Chapter 5. Factor Analysis (요인분석)

▪5.1 개요

*변수들의 상관관계를 이용하여 요인(공통개념)을 구하고 이를 이용하여

1) 변수들을 분류

2) 그룹에 적절한 의미를 부여

Ex) *학생들의 학교 만족도→ 조교, 행정인력,강의실, …

* A기업 지원자 48명의 능력 측정 →15개 항목점수 분류

* 기업 관련 지표 → 20개 항목(매출액, 종업원수, 부채비율…)

▪5.1.1 개념

- Spearman(1904)
  - 6개 분야 능력에 의해 학생들의 intelligence 측정
  - 상관계수에 의해 6개 분야 를 적절히그룹화 가능? (언어, 수리 능력)
  - 그러나 상관계수 복잡하게 얽혀 있음

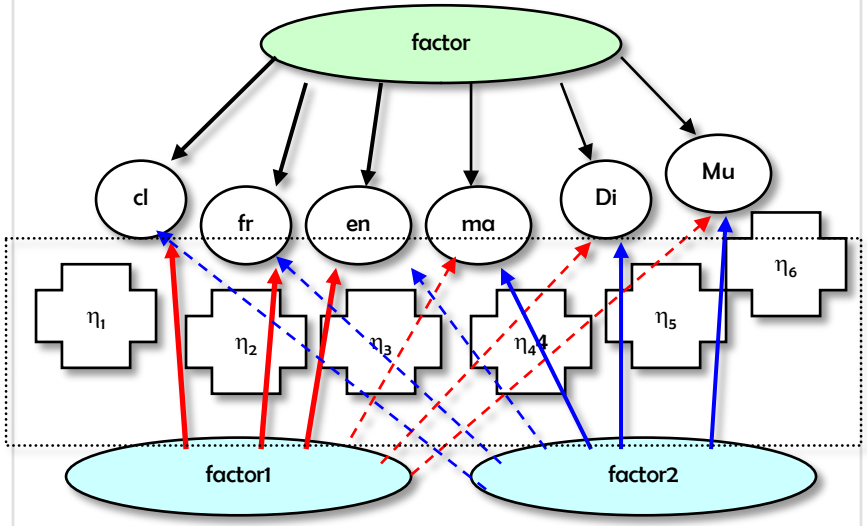
	Classic	French	English	Math	Dis.	Music
Classic	1	**0.83**	**0.78**	0.7	0.66	0.63
French		1	0.67	**0.67**	**0.65**	0.57
English			1	0.64	0.54	0.51
Math				1	0.45	0.51
Discovery					1	0.4
Music						1

Spearman은 변수 간에 내재된 공통개념(f: 이를 factor라 함)부분과 랜덤 부분(η)으로 나눌 수 있다고 생각. 학생들의 지적능력은 일반적 재능으로 해석되는 인자(f), 각 과목에 대한 특별 재능(η) 으로 나눌 수 있다고 믿음.

1) 모형

$$classic = \lambda_1 f + \eta_1,$$
$$french = \lambda_2 f + \eta_2$$
$$english = \lambda_3 f + \eta_3$$
$$math = \lambda_4 f + \eta_4$$
$$discover = \lambda_5 f + \eta_5$$
$$music = \lambda_6 f + \eta_6$$

2) 도식화



▪5.1.2 주성분분석과 비교(p.127)

1) 주성분분석 : p개의 원 변수를 1~2개로 축약
원 변수들의 선형결합으로 만들어짐
모든 원 변수의 영향이 있음

2)요인분석 : p개의 변수들 간의 상호 관계 →
m(< p)개의 변수 그룹으로 나눔
그룹 내 변수들 간에는 상관계수가 높고
다른 그룹 변수 간에는 상관관계가 낮다.

▪요인의 부하(loading)값을 구하는 방법으로 주성분을 이용한 방법을 가장 많이 사용하므로 주성분분석과 유사하다.

주성분 분석	요인 분석
주성분은 원 변수의 직교 선형 결합으로 표현 Y=LX, L은 선형계수 행렬	인자들의 직교 선형 결합으로 원 변수들을 표현 X=LF+η, L 부하행렬, F 관측불가
주성분은 변수들의 변동을 설명	요인은 변수들의 분산-공분산 구조 설명
요인분석이나 주성분 분석의 L을 구하는 방법 유사하다. 공분산 행렬, 상관행렬로부터 고유치 그에 대응하는 고유 벡터를 이용	
선형계수 행렬 L은 변수의 개수 축약하는데 사용되며 는 주성분의 이름을 붙이는데 사용	부하 행렬 L은 변수에 내재된 관계를 알아보는데 사용되며 는 변수들을 그룹화 하는데 사용
▷적절한 주성분의 수를 구하고 주성분의 이름을 부여하고 주성분 점수를 계산한다. ▷데이터 스크린, 이상치 발견, 개체 군집/판별 등에 이용	▷적절한 인자의 수를 구하고 이를 이용하여 변수들을 그룹화 하고 그룹을 이용하여 변수에 내재된 관계를 알아본다. ▷요인점수를 주성분 점수와 유사하게 활용 가능

5.2 모형

5.2.1 목적

(1) 원 변수에 내재된 관계를 설명한 공통 개념 탐색

(2) 공통 개념(요인)의 개수를 선택

(3) 요인의 부하 값을 이용하여 원 변수를 그룹화하고 절적한 이름을 부여한다.

(4) 그룹 내 변수들로 새로운 변수를 만들어 개체들을 평가하고 향후 연구에 활용한다.

5.2.2 모형 및 가정

$\underline{x}' = (x_1, x_2, \cdots, x_p)$의 평균벡터를 $\underline{\mu}$, 공분산 행렬 $\sum$

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\, l_{12} \ldots l_{1m} \\ l_{21}\, l_{22} \ldots l_{2m} \\ \cdots \\ l_{p1}\, l_{p2} \ldots l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_p \end{pmatrix} \Leftrightarrow \underline{x} = L\underline{f} + \underline{\eta}$$

1) 모형에 대한 설명

① x_i는 i번째 원변수 ▷ x: 원변수 벡터

② $f_1, f_2, \cdots f_m$ 공통요인 (common factor) ▷ f: 요인 벡터

③ l_{ij}는 i번째 변수에 j번째 요인이 미치는 영향 ▷ L: 부하 행렬 (factor loading matrix)

④ η_j는 j번째 원변수에 대한 오차항, 특정요인 (specific factor)라 함

2) 모형에 대한 가정

① f_k는 서로 독립이고 평균 0, 분산 1인 동일분포를 따른다.(k=1,2,3...m)

$$E(f_k) = 0_{m\times 1}, Cov(f_k) = E(f_k f_k`) = I_{m\times m}$$

$$\underline{f} \sim iid(\underline{0}_{m\times 1}, I_{m\times m})$$

② η_j서로 독립이고 평균 0, 분산 ψ_i인 동일분포를 따른다. (j=1,2,…,p)

$$E(\underline{\eta}) = 0_{p\times 1}, Cov(\underline{\eta}) = \psi_{p\times p} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \psi_p \end{bmatrix}$$

$$\underline{\eta} \sim iid(\underline{0}_{p\times 1}, \psi_{p\times p} = diag(\psi_1, \psi_2, .., \psi_p)$$

즉, ψ 는 대각행렬이다.

③ f_i와 η_i는 서로 독립이다

$$Cov(\underline{\eta}, \underline{f}) = 0$$

5.2.3 모형 해석

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\ l_{12} \ldots l_{1m} \\ l_{21}\ l_{22} \ldots l_{2m} \\ \cdots \\ l_{p1}\ l_{p2} \ldots l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_p \end{pmatrix} \Leftrightarrow \underline{x} = L\underline{f} + \underline{\eta}$$

1) 공통요인은 원변수 공분산을 완전히 설명한다

$$\Sigma = Cov(\underline{x}) = Cov(L\underline{f} + \underline{\eta})$$
$$= LCov(\underline{f})\ L' + \psi = LL' + \psi$$

그러므로,

$$\Sigma = L\,L' + \psi$$
$$\underline{\eta} \sim (\underline{0}, \psi = diag(\psi_1, \psi_2, .., \psi_p))$$

즉, 인자모형에서 p 개의 분산과 p(p-1)/2개의 공분산구조는 pm개의 인자적재 l_{ij} 와 p 개의 특정분산 ψ_i 에 의해 생성.

→ 왜냐하면 ψ 은 대각행렬으므로 대각원소를 제외하면 0이다.

→행렬로 정리를 하면 다음과 같다.

$$\begin{bmatrix} \sigma_1^{\,2} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^{\,2} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^{\,2} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} l_{1j}^2 + \psi_1 & \sum_{j=1}^{m} l_{1j} l_{2j} & \cdots & \sum_{j=1}^{m} l_{1j} l_{pj} \\ \sum_{j=1}^{m} l_{2j} l_{1j} & \sum_{j=1}^{m} l_{2j}^{\,2} + \psi_2 & \cdots & \sum_{j=1}^{m} l_{2j} l_{pj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} l_{pj} l_{1j} & \sum_{j=1}^{m} l_{pj} l_{2j} & \cdots & \sum_{j=1}^{m} l_{pj}^{\,2} + \psi_p \end{bmatrix}$$

2) i번째 원변수의 변동 (분산)은

$$Var\ (x_i) = \sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2 + \Psi_i = \sum_{k=1}^{m} l_{ik}^2 + \psi_i$$

-공통성(communality)+특정분산(specific variance)

→공분산행렬 대신 상관계수 행렬을 이용한다면 대각 원소가 1이므로

$$\sum_{k=1}^{m} l_{ik}^2 + \Psi_i = 1$$

(3) i번째 변수 x_i(i번째 변수), 원변수 x_j(j번째 변수)의 공분산

$$\mathrm{cov}(x_i, x_j) = \sigma_{ij} = l_{i1}l_{j1} + l_{i2}l_{j2} + \cdots + l_{im}l_{jm}$$

$$= \sum_{k=1}^{m} l_{ik}l_{jk}$$

- 원변수 x_j(j번째 변수), 요인변수 f_k(k번째 요인)의 공분산

$$\mathrm{cov}(x_i, f_k) = l_{ik}$$

5.3 요인 구하기

5.3.1 요인 방정식 풀기

상관계수 행렬(R)을 이용하여 요인을 구하는 방법

(1)상관계수 행렬 R에 대해 R=LL`+Ψ을 만족하는 요인부하 행렬 L과 ψ가 존재할때,
또 다른 직교행렬 P에 대해 다음이 성립하므로
R=LIL`+ψ=(LP)(LP)`+ψ=L*L*`+ψ
요인 부하 행렬은 무수히 존재한다.

(2) L,ψ 의 미지수 개수를 보면 (pm+p)이고 행렬 P로부터 주정할 수 있는 값의 개수는 p(p+1)/2이므로 방정식 수보다 미지수 개수가 많아 해가 무수히 많이 존재한다.

만약 m=p인 경우는 ∑=LL`로 유일하게 분해되고 ψ=0이다.

- (이경우는 주성분 분석 방법이다)

(3) m(<p)인 경우 어떤 L을 이용할 것인가?

얻어진 요인을 해석하는데 용이 하도록 요인변환(factor rotation)을 실시하여 그 값을 이용

(예. SAS에서 ROTATE=VARIMAX 옵션 사용)

5.3.2 해를 구하는 방법

(1) 종류

(1) 주성분 방법(Principal Component method)

원변수 공분산행렬(상관계수 행렬)R 로부터 고유치 고유벡터를 구한다. (주성분과 동일)

즉, i) 만일 m=p $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$, $e_1, e_2, \cdots, e_p$

$$\Sigma = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots \lambda_p e_p e_p'$$

ii) 만일 m < p인경우

$$\Sigma = [\sqrt{\lambda_1} e_1 | \sqrt{\lambda_2} e_2 | \cdots | \sqrt{\lambda_m} e_m] \begin{bmatrix} \sqrt{\lambda_1} e_1' \\ \sqrt{\lambda_2} e_2' \\ \cdots \\ \sqrt{\lambda_m} e_m' \end{bmatrix} + \begin{bmatrix} \Psi_1 & 0 & \cdots & 0 \\ 0 & \Psi_1 & \cdots & 0 \\ --- & & & \\ 0 & \cdots & 0 & \Psi_p \end{bmatrix} = LL' + \Psi$$

그러므로,

i번째 공통요인에 의해 원변수의 설명 변동 크기는 λ_i이다.

i번째 공통요인의 원변수에 대한 설명 능력 다음과 같다.

즉,

상관 계수 행렬:	$\frac{\lambda_i}{p}$
공분산 행렬:	$\frac{\lambda_i}{s_{11} + s_{22} + \cdots + s_{pp}}$

상관계수행렬 (R)로부터 구한 주성분 $y_1, y_2 \cdots, y_p$

$f_1 = \frac{y_1}{\sqrt{\lambda_1}}, \ f_2 = \frac{y_2}{\sqrt{\lambda_2}}, \ \cdots, \ f_p = \frac{y_p}{\sqrt{\lambda_p}}$ 을 정의하여,

- 요인의 L(f_{ij})과 주성분의 L(l_{ij}) 관계 $f_i = y_i \sqrt{\lambda_i}$
  - 주성분 방법을 사용하여 요인 방정식의 해를 구하는 경우 요인분석과 주성분 분석은 동일하다.
  - 다른 점: 주성분 점수(Y=LX)는 원변수의 선형 결합이나 요인 점수는 추정(X=LF)해야 한다 (p126).

(2) 최대우도 추정방법 (Maximum Likelihood Method)

- 원변수가 다변량 정규분포 가정 하에서 f_j와 η_j가 서로 독립이고, x_j 가 다변량 정규분포를 따른다는 가정
  →최대우도추정법으로 L, ψ을 구한다.
  - L의 초기치로 다중 상관계수 제곱을 취하고 큰 공통성을 가진 변수에는 큰 가중치를 주게 되므로 공통성의 추정치가 1 이상이 되는 Heywood가 발생한다. 이 상황에서는 Ψ의 추정치가 음이 된다. (사용불가)

5.4 맛보기(Applicant.txt)

-주성분분석 : 원 변수의 변동을 잘 설명하는 주성분을 찾는 것으로 단위 차이가 없다면 공분산행렬(S)를 이용

-요인분석 : 변수의 내재된 관계를 이용하여 변수를 분류하는 방법으로 상관행렬 이용

-주성분에서 상관계수행렬로 고유치, 고유벡터를 구하면, 요인분석의 고유치, 고유벡터와 동일하다.

예제(3장)

- 예제 데이터의 15개 변수에 대해서 주성분분석(상관계수 행렬이용)과 요인분석 결과를 비교해보자.

지원자 48명 중 우수 지원자 6명을 선발하고자 15개 항목에 대해 평가 - 자료를 SAS로 불러옴

1. ID(지원자 번호)
2. Letter(이력서 X1)
3. Appearance(외모 X2)
4. Academic Abality(학교성적 X3)
5. Likeability(친밀감 X4)
6. Self-Confidence(자신감 x5)
7. Lucidity(명석 x6)
8. Honest (진실 x7)
9. Salesmanship(마케팅능력 x8)

....

16. Suitability(업무 적합성 X15)

(1) SAS 프로그램 (p127)

```
data applicant;
infile 'C:₩다변량₩chapter3 data₩applicant.txt' firstobs=2;
input id I ap aa li sc lc ho sm ex dr am gc po kj su;
run;
proc princomp data=applicant;
var I--su;
run;
proc factor data=applicant;
var I--su;
run;
```

-만약 원 변수들이 다변량 정규분포를 따른다는 가정이 성립한다면 최대우도추정방법을 사용 할 수 있다.

```
PROC FACTOR DATA=APPLICANT METHOD=ML;
```

(2)주성분분석 결과 (Proc princomp ;)

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	7.51379418	5.45749301	0.5009	0.5009
2	2.05630117	0.60048169	0.1371	0.6380
3	1.45581948	0.25792178	0.0971	0.7351
4	1.19789771	0.45874509	0.0799	0.8149
5	0.73915262	0.24457355	0.0493	0.8642
6	0.49457907	0.14331724	0.0330	0.8972
7	0.35126183	0.04135981	0.0234	0.9206
8	0.30990202	0.05294047	0.0207	0.9412
9	0.25696154	0.07205117	0.0171	0.9584
10	0.18491037	0.03223000	0.0123	0.9707
11	0.15268036	0.05511728	0.0102	0.9809
12	0.09756308	0.00874428	0.0065	0.9874
13	0.08881880	0.02418557	0.0059	0.9933
14	0.06463323	0.02890868	0.0043	0.9976
15	0.03572455		0.0024	1.0000

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8
l	0.162440	0.428846	0.315375	-.094347	0.114181	0.621238	0.171116	0.155368
ap	0.213108	-.035266	-.022878	0.262175	0.870203	-.037767	-.009998	-.009056
aa	0.040184	0.236919	-.430470	0.636274	-.212812	0.223389	0.310998	-.043216
ii	0.225078	-.129796	0.465825	0.345375	-.128784	0.111960	-.130674	-.308273
sc	0.290481	-.248896	-.241026	-.172804	0.004916	0.019939	0.143008	0.386488
ic	0.314870	-.130990	-.150037	-.071033	-.206810	0.174643	-.514547	0.023612
ho	0.158117	-.405450	0.283928	0.416491	-.063642	-.303949	0.144365	0.343708
sm	0.324256	-.029492	-.185975	-.198227	0.037393	-.117958	0.010157	-.141580
ex	0.134068	0.553139	0.082591	0.057752	-.103091	-.367209	-.112752	0.584307
dr	0.315071	0.046243	-.079635	-.155987	-.200942	-.250153	0.489632	-.255634
am	0.318024	-.068155	-.208651	-.199291	0.163090	0.113408	0.201079	0.041316
gc	0.331497	-.023150	-.117142	0.074726	-.082414	0.147997	-.408244	0.106225
po	0.333289	0.022257	-.072544	0.188140	-.127323	0.058988	-.016186	-.149205
kj	0.259208	-.082272	0.467206	-.201376	-.111522	0.075376	0.246975	0.051937
su	0.236037	0.420662	0.089152	-.019913	0.080921	-.414317	-.172807	-.382172

(3) 요인분석 결과 (Proc Factor ;)

SAS 시스템 2011년 08월 08일 월요일 오후

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 15 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	7.51379418	5.45749301	0.5009	0.5009
2	2.05630117	0.60048169	0.1371	0.6380
3	1.45581948	0.25792178	0.0971	0.7351
4	1.19789771	0.45874509	0.0799	0.8149
5	0.73915262	0.24457355	0.0493	0.8642
6	0.49457907	0.14331724	0.0330	0.8972
7	0.35126183	0.04135981	0.0234	0.9206
8	0.30990202	0.05294047	0.0207	0.9412
9	0.25696154	0.07205117	0.0171	0.9584
10	0.18491037	0.03223000	0.0123	0.9707
11	0.15268036	0.05511728	0.0102	0.9809
12	0.09756308	0.00874428	0.0065	0.9874
13	0.08881880	0.02418557	0.0059	0.9933
14	0.06463323	0.02890868	0.0043	0.9976
15	0.03572455		0.0024	1.0000

4 factors will be retained by the MINEIGEN criterion.

Factor Pattern

	Factor1	Factor2	Factor3	Factor4
l	0.44527	0.61496	0.38052	-0.10326
ap	0.58416	-0.05057	-0.02760	0.28695
aa	0.11015	0.33974	-0.51939	0.69639
ii	0.61697	-0.18612	0.56205	0.37801
sc	0.79625	-0.35691	-0.29082	-0.18913
ic	0.86310	-0.18784	-0.18103	-0.07774
ho	0.43342	-0.58141	0.34258	0.45584
sm	0.88883	-0.04229	-0.22439	-0.21696
ex	0.36750	0.79319	0.09965	0.07415
dr	0.86365	0.06631	-0.09609	-0.17073
am	0.87175	-0.09773	-0.25175	-0.21812
gc	0.90868	-0.03320	-0.14134	0.08179
po	0.91359	0.03192	-0.08753	0.20592
kj	0.71052	-0.11798	0.56372	-0.22040
su	0.64701	0.60322	0.10757	-0.02179

① 주성분 방법에 의해 구한 요인 분석의 고유치는 주성분 분석의 고유치와 동일하다.

→ 두 방법 모두 상관계수 행렬로 고유치 구해서

② i번째 요인의 j 번째 원 변수에 대한 loading(부하)값은 $f_{ij} = \sqrt{\lambda_i} l_{ij}$(주성분 선형계수) 이다.

Ex) 제일 주성분 y_1의 원 변수 AP의 계수에 고유치 (λ_1)의 제곱근을 곱하면 $(\sqrt{7.5138} \times 0.2131)$ 요인1의 AP부하 값 (f12= 0.5842)와 동일하다.

③ 부하의 의미는 공통개념(요인)이 원 변수에 미치는 영향 정도를 나타내는 값이므로 부하가 크다는 의미는 공통개념에 의해 원 변수가 잘 설명되고 있음을 의미한다.

④ 요인 부하 값의 크기는 변수를 그룹화하는데 사용

⑤ 그룹화

*요인1의 경우 부하 값이 상대적으로 큰 것을 묶으면 그 변수들은 같은 그룹에 속한다.
(LC=명석, SM=마케팅, DR=추진력,AM=야망, GC=개념파악능력,PO=잠재력) → 마케팅 능력

마케팅 능력=(LC+SM+DR+AM+GC+PO)/6

*요인2에서 부하 값이 큰 것을 묶는다.
(L=이력서, EX=경험, SU=업무적합성) → 회사경험

회사경험=(L+EX+SU)/3

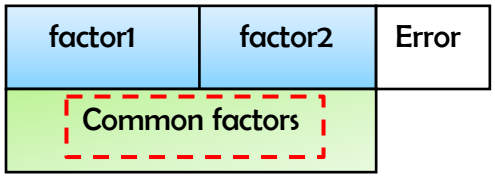
*요인3에서 부하 값이 큰 값을 묶는다.
(LI=친밀감, KJ=화합) → 적응력

*요인4 (AA=학교성적)

5.5 요인 개수 결정

(1) 부하 값의 의미는 각 요인이 원 변수를 설명하는 정도(크기)를 나타내며 요인은 변수들에 내재된 관계에서 공통 부분에 해당한다.

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\ l_{12} \cdots l_{1m} \\ l_{21}\ l_{22} \cdots l_{2m} \\ \cdots \\ l_{p1}\ l_{p2} \cdots l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_p \end{pmatrix}$$



부하 값의 절대값이 큰 것들만(음의 부호는 동일 개념의 반대 척도)선택하여 변수들을 그룹화 하면된다.

(2)요인 개수결정시 고려 사항

① 사소한 요인은 제외하자.

- 원 변수 1-2개에만 부하 값이 큰 요인에 의해 묶을 수 있는 변수는 1-2개이므로 변수그룹의 의미가 없다.

② Kaiser 판단(가장 많이 이용, SAS도 활용)

- 고유치가 1이상인 것만으로 요인의 개수 정한다.

③ SCREE 그림 활용(주관적, 자주 사용 않음)

- 주성분 분산 설명 변동의 크기가 갑자기 줄어들기 바로 전까지의개수로 적절한 인자 개수 사용

④ Large-sample Test(카이제곱 검정)

- MLE에 의해 요인 방정식 해를 구하는 경우 요인 개수를 결정하기 위한 검정 방법으로 적합성 검정을 실시한다.

5.6 요인 회전

(1) 개념

요인 부하 값에 의해 원 변수를 그룹화 한다, but

① 요인의 복합성: 하나의 원 변수에 부하 값이 큰 요인이 2개 이상 존재하거나

② 인자의 크기가 0을 중심으로 ±의 작은 값이 있는 경우 부하 값으로 변수를 그룹화 하는 것은 불가능하다.

(2) 방법

각 요인이 상대적으로 큰 부하 값을 갖도록 요인을 회전(rotate)

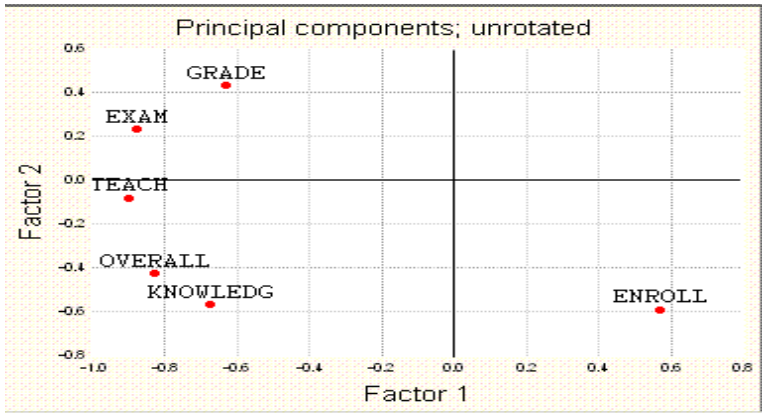
방법: QUARTIMAX rotation, OBLIQUE rotation, PROMAX rotation, VARIMAX(*)

VARIMAX 방법: (Kaiser 제안) 간단한 구조 측정치로 부하행렬의 분산을 최대화 하는 회전 방법

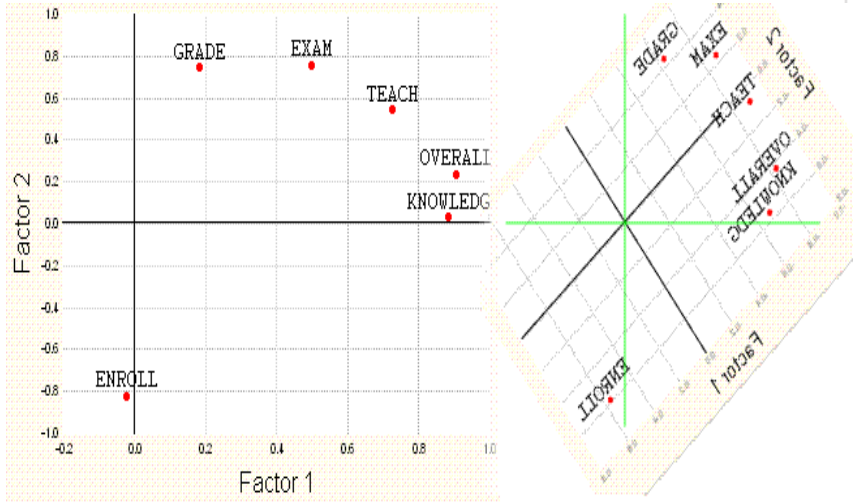
(3) 조건

요인의 개수 m<p인 경우, ∑=LL`+Ψ 을 만족하는 행렬 L은 무수히 많이 존재 하기 때문에 요인회전이 가능하다.

▪예제



▪VARIMAX 방법



▪5.7 예제

5.7.1 데이터 설명 (POLICE.TXT)

경찰에 지원한 50명의 신체적 특성 15개를 측정한 것이다.

ID: 지원자 번호	REACT: 시각적 자극에 대한 반응 시간
HEIGHT (cm)	WEIGHT (kg)
SHLDR: 어깨 넓이(cm)	PELVIC: 골반 넓이(cm)
CHEST: 가슴 넓이(cm)	THIGH: 허벅지 피부 두께 (mm)
PULSE: 맥박	DIAST: 심장 혈압
CHNUP: 턱걸이 회수	BREATH: 폐활량 (liter)

RECVR: 런닝 머신에서(treadmill) 제자리 달리고 5분 후 맥박

SPEED: 런닝 머신에서 제자리 달리기 최대 속력

ENDUR: 런닝 머신에서 달릴 수 있는 최대 시간(분)

FAT: 비만도

15개의 항목을 공통개념(요인)으로 분류 – 요인분석 이용

15개의 항목을 1~2개의 축약된 변수 – 주성분분석 이용

5.7.2 SAS 프로그램

이 프로그램은 요인 방정식 해를 구하는 방법으로 principal factoring 방법을 사용하였고, 요인 회전 방법은 VARIMAX 방법을 사용한 예이다.

```
data police;
infile 'C:₩다변량₩₩chapter5 data₩POLICE.TXT' FIRSTOBS=2;
input ID REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH
PULSE DIAST CHNUP BREATH RECVR SPEED ENDUR FAT;
run;
```

```
proc print data=police;
run;
```

```
proc factor data=police ROTATE=VARIMAX;
VAR REACT -- FAT;
run;
```

Maximum Likelihood 방법을 사용할 경우

```
proc factor data=police ROTATE=VARIMAX METHOD=ML;
```

▪5.7.3요인 개수

Eigenvalues of the Correlation Matrix: Total = 15 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	5.21852549	2.81172746	0.3479	0.3479
2	2.40679803	1.09411978	0.1605	0.5084
3	1.31267825	0.08160302	0.0875	0.5959
4	1.23107523	0.02722958	0.0821	0.6779
5	1.20384565	0.35594143	0.0803	0.7582
6	0.84790423	0.14315618	0.0565	0.8147
7	0.70474804	0.12633945	0.0470	0.8617
8	0.57840860	0.18486182	0.0386	0.9003
9	0.39354677	0.02535508	0.0262	0.9265
10	0.36819169	0.04160632	0.0245	0.9510
11	0.32658537	0.13971542	0.0218	0.9728
12	0.18686996	0.04806356	0.0125	0.9853
13	0.13880640	0.09492503	0.0093	0.9945
14	0.04388137	0.00574644	0.0029	0.9975
15	0.03813492		0.0025	1.0000

5 factors will be retained by the MINEIGEN criterion.

-상관행렬을 사용하면 고유치가 설명하는 분산 변동의 합은 변수 개수 P(15)이고, 각 고유치의 평균 설명력은 1이다.

-Kaiser 제안 방법을 이용한다. 고유치가 1이상인 요인만 출력.

-MINEIGEN(=minimum eigen value)의 의미는 다른 옵션이 설정되지 않았으므로 디폴트로 고유치가 1 이상인 요인들만 출력한다는 의미이다.

-요인 패턴(요인 부하 값)은 특별한 옵션이 없으면 크기가 1이상인 고유치 개수만큼 출력된다. 요인패턴을 원하는 만큼 출력하려면 NFACTORS옵션을 사용하면 된다.

```
proc factor data=police ROTATE=VARIMAX NFACTORS=10;
```

5.7.4 부하값

: 요인이 원 변수를 설명하는 정도의 크기

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\ l_{12}\ \cdots\ l_{1m} \\ l_{21}\ l_{22}\ \cdots\ l_{2m} \\ \cdots \\ l_{p1}\ l_{p2}\ \cdots\ l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_p \end{pmatrix}$$

* 요인회전 전 결과

Factor Pattern

	Factor1	Factor2	Factor3	Factor4	Factor5
REACT	0.11577	0.23649	0.12762	0.06168	0.90082
HEIGHT	0.69783	-0.33725	0.41902	0.05972	0.22820
WEIGHT	0.95187	-0.04446	-0.07678	0.10093	-0.06101
SHLDR	0.68565	-0.32752	0.31506	0.11806	-0.23196
PELVIC	0.67123	-0.29937	0.08798	0.48711	-0.10913
CHEST	0.82416	0.00558	-0.14664	0.15902	-0.18894
THIGH	0.64905	0.47592	-0.26352	-0.22796	0.01317
PULSE	-0.27258	0.59160	0.52991	0.02023	-0.00286
DIAST	-0.08173	0.42916	-0.04908	0.77472	0.05289
CHNUP	-0.66679	-0.36578	0.27478	0.24007	-0.12565
BREATH	0.57632	-0.05036	0.50353	-0.17149	0.17768
RECVR	-0.05822	0.65763	0.45810	-0.11492	-0.41729
SPEED	-0.06704	-0.76236	0.03135	-0.21987	0.04457
ENDUR	-0.46683	-0.08312	-0.14959	0.36143	0.09197
FAT	0.84105	0.34734	-0.26738	-0.07236	0.01546

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4	Factor5
5.2185255	2.4067980	1.3126783	1.2310752	1.2038457

① 요인1 : weight(체중),chest(가슴넓이), fat(비만)

→적절한 그룹 이름 부여 (5.7.6에서)

② 요인 2 : recvr(제자리 달리고 5분 후 맥박)=0.66,

speed(제자리 달리기 최대 속력)=-0.76

두 변수는 반대 개념, 즉 맥박은 속력과 반비례한다고 설명

③ 요인 3: height(키), pulse(맥박), breath(폐활량)

④ 요인4: diast(심장혈압)

⑤ 요인5 : react(반응시간)

5.7.5 공통성

$$Var(x_i) = \sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2 + \Psi_i = \sum_{k=1}^{m} l_{ik}^2 + \psi_i$$

***공통성 결과**

Final Communality Estimates: Total = 11.372923

REACT	HEIGHT	WEIGHT	SHLDR	PELVIC
0.90090440	0.83192324	0.92783038	0.74439622	0.79709996
CHEST	THIGH	PULSE	DIAST	CHNUP
0.76176311	0.76936042	0.70551551	0.79625170	0.72733481
BREATH	RECVR	SPEED	ENDUR	FAT
0.64920781	0.83305264	0.63699092	0.38630835	0.90498318

앞의 요인회전 전 결과에서, 1번째 변수 REACT의 요인1~요인5 부하 값의 제곱이 공통성을 나타낸다.

$$0.11577^2 + 0.23649^2 + 0.12762^2 + 0.06168^2 + 0.90082^2 = 0.9009$$

① 1(100%)에 가까운 값이면 그 변수의 변동이 선택된 요인에 의해 거의 모두 설명된다는 의미이고 낮으면 다른 요인이 존재한다는 것.

② ENDUR 변수는 5요인들에 의해 설명되는 정도가 낮다. 이 변수는 아마 5개의변수 그룹 어디에도 포함되지 않을 가능성이 높다.

*이 변수를 가장 잘 설명할 공통요인은 요인6~요인15 중 하나일 것이다.

5.7.6 변수 그룹

-요인의 부하가 크다는 것 →그 요인에 의해 영향을 많이 받는 다는 것 →부하값이 큰 원 변수 묶음

-요인 회전을 통해 부하 값의 크기를 쉽게 비교

-VARIMAX 방법 사용, REORDER 옵션(크기순)

```
proc factor data=police ROTATE=VARIMAX REORDER;
VAR REACT -- FAT;
   RUN;
```

(1) 요인회전 된 결과

Rotated Factor Pattern

	Factor1	Factor2	Factor3	Factor4	Factor5
FAT	0.89834	0.30408	-0.01710	0.05600	0.04550
THIGH	0.86470	0.07378	0.11051	-0.02796	0.05664
CHEST	0.60652	0.57244	-0.14309	0.11808	-0.17828
ENDUR	-0.38966	-0.26455	-0.16683	0.36931	0.01610
CHNUP	-0.83023	-0.10598	0.00362	0.07366	-0.14626
HEIGHT	0.11446	0.82407	-0.09857	-0.20697	0.29527
SHLDR	0.14604	0.82138	-0.04338	-0.13253	-0.17015
PELVIC	0.16044	0.79465	-0.23908	0.26790	-0.10469
WEIGHT	0.65304	0.68489	-0.17334	0.02459	-0.04056
BREATH	0.19065	0.60670	0.20462	-0.32989	0.30673
RECVR	0.10678	-0.04500	0.88357	0.01179	-0.19694
PULSE	-0.14414	-0.12994	0.78471	0.11661	0.19618
SPEED	-0.38288	0.16941	-0.49297	-0.46286	-0.06663
DIAST	-0.01615	0.01011	0.18516	0.86776	0.09270
REACT	0.11922	-0.00396	-0.00664	0.11263	0.93485

① 직교 변환된 요인들을 이용하여 변수 그룹을 만들수 있는데 적절한 이름을 부여하기 위해서 변수에 대한 지식과 경험이 필요.

② REORDER 옵션은 크기 순으로 정렬될 뿐 아니라 각 요인 별로 크로스 체크 하여 가장 큰 요인에 넣는다.

EX) WEIGHT의 경우 요인1에 0.65인데, 요인2에서 0.68로 더 크기 때문에 요인2에 포함

③ 그룹화

- 요인1 : FAT(비만도), THICHP(허벅지두께), CHEST(가슴둘레), CHNUP(턱걸이)→몸집비대변수

(FAT+THICHP+CHEST+CHNUP)/4 = 비만도지수

- 요인2: HEIGHT(키),SHLDR(어깨 넓이), PELVIC(골반넓이), WEIGHT(체중), BREATH(폐활량) →신체골격 구조 변수

- 요인3: PLUSE(맥박), RECVR(5분후 맥박) →심장지구력

- 요인4 : DIAST(심장혈압)

- 요인5: REACT(반응 시간)

이는 하나씩 분류되기 때문에 변수 분류에 의미가 없다.

-즉 15개의 변수 → 2개의 그룹과 개별변수로 묶음

(2) 2개 그룹(비만지수, 신체골격)에 대한 평균
(CHNUP의 부하값은 음이므로 -를 붙여준다)

* Sas 프로그램

```
DATA POLICE1;
SET POLICE;
CHNUP=-CHNUP;
FAT_INDEX=MEAN(FAT,THIGH,CHEST,CHNUP);
BODY=MEAN(HEIGHT,SHLDR,PELVIC,WEIGHT,BREATH);
RUN;
PROC SORT DATA=POLICE1;
BY DESCENDING FAT_INDEX BODY;
RUN;
PROC PRINT DATA=POLICE1;
VAR ID FAT_INDEX BODY;
RUN;
```

*FAT_INDEX에 의해 정렬된 결과

OBS	ID	INDEX	BODY
1	15	39.6825	109.860
2	37	38.4125	113.792
3	6	37.4050	110.552
4	38	36.5550	110.972
5	19	35.6725	112.472
6	23	35.1050	108.908
7	21	35.0825	123.348
8	18	35.0275	109.332
9	36	34.5925	105.780
10	41	33.3250	105.016
11	4	33.1725	109.704
12	46	32.6350	107.588
13	9	32.4900	110.068
14	24	31.9725	104.704
15	12	31.9125	100.480
16	7	30.4250	123.748
17	3	29.9975	94.472
18	34	29.3525	98.416
19	39	29.3025	107.060
20	47	28.8600	98.320
21	11	28.8325	98.424
22	40	28.4025	94.400
23	17	28.2925	100.876
24	13	28.1900	109.556
25	44	28.1875	102.684
26	48	28.1275	99.444
27	33	28.1175	104.600
28	1	27.8275	96.160
29	27	27.3250	117.368
30	29	27.0625	110.256
31	35	26.9600	105.464
32	43	26.7150	98.852
33	50	26.6500	93.700
34	28	26.3575	87.712
35	25	26.0850	96.860
36	14	26.0825	100.140
37	31	25.8050	103.628
38	22	25.7575	103.324
39	5	25.3600	105.316
40	16	24.9100	96.048
41	32	24.6900	111.296
42	42	24.5425	103.056
43	30	23.1850	101.656
44	20	22.7225	105.592
45	26	22.0150	95.612
46	8	21.2350	102.088
47	45	20.8925	94.660

(3) 회전 전 요인과 회전 후 요인 비교

① 회전된 요인에 의해 원 변수의 변동을 설명하는 부분이 회전되지 않은 경우와 다소 다르다.

② 그러나 선택된 요인들에 의해 설명되는 각 변수의 공통성은 같다.

③ 요인회전에 따라 부하 값이 다르다.

→ 변수 그룹 및 해석이 요이한 방법을 사용

요인 회전 전 부하 값

Factor Pattern

	Factor1	Factor2	Factor3	Factor4	Factor5
REACT	0.11577	0.23649	0.12762	0.06168	0.90082
HEIGHT	0.69783	-0.33725	0.41902	0.05972	0.22820
WEIGHT	0.95187	-0.04446	-0.07678	0.10093	-0.06101
SHLDR	0.68565	-0.32752	0.31506	0.11806	-0.23196
PELVIC	0.67123	-0.29937	0.08798	0.48711	-0.10913
CHEST	0.82416	0.00558	-0.14664	0.15902	-0.18894
THIGH	0.64905	0.47592	-0.26352	-0.22796	0.01317
PULSE	-0.27258	0.59160	0.52991	0.02023	-0.00286
DIAST	-0.08173	0.42916	-0.04908	0.77472	0.05289
CHNUP	-0.66679	-0.36578	0.27478	0.24007	-0.12565
BREATH	0.57632	-0.05036	0.50353	-0.17149	0.17768
RECVR	-0.05822	0.65763	0.45810	-0.11492	-0.41729
SPEED	-0.06704	-0.76236	0.03135	-0.21987	0.04457
ENDUR	-0.46683	-0.08312	-0.14959	0.36143	0.09197
FAT	0.84105	0.34734	-0.26738	-0.07236	0.01546

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4	Factor5
5.2185255	2.4067980	1.3126783	1.2310752	1.2038457

④ 다소 주관적이라는 단점은 있으나 최적의 요인회전방법이 존재하는 것은 아니므로 여러 방법을 적용해 보는 것이 바람직하다.

⑤ 요인 회전방법에 의해 다양한 부하값이 계산되므로 하나의 선형계수를 산출하는 주성분 분석에 비해 요인분석이 선호되는 이유이다.

요인 회전 후 부하 값

Rotated Factor Pattern

	Factor1	Factor2	Factor3	Factor4	Factor5
REACT	0.11922	-0.00396	-0.00664	0.11263	0.93485
HEIGHT	0.11446	0.82407	-0.09857	-0.20697	0.29527
WEIGHT	0.65304	0.68489	-0.17334	0.02459	-0.04056
SHLDR	0.14604	0.82138	-0.04338	-0.13253	-0.17015
PELVIC	0.16044	0.79465	-0.23908	0.26790	-0.10469
CHEST	0.60652	0.57244	-0.14309	0.11808	-0.17828
THIGH	0.86470	0.07378	0.11051	-0.02796	0.05664
PULSE	-0.14414	-0.12994	0.78471	0.11661	0.19618
DIAST	-0.01615	0.01011	0.18516	0.86776	0.09270
CHNUP	-0.83023	-0.10598	0.00362	0.07366	-0.14626
BREATH	0.19065	0.60670	0.20462	-0.32989	0.30673
RECVR	0.10678	-0.04500	0.88357	0.01179	-0.19694
SPEED	-0.38288	0.16941	-0.49297	-0.46286	-0.06663
ENDUR	-0.38966	-0.26455	-0.16683	0.36931	0.01610
FAT	0.89834	0.30408	-0.01710	0.05600	0.04550

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4	Factor5
3.4799455	3.3769250	1.8753118	1.3949543	1.2457860

5.7.7 부하 값 산점도 그리기

(1) 요인분석의 통계량 값을 OUTSTAT 옵션에 의해 F_STAT이름의 SAS data에 저장하고 출력한 것이다.

-SAS 프로그램-

```
DATA police;
INFILE 'C:₩다변량₩chapter5 data₩POLICE.TXT' FIRSTOBS=2;
input REACT -- FAT;
RUN;

proc factor data=police rotate=varimax reorder outstat=F_STAT;
VAR REACT -- FAT;
RUN;

PROC PRINT DATA=F_STAT;
RUN;
```

(2) UNROTATE 회전되지 않은 요인의 부하값
PATTERN 회전된 요인의 부하값

-결과-

OBS	_TYPE_	_NAME_	REACT	HEIGHT	WEIGHT	SHLDR	PELVIC	CHEST	THIGH
1	MEAN		0.3162	177.906	78.3524	40.9520	28.1060	90.6200	16.1300
2	STD		0.0482	6.703	11.4684	1.5861	1.4372	5.9709	6.1019
3	N		50.0000	50.000	50.0000	50.0000	50.0000	50.0000	50.0000
4	CORR	REACT	1.0000	0.222	0.0562	-0.0938	-0.0559	-0.0318	0.1324
5	CORR	HEIGHT	0.2223	1.000	0.6353	0.6543	0.5859	0.4259	0.2232
6	CORR	WEIGHT	0.0562	0.635	1.0000	0.6656	0.6470	0.8887	0.5542
7	CORR	SHLDR	-0.0938	0.654	0.6656	1.0000	0.5824	0.5545	0.2046
8	CORR	PELVIC	-0.0559	0.586	0.6470	0.5824	1.0000	0.5221	0.2075
9	CORR	CHEST	-0.0318	0.426	0.8887	0.5545	0.5221	1.0000	0.3978
10	CORR	THIGH	0.1324	0.223	0.5542	0.2046	0.2075	0.3978	1.0000
11	CORR	PULSE	0.1631	-0.182	-0.2638	-0.1682	-0.3229	-0.2463	-0.0062
12	CORR	DIAST	0.1473	-0.187	-0.0517	-0.1449	0.1477	-0.0084	0.0487
13	CORR	CHNUP	-0.1585	-0.276	-0.5758	-0.2739	-0.1582	-0.4536	-0.6695
14	CORR	BREATH	0.1595	0.588	0.4502	0.3677	0.3536	0.3473	0.2065
15	CORR	RECVR	-0.1296	-0.121	-0.1219	-0.0239	-0.2069	-0.0832	0.2031
16	CORR	SPEED	-0.1493	0.216	-0.0526	0.2018	0.0412	-0.1624	-0.2080
17	CORR	ENDUR	-0.0525	-0.189	-0.3680	-0.2448	-0.2294	-0.3275	-0.3357
18	CORR	FAT	0.1648	0.365	0.8095	0.3306	0.4132	0.7246	0.8442
19	COMMUNAL		0.9009	0.832	0.9278	0.7444	0.7971	0.7618	0.7694
20	PRIORS		1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000
21	EIGENVAL		5.2185	2.407	1.3127	1.2311	1.2038	0.8479	0.7047
22	UNROTATE	Factor1	0.1158	0.698	0.9519	0.6857	0.6712	0.8242	0.6491
23	UNROTATE	Factor2	0.2365	-0.337	-0.0445	-0.3275	-0.2994	0.0056	0.4759
24	UNROTATE	Factor3	0.1276	0.419	-0.0768	0.3151	0.0880	-0.1466	-0.2635
25	UNROTATE	Factor4	0.0617	0.060	0.1009	0.1181	0.4871	0.1590	-0.2280
26	UNROTATE	Factor5	0.9008	0.228	-0.0610	-0.2320	-0.1091	-0.1889	0.0132
27	TRANSFOR	Factor1	0.6985	0.494	-0.4543	-0.2490	-0.0056	.	.
28	TRANSFOR	Factor2	0.7029	-0.363	0.5085	0.3269	-0.0939	.	.
29	TRANSFOR	Factor3	-0.1034	0.695	0.6536	-0.0978	-0.2626	.	.
30	TRANSFOR	Factor4	-0.0722	0.343	-0.2349	0.9064	0.0154	.	.
31	TRANSFOR	Factor5	0.0457	0.152	0.2296	-0.0108	0.9602	.	.
32	PATTERN	Factor1	0.1192	0.114	0.6530	0.1460	0.1604	0.6065	0.8647
33	PATTERN	Factor2	-0.0040	0.824	0.6849	0.8214	0.7947	0.5724	0.0738
34	PATTERN	Factor3	-0.0066	-0.099	-0.1733	-0.0434	-0.2391	-0.1431	0.1105
35	PATTERN	Factor4	0.1126	-0.207	0.0246	-0.1325	0.2679	0.1181	-0.0280
36	PATTERN	Factor5	0.9349	0.295	-0.0406	-0.1701	-0.1047	-0.1783	0.0566

(3) 필요한 데이터를 SUBSET하고 전치를 한다.

TYPE="PATTERN"에 의해 SAS데이터 TEMP를 만들었다.

-SAS 프로그램-

```
DATA TEMP;
SET F_STAT;
IF (_TYPE_='PATTERN');
RUN;
PROC TRANSPOSE DATA=TEMP OUT=TEMPO;
RUN;
PROC PRINT DATA=TEMPO;
RUN;
```

-회전된 요인의 부하값만 출력

OBS	_NAME_	Factor1	Factor2	Factor3	Factor4	Factor5
1	REACT	0.11922	-0.00396	-0.00664	0.11263	0.93485
2	HEIGHT	0.11446	0.82407	-0.09857	-0.20697	0.29527
3	WEIGHT	0.65304	0.68489	-0.17334	0.02459	-0.04056
4	SHLDR	0.14604	0.82138	-0.04338	-0.13253	-0.17015
5	PELVIC	0.16044	0.79465	-0.23908	0.26790	-0.10469
6	CHEST	0.60652	0.57244	-0.14309	0.11808	-0.17828
7	THIGH	0.86470	0.07378	0.11051	-0.02796	0.05664
8	PULSE	-0.14414	-0.12994	0.78471	0.11661	0.19618
9	DIAST	-0.01615	0.01011	0.18516	0.86776	0.09270
10	CHNUP	-0.83023	-0.10598	0.00362	0.07366	-0.14626
11	BREATH	0.19065	0.60670	0.20462	-0.32989	0.30673
12	RECVR	0.10678	-0.04500	0.88357	0.01179	-0.19694
13	SPEED	-0.38288	0.16941	-0.49297	-0.46286	-0.06663
14	ENDUR	-0.38966	-0.26455	-0.16683	0.36931	0.01610
15	FAT	0.89834	0.30408	-0.01710	0.05600	0.04550

(4)요인1과 요인2의 산점도

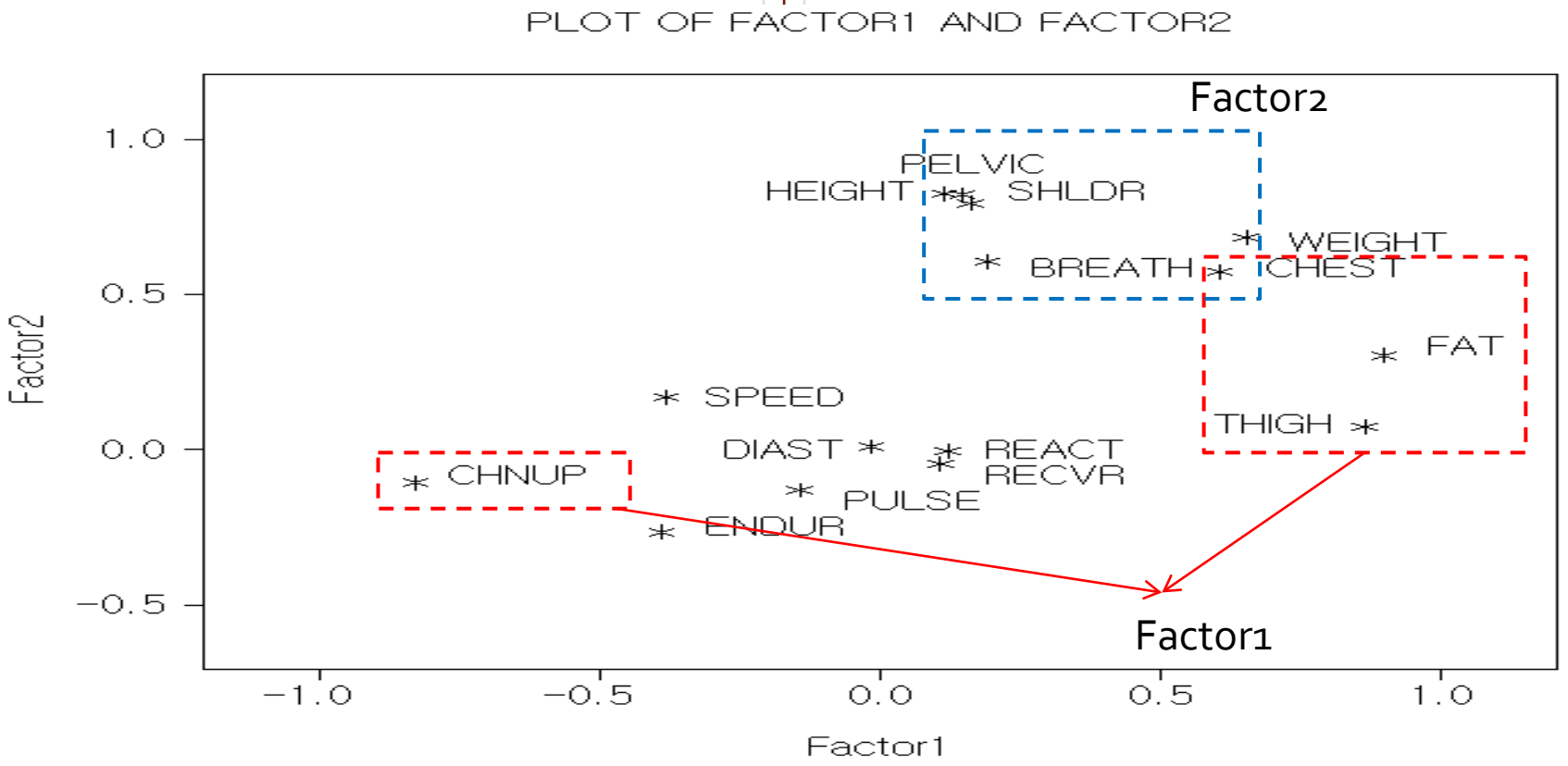
요인3=원 변수 2개

요인4,5= 원 변수 1개

변수가 적기 때문에 요인을 2개로만 선택 적당

-SAS 프로그램(산점도)-

```
TITLE H=1 "PLOT OF FACTOR1 AND FACTOR2";
%PLOTIT(DATA=TEMPO, LABELVAR=_NAME_,
PLOTVARS=FACTOR2 FACTOR1);
RUN;
```



5.8 요인 점수(factor score)

: 각 개체의 요인 값을 요인 점수라 한다.(주성분점수와 유사)

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\ l_{12}\ \cdots\ l_{1m} \\ l_{21}\ l_{22}\ \cdots\ l_{2m} \\ \cdots \\ l_{p1}\ l_{p2}\ \cdots\ l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_p \end{pmatrix} \Leftrightarrow \underline{x} = L\underline{f} + \underline{\eta}$$

- 개체 이상치 진단, 개체 판별 및 군집과 같은 2차 분석
- 주성분분석 : 선형결합이므로 고유벡터에 의해 계산
- 요인분석 : 오차항(η)이 있어서 바로 계산될 수 없다.
  - 2가지 방법을 사용

-Bartlett's Method(Weighted Least Square Method)

$$\underline{z}_r = (\underline{x}_r - \underline{\mu}) \Rightarrow \quad \underline{f}_r = (\hat{L}\hat{\psi}^{-1}\hat{L})^{-1}\hat{L}\hat{\psi}^{-1}\underline{z}_r$$

-Thompson's Method(Regression Method)

$$\begin{bmatrix} \underline{z} \\ \underline{f} \end{bmatrix} \sim N\left(\begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix}, \begin{bmatrix} P & L \\ L' & I \end{bmatrix}\right) \Rightarrow E(\underline{f} \mid \underline{z}) = L'P^{-1}\underline{z} \Rightarrow \quad \underline{f}_r = L'R^{-1}\underline{z}_r$$

▪SAS 이용하기

(1) SCORE옵션을 사용하면 인자 점수를구할 때 사용되는 계수가 출력된다.

```
DATA police;
INFILE 'C:₩다변량₩chapter5 data₩POLICE.TXT' FIRSTOBS=2;
input ID REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH
PULSE DIAST CHNUP BREATH RECVR SPEED ENDUR FAT;
RUN;
proc factor data=police score out=score1 nfactors=2
            ROTATE=VARIMAX;
var REACT -- FAT;
RUN;
PROC PRINT DATA=SCORE1;
RUN;
```

(2) 요인 개수를 반드시 지정해 주어야 한다.

요인의 개수를 2로 지정하면 요인 5개를 사용했을 때의 부하 값과 동일하지 않다.

그러나 각 요인에 의해 설명되는 원 변수 변동은 고유치의 크기와 같으므로 동일하다.

- 요인이 5개 일때

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4	Factor5
5.2185255	2.4067980	1.3126783	1.2310752	1.2038457

- 요인이 2개 일때

Variance Explained by Each Factor

Factor1	Factor2
5.2185255	2.4067980

(3) 요인을 2개만 사용하였으므로 그룹변수는 다름

▪요인이 2개 일 때

Rotated Factor Pattern

	Factor1	Factor2
WEIGHT	0.95012	-0.07276
FAT	0.85101	0.32217
CHEST	0.82396	-0.01894
HEIGHT	0.68749	-0.35786
SHLDR	0.67560	-0.34777
THIGH	0.66293	0.45640
PELVIC	0.66203	-0.31921
BREATH	0.57457	-0.06749
ENDUR	-0.46910	-0.06919
CHNUP	-0.67738	-0.34598
RECVR	-0.03863	0.65907
PULSE	-0.25486	0.59945
DIAST	-0.06893	0.43141
REACT	0.12276	0.23294
SPEED	-0.08969	-0.76002

▪요인이 5개 일때

Rotated Factor Pattern

	Factor1	Factor2	Factor3	Factor4	Factor5
FAT	0.89834	0.30408	-0.01710	0.05600	0.04550
THIGH	0.86470	0.07378	0.11051	-0.02796	0.05664
CHEST	0.60652	0.57244	-0.14309	0.11808	-0.17828
ENDUR	-0.38966	-0.26455	-0.16683	0.36931	0.01610
CHNUP	-0.83023	-0.10598	0.00362	0.07366	-0.14626
HEIGHT	0.11446	0.82407	-0.09857	-0.20697	0.29527
SHLDR	0.14604	0.82138	-0.04338	-0.13253	-0.17015
PELVIC	0.16044	0.79465	-0.23908	0.26790	-0.10469
WEIGHT	0.65304	0.68489	-0.17334	0.02459	-0.04056
BREATH	0.19065	0.60670	0.20462	-0.32989	0.30673
RECVR	0.10678	-0.04500	0.88357	0.01179	-0.19694
PULSE	-0.14414	-0.12994	0.78471	0.11661	0.19618
SPEED	-0.38288	0.16941	0.49297	-0.46286	-0.06663
DIAST	-0.01615	0.01011	0.18516	0.86776	0.09270
REACT	0.11922	-0.00396	-0.00664	0.11263	0.93485

(4) 표준화 점수 계수

Standardized Scoring Coefficients

	Factor1	Factor2
WEIGHT	0.18177	-0.02389
FAT	0.16539	0.13946
CHEST	0.15793	-0.00238
HEIGHT	0.12949	-0.14404
SHLDR	0.12728	-0.13993
THIGH	0.13020	0.19395
PELVIC	0.12487	-0.12816
BREATH	0.10977	-0.02420
ENDUR	-0.09044	-0.03186
CHNUP	-0.13224	-0.14811
RECVR	-0.00302	0.27345
PULSE	-0.04490	0.24725
DIAST	-0.01035	0.17870
REACT	0.02510	0.09756
SPEED	-0.02226	-0.31623

(5) SCORE1을 출력한 결과(요인점수)

OBS	ID	REACT	HEIGHT	WEIGHT	SHLDR	PELVIC	CHEST	THIGH	PULSE	DIAST
1	1	0.310	179.6	74.20	41.7	27.3	82.4	19.0	64	64
2	2	0.345	175.6	62.04	37.5	29.1	84.1	5.5	88	78
3	3	0.293	166.2	72.96	39.4	26.8	88.1	22.0	100	88
4	4	0.254	173.8	85.92	41.2	27.6	97.6	19.5	64	62
5	5	0.384	184.8	65.88	39.8	26.1	88.2	14.5	80	68

CHNUP	BREATH	RECVR	SPEED	ENDUR	FAT	Factor1	Factor2
2	158	108	5.5	4.0	11.91	-0.19401	-0.39645
20	166	108	5.5	4.0	3.13	-1.67745	-0.44907
7	167	116	5.5	4.0	16.89	-0.65264	1.53083
4	220	120	5.5	4.0	19.59	0.62835	-0.11369
9	210	120	5.5	5.0	7.74	-0.69152	0.05442

(6) 요인 점수가 필요한가?

*요인 점수를 주성분 점수와 동일한 개념으로 사용

*요인분석의 주목적은 변수 분류임으로 요인 점수를 구할 필요가 없다. 주성분 점수를 사용하여 2차 분석(회귀분석, 다중 공선성해결, 판별분석 등)에 사용하는 것이 보수적 연구방법론이다.

*요인 점수는 아래와 같은 분석에 사용하기를 권함.

① 속성과 단위가 동일한 변수들 분류, 그룹 내 변수들만 평균을 계산하여 새로운 지표로 사용

② 속성이나 단위가 상이한 변수, 요인점수 계산하여 요인점수의 산점도를 이용한 이상치 진단이나 개체 군집에 활용

5.9 Comment

5.9.1 요인 개수 검정

공통개념인 요인은 고유치가 1이상인경우 선택

각 요인들의 부하 값에 의해 원 변수들이 분류된다.

*경찰데이터에서 요인이 5개가 선택이 됨

EDURE 변수의 경우 5개의 요인 어느 것에도 공통 인자의 역할을 하지 못함→ 더 많은 요인이 필요한지 검정

*요인개수 검정 방법으로 x^2-검정방법을 사용

우도함수가 정의될 수 있어야 함으로 다변량정규분포를 따른다는 가정하에 ML (Maximum Likelihood) 방법에 의해서만 가능

SAS이용하기

```
DATA police;
INFILE 'C:₩다변량₩chapter5 data₩POLICE.TXT' FIRSTOBS=2;
input ID REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH
PULSE DIAST CHNUP BREATH RECVR SPEED ENDUR FAT;
RUN;
```

```
PROC FACTOR DATA=POLICE ROTATE=VARIMAX
NFACTOR=5 METHOD=ML HEYWOOD;
VAR REACT -- FAT;
RUN;
```

HEYWOOD옵션→공통성 값이 1이상인 것을 방지

ML방법 이용하여 추정하는 경우 반드시 함께 사용

Ex) 가설 검정

1) 가설설정

H0: 5개 요인이면 충분하다

Significance Tests Based on 50 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors HA: At least one common factor	105	473.1958	<.0001
H0: 5 Factors are sufficient HA: More factors are needed	40	53.7839	0.0714

2) 검정통계량

X^2= 53.7839, p-value=0.0714

3) 기각역 및 결과

• 유의수준 5% 하에서 귀무가설을 기각 하지 못함

즉, 귀무가설을 기각하지 못하므로 5개 요인으로 충분하다

Ex)요인의 수를 임으로 줄이려 할때

* 경찰 데이터의 경우 요인 5개가 적정 수준이지만 2개로 줄일 수 있는지에 대한 검정

```
DATA police;
INFILE 'C:₩다변량₩chapter5 data₩POLICE.TXT' FIRSTOBS=2;
input ID REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH
PULSE DIAST CHNUP BREATH RECVR SPEED ENDUR FAT;
RUN;
```

```
PROC FACTOR DATA=POLICE SCORE OUT=SCORE1
ROTATE=VARIMAX NFACTOR=2 METHOD=ML HEYWOOD;
VAR REACT -- FAT;
RUN;
```

1) 가설설정

H0: 2개 요인이면 충분하다

Significance Tests Based on 50 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors HA: At least one common factor	105	473.1958	<.0001
H0: 2 Factors are sufficient HA: More factors are needed	76	140.4278	<.0001

2) 검정통계량

X^2= 140.4278 , p-value=0.0001

3) 기각역 및 결과

- 유의수준 5% 하에서 귀무가설을 기각 한다.

  즉, 2개 요인으로 충분하지 않다.

*NAFACTOR=3,4 →기각

즉, 최적의 요인개수는 5개로 Kaiser규칙과 동일

(2) 상관계수와 요인 분석 관계

변수를 그룹화 하는 것은 결국 변수들간의 상관 관계가 높은 것을 묶는 다는 의미이다.

하지만 겹치는 부분이 많아 어려움→요인분석

*SAS 프로그램(POLICE 예제)

```
DATA police;
INFILE 'C:₩다변량₩chapter5 data₩POLICE.TXT' FIRSTOBS=2;
input ID REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH
PULSE DIAST CHNUP BREATH RECVR SPEED ENDUR FAT;
RUN;
PROC corr data=police;
var FAT THIGH CHEST CHNUP HEIGHT SHLDR PELVIC WEIGHT BREATH;
RUN;
```

FACTOR1

FACTOR2

	FAT	THIGH	CHEST	CHNUP	HEIGHT	SHLDR	PELVIC	WEIGHT	BREATH
FAT	1.00000	0.84421 <.0001	0.72462 <.0001	-0.69117 <.0001	0.36511 0.0091	0.33063 0.0190	0.41322 0.0029	0.80951 <.0001	0.29870 0.0351
THIGH	0.84421 <.0001	1.00000	0.39780 0.0042	-0.66953 <.0001	0.22319 0.1192	0.20457 0.1541	0.20749 0.1482	0.55422 <.0001	0.20652 0.1502
CHEST	0.72462 <.0001	0.39780 0.0042	1.00000	-0.45359 0.0009	0.42592 0.0020	0.55449 <.0001	0.52205 0.0001	0.88869 <.0001	0.34730 0.0135
CHNUP	-0.69117 <.0001	-0.66953 <.0001	-0.45359 0.0009	1.00000	-0.27601 0.0524	-0.27387 0.0543	-0.15816 0.2727	-0.57578 <.0001	-0.35762 0.0108
HEIGHT	0.36511 0.0091	0.22319 0.1192	0.42592 0.0020	-0.27601 0.0524	1.00000	0.65429 <.0001	0.58589 <.0001	0.63534 <.0001	0.58783 <.0001
SHLDR	0.33063 0.0190	0.20457 0.1541	0.55449 <.0001	-0.27387 0.0543	0.65429 <.0001	1.00000	0.58244 <.0001	0.66562 <.0001	0.36765 0.0086
PELVIC	0.41322 0.0029	0.20749 0.1482	0.52205 0.0001	-0.15816 0.2727	0.58589 <.0001	0.58244 <.0001	1.00000	0.64702 <.0001	0.35357 0.0118
WEIGHT	0.80951 <.0001	0.55422 <.0001	0.88869 <.0001	-0.57578 <.0001	0.63534 <.0001	0.66562 <.0001	0.64702 <.0001	1.00000	0.45020 0.0010
BREATH	0.29870 0.0351	0.20652 0.1502	0.34730 0.0135	-0.35762 0.0108	0.58783 <.0001	0.36765 0.0086	0.35357 0.0118	0.45020 0.0010	1.00000

5.10 설문분석에 활용

- 요인분석이 언제 설문 분석에 이용될 수 있을 까?

리커트 척도로 조사된 문항을 그룹화하는데 사용

즉, 유사한 여러문항들을 합쳐 하나의 지표점수로 사용 할 수 있느냐를 알아볼 때 요인분석이 사용됨.

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\ l_{12}\ \ldots\ l_{1m} \\ l_{21}\ l_{22}\ \ldots\ l_{2m} \\ \cdots \\ l_{p1}\ l_{p2}\ \ldots\ l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_p \end{pmatrix}$$

그룹의 수 : 고유치가 1이상인 요인에 따라 결정

그룹에 어떤 문항이 묶여지는가 : 부하 값에 의해

요인분석에 의해 변수가 그룹이 가능하려면

① 리커트 척도 문항이어야 한다.

(즉, 변수의 속성과 단위가 모두 동일해야 한다)

② 여러 문항들을 몇 개의 그룹으로 묶으려는 목적에서 실시되어야 한다.

5.10.1 설문예제

대학생 학교 만족도 설문조사 중 일부 내용으로 시설물 만족도 관련 문항들이다(Coding.txt)

요인분석을 통해 9개 시설물 문항의 하위 영역을 구성하려고 한다.

(1) 시설물 관련 설문문항들(9문항)

Q4. 경성대학 건물 안의 공간은?
매우 쾌적하다 7 6 5 4 3 2 1 매우 답답하다
Q5. 경성대학 건물 안팎의 휴식 공간은?
매우 청결하다 7 6 5 4 3 2 1 매우 부족하다
Q6. 강의실 공간은 수업을 하는데 있어~
매우 여유있다 7 6 5 4 3 2 1 매우 비좁다
Q7. 강의실 안의 시설 및 비품은 수업을 하기에~
매우 잘 갖춰있다 7 6 5 4 3 2 1 매우 부족하다
Q8. 강의시간에 보조기자재를 이용하는 것은?
매우 편리하다 7 6 5 4 3 2 1 매우 불편하다
Q9. 경상대학 내에 외국어 공부를 하기 위한 시설은
매우 적절하다 7 6 5 4 3 2 1 매우 부족하다
Q10. 경상대학 내에 컴퓨터 실습을 위한 시설은?
매우 적절하다 7 6 5 4 3 2 1 매우 부족하다
Q11. 경상대학 내에 도서관 시설은?
매우 적절하다 7 6 5 4 3 2 1 매우 부족하다
Q12. 경상대학 화장실 시설은?
매우 청결하다 7 6 5 4 3 2 1 매우 불결하다

(2) SAS 프로그램

```
DATA SURVEY;
INFILE 'C:₩다변량₩chapter5 data₩CODING.TXT';
INPUT (Q1-Q35)(1.);
RUN;
PROC PRINT DATA=SURVEY;
RUN;
```

- Q1~Q35 전체 문항

U									W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
B	Q	Q	Q	Q	Q	Q	Q	Q	Q	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
S	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5
1	1	1	1	1	1	5	2	3	3	2	1	7	1	3	3	1	6	1	1	6	5	5	5	6	4	1	3	2	4	5	1	2	1	2	5
2	1	1	2	1	1	1	1	1	3	1	.	3	3	3	2	1	3	7	1	5	3	5	5	3	2	3	1	4	2	3	5	5	3	2	1	3	.
3	1	1	1	2	4	1	1	2	1	1	1	2	3	1	1	2	3	1	1	1	1	1	1	1	5	2	3	1	5	4	1	1	2	.	.
4	1	1	1	3	3	1	3	3	2	5	3	5	3	3	3	2	3	2	1	1	3	4	4	3	3	3	1	2	3	5	4	1	2	3	5	6
5	1	1	1	3	4	5	5	3	3	5	3	4	4	5	4	2	3	3	5	5	5	5	4	5	4	1	3	2	4	5	1	1	2	4	5
6	1	1	1	3	7	4	4	2	2	1	1	1	4	1	1	1	1	1	1	7	1	3	5	4	4	2	5	1	3	4	2	2	1	.	.
7	1	1	5	4	2	3	3	2	4	3	3	1	3	4	2	1	3	4	4	7	7	3	4	4	4	2	5	1	4	3	2	2	2	5	.
8	1	1	1	3	3	4	3	4	3	4	3	4	3	4	3	2	3	2	3	4	4	4	4	4	4	3	2	3	1	4	5	1	2	3	4	.
9	1	2	1	3	4	3	3	2	4	3	2	3	4	4	4	4	5	7	4	6	3	5	4	4	5	5	2	3	1	5	4	1	1	1	2	3
10	2	1	2	4	4	4	4	4	4	4	4	4	4	3	3	3	4	3	2	3	4	6	4	4	5	1	.	.	.	.	1	1	4	5	6
11	1	1	1	4	4	2	4	4	3	3	2	2	4	5	3	2	5	4	5	3	4	4	4	3	3	4	1	5	2	3	4	1	1	1	2	6
12	2	1	4	3	2	3	3	3	1	6	2	5	4	5	5	3	5	1	2	4	4	6	4	4	4	5	2	3	1	5	4	3	2	1	3	.
13	2	1	1	3	3	1	3	3	3	2	2	1	2	4	3	3	3	6	3	4	5	4	3	2	2	1	3	2	5	4	1	2	1	5	.
14	1	1	1	1	1	2	1	3	1	2	1	1	2	2	1	1	3	1	1	3	4	5	5	5	5	1	3	2	4	5	1	1	2	6	.
15	1	1	1	2	2	3	2	2	2	3	2	1	2	3	3	2	3	2	2	2	2	3	2	2	2	1	3	2	4	5	1	2	3	5	6
16	1	1	4	2	1	1	3	1	1	1	1	1	2	3	1	1	1	1	1	5	4	5	4	7	3	1	3	2	4	5	1	2	2	4	5
17	1	1	1	4	3	3	4	4	4	5	4	5	4	5	4	3	3	4	3	4	4	4	4	3	3	4	1	4	2	3	5	1	2	6	.	.
18	2	1	5	4	4	1	3	2	3	5	3	6	4	5	5	4	1	6	6	5	6	5	4	2	6	1	4	2	3	5	5	3	1	5	6	.
19	2	1	1	2	2	2	3	2	2	2	2	2	2	4	4	3	3	3	3	3	3	4	4	3	3	1	3	2	5	4	2	1	4	.	.
20	1	1	1	4	3	2	2	4	3	5	3	5	4	4	4	2	1	2	2	2	3	3	4	4	4	1	3	2	4	5	1	2	2	3	.

(3) 시설물에 대한 만족도를 측정한 9개 문항이 어떻게 그룹화되어 하위문항이 구성되는지 알아보자

```
DATA SURVEY;
INFILE 'C:₩다변량₩chapter5 data₩CODING.TXT';
INPUT (Q1-Q35)(1.);
RUN;
proc factor data=survey rotate=varimax reorder;
var Q4-Q12;
RUN;
```

-결과

Eigenvalues of the Correlation Matrix: Total = 9 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	3.75509467	2.75103818	0.4172	0.4172
2	1.00405650	0.05050753	0.1116	0.5288
3	0.95354897	0.16659613	0.1059	0.6347
4	0.78695284	0.13707952	0.0874	0.7222
5	0.64987332	0.14284924	0.0722	0.7944
6	0.50702408	0.02290537	0.0563	0.8507
7	0.48411871	0.04536948	0.0538	0.9045
8	0.43874923	0.01816755	0.0487	0.9533
9	0.42058168		0.0467	1.0000

2 factors will be retained by the MINEIGEN criterion.

1. 실험실 데이터나 측정 데이터인 경우 : 고유치가 1이상인 요인만을 택해도 누적 설명비율이 80%이다.

2. 리커트 척도 문항의 경우 : 누적 설명 비율이 낮다.

→문항 분류에 요인분석을 사용하는 경우 원 변수의 내재된 관계를 설명하는 요인을 찾는 것이 중요하므로 요인의 개수는 중요하지 않다.

요인의 개수가 많아지면 변수(문항) 그룹 개수가 많아져 변수그룹의 효과가 미미해 진다.

Rotated Factor Pattern

	Factor1	Factor2
Q7	0.73815	0.21534
Q8	0.69564	-0.03666
Q6	0.64055	0.20458
Q5	0.62266	0.37486
Q9	0.56059	0.49437
Q10	0.13259	0.80343
Q11	0.21992	0.75404
Q12	0.08964	0.55255
Q4	0.51461	0.53426

- 요인1 : 부하의 값의 크기가 0.6 이상을 그룹
  → Q5~Q8을 한 그룹으로 묶는다.
- 요인2 : Q10, Q11

- **요인 1: 강의실 만족도 그룹**
- **요인 2: 정보 시설 만족도 그룹**
- 나머지 문항들은 개별 문항으로 간주하고 분석(Q4,Q12)

5.10.2 내적 일치도(internal consistency)

: 리커트 척도 문항들이 요인분석에 그룹화 되면 그룹 내 문항들이 하나의 개념을 얼마나 잘 표현하는지를 나타내는 지표를 내적 일치도라 한다.

- 내적일치도의 값을 나타낸 것을 Cronbach α 라 한다.
- 이 값을 문항의 신뢰도 계수값으로 나타낼 수 있다.

• Cronbach α

①0과 1사이의 값이고, 1에 가까울수록 내적 일치도가 높다.

② 기준 값이 적절하지 않다. (0.6이상 ?, 0.7이상 ?)

- 문항의 수가 많을수록, 응답자 수가 많을 수록 높아지는 경향이 있기 때문이다.

③ 값의 크기가 판단의 근거가 아니라, 한 문항을 제외했을 때 Cronbach α값이 작아지느냐, 커지느냐를 보고 그 문항을 그대로 두느냐 제외하느냐 를 판단한다.

(1) SAS 프로그램

```
DATA SURVEY;
INFILE 'C:₩다변량₩chapter5 data₩CODING.TXT';
INPUT (Q1-Q35)(1.);
RUN;
proc corr data=survey nosimple nocorr alpha;
var Q5-Q8;
RUN;
proc corr data=survey nosimple nocorr alpha;
var  Q10 Q11;
RUN;
```

(2) Q5-Q8의 신뢰도 계수

크론바흐의 α

변수	α계수
원데이터	0.686085
표준화	0.690973

삭제된 변수를 포함한 크론바흐의 α

삭제된 변수	데이터 변수 합계에 대한 상관계수	데이터 변수 α계수	표준화된 변수 합계에 대한 상관계수	표준화된 변수 α계수
Q5	0.513210	0.596535	0.515436	0.599651
Q6	0.446627	0.634924	0.452792	0.639815
Q7	0.555856	0.574055	0.552425	0.575144
Q8	0.386110	0.670726	0.381783	0.683350

① 4개 문항 모두 사용시 신뢰도 계수 : 0.69

② Q5를 제거시 신뢰도 계수 : 0.60

③ Q6을 제거시 신뢰도 계수 : 0.64

④ Q7을 제거시 신뢰도 계수 : 0.58

⑤ Q8을 제거시 신뢰도 계수 : 0.68

→ 문항 제외시 내적일치도가 떨어지므로 4개 변수를 하나의 그룹으로 묶는 것이 옳다.

(3) Q10, Q11의 신뢰도 계수

크론바흐의 α

변수	α계수
원데이터	0.670557
표준화	0.684502

삭제된 변수를 포함한 크론바흐의 α

삭제된 변수	데이터 변수 합계에 대한 상관계수	데이터 변수 α계수	표준화된 변수 합계에 대한 상관계수	표준화된 변수 α계수
Q10	0.512626	.	0.520337	.
Q11	0.512626	.	0.520337	.

① 2개 문항의 신뢰도 계수 : 0.68

② 변수가 2개인 경우는 제외 신뢰도 계수가 계산될 수 없다.(하나를 제외하면 변수가 하나 밖에 남지 않아)

(4) 신뢰도 계수에 대한 일반적 가이드 라인은 응답자 수 200명, 그룹내 문항이 5개 이면 0.7이상 이되어야 한다.