In [183]:
```
pip install statsmodels
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: statsmodels in c:\programdata\anaconda3\lib\site-packages (0.13.2)
Requirement already satisfied: numpy>=1.17 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (1.21.5)
Requirement already satisfied: scipy>=1.3 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (1.7.3)
Requirement already satisfied: pandas>=0.25 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (1.4.2)
Requirement already satisfied: patsy>=0.5.2 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (0.5.2)
Requirement already satisfied: packaging>=21.3 in c:\programdata\anaconda3\lib\site-packages (from statsmodels) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\programdata\anaconda3\lib\site-packages (from packaging>=21.3->statsmodels) (3.0.4)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->statsmodels) (2021.3)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->statsmodels) (2.8.2)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-packages (from patsy>=0.5.2->statsmodels) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [184]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
```

In [185]:
```
rm=pd.read_csv('CarsDataset.csv')
rm.head()
```

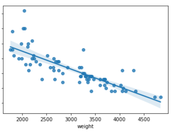
Out[185]:

| | make | price | mpg | rep78 | headroom | trunk | weight | length | turn | displacement | gear_ratio | foreign |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | AMC Concord | 4099 | 22 | 3.0 | 2.5 | 11 | 2930 | 186 | 40 | 121 | 3.58 | Domestic |
| **1** | AMC Pacer | 4749 | 17 | 3.0 | 3.0 | 11 | 3350 | 173 | 40 | 258 | 2.53 | Domestic |
| **2** | AMC Spirit | 3799 | 22 | NaN | 3.0 | 12 | 2640 | 168 | 35 | 121 | 3.08 | Domestic |
| **3** | Buick Century | 4816 | 20 | 3.0 | 4.5 | 16 | 3250 | 196 | 40 | 196 | 2.93 | Domestic |
| **4** | Buick Electra | 7827 | 15 | 4.0 | 4.0 | 20 | 4080 | 222 | 43 | 350 | 2.41 | Domestic |

In [186]:
```
import warnings
warnings.filterwarnings('ignore')
import matplotlib.pyplot as plt
import seaborn as sns
```

In [187]:
```
sns.regplot('weight', 'mpg', data=rm) ##Regression Line and Scatter Plot
```

Out[187]:
```
<AxesSubplot:xlabel='weight', ylabel='mpg'>
```




In [188]:
```
df = rm[["weight", "mpg"]]
df.head()
```

Out[188]:

| | weight | mpg |
|---|---|---|
| **0** | 2930 | 22 |
| **1** | 3350 | 17 |
| **2** | 2640 | 22 |
| **3** | 3250 | 20 |
| **4** | 4080 | 15 |

In [189]:
```
X = df.loc[:, 'weight']
Y = df['mpg']
```

In [190]:
```
X = sm. add_constant(X)       # adding a constant
model = sm.OLS(Y, X).fit()
```

In [191]:
```
model.summary()     ## Model for 1 Dependent Variable('mpg') and 1 Independent Variable('weight')
```

Out[191]:

OLS Regression Results

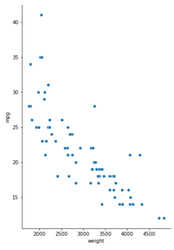
| | | | |
|---|---|---|---|
| **Dep. Variable:** | mpg | **R-squared:** | 0.652 |
| **Model:** | OLS | **Adj. R-squared:** | 0.647 |
| **Method:** | Least Squares | **F-statistic:** | 134.6 |
| **Date:** | Tue, 14 Feb 2023 | **Prob (F-statistic):** | 3.80e-18 |
| **Time:** | 01:55:31 | **Log-Likelihood:** | -195.39 |
| **No. Observations:** | 74 | **AIC:** | 394.8 |
| **Df Residuals:** | 72 | **BIC:** | 399.4 |
| **Df Model:** | 1 | | |
| **Covariance Type:** | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 39.4403 | 1.614 | 24.436 | 0.000 | 36.223 | 42.658 |
| **weight** | -0.0060 | 0.001 | -11.603 | 0.000 | -0.007 | -0.005 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 27.415 | **Durbin-Watson:** | 2.344 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 52.613 |
| **Skew:** | 1.318 | **Prob(JB):** | 3.76e-12 |
| **Kurtosis:** | 6.180 | **Cond. No.** | 1.26e+04 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.26e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

In [192]:
```
p = sns.pairplot(df, x_vars='weight', y_vars='mpg', size=7, aspect=0.7) #Assumption of Linear Regression
```




In [193]:
```
from statsmodels.stats.diagnostic import het_white
from statsmodels.compat import lzip
```

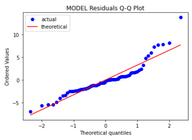
In [194]:
```
#The test for homoscedastic
white_test = het_white(model.resid,  model.model.exog)
#lzipping the array with labels
names = ['lagrange multiplier statistic', 'p-value','f-value', 'f p-value']
lzip(names,white_test)
```

Out[194]:
```
[('lagrange multiplier statistic', 5.889547898475892),
 ('p-value', 0.052613951896037105),
 ('f-value', 3.069704339713459),
 ('f p-value', 0.053264565488889312)]
```

In [195]:
```
## Since it is not having a p-value ≤ 0.05 this indicates that this is not 'Heteroscedasticity'.
```

In [196]:
```
from scipy import stats
```

In [197]:
```
stats.probplot(model.resid, dist="norm", plot= plt)  #Normality errors check using Q-Q plots
plt.title("MODEL Residuals Q-Q Plot")
plt.legend(['actual','theoretical'])
plt.show()
```




In [198]:
```
dg = rm[['weight','displacement','mpg']]
dg.head()
```

Out[198]:

| | weight | displacement | mpg |
|---|---|---|---|
| **0** | 2930 | 121 | 22 |
| **1** | 3350 | 258 | 17 |
| **2** | 2640 | 121 | 22 |
| **3** | 3250 | 196 | 20 |
| **4** | 4080 | 350 | 15 |

In [199]:
```
##result1 = sm.ols(formula = "weight ~ mpg", data = df).fit()
##result2 = sm.ols(formula = "displacement ~ mpg", data = dg).fit()
```

In [200]:
```
##result1.summary()
##result2.summary()
```

In [201]:
```
X = dg.loc[:, 'weight':'displacement']
Y = dg['mpg']
```

In [202]:
```
X = sm. add_constant(X)       # adding a constant
model1 = sm.OLS(Y, X).fit()
```

In [203]:
```
model1.summary() ##Model for 2 Independent Variable and 1 Dependent Variable
```

Out[203]:

OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | mpg | **R-squared:** | 0.653 |
| **Model:** | OLS | **Adj. R-squared:** | 0.643 |
| **Method:** | Least Squares | **F-statistic:** | 66.79 |
| **Date:** | Tue, 14 Feb 2023 | **Prob (F-statistic):** | 4.84e-17 |
| **Time:** | 01:55:32 | **Log-Likelihood:** | -195.24 |
| **No. Observations:** | 74 | **AIC:** | 396.5 |
| **Df Residuals:** | 71 | **BIC:** | 403.4 |
| **Df Model:** | 2 | | |
| **Covariance Type:** | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 40.0845 | 2.020 | 19.843 | 0.000 | 36.057 | 44.113 |
| **weight** | -0.0066 | 0.001 | -5.631 | 0.000 | -0.009 | -0.004 |
| **displacement** | 0.0053 | 0.010 | 0.535 | 0.594 | -0.014 | 0.025 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 28.461 | **Durbin-Watson:** | 2.376 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 55.685 |
| **Skew:** | 1.362 | **Prob(JB):** | 8.10e-13 |
| **Kurtosis:** | 6.262 | **Cond. No.** | 1.57e+04 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.57e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

In [204]:
```
plt.figure(figsize=(10,5))        ##Check for multicollinearity among independent variables
sns.heatmap(dg[['weight','displacement']].corr(),vmin=-1,annot= True)
```

Out[204]:
```
<AxesSubplot:>
```


In [205]:
```
white_test = het_white(model1.resid,  model.model.exog)      #White test
#lzipping the array with labels
names = ['lagrange multiplier statistic', 'p-value','f-value', 'f p-value']
lzip(names,white_test)
```

Out[205]:
```
[('lagrange multiplier statistic', 5.815779915059958),
 ('p-value', 0.054590797328719811),
 ('f-value', 3.027976073165202),
 ('f p-value', 0.054708094983976444)]
```