

## stat\_4

March 19, 2023

```
[ ]: # import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from sklearn.linear_model import LinearRegression
```

```
[ ]: df = pd.read_csv ('cars_clean.csv')
```

```
[ ]: df.shape
```

```
[ ]: # create the linear regression object
lm = LinearRegression()
```

Multiple Linear Regression

```
[ ]: # predictor variables
Z = df [['engine-size', 'highway-mpg']]
```

```
[ ]: # create the model
lm.fit (Z, df['price'])
```

```
[ ]: # get the intercept
print (lm.intercept_)
```

```
[ ]: # value of the coefficients
print (lm.coef_)
```

```
[ ]: # Distribution plot to visually determine the goodness of fit
Y_hat = lm.predict(Z)
```

```
[ ]: # distribution plot
ax1 = sns.distplot (df['price'], hist = False, color = 'r', label = 'Actual_
↪Value')
sns.distplot (Y_hat, hist = False, color = 'b', label = 'Calculated Value', ax =
↪ax1)
```

```
plt.title ('Actual vs Calculated Values for Price')  
plt.xlabel ('Price')  
plt.ylabel ('Proportion of Cars')  
plt.show()
```