Lab for Section 15.5

Use good notation and show appropriate work. Name _____

Definition. The *line of best fit (linear regression)* for a set of data points of the form (x, y) is of the form $y = mx + b$ where

$$m = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2} \text{ and } b = \frac{\left(\sum y\right) - m\left(\sum x\right)}{n}.$$

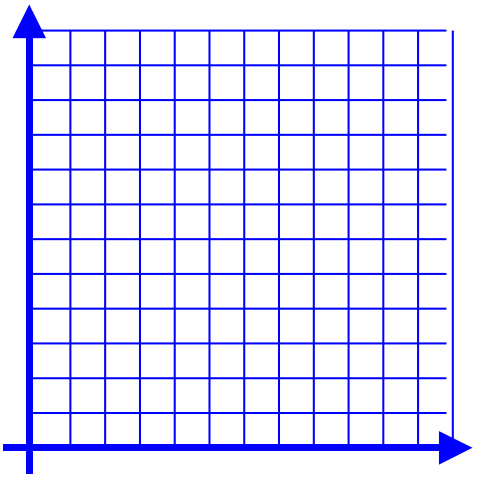
Formula for Computing the Linear Correlation Coefficient. For n data pairs (x, y) the *linear correlation coefficient*, r is given by

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

Example. The following data relates the ages in months, x, of babies to their weights in pounds, y.

x	y	xy	x^2	y^2
2	10			
3	12			
4	14			
6	14			
6	15			
11	22			

(a) Plot the data on a scatterplot.



(b) Determine the line of best fit.

(c) What do the y-intercept and slope represent?

(d) Use the line of best fit to estimate the weight of a 10 month old baby.

(e) Use the line of best fit to estimate the age of a 18 pound baby.

(f) Determine the linear correlation coefficient.

(g) Is the linear correlation a positive or negative correlation?

(h) Is there the linear correlation significant at the
(i) 0.05 level? (ii) 0.01 level?

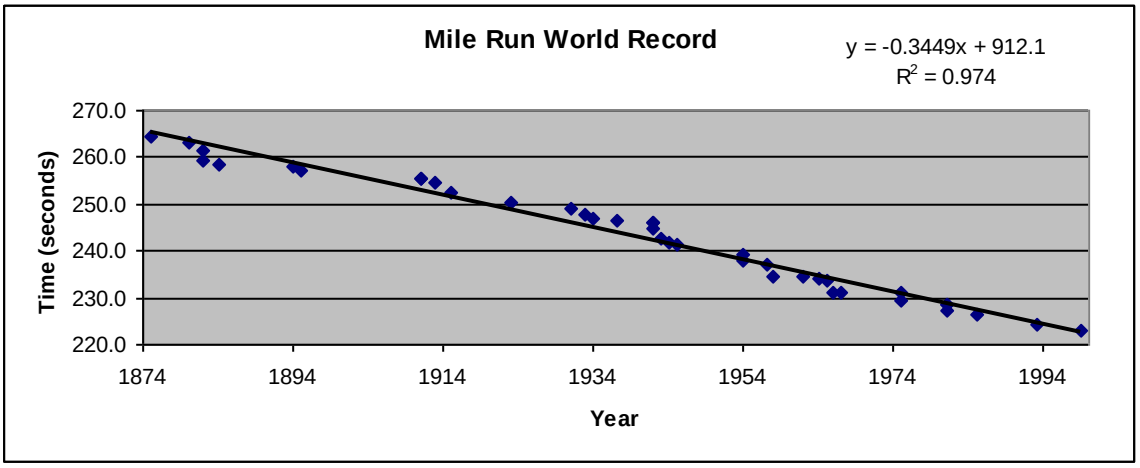
Progress of the World Record in the Mile Run

This activity has been adapted from an activity given in *A sourcebook of Applications of School Mathematics* prepared by a Joint Committee of The Mathematical Association of America and the National Council of Teachers of Mathematics, 1980.

In 1875, Walter Slade of Great Britain ran a mile in 4 minutes, 24.5 seconds. At that time, it was the fastest anyone had ever been timed for the mile. Now, the fastest high school students do better. In 1993, Hicham El Guerouj of Morocco set the current mile world record of 3 minutes 43.13 seconds. Here is how the record has evolved in the years since 1875.

1875	Walter Slade, Britain	4:24.5
1880	Walter George, Britain	4:23.2
1882	George	4:21.4
1882	George	4:19.4
1884	George	4:18.4
1894	Fred Bacon, Scotland	4:18.2
1895	Bacan	4:17.0
1911	Thomas Connett, U. S.	4:15.6
1911	John Paul Jones, U. S.	4:15.4
1913	Jones	4:14.6
1915	Norman Taber, U. S.	4:12.6
1923	Paavo Nurmi, Finland	4:10.4
1931	Jules Ladoumegue, France	4:09.2
1933	Jack Lovelock, New Zealand	4:07.6
1934	Glenn Cunningham, U. S.	4:06.8
1937	Syndey Wooderson, Britain	4:06.4
1942	Gunder Haegg, Sweden	4:06.2
1942	Arne Andersson, Sweden	4:06.2
1942	Haegg	4:04.6
1943	Andersson	4:02.6
1944	Andersson	4:01.6
1945	Haegg	4:01.4
1954	Roger Bannister, Britain	3:59.4
1954	John Landry, Australia	3:58.0
1957	Derek Ibbotson, Britain	3:57.2
1958	Herb Elliott, Australia	3:54.5
1962	Peter Snell, New Zealand	3:54.4
1964	Snell	3:54.1
1965	Michel Jazy, France	3:53.6
1966	Jim Ryun, U. S.	3:51.3
1967	Ryun	3:51.1
1975	Filbert Bayi, Tanzania	3:51.0
1975	John Walker, New Zealand	3:49.4
1981	Sebasten Coe, Britain	3:48.53
1981	Steve Ovett, Britain	3:48.40
1981	Coe	3:47.33
1985	Steve Cram, Britain	3:46.31
1993	Noureddine Morceli, Algeria	3:44.39
1999	Hicham El Guerouj, Morocco	3:43.13

Activities: The line of best fit and square of the correlation are printed in the upper right-hand corner of the scatterplot. The line on the scatterplot is the line of best fit.



1. What does the slope of the equation signify?

2. Use the equation for the line of best fit to predict the mile world record in 2010.

3. Use your equation to predict the year the mile world record may be 3:30.00.

4. Are there limitations to your model? Explain.

5. Find the linear correlation.
   (a) What two quantities are compared by the correlation?

   (b) Is the linear correlation a positive or a negative correlation?

   (c) Is the linear correlation significant ($\alpha = 0.01$)?