

9장 인자분석(FA, factor analysis)

- 인자분석 : 상호연관된 여러 개의 변수들간의 내부적 상호관계를 변수들간의 내재하는 소수의 잠재적인 관측 불가능한 인자를 추출해내는 다변량분석법.
- 대상자료 : 양적자료.
- 종속변수와 독립변수의 구분이 없음.
- 변수 중심의 통계적 기법
- 공분산행렬 또는 상관행렬을 주요 분석 대상
- 차원 축소, 전체 정보를 요약 단순화
- R의 비대각원소들이 거의 0일 때 인자분석에 부적합.
서로 관련되어져 있는 변수들이 있어야 공통 인자를 추출할 수 있음.

- 인자분석의 역사

- 공통인자의 개념은 처음 Galton(1888)에 의해 제안
- 인자분석에 대한 수학적 모형은 Spearman(1904)에 의해 제안
- 다중인자모형 : Garnett(1919), Thurstone(1931)에 의해 제안

- 인자분석 의의

- 다변량 자료

- (i) 서로 상관관계가 있고

- (ii) 직접적으로 해석하기 난해

- 인자분석을 통하여

- (i) 상대적으로 독립적

- (ii) 변수들의 underlying 구조를 이해시키는 개념상 의미 부여

- (iii) 원래 변수들의 개수보다 훨씬 적은 개수의 인자

이들을 통하여 다변량 자료를 분석하고자 한다.

- 공통인자 (common factor)

- 뿌리를 같이하는 인자
- 다변량 자료의 변수들이 구조적 측면에서 서로 공유하고 있는 확률적 변인
- 변수들간의 상관관계를 생성시키는 가설적인, 관찰할 수 없는, 혹은 저변에 깔려있는 인자
- ‘공통’이란 : 공통적인 인자들이 추출되었다면 변수들 간에 더 이상 상관관계가 존재하지 않게 될 것 (Howe, 1955).

(예) (소득수준, 교육정도, 직업의 종류) 의 다변량 자료에 공통인자로 ‘사회적 계층’이라는 가설적 변수를 고려할 수 있다.

- 인자분석의 목적

- ① 차원 축소

- ✓ p차원 (변수의 개수) $\rightarrow$ m차원 (공통인자의 개수)
- ✓ 변동을 최대화 시키는 차원축소 : 주성분분석
- ✓ 공통인자를 통한 차원 축소 : 인자분석

- ② 변수의 그룹화 : 기존 개념의 확인, 새로운 개념의 탐색.

- ✓ 변수 $\rightarrow$ 집단
- ✓ 집단이란 : 그 집단내에 있는 변수들은 서로 높은 상관계수를 가지고, 서로 다른 집단에 속하는 변수들은 상대적으로 낮은 상관관계를 가진다.
- ✓ 집단은 관찰된 상관관계의 생성을 반영하는 기저의 공통인자로 축소될 수 있다.

③ 인자점수(factor scores)를 이용한 회귀분석

- ✓ 다변량자료는 종속변수와 독립변수의 구분이 없다.
- ✓ 원자료를 종속변수, 공통인자변수의 인자점수를 독립변수로 설정
- ✓ 회귀모형 적용 가능
- ✓ 인자점수는 공통요인의 점수로 내재적인 변인의 관찰값을 측정할 수 없을 때, 관찰값을 추정한 값이다.
- ✓ 심리학에서 인자분석이 많이 사용 (Journal : Psychometrics)
- ✓ 어떤 측정할 수 없는 심리요인을 인자분석을 통해 점수화 가능

- 인자분석의 분석 방법

- ① 탐색적 방법 (exploratory method)

- 공통인자와 원래변수들 간에 내재하는 인과관계를 규명.
 - 변수들사이의 관계를 가정하지 않고 적합한 모형을 찾아 나감.

- ② 확증적 방법 (confirmatory method)

- 공통인자들의 개수나 인자들 간의 관계 등 모형에 관한 구체적 가설들을 실제 자료로부터 검정.
 - (i) 가설 설정, 가설에 대한 제약조건 부여
 - (ii) 가설 검정 및 자료의 적합도 검정
 - (iii) 모형 수정

- 주성분분석 vs. 인자분석(factor analysis)의 차이 :

- ① 주성분분석

p 개의 주어진 변수들의 일차결합으로 주성분을 구성하고, 변동의 설명에 대한 기여도가 큰 순서대로 m ($< p$) 개의 주성분을 선택하여, 이들로 전체의 변동을 설명하는 분석법.

통계모형이 없음 (예측할 수 없음).

- ② 인자분석

주어진 변수들을 가상의 공통인자(변수)들의 일차결합으로 나타내고, 이들 인자들 가운데 중요한 m 개의 인자만 선택하여 변수들의 상관구조 설명에 중점.

심리학, 사회과학분야 등에서 응용.

통계모형이 있음 (예측 가능)

(예)다음은 220명의 남학생에 대한 6개 과목 성적의 상관행렬이다
(Johnson & Wichern).

$$\begin{array}{cccccc}
 \text{계일어} & \text{영어} & \text{역사} & \text{산수} & \text{대수} & \text{기하} \\
 R = & \begin{bmatrix}
 1.000 & \mathbf{0.439} & \mathbf{0.410} & 0.288 & 0.329 & 0.248 \\
 & 1.000 & 0.351 & 0.354 & 0.320 & 0.329 \\
 & & 1.000 & 0.164 & 0.190 & 0.181 \\
 & & & 1.000 & \mathbf{0.595} & \mathbf{0.470} \\
 & & & & 1.000 & \mathbf{0.464} \\
 & & & & & 1.000
 \end{bmatrix}
 \end{array}$$

- 계일어, 영어, 역사가 높은 상관관계
 - 산수, 대수, 기하는 서로 높은 상관관계
- (계일어, 영어, 역사), (산수, 대수, 기하) 두 변수 집단, 공통성

$m = 2$ 인 인자분석의 결과

과 목	인자1	인자2
계일어	0.66	0.45
영 어	0.69	0.29
역 사	0.52	0.64
산 수	0.74	-0.41
대 수	0.74	-0.38
기 하	0.67	-0.36

- 인자1에 대응되는 계수(인자적재, factor loading)들은 모두 양수로서 크다. 따라서 인자1은 일반지능(general intelligence)과 같이 지적능력을 나타내는 개념으로 이해할 수 있다.
- 인자2에 대응되는 계수들은 어학과목에는 양수, 수리과목에는 음수가 할당되어 ‘어학과목의 수리과목에 대한 우세성’으로 이해할 수 있다.

- 일반지능의 개념보다는 좀더 구체적인 요인(인자)을 찾기 위해서 인자회전을 시행한다. 다음 표는 varimax 회전에 의한 결과이다.

Varimax 회전의 결과

과목	인자1	인자2
게일어	0.23	0.76
영 어	0.35	0.66
역 사	-0.00	0.82
산 수	0.83	0.15
대 수	0.81	0.18
기 하	0.75	0.15

- 인자회전의 결과 인자1은 수학과목에 높은 계수가, 인자2는 어학과목에 높은 계수가 할당되었다. 따라서 인자1은 ‘수리능력’으로, 인자2는 ‘어학능력’으로 이해할 수 있다.

- m 인자 모형 : 공통인자의 개수가 m 개인 인자 모형

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon},$$

$$x_j = \mu_j + \lambda_{jk}f_k + \epsilon_j, j = 1, \dots, p$$

- $\boldsymbol{\Lambda}$ (인자적재값, factor loading), $p \times m$

λ_{jk} : k 번째 공통인자가 j 번째 변수에 미치는 영향정도

- $\mathbf{f}$ (잠재인자, latent factor, common factor), $m \times 1$

측정불가능한 공통인자

- $\boldsymbol{\mu}$ (평균),

- $\boldsymbol{\epsilon}$ (특정인자, specific factor) : 오차

- 가정 : $\mathbf{f} \sim N_m(0, \mathbf{I})$, 공통인자들은 서로 비상관

$\boldsymbol{\epsilon} \sim N_p(0, \boldsymbol{\Psi})$, 오차 변수들은 서로 비상관이고 유일분산을 가짐

$cov(\mathbf{f}, \boldsymbol{\epsilon}) = \mathbf{0}$ 공통인자와 특정인자는 비상관

- 행렬을 각 변수로 풀어서 모델을 전개하면 (성웅현, 응용다변량분석)
- 가정 : $\mathbf{x}' = (x_1, \dots, x_p) \sim N_p(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu}' = (\mu_1, \dots, \mu_p)$
- 모형 :

$$x_1 = \mu_1 + \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1m}f_m + \epsilon_1$$

$$x_2 = \mu_2 + \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2m}f_m + \epsilon_2$$

$$\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots$$

$$x_p = \mu_p + \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pm}f_m + \epsilon_p$$

$\Rightarrow \mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon}$, with $\Sigma = \text{Cov}(\mathbf{x}) = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$, where

$$\Lambda = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1m} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pm} \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_p \end{bmatrix}, \quad \Psi = \begin{bmatrix} \psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_p \end{bmatrix}$$

위의 모형을 인자패턴(factor pattern) 이라 한다.

- f_j : 공통인자(common factor) – 전체 변수에 공통으로 사용하는 변수(관측 불가능)
- ϵ_i : 유일인자(unique factor) 또는 특정인자(specific factor) – 유일하게 i 번째 변수에만 기여하는 변수로서 오차항과 같은 개념
- λ_{ij} : 인자적재(factor loading).
- $\psi_i = \text{Var}(\epsilon_i)$: 유일분산 또는 특정분산.
- $h_i^2 = \lambda_{i1}^2 + \cdots + \lambda_{im}^2$: 공통인자분산 또는 공통성(communality).

- f 와 ε 에 대하여는 다음과 같은 가정을 한다.
 1. f_j 들은 서로 독립이고, 평균이 0, 분산 1 (즉, $\mathbf{f} \sim (0, I)$)
 2. ε_i 들은 서로 독립이고, 평균이 0, 분산 ψ_i
 3. $\mathbf{f}$ 와 ε 은 서로 독립

참고.

- ① f 와 ε 은 서로 독립이므로 $\text{Var}(x_i) = h_i^2 + \psi_i$
- ② $h_i^2 = \lambda_{i1}^2 + \dots + \lambda_{im}^2$ 은 x_i 의 분산중 m 개의 인자 $f_1, \dots, f_m$ 이 설명하는 부분
- ③ $\text{cov}(x, f) = \Lambda$

- $cov(\mathbf{x}) = \Sigma = \Lambda\Lambda^T + \Psi$: 모수추정의 기본 수식

$\because \mathbf{x} = \mu + \Lambda\mathbf{f} + \epsilon$ 에서 공분산을 취하면

$$\begin{aligned} cov(\mathbf{x}) &= cov(\Lambda\mathbf{f}) + cov(\epsilon) + cov(\Lambda\mathbf{f}, \epsilon) \\ &= \Lambda cov(\mathbf{f}) \Lambda^T + \Psi + \Lambda cov(\mathbf{f}, \epsilon) \\ &= \Lambda\Lambda^T + \Psi \end{aligned}$$

- $\sigma_{jj} = \sum_{k=1}^m \lambda_{jk}^2 + \psi_j = \text{공통성} + \text{특정분산} = x_i \text{와 잠재인자의 분산} + \text{오차}$

$$diag(\Sigma) = diag(\Lambda\Lambda^T) + \Psi = diag\left(\sum_{k=1}^m \lambda_{jk}^2 + \psi_j\right)$$

- Λ 는 직교변환에 유일하지 않다 (비유일성).

$\Lambda_1 = \Lambda D$ (D 는 직교행렬) 이라하면 $\Lambda_1 \Lambda_1^T = \Lambda D D^T \Lambda^T = \Lambda \Lambda^T$ 이다.

따라서 조건을 추가하여 해를 유일하게 만든다.

$\Lambda^T W^{-1} \Lambda = diagonal$ 에서 W 에 따라 추정 방법이 다르다.

● 척도변환의 불변성

– $x_{p \times 1}$ 에 대한 인자분석의 모형은 $x = \mu_x + \Lambda_x f_x + \epsilon_x$ 이다.

– $y_{p \times 1} = Dx$ (D 는 대각행렬) 일 때

$$y = Dx = D(\mu_x + \Lambda_x f_x + \epsilon_x) = D\mu_x + D\Lambda_x f_x + D\epsilon_x \text{ 에서}$$

여기서 $\mu_y = D\mu_x, \Lambda_y = D\Lambda_x, \epsilon_y = D\epsilon_x$ 이다.

– 변환된 y 에 대한 공분산행렬

$$\Sigma_y = D\Sigma_x D = D(\Lambda_x \Lambda_x^T + \Psi_x) D = \Lambda_y \Lambda_y^T + \Psi_y$$

즉 $\Lambda_y = D\Lambda_x, \Psi_y = D\Psi_x D$ 이다.

– 변환된 y 의 인자적재행렬은 D 와 Λ_x 로 구할 수 있다.

– 만약 공분산행렬 Σ 대신에 상관행렬을 이용한 인자분석을 시행하였다면

$$\Lambda_y = \Lambda_x, \Psi_y = \Psi_x$$

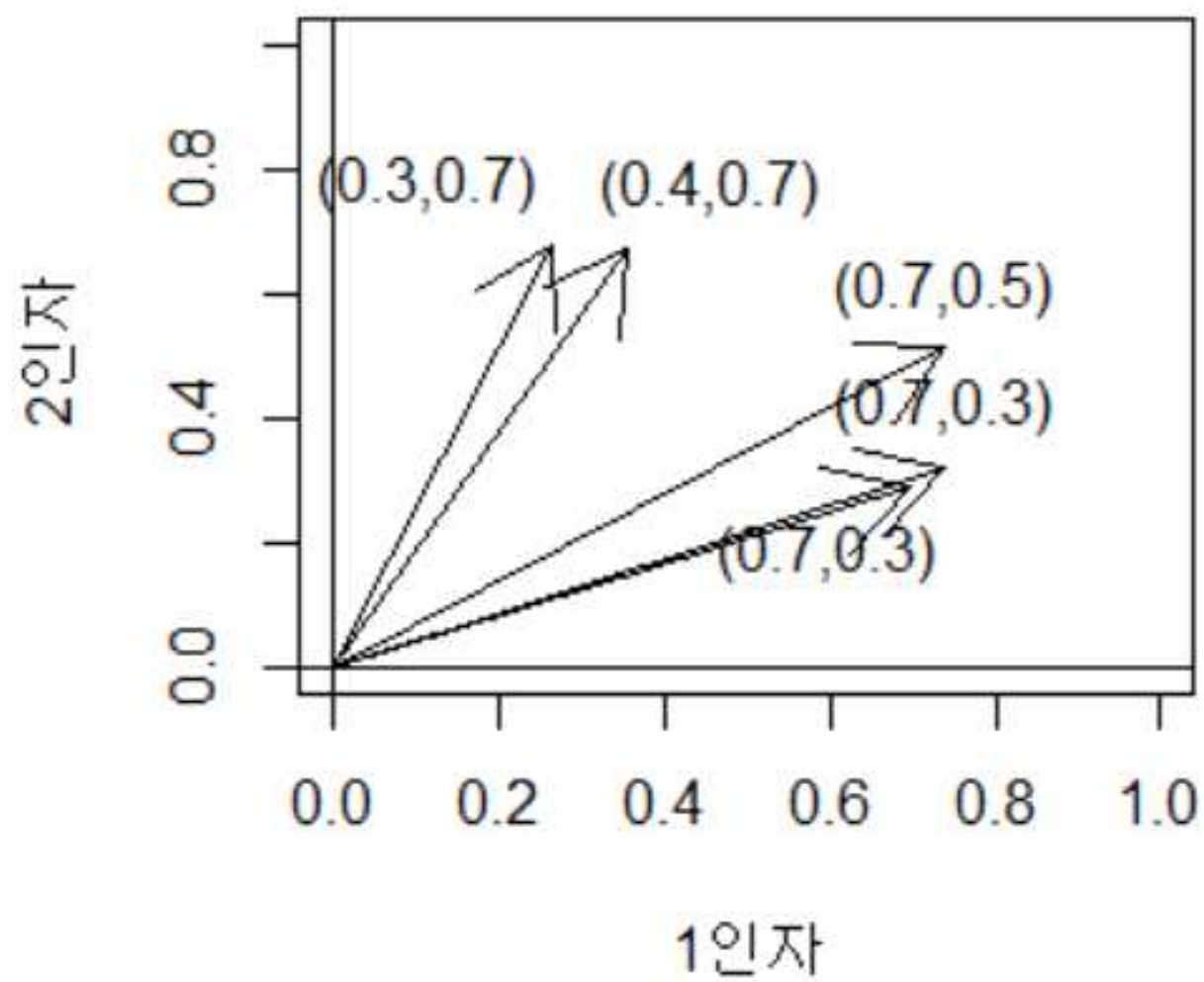
- 인자모형의 기하학적 표현

- 인자를 축으로 변수를 선(또는 벡터)로 표현하고 좌표평면에 그림

(예) oc.d (n=88, p=5) 의 2-인자모형의 인자적재행렬과 유일분산

$$\hat{\Lambda} = \begin{pmatrix} 0.265 & 0.681 \\ 0.356 & 0.674 \\ 0.740 & 0.514 \\ 0.738 & 0.322 \\ 0.696 & 0.290 \end{pmatrix} \quad \hat{\Psi} = \text{diag}(0.465, 0.419, 0.188, 0.351, 0.431)$$

이를 그림으로 표현하면 첫 번째 인자축과 두 번째 인자축의 좌표가 인자적재행렬의 값이다. 첫 번째 변수에 대한 인자적재값 0.265, 0.681이 제 1인자와 제 2인자의 비중이다. 벡터의 길이는 $0.265^2 + 0.681^2 = 0.534$ 가 공통인자에 의한 분산 (공통성)을 나타낸다.



9.2 모수 추정

- 모수의 개수

- $S = \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi}$ 에서

- S 의 자유도는 $p(p+1)/2$,

- $\hat{\Lambda}$ 의 자유도는 pm ,

- $\hat{\Psi}$ 의 자유도는 p ,

- 제약조건의 개수는 $m(m-1)/2$ 이다. 따라서

$$s = p(p+1)/2 - (pm + p - m(m-1)/2) = \frac{1}{2}[(p-m)^2 - (p+m)]$$

이다.

- $s > 0$ 인 경우에 인자분석이 적용될 수 있다.

즉, m 을 1, 2, ... 순차적으로 적용하면서 어느 이상 크게 할 수 없다.

- $s < 0$ 인 경우 : 무수히 많은 해가 존재한다.
- $s = 0$ 인 경우 : 정확한 해를 구할 수 있다. 변수의 상관관계를 설명하는데 Σ 를 더 이상 간단히 할 수 없으므로 변수 축소가 되지 않는다.
- $s > 0$ 인 경우 : 인자분석모형을 고려할 수 있다. 근사적으로 해를 구해야한다. Σ 를 간단히 하여 변수축소가 가능하다.

(예) $p = 5, m = 1$ 인 경우 : $s = \frac{1}{2}[(5-1)^2 - (5+1)] = 5 > 0$

$p = 5, m = 2$ 인 경우 : $s = \frac{1}{2}[(5-2)^2 - (5+2)] = 1 > 0$

$p = 5, m = 3$ 인 경우 : $s = \frac{1}{2}[(5-3)^2 - (5+3)] = -2 < 0$

따라서 $p = 5$ 인 경우 1-인자 직교인자모형, 2-인자 직교인자모형만 모수추정이 가능하다.

(1) 최우추정법 : $\Lambda^T \Psi^{-1} \Lambda = diagonal$ 조건하에서 모수를 추정

- 자료가 다변량 정규분포임을 가정
- 최대우도법을 이용하여 L을 추정하는 방법.
- m-인자모형에 대한 적합도 검정 가능
- 단점 : 추정시간이 오래 소요.
- 공통성의 추정값이 1인 경우 : Heywood case
- 추정방정식 : 다음 식을 반복적으로 수행한다.

$$\Psi^{-1/2}(S_n - \Psi)\Psi^{-1/2} = GLG^T$$

$$\Lambda = \Psi^{1/2} G_1 L_1^{1/2}$$

$$\Psi = diag(S_n - \Lambda \Lambda^T)$$

- 최우추정량은 단위변환에 대하여 불변이다.
공분산행렬, 상관행렬 어느 것을 이용하여도 같은 추정값을 가짐.

```
> a=factanal(oc.d,factors=2,rotation="none") #표본상관행렬이용
> print(a)
```

Call:

```
factanal(x = oc.d, factors = 2, rotation = "none")
```

Uniquenesses: (유일성 Ψ)

mechanics	vector	algebra	analysis	statistics
0.466	0.419	0.189	0.352	0.431

Loadings: (적재인자행렬, Λ)

	Factor1	Factor2
mechanics	0.628	0.373

vector	0.695	0.312
algebra	0.899	
analysis	0.780	-0.201
statistics	0.727	-0.200

	Factor1	Factor2	
SS loadings	2.824	0.319	
Proportion Var	0.565	0.064	(첫번째 인자 설명률이 높음)
Cumulative Var	0.565	0.629	(2인자 모형이 변동의 62.9% 설명)

Test of the hypothesis that 2 factors are sufficient. (m-인자 적합도)

The chi square statistic is 0.07 on 1 degree of freedom.

The p-value is 0.785 (H_0 채택)

- 2인자 모형은 유의하다.

> a\$loadings # Λ

- 첫 번째 공통인자는 일반적인 학습능력
- 두 번째 공통인자는 (mechanics, vector), (analysis, statistics)의 대비. 공간능력과 대수능력의 대비.

> a\$unique # Ψ

- algebra는 변이가 적다
- 그 이외의 변수의 변이는 비슷하다.

$\Sigma = \Lambda\Lambda^T + \Psi$ 의 확인

> L=a\$loadings

> P=diag(a\$unique)

> L%*%t(L)+P - cor(oc.d)

	mechanics	vector	algebra	analysis	statistics
mechanics	-2.671292e-07	1.192266e-05	-2.095624e-04	5.481777e-03	-6.618074e-03
vector	1.192266e-05	-2.637076e-07	1.914151e-04	-5.713624e-03	6.953058e-03
algebra	-2.095624e-04	1.914151e-04	-2.065223e-08	4.192968e-04	-5.713067e-04
analysis	5.481777e-03	-5.713624e-03	4.192968e-04	-4.011667e-08	5.191412e-05
statistics	-6.618074e-03	6.953058e-03	-5.713067e-04	5.191412e-05	-9.569022e-07

- 대각원소는 거의 차이가 없으나 비대각원소는 약간의 차이가 있음.
- 대각원소는 특정분산에 의하여 보정이 되지만, 비대각원소는 공통인자 모형에 의하여 설명되어야 하므로 오차가 발생할 수 밖에 없다.

- 공통인자의 기여도

SS loadings : 2.824 0.319

```
> sum(L[,1]^2)
```

```
[1] 2.824171 # 첫 번째 인자가 5개 변수를 설명하는 양(전체:p)
```

```
> sum(L[,2]^2)
```

```
[1] 0.3194919
```

- 2인자 모형의 변동에 대한 전체 설명력이 62.9%이다.
- 전체변동은 p 이므로 5이다.
- $2.824/5 = 0.565$, $0.319/5 = 0.064$
- 5차원을 2차원으로 설명하므로 정보손실은 있을 수 밖에 없다.
- 저차원의 정돈된 자료를 가지고 분석하느냐,
고차원의 혼재된 자료를 가지고 분석하느냐?

- m=1인 경우

```
> factanal(oc.d,factors=1,rotation="none")
```

Test of the hypothesis that 1 factor is sufficient.

The chi square statistic is 8.65 on 5 degrees of freedom.

The p-value is 0.124

1-인자 모형 충분하지만, 변동설명력이 55.3% 이다.

```
> a=factanal(oc.d,factors=1,rotation="none")
```

```
> round(a$load%*%t(a$load)+diag(a$uniq)-cor(oc.d),2)
```

	mechanics	vector	algebra	analysis	statistics
mechanics	0.00	-0.15	0	0.05	0.04
vector	-0.15	0.00	0	0.03	0.05
algebra	0.00	0.00	0	0.00	0.00

```
> a=factanal(oc.d,factors=2,rotation="none")
> round(a$load%*%t(a$load)+diag(a$uniq)-cor(oc.d),2)
```

	mechanics	vector	algebra	analysis	statistics
mechanics	0.00	0.00	0	0.01	-0.01
vector	0.00	0.00	0	-0.01	0.01
algebra	0.00	0.00	0	0.00	0.00
analysis	0.01	-0.01	0	0.00	0.00
statistics	-0.01	0.01	0	0.00	0.00

-m=1인 경우는 m=2인 경우에 비하여 정확도가 떨어진다.

● m=3인 경우

```
> factanal(oc.d,factors=3,rotation="none")
```

Error in factanal(oc.d, factors = 3, rotation = "none") :

3 factors are too many for 5 variables

3-인자 모형 추정이 불가능하다.

- 직교변환에 대하여 유일하지 않다.
- 직교행렬 $D = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \theta = \pi/4$

```
> A=matrix(c(1/sqrt(2),1/sqrt(2),-1/sqrt(2),1/sqrt(2)),2)
```

```
> A
```

```
      [,1]      [,2]
```

```
[1,] 0.7071068 -0.7071068
```

```
[2,] 0.7071068  0.7071068
```

```
> solve(A)
```

```
      [,1]      [,2]
```

```
[1,] 0.7071068 0.7071068
```

```
[2,] -0.7071068 0.7071068
```

```
> L1=L%*%A
> P1=A%*%P%*%A
> L1%*%t(L1)+P1 - cor(oc.d)
```

	mechanics	vector	algebra	analysis	statistics
mechanics	-2.671292e-07	1.192266e-05	-2.095624e-04	5.481777e-03	-6.618074e-03
vector	1.192266e-05	-2.637076e-07	1.914151e-04	-5.713624e-03	6.953058e-03
algebra	-2.095624e-04	1.914151e-04	-2.065223e-08	4.192968e-04	-5.713067e-04
analysis	5.481777e-03	-5.713624e-03	4.192968e-04	-4.011667e-08	5.191412e-05
statistics	-6.618074e-03	6.953058e-03	-5.713067e-04	5.191412e-05	-9.569022e-07

- 직교변환에 대하여 직교인자모형의 추정량은 불변이다.

● Λ, Ψ : 척도변환에 대하여 불변

```
> a=factanal(oc.d,factors=2,rotation="none")
```

```
> b=factanal(covmat=cov(oc.d),factors=2,rotation="none") # S사용
```

```
> b$unique - a$unique
```

mechanics	vector	algebra	analysis	statistics
2.053913e-15	8.326673e-16	2.831069e-15	-4.440892e-16	-2.775558e-15

```
> b$loadings - a$loadings
```

Loadings:

	Factor1	Factor2
--	---------	---------

mechanics		
-----------	--	--

vector		
--------	--	--

algebra		
---------	--	--

analysis		
----------	--	--

statistics

	Factor1	Factor2
SS loadings	0	0
Proportion Var	0	0
Cumulative Var	0	0

(2) 주성분법

- 공분산행렬을 스펙트럴분해하면 $\Sigma = GLG^T$
- $m = p$ 이면 $\Psi = O$ 이다.

$$\Sigma = \Lambda\Lambda^T + O = \Lambda\Lambda^T$$

이 되어 인자적재값과 주성분계수의 값이 일치하게 된다.

- $m < p$ 이면 근사적으로

$$\Sigma \approx \Lambda\Lambda^T, \text{ 단 } \Lambda = G_1 L_1^{1/2}, G_1 : p \times m, L_1 : m \times m \text{ 행렬}$$

$$\psi_{ii} = \sigma_{ii} - \sum_{j=1}^m \lambda_{ij}^2$$

● 행렬에 의한 추정 절차

1. 초기값으로 $\hat{\Psi} = 0$ 그리고 $\hat{\Lambda}$ 구하고 $\hat{\Psi}$ 구한다.
2. 추정방정식

$$S - \Psi = GLG^T$$

$$\Lambda = G_1 L_1^{1/2}$$

$$\Psi = \text{diag}(S - \Lambda \Lambda^T)$$

● 주성분인자법의 성질

- S 를 주성분분석에 의해 분해하므로 주성분법이라 한다.
- S 대신 S_n 을 이용하여도 같은 결과를 얻는다.
- 변수사이의 단위변화가 큰 경우에 상관행렬을 이용한다.
- $\Lambda^T \Lambda = L_1^{1/2} G_1^T G_1 L_1^{1/2} = L_1$ 는 대각행렬이므로 식별성 조건 만족
- S 또는 R 이 정칙행렬(non-singular)을 요구하지 않는다.
- i 번째 인자에 의해 설명되는 분산비율 : $\frac{\hat{l}_i}{s_{11} + \dots + s_{pp}}$
- 상관행렬이용 : i 번째 인자에 의해 설명되는 분산비율 : $\hat{l}_i/p$

(3) 주축인자법 (김제희 교수 강의노트 발췌)

- 주성분법에 의해 Ψ 를 추정한 후 인자적재행렬을 구한다.
- 추정 절차 (S 이용)
 - ① S 을 이용한 주성분법에 의해 $\hat{\Psi}$ 를 구한다.
 - ② $S - \hat{\Psi} \approx \Lambda \Lambda^T$ 관계식을 이용하여 $\Lambda = G_1 L_1$ 을 구한다. 단, $S - \hat{\Psi} = GLG^T$
 - ③ 공통성 : $h_i^2 = s_{ii} - \hat{\psi}_i$ 으로부터 $\hat{\psi}_i$ 를 추정한다.
 - ④ 공통성이 수렴할 때까지 2, 3 절차 반복 수행.

단, 공통성 초기 추정값은 $h_i^2 = s_{ii} - \frac{1}{s^{ii}} = s_{ii} R_i^2$, $S^{-1} = (s^{ii})$, R_i^2 은

X_i 와 나머지 변수들의 다중상관계수이다.

- 반복수렴이 안 될 경우 : $h_i^2 < 0$ 인 경우 발생.

- 추정 절차 (R 이용)

- ① R 을 이용한 주성분법에 의해 $\hat{\Psi}$ 를 구한다.
- ② $R - \hat{\Psi} \approx \Lambda \Lambda^T$ 관계식을 이용하여 $\Lambda = G_1 L_1$ 을 구한다. 단, $R - \hat{\Psi} = GLG^T$
- ③ 공통성 : $h_i^2 = 1 - \hat{\psi}_i$ 으로부터 $\hat{\psi}_i$ 를 추정한다.
- ④ 공통성이 수렴할 때까지 2, 3 절차 반복 수행.

단, 공통성 초기 추정값은 $h_i^2 = R_i^2 = 1 - \frac{1}{r^{ii}}$, $R^{-1} = (r^{ii})$

- 다중상관계수제곱 (squared multiple correlation, SMC) : R_i^2
- 주성분방법보다 인자의 수가 많아지는 경향이 있다.

- Heywood case

- 인자에 대한 공통분산은 다중상관계수의 제곱. $0 \leq h_i^2 \leq 1$
- Heywood case : $h_i^2 = 1$ 인 상황
- ultra-Heywood case : $h_i^2 > 1$ 인 상황
- 결과 : 특수분산의 추정값 $\psi_i \leq 0$, 인자분석의 결과가 유의성이 없다.

- Heywood case 발생 원인

- 공통성에 관한 부적절한 초기값
- 공통인자수가 너무 많거나
- 공통인자수가 너무 적거나
- 안정적인 추정값을 계산하기 부족한 양의 데이터가 있거나
- 공통인자모형 적합이 부적절한 데이터인 경우

● 인자분석의 단계

1. 인자적재행렬 Λ 추정
2. 인자의 수 결정 (m)
3. 모수 Λ, Ψ 추정
4. 인자점수 추정

9.3 적합도 검정과 인자 수의 결정

(1) 검정에 의한 결정

$H_m : \Sigma = \Lambda \Lambda^T + \Psi$: m-인자 모형이 적합하다.

대립가설 : Not H_m

– Bartlett(1951)가 우도비검정법을 이용하여 검정통계량은

$$T = \left(n - \frac{2p + 4m + 11}{6} \right) \ln \left(\frac{|\hat{\Lambda} \hat{\Lambda}^T + \Psi|}{S} \right)$$

– H_0 하에서 $T \sim \chi^2((p-m)^2 - (p+m))/2$ 이다.

– 검정통계량 : $-2 \log \lambda = -n \sum_{i=k+1}^n \log(1 + l_i)$,

l_i 는 $\hat{\Psi}^{-1/2}(S - \hat{\Psi})\hat{\Psi}^{-1/2}$ 의 고유치

- R에서 `factanal()`의 p-value 이다.
- p-value가 유의수준 보다 작으면 m 크기를 늘려 인자분석 적용

(2) 총 변이에 대한 공헌도 (percentage of total variance)

공통인자의 총 설명비율이 70~90%가 되도록 m 을 결정한다.

(3) 평균고유값 (average of eigenvalues)

- 공분산행렬 이용 : 고유값의 평균보다 큰 고유치들의 개수.
- 상관행렬 이용: 고유값 평균=1, 1보다 큰 고유치들의 개수

(4) 스크리도표

- j 번째 인자의 설명비율 : $\frac{l_j}{\sum_{i=1}^p s_{ii}}$;; scree plot에서 급격히 변화

(예) oc.d (n=88, p=5)

#검정에 의한 인자수 결정

```
> factanal(oc.d,2,rotation="none")
```

The chi square statistic is 0.07 on 1 degree of freedom.

The p-value is 0.785

```
> oc.fa2=factanal(oc.d,factors=2,rotation="none")
```

```
> psi=oc.fa2$unique
```

```
> A=diag(psi^(-0.5))%*%(cor(oc.d)-diag(psi))%*%diag(psi^(-0.5))
```

```
> l=eigen(A)$val
```

```
> T0=-88*sum(log(1+l[3:5]))
```

```
> 1-pchisq(T0,1)
```

```
[1] 0.7790251
```

```
# 공통인자의 공헌도
```

```
> oc.fa2$load
```

```
actor1 Factor2
```

```
SS loadings      2.824    0.319
```

```
Proportion Var   0.565    0.064
```

```
Cumulative Var   0.565    0.629
```

```
> oc.fa1=factanal(oc.d,1,rotation="none")
```

```
> oc.fa1$load
```

```
Proportion Var   0.553
```

```
# 평균 고유값
```

```
> eigen(cor(oc.d))$val
```

```
[1] 3.1809801 0.7395718 0.4449651 0.3878924 0.2465905
```

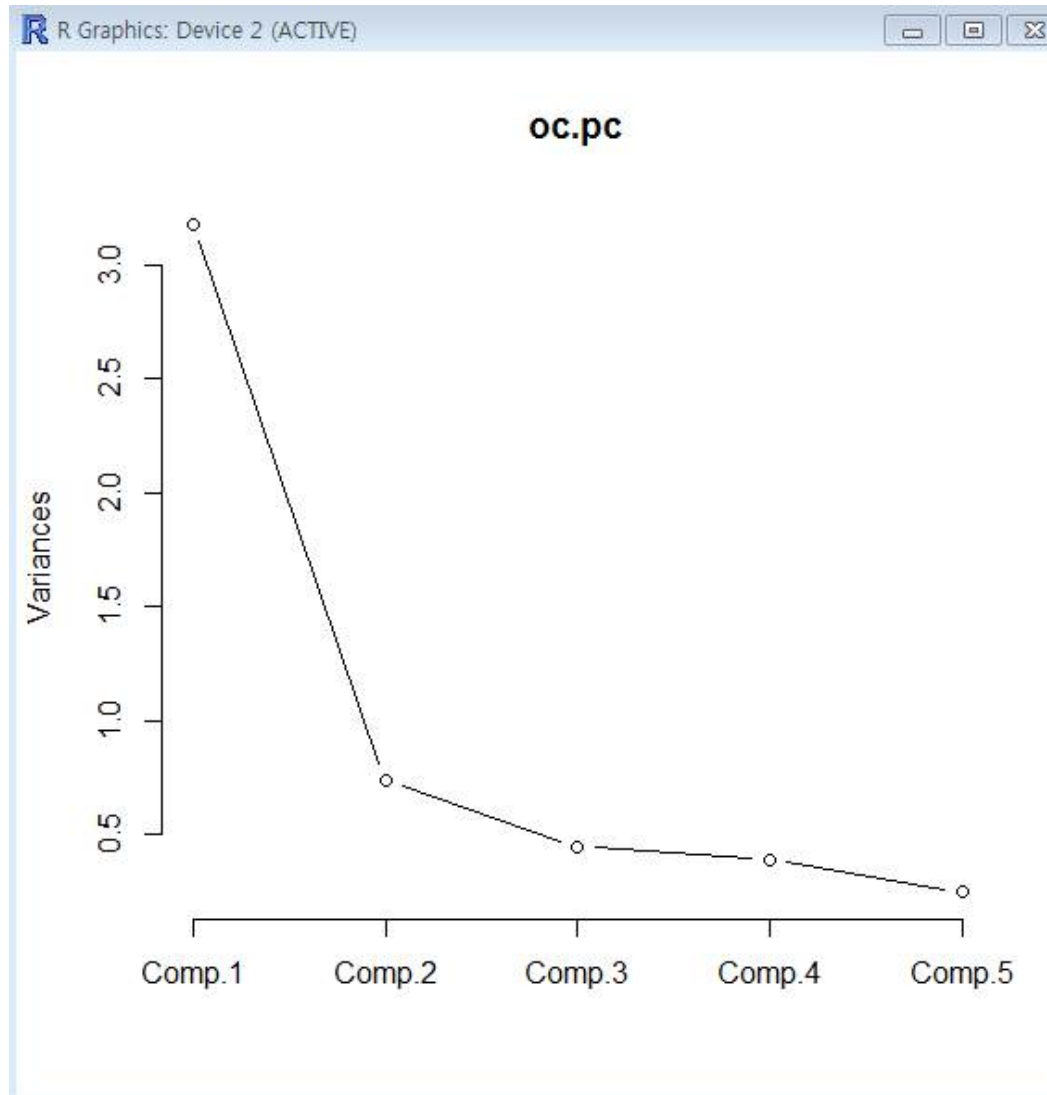
*상관행렬을 이용하므로 1보다 큰 것의 개수 : 1 공통인자

```
# 스크리도표
```

```
> oc.pc=princomp(oc.d,cor=T)
```

```
> plot(oc.pc,type="l")
```

그림에서 1개 이후에 떨어지는 정도가 약간 작아진다. 그러나 인자 3이 후에는 경사도가 0에 매우 가까우므로 인자의 수를 2로 결정하는 것이 타당하다.



9.4 인자의 회전

- 직교인자모형에서 공통인자의 인자적재행렬의 이상적 형태

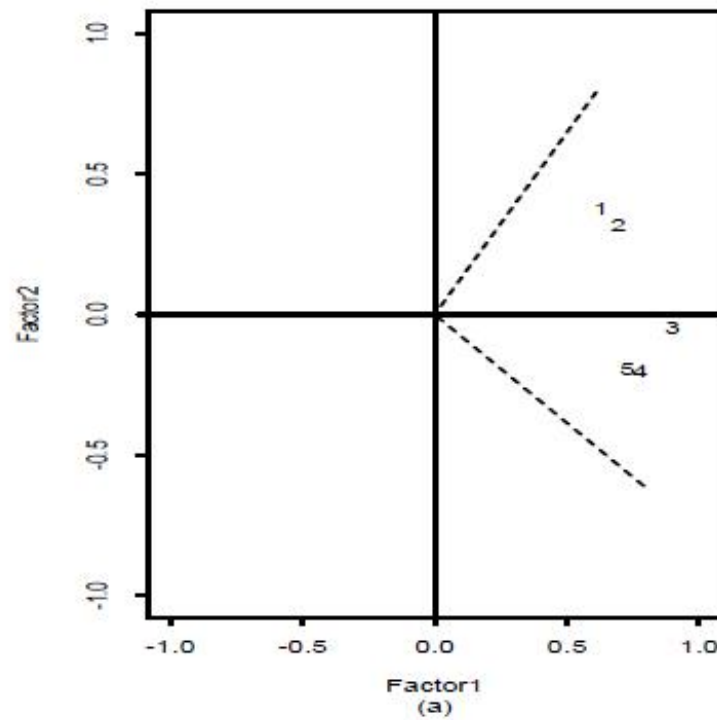
$$p=5, m=2 \text{ 인 경우 : } \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}$$

- 하나의 공통인자가 여러 개의 원래변수에 영향을 주고,
- 하나의 원래변수는 하나의 공통인자에 영향을 받는다.

Q : 인자적재행렬을 어떻게 이상적인 형태로 변환할 수 있는가?

A : 직교변환에 의한 인자적재행렬은 여전히 인자적재행렬이므로 적절한 변환행렬을 찾으면, 이상적인 적재행렬을 구할 수 있다.

- 인자회전 : 인자적재행렬을 적절한 변환으로 해석이 용이한 단순한 구조로 만드는 것.
- 아이디어 : 축을 이동



(1) 직교회전

(1-1) 기하적 회전

- 두 개의 주요 인자축이 서로 직교하도록 원래 인자축을 θ 만큼 회전
- 각 θ 회전형렬 : θ 가 주어진 경우

$$\Lambda^* = \Lambda G, G = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

$$(\lambda_{i1}^*, \lambda_{i2}^*) = (\lambda_{i1}\cos\theta - \lambda_{i2}\sin\theta, \lambda_{i1}\sin\theta + \lambda_{i2}\cos\theta)$$

- 원래인자축 vs 주요인자축 (변환후 인자축)
- 주요인자축은 직교하므로 서로 비상관
- $m > 2$ 인 경우에는 적용하기 어렵다.

(1-2) 해석적 방법

$$\Lambda^* = \Lambda G, G^T G = G G^T = I$$

- 변환행렬 G : 직교행렬 $\rightarrow$ 직교회전
- 변환 행렬 G 를 찾는 방법

$$\sum_{i=1}^p \left[\sum_{j=1}^m \lambda_{ij}^{*2} \right]^2 = \sum_{i=1}^p \sum_{j=1}^m \lambda_{ij}^2 + 2 \sum_i \sum_{k>j} \lambda_{ij}^2 \lambda_{ik}^2$$

- Neuhaus & Wringley (1954) : 우변 끝항을 최대로 하는 G
- Kaiser(1958) : varimax : 인자적재값들의 제곱의 분산을 최대화

$$V^* = \sum_i \left[\sum_j \lambda_{ij}^{*4} - \left(\sum_j \lambda_{ij}^{*2} \right)^2 / p \right]$$

V^* 를 최대로 하는 G 선택

- 각 인자에 관한 인자적재계수의 제곱을 가장 많이 흐트러지게
- 많은 연구자들에 의해 추천되는 인자회전 방법 : 베리맥스법

(2) 사각회전

- 해석적 편리를 위해 비직교회전 (nonorthogonal : oblique rotation)
- 변환된 두 축이 직교일 필요가 없다.
- 변환된 인자적재행렬만 이상적인 경우로 변환한다. → 해석 이점
- 사각회전에서 변환행렬 Q (비직교행렬, 비정칙치행렬)

$$\begin{aligned} f^* &= Q^T f \\ \Lambda^* &= \Lambda Q \end{aligned}$$

이므로

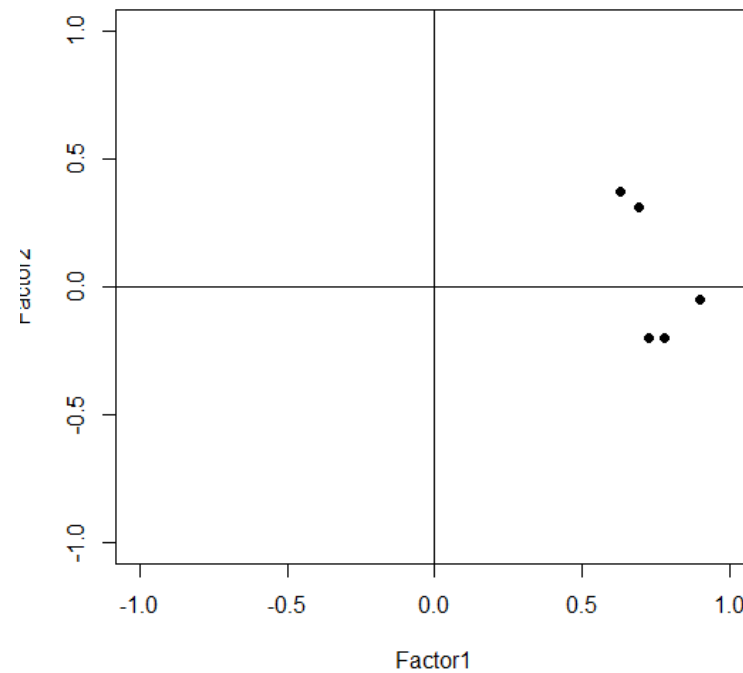
$$\text{cov}(f^*) = Q^T I Q = Q^T Q \neq I$$

이다. 사각회전된 인자 f^* 는 비상관되어 있지 않다.

- 변환후 내재된 인자들의 상관성을 설명해야 하는 문제가 대두
- 최근에는 유도된 인자들이 서로 비상관일 필요가 없다
- 사각회전 : Oblimax, Promax 방법 등

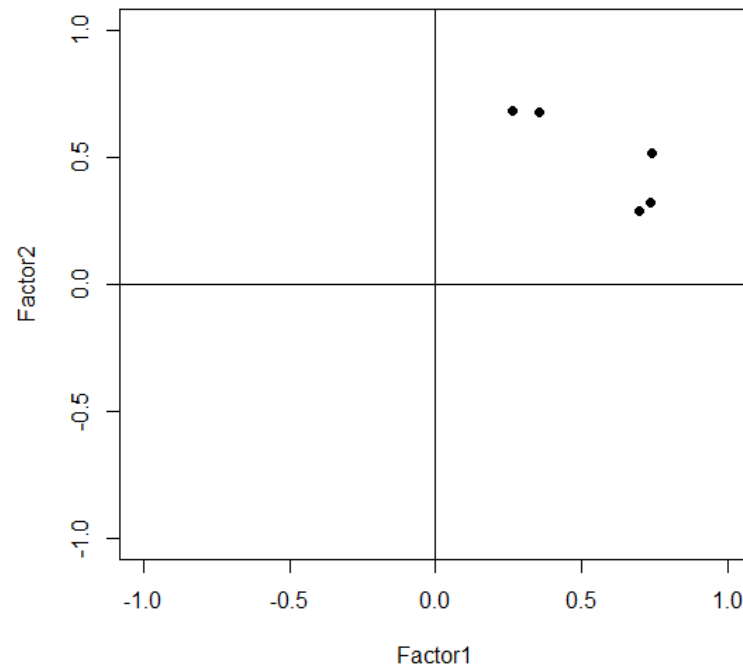
(예) 무회전

```
> oc.fa=factanal(oc.d,factors=2,rotation="none")  
> plot(oc.fa$load,xlim=c(-1,1),ylim=c(-1,1),pch=16)  
> abline(h=0,v=0)
```



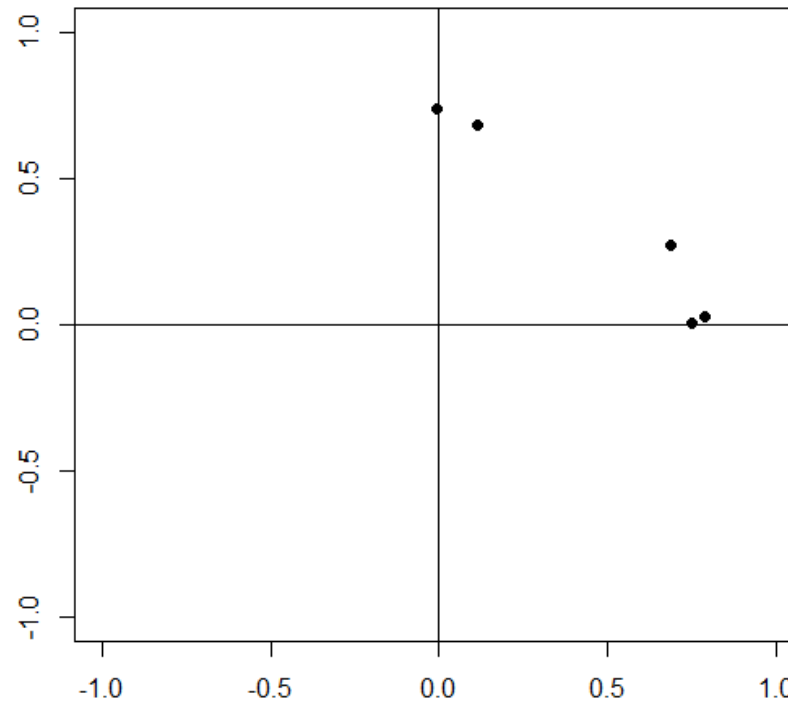
varimax 회전

```
> oc.fa1=factanal(oc.d,factors=2,rotation="varimax")  
> plot(oc.fa1$load,xlim=c(-1,1),ylim=c(-1,1),pch=16)  
> abline(h=0,v=0)
```



promax회전

```
> oc.fa2=factanal(oc.d,factors=2,rotation="promax")  
> plot(oc.fa2$load,xlim=c(-1,1),ylim=c(-1,1),pch=16)  
> abline(h=0,v=0)
```



- 회전행렬 구하기

```
> varimax(oc.fa$load)
```

```
$loadings
```

Loadings:

	Factor1	Factor2
mechanics	0.265	0.681
vector	0.356	0.674
algebra	0.740	0.514
analysis	0.738	0.322
statistics	0.696	0.290

Factor1 Factor2

SS loadings	1.774	1.370
Proportion Var	0.355	0.274
Cumulative Var	0.355	0.629

\$rotmat # 변환행렬

	[,1]	[,2]
--	------	------

[1,]	0.7880178	0.6156524
------	-----------	-----------

[2,]	-0.6156524	0.7880178
------	------------	-----------

>promax(oc.fa\$load)

9.5 잠재인자값 (인자점수, factor score)

- 잠재인자 f 의 추정 : 관측불가능한 독립변수 추정
- 추정된 인자점수를 통하여 후속 통계분석 가능
- 회귀분석방법과 Bartlett 방법이 있음.

9.5.1 회귀분석방법

- Thompson(1951) 제안
- $cov(x, f) = \begin{pmatrix} var(x) & cov(x, f) \\ cov(x, f) & var(f) \end{pmatrix} = \begin{pmatrix} \Lambda\Lambda^T + \Psi & \Lambda \\ \Lambda & I \end{pmatrix}$
- 다변량정규분포 가정하에서

$$E(f|x) = \mu_f + \Sigma_{fx}\Sigma_{xx}^{-1}(x - \mu_x)$$

- 인자모형으로부터 $\mu_f = 0, \Sigma_{fx} = \text{cov}(f, x) = \Lambda^T$ 이므로

$$E(f|x) = \Lambda^T \Sigma_{xx}^{-1} (x - \mu_x)$$

- 따라서 인자점수는

$$\hat{f}_j = \hat{\Lambda}^T S^{-1} (x_j - \bar{x}), j = 1, \dots, n$$

9.5.2 가중최소제곱법 (Bartlett, 1937)

- $\sum_{i=1}^p \epsilon_i^2 / \psi_i$ 를 최소화

- $\sum_{i=1}^p \epsilon_i^2 / \psi_i = \epsilon^T \Psi^{-1} \epsilon$

$$= (x - \mu - \Lambda f)^T \Psi^{-1} (x - \mu - \Lambda f)$$

$$= f^T \Lambda \Psi^{-1} \Lambda f + f^T \Lambda^T \Psi^{-1} (x - \mu) + (x - \mu)^T \Psi^{-1} \Lambda f - (x - \mu)^T \Psi^{-1} (x - \mu)$$

- $\hat{f} = (\hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^T \hat{\Psi}^{-1} (x - \hat{\mu})$

(예)

```
> f1=factanal(oc.d, factors=2, scores="regression")$scores
```

```
> f2=factanal(oc.d, factors=2, scores="Bartlett")$scores
```

```
> f1[1:3,]
```

	Factor1	Factor2
--	---------	---------

[1,]	1.167431	1.846969
------	----------	----------

[2,]	1.988234	1.541268
------	----------	----------

[3,]	1.427669	1.570329
------	----------	----------

```
> f2[1:3,]
```

	Factor1	Factor2
--	---------	---------

[1,]	0.739673	2.755845
------	----------	----------

[2,]	2.210417	1.698799
------	----------	----------

[3,]	1.312477	2.085236
------	----------	----------

- 1번 학생의 경우 factor1보다 factor2의 점수가 더 좋다. 이는 open book 일 때보다 closed book일 때 더 성적이 뛰어난을 나타낸다. 이는 적용능력보다 암기능력이 더 우수하다고 할 수 있다.
- 회귀분석방법 보다 가중최소제곱법의 방법이 요인 점수의 차이가 더 크게 나타난다. 이는 가중치를 부여함으로써 두 잠재요인에 대한 간격을 차별성을 두기 때문이다.

인자분석전략

1. 공분산행렬을 분석하여 공통인자가 존재하는지를 검사한다.
2. 주성분을 이용한 인자분석을 실행한다. (변환 고려)
3. 최대우도를 이용한 인자분석을 실행한다. (변환 고려)
4. 위 두 결과를 비교하여 차이가 크면 이유를 찾아야한다. 특히, 이과정에서 다변량 정규성에 대한 검정을 필요로 한다.
5. 공통인자의 개수 m 을 정하여 순차적으로 1, 2, 3을 실행한다. 자료가 가지는 특징에 의해 상정할 수 있는 인자의 개수와 비교한다. (사회과학분야에서 특히 이런 경우가 많다)
6. 인자적재행렬을 해석하여 실제 자료가 가지는 의미와 비교하여 분석한다.