

## 공분산 행렬

```
: import numpy as np # 넘파이  
import pandas as pd # 판다스  
import matplotlib.pyplot as plt #매트립
```

```
: data=pd.read_csv("test2.csv") # csv불러오기  
data
```

```
:      height  weight  
0         170      65  
1         160      55  
2         180      70  
3         170      60  
4         175      62  
5         180      72  
6         160      60  
7         165      58  
8         186     100  
9         172      67
```

```
: height=data['height'] # height 에있는 데이터만  
weight=data['weight'] # weight 에 있는 데이터만
```

```
: height
```

```
: 0    170  
1    160  
2    180  
3    170  
4    175  
5    180  
6    160  
7    165  
8    186  
9    172  
Name: height, dtype: int64
```

일단 위에 코드를보면 test2.csv 파일에 height weight를 엑셀로 만들어서 저장합니다.  
height와 weight의 column은 feature 특징으로 feature data 라고 불립니다.  
height=data['height']에서 height 만 빼서 쭉나열하고 weight=data['weight']에서  
weight 만쭉나열합니다.

```

: weight
:
0    65
1    55
2    70
3    60
4    62
5    72
6    60
7    58
8   100
9    67
Name: weight, dtype: int64

: height=height # 배열로 변경
weight=weight # 배열로 변경

: height
: array([170, 160, 180, 170, 175, 180, 160, 165, 186, 172], dtype=int64)

: weight
: array([ 65,  55,  70,  60,  62,  72,  60,  58, 100,  67], dtype=int64)

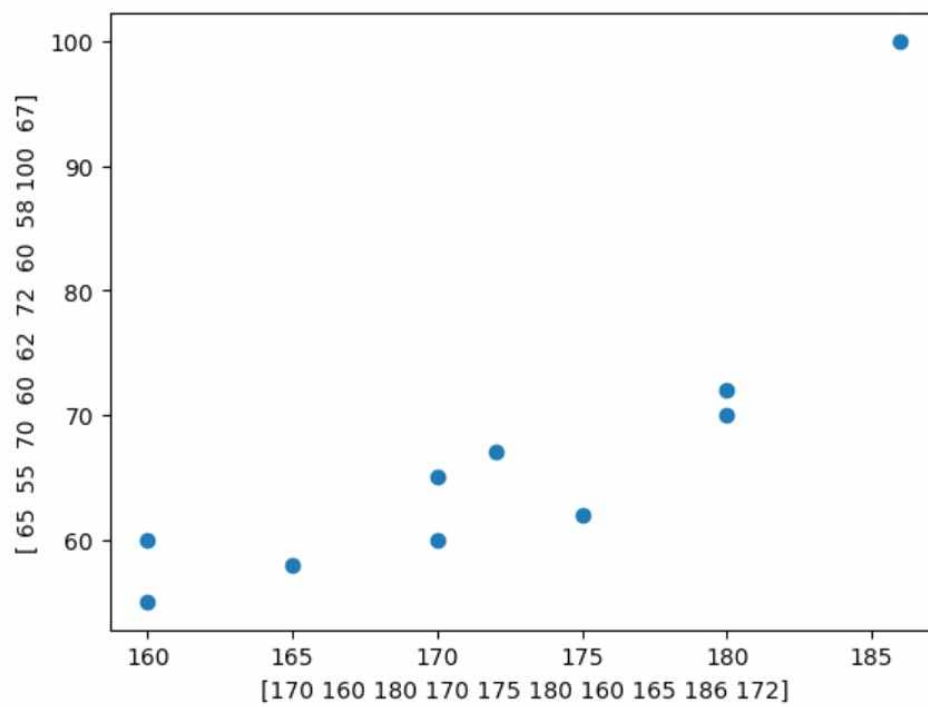
: cov_matrix=np.cov(height,weight) # 공분산행렬로
cov_matrix
: array([[ 75.28888889,  91.08888889],
        [ 91.08888889, 163.87777778]])

: np.corrcoef(height,weight) # 상관계수
: array([[1.          ,  0.82004923],
        [ 0.82004923,  1.          ]])

: plt.scatter(height,weight) # 두변수의 상관관계를 직교좌표계에 평면에 점으로 표현하는 그래프 입니다.
plt.xlabel(height) # x좌표에 height
plt.ylabel(weight) # y좌표에 weight
plt.show() # 보여주기

```

그리고 height=np.array(height)에서 특징들만모아서 배열로 나열합니다.  
weight=np.array(weight) 에서도 특징들만 모아서 배열로 나열합니다.  
그리고 공분산행렬로 만듭니다. 그러면 결과가 나오는데요 여기에서 주목할점은  
(height의 분산, height,weight공분산)  
(weight,height의공분산,weight의분산)  
입니다. 사실 분산은 신경 안써도 되고 공분산만 신경쓰면됩니다.  
그리고 스케일링 한후에 상관계수를 구하면 0.8204923으로 1에 가깝습니다 .  
그리고 plt.scatter로 점으로 표현을하면



이렇게 결과가 나옵니다. x가증가할때에 y도 같이 증가합니다