Tristan Shah

STAT 360

Due: 11/20/19

*Domestic vs International Fuel Cost*

\*This project was done in python.

Source code: https://github.com/gladisor/STAT360\_Project

1.

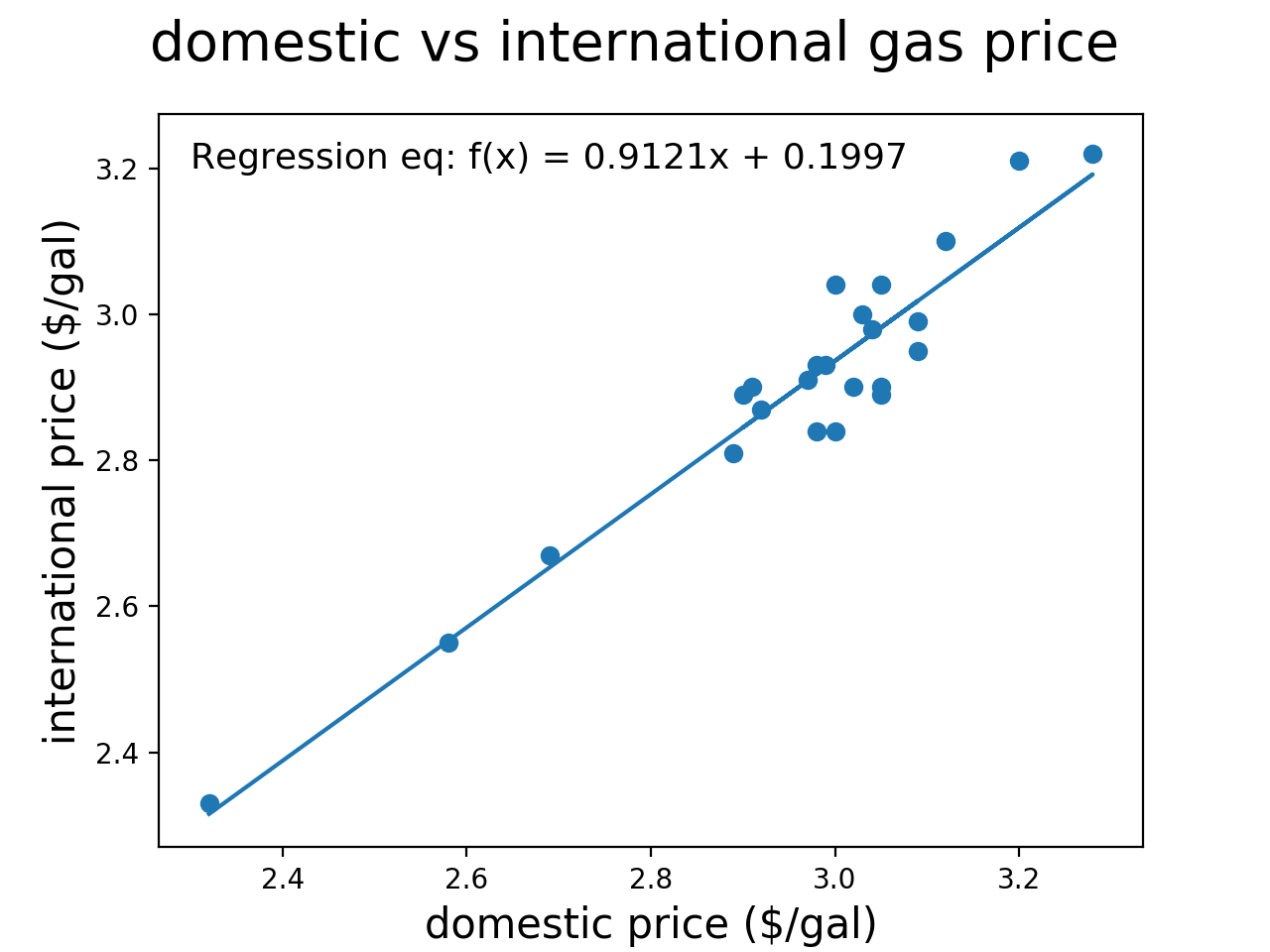
Explanatory Variable:

The explanatory variable is domestic price of fuel. It was plotted on the x axis.

Response Variable:

The response variable is international price of fuel. It was plotted on the y axis.

2a. Scatter plot & Regression line:



Regression equation: f(x) = 0.9121x + 0.1997

2b. Correlation coefficient r: 0.95524514

2c.

Form:

The form of the relationship appears to be linear. For this project I used the NumPy function polyfit(). By restricting it to a first degree polynomial the regression line will always be f(x) = mx + b (linear). However If I chose to use more than a first degree polynomial the model would fit the data more accurately but may not be as good for extrapolating new data.

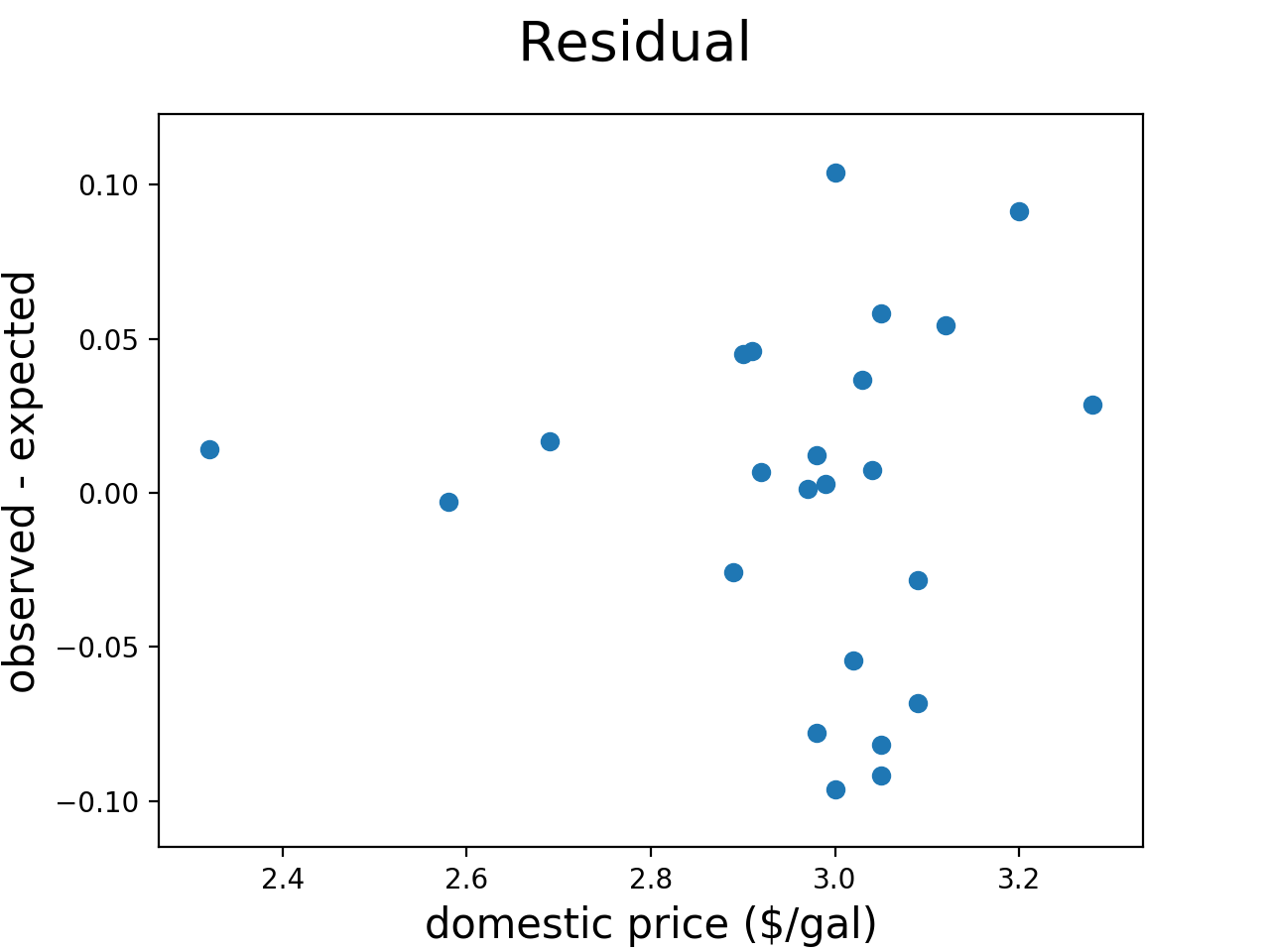
Strength:

Strength of the relationship can be described using the correlation coefficient. For this relationship the correlation coefficient was 0.95524514. A value of r between 0.7 and 1 is considered strong. Therefore the relationship between domestic fuel and international fuel is strong.

Direction:

The direction of this relationship is positive. As the price of domestic fuel increases so does the price of international fuel. This relationship is not inverse (which would yield a negative value of r).

3a. Residual Plot:



3b. Predict the international carrier fuel cost if the domestic carrier fuel cost is $2.90/gallon.:

In order to predict the value of the response variable evaluated at an input of 2.90 all we need to do is to plug that value into our regression equation:

According to our model the expected price of international fuel given a price of $2.90 for domestic fuel is $2.84479.

3c. What do you call this prediction, interpolation or extrapolation?:

This prediction is an interpolation because the value is within the data range. If the value was outside of the data range then the prediction would be an extrapolation.

3d. How reliable is this prediction:

Since the correlation coefficient r is greater than 0.7 the prediction is very reliable.

4. Descriptive Statistics:

|  |  |  |
| --- | --- | --- |
|  | Domestic | International |
| Average ($/gal) | 2.9645833 | 2.9037500 |
| Max ($/gal) | 3.28 | 3.22 |
| Min ($/gal) | 2.32 | 2.33 |