1x$ .

Because the variable  $b_0$  in (1) remains constant throughout this data, we replace the summation in front of  $b_0$  sign with  $n$ . Hence

$$b_0 = \frac{\sum y - b_1 \sum x}{n}. \quad (3)$$

The second equation that is derived with respect to  $b_1$  will be simplified. We distribute the  $x$  to each term in the parentheses. Since both  $b_0$  and  $b_1$  are constants they do not have to be multiplied each time by  $x$  and  $x^2$  respectively, thus we take out and move the  $b$  variables in front of the summation sign:  $0 = \sum xy - b_0 \sum x - b_1 \sum x^2$ .

Now, we substitute the equation (3) for  $b_0$  into the above equation to obtain:

$$b_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}. \quad (4)$$

Formula (4) allows us to find the variable  $b_1$  using the known data points. Once  $b_1$  is found we can use (3) to find  $b_0$ . The sum of  $y$  and the sum of  $x$  divided by  $n$  is equal to the mean of the  $y$  and  $x$  values, respectively. Therefore equation (3) can be used in the following form:

$$b_0 = \bar{y} - b_1 \bar{x}. \quad (5)$$

Using the data above we find that the sum of the  $x$ -values is 9693, the sum of the  $y$ -values is 133.646693, the sum of  $x^2$ -values is 5247311, and the sum of  $xy$  is 53528.0686.

From (4) it follows that:

$$b_1 = \frac{24(53528.0686) - (9693)(133.646693)}{24(5247311) - (9693)^2} = -0.00033656635. \quad (6)$$

We can restrict ourselves in (6) to four significant figures, obtaining eventually:

$$b_1 \approx -3.3657 \times 10^{-4}. \quad (7)$$

Now the value of  $b_1$  can be used in (5) to find the value of  $b_0$ . The mean of the  $y$ -values is 5.568612472 and the mean of the  $x$ -values is 403.875. We obtain:

$$b_0 \approx 5.7045 \quad (8)$$

By (7) and (8) the line of best fit is the following:

$$y = -3.3657 \times 10^{-4}x + 5.7045. \quad (9)$$

Observe that equation (9) is a line in semi-logarithmic coordinates. In order to find the line of best fit for the exponential curve, we apply the exponential function to both sides of (9). We obtain:  $e^y = 300.2153e^{-3.3657 \times 10^{-4}x}$ . Hence the following function approximately describes how the women's 1500-meter world record changes over time  $t$ :

$$f(t) = 300.2153e^{-3.3657 \times 10^{-4}t}. \quad (10)$$

## DISCUSSION

The function (10) can be used to predict future world records from the data known so far.

However the function can produce the results of varying accuracies. For example, when we consider month 482 and substitute it into the function, we obtain  $f(482)=255.3s$ , which is extremely close to 4:15.6 (255.6s) the actual world record that was run on 24. October 1967. On the other hand, when we estimate the latest world record, the result is much less accurate:  $f(1055)=210.5s$ . While the world record run by Genzebe Dibaba on 17. July 2015 was 3:50.1 (230.1s); the function provides a time that is almost 20s lower than the actual world record. This indicates that exponent of  $e$  is too big in absolute value, and thus causes the function to decrease too rapidly, predicting the world records to happen faster and to be substantially lower times than what will most probably occur.

The domain of this function is also relatively small, as the 1500-meter run did not become an official event until the year of 1927, so we have available data for only 90 years. As

the domain is small, it allows for less data to be collected, fewer opportunities to break the world record. Another limitation of this function is that it approaches zero, as time approaches infinity:

$$\lim_{t \rightarrow \infty} 300.2153e^{-3.3657 \times 10^{-4}t} = 0.$$

However, no matter how much time passes a human being will never be able to run 1500 meters for a second, it is impossible. The human limit is unknown, and therefore the limit to the real time of this function is unknown. This unknown time limit is what drives athletes to continuously test their limits and attempt to break the world records. For this reason, the world record will much likely keep to decrease over time, but the decreases will become smaller and smaller, without ever going beyond the human limit.

## CONCLUSION AND RECOMMENDATIONS

The recommendations for this project to calculate a more accurate function of time include using a nonlinear regression, using the unit of days rather than months for the time elapsed, using the world record times measured to the hundredth of a second rather than the tenth of a second, and focusing on the accuracy of approximation/rounding of regression coefficients.

To have more manageable numbers, months rather than days were used for the measurement. And by choosing months as independent variables, we were forced to combine world records, and find the average, that was set in the same month, therefore decreasing the actual resemblance to the real time.

Today, the race times are usually measured to a precision which is hundredth a second, but can even be determined to a thousandth of a second if necessary. We only use data measured to a tenth of a second because the records set in the earlier 20<sup>th</sup> century were only measured to a tenth of a second. However, as the world record continues to get faster, it will do so in smaller

increments, and eventually the tenth increment will not be large enough to recognize a new world record and the hundredth of a second will be needed.

Furthermore, the rounding of the coefficient and constant in the line of best fit to the fourth decimal place, while we had numbers to the sixth decimal place, decreases the preciseness of the function. Therefore, a less accurate estimation of the potential world record times is found.

In conclusion, the function  $f(t) = 300.2153e^{-3.3657 \times 10^{-4}t}$  is a useful equation to predict the women's 1500-meter world record time. It can be used as a guide to compare future times to previous times.

A possible limitation may occur if one or multiple world records are found to be invalid because athletes were consuming Performance-Enhancing Drugs (PEDs). These times would be removed, and this could potentially change the function (10).

## Nomenclature

$s$ : second

$min$ : minute

$m$ : meter

$t$ : time

$x$ : values on the  $x$ -axis or an independent value in the data set

$y$ : values on the  $y$ -axis or the dependent value in the data set

$n$ : number of terms in the data set

$b_0$ : the constant in the line of best fit

$b_1$ : the coefficient in the line of best fit

$\bar{y}$  = the mean of the  $y$  values

$\bar{x}$  = the mean of the  $x$  values

## REFERENCES

“12.5- Logarithmic Graphs.” *12.5 - Logarithmic Graphs*,

[mathonweb.com/help\\_ebook/html/expoapps\\_3.htm](http://mathonweb.com/help_ebook/html/expoapps_3.htm).

“1500 Metres World Record Progression.” *Wikipedia*, Wikimedia Foundation, 6 Dec. 2017,

[en.wikipedia.org/wiki/1500\\_metres\\_world\\_record\\_progression](https://en.wikipedia.org/wiki/1500_metres_world_record_progression).

“Proof of Why the Linear Regression Equation Is the ‘Best Fit’ Equation.”

[coccweb.cocc.edu/srule/MTH244/other/LRJ.PDF](http://coccweb.cocc.edu/srule/MTH244/other/LRJ.PDF).

James Stewart, Troy Day, *Biocalculus: Calculus Probability, and Statistics for the Life Sciences*,

Cengage Learning (2016).