

Basic Statistics & Visualizations of Data with



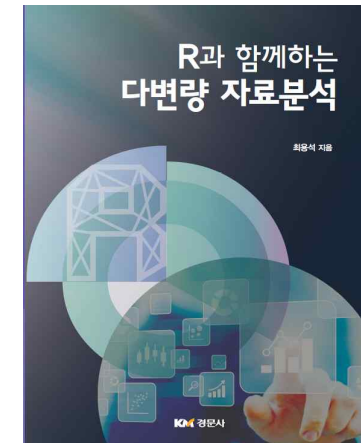
주제 2: 자료의 시각화



최 용 석 (부산대학교 통계학과)

yschoi.pusan.ac.kr

2018.01.09.~10.



부산대학교 간호과학연구소 워크숍

주제 2: 자료의 시각화

Prolog : Unstructured Data vs. Structured Data

2.0 다변량 자료 Multivariate Data

- 범주형자료의 시각화 (Visualizations for CDA) :

2.1 단순 대응분석 그림

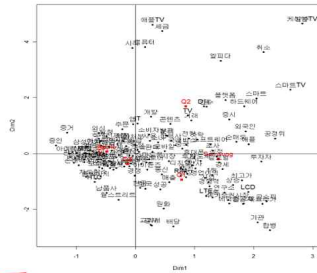
2.2 다중 대응분석 그림

- 다변량자료의 시각화 (Visualizations for MDA) :

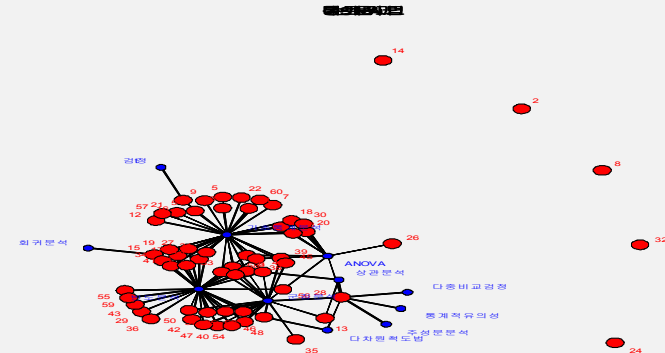
2.3 행렬도

Prolog : Unstructured Data(Text Data) $\Rightarrow$ Structured Data(EDA, CDA, MDA)

통계적 기법 적용을 위한 비정형 빅 데이터의 정형화와 분석 및 시각화



MapReduce



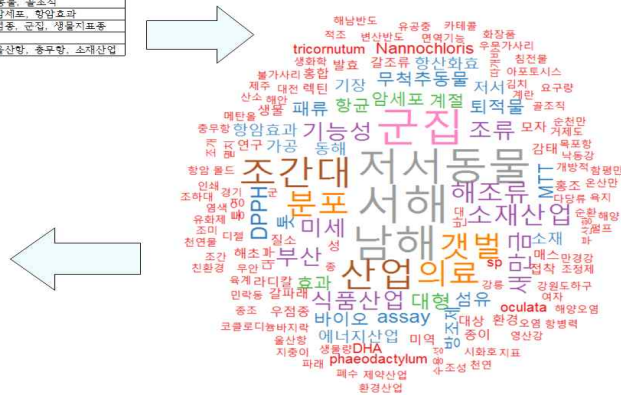
워드 클라우드(Word Cloud)

a. row data

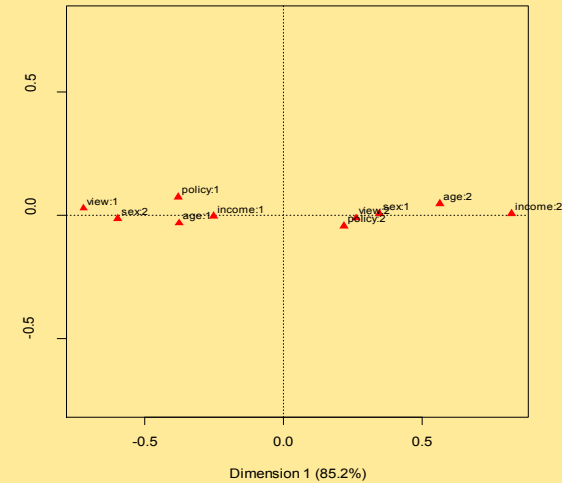
번호	년도	키워드
논문1	2015	남해, 통영, 화성물산업, 굴, 라디칼, DPPH
논문2	2015	의료산업, 해양유적추출물, 굴조각
논문3	2014	미세조류, 의료산업, 암세포, 항암효과
논문4	2014	포도산대, 해초류, 오징어, 간접, 생물기표종
...	...	...
논문60	1989	생물유화제, 부산항, 울산항, 충무항, 소재산업

b. 빈도표

키워드	빈도
1. 서해	18
2. 저서동물	17
3. 남해	17
4. 군집	16
5. 포산대	13
6. 산업	13
7. 의료	10
8. 분포	10
9. 갯벌	10
10. 해초류	8



MDS 분석결과도:

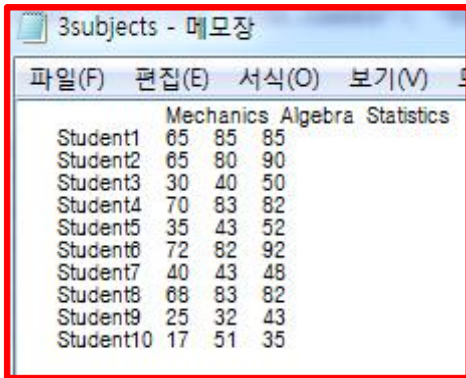


2.0 다변량자료 Multivariate Data

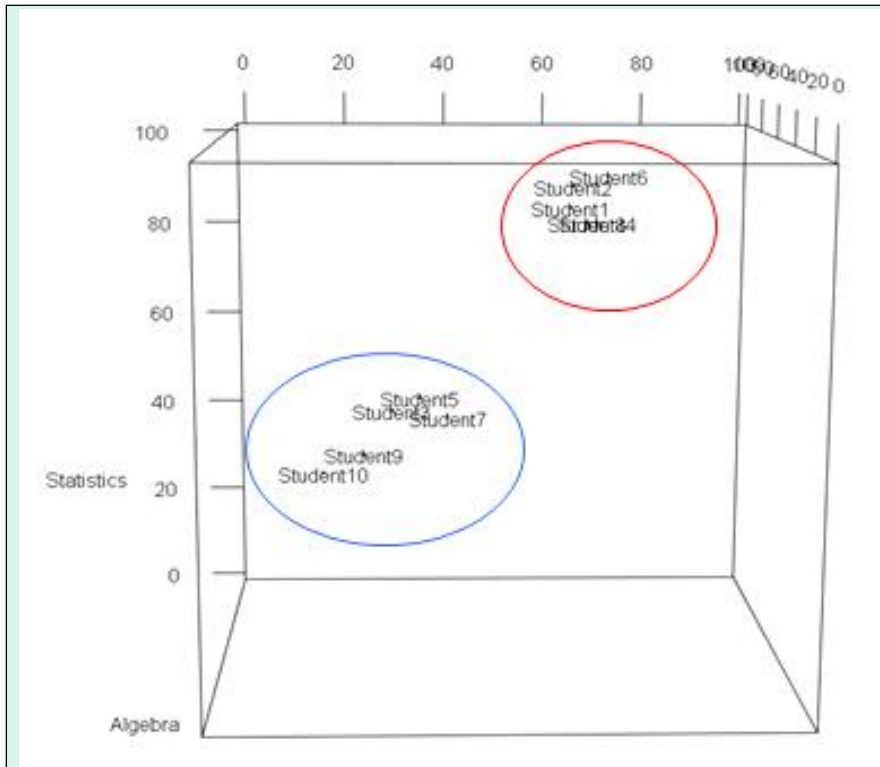
- 자연과학, 사회과학, 의학, 경제학, 경영학, 교육학 등 여러 분야에서 관심이 되는 현상을 설명하기 위해 하나의 측정변수에 의해서 결과를 관찰할 수도 있지만 대부분 많은 변수로 다양한 관점에서 자료를 수집한다.
- 이러한 자료는 일반적으로 n 개의 개체(object)에 대해 p 개의 변수(variable)로 측정하며 이를 다변량 자료(multivariate data)라 한다.

[자료 2.0.1] 10명의 학생들의 3과목 시험성적 : 3차원 공간의 기하적 표현

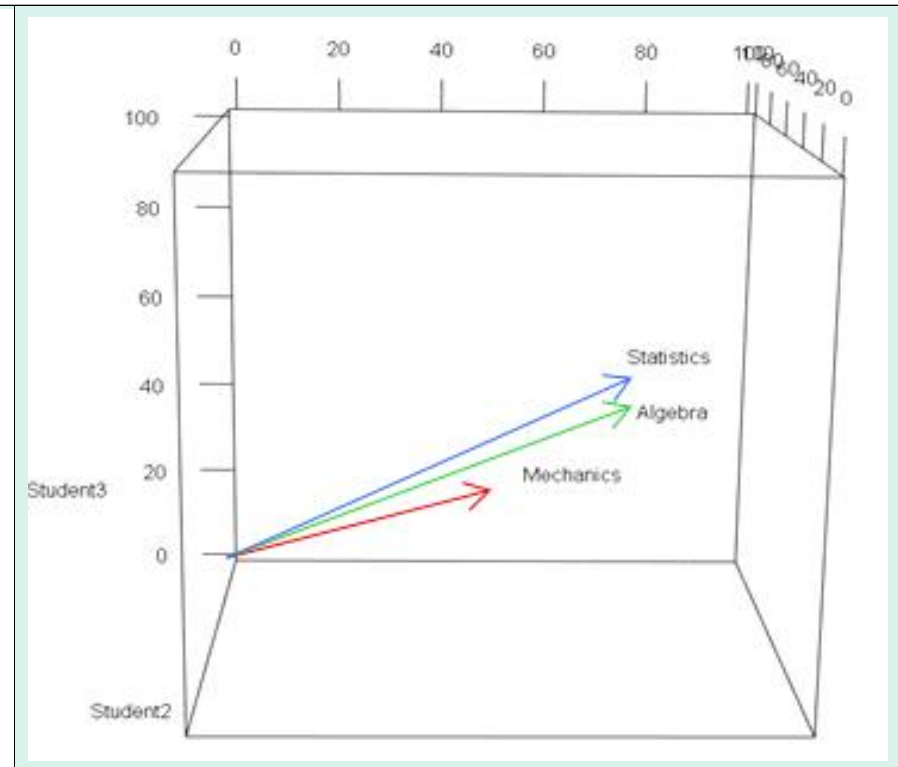
3subjects.txt				3subjects-3d.R
Students	Mechanics	Algebra	Statistics	<pre> setwd("C:/Users/stat/Desktop/개인_WORK/외부초청강의/부산대간호 과학연구소/R-codes") Data1.1.1<-read.table("3subjects.txt", header=T) X<-Data1.1.1 library(rgl) # Observations in Variables Space lim<-c(0, 100) plot3d(X[,1], X[,2], X[,3],xlim=lim, ylim=lim, zlim=lim, xlab="Mechanics", ylab="Algebra", zlab="Statistics") text3d(X[,1], X[,2], X[,3],rownames(X)) # Variables in Observations Space plot3d(X[1,], X[2,], X[3,], xlim=lim, ylim=lim, zlim=lim, xlab="Student1", ylab="Student2", zlab="Student3") text3d(X[1,], X[2,], X[3,], colnames(X)) </pre>
Student1	65	85	85	
Student2	65	80	90	
Student3	30	40	50	
Student4	70	83	82	
Student5	35	43	52	
Student6	72	82	92	
Student7	40	43	48	
Student8	68	83	82	
Student9	25	32	43	
Student10	17	51	35	



Students	Mechanics	Algebra	Statistics
Student1	65	85	85
Student2	65	80	90
Student3	30	40	50
Student4	70	83	82
Student5	35	43	52
Student6	72	82	92
Student7	40	43	48
Student8	68	83	82
Student9	25	32	43
Student10	17	51	35



변수 공간의 개체 : $Q^p-Space$



개체 공간의 변수 : $R^n-Space$

범주형자료의 시각화 (Visualizations for CDA)

2.1 단순 대응분석 그림 Simple Correspondence Analysis Map

Correspondence Analysis

❖ Definition

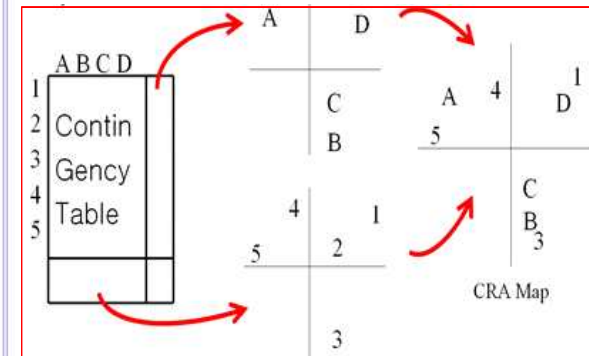
- ✓ Technique for displaying the associations among the rows and columns of contingency table in a type of scatterplot or map

❖ In general, chi-square statistic providing a testing for the independence bt rows and columns

→ In addition, CRA accounts more clearly for the relationship bt rows and columns

❖ Applications

- ✓ Educational Psychology, Medicine, Survey, Social Science, etc.
- ✓ Categorical Data Analysis



❖ Characters represent the coordinate points of rows and columns categories.

❖ Distance bt two rows (or columns) is Euclidean distance and their closeness.

❖ Distance bt row and column can't be defined and just their the direction can be interpreted.

❖ If there are some rows in the direction of a column, they are matched or corresponded.

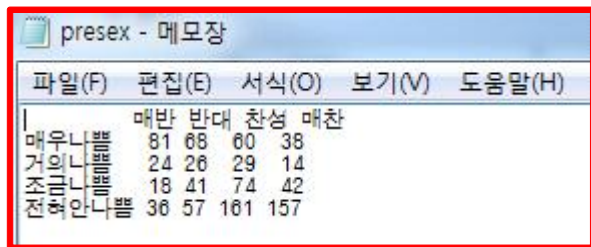
References: 최용석(2001). SAS 대응분석의 이해와 응용, 자유아카데미, 서울.

1) 이원분할표 Tow-Way Contingency Table

presex.txt

[자료 2.1.1] 결혼 전 성관계에 대한 반응과 피임허용
이원분할표

혼전 성관계	10대 임신조절			
	매우 반대	반	대	찬성
매우 나쁘다	81	68	60	38
거의 나쁘다	24	26	29	14
조금 나쁘다	18	41	74	42
전혀 나쁘지 않다	36	57	161	157



	매반	반대	찬성	매찬
매우나쁨	81	68	60	38
거의나쁨	24	26	29	14
조금나쁨	18	41	74	42
전혀안나쁨	36	57	161	157

presex-SCRA.R

```
# Simple CRA ca() : Matrix for Two-Way Table
```

```
O<-matrix(c(81, 68, 60, 38,
             24, 26, 29, 14,
             18, 41, 74, 42,
             36, 57, 161, 157), byrow=T, nrow=4)
```

```
rownames(O)<-c("매우나쁨", "거의나쁨", "조금나쁨", "전혀안나쁨")
```

```
colnames(O)<-c("매반", "반대", "찬성", "매찬")
```

```
library(ca)
```

```
sca=ca(O)
```

```
par(pty="s")
```

```
plot(sca, main="SCRA package ca : 이원분할표")
```

```
# Simple CRA ca() : Text for Two-Way Table
```

```
setwd("C:/Users/stat/Desktop/개인_WORK/외부초청강의/부산대간  
호과학연구소/R-codes")
```

```
O<-read.table("presex.txt", header=T)
```

```
library(ca)
```

```
sca=ca(O)
```

```
sca
```

```
win.graph()
```

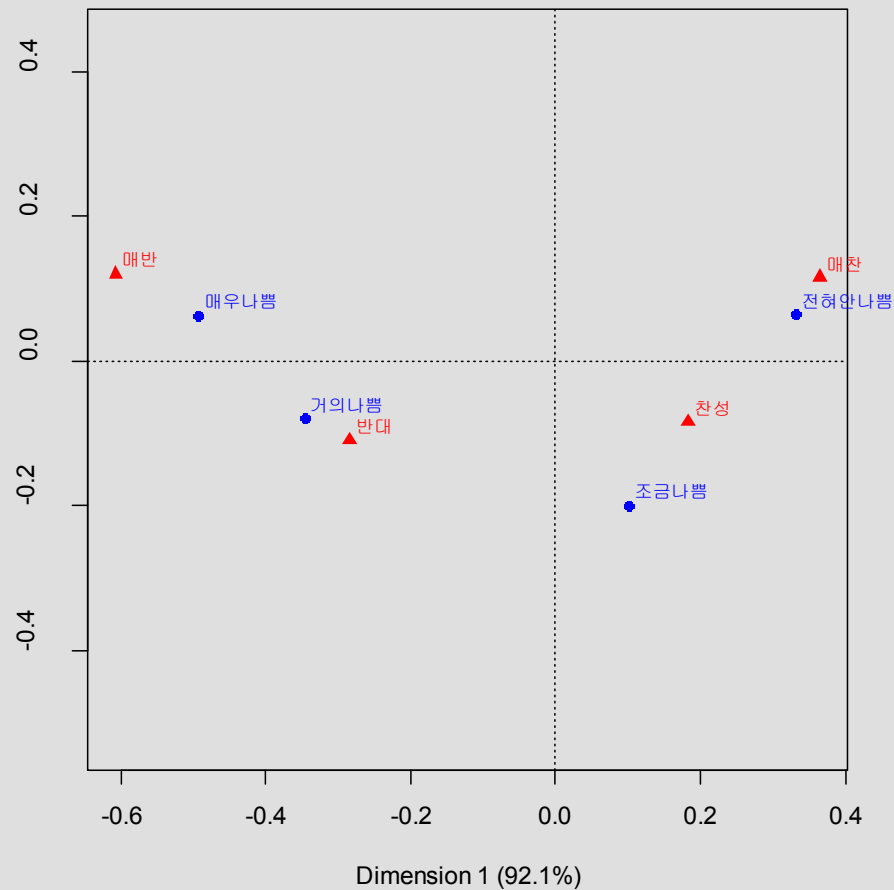
```
par(pty="s")
```

```
plot(sca, main="SCRA package ca : 이원분할표")
```

presex - 메모장

파일(F)	편집(E)	서식(O)	보기(V)	도움말(H)
매우나쁨	81	68	60	38
거의나쁨	24	26	29	14
조금나쁨	18	41	74	42
전혀안나쁨	36	57	161	157

SCRA package ca :



- 수평축(Dim1)에 대하여 왼편의 행범주 (매우 나쁨, 거의나쁨)은 열범주 (매반, 반대)방향으로 놓여져 있다. 이는 혼전 성관계가 나쁘다고 여기는 사람들은 10대에 피임을 반대하는 입장에 있음을 나타낸다.

- 오른편의 행범주 (조금나쁨, 전혀안나쁨)은 열범주 (찬성, 매찬) 방향으로 놓여져 있어 이를 서로 대응한다고 한다. 이는 혼전 성관계에 대해 부정적이지 않은 사람들은 10대에 피임을 찬성하고 있다.

- 혼전 성경험에 대해서도 관대한 사람들은 10대에게 피임을 허용해야 한다고 좀 더 강하게 느끼는 입장을 보이고 있다.

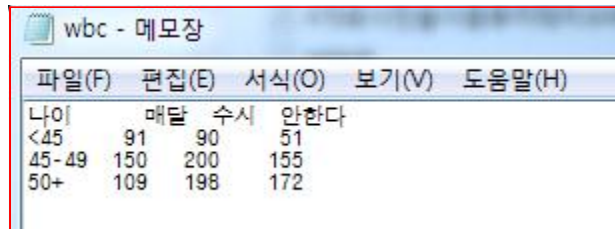
단순 대응분석 그림 : library(ca), ca()

해석

wbc.txt

여성들의 유방자기 진단 빈도와 나이의 이원분할표
(최용석, 정광모, 2009).

나 이	유방자기진단 빈도		
	매 달	수시로	안한다
< 45	91	90	51
45 - 49	150	200	155
50 +	109	198	172

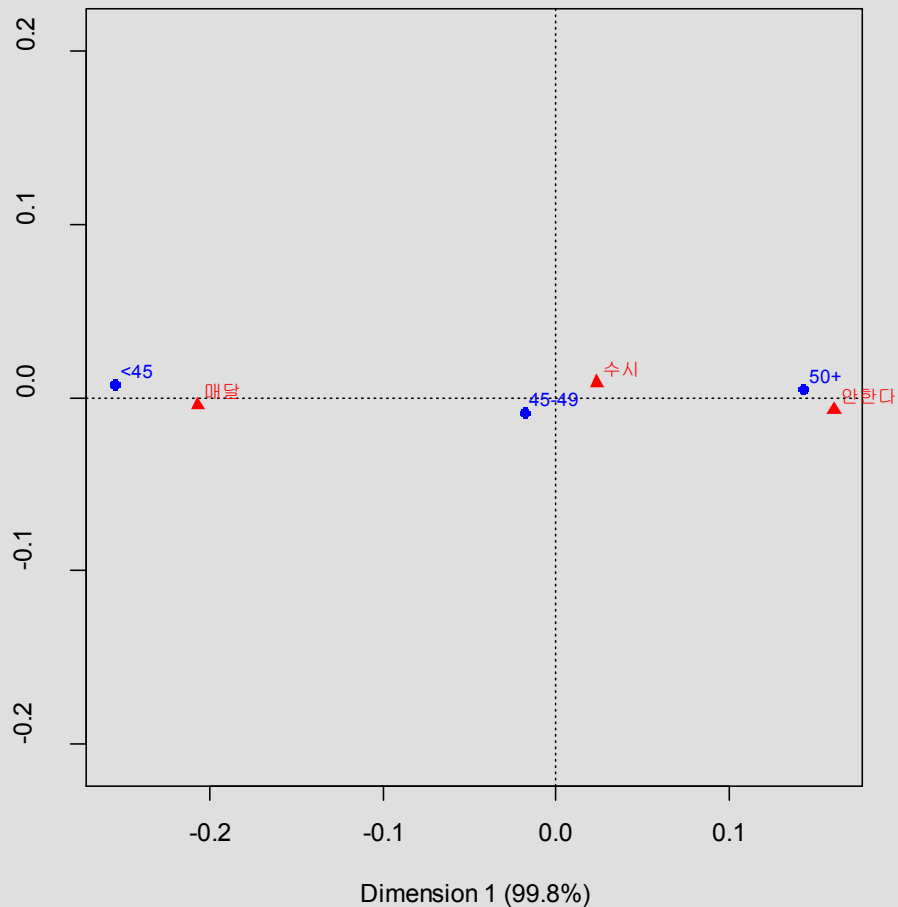


파일(F)	편집(E)	서식(O)	보기(V)	도움말(H)
나이	매달	수시	안한다	
<45	91	90	51	
45-49	150	200	155	
50+	109	198	172	

wbc-SCA.R

```
setwd("C:/Users/stat/Desktop/개인_WORK/
외부초청강의/부산대간호과학연구소
/R-codes")
wbc<-read.table("wbc.txt", header=T)
rownames(wbc)<-wbc[,1]
colnames(wbc)
wbc<-wbc[,-1]
library("ca")
ca(wbc)
par(pty="s")
plot(ca(wbc), main="Simple CA for Women
Breast Cancer")
```

Simple CA for Women Breast Cancer



- 수평축(Dim1)에 대하여 왼편의 행범주 나이 <45는 열범주 매달과 같은 방향으로 놓여져 있다.

- 오른편의 행범주 나이 50+는 열범주 안한다 방향으로 놓여져 있어 이를 서로 대응한다고 한다.

- 나이 45-49는 다소 열범주 매달 방향에 있으나 중앙에 있어 특별히 대응관계를 보여 주고 있지 않다.

단순 대응분석 그림 : library(ca), ca()

해석

Algorithm for Simple CA

[Step 1] **Data Matrix** : $O = (o_{ij})$, $o_{ij} \geq 0$, $i = 1, \dots, n; j = 1, \dots, p (n > p)$

[Step 2] **Correspondence Matrix** : $F = (f_{ij}) = (o_{ij}/N)$

$$\text{with grand total } N = \sum_{i=1}^n \sum_{j=1}^p o_{ij}.$$

[Step 3] **Centred Correspondence Matrix** : $F - \mathbf{r}\mathbf{c}^t$

where the column and row mass vectors $\mathbf{r} = F\mathbf{1}_p$ and $\mathbf{c} = F^t\mathbf{1}_n$, respectively.

$$\text{And } r_i = \sum_{j=1}^p f_{ij}, \quad i = 1, \dots, n \quad \text{and} \quad c_j = \sum_{i=1}^n f_{ij}, \quad j = 1, \dots, p.$$

[Step 4] **Residual Matrix** : $Y = D_r^{-1/2}(F - \mathbf{r}\mathbf{c}^t)D_c^{-1/2} = ((f_{ij} - r_i c_j) / \sqrt{r_i c_j})$

$$\text{where } D_r^{1/2} = \text{diag}(\sqrt{r_1}, \dots, \sqrt{r_n}) \quad \text{and} \quad D_c^{1/2} = \text{diag}(\sqrt{c_1}, \dots, \sqrt{c_p}).$$

[Step 5] **Singular Value Decomposition** : $Y = U D_\lambda V^t$

[Step 6] **Coordinate Matrices for SCA** : $C_R = D_r^{-1/2} U D_\lambda$, $C_C = D_c^{-1/2} V D_\lambda$.

[Step 7] **Goodness-of-fit** for s-dimensional SCA Map(Biplot) :

$$fit = \frac{(\lambda_1^2 + \lambda_2^2 + \dots + \lambda_s^2)}{\sum_{k=1}^{p-1} \lambda_k^2} \times 100\%.$$

Geometric Concept of Profiles

- row profiles :

$$\mathbf{r}_i = (f_{i1}/f_{i+}, \dots, f_{ip}/f_{i+})', \quad i = 1, \dots, n \in \mathcal{E}^p \text{ weighted by } D_c^{-1} = \text{diag}(1/c_1, \dots, 1/c_p).$$

- $d_p^2(\mathbf{r}_i, \mathbf{r}_k) = \|\mathbf{r}_i - \mathbf{r}_k\|_{D_c^{-1}}^2 = (\mathbf{r}_i - \mathbf{r}_k)' D_c^{-1} (\mathbf{r}_i - \mathbf{r}_k)$: X^2 -distance bt two row profiles.

$$= \sum_j (f_{ij}/f_{i+} - f_{kj}/f_{k+})^2 / f_{+j}$$

- column profiles :

$$\mathbf{c}_j = (f_{1j}/f_{+j}, \dots, f_{nj}/f_{+j})', \quad j = 1, \dots, p \in \mathcal{E}^n \text{ weighted by } D_r^{-1} = \text{diag}(1/r_1, \dots, 1/r_n).$$

- $d_n^2(\mathbf{c}_j, \mathbf{c}_l) = \|\mathbf{c}_j - \mathbf{c}_l\|_{D_r^{-1}}^2 = (\mathbf{c}_j - \mathbf{c}_l)' D_r^{-1} (\mathbf{c}_j - \mathbf{c}_l)$: X^2 -distance bt two column profiles.

$$= \sum_i (f_{ij}/f_{+j} - f_{il}/f_{+l})^2 / f_{i+}$$

Algebraic Relationship :

X^2 -Statistics and Total inertia

- $X^2 = \sum_{i=1}^n \sum_{j=1}^p \left(\frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}}} \right)^2$ where $e_{ij} = \frac{o_{i+} o_{+j}}{N} = N r_i c_j$ under $H_0 : \pi_{ij} = \pi_{i+} \pi_{+j}$.

$$- X^2 = N \sum_{i=1}^n f_{i+} (\mathbf{r}_i - \mathbf{c})' D_c^{-1} (\mathbf{r}_i - \mathbf{c})$$

$$- \sum_{i=1}^n \sum_{j=1}^p y_{ij}^2 = \sum_{i=1}^n \sum_{j=1}^p \left(\frac{f_{ij} - r_i c_j}{\sqrt{r_i c_j}} \right)^2 = \frac{X^2}{N} = \sum_{k=1}^{p-1} \lambda_k^2 : \text{Total inertia}$$

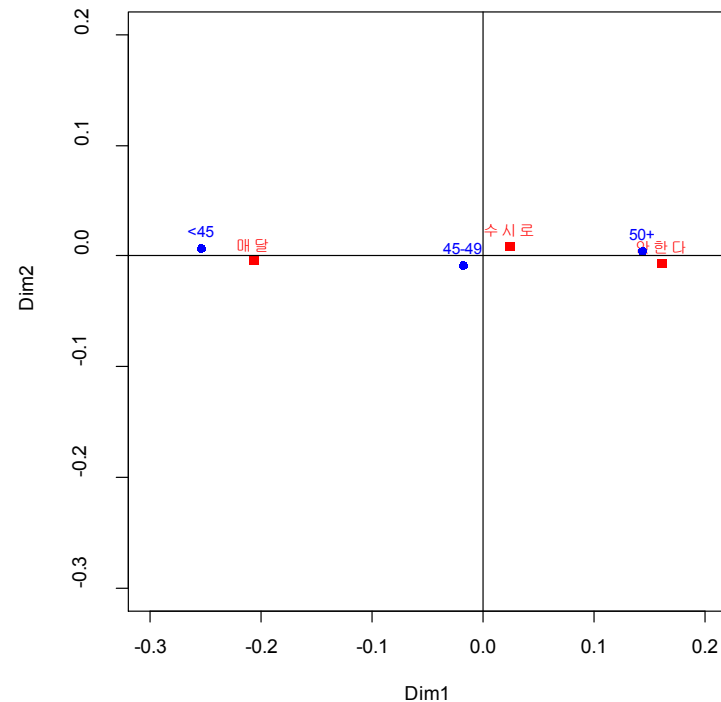
- Total inertia : a measure of the variation in the count data.

[R-code : wbc-SCAalgorithm.R]

```
O<-matrix(c(91, 90, 51,
            150, 200, 155,
            109, 198, 172), byrow=T, nrow=3)
F <- O/sum(O)
r <- apply(F,1,sum)
c <- apply(F,2,sum)
Dr<- diag(1/sqrt(r))
Dc<- diag(1/sqrt(c))
Y <- Dr%*(F-r%*(t(c))%*Dc
svd.Y <- svd(Y)
U <- svd.Y$u
V <- svd.Y$v
D <- diag(svd.Y$d)
Cr <- (Dr%*U%*D)[,1:2]
Cc <- (Dc%*V%*D)[,1:2]
rownames(Cr) <- c("<45", "45-49", "50+")
rownames(Cc) <- c("매달", "수시로", "안한다")

eig <- (svd.Y$d)^2
per <- eig/sum(eig)*100
gof <- sum(per[1:2])
rbind(round(eig, 3),round(per, 3))
par(pty="s")
lim <-range(pretty(Cr))
plot(Cc[, 1:2],
xlab="Dim1",ylab="Dim2",xlim=lim,ylim=lim,pch=15,col=2,
Biplot of WBC : SCA Algorithm)
text(Cc[, 1:2],rownames(Cc),cex=0.8,col=2,pos=3)
points(Cr[, 1:2],pch=16, col=4)
text(Cr[, 1:2],rownames(Cr),cex=0.8,pos=3, col=4)
abline(v=0,h=0)
```

SCA Biplot of WBC : SCA Algorithm



나 이	유방자기진단 빈도		
	매 달	수시로	안한다
< 45	91	90	51
45 - 49	150	200	155
50 +	109	198	172

2.2 다중 대응분석 그림 Multiple Correspondence Analysis Map

1) 삼원분할표 Three-Way Contingency Table

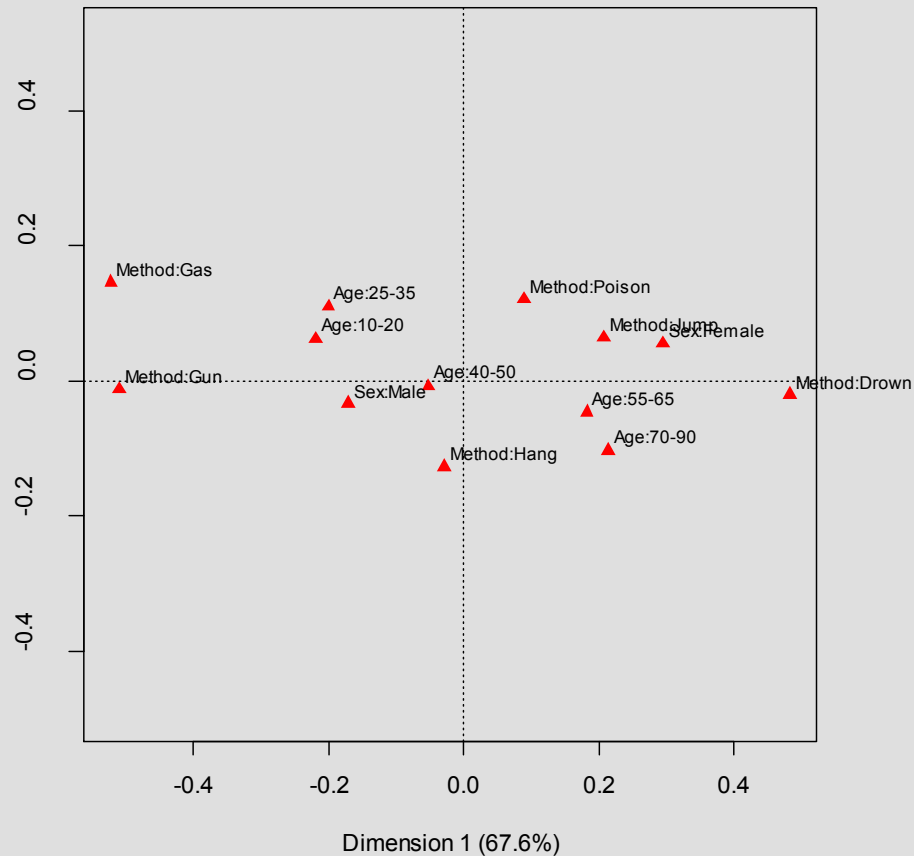
sucidefreq.txt								sucidefreq-MCRA.R	
[자료 2.2.1] 2×5×6 성별과 연령별 자살유형의 분할표								<pre> setwd("C:/Users/stat/Desktop/개인_WORK/외부초 청강의/부산대간호과학연구소/R-codes") Data0.2.1<-read.table("sucidefreq.txt") Freq<-as.matrix(Data0.2.1) sex<-c("Male", "Female") age<- c("10-20", "25-35", "40-50", "55-65", "70-90") method<-c("Poison", "Gas", "Hang", "Drown", "Gun", "Jump") data<-expand.grid(Sex=sex, Age=age, Method=method) data<-cbind(data, Freq) # Makeing a 3-way table table<-xtabs(Freq~Sex+Age+Method, data=data) plot(mjca(table), main="MCRA : 성별, 나이, 자살유형의 삼원분할표") </pre>	
Sex	Age	Poison	Gas	Hang	Drown	Gun	Jump		
M	10-20	1160	335	1524	67	512	189		
M	25-35	2823	883	2751	213	852	366		
M	40-50	2465	625	3936	247	875	244		
M	55-65	1531	201	3581	207	477	273		
M	70-90	938	45	2948	212	229	268		
F	10-20	921	40	212	30	25	131		
F	25-35	1672	113	575	139	64	276		
F	40-50	2224	91	1481	354	52	327		
F	55-65	2283	45	2014	679	29	388		
F	70-90	1548	29	1355	501	3	383		

sucidefreq - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도

1160
921
2823
1672
2465
2224
1531
2283
938
1548
335
40
883
113
625
91
201
45
45
29

자살유형별의
다중 대응분석 결과



- 수평축(제1 축) : 연령별(Age)로 살펴보면 축의 왼편에 10-20대와 25-35대가 같은 위치에 있어서 서로 대응하여 동질하며 40-50대는 다소 원점에 가깝지만 이들과 같이 대응한다고 볼 수 있다. 오른편에는 고 연령대(55-65, 70-90)가 대응하고 있다.

- 성별(Sex)은 제1축에 대하여 왼편이 남자(Male)이고 오른편에 여자(Female)가 있어 서로 동질하지 않음을 보여준다.

- 자살유형(Method)을 보면 제1축의 왼편에 Gun(총기)와 Gas(가스)가 있고 Hang(목매기)도 이들과 대응하나 뚜렷하지는 않다. 오른편에는 Poison(독약), Jump(뛰어내림), Drown(익사)가 서로 대응한다.

- 전체적으로 나이가 젊은 남성들은 자살 유형으로 Gun(총기)와 Gas(가스)가 있고 Hang(목매기)를 선택하고 나이가 많은 여성들은 Poison(독약), Jump(뛰어내림), Drown(익사)의 자살유형을 선택하고 있음을 보여주고 있다.

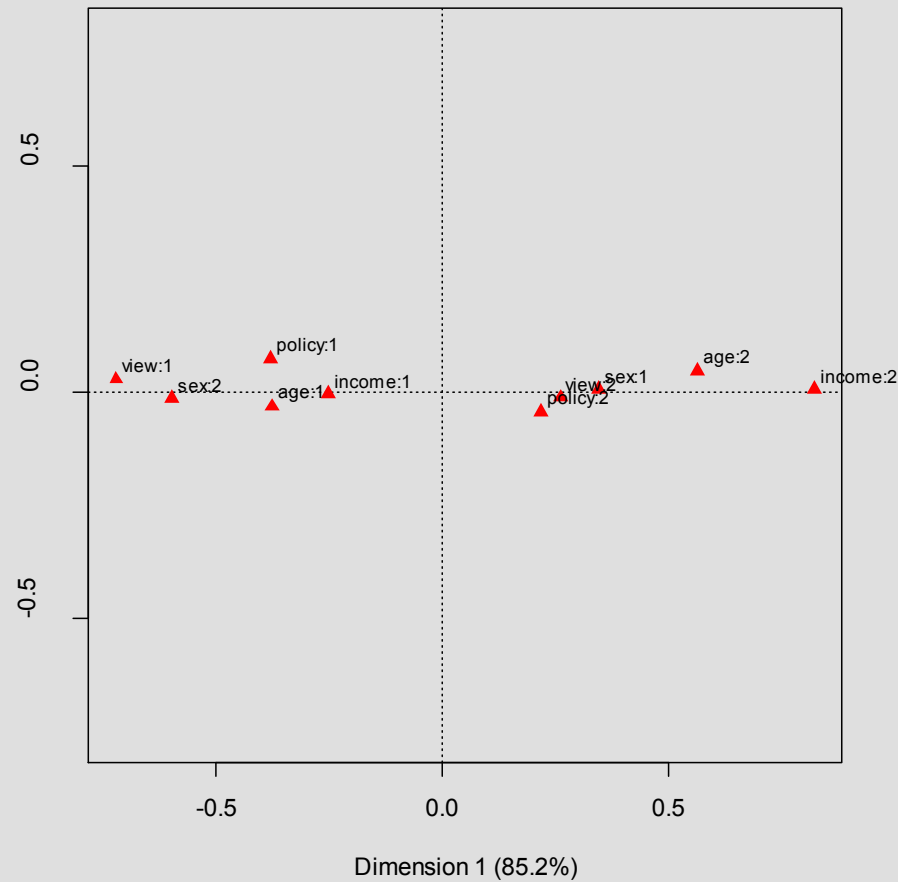
삼원 분할표의 다중 대응분석 그림 : library(ca), mjca()

해 석

2) 오원 분할표 Five-Way Contingency Table

view.txt	view-MCRA.R																																																																																																																																																																																																																																										
[자료 2.2.2] 경제전망과 경제정책선택 <table><thead><tr><th>id</th><th>sex</th><th>age</th><th>income</th><th>view</th><th>policy</th></tr></thead><tbody><tr><td>1</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td><td>2</td></tr><tr><td>3</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>4</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>5</td><td>2</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>6</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td></tr><tr><td>7</td><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>8</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>9</td><td>1</td><td>1</td><td>1</td><td>1</td><td>2</td></tr><tr><td>10</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td colspan="6">----- 중 간 생 략 -----</td></tr><tr><td>115</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>116</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>117</td><td>2</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>118</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td></tr><tr><td>119</td><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>120</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></tbody></table> view - 메모장 <table><thead><tr><th>id</th><th>sex</th><th>age</th><th>income</th><th>view</th><th>policy</th></tr></thead><tbody><tr><td>1</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td><td>2</td></tr><tr><td>3</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>4</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>5</td><td>2</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>6</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td></tr><tr><td>7</td><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>8</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>9</td><td>1</td><td>1</td><td>1</td><td>1</td><td>2</td></tr><tr><td>10</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>11</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>12</td><td>2</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>13</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td></tr><tr><td>14</td><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>15</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>16</td><td>1</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>17</td><td>2</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>18</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td></tr><tr><td>19</td><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>20</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></tbody></table>	id	sex	age	income	view	policy	1	1	2	2	2	2	2	1	1	1	1	2	3	2	1	1	1	1	4	1	2	2	2	2	5	2	1	1	2	2	6	1	2	1	2	1	7	1	1	1	2	2	8	1	2	2	2	2	9	1	1	1	1	2	10	2	1	1	1	1	----- 중 간 생 략 -----						115	2	1	1	1	1	116	1	2	2	2	2	117	2	1	1	2	2	118	1	2	1	2	1	119	1	1	1	2	2	120	2	1	1	1	1	id	sex	age	income	view	policy	1	1	2	2	2	2	2	1	1	1	1	2	3	2	1	1	1	1	4	1	2	2	2	2	5	2	1	1	2	2	6	1	2	1	2	1	7	1	1	1	2	2	8	1	2	2	2	2	9	1	1	1	1	2	10	2	1	1	1	1	11	1	2	2	2	2	12	2	1	1	2	2	13	1	2	1	2	1	14	1	1	1	2	2	15	2	1	1	1	1	16	1	2	2	2	2	17	2	1	1	2	2	18	1	2	1	2	1	19	1	1	1	2	2	20	2	1	1	1	1	<pre>setwd("C:/Users/stat/Desktop/개인_WORK/외부초 청강의/부산대간호과학연구소/R-codes") view<-read.table("view.txt", header=T) view<-view[, -1] ##Multiple CA Plot library(ca) mjca(view) par(pty="s") plot(mjca(view), main="MCRA : 성별, 나이 수입에 따른 경제전망과 정책선택도")</pre>
id	sex	age	income	view	policy																																																																																																																																																																																																																																						
1	1	2	2	2	2																																																																																																																																																																																																																																						
2	1	1	1	1	2																																																																																																																																																																																																																																						
3	2	1	1	1	1																																																																																																																																																																																																																																						
4	1	2	2	2	2																																																																																																																																																																																																																																						
5	2	1	1	2	2																																																																																																																																																																																																																																						
6	1	2	1	2	1																																																																																																																																																																																																																																						
7	1	1	1	2	2																																																																																																																																																																																																																																						
8	1	2	2	2	2																																																																																																																																																																																																																																						
9	1	1	1	1	2																																																																																																																																																																																																																																						
10	2	1	1	1	1																																																																																																																																																																																																																																						
----- 중 간 생 략 -----																																																																																																																																																																																																																																											
115	2	1	1	1	1																																																																																																																																																																																																																																						
116	1	2	2	2	2																																																																																																																																																																																																																																						
117	2	1	1	2	2																																																																																																																																																																																																																																						
118	1	2	1	2	1																																																																																																																																																																																																																																						
119	1	1	1	2	2																																																																																																																																																																																																																																						
120	2	1	1	1	1																																																																																																																																																																																																																																						
id	sex	age	income	view	policy																																																																																																																																																																																																																																						
1	1	2	2	2	2																																																																																																																																																																																																																																						
2	1	1	1	1	2																																																																																																																																																																																																																																						
3	2	1	1	1	1																																																																																																																																																																																																																																						
4	1	2	2	2	2																																																																																																																																																																																																																																						
5	2	1	1	2	2																																																																																																																																																																																																																																						
6	1	2	1	2	1																																																																																																																																																																																																																																						
7	1	1	1	2	2																																																																																																																																																																																																																																						
8	1	2	2	2	2																																																																																																																																																																																																																																						
9	1	1	1	1	2																																																																																																																																																																																																																																						
10	2	1	1	1	1																																																																																																																																																																																																																																						
11	1	2	2	2	2																																																																																																																																																																																																																																						
12	2	1	1	2	2																																																																																																																																																																																																																																						
13	1	2	1	2	1																																																																																																																																																																																																																																						
14	1	1	1	2	2																																																																																																																																																																																																																																						
15	2	1	1	1	1																																																																																																																																																																																																																																						
16	1	2	2	2	2																																																																																																																																																																																																																																						
17	2	1	1	2	2																																																																																																																																																																																																																																						
18	1	2	1	2	1																																																																																																																																																																																																																																						
19	1	1	1	2	2																																																																																																																																																																																																																																						
20	2	1	1	1	1																																																																																																																																																																																																																																						
(1) sex(성별) : 1(male), 2(female) (2) age(나이) : 1(30세 이하), 2(30세 이상) (3) income(연수입) : 1(1000만원 이하), 2(1000만원 이상) (4) view(경제전망) : 1(낙관), 2(비관) (5) policy(정부의 경제정책) : 1(찬성), 2(반대)																																																																																																																																																																																																																																											

경제전망



- 제1축의 왼편에 연령(age:1)과 수입(income:1)이 낮은 여성(sex:2)들이 경제전망을 낙관(view:1)하고 정부의 경제정책을 찬성(policy:1)하고 있다.

- 제1축의 오른편에는 연령(age:2)과 수입(income:2)이 높은 남성(sex:1)들이 경제전망을 비판(view:2)하고 경제 정책을 반대(policy:2)하고 있음을 보여주고 있다.

오원분할표의 다중 대응분석 그림 : library(ca), mja()

해 석

3) 다원 분할표 Multi-Way Contingency Table

ptsd.txt

[자료 2.2.3] 외상 후 스트레스증후군에 대한 다시점자료

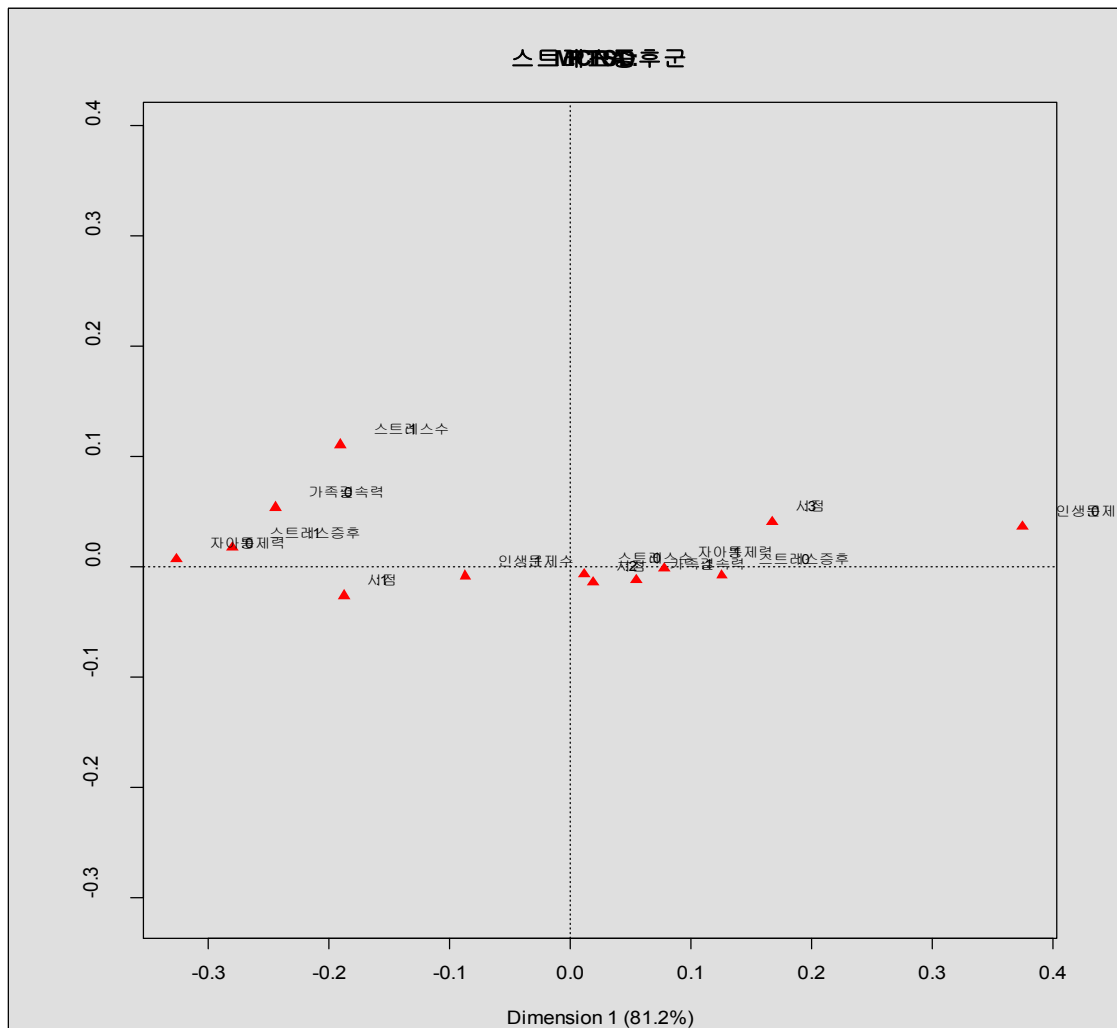
subject	time	ptsd	control	problems	sevent	cohes
15	1	0	3.22222	5.625	1	8
15	2	0	3.16667	5.375	0	8
15	3	0	3.27778	3.750	1	8
18	1	1	2.55556	9.250	0	8
18	2	0	3.44444	4.375	0	8
18	3	0	3.33333	2.375	0	8
----- 중 간 생 략 -----						
571	1	0	3.55556	3.000	0	7
571	2	0	2.94444	1.875	0	7
571	3	0	3.50000	2.750	0	7

ptsd - 메모장

파일(F)	편집(E)	서식(O)	보기(V)	도움말(H)		
개체	시점	스트레스증후	자아통제력	인생문제수	스트레스수	가족결속력
15	1	0	3.22222	5.625	1	8
15	2	0	3.16667	5.375	0	8
15	3	0	3.27778	3.750	1	8
18	1	1	2.55556	9.250	0	8
18	2	0	3.44444	4.375	0	8
18	3	0	3.33333	2.375	0	8
19	1	1	2.72222	7.750	1	7
19	2	1	2.77778	7.750	1	7
19	3	0	2.77778	7.500	1	7

ptsd-MCRA.R

setwd("C:/Users/stat/Desktop/개인_WORK/외부초청강의/부산대간호과학연구소/R-codes")
ptsd<-read.table("ptsd.txt", header=T)
v1 <- as.numeric(ptsd\$자아통제력>=3)
v2 <- as.numeric(ptsd\$인생문제수>=3)
v3<- as.numeric(ptsd\$스트레스수>=3)
v4 <- as.numeric(ptsd\$가족결속력>=6)
nptsd <- data.frame(ptsd[,1:3],
자아통제력=v1,인생문제수=v2, 스트레스수=v3, 가족결속력=v4)
nptsd<-nptsd[, -1]
library("ca")
mjca(nptsd)
par(pty="s")
plot(mjca(nptsd), main="MCRA : 스트레스증후군 PTSD")



다원분할표의 다중 대응분석 그림 : library(ca), mja()

- 제1축 왼편 : 시점:1(3개월 후), 자아통제력:0과 가족결속력:0, 인생문제수:1과 스트레스수:1이 서로 대응한다.
- 환자가 외상 3개월 후 자아통제력과 가족결속력이 낮으며 인생문제와 스트레스가 높음을 보여주고 있다.
- 제2축의 오른편 : 외상 후 환자가 시점 2-3(3-12개월 후)에서 자아통제력(1)과 가족결속력(1)이 높고, 인생문제수(0)과 스트레스(0)가 낮아졌음을 보여주고 있다.

해 석

[자료 2.2.1] 조건대 저널의 키워드와 통계기법 자료

[illegible]

origin.csv

	A	B	C	D	E	F
1	1	2	3	4	5	6
2	1	0	1	1	1	1
3	0	0	1	1	0	1
4	0	0	0	1	0	1
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	1	0	1	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0

	A	B	C	D	E	F	G	H	I	J	K	L
1	기초통계	원빈도분석	군집분석	다차원척도	주성분분석	ANOVA	통계적유의	다중비교	상관분석	회귀분석	t검정	서해

origin-MCRA.R

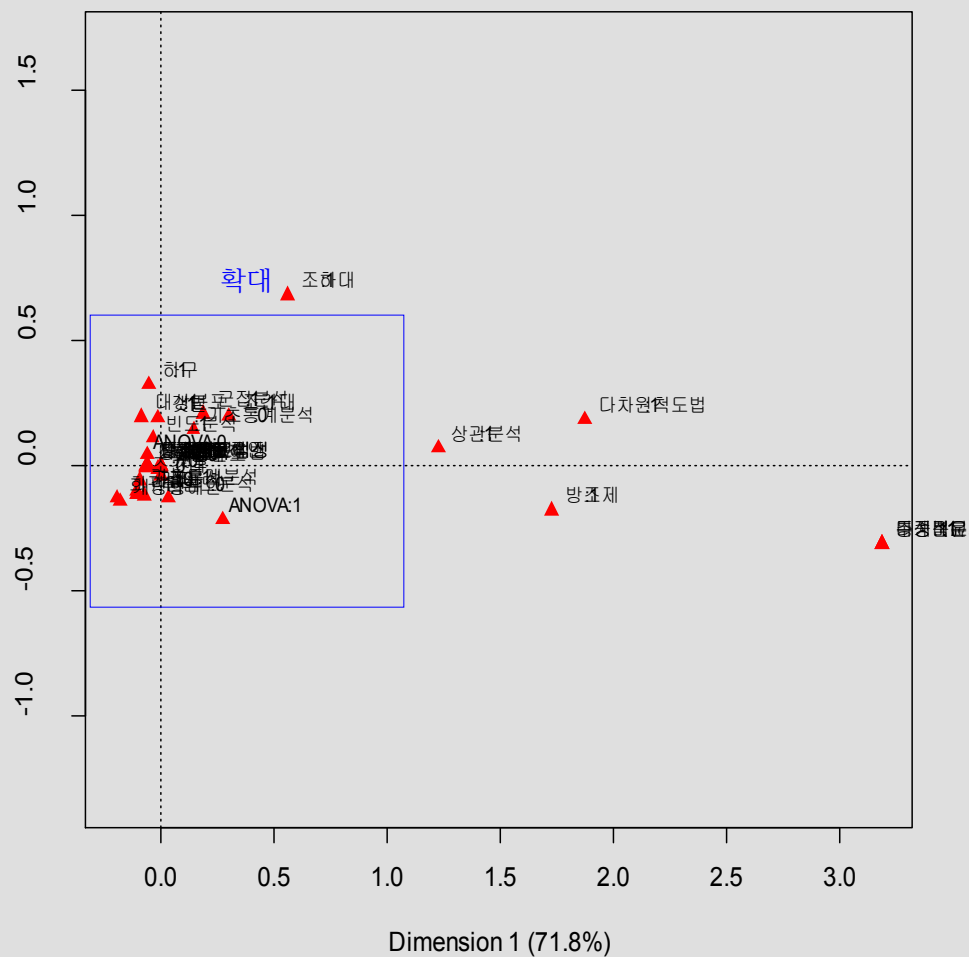
```
# 데이터 불러오기(파일명 : origin.data.csv, colname.csv)
setwd("C:/Users/stat/Desktop/개인_WORK/외부초
청강의/부산대간호과학연구소/R-codes")
origin.data <- read.csv("origin.data.csv",header=TRUE)
dim(origin.data) #138,60
```

```
data <- t(origin.data)
rownames(data) <- c(1:60)
colnames(data)<-
as.matrix(read.csv("colname.csv",header=FALSE))
data <- data.frame(data)
dim(data) #60,138
```

```
library(ca)
```

```
# 통계기법과 Zonation의 연관성
#먼저 합이 0이 되는 열이 있는지 확인, 없으면 그 데이터
그대로 사용
apply(data[,c(1:11,45:51)],2,sum)
x9 <- data[,c(1:11,45:51)] #60,18
mjca(x9)
plot(mjca(x9))
```

```
# 영역 확대
plot(mjca(x9), xlim=c(-0.3,0.3), ylim=c(-0.3,0.3))
```



통계기법과 Zonation의 연관성을 위한 다중 대응분석 그림

```
> apply(data[,c(1:11,45:51)],2,sum)
```

기초통계분석	빈도분석	군집분석	다차원척도법	주성분분석
35	32	21	2	1
ANOVA	통계적유의성	다중비교검정	상관분석	회귀분석
10	1	1	3	1
t검정	대상분포	조간대	조하대	방조제
1	2	14	1	2
갯벌	개방형해안	하구		
10	1	1		

빈도 : apply(, sum)

5) 버트행렬 자료 : 교육수준, 나이, 거주지역의 규모 등 인구통계학적 변인과 사회경제적 변인간의 관계

socioeconburt.txt

–EDU(교육수준): 1(none),2(초등),3(고중태),4(고졸),5(대졸)

–LOD(주거상태): 1(대출), 2(소유), 3(전세), 4(기숙)

–STO(주식): 1(소유), 2(무소유)

–HOU(부동산): 1(소유), 2(무소유)

–AGE(나이) : 1(<25), 2(25–34), 3(35–49),4(50–64), 5(65<)

–SIZ(지역규모) : 1(<2천), 2(2천–2만), 3(5만–10만), 4(10만<)

–SEX(성별) : F, M

인구통계적 변인과 사회 경제적 변인에 의한 버트표

	SEXM	SEXF	EDU1	EDU2	EDU3	EDU4	EDU5	LOD1	LOD2	LOD3	LOD4	STO1	STO2	HOU1	HOU2	AGE1	AGE2	AGE3	AGE4	AGE5	SIZ1	SIZ2	SIZ3	SIZ4	SIZ5
SEXM	469	0	102	164	69	65	69	62	151	224	32	54	415	35	434	18	111	169	84	87	42	40	81	161	145
SEXF	0	531	94	163	91	102	81	58	142	302	29	67	464	47	484	22	136	187	104	82	41	47	94	168	181
EDU1	102	94	196	0	0	0	0	17	59	105	15	12	184	11	185	9	27	54	47	59	19	17	35	70	55
EDU2	164	163	0	327	0	0	0	45	116	151	15	26	301	21	306	0	55	125	84	63	43	34	67	110	73
EDU3	69	91	0	0	160	0	0	16	37	98	9	19	141	14	146	14	47	60	27	12	12	13	34	52	49
EDU4	65	102	0	0	0	167	0	27	39	87	14	27	140	19	148	16	62	63	16	10	8	15	21	57	66
EDU5	69	81	0	0	0	0	150	15	42	85	8	37	113	17	133	1	56	54	14	25	1	8	18	40	83
LOD1	62	58	17	45	16	27	15	120	0	0	0	11	109	7	113	3	23	68	20	6	7	20	27	48	18
LOD2	151	142	59	116	37	39	42	0	293	0	0	60	233	48	245	9	24	91	80	89	59	33	62	72	67
LOD3	224	302	105	151	98	87	85	0	0	526	0	45	481	23	503	22	180	182	79	63	11	29	80	191	215
LOD4	32	29	15	15	9	14	8	0	0	0	61	5	56	4	57	6	20	15	9	11	6	5	6	18	26
STO1	54	67	12	26	19	27	37	11	60	45	5	121	0	39	82	2	18	35	27	39	4	9	22	36	50
STO2	415	464	184	301	141	140	113	109	233	481	56	0	879	43	836	38	229	321	161	130	79	78	153	293	276
HOU1	35	47	11	21	14	19	17	7	48	23	4	39	43	82	0	3	11	27	20	21	7	12	12	25	26
HOU2	434	484	185	306	146	148	133	113	245	503	57	82	836	0	918	37	236	329	168	148	76	75	163	304	300
AGE1	18	22	9	0	14	16	1	3	9	22	6	2	38	3	37	40	0	0	0	0	2	4	8	15	11
AGE2	111	136	27	55	47	62	56	23	24	180	20	18	229	11	236	0	247	0	0	0	13	16	34	102	82
AGE3	169	187	54	125	60	63	54	68	91	182	15	35	321	27	329	0	0	356	0	0	31	34	71	100	120
AGE4	84	104	47	84	27	16	14	20	80	79	9	27	161	20	168	0	0	0	188	0	25	20	26	60	57
AGE5	87	82	59	63	12	10	25	6	89	63	11	39	130	21	148	0	0	0	0	169	12	13	36	52	56
SIZ1	42	41	19	43	12	8	1	7	59	11	6	4	79	7	76	2	13	31	25	12	83	0	0	0	0
SIZ2	40	47	17	34	13	15	8	20	33	29	5	9	78	12	75	4	16	34	20	13	0	87	0	0	0
SIZ3	81	94	35	67	34	21	18	27	62	80	6	22	153	12	163	8	34	71	26	36	0	0	175	0	0
SIZ4	161	168	70	110	52	57	40	48	72	191	18	36	293	25	304	15	102	100	60	52	0	0	0	329	0
SIZ5	145	181	55	73	49	66	83	18	67	215	26	50	276	26	300	11	82	120	57	56	0	0	0	0	326

socioeconburt-MCAalgorithm.R

setwd("C:/Users/stat/Desktop/개인_WORK/외부초청강의/부산대간호과학연구소 /R-codes")

burt<-read.table("socioeconburt.txt")

B<-as.matrix(burt)

colnames<-c("sexm","sexf", "edu1", "edu2", "edu3", "edu4", "edu5",
"lod1", "lod2", "lod3", "lod4", "sto1", "sto2", "hou1", "hou2",
"age1", "age2", "age3", "age4", "age5", "siz1", "siz2", "siz3", "siz4",
"siz5")

n=1000

q=25

P <- B / sum(B)

cm <- apply(P, 2, sum)

Dc<-diag(1/sqrt(cm))

eP <- cm %%% t(cm)

Y <- (P - eP) / sqrt(eP)

Spectral Decomposition

S<-t(Y)%%Y

dec <- eigen(S)

V<-dec\$vectors

delt <- dec\$values

expl <- 100*(delt/ sum(delt))

lam <- delt^2

fit <- 100*(lam/ sum(lam))

fit

rbind(round(lam, 3),round(fit, 1))

Dl<-diag(delt)

Cb<- Dc%%V%%Dl

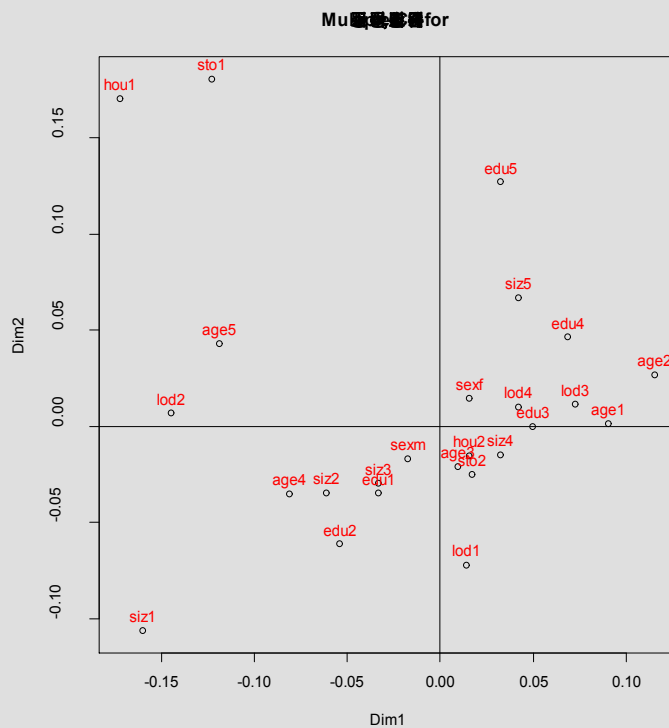
Cb

lim<-range(pretty(Cb))

plot(Cb[, 1:2], xlab="Dim1", ylab="Dim2", main="Multiple CA for Burt Matrix")

text(Cb[, 1:2], colnames, col=2, pos=3)

abline(v=0, h=0)



: Goodness-of-fit

```
> rbind(round(lam, 3), round(fit, 1))
      [,1] [,2] [,3] [,4] [,5]
[1,]  0.004 0.002 0.001 0.001 0.001
[2,] 31.200 18.100 9.300 7.400 5.400
```

사분면	대응 관계	해 석
1	age1-2, edu3-5, siz5, lod3-4, sexf	35세미만/50만명대도시 교육수준상/임대주택/여
2	age5, hou1, lod2, sto1	65세이상/주식 부동산소유
3	age4, edu1-2, siz1-3, sexm	5-60대/교육수준하, 주민수 10만명이하/남
4	age3, hou2, siz4, lod1, sto2	3-40대/10만이상도시 주식동산무소유

다중 대응분석 그림 : library(ca), mja()

해 석

Algorithms for Multiple CA

[Step 1] **Indicator Matrix** : $Z = [Z_1 | Z_2 | \dots | Z_p]$, $n \times q$ indicator matrix.

where Z_j : $n \times q_j$ indicator matrix of the j th categorical variable with q_j categories for the n samples and $q = (q_1 + q_2 + \dots + q_p)$.

[Step 2] **Burt Matrix** : $B = Z^t Z = (b_{ij})$, $i, j = 1, \dots, q$.

[Step 3] **Correspondence Matrix** : $P = (p_{ij}) = (b_{ij}/N)$

with grand total $N = \sum_{i=1}^n \sum_{j=1}^q b_{ij}$.

[Step 4] **Centred Correspondence Matrix** : $P - \mathbf{r}\mathbf{c}^t$

where the column and row mass vectors $\mathbf{r} = P\mathbf{1}_q$ and $\mathbf{c} = P^t\mathbf{1}_n$, respectively.

And $r_i = \sum_{j=1}^q p_{ij}$, $i = 1, \dots, q$ and $c_j = \sum_{i=1}^n p_{ij}$, $j = 1, \dots, q$. In fact, $\mathbf{r} = \mathbf{c}$

[Step 5] **Residual Matrix** : $Y = D_r^{-1/2} (P - \mathbf{r}\mathbf{c}^t) D_c^{-1/2} = \left(\frac{p_{ij} - r_i c_j}{\sqrt{r_i c_j}} \right)$

where $D_r^{1/2} = \text{diag}(\sqrt{r_1}, \dots, \sqrt{r_q})$ and $D_c^{1/2} = \text{diag}(\sqrt{c_1}, \dots, \sqrt{c_q})$.

[Step 6] **Singular Value Decomposition** : $Y = U D_\lambda V^t$

[Step 7] **Coordinate Matrix for MCA** : $C_B = D_c^{-1/2} V D_\lambda$.

[Step 8] **Goodness-of-fit** for s-dimensional MCA Map(Biplot) :

$$fit = \frac{(\lambda_1^2 + \lambda_2^2 + \dots + \lambda_s^2)}{\sum_{k=1}^r \lambda_k^2} \times 100\%,$$

Indicator Matrix/ Burt Matrix

	c1	c2	Total
r1	2	2	4
r2	1	3	4
Total	3	5	8

	Rows		Columns		Total
	r1	r2	c1	c2	
	1	0	1	0	2
	1	0	1	0	2
	1	0	0	1	2
	1	0	0	1	2
	0	1	1	0	2
	0	1	0	1	2
	0	1	0	1	2
	0	1	0	1	2
Total	4	4	3	5	16

$Z = [Z_1 | Z_2]$: Indicator Matrix

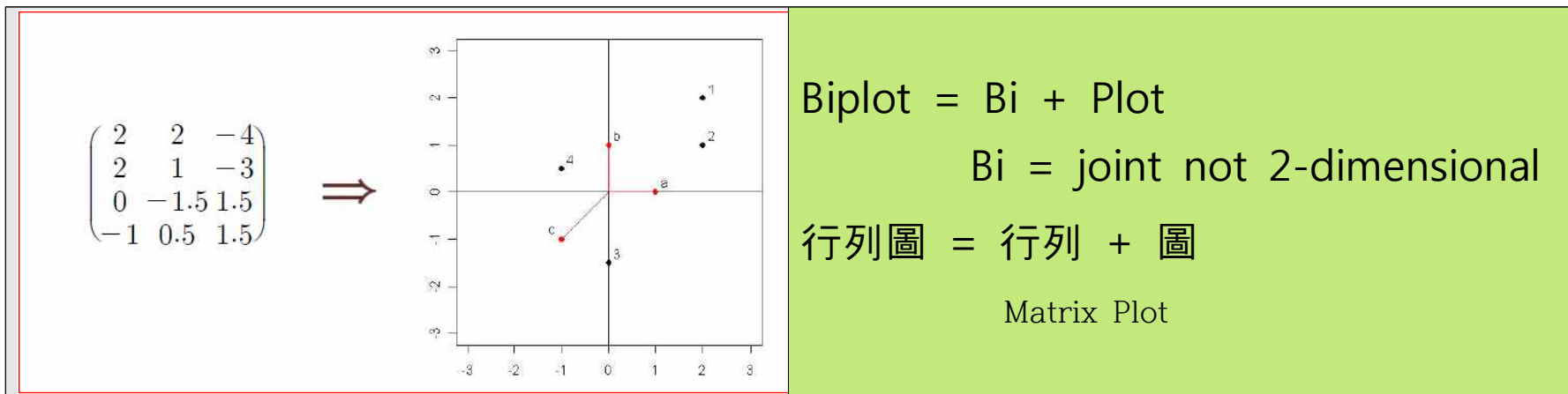
$B = Z Z^t = \begin{pmatrix} Z_1^t Z_1 & Z_1^t Z_2 \\ Z_2^t Z_1 & Z_2^t Z_2 \end{pmatrix} = \begin{bmatrix} 4 & 0 & 2 & 2 \\ 0 & 4 & 1 & 3 \\ 2 & 1 & 3 & 0 \\ 2 & 3 & 0 & 5 \end{bmatrix}$: Burt Matrix

$Z_1^t Z_2 = N$

다변량자료의 시각화 (Visualizations for MDA)

2.3 행렬도 Biplot

Geometric Visualization of Biplot



◇ 행렬도 개발

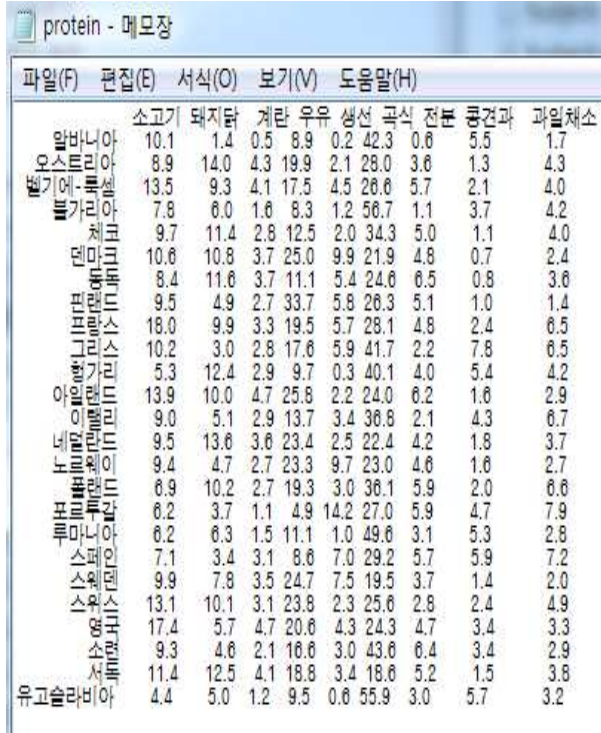
Gabriel(1971, 1982) : 주성분행렬도, JK' 행렬도, 대칭행렬도

◇ 국내 행렬도 연구

Choi(1991) : 행렬도의 저항성 버전 연구, 행렬도를 처음 소개

허명희(1993) : Biplot을 行列圖라 처음 명명함

1) [자료 0.1.3] 유럽 25개국의 단백질 섭취원(단위: 그램/일) 자료

protein.txt	protein-PCbiplot.R
	<pre> setwd("C:/Users/stat/Desktop/개인_WORK/외부초청강의/부산대간호과학연구 소/R-codes") Y<-read.table("protein.txt", header=T) pcasvd<-prcomp(Y, scale=T) summary(pcasvd) biplot(pcasvd, pc.biplot=T, xlab="Dim1", ylab="Dim2", main=" PC Biplot : pc.biplot() ", cex=1.0, pch=16) abline(v=0,h=0) </pre>

1) [자료 0.1.3] 유럽 25개국의 단백질 섭취원(단위: 그램/일) 자료 : Algebraic Algorithm

protein-PCbiplotAlgorithm.R

```
setwd("C:/Users/stat/Desktop/개인_WORK/외부초청강의/부산대간호학연구
소/R-codes")
X<-read.table("protein.txt", header=T)
colnames(X)
rownames(X)

##### Functions #####
PCbiplot <- function(X,std=0) {
  if(!is.matrix(X)) { X <- as.matrix(X) }
  if(is.na(match(std,c(1,0)))){stop("std : Select One of
1:Standardized / 0:Centred","\n")}
  n <- nrow(X)
  p <- ncol(X)
  if(std==0) { Y <- scale(X,scale=F) ;std <- "Centred"}
  if(std==1) { Y <- scale(X) ;std <- "Standardized"}
  svd.Y <- svd(Y)
  U <- svd.Y$u
  V <- svd.Y$v
  D <- diag(svd.Y$d)
  r <- length(svd.Y$d)
  A <- U%*%D
  B <- V%*%D
  dimnames(A) <- list(rownames(X),paste("Dim",c(1:r),sep=""))
  dimnames(B) <- list(colnames(X),paste("Dim",c(1:r),sep=""))
  method <- paste(std,"PCplot")
  gof <- gof.2d(svd.Y$d)

  return(list(method=method,Eigen=gof$per,GOF=gof$gof,A=A[,1:2],B=B[,1:2]))
}

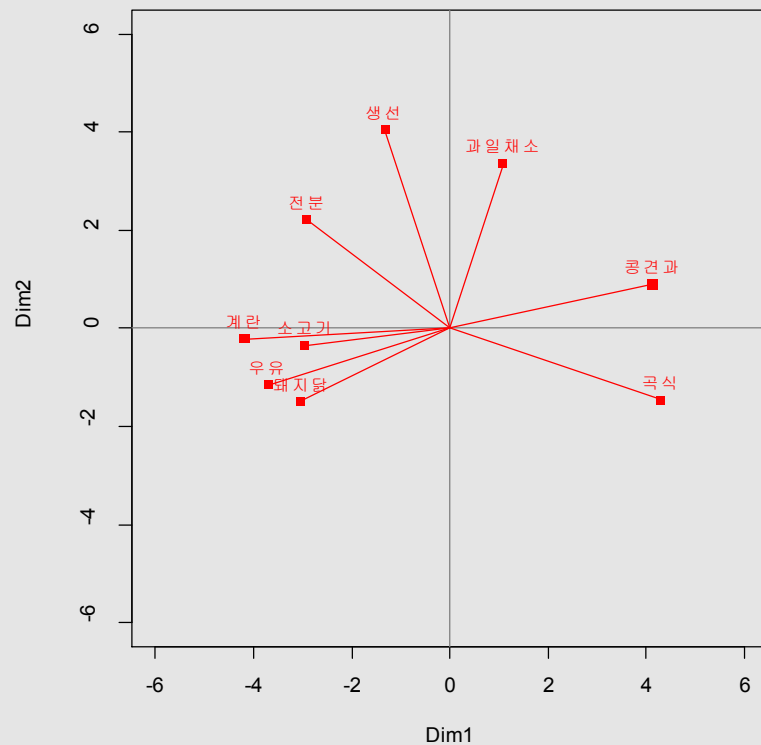
gof.2d <- function(s.val){
  e.val <- s.val^2
  per <- e.val/sum(e.val)*100
  e.mat <- round(cbind(e.val,per),4)
  colnames(e.mat) <- c("Eigenvalue", "Percent")
  gof <- paste("Goodness of Fit :
",round(sum(per[1:2]),3),"%",sep="")
  return(list(per=e.mat,gof=gof))
}

plot.2d <-
function(OBS=NULL,VAR=NULL,Line=T,v.lab=T,o.lab=T,v.pos=3,o.pos=3,lim
=NULL,Title=NULL){
  if(!is.logical(Line)){ stop("Line : Select One of T:Line Draw /
F:Only Points","\n") }
```

```

  if(!is.logical(v.lab)){ stop("v.lab : Select One of T:Var label write / F:Not","\n")
}
  if(!is.logical(o.lab)){ stop("o.lab : Select One of T:Obs label write / F:Not","\n")
}
  if(is.null(lim)){
    if(is.null(OBS)){ lim <- range(pretty(VAR)) }
    if(is.null(VAR)){ lim <- range(pretty(OBS)) }
    if(!is.null(OBS)&!is.null(VAR)){ lim <- range(pretty(range(OBS,VAR))) }
  }
  par(pty="s")
  if(!is.null(OBS)){
    n <- nrow(OBS)
    plot(OBS[,1],OBS[,2],xlab="Dim1",ylab="Dim2",xlim=lim,ylim=lim,pch=16,main=Title)
    abline(v=0,h=0,col="gray50")
    if(o.lab) {
      if(is.null(rownames(OBS))) {
        text(OBS[,1],OBS[,2],c(1:n),cex=0.8,pos=o.pos) }
      if(!is.null(rownames(OBS))){
        text(OBS[,1],OBS[,2],rownames(OBS),cex=0.8,p=o.pos) }
    }
    if(!is.null(VAR)){
      points(VAR[,1],VAR[,2],pch=15,col=2)
      if(v.lab){
        text(VAR[,1],VAR[,2],rownames(VAR),cex=0.8,col=2,pos=v.pos) }
      if(Line){ arrows(0,0,VAR[,1],VAR[,2],col=2,code=0) }
    }
  }
  if(is.null(OBS)){
    plot(VAR[,1],VAR[,2],xlab="Dim1",ylab="Dim2",xlim=lim,ylim=lim,pch=15,col=2,main=Title)
    abline(v=0,h=0,col="gray50")
    if(v.lab){
      text(VAR[,1],VAR[,2],rownames(VAR),cex=0.8,col=2,pos=v.pos) }
    if(Line){ arrows(0,0,VAR[,1],VAR[,2],col=2,code=0) }
  }
}

##### PCA, PCplot #####
standard <- PCbiplot(X, std=1)
plot.2d(OBS=standard$A, VAR=standard$B)
plot.2d(OBS=standard$A)
plot.2d(VAR=standard$B)
PCbiplot(X, std=1)
```

변수 행렬도 : plot.2d(VAR=standard\$B)

Eigenvalue Percent		
[1,]	96.1545	44.5160
[2,]	39.2400	18.1667
[3,]	27.0701	12.5324
[4,]	22.9119	10.6074
[5,]	11.1321	5.1538
[6,]	7.8031	3.6126
[7,]	6.5186	3.0178
[8,]	2.7910	1.2921
[9,]	2.3787	1.1012

\$GOF		
[1]	"Goodness of Fit : 62.683%"	

\$A		
	Dim1	Dim2
알바니아	3.4853673	-1.63047985
오스트리아	-1.4226694	-1.04123130
벨기에-룩셈	-1.6220323	0.15949557
불가리아	3.1340813	-1.30106563
체코	-0.3704646	-0.60266842
덴마크	-2.3652688	0.28544582
동독	-1.4222108	0.45030085
핀란드	-1.5638563	-0.59600255
프랑스	-1.4879824	0.78536517
그리스	2.2397000	1.00105887
헝가리	1.4574398	-0.81595115
아일랜드	-2.6634775	-0.76370648
이탈리아	1.5345653	0.39898708
네덜란드	-1.6414454	-0.91199089
노르웨이	-0.9747029	0.82202867
폴란드	-0.1218695	0.53174194
포르투갈	1.7058540	4.28893399
루마니아	2.7568124	-1.11878536
스페인	1.3118074	2.55352416
스웨덴	-1.6337300	-0.20738445
스위스	-0.9123182	-0.75105865
영국	-1.7353682	-0.09397944
소련	0.7825965	-0.11077014
서독	-2.0938353	-0.29377901
유고슬라비아	3.6230077	-1.03802883

\$B		
	Dim1	Dim2
소고기	-2.967339	-0.3523707
돼지갈비	-3.045265	-1.4836926
계란	-4.183941	-0.2213496
우유	-3.703933	-1.1562978
생선	-1.330161	4.0517969
곡식	4.292442	-1.4625932
견과	-2.914763	2.2101643
콩견과	4.121732	0.8977236
과일채소	1.082760	3.3587925

행렬도를 위한 결과 : PCbiplot(X, std=1)

2) 일본과 구미 선진국의 고령화율 추이 (단위 : %) : 대도시 고령사회의 현상과 대응(금성근 · 한승옥 · 한성일(창의연구 2015-12-657))

agingt.txt											agingt-PCbiplot.R	
연도	일본	캐나다	미국	이탈리아	네덜란드	스웨덴	스위스	영국	독일			
1970	7.1	7.9	9.8	11.2	10.2	13.7	11.4	13.0	13.7			
1980	9.1	9.4	11.2	13.5	11.5	16.3	13.8	14.9	15.6			
1990	12.1	11.3	12.3	15.2	12.8	17.8	14.6	15.7	15.0			
2000	17.4	12.6	12.4	18.4	13.6	17.2	15.4	15.9	16.4			
2005	20.2	13.1	12.4	19.7	14.2	17.2	16.0	16.1	18.9			
2010	23.1	14.1	13.0	20.4	15.4	18.3	17.3	16.6	20.5			
2020	29.3	18.1	16.1	23.0	19.8	21.0	20.2	18.5	23.0			
2030	31.8	22.7	19.8	26.8	23.8	22.6	24.1	20.9	28.2			
2040	36.5	24.5	21.0	31.8	26.3	24.1	26.3	22.6	31.8			
2050	39.6	25.5	21.6	33.3	25.6	24.1	26.0	22.9	32.5			

agingt - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

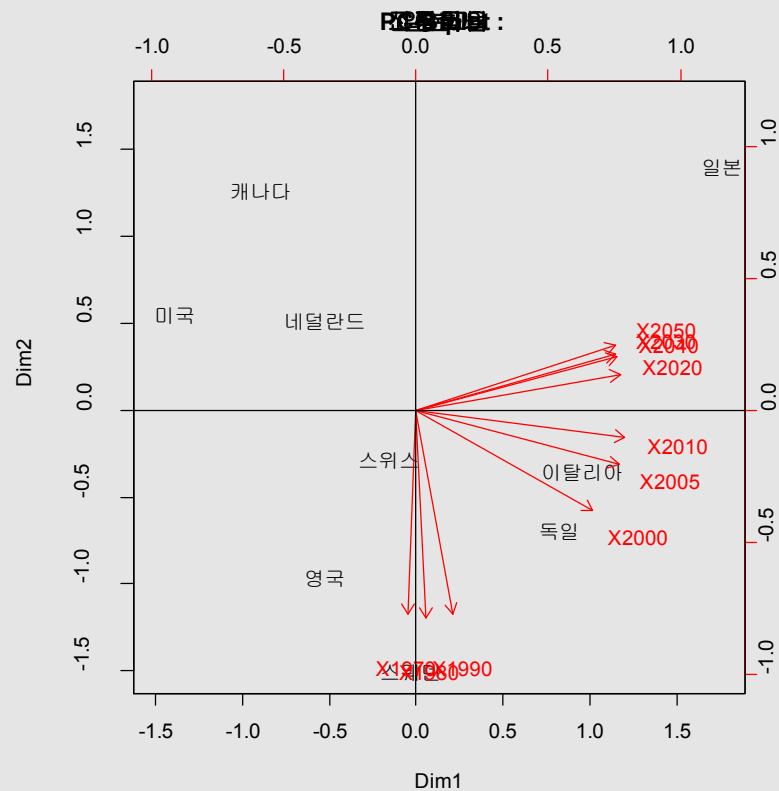
	1970	1980	1990	2000	2005	2010	2020	2030	2040	2050
일본	7.1	9.1	12.1	17.4	20.2	23.1	29.3	31.8	36.5	39.6
캐나다	7.9	9.4	11.3	12.6	13.1	14.1	18.1	22.7	24.5	25.5
미국	9.8	11.2	12.3	12.4	12.4	13.0	16.1	19.8	21.0	21.6
이탈리아	11.2	13.5	15.2	18.4	19.7	20.4	23.0	26.8	31.8	33.3
네덜란드	10.2	11.5	12.8	13.6	14.2	15.4	19.8	23.8	26.3	25.6
스웨덴	13.7	16.3	17.8	17.2	17.2	18.3	21.0	22.6	24.1	24.1
스위스	11.4	13.8	14.6	15.4	16.0	17.3	20.2	24.1	26.3	26.0
영국	13.0	14.9	15.7	15.9	16.1	16.6	18.5	20.9	22.6	22.9
독일	13.7	15.6	15.0	16.4	18.9	20.5	23.0	28.2	31.8	32.5

```

setwd("C:/Users/stat/Desktop/개인_WORK/외부초청강의/부산대간호과학연구소/R-codes")
Y<-read.table("agingt.txt", header=T)

pcasvd<-prcomp(Y, scale=T)
summary(pcasvd)
biplot(pcasvd, pc.biplot=T, xlab="Dim1", ylab="Dim2",
        main=" PC Biplot :일본과 구미 선진국의 고령
        화율 추이 ", cex=1.0, pch=16)
abline(v=0,h=0)

```



```
> summary(pcasvd)
Importance of components:
               PC1      PC2      PC3      PC4      PC5      PC6      PC7
Standard deviation  2.5011 1.8488 0.48344 0.25675 0.1343 0.07379 0.05536
Proportion of Variance 0.6255 0.3418 0.02337 0.00659 0.0018 0.00054 0.00031
Cumulative Proportion 0.6255 0.9673 0.99072 0.99731 0.9991 0.99966 0.99996

               PC8      PC9
Standard deviation  0.01917 3.824e-16
Proportion of Variance 0.00004 0.000e+00
Cumulative Proportion 1.00000 1.000e+00
```

References

Correspondence Analysis

Books

- 정광모, 최용석(2009). 범주형자료분석-SAS의 응용 및 해석-제2판, 자유아카데미.
- 최용석(2001). SAS 대응분석의 이해와 응용. 서울: 자유아카데미.
- 최용석(2017). R과 함께하는 다변량자료분석, 경문사(10월 출간 예정)
- 허명희(2010). R을 활용한 사회네트워크분석 입문, 자유아카데미- 유충현, 홍성학(2015). R을 활용한 데이터시각화, 인사이트.
- Keen, K.J. (2010). Graphics for Statistics and Data Analysis with R, Chapman&Hall/CRC, New York.
- Wickham, H.(2016). ggplot2 - Elegant Graphics for Data Analysis, Second Ed., Springer, AG.
- Sarkar, D.(2008). Lattice - Multivariate Data Visualization with R, Springer, New York.

Papers

- Choi, J. and Choi, Y.S. (2014). Standardizing **Unstructured Big Data** and Visual Interpretation Using MapReduce and **Correspondence Analysis**, *The Korean Journal of Applied Statistics*, 27(2), 169-183.
- Lee, C. and Choi, Y.S. (2014). Visualizations for Matched Pairs Models Using Modified **Correspondence Analysis**, *Communications for Statistical Applications and Methods*, 21(4), 275-284 .
- Kim, J.G., Choi, Y.S. and Choi, S.K.(2012). **Social Network Analysis** and It's Applications for Co-authors and Key-words Networks in the JKSS, *Communications of the Korean Statistical Society* , 19(4), 547-558.
- 2005, Dynamic **Simple Correspondence Analysis** (with G.H. Hyun and M.R. Seo), *The Korean Communications in Statistics*, 12(1), 199-205 .
- 2005, A Numerical Comparison of Map Variability in **SCA** Using the Procrustes Analysis (with G.H. Hyun and J.G. Kim), *Journal of the Korean Data Analysis Society*, 7(5), 1531-1538.
- 2002, Reinterpretation of **Multiple Correspondence Analysis** using the K-means Cluster Analysis(with G.H. Hyun and K. H. Kim), *The Korean Communications in Statistics*, 9(2), 505-514.
- 1999, Robust **Simple Correspondence Analysis** (with M.H. Huh), *Journal of the Korean Statistical Society*, 28(3), 337-346.

Biplots

Books

- 정민지(2017). A Study on Clustering Methods for Proximity Data in Text Mining, 석사학위논문, 부산대학교.
- 최용석(2018). R과 함께하는 다변량자료분석, 경문사(1월 출간 예정)
- 최용석(2014). 다차원척도법의 산책, 자유아카데미.
- 최용석, 신상민 (2013). R을 활용한 행렬도 분석, 자유아카데미.
- Keen, K.J. (2010). Graphics for Statistics and Data Analysis with R, Chapman&Hall/CRC, New York.
- Wickham, H.(2016). ggplot2 - Elegant Graphics fir Data Analysis, Second Ed., Springer, AG.
- Sarkar, D.(2008). Lattice - Multivariate Data Visualization with R, Springer, New York.

Papers

- Shin, S.M., Chun, S.K. and Choi, Y.S.(2017). [Multidimensional scaling](#) of categorical data using the partition method, The Korean Journal of Applied Statistics, To appear
- Nam, S.C. and Choi, Y.S.(2017). Non-parametric Approach for the Grouped Dissimilarities Using the [Multidimensional Scaling](#) and Analysis of Distance, *The Korean Journal of Applied Statistics*, 27(1), The Korean Statistical Society.
- Shin, S.M., Park , J.S., Choi, Y.S. (2015). The mixed [h-plot](#) for continuous and categorical data, 18(1), Journal of the Korean Data Analysis Society
- Shin, S.M., Kim, E.S. and Choi, Y.S. (2015). [Multidimensional scaling](#) using the pseudo-points based on partition method, The Korean Journal of Applied Statistics, 28(6), 1171-1180.DOI: <http://dx.doi.org/10.5351/KJAS.2015.28.6.1171> (2015.12)
- Lee, S.G., Choi, Y.S. and Lee, B.H. (2014). Visualizations of Asymmetric [Multidimensional Scaling](#), The Korean Journal of Applied Statistics, 27(4), 619-627.
- Lee, S.J. and Choi, Y.S. (2014). [Independent Component Biplot](#), The Korean Journal of Applied Statistics, 27(1), 31-41 .
- Choi, T.H. and Choi, Y.S. (2013). Relationship between Physical Fitness and Basic Skill Factors for KTA Tennis Players Using the [Semi-Partial Cannonical Correlation Biplot](#) That Removed the Linear Effect of the Set of Covariate Variables from Physical Fitness

Factor and Procrustes Analysis, Journal of the Korean Data Analysis Society , 15(6), 3067-3076.

- Choi, T.H. and Choi, Y.S. (2012). Relationship between Physical Fitness and Basic Skill Factors for KTA Players Using the Partial [Canonical Correlation Biplot](#) Removing the Linear Effect of the Set of Covariate Variables and Procrustes Analysis, Communications of the Korean Statistical Society , 19(1), 97-106 .
- Lee, B.H., Choi, Y.S. and Shin, S.M.(2012). [Semi-Partial Canonical Correlation Biplot](#), The Korean Journal of Applied Statistics , 25(3), 521-529.
- Yeom, A.R . and Choi, Y.S. (2011). [Partial Canonical Correlation Biplot](#), The Korean Journal of Applied Statistics , 24(3),559-566 .
- Choi, T.H. and Choi, Y.S. (2010). A Study on the Relationship between Physique, Physical Fitness and Basic Skill Factors of Tennis Players of Korea Tennis Association Using the [Generalized Canonical Correlation Biplot](#) and Procrustes Analysis, The Korean Communications in Statistics , 17(6), 1-9.
- Hong, H.U., Choi, Y.S., Shin, S.M. and Kang, C.W . (2010). A Study on Shape Variability in [Canonical Correlation Biplot](#) with Missing Values, The Korean Journal of Applied Statistics , 23(5), 955-966
- Choi, T.H., Choi, Y.S. and Shin, S.M.(2009). A Study on the Relationship between Player Characteristic Factors and Competitive Factors of Tennis Grand Slams Competition Using [Canonical Correlation Biplot](#) and Procrustes Analysis, The Korean Journal of Applied Statistics, 22(4), 855-864.
- Shin, S.M., Choi, Y.S. and Lee, N.Y. (2008). Comparison of Shape Variability in [Principal Component Biplot](#) with Missing Values, The Korean Journal of Applied Statistics, 21(6), 1109-1116.
- Kim, H.Y. and Choi, Y.S (2008). Applications of Cluster Analysis in [Biplots](#), Communications of The Korean Statistical Society, 15(1), 65-76.
- Choi, T.H. and Choi, Y.S. (2008). A Study on the Relationship between Skill and Competition Score Factors of KLPGA Players Using [Canonical Correlation Biplot](#) and Cluster Analysis, The Korean Journal of Applied Statistics, 21(3), 429-439.
- 2005, [Biplot](#) Analysis of Multinomial Logistic Regression Models (with C.W. Kang and K.D. Kim), Journal of the Korean Data Analysis Society, 7(3), 839-849.
- 2005, [Biplot](#) of the Parameter Matrix in MANOVA (with G.H. Hyun and S.M. Jung), Journal of the Korean Data Analysis Society, 7(3), 851-858.
- 2005, [Biplots'](#) Variability based on the Procrustes Analysis (with G.H. Hyun and W.J. Woo), Journal of the Korean Data Analysis

Society, 7(6), 1925-1933.

- 2005, [MANCOVA Biplot](#) (with G.H. Hyun and S.M. Jung), The Korean Communications in Statistics, 12(3), 705-712 .
- 2000, The Development of [Biplots](#) System (with G.H. Hyun), The Korean Journal of Applied Statistics, 3(2), 297-306.
- 2000, The Exploratory Approach of Baseline-Category Logit Models using the [Biplot](#) (with K.D. Kim), Journal of Data Sciences and Classifications, 4, 35-47.
- 1998, Robust Weighted Nonmetric [Multidimensional Scaling](#) (with I.T. Jolliffe and J.R. Lee), Journal of Data Sciences and Classifications, 2, 33-47.
- 1995, Resistant [h-plot](#) for a Sample Variance-Covariance Matrix. Journal of The Korean Statistical Society, 24(2), 407-417.