# Calculating Best Line of Fit

**(level 2.2)**

**PROBLEM:** Often when we collect data, we know the relationship between two parameters is linear but because of measurement error, the data points don't quite line up. If we want to do analysis on the data which requires the assumption that it is linear, we must find a best line of fit. The equation of a line is given by:

$$y = mx+b \tag{1}$$

where $m$ represents the slope of the line and $b$ represent the $y$-intercept of the line. The following 2 equations define these parameters according to a "least squares fit of the data":

$$m = \frac{n\sum_{i=1}^{n} x_i y_i - \left[\sum_{i=1}^{n} x_i\right] * \left[\sum_{i=1}^{n} y_i\right]}{n\sum_{i=1}^{n} x_i^2 - \left[\sum_{i=1}^{n} x_i\right]^2} \tag{2}$$

and

$$b = \frac{\left[\sum_{i=1}^{n} y_i\right] * \left[\sum_{i=1}^{n} x_i^2\right] - \left[\sum_{i=1}^{n} x_i y_i\right] * \left[\sum_{i=1}^{n} x_i\right]}{n\sum_{i=1}^{n} x_i^2 - \left[\sum_{i=1}^{n} x_i\right]^2} \tag{3}$$

1. **BASIC** Design and code a program that finds the best line of fit for the following data set which represents the distance an object travels in 15 seconds. Plot the line on the same plot as the original data.

| *t*= | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *d*= | 0 | 3.5 | 11.1 | 14.2 | 20.6 | 22.3 | 31.4 | 38.5 | 40.3 | 42.2 | 53.1 | 56.4 | 61.1 | 64.8 | 70.1 | 75.6 |

*created: 2006-09-03*

2. **BASIC** Find a second best line of fit using Matlab's `polyfit()` function. Modify your program in 1 to plot the result on the same plot you created in 1, in a different color and with a dotted line (HINT: use the `hold` command in Matlab).
3. **ADVANCED** Modify your program in 2 to look for the data stored as a file before creating it. If it doesn't exist, create it and store it, and then do the analysis. If it does exist, load it and do the analysis (HINT: use `exist` command in Matlab, within an if…else…end block).
4. **CHALLENGE** Modify your program to include the following function(s) and integrate these function into the program you created in 3:

   a. slope() – calculates slope of a data set
   b. intercept() – calculates y-intercept of a data set
   c. blf – calculates a best line of fit of a data set

   Make the functions generic by passing the data sets, and store them in your library so that they may be reused in other programs.