

```
In [29]: import numpy as np
import pandas as pd
from scipy import stats
import statsmodels.api as sm
import matplotlib.pyplot as plt
```

```
In [36]: df = pd.read_excel('Housing.xlsx')
```

```
In [50]: print(df)
```

	House Price	House Size (sq.ft.)	State	Number of Rooms	\
0	1116000	1940	IN	8	
1	860000	1300	IN	5	
2	818400	1420	IN	6	
3	1000000	1680	IN	7	
4	640000	1270	IN	5	
5	1010000	1850	IN	7	
6	600000	1000	IN	4	
7	700000	1100	LA	4	
8	1100000	1600	LA	7	
9	570000	1000	NY	5	
10	860000	2150	NY	9	
11	1085000	1900	NY	9	
12	1250000	2200	NY	9	
13	850000	1100	TX	4	
14	640000	860	TX	4	
15	900000	1325	TX	6	
16	730000	1350	TX	6	
17	750000	1600	TX	6	
18	650000	950	TX	2	
19	680000	1250	TX	4	

	Year of Construction
0	2002
1	1992
2	1987
3	2000
4	1995
5	1998
6	2015
7	2014
8	2017
9	1997
10	1997
11	2000
12	2014
13	2017
14	1997
15	1997
16	2000
17	1992
18	1987
19	2000

```
In [40]: X = df[['House Size (sq.ft.)', 'Number of Rooms']]
Y = df['House Price']
```

```
In [45]: X1 = sm.add_constant(X,prepend=True)
reg = sm.OLS(Y, X1).fit()
reg.summary()
```

```
/anaconda3/lib/python3.7/site-packages/numpy/core/fromnumeric.py:2389: FutureWarning: Method .ptp is deprecated and will be removed in a future version. Use numpy.ptp instead.
  return ptp(axis=axis, out=out, **kwargs)
```

Out[45]: OLS Regression Results

Dep. Variable:	House Price	R-squared:	0.683
Model:	OLS	Adj. R-squared:	0.645
Method:	Least Squares	F-statistic:	18.30
Date:	Sat, 03 Aug 2019	Prob (F-statistic):	5.77e-05
Time:	13:31:44	Log-Likelihood:	-260.28
No. Observations:	20	AIC:	526.6
Df Residuals:	17	BIC:	529.6
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	2.737e+05	1.03e+05	2.655	0.017	5.62e+04	4.91e+05
House Size (sq.ft.)	314.1363	190.485	1.649	0.117	-87.752	716.025
Number of Rooms	1.944e+04	3.95e+04	0.492	0.629	-6.39e+04	1.03e+05

Omnibus:	1.326	Durbin-Watson:	1.852
Prob(Omnibus):	0.515	Jarque-Bera (JB):	0.810
Skew:	-0.487	Prob(JB):	0.667
Kurtosis:	2.853	Cond. No.	5.89e+03

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 5.89e+03. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [46]: X = df[['House Size (sq.ft.)', 'Year of Construction']]
Y = df['House Price']
```

```
In [47]: X1 = sm.add_constant(X,prepend=True)
reg = sm.OLS(Y, X1).fit()
reg.summary()
```

Out[47]: OLS Regression Results

Dep. Variable:	House Price	R-squared:	0.735
Model:	OLS	Adj. R-squared:	0.704
Method:	Least Squares	F-statistic:	23.55
Date:	Sat, 03 Aug 2019	Prob (F-statistic):	1.26e-05
Time:	13:36:22	Log-Likelihood:	-258.49
No. Observations:	20	AIC:	523.0
Df Residuals:	17	BIC:	526.0
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-9.654e+06	5.21e+06	-1.852	0.081	-2.07e+07	1.34e+06
House Size (sq.ft.)	394.0417	61.098	6.449	0.000	265.137	522.947
Year of Construction	4960.9407	2607.443	1.903	0.074	-540.283	1.05e+04

Omnibus:	2.064	Durbin-Watson:	1.926
Prob(Omnibus):	0.356	Jarque-Bera (JB):	1.689
Skew:	-0.663	Prob(JB):	0.430
Kurtosis:	2.480	Cond. No.	5.36e+05

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 5.36e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [48]: X = df[['Number of Rooms', 'Year of Construction']]
Y = df['House Price']
```

```
In [49]: X1 = sm.add_constant(X,prepend=True)
reg = sm.OLS(Y, X1).fit()
reg.summary()
```

Out[49]: OLS Regression Results

Dep. Variable:	House Price	R-squared:	0.677
Model:	OLS	Adj. R-squared:	0.639
Method:	Least Squares	F-statistic:	17.79
Date:	Sat, 03 Aug 2019	Prob (F-statistic):	6.79e-05
Time:	13:37:32	Log-Likelihood:	-260.47
No. Observations:	20	AIC:	526.9
Df Residuals:	17	BIC:	529.9
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-8.471e+06	5.77e+06	-1.468	0.160	-2.06e+07	3.7e+06
Number of Rooms	7.824e+04	1.4e+04	5.574	0.000	4.86e+04	1.08e+05
Year of Construction	4424.7160	2887.793	1.532	0.144	-1667.996	1.05e+04

Omnibus:	2.115	Durbin-Watson:	1.959
Prob(Omnibus):	0.347	Jarque-Bera (JB):	1.400
Skew:	-0.407	Prob(JB):	0.497
Kurtosis:	1.991	Cond. No.	4.34e+05

Warnings:

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

[2] The condition number is large, 4.34e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [ ]:
```