과제1

- 컴퓨터공학과 12171676 이종법
- 언어: python / 문제풀이는 손으로 풀고 pdf 합치기했습니다.

```
In [1]:

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import os
from glob import glob
import sys

In [2]:

# 시 각화 패키지를
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

plt.style.use("ggplot")
```

문제 1.

- (1) scatter plot 및 해석
- (2) x_bar, 공분산 행렬, 상관계수 행렬

```
In [3]: # 파일 불러오기
X = pd.DataFrame( np.loadtxt('../data/P1-4.DAT', unpack = True).T,columns=['x1','x2','x3']
X
```

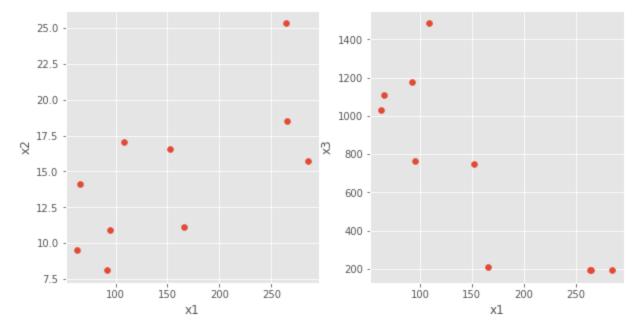
```
Out[3]:
               х1
                     x2
                             х3
         0 108.28 17.05 1484.10
         1 152.36 16.59
                          750.33
            95.04 10.91
                          766.42
            65.45 14.14 1110.46
            62.97 9.52 1031.29
         5 263.99 25.33
                         195.26
         6 265.19 18.54
                         193.83
         7 285.06 15.73
                         191.11
             92.01
                  8.10 1175.16
         9 165.68 11.13
                          211.15
```

```
In [4]: plt.figure(figsize=(10,5))
    plt.subplot(1,2,1)
```

```
plt.scatter(X['x1'],X['x2'])
plt.xlabel("x1")
plt.ylabel("x2")

plt.subplot(1,2,2)

plt.scatter(X['x1'],X['x3'])
plt.xlabel("x1")
plt.ylabel("x3")
plt.ylabel("x3")
```



x1 과 x2는 양의 상관관계를 가지고, x1과 x3는 음의 상관관계를 가진다.

```
In [5]:
        print("x bar : ", np.array(X).mean(axis=0))
       x bar : [155.603 14.704 710.911]
In [6]:
        print("S n(공분산 행렬) : ")
        print(X.cov())
       S n(공분산 행렬) :
                                  x2
            7476.453246
                          303.618620 -35575.959570
       x1
             303.618620
                           26.190316
                                      -1053.827393
       x3 -35575.959570 -1053.827393 237054.269832
In [7]:
        print("Sample Correlation R : ")
        print(X.corr())
       Sample Correlation R :
                          x2
       x1 1.000000 0.686136 -0.845055
       x2 0.686136 1.000000 -0.422937
       x3 -0.845055 -0.422937 1.000000
```

문제 2

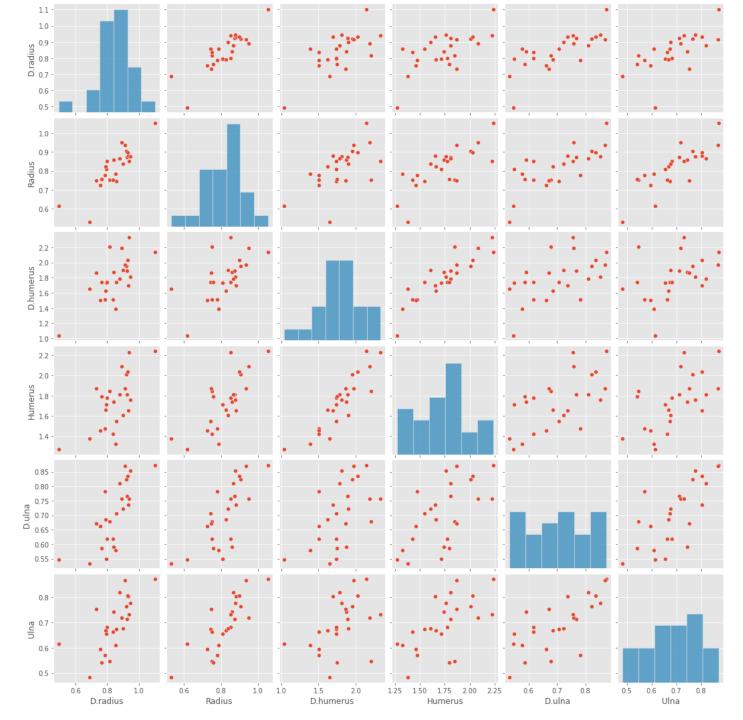
- (1) Pairwise Scatter Plot
- (2) x_bar, 공분산 행렬, 상관계수 행렬

```
In [8]: # 파일 불러오기
X = pd.DataFrame(np.loadtxt('../data/T1-8.DAT', unpack = True).T,columns=['D.radius','Rac
```

Out[8]:		D.radius	Radius	D.humerus	Humerus	D.ulna	Ulna
	0	1.103	1.052	2.139	2.238	0.873	0.872
	1	0.842	0.859	1.873	1.741	0.590	0.744
	2	0.925	0.873	1.887	1.809	0.767	0.713
	3	0.857	0.744	1.739	1.547	0.706	0.674
	4	0.795	0.809	1.734	1.715	0.549	0.654
	5	0.787	0.779	1.509	1.474	0.782	0.571
	6	0.933	0.880	1.695	1.656	0.737	0.803
	7	0.799	0.851	1.740	1.777	0.618	0.682
	8	0.945	0.876	1.811	1.759	0.853	0.777
	9	0.921	0.906	1.954	2.009	0.823	0.765
	10	0.792	0.825	1.624	1.657	0.686	0.668
	11	0.815	0.751	2.204	1.846	0.678	0.546
	12	0.755	0.724	1.508	1.458	0.662	0.595
	13	0.880	0.866	1.786	1.811	0.810	0.819
	14	0.900	0.838	1.902	1.606	0.723	0.677
	15	0.764	0.757	1.743	1.794	0.586	0.541
	16	0.733	0.748	1.863	1.869	0.672	0.752
	17	0.932	0.898	2.028	2.032	0.836	0.805
	18	0.856	0.786	1.390	1.324	0.578	0.610
	19	0.890	0.950	2.187	2.087	0.758	0.718
	20	0.688	0.532	1.650	1.378	0.533	0.482
	21	0.940	0.850	2.334	2.225	0.757	0.731
	22	0.493	0.616	1.037	1.268	0.546	0.615
	23	0.835	0.752	1.509	1.422	0.618	0.664
	24	0.915	0.936	1.971	1.869	0.869	0.868

```
In [9]: sns.pairplot(X)
```

ut[9]: <seaborn.axisgrid.PairGrid at 0x20a4684b788>



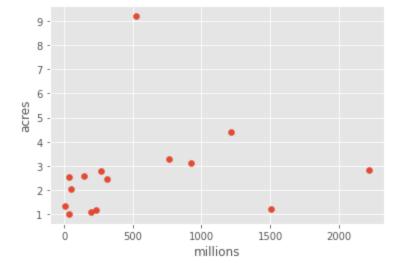
RADIUS, HUMERUS, uLna 간에 모두 양의 상관관계를 보인다

```
In [10]:
         print("x bar : ",np.array(X).mean(axis=0))
         x bar : [0.8438  0.81832  1.79268  1.73484  0.7044  0.69384]
In [11]:
         print("S n(공분산 행렬) : ")
         print(X.cov())
         S n(공분산 행렬) :
                    D.radius
                                 Radius
                                         D.humerus
                                                      Humerus
                                                                 D.ulna
                                                                              Ulna
         D.radius
                    0.013002
                               0.010378
                                                     0.020086
                                                               0.009121
                                                                          0.007958
                                          0.022350
         Radius
                    0.010378
                               0.011418
                                          0.018535
                                                     0.021100
                                                               0.008530
                                                                          0.008909
         D.humerus
                    0.022350
                               0.018535
                                          0.080357
                                                     0.066776
                                                               0.016837
                                                                          0.012847
         Humerus
                    0.020086
                               0.021100
                                          0.066776
                                                     0.069484
                                                               0.017735
                                                                          0.016794
         D.ulna
                    0.009121
                               0.008530
                                          0.016837
                                                     0.017735
                                                               0.011568
                                                                          0.008071
                                                     0.016794
                                                               0.008071
                                                                          0.010599
         Ulna
                    0.007958
                               0.008909
                                          0.012847
```

```
print("Sample Correlation R : ")
        print(X.corr())
        Sample Correlation R :
                   D.radius Radius D.humerus Humerus
                                                             D.ulna
                                                                          Ulna
                  1.000000 0.851807 0.691459 0.668258 0.743693 0.677894
        D.radius
        Radius
                  0.851807 1.000000 0.611916 0.749093 0.742178 0.809798
        D.humerus 0.691459 0.611916 1.000000 0.893646 0.552222 0.440205
                  0.668258 0.749093 0.893646 1.000000 0.625550 0.618820
        Humerus
        D.ulna
                  0.743693  0.742178  0.552222  0.625550  1.000000  0.728892
        Ulna
                  0.677894 0.809798
                                       0.440205 0.618820 0.728892 1.000000
       문제 3.
        • (a) scatter plot과 r
        • (b) 세 개의 outlier 제거와, r
        • (c) m^2 로 바뀔시 상관계수와 그 이유
In [4]:
        # 파일 불러오기
        X = pd.DataFrame( np.loadtxt('../data/T1-11.DAT', unpack = True).T,columns=['millions','ac
           millions acres
Out[4]:
         0
              47.4
                   2.05
         1
              35.8
                   1.02
         2
              32.9
                   2.53
         3
            1508.5
                   1.23
            1217.4
                   4.40
         5
             310.0
                   2.46
         6
             521.8
                   9.19
         7
              5.6
                   1.34
         8
             922.7
                   3.14
         9
             235.6
                   1.17
        10
             265.8
                   2.80
        11
             199.0
                   1.09
            2219.8
                   2.84
        12
             761.3
        13
                   3.30
        14
             146.6
                   2.59
```

```
In [5]: plt.scatter(X['millions'],X['acres'])
   plt.xlabel("millions")
   plt.ylabel("acres")
```

Out[5]: Text(0, 0.5, 'acres')



```
In [6]:
r=X.corr()["acres"][0]
print("<mark>상관계수 : ",</mark>r)
#이상치 때문에 상관계수가 낮은것으로 생각된다.
#scatter plot에서는 outlier 1개, leverage 2개로 관측됨.
```

상관계수 : 0.1725274227916501

• outlier 제거와 r

```
In [7]:
         plt.subplot(1,2,1)
         plt.boxplot(X['millions'])
         plt.subplot(1,2,2)
         plt.boxplot(X['acres'])
        {'whiskers': [<matplotlib.lines.Line2D at 0x2f91f6f0448>,
Out[7]:
          <matplotlib.lines.Line2D at 0x2f91f6f0c08>],
         'caps': [<matplotlib.lines.Line2D at 0x2f91f6f0388>,
          <matplotlib.lines.Line2D at 0x2f91f6b3fc8>],
         'boxes': [<matplotlib.lines.Line2D at 0x2f91f6e9ac8>],
         'medians': [<matplotlib.lines.Line2D at 0x2f91f69ed88>],
         'fliers': [<matplotlib.lines.Line2D at 0x2f91f6f05c8>],
         'means': []}
                                    9
         2000
                                    8
                                    7 -
        1500 -
                                    6
                                    5 -
         1000
                                    3 -
         500
                                    2 -
           0 -
```

```
In [8]: term1= X['millions']>1500
  term2 =X['acres']>5

X[ term1 | term2]
  #outlier 37#
```

```
    millions
    acres

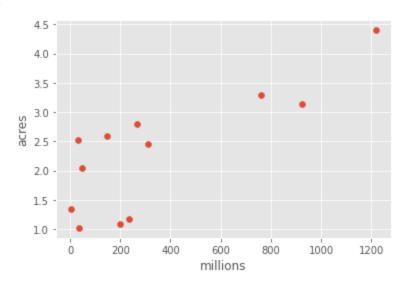
    3
    1508.5
    1.23

    6
    521.8
    9.19

    12
    2219.8
    2.84
```

```
In [9]: X=X.loc[(X['millions']<1500) & (X['acres']<5) ]
    X=X.reset_index(drop=True)
    plt.scatter(X['millions'], X['acres'])
    plt.xlabel("millions")
    plt.ylabel("acres")</pre>
```

Out[9]: Text(0, 0.5, 'acres')



```
In [10]: r=X.corr()['acres'][0] print("outlier가 없는 상관계수 r : ",r)
```

outlier가 없는 상관계수 r : 0.8024842602787746

• acres^2 시 상관계수

$$r = r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

상관계수는 scale(단위)에는 불변하다.

scale은 서로 상쇄돼기 때문이다.

acres와 m^2이 서로간에 상수를 곱하여 변환이 된다면, 상쇄될 것이다.

그래서, 변하지 않을 것으로 생각된다.

```
In [20]: X['acres']=X['acres']*4046.9 r=X.corr()['acres'][0] print("데이터 변환시 상관계수 r : ",r)
```

데이터 변환시 상관계수 r: 0.8024842602787746

결과 변하지 않았다.

문제 4번 부터는 손풀이

INHA UNIVERSITY

100, INHA-RO, NAM-GU, INCHEON 22212, KOREA

4.

(i) ||x||

$$Cos 0 = \frac{1}{\sqrt{36} \times \sqrt{11}} = 0.0509$$

(iii) YEXOI Projection

$$\frac{1}{35} \cdot \mathcal{X} = \frac{1}{35} \left[5, 1, 37 \right] \rightarrow \text{Projection of } 4 \cdot 0.0 \times 1$$

$$A = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}$$

(a) A) & STWMCHIC 22 GOICK.

Positive Letilite 94?

$$[x_1, x_2][q-2][x_1] = [qx_1-2x_2, -2x_1+6x_2][x_1] = qx_1^2 - 2x_1x_2 - 2x_1x_1+6x_2^2$$

$$[-2, 6][x_1] = [qx_1-2x_2, -2x_1+6x_2][x_1] = qx_1^2 - 4x_1x_2 + 6x_2^2$$

, Az Positive definite ofet



$$\det(A-\lambda I)=|A-\lambda I|=0 \rightarrow A-\lambda I=\begin{pmatrix} q-\lambda & -2\\ -2 & 6-\lambda \end{pmatrix}$$

$$|A-\lambda I| = (9-\lambda)(6-\lambda)-4 = (\lambda-10)(\lambda-5) = 0.$$

.. eigen Value:
$$\lambda_1=10$$
, $\lambda_2=5$.

$$(A-10.I)\chi = \begin{pmatrix} -1 & -2 \\ -2 & -4 \end{pmatrix} \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} -\chi_1 - \chi_2 \\ -2\chi_1 - 4\chi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$(A-5I)\chi = \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} 4\chi_1 - 2\chi_2 \\ -2\chi_1 + \chi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\frac{1}{2} \chi_{1} = \frac{1}{2} \chi_{2} / \chi_{1}^{2} + 4 \chi_{1}^{2} = 1 \\
\frac{1}{2} e_{2} = \left(\frac{1}{\sqrt{5}}, \sqrt{5} \right)$$

(-) () 3×3 5×1



INHA UNIVERSITY

100 INHA-RO, NAM-GU, INCHEON 22212, KOREA

5-d). An Gloc spectral decomposition

$$A = 10 \cdot \left(-\frac{1}{5}\right) \left(-\frac{2}{5}\right) + 5 \cdot \left(-\frac{2}{5}\right) + 5 \cdot \left(-\frac{2}{5}\right) = 10 \cdot \left(-\frac{2}{5}\right) + 5 \cdot \left(-\frac{2}{5}\right) = 10 \cdot \left(-\frac{2}{5}\right) + 5 \cdot \left(-\frac{2}{5}\right) = 10 \cdot \left(-\frac{2$$

$$A \pm \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}, A^{-1} = 1 \begin{bmatrix} d - b \\ -2 & 0 \end{bmatrix} = 1 \begin{bmatrix} 6 & 2 \\ 50 \begin{bmatrix} 2 & 9 \end{bmatrix}$$

$$\lambda^{2} - \frac{15}{50}\lambda + \frac{59}{50^{2}} - \frac{4}{50^{2}}$$

$$=(\lambda - \frac{1}{50})(\lambda - \frac{5}{50}) = (\lambda - \frac{1}{5})(\lambda - \frac{1}{10})$$

=)
$$2x_{12}x_{2}$$
 => $e_{1}=\left(\frac{1}{15},\frac{2}{15}\right)^{1/(2)}$

2)
$$\lambda_{2} = \frac{1}{10}$$
 $\begin{bmatrix} \frac{6}{50} & \frac{1}{25} \\ \frac{1}{25} & \frac{4}{50} \end{bmatrix} - \begin{bmatrix} \frac{1}{0} & \frac{1}{0} \\ \frac{1}{25} & \frac{1}{25} \end{bmatrix} = \begin{bmatrix} \frac{1}{50} & \frac{1}{25} \\ \frac{1}{25} & \frac{1}{25} \end{bmatrix} = \begin{bmatrix} \frac{1}{50} & \frac{1}{25} \\ \frac{1}{25} & \frac{1}{25} \end{bmatrix} = \begin{bmatrix} \frac{1}{50} & \frac{1}{25} \\ \frac{1}{25} & \frac{1}{25} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$$\Rightarrow \quad \chi_{1} = -\frac{1}{2}\chi_{2} \Rightarrow e_{1} = (\frac{2}{55}, \frac{1}{55}), e_{2} = (\frac{2}{55}, \frac{1}{55}), e_{2} = (\frac{2}{55}, \frac{1}{55})$$

$$\therefore \quad \lambda_{1} = \frac{1}{5}, \quad \lambda_{2} = \frac{1}{10}, \quad \lambda_{2} = \frac{1}{10}, \quad \xi_{1} = (\frac{1}{55}, \frac{2}{55}), e_{2} = (\frac{2}{55}, \frac{1}{55})$$



(a)
$$E(X^{(1)})$$
, $Cov(X^{(1)})$

$$E\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, Cov\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

$$E(\begin{bmatrix} \times 3 \end{bmatrix}) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, COV(\begin{bmatrix} \times 3 \\ \times 4 \end{bmatrix}) = \begin{bmatrix} 9 & -2 \\ -2 & 4 \end{bmatrix}$$

$$\mathcal{Z}_{\mathsf{X}} = \left(\begin{array}{c} \mathcal{L}_{\mathsf{A}}(\mathsf{X}^{(1)}) \\ \mathcal{L}_{\mathsf{A}}(\mathsf{A}^{(1)}) \\ \mathcal{L}_{\mathsf{A}}(\mathsf{A}^{($$

$$E(AX^{(1)}) = AE(X^{(1)}) = [1,2][4] = 10$$

$$E(A|X^{(1)}) = AE(X^{(1)}) = [1,2][4] = 10$$

$$Cov(A|X^{(1)}) = ACov(X^{(1)})A^{T} = [1,2][3][3] = [3,2][1]$$

$$E(B \times (2)) = B \cdot E(\times (2)) = \begin{bmatrix} 1 & -2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$Cov(Bx^{(2)}) = BCov(x^{(2)})B^{T} = \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 9 & -2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} - \begin{bmatrix} 33 & 36 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -2 & 4 \end{bmatrix} \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 36 & 48 \end{bmatrix}$$



INHA UNIVERSITY

100, INHA-RO, NAM-GU, INCHEON 22212, KOREA

$$(6-+), \quad Cov(AX^{(1)}, \beta X^{(2)})$$

$$Cov(Ax^{(1)}, Bx^{(2)}) = E(A(x^{(1)} - E(x^{(1)}))(x^{(2)} - E(x^{(2)}))B^{T})$$

$$= A(Cov(x^{(1)}, x^{(2)})B^{T})$$

$$= [1,2](2 2)[1 2]$$

$$= [1,2](2 2)[2 -1]$$

$$= [4,2][1 2] = [06]$$



INHA-RO, NAM-GU, INCHEON 22212, KOREA

图刊 7.)

(a) Art Veal eigen Value ? 7+ 3 32

$$(3)$$
 (3) $(4+d)$ $(4+d)$

a2-2ad + 46+ d 2 0 (b)

(=) (a-1)2+ (26)270 ... orly a, d, b = 2001 300

44 cisas Valuet द्रयांच्य

(C)

(=)

5:60

:. 0=0, 1,21 get Afeifen value 24.

