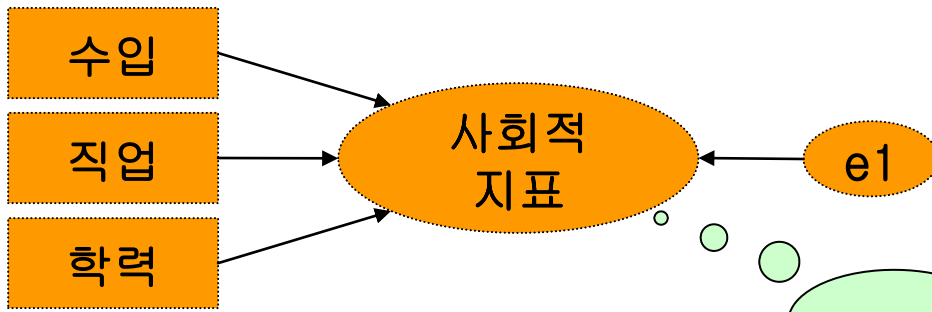
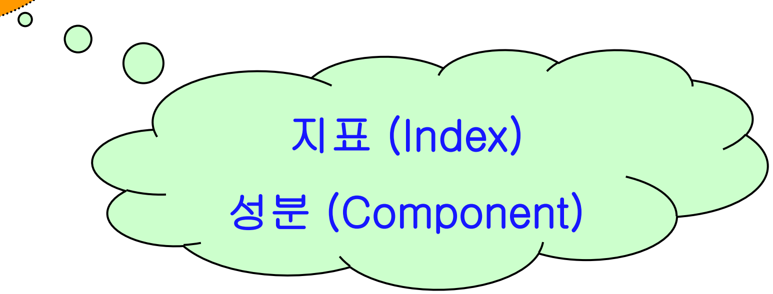
제 3 장
인자분석
Factor Analysis

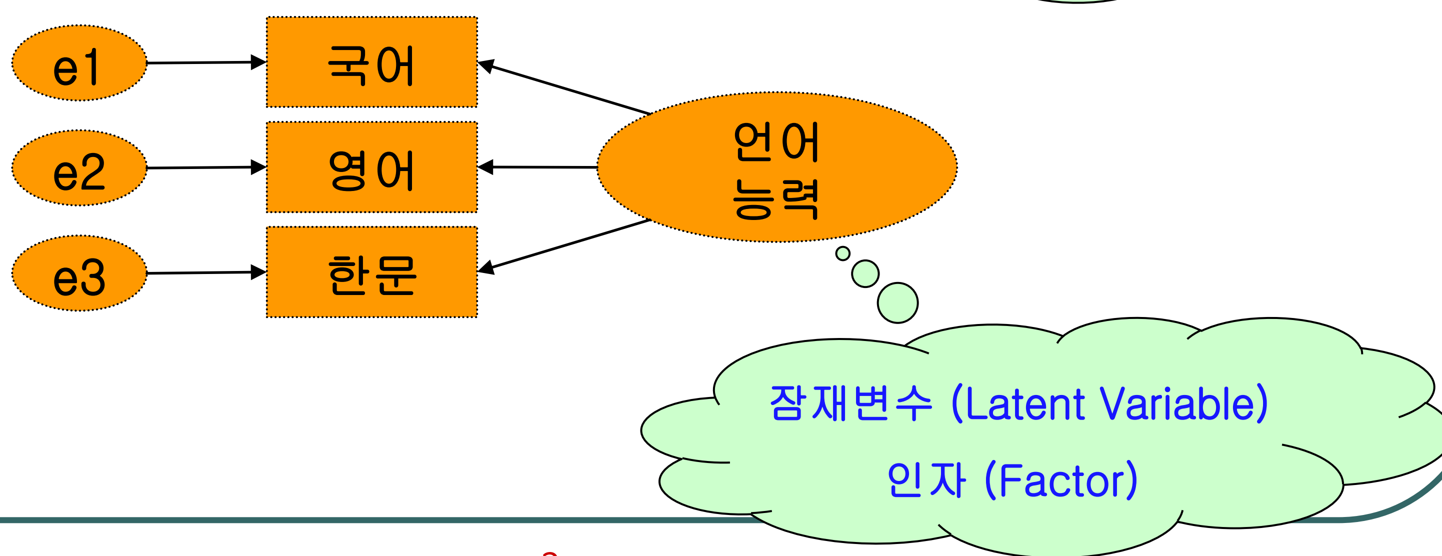
3.1 인자분석의 개념

- 형성적 지표 (Formative indicator)

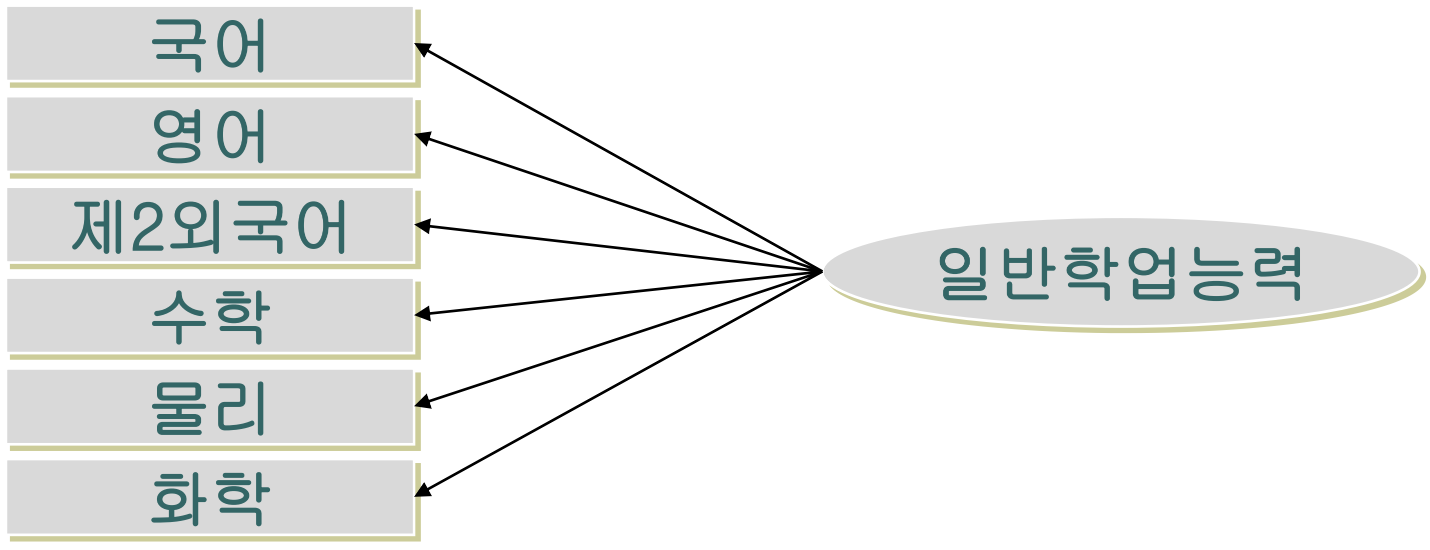




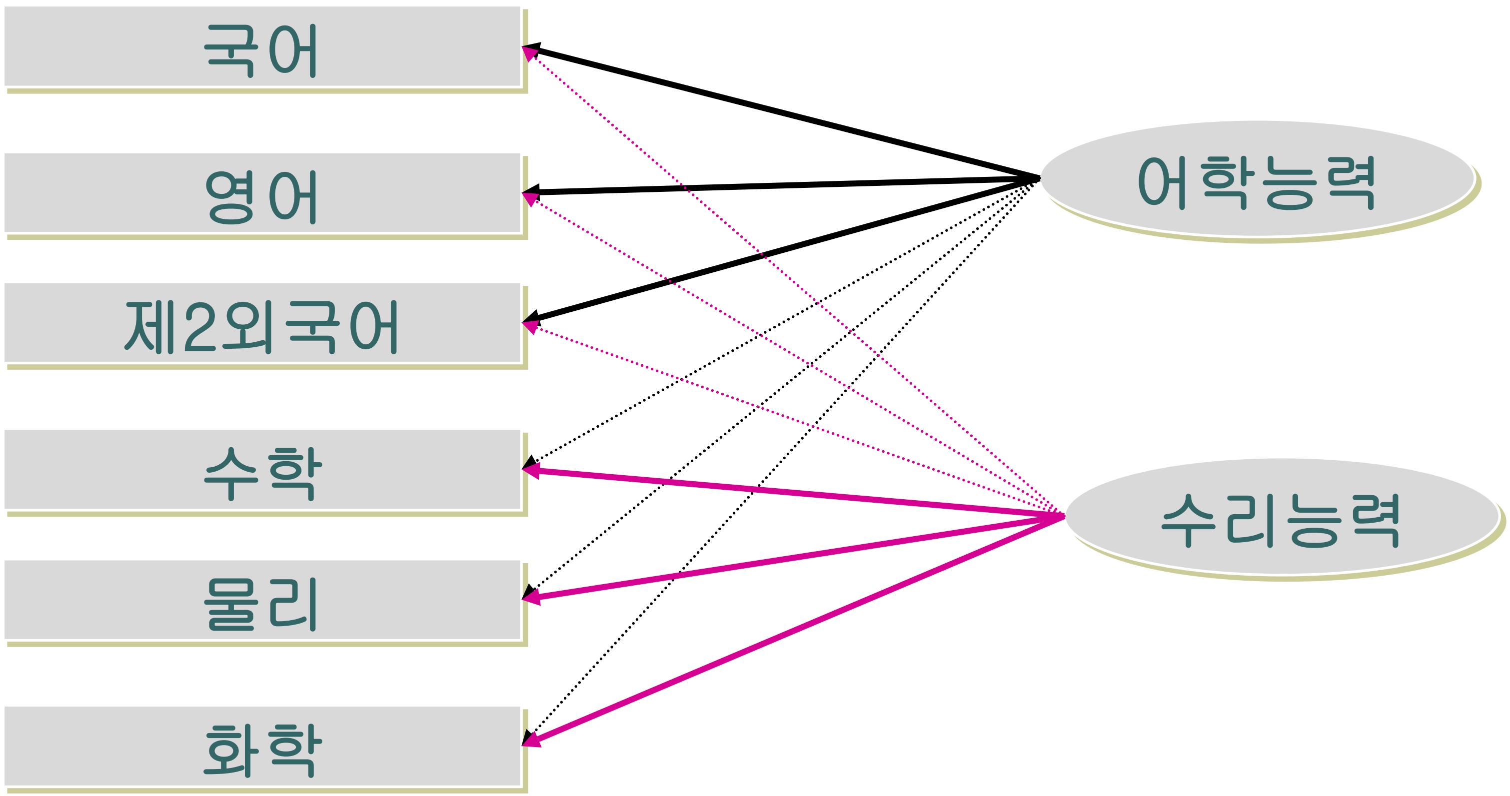
- 반영적 지표 (Reflective indicators)



3.1 인자분석의 개념



3.1 인자분석의 개념



3.1 인자분석의 개념 I

(1) 서로 상관관계를 맺고 있어서,
(2) 직접적으로 해석하기 어려운,
(3) 여러 관찰(측정)변수들 간의 구조적 연관관계를,

(a) 상대적으로 독립적이면서,
(b) 변수들의 저변구조를 이해하기 위해 개념상 의미를 부여할 수 있는,
(c) 원래 변수들의 개수보다 훨씬 적은 개수의 공통인자들을 상정하여,
이들을 통해 분석하고자 하는 통계적 방법.

- 공통인자 (Common Factor)
  - ✓ 변수들이 그들의 구조적 측면에서 서로 공유하고 있는 확률적 인자로서, 변수들 간의 상관관계를 생성시키는 가설적인, 관찰할 수 없는, 혹은 저변에 깔려 있는 인자를 의미.
- 인자분석은 상관(혹은 공분산)행렬의 구조에 관한 통계적 모형을 구축하고, 그와 같은 구조를 생성시키는 소수 몇 개의 인자를 유도하여, 변수들 간의 구조적 관계를 해석하는 공분산 내지 상관중심의 기법.

3.1 인자분석의 개념 II

- 인자분석(factor analysis)은 공분산 구조를 몇 개의 관측 불가능한 ‘인자(factor)'로 설명하려는 것으로 변수들 간에 내재하고 있는 공통의 구조를 파악하고, 데이터의 특성을 몇 개의 인자로 축약하여 설명하고자 하는 것이 인자분석의 목적이라 할 수 있다.

- 변수들에 내재되어 있는 공통의 인자로서 데이터를 설명함으로써 분석에 필요한 변수의 차원을 줄일 수 있다.

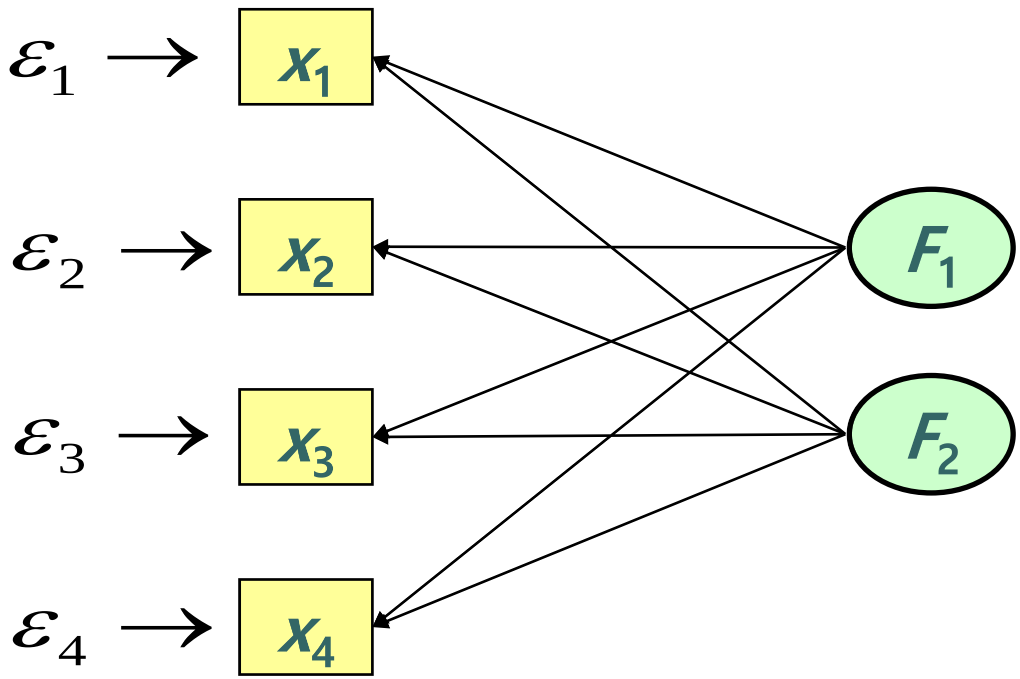
- 인자분석은 특히 경영학, 사회학 관련분야에서 활발히 사용되며 이러한 분야에서는 변수의 개수가 많고 관계가 연계되어 있어, 변수간의 구조(structure)와 변수 내 연관성에 대한 집중적 해석이 매우 중요하다.

- 인자분석은 Spearman(1904)에 의해 시작. 단일인자모형 고려. Garnett(1919), Thurstone(1931) 등에 의해 여러 개의 인자를 동시에 고려하는 다중 인자모형 (multiple common factor model)으로 확장.

3.1 인자분석의 개념 III

- 공통인자
  - 변수들이 구조적 측면에서 서로 공유하고 있는 확률적 인자
  - 변수들간의 상관관계를 생성시키는 가설적인, 관찰할 수 없는, 혹은 저변에 깔려 있는 요인
  - 잠재변수(latent variable)
- 인자분석은 상관(혹은 공분산)행렬의 구조에 관한 통계적 모형을 구축하고, 그와 같은 구조를 생성시키는 소수 몇 개의 인자를 유도하여, 변수들 간의 구조적 관계를 해석하는 공분산 내지 상관중심의 기법

❖ 인자모형 개념도


ε_1
ε_2
ε_3
ε_4
x_1
x_2
x_3
x_4
F_1
F_2

3.1.2 인자분석의 목적

- 인자분석의 목적
  - 변수들의 군집화
    - 변수들 간의 상관(공분산) 행렬의 구조에 관한 통계적 모형을 구축
    - 그 구조를 생성시키는 소수의 인자 유도
    - 변수들간의 상관계수가 높은 집단 생성
    - 변수 군집간에는 변수들간의 상관관계가 낮도록 유지

3.2 예시: TV 프로그램 유형별 시청 정도

ID	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	gender	age	$\cdots$
1	2	1	1	1	4	2	1	1	1	2	$\cdots$
2	5	.	2	.	1	1	1	1	1	2	$\cdots$
3	5	2	2	2	6	5	5	1	1	3	$\cdots$
4	4	4	3	4	6	6	4	5	1	3	$\cdots$
5	5	2	3	2	3	4	4	1	1	3	$\cdots$
6	4	5	5	4	4	5	4	1	1	3	$\cdots$
7	4	5	3	5	3	5	5	3	1	3	$\cdots$
8	5	5	5	5	5	5	4	1	1	3	$\cdots$
9	5	4	4	3	5	5	3	1	1	4	$\cdots$
10	5	4	2	1	3	4	1	1	1	4	$\cdots$
$\cdots$					$\cdots$					$\cdots$	
1000	4	4	5	4	6	4	3	1	1	3	$\cdots$

- 분석변수:

  x_1(뉴스/보도), x_2(드라마), x_3(영화), x_4(쇼/오락), x_5(스포츠), x_6(다큐멘터리), x_7(생활정보), x_8(어린이/만화)

- 변수값:

  1='전혀 안본다.', 2='안본다.', 3='별로 안본다.', 4='약간 본다.', 5='어느 정도 본다.', 6='매우 많이 본다.', .='모름/거절'

3.1 자료행렬을 이용한 인자분석 in R. TV program data

SAS 데이터를 R에서 불러오기

```
> install.packages("sas7bdat")
> library(sas7bdat)
> tvprog<- read.sas7bdat("tvprog.sas7bdat")
> dim(tvprog)
[1] 1000   15
> head(tvprog,5) # 첫 5개 데이터 보기.
  ID x1  x2 x3  x4 x5 x6 x7 x8 gender married age region job house
1  1  2   1  1   1  4  2  1  1      1       1   2      1   3     1
2  2  5 NaN  2 NaN  1  1  1  1      1       1   2      1   3     2
3  3  5   2  2   2  6  5  5  1      1       2   3      1   1     1
4  4  4   4  3   4  6  6  4  5      1       2   3      1   4     3
5  5  5   2  3   2  3  4  4  1      1       2   3      1   4     3
> tvprog<- tvprog[,2:9]
> dim(tvprog)
[1] 1000    8
```

데이터 클리닝 (NaN 을 찾아내고 처리하기)

```
> length(which(tvprog=="NaN"))
[1] 61
> which(apply(tvprog,1,mean)=="NaN")
 [1]   2  67  70  95 238 309 315 412 427 435
510 512 540 561 562 582 584 600 722 742 852 904
942
> length(which(apply(tvprog,1,mean)=="NaN"))
[1] 23
> row.id.NaN<-
which(apply(tvprog,1,mean)=="NaN")

> tv.data<- tvprog[-row.id.NaN,]
> dim(tv.data)  # tv.data 를 이용하여 분석을 실시한다.
[1] 977  8
```

R 명령어 factanal 을 이용한 인자분석

1. 최대 우도법을 이용한 인자의 개수가 몇개가 좋은가 파악한다. 이때 아래에 제공되는 p-value 보고 결정.

귀무가설: 현재 가정된 인자의 개수가 적당하다.
대립가설: 현재 가정된 인자의 개수가 부족하다.

즉, p-value 가 작다면 귀무가설을 기각하고, 인자의 개수가 부족하다는 결론을 내리게 되며, 인자의 수를 늘려봐야 한다.

```
factanal(tv.data, factors = 1, method ="mle")$PVAL
factanal(tv.data, factors = 2, method ="mle")$PVAL
factanal(tv.data, factors = 3, method ="mle")$PVAL
factanal(tv.data, factors = 4, method ="mle")$PVAL
factanal(tv.data, factors = 5, method ="mle")$PVAL
```

R 명령어 factanal 을 이용한 인자분석

```
> factanal(tv.data, factors = 1, method ="mle")$PVAL
  objective
9.62098e-72
> factanal(tv.data, factors = 2, method ="mle")$PVAL
   objective
1.643972e-10
> factanal(tv.data, factors = 3, method ="mle")$PVAL
   objective
0.0006304031
> factanal(tv.data, factors = 4, method ="mle")$PVAL
 objective
0.00251929
> factanal(tv.data, factors = 5, method ="mle")$PVAL
Error in factanal(tv.data, factors = 5, method =
"mle"):  5 factors are too many for 8 variables
```

위의 결과를 보면 명확한 인자의 개수가 결정되기 힘듣다.

R 명령어 factanal 을 이용한 인자분석

```
>  my.model<- factanal(tv.data, factors = 3, method ="mle")
> my.model$loadings
Loadings:
   Factor1 Factor2 Factor3
x1 0.478   0.149   0.451
x2 0.139   0.546   0.408    # 빈칸은 인자 적재(loadings) 중에서
x3 0.272   0.674            # 무시할 만큼 작은 숫자를 의미한다.
x4 0.167   0.684   0.289
x5 0.481   0.341   0.212
x6 0.920   0.163
x7 0.538   0.328   0.372
x8 0.116   0.473

               Factor1 Factor2 Factor3
SS loadings      1.729   1.717   0.647
Proportion Var   0.216   0.215   0.081
Cumulative Var   0.216   0.431   0.512
# 3개의 인자가 8개 변수로 이루어진 데이터의 약 51.2% 정도 변동을 설명
```

R 명령어 factanal 을 이용한 인자분석

전체 인자의 적재(loadings)를 보고 싶을 경우.

```
> my.model$loadings[]
     Factor1   Factor2    Factor3
x1 0.4775037 0.1494263 0.45140122
x2 0.1390126 0.5462793 0.40804590
x3 0.2724471 0.6739272 0.04719797
x4 0.1667688 0.6841201 0.28890533
x5 0.4812852 0.3406952 0.21199301
x6 0.9195360 0.1632219 0.06049769
x7 0.5378964 0.3283148 0.37171067
x8 0.1164471 0.4725950 0.06198297
```

R 명령어 factanal 을 이용한 인자분석

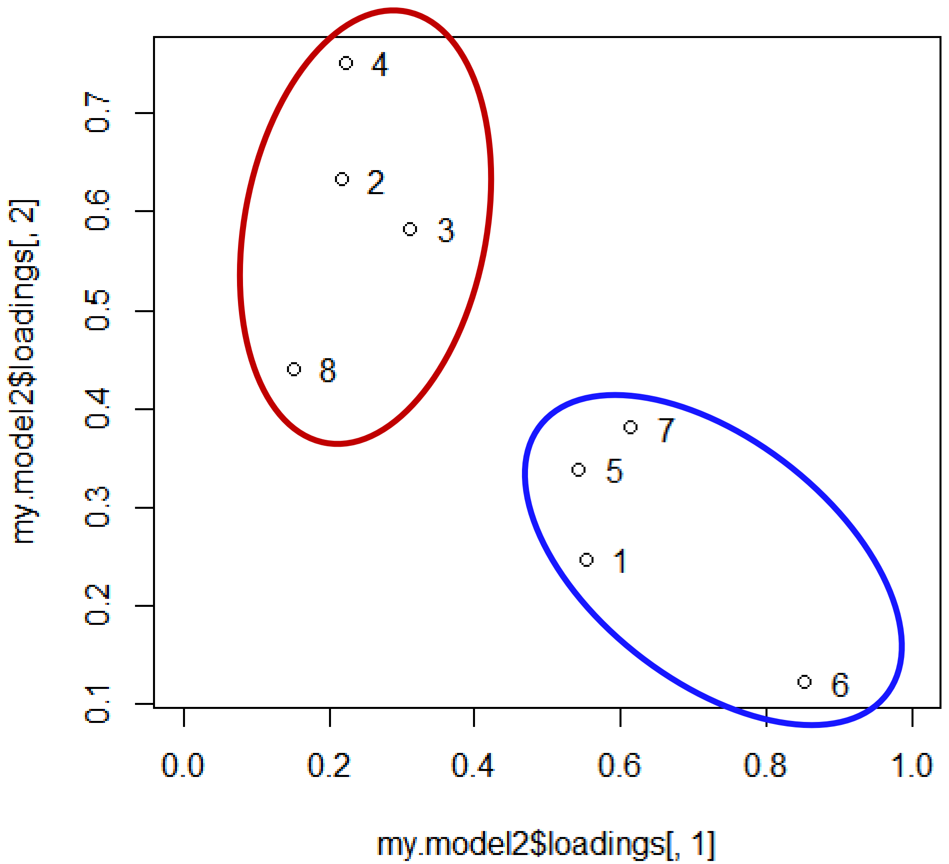
해석의 용이를 위해 2개의 인자를 선택한 모형

```
>  my.model2<-factanal(tv.data,factors=2,method ="mle")
> loadings(my.model2)
Loadings:
   Factor1 Factor2
x1 0.553   0.247          # 인자가 3개인 모형과 비교해보면
X2 0.218   0.632          # 인자 적재가 좀더 명확히 드러난다.
x3 0.312   0.583          # 즉, 로딩 값들의 대비가 더 뚜렷함.
x4 0.224   0.752
x5 0.543   0.339
x6 0.854   0.123
x7 0.615   0.380
x8 0.151   0.441
               Factor1 Factor2
SS loadings      1.925   1.835
Proportion Var   0.241   0.229 # 2개의 인자로 데이터
Cumulative Var   0.241   0.470 # 변동의 47% 정도를 설명.
```

그래프를 통해 두개의 인자를 구성하는 변수들 파악하기

각각의 인자를 하나의 축으로 놓고 2D로 파악 (Biplot)

```
> plot(my.model2$loadings[,1],my.model2$loadings[,2],xlim=c(0,1))
> text(x=my.model2$loadings[,1]+0.05,y=my.model2$loadings[,2])
```



Intuition

- 두 변수 X_1과 X_2가 다음과 같은 선형관계를 이루고 있다고 가정하자.

$$X_2 = 2X_1 - 1$$

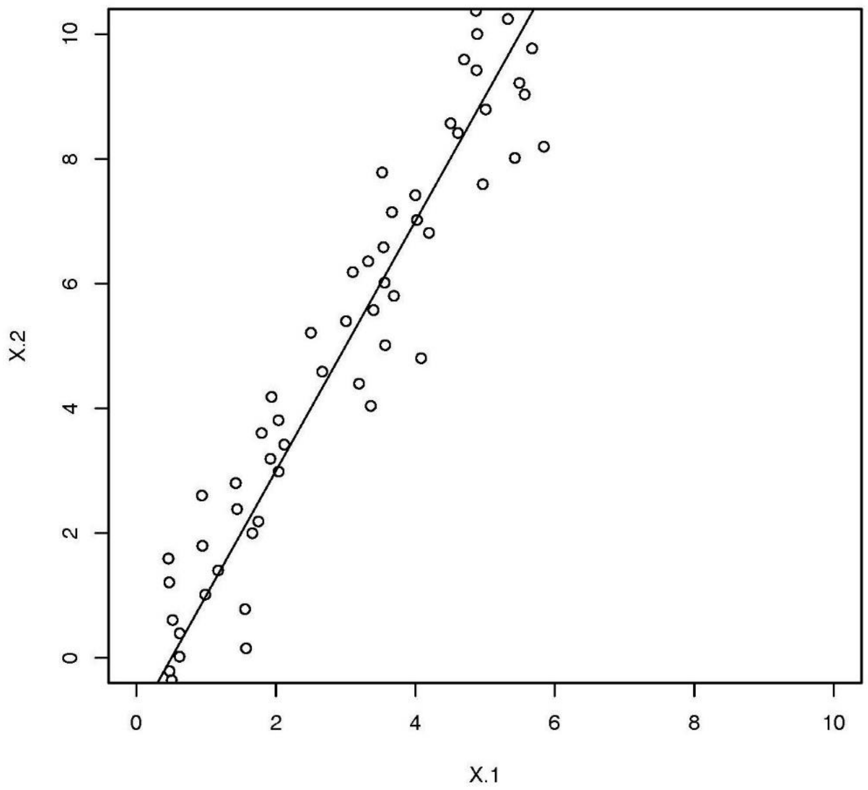
- X_1과 X_2의 관계는 다음의 공통인자 f 를 통해 표현 할 수 있다.

$$X_1 = 2 + f;\ X_2 = 3 + 2f$$
$$\vdots$$

- X_1과 X_2 의 관계는 공통인자 f를 통해 설명되어진다.

Intuition

- 두 변수 X_1과 X_2의 대부분의 변량 (*variability*)은 공통인자 f를 통해 설명할 수 있다.



3.3 인자분석모형

- 만약 p개의 변수 $(X_1, X_2, \cdots, X_p)$가 있다면 다음과 같은 선형식으로 표현 할 수 있다.

$$X_j = \mu_j + \lambda_j f + \varepsilon_j, \qquad j = 1, 2, \cdots, p.$$

- p개의 변수 $(X_1, X_2, \cdots, X_p)$ 의 변량은 공통인자 f를 통해 설명할 수 있다.
- 특수인자 ε_j 는 공통인자 f를 통해 설명한 X_j의 변량에 대한 오차항이다.
- 인자적재값 λ_j는 X_j와 공통인자 f의 상관관계를 나타낸다.

3.3 인자분석모형

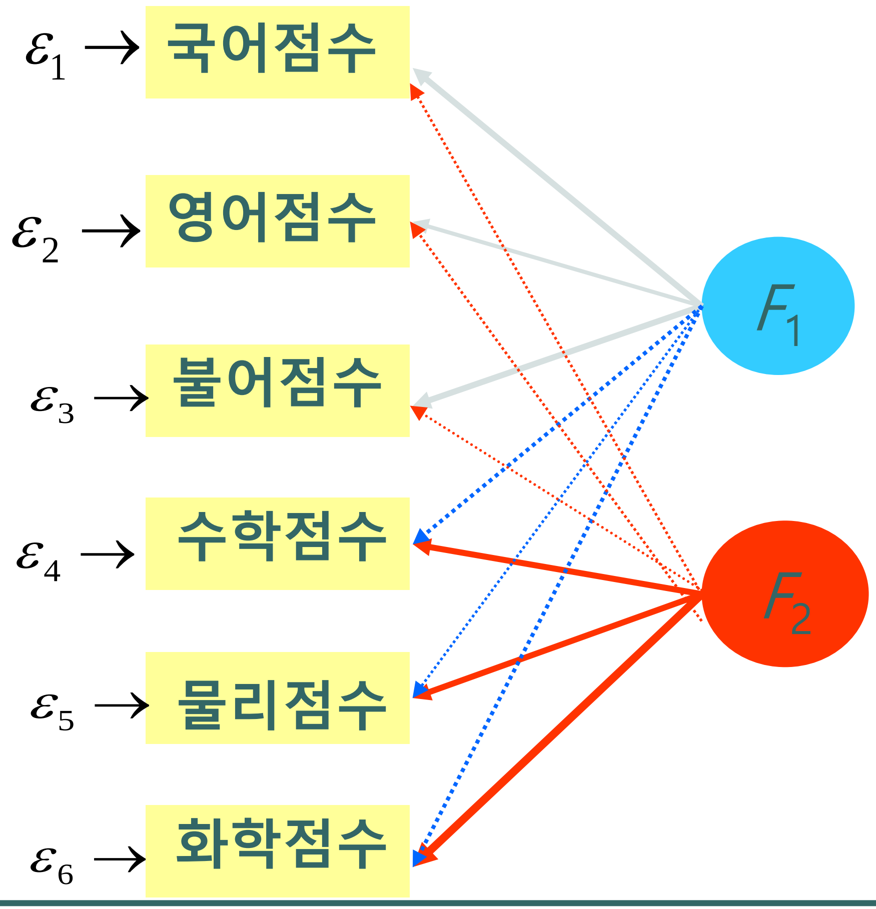
- ***m*-인자모형**
  - ✓ 각 변수는 *m*개의 공통인자(F_1, F_2, …, F_m)의 선형결합과 그 변수에만 영향을 미치는 특수한 인자의 합으로 표현될 수 있다.

$$\mathbf{X}_{p\times1} = \boldsymbol{\mu}_{p\times1} + \boldsymbol{\Lambda}_{p\times m}\mathbf{F}_{m\times1} + \boldsymbol{\varepsilon}_{p\times1}$$

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} + \begin{pmatrix} \lambda_{11} \cdots \lambda_{1m} \\ \lambda_{21} \cdots \lambda_{2m} \\ \vdots \quad \ddots \quad \vdots \\ \lambda_{p1} \cdots \lambda_{pm} \end{pmatrix} \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

3.3 인자분석모형의 예

- 2-인자모형



$$X_1 = \mu_1 + \lambda_{11}F_1 + \lambda_{12}F_2 + \varepsilon_1$$

$$X_2 = \mu_2 + \lambda_{21}F_1 + \lambda_{22}F_2 + \varepsilon_2$$

$$\vdots$$

$$X_p = \mu_p + \lambda_{p1}F_1 + \lambda_{p2}F_2 + \varepsilon_p$$

3.3 인자분석모형

$$\begin{cases} x_1 - \mu_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \cdots + \lambda_{1m}F_m + \varepsilon_1 \\ x_2 - \mu_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \cdots + \lambda_{2m}F_m + \varepsilon_2 \\ \quad \cdots \\ x_p - \mu_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \cdots + \lambda_{pm}F_m + \varepsilon_p \end{cases}$$

$p \times 1$ 확률벡터 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p)'$가 위와 같이 모평균벡터 μ와 모공분산행렬 Σ을 갖는다.
$\mathbf{X}$ 는 m개의 공통인자(common factor) $F_1, F_2, \ldots, F_m$과 특수인자(specific factor) $\epsilon_1, \epsilon_2, \ldots, \epsilon_p$
의 선형결합으로 표현된다고 가정한다.

3.3 인자분석모형

$$\begin{cases} x_1 - \mu_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \cdots + \lambda_{1m}F_m + \varepsilon_1 \\ x_2 - \mu_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \cdots + \lambda_{2m}F_m + \varepsilon_2 \\ \quad \cdots \\ x_p - \mu_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \cdots + \lambda_{pm}F_m + \varepsilon_p \end{cases}$$

여기서 λ_{ij} 는 *i* 번째 변수의 *j* 번째 인자의 적재값이며, $F_1, F_2, \cdots, F_m$ 는 *m* 개의, $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$ 는 *p* 개의 관측불가능한 (unobservable) 확률변수이다. (전체 관측불가능한 확률변수의 개수는 *m*+*p*개) 이와 같이 *m*개 공통인자를 포함한 인자모형을 *m*–인자모형이라고 부르기도 한다.

$F_1, F_2, \cdots, F_m$: 공통인자 (common factor)
$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$: 특수인자 (specific factor)

3.3.1 인자분석모형의 가정

$$\begin{cases} x_1 - \mu_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \cdots + \lambda_{1m}F_m + \varepsilon_1 \\ x_2 - \mu_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \cdots + \lambda_{2m}F_m + \varepsilon_2 \\ \quad \cdots \\ x_p - \mu_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \cdots + \lambda_{pm}F_m + \varepsilon_p \end{cases}$$

(1) $E(F_i) = 0,\ \ Var(F_i) = 1,\ \ Cov(F_i, F_j) = 0$ for $i \neq j$

(2) $E(\varepsilon_i) = 0,\ \ Var(\varepsilon_i) = \varphi_i\,,\ \ Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$

(3) $Cov(F_i, \varepsilon_j) = 0$ for all i, j

i 번째 변수의 분산 중 m개의 공통인자에 의해 설명되는 부분

- $Var(x_i) = \boxed{\left\{\lambda_{i1}^2 + \lambda_{i2}^2 + \ldots + \lambda_{im}^2\right\}} + \varphi_i$ =공통성(communality)+특수성
- $\lambda_{ik} = Cov(x_i, F_k)$, i.e. $\lambda_{ik} = Corr(z_i, F_k)$

3.3.1 인자분석모형의 용어 정리

- 용어정리

$$\mathbf{X}_{p\times 1} = \boldsymbol{\mu}_{p\times 1} + \boldsymbol{\Lambda}_{p\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\varepsilon}_{p\times 1}$$

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} + \begin{pmatrix} \boxed{\begin{matrix} \lambda_{11} & \cdots & \lambda_{1m} \\ \lambda_{21} & \cdots & \lambda_{2m} \\ \vdots & & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pm} \end{matrix}} \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix} F_1 \\ \vdots \\ F_m \end{matrix}} \end{pmatrix} + \begin{pmatrix} \boxed{\begin{matrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{matrix}} \end{pmatrix}$$

인자적재(부하)행렬

공통인자

특수인자

3.4 인자적재와 특수분산의 추정

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{pmatrix} \qquad \Psi = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}$$

인자패턴(적재)행렬 특수분산행렬

- 주성분인자법 : 주성분분석 이용

$$F_1 = P_1 / \sqrt{l_1}, \quad F_2 = P_2 / \sqrt{l_2}, \quad \ldots$$

- 주축인자법 : 특수분산의 초기값을 정하고 반복적으로 계산

$$\mathrm{R} = \Lambda\Lambda' + \Psi \qquad \text{R: X의 상관행렬}$$

- 최대우도법 : 다변량정규분포의 가정 하에 우도(가능도)함수를 최대화

3.4 인자적재와 특수분산의 추정

- 주성분인자법 : 상관행렬 R의 경우
  - ✓ i 번째 변수의 공통성에 관한 초기추정값 (C_i^0)을 선택
  - ✓ Ψ의 초기추측값: $\Psi^0 = \text{diag}(1- C_1^0, 1- C_2^0, \cdots, 1- C_p^0)$
  - ✓ $R - \Psi^0$: 조정된 상관행렬 (adjusted correlation matrix)
  - ✓ 분광분해 (spectral decomposition)를 적용

$$R - \Psi^0 = \delta_1 e_1 e_1' + \delta_2 e_2 e_2' + \cdots + \delta_p e_p e_p'$$

여기서 $\delta_1, \delta_2, \cdots, \delta_p$는 $R - \Psi^0$의 고유값들로서 $\delta_1 > \delta_2 > \cdots > \delta_p$,
$e_1, e_2, \cdots, e_p$는 $R - \Psi^0$의 고유벡터들

  - ✓ 보유인자 개수 ($m \le p$)만큼의 주성분만을 이용하여 공통성을 추정

$$\Lambda\Lambda' = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots + \lambda_m e_m e_m'$$

  - ✓ 특수분산행렬 Ψ은 1에서 추정된 공통성을 빼서 추정

$$\psi_{ii} = 1 - (\lambda_1 e_{1i}^2 + \lambda_2 e_{2i}^2 + \cdots + \lambda_m e_{mi}^2)$$

  - ✓ 공통성에 관한 초기추정값의 적절한 선택이 추정의 정확성의 관건

3.4 인자적재와 특수분산의 추정

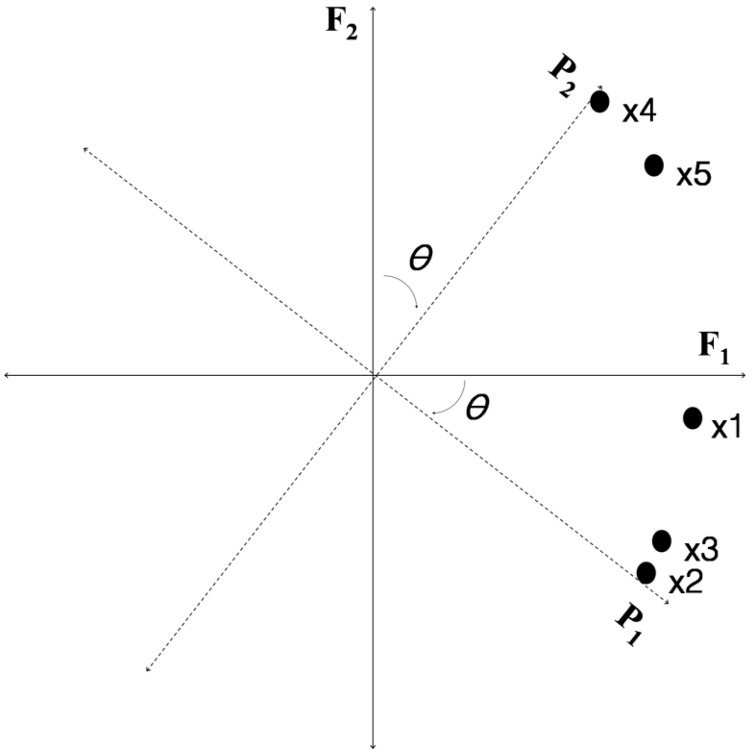
- 반복적 주축인자법: 상관행렬 R의 경우
  - 비반복적 주축인자추정법과 동일하나 다음의 과정을 수렴할 때까지 반복 실행
    1.조정된 상관행렬에 분광분해를 적용한 후 보유인자 개수 ($m \le p$)만큼의 주성분만을 이용하여 Λ을 추정
    2. 1에서 추정된 공통성을 빼서 Ψ을 추정
  - 공통성 $\Lambda\Lambda'$의 초기 추정값 (initial value)
    1. X_i와 나머지 변수들간의 다중상관계수제곱 (squared multiple correlation coefficient)
    2. X_i와 다른 변수들간 상관계수의 절대값 중 최대를 이용

3.4 인자적재와 특수분산의 추정

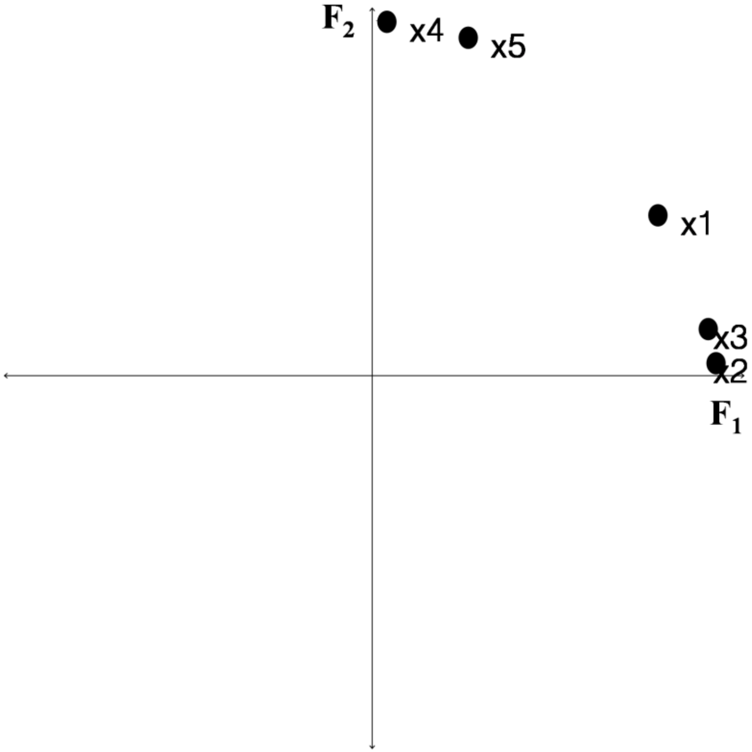
- 최대우도추정법 (maximum likelihood method)
  - ✓ 공통인자 F와 특수인자변수 ε이 정규분포를 따른다고 가정
  - ✓ 자료 X는 다변량 정규 분포 $N_p(\mu, \Sigma)$를 따르게 되며 $\Sigma = \Lambda\Lambda' + \Psi$ 가정하에서 우도함수(likelihood function)를 계산할 수 있다.
  - ✓ 우도를 최대화하는 모수 Λ와 Ψ를 반복적 처리절차를 통하여 추정
  - ✓ 이 추정법은 상대적으로 큰 공통성을 가지는 변수에 큰 가중값을 주는 효과를 가지고 있음

3.5 인자의 회전 (해석의 편이를 위해 회전)

3.5.1 직교회전 (Orthogonal Rotation)



(a) 초기 인자적재행렬



(b) 회전된 인자적재행렬

3.5 인자의 회전 (해석의 편이를 위해 회전)

일반적으로 해석의 편이를 위해 인자를 회전하기 때문에 R 에서 디폴트로 인자를 직교회전한다.

직교회전의 대표적인 방법은 VARIMAX 방법이다.

> ? factanal

Perform maximum-likelihood factor analysis on a covariance matrix or data matrix.

Usage

```
factanal(x, factors, data = NULL, covmat = NULL, n.obs = NA,
subset, na.action, start = NULL, scores = c("none",
"regression", "Bartlett"), rotation = "varimax", control =
NULL, ...)
```

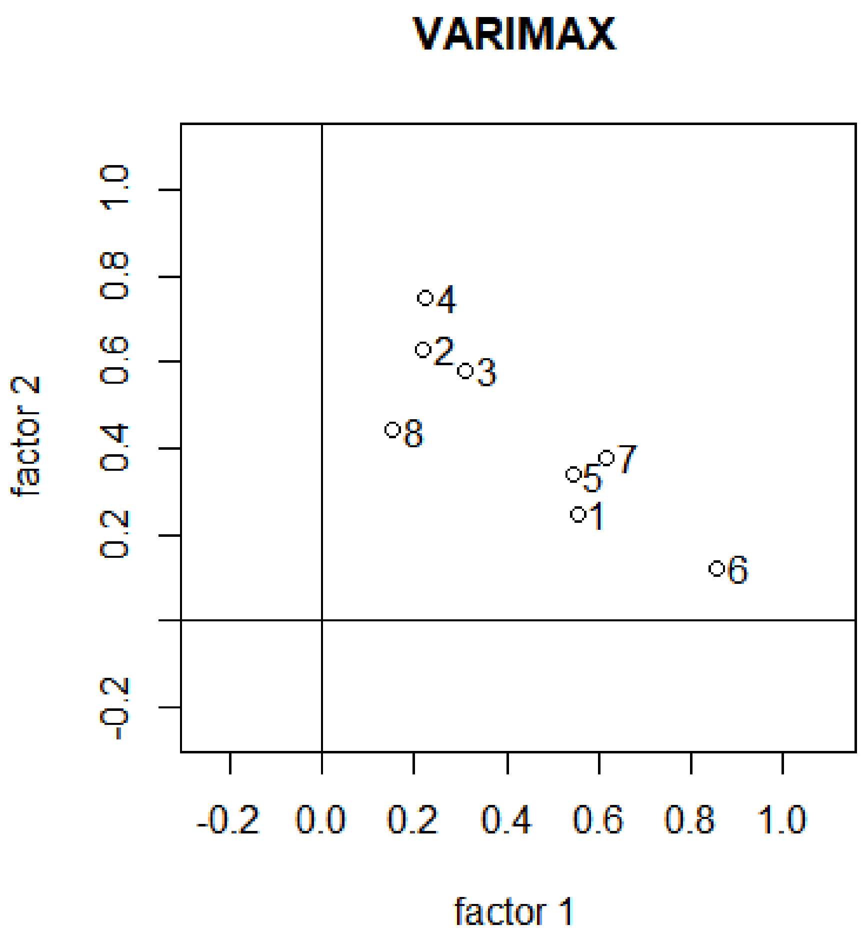
3.5.2 사각회전 또는 비직교회전(Oblique Rotation)

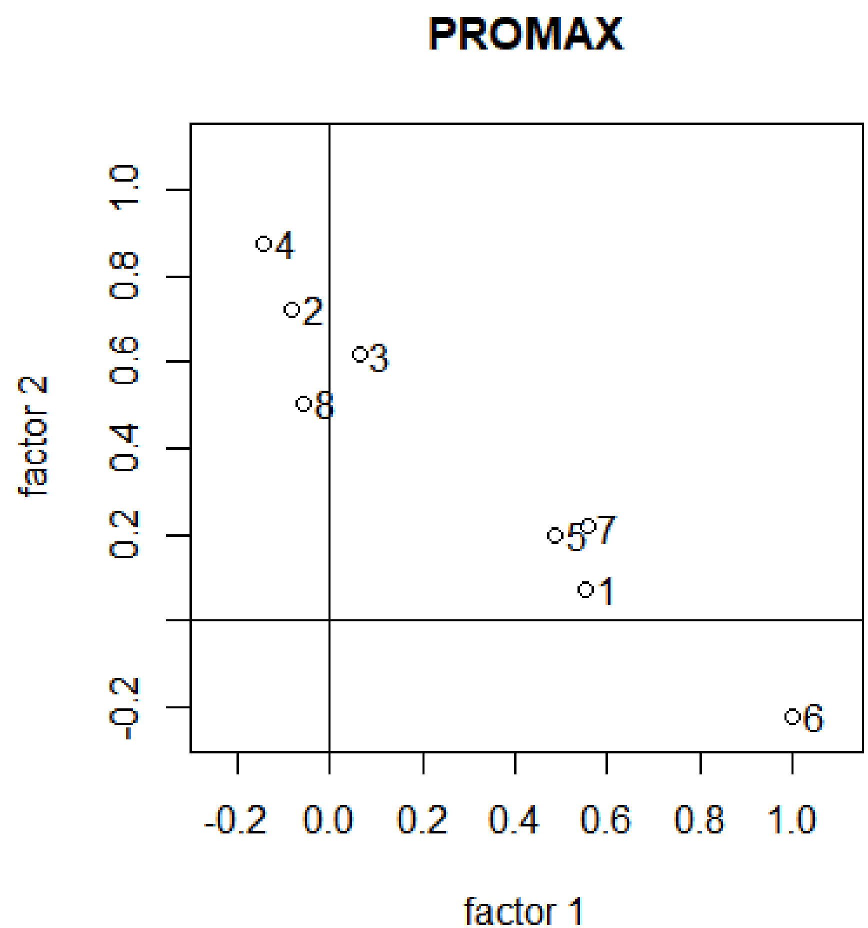
사각회전의 대표적인 Promax Rotation

```
> promax(loadings(my.model2))

 Loadings:
    Factor1 Factor2
 x1  0.553
 x2          0.724
 x3          0.617
 x4 -0.144   0.876
 x5  0.488   0.198
 x6  1.000  -0.223
 x7  0.555   0.219
 x8          0.505

                Factor1 Factor2
 SS loadings      1.888   2.069
 Proportion Var   0.236   0.259
 Cumulative Var   0.236   0.495  # 49.5% 설명됨.
```

R 실행 파일 참고하기.
"Ch03(Rcode).R"

직교회전과 비직교 회전의 로딩 비교





3.6 인자의 개수 결정하는 방법

3.6.1 고유값의 크기를 보고 결정.

- 표본상관행렬 R에 기초하여 분석을 수행하는 경우, 표본상관행렬의 고유값 중 1보다 큰 개수만큼 인자를 보유. (Kaiser의 규칙)
- 일반적으로 연구자가 기대하는 인자의 개수와 일치하는 경우가 많음.

```
> eigen(cor(tv.data))$value
3.5972341 1.1426129 0.7691740 0.6395068
0.5567417 0.5062387 0.4424260 0.3460658
```

3.6.2 인자의 공헌도를 보고 결정

- 관찰변수의 공통성: $c_i = \sum_{j=1}^{m} \lambda_{ij}^2 = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2$
  - ✓ m 개의 인자들에 의해 설명되는 변수 x_i의 분산
- 인자의 공헌도 (설명분산) : $v_k = \sum_{i=1}^{p} \lambda_{ij}^2 = \lambda_{1k}^2 + \lambda_{2k}^2 + \cdots + \lambda_{pk}^2$
  - ✓ 원래변수의 전체분산 중 k 번째 인자에 의해서 설명되는 분산의 양

```
> my.model$loadings
......
......
......
                Factor1 Factor2 Factor3
SS loadings       1.729   1.717   0.647
Proportion Var    0.216   0.215   0.081
Cumulative Var    0.216   0.431   0.512
```

3.6.3 카이제곱 적합도검정을 통해 결정

귀무가설: 현재의 모형이 충분하다.
대립가설: 현재의 모형이 충분치 못하다.

```
> my.model
```

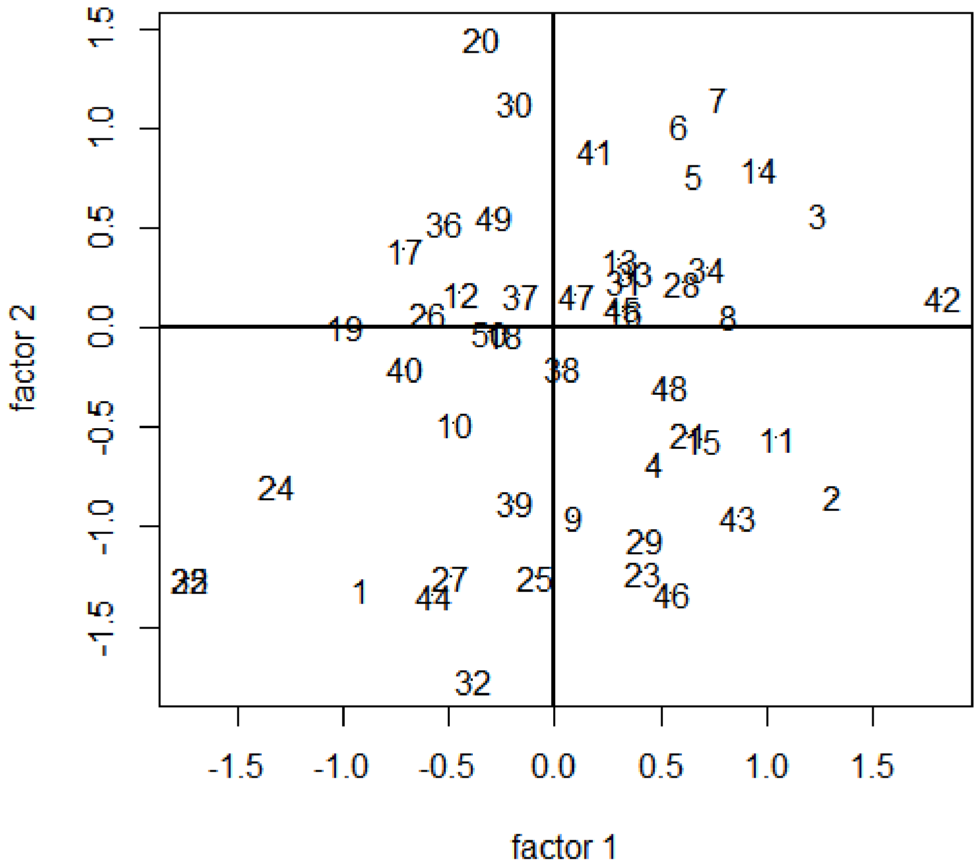
Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 25.45 on 7 degrees of freedom.
The p-value is 0.00063

```
> my.model2
```

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 73.72 on 13 degrees of freedom.
The p-value is 1.64e-10

3.7 인자의 점수화

개개의 자료 (observation)를 추출된 인자 점수를 이용하여 쉽게 파악할 수 있다. (예: 첫 50명의 TV 선호도)



예) ‘인자분석’ 요약

✓ 이충란, 권낙원 (2008). 보건교사의 자기효능감 척도 개발, 지역사회간호학회지, 제19권, 제2호, 247-259.

〈Table 3〉 The Result of the Second Exploratory Factor Analysis (N=378)

Item	Factor structure coefficient 1	2	3	4	Communality
13	.816				.667
11	.799				.667
10	.796				.665
31	.785				.670
4	.779				.621
17	.778				.614
21	.764				.631
25	.744				.565
8	.705				.514
1	.681				.484
34		.840			.711
30		.816			.675
26		.807			.670
23		.800			.655
29		.787			.648
15		.717			.599
32		.650			.504
12			.857		.740
9			.831		.702
7			.813		.668
20			.718		.581
16			.651		.505
14			.587		.561
2				.858	.743
18				.672	.591
28				.635	.471
5				.577	.475
Eigenvalue	11.86	1.91	1.65	1.78	
Explained variable(%)	43.91	7.09	6.12	4.36	
Commutative variable(%)	43.91	50.99	57.12	61.48	(Total)
Number of items	10	7	5	4	26
Cronbach's α	.92	.90	.85	.72	.95
Factor name	Instruction	Task-interpersonal relationship	Community connection	Health-service	

확증적 인자분석 (Confirmatory factor analysis)