**Goals:** Developing problem-solving skills, declaring variables, reading data from a file, using loops, and using arrays and functions.

**Problem:** Many engineering and scientific relationships can be manipulated to a format of a straight line, y = mx + b, where m is the slope of the line and b is the y intercept. However, because most measurements have some inherent error, the formula for the line may not be obvious. Therefore, linear regression may be used to determine the best fit of a line to the measured data. Linear regression uses the following two equations for calculations of slope and y-intercept.

# Σx ∗ Σy - n Σ(x \* y) (Σx)2 – n Σ(x)2

**Slope =**

# Σy – slope \* Σx

**y-intercept =**

**n**

where n is the number of (x,y) data sets that were measured. The “goodness of fit” of this straight line can be estimated by summing the square of the residuals. The residual is defined as the difference between the measured y value and the calculated y value (using the calculated slope and y-intercept) for a given x. The equation is shown below where ym is the measured

# “goodness of fit” = Σ(ym - yc)2

y value and yc is the calculated y value. A lower “goodness of fit” value indicates a better match of the line to the data.

For this assignment, you write the C++ source code that uses a set of data to determine the slope and y-intercept of the line the best fits this data. The data is stored in a file called **Proj4Linear.txt**. The x-values are in the left column and the y-values are in the right column. Then you are to determine the “goodness of fit” of this line to the data. The data may be found in a file on ANGEL with the name linear.txt. Your program should output the calculated values of slope and y-intercept; output the x and y values from the file; and the measured y value for x = 2.42. Your program should employ the following:

♦ a function that reads the data file and inputs the data into parallel 1-dimensional arrays for measured x and y values. The left column should be stored in the array for x-values and the right column should be stored in the array for y-values. The two 1dimensional arrays should be **sent back** to the function call in main.

♦ a function that accepts the x and y arrays and calculates the slope and y intercept which are **sent back** to main.

♦ a function accepts the x and y arrays, the slope and y intercept then determines the goodness of fit, and **returns** the goodness of fit to the function call.

♦ a function that searches the original x data for a specific x value and returns the index of that x value so main can output the y value that corresponds to that x value.

.

Your main function should call the function to read the file, call the function to determine the slope and y-intercept, output the formula of the line, call the function to determine the goodness of fit, display the goodness of fit ask the user to input an x value for searching, and output the corresponding y value.

Remember to include comments for the overall program specifying the purpose of the program, input, output and processing. In addition, write comments **immediately before** each function other than main which specifies the purpose of the function, input to the function (what is passed to the function or input within the function), output form the function (what is returned from the function or displayed by the function) and processing needed to complete the task of the function.

**Do not use pointers, vectors or any concepts beyond Chapter 8 of your text book. Remember using material obtained by internet searches constitutes a violation of academic integrity.**