Techniques and Applications of Multivariate Analysis

8. Correspondence Analysis (CRA)

Contents

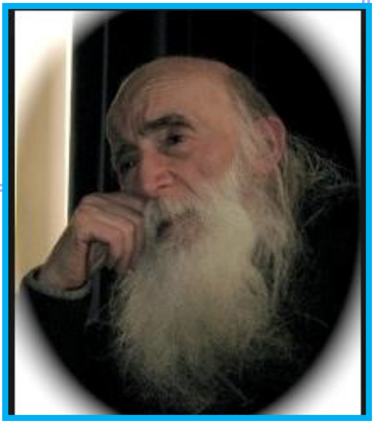
8.1 Comprehension of CRA

- **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 statistics provide tests for the independence or homogeniety bt rows and columns.**

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

- **Applications**
  - **Educational Psychology, Medicine, Survey, Social Science, etc.**
  - **Categorical Data Analysis, Text Mining, Big Data**



Jean-Paul Benzécri is a French statistician and was born in 1932. He is most famous for the development of the Correspondence analysis, a statistical technique for analyzing contingency tables. It is widely used in sociological studies of categorical data. (Wikipedia)

8.1 Comprehension of CRA

Types of CRA : Simple CRA and Multiple CRA

- Simple CRA is the CRA of the two-way contingency table

3 x 3 Table for the preference of the sizes of car in 3 countries

	Large	Medium	Small
American	37	61	30
European	4	20	21
Japanese	2	61	102

O.. = 338

- Multiple CRA is the CRA of the multi-way contingency table

[Data 8.4.1] 2 x 2 x 2 Table for customers

Sex	Age	Height	
		tall	small
Male	old	1	3
	young	2	1
Female	old	1	2
	young	1	1

O... = 12

8.1 Comprehension of CRA

[Example 8.2.1] 4 x 4 Table for Premarital Sex and Ability of Teenage(14~16 ages) Birth Control

- **[Data 8.2.1] Two-Way Contingency Table(presex.txt)**

Ask about sexual relations before marriage

Premarital Sex	Teenage Birth Control			
	매우 반대	반 대	찬 성	매우 찬성
Always wrong	81	68	60	38
Almost wrong	24	26	29	14
Sometimes wrong	18	41	74	42
Not wrong at all	36	57	161	157

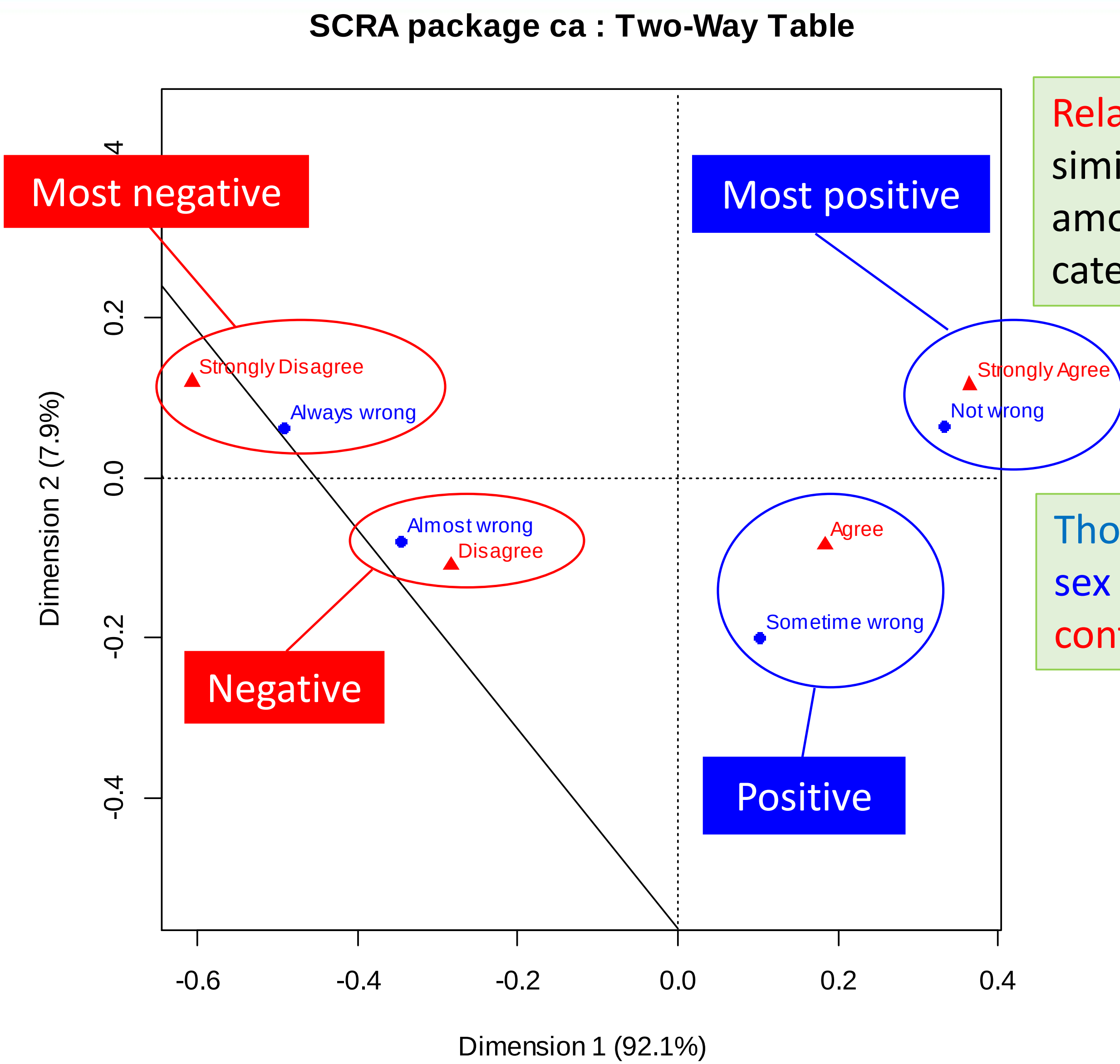
Strongly Disagree ~ Strongly Agree

O.. = 926

[R-code 8.2.2] 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("Always wrong", "Almost wrong", "Sometime wrong", "Not wrong")
colnames(O)<- c("Strongly Disagree", "Disagree", "Agree", "Strongly Agree")
library(ca)
sca=ca(O)
par(pty="s")
plot(sca, main="SCRA package ca : Two-Way Table")
```

8.1 Comprehension of CRA



Relative positions of the points : similarities and differences among rows and columns categories.

Those who consider premarital sex to be bad(wrong) are against contraception in their teens.

8.1 Comprehension of CRA

- **[Example 8.2.3] 5 x 4 Table for Smoking Habits of 193 Staff members**
  - **[Data 8.2.2] Two-Way Contingency Table : data(smoke)**

Staff group
SM: Senior managers
JM: Junior managers
SE: Senior employees
JE: Junior employees
SC: Secretaries

Contingency Table

	none	light	medium	heavy	Sum
SM	4	2	3	2	11
JM	4	3	7	4	18
SE	25	10	12	4	51
JE	18	24	33	13	88
SC	10	6	7	2	25
Sum	61	45	62	25	193

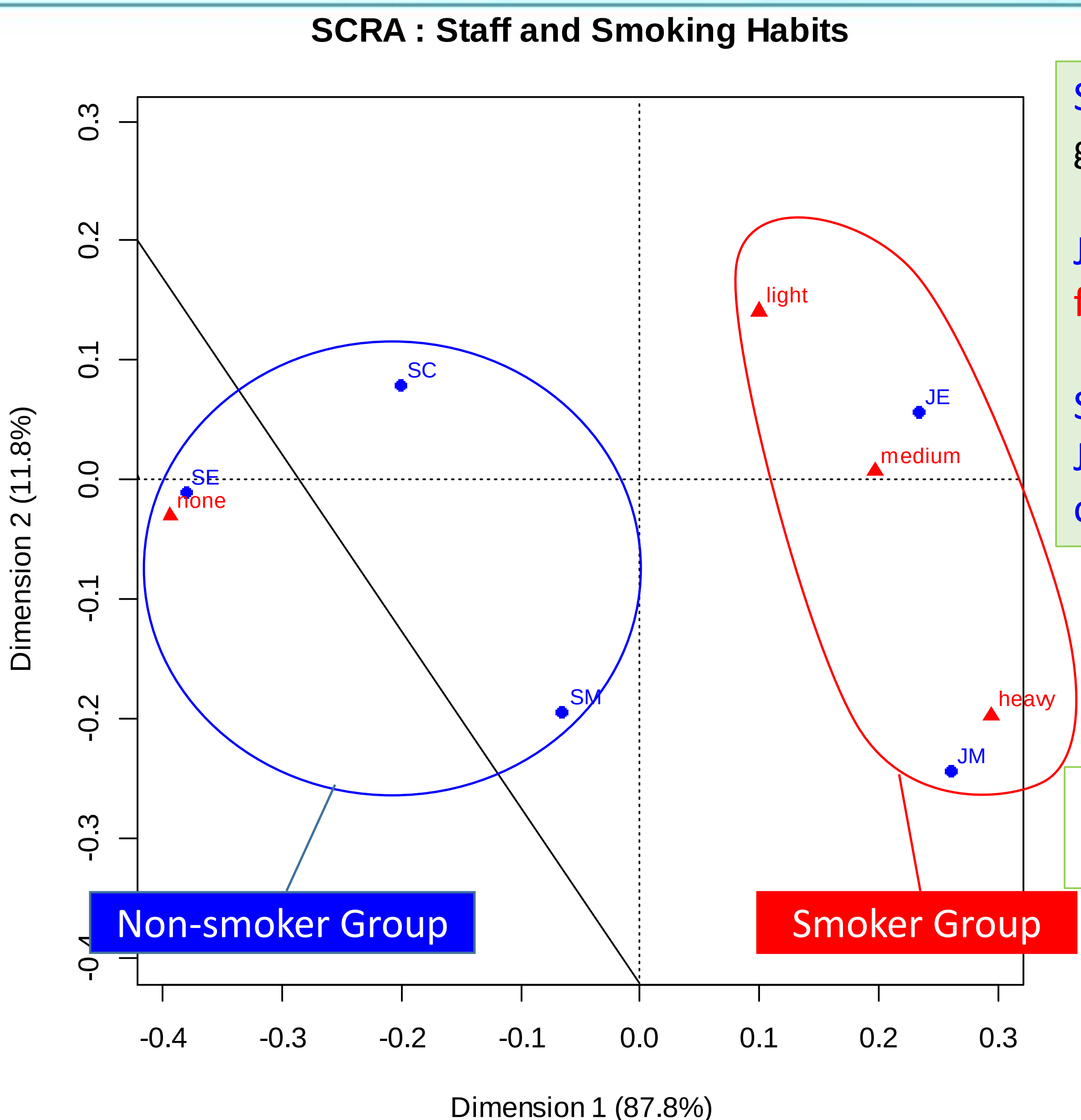
Smoking category
None : does not smoke
Light : smokes 1-10 cigarettes a day
Medium : smokes 11-20 cigarettes a day
Heavy : smokes more than 20 cigarettes a day

$O_{..} = 193$

[R-code 8.2.3] smoke-SCRA.R

```
# Simple CRA : smoker data
library(ca)
data(smoke)
O=smoke
sca=ca(O)
sca
win.graph()
par(pty="s")
plot(sca, main="SCRA : Staff and Smoking Habits")
```

8.1 Comprehension of CRA



SE and SC are relatively similar group corresponding to None.

JM and JE are relatively far from those group.

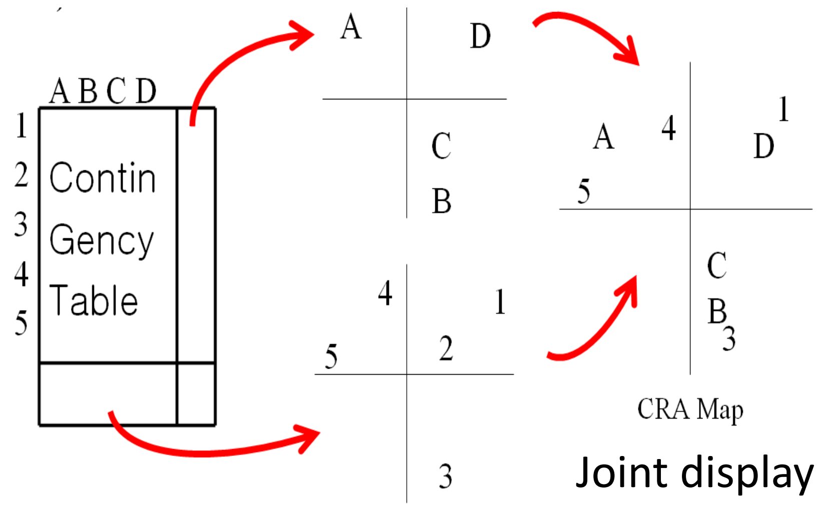
SM is lying almost midway bt JM and SE. Anyway SM correspond to None.

According to staff group, smoking habits are different.

8.1 Comprehension of CRA

Interpretations of CRA Map

- Characters represent the coordinate points of rows and columns categories.
- Distance b.t. two rows (or columns) is weighted Euclidean distance and their closeness.
- Distance b.t. 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.



Joint display

8.1 Comprehension of CRA

Geometric Concept of Simple CRA

(1) Contingency Table Data Matrix : $O = (o_{ij}),\ o_{ij} \geq 0,\ i = 1,\ldots,n\ ;\ j = 1,\ldots,p$

$n \times p$ Contingency Table

Rows Categories	Column Categories				Sum
	1	2	$\cdots$	p	
1	o_{11}	o_{21}	$\cdots$	o_{1p}	$o_{1.}$
2	o_{21}	o_{22}	$\cdots$	o_{2p}	$o_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	o_{n1}	o_{n2}	$\cdots$	o_{np}	$o_{n.}$
Sum	$o_{.1}$	$o_{.2}$	$\cdots$	$o_{.p}$	$o_{..}$

(2) Correspondence Matrix : $F = (f_{ij}),\ f_{ij} = \dfrac{o_{ij}}{o_{..}},\ \ i = 1,\ldots,n\ ;\ j = 1,\ldots,p$

- Total sum : $o_{..} = \sum_{i=1}^{n}\sum_{j=1}^{p} o_{ij} = \mathbf{1}_n^T O \mathbf{1}_p$

$n \times p$ Two-ways Contingency Table

Rows Categories	Column Categories				Sum
	1	2	$\cdots$	p	
1	f_{11}	f_{21}	$\cdots$	f_{1p}	$f_{1.}$
2	f_{21}	f_{22}	$\cdots$	f_{2p}	$f_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	f_{n1}	f_{n2}	$\cdots$	f_{np}	$f_{n.}$
Sum	$f_{.1}$	$f_{.2}$	$\cdots$	$f_{.p}$	$f_{..}$

8.1 Comprehension of CRA

(3) Profiles :

- ith row profile : $\boldsymbol{r}_i = \frac{1}{f_{i\,.}}(f_{i\,1}, \ldots, f_{i\,p})^T = \frac{1}{o_{i.}}(o_{i1}, \ldots, o_{ip})^T$, $i = 1, \ldots, n$

$\Rightarrow$ Data points : $\boldsymbol{r}_1$, ... , $\boldsymbol{r}_n$ $\in WE^p$ by D_c^{-1}(E^p weighted by D_c^{-1})

with $D_c = diag(f_{.1}, \ldots, f_{.p})$: $p \times p$ diagonal matrix

$\Rightarrow$ Centroid : $\boldsymbol{c} = (f_{.1}, \cdots, f_{.p})^T \in WE^p$ by D_c^{-1}

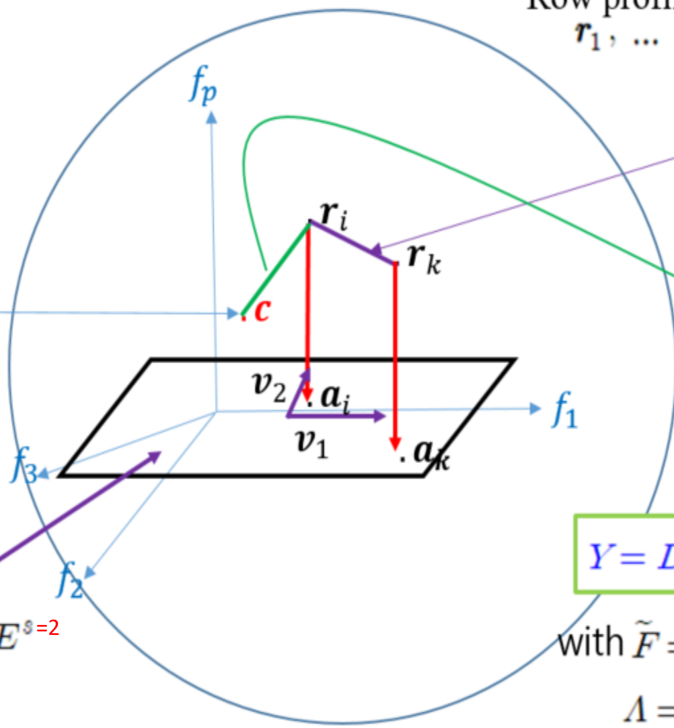
$$= \sum_{i=1}^{n} f_{i.}\boldsymbol{r}_i = \sum_{i=1}^{n} (f_{i1}, \cdots, f_{ip})^T = F^T \mathbf{1}_n$$

Geometric Concept of Simple CRA

Row profiles : $\boldsymbol{r}_1, \dots, \boldsymbol{r}_n \in WE^p$ by D_c^{-1} with $D_c = diag(f_{.1}, \dots, f_{.p})$

Centroid :

$\boldsymbol{c} = (f_{.1}, \cdots, f_{.p})^T$



$Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = WE^{s=2}$

$$d^2(\boldsymbol{r}_i, \boldsymbol{r}_k) = \| \boldsymbol{r}_i - \boldsymbol{r}_k \|^2_{D_c^{-1}} = (\boldsymbol{r}_i - \boldsymbol{r}_k)^T D_c^{-1} (\boldsymbol{r}_i - \boldsymbol{r}_k)$$

$$X^2 = o_{..} \sum_{i=1}^{n} f_{i.} d^2(\boldsymbol{r}_i, \boldsymbol{c}) = o_{..} \sum_{i=1}^{n} f_{i.} \| \boldsymbol{r}_i - \boldsymbol{c} \|^2_{D_c^{-1}}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}}} \right)^2$$

How to find the optimal space WE^s ?

$Y = D_r^{-1/2} \tilde{F} D_c^{-1/2} = U \Lambda V^T$: SVD of residual matrix

with $\tilde{F} = F - \boldsymbol{r}\boldsymbol{c}^T$, $\boldsymbol{c} = F^T \boldsymbol{1}_n$, $\boldsymbol{r} = F \boldsymbol{1}_p$, $D_r = diag(f_{1.}, \dots, f_{n.})$

$\Lambda = diag(\lambda_1, \dots, \lambda_p)$, $U^T U = I_p$, $V^T V = I_p$

Row Coordinates : $A_{(s)} = D_{r(s)}^{-1/2} U_{(s)} \Lambda_{(s)} \in WE^s$ $\Longleftrightarrow$ $\boldsymbol{a}_i, \dots, \boldsymbol{a}_n \in WE^s$

Columns of V define the principal axexs for row coordinates $A_{(s)}$ and the rows of $U_{(s)} \Lambda_{(s)}$ define the coordinates

8.1 Comprehension of CRA

(4) Some Properties of Profiles :

1) Squared chi-squared distance weighted by weights $f_{.j}$, $j = 1, \ldots, p$.

- $d^2(\boldsymbol{r}_i, \boldsymbol{r}_k) = \|\boldsymbol{r}_i - \boldsymbol{r}_k\|^2_{D_c^{-1}} = (\boldsymbol{r}_i - \boldsymbol{r}_k)^T D_c^{-1} (\boldsymbol{r}_i - \boldsymbol{r}_k) = \sum_{j=1}^{p} \frac{(f_{ij}/f_{i.} - f_{kj}/f_{k.})^2}{f_{.j}}$

  : squared chi-squared distance between $\boldsymbol{r}_i$ and $\boldsymbol{r}_k$

- $d^2(\boldsymbol{r}_i, \boldsymbol{c}) = \|\boldsymbol{r}_i - \boldsymbol{c}\|^2_{D_c^{-1}} = (\boldsymbol{r}_i - \boldsymbol{c})^T D_c^{-1} (\boldsymbol{r}_i - \boldsymbol{c})$

  : squared chi-squared distance between $\boldsymbol{r}_i$ and centroid $\boldsymbol{c}$

2) X^2 − test statistics

- $$X^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}}} \right)^2 = o_{..} \sum_{i=1}^{n} f_{i.} \|\boldsymbol{r}_i - \boldsymbol{c}\|^2_{D_c^{-1}}$$

  $$= o_{..} \sum_{i=1}^{n} f_{i.} d^2(\boldsymbol{r}_i, \boldsymbol{c}) = o_{..} \sum_{i=1}^{n} f_{i.} (\boldsymbol{r}_i - \boldsymbol{c})^T D_c^{-1} (\boldsymbol{r}_i - \boldsymbol{c})$$

8.1 Comprehension of CRA

(5) Optimal Space WE^s, $s \ll p$ and Coordinates Matrix :

- $Y = D_r^{-1/2}\widetilde{F}D_c^{-1/2} = U\Lambda V^T$: SVD of residual matrix
  - Centered correspondence matrix : $\widetilde{F} = F - \boldsymbol{rc}^T$
  - Centroid : $\boldsymbol{c} = \boldsymbol{F}^T\boldsymbol{1}_n$, $\boldsymbol{r} = F\boldsymbol{1}_p = (f_{1.}, \cdots, f_{n.})^T$
  - Diagonal matrix : $D_c = diag(f_{.1}, \ldots, f_{.p})$, $D_r = diag(f_{1.}, \ldots, f_{n.})$
  - Diagonal matrix with singula values : $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$
  - $U^TU = I_p$, $V^TV = I_p$

- Row Coordinates : $\widetilde{C}_{R(s)} = D_{r(s)}^{-1/2}U_{(s)}\Lambda_{(s)} \in WE^s$

  Column Coordinates : $\widetilde{C}_{C(s)} = D_{c(s)}^{-1/2}V_{(s)}\Lambda_{(s)} \in WE^s$

- Goodness-of fit : $fit = \dfrac{\sum_{k=1}^{s} l_k}{\sum_{k=1}^{q} l_k} \times 100\%$, $l_k = \lambda_k^2$, $q = \min(n-1, p-1)$.

singular value

principal inertia(eigenvalue)

8.2 Simple CRA

[Table 8.2.2] Algorithm for Simple CRA

[Step 1] $n \times p$ two-way table data matrix: $O = (o_{ij})$,

$o_{ij} \geq 0$, $i = 1, \ldots, n$; $j = 1, \ldots, p$

[Step 2] Correspondence Matrix : $F = (f_{ij})$, $f_{ij} = \dfrac{o_{ij}}{o_{..}}$, $i = 1, \ldots, n$; $j = 1, \ldots, p$

- $o_{..} = \sum_{i=1}^{n} \sum_{j=1}^{p} o_{ij}$: total sum of all elements of O

$\Rightarrow$ Centred correspondence matrix : $\tilde{F} = F - \boldsymbol{r}\boldsymbol{c}^t$

- Row and Column centroids : $\boldsymbol{c} = F^t \mathbf{1}_n$, $\boldsymbol{r} = F\mathbf{1}_p$

[Step 3] SVD of residual matrix : $Y = D_r^{-1/2} \tilde{F} D_c^{-1/2} = U \Lambda V^t$

- $D_r = diag(f_{1.}, \ldots, f_{n.})$, marginal sum of rows $f_{i.} = \sum_{j=1}^{p} f_{ij}$, $i = 1, \ldots, n$
- $D_c = diag(f_{.1}, \ldots, f_{.p})$, marginal sum of columns$f_{.j} = \sum_{i=1}^{n} f_{ij}$, $j = 1, \ldots, p$
- $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$: diagonal matrix with singular values $\lambda_1, \ldots, \lambda_p$
- U and V: orthonormal matrix st $U^t U = I_p$ and $V^t V = I_p$.

[Step 4] Coordinates of rows and columns of Simple CRA Map :

$$\boxed{A = D_r^{-1/2} U \Lambda, \quad B = D_c^{-1/2} V \Lambda}$$

$\Rightarrow$ Coordinates of s-dimensional Simple CRA Map :

$$A_{(s)} = D_{r(s)}^{-1/2} U_{(s)} \Lambda_{(s)}, \quad B = D_{c(s)}^{-1/2} V_{(s)} \Lambda_{(s)}$$

[Step 5] Goodness-of-fit of $s(\geq 2)$-dimensional Simple CRA Map :

$$fit = \frac{\sum_{k=1}^{s} \lambda_k^2}{\sum_{k=1}^{p-1} \lambda_k^2} \times 100$$

8.2 Simple CRA

[R-code 8.2.1] presex-SCRAalgorithm.R

```
# Simple CRA : 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)
F <- O/sum(O)
         r <- apply(F,1,sum); c <- apply(F,2,sum)
         Dr<- diag(1/sqrt(r)); Dc<- diag(1/sqrt(c))
r;c;Dr;Dc
cF<- F-r%*%t(c); Y <- Dr%*%(cF)%*%Dc
svd.Y <- svd(Y)
         U <- svd.Y$u; V <- svd.Y$v; D <- diag(svd.Y$d)
A <- (Dr%*%U%*%D)[,1:2]; B <- (Dc%*%V%*%D)[,1:2]
rownames(A) <- c("Always wrong", "Almost wrong", "Sometime wrong", "Not wrong")
rownames(B) <- c("Strongly Disagree", "Disagree", "Agree", "Strongly Agree")
A;B
eig <- (svd.Y$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(A))
         plot(B[, 1:2], xlab="Dim1",ylab="Dim2",xlim=lim,ylim=lim,pch=15,col=2, main="SCRA
Algorithm : Two-Way Table")
         text(B[, 1:2],rownames(B),cex=0.8,col=2,pos=3)
         points(A[, 1:2],pch=16, col=4)
         text(A[, 1:2],rownames(A),cex=0.8,pos=3, col=4)
         abline(v=0,h=0)
```

8.3 Independence and Homogeneity in CRA

Probability of (I, j) cell

Independence : $H_0 \ : \ \pi_{ij} = \pi_{i.} \times \pi_{.j}$ Test $X^2 = \sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{o_{ij}-e_{ij}}{\sqrt{e_{ij}}}\right)^2 \sim X^2_{df=(n-1)(p-1)}$

$\frac{\pi_{1j}}{\pi_{1.}} = \cdots = \frac{\pi_{nj}}{\pi_{n.}}, \ j=1, \ ..., \ p$: Homogeneity of rows

test

$\frac{\pi_{i1}}{\pi_{.1}} = \cdots = \frac{\pi_{ip}}{\pi_{.p}}, \ \ i=1, \ ..., \ n$: Homogeneity of columns

- **[Example 8.2.1] 4 x 4 Table for Premarital Sex and Ability of Teenage Birth Control**

Ho : Premarital Sex and Ability of Teenage Birth Control are **independent**

⟺ Ho : There is no association b.t. Premarital Sex and Ability of Teenage Birth Control

⟺ Ho : Premarital Sex categories are homogenous

Ho : Ability of Teenage Birth Control categories are homogenous

Chi-square test : X-squared = 128.6836, df = 9, p-value < 2.2e-16 ⬅ > chisq.test(O)

8.4 Multiple CRA

✓**Multiple CRA** is the CRA of the multi-way contingency table or Burt matrix

- **[Example 8.4.1] and [Example 8.4.2]**

[Data 8.4.1] Customers Data : O.. = 12

Sex	Age	Height	
		Tall	Short
Male	Old	1	3
	Young	2	1
Female	Old	1	2
	Young	1	1

[Data 8.4.2] Driver Injury Data : O... = 82,769

Condition	Belt	Injury			
		None	Minimal	Minor	Major
Normal	Yes	12500	604	344	38
	No	61971	3519	2272	237
Drinking	Yes	313	43	15	4
	No	3992	481	370	66

8.4 Multiple CRA

Relations of Table, Indicator Matrix and Burt Matrix

Sex	Age	Height	
		Tall	Short
Male	Old	1	3
	Young	2	1
Female	Old	1	2
	Young	1	1

Indicator Matrix

$$Z = \left[\, Z_1 \mid Z_2 \mid Z_3 \,\right]$$

Male	Female	Old	Young	Tall	Short
1	0	1	0	1	0
1	0	1	0	0	1
1	0	1	0	0	1
1	0	1	0	0	1
1	0	0	1	1	0
1	0	0	1	1	0
1	0	0	1	0	1
0	1	1	0	1	0
0	1	1	0	0	1
0	1	1	0	0	1
0	1	0	1	1	0
0	1	0	1	0	1

$$B = Z^t Z = \begin{pmatrix} Z_1^t Z_1 & Z_1^t Z_2 & Z_1^t Z_3 \\ Z_2^t Z_1 & Z_2^t Z_2 & Z_2^t Z_3 \\ Z_3^t Z_1 & Z_3^t Z_2 & Z_3^t Z_3 \end{pmatrix}$$

Burt Matrix

	Male	Female	Old	Young	Tall	Short
Male	7	0	4	3	3	4
Female	0	5	3	2	2	3
Old	4	3	7	0	2	5
Young	3	2	0	5	3	2
Tall	3	2	2	3	5	0
Short	4	3	5	2	0	7

8.4 Multiple CRA

Driver Condition	Seat Belt Usages	Driver Injury Level			
		None	Minimal	Minor	Major
Normal	Yes	12500	604	344	38
	No	61971	3519	2272	237
Drinking	Yes	313	43	15	4
	No	3992	481	370	66

$$Z = \left[\, Z_1 \mid Z_2 \mid Z_3 \,\right]$$

O.. = 82,769

$B =$

	음주	정상	미착용	착용	중상	중경상	경상	없음
음주	5284	0	4909	375	70	524	385	4305
정상	0	81485	67999	13486	275	4123	2616	74471
미착용	4909	67999	72908	0	303	4000	2642	65963
착용	375	13486	0	13861	42	647	359	12813
중상	70	275	303	42	345	0	0	0
중경상	524	4123	4000	647	0	4647	0	0
경상	385	2616	2642	359	0	0	3001	0
없음	4305	74471	65963	12813	0	0	0	78776

Algorithm for MCRA Map

[Table 8.4.1] Algorithm for Multiple CRA

[Step 1] $n \times q$ Indicator Matrix : $\boxed{Z = [\, Z_1 \mid Z_2 \mid \cdots \mid Z_p \,]}$,

- Z_j : $n \times q_j$, $j = 1, \ldots, p$
- $q = q_1 + q_2 + \cdots + q_p$

[Step 2] $q \times q$ Burt Matrix : $\boxed{B = Z^T Z = (b_{ij})}$, $i, j = 1, \ldots, q$

[Step 3] Correspondence Matrix : $\boxed{P = (p_{ij}),\ p_{ij} = \dfrac{b_{ij}}{b_{..}}}$, $i, j = 1, \ldots, q$

- $b_{..} = \sum_{i=1}^{q} \sum_{j=1}^{q} b_{ij}$: total sum of all elements of B
- Row and Column Centroids : $\boldsymbol{c} = P^T \mathbf{1}_q$

$\Rightarrow$ • Centred correspondence matrix : $\widetilde{P} = P - \boldsymbol{c}\boldsymbol{c}^T$

[Step 4] SVD of residual matrix : $\boxed{Y = D^{-1/2} \widetilde{P} D^{-1/2} = U \Lambda V^T}$

- $D = diag(\boldsymbol{c}) = diag(p_{.1}, \ldots, p_{.q})$ with $p_{.j} = \sum_{i=1}^{q} p_{ij}$, $j = 1, \ldots, q$
- $\Lambda = diag(\lambda_1, \ldots, \lambda_q)$: diagonal matrix with singular values $\lambda_1, \ldots, \lambda_p$
- U and V : ortho-normal matrices st $U^t U = I_q$ and $V^t V = I_q$.

Algorithm for MCRA Map

[Step 5] Coordinates of MCRA Map : $\boxed{A_B = D_c^{-1/2} V \Lambda}$

$\Rightarrow$ Coordinates of s-dimensional MCRA Map : $A_{B(s)} = D_{c(s)}^{-1/2} V_{(s)} \Lambda_{(s)}$

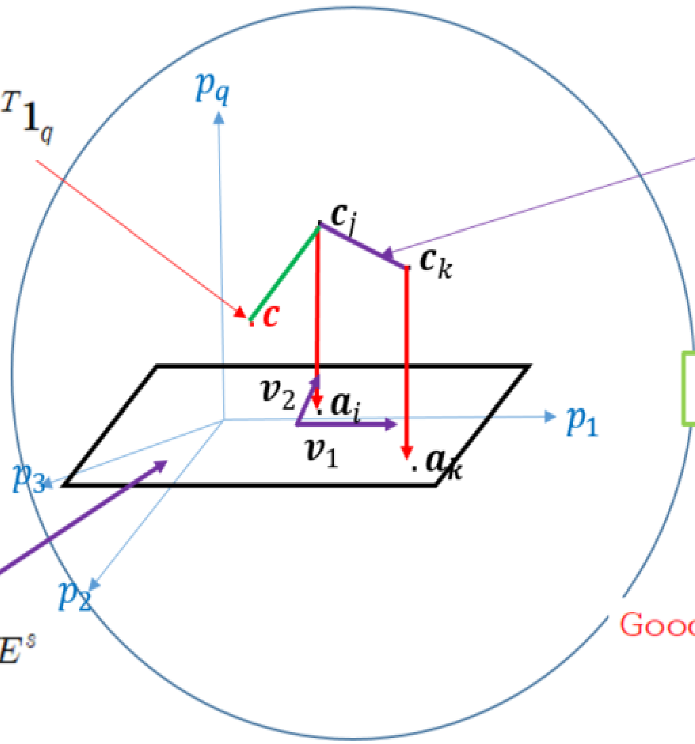
[Step 6] Goodness-of-fit of $s(\geq 2)$-dimensional MCRA Map :

$$\boxed{fit = \frac{\sum_{k=1}^{s} \lambda_k^2}{\sum_{k=1}^{q} \lambda_k^2} \times 100}$$

Geometric Concept of MCRA

Column profiles : $\boldsymbol{c}_j = \frac{1}{p_{.j}}(p_{1j},\dots,p_{qj})^T = \frac{1}{b_{.j}}(b_{1j},\dots,b_{qj})^T,\ j=1,\dots,q\ \in WE^q$ by D^{-1}
with $D = diag(\mathbf{c}) = diag(p_{.1},\dots,p_{.q})$

Centroid : $\boldsymbol{c} = P^T \mathbf{1}_q$



$d^2(\boldsymbol{c}_j, \boldsymbol{c}_k) = (\boldsymbol{c}_j - \boldsymbol{c}_k)^T D^{-1}(\boldsymbol{c}_j - \boldsymbol{c}_k)$

$Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = WE^s$

How to find the optimal space WE^s ?

$Y = D^{-1/2}\widetilde{P}D^{-1/2} = U\Lambda V^T$: SVD of residual matrix

with $\widetilde{P} = P - \boldsymbol{c}\boldsymbol{c}^T,\ \boldsymbol{c} = P^T\mathbf{1}_q,\ D = diag(p_{.1},\dots,p_{.q}),$

$\Lambda = diag(\lambda_1,\dots,\lambda_q),\ U^t U = I_q,\ V^t V = I_q$

Goodness-of-fit of $s(\geq 2)$-dimensional MCRA Map :

$$fit = \frac{\sum_{k=1}^{s}\lambda_k^2}{\sum_{k=1}^{q}\lambda_k^2} \times 100$$

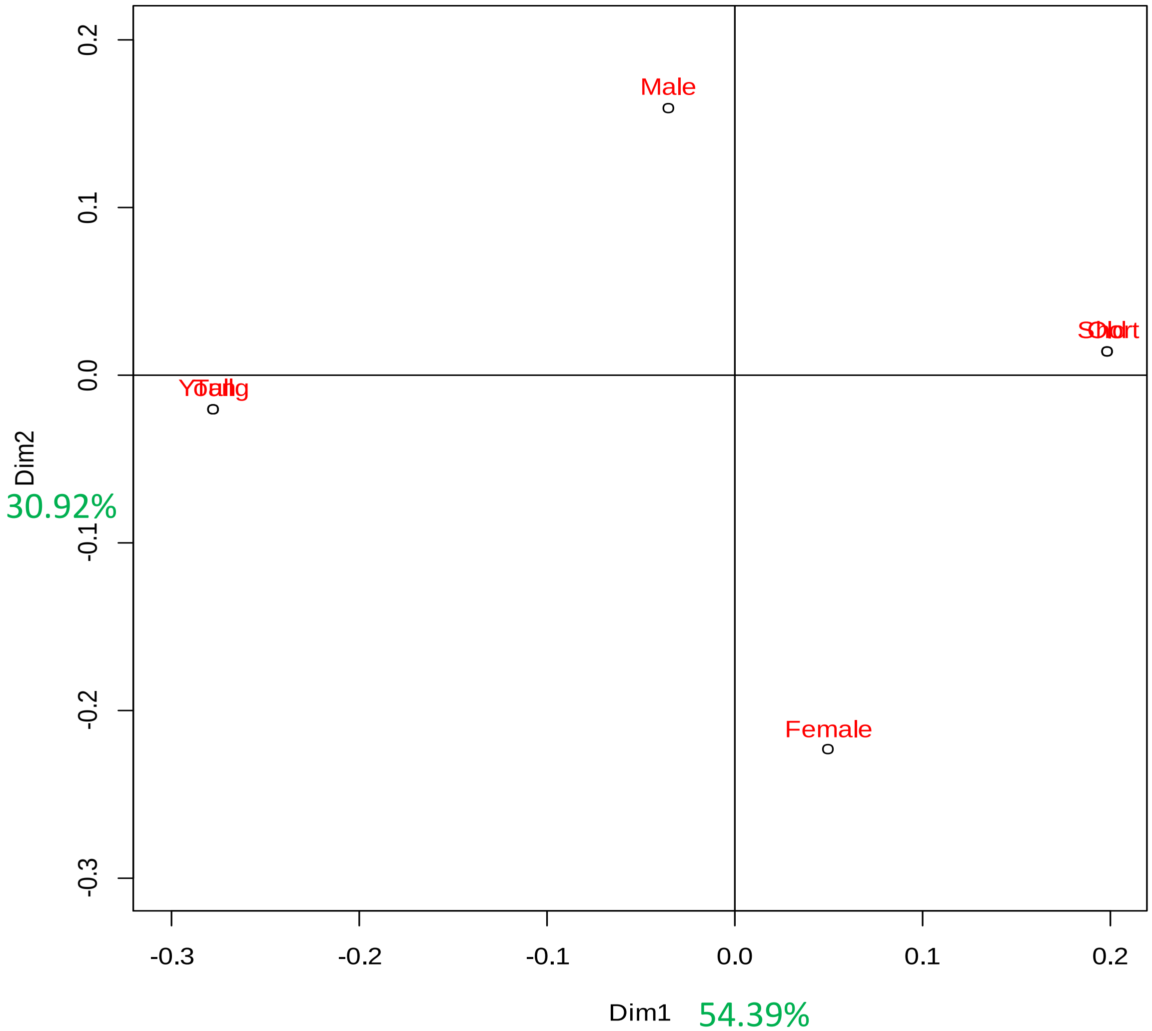
Row Coordinates : $A_B = D_c^{-1/2} V\Lambda \in WE^s \Longleftrightarrow \boldsymbol{a}_i,\dots,\boldsymbol{a}_q \in WE^s$

8.4 Multiple CRA

- **[Example 8.4.1] 2 x 2 x 2 Table for Customers**

GOF = 85.31%

MCRA algorithm : Burt Matrix




```
> B
       Male Female Old Young Tall Short
Male      7      0   4     3    3     4
Female    0      5   3     2    2     3
Old       4      3   7     0    2     5
Young     3      2   0     5    3     2
Tall      3      2   2     3    5     0
Short     4      3   5     2    0     7
> P <- B / sum(B)
> cm <- apply(P, 2, sum)
> Dc<-diag(1/sqrt(cm))
> eP <- cm %*% t(cm)
> Y <- (P - eP) / sqrt(eP)
> ## Singular Value Decomposition
> svd.Y<-svd(Y)
> V<-svd.Y$v
> Dl<-diag((svd.Y$d)^2)
> lam<-(svd.Y$d)^2
> fit<-lam/sum(lam)*100
> rbind(round(lam, 3),round(fit, 3))
       [,1]   [,2]   [,3] [,4] [,5] [,6]
[1,]  0.193  0.110  0.052    0    0    0
[2,] 54.386 30.923 14.690    0    0    0
> Cb<- Dc%*%V%*%Dl
> rownames(Cb)<-colnames(Z)
> Cb2=Cb[, 1:2]
> Cb2
              [,1]        [,2]
Male   -0.03553638  0.15971843
Female  0.04975093 -0.22360581
Old     0.19862898  0.01428748
Young  -0.27808057 -0.02000247
Tall   -0.27808057 -0.02000247
Short   0.19862898  0.01428748
```

8.4 Multiple CRA

- **[Example 8.4.3] Burt Table for Demographic and Socioeconomic Variables**

- Living conditions and aspiration of the French
- 1000 subsamples over 18 years

[Data 8.4.3](sociecono.txt)

	성별		교육수준					주거상태				주식		동산		나이					거주지규모				
	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

Sex : Male, Female SEXM SEXF

Level of education :
None/Grade School/Some High School/High school/College:
EDU1 ~ EDU5

Home ownership :
Mortgage/Owner/Tenant/Rent Free :
LOD1 ~ LOD4

Ownership of stocks and shares :
Yes/No : STO1~ STO2

Ownership of real estate :
Yes/No : HOU1 ~ HOU2

Age :
<25, 25-34, 35-49, 50-64, 65< : AGE1 ~ AGE5

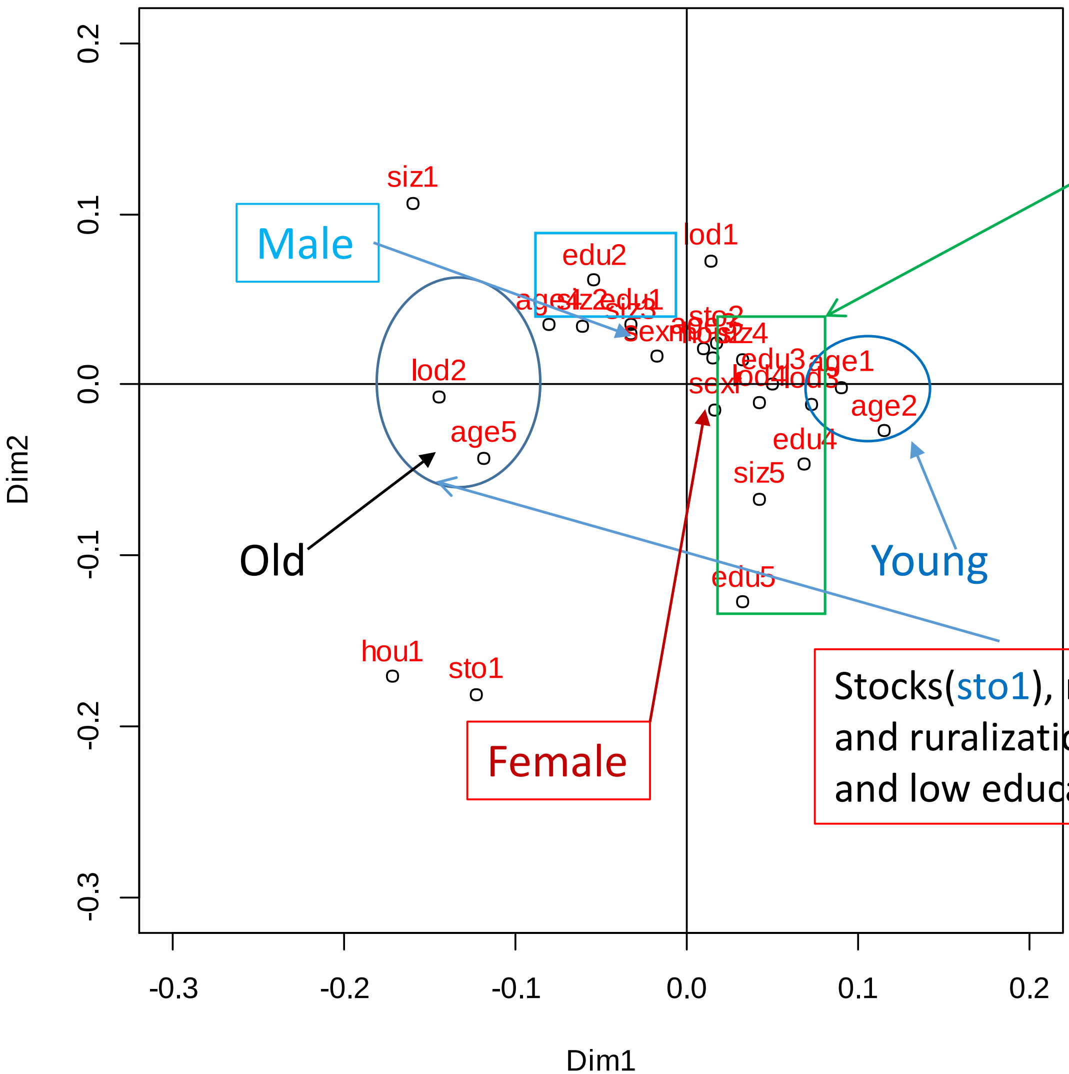
Size of town (Number of Inhabitants):
<2,000, 2,000-20,000, 50,000-100,000, 100,000< :
SIZ1 ~ SIZ5

Ruralization ⟷ Urbanization

8.4 Multiple CRA

GOF = 27.39%

MCRA algorithm : Burt Matrix



High educations(edu3-edu5) are related with rather young(age1-age3) and urban part(siz4-siz5).

Stocks(sto1), real estate(hou1) homeownership(lod2) and ruralization(siz1) are related with old ages(4-5) and low educations(edu1-edu2).

8.4 Multiple CRA

[Example 8.4.4] Input Data : Three-Way Table

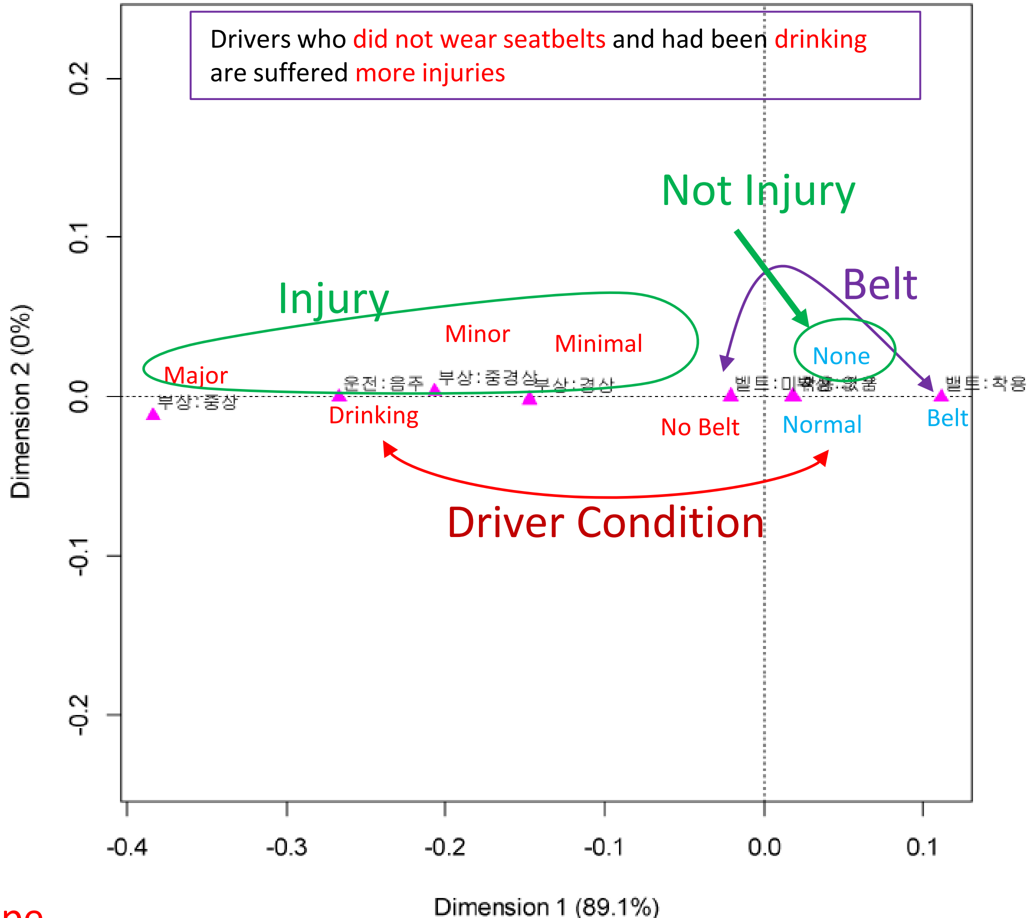
Condition 운전상태	Belt 벨트	Injury 부상			
		None 없음	Minimal 경상	Minor 중경상	Major 중상
Normal 정상	Yes 착용	12500	604	344	38
	No 미착용	61971	3519	2272	237
Drinking 음주	Yes 착용	313	43	15	4
	No 미착용	3992	481	370	66

운전	벨트	부상	빈도
정상	착용	없음	12500
정상	착용	경상	604
정상	착용	중경상	344
정상	착용	중상	38
정상	미착용	없음	61971
정상	미착용	경상	3519
정상	미착용	중경상	2272
정상	미착용	중상	237
음주	착용	없음	313
음주	착용	경상	43
음주	착용	중경상	15
음주	착용	중상	4
음주	미착용	없음	3992
음주	미착용	경상	481
음주	미착용	중경상	370
음주	미착용	중상	66

[Data 8.4.4] accidenttable.txt : Cell position type

[R-code 8.4.4] accidenttable-MCRA.R

```
acdtable<-read.table("accidenttable.txt", header=T)
table=xtabs(빈도 ~ 벨트+부상+운전, data=acdtable)
mjca(table)
plot(mjca(table), main="MCRA: 삼원분할표")
```



8.5 MCRA of Classification Variables Data

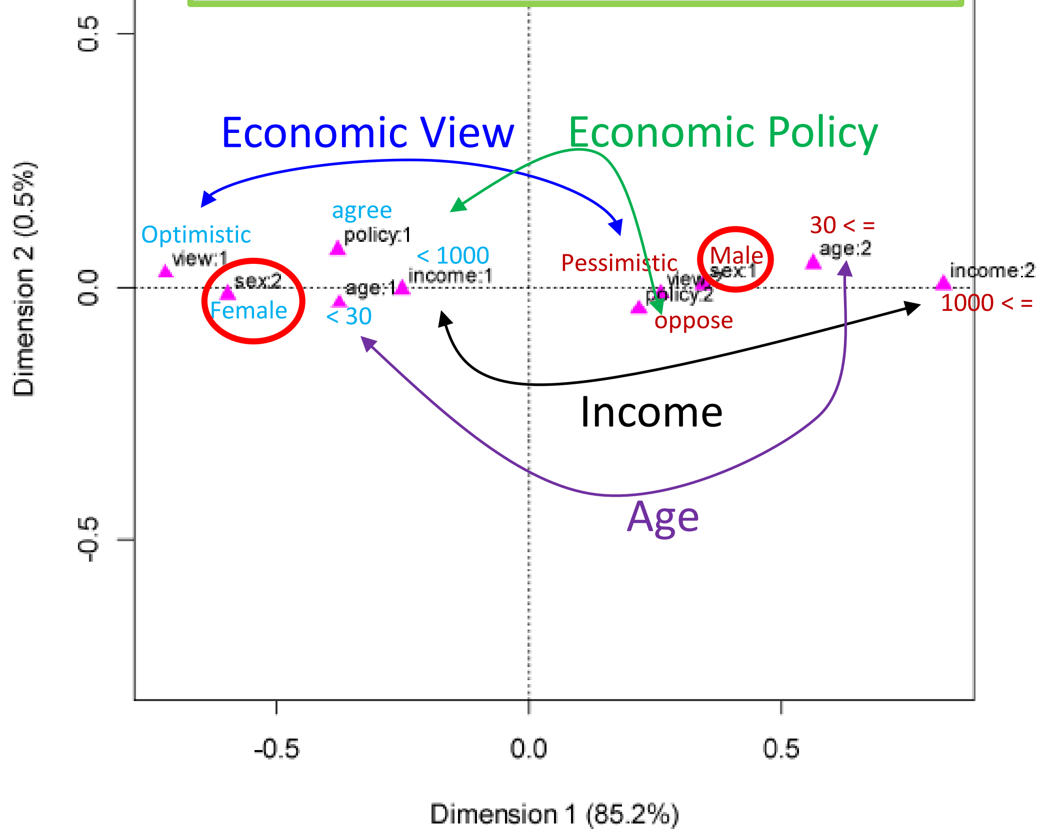
- **[Example 8.5.1] Economic View and Policy Preference**

•120 subsamples over 19 years with questions for Survey

[Data 8.5.1](view.txt) : Coding type of Survey Data

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
		------중간생략------			
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

Sex : 1(Male), 2(Female)
Age : 1(< 30), 2(30< =)
Income : 1(< 1000만), 2(1000만 < =)
Economic View : 1(Optimistic), 2(Pessimistic)
Economic Policy : 1(agree), 2(oppose)



8.6 R for CRA : Practice Time

R-Code:

Simple CRA	Multiple CRA
library(ca) ca()	library(ca) mjca()
SCRA Algorithm : Two-Way Table	MCRA Algorithm : Z: Indicator Matrix B: Burt Table Classification Variables Data Three-Way Table

8.6 R for CRA : Practice Time

R-code list of Chapter 8 Correspondence Analysis

presex-SCRAalgorithm.R	[R-코드 8.2.1]	결혼 전 성관계에 대한 반응과 피임허용의 이원분할표의 단순 CRA 수행단계
presex-SCRA.R	[R-코드 8.2.2]	[R-코드 8.2.1]에서 R의 함수 ca()의 활용
smoke-SCRA.R	[R-코드 8.2.3]	지위에 따른 흡연습관에 대한 이원분할표의 단순 CRA
customer-MCRAalgorithm.R	[R-코드 8.4.1]	성별, 나이, 키에 따른 삼원분할표의 다중 CRA 알고리즘
accident-MCRAalgorithm.R	[R-코드 8.4.2]	사고운전자 부상정도의 삼원분할표에 대한 다중 CRA 알고리즘
sociecono-MCRAalgorithm.R	[R-코드 8.4.3]	인구통계적 변인과 사회경제적 변인에 의한 버트 행렬에 대한 다중 CRA 알고리즘
accidenttable-MCRA.R	[R-코드 8.4.4]	삼원분할표 입력자료에 대한 다중 CRA
sucidefreq-MCRA.R	[R-코드 8.4.5]	[자료 1.3.6]의 삼원분할표에 대한 다중 CRA
view-MCRA.R	[R-코드 8.5.1]	[자료 8.5.1] 5가지 범주의 분류변수 자료의 다중 CRA
ptsd-MCRA.R	[R-코드 8.5.2]	외상 후 스트레스 증후군에 대한 다시점자료의 다중 CRA

8.6 R for CRA : Practice Time

[R-code 8.2.1] SCRA Algorithm for Premarital Sex and Ability of Teenage Birth Control(presex-SCRAalgorithm.R)

```
# Simple CRA : 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)
F <- O/sum(O)
        r <- apply(F,1,sum)
        c <- apply(F,2,sum)
        Dr<- diag(1/sqrt(r))
        Dc<- diag(1/sqrt(c))
r;c;Dr;Dc
cF<- F-r%*%t(c)
Y <- Dr%*%(cF)%*%Dc
svd.Y <- svd(Y)
        U <- svd.Y$u
        V <- svd.Y$v
        D <- diag(svd.Y$d)
A <- (Dr%*%U%*%D)[,1:2]
B <- (Dc%*%V%*%D)[,1:2]
        rownames(A) <- c("매우나쁨", "거의나쁨", "조금나쁨", "전혀안나쁨")
        rownames(B) <- c("매반", "반대", "찬성", "매찬")
A;B
eig <- (svd.Y$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(A))
        plot(B[, 1:2], xlab="Dim1",ylab="Dim2",xlim=lim,ylim=lim,pch=15,col=2,
        main="SCRA Algorithm : 이원분할표")
        text(B[, 1:2],rownames(B),cex=0.8,col=2,pos=3)
        points(A[, 1:2],pch=16, col=4)
        text(A[, 1:2],rownames(A),cex=0.8,pos=3, col=4)
        abline(v=0,h=0)
```

presex-SCRA.R

```
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")
```

[Step 1]

[Step 2]

[Step 3]

[Step 4]

[Step 5]

8.6 R for CRA : Practice Time

[R-code 8.4.1] MCRA Algorithm for Customers Data(customer-MCRAalogrithm.R)

```
# MCRA : Burt Table from Indicator Matrix
Z<-matrix(c(1, 0, 1, 0, 1, 0,
            1, 0, 1, 0, 0, 1,
            1, 0, 1, 0, 0, 1,
            1, 0, 1, 0, 0, 1,
            1, 0, 0, 1, 1, 0,
            1, 0, 0, 1, 1, 0,
            1, 0, 0, 1, 0, 1,
            0, 1, 1, 0, 1, 0,
            0, 1, 1, 0, 0, 1,
            0, 1, 1, 0, 0, 1,
            0, 1, 0, 1, 1, 0,
            0, 1, 0, 1, 0, 1), byrow=T, nrow=12)
colnames(Z)<-c("남", "여", "중장년", "청소년", "키큼", "키작음")
Z
B <- t(Z) %*% Z
B
P <- B / sum(B)
cm <- apply(P, 2, sum)
Dc<-diag(1/sqrt(cm))
eP <- cm %*% t(cm)
Y <- (P - eP) / sqrt(eP)
```

```
## Singular Value Decomposition
svd.Y<-svd(Y)
V<-svd.Y$v
Dl<-diag((svd.Y$d)^2)
lam<-(svd.Y$d)^2
fit<-lam/sum(lam)*100
rbind(round(lam, 3),round(fit, 3))
Cb<- Dc%*%V%*%Dl
rownames(Cb)<-colnames(Z)
Cb2=Cb[, 1:2]
Cb2
par(pty="s")
lim<-range(pretty(Cb))
plot(Cb2, xlab="Dim1", ylab="Dim2", xlim=lim, ylim=lim, main="MCRA algorithm :
Burt Matrix")
text(Cb2, colnames(Z), pos=3, col=1)
abline(v=0, h=0)
```

8.6 R for CRA : Practice Time

[R-code 8.4.2] MCRA Algorithm for Driver Injury Data(accident-MCRAalogrithm.R)

```
accident<-read.table("accident.txt", head=T)
B<-as.matrix(accident)
rownames(B)<-colnames(B)
B
P <-B / sum(B)
cm <-apply(P, 2, sum)
Dc<-diag(1/sqrt(cm))
eP <-cm %*% t(cm)
Y <-(P -eP) / sqrt(eP)

## Singular Value Decomposition
svd.Y<-svd(Y)
V<-svd.Y$v
Dl<-diag((svd.Y$d)^2)
lam<-(svd.Y$d)^2
fit<-lam/sum(lam)*100
rbind(round(lam, 3), round(fit, 3))
Cb<-Dc%*%V%*%Dl
rownames(Cb)<-colnames(B)
Cb2=Cb[, 1:2]
Cb2
limy<-range(pretty(Cb2))
limx<-c(-1, 1)
plot(Cb2, xlab="Dim1", ylab="Dim2", xlim=limx, ylim=limy, main="MCRA algorithm: Burt Matrix")
text(Cb2, rownames(Cb), col=1, pos=3)
```

accident - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

음주	정상	미착용	착용	중상	중경상	경상	없음
5284	0	4909	375	70	524	385	4305
0	81485	67999	13486	275	4123	2616	74471
4909	67999	72908	0	303	4000	2642	65963
375	13486	0	13861	42	647	359	12813
70	275	303	42	345	0	0	0
524	4123	4000	647	0	4647	0	0
385	2616	2642	359	0	0	3001	0
4305	74471	65963	12813	0	0	0	78776

accidenttable - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

운전	벨트	부상	빈도
정상	착용	없음	12500
정상	착용	경상	604
정상	착용	중경상	344
정상	착용	중상	38
정상	미착용	없음	61971
정상	미착용	경상	3519
정상	미착용	중경상	2272
정상	미착용	중상	237
음주	착용	없음	313
음주	착용	경상	43
음주	착용	중경상	15
음주	착용	중상	4

accidenttable-MCRA.R

```
acdtable<-read.table("accidenttable.txt", header=T)
table=xtabs(빈도 ~ 벨트+부상+운전, data=acdtable)
mjca(table)
plot(mjca(table), main="MCRA: 삼원분할표")
```

8.6 R for CRA : Practice Time

[R-code 8.4.1] MCRA Algorithm for Customers Data(customer-MCRAalogrithm.R)

```
# MCRA : Burt Table from Indicator Matrix
Z<-matrix(c(1, 0, 1, 0, 1, 0,
            1, 0, 1, 0, 0, 1,
            1, 0, 1, 0, 0, 1,
            1, 0, 1, 0, 0, 1,
            1, 0, 0, 1, 1, 0,
            1, 0, 0, 1, 1, 0,
            1, 0, 0, 1, 0, 1,
            0, 1, 1, 0, 1, 0,
            0, 1, 1, 0, 0, 1,
            0, 1, 1, 0, 0, 1,
            0, 1, 0, 1, 1, 0,
            0, 1, 0, 1, 0, 1), byrow=T, nrow=12)
colnames(Z)<-c("남", "여", "중장년", "청소년", "키큼", "키작음")
Z
B <- t(Z) %*% Z
B
P <- B / sum(B)
cm <- apply(P, 2, sum)
Dc<-diag(1/sqrt(cm))
eP <- cm %*% t(cm)
Y <- (P - eP) / sqrt(eP)
```

```
## Singular Value Decomposition
svd.Y<-svd(Y)
V<-svd.Y$v
Dl<-diag((svd.Y$d)^2)
lam<-(svd.Y$d)^2
fit<-lam/sum(lam)*100
rbind(round(lam, 3),round(fit, 3))
Cb<- Dc%*%V%*%Dl
rownames(Cb)<-colnames(Z)
Cb2=Cb[, 1:2]
Cb2
par(pty="s")
lim<-range(pretty(Cb))
plot(Cb2, xlab="Dim1", ylab="Dim2", xlim=lim, ylim=lim, main="MCRA algorithm :
Burt Matrix")
text(Cb2, colnames(Z), pos=3, col=1)
abline(v=0, h=0)
```