

回归分析

```
import scipy.stats as stats
import numpy as np
import statsmodels.api as sm
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn import model_selection

from patsy import dmatrices
```

1.数据选择

数据选取例2.3

```
data=pd.read_csv("chapter2.csv")
```

```
data
```

	year	Y	X1	X2	X3
0	1994.0	1023.5	48637.5	5.90	111.78
1	1995.0	1375.7	61339.9	6.24	115.70
2	1996.0	1638.4	71813.6	6.49	118.58
3	1997.0	2112.7	79715.0	6.60	122.64
4	1998.0	2391.2	85195.5	6.64	127.85
5	1999.0	2831.9	90564.4	6.74	135.17
6	2000.0	3175.5	100280.1	6.87	167.98
7	2001.0	3522.4	110863.1	7.01	169.80
8	2002.0	3878.4	121717.4	7.19	176.52
9	2003.0	3442.3	137422.0	7.30	180.98
10	2004.0	4710.7	161840.2	7.44	187.07
11	2005.0	5285.9	187318.9	7.54	334.52
12	2006.0	6229.7	219438.5	7.71	345.70
13	2007.0	7770.6	270232.3	7.80	358.37
14	2008.0	8749.3	319515.5	7.97	373.02
15	2009.0	10183.7	349081.4	8.55	386.08
16	2010.0	12579.8	413030.3	9.12	400.82
17	2011.0	19305.4	489300.6	9.32	410.64
18	2012.0	22706.2	540367.4	9.76	423.75
19	2013.0	26276.1	595244.4	10.31	435.62
20	2014.0	30311.9	643974.0	11.18	446.39
21	2015.0	34195.1	689052.1	12.10	457.73
22	2016.0	39390.0	743585.5	12.40	469.63

```
from statsmodels.formula.api import ols
```

2. 进行多元线性回归

```
data.columns=[ 'year', "Y", "X1", "X2", "X3" ]
data=data[ [ "Y", "X1", "X2", "X3" ] ]
```

```
data=data.astype(float)
```

```
fit = ols('Y~ X1 + X2 + X3', data=data).fit()

fit.summary()
```

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.989
Model:	OLS	Adj. R-squared:	0.988
Method:	Least Squares	F-statistic:	584.9
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	6.92e-19
Time:	23:34:41	Log-Likelihood:	-195.34
No. Observations:	23	AIC:	398.7
Df Residuals:	19	BIC:	403.2
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t 	[0.025	0.975]
Intercept	-1.552e+04	5801.877	-2.675	0.015	-2.77e+04	-3379.285
X1	0.0449	0.009	4.871	0.000	0.026	0.064
X2	2724.8230	932.268	2.923	0.009	773.563	4676.083
X3	-30.2989	5.623	-5.388	0.000	-42.069	-18.529

Omnibus:	2.781	Durbin-Watson:	1.186
Prob(Omnibus):	0.249	Jarque-Bera (JB):	1.307
Skew:	-0.508	Prob(JB):	0.520
Kurtosis:	3.575	Cond. No.	7.82e+06

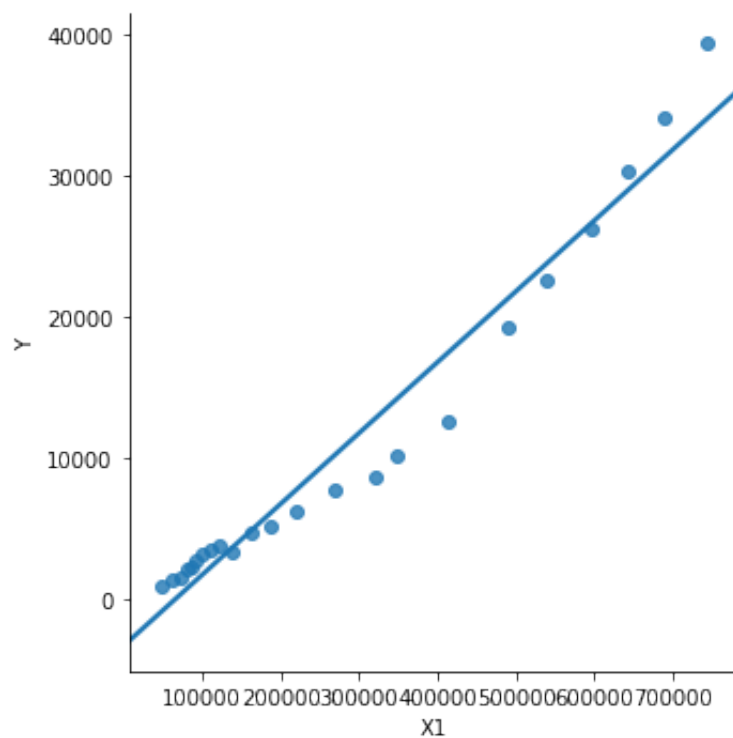
Warnings:

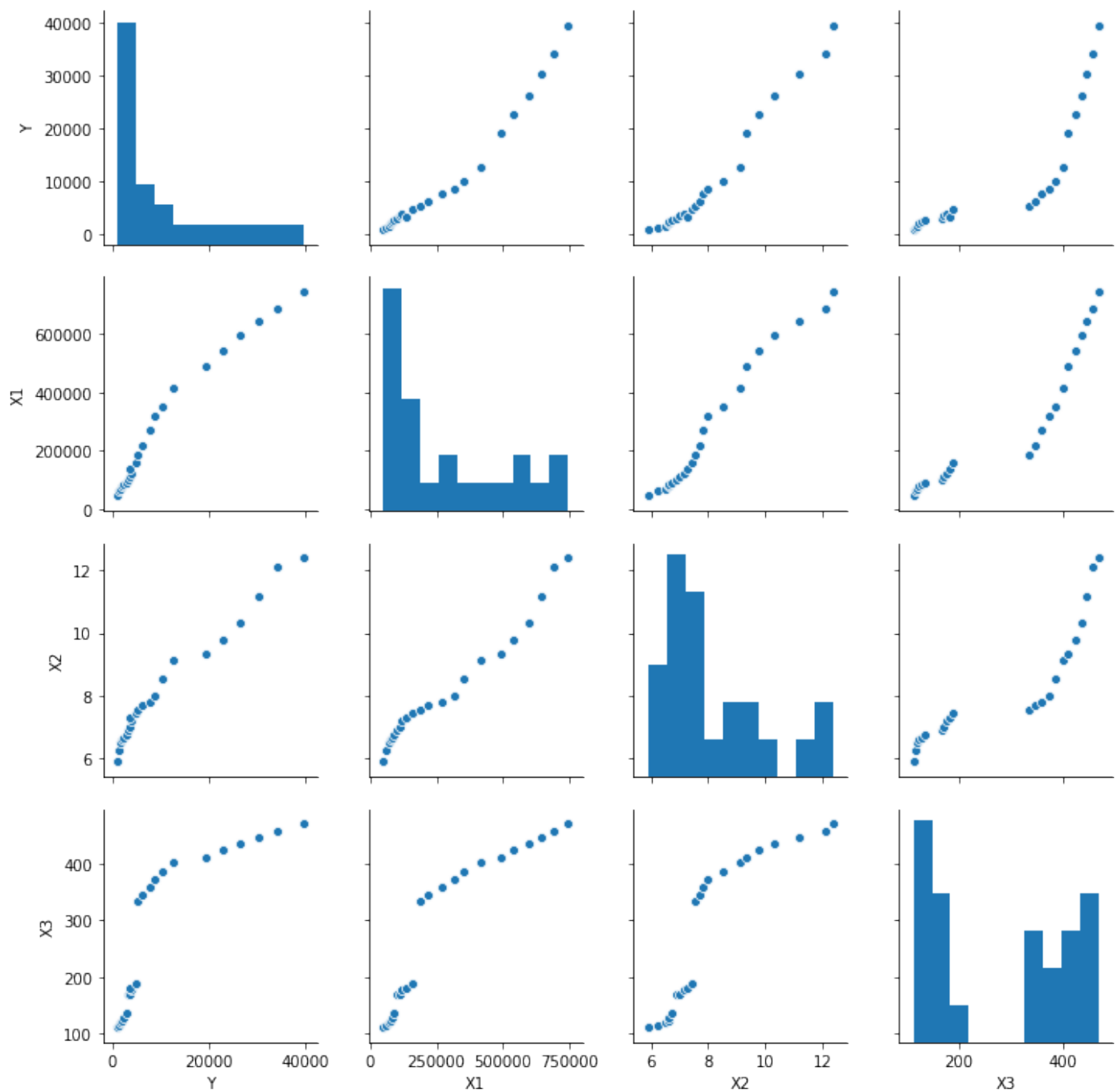
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 7.82e+06. This might indicate that there are strong multicollinearity or other numerical problems.

```
sns.lmplot(x='X1', y='Y', data=data, ci=None)
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False
plt.show()
sns.pairplot(data.loc[:, ['Y', 'X1', 'X2', 'X3']])
# 显示图形
plt.show()
```

findfont: Font family ['sans-serif'] not found. Falling back to DejaVu Sans.





```
(9.952110091358767,
0.35437608786573616,
1.1017314088803234,
0.42370237871794003)
```

