**Linear Regression Practice on Marathon Data**

Starters-

· Make sure the Data Analysis ToolPack is turned on in Excel.

· Open the document containing the data from the Google Classroom.

· Look at the 2010 Boston Marathon Data Tab.

For this project you will be performing linear regression on three sets of runners’ split data for the Boston Marathon. For each set you will answer/complete the following below. The splits data you will be performing the regression on is:

a. 5k split and 10k split

b. 5k split and half marathon

c. Half marathon split and total marathon time

1. Create a scatterplot between each split data above. Be sure to include the following in each scatterplot.

· Plot title, axes labeled

· Tic marks appropriately labeled and spaced

· Markers appropriate size (not too large/small)

· Define x and y variables

2. Based on the scatterplot, describe the relationship between *x* and *y*. Remember to describe strength, form, direction, outliers and influential points.

3. Using the Data Analysis ToolPak in excel, calculate the Least-Squares Regression Line (LSRL). Be sure to include the following:

· Variable definitions

· Interpretations of slope and intercept in context.

· Equation for LSRL, using the context of the data set.

4. Using the results from the regression in *step 3*, find and interpret (in context) the correlation coefficient and coefficient of determination. Be sure to comment if the correlation coefficient match your description of the scatterplot.

5. Determine if the LSRL is a good model for the data. Be sure to calculate the residuals (include the equation for finding residuals-in context), make and interpret the residual plot.