**Linear Regression Lab  
Intro to Data Science**

In this lab, you will use Knime to find a regression line for predicting **petal length** in the Iris dataset.

In our previous examples, the one feature to use to predict another feature was obvious because there were only two numerical features within the dataset. In this data, there are **three possible features** to use as predictors of the petal length (petal width, sepal length, and sepal width).

Although it is possible to use multiple features for predictors in linear regression, it turns out **that only one of these three features has a linear relationship with petal length**.

Use a Scatter Matrix node to examine the relationships and choose ***the one feature*** that would be best to use as the predictor. Then find the regression line using KNIME.

Paste a screenshot of your KNIME work flow here:

Paste a screenshot of the “Coefficients and Statistics” window here:  
 

Paste a screenshot of the window that shows the **correlation coefficient, r,** here:

What is the equation of your regression line? Since you will be using only one feature as the predictor, there will be only ***one regression line***. Specify what features the x (or x1 , x2 etc.) and y values represent.

**Using your equation**, show how to find a prediction for the petal length of the following flower.

|  |  |  |  |
| --- | --- | --- | --- |
| sepal\_length | sepal\_width |  | petal\_width |
| 5.3 | 3.2 |  | 0.4 |

Since you are only using one predictor, only **one of the values** above will be used.