

제 10 장 주성분분석

10.1 개념 및 목적

주성분분석(principal component analysis)의 idea :

여러 개($p \geq 2$)의 다변량 양적변수

중요한 $m(\leq p)$ 개의 (보통 2~3개) 주성분으로 표현

즉, p 차원의 자료를 2차원 또는 3차원의 주성분 공간으로 사영
이들을 그래프로 나타내어 자료가 갖는 특성을 찾아냄

[예제 10.1] 인체의 측정자료로서, $n = 110$ 명에 대하여

$$x_1 = \text{키}, \quad x_2 = \text{몸무게}, \quad x_3 = \text{가슴둘레}$$

상관행렬의 고유값은 $(2.1, 0.6, 0.3)$

$$y_1 = 0.8x_1 + 0.5x_2 + 0.3x_3$$

$$y_2 = -0.5x_1 + 0.8x_2 + 0.3x_3$$

$$y_3 = -0.3x_1 - 0.3x_2 + 0.9x_3$$

제1주성분 : '크기(size)'; y_2 와 y_3 는 '모양(shape)'

(y_1, y_2)로는 전체의 $(2.1 + 0.6)/3 = 0.9$ (즉, 90%)를 설명

10.2 이론적 배경

10.2.1 주성분 모형

$\mathbf{x}' = (x_1, x_2, \dots, x_p)$: 평균 μ , 공분산행렬 Σ

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$: Σ 의 고유값(eigenvalue)

$\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$: $\lambda_1, \lambda_2, \dots, \lambda_p$ 에 대응되는

고유벡터(eigenvector)

(고유벡터들은 모두 단위벡터이며 서로 수직)

$$y_1 = \ell_1' \mathbf{X} = \ell_{11} X_1 + \ell_{12} X_2 + \cdots + \ell_{1p} X_p$$

$$y_2 = \ell_2' \mathbf{X} = \ell_{21} X_1 + \ell_{22} X_2 + \cdots + \ell_{2p} X_p$$

$$\vdots$$

$$y_p = \ell_p' \mathbf{X} = \ell_{p1} X_1 + \ell_{p2} X_2 + \cdots + \ell_{pp} X_p$$

$\ell_i' = (\ell_{i1}, \ell_{i2}, \cdots, \ell_{ip}), \quad i=1, \cdots, p$ 크기가 1인 단위벡터

1. y_1 은 x 들의 선형결합 가운데 최대분산
2. y_2 는 y_1 과 무상관이면서 x 들의 선형결합 가운데 최대분산
- \vdots
- p. y_p 는 y_1, \dots, y_{p-1} 과 무상관이면서 x 들의 선형결합 가운데 최대분산

y_1 : 제1주성분 (1st principal component)

y_2 : 제2주성분

...

y_p : 제 p 주성분

1. ℓ_1 은 제일 큰 고유값 λ_1 에 대응되는 고유벡터 e_1

2. ℓ_2 는 λ_2 에 대응되는 고유벡터 e_2

\vdots

p. ℓ_p 는 λ_p 에 대응되는 고유벡터 e_p

주성분들의 성질 :

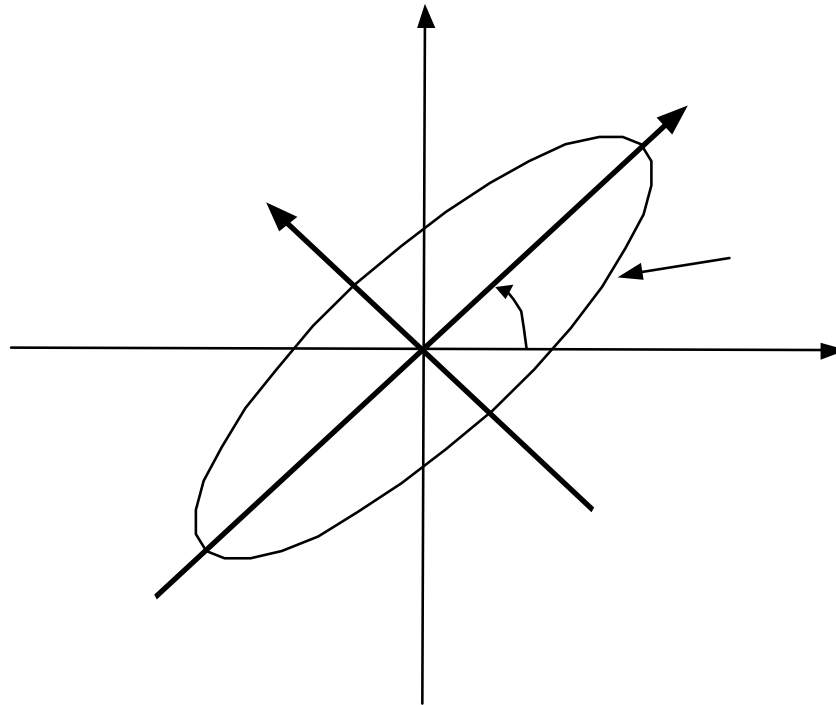
1. 주성분들은 서로 무상관(orthogonal)

$$\text{Cov}(y_i, y_j) = 0, \quad \forall i \neq j.$$

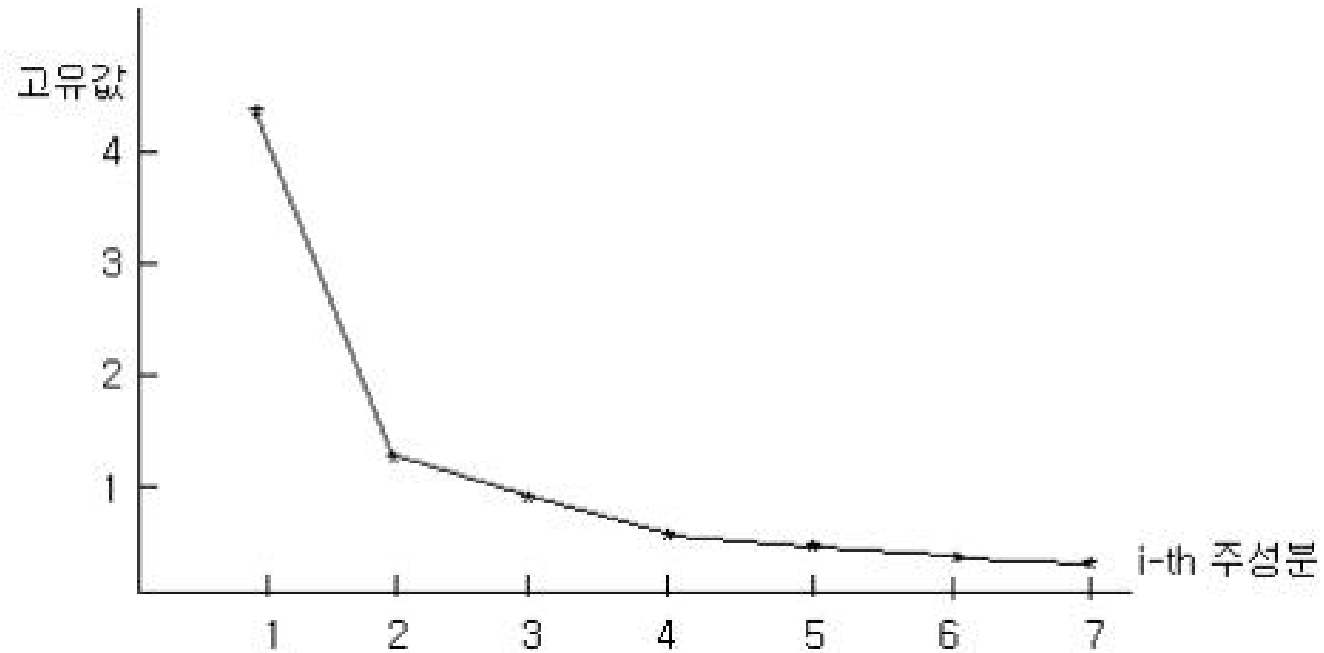
2. $\text{Var}(y_i) = \lambda_i, \quad i = 1, \dots, p$

$$\left(\begin{array}{l} \text{처음 } m \text{개의 주성분에 의해} \\ \text{설명되는 변동의 비율} \end{array} \right) = \frac{\lambda_1 + \dots + \lambda_m}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

참고 주성분의 기하학적 의미: $p=2$ 인 경우



스크리 산점도(Scree plot) :



<그림 10.2> 스크리 산점도(scree plot)

10.2.2 행렬도(Biplot)

자료행렬(data matrix)에서

행(row)들은 n 개 개체의 관측값을

열(column)들은 p 개의 변수를 나타낼 때,

행렬도(biplot) : 행과 열을 함께 저차원 공간에 나타낸 그림

n 개 관측값 : $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$$

표준화된 자료 : z_1, z_2, \dots, z_n

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i = 1, \dots, n, \quad j = 1, \dots, p$$

e_j ($j = 1, \dots, p$) : 제 j 주성분 방향으로의 단위벡터

z_i 의 e_1 으로의 사영(projection) :

$$z_i' e_1 = z_{i1} e_{11} + z_{i2} e_{12} + \dots + z_{ip} e_{1p}$$

(제1주성분에 대한 z_i 의 주성분점수)

따라서, z_i 의 주성분 점수 :

$$(z_i'e_1, z_i'e_2, \cdots, z_i'e_p)$$

z_1, z_2, \cdots, z_n 을 m -차원 주성분 공간에서

$$(z_i'e_1, z_i'e_2, \cdots, z_i'e_m), \quad i = 1, \cdots, n$$

으로 근사

\Rightarrow 이들의 그림 : 개체 플롯(observation plot)

변수 플롯(variable plot) : 변수들을 같은 주성분 공간에 나타냄

변수 $z = (z_1, z_2, \dots, z_p)'$ 에서

j 번째 변수인 z_j 를 $c_j = (0, \dots, 1, \dots, 0)'$ 으로 나타냄

c_j 를 주성분 공간에 사영한 결과 :

j 번째 변수 z_j 의 특성을 주성분 공간에 나타낸 것으로 이해

$$c_j' e_k = e_{jk}, \quad k = 1, \dots, p \quad (e_k \text{의 } j\text{번째 성분})$$

c_j 는 m -차원 주성분 공간 사영 결과

$$(c_j'e_1, c_j'e_2, \dots, c_j'e_m) = (e_{j1}, e_{j2}, \dots, e_{jm})$$

이것은 고유벡터 행렬

$$(e_1, e_2, \dots, e_m)$$

에서 j 번째 행(row)에 해당

행렬도(biplot) :

행(관측값)과 열(변수)을 함께 주성분 공간에 나타낸 그림

10.3. PRINCOMP 절차

(1) PRINCOMP 절차

```
PROC PRINCOMP <options>;  
  VAR variables;  
  PARTIAL variables;  
  WEIGHT variables;  
  BY variables;
```

① PROC PRINCOMP <options>;

㉞ dsn(TYPE=CORR) : 이미 생성된 'CORR' 형태 이용

㉟ dsn(TYPE=COV) : 이미 생성된 'COV' 형태 이용

<예 10.1> PROC CORR DATA=origin COV OUT=corrout;

PROC PRINCOMP DATA=corrout(TYPE=COV);

origin이란 자료의 공분산행렬(COV)을 사용하여 분석

COV : 공분산행렬을 이용하여 주성분을 구한다. 디폴트(default)는 상관행렬

N=n : 주성분의 개수를 지정한다. 디폴트는 변수의 개수

(2) 예제

[예제 10.3] 미국의 50개 주에서 7종류 범죄에 대해
10만 명당 범죄율을 조사한 자료
변수명은

살인(murder),	강간(rape),
약탈(robbery),	폭행(assault),
가택침입(burglary),	절도(larceny),
자동차 절도(auto)	

지역에 따른 범죄의 특징은?

<표 10.1> 미국 50개주의 범죄자료

주	살인	강간	약탈	폭행	가택침입	절도	자동차절도
Alabama	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7
Alaska	10.8	51.6	96.8	284.0	1331.7	3369.8	753.3
Arizona	9.5	34.2	138.2	312.3	2346.1	4467.4	439.5
Arkansas	8.8	27.6	83.2	203.4	972.6	1862.1	183.4
California	11.5	49.4	287.0	358.0	2139.4	3499.8	663.5
Colorado	6.3	42.0	170.7	292.9	1935.2	3903.2	477.1
Connecticut	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2
Delaware	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0
Florida	10.2	39.6	187.9	449.1	1859.9	3840.5	351.4
Georgia	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9
Hawaii	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4
Idaho	5.5	19.4	39.6	172.5	1050.8	2599.6	237.6
Illinois	9.9	21.8	211.3	209.0	1085.0	2828.5	528.6
Indiana	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4
Iowa	2.3	10.6	41.2	89.8	812.5	2685.1	219.9
Kansas	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3
Kentucky	10.1	19.1	81.1	123.3	872.2	1662.1	245.4
Louisiana	15.5	30.9	142.9	335.5	1165.5	2469.9	337.7
Maine	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9
Maryland	8.0	34.8	292.1	358.9	1400.0	3177.7	428.5
Massachusetts	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1
Michigan	9.3	38.9	261.9	274.6	1522.7	3159.0	545.5
Minnesota	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1
Mississippi	14.3	19.6	65.7	189.1	915.6	1239.9	144.4
Missouri	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4
Montana	5.4	16.7	39.2	156.8	804.9	2773.2	309.2
Nebraska	3.9	18.1	64.7	112.7	760.0	2316.1	249.1
Nevada	15.8	49.1	323.1	355.0	2453.1	4212.6	559.2
New Hampshire	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4
New Jersey	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5
New Mexico	8.8	39.1	109.6	343.4	1418.7	3008.6	259.5
New York	10.7	29.4	472.6	319.1	1728.0	2782.0	745.8
North Carolina	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1
North Dakota	0.9	9.0	13.3	43.8	446.1	1843.0	144.7
Ohio	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4
Oklahoma	8.6	29.2	73.8	205.0	1288.2	2228.1	326.8
Oregon	4.9	39.9	124.1	286.9	1636.4	3506.1	388.9
Pennsylvania	5.6	19.0	130.3	128.0	877.5	1624.1	333.2
Rhode Island	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4
South Carolina	11.9	33.0	105.9	485.3	1613.6	2342.4	245.1
South Dakota	2.0	13.5	17.9	155.7	570.5	1704.4	147.5
Tennessee	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0
Texas	13.3	33.8	152.4	208.2	1603.1	2988.7	397.6
Utah	3.5	20.3	68.8	147.3	1171.6	3004.6	334.5
Vermont	1.4	15.9	30.8	101.2	1348.2	2201.0	265.2
Virginia	9.0	23.3	92.1	165.7	986.2	2521.2	226.7
Washington	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3
West Virginia	6.0	13.2	42.2	90.9	597.4	1341.7	163.3
Wisconsin	2.8	12.9	52.2	63.7	846.9	2614.2	220.7
Wyoming	5.4	21.9	39.7	173.9	811.6	2772.2	282.0

```

/* PRINC01.SAS : PRINCIPAL COMPONENT ANALYSIS OF CRIME DATA */
  OPTIONS LS=80;
DATA CRIME;
  TITLE 'CRIME RATES PER 100,000 POPULATION BY STATE';
  INFILE 'C:\SASPGM\CRIME.DAT';    ①
  INPUT STATE $1-15 MURDER RAPE ROBBERY ASSAULT
          BURGLARY LARCENY AUTO;

RUN;
PROC PRINCOMP OUT=CRIMCOMP;        ②
RUN;
PROC SORT DATA=CRIMCOMP;
  BY PRIN1;
RUN;
PROC PRINT;
  ID STATE;                        ③
  VAR PRIN1 PRIN2 MURDER RAPE ROBBERY ASSAULT
          BURGLARY LARCENY AUTO;
  TITLE2 'STATES LISTED IN ORDER OF OVERALL CRIME RATE';

```

```

    TITLE3 'AS DETERMINED BY THE FIRST PRINCIPAL COMPONENT';
RUN;
PROC SORT;
    BY PRIN2;
PROC PRINT;
    ID STATE;                                ④
    VAR PRIN1 PRIN2 MURDER RAPE ROBBERY ASSAULT
        BURGLARY LARCENY AUTO;
    TITLE2 'STATES LISTED IN ORDER OF PROPERTY VS. VIOLENT CRIME';
    TITLE3 'AS DETERMINED BY THE SECOND PRINCIPAL COMPONENT';
RUN;
PROC PLOT;
    PLOT PRIN2*PRIN1=STATE / VPOS=23;        ⑤
    TITLE2 'PLOT OF THE FIRST TWO PRINCIPAL COMPONENTS';
RUN;
QUIT;

```

<프로그램 및 출력결과 설명>

① <표 10.1>의 자료가 CRIME에 기억

② 상관행렬을 이용하여 주성분을 구한다.

제1주성분 : 전반적인 범죄율

제2주성분 : 재산범죄가 강력범죄에 비하여 우세한 정도

The PRINCOMP Procedure
Simple Statistics

	MURDER	RAPE	ROBBERY	ASSAULT
Mean	7.444000000	25.73400000	124.0920000	211.3000000
Std	3.866768941	10.75962995	88.3485672	100.2530492

	BURGLARY	LARCENY	AUTO
Mean	1291.904000	2671.288000	377.5260000
Std	432.455711	725.908707	193.3944175

Correlation Matrix

	MURDER	RAPE	ROBBERY	ASSAULT	BURGLARY	LARCENY	AUTO
MURDER	1.0000	0.6012	0.4837	0.6486	0.3858	0.1019	0.0688
RAPE	0.6012	1.0000	0.5919	0.7403	0.7121	0.6140	0.3489
ROBBERY	0.4837	0.5919	1.0000	0.5571	0.6372	0.4467	0.5907
ASSAULT	0.6486	0.7403	0.5571	1.0000	0.6229	0.4044	0.2758
BURGLARY	0.3858	0.7121	0.6372	0.6229	1.0000	0.7921	0.5580
LARCENY	0.1019	0.6140	0.4467	0.4044	0.7921	1.0000	0.4442
AUTO	0.0688	0.3489	0.5907	0.2758	0.5580	0.4442	1.0000

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	4.11495951	2.87623768	0.5879	0.5879
2	1.23872183	0.51290521	0.1770	0.7648
3	0.72581663	0.40938458	0.1037	0.8685
4	0.31643205	0.05845759	0.0452	0.9137
5	0.25797446	0.03593499	0.0369	0.9506
6	0.22203947	0.09798342	0.0317	0.9823
7	0.12405606		0.0177	1.0000

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
MURDER	0.300279	-.629174	0.178245	-.232114	0.538123	0.259117	0.267593
RAPE	0.431759	-.169435	-.244198	0.062216	0.188471	-.773271	-.296485
ROBBERY	0.396875	0.042247	0.495861	-.557989	-.519977	-.114385	-.003903
ASSAULT	0.396652	-.343528	-.069510	0.629804	-.506651	0.172363	0.191745
BURGLARY	0.440157	0.203341	-.209895	-.057555	0.101033	0.535987	-.648117
LARCENY	0.357360	0.402319	-.539231	-.234890	0.030099	0.039406	0.601690
AUTO	0.295177	0.502421	0.568384	0.419238	0.369753	-.057298	0.147046

<그림 10.3> PRINCOMP 절차에 의한 범 죄 자료 분석 결과

③ 50개주를 첫 번째 주성분 점수의 크기순으로 배열한 결과

Nevada, California, Florida, New York : 높은 범죄 발생률

North Dakota나 South Dakota : 낮은 범죄 발생률

CRIME RATES PER 100,000 POPULATION BY STATE
STATES LISTED IN ORDER OF OVERALL CRIME RATE
AS DETERMINED BY THE FIRST PRINCIPAL COMPONENT

S T A T E	P r i n 1	P r i n 2	M U R D E R		R O B B E R Y	A S S A U L T	B U R G L A R Y	L A R C E N Y	A U T O
North Dakota	-3.96408	0.38767	0.9	9.0	13.3	43.8	446.1	1843.0	144.7
South Dakota	-3.17203	-0.25446	2.0	13.5	17.9	155.7	570.5	1704.4	147.5
West Virginia	-3.14772	-0.81425	6.0	13.2	42.2	90.9	597.4	1341.7	163.3
Iowa	-2.58156	0.82475	2.3	10.6	41.2	89.8	812.5	2685.1	219.9
Wisconsin	-2.50296	0.78083	2.8	12.9	52.2	63.7	846.9	2614.2	220.7
New Hampshire	-2.46562	0.82503	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4

Nebraska	-2.15071	0.22574	3.9	18.1	64.7	112.7	760.0	2316.1	249.1
Vermont	-2.06433	0.94497	1.4	15.9	30.8	101.2	1348.2	2201.0	265.2
Maine	-1.82631	0.57878	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9
Kentucky	-1.72691	-1.14663	10.1	19.1	81.1	123.3	872.2	1662.1	245.4
Pennsylvania	-1.72007	-0.19590	5.6	19.0	130.3	128.0	877.5	1624.1	333.2
Montana	-1.66801	0.27099	5.4	16.7	39.2	156.8	804.9	2773.2	309.2
Minnesota	-1.55434	1.05644	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1
Mississippi	-1.50736	-2.54671	14.3	19.6	65.7	189.1	915.6	1239.9	144.4
Idaho	-1.43245	-0.00801	5.5	19.4	39.6	172.5	1050.8	2599.6	237.6
Wyoming	-1.42463	0.06268	5.4	21.9	39.7	173.9	811.6	2772.2	282.0
Arkansas	-1.05441	-1.34544	8.8	27.6	83.2	203.4	972.6	1862.1	183.4
Utah	-1.04996	0.93656	3.5	20.3	68.8	147.3	1171.6	3004.6	334.5
Virginia	-0.91621	-0.69265	9.0	23.3	92.1	165.7	986.2	2521.2	226.7
North Carolina	-0.69925	-1.67027	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1
Kansas	-0.63407	-0.02804	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3
Connecticut	-0.54133	1.50123	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2
Indiana	-0.49990	0.00003	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4
Oklahoma	-0.32136	-0.62429	8.6	29.2	73.8	205.0	1288.2	2228.1	326.8
Rhode Island	-0.20156	2.14658	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4
Tennessee	-0.13660	-1.13498	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0
Alabama	-0.04988	-2.09610	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7
New Jersey	0.21787	0.96421	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5
Ohio	0.23953	0.09053	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4
Georgia	0.49041	-1.38079	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9
Illinois	0.51290	0.09423	9.9	21.8	211.3	209.0	1085.0	2828.5	528.6
Missouri	0.55637	-0.55851	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4
Hawaii	0.82313	1.82392	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4
Washington	0.93058	0.73776	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3
Delaware	0.96458	1.29674	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0
Massachusetts	0.97844	2.63105	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1

Louisiana	1.	12020	-2.	08327	15.	5	30.	9	142.	9	335.	5	1165.	5	2469.	9	337.	7
New Mexico	1.	21417	-0.	95076	8.	8	39.	1	109.	6	343.	4	1418.	7	3008.	6	259.	5
Texas	1.	39696	-0.	68131	13.	3	33.	8	152.	4	208.	2	1603.	1	2988.	7	397.	6
Oregon	1.	44900	0.	58603	4.	9	39.	9	124.	1	286.	9	1636.	4	3506.	1	388.	9
South Carolina	1.	60336	-2.	16211	11.	9	33.	0	105.	9	485.	3	1613.	6	2342.	4	245.	1
Marvland	2.	18280	-0.	19474	8.	0	34.	8	292.	1	358.	9	1400.	0	3177.	7	428.	5
Michigan	2.	27333	0.	15487	9.	3	38.	9	261.	9	274.	6	1522.	7	3159.	0	545.	5
Alaska	2.	42151	0.	16652	10.	8	51.	6	96.	8	284.	0	1331.	7	3369.	8	753.	3
Colorado	2.	50929	0.	91660	6.	3	42.	0	170.	7	292.	9	1935.	2	3903.	2	477.	1
Arizona	3.	01414	0.	84495	9.	5	34.	2	138.	2	312.	3	2346.	1	4467.	4	439.	5
Florida	3.	11175	-0.	60392	10.	2	39.	6	187.	9	449.	1	1859.	9	3840.	5	351.	4
New York	3.	45248	0.	43289	10.	7	29.	4	472.	6	319.	1	1728.	0	2782.	0	745.	8
California	4.	28380	0.	14319	11.	5	49.	4	287.	0	358.	0	2139.	4	3499.	8	663.	5
Nevada	5.	26699	-0.	25262	15.	8	49.	1	323.	1	355.	0	2453.	1	4212.	6	559.	2

<그림 10.4> 첫 번째 주성분 점수의 크기순으로 배열된 자료

④ 두 번째 주성분점수의 크기순 배열

Massachusetts, Rhode Island, Connecticut :

재산과 관련된 범죄가 많이 발생

Mississippi, Alabama : 강력범죄가 많이 발생

CRIME RATES PER 100,000 POPULATION BY STATE
STATES LISTED IN ORDER OF PROPERTY VS. VIOLENT CRIME
AS DETERMINED BY THE SECOND PRINCIPAL COMPONENT

S T A T E	P r i n 1	P r i n 2	M U R D E R	R A P E	R O B B E R Y	A S S A U L T	B U R G L A R Y	L A R C E N Y	A U T O
Mississippi	-1.50736	-2.54671	14.3	19.6	65.7	189.1	915.6	1239.9	144.4
South Carolina	1.60336	-2.16211	11.9	33.0	105.9	485.3	1613.6	2342.4	245.1
Alabama	-0.04988	-2.09610	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7
Louisiana	1.12020	-2.08327	15.5	30.9	142.9	335.5	1165.5	2469.9	337.7
North Carolina	-0.69925	-1.67027	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1
Georgia	0.49041	-1.38079	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9
Arkansas	-1.05441	-1.34544	8.8	27.6	83.2	203.4	972.6	1862.1	183.4
Kentucky	-1.72691	-1.14663	10.1	19.1	81.1	123.3	872.2	1662.1	245.4
Tennessee	-0.13660	-1.13498	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0
New Mexico	1.21417	-0.95076	8.8	39.1	109.6	343.4	1418.7	3008.6	259.5
West Virginia	-3.14772	-0.81425	6.0	13.2	42.2	90.9	597.4	1341.7	163.3
Virginia	-0.91621	-0.69265	9.0	23.3	92.1	165.7	986.2	2521.2	226.7
Texas	1.39696	-0.68131	13.3	33.8	152.4	208.2	1603.1	2988.7	397.6
Oklahoma	-0.32136	-0.62429	8.6	29.2	73.8	205.0	1288.2	2228.1	326.8
Florida	3.11175	-0.60392	10.2	39.6	187.9	449.1	1859.9	3840.5	351.4

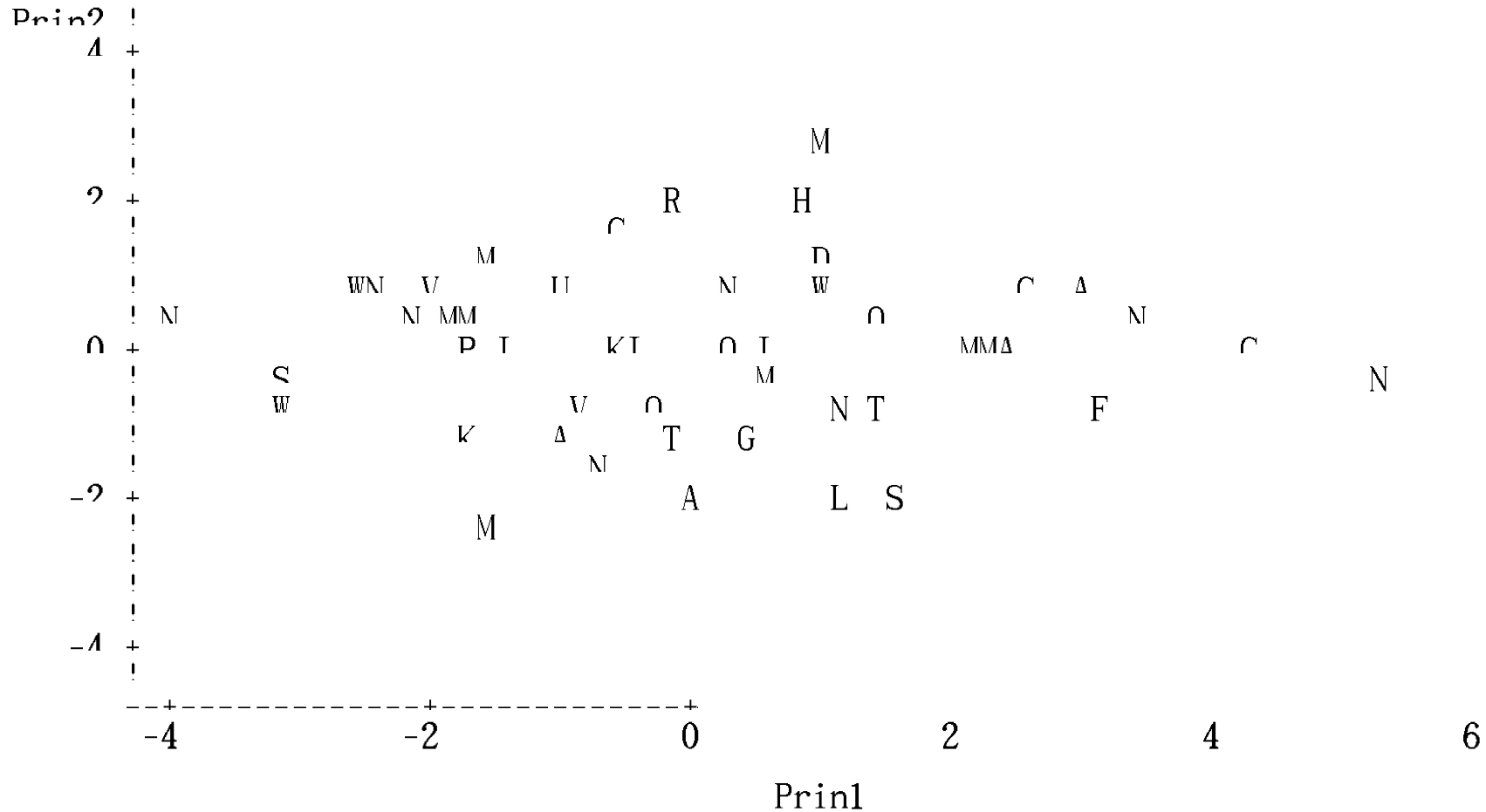
Missouri	0.55637	-0.55851	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4
South Dakota	-3.17203	-0.25446	2.0	13.5	17.9	155.7	570.5	1704.4	147.5
Nevada	5.26699	-0.25262	15.8	49.1	323.1	355.0	2453.1	4212.6	559.2
Pennsylvania	-1.72007	-0.19590	5.6	19.0	130.3	128.0	877.5	1624.1	333.2
Maryland	2.18280	-0.19474	8.0	34.8	292.1	358.9	1400.0	3177.7	428.5
Kansas	-0.63407	-0.02804	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3
Idaho	-1.43245	-0.00801	5.5	19.4	39.6	172.5	1050.8	2599.6	237.6
Indiana	-0.49990	0.00003	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4
Wyoming	-1.42463	0.06268	5.4	21.9	39.7	173.9	811.6	2772.2	282.0
Ohio	0.23953	0.09053	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4
Illinois	0.51290	0.09423	9.9	21.8	211.3	209.0	1085.0	2828.5	528.6
California	4.28380	0.14319	11.5	49.4	287.0	358.0	2139.4	3499.8	663.5
Michigan	2.27333	0.15487	9.3	38.9	261.9	274.6	1522.7	3159.0	545.5
Alaska	2.42151	0.16652	10.8	51.6	96.8	284.0	1331.7	3369.8	753.3
Nebraska	-2.15071	0.22574	3.9	18.1	64.7	112.7	760.0	2316.1	249.1
Montana	-1.66801	0.27099	5.4	16.7	39.2	156.8	804.9	2773.2	309.2
North Dakota	-3.96408	0.38767	0.9	9.0	13.3	43.8	446.1	1843.0	144.7
New York	3.45248	0.43289	10.7	29.4	472.6	319.1	1728.0	2782.0	745.8
Maine	-1.82631	0.57878	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9
Oregon	1.44900	0.58603	4.9	39.9	124.1	286.9	1636.4	3506.1	388.9
Washington	0.93058	0.73776	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3
Wisconsin	-2.50296	0.78083	2.8	12.9	52.2	63.7	846.9	2614.2	220.7
Iowa	-2.58156	0.82475	2.3	10.6	41.2	89.8	812.5	2685.1	219.9
New Hampshire	-2.46562	0.82503	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4
Arizona	3.01414	0.84495	9.5	34.2	138.2	312.3	2346.1	4467.4	439.5
Colorado	2.50929	0.91660	6.3	42.0	170.7	292.9	1935.2	3903.2	477.1
Utah	-1.04996	0.93656	3.5	20.3	68.8	147.3	1171.6	3004.6	334.5
Vermont	-2.06433	0.94497	1.4	15.9	30.8	101.2	1348.2	2201.0	265.2
New Jersey	0.21787	0.96421	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5
Minnesota	-1.55434	1.05644	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1

Delaware	0.96458	1.29674	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0
Connecticut	-0.54133	1.50123	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2
Hawaii	0.82313	1.82392	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4
Rhode Island	-0.20156	2.14658	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4
Massachusetts	0.97844	2.63105	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1

<그림 10.5> 두 번째 주성분 점수의 크기순으로 배열된 자료

⑤ 처음 두 개의 주성분의 산점도

Plot of Prin2*Prin1. Symbol is value of STATE.



<그림 10.6> 첫 번째 주성분과 두 번째 주성분의 산점도

(3) 메뉴의 이용

분석창에서 **통계량** → **다변량** → **주성분 분석**을 선택

[예제 10.4] <표 10.2>의 자료는 우리 나라의 주요 도시에서 월별 평균기온을 나타내는 자료로서, 2000년 1월부터 12월까지 12개월간 25개 도시에서 관측된 자료이다. 이 자료를 통하여 도시별 기온의 특성을 결정하는 형태를 찾아보자.(KOSIS 자료)

<표 10.2> 도시별 월 평균 기온

CITY	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Seoul	-2.1	-1.7	6.3	11.9	17.5	23.7	26.8	26.2	20.7	14.9	7.0	0.9
Busan	4.2	3.2	9.1	13.2	17.3	20.4	25.3	26.7	21.9	18.2	12.1	7.3
Incheon	-1.5	-1.2	6.0	11.5	16.5	22.3	26.1	26.5	21.3	15.6	7.4	1.5
Suwon	-2.0	-2.1	5.6	11.6	17.5	23.3	26.9	25.9	20.1	14.1	6.1	0.3
Sokcho	-0.4	-4.4	6.9	11.2	14.3	19.2	24.8	24.0	18.6	14.5	7.8	3.2
Chunchon	-3.0	-3.2	4.9	10.8	16.7	22.7	26.0	24.6	18.5	12.1	4.1	-1.5
Kangnung	0.8	0.9	8.1	12.6	17.0	21.7	27.2	25.4	19.7	15.5	8.6	4.1
Chongju	-1.2	-1.6	5.7	11.5	17.2	22.7	26.2	26.1	19.7	14.4	6.4	0.9
Chupungnyon	-2.8	-1.6	5.5	11.3	17.0	21.5	24.8	24.2	18.2	13.0	5.4	0.0
Daejeon	-1.2	-1.1	6.1	11.9	17.2	22.2	25.6	25.8	19.9	14.5	6.6	1.4
Seosan	-1.0	-2.4	3.9	9.7	15.1	20.8	25.0	25.2	19.2	13.7	5.8	0.6
Jeonju	0.8	0.2	6.9	12.8	18.2	22.9	26.9	26.9	20.8	15.6	8.0	2.7
Gunsan	0.3	0.0	5.9	11.2	16.3	21.5	25.3	25.7	20.2	15.3	8.1	2.8
Gwangju	1.0	0.6	6.9	12.6	17.8	22.5	26.5	26.3	20.4	15.9	8.4	3.2
Mokpo	1.9	1.1	6.7	11.9	16.8	21.7	25.8	26.6	20.7	16.4	9.0	4.3
Yosu	3.1	2.5	8.3	12.8	17.5	20.9	24.9	26.2	21.3	17.3	10.8	6.0
Daegu	1.1	1.3	7.9	13.0	19.5	23.3	26.9	26.5	21.2	16.7	9.4	3.9
Pohang	2.7	2.3	8.9	13.9	18.7	21.9	26.5	26.1	20.9	16.8	10.1	5.0
Ullungdo	1.9	0.8	6.6	11.9	17.2	19.6	23.9	23.5	18.7	15.1	9.2	4.7
Ulsan	3.1	2.7	8.7	13.6	18.1	21.4	26.0	26.0	20.6	16.5	10.0	4.9
Tongyoung	3.9	3.0	8.6	13.1	17.2	20.9	25.2	26.6	21.5	17.2	10.6	5.9
Jinju	1.1	1.2	7.0	12.7	18.4	22.3	26.5	26.4	20.5	15.2	7.5	1.6
Jeju	6.0	4.6	9.4	13.5	17.2	21.6	26.4	28.0	22.2	18.2	12.4	8.4
Seogwipo	8.1	6.4	11.0	14.8	18.6	22.4	27.2	28.3	23.8	20.3	14.7	10.6
Uljin	1.9	1.4	7.7	12.2	16.0	19.0	23.9	23.4	18.4	14.6	8.6	4.2

The PRINCOMP Procedure

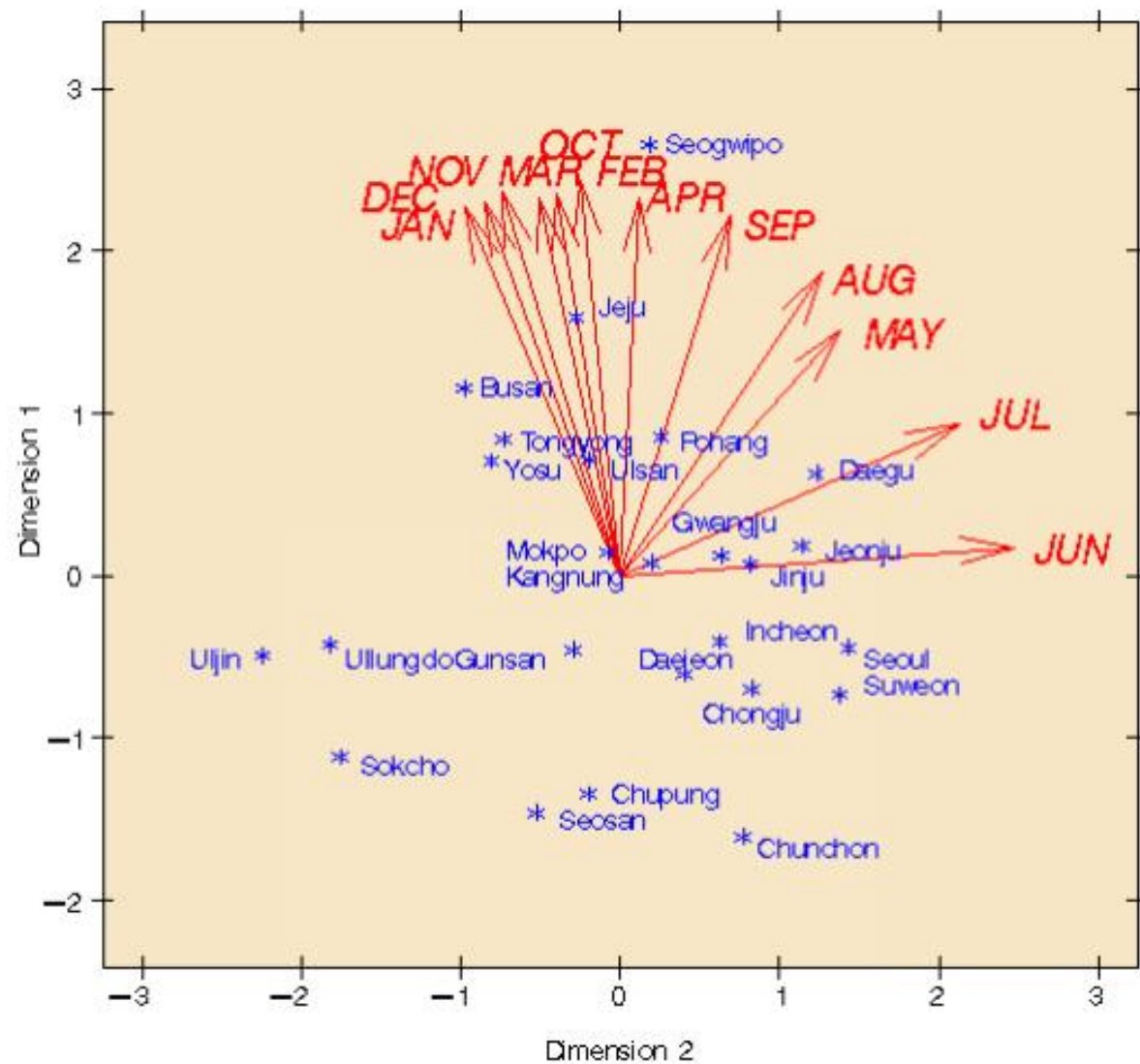
Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	7.98817633	5.21186255	0.6657	0.6657
2	2.77631377	2.09087817	0.2314	0.8970
3	0.68543560	0.40071721	0.0571	0.9542
4	0.28471839	0.17511880	0.0237	0.9779
5	0.10959959	0.04184350	0.0091	0.9870
6	0.06775610	0.03136260	0.0056	0.9927
7	0.03639350	0.01115707	0.0030	0.9957
8	0.02523643	0.01057010	0.0021	0.9978
9	0.01466633	0.00712181	0.0012	0.9990
10	0.00754452	0.00464590	0.0006	0.9997
11	0.00289862	0.00163779	0.0002	0.9999
12	0.00126082		0.0001	1.0000

<그림 10.7> 고유값

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
JAN	0.326022	-.203575	-.036764	-.027091	0.455278	0.057542
FEB	0.332024	-.096383	0.155419	-.301333	0.583925	-.159593
MAR	0.329570	-.122337	0.191627	0.407721	-.235098	-.296897
APR	0.329329	0.027622	0.371424	0.227373	-.191349	-.456574
MAY	0.213072	0.331984	0.633402	-.401271	-.130077	0.232737
JUN	0.024698	0.592189	0.000003	-.098256	0.004482	0.162447
JUL	0.131832	0.510337	-.059120	0.642263	0.313060	0.225335
AUG	0.265469	0.305379	-.458890	-.202309	0.110358	-.430482
SEP	0.314830	0.165455	-.372568	-.221192	-.359774	-.077637
OCT	0.344333	-.058852	-.185490	-.060897	-.283433	0.276671
NOV	0.334570	-.180098	-.072862	0.033011	-.144671	0.351117
DEC	0.320799	-.235930	-.077602	0.112700	0.045178	0.390010

<그림 10.8> 고유벡터



<그림 10.9> 두 주성분에 의한 행렬도(biplot)

세로축은 제1주성분 ; 가로축은 제2주성분

변수 플롯 : 세로축 11, 12, 1월을 비롯한 추운 달 (겨울 온도)

가로축 6, 7월의 더운 달 (여름 온도)

개체 플롯 :

제1축

서귀포, 제주 등 : 겨울 온도가 높은 도시

춘천, 서산, 추풍령 등 : 겨울 온도가 낮은 도시

제2축

대구, 전주, 진주, 서울, 수원 등 : 여름 온도가 높은 도시

울릉도, 속초 등 : 여름 온도가 낮은 도시