

# REPORT



|      |   |            |
|------|---|------------|
| 수강과목 | : | 다변량통계학(I)  |
| 담당교수 | : | 최용석        |
| 학 과  | : | 통계학과       |
| 학 번  | : | 201611531  |
| 이 름  | : | 정호재        |
| 제출일자 | : | 2020.05.26 |

# HW2 for Multivariate Statistics I

May 19, 2020

## Chapter 2. Principal Component Analysis

Consider the Exercise 2.8 and Exercise 2.9 in page 126.

1. [Data 2.8.2](airpollution.txt) indicates 42 measurements recorded on the 7 air-pollution variables(wind, Solar radiation,  $CO$ ,  $NO$ ,  $NO_2$ ,  $O_3$ ,  $HC$ ) in the Los Angeles at 12:00 noon on different days.

(1) Perform PCA on a given data using the covariance matrix  $S$  and correlation matrix  $R$ , and compare the two results.

[1단계] 다변량 자료행렬  $X$ 를 준비한다.

```
> head(data)
  Wind Solar CO NO NO2 O3 HC
1    8   98  7  2  12  8  2
2    7  107  4  3   9  5  3
3    7  103  4  3   5  6  3
4   10   88  5  2   8 15  4
5    6   91  4  2   8 10  3
6    8   90  5  2  12 12  4

> dim(data)
[1] 42  7
```

[자료 2.8.2](airpollution)의 자료는 변수는  $p=7$ 이고 개체 수는  $n=42$ 인  $42 \times 7$  다변량 자료행렬이다.

[2단계] 공분산행렬 S를 구한다.

```
> S
      Wind      Solar      CO      NO      NO2      O3      HC
Wind  2.5000000 -2.7804878 -0.3780488 -0.4634146 -0.5853659 -2.2317073 0.1707317
Solar -2.7804878 300.5156794  3.9094077 -1.3867596  6.7630662 30.7909408 0.6236934
CO    -0.3780488  3.9094077  1.5220674  0.6736353  2.3147503  2.8217189 0.1416957
NO    -0.4634146 -1.3867596  0.6736353  1.1823461  1.0882695 -0.8106852 0.1765389
NO2   -0.5853659  6.7630662  2.3147503  1.0882695 11.3635308  3.1265970 1.0441347
O3    -2.2317073 30.7909408  2.8217189 -0.8106852  3.1265970 30.9785134 0.5946574
HC     0.1707317  0.6236934  0.1416957  0.1765389  1.0441347  0.5946574 0.4785134
```

다른 변수들에 비해 Solar의 분산이 매우 크다.

[2단계] 상관행렬 R를 구한다.

```
> R
      Wind      Solar      CO      NO      NO2      O3      HC
Wind  1.0000000 -0.10144191 -0.1938032 -0.26954261 -0.1098249 -0.2535928 0.15609793
Solar -0.1014419  1.00000000  0.1827934 -0.07356907  0.1157320  0.3191237 0.05201044
CO    -0.1938032  0.18279338  1.0000000  0.50215246  0.5565838  0.4109288 0.16603235
NO    -0.2695426 -0.07356907  0.5021525  1.00000000  0.2968981 -0.1339521 0.23470432
NO2   -0.1098249  0.11573199  0.5565838  0.29689814  1.0000000  0.1666422 0.44776780
O3    -0.2535928  0.31912373  0.4109288 -0.13395214  0.1666422  1.0000000 0.15445056
HC     0.1560979  0.05201044  0.1660323  0.23470432  0.4477678  0.1544506 1.00000000
```

[3단계] S의 고윳값과 고유벡터를 구한다.

고윳값

```
> D_S  
[1] 304.26 28.28 11.46 2.52 1.28 0.53 0.21
```

고유벡터

```
> V_S  
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]  
[1,] 0.01 0.08 0.03 0.92 0.34 0.01 -0.17  
[2,] -0.99 0.12 0.01 0.00 0.00 0.00 0.00  
[3,] -0.01 -0.10 -0.18 -0.14 0.65 -0.56 0.44  
[4,] 0.00 0.01 -0.13 -0.33 0.64 0.50 -0.46  
[5,] -0.02 -0.15 -0.96 0.10 -0.21 -0.01 -0.11  
[6,] -0.11 -0.97 0.17 0.06 0.00 0.05 -0.07  
[7,] 0.00 -0.02 -0.09 0.11 0.06 0.66 0.74
```

[3단계] R의 고윳값과 고유벡터를 구한다.

고윳값

```
> D_R  
[1] 2.34 1.39 1.20 0.73 0.65 0.54 0.16
```

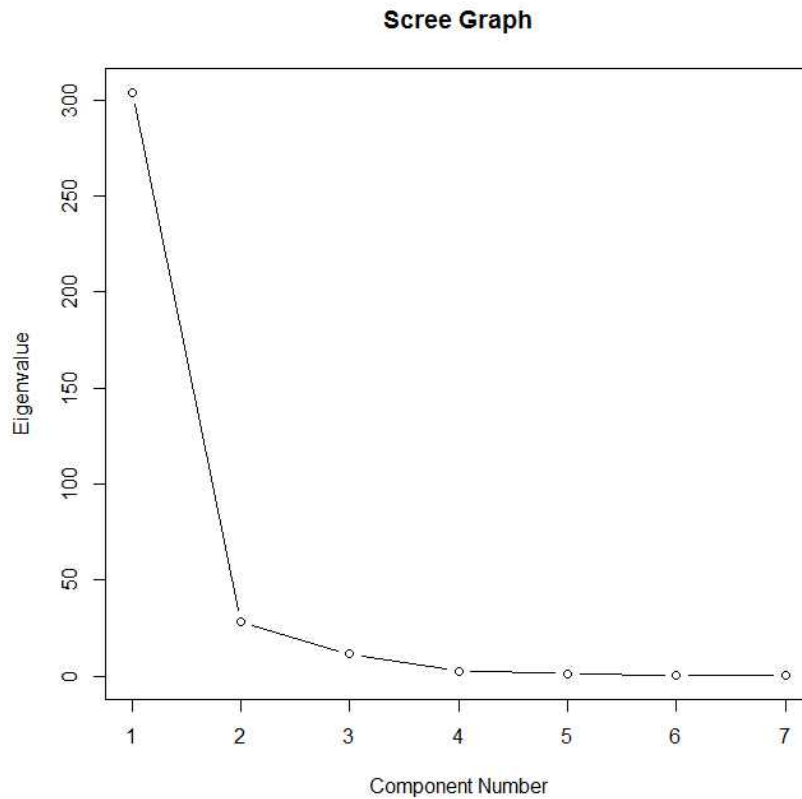
고유벡터

```
> V_R  
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]  
[1,] 0.24 0.28 0.64 0.17 0.56 -0.22 0.24  
[2,] -0.21 -0.53 0.22 0.78 -0.16 -0.01 0.01  
[3,] -0.55 -0.01 -0.11 0.01 0.57 -0.11 -0.59  
[4,] -0.38 0.43 -0.41 0.29 -0.06 -0.45 0.46  
[5,] -0.50 0.20 0.20 -0.04 0.05 0.74 0.34  
[6,] -0.32 -0.57 0.16 -0.51 0.08 -0.33 0.42  
[7,] -0.32 0.31 0.54 -0.14 -0.57 -0.27 -0.31
```

[4단계] 고윳값의 총합에서 70% 이상 설명비율의 합을 갖는  $m(\leq p)$ 개 고윳값을 선택한다.(S)  
설명비율

```
> round(gof_S, 2)
[1] 87.30  8.11  3.29  0.72  0.37  0.15  0.06
```

$m=2$ 개 고윳값( $\lambda_1=304.26$ ,  $\lambda_2=28.28$ )의 설명비율 합은  $87.30\%+8.11\%=95.41\%$ 이다.



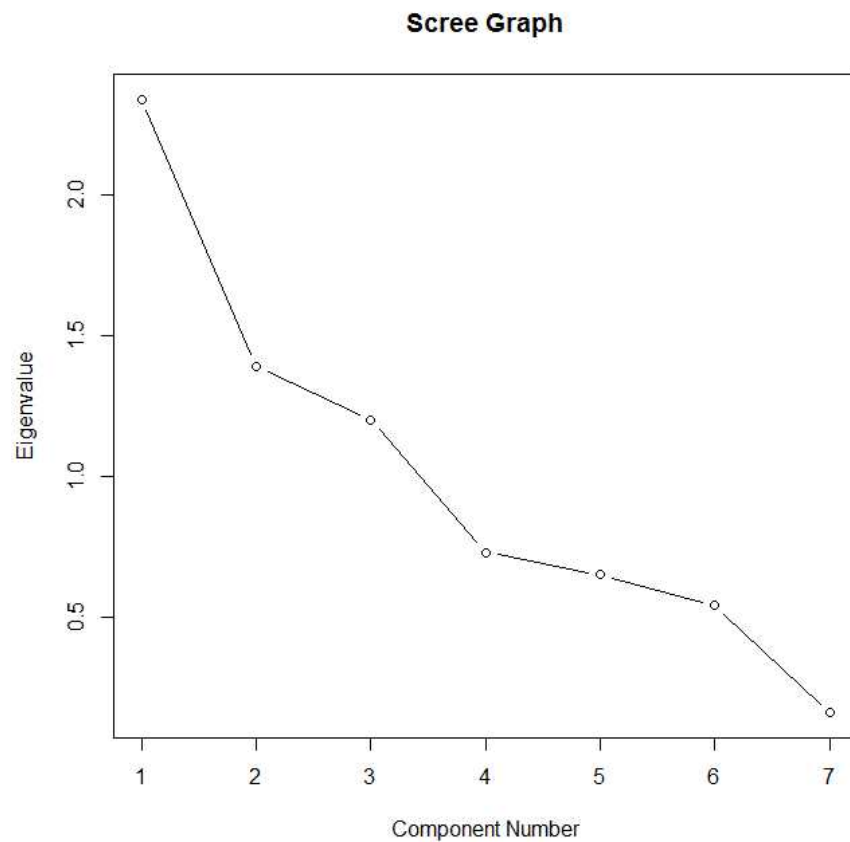
scree graph를 보면 팔꿈치가 2에서 이루어진다. 주성분의 개수는 1개로 하는 것이 타당해 보인다.  
제 1주성분만으로도 약 87% 설명력을 확보하여 제 1주성분에 대한 충분한 해석이 이루어지게 된다.  
하지만, [7단계]에서 시각적 편의상 주성분점수를 보여줄 2차원 주성분공간을 얻기 위해 제 2주성분도 함께 고려하였다.

[4단계] 고춧값의 총합에서 70% 이상 설명비율의 합을 갖는  $m(\leq p)$ 개 고춧값을 선택한다.(R)

설명비율

```
> round(gof_R, 2)
[1] 33.38 19.83 17.12 10.41  9.27  7.70  2.28
```

$m=3$ 개 고춧값( $l_1=2.34$ ,  $l_2=1.39$ ,  $l_3=1.20$ )의 설명비율 합은  $33.38\%+19.83\%+17.12\% = 70.33\%$ 이다.



scree graph를 보면 팔꿈치가 4에서 이루어진다. 주성분의 개수는 3개로 하는 것이 타당해 보인다. 3개의 eigenvalues의 합이 70%를 넘기 때문에 3개의 고유치를 선택한다.

[5단계] 2개 고윳값 11과 12에 대응하는 고유벡터 v1과 v2를 활용하여 중심화 변수의 선형결합인 주성분 p1과 p2를 구한다.(S)

```
> v_s
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]  0.01  0.08  0.03  0.92  0.34  0.01 -0.17
[2,] -0.99  0.12  0.01  0.00  0.00  0.00  0.00
[3,] -0.01 -0.10 -0.18 -0.14  0.65 -0.56  0.44
[4,]  0.00  0.01 -0.13 -0.33  0.64  0.50 -0.46
[5,] -0.02 -0.15 -0.96  0.10 -0.21 -0.01 -0.11
[6,] -0.11 -0.97  0.17  0.06  0.00  0.05 -0.07
[7,]  0.00 -0.02 -0.09  0.11  0.06  0.66  0.74
```

원 변수 :  $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)^t = (\text{wind, solar radiation, } CO, NO, NO_2, O_3, HC)^t$

중심화 변수 :  $y = (y_1, y_2, y_3, y_4, y_5, y_6, y_7)^t$

제 1주성분 :  $p_1 = 0.01 \times y_1 - 0.99 \times y_2 - 0.01 \times y_3 + 0.00 \times y_4 - 0.02 \times y_5 - 0.11 \times y_6 + 0.00 \times y_7$

제 2주성분 :  $p_2 = 0.08 \times y_1 + 0.12 \times y_2 - 0.10 \times y_3 + 0.01 \times y_4 - 0.15 \times y_5 - 0.97 \times y_6 - 0.02 \times y_7$

[5단계] 3개 고윳값 11, 12, 13에 대응하는 고유벡터 v1, v2, v3를 활용하여 표준화 변수의 선형결합인 주성분 p1과 p2, p3를 구한다.(R)

```
> v_R
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]  0.24  0.28  0.64  0.17  0.56 -0.22  0.24
[2,] -0.21 -0.53  0.22  0.78 -0.16 -0.01  0.01
[3,] -0.55 -0.01 -0.11  0.01  0.57 -0.11 -0.59
[4,] -0.38  0.43 -0.41  0.29 -0.06 -0.45  0.46
[5,] -0.50  0.20  0.20 -0.04  0.05  0.74  0.34
[6,] -0.32 -0.57  0.16 -0.51  0.08 -0.33  0.42
[7,] -0.32  0.31  0.54 -0.14 -0.57 -0.27 -0.31
```

원 변수 :  $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)^t = (\text{wind, solar radiation, } CO, NO, NO_2, O_3, HC)^t$

표준화 변수 :  $z = (z_1, z_2, z_3, z_4, z_5, z_6, z_7)^t$

제 1주성분 :  $p_1 = 0.24 \times z_1 - 0.21 \times z_2 - 0.55 \times z_3 - 0.38 \times z_4 - 0.50 \times z_5 - 0.32 \times z_6 - 0.32 \times z_7$

제 2주성분 :  $p_2 = 0.28 \times z_1 - 0.53 \times z_2 - 0.01 \times z_3 + 0.43 \times z_4 + 0.20 \times z_5 - 0.57 \times z_6 + 0.31 \times z_7$

제 3주성분 :  $p_3 = 0.64 \times z_1 + 0.22 \times z_2 - 0.11 \times z_3 - 0.41 \times z_4 + 0.20 \times z_5 + 0.16 \times z_6 + 0.54 \times z_7$



[6단계] 중심화 자료를 주성분에 삽입하여 주성분점수를 얻는다.

중심화 자료 : Y

> Y

|    | Wind | Solar      | CO        | NO         | NO2         | O3         | HC         |
|----|------|------------|-----------|------------|-------------|------------|------------|
| 1  | 0.5  | 24.142857  | 2.452381  | -0.1904762 | 1.95238095  | -1.4047619 | -1.0952381 |
| 2  | -0.5 | 33.142857  | -0.547619 | 0.8095238  | -1.04761905 | -4.4047619 | -0.0952381 |
| 3  | -0.5 | 29.142857  | -0.547619 | 0.8095238  | -5.04761905 | -3.4047619 | -0.0952381 |
| 4  | 2.5  | 14.142857  | 0.452381  | -0.1904762 | -2.04761905 | 5.5952381  | 0.9047619  |
| 5  | -1.5 | 17.142857  | -0.547619 | -0.1904762 | -2.04761905 | 0.5952381  | -0.0952381 |
| 6  | 0.5  | 16.142857  | 0.452381  | -0.1904762 | 1.95238095  | 2.5952381  | 0.9047619  |
| 7  | 1.5  | 10.142857  | 2.452381  | 1.8095238  | 1.95238095  | 5.5952381  | 1.9047619  |
| 8  | -2.5 | -1.857143  | 1.452381  | 1.8095238  | 10.95238095 | 4.5952381  | 0.9047619  |
| 9  | -0.5 | 8.142857   | 0.452381  | -1.1904762 | 0.95238095  | 1.5952381  | -0.0952381 |
| 10 | 0.5  | -9.857143  | 0.452381  | -0.1904762 | 2.95238095  | -0.4047619 | 0.9047619  |
| 11 | -1.5 | -2.857143  | 0.452381  | 1.8095238  | -0.04761905 | -6.4047619 | -0.0952381 |
| 12 | -1.5 | 17.142857  | -0.547619 | -0.1904762 | 1.95238095  | -2.4047619 | -0.0952381 |
| 13 | -0.5 | -1.857143  | 2.452381  | 1.8095238  | 7.95238095  | 0.5952381  | -0.0952381 |
| 14 | 2.5  | -3.857143  | -0.547619 | -0.1904762 | 0.95238095  | -2.4047619 | -0.0952381 |
| 15 | 2.5  | -1.857143  | -0.547619 | -1.1904762 | -2.04761905 | 0.5952381  | -0.0952381 |
| 16 | 1.5  | 3.142857   | -0.547619 | -1.1904762 | -1.04761905 | 0.5952381  | -0.0952381 |
| 17 | 0.5  | 2.142857   | -0.547619 | -1.1904762 | -3.04761905 | -2.4047619 | -0.0952381 |
| 18 | 0.5  | -2.857143  | 0.452381  | 0.8095238  | 5.95238095  | -5.4047619 | 0.9047619  |
| 19 | 1.5  | -6.857143  | -0.547619 | -0.1904762 | 2.95238095  | -7.4047619 | -0.0952381 |
| 20 | 1.5  | -4.857143  | -1.547619 | 0.8095238  | -1.04761905 | -4.4047619 | -0.0952381 |
| 21 | 2.5  | -11.857143 | 0.452381  | 0.8095238  | 3.95238095  | -5.4047619 | 0.9047619  |
| 22 | 1.5  | 14.142857  | -0.547619 | -0.1904762 | -3.04761905 | -3.4047619 | -0.0952381 |
| 23 | 0.5  | 6.142857   | -0.547619 | -0.1904762 | 2.95238095  | 1.5952381  | 0.9047619  |
| 24 | -2.5 | -43.857143 | -1.547619 | 0.8095238  | -5.04761905 | -7.4047619 | -0.0952381 |
| 25 | -1.5 | 9.142857   | 0.452381  | -1.1904762 | -0.04761905 | 13.5952381 | 0.9047619  |
| 26 | 0.5  | 10.142857  | -1.547619 | -0.1904762 | -3.04761905 | -3.4047619 | -0.0952381 |
| 27 | -1.5 | 4.142857   | -0.547619 | -0.1904762 | 0.95238095  | 1.5952381  | -0.0952381 |
| 28 | 0.5  | 5.142857   | -2.547619 | -1.1904762 | -3.04761905 | 0.5952381  | -0.0952381 |
| 29 | -1.5 | -11.857143 | -0.547619 | 0.8095238  | -1.04761905 | -1.4047619 | -0.0952381 |
| 30 | 2.5  | -36.857143 | -1.547619 | -1.1904762 | -3.04761905 | -7.4047619 | -0.0952381 |
| 31 | 0.5  | -2.857143  | -0.547619 | -1.1904762 | -0.04761905 | -2.4047619 | -0.0952381 |
| 32 | -0.5 | -21.857143 | -0.547619 | -1.1904762 | 1.95238095  | -1.4047619 | 0.9047619  |
| 33 | -2.5 | -25.857143 | 1.452381  | 2.8095238  | -2.04761905 | -5.4047619 | -0.0952381 |
| 34 | -1.5 | 1.142857   | -0.547619 | -1.1904762 | -0.04761905 | 14.5952381 | -0.0952381 |
| 35 | 2.5  | -38.857143 | -0.547619 | -1.1904762 | -4.04761905 | -0.4047619 | -1.0952381 |
| 36 | 0.5  | 11.142857  | -0.547619 | -1.1904762 | -1.04761905 | 0.5952381  | -1.0952381 |
| 37 | -2.5 | 12.142857  | -1.547619 | -1.1904762 | -4.04761905 | 2.5952381  | -1.0952381 |
| 38 | -2.5 | 12.142857  | 2.452381  | -0.1904762 | 2.95238095  | 8.5952381  | -1.0952381 |
| 39 | -0.5 | 5.142857   | 2.452381  | 1.8095238  | -1.04761905 | 15.5952381 | -0.0952381 |
| 40 | -0.5 | 5.142857   | 0.452381  | -0.1904762 | -2.04761905 | -3.4047619 | -1.0952381 |
| 41 | -1.5 | -5.857143  | 1.452381  | -0.1904762 | 0.95238095  | 4.5952381  | -0.0952381 |
| 42 | 0.5  | -33.857143 | -0.547619 | 0.8095238  | -4.04761905 | -4.4047619 | -1.0952381 |

attr(,"scaled:center")

|          | Wind      | Solar    | CO       | NO        | NO2      | O3       | HC |
|----------|-----------|----------|----------|-----------|----------|----------|----|
| 7.500000 | 73.857143 | 4.547619 | 2.190476 | 10.047619 | 9.404762 | 3.095238 |    |



주성분 점수 행렬 :  $P=YV(2)$  여기서  $V(2)=(v1, v2)$

```
> P_S
      [,1]      [,2]
1 -23.805476  3.78166667
2 -32.305476  8.43166667
3 -28.375476  7.58166667
4 -14.555476 -3.28833333
5 -17.005476  1.72166667
6 -16.305476 -0.89833333
7 -10.705476 -4.64833333
8  1.074524 -6.66833333
9 -8.265476 -0.80833333
10  9.744524 -1.25833333
11  3.514524  5.73166667
12 -16.755476  4.03166667
13  1.584524 -2.25833333
14  4.094524  1.98166667
15  1.844524 -0.24833333
16 -3.135476  0.12166667
17 -1.785476  3.13166667
18  3.304524  3.99166667
19  7.564524  6.09166667
20  5.344524  4.13166667
21 12.274524  3.37166667
22 -13.545476  5.63166667
23 -6.305476 -1.17833333
24 44.324524  2.64166667
25 -10.565476 -12.27833333
26 -9.585476  5.17166667
27 -4.305476 -1.25833333
28 -5.065476  0.78166667
29 11.904524  0.04166667
30 37.404524  3.56166667
31  3.104524  2.08166667
32 21.754524 -1.56833333
33 26.194524  2.13166667
34 -2.745476 -14.08833333
35 38.624524 -3.39833333
36 -11.065476  1.02166667
37 -12.235476 -0.48833333
38 -13.075476 -7.74833333
39 -6.815476 -14.61833333
40 -4.685476  4.16166667
41  5.244524 -5.56833333
42 34.094524  0.94166667
```

```
> round(V_S2,2)
      [,1] [,2]
[1,]  0.01 0.08
[2,] -0.99 0.12
[3,] -0.01 -0.10
[4,]  0.00 0.01
[5,] -0.02 -0.15
[6,] -0.11 -0.97
[7,]  0.00 -0.02
```

[6단계] 표준화 자료를 주성분에 삽입하여 주성분점수를 얻는다.

표준화 자료 : Z

> Z

|    | Wind       | Solar       | CO         | NO         | NO2         | O3          | HC         |
|----|------------|-------------|------------|------------|-------------|-------------|------------|
| 1  | 0.3162278  | 1.39269205  | 1.9877923  | -0.1751735 | 0.57917248  | -0.25239017 | -1.5832938 |
| 2  | -0.3162278 | 1.91186128  | -0.4438759 | 0.7444873  | -0.31077548 | -0.79139290 | -0.1376777 |
| 3  | -0.3162278 | 1.68111940  | -0.4438759 | 0.7444873  | -1.49737275 | -0.61172533 | -0.1376777 |
| 4  | 1.5811388  | 0.81583735  | 0.3666801  | -0.1751735 | -0.60742479 | 1.00528288  | 1.3079384  |
| 5  | -0.9486833 | 0.98889376  | -0.4438759 | -0.1751735 | -0.60742479 | 0.10694499  | -0.1376777 |
| 6  | 0.3162278  | 0.93120829  | 0.3666801  | -0.1751735 | 0.57917248  | 0.46628014  | 1.3079384  |
| 7  | 0.9486833  | 0.58509548  | 1.9877923  | 1.6641482  | 0.57917248  | 1.00528288  | 2.7535544  |
| 8  | -1.5811388 | -0.10713016 | 1.1772362  | 1.6641482  | 3.24901634  | 0.82561530  | 1.3079384  |
| 9  | -0.3162278 | 0.46972454  | 0.3666801  | -1.0948343 | 0.28252316  | 0.28661257  | -0.1376777 |
| 10 | 0.3162278  | -0.56861391 | 0.3666801  | -0.1751735 | 0.87582179  | -0.07272259 | 1.3079384  |
| 11 | -0.9486833 | -0.16481563 | 0.3666801  | 1.6641482  | -0.01412616 | -1.15072806 | -0.1376777 |
| 12 | -0.9486833 | 0.98889376  | -0.4438759 | -0.1751735 | 0.57917248  | -0.43205775 | -0.1376777 |
| 13 | -0.3162278 | -0.10713016 | 1.9877923  | 1.6641482  | 2.35906838  | 0.10694499  | -0.1376777 |
| 14 | 1.5811388  | -0.22250110 | -0.4438759 | -0.1751735 | 0.28252316  | -0.43205775 | -0.1376777 |
| 15 | 1.5811388  | -0.10713016 | -0.4438759 | -1.0948343 | -0.60742479 | 0.10694499  | -0.1376777 |
| 16 | 0.9486833  | 0.18129719  | -0.4438759 | -1.0948343 | -0.31077548 | 0.10694499  | -0.1376777 |
| 17 | 0.3162278  | 0.12361172  | -0.4438759 | -1.0948343 | -0.90407411 | -0.43205775 | -0.1376777 |
| 18 | 0.3162278  | -0.16481563 | 0.3666801  | 0.7444873  | 1.76576975  | -0.97106048 | 1.3079384  |
| 19 | 0.9486833  | -0.39555751 | -0.4438759 | -0.1751735 | 0.87582179  | -1.33039564 | -0.1376777 |
| 20 | 0.9486833  | -0.28018657 | -1.2544320 | 0.7444873  | -0.31077548 | -0.79139290 | -0.1376777 |
| 21 | 1.5811388  | -0.68398485 | 0.3666801  | 0.7444873  | 1.17247111  | -0.97106048 | 1.3079384  |
| 22 | 0.9486833  | 0.81583735  | -0.4438759 | -0.1751735 | -0.90407411 | -0.61172533 | -0.1376777 |
| 23 | 0.3162278  | 0.35435360  | -0.4438759 | -0.1751735 | 0.87582179  | 0.28661257  | 1.3079384  |
| 24 | -1.5811388 | -2.52991988 | -1.2544320 | 0.7444873  | -1.49737275 | -1.33039564 | -0.1376777 |
| 25 | -0.9486833 | 0.52741001  | 0.3666801  | -1.0948343 | -0.01412616 | 2.44262350  | 1.3079384  |
| 26 | 0.3162278  | 0.58509548  | -1.2544320 | -0.1751735 | -0.90407411 | -0.61172533 | -0.1376777 |
| 27 | -0.9486833 | 0.23898266  | -0.4438759 | -0.1751735 | 0.28252316  | 0.28661257  | -0.1376777 |
| 28 | 0.3162278  | 0.29666813  | -2.0649881 | -1.0948343 | -0.90407411 | 0.10694499  | -0.1376777 |
| 29 | -0.9486833 | -0.68398485 | -0.4438759 | 0.7444873  | -0.31077548 | -0.25239017 | -0.1376777 |
| 30 | 1.5811388  | -2.12612159 | -1.2544320 | -1.0948343 | -0.90407411 | -1.33039564 | -0.1376777 |
| 31 | 0.3162278  | -0.16481563 | -0.4438759 | -1.0948343 | -0.01412616 | -0.43205775 | -0.1376777 |
| 32 | -0.3162278 | -1.26083955 | -0.4438759 | -1.0948343 | 0.57917248  | -0.25239017 | 1.3079384  |
| 33 | -1.5811388 | -1.49158143 | 1.1772362  | 2.5838090  | -0.60742479 | -0.97106048 | -0.1376777 |
| 34 | -0.9486833 | 0.06592625  | -0.4438759 | -1.0948343 | -0.01412616 | 2.62229108  | -0.1376777 |
| 35 | 1.5811388  | -2.24149253 | -0.4438759 | -1.0948343 | -1.20072343 | -0.07272259 | -1.5832938 |
| 36 | 0.3162278  | 0.64278095  | -0.4438759 | -1.0948343 | -0.31077548 | 0.10694499  | -1.5832938 |
| 37 | -1.5811388 | 0.70046642  | -1.2544320 | -1.0948343 | -1.20072343 | 0.46628014  | -1.5832938 |
| 38 | -1.5811388 | 0.70046642  | 1.9877923  | -0.1751735 | 0.87582179  | 1.54428561  | -1.5832938 |
| 39 | -0.3162278 | 0.29666813  | 1.9877923  | 1.6641482  | -0.31077548 | 2.80195866  | -0.1376777 |
| 40 | -0.3162278 | 0.29666813  | 0.3666801  | -0.1751735 | -0.60742479 | -0.61172533 | -1.5832938 |
| 41 | -0.9486833 | -0.33787204 | 1.1772362  | -0.1751735 | 0.28252316  | 0.82561530  | -0.1376777 |
| 42 | 0.3162278  | -1.95306518 | -0.4438759 | 0.7444873  | -1.20072343 | -0.79139290 | -1.5832938 |

attr(,"scaled:center")

| Wind     | Solar     | CO       | NO       | NO2       | O3       | HC       |
|----------|-----------|----------|----------|-----------|----------|----------|
| 7.500000 | 73.857143 | 4.547619 | 2.190476 | 10.047619 | 9.404762 | 3.095238 |

attr(,"scaled:scale")

| Wind      | Solar      | CO        | NO        | NO2       | O3        | HC        |
|-----------|------------|-----------|-----------|-----------|-----------|-----------|
| 1.5811388 | 17.3353881 | 1.2337209 | 1.0873574 | 3.3709837 | 5.5658345 | 0.6917466 |

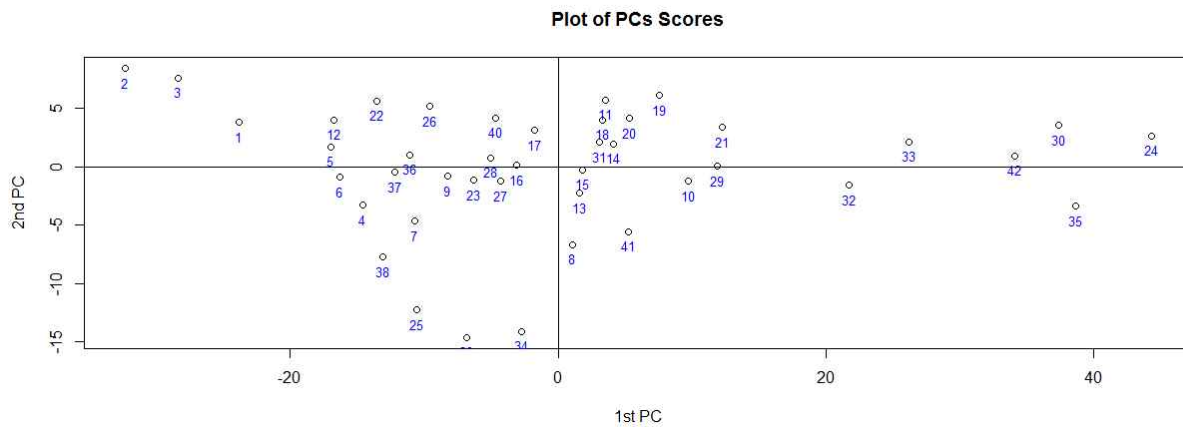


주성분 점수 행렬 :  $P=ZV(3)$  여기서  $V(3)=(v1, v2, v3)$

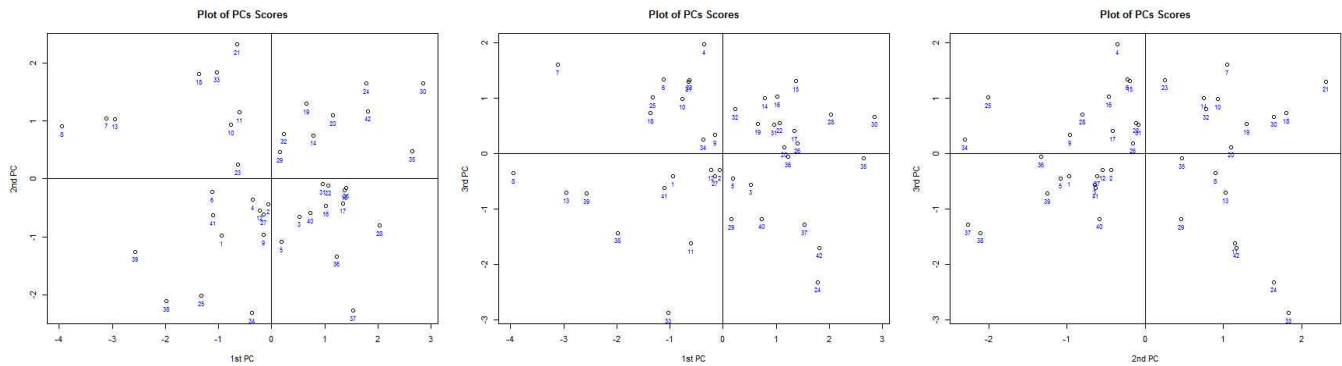
```
> P_R
      [,1]      [,2]      [,3]
1 -0.94545786 -0.97590972 -0.41758458
2 -0.06346861 -0.43100317 -0.30131367
3  0.52079219 -0.64843995 -0.56064953
4 -0.36347907 -0.35770164  1.96854640
5  0.18889289 -1.08575455 -0.44767293
6 -1.11210338 -0.22847232  1.33546394
7 -3.11326228  1.04766163  1.59851956
8 -3.94407753  0.90253489 -0.35910579
9 -0.14909458 -0.96148796  0.33751756
10 -0.77298450  0.93299487  0.97859248
11 -0.60777015  1.14404753 -1.62734000
12 -0.23192488 -0.54120354 -0.29659391
13 -2.94875910  1.03211593 -0.71233339
14  0.77795002  0.74985607  0.99665552
15  1.37568633 -0.19196583  1.30734891
16  1.01500260 -0.46259001  1.02536125
17  1.34445742 -0.42053243  0.40299874
18 -1.36475910  1.80447812  0.72462269
19  0.65332134  1.29520074  0.52873723
20  1.14624593  1.09306284  0.11514005
21 -0.65550559  2.31515318  1.30128882
22  1.05890187 -0.11245980  0.55425217
23 -0.63598909  0.24710661  1.32830013
24  1.77731212  1.64698341 -2.32244733
25 -1.31719384 -2.00926392  1.01170138
26  1.40137418 -0.15914859  0.18787858
27 -0.15609338 -0.61272260 -0.40591697
28  2.02724637 -0.80327277  0.70563392
29  0.15738887  0.46047617 -1.19093092
30  2.85375403  1.64615948  0.66302760
31  0.96005318 -0.08967635  0.51753431
32  0.22168879  0.77851897  0.79967673
33 -1.03705992  1.83642421 -2.88013503
34 -0.36937289 -2.30712347  0.24745026
35  2.64064252  0.47485603 -0.09025003
36  1.22889883 -1.33240494 -0.05851654
37  1.53720946 -2.26894595 -1.29147516
38 -1.97871937 -2.10506811 -1.43739085
39 -2.56103918 -1.25202372 -0.72626436
40  0.73281540 -0.58839188 -1.17997212
41 -1.09904645 -0.63043229 -0.62490690
42  1.80752639  1.16836483 -1.70544822
```

```
> round(V_R2,2)
      [,1] [,2] [,3]
[1,]  0.24  0.28  0.64
[2,] -0.21 -0.53  0.22
[3,] -0.55 -0.01 -0.11
[4,] -0.38  0.43 -0.41
[5,] -0.50  0.20  0.20
[6,] -0.32 -0.57  0.16
[7,] -0.32  0.31  0.54
```

[7단계] 주성분점수를 차원 축소된 새로운 다변량 자료로 고려한다.(S)



[7단계] 주성분점수를 차원 축소된 새로운 다변량 자료로 고려한다.(R)



주성분점수의 산점도를 비교해볼 때 공분산행렬을 사용하였을 때의 데이터들의 분산이 크다. 이는 변수 solar radiation의 분산이 다른 변수들에 비해 크기 때문에 이러한 결과가 나온 것으로 보인다. 따라서 공분산행렬을 이용한 주성분 분석보다 상관행렬을 이용하는 것이 적절해 보인다.

(2) Find the appropriate principal components with the goodness-of-fit and interpret them.

```
> round(V_R2, 2)
      [,1] [,2] [,3]
[1,]  0.24  0.28  0.64
[2,] -0.21 -0.53  0.22
[3,] -0.55 -0.01 -0.11
[4,] -0.38  0.43 -0.41
[5,] -0.50  0.20  0.20
[6,] -0.32 -0.57  0.16
[7,] -0.32  0.31  0.54
```

원 변수 :  $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)^t = (\text{wind}, \text{solar radiation}, CO, NO, NO_2, O_3, HC)^t$

표준화 변수 :  $z = (z_1, z_2, z_3, z_4, z_5, z_6, z_7)^t$

제 1주성분 :  $p_1 = 0.24 \times z_1 - 0.21 \times z_2 - 0.55 \times z_3 - 0.38 \times z_4 - 0.50 \times z_5 - 0.32 \times z_6 - 0.32 \times z_7$

제 2주성분 :  $p_2 = 0.28 \times z_1 - 0.53 \times z_2 - 0.01 \times z_3 + 0.43 \times z_4 + 0.20 \times z_5 - 0.57 \times z_6 + 0.31 \times z_7$

제 3주성분 :  $p_3 = 0.64 \times z_1 + 0.22 \times z_2 - 0.11 \times z_3 - 0.41 \times z_4 + 0.20 \times z_5 + 0.16 \times z_6 + 0.54 \times z_7$

해석 : 제 1주성분  $p_1$ 의 주성분계수의 부호가 양인  $z_1(\text{wind})$ 를 제외하면 모두 음(-)의 값으로 wind에 대한 일반적인 성분을 나타내고 있다. 제 1주성분의 설명력은 33.38%이다.

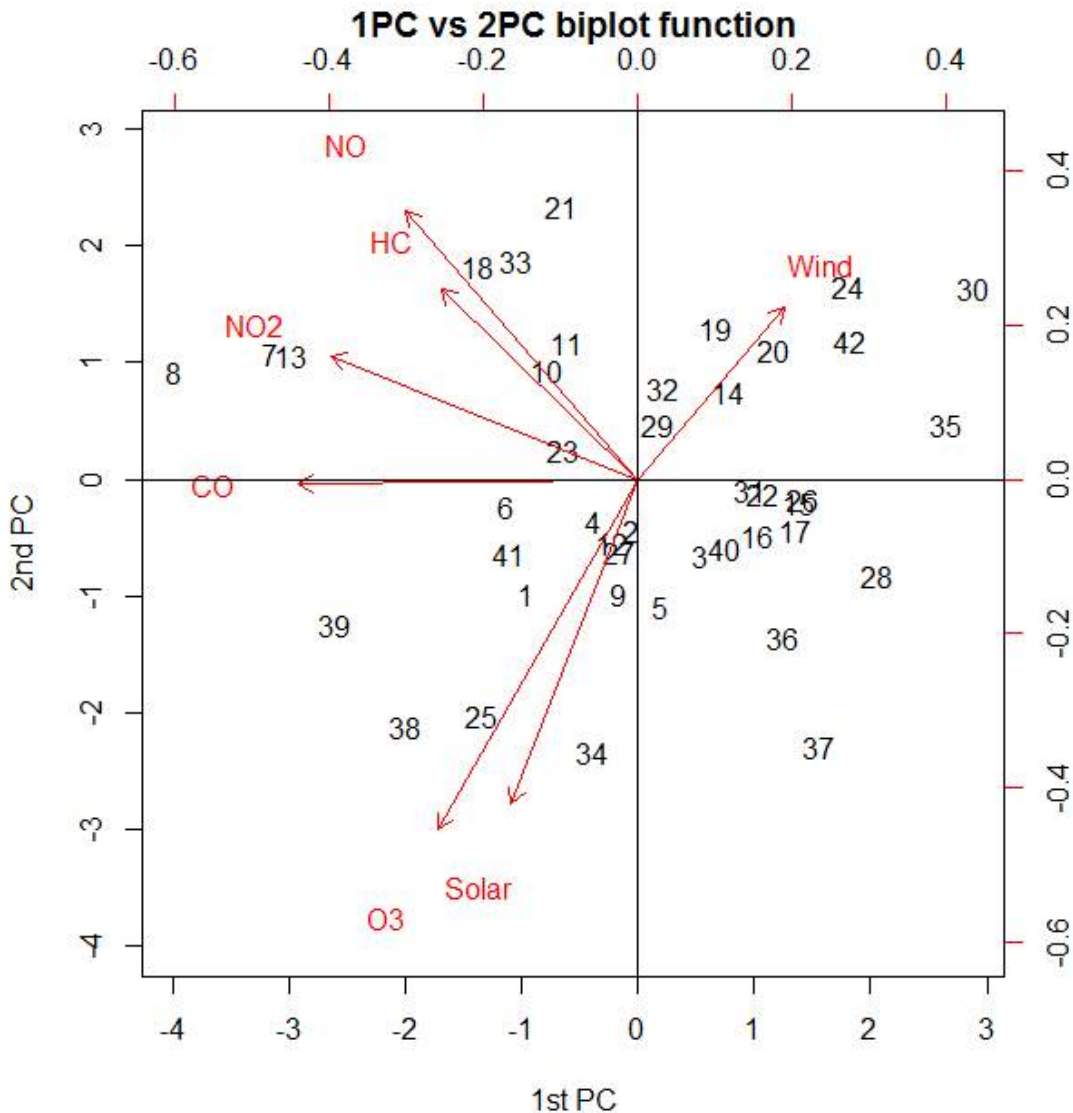
제 2주성분은 주성분계수 부호가 양인 wind,  $NO$ ,  $NO_2$ ,  $HC$ 와 주성분계수 부호가 음인 solar radiation,  $CO$ ,  $O_3$ 의 대비를 나타낸다. solar radiation과  $O_3$ 은 햇빛과 연관이 있고  $CO$ 는 주성분에 큰 영향을 미치지 않으므로 햇빛에 대한 성분을 나타내고 있다. 제 2주성분의 설명력은 19.83%이다.

제 3주성분은 주성분계수 부호가 양인 wind, solar radiation,  $NO_2$ ,  $O_3$ ,  $HC$ 와 주성분계수 부호가 음인  $CO$ ,  $NO$ 의 대비를 나타낸다. 일산화 화합물과 나머지 변수들에서 오염도의 관계에 대하여 설명한다. 바람이 가장 많이 영향을 끼치고 있다. 제 3주성분의 설명력은 17.12%이다.

따라서 세 주성분의 설명력은 70.33%이다.



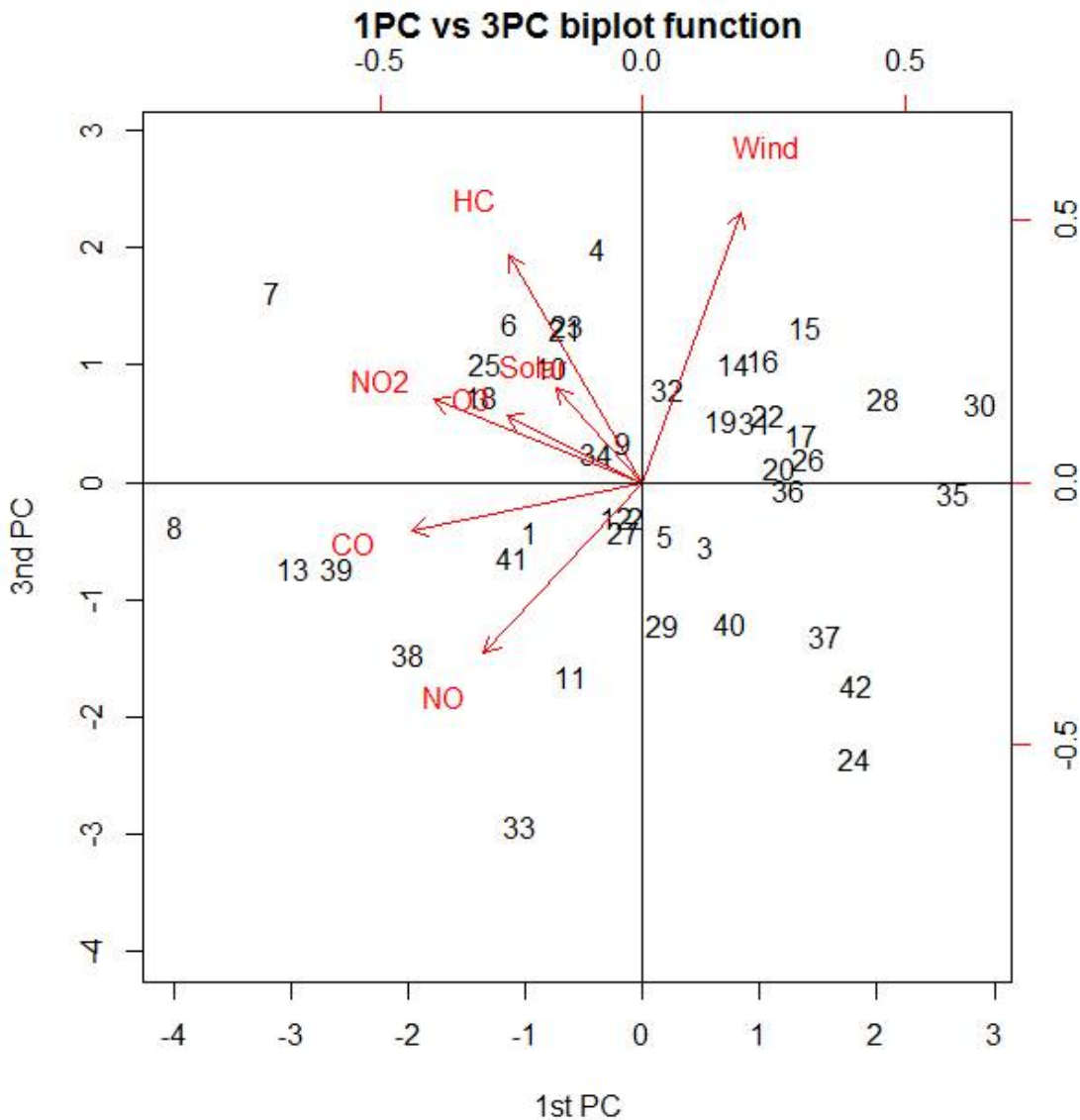
- (3) Check the correlations between the variables through the principal component biplot and compare them with interpretations of principal components in (2).



제 1주성분과 제 2주성분을 비교한 주성분행렬도에서 solar radiation과  $O_3$ 는 큰 상관관계를 보이며, 다른 오염물질과는 거의 수직으로 상관관계가 크게 없어보인다. 다른 오염물질과 wind도 큰 영향은 없어보인다. 반면 wind와 solar radiation과  $O_3$ 는 서로 방향이 달라 음의 상관을 가지고 있다. 제 1 주성분에서 풍속은 양, 나머지 요인들은 음으로 서로 반대 방향으로 화살표가 뻗어있다. 이는 2)의 해석과 같다.

제 2주성분에서 wind, NO,  $NO_2$ , HC는 양, solar radiation, CO,  $O_3$ 은 음으로 서로 반대 방향으로 다른 부호를 가지고 있어 2)의 해석과 같다. 또한 CO에서는 화살표의 방향이 0에 가까워 제 2주성분에 큰 영향을 미치지 않는데 이는 2)에서 CO은 -0.01의 영향을 주는 것으로 해석이 같음을 알 수 있다.

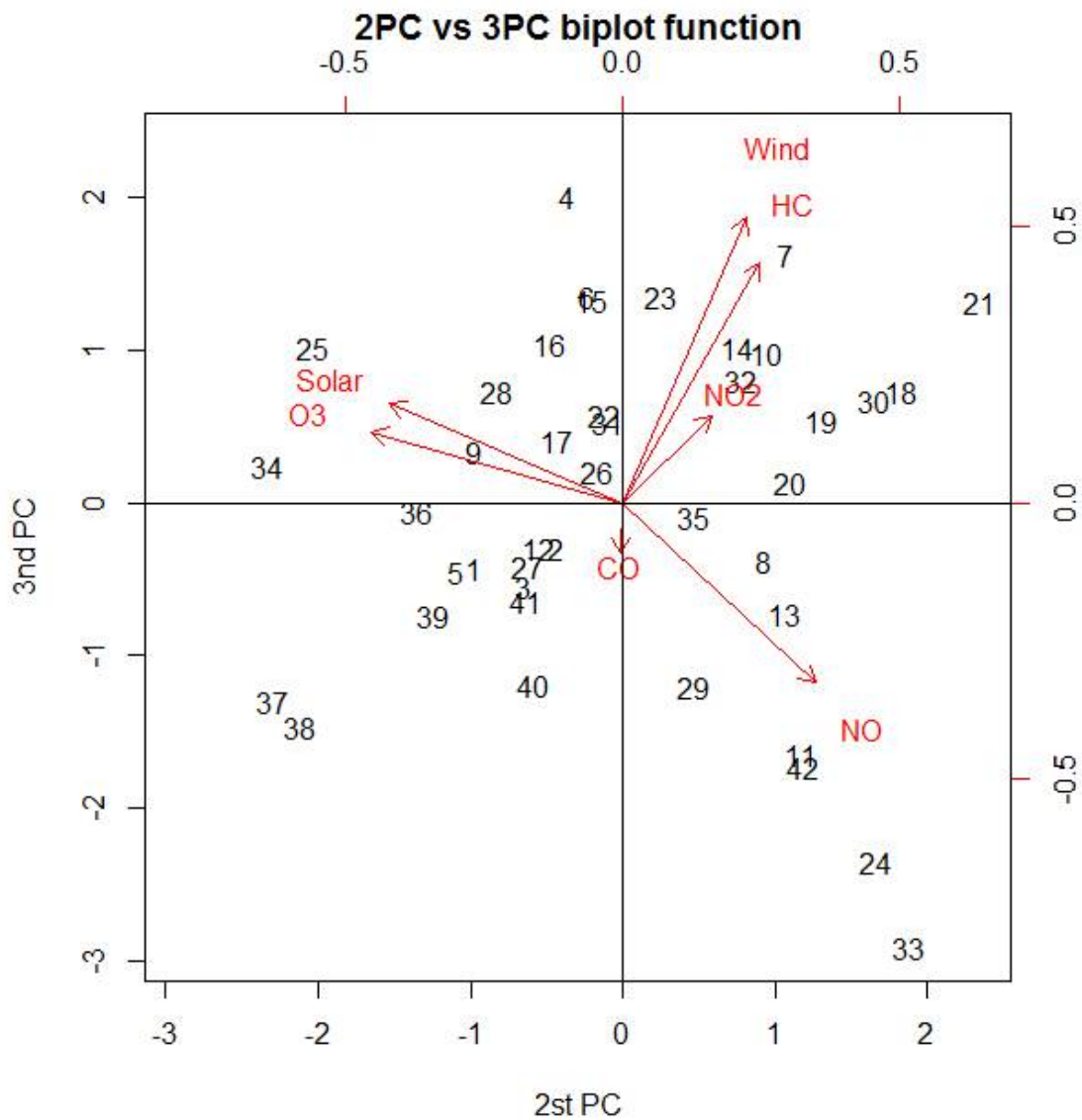
위의 행렬도는  $33.38\% + 19.83\% = 53.21\%$ 의 설명력을 가진다.



제 1주성분과 제 3주성분을 비교한 주성분행렬도에서도 solar radiation과  $O_3$ 는 큰 상관관계를 보이며, 다른 오염물질과도 상관이 있는 듯 보인다.  $NO$ 와  $HC$ 는 거의 수직으로 상관관계가 크게 없어 보인다. wind와  $O_3$ 도 큰 상관이 없어 보인다.  $NO$ 와 wind는 서로 방향이 달라 음의 상관을 가지고 있다.

제 3주성분에서  $NO$ ,  $CO$ 는 음, wind, solar radiation,  $NO_2$ ,  $O_3$ ,  $HC$ 은 양으로 서로 반대 방향으로 다른 부호를 가지고 있다. 그리고 wind가 가장 많은 영향을 끼치고 있다. 이는 2)의 해석과 같다.

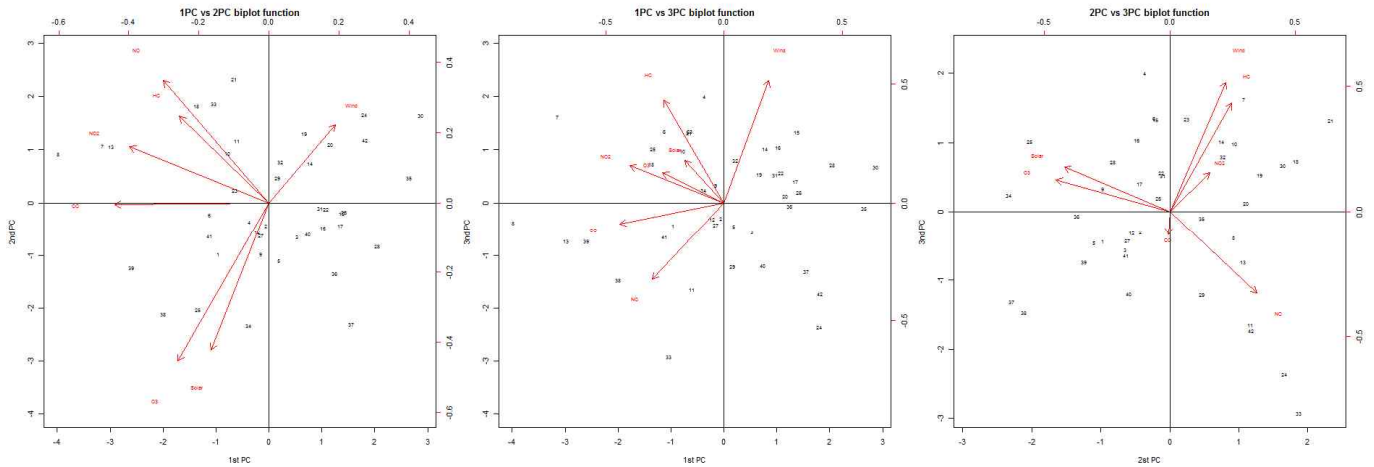
위의 행렬도는  $33.38\% + 17.12\% = 50.5\%$ 의 설명력을 가진다.



제 2주성분과 제 3주성분을 비교한 주성분행렬도에서도 solar radiation과  $O_3$ 는 큰 상관관계를 보이며, 다른 오염물질( $NO$ 를 제외한)는 상관관계가 없는 듯 보인다. wind도 마찬가지이다. 하지만  $NO$ 는 solar radiation과  $O_3$ 과 방향이 달라 음의 상관을 띄고 있다.

위의 행렬도는  $19.83\% + 17.12\% = 36.95\%$ 의 설명력을 가진다.

(4) Describe some relationships between the clusters of air pollution observations and variables in the biplot.



오존이 가정, 자동차, 사업장 등 대기 중으로 직접 배출된 질소산화물(NOX), 탄화수소(HC), 일산화탄소(CO) 등과 같은 대기오염물질들이 햇빛에 의해 광화학 반응을 일으켜 오염물질을 만들어낸다.

따라서 wind의 영향성분인 제 1주성분과 관련된 '제 1주성분과 제 2주성분을 비교한 주성분행렬도'와 '제 1주성분과 제 3주성분을 비교한 주성분행렬' 모두 왼편에는 대기오염이 심한 곳이고 오른편에는 대기오염이 덜 심한 곳으로 그룹을 형성한다. 태양 방사능과 오존의 그룹과 나머지 변수의 대비를 나타내는 제 2주성분과 관련된 2nd PC에서는 '제 1주성분과 제 2주성분을 비교한 주성분행렬도'에서는 아래편에 오존과 태양 방사능 그룹이 '제 2주성분과 제 3주성분을 비교한 주성분행렬도' 왼편에 형성되고 있다.

또한 세 주성분행렬도에서 wind는 모두 각각의 성분에 양의 상관관계를 가지므로 대기오염 관찰치들의 군집들이 wind 변수에 영향을 많이 받는다고 할 수 있다.

```

setwd("D:/2020 1학기 정호재/다변량통계학(1)/200402 다변량 실습1/Rdata")
data<-read.table("airpollution.txt", header=T)
rownames<-rownames(data)
head(data)
dim(data)

S<-cov(data)
S
R<-cor(data)
R
eigen.S<-eigen(S)
D_S<-round(eigen.S$values,2)
V_S<-round(eigen.S$vectors,2)

eigen.R<-eigen(R)
D_R<-round(eigen.R$values,2)
V_R<-round(eigen.R$vectors,2)

gof_S<-D_S/sum(D_S)*100
round(gof_S, 2)
plot(D_S, type="b", main="Scree Graph", xlab="Component Number", ylab="Eigenvalue")
gof_R<-D_R/sum(D_R)*100
round(gof_R,2)
plot(D_R, type="b", main="Scree Graph", xlab="Component Number", ylab="Eigenvalue")

V_S2<-V_S[,1:2]
round(V_S2,2)

V_R2<-V_R[,1:3]
round(V_R2,2)

Y<-scale(data, scale=F)
P_S<-Y%*%V_S2
P_S
Z<-scale(data,scale=T)
P_R<-Z%*%V_R2

plot(P_S[,1], P_S[, 2], main="Plot of PCs Scores", xlab="1st PC", ylab="2nd PC")
text(P_S[,1], P_S[, 2], labels=rownames, cex=0.8, col="blue", pos=1)
abline(v=0, h=0)

```



```

par(mfrow=c(1,3))
plot(P_R[,1], P_R[,2], main="Plot of PCs Scores", xlab="1st PC", ylab="2nd PC")
text(P_R[,1], P_R[,2], labels=rownames, cex=0.8, col="blue", pos=1)
abline(v=0, h=0)

plot(P_R[,1], P_R[,3], main="Plot of PCs Scores", xlab="1st PC", ylab="3rd PC")
text(P_R[,1], P_R[,3], labels=rownames, cex=0.8, col="blue", pos=1)
abline(v=0, h=0)

plot(P_R[,2], P_R[,3], main="Plot of PCs Scores", xlab="2nd PC", ylab="3rd PC")
text(P_R[,2], P_R[,3], labels=rownames, cex=0.8, col="blue", pos=1)
abline(v=0, h=0)
#####
X=data
pca.R<-princomp(X, cor=T)
summary(pca.R, loadings=T)
round(pca.R$scores, 3)
screeplot(pca.R, type="lines")

# Principle component biplot (SVD)
par(mfrow=c(1,3))
biplot(pca.R, choices=c(1,2), scale=0, xlab="1st PC",ylab="2nd PC",
       main="1PC vs 2PC biplot function")
abline(v=0, h=0)
biplot(pca.R, choices=c(1,3), scale=0, xlab="1st PC",ylab="3nd PC",
       main="1PC vs 3PC biplot function")
abline(v=0, h=0)
biplot(pca.R, choices=c(2,3), scale=0, xlab="2st PC",ylab="3nd PC",
       main="2PC vs 3PC biplot function")
abline(v=0, h=0)
#####3
X=data
n <- nrow(X)
rownames(X)
colnames(X)

Y <- scale(X,scale=T)

svd.Y <- svd(Y)
U <- svd.Y$u
V <- svd.Y$v
D <- diag(svd.Y$d)

```

```

G <- (sqrt(n-1)*U)[,1:2]
H <- (sqrt(1/(n-1))*V**%D)[,1:2]
rownames(G)<-rownames(X)
rownames(H)<-colnames(X)

biplot(G,H, xlab="1st PC",ylab="2nd PC", main="1PC vs 2PC biplot function",
        xlim=lim2,ylim=lim2,cex=0.8,pch=16)
abline(v=0,h=0)

G <- (sqrt(n-1)*U)[,c(1,3)]
H <- (sqrt(1/(n-1))*V**%D)[,c(1,3)]
rownames(G)<-rownames(X)
rownames(H)<-colnames(X)

biplot(G,H, xlab="1st PC",ylab="3nd PC", main="1PC vs 3PC biplot function",
        xlim=lim2,ylim=lim2,cex=0.8,pch=16)
abline(v=0,h=0)

G <- (sqrt(n-1)*U)[,2:3]
H <- (sqrt(1/(n-1))*V**%D)[,2:3]
rownames(G)<-rownames(X)
rownames(H)<-colnames(X)

biplot(G,H, xlab="2st PC",ylab="3nd PC", main="2PC vs 3PC biplot function",
        xlim=lim2,ylim=lim2,cex=0.8,pch=16)
abline(v=0,h=0)

```

## 2. Solve the problems (1) ~ (6) in Exercise 2.9.

(1) 공분산행렬 S와 상관행렬 R에 대해 PCA 수행단계를 적용하고 결과를 구하라.

[1단계] 다변량 자료행렬 X를 준비한다.

```
> head(data)
      x100  x200  x400 x800 x1500 x5000 x10000 Marathon
Argentina 10.23 20.37 46.18 1.77  3.68 13.33  27.65   129.57
Australia  9.93 20.06 44.38 1.74  3.53 12.93  27.53   127.51
Austria    10.15 20.45 45.80 1.77  3.58 13.26  27.72   132.22
Belgium    10.14 20.19 45.02 1.73  3.57 12.83  26.87   127.20
Bermuda    10.27 20.30 45.26 1.79  3.70 14.64  30.49   146.37
Brazil     10.00 19.89 44.29 1.70  3.57 13.48  28.13   126.05
> dim(data)
[1] 54  8
```

[자료 2.8.3](trackrecord2005-men)의 자료는 변수는  $p=8$ 이고 개체 수는  $n=54$ 인  $54 \times 8$  다변량 자료행렬이다

[2단계] 공분산행렬 S를 구한다.

```
> S
      x100      x200      x400      x800      x1500      x5000      x10000  Marathon
x100    0.048972921 0.11104437 0.25602156 0.008263871 0.025720126 0.12457530 0.26561286 1.3401386
x200    0.111044375 0.30090342 0.66681838 0.022929210 0.066193082 0.31773382 0.68893557 3.5410381
x400    0.256021558 0.66681838 2.06995573 0.057937876 0.168472956 0.85348641 1.84994074 9.1788571
x800    0.008263871 0.02292921 0.05793788 0.002751223 0.007130818 0.03434829 0.07425695 0.3789048
x1500   0.025720126 0.06619308 0.16847296 0.007130818 0.023033962 0.10583270 0.22970126 1.1925635
x5000   0.124575297 0.31773382 0.85348641 0.034348288 0.105832704 0.57887523 1.26253347 6.4304888
x10000  0.265612858 0.68893557 1.84994074 0.074256953 0.229701258 1.26253347 2.81956883 14.3425380
Marathon 1.340138644 3.54103809 9.17885709 0.378904752 1.192563522 6.43048882 14.34253802 80.1353563
```

다른 변수들에 비해 Marathon의 분산이 매우 크다.

[2단계] 상관행렬 R를 구한다.

```
> R
      x100      x200      x400      x800      x1500      x5000      x10000  Marathon
x100    1.0000000 0.9147554 0.8041147 0.7119388 0.7657919 0.7398803 0.7147921 0.6764873
x200    0.9147554 1.0000000 0.8449159 0.7969162 0.7950871 0.7613028 0.7479519 0.7211157
x400    0.8041147 0.8449159 1.0000000 0.7677488 0.7715522 0.7796929 0.7657481 0.7126823
x800    0.7119388 0.7969162 0.7677488 1.0000000 0.8957609 0.8606959 0.8431074 0.8069657
x1500   0.7657919 0.7950871 0.7715522 0.8957609 1.0000000 0.9165224 0.9013380 0.8777788
x5000   0.7398803 0.7613028 0.7796929 0.8606959 0.9165224 1.0000000 0.9882324 0.9441466
x10000  0.7147921 0.7479519 0.7657481 0.8431074 0.9013380 0.9882324 1.0000000 0.9541630
Marathon 0.6764873 0.7211157 0.7126823 0.8069657 0.8777788 0.9441466 0.9541630 1.0000000
```

[3단계] S의 고윳값과 고유벡터를 구한다.

고윳값

```
> D_S
[1] 84.51  1.14  0.23  0.08  0.01  0.01  0.00  0.00
```

고유벡터

```
> V_S
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] -0.02 -0.10  0.01 -0.32 -0.31  0.88  0.03 -0.09
[2,] -0.04 -0.25  0.08 -0.90  0.17 -0.29 -0.07  0.04
[3,] -0.11 -0.92  0.25  0.29  0.01  0.00  0.00  0.00
[4,]  0.00 -0.01 -0.01 -0.03 -0.04 -0.13  0.19 -0.97
[5,] -0.01 -0.03 -0.04 -0.06 -0.21 -0.11  0.94  0.22
[6,] -0.08 -0.12 -0.38 -0.04 -0.83 -0.30 -0.25  0.04
[7,] -0.18 -0.21 -0.87  0.02  0.38  0.12  0.06 -0.01
[8,] -0.97  0.17  0.15  0.01  0.00  0.00  0.00  0.00
```

[3단계] R의 고윳값과 고유벡터를 구한다.

고윳값

```
> D_R
[1] 6.70 0.64 0.23 0.21 0.10 0.07 0.05 0.01
```

고유벡터

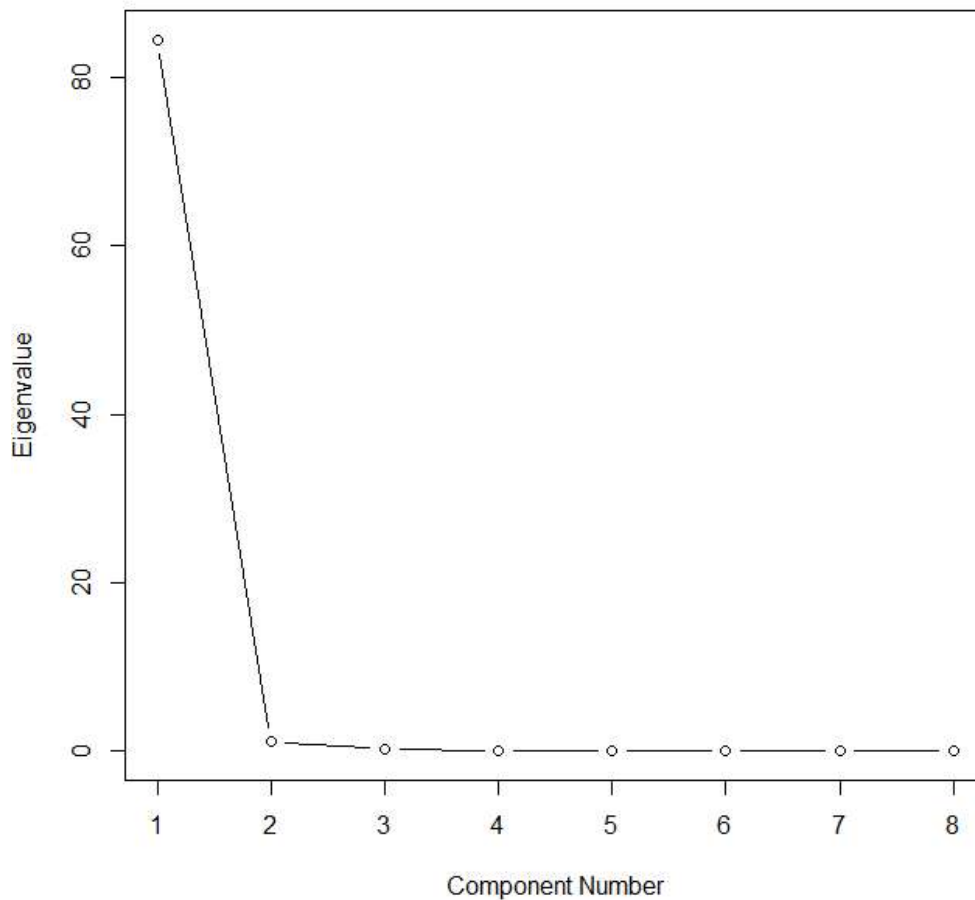
```
> V_R
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] -0.33 -0.53  0.34  0.38 -0.30 -0.36  0.35  0.07
[2,] -0.35 -0.47  0.00  0.22  0.54  0.35 -0.44 -0.06
[3,] -0.34 -0.35 -0.07 -0.85 -0.13  0.08  0.11  0.00
[4,] -0.35  0.09 -0.78  0.13  0.23 -0.34  0.26  0.04
[5,] -0.37  0.15 -0.24  0.23 -0.65  0.53 -0.15  0.04
[6,] -0.37  0.29  0.18 -0.05 -0.07 -0.36 -0.33 -0.71
[7,] -0.37  0.33  0.24 -0.09  0.06 -0.27 -0.35  0.70
[8,] -0.35  0.39  0.33  0.02  0.34  0.38  0.59 -0.07
```

[4단계] 고윳값의 총합에서 70% 이상 설명비율의 합을 갖는  $m(\leq p)$ 개 고윳값을 선택한다.(S)  
설명비율

```
> round(gof_S, 2)
[1] 98.29  1.33  0.27  0.09  0.01  0.01  0.00  0.00
```

$m=2$ 개 고윳값( $\lambda_1=84.51$ ,  $\lambda_2=1.14$ )의 설명비율 합은  $98.29\%+1.33\%=99.62\%$ 이다.

**Scree Graph**



scree graph를 보면 팔꿈치가 2에서 이루어진다. 주성분의 개수는 1개로 하는 것이 타당해 보인다. 제 1주성분만으로도 약 98.29% 설명력을 확보하여 제 1주성분에 대한 충분한 해석이 이루어지게 된다. 하지만, [7단계]에서 시각적 편의상 주성분점수를 보여줄 2차원 주성분공간을 얻기 위해 제 2주성분도 함께 고려하였다.



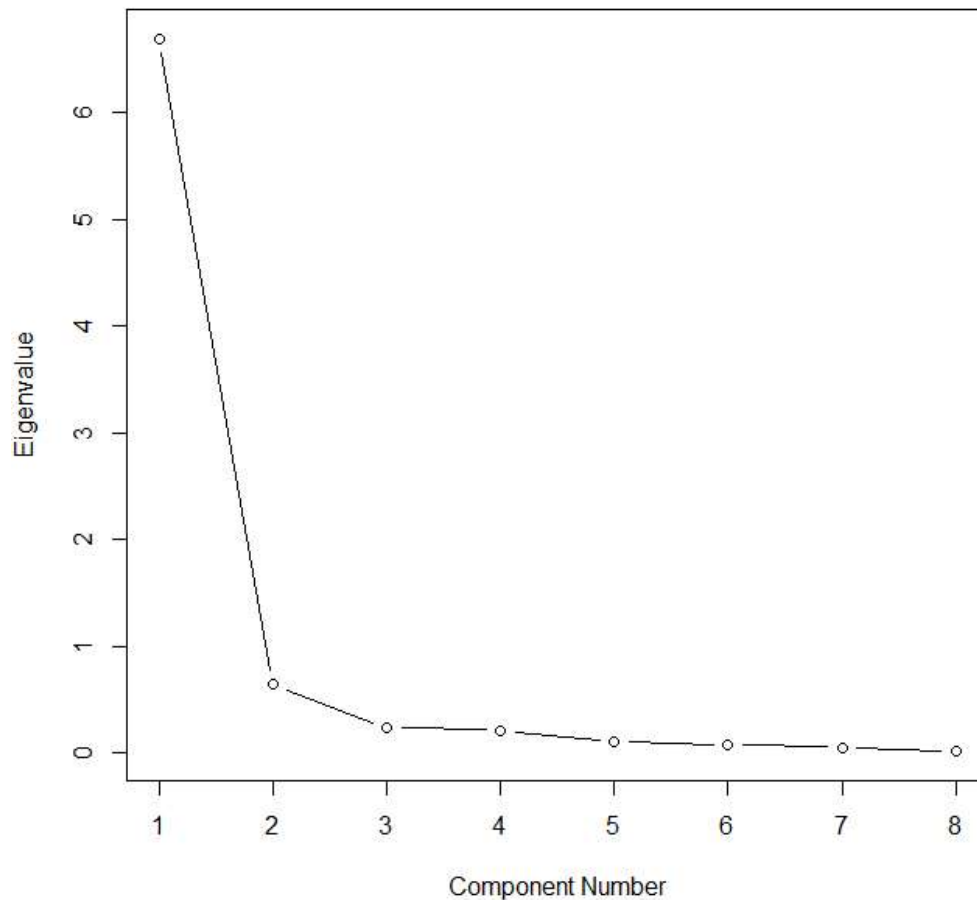
[4단계] 고춧값의 총합에서 70% 이상 설명비율의 합을 갖는  $m(\leq p)$ 개 고춧값을 선택한다.(R)

설명비율

```
> round(gof_R, 2)
[1] 83.65  7.99  2.87  2.62  1.25  0.87  0.62  0.12
```

$m=2$ 개 고춧값( $l_1=6.70$ ,  $l_2=0.64$ )의 설명비율 합은  $83.65\%+7.99\% = 91.64\%$ 이다.

**Scree Graph**



scree graph를 보면 팔꿈치가 2에서 이루어진다. 주성분의 개수는 1개로 하는 것이 타당해 보인다.

제 1주성분만으로도 약 83.65% 설명력을 확보하여 제 1주성분에 대한 충분한 해석이 이루어지게 된다.

하지만, [7단계]에서 시각적 편의상 주성분점수를 보여줄 2차원 주성분공간을 얻기 위해 제 2주성분도 함께 고려하였다.

[5단계] 2개 고ؤ값 11과 12에 대응하는 고유벡터 v1과 v2를 활용하여 중심화 변수의 선형결합인 주성분 p1과 p2를 구한다.(S)

```
> v_s
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] -0.02 -0.10  0.01 -0.32 -0.31  0.88  0.03 -0.09
[2,] -0.04 -0.25  0.08 -0.90  0.17 -0.29 -0.07  0.04
[3,] -0.11 -0.92  0.25  0.29  0.01  0.00  0.00  0.00
[4,]  0.00 -0.01 -0.01 -0.03 -0.04 -0.13  0.19 -0.97
[5,] -0.01 -0.03 -0.04 -0.06 -0.21 -0.11  0.94  0.22
[6,] -0.08 -0.12 -0.38 -0.04 -0.83 -0.30 -0.25  0.04
[7,] -0.18 -0.21 -0.87  0.02  0.38  0.12  0.06 -0.01
[8,] -0.97  0.17  0.15  0.01  0.00  0.00  0.00  0.00
```

원 변수 :  $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)^t = (x_{100}, x_{200}, x_{400}, x_{800}, x_{1500}, x_{5000}, x_{10000}, \text{Marathon})^t$

중심화 변수 :  $y = (y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8)^t$

제 1주성분 :  $p_1 = -0.02 \times y_1 - 0.04 \times y_2 - 0.11 \times y_3 + 0.00 \times y_4 - 0.01 \times y_5 - 0.08 \times y_6 - 0.18 \times y_7 - 0.97 \times y_8$

제 2주성분 :  $p_2 = -0.10 \times y_1 - 0.25 \times y_2 - 0.92 \times y_3 - 0.01 \times y_4 - 0.03 \times y_5 - 0.12 \times y_6 - 0.21 \times y_7 + 0.17 \times y_8$

[5단계] 2개 고ؤ값 11과 12에 대응하는 고유벡터 v1과 v2를 활용하여 중심화 변수의 선형결합인 주성분 p1과 p2를 구한다.(R)

```
> v_r
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] -0.33 -0.53  0.34  0.38 -0.30 -0.36  0.35  0.07
[2,] -0.35 -0.47  0.00  0.22  0.54  0.35 -0.44 -0.06
[3,] -0.34 -0.35 -0.07 -0.85 -0.13  0.08  0.11  0.00
[4,] -0.35  0.09 -0.78  0.13  0.23 -0.34  0.26  0.04
[5,] -0.37  0.15 -0.24  0.23 -0.65  0.53 -0.15  0.04
[6,] -0.37  0.29  0.18 -0.05 -0.07 -0.36 -0.33 -0.71
[7,] -0.37  0.33  0.24 -0.09  0.06 -0.27 -0.35  0.70
[8,] -0.35  0.39  0.33  0.02  0.34  0.38  0.59 -0.07
```

원 변수 :  $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)^t = (x_{100}, x_{200}, x_{400}, x_{800}, x_{1500}, x_{5000}, x_{10000}, \text{Marathon})^t$

표준화 변수 :  $z = (z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8)^t$

제 1주성분 :  $p_1 = -0.33 \times z_1 - 0.35 \times z_2 - 0.34 \times z_3 - 0.35 \times z_4 - 0.37 \times z_5 - 0.37 \times z_6 - 0.37 \times z_7 - 0.35 \times z_8$

제 2주성분 :  $p_2 = -0.53 \times z_1 - 0.47 \times z_2 - 0.35 \times z_3 + 0.09 \times z_4 + 0.15 \times z_5 + 0.29 \times z_6 + 0.33 \times z_7 + 0.39 \times z_8$

[6단계] 중심화 자료를 주성분에 삽입하여 주성분점수를 얻는다.

중심화 자료 : Y

> Y

|                        | x100         | x200         | x400        | x800         | x1500        | x5000       | x10000      | Marathon   |
|------------------------|--------------|--------------|-------------|--------------|--------------|-------------|-------------|------------|
| Argentina              | 0.013148148  | -0.171481481 | 0.35092593  | 0.001851852  | 0.026666667  | -0.28759259 | -0.88518519 | -3.9085185 |
| Australia              | -0.286851852 | -0.481481481 | -1.44907407 | -0.028148148 | -0.123333333 | -0.68759259 | -1.00518519 | -5.9685185 |
| Austria                | -0.066851852 | -0.091481481 | -0.02907407 | 0.001851852  | -0.073333333 | -0.35759259 | -0.81518519 | -1.2585185 |
| Belgium                | -0.076851852 | -0.351481481 | -0.80907407 | -0.038148148 | -0.083333333 | -0.78759259 | -1.66518519 | -6.2785185 |
| Bermuda                | 0.053148148  | -0.241481481 | -0.56907407 | 0.021851852  | 0.046666667  | 1.02240741  | 1.95481481  | 12.8914815 |
| Brazil                 | -0.216851852 | -0.651481481 | -1.53907407 | -0.068148148 | -0.083333333 | -0.13759259 | -0.40518519 | -7.4285185 |
| Canada                 | -0.376851852 | -0.371481481 | -1.10907407 | -0.018148148 | -0.123333333 | -0.38759259 | -0.93518519 | -3.3885185 |
| Chile                  | -0.116851852 | -0.391481481 | 0.09092593  | -0.008148148 | -0.003333333 | -0.22759259 | -0.44518519 | -1.2885185 |
| China                  | -0.046851852 | -0.121481481 | -0.57907407 | 0.001851852  | -0.043333333 | -0.19759259 | -0.36518519 | -4.2985185 |
| Columbia               | 0.073148148  | 0.308518519  | 0.01092593  | 0.031851852  | 0.066666667  | -0.12759259 | -0.65518519 | -2.3085185 |
| CookIslands            | 0.753148148  | 1.918518519  | 5.57092593  | 0.171851852  | 0.586666667  | 3.08240741  | 6.84481481  | 37.7814815 |
| CostaRica              | 0.103148148  | 0.418518519  | 0.59092593  | 0.101851852  | 0.186666667  | 0.13240741  | 0.27481481  | -0.2485185 |
| CzechRepublic          | 0.023148148  | 0.068518519  | -0.05907407 | -0.018148148 | -0.073333333 | -0.19759259 | -0.73518519 | -1.9085185 |
| Denmark                | 0.073148148  | -0.021481481 | 0.06092593  | -0.078148148 | -0.133333333 | -0.19759259 | -0.62518519 | -4.0485185 |
| DominicanRepub         | -0.056851852 | 0.108518519  | -0.92907407 | 0.041851852  | 0.076666667  | 0.69240741  | 1.89481481  | 12.5214815 |
| Finland                | -0.006851852 | -0.071481481 | -0.33907407 | -0.028148148 | -0.043333333 | -0.34759259 | -1.01518519 | -2.3285185 |
| France                 | -0.196851852 | -0.381481481 | -1.18907407 | -0.048148148 | -0.173333333 | -0.63759259 | -1.15518519 | -7.1185185 |
| Germany                | -0.156851852 | -0.311481481 | -1.49907407 | -0.038148148 | -0.123333333 | -0.70759259 | -1.17518519 | -5.0085185 |
| GreatBritain           | -0.346851852 | -0.601481481 | -1.46907407 | -0.068148148 | -0.163333333 | -0.60759259 | -1.23518519 | -6.3485185 |
| Greece                 | -0.106851852 | -0.691481481 | -0.25907407 | -0.018148148 | -0.043333333 | -0.13759259 | -0.41518519 | -1.4385185 |
| Guatemala              | 0.103148148  | 0.548518519  | 2.61092593  | 0.051851852  | 0.086666667  | 0.36240741  | 0.80481481  | -0.9485185 |
| Hungary                | -0.136851852 | -0.431481481 | -0.39907407 | -0.008148148 | -0.063333333 | -0.16759259 | -0.50518519 | -1.3785185 |
| India                  | 0.113148148  | 0.188518519  | -0.34907407 | -0.008148148 | -0.023333333 | -0.11759259 | 0.27481481  | -1.4785185 |
| Indonesia              | -0.016851852 | 0.388518519  | 0.54092593  | 0.061851852  | 0.116666667  | 0.59240741  | 1.11481481  | 5.7014815  |
| Ireland                | 0.133148148  | -0.001481481 | -0.24907407 | -0.018148148 | -0.093333333 | -0.54759259 | -0.75518519 | -4.3285185 |
| Israel                 | -0.016851852 | 0.348518519  | 0.76092593  | 0.031851852  | 0.046666667  | 0.04240741  | 0.18481481  | 0.7314815  |
| Italy                  | -0.206851852 | -0.821481481 | -0.56907407 | -0.038148148 | -0.103333333 | -0.52759259 | -1.25518519 | -6.1885185 |
| Japan                  | -0.216851852 | -0.511481481 | -1.04907407 | 0.001851852  | -0.033333333 | -0.39759259 | -0.95518519 | -7.3185185 |
| Kenya                  | 0.063148148  | -0.111481481 | -1.64907407 | -0.068148148 | -0.213333333 | -0.95759259 | -2.07518519 | -8.9285185 |
| Korea, South           | 0.123148148  | -0.131481481 | -0.45907407 | -0.028148148 | -0.013333333 | 0.22240741  | -0.02518519 | -6.2785185 |
| Korea, North           | 0.383148148  | 0.688518519  | 1.12092593  | 0.051851852  | 0.116666667  | 0.28240741  | -0.08518519 | -4.2185185 |
| Luxembourg             | 0.193148148  | 0.228518519  | 2.07092593  | -0.008148148 | 0.016666667  | 0.02240741  | 0.23481481  | 0.5514815  |
| Malaysia               | 0.083148148  | 0.378518519  | 0.58092593  | 0.021851852  | 0.106666667  | 0.49240741  | 0.96481481  | 15.7914815 |
| Mauritius              | -0.086851852 | -0.481481481 | -1.13907407 | 0.031851852  | 0.176666667  | 0.53240741  | 1.30481481  | 9.5914815  |
| Mexico                 | -0.006851852 | -0.141481481 | -1.51907407 | 0.011851852  | -0.023333333 | -0.48759259 | -1.39518519 | -6.2885185 |
| Myanmar(Burma)         | 0.423148148  | 0.978518519  | 2.80092593  | 0.031851852  | 0.146666667  | 0.57240741  | 1.08481481  | 6.0914815  |
| Netherlands            | -0.026851852 | -0.351481481 | -0.14907407 | -0.038148148 | -0.103333333 | -0.39759259 | -1.09518519 | -5.1685185 |
| NewZealand             | -0.106851852 | -0.121481481 | 0.26092593  | -0.028148148 | -0.113333333 | -0.40759259 | -0.83518519 | -4.8885185 |
| Norway                 | -0.136851852 | -0.371481481 | 0.28092593  | -0.058148148 | -0.033333333 | -0.50759259 | -0.99518519 | -3.3085185 |
| PapuaNewGuinea         | 0.183148148  | 0.638518519  | 0.94092593  | 0.031851852  | 0.346666667  | 1.10240741  | 2.82481481  | 14.6514815 |
| Philippines            | 0.353148148  | 0.888518519  | -0.25907407 | 0.031851852  | 0.166666667  | 0.35240741  | 0.50481481  | 4.9614815  |
| Poland                 | -0.216851852 | -0.561481481 | -1.20907407 | -0.048148148 | -0.063333333 | -0.32759259 | -0.64518519 | -4.2485185 |
| Portugal               | -0.356851852 | -0.421481481 | 0.28092593  | -0.018148148 | -0.153333333 | -0.56759259 | -1.32518519 | -7.1185185 |
| Romania                | -0.006851852 | 0.208518519  | -0.05907407 | -0.008148148 | -0.083333333 | -0.36759259 | -0.86518519 | -1.1785185 |
| Russia                 | -0.106851852 | -0.311481481 | -1.22907407 | -0.058148148 | -0.113333333 | -0.41759259 | -0.63518519 | -4.3185185 |
| Samoa                  | 0.563148148  | 1.318518519  | 4.15092593  | 0.171851852  | 0.356666667  | 2.66240741  | 6.17481481  | 28.0214815 |
| Singapore              | 0.153148148  | 0.598518519  | 1.77092593  | 0.071851852  | 0.206666667  | 1.34240741  | 2.78481481  | 10.7414815 |
| Spain                  | -0.046851852 | 0.048518519  | -0.86907407 | -0.038148148 | -0.173333333 | -0.57759259 | -1.29518519 | -6.2485185 |
| Sweden                 | -0.036851852 | -0.111481481 | -0.28907407 | -0.008148148 | -0.043333333 | -0.32759259 | -0.60518519 | -3.0985185 |
| Switzerland            | -0.056851852 | -0.131481481 | -0.83907407 | -0.058148148 | -0.123333333 | -0.48759259 | -0.63518519 | -3.9185185 |
| Taiwan                 | 0.143148148  | 0.268518519  | 0.89092593  | 0.021851852  | 0.116666667  | 0.29240741  | 0.66481481  | 0.8714815  |
| Turkey                 | 0.163148148  | 0.498518519  | 0.80092593  | 0.011851852  | -0.063333333 | -0.16759259 | -0.20518519 | -3.2285185 |
| U.S.A.                 | -0.436851852 | -1.221481481 | -2.64907407 | -0.058148148 | -0.193333333 | -0.64759259 | -1.30518519 | -8.0985185 |
| attr(,"scaled:center") |              |              |             |              |              |             |             |            |
|                        | x100         | x200         | x400        | x800         | x1500        | x5000       | x10000      | Marathon   |
|                        | 10.216852    | 20.541481    | 45.829074   | 1.768148     | 3.653333     | 13.617593   | 28.535185   | 133.478519 |



주성분 점수 행렬 :  $P=YV(2)$  여기서  $V(2)=(v_1, v_2)$

```
> P_S
      [,1]      [,2]
Argentina  3.94133148 -0.72616296
Australia  6.21103148  0.76513704
Austria    1.40503148  0.05863704
Belgium     6.55833148  0.21963704
Bermuda    -12.86766852  2.23533704
Brazil      7.49013148  0.44243704
Canada      3.63183148  0.82163704
Chile       1.35623148 -0.07216296
China       4.32103148 -0.06126296
Columbia    2.35173148 -0.33636296
CookIslands -38.83716852 -1.08396296
CostaRica   0.09533148 -0.78106296
CzechRepublic 2.00343148 -0.10906296
Denmark     4.04943148 -0.58646296
DominicanRepub -12.44406852  2.47823704
Finland     2.50993148  0.19113704
France      7.31563148  0.32363704
Germany     5.30813148  0.95703704
GreatBritain 6.62323148  0.79523704
Greece      1.53983148  0.28253704
Guatemala   0.43413148 -2.92636296
Hungary     1.50603148  0.38253704
India       1.42293148 -0.03146296
Indonesia   -5.85436852  0.06683704
Ireland     4.40413148 -0.29236296
Israel      -0.84396852 -0.70676296
Italy       6.37163148  0.02793704
Japan       7.44323148  0.11983704
Kenya       9.29753148  0.57863704
Korea,South 6.13033148 -0.64516296
Korea,North 3.92503148 -1.97886296
Luxembourg  -0.81996852 -1.94036296
Malaysia    -15.61256852  1.78203704
Mauritius   -9.43666852  2.46403704
Mexico      6.56313148  0.71663704
Myanmar(Burma) -6.50696852 -2.12946296
Netherlands 5.27443148 -0.36976296
NewZealand  4.90423148 -0.80206296
Norway      3.41603148 -0.44286296
PapuaNewGuinea -14.94476852  0.71093704
Philippines -4.94746852  0.67073704
Poland      4.42383148  0.72933704
Portugal    7.18353148 -0.97636296
Romania     1.32743148  0.03093704
Russia      4.48763148  0.67263704
Samoa       -29.02946852 -1.06976296
Singapore   -11.25176852 -0.72096296
Spain       6.43673148  0.07673704
Sweden      3.17813148 -0.06146296
Switzerland 4.05423148  0.34053704
Taiwan      -1.10116852 -0.93136296
Thailand    -5.96246852  0.43493704
Turkey      3.07133148 -1.36166296
U.S.A.      8.49323148  1.76763704
```

```
> round(V_S2,2)
      [,1] [,2]
[1,] -0.02 -0.10
[2,] -0.04 -0.25
[3,] -0.11 -0.92
[4,]  0.00 -0.01
[5,] -0.01 -0.03
[6,] -0.08 -0.12
[7,] -0.18 -0.21
[8,] -0.97  0.17
```



[6단계] 표준화 자료를 주성분에 삽입하여 주성분점수를 얻는다.

표준화 자료 : Z

```
'> z
      x100      x200      x400      x800      x1500      x5000      x10000      Marathon
Argentina  0.05941370 -0.312610573  0.243912988  0.03530558  0.17570512 -0.37799427 -0.52716041 -0.43661644
Australia -1.29622279 -0.877740269 -1.007186875 -0.53664481 -0.81263616 -0.90373002 -0.59862483 -0.66673685
Austria   -0.30208937 -0.166770651 -0.020208094  0.03530558 -0.48318907 -0.46999802 -0.48547283 -0.14058776
Belgium   -0.34727725 -0.640750396 -0.562351368 -0.72729495 -0.54907849 -1.03516395 -0.99167916 -0.70136662
Bermuda    0.24016523 -0.440220504 -0.395538053  0.41660584  0.30748395  1.34379031  1.16416428  1.44009368
Brazil     -0.97990761 -1.187650102 -1.069741868 -1.29924534 -0.54907849 -0.18084336 -0.24130271 -0.82983190
Canada     -1.70291374 -0.677210376 -0.770868012 -0.34599468 -0.81263616 -0.50942820 -0.55693725 -0.37852780
Chile      -0.52802878 -0.713670357  0.063198563 -0.15534455 -0.02196314 -0.29913391 -0.26512419 -0.14393903
China      -0.21171360 -0.221460622 -0.402488608  0.03530558 -0.28552081 -0.25970372 -0.21748124 -0.48018293
Columbia   0.33054100  0.562428957  0.007594125  0.60725597  0.43926279 -0.16769997 -0.39018693 -0.25788215
CookIslands 3.40331704  3.497457378  3.872102589  3.27635782  3.86551255  4.05132942  4.07633955  4.22052908
CostaRica  0.46610465  0.762958849  0.410726303  1.94180690  1.22993581  0.17402827  0.16366235 -0.02776174
CzechRepublic 0.10460158  0.124909192 -0.041059759 -0.34599468 -0.48318907 -0.25970372 -0.43782988 -0.21319857
Denmark    0.33054100 -0.039160720  0.042346899 -1.48989547 -0.87852558 -0.25970372 -0.37232082 -0.45225569
DominicanRepub -0.25690148  0.197829153 -0.645758026  0.79790611  0.50515221  0.91005832  1.12843207  1.39876137
Finland    -0.03096207 -0.130310671 -0.235675293 -0.53664481 -0.28552081 -0.45685463 -0.60458020 -0.26011633
France     -0.88953185 -0.695440367 -0.826472450 -0.91794508 -1.14208325 -0.83801305 -0.68795536 -0.79520213
Germany    -0.70878031 -0.567830435 -1.041939649 -0.72729495 -0.81263616 -0.93001680 -0.69986610 -0.55949627
GreatBritain -1.56735009 -1.096500151 -1.021087985 -1.29924534 -1.07619383 -0.79858287 -0.73559831 -0.70918625
Greece     -0.48284090 -1.260570063 -0.180070855 -0.34599468 -0.28552081 -0.18084336 -0.24725808 -0.16069537
Guatemala  0.46610465  0.999948721  1.814738371  0.98855624  0.57104163  0.47632632  0.47929689 -0.10595799
Hungary    -0.61840455 -0.786590318 -0.277378622 -0.15534455 -0.41729965 -0.22027354 -0.30085640 -0.15399284
India      0.51129253  0.343669074 -0.242625848 -0.15534455 -0.15374198 -0.15455657  0.16366235 -0.16516373
Indonesia -0.07614995  0.708268878  0.375973529  1.17920637  0.76870988  0.77862438  0.66391332  0.63690643
Ireland    0.60166829 -0.002700739 -0.173120300 -0.34599468 -0.61496791 -0.71972250 -0.44974062 -0.48353420
Israel     -0.07614995  0.635348918  0.528885734  0.60725597  0.30748395  0.05573773  0.11006404  0.08171302
Italy      -0.93471973 -1.497559935 -0.395538053 -0.72729495 -0.68085732 -0.69343572 -0.74750905 -0.69131282
Japan      -0.97990761 -0.932430239 -0.729164683  0.03530558 -0.21963139 -0.52257160 -0.56884799 -0.81754391
Kenya      0.28535311 -0.203230631 -1.146197971 -1.29924534 -1.40564093 -1.25860165 -1.23584928 -0.99739530
Korea, South 0.55648041 -0.239690612 -0.319081951 -0.53664481 -0.08785256  0.29231881 -0.01499871 -0.70136662
Korea, North 1.73136537  1.255168584  0.779105707  0.98855624  0.76870988  0.37117917 -0.05073092 -0.47124621
Luxembourg 0.87279559  0.416589035  1.439408412 -0.15534455  0.10981570  0.02945094  0.13984088  0.06160541
Malaysia   0.37572888  0.690038888  0.403775748  0.41660584  0.70282046  0.64719044  0.57458279  1.76404959
Mauritius  -0.39246513 -0.877740269 -0.791719676  0.60725597  1.16404639  0.69976402  0.77706532  1.07145419
Mexico     -0.03096207 -0.257920602 -1.055840759  0.22595571 -0.15374198 -0.64086214 -0.83088421 -0.70248371
Myanmar (Burma) 1.91211690  1.783838300  1.946798912  0.60725597  0.96637814  0.75233759  0.64604721  0.68047291
Netherlands -0.12133783 -0.640750396 -0.103614752 -0.72729495 -0.68085732 -0.52257160 -0.65222315 -0.57736970
NewZealand -0.48284090 -0.221460622  0.181357994 -0.53664481 -0.74674674 -0.53571499 -0.49738357 -0.54609120
Norway     -0.61840455 -0.677210376  0.195259104 -1.10859521 -0.21963139 -0.66714893 -0.59266946 -0.36959108
PapuaNewGuinea 0.82760771  1.164018633  0.653995720  0.60725597  2.28416651  1.44893746  1.68228135  1.63670140
Philippines 1.59580172  1.619768388 -0.180070855  0.60725597  1.09815697  0.46318293  0.30063583  0.55424182
Poland     -0.97990761 -1.023580190 -0.840373560 -0.91794508 -0.41729965 -0.43056784 -0.38423156 -0.47459748
Portugal   -1.61253797 -0.768360327  0.195259104 -0.34599468 -1.01030442 -0.74600929 -0.78919663 -0.79520213
Romania    -0.03096207  0.380129055 -0.041059759 -0.15534455 -0.54907849 -0.48314142 -0.51524967 -0.13165105
Russia     -0.48284090 -0.567830435 -0.854274670 -1.10859521 -0.74674674 -0.54885839 -0.37827619 -0.48241711
Samoa      2.54474726  2.403657966  2.885123809  3.27635782  2.35005593  3.49930688  3.67732985  3.13024987
Singapore  0.69204406  1.091098672  1.230891768  1.36985650  1.36171465  1.76437891  1.65845988  1.19991946
Spain      -0.21171360  0.088449212 -0.604054697 -0.72729495 -1.14208325 -0.75915269 -0.77133052 -0.69801535
Sweden     -0.16652572 -0.203230631 -0.200922519 -0.15534455 -0.28552081 -0.43056784 -0.36041009 -0.34613221
Switzerland -0.25690148 -0.239690612 -0.583203033 -1.10859521 -0.81263616 -0.64086214 -0.37827619 -0.43773353
Taiwan     0.64685618  0.489508996  0.619242946  0.41660584  0.76870988  0.38432257  0.39592173  0.09735227
Thailand    0.05941370  0.270749114  0.153555775  0.79790611  0.76870988  0.83119795  0.67582406  0.65366277
Turkey     0.73723194  0.908798770  0.556687953  0.22595571 -0.41729965 -0.22027354 -0.12219534 -0.36065437
U.S.A.     -1.97404104 -2.226759543 -1.841253450 -1.10859521 -1.27386209 -0.85115644 -0.77728589 -0.90467688
attr(,"scaled:center")
      x100      x200      x400      x800      x1500      x5000      x10000      Marathon
10.216852 20.541481 45.829074 1.768148 3.653333 13.617593 28.535185 133.478519
attr(,"scaled:scale")
      x100      x200      x400      x800      x1500      x5000      x10000      Marathon
0.2212983 0.5485466 1.4387341 0.0524521 0.1517694 0.7608385 1.6791572 8.9518354
```

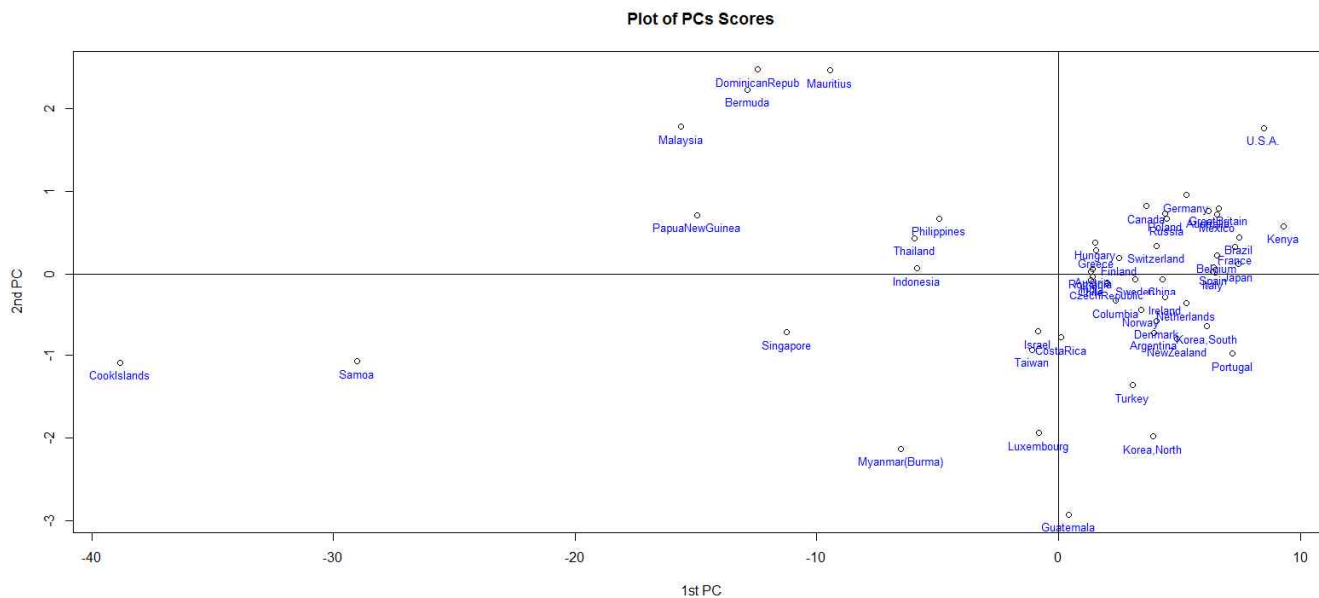


주성분 점수 행렬 :  $P=ZV(2)$  여기서  $V(2)=(v1, v2)$

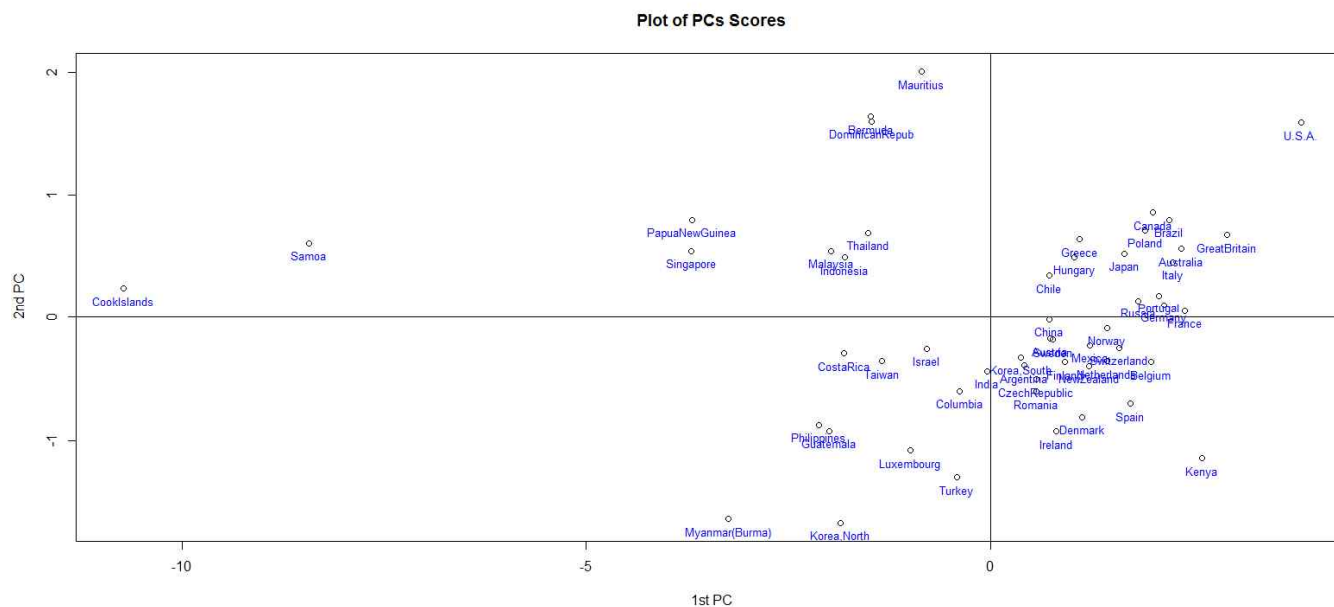
```
> P_R
      [,1]      [,2]
Argentina  0.41723190 -0.39426025
Australia  2.35513641  0.56220268
Austria    0.73408291 -0.17507314
Belgium     1.98318614 -0.36677036
Bermuda    -1.48225151  1.63718144
Brazil      2.20728940  0.79695347
Canada      2.00989376  0.85845270
Chile       0.72419772  0.34550931
China       0.72213104 -0.01683914
Columbia    -0.38690059 -0.59961254
CookIslands -10.72510716  0.23798410
CostaRica   -1.81043466 -0.29647764
CzechRepublic 0.56830861 -0.50633851
Denmark     1.12888623 -0.81203171
DominicanRepub -1.47488636  1.59859420
Finland     0.91319580 -0.36442857
France      2.40473091  0.05346878
Germany     2.14100662  0.10099655
GreatBritain 2.91696031  0.67415115
Greece      1.10314286  0.64071982
Guatemala   -1.99458301 -0.92256450
Hungary     1.02917588  0.49473927
India       -0.04082463 -0.43985945
Indonesia   -1.80439671  0.49060339
Ireland     0.81183021 -0.92611972
Israel      -0.79331865 -0.25823574
Italy       2.24866592  0.45272146
Japan       1.65650837  0.52492595
Kenya       2.61352971 -1.14413659
Korea,South 0.37194205 -0.32578736
Korea,North -1.85960254 -1.66884860
Luxembourg  -0.99368865 -1.08096574
Malaysia    -1.97811696  0.54341839
Mauritius   -0.85876530  2.00413658
Mexico      1.22769103 -0.22955890
Myanmar(Burma) -3.24292101 -1.63683790
Netherlands 1.44275703 -0.35781183
NewZealand  1.21268736 -0.39626267
Norway      1.43947066 -0.08821031
PapuaNewGuinea -3.69195326  0.79631693
Philippines -2.12776466 -0.87497749
Poland      1.91061613  0.71260036
Portugal    2.07593326  0.17784220
Romania     0.56414483 -0.60570994
Russia      1.82472194  0.13785379
Samoa       -8.42917795  0.60826726
Singapore   -3.69846864  0.54406267
Spain       1.73199506 -0.70163215
Sweden      0.76821926 -0.18150079
Switzerland 1.58592986 -0.25013280
Taiwan      -1.34833167 -0.35676192
Thailand    -1.51664758  0.69362999
Turkey      -0.42238113 -1.29982706
U.S.A.      3.83532342  1.59024086
```

```
> round(V_R2, 2)
      [,1] [,2]
[1,] -0.33 -0.53
[2,] -0.35 -0.47
[3,] -0.34 -0.35
[4,] -0.35  0.09
[5,] -0.37  0.15
[6,] -0.37  0.29
[7,] -0.37  0.33
[8,] -0.35  0.39
```

[7단계] 주성분점수를 차원 축소된 새로운 다변량 자료로 고려한다.(S)



[7단계] 주성분점수를 차원 축소된 새로운 다변량 자료로 고려한다.(R)



주성분점수의 산점도를 비교해볼 때 공분산행렬을 사용하였을 때의 데이터들의 분산이 크다. 이는 변수 Marathon의 분산이 다른 변수들에 비해 크기 때문에 이러한 결과가 나온 것으로 보인다. 따라서 공분산행렬을 이용한 주성분 분석보다 상관행렬을 이용하는 것이 적절해 보인다.

(2) 두 개의 주성분을 결정하고 주성분과 변수 간의 상관계수를 이용하여 주성분을 해석하라.

```
> round(V_R2, 2)
      [,1] [,2]
[1,] -0.33 -0.53
[2,] -0.35 -0.47
[3,] -0.34 -0.35
[4,] -0.35  0.09
[5,] -0.37  0.15
[6,] -0.37  0.29
[7,] -0.37  0.33
[8,] -0.35  0.39
```

원 변수 :  $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)^t = (x_{100}, x_{200}, x_{400}, x_{800}, x_{1500}, x_{5000}, x_{10000}, \text{Marathon})^t$

표준화 변수 :  $z = (z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8)^t$

제 1주성분 :  $p_1 = -0.33 \times z_1 - 0.35 \times z_2 - 0.34 \times z_3 - 0.35 \times z_4 - 0.37 \times z_5 - 0.37 \times z_6 - 0.37 \times z_7 - 0.35 \times z_8$

제 2주성분 :  $p_2 = -0.53 \times z_1 - 0.47 \times z_2 - 0.35 \times z_3 + 0.09 \times z_4 + 0.15 \times z_5 + 0.29 \times z_6 + 0.33 \times z_7 + 0.39 \times z_8$

$m=2$ 개 고윳값( $\lambda_1=6.70, \lambda_2=0.64$ )의 설명비율 합은  $83.65\%+7.99\% = 91.64\%$ 이다.

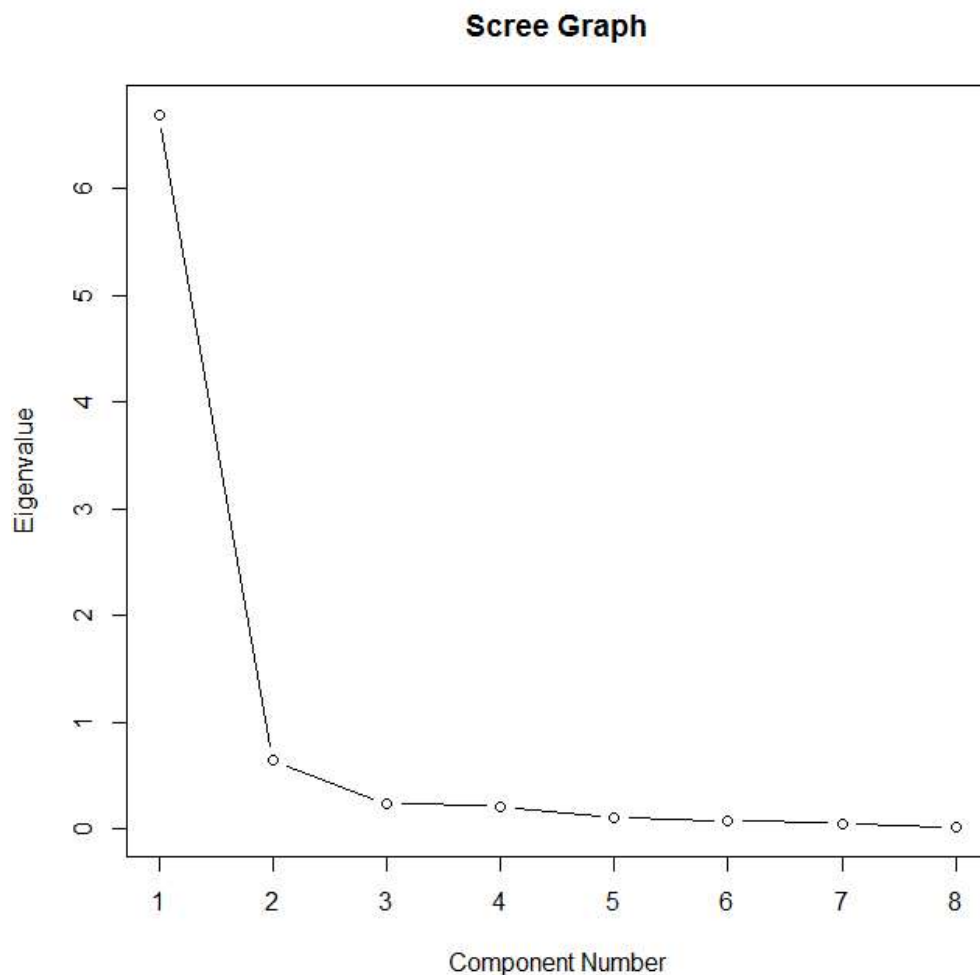
해석 : 제 1주성분  $p_1$ 의 주성분계수가 모두 음의 값이다. 이는 육상기록에 대한 평균적인 성분을 나타내고 있다.

제 1주성분의 설명력은 83.65%이다.

제 2주성분은 주성분계수 부호가 양인  $x_{100}, x_{200}, x_{400}$ 와 주성분계수 부호가 음인  $x_{800}, x_{1500}, x_{5000}, x_{10000}, \text{Marathon}$ 의 대비를 나타낸다. 즉 단거리와 단거리 육상종목에 대한 성분을 나타내고 있다. 제 2주성분의 설명력은 7.99%이다.

따라서 두 주성분의 설명력은 91.64%이다.

(3). (2)에서 선택한 주성분의 설명력과 스크리그림을 통해 주성분 2개가 타당함을 살펴보자.



```
> round(gof_R, 2)
[1] 83.65  7.99  2.87  2.62  1.25  0.87  0.62  0.12
```

m=2개 고윳값( $\lambda_1=6.70$ ,  $\lambda_2=0.64$ )의 설명비율 합은  $83.65\%+7.99\% = 91.64\%$ 이다.

scree graph를 보면 팔꿈치가 2에서 이루어진다. 주성분의 개수는 1개로 하는 것이 타당해 보인다.

제 1주성분만으로도 약 83.65% 설명력을 확보하여 제 1주성분에 대한 충분한 해석이 이루어지게 된다.

하지만, [7단계]에서 시각적 편의상 주성분점수를 보여줄 2차원 주성분공간을 얻기 위해 제 2주성분도 함께 고려하였다. 그러면 설명력은 91.64%이다. 따라서 주성분 2개를 선택하여도 타당하다.

(4) 제 1주성분 점수에 의해 모든 국가들의 순위를 부여하고 이러한 순위가 그들 국가들에 대한 육상의 선진국과 후진국으로 알려진 것과 일치하는지를 살펴보자.

제 1 주성분 점수에 의한 국가 순위

```
> rank(-sort(P_R[,1],decreasing = T))
```

|             |                 |                |
|-------------|-----------------|----------------|
| U.S.A.      | GreatBritain    | Kenya          |
| 1           | 2               | 3              |
| France      | Australia       | Italy          |
| 4           | 5               | 6              |
| Brazil      | Germany         | Portugal       |
| 7           | 8               | 9              |
| Canada      | Belgium         | Poland         |
| 10          | 11              | 12             |
| Russia      | Spain           | Japan          |
| 13          | 14              | 15             |
| Switzerland | Netherlands     | Norway         |
| 16          | 17              | 18             |
| Mexico      | NewZealand      | Denmark        |
| 19          | 20              | 21             |
| Greece      | Hungary         | Finland        |
| 22          | 23              | 24             |
| Ireland     | Sweden          | Austria        |
| 25          | 26              | 27             |
| Chile       | China           | CzechRepublic  |
| 28          | 29              | 30             |
| Romania     | Argentina       | Korea,South    |
| 31          | 32              | 33             |
| India       | Columbia        | Turkey         |
| 34          | 35              | 36             |
| Israel      | Mauritius       | Luxembourg     |
| 37          | 38              | 39             |
| Taiwan      | DominicanRepub  | Bermuda        |
| 40          | 41              | 42             |
| Thailand    | Indonesia       | CostaRica      |
| 43          | 44              | 45             |
| Korea,North | Malaysia        | Guatemala      |
| 46          | 47              | 48             |
| Philippines | Myanmar (Burma) | PapuaNewGuinea |
| 49          | 50              | 51             |
| Singapore   | Samoa           | CookIslands    |
| 52          | 53              | 54             |

2019 세계육상선수권대회 메달집계표

Medal Table

Select a championship

IAAF World Athletics Championships, DOHA 2019

Medal Table Placing Table

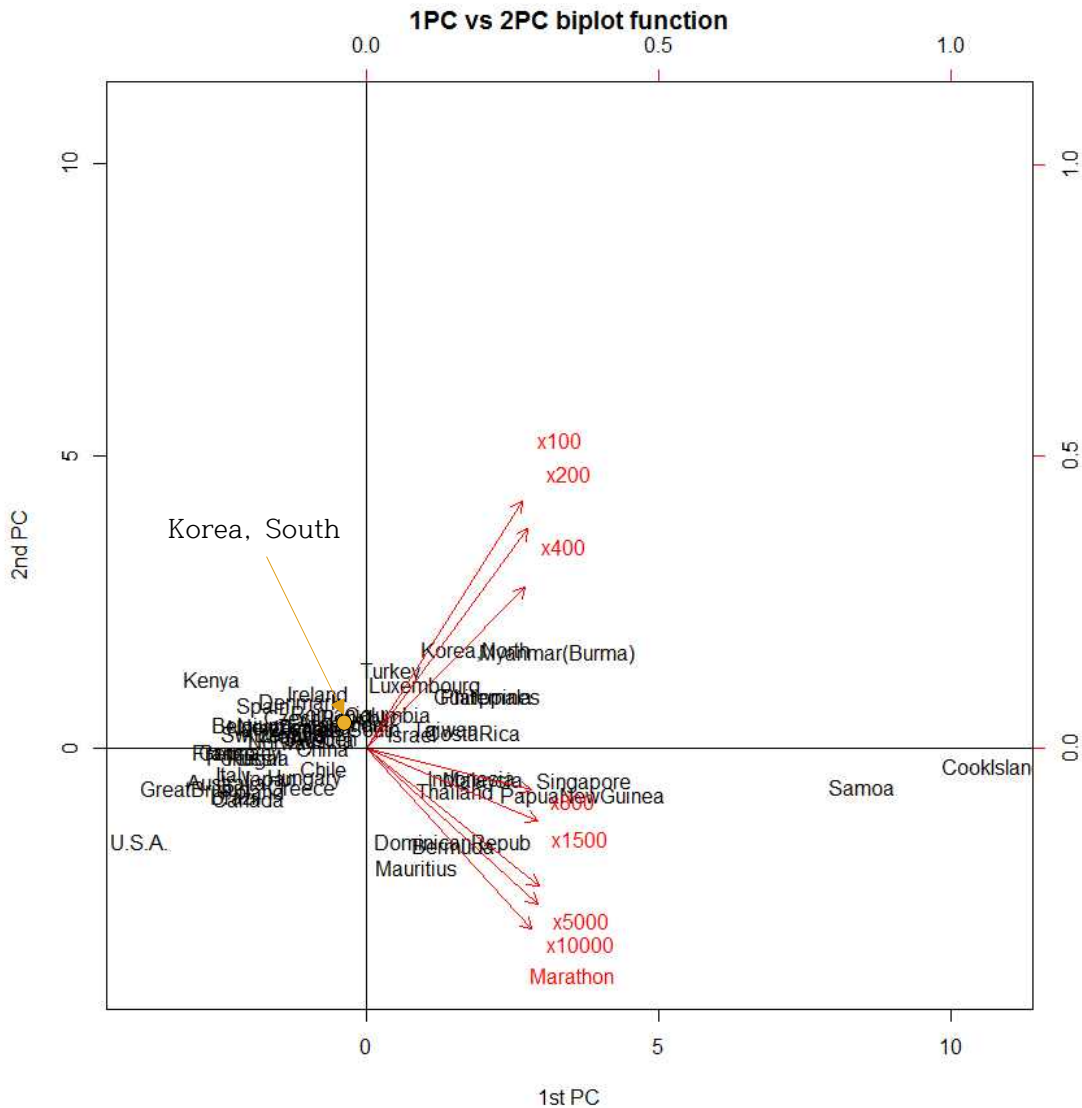
Medal Table Please click on a row below to view Medal Winners

VIEW DOWNLOAD

| Rank | Country            | Gold | Silver | Bronze | Total |
|------|--------------------|------|--------|--------|-------|
| 1    | UNITED STATES      | 14   | 11     | 4      | 29    |
| 2    | KENYA              | 5    | 2      | 4      | 11    |
| 3    | JAMAICA            | 3    | 5      | 4      | 12    |
| 4    | PR OF CHINA        | 3    | 3      | 3      | 9     |
| 5    | ETHIOPIA           | 2    | 5      | 1      | 8     |
| 6    | GREAT BRITAIN & NI | 2    | 3      | 0      | 5     |
| 7    | GERMANY            | 2    | 0      | 4      | 6     |
| 8    | JAPAN              | 2    | 0      | 1      | 3     |
| 9    | NETHERLANDS        | 2    | 0      | 0      | 2     |
| 9    | UGANDA             | 2    | 0      | 0      | 2     |
| 11   | POLAND             | 1    | 2      | 3      | 6     |
| 12   | BAHRAIN            | 1    | 1      | 1      | 3     |
| 12   | CUBA               | 1    | 1      | 1      | 3     |
| 12   | SWEDEN             | 1    | 1      | 1      | 3     |
| 15   | BAHAMAS            | 1    | 1      | 0      | 2     |
| 16   | QATAR              | 1    | 0      | 1      | 2     |
| 17   | AUSTRALIA          | 1    | 0      | 0      | 1     |
| 17   | GRENADA            | 1    | 0      | 0      | 1     |
| 17   | NORWAY             | 1    | 0      | 0      | 1     |
| 17   | VENEZUELA          | 1    | 0      | 0      | 1     |
| 21   | ESTONIA            | 0    | 2      | 0      | 2     |
| 21   | UKRAINE            | 0    | 2      | 0      | 2     |
| 23   | CANADA             | 0    | 1      | 4      | 5     |
| 24   | BELGIUM            | 0    | 1      | 1      | 2     |
| 24   | COLOMBIA           | 0    | 1      | 1      | 2     |
| 24   | FRANCE             | 0    | 1      | 1      | 2     |
| 27   | ALGERIA            | 0    | 1      | 0      | 1     |
| 27   | BOSNIA-HERZEGOVINA | 0    | 1      | 0      | 1     |
| 27   | PORTUGAL           | 0    | 1      | 0      | 1     |
| 30   | AUSTRIA            | 0    | 0      | 2      | 2     |
| 31   | BURKINA FASO       | 0    | 0      | 1      | 1     |
| 31   | COTE D'IVOIRE      | 0    | 0      | 1      | 1     |
| 31   | CROATIA            | 0    | 0      | 1      | 1     |
| 31   | EQUADOR            | 0    | 0      | 1      | 1     |
| 31   | SPAIN              | 0    | 0      | 1      | 1     |
| 31   | GREECE             | 0    | 0      | 1      | 1     |
| 31   | HUNGARY            | 0    | 0      | 1      | 1     |
| 31   | ITALY              | 0    | 0      | 1      | 1     |
| 31   | MOROCCO            | 0    | 0      | 1      | 1     |
| 31   | NAMIBIA            | 0    | 0      | 1      | 1     |
| 31   | NIGERIA            | 0    | 0      | 1      | 1     |
| 31   | NEW ZEALAND        | 0    | 0      | 1      | 1     |
| 31   | SWITZERLAND        | 0    | 0      | 1      | 1     |

세계육상선수권대회 자료를 참고하여 제 1 주성분 점수에 의한 국가 순위와 비교해봤을 때 실제로 육상의 선진국과 후진국으로 알려진 나라들이 일치한다.

(5) 주성분행렬도에서 육상기록변수들의 관계와 (4)에서 살펴본 국가들과의 관계를 시각적으로 살펴보고 설명하라.



육상선진국들은 제 1주성분을 기준으로 왼편에 육상후진국들은 오른편에 위치한다.

(x100, x200, x400, x800, x1500, x5000, x10000, Marathon)<sup>t</sup>

제 1주성분과 제 2주성분을 비교한 주성분행렬도에서 단거리종목인 변수 x100, x200, x400간에 큰 상관관계를 보이며, 장거리종목인 변수 x800, x1500, x5000, x10000, Marathon간에 큰 상관관계를 보인다. 또한 단거리종목과 장거리종목은 서로 90에 가까워 큰 상관이 없어보인다.

제 1 주성분에서 모든 요인들은 양으로 화살표가 뻗어있다. 4)에서 본 나라들과 비교해봤을 때 육상선진국들은 제 1주성분을 기준으로 왼편에 육상후진국들은 오른편에 위치한다.

제 2주성분에서 x100, x200, x400는 양, x800, x1500, x5000, x10000, Marathon은 음으로 서로 반대 방향으로 다른 부호를 가지고 있다. 이를 바탕으로 제 2주성분을 기준으로 위쪽에 나라들은 장거리종목에 강하고 아래쪽에 있는 나라들은 단거리 종목에 강하다.



(6) 우리나라는 주성분행렬도에서 육상 후진국인지를 살펴보고, 특히 어느 종목에 취약한지를 살펴보라.

주성분 행렬도를 보면 우리나라는 제 1주성분에서 오른쪽에 있지 않으므로 육상 후진국이라고 할 수 없다. 우리나라의 점과 원점 간의 직선을 그었을 때, 직선과 다른 변수들과 각도를 보면 100m 기록은 각도가 가장 좁아  $\cos$ 이 가장 큰 값을 가진다. 큰 값을 가진다는 것은 기록이 크다는 의미이므로 우리나라가 100m에서 가장 취약함을 알 수 있다. 그리고 대체로 단거리 종목에 취약하다. 반대로 장거리 기록에서는 마라톤이 가장 각도가 크므로  $\cos$ 이 작은 값을 가진다. 이는 우리나라가 마라톤종목에 강하다는 것을 알 수 있다. 따라서 우리나라는 육상에서 중간정도의 성적을 가지고 있다.



```

setwd("D:/2020 1학기 정호재/다변량통계학(1)/200402 다변량 실습1/Rdata")
data<-read.table("trackrecord2005-men.txt", header=T)
rownames<-rownames(data)
head(data)
dim(data)

S<-cov(data)
S
R<-cor(data)
R
eigen.S<-eigen(S)
D_S<-round(eigen.S$values,2) # Eigenvalue
V_S<-round(eigen.S$vectors,2) # Eigenvectors

eigen.R<-eigen(R)
D_R<-round(eigen.R$values,2) # Eigenvalue
V_R<-round(eigen.R$vectors,2) # Eigenvectors

gof_S<-D_S/sum(D_S)*100 # Goodness-of fit
round(gof_S, 2)
plot(D_S, type="b", main="Scree Graph", xlab="Component Number", ylab="Eigenvalue")
gof_R<-D_R/sum(D_R)*100
round(gof_R,2)
plot(D_R, type="b", main="Scree Graph", xlab="Component Number", ylab="Eigenvalue")

V_S2<-V_S[,1:2] #one selected eigen vectors corresponding eigen values
round(V_S2,2)

V_R2<-V_R[,1:2]
round(V_R2,2)

Y<-scale(data, scale=F) # Centred Data Matrix
P_S<-Y%*%V_S2 # PCs Scores
P_S
Z<-scale(data,scale=T)
P_R<-Z%*%V_R2

plot(P_S[,1], P_S[, 2], main="Plot of PCs Scores", xlab="1st PC", ylab="2nd PC")
text(P_S[,1], P_S[, 2], labels=rownames, cex=0.8, col="blue", pos=1)
abline(v=0, h=0)

```

```
plot(P_R[,1], P_R[,2], main="Plot of PCs Scores", xlab="1st PC", ylab="2nd PC")
text(P_R[,1], P_R[,2], labels=rownames, cex=0.8, col="blue", pos=1)
abline(v=0, h=0)
```

```
#####
```

```
rank(-sort(P_R[,1],decreasing = T))
```

```
X=data
```

```
pca.R<-princomp(X, cor=T)
```

```
summary(pca.R, loadings=T)
```

```
round(pca.R$scores, 3)
```

```
screeplot(pca.R, type="lines")
```

```
# Principle component biplot (SVD)
```

```
biplot(pca.R, scale=0, xlab="1st PC",ylab="2nd PC",
      main="1PC vs 2PC biplot function")
```

```
abline(v=0, h=0)
```

```
#####
```

```
X=data
```

```
n <- nrow(X)
```

```
rownames(X)
```

```
colnames(X)
```

```
Y <- scale(X,scale=T)
```

```
# Biplot based on the Singular Value Decomposition
```

```
svd.Y <- svd(Y)
```

```
U <- svd.Y$u
```

```
V <- svd.Y$v
```

```
D <- diag(svd.Y$d)
```

```
G <- (sqrt(n-1)*U)[,1:2]
```

```
H <- (sqrt(1/(n-1))*V%*%D)[,1:2]
```

```
rownames(G)<-rownames(X)
```

```
rownames(H)<-colnames(X)
```

```
biplot(G,H, xlab="1st PC",ylab="2nd PC", main="1PC vs 2PC biplot function",
      xlim=lim2,ylim=lim2,cex=0.8,pch=16)
```

```
abline(v=0,h=0)
```