

[ジェンダー,身長,体重]のデータ

```
In [1]: import numpy as np
import pandas as pd
```

```
In [2]: import os

os.getcwd()
```

```
Out[2]: 'C:\Users\david\Documents\PYTHON_NOTEBOOKS\Spring2021'
```

```
In [3]: nRowsRead = None # specify 'None' if want to read whole file
# weight-height.csv has 10000 rows in reality, but we are only loading/previewing the first 1000 rows
df1 = pd.read_csv('data_h_w.csv', delimiter=',', nrows = nRowsRead)
df1.dataframeName = 'gender-weight-height'
nRow, nCol = df1.shape
print(f'There are {nRow} rows and {nCol} columns')
```

There are 1000 rows and 4 columns

```
In [4]: df1.head(5)
```

```
Out[4]:
```

	Unnamed: 0	Gender	Height	Weight
0	9953	Female	61.278349	132.534229
1	3850	Male	70.454991	190.824938
2	4962	Male	70.144763	194.831396
3	3886	Male	69.231469	173.829606
4	5437	Female	68.759636	153.689739

```
In [5]: df1=df1.drop('Unnamed: 0', axis=1)
```

```
In [6]: df1.tail()
```

```
Out[6]:
```

	Gender	Height	Weight
995	Male	72.796730	196.504136
996	Female	64.258342	140.236766
997	Male	67.903612	177.352590
998	Male	71.189384	202.782736
999	Male	66.631041	170.421149

```
In [7]: df1.describe(include='all')
```

```
Out[7]:
```

	Gender	Height	Weight
count	1000	1000.000000	1000.000000
unique	2	NaN	NaN
top	Male	NaN	NaN
freq	530	NaN	NaN
mean	NaN	66.475941	162.186259
std	NaN	3.981159	33.052238
min	NaN	55.668202	68.982530
25%	NaN	63.472736	137.364286
50%	NaN	66.438081	163.870956
75%	NaN	69.563558	188.352381
max	NaN	78.095867	255.690835

```
In [8]: df2 = df1.sample(frac=0.1, random_state=1)
df2.describe()
```

```
Out[8]:
```

	Height	Weight
count	100.000000	100.000000
mean	67.263896	168.845554
std	3.705571	31.735864

	Height	Weight
min	60.955487	105.880581
25%	63.827713	139.702536
50%	67.382045	176.407739
75%	70.158781	195.057177
max	75.259457	228.800516

In [9]: `df2.head()`

Out[9]:

	Gender	Height	Weight
507	Male	63.724727	165.979417
818	Female	62.968382	124.773903
452	Male	70.643193	198.491168
368	Male	70.035068	181.324779
242	Male	75.259457	227.437138

In [10]: `df2.tail()`

Out[10]:

	Gender	Height	Weight
671	Male	68.122768	196.414415
559	Female	61.705576	126.787299
593	Male	70.052511	194.974116
258	Male	71.213913	195.923416
154	Female	66.199283	180.698317

In [11]:

In [12]: `df2.head()`

Out[12]:

	Gender	Height	Weight	Height(cm)	Weight(kg)
507	Male	63.724727	165.979417	162	75.3
818	Female	62.968382	124.773903	160	56.6
452	Male	70.643193	198.491168	179	90.0
368	Male	70.035068	181.324779	178	82.2
242	Male	75.259457	227.437138	191	103.2

In [13]: `df2.describe()`

Out[13]:

	Height	Weight	Height(cm)	Weight(kg)
count	100.000000	100.000000	100.000000	100.000000
mean	67.263896	168.845554	170.890000	76.583000
std	3.705571	31.735864	9.439756	14.394816
min	60.955487	105.880581	155.000000	48.000000
25%	63.827713	139.702536	162.000000	63.350000
50%	67.382045	176.407739	171.000000	80.000000
75%	70.158781	195.057177	178.250000	88.450000
max	75.259457	228.800516	191.000000	103.800000

In [14]: `df2['BMI']=df2.apply(lambda x:round(x['Weight(kg)']/(x['Height(cm)']/100)**2,1),axis=1)`

In [15]: `df2.head()`

Out[15]:

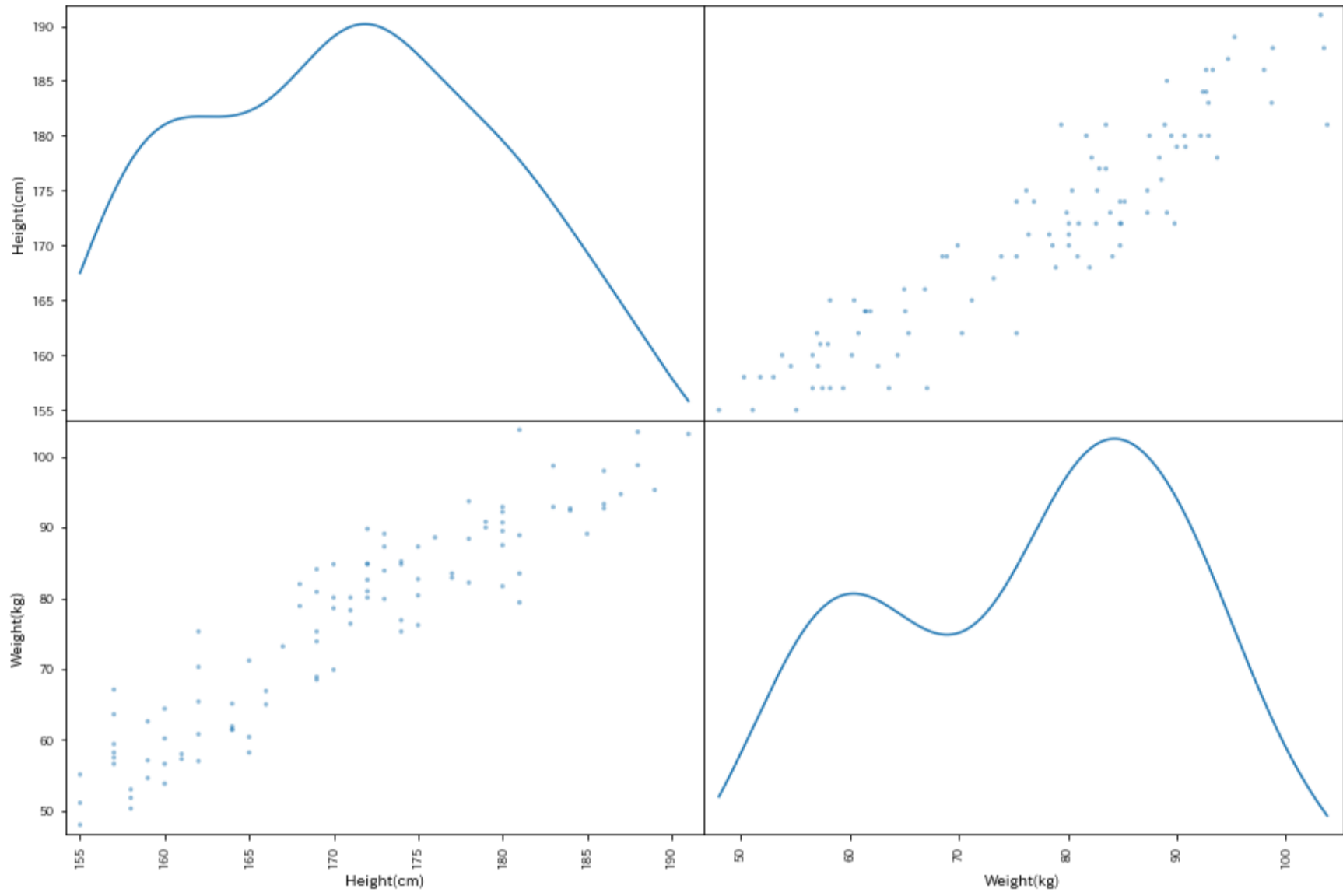
	Gender	Height	Weight	Height(cm)	Weight(kg)	BMI
507	Male	63.724727	165.979417	162	75.3	28.7
818	Female	62.968382	124.773903	160	56.6	22.1
452	Male	70.643193	198.491168	179	90.0	28.1
368	Male	70.035068	181.324779	178	82.2	25.9

	Gender	Height	Weight	Height(cm)	Weight(kg)	BMI
242	Male	75.259457	227.437138	191	103.2	28.3

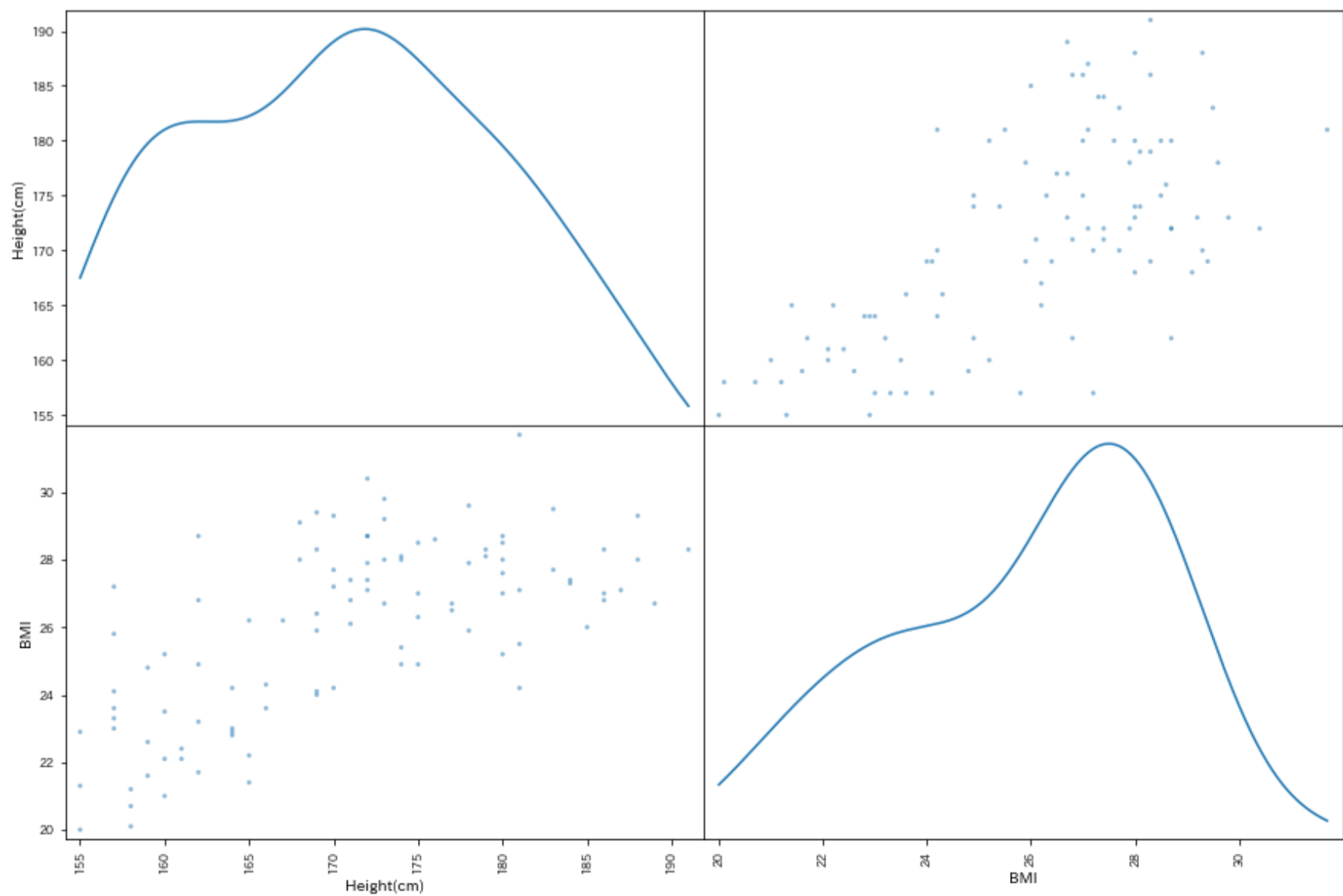
```
In [16]: %matplotlib inline
```

```
In [17]: import japanize_matplotlib
import matplotlib.pyplot as plt

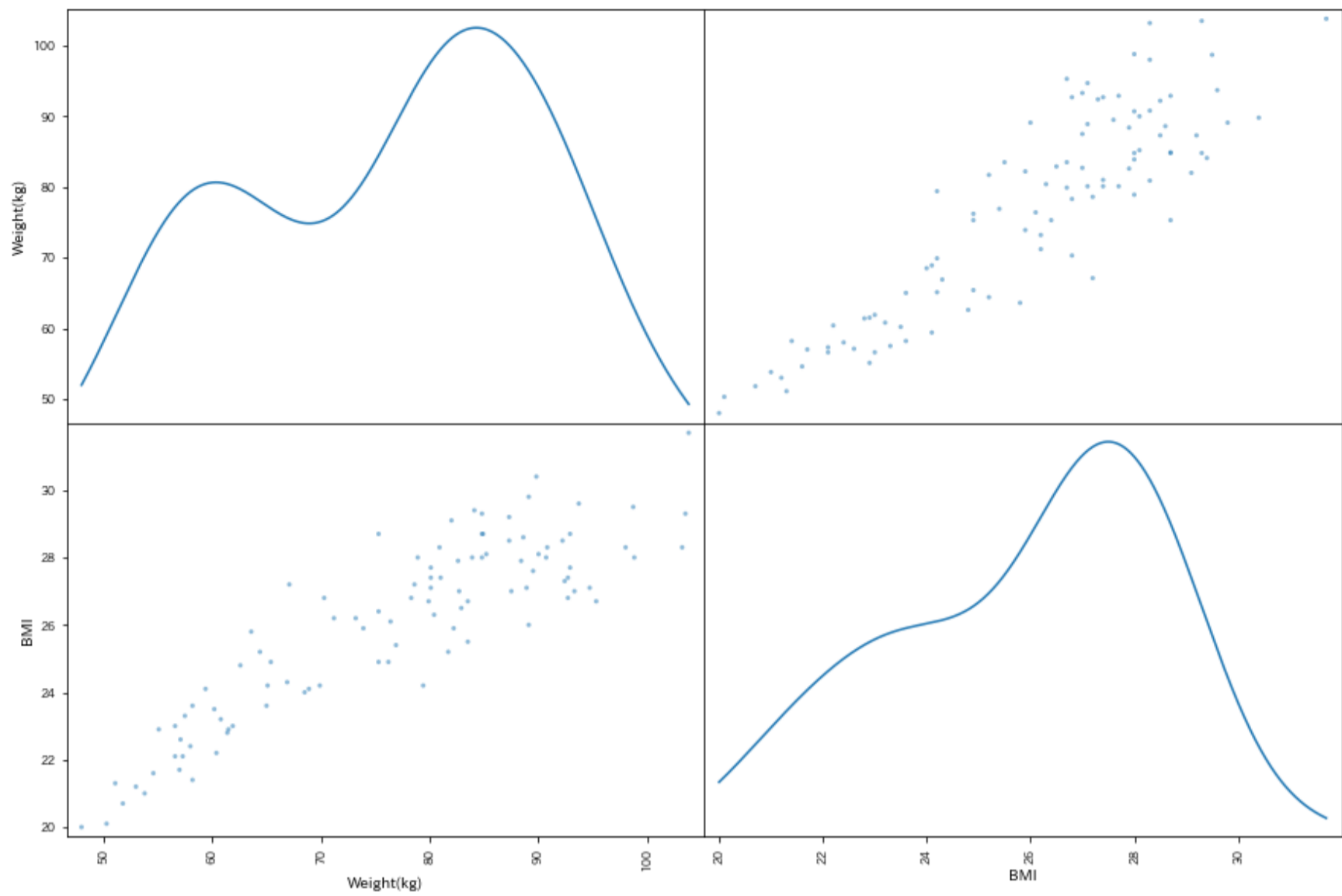
ax = pd.plotting.scatter_matrix(df2[['Height(cm)', 'Weight(kg)']], figsize=[15, 10], diagonal='kde')
```



```
In [22]:
```



```
In [21]:
```



In []:

In [18]:

```
df2[['Height(cm)', 'Weight(kg)']].head()
```

Out[18]:

	Height(cm)	Weight(kg)
507	162	75.3
818	160	56.6
452	179	90.0
368	178	82.2
242	191	103.2

線形回帰モデルの作成

- X = 体重
- Y = BMI
- a = 傾きの値
- b = 切片

In [24]:

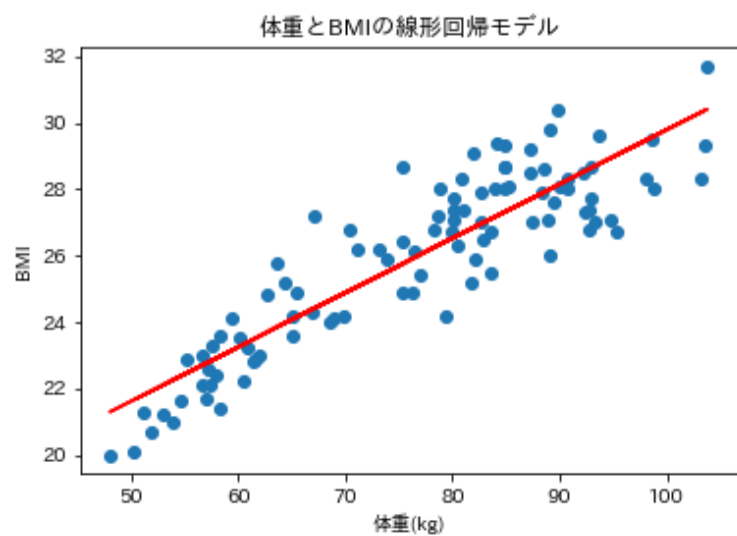
```
#コードを入力する
```

a 0.16350027835569186 b 13.438658182686044

モデルのプロット

In [26]:

```
# グラフを描く
```



同様に1000個のデータを分析しましょう

- df1のデータフレーム

In []: