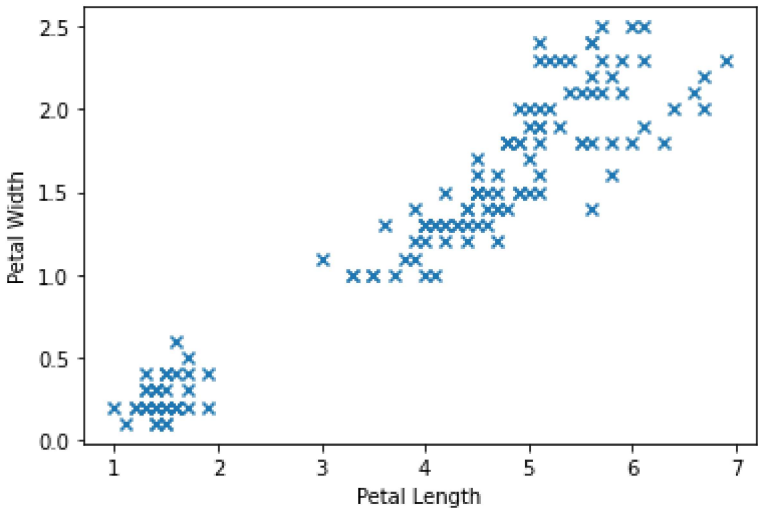
In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats
import seaborn as sns
import statsmodels.api as sm
from sklearn import linear_model
```

In [2]:
```
#read in the data for iris
iris = sns.load_dataset('iris')
iris.head()
```

Out[2]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

In [5]:
```
x = iris['petal_length']
y = iris['petal_width']
plt.scatter(x,y,marker = 'x')
plt.xlabel('Petal Length')
plt.ylabel('Petal Width')
plt.show()
```



In [6]:
```
#Linear Regression
model = sm.OLS(y,x)
result = model.fit()
print(result.summary())
```

```
                            OLS Regression Results
==============================================================================
==
Dep. Variable:            petal_width   R-squared (uncentered):                   0.9
67
Model:                            OLS   Adj. R-squared (uncentered):              0.9
67
Method:                 Least Squares   F-statistic:                              441
7.
Date:                Wed, 03 Aug 2022   Prob (F-statistic):                    1.22e-1
12
Time:                        20:54:28   Log-Likelihood:                         -8.71
79
No. Observations:                 150   AIC:                                      19.
44
Df Residuals:                     149   BIC:                                      22.
45
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
petal_length     0.3365      0.005     66.463      0.000       0.327       0.347
==============================================================================
Omnibus:                       19.720   Durbin-Watson:                   0.857
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               23.498
Skew:                           0.957   Prob(JB):                     7.90e-06
Kurtosis:                       3.311   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a
constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly spec
ified.
```

In [9]:

```
 # Linear Regression with intercept i.e with constant term
# we -need to add a conatant column
x1 = iris['petal_length']
y1 = iris['petal_width']
x1 = sm.add_constant(x1)#added a constant column
x1.head()
```

Out[9]:

	const	petal_length
0	1.0	1.4
1	1.0	1.4
2	1.0	1.3
3	1.0	1.5
4	1.0	1.4

In [11]:

```
#fittinf regression model
model1 = sm.OLS(y1,x1)
result1 = model1.fit()
print(result1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            petal_width   R-squared:                       0.927
Model:                            OLS   Adj. R-squared:                  0.927
Method:                 Least Squares   F-statistic:                     1882.
Date:                Wed, 03 Aug 2022   Prob (F-statistic):           4.68e-86
Time:                        22:16:59   Log-Likelihood:                 24.796
No. Observations:                 150   AIC:                            -45.59
Df Residuals:                     148   BIC:                            -39.57
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          -0.3631      0.040     -9.131      0.000      -0.442      -0.285
petal_length    0.4158      0.010     43.387      0.000       0.397       0.435
==============================================================================
Omnibus:                        5.765   Durbin-Watson:                   1.455
Prob(Omnibus):                  0.056   Jarque-Bera (JB):                5.555
Skew:                           0.359   Prob(JB):                       0.0622
Kurtosis:                       3.611   Cond. No.                         10.3
==============================================================================

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

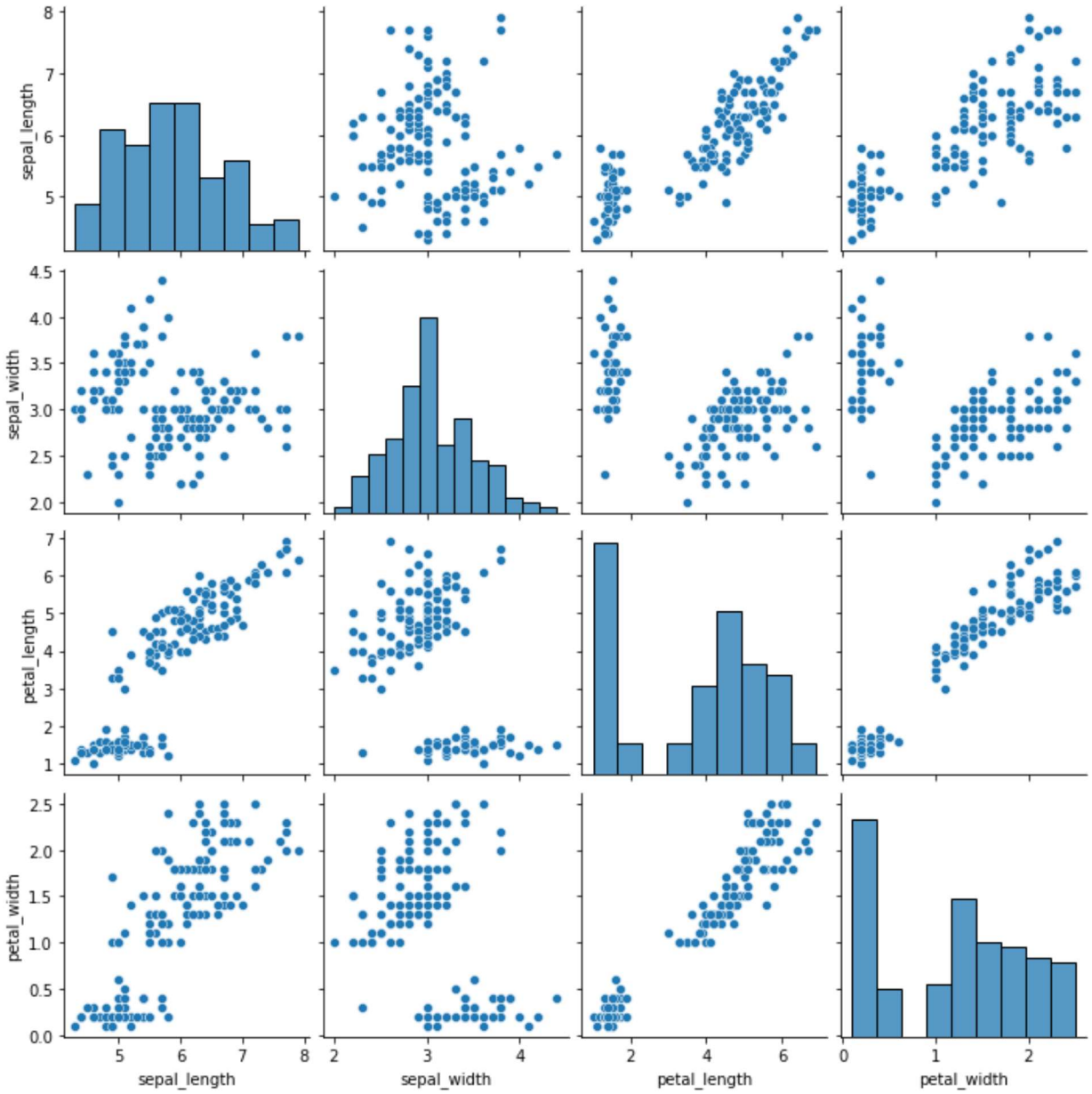
In [12]:
```
#conditions for linear regression
#loading iris data set
iris = sns.load_dataset('iris')
iris.head()
```

Out[12]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

In [13]:
```
sns.pairplot(iris[['sepal_length','sepal_width','petal_length','petal_width']].dropna(
```

Out[13]:
```
<seaborn.axisgrid.PairGrid at 0x1af4e0f4fa0>
```



In [14]:
```
#multi linear regression
x2= iris[['sepal_length','sepal_width','petal_length']]
y2= iris[['petal_width']]

#Adding constant column to my predictor variable

x2 = sm.add_constant(x2)
x2.head()
```

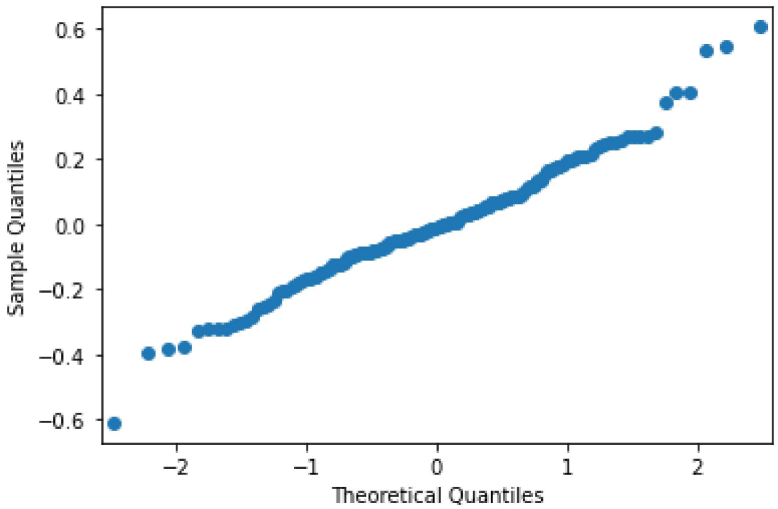
Out[14]:

	const	sepal_length	sepal_width	petal_length
0	1.0	5.1	3.5	1.4
1	1.0	4.9	3.0	1.4
2	1.0	4.7	3.2	1.3
3	1.0	4.6	3.1	1.5
4	1.0	5.0	3.6	1.4

In [15]:
```
#regression model fitting
model3 = sm.OLS(y2,x2)
result3 = model3.fit()

print(result3.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            petal_width   R-squared:                       0.938
Model:                            OLS   Adj. R-squared:                  0.937
Method:                 Least Squares   F-statistic:                     734.4
Date:                Wed, 03 Aug 2022   Prob (F-statistic):           7.83e-88
Time:                        23:09:40   Log-Likelihood:                 36.751
No. Observations:                 150   AIC:                            -65.50
Df Residuals:                     146   BIC:                            -53.46
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          -0.2403      0.178     -1.347      0.180      -0.593       0.112
sepal_length   -0.2073      0.048     -4.363      0.000      -0.301      -0.113
sepal_width     0.2228      0.049      4.553      0.000       0.126       0.320
petal_length    0.5241      0.024     21.399      0.000       0.476       0.572
==============================================================================
Omnibus:                        5.609   Durbin-Watson:                   1.573
Prob(Omnibus):                  0.061   Jarque-Bera (JB):                6.811
Skew:                           0.223   Prob(JB):                       0.0332
Kurtosis:                       3.944   Cond. No.                         90.1
==============================================================================

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

In [16]:
```
#for linear regression to work the residuals are normally distributed to check this we

residuals = result3.resid
sm.qqplot(residuals)
plt.show()#since graph is a straignt line so residuals are normally distributed
```



In [18]:
```
#evaluation of milti colinearity
corelation = x2.corr()
corelation
```

Out[18]:

	const	sepal_length	sepal_width	petal_length
const	NaN	NaN	NaN	NaN
sepal_length	NaN	1.000000	-0.11757	0.871754
sepal_width	NaN	-0.117570	1.00000	-0.428440
petal_length	NaN	0.871754	-0.42844	1.000000

In [ ]:
```

```