Lab 1: Linear Regression

Stephen Kemp
EE 145L
Section: Tuesday 9:00-11:00am

October 15, 2018

1 Abstract

The purpose of this lab was to review linear regression and the least-squares method, as well as to practice characterizing and analyzing data sets. Three provided data sets were characterized. For a fourth data set, a manually generated best-fit line was compared to a least-squares regression line.

2 Introduction

Linear regression is a technique for analyzing linear relationships between variables in sets of data. Performing linear regression allows one to fit, to a set of data, a linear equation that most closely resembles the data in question. This is useful for extrapolating trends from data, and predicting where additional data will fall.

One method of linear regression is called 'least squares' regression. This is performed by minimizing the sum of the squared errors, where 'error' is defined as the difference between the y-value of an (x,y) data point, and the y-value of the regression equation at the same corresponding x-value. This is the most common linear regression method used, and is available most, if not all, spreadsheet and data analysis software.

The purpose of this lab is to perform several exercises to practice observing and characterizing data, and to compare the accuracy of linear regression with that of a manually-fit best fit line.

3 Materials and Methods

Materials:

No materials were used during this lab

Methods:

No data was collected during this lab, the excises were performed in 'Open Office' spreadsheet software.

4 Results and Analysis

Exercise 0:

1. The plot for this exercise is shown in Figure 1. In this example, the month is the independent variable and the temperature is the dependent variable. By observing the data, it does not appear that there is a linear relationship between the variables. Nor does it appear that there is a way to

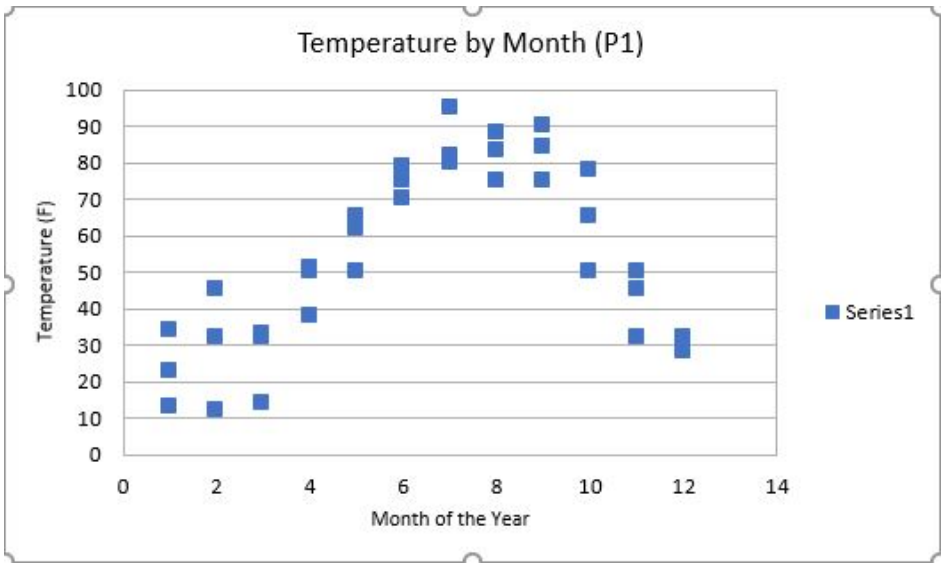

Figure 1: Exercise 0 Part 1 Plot

transform the data to make the relationship linear. There is not a simple mathematical relationship between the month and temperature that can be leveraged to transformed the data.

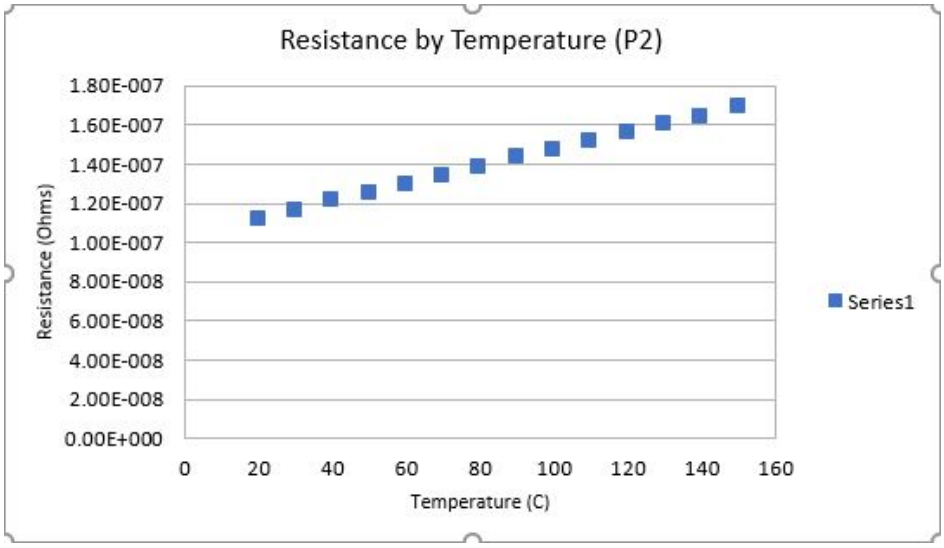

Figure 2: Exercise 0 Part 2 Plot

2. The plot for this exercise is shown in Figure 2. In this exercise, temperature is the independent variable and the wire's resistance is the dependant variable. By observation, the data clearly has a linear relationship without transformation.

3. The raw data plot for this exercise is shown in Figure 3. In this exercise, the energy of the bullet is the dependent variable and velocity is the independent variable. By observation, the relationship between the variables appears to be parabolic. This is supported by the knowledge that $KE = \frac{1}{2}mv^2$. As shown in the second plot, the data can be transformed by taking the square root of the y-data to show a linear relationship.

Exercises 1 and 2:

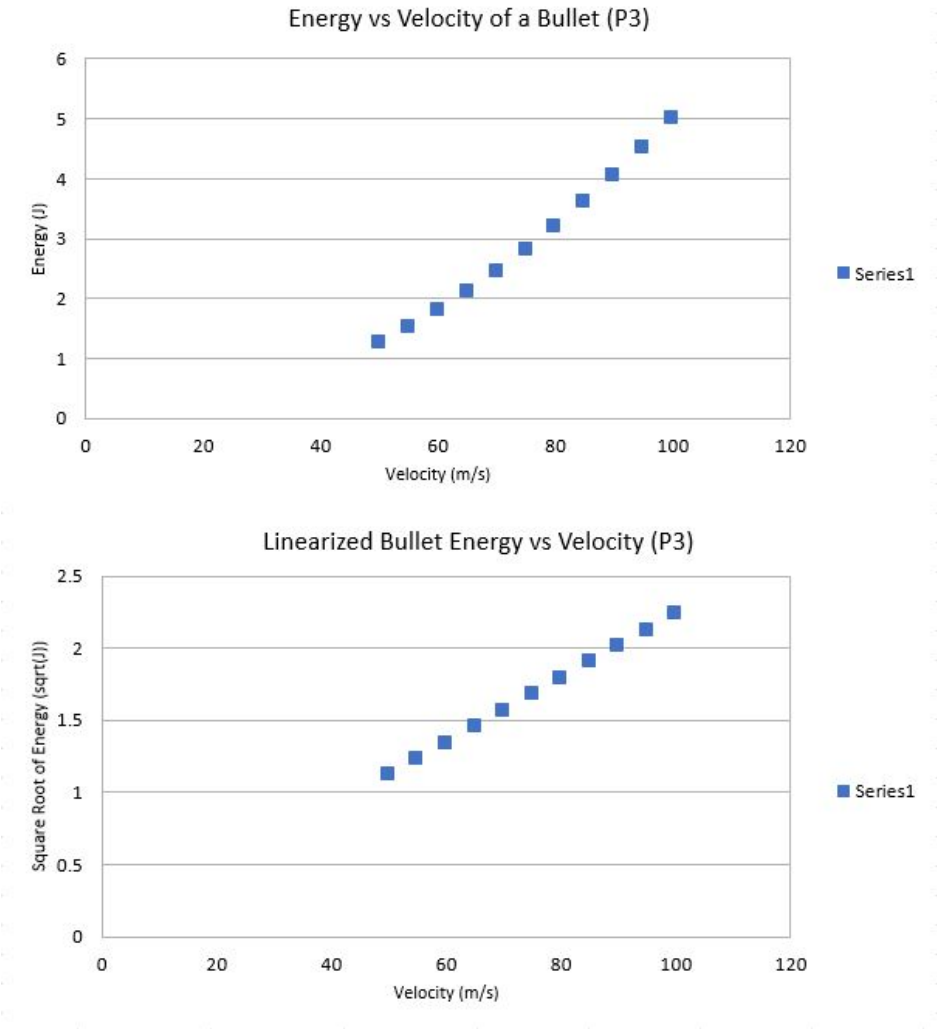

Figure 3: Exercise 0 Part 3 Raw Data and Transformed Plots

The data for Exercises 1 and 2 is shown in Figure 4. The manually fit line was placed in the spreadsheet software using the line-placement tool and visual estimation. The approximate equation for this line is $y = 0.0023x + 1.50$. Using this equation, the specific heat at 75C is estimated as $1.67\frac{J}{mol*C}$). The linear regression equation was generated as $y = 0.00225714x + 1.5107143$. Using the linear regression equation, the specific heat at 75C is estimated as $1.6799998\frac{J}{mol*C}$). The percentage difference between these two values is 0.44%. It is difficult to know which value is more accurate because the chemical in question is unknown and so the true value of it's specific heat at 75C is unknown. The value derived from the linear regression is likely more accurate because the regression equation was generated by minimizing total error from the data. So if the data is to be trusted, then the regression equation should be more accurate than the manually fit line, which was generated using the error-prone method of human visual estimation.

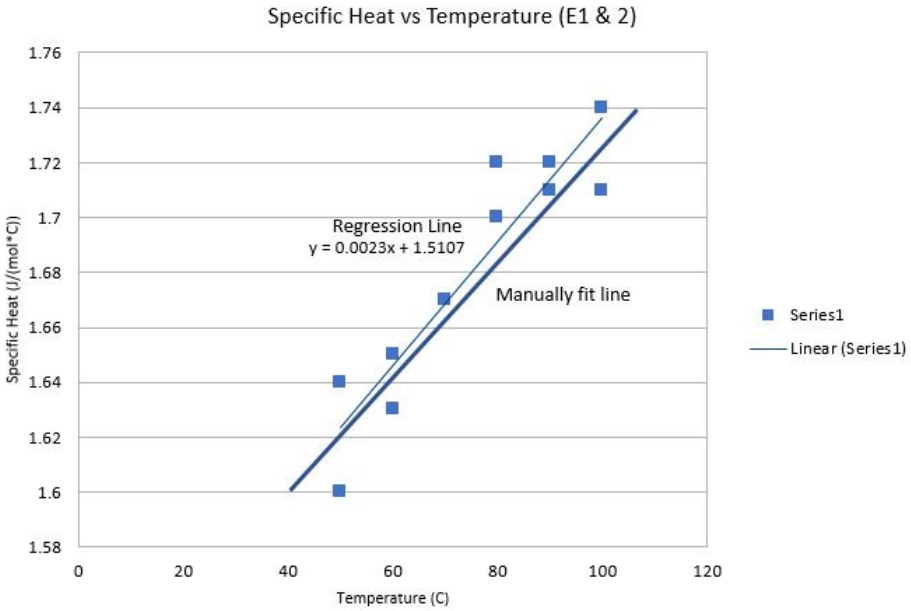

Figure 4: Exercise 1 and 2 Plot

5 Conclusions

The exercises were carried out sufficiently, and everything went as expected. The goals of practicing data characterization and creating and comparing best fit lines were met. The lab briefing offered a good review of how linear regression is performed using the least squares method, as well as why it is performed.