

# Regression Analysis MATLAB GUI Assignment

This document contains exercises to be performed with the Regression Analysis MATLAB GUI.

## DISPLACEMENT RECORD DATA SET

1. Using the “Select Data Set” drop down menu at the right side of the window, select “Displacement Record” to be the active data set.
2. Plot a first-order regression line through all of the data points. Look at the equation of the line, and the  $R^2$  value. Is this line a good fit to this data? Given the overall appearance of the data set, does a first-order regression seem logical?
3. Fit second-, third-, and fourth-order regression lines through the data. How much does the  $R^2$  improve over the result for the first-order regression line? In this case, does it make sense to use a higher-order regression line?

## STRAIN DATA SET

This is calibration data for a strain gage. It is expected that the output from the gage will be linear over a certain range, and outside of that range it is non-linear. When actually taking measurements, the gage would be used only within its linear range.

1. Select “Strain” to be the active data set.
2. Fit a first-order regression line through all of the data points. Given what you know about the data, as explained above, does this make sense?
3. Try fitting a higher-order regression line through all of the points. How much does the  $R^2$  value improve? If the gage will only be used in its linear range, is this reasonable?
4. The final data points collected are clearly outside of the linear range of the gage. Return to a first-order regression line, and eliminate points (by clicking on the radio buttons at the left of the window) until you are only left with points that appear to be within the linear range. Now what is the  $R^2$  value? Is this a good regression line?

## SHOCK SPRING DATA SET

1. Select “Shock Spring” to be the active data set.
2. Fit a first-order regression line through only the middle three data points ( $x = 8.00, 11.00,$  and  $14.00$ ). Look at the  $R^2$  value. In general, it is virtually impossible to have an  $R^2$  of 1 with real data—in this case, is this correct?
3. Plot first- through fourth-order regression lines using all of the data points. Which regression line seems to best fit the trend of the data?
4. Plot a third-order regression line using all of the data points. Then plot the y-variance of the data points (see the Regression Analysis Tutorial for more information on calculating the standard deviation of data points). Are there any outliers? If so, eliminate them from the regression calculation, and see how this changes the overall variance of the data. If you were actually using this regression line for further calculations, would it make sense to eliminate the outlying point?
5. Fit a second-order regression line through roughly the first half of the data points. Then fit a second-order regression line using only the second half of the data points. If the spring were only going to be used over one of these ranges, would the second-order or third-order regression line be better?