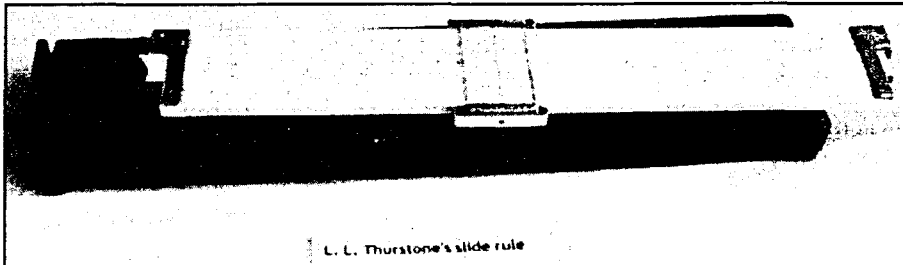


튜토리얼

Scaling of Psychological Attribution : MDS를 중심으로
김청택(서울대 심리학과)




L. L. Thurstone's slide rule

Scaling of Psychological Attribution - MDS를 중심으로 -


김창택
서울대학교
심리학과 / 인지과학협동과정

Scaling

- ❖ 척도란 인간의 심리적 속성을 측정하기 위한 도구이다. 예컨대 물리적인 속성인 길이를 잴 때 자가 필요하듯이 심리적 속성 지능을 측정하기 위해서는 심리적 척도가 필요하다.
- ❖ 여기서 측정(measurement)이란 전통적으로 “특정한 규칙에 따라서 대상이나 사건 혹은 속성에 대해 숫자를 할당하는 것”이다.
 - 키를 측정한다고 할 때 머리끝부터 발끝까지의 길이라는 속성에 대하여 미터 단위로 숫자를 할당하는 것이다.
 - 마찬가지로 공격성을 측정한다고 하면, 공격성에 어떤 숫자를 할당해야 한다.



물리적 특성을
측정하는 척도



심리적 특성을
측정하는 척도

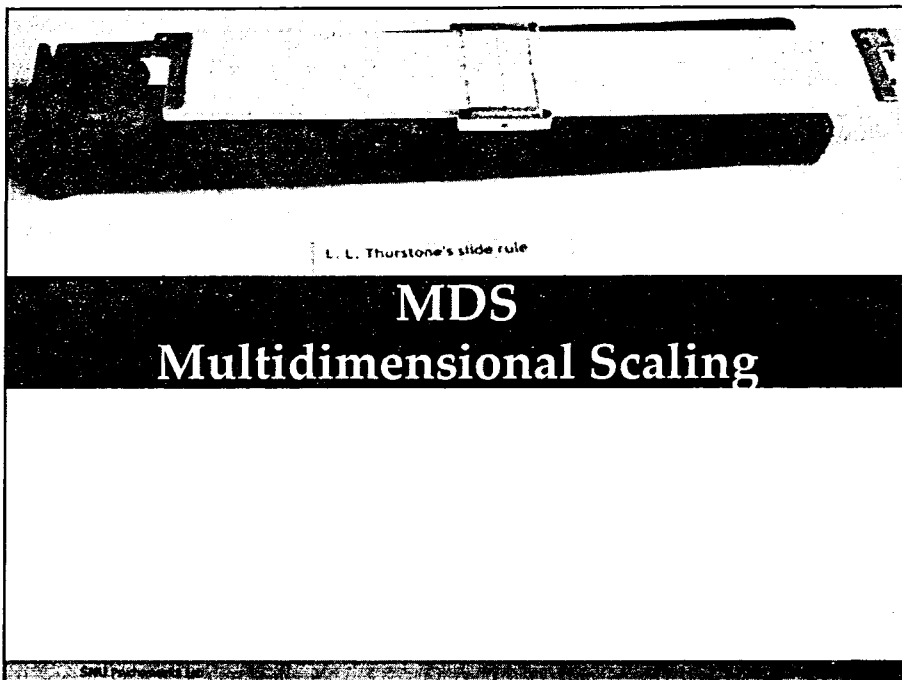
SNU Psychometrics Lab 2005-12-05

측정의 방법

- ❖ 심리적 속성에 대한 척도는 다음과 같은 문제가 있다. (예컨대 “공격성”이라는 속성에 대하여 생각해보자)
 - 공격성에 숫자를 부여하였을 때, 그 숫자에 대한 해석이 어렵다. 즉 공격성에 해당하는 객관적 지표가 잘 존재하지 않는다.
 - Likert Scale을 사용하여 공격성 문항에 1,2,3,4,5의 숫자를 부여하였을 때, 그 숫자가 어떤 scale인지, 즉 ordinal scale인지 interval scale인지를 확인하기 힘들다.
- ❖ 심리척도의 개발에 있어서 문항들이나 **composite score**에 해당하는 심리적인 크기를 반영할 수 있는 **scale**을 개발하려 시도한다.
- ❖ 이때 심리적인 크기는 항상 상대적인 크기이다. 절대적인 크기는 있을 수 없다.
- ❖ 여러 대상에게 여러 문항에 대한 질문을 함으로써 대상들간의 관계와 문항들간의 관계를 파악할 수 있다.

Scale의 종류

- ❖ Scale은 다음과 같은 네 가지로 분류할 수 있다.
 - 명명척도 (nominal scale)
 - 다른 숫자가 다른 대상/속성을 지칭하고 있음을 의미
 - 성별 (1/2), 번호판
 - 가능한 계산, 통계치: counting, frequency analysis
 - 서열척도 (ordinal scale)
 - 숫자의 순서만 의미를 지님, 차이나 비율은 의미가 없음
 - 학교 등수, Likert Scale 점수
 - Order statistics
 - 등간척도 (interval scale)
 - 간격이 의미가 있으나 비율은 의미가 없음
 - 절대 영점이 없음
 - 온도, IQ
 - 비율척도 (ratio scale)
 - 유리수의 모든 속성을 지니고 있음
 - 키, 몸무게 등, 심리척도에는 비율척도가 없음
- ❖ Scaling Method에서 가장 많이 다루어지는 것은 서열척도를 이용하여 등간척도의 의미를 가지는 새로운 scaling을 만들어 내는 것임



MDS

- ❖ 여러 대상들의 (예, 상품, 개인) 유사성(비유사성) 정보를 이용하여 대상들을 축소된 다차원 공간 속에 표상하는 방법
- ❖ 두 가지 유형의 MDS기법
 - Metric MDS: 유사성 정보가 interval scale의 의미를 가지고 있을 때 사용
 - Nonmetric MDS: 유사성 정보가 ordinal scale의 의미만을 가지고 있을 때 사용
- ❖ Nonmetric MDS는 Shepard (1962)에 의하여 개발, Kruskal (1964)에 의하여 통계적으로 정립되고, Carrol & Chang (1970)에 의해 발전된 기법
 - Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 27, 125-140, 219-246.
 - Kruskal, J. B. (1964). Multidimensional Scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika*, 29, 1-27.
 - Kruskal, J. B. (1964). Nonmetric Multidimensional Scaling: A numerical Methods, *Psychometrika*, 29, 115-129.
 - Carrol, J. D. & Chang, J-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283-319.

SNU Psychometrics Lab
2005-12-05

Singular Value Decomposition

❖ Any matrix A can be expressed as

$$A=UDV'$$

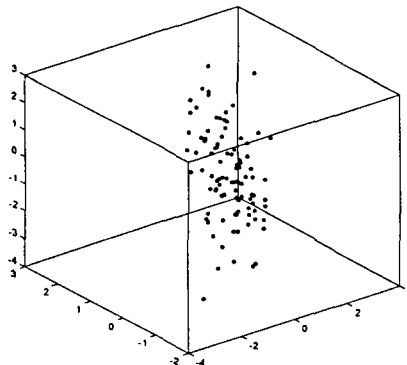
where

- D is a diagonal matrix whose elements are singular value of A
- U consists of eigen vectors of AA'
Describe variables
- V consists of eigen vectors of $A'A$
Describe individuals

Multivariate Data

DATA=

| | | |
|---------|---------|---------|
| -1.5055 | -0.5715 | -0.5950 |
| 0.0709 | -1.0141 | -0.5628 |
| -0.8342 | -0.8800 | -0.0216 |
| -0.2889 | -0.5682 | 0.1253 |
| 1.1497 | 0.2877 | -0.3303 |
| 0.2976 | -0.0602 | 0.2232 |
| -0.4770 | -0.6379 | -1.5234 |
| -0.9111 | -1.5496 | -1.9811 |
| 0.7204 | 0.2491 | -0.6949 |
| -0.4486 | 0.0415 | -0.4129 |
| -0.8658 | -1.4450 | 0.4076 |
| -0.8977 | -0.9921 | -0.0257 |
| ... | ... | ... |
| 2.2936 | 2.5036 | 1.6799 |
| -0.5284 | -0.1091 | 0.0648 |
| -0.6358 | -0.7895 | -0.2454 |
| 0.8953 | 0.1840 | 0.7584 |



Singular Value Decomposition

❖ DATA=USV'

U(100x3)=

| | | |
|---------|---------|---------|
| -0.1070 | 0.0272 | -0.1464 |
| -0.0586 | -0.0243 | 0.1678 |
| -0.0729 | 0.0709 | 0.0107 |
| -0.0320 | 0.0494 | 0.0466 |
| 0.0494 | -0.1002 | 0.1295 |
| 0.0173 | 0.0130 | 0.0572 |
| -0.0975 | -0.1185 | 0.0161 |
| -0.1694 | -0.1123 | 0.0904 |
| 0.0177 | -0.1186 | 0.0665 |
| -0.0307 | -0.0259 | -0.0790 |
| -0.0840 | 0.1414 | 0.0988 |
| -0.0805 | 0.0778 | 0.0187 |
| ... | ... | ... |
| ... | ... | ... |

S(3x3)=

| | | |
|---------|--------|--------|
| 14.6706 | 0 | 0 |
| 0 | 8.0274 | 0 |
| 0 | 0 | 4.5097 |

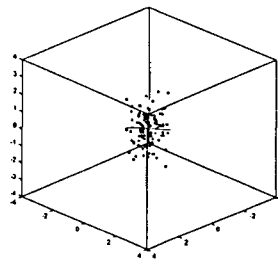
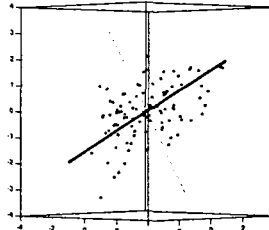
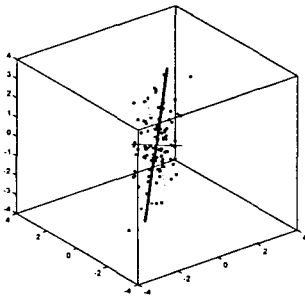
V(3,3)=

| | | |
|--------|---------|---------|
| 0.6147 | -0.3693 | 0.6969 |
| 0.6212 | -0.3177 | -0.7163 |
| 0.4860 | 0.8733 | 0.0341 |

SNU Psychometrics Lab

2005-12-05

Singular Value Decomposition



SNU Psychometrics Lab

2005-12-05

Spectral Decomposition

❖ Any symmetric matrix A can be expressed as

$$A=UDU'$$

$$A=U(D^{1/2}D^{1/2})U'$$

$$A=XX' \text{ where } X=UD^{1/2}$$

where

- D is a diagonal matrix whose elements are eigen values of A
- U consists of eigen vectors of A

Metric MDS

❖ 용어정리

- δ_{ij} : i 항목과 j 항목의 비유사성 (관찰된 자료)
- d_{ij} : i 항목과 j 항목 사이의 거리
- D : $n \times n$ 차원의 행렬, 각 항목은 d_{ij} 로 구성
- $x_{i,r}$: i 번째 항목의 r 번째 차원의 coordinate임
항목이 2차원상에서 표현될 때, i 번째 항목은 (x_{i1}, x_{i2}) 로 표현될 수 있다.
- X : $n \times r$ 원의 행렬

Longcoren (1958)의 방법

- ❖ 우리가 가지고 있는 자료는 비유사성 (유사성) 자료임
 - 유사성 자료는 비유사성 자료로 변환
- ❖ 비유사성 자료를 이용하여 항목간 거리를 계산: $d = c + \delta$
여기서 d 는 거리, δ 는 비유사성, c 는 상수
- ❖ 상수는 거리가 삼각 규칙을 만족시키는 가장 작은 값을 정함
 $c = \max(\delta_{ij} - \delta_{ik} - \delta_{jk})$
- ❖ 이렇게 구성된 $p \times p$ 거리 행렬 D 를 사용하여 새로운 $p \times r$ 행렬 X 를 구성하는 것이 MDS이다. 이때 r 은 항목들이 배열되어 있는 공간의 차원이다.

SNU Psychometrics Lab

2005-12-05

Estimation

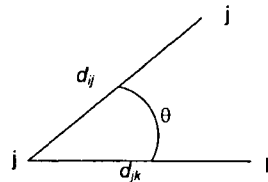
- ❖ MDS의 분석 방법은 항목들 간의 거리정보를 사용하여 위치정보를 찾아내는 것이다. 만약 항목들이 Euclidian 공간에 놓여 있다면 다음과 같은 성질을 이용하여 위치를 찾을 수 있다.
- ❖ i 번째 항목이 영점에 놓여있다고 가정하자.

$$d_{jk}^2 = d_{ik}^2 + d_{ij}^2 - 2d_{ik}d_{ij} \cos \theta$$

$$-\frac{1}{2}(d_{jk}^2 - d_{ik}^2 - d_{ij}^2) = d_{ik}d_{ij} \cos \theta$$

$$\mathbf{x}'_j \mathbf{x}_k = |\mathbf{x}_j| |\mathbf{x}_k| \cos \theta$$

$$-\frac{1}{2}(d_{jk}^2 - d_{ik}^2 - d_{ij}^2) = \mathbf{x}'_j \mathbf{x}_k$$



SNU Psychometrics Lab

2005-12-05

Distortion

- ❖ i 를 원점으로 삼고, $(n-1) \times (n-1)$ 크기의 B 행렬을 구성한다. 이 행렬의 원소는 다음과 같다.

$$b_{jk} = -\frac{1}{2}(d_{jk}^2 - d_{ij}^2 - d_{ik}^2)$$

- ❖ $B = XX'$
- ❖ B 는 대칭행렬이므로 $B = UDU'$.
따라서 $X = UD^{1/2}$ 이다.
- ❖ X 는 i 번째 항목을 원점으로 삼았을 때 $(n-1)$ 개 항목의 좌표를 나타내고 있다.

Double Centering

- ❖ Configuration의 centroid를 원점으로 삼는 것이 display와 해석을 돕는다.
- ❖ Centroid를 원점으로 삼고 B 행렬을 계산하면 다음과 같이 되는데 이에 근거하여 X matrix를 구함

$$b_{jk} = -\frac{1}{2}d_{jk}^2 + \frac{1}{2}\sum_i \frac{d_{jk}^2}{n} + \frac{1}{2}\sum_k \frac{d_{jk}^2}{n} + \frac{1}{2}\sum_i \sum_j \frac{d_{jk}^2}{n^2}$$

Nonmetric MDS

❖ History

- 인지심리학자인 Roger Shepard (1962)가 처음으로 MDS기법을 개발
- Kruskal (1964)에 의하여 이론과 계산 알고리즘이 정립
- Carrol & Chang (1970)에 INDSCAL (Individual differences scale) model이 개발

Kruskal 방식

❖ 기본 아이디어

- loss function (Stress)를 정의
- stress를 최소화 하는 configuration을 찾아냄

$$\text{Raw Stress } S^* = \sum_{i < j} (d_{ij} - \hat{d}_{ij})^2$$

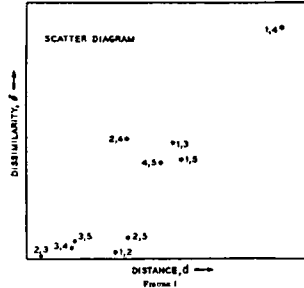
$$\text{Stress} = \frac{S^*}{T^*} \quad \text{where } T^* = \sum_{i < j} d_{ij}^2$$

❖ Kruskal이 제안하는 stress에 의한 모형의 평가

| stress | Goodness of fit |
|--------|-----------------|
| 20% | poor |
| 10% | fair |
| 5% | good |
| 2.5% | excellent |
| 0% | perfect |

δ와 d

- ❖ 5개 대상들간의 거리를 나타내는 경우를 생각해 보자.
- ❖ 어떤 방식으로든 5개의 대상을 공간상에서 표상하면, x_1, \dots, x_5 사이의 거리들을 계산할 수 있다. 이를 d 라 한다.
- ❖ 사람들이 반응한 dissimilarity, δ 와 d 의 관계를 scatter diagram으로 나타낼 수 있다.

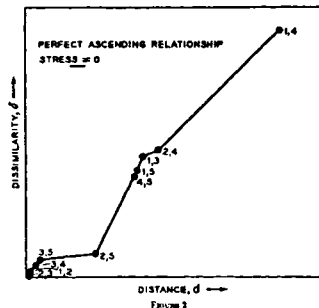


SNU Psychometrics Lab

2005-12-05

δ와 d

- ❖ 가장 자료를 잘 설명하는 δ 는 d 의 순서를 그대로 유지하는 것이다.
- ❖ 오른쪽 그림에서 점들을 연결하면 단조증가함수 형태를 형성해야 한다.
- ❖ 오른쪽 그림은 stress가 0인 완벽한 관계이다.



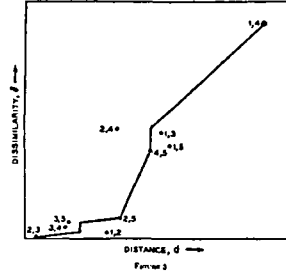
SNU Psychometrics Lab

2005-12-05

δ, d, \hat{d} 의 관계

- ❖ 그러나 오른쪽 그림에서는 각 점들을 연결하면 단조증가함수의 형태를 띠지 않는다.
- ❖ 이때 단조증가함수의 형태를 유지하면서 가장 자료에 근접하는 선을 그을 수 있는데 이를 \hat{d} 이라 한다.
- ❖ 이는 δ 로 d 를 예측하는 회귀모형으로 볼 수 있다. 단 여기에서의 제약은 δ 의 순서를 위배하지 않는 방식으로 \hat{d} 이 결정되어야 한다. 즉

$$\delta_{ij} > \delta_{kl} \Rightarrow \hat{d}_{ij} > \hat{d}_{kl}$$

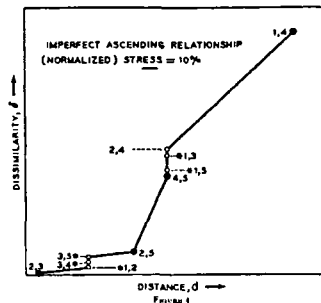


SNU Psychometrics Lab

2005-12-05

$$s^* = \sum_{i < j} (d_{ij} - \hat{d}_{ij})^2$$

- ❖ d 와 \hat{d} 의 차이는 오른쪽 그림에서 점선으로 표시된다.
- ❖ 이 점선 길이의 제곱합이 raw stress가 된다.
- ❖ 이 차이들을 최소화하는 방식으로 x 를 구한다.
- ❖ d 는 x 에 의하여 결정됨을 명심하여야 한다.



SNU Psychometrics Lab

2005-12-05

d_hat을 구하는 방법

❖ 기본원리: 이 절차는 δ 로 d 를 predict 하는 것이다. 이때 δ 의 순서가 d 에도 유지되어야 한다.

❖ 절차

- 먼저 δ 를 $n-1$ (대상의 수 -1)개의 partition으로 나눈다. 각 partition에는 동일한 d_hat 을 할당한다.
- 특정한 하나의 partition을 택한다.
 - 이 partition의 위 partition의 d 가 현재 partition의 d 보다 크면 이 partition의 상한계를 유지한다.
 - 만약 그렇지 않으면, 두 partition을 합치고 d_hat 을 두 d 의 평균으로 정한다.
- 이러한 과정을 현 partition의 상하 partition에 대하여 반복적으로 적용한다.
- 이러한 절차를 반복한다.

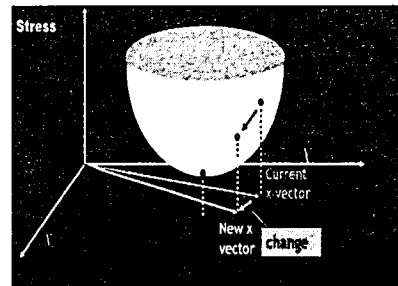
Configuration x 를 구하는 algorithm

❖ Gradient descent algorithm (Newton-Raphson method)

❖ 현재 d_hat 이 정해져 있다.
Stress를 x 에 대하여 미분하여 기울기를 얻음

$$x_{new} = x_{old} - c \frac{d(stress)}{dx}$$

❖ 위의 절차를 반복함



INDSCAL

Individual Differences Scales

- ❖ Individual difference를 반영하는 MDS기법
- ❖ Individual difference를 접근하는 두 가지 방식을 생각할 수 있음.
 - 모든 사람들에게 공통적인 map이 있다고 가정함
 - 각 개인마다 모두 다른 map이 있다는 가정
- ❖ INDSCAL은 common underlying configuration이 있다고 가정하지만, 그 configuration이 similarity에 영향을 주는 weight가 다르다고 가정

$$y_{jd}^{(i)} = w_{id}^{1/2} x_{jd}$$

$$d_{jk}^{(i)} = \sqrt{\sum_{d=1}^r (y_{jd}^{(i)} - y_{kd}^{(i)})^2} = \sqrt{\sum_{d=1}^r w_{id} (x_{jd} - x_{kd})^2}$$

- ❖ 위의 공식에 따르면 $d_{jk}^{(i)}$ 는 x space가 w 에 의해 확장 또는 축소되는 y 들간의 거리로 볼 수 있다.

Estimation

- ❖ Metric MDS에서 $B=XX'$ 을 만족하였다.
- ❖ 여기서는 X 대신 y 를 사용

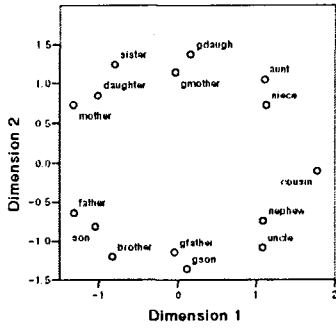
$$b_{jk}^{(i)} = \sum_{d=1}^r y_{jd}^{(i)} y_{kd}^{(i)} = \sum_{d=1}^r w_{id} x_{jd} x_{kd}$$

- ❖ Canonical decomposition를 사용하여 w 를 구함

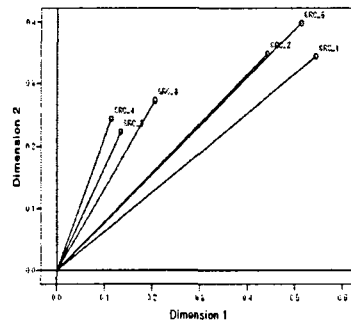
Configuration & Ws

Object Points

Common Space

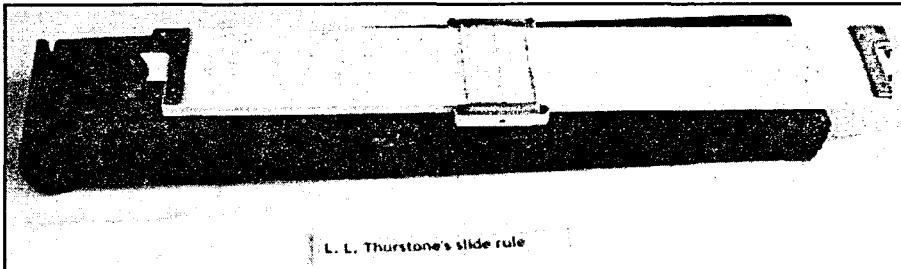


Dimension Weights



SNU Psychometrics Lab

2005-12-05



L. L. Thurstone's slide rule

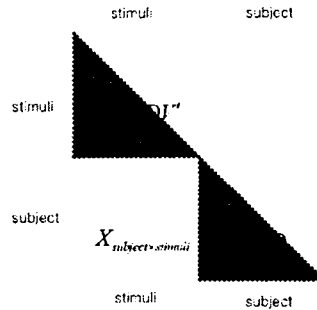
MDS의 확장

SNU Psychometrics Lab

Unfolding 기법 (MDPref)

| | 평론가1 | 평론가2 | 평론가3 |
|-----|------|------|------|
| 영화1 | . | . | . |
| 영화2 | . | . | . |
| 영화3 | . | . | . |
| 영화4 | . | . | . |

- ❖ 영화1, 영화2, 영화3, 영화4와 평론가1, 평론가2, 평론가3을 동일한 그래프에서 표현하고자 함.
- ❖ 이를 위해서는 7 by 7 유사성 행렬이 필요함. 그러나 대체적으로 이를 계산하는 것은 가능하지 않음.
- ❖ 오른쪽 그림에서 우리가 가지고 있는 자료는 stimulus(영화) by subject(평론가) 행렬임.
- ❖ 즉 그림에서 빨간색 부분과 파란색부분의 두 삼각형에 대한 정보가 주어지지 않음. 수학적 방법으로 이 두 부분을 펴(unfolding).



MDPref

- ❖ MDPref는 singular value decomposition에 기초한 metric 모형이다.
 - ❖ 즉 관찰된 반응(주로 선호도) S_{ij} 는 오른쪽과 같이 가정된다. 여기서 y_i 는 i 번째 subject의 선호도를 나타내고, x_j 는 다차원 공간에서 자극의 위치를 나타낸다.
 - ❖ 모든 개인을 포함하는 Matrix는 오른쪽과 같다.
 - ❖ Singular Value Decomposition에 의하여 S는 오른쪽과 같이 분해되고
- $Y_r = U_r D_r$, 그리고 X를 V_r 로 할당한다. 이는 OLS estimator 와 일치한다.

$$S_{ij} = y_i' x_j$$

$$S = YX'$$

$$S = UD V'$$

$$\hat{S} = U_r D_r V_r'$$

Correspondence Analysis

❖ Cross-table (contingency table)에 대한 graphical representation.

- Contingency table의 행과 열의 연합(association)을 표현하는데 그래프로 표현하는 기법

❖ 용어정리

- Profile: 행이나 열의 반응 비율: row profile, column profile, average profile
- Mass: 각 행이나 열에서 profile을 산출한 사례수
- Distance: Euclidian Distance & Chi-square Distance
- Inertia

$$\begin{aligned}
 \text{Inertia} &= \text{mass} \times \text{distance}^2 \\
 &= \text{mass} \times \frac{(p_i - p_{ave-i})^2}{p_{ave-i}}
 \end{aligned}$$

SNU Psychometrics Lab

2005-12-05

Vertex

- Vertex는 가상적인 행 profiles을 나타내고 다음과 같은 형식을 띤다.
[1 0 0] [0 1 0] [0 0 1]
이때 profile은 하나의 범주에 전적으로 집중해 있다.
- 삼각형안에 profile의 모든 점들이 놓이게 된다.
- profile point가 vertex로 거리가, 해당 행이나 열의 연상의 정도를 나타낸다.
- Profile의 점들은 vertex 점들의 가중화된 합으로 기술될 수 있다.

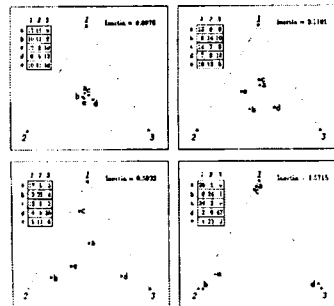


FIGURE 1.4 Four 3 x 3 tables, with increasing total inertia, and the graphical displays of the row profiles and column vertices.

SNU Psychometrics Lab

2005-12-05

Asymmetric Map

- ❖ Vertex를 경계로 그 안에 대상들을 표현함.
- ❖ 열의 한 범주가 행의 어떤 범주와 가장 연합되었는지를 판단할 수 있음.

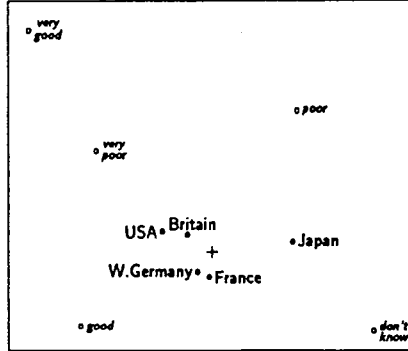


FIGURE 1.7 Asymmetric two-dimensional CA map of the data of Example 3 in Figure 1.3, showing the projection of the row profiles (countries) and column vertices (response categories).

Symmetric Map

- ❖ 행과 열에 대한 두 표상을 결합시킨 map
- ❖ 두 개의 mapping이 overlapping된 것으로 해석하여야 함.
- ❖ 차원간의 거리 비교는 하지 못함.

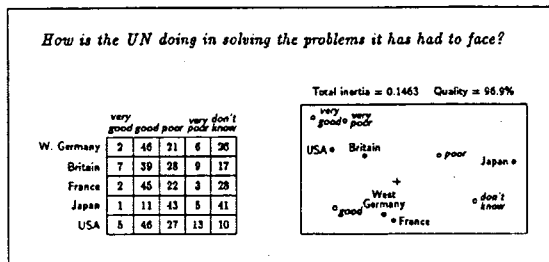


FIGURE 1.3 Example 3: A 5 x 5 table of percentages (rows sum to 100%) from a series of surveys in five different countries, together with the two-dimensional symmetric CA map and the amount of inertia accounted for by the map.

Symmetric Map을 사용할 때 유의할 점

- ❖ row profile과 column profile이 비록 하나의 map에 제시되어 있지만 다른 공간을 차지하고 있다.
- ❖ Row와 column points는 map에서 횡축과 종축에서 동일한 정도로 퍼져있다.
- ❖ 두 가지 map을 동시에 표현할 수 있는 근거는 asymmetric map을 매개로 해서이다.
- ❖ 직접적인 row-to-column distance 해석은 가능하지 않다. 그러나 주축에 대하여서는 행과 열을 동시에 해석할 수 있다.
- ❖ 엄격하게 말하면, 행과 열의 연합은 symmetric map에서는 관찰할 수 없다.

Inertia

- ❖ 만약 행과 열이 완전히 독립적이면, 수반표의 각 cell의 수치는 (즉 mass의 분포) 행과 열의 profile로부터 재생할 수 있다.
- ❖ 행과 열 profile에서 예언되는 빈도에서 벗어나는 정도가 overall chi-square의 값을 결정한다.
- ❖ 여기에서는 overall chi-square 분해한다.

$$Inertia = \frac{\chi^2}{N}$$

Quality

- ❖ **Quality:** 정해진 차원에 의해 정의된 해당 행 coordinate system에서 row point들의 표상의 quality이다
- ❖ 한 점의 **quality**는 정해진 차원의 원점에서부터의 거리 제곱을 최대차원의 공간에서의 원점에서의 거리로 나눈 것이다.
- ❖ **Relative inertia:** 각 점에 의하여 설명할 수 있는 전체 inertia의 비율이다.
- ❖ 상대적 inertia is 선택된 차원의 수와 무관하다. 특정한 solution은 하나의 점을 매우 잘 설명할 수 있으나 (즉 high quality를 가지고 있느냐), 그 점이 전체 inertia에는 공헌하는 바가 적을 수 있다. (예, a row point with a pattern of relative frequencies across the columns that is similar to the average pattern across all rows).

Illustration

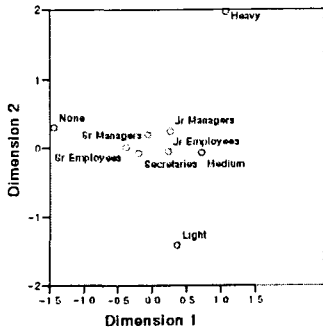
| Staff group | Smoking category | | | | row total |
|------------------|------------------|-------|--------|-------|-----------|
| | None | Light | medium | heavy | |
| Senior Managers | 4 | 2 | 3 | 2 | 11 |
| Junior Managers | 4 | 3 | 7 | 4 | 18 |
| Senior Employees | 25 | 10 | 12 | 4 | 51 |
| Junior Employees | 18 | 24 | 33 | 13 | 88 |
| Secretaries | 10 | 6 | 7 | 2 | 25 |
| Column Totals } | 61 | 45 | 62 | 25 | 193 |

Percentage of Row Totals

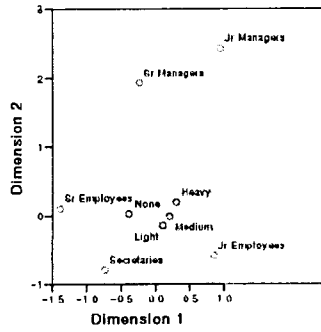
| Staff group | Smoking category | | | | row total |
|------------------|------------------|-------|--------|-------|-----------|
| | None | Light | medium | heavy | |
| Senior Managers | 36.36 | 18.18 | 27.27 | 18.18 | 100.00 |
| Junior Managers | 22.22 | 16.67 | 38.89 | 22.22 | 100.00 |
| Senior Employees | 49.02 | 19.61 | 23.53 | 7.84 | 100.00 |
| Junior Employees | 20.45 | 27.27 | 37.50 | 14.77 | 100.00 |
| Secretaries | 40.00 | 24.00 | 28.00 | 8.00 | 100.00 |

Asymmetric Map

Row Principal Normalization



Column Principal Normalization

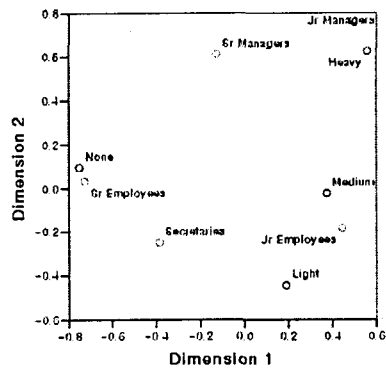


SNU Psychometrics Lab

2005-12-05

Asymmetric Map

Symmetrical Normalization



SNU Psychometrics Lab

2005-12-05

SVD & Statistics

Eigenvalues and Inertia for all Dimensions

Total Inertia = .08519 & $\chi^2 = 16.442$

| No. of Dims | Singular Values | Eigen-Values | Perc. of Inertia | Cumulatv Percent | Chi Squares |
|-------------|-----------------|--------------|------------------|------------------|-------------|
| 1 | 0.2734 | 0.0748 | 87.8 | 87.8 | 14.4285 |
| 2 | 0.1001 | 0.0100 | 11.8 | 99.5 | 1.9333 |
| 3 | 0.0203 | 0.0004 | 0.5 | 100.0 | 0.0798 |

Multiple Correspondence Analysis

- ❖ Multiple correspondence analysis: correspondence analysis를 두 개 이상의 변수가 있을 때로 확장한 경우임
- ❖ MCA는 관찰수를 행으로 하고 범주를 열로 하는 indicator (or design) matrix 상에서 단순 대응분석을 하는 것과 같다.
- ❖ 실제로는 Burt matrix를 사용하여 분석하게 된다.

Burt Matrix

Burt Matrix

| | Survival | | Age | | | Location | | |
|-------------------|----------|-----|--------|-------|-----|----------|--------|----------|
| | NO | YES | less50 | 50-69 | 69+ | TOKYO | BOSTON | GLAMORGN |
| SURVIVAL:NO | 210 | 0 | 68 | 93 | 49 | 60 | 82 | 68 |
| SURVIVAL:YES | 0 | 554 | 212 | 258 | 84 | 230 | 171 | 153 |
| AGE:UNDER 50 | 68 | 212 | 280 | 0 | 0 | 151 | 58 | 71 |
| AGE:50TO69 | 93 | 258 | 0 | 351 | 0 | 120 | 122 | 109 |
| AGE:OVER 69 | 49 | 84 | 0 | 0 | 133 | 19 | 73 | 41 |
| LOCATION:TOKYO | 60 | 230 | 151 | 120 | 19 | 290 | 0 | 0 |
| LOCATION:BOSTON | 82 | 171 | 58 | 122 | 73 | 0 | 253 | 0 |
| LOCATION:GLAMORGN | 68 | 153 | 71 | 109 | 41 | 0 | 0 | 221 |

맺는말

- ❖ 이 강의에서는 MDS기법을 소개하는 데 목적을 두었다.
- ❖ MDS는 많이 사용되는 기법이기는 하나 그 원리를 기술하는 책들은 많지 않다.
- ❖ 현재 많은 user-friendly한 소프트웨어가 개발되어 있으므로 소프트웨어의 사용법보다는 원리를 이해하는데 중점을 두었다.
- ❖ PC-MDS등과 같은 많은 stand-alone 프로그램과 SPSS, SAS등이 MDS 분석법을 제공하고 있다.