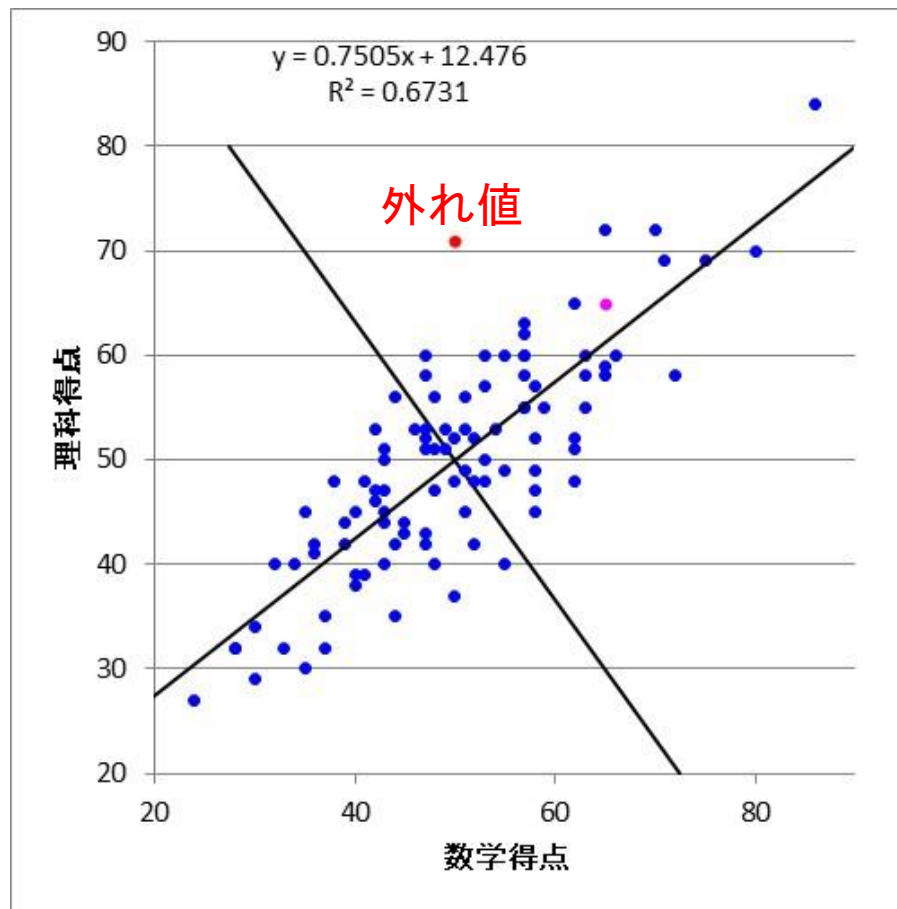


# マハラノビス距離をpythonを用いて計算

## マハラノビス距離の一般式

$$D_M(x, y) = \sqrt{(x - \mu_x, y - \mu_y)^T \Sigma^{-1} (x - \mu_x, y - \mu_y)}$$

$\mu_x$  及び  $\mu_y$  : 平均値  
 $(x - \mu_x, y - \mu_y)^T$  :  $(x - \mu_x, y - \mu_y)$  の転置行列  
 $\Sigma^{-1}$  : 共分散行列の逆行列



| 数学 | 理科 |
|----|----|
| M  | S  |
| 39 | 44 |
| 63 | 58 |
| 63 | 60 |
| 43 | 50 |
| 66 | 60 |
| 71 | 69 |
| 55 | 60 |
| 48 | 47 |

}

|    |    |
|----|----|
| 41 | 48 |
| 43 | 47 |
| 28 | 32 |
| 57 | 55 |
| 53 | 60 |
| 65 | 58 |
| 48 | 40 |
| 53 | 57 |

- ①pythonで処理する準備として  
左記Excelファイルを  
名前を付けて保存  
CSV(コンマ区切り) (\*.CSV)

今回、MT\_data.csvで保存

```
# CSV読み込み用
import pandas as pd

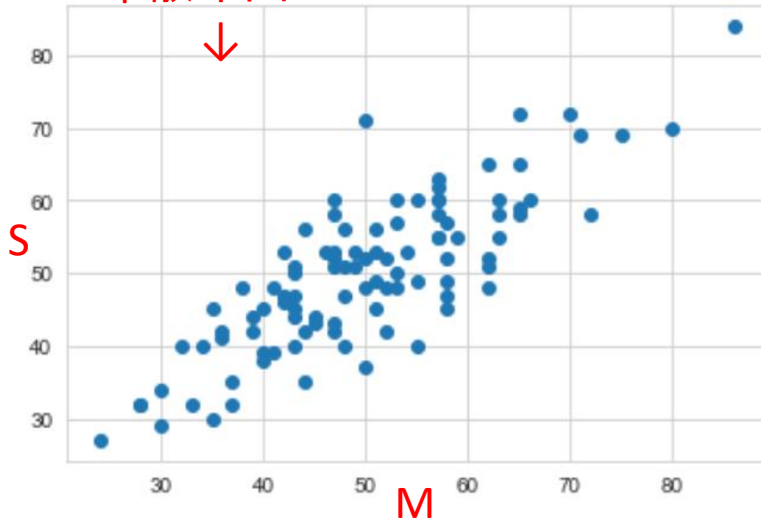
# 計算用
import numpy as np
from scipy.spatial import distance
```

```
# グラフ用
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set_style("whitegrid")
```

```
# データの読み込み
data = pd.read_csv("MT_data.csv") ←今回pandasで読み込む
data.head()| ←1行目をヘッダーとする→
```

```
# データをMとS絞る
data = data[['M', 'S']]
plt.scatter(data['M'].values, data['S'].values)
```

↑散布図



| M  | S  |
|----|----|
| 39 | 44 |
| 63 | 58 |
| 63 | 60 |
| 43 | 50 |
| 66 | 60 |
| 71 | 69 |
| 55 | 60 |
| 48 | 47 |

```
# 平均
mean = np.mean(data, axis=0)
print(mean)
#標準偏差
#sigma=np.std(data, axis=0)
#print(sigma)
# 平均引いた値
data_m = data - mean
print(data_m)
```

```
# 共分散
cov = np.cov(data.T)
print(cov)
print(np.linalg.pinv(cov))
```

```
M    50.0
S    50.0
dtype: float64
      M    S
0  -11.0  -6.0
1   13.0   8.0
2   13.0  10.0
3   -7.0   0.0
4   16.0  10.0
..    ..
95    7.0   5.0
96    3.0  10.0
97   15.0   8.0
98   -2.0 -10.0
99    3.0   7.0
```

共分散行列

逆行列

```
[[134.88888889 101.23232323]
 [101.23232323 112.86868687]]
[[ 0.0226792 -0.02034106]
 [-0.02034106  0.02710382]]
```

# Numpyをそのまま用いた場合

`np. dot(a,b,axis=1)` ←ベクトルaとbの内積 axis=1は列

# pythonの関数の計算

```
result_1 = np.sqrt(np.sum(np.dot(data_m, np.linalg.pinv(cov)) * data_m, axis=1))
```

```
print(result_1)
```

```
plt.plot(result_1)
```

$$D_M(x, y) = \sqrt{(x - \mu_x, y - \mu_y)^T \Sigma^{-1} (x - \mu_x, y - \mu_y)}$$

pandasの配列を  
numpy配列に変換

```
0    1.017301
1    1.156067
2    1.120041
3    1.054173
4    1.416728
...
95   0.604154
96   1.301550
97   1.398432
98   1.409772
99   0.823332
Length: 100, dtype: float64
```

## Scipyにあるdistance.mahalanobisを用いた場合

# distance.mahalanobisを使ってみた場合

```
data = data.to_numpy()
```

```
data_m = data_m.to_numpy()
```

```
result_2 = []
```

```
# データ: data[i], 平均値: mean, 共分散行列の逆行列: np.linalg.pinv(cov) から距離を計算
for i in range(len(data_m)):
```

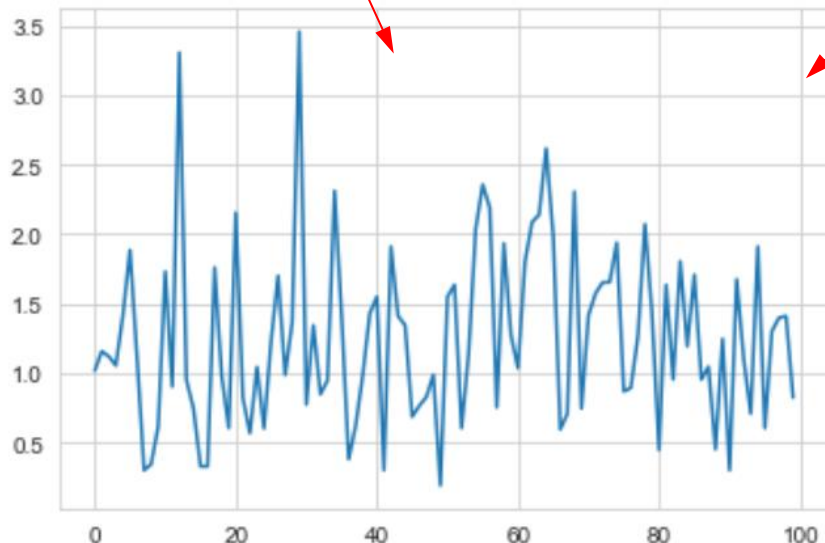
```
    result_2.append(distance.mahalanobis(data[i], mean, np.linalg.pinv(cov)))
```

```
# 結果
```

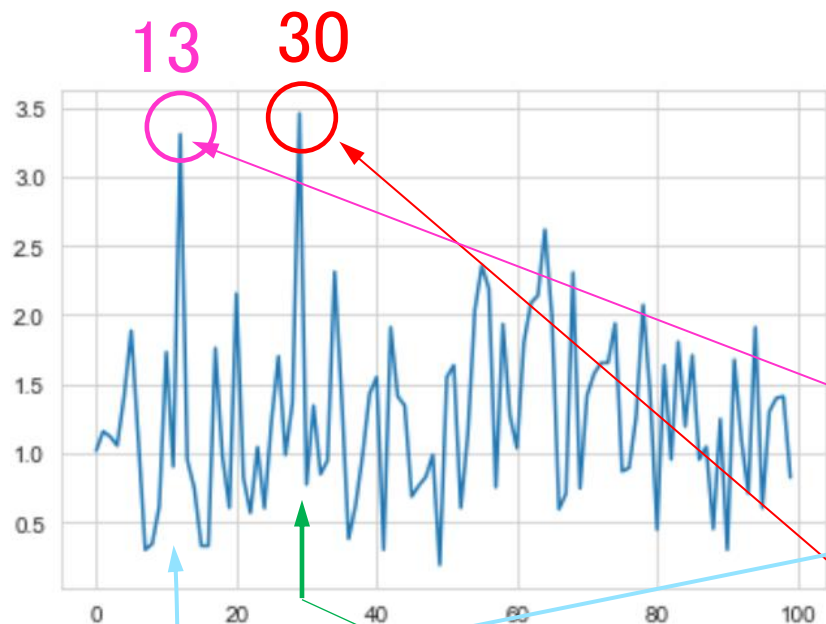
```
print(result_2)
```

```
plt.plot(result_2)
```

```
[1.0173010824347304, 1.1560665324461294, 1.1200412282093855, 1.0541730987294882, 1.4167284738800983, 1.885164081
5041744, 1.1150140193317954, 0.30092939040628164, 0.3444070959181984, 0.6151175771282511, 1.7297347499381068, 0.
9023226756224968, 3.3059573619482276, 0.9506236040795093, 0.7460702228237485, 0.32926474875193545, 0.32926474875
193545, 1.7592066251834595, 0.9810129958340037, 0.6041543028296548, 2.1528994999527686, 0.823331985106916, 0.567
6743206056694, 1.0435747412113976, 0.6015484504149979, 1.2306854669082852, 1.6991762965610149, 0.986871245808530
5, 1.3660319614349778, 3.457279861895322, 0.7735336091645513, 1.34314172188579, 0.8479579039861176, 0.9480516450
532579, 2.3110943685712004, 1.4135530838101287, 0.3802199766327757, 0.623425946540941, 0.9868712458085305, 1.430
9801743300157, 1.5484991807526158, 0.30077422520749897, 1.9100404105137165, 1.4105863904537674, 1.34392878175476
67, 0.6864351465760886, 0.7637423851413677, 0.8269919997840295, 0.9868712458085305, 0.19079735657733543, 1.54849
91807526158, 1.6345562237045403, 0.6041543028296548, 1.1353486412113387, 2.033459636343894, 2.356125304391862,
2.18997063848002, 0.7534765888713665, 1.9338340229113782, 1.2741159240012347, 1.0355915414963488, 1.803724006797
585, 2.08580707504559, 2.14022086688758, 2.6160163800060268, 1.9844337798953744, 0.5943593031250974, 0.708066188
2621826, 2.3046621443942223, 0.7460702228237485, 1.4167284738800983, 1.573618448079614, 1.6520464659136547, 1.65
3983999568059, 1.938130274613531, 0.8685182757074063, 0.8930254463372057, 1.263698259542792, 2.0711830829926976,
1.4287700861004522, 0.44630130560948644, 1.6345562237045403, 0.9546779814267096, 1.8046453512449936, 1.192962503
3220506, 1.7073745750461218, 0.9533994023435154, 1.0464162501023033, 0.4517884708840664, 1.246851893081882, 0.30
077422520749897, 1.674730908860643, 1.1014320841590834, 0.7077364690720067, 1.9100404105137165, 0.60415430282965
48, 1.3015495729097084, 1.3984317477546344, 1.4097717588083294, 0.823331985106916]
```



マハラビス距離



| No. | 数学M | 理科S |
|-----|-----|-----|
| 1   | 39  | 44  |
| 2   | 63  | 58  |
| 3   | 63  | 60  |
| 4   | 43  | 50  |
| 5   | 66  | 60  |
| 6   | 71  | 69  |
| 7   | 55  | 60  |
| 8   | 48  | 47  |
| 9   | 54  | 53  |
| 10  | 43  | 44  |
| 11  | 30  | 34  |
| 12  | 47  | 53  |
| 13  | 86  | 84  |
| 14  | 51  | 45  |

|    |    |    |
|----|----|----|
| 28 | 57 | 60 |
| 29 | 63 | 55 |
| 30 | 50 | 71 |
| 31 | 44 | 42 |
| 32 | 58 | 49 |

}

