Understanding Regression NAME:

The following table gives the airfare from Baltimore, Maryland to several destinations. We are interested in finding the relationship between distance and airfare. We will graph these points and determine the line of best fit that mimics the pattern of the points. Follow the instructions below. Then answer the questions that follow.

Destination	**Distance (miles)**	**Airfare ($)**
Atlanta	576	178
Boston	370	138
Chicago	612	94
Dallas	1216	278
Detroit	409	158
Denver	1502	258
Miami	946	198
New Orleans	998	188
New York	189	98
Orlando	787	179
Pittsburgh	210	138
St. Louis	737	98

1. Use Minitab to draw the scatter plot and figure the line of best fit for this data.
a.) Put the distances in **C1** and the airfares in **C2**. Label the columns appropriately in the topmost, unnumbered row. You do not need to enter the city names.

b.) Let's draw a scatter plot of the data first. Minitab will draw this scatterplot with the line of best fit drawn in. We will later find the equation of this line.

Select **Graph**, and then **Scatterplot** from the top menu. Select **With Regression** from this dialog box and click **OK**.

Click in the first row under **Y variables** if your cursor is not already there. Then double-click on **C2** to assign C2 to the Y variable.

Then click under the **X variables** and double-click **C1** to assign it there. Click **OK** to make it graph.

Notice how nicely the line mimics the pattern of points. Now, let's find the equation of this line.

b.) Choose **Stat** from the top menu. Select **Regression** from this pull-down menu. Choose **Regression** again.

The dialog box that opens up asks you for the **Response** and **Predictors** variables. We want an equation that outputs the airfare if we input the distance. So the airfares will be the **Response** (output) variable and the distances will serve as the **Predictors** (input) variable. The response and predictor variables are the same as the dependent (y) and independent variables (x) you may remember from algebra. Once you assign the Response and Predictors variables, click **OK** to make it calculate the regression line.

You'll probably need to scroll up the **Session** window to find the regression equation. Write down the regression equation here.

c.) Find **R-Sq** somewhat below the equation. This is the square of the correlation. Find the square root of this (use the decimal form) to figure r, the correlation coefficient. Round it to two decimal places.

d.) Using your value of r, discuss the strength and direction of the linear correlation between distance and airfare.

e.) Bloomington/ Normal is 758 miles from Baltimore, MD. Use your regression equation to estimate the price of a plane ticket from Baltimore to Bloomington. Round your answer to a whole number and label it as dollars.

2. Look at your scatter plot. There are a couple of points that do not really fit the overall pattern of points.
a.) Which two destinations do you think I am talking about?

b.) Delete the data for Chicago and St. Louis, the two outliers. Simply highlight them in the Data window and press **Delete**. Now you should have ten data points. Regraph the scatterplot and refigure the regression line for these ten points. What is the regression equation? What is the value of *r*? How does removing the outliers affect the strength and direction of your regression line?