ExercisesAutoMulti

February 21, 2025

1 Multilinear Regression: Auto dataset

1.1 Import notebook functions

[1]:
```
from notebookfuncs import *
```

1.2 Import standard libraries

[2]:
```
import numpy as np
import pandas as pd

pd.set_option("display.max_rows", 1000)
pd.set_option("display.max_columns", 1000)
pd.set_option("display.width", 1000)
pd.set_option("display.max.colwidth", None)
import matplotlib.pyplot as plt
from matplotlib.pyplot import subplots
import seaborn as sns
import itertools
```

1.3 New imports

[3]:
```
import statsmodels.api as sm
```

1.4 Import statsmodels.objects

[4]:
```
from statsmodels.stats.outliers_influence import variance_inflation_factor as␣
↪VIF
from statsmodels.stats.outliers_influence import summary_table
from statsmodels.stats.anova import anova_lm
import statsmodels.formula.api as smf
```

1.5 Import ISLP objects

[5]:
```
import ISLP
from ISLP import models
from ISLP import load_data
```

```python
from ISLP.models import ModelSpec as MS, summarize, poly
```

Import user functions

[6]:
```python
from userfuncs import *
```

Set level of significance (alpha)

[7]:
```python
LOS_Alpha = 0.01
```

[7]: 0.01

[8]:
```python
Auto = load_data("Auto")
Auto = Auto.sort_values(by=["year"], ascending=True)
Auto.head()
Auto.columns
```

[8]:
```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
'acceleration', 'year', 'origin'], dtype='object')
```

[9]:
```python
Auto.shape
```

[9]: (392, 8)

[10]:
```python
Auto.describe()
```

[10]:
```
              mpg   cylinders  displacement  horsepower       weight  
acceleration        year      origin  
count  392.000000  392.000000    392.000000  392.000000   392.000000  
392.000000  392.000000  392.000000  
mean    23.445918    5.471939    194.411990  104.469388  2977.584184  
15.541327   75.979592    1.576531  
std      7.805007    1.705783    104.644004   38.491160   849.402560  
2.758864    3.683737    0.805518  
min      9.000000    3.000000     68.000000   46.000000  1613.000000  
8.000000   70.000000    1.000000  
25%     17.000000    4.000000    105.000000   75.000000  2225.250000  
13.775000   73.000000    1.000000  
50%     22.750000    4.000000    151.000000   93.500000  2803.500000  
15.500000   76.000000    1.000000  
75%     29.000000    8.000000    275.750000  126.000000  3614.750000  
17.025000   79.000000    2.000000  
max     46.600000    8.000000    455.000000  230.000000  5140.000000  
24.800000   82.000000    3.000000  
```

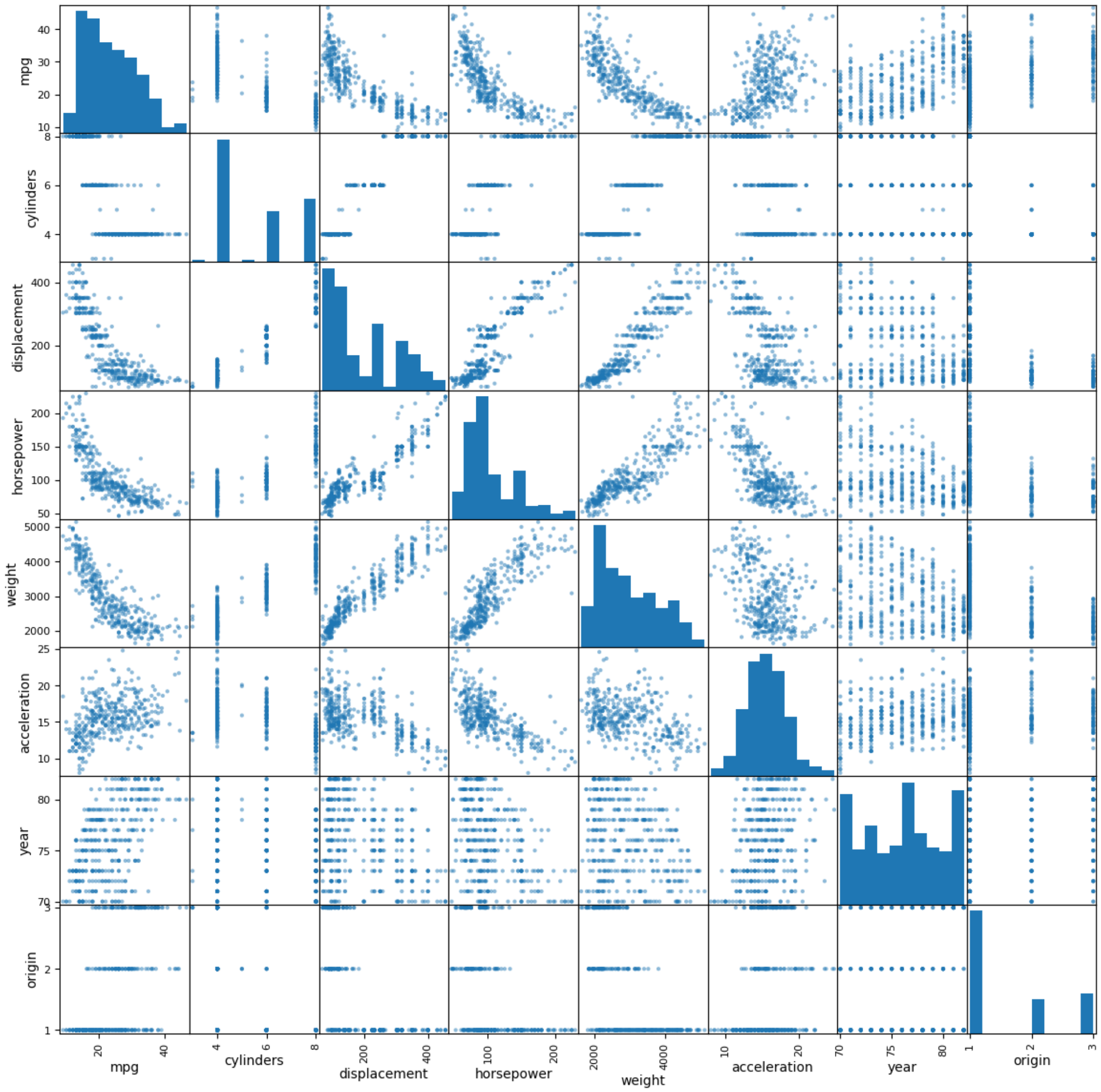
1.6 9. This question involves the use of multiple linear regression on the Auto data set.

1.6.1 (a) Produce a scatterplot matrix which includes all of the variables in the data set.

[11]:
```
pd.plotting.scatter_matrix(Auto, figsize=(14, 14))
```

[11]:
```
array([[<Axes: xlabel='mpg', ylabel='mpg'>,
        <Axes: xlabel='cylinders', ylabel='mpg'>,
        <Axes: xlabel='displacement', ylabel='mpg'>,
        <Axes: xlabel='horsepower', ylabel='mpg'>,
        <Axes: xlabel='weight', ylabel='mpg'>,
        <Axes: xlabel='acceleration', ylabel='mpg'>,
        <Axes: xlabel='year', ylabel='mpg'>,
        <Axes: xlabel='origin', ylabel='mpg'>],
       [<Axes: xlabel='mpg', ylabel='cylinders'>,
        <Axes: xlabel='cylinders', ylabel='cylinders'>,
        <Axes: xlabel='displacement', ylabel='cylinders'>,
        <Axes: xlabel='horsepower', ylabel='cylinders'>,
        <Axes: xlabel='weight', ylabel='cylinders'>,
        <Axes: xlabel='acceleration', ylabel='cylinders'>,
        <Axes: xlabel='year', ylabel='cylinders'>,
        <Axes: xlabel='origin', ylabel='cylinders'>],
       [<Axes: xlabel='mpg', ylabel='displacement'>,
        <Axes: xlabel='cylinders', ylabel='displacement'>,
        <Axes: xlabel='displacement', ylabel='displacement'>,
        <Axes: xlabel='horsepower', ylabel='displacement'>,
        <Axes: xlabel='weight', ylabel='displacement'>,
        <Axes: xlabel='acceleration', ylabel='displacement'>,
        <Axes: xlabel='year', ylabel='displacement'>,
        <Axes: xlabel='origin', ylabel='displacement'>],
       [<Axes: xlabel='mpg', ylabel='horsepower'>,
        <Axes: xlabel='cylinders', ylabel='horsepower'>,
        <Axes: xlabel='displacement', ylabel='horsepower'>,
        <Axes: xlabel='horsepower', ylabel='horsepower'>,
        <Axes: xlabel='weight', ylabel='horsepower'>,
        <Axes: xlabel='acceleration', ylabel='horsepower'>,
        <Axes: xlabel='year', ylabel='horsepower'>,
        <Axes: xlabel='origin', ylabel='horsepower'>],
       [<Axes: xlabel='mpg', ylabel='weight'>,
        <Axes: xlabel='cylinders', ylabel='weight'>,
        <Axes: xlabel='displacement', ylabel='weight'>,
        <Axes: xlabel='horsepower', ylabel='weight'>,
        <Axes: xlabel='weight', ylabel='weight'>,
        <Axes: xlabel='acceleration', ylabel='weight'>,
        <Axes: xlabel='year', ylabel='weight'>,
        <Axes: xlabel='origin', ylabel='weight'>],
```

```
[<Axes: xlabel='mpg', ylabel='acceleration'>,
 <Axes: xlabel='cylinders', ylabel='acceleration'>,
 <Axes: xlabel='displacement', ylabel='acceleration'>,
 <Axes: xlabel='horsepower', ylabel='acceleration'>,
 <Axes: xlabel='weight', ylabel='acceleration'>,
 <Axes: xlabel='acceleration', ylabel='acceleration'>,
 <Axes: xlabel='year', ylabel='acceleration'>,
 <Axes: xlabel='origin', ylabel='acceleration'>],
[<Axes: xlabel='mpg', ylabel='year'>,
 <Axes: xlabel='cylinders', ylabel='year'>,
 <Axes: xlabel='displacement', ylabel='year'>,
 <Axes: xlabel='horsepower', ylabel='year'>,
 <Axes: xlabel='weight', ylabel='year'>,
 <Axes: xlabel='acceleration', ylabel='year'>,
 <Axes: xlabel='year', ylabel='year'>,
 <Axes: xlabel='origin', ylabel='year'>],
[<Axes: xlabel='mpg', ylabel='origin'>,
 <Axes: xlabel='cylinders', ylabel='origin'>,
 <Axes: xlabel='displacement', ylabel='origin'>,
 <Axes: xlabel='horsepower', ylabel='origin'>,
 <Axes: xlabel='weight', ylabel='origin'>,
 <Axes: xlabel='acceleration', ylabel='origin'>,
 <Axes: xlabel='year', ylabel='origin'>,
 <Axes: xlabel='origin', ylabel='origin'>]], dtype=object)
```



1.6.2 (b) Compute the matrix of correlations between the variables using the DataFrame.corr() method.

[12]:
```
Auto.corr()
```

[12]:
```
                   mpg  cylinders  displacement  horsepower    weight
acceleration      year    origin
mpg           1.000000  -0.777618     -0.805127   -0.778427 -0.832244
0.423329  0.580541  0.565209
cylinders    -0.777618   1.000000      0.950823    0.842983  0.897527
-0.504683 -0.345647 -0.568932
displacement -0.805127   0.950823      1.000000    0.897257  0.932994
-0.543800 -0.369855 -0.614535
```

```
horsepower   -0.778427   0.842983      0.897257    1.000000  0.864538
-0.689196 -0.416361 -0.455171
weight       -0.832244   0.897527      0.932994    0.864538  1.000000
-0.416839 -0.309120 -0.585005
acceleration  0.423329  -0.504683     -0.543800   -0.689196 -0.416839
1.000000  0.290316  0.212746
year          0.580541  -0.345647     -0.369855   -0.416361 -0.309120
0.290316  1.000000  0.181528
origin        0.565209  -0.568932     -0.614535   -0.455171 -0.585005
0.212746  0.181528  1.000000
```

1.6.3 (c) Use the sm.OLS() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summarize() function to print the results. Comment on the output. For instance:

1.7 Convert year and origin columns to categorical types

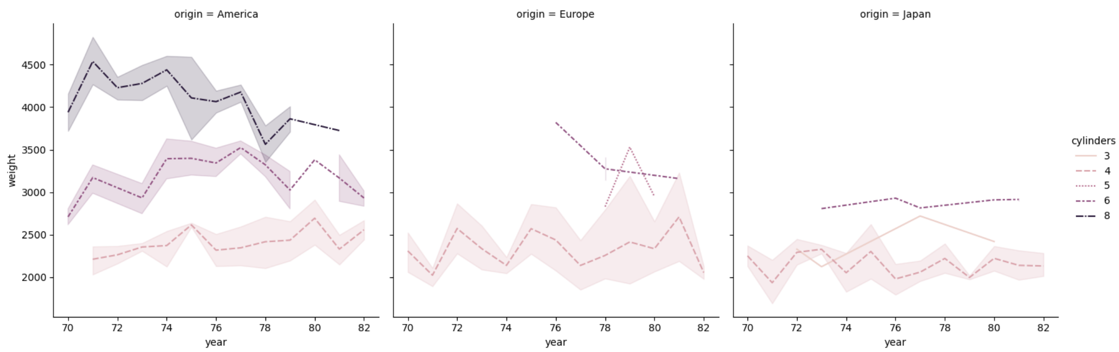
[13]:
```
Auto["origin"] = Auto["origin"].astype("category")
Auto["origin"] = Auto["origin"].cat.rename_categories(
    {1: "America", 2: "Europe", 3: "Japan"}
)
Auto["year"] = Auto["year"].astype("category")
Auto.describe()
```

[13]:
```
              mpg   cylinders  displacement  horsepower       weight
acceleration
count  392.000000  392.000000    392.000000  392.000000   392.000000
392.000000
mean    23.445918    5.471939    194.411990  104.469388  2977.584184
15.541327
std      7.805007    1.705783    104.644004   38.491160   849.402560
2.758864
min      9.000000    3.000000     68.000000   46.000000  1613.000000
8.000000
25%     17.000000    4.000000    105.000000   75.000000  2225.250000
13.775000
50%     22.750000    4.000000    151.000000   93.500000  2803.500000
15.500000
75%     29.000000    8.000000    275.750000  126.000000  3614.750000
17.025000
max     46.600000    8.000000    455.000000  230.000000  5140.000000
24.800000
```

[14]:
```
sns.relplot(
    Auto,
    x="year",
    y="weight",
```

```
    col="origin",
    hue="cylinders",
    style="cylinders",
    estimator="mean",
    kind="line",
)
```

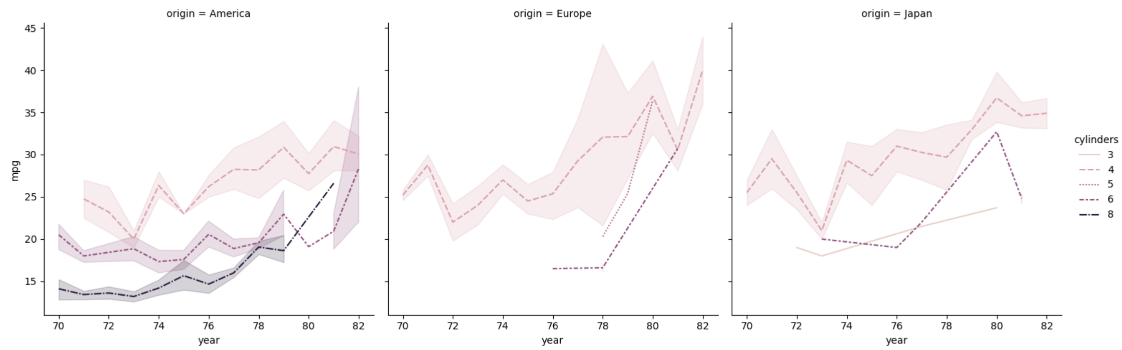
[14]: <seaborn.axisgrid.FacetGrid at 0x74027bb74cb0>



The weight of the 8-cylinder American made models show a decline from the highs of 1972. It can also be seen that American made cars are heavier than their European and Japanese counterparts especially in the most common models with 4 cylinders.

[15]:
```
sns.relplot(
    Auto,
    x="year",
    y="mpg",
    col="origin",
    hue="cylinders",
    style="cylinders",
    estimator="mean",
    kind="line",
)
```

[15]: <seaborn.axisgrid.FacetGrid at 0x74027e14ef60>



It can be seen that after the oil shock of 1973 and the regulations and actions taken by the US government, the mileage for American made cars rose across all models. This was, however, matched by the European and Japanese models which were already lighter and more fuel efficient.

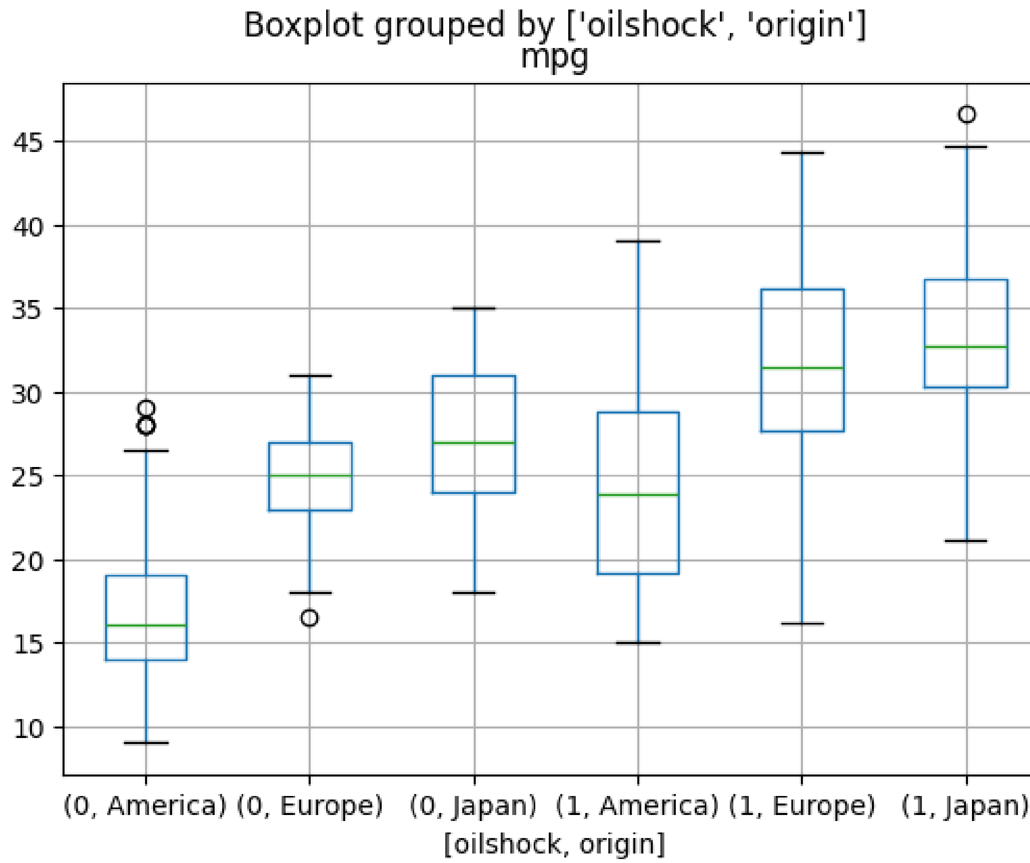
1.7.1 Encode categorical variables as dummy variables dropping the first to remove multicollinearity.

[16]:
```
def categorize_for_oil_shock(row):
    # we add 3 years because it takes approximately that long for car␣
 ↪manufacturers to introduce a new model
    if row["year"] in (70, 71, 72, 73, 74, 75, 76):
        return 0
    return 1


Auto["oilshock"] = Auto.apply(categorize_for_oil_shock, axis=1)
```

[17]:
```
Auto.boxplot(column="mpg", by=["oilshock", "origin"])
```

[17]: `<Axes: title={'center': 'mpg'}, xlabel='[oilshock, origin]'>`



[18]:
```
Auto_os = Auto.drop(["year"], axis=1)
Auto_os.columns
```

[18]:
```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
'acceleration', 'origin', 'oilshock'], dtype='object')
```

[19]:
```
# standardizing dataframes
Auto_os["oilshock"] = Auto_os["oilshock"].astype("category")
Auto_os = Auto_os.apply(standardize)
Auto_os.describe()
```

[19]:
```
                mpg     cylinders  displacement    horsepower        weight
acceleration
count  3.920000e+02  3.920000e+02  3.920000e+02  3.920000e+02  3.920000e+02
3.920000e+02
mean   1.812609e-16 -1.087565e-16 -7.250436e-17 -1.812609e-16 -3.625218e-17
-8.519262e-16
std    1.001278e+00  1.001278e+00  1.001278e+00  1.001278e+00  1.001278e+00
1.001278e+00
```

```
min   -1.853218e+00 -1.451004e+00 -1.209563e+00 -1.520975e+00 -1.608575e+00
-2.736983e+00
25%   -8.269250e-01 -8.640136e-01 -8.555316e-01 -7.665929e-01 -8.868535e-01
-6.410551e-01
50%   -8.927701e-02 -8.640136e-01 -4.153842e-01 -2.853488e-01 -2.052109e-01
-1.499869e-02
75%    7.125143e-01  1.483947e+00  7.782764e-01  5.600800e-01  7.510927e-01
5.384714e-01
max    2.970359e+00  1.483947e+00  2.493416e+00  3.265452e+00  2.549061e+00
3.360262e+00
```

[20]:
```
Auto_os = pd.get_dummies(
    Auto_os, columns=list(["origin"]), drop_first=True, dtype=np.uint8
)
Auto_os.columns
```

[20]:
```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
'acceleration', 'oilshock', 'origin_Europe', 'origin_Japan'], dtype='object')
```

[21]:
```
y = Auto_os["mpg"]
```

[21]:
```
name
chevrolet chevelle malibu              -0.698638
hi 1200d                               -1.853218
dodge d200                             -1.596645
chevy c20                              -1.724931
ford f250                              -1.724931
amc gremlin                            -0.313779
bmw 2002                                0.327654
saab 99e                                0.199368
audi 100 ls                             0.071081
volkswagen 1131 deluxe sedan            0.327654
datsun pl510                            0.455941
ford maverick                          -0.313779
amc hornet                             -0.698638
plymouth duster                        -0.185492
peugeot 504                             0.199368
buick estate wagon (sw)                -1.211785
toyota corona mark ii                   0.071081
plymouth satellite                     -0.698638
amc rebel sst                          -0.955212
ford torino                            -0.826925
ford galaxie 500                       -1.083498
chevrolet impala                       -1.211785
buick skylark 320                      -1.083498
pontiac catalina                       -1.211785
amc ambassador dpl                     -1.083498
```

dodge challenger se	-1.083498
plymouth 'cuda 340	-1.211785
chevrolet monte carlo	-1.083498
plymouth fury iii	-1.211785
amc hornet sportabout (sw)	-0.698638
chevrolet vega (sw)	-0.185492
pontiac firebird	-0.570352
ford mustang	-0.698638
mercury capri 2000	-0.057205
toyota corolla 1200	0.969088
peugeot 304	0.840801
datsun 1200	1.482234
volkswagen model 111	0.455941
plymouth cricket	0.327654
pontiac safari (sw)	-1.340071
opel 1900	0.584228
ford country squire (sw)	-1.340071
fiat 124b	0.840801
plymouth fury iii	-1.211785
dodge monaco (sw)	-1.468358
chevrolet vega 2300	0.584228
toyota corona	0.199368
amc gremlin	-0.570352
plymouth satellite custom	-0.955212
datsun pl510	0.455941
ford torino 500	-0.570352
amc matador	-0.698638
chevrolet impala	-1.211785
pontiac catalina brougham	-1.211785
ford galaxie 500	-1.211785
chevrolet chevelle malibu	-0.826925
chevrolet chevelle concours (sw)	-1.340071
plymouth satellite custom (sw)	-1.211785
volvo 145e (sw)	-0.698638
volkswagen 411 (sw)	-0.185492
peugeot 504 (sw)	-0.313779
ford pinto (sw)	-0.185492
datsun 510 (sw)	0.584228
toyouta corona mark ii (sw)	-0.057205
dodge colt (sw)	0.584228
amc matador (sw)	-1.083498
toyota corolla 1600 (sw)	0.455941
renault 12 (sw)	0.327654
mazda rx2 coupe	-0.570352
ford gran torino (sw)	-1.340071
oldsmobile delta 88 royale	-1.468358
chrysler newport royal	-1.340071

toyota corona hardtop	0.071081
volkswagen type 3	-0.057205
chevrolet vega	-0.442065
ford pinto runabout	-0.313779
chevrolet impala	-1.340071
dodge colt hardtop	0.199368
plymouth fury iii	-1.083498
ford galaxie 500	-1.211785
amc ambassador sst	-0.826925
mercury marquis	-1.596645
buick lesabre custom	-1.340071
pontiac catalina	-1.211785
fiat 124 sport coupe	0.327654
amc gremlin	-0.698638
toyota carina	-0.442065
chevrolet vega	-0.313779
datsun 610	-0.185492
maxda rx3	-0.698638
ford pinto	-0.570352
mercury capri v6	-0.313779
chevrolet monte carlo s	-1.083498
saab 99le	0.071081
fiat 128	0.712514
opel manta	0.071081
audi 100ls	-0.442065
volvo 144ea	-0.570352
dodge dart custom	-1.083498
toyota mark ii	-0.442065
oldsmobile omega	-1.596645
oldsmobile vista cruiser	-1.468358
pontiac grand prix	-0.955212
plymouth custom suburb	-1.340071
amc hornet	-0.698638
chevrolet impala	-1.596645
buick century 350	-1.340071
amc matador	-1.211785
chevrolet malibu	-1.340071
dodge coronet custom	-1.083498
mercury marquis brougham	-1.468358
chevrolet caprice classic	-1.340071
ford ltd	-1.340071
plymouth fury gran sedan	-1.211785
ford gran torino	-1.211785
buick electra 225 custom	-1.468358
amc ambassador brougham	-1.340071
plymouth valiant	-0.698638
chevrolet nova custom	-0.955212

ford country	-1.468358
ford maverick	-0.698638
plymouth duster	-0.057205
volkswagen super beetle	0.327654
chrysler new yorker brougham	-1.340071
audi fox	0.712514
volkswagen dasher	0.327654
opel manta	0.327654
toyota corona	0.969088
datsun 710	1.097374
subaru	0.327654
fiat 128	0.071081
fiat 124 tc	0.327654
honda civic	0.071081
fiat x1.9	0.969088
amc matador (sw)	-1.211785
dodge colt	0.584228
ford gran torino (sw)	-1.211785
ford gran torino	-0.955212
buick century luxus (sw)	-1.340071
dodge coronet custom (sw)	-1.211785
plymouth duster	-0.442065
amc hornet	-0.570352
chevrolet nova	-1.083498
ford pinto	0.327654
datsun b210	0.969088
chevrolet vega	0.199368
chevrolet chevelle malibu classic	-0.955212
amc matador	-0.955212
plymouth satellite sebring	-0.698638
toyota corolla 1200	1.097374
datsun 710	0.071081
pontiac astro	-0.057205
amc gremlin	-0.442065
toyota corona	0.071081
volkswagen dasher	0.199368
ford pinto	-0.698638
saab 99le	0.199368
amc pacer	-0.570352
audi 100ls	-0.057205
peugeot 504	-0.057205
volvo 244dl	-0.185492
honda civic cvcc	1.225661
ford pinto	-0.057205
volkswagen rabbit	0.712514
toyota corolla	0.712514
plymouth valiant custom	-0.570352

chevrolet monza 2+2	-0.442065
ford mustang ii	-1.340071
chevrolet nova	-0.698638
mercury monarch	-1.083498
pontiac catalina	-0.955212
chevrolet bel air	-1.083498
plymouth grand fury	-0.955212
ford maverick	-1.083498
buick century	-0.826925
chevroelt chevelle malibu	-0.955212
amc matador	-1.083498
plymouth fury	-0.698638
buick skyhawk	-0.313779
ford ltd	-1.211785
ford pinto	0.391798
pontiac ventura sj	-0.634495
amc pacer d/l	-0.762782
volkswagen rabbit	0.776658
datsun b-210	1.097374
toyota corolla	0.584228
volvo 245	-0.442065
ford f108	-1.340071
peugeot 504	-0.570352
toyota mark ii	-0.570352
mercedes-benz 280s	-0.891068
cadillac seville	-0.891068
chevy c10	-1.340071
dodge d100	-1.340071
ford granada ghia	-0.698638
plymouth volare premier v8	-1.340071
dodge aspen se	-0.442065
vw rabbit	0.712514
opel 1900	0.199368
honda civic	1.225661
fiat 131	0.584228
capri ii	0.199368
dodge colt	0.327654
renault 12tl	0.455941
dodge coronet brougham	-0.955212
amc matador	-1.019355
chevrolet chevelle malibu classic	-0.762782
plymouth valiant	-0.185492
chevrolet nova	-0.185492
ford maverick	0.071081
amc hornet	-0.121349
chevrolet chevette	0.712514
chevrolet woody	0.135225

ford gran torino	-1.147642
ford mustang ii 2+2	0.263511
volkswagen rabbit custom	0.712514
pontiac sunbird coupe	0.135225
toyota corolla liftback	0.327654
chevrolet chevette	0.904944
bmw 320i	-0.249635
subaru dl	0.840801
volkswagen dasher	0.904944
datsun 810	-0.185492
mazda rx-4	-0.249635
ford thunderbird	-0.955212
dodge colt m/m	1.289804
chrysler cordoba	-1.019355
chevrolet monte carlo landau	-1.019355
plymouth arrow gs	0.263511
buick opel isuzu deluxe	0.840801
renault 5 gtl	1.610521
datsun f-10 hatchback	1.289804
pontiac grand prix lj	-0.955212
oldsmobile cutlass supreme	-0.826925
chevrolet caprice classic	-0.762782
mercury cougar brougham	-1.083498
chevrolet concours	-0.762782
buick skylark	-0.377922
plymouth volare custom	-0.570352
ford granada	-0.634495
dodge monaco brougham	-1.019355
honda accord cvcc	1.033231
datsun 510	0.481598
toyota corona	0.520084
chevrolet chevette	0.840801
buick regal sport coupe (turbo)	-0.737124
ford futura	-0.685810
dodge omni	0.956259
dodge magnum xe	-0.762782
toyota celica gt liftback	-0.300950
peugeot 604sl	-0.929554
oldsmobile starfire sx	0.045424
datsun 200-sx	0.058253
audi 5000	-0.403579
volvo 264gl	-0.826925
saab 99gle	-0.236807
volkswagen scirocco	1.033231
honda accord lx	0.776658
plymouth sapporo	-0.031548
chevrolet monte carlo landau	-0.544694

mazda glc deluxe	1.200003
dodge aspen	-0.621666
volkswagen rabbit custom diesel	2.521356
ford fiesta	1.623349
datsun b210 gx	2.046695
honda civic cvcc	1.623349
amc concord d/l	-0.685810
dodge diplomat	-0.519037
mercury monarch ghia	-0.416408
oldsmobile cutlass salon brougham	-0.454894
chevrolet malibu	-0.377922
ford fairmont (auto)	-0.416408
ford fairmont (man)	0.212197
plymouth volare	-0.377922
amc concord	-0.519037
buick century special	-0.365093
mercury zephyr	-0.339436
pontiac phoenix lj	-0.544694
plymouth horizon	1.379605
mercedes benz 300d	0.250683
cadillac eldorado	-0.057205
peugeot 504	0.481598
oldsmobile cutlass salon brougham	0.058253
plymouth horizon tc3	1.418091
amc spirit dl	0.507256
fiat strada custom	1.777293
buick skylark limited	0.635542
chevrolet citation	0.686857
oldsmobile omega brougham	0.430284
pontiac phoenix	1.289804
datsun 210	1.071717
dodge colt hatchback custom	1.572035
dodge st. regis	-0.672981
vw rabbit custom	1.084545
mercury zephyr 6	-0.467723
ford fairmont 4	-0.147006
amc concord dl 6	-0.416408
dodge aspen 6	-0.365093
chevrolet caprice classic	-0.826925
ford ltd landau	-0.749953
pontiac lemans v6	-0.249635
maxda glc deluxe	1.366776
buick estate wagon (sw)	-0.839754
ford country squire (sw)	-1.019355
chevrolet malibu classic (sw)	-0.544694
chrysler lebaron town @ country (sw)	-0.634495
mercury grand marquis	-0.891068

vw rabbit c (diesel)	2.675299
vw dasher (diesel)	2.559841
audi 5000s (diesel)	1.661835
mercedes-benz 240d	0.840801
honda civic 1500 gl	2.713785
datsun 280-zx	1.187175
vokswagen rabbit	0.815144
mazda rx-7 gs	0.032595
triumph tr7 coupe	1.482234
honda accord	1.148689
datsun 210	2.226296
subaru dl	1.328290
dodge colt	0.571399
mazda glc	2.970359
toyota corolla	1.123031
vw rabbit	2.316097
toyota corolla tercel	1.879922
chevrolet chevette	1.110203
chevrolet citation	0.584228
ford fairmont	0.378969
datsun 310	1.764465
dodge aspen	-0.557523
audi 4000	1.392433
toyota corona liftback	0.815144
mazda 626	1.007574
datsun 510 hatchback	1.738807
amc concord	0.109567
peugeot 505s turbo diesel	0.597056
honda prelude	1.315461
toyota corolla	1.148689
datsun 200sx	1.212832
mazda 626	1.046059
volvo diesel	0.930602
chrysler lebaron salon	-0.749953
datsun 810 maxima	0.096739
buick century	-0.134177
oldsmobile cutlass ls	0.404626
ford granada gl	-0.416408
volkswagen jetta	1.225661
toyota cressida	0.250683
ford escort 2h	0.827972
plymouth reliant	0.481598
plymouth horizon 4	1.443748
ford escort 4w	1.405262
buick skylark	0.404626
dodge aries wagon (sw)	0.301997
plymouth reliant	0.840801

```
toyota starlet                        2.008209
chevrolet citation                    0.006938
honda civic 1300                      1.495063
subaru                                1.135860
datsun 210 mpg                        1.738807
toyota tercel                         1.828608
mazda glc 4                           1.366776
plymouth champ                        1.995380
chrysler lebaron medallion            0.327654
honda civic (auto)                    1.097374
datsun 310 gx                         1.867094
buick century limited                 0.199368
oldsmobile cutlass ciera (diesel)     1.867094
ford granada l                       -0.185492
dodge rampage                         1.097374
dodge charger 2.2                     1.610521
chevrolet camaro                      0.455941
ford mustang gl                       0.455941
vw pickup                             2.636813
honda civic                           1.867094
toyota celica gt                      1.097374
toyota corolla                        1.353947
ford ranger                           0.584228
nissan stanza xe                      1.610521
mercury lynx l                        1.610521
plymouth horizon miser                1.867094
mazda glc custom                      0.969088
mazda glc custom l                    1.738807
volkswagen rabbit l                   1.610521
ford fairmont futura                  0.071081
pontiac phoenix                       0.455941
dodge aries se                        0.712514
pontiac j2000 se hatchback            0.969088
chevrolet cavalier 2-door             1.353947
chevrolet cavalier wagon              0.455941
chevrolet cavalier                    0.584228
honda accord                          1.610521
chevy s-10                            0.969088
Name: mpg, dtype: float64
```

[22]:
```
cols = list(Auto_os.columns)
cols.remove("mpg")
formula = " + ".join(cols)
model = smf.ols(f"mpg ~ {formula}", data=Auto_os)
results = model.fit()
results.summary()
```

[22]:

Dep. Variable:	mpg	**R-squared:**	0.808
Model:	OLS	**Adj. R-squared:**	0.804
Method:	Least Squares	**F-statistic:**	201.7
Date:	Fri, 21 Feb 2025	**Prob (F-statistic):**	3.05e-132
Time:	19:20:21	**Log-Likelihood:**	-232.60
No. Observations:	392	**AIC:**	483.2
Df Residuals:	383	**BIC:**	518.9
Df Model:	8		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-0.4186	0.041	-10.263	0.000	-0.499	-0.338
oilshock[T.1]	0.6363	0.048	13.204	0.000	0.542	0.731
cylinders	-0.1382	0.073	-1.885	0.060	-0.282	0.006
displacement	0.2845	0.107	2.659	0.008	0.074	0.495
horsepower	-0.2213	0.069	-3.192	0.002	-0.358	-0.085
weight	-0.5923	0.073	-8.085	0.000	-0.736	-0.448
acceleration	0.0053	0.036	0.146	0.884	-0.066	0.076
origin_Europe	0.3038	0.076	4.015	0.000	0.155	0.453
origin_Japan	0.3819	0.074	5.156	0.000	0.236	0.528

Omnibus:	20.039	**Durbin-Watson:**	1.331
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	27.583
Skew:	0.413	**Prob(JB):**	1.02e-06
Kurtosis:	4.004	**Cond. No.**	11.9

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

i. Is there a relationship between the predictors and the response? Use the anova_lm() function from statsmodels to answer this question.

ii. Which predictors appear to have a statistically significant relationship to the response?

iii. What does the coefficient for the year variable suggest?

[23]:
```
anova_lm(results)
```

[23]:

	df	sum_sq	mean_sq	F	PR(>F)
oilshock	1.0	106.483141	106.483141	542.347509	2.293307e-75
cylinders	1.0	170.845795	170.845795	870.163964	1.267122e-100
displacement	1.0	11.934469	11.934469	60.785485	6.078862e-14
horsepower	1.0	3.951021	3.951021	20.123619	9.610639e-06
weight	1.0	17.796189	17.796189	90.640818	1.988543e-19
acceleration	1.0	0.009116	0.009116	0.046430	8.295108e-01
origin_Europe	1.0	0.564108	0.564108	2.873155	9.088094e-02
origin_Japan	1.0	5.218909	5.218909	26.581317	4.058867e-07
Residual	383.0	75.197253	0.196337	NaN	NaN

There seems to be a statistical relationship between all of the predictors and the response variable, mpg, except for acceleration.

Even though some of the categorical variables are insignificant, even if one of the levels is significant, it is advisable to retain them all in the model. https://stats.stackexchange.com/questions/24298/can-i-ignore-coefficients-for-non-significant-levels-of-factors-in-a-linear-mode

Note: Year has been converted to a categorical variable oilshock to better capture the effects of the oil shock of 1973 on the mileage.

1.7.2 (d) Produce some of diagnostic plots of the linear regression fit as described in the lab. Comment on any problems you see with the fit. Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage?

Before producing the diagnostic plots, let's first test for collinearity using correlation matrix and variance inflation factors.

[24]:
```
Auto_os.corr(numeric_only=True)
```

[24]:
```
                     mpg  cylinders  displacement  horsepower    weight  \
acceleration  origin_Europe  origin_Japan
mpg             1.000000  -0.777618     -0.805127   -0.778427 -0.832244
0.423329        0.244313      0.451454
cylinders      -0.777618   1.000000      0.950823    0.842983  0.897527
-0.504683       -0.352324     -0.404209
displacement   -0.805127   0.950823      1.000000    0.897257  0.932994
-0.543800       -0.371633     -0.440825
horsepower     -0.778427   0.842983      0.897257    1.000000  0.864538
-0.689196       -0.284948     -0.321936
weight         -0.832244   0.897527      0.932994    0.864538  1.000000
-0.416839       -0.293841     -0.447929
acceleration    0.423329  -0.504683     -0.543800   -0.689196 -0.416839
1.000000        0.208298      0.115020
origin_Europe   0.244313  -0.352324     -0.371633   -0.284948 -0.293841
0.208298        1.000000     -0.230157
origin_Japan    0.451454  -0.404209     -0.440825   -0.321936 -0.447929
0.115020       -0.230157      1.000000
```

[25]:
```
vifdf = calculate_VIFs("mpg ~ " + " + ".join(Auto_os.columns) + " - mpg",␣
↪Auto_os)
vifdf
```

[25]:
```
                     VIF
Feature
oilshock[T.1]   1.149269
cylinders      10.737464
displacement   22.861475
```

```
horsepower      9.594564
weight         10.715246
acceleration    2.614133
origin_Europe   1.639338
origin_Japan    1.762590
```

[26]:
```python
identify_highest_VIF_feature(vifdf)
```

```
We find the highest VIF in this model is displacement with a VIF of
22.861474853464927
Hence, we drop displacement from the model to be fitted.
```

[26]:
```
('displacement', 22.861474853464927)
```

[27]:
```python
vifdf = calculate_VIFs(
    "mpg ~ " + " + ".join(Auto_os.columns) + " - mpg - displacement", Auto_os
)
vifdf
```

[27]:
```
                    VIF
Feature
oilshock[T.1]  1.139339
cylinders      6.190903
horsepower     8.641303
weight         9.024884
acceleration   2.591157
origin_Europe  1.450726
origin_Japan   1.591434
```

[28]:
```python
identify_highest_VIF_feature(vifdf)
```

```
No variables are significantly collinear.
```

1.7.3 Linear Regression for mpg ~ cylinders + horsepower + weight + acceleration + oilshock + origin_Europe + origin_Japan

[29]:
```python
cols = list(Auto_os.columns)
cols.remove("mpg")
cols.remove("displacement")
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_os)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.805
Model:                            OLS   Adj. R-squared:                  0.801
Method:                 Least Squares   F-statistic:                     225.9
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          6.41e-132
Time:                        19:20:22   Log-Likelihood:                -236.18
```

```
No. Observations:                 392   AIC:                             488.4
Df Residuals:                     384   BIC:                             520.1
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
=
                    coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
Intercept        -0.3890      0.040     -9.837      0.000      -0.467
-0.311
oilshock[T.1]     0.6243      0.048     12.911      0.000       0.529
0.719
cylinders        -0.0113      0.056     -0.202      0.840      -0.122
0.099
horsepower       -0.1632      0.066     -2.461      0.014      -0.294
-0.033
weight           -0.5149      0.068     -7.599      0.000      -0.648
-0.382
acceleration     -0.0038      0.036     -0.103      0.918      -0.075
0.068
origin_Europe     0.2356      0.072      3.283      0.001       0.095
0.377
origin_Japan      0.3205      0.071      4.518      0.000       0.181
0.460
==============================================================================
Omnibus:                       25.646   Durbin-Watson:                   1.305
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.287
Skew:                           0.456   Prob(JB):                     1.79e-09
Kurtosis:                       4.278   Cond. No.                         7.67
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                 df      sum_sq     mean_sq           F         PR(>F)
oilshock        1.0  106.483141  106.483141  533.906720   1.149811e-74
cylinders       1.0  170.845795  170.845795  856.621225  7.985118e-100
horsepower      1.0   12.927972   12.927972   64.820882   1.039468e-14
weight          1.0   20.649905   20.649905  103.538670   1.085729e-21
acceleration    1.0    0.003626    0.003626    0.018183   8.928058e-01
origin_Europe   1.0    0.432514    0.432514    2.168627   1.416711e-01
origin_Japan    1.0    4.071523    4.071523   20.414626   8.312108e-06
Residual      384.0   76.585524    0.199441         NaN            NaN
```

[29]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b9a42c0>

[30]:
```
identify_least_significant_feature(results, alpha=LOS_Alpha)
```

```
We find the least significant variable in this model is acceleration with a
p-value of 0.9177499426057751 and a coefficient of -0.003752017674274216
Using the backward methodology, we suggest dropping acceleration from the new
model
```

Linear Regression after dropping acceleration. The model now is mpg ~ cylinders + horsepower + weight + oilshock + origin_Europe + origin_Japan

[31]:
```
cols.remove("acceleration")
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_os)
simple_model = results
models = []
models.append(
    {
        "name": "simple_model",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
    }
)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.805
Model:                            OLS   Adj. R-squared:                  0.802
Method:                 Least Squares   F-statistic:                     264.3
Date:                Fri, 21 Feb 2025   Prob (F-statistic):         3.80e-133
Time:                        19:20:22   Log-Likelihood:                -236.19
No. Observations:                 392   AIC:                             486.4
Df Residuals:                     385   BIC:                             514.2
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
=
                    coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
Intercept        -0.3889      0.039     -9.849      0.000      -0.467
-0.311
oilshock[T.1]     0.6245      0.048     12.935      0.000       0.530
0.719
cylinders        -0.0105      0.055     -0.189      0.850      -0.120
0.099
horsepower       -0.1585      0.048     -3.285      0.001      -0.253
-0.064
weight           -0.5182      0.060     -8.704      0.000      -0.635
```

```
-0.401
origin_Europe      0.2352      0.072      3.287      0.001      0.095
0.376
origin_Japan       0.3202      0.071      4.524      0.000      0.181
0.459
==============================================================================
Omnibus:                       25.330   Durbin-Watson:                   1.305
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               39.508
Skew:                           0.454   Prob(JB):                     2.64e-09
Kurtosis:                       4.263   Cond. No.                         6.84
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                  df      sum_sq     mean_sq           F         PR(>F)
oilshock         1.0  106.483141  106.483141  535.282217   7.450502e-75
cylinders        1.0  170.845795  170.845795  858.828126  4.453195e-100
horsepower       1.0   12.927972   12.927972   64.987879   9.610211e-15
weight           1.0   20.649905   20.649905  103.805416   9.634452e-22
origin_Europe    1.0    0.434982    0.434982    2.186618   1.400323e-01
origin_Japan     1.0    4.070552    4.070552   20.462339   8.111817e-06
Residual       385.0   76.587654    0.198929         NaN            NaN
```

Linear Regression after dropping cylinders. The model now is mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan

[32]:

```
cols.remove("cylinders")
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_os)
simple_model = results
models = []
models.append(
    {
        "name": "simple_model",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
    }
)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.805
Model:                            OLS   Adj. R-squared:                  0.802
Method:                 Least Squares   F-statistic:                     317.9
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          2.06e-134
Time:                        19:20:22   Log-Likelihood:                -236.21
No. Observations:                 392   AIC:                             484.4
Df Residuals:                     386   BIC:                             508.2
```

```
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
=
                    coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
Intercept        -0.3901      0.039    -10.030      0.000      -0.467
-0.314
oilshock[T.1]     0.6250      0.048     12.983      0.000       0.530
0.720
horsepower       -0.1613      0.046     -3.510      0.001      -0.252
-0.071
weight           -0.5245      0.050    -10.576      0.000      -0.622
-0.427
origin_Europe     0.2386      0.069      3.448      0.001       0.103
0.375
origin_Japan      0.3222      0.070      4.611      0.000       0.185
0.460
==============================================================================
Omnibus:                       24.971   Durbin-Watson:                   1.304
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               38.456
Skew:                           0.453   Prob(JB):                     4.46e-09
Kurtosis:                       4.239   Cond. No.                         5.60
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                  df      sum_sq     mean_sq           F        PR(>F)
oilshock         1.0  106.483141  106.483141  536.622900  4.863116e-75
horsepower       1.0  165.048555  165.048555  831.763917  2.445119e-98
weight           1.0   39.079210   39.079210  196.940090  1.939884e-36
origin_Europe    1.0    0.574647    0.574647    2.895939  8.960825e-02
origin_Japan     1.0    4.219706    4.219706   21.265252  5.446537e-06
Residual       386.0   76.594741    0.198432         NaN           NaN
```

We can now try and plot the diagnostics for the model.

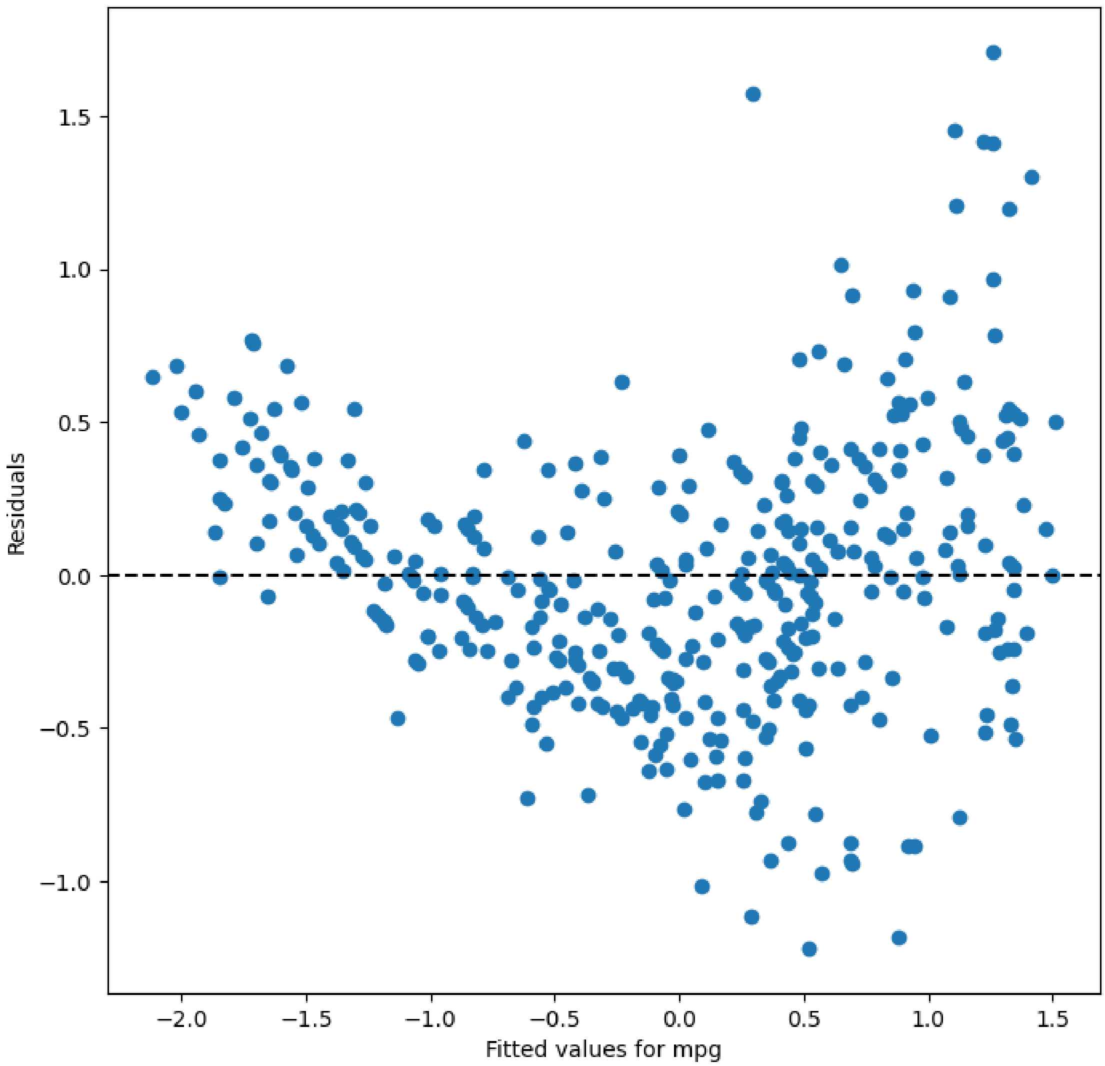
[33]:
```python
TSS = np.sum((y - np.mean(y)) ** 2)
TSS
RSS = np.sum((y - results.fittedvalues) ** 2)
RSS
RSE = np.sqrt(RSS / results.df_model)
display("RSE " + str(RSE))
display("R-squared adjusted : " + str(results.rsquared_adj))
display("F-statistic : " + str(results.fvalue))
```

'RSE 3.9139428061794668'

'R-squared adjusted : 0.8020742313429469'

'F-statistic : 317.8976193276657'

[34]:
```
display_residuals_plot(results)
```



There is some evidence of non-linearity and heteroskedasticity from the residuals plot above.

1.7.4 (e) Fit some models with interactions as described in the lab. Do any interactions appear to be statistically significant?

[35]:
```python
formula = " + ".join(cols)
formula += " + " + "horsepower: weight"
results = perform_analysis("mpg", formula, Auto_os)
numeric_interactions = results
models.append(
    {
        "name": "numeric_interactions",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
    }
)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.847
Model:                            OLS   Adj. R-squared:                  0.845
Method:                 Least Squares   F-statistic:                     355.8
Date:                Fri, 21 Feb 2025   Prob (F-statistic):         1.15e-153
Time:                        19:20:22   Log-Likelihood:                -187.98
No. Observations:                 392   AIC:                             390.0
Df Residuals:                     385   BIC:                             417.8
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
=====
                        coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----
Intercept            -0.5631      0.038    -14.715      0.000      -0.638
-0.488
oilshock[T.1]         0.6508      0.043     15.243      0.000       0.567
0.735
horsepower           -0.3723      0.045     -8.185      0.000      -0.462
-0.283
weight               -0.4926      0.044    -11.192      0.000      -0.579
-0.406
origin_Europe         0.1565      0.062      2.535      0.012       0.035
0.278
origin_Japan          0.2061      0.063      3.278      0.001       0.082
0.330
horsepower:weight     0.2300      0.022     10.364      0.000       0.186
0.274
==============================================================================
Omnibus:                       27.116   Durbin-Watson:                   1.364
```

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               45.242
Skew:                           0.457   Prob(JB):                     1.50e-10
Kurtosis:                       4.390   Cond. No.                         7.09
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                      df      sum_sq     mean_sq            F         PR(>F)
oilshock             1.0  106.483141  106.483141   684.553356   1.927144e-87
horsepower           1.0  165.048555  165.048555  1061.055689  1.099814e-112
weight               1.0   39.079210   39.079210   251.230425   6.522402e-44
origin_Europe        1.0    0.574647    0.574647     3.694260   5.533771e-02
origin_Japan         1.0    4.219706    4.219706    27.127429   3.109409e-07
horsepower:weight    1.0   16.707504   16.707504   107.408348   2.313061e-22
Residual           385.0   59.887237    0.155551          NaN            NaN
```

[36]:

```
formula = " + ".join(cols)
formula += " + " + "horsepower: weight"
formula += " + " + "oilshock: weight"
formula += " + " + "oilshock: horsepower"
results = perform_analysis("mpg", formula, Auto_os)
oilshock_interactions = results
models.append(
    {
        "name": "oilshock_interactions",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
    }
)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.861
Model:                            OLS   Adj. R-squared:                  0.858
Method:                 Least Squares   F-statistic:                     297.5
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          3.50e-159
Time:                        19:20:22   Log-Likelihood:                -168.92
No. Observations:                 392   AIC:                             355.8
Df Residuals:                     383   BIC:                             391.6
Df Model:                           8
Covariance Type:            nonrobust
================================================================================
============
                               coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------------
```

```
Intercept                     -0.5350      0.037    -14.531      0.000      -0.607
-0.463
oilshock[T.1]                  0.5913      0.042     14.071      0.000       0.509
0.674
horsepower                    -0.2801      0.051     -5.456      0.000      -0.381
-0.179
oilshock[T.1]:horsepower      -0.2276      0.088     -2.598      0.010      -0.400
-0.055
weight                        -0.4591      0.050     -9.226      0.000      -0.557
-0.361
oilshock[T.1]:weight          -0.0963      0.082     -1.181      0.238      -0.257
0.064
origin_Europe                  0.1804      0.059      3.039      0.003       0.064
0.297
origin_Japan                   0.1929      0.060      3.209      0.001       0.075
0.311
horsepower:weight              0.1715      0.023      7.403      0.000       0.126
0.217
==============================================================================
Omnibus:                       23.934   Durbin-Watson:                   1.456
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.610
Skew:                           0.377   Prob(JB):                     3.39e-10
Kurtosis:                       4.450   Cond. No.                         11.3
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                        df      sum_sq     mean_sq            F         PR(>F)
oilshock               1.0  106.483141  106.483141   750.563317   2.852400e-92
horsepower             1.0  165.048555  165.048555  1163.370934  3.985476e-118
oilshock:horsepower    1.0   19.155472   19.155472   135.020381   6.081794e-27
weight                 1.0   35.071635   35.071635   247.207986   2.471356e-43
oilshock:weight        1.0    0.468549    0.468549     3.302640   6.994992e-02
origin_Europe          1.0    0.971997    0.971997     6.851278   9.208817e-03
origin_Japan           1.0    2.687942    2.687942    18.946386   1.727234e-05
horsepower:weight      1.0    7.776131    7.776131    54.811293   8.507915e-13
Residual             383.0   54.336579    0.141871          NaN            NaN
```

[37]:

```
formula = " + ".join(cols)
formula += " + " + "oilshock: horsepower"
formula += " + " + "origin_Europe: horsepower"
formula += " + " + "origin_Japan: horsepower"
formula += " + " + "origin_Europe: weight"
formula += " + " + "origin_Japan: weight"
formula += " + " + "oilshock: weight"
formula += " + " + "oilshock: horsepower"
```

```python
results = perform_analysis("mpg", formula, Auto_os)
origin_interactions = results
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.855
Model:                            OLS   Adj. R-squared:                  0.851
Method:                 Least Squares   F-statistic:                     204.1
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          6.22e-152
Time:                        19:20:22   Log-Likelihood:                -177.44
No. Observations:                 392   AIC:                             378.9
Df Residuals:                     380   BIC:                             426.5
Df Model:                          11
Covariance Type:            nonrobust
================================================================================
============
                              coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------------
Intercept                  -0.4245      0.035    -12.121      0.000      -0.493
-0.356
oilshock[T.1]               0.5690      0.044     13.054      0.000       0.483
0.655
horsepower                 -0.0708      0.048     -1.470      0.142      -0.166
0.024
oilshock[T.1]:horsepower   -0.1615      0.096     -1.687      0.092      -0.350
0.027
weight                     -0.4713      0.054     -8.712      0.000      -0.578
-0.365
oilshock[T.1]:weight       -0.2333      0.087     -2.694      0.007      -0.404
-0.063
origin_Europe               0.0297      0.082      0.363      0.717      -0.131
0.191
origin_Japan               -0.0010      0.131     -0.007      0.994      -0.259
0.257
origin_Europe:horsepower   -0.5852      0.130     -4.515      0.000      -0.840
-0.330
origin_Japan:horsepower    -0.2801      0.204     -1.370      0.172      -0.682
0.122
origin_Europe:weight        0.1640      0.120      1.370      0.171      -0.071
0.399
origin_Japan:weight        -0.1326      0.245     -0.541      0.589      -0.614
0.349
==============================================================================
Omnibus:                       20.717   Durbin-Watson:                   1.585
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               32.838
Skew:                           0.373   Prob(JB):                     7.40e-08
```

```
Kurtosis:                       4.205   Cond. No.                         25.3
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                             df       sum_sq     mean_sq            F
PR(>F)
oilshock                    1.0  106.483141  106.483141   712.972849
3.365342e-89
horsepower                  1.0  165.048555  165.048555  1105.105820
1.586816e-114
oilshock:horsepower         1.0   19.155472   19.155472   128.258155
8.134152e-26
weight                      1.0   35.071635   35.071635   234.827067
1.302452e-41
oilshock:weight             1.0    0.468549    0.468549     3.137234
7.732495e-02
origin_Europe               1.0    0.971997    0.971997     6.508146
1.112920e-02
origin_Japan                1.0    2.687942    2.687942    17.997494
2.781984e-05
origin_Europe:horsepower    1.0    3.024522    3.024522    20.251113
9.040705e-06
origin_Japan:horsepower     1.0    1.977640    1.977640    13.241566
3.116551e-04
origin_Europe:weight        1.0    0.313437    0.313437     2.098659
1.482531e-01
origin_Japan:weight         1.0    0.043767    0.043767     0.293050
5.885897e-01
Residual                  380.0   56.753344    0.149351          NaN
NaN
```

[37]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b6c24e0>

- From the above analysis, we can see that there is no significant interaction between origin and weight.
- So we can omit them from the model.

[38]:
```python
formula = " + ".join(cols)
formula += " + " + "oilshock: horsepower"
formula += " + " + "origin_Europe: horsepower"
formula += " + " + "origin_Japan: horsepower"
formula += " + " + "oilshock: weight"
formula += " + " + "oilshock: horsepower"
results = perform_analysis("mpg", formula, Auto_os)
origin_interactions = results
```

```
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.854
Model:                            OLS   Adj. R-squared:                  0.851
Method:                 Least Squares   F-statistic:                     248.9
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          8.18e-154
Time:                        19:20:22   Log-Likelihood:                -178.67
No. Observations:                 392   AIC:                             377.3
Df Residuals:                     382   BIC:                             417.1
Df Model:                           9
Covariance Type:            nonrobust
================================================================================
============
                            coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------------
Intercept                -0.4296      0.034    -12.650      0.000      -0.496
-0.363
oilshock[T.1]             0.5693      0.043     13.280      0.000       0.485
0.654
horsepower               -0.0797      0.047     -1.689      0.092      -0.172
0.013
oilshock[T.1]:horsepower -0.1958      0.092     -2.133      0.034      -0.376
-0.015
weight                   -0.4571      0.051     -8.947      0.000      -0.558
-0.357
oilshock[T.1]:weight     -0.2019      0.084     -2.408      0.016      -0.367
-0.037
origin_Europe             0.0044      0.080      0.055      0.956      -0.153
0.162
origin_Japan              0.0750      0.084      0.892      0.373      -0.090
0.240
origin_Europe:horsepower -0.4667      0.096     -4.884      0.000      -0.655
-0.279
origin_Japan:horsepower  -0.3682      0.101     -3.637      0.000      -0.567
-0.169
==============================================================================
Omnibus:                       20.114   Durbin-Watson:                   1.577
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               31.613
Skew:                           0.366   Prob(JB):                     1.37e-07
Kurtosis:                       4.183   Cond. No.                         10.2
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                            df      sum_sq     mean_sq            F
PR(>F)
```

```
oilshock                       1.0  106.483141  106.483141   712.242504
2.550768e-89
horsepower                     1.0  165.048555  165.048555  1103.973788
9.879451e-115
oilshock:horsepower            1.0   19.155472   19.155472   128.126771
8.220841e-26
weight                         1.0   35.071635   35.071635   234.586518
1.270254e-41
oilshock:weight                1.0    0.468549    0.468549     3.134021
7.747196e-02
origin_Europe                  1.0    0.971997    0.971997     6.501479
1.116822e-02
origin_Japan                   1.0    2.687942    2.687942    17.979058
2.804561e-05
origin_Europe:horsepower       1.0    3.024522    3.024522    20.230368
9.121050e-06
origin_Japan:horsepower        1.0    1.977640    1.977640    13.228002
3.136339e-04
Residual                     382.0   57.110548    0.149504          NaN
NaN
```

[38]:
```
<statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b6d7260>
```

- From the above analysis, it is evident that with the interaction between origin and horsepower, the interaction between oilshock and weight and horsepower is insignificant. We can drop these from the model as well.

[39]:
```python
formula = " + ".join(cols)
formula += " + " + "oilshock: horsepower"
formula += " + " + "origin_Europe: horsepower"
formula += " + " + "origin_Japan: horsepower"
results = perform_analysis("mpg", formula, Auto_os)
origin_interactions = results
models.append(
    {
        "name": "origin_interactions",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
    }
)
```

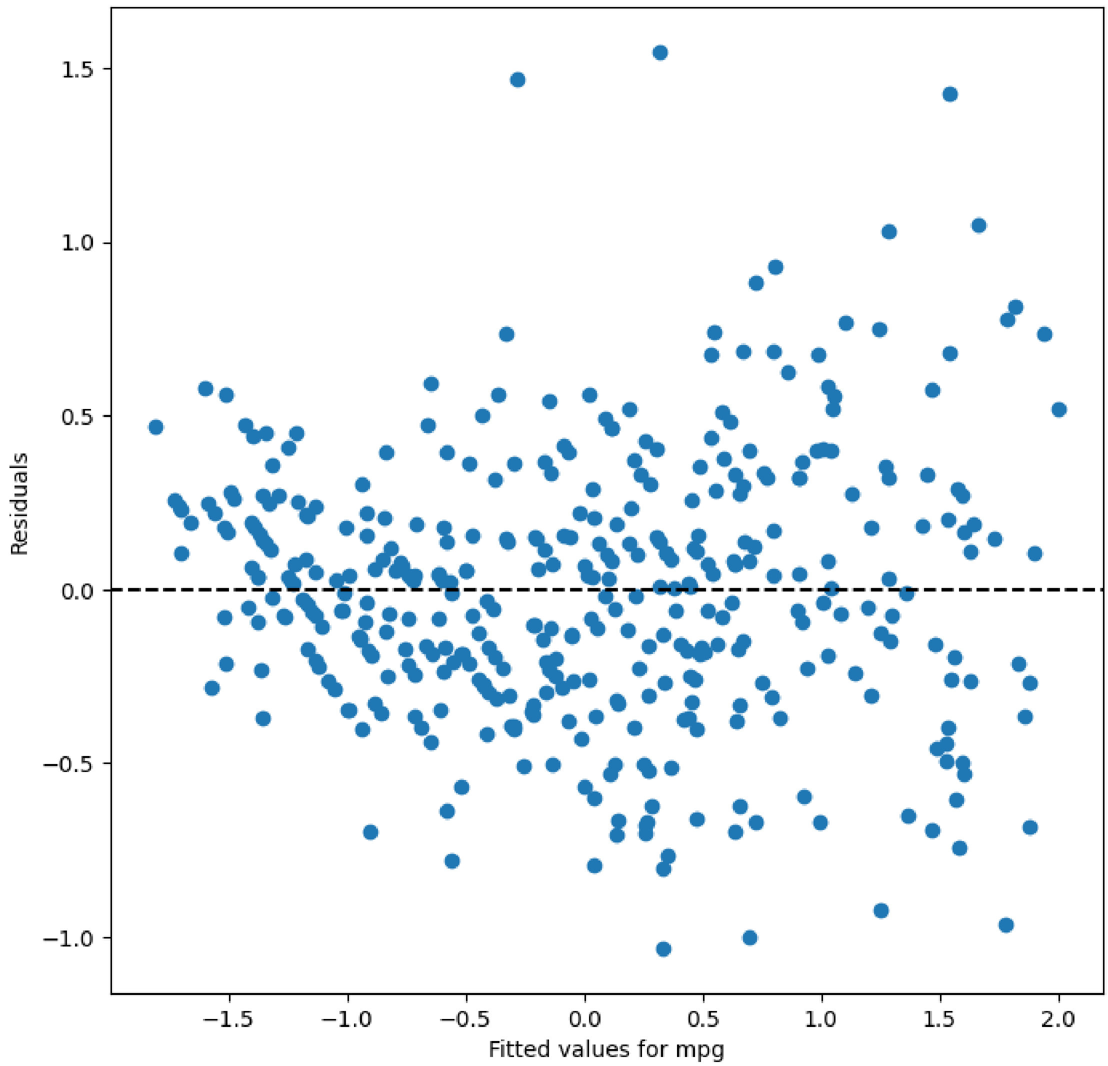
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.852
Model:                            OLS   Adj. R-squared:                  0.849
Method:                 Least Squares   F-statistic:                     275.8
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          8.41e-154
Time:                        19:20:23   Log-Likelihood:                -181.63
No. Observations:                 392   AIC:                             381.3
```

```
Df Residuals:                     383   BIC:                             417.0
Df Model:                           8
Covariance Type:            nonrobust
================================================================================
============
                              coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------------
Intercept                  -0.4267      0.034    -12.495      0.000      -0.494
-0.360
oilshock[T.1]               0.5628      0.043     13.073      0.000       0.478
0.647
horsepower                 -0.0265      0.042     -0.633      0.527      -0.109
0.056
oilshock[T.1]:horsepower   -0.3804      0.051     -7.497      0.000      -0.480
-0.281
weight                     -0.5231      0.043    -12.051      0.000      -0.608
-0.438
origin_Europe               0.0041      0.081      0.051      0.959      -0.155
0.163
origin_Japan                0.0877      0.084      1.039      0.299      -0.078
0.254
origin_Europe:horsepower   -0.4423      0.096     -4.626      0.000      -0.630
-0.254
origin_Japan:horsepower    -0.3589      0.102     -3.526      0.000      -0.559
-0.159
==============================================================================
Omnibus:                       19.159   Durbin-Watson:                   1.576
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               29.046
Skew:                           0.362   Prob(JB):                     4.93e-07
Kurtosis:                       4.119   Cond. No.                         9.30
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                            df      sum_sq     mean_sq            F
PR(>F)
oilshock                   1.0  106.483141  106.483141   703.425918
9.826044e-89
horsepower                 1.0  165.048555  165.048555  1090.308104
4.259700e-114
oilshock:horsepower        1.0   19.155472   19.155472   126.540737
1.468876e-25
weight                     1.0   35.071635   35.071635   231.682658
2.992243e-41
origin_Europe              1.0    0.792887    0.792887     5.237801
```

```
2.264506e-02
origin_Japan                1.0     2.840217     2.840217     18.762427
1.893486e-05
origin_Europe:horsepower    1.0     2.748574     2.748574     18.157033
2.563625e-05
origin_Japan:horsepower     1.0     1.881783     1.881783     12.431031
4.733982e-04
Residual                  383.0    57.977737     0.151378           NaN
NaN
```

[40]:
```
display_residuals_plot(results)
```



[41]:
```
anova_lm(simple_model, numeric_interactions, oilshock_interactions,␣
 ↪origin_interactions)
```

[41]:
```
   df_resid        ssr  df_diff    ss_diff           F        Pr(>F)
0     386.0  76.594741      0.0        NaN         NaN           NaN
1     385.0  59.887237      1.0  16.707504  110.369506  7.218441e-23
2     383.0  54.336579      2.0   5.550658   18.333780  2.489897e-08
3     383.0  57.977737     -0.0  -3.641158         inf           NaN
```

[42]:
```python
pd.DataFrame(models)
```

[42]:
```
                     name
model  R-squared adjusted
0           simple_model
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan
0.802074
1   numeric_interactions
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan +
horsepower: weight            0.844846
2  oilshock_interactions                mpg ~ horsepower + weight + oilshock +
origin_Europe + origin_Japan + horsepower: weight + oilshock: weight + oilshock:
horsepower            0.858491
3    origin_interactions  mpg ~ horsepower + weight + oilshock + origin_Europe +
origin_Japan + oilshock: horsepower + origin_Europe: horsepower + origin_Japan:
horsepower            0.849008
```

1.7.5 (f) Try a few different transformations of the variables, such as log(X), √X, X2 . Comment on your findings.

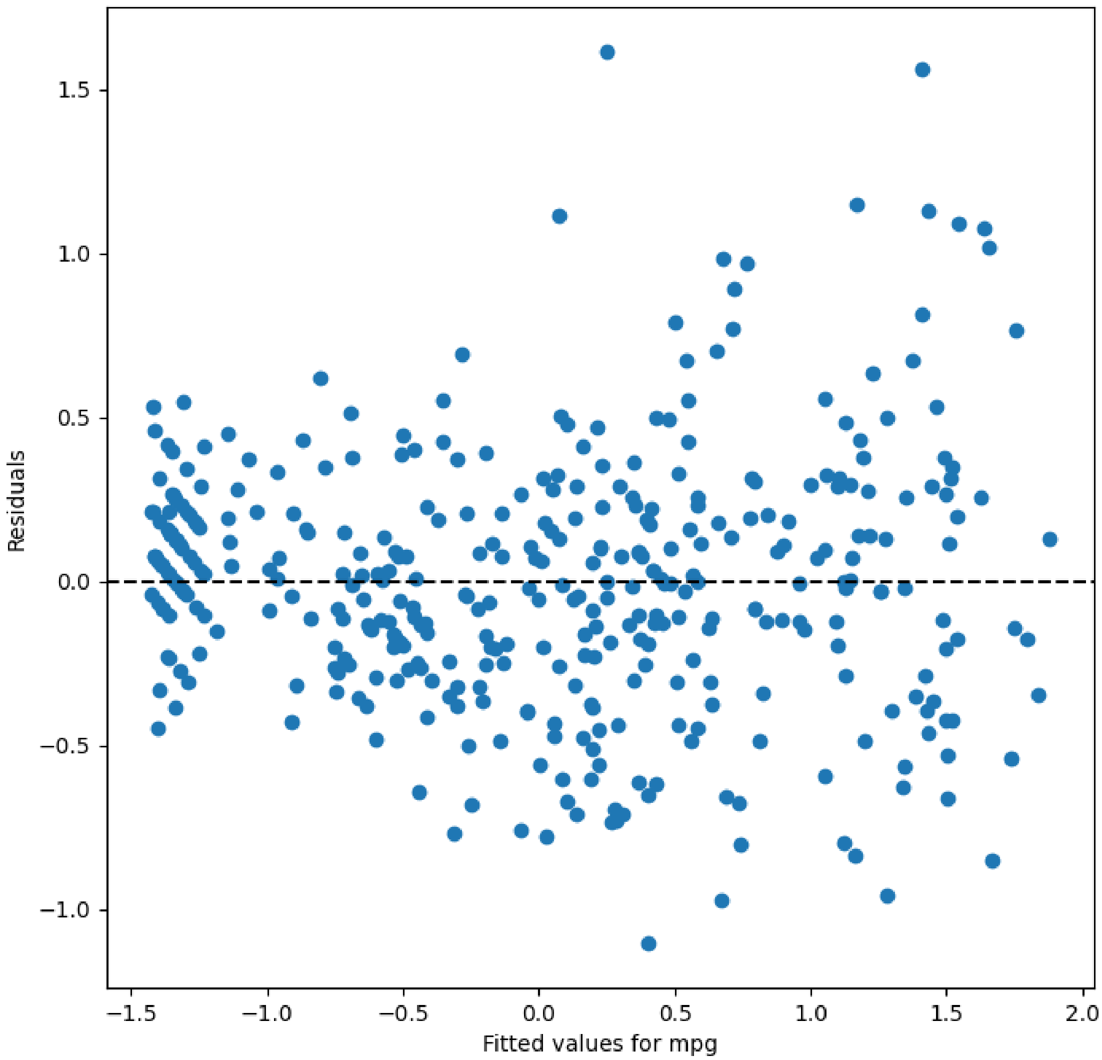
[43]:
```python
formula = simple_model.model.formula
formula = formula[formula.rindex("~") + 1 :]
# Add higher order transformations for horsepower and weight
formula += " + " + "I(horsepower**2)"
formula += " + " + "I(weight**2)"
results = perform_analysis("mpg", formula, Auto_os)
squared_transformations = results
models.append(
    {
        "name": "squared_transformation",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
    }
)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.848
Model:                            OLS   Adj. R-squared:                  0.846
Method:                 Least Squares   F-statistic:                     306.8
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          5.75e-153
Time:                        19:20:23   Log-Likelihood:                -186.58
```

```
No. Observations:                 392   AIC:                             389.2
Df Residuals:                     384   BIC:                             420.9
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
======
                         coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------
Intercept             -0.6020      0.040    -15.072      0.000      -0.681
-0.524
oilshock[T.1]          0.6580      0.043     15.334      0.000       0.574
0.742
horsepower            -0.3798      0.054     -7.058      0.000      -0.486
-0.274
weight                -0.5069      0.050    -10.044      0.000      -0.606
-0.408
origin_Europe          0.1436      0.062      2.324      0.021       0.022
0.265
origin_Japan           0.1959      0.064      3.047      0.002       0.069
0.322
I(horsepower ** 2)     0.0976      0.021      4.667      0.000       0.056
0.139
I(weight ** 2)         0.1413      0.026      5.382      0.000       0.090
0.193
==============================================================================
Omnibus:                       26.672   Durbin-Watson:                   1.394
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               46.435
Skew:                           0.436   Prob(JB):                     8.26e-11
Kurtosis:                       4.443   Cond. No.                         10.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                       df      sum_sq     mean_sq            F         PR(>F)
oilshock              1.0  106.483141  106.483141   687.677177   1.332036e-87
horsepower            1.0  165.048555  165.048555  1065.897602  7.775452e-113
weight                1.0   39.079210   39.079210   252.376864   4.864770e-44
origin_Europe         1.0    0.574647    0.574647     3.711118   5.478894e-02
origin_Japan          1.0    4.219706    4.219706    27.251220   2.932534e-07
I(horsepower ** 2)    1.0   12.648520   12.648520    81.685220   7.946432e-18
I(weight ** 2)        1.0    4.485871    4.485871    28.970134   1.281966e-07
Residual            384.0   59.460351    0.154845          NaN            NaN
```

[44]:
```
display_residuals_plot(results)
```



[45]:
```
anova_lm(simple_model, squared_transformations)
```

[45]:

	df_resid	ssr	df_diff	ss_diff	F	Pr(>F)
0	386.0	76.594741	0.0	NaN	NaN	NaN
1	384.0	59.460351	2.0	17.134391	55.327677	7.681995e-22

[46]:
```
pd.DataFrame(models)
```

[46]:
```
                        name
model  R-squared adjusted
0              simple_model
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan
0.802074
1     numeric_interactions
```

```
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan +
horsepower: weight            0.844846
2   oilshock_interactions                  mpg ~ horsepower + weight + oilshock +
origin_Europe + origin_Japan + horsepower: weight + oilshock: weight + oilshock:
horsepower            0.858491
3     origin_interactions  mpg ~ horsepower + weight + oilshock + origin_Europe
+ origin_Japan + oilshock: horsepower + origin_Europe: horsepower +
origin_Japan: horsepower            0.849008
4  squared_transformation                                                   mpg ~
horsepower + weight + oilshock + origin_Europe + origin_Japan + I(horsepower**2)
+ I(weight**2)            0.845550
```

- Since we've standardized the variables, we cannot run log or square root transformations on the negative valued columns.
- We can reload the data and run the log and sqrt transformations on the original un-standardized data.

[47]:
```
Auto = load_data("Auto")
Auto = Auto.sort_values(by=["year"], ascending=True)
Auto.columns
```

[47]:
```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
'acceleration', 'year', 'origin'], dtype='object')
```

[48]:
```
print("Minimums:")
print(Auto.min())
print("Maximums:")
print(Auto.max())
```

```
Minimums:
mpg                9.0
cylinders          3.0
displacement      68.0
horsepower        46.0
weight          1613.0
acceleration       8.0
year              70.0
origin             1.0
dtype: float64
Maximums:
mpg               46.6
cylinders          8.0
displacement     455.0
horsepower       230.0
weight          5140.0
acceleration      24.8
year              82.0
origin             3.0
```

```
dtype: float64
```

- From the above, we can see that the values for displacement, horsepower and weight are quite large.
- Hence, we log or square root transform only these variables.

1.7.6 Now let's categorize the variables

[49]:
```
Auto["origin"] = Auto["origin"].astype("category")
Auto["origin"] = Auto["origin"].cat.rename_categories(
    {1: "America", 2: "Europe", 3: "Japan"}
)
Auto["year"] = Auto["year"].astype("category")
Auto["oilshock"] = Auto.apply(categorize_for_oil_shock, axis=1)
```

1.8 Log Transformed Model

[50]:
```
Auto_log = Auto.copy(deep=True)
```

[50]:
```
                                mpg  cylinders  displacement  horsepower  
weight  acceleration year  origin  oilshock
name
chevrolet chevelle malibu      18.0          8         307.0         130
3504          12.0   70  America        0
hi 1200d                        9.0          8         304.0         193
4732          18.5   70  America        0
dodge d200                     11.0          8         318.0         210
4382          13.5   70  America        0
chevy c20                      10.0          8         307.0         200
4376          15.0   70  America        0
ford f250                      10.0          8         360.0         215
4615          14.0   70  America        0
amc gremlin                    21.0          6         199.0          90
2648          15.0   70  America        0
bmw 2002                       26.0          4         121.0         113
2234          12.5   70   Europe        0
saab 99e                       25.0          4         104.0          95
2375          17.5   70   Europe        0
audi 100 ls                    24.0          4         107.0          90
2430          14.5   70   Europe        0
volkswagen 1131 deluxe sedan   26.0          4          97.0          46
1835          20.5   70   Europe        0
datsun pl510                   27.0          4          97.0          88
2130          14.5   70    Japan        0
ford maverick                  21.0          6         200.0          85
2587          16.0   70  America        0
amc hornet                     18.0          6         199.0          97
2774          15.5   70  America        0
```

```
plymouth duster                      22.0          6          198.0          95
2833         15.5   70  America         0
peugeot 504                          25.0          4          110.0          87
2672         17.5   70   Europe         0
buick estate wagon (sw)              14.0          8          455.0         225
3086         10.0   70  America         0
toyota corona mark ii                24.0          4          113.0          95
2372         15.0   70    Japan         0
plymouth satellite                   18.0          8          318.0         150
3436         11.0   70  America         0
amc rebel sst                        16.0          8          304.0         150
3433         12.0   70  America         0
ford torino                          17.0          8          302.0         140
3449         10.5   70  America         0
ford galaxie 500                     15.0          8          429.0         198
4341         10.0   70  America         0
chevrolet impala                     14.0          8          454.0         220
4354          9.0   70  America         0
buick skylark 320                    15.0          8          350.0         165
3693         11.5   70  America         0
pontiac catalina                     14.0          8          455.0         225
4425         10.0   70  America         0
amc ambassador dpl                   15.0          8          390.0         190
3850          8.5   70  America         0
dodge challenger se                  15.0          8          383.0         170
3563         10.0   70  America         0
plymouth 'cuda 340                   14.0          8          340.0         160
3609          8.0   70  America         0
chevrolet monte carlo                15.0          8          400.0         150
3761          9.5   70  America         0
plymouth fury iii                    14.0          8          440.0         215
4312          8.5   70  America         0
amc hornet sportabout (sw)           18.0          6          258.0         110
2962         13.5   71  America         0
chevrolet vega (sw)                  22.0          4          140.0          72
2408         19.0   71  America         0
pontiac firebird                     19.0          6          250.0         100
3282         15.0   71  America         0
ford mustang                         18.0          6          250.0          88
3139         14.5   71  America         0
mercury capri 2000                   23.0          4          122.0          86
2220         14.0   71  America         0
toyota corolla 1200                  31.0          4           71.0          65
1773         19.0   71    Japan         0
peugeot 304                          30.0          4           79.0          70
2074         19.5   71   Europe         0
datsun 1200                          35.0          4           72.0          69
```

1613	18.0	71	Japan	0
volkswagen model 111	27.0	4	97.0	60
1834	19.0	71	Europe	0
plymouth cricket	26.0	4	91.0	70
1955	20.5	71	America	0
pontiac safari (sw)	13.0	8	400.0	175
5140	12.0	71	America	0
opel 1900	28.0	4	116.0	90
2123	14.0	71	Europe	0
ford country squire (sw)	13.0	8	400.0	170
4746	12.0	71	America	0
fiat 124b	30.0	4	88.0	76
2065	14.5	71	Europe	0
plymouth fury iii	14.0	8	318.0	150
4096	13.0	71	America	0
dodge monaco (sw)	12.0	8	383.0	180
4955	11.5	71	America	0
chevrolet vega 2300	28.0	4	140.0	90
2264	15.5	71	America	0
toyota corona	25.0	4	113.0	95
2228	14.0	71	Japan	0
amc gremlin	19.0	6	232.0	100
2634	13.0	71	America	0
plymouth satellite custom	16.0	6	225.0	105
3439	15.5	71	America	0
datsun pl510	27.0	4	97.0	88
2130	14.5	71	Japan	0
ford torino 500	19.0	6	250.0	88
3302	15.5	71	America	0
amc matador	18.0	6	232.0	100
3288	15.5	71	America	0
chevrolet impala	14.0	8	350.0	165
4209	12.0	71	America	0
pontiac catalina brougham	14.0	8	400.0	175
4464	11.5	71	America	0
ford galaxie 500	14.0	8	351.0	153
4154	13.5	71	America	0
chevrolet chevelle malibu	17.0	6	250.0	100
3329	15.5	71	America	0
chevrolet chevelle concours (sw)	13.0	8	307.0	130
4098	14.0	72	America	0
plymouth satellite custom (sw)	14.0	8	318.0	150
4077	14.0	72	America	0
volvo 145e (sw)	18.0	4	121.0	112
2933	14.5	72	Europe	0
volkswagen 411 (sw)	22.0	4	121.0	76
2511	18.0	72	Europe	0

```
peugeot 504 (sw)                    21.0          4        120.0          87
2979          19.5   72   Europe       0
ford pinto (sw)                     22.0          4        122.0          86
2395          16.0   72  America       0
datsun 510 (sw)                     28.0          4         97.0          92
2288          17.0   72    Japan       0
toyouta corona mark ii (sw)         23.0          4        120.0          97
2506          14.5   72    Japan       0
dodge colt (sw)                     28.0          4         98.0          80
2164          15.0   72  America       0
amc matador (sw)                    15.0          8        304.0         150
3892          12.5   72  America       0
toyota corolla 1600 (sw)            27.0          4         97.0          88
2100          16.5   72    Japan       0
renault 12 (sw)                     26.0          4         96.0          69
2189          18.0   72   Europe       0
mazda rx2 coupe                     19.0          3         70.0          97
2330          13.5   72    Japan       0
ford gran torino (sw)               13.0          8        302.0         140
4294          16.0   72  America       0
oldsmobile delta 88 royale          12.0          8        350.0         160
4456          13.5   72  America       0
chrysler newport royal              13.0          8        400.0         190
4422          12.5   72  America       0
toyota corona hardtop               24.0          4        113.0          95
2278          15.5   72    Japan       0
volkswagen type 3                   23.0          4         97.0          54
2254          23.5   72   Europe       0
chevrolet vega                      20.0          4        140.0          90
2408          19.5   72  America       0
ford pinto runabout                 21.0          4        122.0          86
2226          16.5   72  America       0
chevrolet impala                    13.0          8        350.0         165
4274          12.0   72  America       0
dodge colt hardtop                  25.0          4         97.5          80
2126          17.0   72  America       0
plymouth fury iii                   15.0          8        318.0         150
4135          13.5   72  America       0
ford galaxie 500                    14.0          8        351.0         153
4129          13.0   72  America       0
amc ambassador sst                  17.0          8        304.0         150
3672          11.5   72  America       0
mercury marquis                     11.0          8        429.0         208
4633          11.0   72  America       0
buick lesabre custom                13.0          8        350.0         155
4502          13.5   72  America       0
pontiac catalina                    14.0          8        400.0         175
```

4385	12.0	72	America	0			
fiat 124 sport coupe				26.0	4	98.0	90
2265	15.5	73	Europe	0			
amc gremlin				18.0	6	232.0	100
2789	15.0	73	America	0			
toyota carina				20.0	4	97.0	88
2279	19.0	73	Japan	0			
chevrolet vega				21.0	4	140.0	72
2401	19.5	73	America	0			
datsun 610				22.0	4	108.0	94
2379	16.5	73	Japan	0			
maxda rx3				18.0	3	70.0	90
2124	13.5	73	Japan	0			
ford pinto				19.0	4	122.0	85
2310	18.5	73	America	0			
mercury capri v6				21.0	6	155.0	107
2472	14.0	73	America	0			
chevrolet monte carlo s				15.0	8	350.0	145
4082	13.0	73	America	0			
saab 99le				24.0	4	121.0	110
2660	14.0	73	Europe	0			
fiat 128				29.0	4	68.0	49
1867	19.5	73	Europe	0			
opel manta				24.0	4	116.0	75
2158	15.5	73	Europe	0			
audi 100ls				20.0	4	114.0	91
2582	14.0	73	Europe	0			
volvo 144ea				19.0	4	121.0	112
2868	15.5	73	Europe	0			
dodge dart custom				15.0	8	318.0	150
3399	11.0	73	America	0			
toyota mark ii				20.0	6	156.0	122
2807	13.5	73	Japan	0			
oldsmobile omega				11.0	8	350.0	180
3664	11.0	73	America	0			
oldsmobile vista cruiser				12.0	8	350.0	180
4499	12.5	73	America	0			
pontiac grand prix				16.0	8	400.0	230
4278	9.5	73	America	0			
plymouth custom suburb				13.0	8	360.0	170
4654	13.0	73	America	0			
amc hornet				18.0	6	232.0	100
2945	16.0	73	America	0			
chevrolet impala				11.0	8	400.0	150
4997	14.0	73	America	0			
buick century 350				13.0	8	350.0	175
4100	13.0	73	America	0			

amc matador				14.0	8	304.0	150
3672	11.5	73	America	0			
chevrolet malibu				13.0	8	350.0	145
3988	13.0	73	America	0			
dodge coronet custom				15.0	8	318.0	150
3777	12.5	73	America	0			
mercury marquis brougham				12.0	8	429.0	198
4952	11.5	73	America	0			
chevrolet caprice classic				13.0	8	400.0	150
4464	12.0	73	America	0			
ford ltd				13.0	8	351.0	158
4363	13.0	73	America	0			
plymouth fury gran sedan				14.0	8	318.0	150
4237	14.5	73	America	0			
ford gran torino				14.0	8	302.0	137
4042	14.5	73	America	0			
buick electra 225 custom				12.0	8	455.0	225
4951	11.0	73	America	0			
amc ambassador brougham				13.0	8	360.0	175
3821	11.0	73	America	0			
plymouth valiant				18.0	6	225.0	105
3121	16.5	73	America	0			
chevrolet nova custom				16.0	6	250.0	100
3278	18.0	73	America	0			
ford country				12.0	8	400.0	167
4906	12.5	73	America	0			
ford maverick				18.0	6	250.0	88
3021	16.5	73	America	0			
plymouth duster				23.0	6	198.0	95
2904	16.0	73	America	0			
volkswagen super beetle				26.0	4	97.0	46
1950	21.0	73	Europe	0			
chrysler new yorker brougham				13.0	8	440.0	215
4735	11.0	73	America	0			
audi fox				29.0	4	98.0	83
2219	16.5	74	Europe	0			
volkswagen dasher				26.0	4	79.0	67
1963	15.5	74	Europe	0			
opel manta				26.0	4	97.0	78
2300	14.5	74	Europe	0			
toyota corona				31.0	4	76.0	52
1649	16.5	74	Japan	0			
datsun 710				32.0	4	83.0	61
2003	19.0	74	Japan	0			
subaru				26.0	4	108.0	93
2391	15.5	74	Japan	0			
fiat 128				24.0	4	90.0	75

2108	15.5	74	Europe	0
fiat 124 tc	26.0	4	116.0	75
2246	14.0	74	Europe	0
honda civic	24.0	4	120.0	97
2489	15.0	74	Japan	0
fiat x1.9	31.0	4	79.0	67
2000	16.0	74	Europe	0
amc matador (sw)	14.0	8	304.0	150
4257	15.5	74	America	0
dodge colt	28.0	4	90.0	75
2125	14.5	74	America	0
ford gran torino (sw)	14.0	8	302.0	140
4638	16.0	74	America	0
ford gran torino	16.0	8	302.0	140
4141	14.0	74	America	0
buick century luxus (sw)	13.0	8	350.0	150
4699	14.5	74	America	0
dodge coronet custom (sw)	14.0	8	318.0	150
4457	13.5	74	America	0
plymouth duster	20.0	6	198.0	95
3102	16.5	74	America	0
amc hornet	19.0	6	232.0	100
2901	16.0	74	America	0
chevrolet nova	15.0	6	250.0	100
3336	17.0	74	America	0
ford pinto	26.0	4	122.0	80
2451	16.5	74	America	0
datsun b210	31.0	4	79.0	67
1950	19.0	74	Japan	0
chevrolet vega	25.0	4	140.0	75
2542	17.0	74	America	0
chevrolet chevelle malibu classic	16.0	6	250.0	100
3781	17.0	74	America	0
amc matador	16.0	6	258.0	110
3632	18.0	74	America	0
plymouth satellite sebring	18.0	6	225.0	105
3613	16.5	74	America	0
toyota corolla 1200	32.0	4	71.0	65
1836	21.0	74	Japan	0
datsun 710	24.0	4	119.0	97
2545	17.0	75	Japan	0
pontiac astro	23.0	4	140.0	78
2592	18.5	75	America	0
amc gremlin	20.0	6	232.0	100
2914	16.0	75	America	0
toyota corona	24.0	4	134.0	96
2702	13.5	75	Japan	0

```
volkswagen dasher                     25.0          4         90.0          71
2223          16.5   75   Europe         0
ford pinto                            18.0          6        171.0          97
2984          14.5   75  America         0
saab 99le                             25.0          4        121.0         115
2671          13.5   75   Europe         0
amc pacer                             19.0          6        232.0          90
3211          17.0   75  America         0
audi 100ls                            23.0          4        115.0          95
2694          15.0   75   Europe         0
peugeot 504                           23.0          4        120.0          88
2957          17.0   75   Europe         0
volvo 244dl                           22.0          4        121.0          98
2945          14.5   75   Europe         0
honda civic cvcc                      33.0          4         91.0          53
1795          17.5   75    Japan         0
ford pinto                            23.0          4        140.0          83
2639          17.0   75  America         0
volkswagen rabbit                     29.0          4         90.0          70
1937          14.0   75   Europe         0
toyota corolla                        29.0          4         97.0          75
2171          16.0   75    Japan         0
plymouth valiant custom               19.0          6        225.0          95
3264          16.0   75  America         0
chevrolet monza 2+2                   20.0          8        262.0         110
3221          13.5   75  America         0
ford mustang ii                       13.0          8        302.0         129
3169          12.0   75  America         0
chevrolet nova                        18.0          6        250.0         105
3459          16.0   75  America         0
mercury monarch                       15.0          6        250.0          72
3432          21.0   75  America         0
pontiac catalina                      16.0          8        400.0         170
4668          11.5   75  America         0
chevrolet bel air                     15.0          8        350.0         145
4440          14.0   75  America         0
plymouth grand fury                   16.0          8        318.0         150
4498          14.5   75  America         0
ford maverick                         15.0          6        250.0          72
3158          19.5   75  America         0
buick century                         17.0          6        231.0         110
3907          21.0   75  America         0
chevroelt chevelle malibu             16.0          6        250.0         105
3897          18.5   75  America         0
amc matador                           15.0          6        258.0         110
3730          19.0   75  America         0
plymouth fury                         18.0          6        225.0          95
```

3785	19.0	75	America	0			
buick skyhawk				21.0	6	231.0	110
3039	15.0	75	America	0			
ford ltd				14.0	8	351.0	148
4657	13.5	75	America	0			
ford pinto				26.5	4	140.0	72
2565	13.6	76	America	0			
pontiac ventura sj				18.5	6	250.0	110
3645	16.2	76	America	0			
amc pacer d/l				17.5	6	258.0	95
3193	17.8	76	America	0			
volkswagen rabbit				29.5	4	97.0	71
1825	12.2	76	Europe	0			
datsun b-210				32.0	4	85.0	70
1990	17.0	76	Japan	0			
toyota corolla				28.0	4	97.0	75
2155	16.4	76	Japan	0			
volvo 245				20.0	4	130.0	102
3150	15.7	76	Europe	0			
ford f108				13.0	8	302.0	130
3870	15.0	76	America	0			
peugeot 504				19.0	4	120.0	88
3270	21.9	76	Europe	0			
toyota mark ii				19.0	6	156.0	108
2930	15.5	76	Japan	0			
mercedes-benz 280s				16.5	6	168.0	120
3820	16.7	76	Europe	0			
cadillac seville				16.5	8	350.0	180
4380	12.1	76	America	0			
chevy c10				13.0	8	350.0	145
4055	12.0	76	America	0			
dodge d100				13.0	8	318.0	150
3755	14.0	76	America	0			
ford granada ghia				18.0	6	250.0	78
3574	21.0	76	America	0			
plymouth volare premier v8				13.0	8	318.0	150
3940	13.2	76	America	0			
dodge aspen se				20.0	6	225.0	100
3651	17.7	76	America	0			
vw rabbit				29.0	4	90.0	70
1937	14.2	76	Europe	0			
opel 1900				25.0	4	116.0	81
2220	16.9	76	Europe	0			
honda civic				33.0	4	91.0	53
1795	17.4	76	Japan	0			
fiat 131				28.0	4	107.0	86
2464	15.5	76	Europe	0			

capri ii				25.0	4	140.0	92
2572	14.9	76	America	0			
dodge colt				26.0	4	98.0	79
2255	17.7	76	America	0			
renault 12tl				27.0	4	101.0	83
2202	15.3	76	Europe	0			
dodge coronet brougham				16.0	8	318.0	150
4190	13.0	76	America	0			
amc matador				15.5	8	304.0	120
3962	13.9	76	America	0			
chevrolet chevelle malibu classic				17.5	8	305.0	140
4215	13.0	76	America	0			
plymouth valiant				22.0	6	225.0	100
3233	15.4	76	America	0			
chevrolet nova				22.0	6	250.0	105
3353	14.5	76	America	0			
ford maverick				24.0	6	200.0	81
3012	17.6	76	America	0			
amc hornet				22.5	6	232.0	90
3085	17.6	76	America	0			
chevrolet chevette				29.0	4	85.0	52
2035	22.2	76	America	0			
chevrolet woody				24.5	4	98.0	60
2164	22.1	76	America	0			
ford gran torino				14.5	8	351.0	152
4215	12.8	76	America	0			
ford mustang ii 2+2				25.5	4	140.0	89
2755	15.8	77	America	1			
volkswagen rabbit custom				29.0	4	97.0	78
1940	14.5	77	Europe	1			
pontiac sunbird coupe				24.5	4	151.0	88
2740	16.0	77	America	1			
toyota corolla liftback				26.0	4	97.0	75
2265	18.2	77	Japan	1			
chevrolet chevette				30.5	4	98.0	63
2051	17.0	77	America	1			
bmw 320i				21.5	4	121.0	110
2600	12.8	77	Europe	1			
subaru dl				30.0	4	97.0	67
1985	16.4	77	Japan	1			
volkswagen dasher				30.5	4	97.0	78
2190	14.1	77	Europe	1			
datsun 810				22.0	6	146.0	97
2815	14.5	77	Japan	1			
mazda rx-4				21.5	3	80.0	110
2720	13.5	77	Japan	1			
ford thunderbird				16.0	8	351.0	149

```
4335           14.5   77  America         1
dodge colt m/m                           33.5           4          98.0           83
2075           15.9   77  America         1
chrysler cordoba                         15.5           8         400.0          190
4325           12.2   77  America         1
chevrolet monte carlo landau             15.5           8         350.0          170
4165           11.4   77  America         1
plymouth arrow gs                        25.5           4         122.0           96
2300           15.5   77  America         1
buick opel isuzu deluxe                  30.0           4         111.0           80
2155           14.8   77  America         1
renault 5 gtl                            36.0           4          79.0           58
1825           18.6   77   Europe         1
datsun f-10 hatchback                    33.5           4          85.0           70
1945           16.8   77    Japan         1
pontiac grand prix lj                    16.0           8         400.0          180
4220           11.1   77  America         1
oldsmobile cutlass supreme               17.0           8         260.0          110
4060           19.0   77  America         1
chevrolet caprice classic                17.5           8         305.0          145
3880           12.5   77  America         1
mercury cougar brougham                  15.0           8         302.0          130
4295           14.9   77  America         1
chevrolet concours                       17.5           6         250.0          110
3520           16.4   77  America         1
buick skylark                            20.5           6         231.0          105
3425           16.9   77  America         1
plymouth volare custom                   19.0           6         225.0          100
3630           17.7   77  America         1
ford granada                             18.5           6         250.0           98
3525           19.0   77  America         1
dodge monaco brougham                    15.5           8         318.0          145
4140           13.7   77  America         1
honda accord cvcc                        31.5           4          98.0           68
2045           18.5   77    Japan         1
datsun 510                               27.2           4         119.0           97
2300           14.7   78    Japan         1
toyota corona                            27.5           4         134.0           95
2560           14.2   78    Japan         1
chevrolet chevette                       30.0           4          98.0           68
2155           16.5   78  America         1
buick regal sport coupe (turbo)          17.7           6         231.0          165
3445           13.4   78  America         1
ford futura                              18.1           8         302.0          139
3205           11.2   78  America         1
dodge omni                               30.9           4         105.0           75
2230           14.5   78  America         1
```

```
dodge magnum xe                       17.5          8          318.0          140
4080          13.7   78  America         1
toyota celica gt liftback             21.1          4          134.0           95
2515          14.8   78    Japan         1
peugeot 604sl                         16.2          6          163.0          133
3410          15.8   78   Europe         1
oldsmobile starfire sx                23.8          4          151.0           85
2855          17.6   78  America         1
datsun 200-sx                         23.9          4          119.0           97
2405          14.9   78    Japan         1
audi 5000                             20.3          5          131.0          103
2830          15.9   78   Europe         1
volvo 264gl                           17.0          6          163.0          125
3140          13.6   78   Europe         1
saab 99gle                            21.6          4          121.0          115
2795          15.7   78   Europe         1
volkswagen scirocco                   31.5          4           89.0           71
1990          14.9   78   Europe         1
honda accord lx                       29.5          4           98.0           68
2135          16.6   78    Japan         1
plymouth sapporo                      23.2          4          156.0          105
2745          16.7   78  America         1
chevrolet monte carlo landau          19.2          8          305.0          145
3425          13.2   78  America         1
mazda glc deluxe                      32.8          4           78.0           52
1985          19.4   78    Japan         1
dodge aspen                           18.6          6          225.0          110
3620          18.7   78  America         1
volkswagen rabbit custom diesel       43.1          4           90.0           48
1985          21.5   78   Europe         1
ford fiesta                           36.1          4           98.0           66
1800          14.4   78  America         1
datsun b210 gx                        39.4          4           85.0           70
2070          18.6   78    Japan         1
honda civic cvcc                      36.1          4           91.0           60
1800          16.4   78    Japan         1
amc concord d/l                       18.1          6          258.0          120
3410          15.1   78  America         1
dodge diplomat                        19.4          8          318.0          140
3735          13.2   78  America         1
mercury monarch ghia                  20.2          8          302.0          139
3570          12.8   78  America         1
oldsmobile cutlass salon brougham     19.9          8          260.0          110
3365          15.5   78  America         1
chevrolet malibu                      20.5          6          200.0           95
3155          18.2   78  America         1
ford fairmont (auto)                  20.2          6          200.0           85
```

2965	15.8	78	America	1
ford fairmont (man)	25.1	4	140.0	88
2720	15.4	78	America	1
plymouth volare	20.5	6	225.0	100
3430	17.2	78	America	1
amc concord	19.4	6	232.0	90
3210	17.2	78	America	1
buick century special	20.6	6	231.0	105
3380	15.8	78	America	1
mercury zephyr	20.8	6	200.0	85
3070	16.7	78	America	1
pontiac phoenix lj	19.2	6	231.0	105
3535	19.2	78	America	1
plymouth horizon	34.2	4	105.0	70
2200	13.2	79	America	1
mercedes benz 300d	25.4	5	183.0	77
3530	20.1	79	Europe	1
cadillac eldorado	23.0	8	350.0	125
3900	17.4	79	America	1
peugeot 504	27.2	4	141.0	71
3190	24.8	79	Europe	1
oldsmobile cutlass salon brougham	23.9	8	260.0	90
3420	22.2	79	America	1
plymouth horizon tc3	34.5	4	105.0	70
2150	14.9	79	America	1
amc spirit dl	27.4	4	121.0	80
2670	15.0	79	America	1
fiat strada custom	37.3	4	91.0	69
2130	14.7	79	Europe	1
buick skylark limited	28.4	4	151.0	90
2670	16.0	79	America	1
chevrolet citation	28.8	6	173.0	115
2595	11.3	79	America	1
oldsmobile omega brougham	26.8	6	173.0	115
2700	12.9	79	America	1
pontiac phoenix	33.5	4	151.0	90
2556	13.2	79	America	1
datsun 210	31.8	4	85.0	65
2020	19.2	79	Japan	1
dodge colt hatchback custom	35.7	4	98.0	80
1915	14.4	79	America	1
dodge st. regis	18.2	8	318.0	135
3830	15.2	79	America	1
vw rabbit custom	31.9	4	89.0	71
1925	14.0	79	Europe	1
mercury zephyr 6	19.8	6	200.0	85
2990	18.2	79	America	1

ford fairmont 4	22.3	4	140.0	88
2890 17.3 79 America	1			
amc concord dl 6	20.2	6	232.0	90
3265 18.2 79 America	1			
dodge aspen 6	20.6	6	225.0	110
3360 16.6 79 America	1			
chevrolet caprice classic	17.0	8	305.0	130
3840 15.4 79 America	1			
ford ltd landau	17.6	8	302.0	129
3725 13.4 79 America	1			
pontiac lemans v6	21.5	6	231.0	115
3245 15.4 79 America	1			
maxda glc deluxe	34.1	4	86.0	65
1975 15.2 79 Japan	1			
buick estate wagon (sw)	16.9	8	350.0	155
4360 14.9 79 America	1			
ford country squire (sw)	15.5	8	351.0	142
4054 14.3 79 America	1			
chevrolet malibu classic (sw)	19.2	8	267.0	125
3605 15.0 79 America	1			
chrysler lebaron town @ country (sw)	18.5	8	360.0	150
3940 13.0 79 America	1			
mercury grand marquis	16.5	8	351.0	138
3955 13.2 79 America	1			
vw rabbit c (diesel)	44.3	4	90.0	48
2085 21.7 80 Europe	1			
vw dasher (diesel)	43.4	4	90.0	48
2335 23.7 80 Europe	1			
audi 5000s (diesel)	36.4	5	121.0	67
2950 19.9 80 Europe	1			
mercedes-benz 240d	30.0	4	146.0	67
3250 21.8 80 Europe	1			
honda civic 1500 gl	44.6	4	91.0	67
1850 13.8 80 Japan	1			
datsun 280-zx	32.7	6	168.0	132
2910 11.4 80 Japan	1			
vokswagen rabbit	29.8	4	89.0	62
1845 15.3 80 Europe	1			
mazda rx-7 gs	23.7	3	70.0	100
2420 12.5 80 Japan	1			
triumph tr7 coupe	35.0	4	122.0	88
2500 15.1 80 Europe	1			
honda accord	32.4	4	107.0	72
2290 17.0 80 Japan	1			
datsun 210	40.8	4	85.0	65
2110 19.2 80 Japan	1			
subaru dl	33.8	4	97.0	67

```
2145          18.0   80    Japan         1
dodge colt                              27.9          4         156.0          105
2800          14.4   80  America         1
mazda glc                               46.6          4          86.0           65
2110          17.9   80    Japan         1
toyota corolla                          32.2          4         108.0           75
2265          15.2   80    Japan         1
vw rabbit                               41.5          4          98.0           76
2144          14.7   80   Europe         1
toyota corolla tercel                   38.1          4          89.0           60
1968          18.8   80    Japan         1
chevrolet chevette                      32.1          4          98.0           70
2120          15.5   80  America         1
chevrolet citation                      28.0          4         151.0           90
2678          16.5   80  America         1
ford fairmont                           26.4          4         140.0           88
2870          18.1   80  America         1
datsun 310                              37.2          4          86.0           65
2019          16.4   80    Japan         1
dodge aspen                             19.1          6         225.0           90
3381          18.7   80  America         1
audi 4000                               34.3          4          97.0           78
2188          15.8   80   Europe         1
toyota corona liftback                  29.8          4         134.0           90
2711          15.5   80    Japan         1
mazda 626                               31.3          4         120.0           75
2542          17.5   80    Japan         1
datsun 510 hatchback                    37.0          4         119.0           92
2434          15.0   80    Japan         1
amc concord                             24.3          4         151.0           90
3003          20.1   80  America         1
peugeot 505s turbo diesel               28.1          4         141.0           80
3230          20.4   81   Europe         1
honda prelude                           33.7          4         107.0           75
2210          14.4   81    Japan         1
toyota corolla                          32.4          4         108.0           75
2350          16.8   81    Japan         1
datsun 200sx                            32.9          4         119.0          100
2615          14.8   81    Japan         1
mazda 626                               31.6          4         120.0           74
2635          18.3   81    Japan         1
volvo diesel                            30.7          6         145.0           76
3160          19.6   81   Europe         1
chrysler lebaron salon                  17.6          6         225.0           85
3465          16.6   81  America         1
datsun 810 maxima                       24.2          6         146.0          120
2930          13.8   81    Japan         1
```

buick century				22.4	6	231.0	110
3415	15.8	81	America	1			
oldsmobile cutlass ls				26.6	8	350.0	105
3725	19.0	81	America	1			
ford granada gl				20.2	6	200.0	88
3060	17.1	81	America	1			
volkswagen jetta				33.0	4	105.0	74
2190	14.2	81	Europe	1			
toyota cressida				25.4	6	168.0	116
2900	12.6	81	Japan	1			
ford escort 2h				29.9	4	98.0	65
2380	20.7	81	America	1			
plymouth reliant				27.2	4	135.0	84
2490	15.7	81	America	1			
plymouth horizon 4				34.7	4	105.0	63
2215	14.9	81	America	1			
ford escort 4w				34.4	4	98.0	65
2045	16.2	81	America	1			
buick skylark				26.6	4	151.0	84
2635	16.4	81	America	1			
dodge aries wagon (sw)				25.8	4	156.0	92
2620	14.4	81	America	1			
plymouth reliant				30.0	4	135.0	84
2385	12.9	81	America	1			
toyota starlet				39.1	4	79.0	58
1755	16.9	81	Japan	1			
chevrolet citation				23.5	6	173.0	110
2725	12.6	81	America	1			
honda civic 1300				35.1	4	81.0	60
1760	16.1	81	Japan	1			
subaru				32.3	4	97.0	67
2065	17.8	81	Japan	1			
datsun 210 mpg				37.0	4	85.0	65
1975	19.4	81	Japan	1			
toyota tercel				37.7	4	89.0	62
2050	17.3	81	Japan	1			
mazda glc 4				34.1	4	91.0	68
1985	16.0	81	Japan	1			
plymouth champ				39.0	4	86.0	64
1875	16.4	81	America	1			
chrysler lebaron medallion				26.0	4	156.0	92
2585	14.5	82	America	1			
honda civic (auto)				32.0	4	91.0	67
1965	15.7	82	Japan	1			
datsun 310 gx				38.0	4	91.0	67
1995	16.2	82	Japan	1			
buick century limited				25.0	6	181.0	110

2945	16.4	82	America	1
oldsmobile cutlass ciera (diesel)	38.0	6	262.0	85
3015	17.0	82	America	1
ford granada l	22.0	6	232.0	112
2835	14.7	82	America	1
dodge rampage	32.0	4	135.0	84
2295	11.6	82	America	1
dodge charger 2.2	36.0	4	135.0	84
2370	13.0	82	America	1
chevrolet camaro	27.0	4	151.0	90
2950	17.3	82	America	1
ford mustang gl	27.0	4	140.0	86
2790	15.6	82	America	1
vw pickup	44.0	4	97.0	52
2130	24.6	82	Europe	1
honda civic	38.0	4	91.0	67
1965	15.0	82	Japan	1
toyota celica gt	32.0	4	144.0	96
2665	13.9	82	Japan	1
toyota corolla	34.0	4	108.0	70
2245	16.9	82	Japan	1
ford ranger	28.0	4	120.0	79
2625	18.6	82	America	1
nissan stanza xe	36.0	4	120.0	88
2160	14.5	82	Japan	1
mercury lynx l	36.0	4	98.0	70
2125	17.3	82	America	1
plymouth horizon miser	38.0	4	105.0	63
2125	14.7	82	America	1
mazda glc custom	31.0	4	91.0	68
1970	17.6	82	Japan	1
mazda glc custom l	37.0	4	91.0	68
2025	18.2	82	Japan	1
volkswagen rabbit l	36.0	4	105.0	74
1980	15.3	82	Europe	1
ford fairmont futura	24.0	4	140.0	92
2865	16.4	82	America	1
pontiac phoenix	27.0	4	151.0	90
2735	18.0	82	America	1
dodge aries se	29.0	4	135.0	84
2525	16.0	82	America	1
pontiac j2000 se hatchback	31.0	4	112.0	85
2575	16.2	82	America	1
chevrolet cavalier 2-door	34.0	4	112.0	88
2395	18.0	82	America	1
chevrolet cavalier wagon	27.0	4	112.0	88
2640	18.6	82	America	1

```
chevrolet cavalier                        28.0          4         112.0           88
2605          19.6    82   America           1
honda accord                              36.0          4         107.0           75
2205          14.5    82     Japan           1
chevy s-10                                31.0          4         119.0           82
2720          19.4    82   America           1
```

[51]:

```
Auto_log["log_displacement"] = np.log(Auto_log["displacement"])
Auto_log["log_horsepower"] = np.log(Auto_log["horsepower"])
Auto_log["log_weight"] = np.log(Auto_log["weight"])
Auto_log = Auto_log.drop(
    columns=[
        "displacement",
        "weight",
        "horsepower",
        "year",
    ]
)
Auto_log.columns
```

[51]:

```
Index(['mpg', 'cylinders', 'acceleration', 'origin', 'oilshock',
'log_displacement', 'log_horsepower', 'log_weight'], dtype='object')
```

[52]:

```
Auto_log.corr(numeric_only=True)
```

[52]:

```
                       mpg  cylinders  acceleration  oilshock  log_displacement
log_horsepower  log_weight
mpg               1.000000  -0.777618      0.423329  0.521192         -0.828453
-0.817517   -0.844194
cylinders        -0.777618   1.000000     -0.504683 -0.273703          0.942814
0.843204    0.884303
acceleration      0.423329  -0.504683      1.000000  0.195892         -0.497107
-0.698328   -0.401563
oilshock          0.521192  -0.273703      0.195892  1.000000         -0.268161
-0.299037   -0.250520
log_displacement -0.828453   0.942814     -0.497107 -0.268161          1.000000
0.872149    0.942850
log_horsepower   -0.817517   0.843204     -0.698328 -0.299037          0.872149
1.000000    0.873956
log_weight       -0.844194   0.884303     -0.401563 -0.250520          0.942850
0.873956    1.000000
```

[53]:

```
Auto_log = pd.get_dummies(
    Auto_log, columns=list(["origin"]), drop_first=True, dtype=np.uint8
)
Auto_log.columns
```

[53]:
```
Index(['mpg', 'cylinders', 'acceleration', 'oilshock', 'log_displacement',
'log_horsepower', 'log_weight', 'origin_Europe', 'origin_Japan'],
dtype='object')
```

[54]:
```
cols = list(Auto_log.columns)
cols.remove("mpg")
```

[55]:
```
vifdf = calculate_VIFs("mpg ~ " + " + ".join(cols), Auto_log)
vifdf
```

[55]:

Feature	VIF
cylinders	9.828626
acceleration	3.304749
oilshock	1.147770
log_displacement	25.969595
log_horsepower	11.414709
log_weight	16.146573
origin_Europe	1.876698
origin_Japan	2.097688

[56]:
```
identify_highest_VIF_feature(vifdf)
```

```
We find the highest VIF in this model is log_displacement with a VIF of
25.96959512578754
Hence, we drop log_displacement from the model to be fitted.
```

[56]:
```
('log_displacement', 25.96959512578754)
```

[57]:
```
cols.remove("log_displacement")
vifdf = calculate_VIFs("mpg ~ " + " + ".join(cols), Auto_log)
vifdf
```

[57]:

Feature	VIF
cylinders	5.535070
acceleration	3.179336
oilshock	1.142791
log_horsepower	11.411764
log_weight	10.608718
origin_Europe	1.451961
origin_Japan	1.652749

[58]:
```
identify_highest_VIF_feature(vifdf)
```

```
We find the highest VIF in this model is log_horsepower with a VIF of
11.411764499222897
Hence, we drop log_horsepower from the model to be fitted.
```

[58]: ('log_horsepower', 11.411764499222897)

[59]:
```
cols.remove("log_horsepower")
vifdf = calculate_VIFs("mpg ~ " + " + ".join(cols), Auto_log)
vifdf
```

[59]:

	VIF
Feature	
cylinders	5.517868
acceleration	1.377517
oilshock	1.118666
log_weight	5.014899
origin_Europe	1.451265
origin_Japan	1.608682

[60]:
```
identify_highest_VIF_feature(vifdf)
```

```
No variables are significantly collinear.
```

[61]:
```
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_log)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.823
Model:                            OLS   Adj. R-squared:                  0.821
Method:                 Least Squares   F-statistic:                     299.3
Date:                Fri, 21 Feb 2025   Prob (F-statistic):         1.36e-141
Time:                        19:20:24   Log-Likelihood:                -1021.3
No. Observations:                 392   AIC:                             2057.
Df Residuals:                     385   BIC:                             2084.
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
=
                    coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
Intercept       168.2416      9.696     17.351      0.000     149.177
187.306
cylinders         0.0129      0.230      0.056      0.955      -0.440
0.465
acceleration      0.1805      0.071      2.538      0.012       0.041
0.320
oilshock          5.1312      0.355     14.470      0.000       4.434
5.828
log_weight      -18.9156      1.331    -14.211      0.000     -21.533
-16.299
```

```
origin_Europe     1.3692      0.531      2.578      0.010      0.325
2.413
origin_Japan      1.5602      0.528      2.956      0.003      0.522
2.598
==============================================================================
Omnibus:                       30.158   Durbin-Watson:                   1.253
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               51.811
Skew:                           0.493   Prob(JB):                     5.62e-12
Kurtosis:                       4.484   Cond. No.                     1.08e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.08e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
                  df        sum_sq       mean_sq            F         PR(>F)
cylinders        1.0  14403.083079  14403.083079  1318.559127  2.133814e-126
acceleration     1.0     30.471304     30.471304     2.789557   9.569322e-02
oilshock         1.0   2422.051542   2422.051542   221.731566   6.308582e-40
log_weight       1.0   2639.878573   2639.878573   241.672978   1.215817e-42
origin_Europe    1.0     22.569818     22.569818     2.066199   1.514086e-01
origin_Japan     1.0     95.449335     95.449335     8.738101   3.308129e-03
Residual       385.0   4205.489819     10.923350          NaN            NaN
```

[61]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b6c2d50>

[62]:
```
identify_least_significant_feature(results, alpha=LOS_Alpha)
```

```
We find the least significant variable in this model is cylinders with a p-value
of 0.9554479997905154 and a coefficient of 0.012867295201156259
Using the backward methodology, we suggest dropping cylinders from the new model
```

[63]:
```
cols.remove("cylinders")
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_log)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.823
Model:                            OLS   Adj. R-squared:                  0.821
Method:                 Least Squares   F-statistic:                     360.0
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          6.83e-143
Time:                        19:20:25   Log-Likelihood:                -1021.3
No. Observations:                 392   AIC:                             2055.
Df Residuals:                     386   BIC:                             2078.
Df Model:                           5
Covariance Type:            nonrobust
===============================================================================
```

```
=
                    coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
Intercept       167.8689      7.032     23.870      0.000     154.042
181.696
acceleration      0.1792      0.067      2.673      0.008       0.047
0.311
oilshock          5.1289      0.352     14.574      0.000       4.437
5.821
log_weight      -18.8571      0.820    -22.992      0.000     -20.470
-17.245
origin_Europe     1.3628      0.518      2.631      0.009       0.344
2.381
origin_Japan      1.5576      0.525      2.967      0.003       0.525
2.590
==============================================================================
Omnibus:                       30.308   Durbin-Watson:                   1.253
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               52.282
Skew:                           0.493   Prob(JB):                     4.44e-12
Kurtosis:                       4.492   Cond. No.                         751.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                  df        sum_sq       mean_sq           F         PR(>F)
acceleration     1.0   4268.531557   4268.531557  391.783092   1.076057e-60
oilshock         1.0   4757.627552   4757.627552  436.674301   2.076230e-65
log_weight       1.0  10466.602734  10466.602734  960.667136  8.818191e-107
origin_Europe    1.0     24.823504     24.823504    2.278402   1.320051e-01
origin_Japan     1.0     95.884166     95.884166    8.800637   3.198611e-03
Residual       386.0   4205.523957     10.895140         NaN            NaN
```

[63]:
```
<statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b6c1a30>
```

[64]:
```python
identify_least_significant_feature(results, alpha=LOS_Alpha)
```

```
No variables are statistically insignificant.
The model mpg ~ acceleration + oilshock + log_weight + origin_Europe +
origin_Japan cannot be pruned further.
```

[65]:
```python
models.append(
    {
        "name": "log_transformation",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
```

```
    }
)
```

[66]:

```
pd.DataFrame(models)
```

[66]:

```
                     name
model  R-squared adjusted
0              simple_model
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan
0.802074
1    numeric_interactions
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan +
horsepower: weight            0.844846
2   oilshock_interactions                    mpg ~ horsepower + weight + oilshock +
origin_Europe + origin_Japan + horsepower: weight + oilshock: weight + oilshock:
horsepower             0.858491
3     origin_interactions  mpg ~ horsepower + weight + oilshock + origin_Europe
+ origin_Japan + oilshock: horsepower + origin_Europe: horsepower +
origin_Japan: horsepower            0.849008
4   squared_transformation                                            mpg ~
horsepower + weight + oilshock + origin_Europe + origin_Japan + I(horsepower**2)
+ I(weight**2)            0.845550
5      log_transformation
mpg ~ acceleration + oilshock + log_weight + origin_Europe + origin_Japan
0.821151
```

1.9 Square Root Transformed Model

[67]:

```
Auto_sqrt = Auto.copy(deep=True)
```

[67]:

```
                                          mpg  cylinders  displacement  horsepower
weight  acceleration year   origin  oilshock
name
chevrolet chevelle malibu              18.0          8         307.0         130
3504          12.0    70  America        0
hi 1200d                                9.0          8         304.0         193
4732          18.5    70  America        0
dodge d200                             11.0          8         318.0         210
4382          13.5    70  America        0
chevy c20                              10.0          8         307.0         200
4376          15.0    70  America        0
ford f250                              10.0          8         360.0         215
4615          14.0    70  America        0
amc gremlin                            21.0          6         199.0          90
2648          15.0    70  America        0
bmw 2002                               26.0          4         121.0         113
2234          12.5    70   Europe        0
```

```
saab 99e                               25.0          4        104.0           95
2375          17.5   70   Europe        0
audi 100 ls                            24.0          4        107.0           90
2430          14.5   70   Europe        0
volkswagen 1131 deluxe sedan           26.0          4         97.0           46
1835          20.5   70   Europe        0
datsun pl510                           27.0          4         97.0           88
2130          14.5   70    Japan        0
ford maverick                          21.0          6        200.0           85
2587          16.0   70  America        0
amc hornet                             18.0          6        199.0           97
2774          15.5   70  America        0
plymouth duster                        22.0          6        198.0           95
2833          15.5   70  America        0
peugeot 504                            25.0          4        110.0           87
2672          17.5   70   Europe        0
buick estate wagon (sw)                14.0          8        455.0          225
3086          10.0   70  America        0
toyota corona mark ii                  24.0          4        113.0           95
2372          15.0   70    Japan        0
plymouth satellite                     18.0          8        318.0          150
3436          11.0   70  America        0
amc rebel sst                          16.0          8        304.0          150
3433          12.0   70  America        0
ford torino                            17.0          8        302.0          140
3449          10.5   70  America        0
ford galaxie 500                       15.0          8        429.0          198
4341          10.0   70  America        0
chevrolet impala                       14.0          8        454.0          220
4354           9.0   70  America        0
buick skylark 320                      15.0          8        350.0          165
3693          11.5   70  America        0
pontiac catalina                       14.0          8        455.0          225
4425          10.0   70  America        0
amc ambassador dpl                     15.0          8        390.0          190
3850           8.5   70  America        0
dodge challenger se                    15.0          8        383.0          170
3563          10.0   70  America        0
plymouth 'cuda 340                     14.0          8        340.0          160
3609           8.0   70  America        0
chevrolet monte carlo                  15.0          8        400.0          150
3761           9.5   70  America        0
plymouth fury iii                      14.0          8        440.0          215
4312           8.5   70  America        0
amc hornet sportabout (sw)             18.0          6        258.0          110
2962          13.5   71  America        0
chevrolet vega (sw)                    22.0          4        140.0           72
```

```
2408          19.0   71  America        0
pontiac firebird                     19.0          6         250.0          100
3282          15.0   71  America        0
ford mustang                         18.0          6         250.0           88
3139          14.5   71  America        0
mercury capri 2000                   23.0          4         122.0           86
2220          14.0   71  America        0
toyota corolla 1200                  31.0          4          71.0           65
1773          19.0   71    Japan        0
peugeot 304                          30.0          4          79.0           70
2074          19.5   71   Europe        0
datsun 1200                          35.0          4          72.0           69
1613          18.0   71    Japan        0
volkswagen model 111                 27.0          4          97.0           60
1834          19.0   71   Europe        0
plymouth cricket                     26.0          4          91.0           70
1955          20.5   71  America        0
pontiac safari (sw)                  13.0          8         400.0          175
5140          12.0   71  America        0
opel 1900                            28.0          4         116.0           90
2123          14.0   71   Europe        0
ford country squire (sw)             13.0          8         400.0          170
4746          12.0   71  America        0
fiat 124b                            30.0          4          88.0           76
2065          14.5   71   Europe        0
plymouth fury iii                    14.0          8         318.0          150
4096          13.0   71  America        0
dodge monaco (sw)                    12.0          8         383.0          180
4955          11.5   71  America        0
chevrolet vega 2300                  28.0          4         140.0           90
2264          15.5   71  America        0
toyota corona                        25.0          4         113.0           95
2228          14.0   71    Japan        0
amc gremlin                          19.0          6         232.0          100
2634          13.0   71  America        0
plymouth satellite custom            16.0          6         225.0          105
3439          15.5   71  America        0
datsun pl510                         27.0          4          97.0           88
2130          14.5   71    Japan        0
ford torino 500                      19.0          6         250.0           88
3302          15.5   71  America        0
amc matador                          18.0          6         232.0          100
3288          15.5   71  America        0
chevrolet impala                     14.0          8         350.0          165
4209          12.0   71  America        0
pontiac catalina brougham            14.0          8         400.0          175
4464          11.5   71  America        0
```

```
ford galaxie 500                    14.0          8         351.0          153
4154          13.5   71  America       0
chevrolet chevelle malibu           17.0          6         250.0          100
3329          15.5   71  America       0
chevrolet chevelle concours (sw)    13.0          8         307.0          130
4098          14.0   72  America       0
plymouth satellite custom (sw)      14.0          8         318.0          150
4077          14.0   72  America       0
volvo 145e (sw)                     18.0          4         121.0          112
2933          14.5   72   Europe       0
volkswagen 411 (sw)                 22.0          4         121.0           76
2511          18.0   72   Europe       0
peugeot 504 (sw)                    21.0          4         120.0           87
2979          19.5   72   Europe       0
ford pinto (sw)                     22.0          4         122.0           86
2395          16.0   72  America       0
datsun 510 (sw)                     28.0          4          97.0           92
2288          17.0   72    Japan       0
toyouta corona mark ii (sw)         23.0          4         120.0           97
2506          14.5   72    Japan       0
dodge colt (sw)                     28.0          4          98.0           80
2164          15.0   72  America       0
amc matador (sw)                    15.0          8         304.0          150
3892          12.5   72  America       0
toyota corolla 1600 (sw)            27.0          4          97.0           88
2100          16.5   72    Japan       0
renault 12 (sw)                     26.0          4          96.0           69
2189          18.0   72   Europe       0
mazda rx2 coupe                     19.0          3          70.0           97
2330          13.5   72    Japan       0
ford gran torino (sw)               13.0          8         302.0          140
4294          16.0   72  America       0
oldsmobile delta 88 royale          12.0          8         350.0          160
4456          13.5   72  America       0
chrysler newport royal              13.0          8         400.0          190
4422          12.5   72  America       0
toyota corona hardtop               24.0          4         113.0           95
2278          15.5   72    Japan       0
volkswagen type 3                   23.0          4          97.0           54
2254          23.5   72   Europe       0
chevrolet vega                      20.0          4         140.0           90
2408          19.5   72  America       0
ford pinto runabout                 21.0          4         122.0           86
2226          16.5   72  America       0
chevrolet impala                    13.0          8         350.0          165
4274          12.0   72  America       0
dodge colt hardtop                  25.0          4          97.5           80
```

```
2126          17.0   72  America         0
plymouth fury iii                     15.0          8         318.0          150
4135          13.5   72  America         0
ford galaxie 500                      14.0          8         351.0          153
4129          13.0   72  America         0
amc ambassador sst                    17.0          8         304.0          150
3672          11.5   72  America         0
mercury marquis                       11.0          8         429.0          208
4633          11.0   72  America         0
buick lesabre custom                  13.0          8         350.0          155
4502          13.5   72  America         0
pontiac catalina                      14.0          8         400.0          175
4385          12.0   72  America         0
fiat 124 sport coupe                  26.0          4          98.0           90
2265          15.5   73   Europe         0
amc gremlin                           18.0          6         232.0          100
2789          15.0   73  America         0
toyota carina                         20.0          4          97.0           88
2279          19.0   73    Japan         0
chevrolet vega                        21.0          4         140.0           72
2401          19.5   73  America         0
datsun 610                            22.0          4         108.0           94
2379          16.5   73    Japan         0
maxda rx3                             18.0          3          70.0           90
2124          13.5   73    Japan         0
ford pinto                            19.0          4         122.0           85
2310          18.5   73  America         0
mercury capri v6                      21.0          6         155.0          107
2472          14.0   73  America         0
chevrolet monte carlo s               15.0          8         350.0          145
4082          13.0   73  America         0
saab 99le                             24.0          4         121.0          110
2660          14.0   73   Europe         0
fiat 128                              29.0          4          68.0           49
1867          19.5   73   Europe         0
opel manta                            24.0          4         116.0           75
2158          15.5   73   Europe         0
audi 100ls                            20.0          4         114.0           91
2582          14.0   73   Europe         0
volvo 144ea                           19.0          4         121.0          112
2868          15.5   73   Europe         0
dodge dart custom                     15.0          8         318.0          150
3399          11.0   73  America         0
toyota mark ii                        20.0          6         156.0          122
2807          13.5   73    Japan         0
oldsmobile omega                      11.0          8         350.0          180
3664          11.0   73  America         0
```

```
oldsmobile vista cruiser            12.0          8          350.0          180
4499          12.5   73  America        0
pontiac grand prix                  16.0          8          400.0          230
4278           9.5   73  America        0
plymouth custom suburb              13.0          8          360.0          170
4654          13.0   73  America        0
amc hornet                          18.0          6          232.0          100
2945          16.0   73  America        0
chevrolet impala                    11.0          8          400.0          150
4997          14.0   73  America        0
buick century 350                   13.0          8          350.0          175
4100          13.0   73  America        0
amc matador                         14.0          8          304.0          150
3672          11.5   73  America        0
chevrolet malibu                    13.0          8          350.0          145
3988          13.0   73  America        0
dodge coronet custom                15.0          8          318.0          150
3777          12.5   73  America        0
mercury marquis brougham            12.0          8          429.0          198
4952          11.5   73  America        0
chevrolet caprice classic           13.0          8          400.0          150
4464          12.0   73  America        0
ford ltd                            13.0          8          351.0          158
4363          13.0   73  America        0
plymouth fury gran sedan            14.0          8          318.0          150
4237          14.5   73  America        0
ford gran torino                    14.0          8          302.0          137
4042          14.5   73  America        0
buick electra 225 custom            12.0          8          455.0          225
4951          11.0   73  America        0
amc ambassador brougham             13.0          8          360.0          175
3821          11.0   73  America        0
plymouth valiant                    18.0          6          225.0          105
3121          16.5   73  America        0
chevrolet nova custom               16.0          6          250.0          100
3278          18.0   73  America        0
ford country                        12.0          8          400.0          167
4906          12.5   73  America        0
ford maverick                       18.0          6          250.0           88
3021          16.5   73  America        0
plymouth duster                     23.0          6          198.0           95
2904          16.0   73  America        0
volkswagen super beetle             26.0          4           97.0           46
1950          21.0   73   Europe        0
chrysler new yorker brougham        13.0          8          440.0          215
4735          11.0   73  America        0
audi fox                            29.0          4           98.0           83
```

2219	16.5	74	Europe	0
volkswagen dasher	26.0	4	79.0	67
1963	15.5	74	Europe	0
opel manta	26.0	4	97.0	78
2300	14.5	74	Europe	0
toyota corona	31.0	4	76.0	52
1649	16.5	74	Japan	0
datsun 710	32.0	4	83.0	61
2003	19.0	74	Japan	0
subaru	26.0	4	108.0	93
2391	15.5	74	Japan	0
fiat 128	24.0	4	90.0	75
2108	15.5	74	Europe	0
fiat 124 tc	26.0	4	116.0	75
2246	14.0	74	Europe	0
honda civic	24.0	4	120.0	97
2489	15.0	74	Japan	0
fiat x1.9	31.0	4	79.0	67
2000	16.0	74	Europe	0
amc matador (sw)	14.0	8	304.0	150
4257	15.5	74	America	0
dodge colt	28.0	4	90.0	75
2125	14.5	74	America	0
ford gran torino (sw)	14.0	8	302.0	140
4638	16.0	74	America	0
ford gran torino	16.0	8	302.0	140
4141	14.0	74	America	0
buick century luxus (sw)	13.0	8	350.0	150
4699	14.5	74	America	0
dodge coronet custom (sw)	14.0	8	318.0	150
4457	13.5	74	America	0
plymouth duster	20.0	6	198.0	95
3102	16.5	74	America	0
amc hornet	19.0	6	232.0	100
2901	16.0	74	America	0
chevrolet nova	15.0	6	250.0	100
3336	17.0	74	America	0
ford pinto	26.0	4	122.0	80
2451	16.5	74	America	0
datsun b210	31.0	4	79.0	67
1950	19.0	74	Japan	0
chevrolet vega	25.0	4	140.0	75
2542	17.0	74	America	0
chevrolet chevelle malibu classic	16.0	6	250.0	100
3781	17.0	74	America	0
amc matador	16.0	6	258.0	110
3632	18.0	74	America	0

plymouth satellite sebring	18.0	6	225.0	105
3613 16.5 74 America	0			
toyota corolla 1200	32.0	4	71.0	65
1836 21.0 74 Japan	0			
datsun 710	24.0	4	119.0	97
2545 17.0 75 Japan	0			
pontiac astro	23.0	4	140.0	78
2592 18.5 75 America	0			
amc gremlin	20.0	6	232.0	100
2914 16.0 75 America	0			
toyota corona	24.0	4	134.0	96
2702 13.5 75 Japan	0			
volkswagen dasher	25.0	4	90.0	71
2223 16.5 75 Europe	0			
ford pinto	18.0	6	171.0	97
2984 14.5 75 America	0			
saab 99le	25.0	4	121.0	115
2671 13.5 75 Europe	0			
amc pacer	19.0	6	232.0	90
3211 17.0 75 America	0			
audi 100ls	23.0	4	115.0	95
2694 15.0 75 Europe	0			
peugeot 504	23.0	4	120.0	88
2957 17.0 75 Europe	0			
volvo 244dl	22.0	4	121.0	98
2945 14.5 75 Europe	0			
honda civic cvcc	33.0	4	91.0	53
1795 17.5 75 Japan	0			
ford pinto	23.0	4	140.0	83
2639 17.0 75 America	0			
volkswagen rabbit	29.0	4	90.0	70
1937 14.0 75 Europe	0			
toyota corolla	29.0	4	97.0	75
2171 16.0 75 Japan	0			
plymouth valiant custom	19.0	6	225.0	95
3264 16.0 75 America	0			
chevrolet monza 2+2	20.0	8	262.0	110
3221 13.5 75 America	0			
ford mustang ii	13.0	8	302.0	129
3169 12.0 75 America	0			
chevrolet nova	18.0	6	250.0	105
3459 16.0 75 America	0			
mercury monarch	15.0	6	250.0	72
3432 21.0 75 America	0			
pontiac catalina	16.0	8	400.0	170
4668 11.5 75 America	0			
chevrolet bel air	15.0	8	350.0	145

4440	14.0	75	America	0
plymouth grand fury	16.0	8	318.0	150
4498	14.5	75	America	0
ford maverick	15.0	6	250.0	72
3158	19.5	75	America	0
buick century	17.0	6	231.0	110
3907	21.0	75	America	0
chevroelt chevelle malibu	16.0	6	250.0	105
3897	18.5	75	America	0
amc matador	15.0	6	258.0	110
3730	19.0	75	America	0
plymouth fury	18.0	6	225.0	95
3785	19.0	75	America	0
buick skyhawk	21.0	6	231.0	110
3039	15.0	75	America	0
ford ltd	14.0	8	351.0	148
4657	13.5	75	America	0
ford pinto	26.5	4	140.0	72
2565	13.6	76	America	0
pontiac ventura sj	18.5	6	250.0	110
3645	16.2	76	America	0
amc pacer d/l	17.5	6	258.0	95
3193	17.8	76	America	0
volkswagen rabbit	29.5	4	97.0	71
1825	12.2	76	Europe	0
datsun b-210	32.0	4	85.0	70
1990	17.0	76	Japan	0
toyota corolla	28.0	4	97.0	75
2155	16.4	76	Japan	0
volvo 245	20.0	4	130.0	102
3150	15.7	76	Europe	0
ford f108	13.0	8	302.0	130
3870	15.0	76	America	0
peugeot 504	19.0	4	120.0	88
3270	21.9	76	Europe	0
toyota mark ii	19.0	6	156.0	108
2930	15.5	76	Japan	0
mercedes-benz 280s	16.5	6	168.0	120
3820	16.7	76	Europe	0
cadillac seville	16.5	8	350.0	180
4380	12.1	76	America	0
chevy c10	13.0	8	350.0	145
4055	12.0	76	America	0
dodge d100	13.0	8	318.0	150
3755	14.0	76	America	0
ford granada ghia	18.0	6	250.0	78
3574	21.0	76	America	0

```
plymouth volare premier v8          13.0           8          318.0          150
3940         13.2   76  America        0
dodge aspen se                      20.0           6          225.0          100
3651         17.7   76  America        0
vw rabbit                           29.0           4           90.0           70
1937         14.2   76   Europe        0
opel 1900                           25.0           4          116.0           81
2220         16.9   76   Europe        0
honda civic                         33.0           4           91.0           53
1795         17.4   76    Japan        0
fiat 131                            28.0           4          107.0           86
2464         15.5   76   Europe        0
capri ii                            25.0           4          140.0           92
2572         14.9   76  America        0
dodge colt                          26.0           4           98.0           79
2255         17.7   76  America        0
renault 12tl                        27.0           4          101.0           83
2202         15.3   76   Europe        0
dodge coronet brougham              16.0           8          318.0          150
4190         13.0   76  America        0
amc matador                         15.5           8          304.0          120
3962         13.9   76  America        0
chevrolet chevelle malibu classic   17.5           8          305.0          140
4215         13.0   76  America        0
plymouth valiant                    22.0           6          225.0          100
3233         15.4   76  America        0
chevrolet nova                      22.0           6          250.0          105
3353         14.5   76  America        0
ford maverick                       24.0           6          200.0           81
3012         17.6   76  America        0
amc hornet                          22.5           6          232.0           90
3085         17.6   76  America        0
chevrolet chevette                  29.0           4           85.0           52
2035         22.2   76  America        0
chevrolet woody                     24.5           4           98.0           60
2164         22.1   76  America        0
ford gran torino                    14.5           8          351.0          152
4215         12.8   76  America        0
ford mustang ii 2+2                 25.5           4          140.0           89
2755         15.8   77  America        1
volkswagen rabbit custom            29.0           4           97.0           78
1940         14.5   77   Europe        1
pontiac sunbird coupe               24.5           4          151.0           88
2740         16.0   77  America        1
toyota corolla liftback             26.0           4           97.0           75
2265         18.2   77    Japan        1
chevrolet chevette                  30.5           4           98.0           63
```

2051	17.0	77	America	1			
bmw 320i				21.5	4	121.0	110
2600	12.8	77	Europe	1			
subaru dl				30.0	4	97.0	67
1985	16.4	77	Japan	1			
volkswagen dasher				30.5	4	97.0	78
2190	14.1	77	Europe	1			
datsun 810				22.0	6	146.0	97
2815	14.5	77	Japan	1			
mazda rx-4				21.5	3	80.0	110
2720	13.5	77	Japan	1			
ford thunderbird				16.0	8	351.0	149
4335	14.5	77	America	1			
dodge colt m/m				33.5	4	98.0	83
2075	15.9	77	America	1			
chrysler cordoba				15.5	8	400.0	190
4325	12.2	77	America	1			
chevrolet monte carlo landau				15.5	8	350.0	170
4165	11.4	77	America	1			
plymouth arrow gs				25.5	4	122.0	96
2300	15.5	77	America	1			
buick opel isuzu deluxe				30.0	4	111.0	80
2155	14.8	77	America	1			
renault 5 gtl				36.0	4	79.0	58
1825	18.6	77	Europe	1			
datsun f-10 hatchback				33.5	4	85.0	70
1945	16.8	77	Japan	1			
pontiac grand prix lj				16.0	8	400.0	180
4220	11.1	77	America	1			
oldsmobile cutlass supreme				17.0	8	260.0	110
4060	19.0	77	America	1			
chevrolet caprice classic				17.5	8	305.0	145
3880	12.5	77	America	1			
mercury cougar brougham				15.0	8	302.0	130
4295	14.9	77	America	1			
chevrolet concours				17.5	6	250.0	110
3520	16.4	77	America	1			
buick skylark				20.5	6	231.0	105
3425	16.9	77	America	1			
plymouth volare custom				19.0	6	225.0	100
3630	17.7	77	America	1			
ford granada				18.5	6	250.0	98
3525	19.0	77	America	1			
dodge monaco brougham				15.5	8	318.0	145
4140	13.7	77	America	1			
honda accord cvcc				31.5	4	98.0	68
2045	18.5	77	Japan	1			

```
datsun 510                          27.2          4          119.0          97
2300          14.7   78    Japan       1
toyota corona                       27.5          4          134.0          95
2560          14.2   78    Japan       1
chevrolet chevette                  30.0          4           98.0          68
2155          16.5   78  America       1
buick regal sport coupe (turbo)     17.7          6          231.0         165
3445          13.4   78  America       1
ford futura                         18.1          8          302.0         139
3205          11.2   78  America       1
dodge omni                          30.9          4          105.0          75
2230          14.5   78  America       1
dodge magnum xe                     17.5          8          318.0         140
4080          13.7   78  America       1
toyota celica gt liftback           21.1          4          134.0          95
2515          14.8   78    Japan       1
peugeot 604sl                       16.2          6          163.0         133
3410          15.8   78   Europe       1
oldsmobile starfire sx              23.8          4          151.0          85
2855          17.6   78  America       1
datsun 200-sx                       23.9          4          119.0          97
2405          14.9   78    Japan       1
audi 5000                           20.3          5          131.0         103
2830          15.9   78   Europe       1
volvo 264gl                         17.0          6          163.0         125
3140          13.6   78   Europe       1
saab 99gle                          21.6          4          121.0         115
2795          15.7   78   Europe       1
volkswagen scirocco                 31.5          4           89.0          71
1990          14.9   78   Europe       1
honda accord lx                     29.5          4           98.0          68
2135          16.6   78    Japan       1
plymouth sapporo                    23.2          4          156.0         105
2745          16.7   78  America       1
chevrolet monte carlo landau        19.2          8          305.0         145
3425          13.2   78  America       1
mazda glc deluxe                    32.8          4           78.0          52
1985          19.4   78    Japan       1
dodge aspen                         18.6          6          225.0         110
3620          18.7   78  America       1
volkswagen rabbit custom diesel     43.1          4           90.0          48
1985          21.5   78   Europe       1
ford fiesta                         36.1          4           98.0          66
1800          14.4   78  America       1
datsun b210 gx                      39.4          4           85.0          70
2070          18.6   78    Japan       1
honda civic cvcc                    36.1          4           91.0          60
```

1800	16.4	78	Japan	1			
amc concord d/l				18.1	6	258.0	120
3410	15.1	78	America	1			
dodge diplomat				19.4	8	318.0	140
3735	13.2	78	America	1			
mercury monarch ghia				20.2	8	302.0	139
3570	12.8	78	America	1			
oldsmobile cutlass salon brougham				19.9	8	260.0	110
3365	15.5	78	America	1			
chevrolet malibu				20.5	6	200.0	95
3155	18.2	78	America	1			
ford fairmont (auto)				20.2	6	200.0	85
2965	15.8	78	America	1			
ford fairmont (man)				25.1	4	140.0	88
2720	15.4	78	America	1			
plymouth volare				20.5	6	225.0	100
3430	17.2	78	America	1			
amc concord				19.4	6	232.0	90
3210	17.2	78	America	1			
buick century special				20.6	6	231.0	105
3380	15.8	78	America	1			
mercury zephyr				20.8	6	200.0	85
3070	16.7	78	America	1			
pontiac phoenix lj				19.2	6	231.0	105
3535	19.2	78	America	1			
plymouth horizon				34.2	4	105.0	70
2200	13.2	79	America	1			
mercedes benz 300d				25.4	5	183.0	77
3530	20.1	79	Europe	1			
cadillac eldorado				23.0	8	350.0	125
3900	17.4	79	America	1			
peugeot 504				27.2	4	141.0	71
3190	24.8	79	Europe	1			
oldsmobile cutlass salon brougham				23.9	8	260.0	90
3420	22.2	79	America	1			
plymouth horizon tc3				34.5	4	105.0	70
2150	14.9	79	America	1			
amc spirit dl				27.4	4	121.0	80
2670	15.0	79	America	1			
fiat strada custom				37.3	4	91.0	69
2130	14.7	79	Europe	1			
buick skylark limited				28.4	4	151.0	90
2670	16.0	79	America	1			
chevrolet citation				28.8	6	173.0	115
2595	11.3	79	America	1			
oldsmobile omega brougham				26.8	6	173.0	115
2700	12.9	79	America	1			

```
pontiac phoenix                       33.5          4         151.0          90
2556          13.2   79  America        1
datsun 210                            31.8          4          85.0          65
2020          19.2   79    Japan        1
dodge colt hatchback custom           35.7          4          98.0          80
1915          14.4   79  America        1
dodge st. regis                       18.2          8         318.0         135
3830          15.2   79  America        1
vw rabbit custom                      31.9          4          89.0          71
1925          14.0   79   Europe        1
mercury zephyr 6                      19.8          6         200.0          85
2990          18.2   79  America        1
ford fairmont 4                       22.3          4         140.0          88
2890          17.3   79  America        1
amc concord dl 6                      20.2          6         232.0          90
3265          18.2   79  America        1
dodge aspen 6                         20.6          6         225.0         110
3360          16.6   79  America        1
chevrolet caprice classic             17.0          8         305.0         130
3840          15.4   79  America        1
ford ltd landau                       17.6          8         302.0         129
3725          13.4   79  America        1
pontiac lemans v6                     21.5          6         231.0         115
3245          15.4   79  America        1
maxda glc deluxe                      34.1          4          86.0          65
1975          15.2   79    Japan        1
buick estate wagon (sw)               16.9          8         350.0         155
4360          14.9   79  America        1
ford country squire (sw)              15.5          8         351.0         142
4054          14.3   79  America        1
chevrolet malibu classic (sw)         19.2          8         267.0         125
3605          15.0   79  America        1
chrysler lebaron town @ country (sw)  18.5          8         360.0         150
3940          13.0   79  America        1
mercury grand marquis                 16.5          8         351.0         138
3955          13.2   79  America        1
vw rabbit c (diesel)                  44.3          4          90.0          48
2085          21.7   80   Europe        1
vw dasher (diesel)                    43.4          4          90.0          48
2335          23.7   80   Europe        1
audi 5000s (diesel)                   36.4          5         121.0          67
2950          19.9   80   Europe        1
mercedes-benz 240d                    30.0          4         146.0          67
3250          21.8   80   Europe        1
honda civic 1500 gl                   44.6          4          91.0          67
1850          13.8   80    Japan        1
datsun 280-zx                         32.7          6         168.0         132
```

```
2910          11.4   80    Japan         1
vokswagen rabbit                      29.8          4         89.0          62
1845          15.3   80   Europe         1
mazda rx-7 gs                         23.7          3         70.0         100
2420          12.5   80    Japan         1
triumph tr7 coupe                     35.0          4        122.0          88
2500          15.1   80   Europe         1
honda accord                          32.4          4        107.0          72
2290          17.0   80    Japan         1
datsun 210                            40.8          4         85.0          65
2110          19.2   80    Japan         1
subaru dl                             33.8          4         97.0          67
2145          18.0   80    Japan         1
dodge colt                            27.9          4        156.0         105
2800          14.4   80  America         1
mazda glc                             46.6          4         86.0          65
2110          17.9   80    Japan         1
toyota corolla                        32.2          4        108.0          75
2265          15.2   80    Japan         1
vw rabbit                             41.5          4         98.0          76
2144          14.7   80   Europe         1
toyota corolla tercel                 38.1          4         89.0          60
1968          18.8   80    Japan         1
chevrolet chevette                    32.1          4         98.0          70
2120          15.5   80  America         1
chevrolet citation                    28.0          4        151.0          90
2678          16.5   80  America         1
ford fairmont                         26.4          4        140.0          88
2870          18.1   80  America         1
datsun 310                            37.2          4         86.0          65
2019          16.4   80    Japan         1
dodge aspen                           19.1          6        225.0          90
3381          18.7   80  America         1
audi 4000                             34.3          4         97.0          78
2188          15.8   80   Europe         1
toyota corona liftback                29.8          4        134.0          90
2711          15.5   80    Japan         1
mazda 626                             31.3          4        120.0          75
2542          17.5   80    Japan         1
datsun 510 hatchback                  37.0          4        119.0          92
2434          15.0   80    Japan         1
amc concord                           24.3          4        151.0          90
3003          20.1   80  America         1
peugeot 505s turbo diesel             28.1          4        141.0          80
3230          20.4   81   Europe         1
honda prelude                         33.7          4        107.0          75
2210          14.4   81    Japan         1
```

toyota corolla				32.4	4	108.0	75
2350	16.8	81	Japan	1			
datsun 200sx				32.9	4	119.0	100
2615	14.8	81	Japan	1			
mazda 626				31.6	4	120.0	74
2635	18.3	81	Japan	1			
volvo diesel				30.7	6	145.0	76
3160	19.6	81	Europe	1			
chrysler lebaron salon				17.6	6	225.0	85
3465	16.6	81	America	1			
datsun 810 maxima				24.2	6	146.0	120
2930	13.8	81	Japan	1			
buick century				22.4	6	231.0	110
3415	15.8	81	America	1			
oldsmobile cutlass ls				26.6	8	350.0	105
3725	19.0	81	America	1			
ford granada gl				20.2	6	200.0	88
3060	17.1	81	America	1			
volkswagen jetta				33.0	4	105.0	74
2190	14.2	81	Europe	1			
toyota cressida				25.4	6	168.0	116
2900	12.6	81	Japan	1			
ford escort 2h				29.9	4	98.0	65
2380	20.7	81	America	1			
plymouth reliant				27.2	4	135.0	84
2490	15.7	81	America	1			
plymouth horizon 4				34.7	4	105.0	63
2215	14.9	81	America	1			
ford escort 4w				34.4	4	98.0	65
2045	16.2	81	America	1			
buick skylark				26.6	4	151.0	84
2635	16.4	81	America	1			
dodge aries wagon (sw)				25.8	4	156.0	92
2620	14.4	81	America	1			
plymouth reliant				30.0	4	135.0	84
2385	12.9	81	America	1			
toyota starlet				39.1	4	79.0	58
1755	16.9	81	Japan	1			
chevrolet citation				23.5	6	173.0	110
2725	12.6	81	America	1			
honda civic 1300				35.1	4	81.0	60
1760	16.1	81	Japan	1			
subaru				32.3	4	97.0	67
2065	17.8	81	Japan	1			
datsun 210 mpg				37.0	4	85.0	65
1975	19.4	81	Japan	1			
toyota tercel				37.7	4	89.0	62

```
2050           17.3   81    Japan         1
mazda glc 4                             34.1          4          91.0            68
1985           16.0   81    Japan         1
plymouth champ                          39.0          4          86.0            64
1875           16.4   81  America         1
chrysler lebaron medallion              26.0          4         156.0            92
2585           14.5   82  America         1
honda civic (auto)                      32.0          4          91.0            67
1965           15.7   82    Japan         1
datsun 310 gx                           38.0          4          91.0            67
1995           16.2   82    Japan         1
buick century limited                   25.0          6         181.0           110
2945           16.4   82  America         1
oldsmobile cutlass ciera (diesel)       38.0          6         262.0            85
3015           17.0   82  America         1
ford granada l                          22.0          6         232.0           112
2835           14.7   82  America         1
dodge rampage                           32.0          4         135.0            84
2295           11.6   82  America         1
dodge charger 2.2                       36.0          4         135.0            84
2370           13.0   82  America         1
chevrolet camaro                        27.0          4         151.0            90
2950           17.3   82  America         1
ford mustang gl                         27.0          4         140.0            86
2790           15.6   82  America         1
vw pickup                               44.0          4          97.0            52
2130           24.6   82   Europe         1
honda civic                             38.0          4          91.0            67
1965           15.0   82    Japan         1
toyota celica gt                        32.0          4         144.0            96
2665           13.9   82    Japan         1
toyota corolla                          34.0          4         108.0            70
2245           16.9   82    Japan         1
ford ranger                             28.0          4         120.0            79
2625           18.6   82  America         1
nissan stanza xe                        36.0          4         120.0            88
2160           14.5   82    Japan         1
mercury lynx l                          36.0          4          98.0            70
2125           17.3   82  America         1
plymouth horizon miser                  38.0          4         105.0            63
2125           14.7   82  America         1
mazda glc custom                        31.0          4          91.0            68
1970           17.6   82    Japan         1
mazda glc custom l                      37.0          4          91.0            68
2025           18.2   82    Japan         1
volkswagen rabbit l                     36.0          4         105.0            74
1980           15.3   82   Europe         1
```

```
ford fairmont futura                 24.0          4         140.0          92
2865          16.4    82   America      1
pontiac phoenix                      27.0          4         151.0          90
2735          18.0    82   America      1
dodge aries se                       29.0          4         135.0          84
2525          16.0    82   America      1
pontiac j2000 se hatchback           31.0          4         112.0          85
2575          16.2    82   America      1
chevrolet cavalier 2-door            34.0          4         112.0          88
2395          18.0    82   America      1
chevrolet cavalier wagon             27.0          4         112.0          88
2640          18.6    82   America      1
chevrolet cavalier                   28.0          4         112.0          88
2605          19.6    82   America      1
honda accord                         36.0          4         107.0          75
2205          14.5    82     Japan      1
chevy s-10                           31.0          4         119.0          82
2720          19.4    82   America      1
```

[68]:
```
Auto_sqrt["sqrt_displacement"] = np.sqrt(Auto_sqrt["displacement"])
Auto_sqrt["sqrt_horsepower"] = np.sqrt(Auto_sqrt["horsepower"])
Auto_sqrt["sqrt_weight"] = np.sqrt(Auto_sqrt["weight"])
Auto_sqrt = Auto_sqrt.drop(
    columns=[
        "displacement",
        "weight",
        "horsepower",
        "year",
    ]
)
Auto_sqrt.columns
```

[68]:
```
Index(['mpg', 'cylinders', 'acceleration', 'origin', 'oilshock',
'sqrt_displacement', 'sqrt_horsepower', 'sqrt_weight'], dtype='object')
```

[69]:
```
Auto_sqrt.corr(numeric_only=True)
```

[69]:
```
                        mpg  cylinders  acceleration  oilshock
sqrt_displacement  sqrt_horsepower  sqrt_weight
mpg                1.000000  -0.777618      0.423329  0.521192
-0.821331          -0.802311    -0.840095
cylinders          -0.777618   1.000000     -0.504683 -0.273703
0.953208           0.849266     0.893465
acceleration       0.423329  -0.504683      1.000000  0.195892
-0.521812          -0.696702    -0.409829
oilshock           0.521192  -0.273703      0.195892  1.000000
-0.284587          -0.306247    -0.260664
```

```
sqrt_displacement -0.821331   0.953208     -0.521812 -0.284587
1.000000          0.886470    0.939868
sqrt_horsepower    -0.802311  0.849266     -0.696702 -0.306247
0.886470          1.000000    0.872045
sqrt_weight        -0.840095  0.893465     -0.409829 -0.260664
0.939868          0.872045    1.000000
```

[70]:
```
Auto_sqrt = pd.get_dummies(
    Auto_sqrt, columns=list(["origin"]), drop_first=True, dtype=np.uint8
)
Auto_sqrt.columns
```

[70]:
```
Index(['mpg', 'cylinders', 'acceleration', 'oilshock', 'sqrt_displacement',
'sqrt_horsepower', 'sqrt_weight', 'origin_Europe', 'origin_Japan'],
dtype='object')
```

[71]:
```
cols = list(Auto_sqrt.columns)
cols.remove("mpg")
```

[72]:
```
vifdf = calculate_VIFs("mpg ~ " + " + ".join(cols), Auto_sqrt)
vifdf
```

[72]:
```
                         VIF
Feature
cylinders          11.465746
acceleration        3.010771
oilshock            1.151324
sqrt_displacement  27.042946
sqrt_horsepower    10.615281
sqrt_weight        13.450552
origin_Europe       1.774827
origin_Japan        1.944729
```

[73]:
```
identify_highest_VIF_feature(vifdf)
```

```
We find the highest VIF in this model is sqrt_displacement with a VIF of
27.042946454149405
Hence, we drop sqrt_displacement from the model to be fitted.
```

[73]:
```
('sqrt_displacement', 27.042946454149405)
```

[74]:
```
cols.remove("sqrt_displacement")
vifdf = calculate_VIFs("mpg ~ " + " + ".join(cols), Auto_sqrt)
vifdf
```

[74]:
```
                       VIF
Feature
cylinders         5.974510
```

```
acceleration      2.934605
oilshock          1.141428
sqrt_horsepower  10.446261
sqrt_weight       9.963350
origin_Europe     1.450840
origin_Japan      1.623907
```

[75]:
```python
identify_highest_VIF_feature(vifdf)
```

```
We find the highest VIF in this model is sqrt_horsepower with a VIF of
10.446261176837464
Hence, we drop sqrt_horsepower from the model to be fitted.
```

[75]:
```
('sqrt_horsepower', 10.446261176837464)
```

[76]:
```python
cols.remove("sqrt_horsepower")
vifdf = calculate_VIFs("mpg ~ " + " + ".join(cols), Auto_sqrt)
vifdf
```

[76]:
```
                    VIF
Feature
cylinders      5.907717
acceleration   1.377206
oilshock       1.119726
sqrt_weight    5.331435
origin_Europe  1.446456
origin_Japan   1.581460
```

[77]:
```python
identify_highest_VIF_feature(vifdf)
```

```
No variables are significantly collinear.
```

[78]:
```python
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_sqrt)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.814
Model:                            OLS   Adj. R-squared:                  0.811
Method:                 Least Squares   F-statistic:                     281.0
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          2.76e-137
Time:                        19:20:26   Log-Likelihood:                -1031.4
No. Observations:                 392   AIC:                             2077.
Df Residuals:                     385   BIC:                             2105.
Df Model:                           6
Covariance Type:            nonrobust
================================================================================
=
                    coef    std err          t      P>|t|      [0.025
```

```
0.975]
--------------------------------------------------------------------------------
-
Intercept        54.3622      2.396     22.687      0.000      49.651
59.073
cylinders         0.0148      0.244      0.061      0.952      -0.466
0.495
acceleration      0.1748      0.073      2.395      0.017       0.031
0.318
oilshock          5.0506      0.364     13.873      0.000       4.335
5.766
sqrt_weight      -0.6785      0.052    -13.130      0.000      -0.780
-0.577
origin_Europe     1.5511      0.544      2.851      0.005       0.481
2.621
origin_Japan      1.9033      0.537      3.544      0.000       0.847
2.959
==============================================================================
Omnibus:                       25.773   Durbin-Watson:                   1.269
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               41.514
Skew:                           0.449   Prob(JB):                     9.67e-10
Kurtosis:                       4.317   Cond. No.                         803.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                  df        sum_sq       mean_sq            F         PR(>F)
cylinders        1.0  14403.083079  14403.083079  1252.209748  4.496819e-123
acceleration     1.0     30.471304     30.471304     2.649187   1.044209e-01
oilshock         1.0   2422.051542   2422.051542   210.574120   2.284075e-38
sqrt_weight      1.0   2365.801178   2365.801178   205.683691   1.125346e-37
origin_Europe    1.0     24.780703     24.780703     2.154444   1.429743e-01
origin_Japan     1.0    144.484455    144.484455    12.561536   4.421781e-04
Residual       385.0   4428.321210     11.502133          NaN            NaN
```

[78]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b6563f0>

[79]:
```
identify_least_significant_feature(results, alpha=LOS_Alpha)
```

```
We find the least significant variable in this model is cylinders with a p-value
of 0.9516256870315071 and a coefficient of 0.014835798307379022
Using the backward methodology, we suggest dropping cylinders from the new model
```

[80]:
```
cols.remove("cylinders")
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_sqrt)
```

```
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.814
Model:                            OLS   Adj. R-squared:                  0.812
Method:                 Least Squares   F-statistic:                     338.0
Date:                Fri, 21 Feb 2025   Prob (F-statistic):         1.43e-138
Time:                        19:20:26   Log-Likelihood:                -1031.4
No. Observations:                 392   AIC:                             2075.
Df Residuals:                     386   BIC:                             2099.
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
=
                    coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
Intercept        54.3324      2.342     23.198      0.000      49.728
58.937
acceleration      0.1733      0.069      2.512      0.012       0.038
0.309
oilshock          5.0486      0.362     13.946      0.000       4.337
5.760
sqrt_weight      -0.6760      0.031    -21.968      0.000      -0.737
-0.616
origin_Europe     1.5438      0.530      2.913      0.004       0.502
2.586
origin_Japan      1.8999      0.534      3.561      0.000       0.851
2.949
==============================================================================
Omnibus:                       25.925   Durbin-Watson:                   1.269
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               41.952
Skew:                           0.450   Prob(JB):                     7.77e-10
Kurtosis:                       4.326   Cond. No.                         783.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                  df        sum_sq       mean_sq           F         PR(>F)
acceleration     1.0   4268.531557   4268.531557  372.068179   1.546709e-58
oilshock         1.0   4757.627552   4757.627552  414.700418   3.906846e-63
sqrt_weight      1.0  10191.062423  10191.062423  888.307839  3.798699e-102
origin_Europe    1.0     27.910362     27.910362    2.432817   1.196386e-01
origin_Japan     1.0    145.497978    145.497978   12.682387   4.152294e-04
Residual       386.0   4428.363596     11.472445         NaN            NaN
```

[80]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b67f080>

[81]:
```
identify_least_significant_feature(results, alpha=LOS_Alpha)
```

```
We find the least significant variable in this model is acceleration with a
p-value of 0.01239680442008935 and a coefficient of 0.17332821877689886
Using the backward methodology, we suggest dropping acceleration from the new
model
```

[82]:
```
cols.remove("acceleration")
formula = " + ".join(cols)
results = perform_analysis("mpg", formula, Auto_sqrt)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.811
Model:                            OLS   Adj. R-squared:                  0.809
Method:                 Least Squares   F-statistic:                     415.3
Date:                Fri, 21 Feb 2025   Prob (F-statistic):          1.50e-138
Time:                        19:20:26   Log-Likelihood:                -1034.6
No. Observations:                 392   AIC:                             2079.
Df Residuals:                     387   BIC:                             2099.
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
=
                    coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-
Intercept        58.2668      1.753     33.230      0.000      54.819
61.714
oilshock          5.1556      0.362     14.244      0.000       4.444
5.867
sqrt_weight      -0.6999      0.029    -23.759      0.000      -0.758
-0.642
origin_Europe     1.6530      0.532      3.108      0.002       0.607
2.699
origin_Japan      1.8263      0.536      3.405      0.001       0.772
2.881
==============================================================================
Omnibus:                       31.883   Durbin-Watson:                   1.241
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               60.472
Skew:                           0.483   Prob(JB):                     7.39e-14
Kurtosis:                       4.664   Cond. No.                         574.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

```
                  df        sum_sq       mean_sq            F         PR(>F)
oilshock         1.0   6470.207217   6470.207217   556.341155   7.004177e-77
sqrt_weight      1.0  12674.256790  12674.256790  1089.796728  1.354903e-114
origin_Europe    1.0     38.907051     38.907051     3.345425   6.816142e-02
origin_Japan     1.0    134.840521    134.840521    11.594270   7.307851e-04
Residual       387.0   4500.781890     11.629927          NaN            NaN
```

[82]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x74027b67f9b0>

[83]:
```
models.append(
    {
        "name": "sqrt_transformation",
        "model": results.model.formula,
        "R-squared adjusted": results.rsquared_adj,
    }
)
```

[84]:
```
pd.DataFrame(models)
```

[84]:
```
                      name
model  R-squared adjusted
0            simple_model
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan
0.802074
1    numeric_interactions
mpg ~ horsepower + weight + oilshock + origin_Europe + origin_Japan +
horsepower: weight            0.844846
2   oilshock_interactions                        mpg ~ horsepower + weight + oilshock +
origin_Europe + origin_Japan + horsepower: weight + oilshock: weight + oilshock:
horsepower             0.858491
3     origin_interactions  mpg ~ horsepower + weight + oilshock + origin_Europe
+ origin_Japan + oilshock: horsepower + origin_Europe: horsepower +
origin_Japan: horsepower            0.849008
4  squared_transformation                                                   mpg ~
horsepower + weight + oilshock + origin_Europe + origin_Japan + I(horsepower**2)
+ I(weight**2)            0.845550
5      log_transformation
mpg ~ acceleration + oilshock + log_weight + origin_Europe + origin_Japan
0.821151
6     sqrt_transformation
mpg ~ oilshock + sqrt_weight + origin_Europe + origin_Japan            0.809089
```

[85]:
```
allDone()
```

<IPython.lib.display.Audio object>