(Sample) Curve-Fitting Project - Linear Model: Men's 400 Meter Dash

Submitted by Suzanne Sands

(LR-1) **Purpose**: To analyze the winning times for the Olympic Men's 400 Meter Dash using a linear model

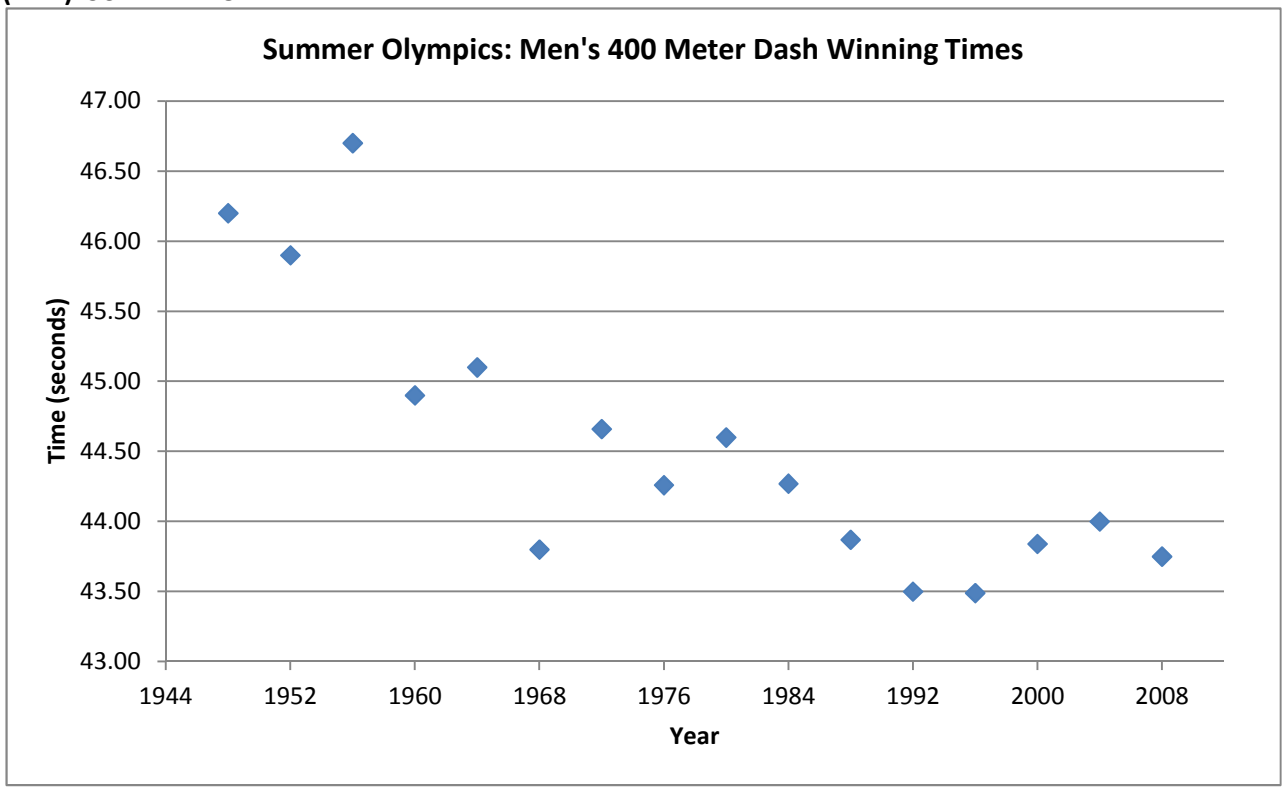
Data: The winning times were retrieved from http://www.databaseolympics.com/sport/sportevent.htm?sp=ATH&enum=130

The winning times were gathered for the most recent 16 Summer Olympics, post-WWII. (More data was available, back to 1896.)

DATA:

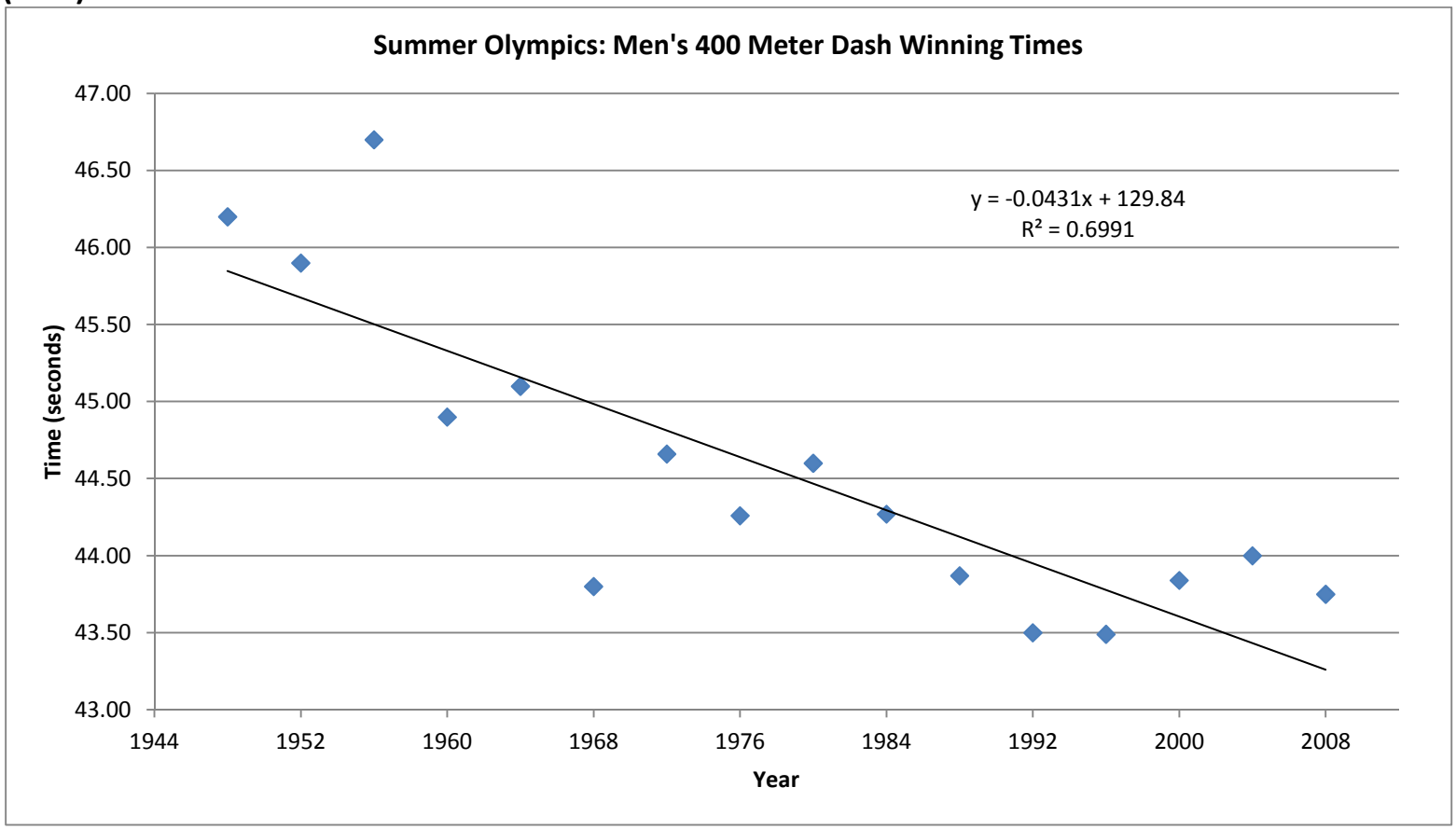
Summer Olympics: Men's 400 Meter Dash Winning Times	
Year	**Time (seconds)**
1948	46.20
1952	45.90
1956	46.70
1960	44.90
1964	45.10
1968	43.80
1972	44.66
1976	44.26
1980	44.60
1984	44.27
1988	43.87
1992	43.50
1996	43.49
2000	43.84
2004	44.00
2008	43.75

(LR-2) SCATTERPLOT:



As one would expect, the winning times generally show a downward trend, as stronger competition and training methods result in faster speeds. The trend is somewhat linear.

(LR-3)



Line of Best Fit (Regression Line)

$y = -0.0431x + 129.84$ where x = Year and y = Winning Time (in seconds)

(LR-4) The slope is –0.0431 and is negative since the winning times are generally decreasing.
The slope indicates that in general, the winning time decreases by 0.0431 second a year, and so the winning time decreases at an average rate of 4(0.0431) = 0.1724 second each 4-year Olympic interval.

(LR-5) Values of r^2 and r:

$r^2 = 0.6991$

We know that the slope of the regression line is negative so the correlation coefficient r must be negative.

$$r = -\sqrt{0.6991} = -0.84$$

Recall that $r = -1$ corresponds to perfect negative correlation, and so $r = -0.84$ indicates moderately strong negative correlation (relatively close to -1 but not very strong).

(LR-6) Prediction: For the 2012 Summer Olympics, substitute $x = 2012$ to get $y = -0.0431(2012) + 129.84 \approx 43.1$ seconds.

The regression line predicts a winning time of 43.1 seconds for the Men's 400 Meter Dash in the 2012 Summer Olympics in London.

(LR-7) Narrative:

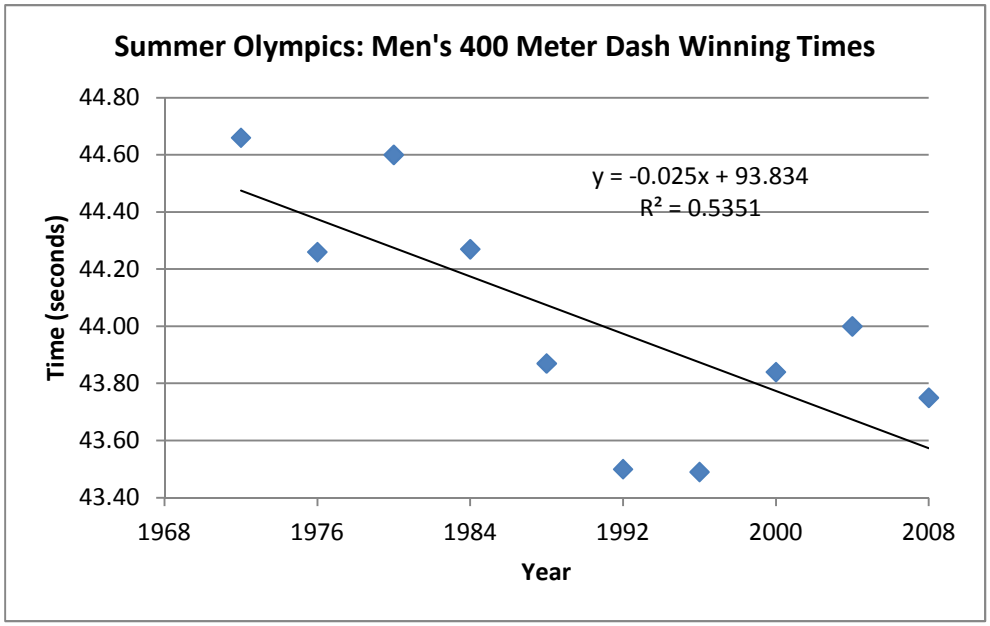
The data consisted of the winning times for the men's 400m event in the Summer Olympics, for 1948 through 2008. The data exhibit a moderately strong downward linear trend, looking overall at the 60 year period.

The regression line predicts a winning time of 43.1 seconds for the 2012 Summer Olympics, which would be nearly 0.4 second less than the existing Olympic record of 43.49 seconds, quite a feat!

Will the regression line's prediction be accurate? In the last two decades, there appears to be more of a cyclical (up and down) trend. Could winning times continue to drop at the same average rate? Extensive searches for talented potential athletes and improved full-time training methods can lead to decreased winning times, but ultimately, there will be a physical limit for humans.

Note that there were some unusual data points of 46.7 seconds in 1956 and 43.80 in 1968, which are far above and far below the regression line.

If we restrict ourselves to looking just at the most recent winning times, beyond 1968, for Olympic winning times in 1972 and beyond (10 winning times), we have the following scatterplot and regression line.



Using the most recent ten winning times, our regression line is $y = -0.025x + 93.834$.
When $x = 2012$, the prediction is $y = -0.025(2012) + 93.834 \approx 43.5$ seconds. This line predicts a winning time of 43.5 seconds for 2012 and that would indicate an excellent time close to the existing record of 43.49 seconds, but not dramatically below it.

Note too that for $r^2 = 0.5351$ and for the negatively sloping line, the correlation coefficient is $r = -\sqrt{0.5351} = -0.73$, not as strong as when we considered the time period going back to 1948. The most recent set of 10 winning times do not visually exhibit as strong a linear trend as the set of 16 winning times dating back to 1948.

CONCLUSION:

I have examined two linear models, using different subsets of the Olympic winning times for the men's 400 meter dash and both have moderately strong negative correlation coefficients. One model uses data extending back to 1948 and predicts a winning time of 43.1 seconds for the 2012 Olympics, and the other model uses data from the most recent 10 Olympic games and predicts 43.5 seconds. My guess is that 43.5 will be closer to the actual winning time. We will see what happens later this summer!

UPDATE: When the race was run in August, 2012, the winning time was 43.94 seconds.