04-05-multi-linear-regression

February 13, 2025

[ ]:
```
import pandas as pd
from google.colab import files
uploaded = files.upload()
file_name = list(uploaded.keys())[0]
cars = pd.read_csv(file_name)
```

```
<IPython.core.display.HTML object>
```

```
Saving Cars.csv to Cars.csv
```

[ ]:
```
smf = pd.read_csv(file_name)
```

[ ]:
```
ols = pd.read_csv(file_name)
```

[ ]:
```
DataFrame = pd.read_csv(file_name)
```

[ ]:
```
smf.ols = pd.read_csv(file_name)
```

```
<ipython-input-17-cbb3ff838e8d>:1: UserWarning: Pandas doesn't allow columns to
be created via a new attribute name - see https://pandas.pydata.org/pandas-
docs/stable/indexing.html#attribute-access
  smf.ols = pd.read_csv(file_name)
```

[ ]:
```

```

[ ]:
```
#vif=varience inflamation factor
rsq_hp=smf.ols('HP~WT+VOL+SP',data=cars).fit().rsquared
vif_hp=1/(1-rsq_hp)
rsq_wt=smf.ols('WT~HP+VOL+SP',data=cars).fit()
vif_wt=1/(1-rsq_wt.rsquared)
rsq_vol=smf.ols('VOL~HP+WT+SP',data=cars).fit()
vif_vol=1/(1-rsq_vol.rsquared)
rsq_sp=smf.ols('SP~HP+VOL+WT',data=cars).fit().rsquared
vif_sp=1/(1-rsq_sp)
d1={'Variables':['HP','WT','VOL','SP'],'VIF':[vif_hp,vif_wt,vif_vol,vif_sp]}
Vif_frame=pd.DataFrame(d1)
Vif_frame
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-20-72140651ea32> in <cell line: 0>()
      1 #vif=varience inflamation factor
----> 2 rsq_hp=smf.ols('HP~WT+VOL+SP',data=cars).fit().rsquared
      3 vif_hp=1/(1-rsq_hp)
      4 rsq_wt=smf.ols('WT~HP+VOL+SP',data=cars).fit()
      5 vif_wt=1/(1-rsq_wt.rsquared)

TypeError: 'DataFrame' object is not callable
```

[ ]:
```
ml_v=smf.ols('MPG~VOL',data =cars).fit()
print(ml_v.tvalues, '\n', ml_v.pvalues)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-4-0317ca6b38d9> in <cell line: 0>()
----> 1 ml_v=smf.ols('MPG~VOL',data =cars).fit()
      2 print(ml_v.tvalues, '\n', ml_v.pvalues)

NameError: name 'smf' is not defined
```

[ ]:
```
import statsmodels.formula.api as smf
```

[ ]:
```
model=smf.ols('MPG~WT+VOL+SP',data =cars).fit()
```

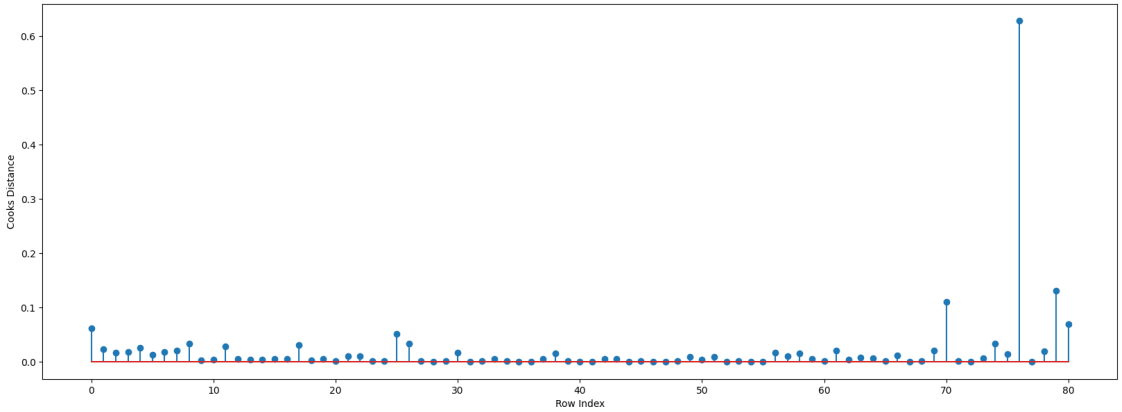
[ ]:
```
list(np.where(model.resid>10))
```

[ ]:
```
[array([ 0, 76])]
```

[ ]:
```
def get_standardized_values(vals):
  return (vals-vals.mean())/vals.std()
```

[ ]:
```
model_influence = model.get_influence()
(c, _) = model_influence.cooks_distance
```

[ ]:
```
model_influence=model.get_influence()
(c,_)=model_influence.cooks_distance
fig=plt.subplots(figsize=(20,7))
plt.stem(np.arange(len(cars)),np.round(c,3))
plt.xlabel('Row Index')
plt.ylabel('Cooks Distance')
plt.show()
```



[ ]:

[ ]:
```
import statsmodels.api as sm
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

[ ]:
```
# Example data (replace with your own dataset)
data = pd.DataFrame({
    'X': np.random.rand(100),
    'Y': np.random.rand(100)
})
```

[ ]:
```
x = cars.iloc[:,1:]
y = cars.iloc[:,0]
```

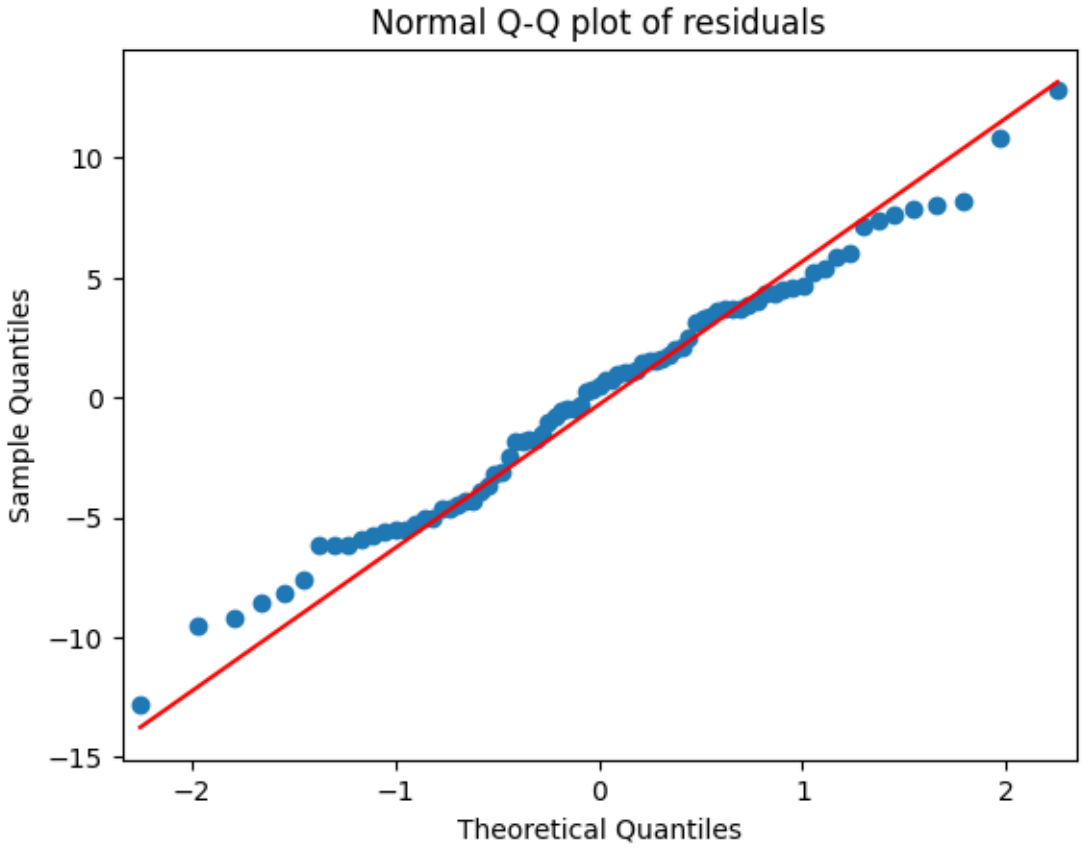
[ ]:
```
x = sm.add_constant(x)
```

[ ]:
```
model.fittedvalues
```

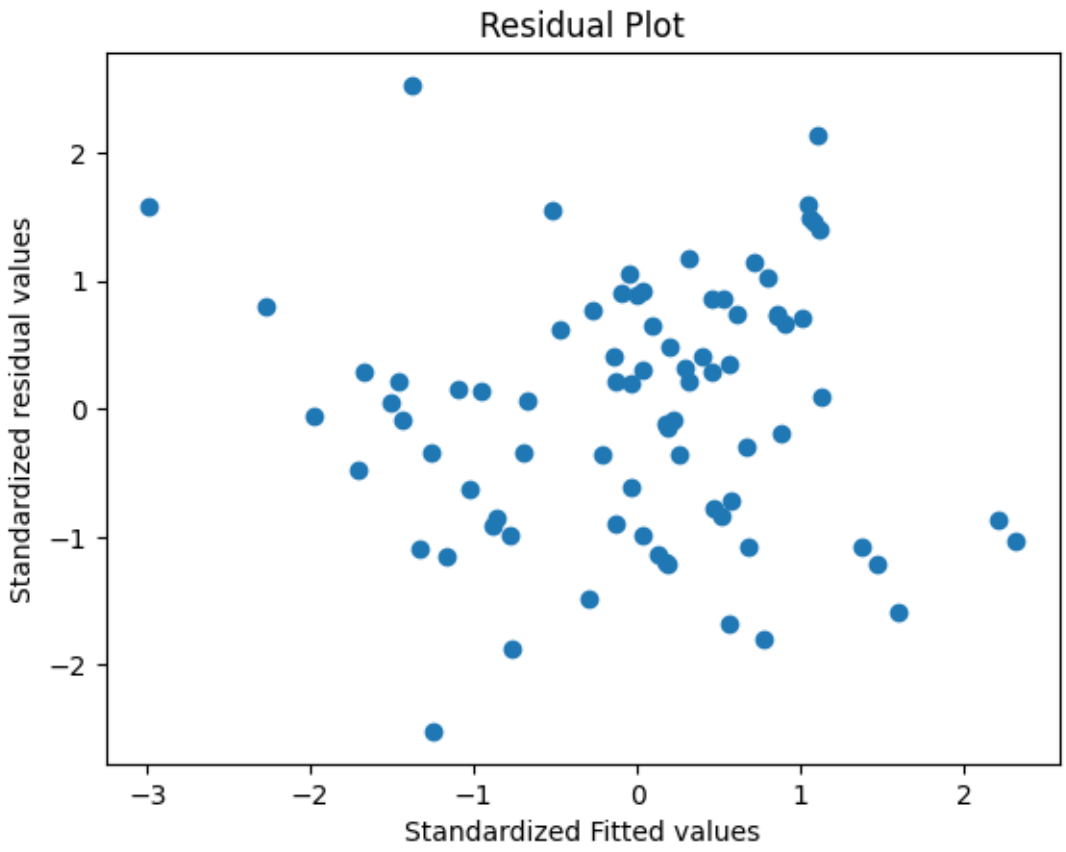
[ ]:
```
0     42.832365
1     42.857708
2     42.418722
3     39.825362
4     42.341828
        ...
76    24.015336
77    19.467343
78    30.851867
79    11.800779
80    24.947729
Length: 81, dtype: float64
```

[ ]:

[ ]:
```
sm.qqplot(model.resid, line='q')
plt.title("Normal Q-Q plot of residuals")
plt.show()
```



[ ]:
```
plt.scatter(get_standardized_values(model.fittedvalues),
            get_standardized_values(model.resid))
plt.title('Residual Plot')
plt.xlabel('Standardized Fitted values')
plt.ylabel('Standardized residual values')
plt.show()
```



[ ]:
```
#new datav for prediction
new_data=pd.DataFrame({'HP':40,"VOL":95,"SP":102,"WT":35},index=[1 ])
```

[ ]:
```
from typing_extensions import final
final_ml_V=smf.ols('MPG~VOL',data =cars).fit()
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-22-fe2655658679> in <cell line: 0>()
      1 from typing_extensions import final
----> 2 final_ml_V=smf.ols('MPG~VOL',data =cars).fit()

TypeError: 'DataFrame' object is not callable
```

[ ]:
```
final_ml_V.predict(new_data)
final_ml_V.predict(cars_new.iloc[0:5,])
pred_y=final_ml_v.predict(cars.iloc[:,1:])
```

```
---------------------------------------------------------------------------
```

```
NameError                                 Traceback (most recent call last)
<ipython-input-8-711a6f3df059> in <cell line: 0>()
----> 1 final_ml_V.predict(new_data)
      2 final_ml_V.predict(cars_new.iloc[0:5,])
      3 pred_y=final_ml_v.predict(cars.iloc[:,1:])

NameError: name 'final_ml_V' is not defined
```

[ ]: