1. Consider the following data set:

x	1	3	5	8
y	2	3	6	7

A proposed model for these data is given by the equation

$$y = 0.75x + 1.25.$$

Find the errors, e_1, e_2, e_3, and e_4, between the y values of each of the points and the proposed model. Also, give the sum of the squares of the errors.

2. Consider the following data set:

x	1	2	5	6
y	2	3	5	7

a. Find the value for the slope, a, and the y-intercept, b of the line

$$y = ax + b,$$

b. Find the actual errors , e_1, e_2, e_3, and e_4, between the y values of each of the points and the proposed model. Give the sum of the square errors.

3. Consider the following data set:

x	6.7	6.9	7.1	7.3	7.5
y	0.9	1.2	1.4	1.2	1.5

a. Find the value for the slope, a, and the y-intercept, b of the line

$$y = ax + b,$$

that best approximates these data.

b. Find the sum of the square errors with the data of the table and the line you have found.

4. A limited set of data is collected and shown in the table below:

t	1	3	5	8
y	4.2	2.9	5.7	5.1

Two researchers interpreted these data differently. Researcher A felt that a good model (A) is given by

$$y = 0.4t + 2.7,$$

while Researcher B thought the biological evidence suggests a better model (B) satisfies the model

$$y = -0.4t + 6.1.$$

a. Which model shows an increasing relationship between the variables? Which one shows a decreasing relationship?

b. Find the sum of squares errors for each of the models. Which Model is better according to the data?

c. Use the formula for finding the least squares best fit line for the data in this problem. Which researcher had the right understanding of how y related to t?

5. A research project on the plankton examines the light intensity filtered by the plankton as a function of the depth of the water. The data are shown in the table below:

depth (m)	1	1.5	2	3	4	5
intensity	0.32	0.29	0.27	0.27	0.15	0.11

a. The least squares best fit to this data set is given by the equation

$$I = -0.0524d + 0.3792,$$

where d is the depth in meters and I is the intensity of light filtered by the plankton. Find the sum of squares error. (You should graph the line and the data on paper to visualize the fit of the data.)

b. On observing the graph of the data, one point seemed obviously erroneous. Give the d value of the point that is most likely erroneous. When this point is removed, then the new least squares best fit model is given by

$$I = -0.0536d + 0.3728.$$

Find the sum of squares error for this model.(Don't forget to compute with the erroneous point removed.)

c. If the model in Part b. is taken to be the actual model, then find the percent error between the slopes of the models in Parts a. and b.

6. The growth of the U.S. population, $P(t)$, over the first part of the twentieth century was almost linear. Below is a table of the approximate populations in millions associated with the years after 1900.

t	0	10	20	30	40	50
P	75.7	91.9	105.9	122.4	131.5	151.4

a. Use the general formula to find the best linear fit model (least squares best fit line) through the data. From the linear model created, compute the sum of square error with the data.

b. Find the population predicted by the model for 1940. Use the value in the Table as the best value to computer the absolute error, relative error, and percent error.