4. (20 points) For each of the datasets below, fit at least two models each by transforming the data  
appropriately. Graph the residuals and analyze how well the model fits the data, indicating what you  
believe to be the best model, with explanation.  
x 1 2 3 4 5 6  
y 16.3 23.1 37.4 46.9 58.7 91.0  
x 1 2 3 4 5 6  
y 1.34 1.61 -0.98 -3.80 -4.55 -2.30

A graph on a screen

Description automatically generated

The points seem to be sinusoidal, lets look at the model

A screenshot of a computer

Description automatically generated

Y = -1.1286x + 2.5033

With an r squared value of 0.6709 and a pretty low p value which suggests that the linear model does represent somewhat of the relationship between x and y.

Let’s find the new x values after a sin(x) transformation has been applied and see how the model improves

A graph with a red line

Description automatically generated

You can see that the relationship between the transformed x values and the y values are very close. Looking at the linear model you can see the following.

A white sheet with numbers and lines

Description automatically generated

You can see how high of an R^2 value and how high of an F-statistic value there is. This represents a very close to linear fit. You can also observe how small the residual values are which suggests that the linear model is very close to the data that was observed following the sinusoidal transformation.

A graph with a red line

Description automatically generated

You can see that there is no clear pattern with this transformation. That suggests that the transformation did not do as well as a job as applying the sinusoidal transformation to the x values.

A screenshot of a computer program

Description automatically generated

From the linear model, the performance metrics do do as well, with a smaller f statistic and a lower r squared value. Therefore the best model is going to be the first one where the sinusoidal transformation to the x values was done.

A graph with a red line

Description automatically generated

The graph shows a graph with an increasing trend which gradient increases at a faster rate. Let’s observe the model that is associated with this set of data.

A screenshot of a computer

Description automatically generated

There is a high r squared value with a high f statistic which suggests a pretty strong correlation between the change in values of x resulting in the change in y values.

The first transformation is going to be taking the log values of response, y values.

A graph with a red line

Description automatically generated

You can observe that the points on this line following a linear trend much closer than before. This suggests that the values of y were exponential. Let’s review the linear model, post transformation.

A screenshot of a computer code

Description automatically generated

Look at how small the residual values are, this suggests that the values only differ by a small amount from the linear model that is fit. The resulting R^2 value is very high at almost 0.99 which suggests that the trend almost fits exactly on a linear trend. This is likely going to be the best model for the data that is provided.

The next transformation is going to be a modification to the y values where you take the reciprocal of those values. This can be seen in the graph below.

A graph with a red line

Description automatically generated

For most points it follows a pretty strong trend downwards however it does not seem to make a difference in the correlation between x and y with the transformation of 1/y to the y values. Let’s observe the linear model evaluation.

A white sheet with black numbers and symbols

Description automatically generated with medium confidence

While the R^2 is high, it is not an improvement on the previous model. This means that the model that you want to choose is taking the log(y) of the original y values to best find a trend line.