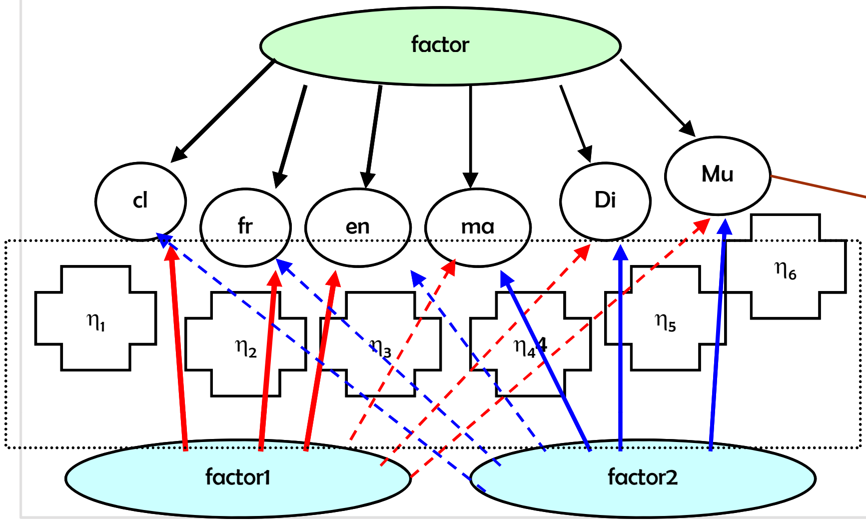
FA Concept (페이지 120)

▪History 페이지 120

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

	Classic	French	English	Math	Dis.	Music
Classic	1	.83	.78	.7	.66	.63
French		1	.67	.67	.65	.57
English			1	.64	.54	.51
Math				1	.45	.51
Discovery					1	.4
Music						1



▪Concept 개념

- 개체의 특성을 측정한 변수들 간에는 구조적 연관 관계(상관 관계)가 존재한다. ▷ 공분산 구조
- 내재된 공분산 구조를 설명하는 unobserved, latent 변인 존재
- 요인을 이용하여 구조적 연관 관계 크기(상관계수)에 따라 변수들을 그룹화

- 모형
  - 원 변수를 요인에 의해 표현 $X = LF$

- 요인(인자, factor)을 이용하여
  - 원변수의 내재된 구조적 관계를 이해한다.
  - 요인의 개수를 결정한다. 그룹의 크기
  - 변수 그룹의 이름 부여, 부하 행렬 이용

$S\ (R)$

factor1	factor2	Error
Common factors		

▪활용

- 설문지 리커트 척도 문항 하위개념 분류
- 주성분 점수와 유사하게, 실제로는 동일 속성, 동일 단위 변수에 대해서만 요인점수 사용

개념 2

▪상관분석과 관계 (페이지 145)

- Football.xls
  - 변수: RO_YDS, RD_YDS, PO_YDS, PD_YDS, TO_YDS, TD_YDS
  - 변수의 구조적 관계 상관관계에 따라 요인의 loading 값 크기가 결정된다.
  - 변수의 그룹화는 상관관계 정도에 의존

```
상관: RO_YDS, RD_YDS, PO_YDS, PD_YDS,

        RO_YDS  RD_YDS  PO_YDS  PD_YDS  TO_YDS
RD_YDS  -0.738
         0.037

PO_YDS  -0.469  -0.078
         0.241   0.854

PD_YDS  -0.194   0.668  -0
         0.645   0.070   0

TO_YDS   0.840  -0.876   0
         0.009   0.004   0

TD_YDS  -0.426   0.884  -0
         0.292   0.004   0
```

```
인자 분석: RO_YDS, RD_YI

상관 행렬에 대한 주성분 인자 분

비회전 인자 적재 및 공통성

변수      요인1    요인2    요인3
RO_YDS   -0.727   -0.674   -0.104
RD_YDS    0.978    0.046    0.051
PO_YDS   -0.096    0.874   -0.475
PD_RAT    0.760   -0.492   -0.380
TO_YDS   -0.890   -0.219   -0.380
TD_YDS    0.912   -0.300   -0.242

분산      3.6961   1.6013   0.5860
% Var     0.616    0.267    0.098
```

▪주성분분석과 비교 (페이지 122)

주성분 분석	요인 분석
주성분은 원 변수의 직교 선형 결합으로 표현	인자들의 직교 선형 결합으로 원 변수들을 표현
Y=LX, L은 선형계수 행렬	X=LF, L 부하행렬, F 관측불가
주성분은 변수들의 변동을 설명	요인은 변수들의 분산-공분산 구조 설명
요인분석이나 주성분 분석의 L을 구하는 방법 유사하다. 공분산 행렬, 상관 행렬로부터 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

주성분 분석과 요인분석 비교2 (페이지 128)

주성분 분석: 이력서, 외모, 대

공분산 행렬에 대한 고유 분석

고유값	66.536	18.181	10.591	6.7
비율	0.543	0.148	0.086	0.0
누적	0.543	0.691	0.778	0.8

고유값	1.136	0.873	0.707	0.509
비율	0.009	0.007	0.006	0.004
누적	0.980	0.988	0.993	0.998

l_{ij}

변수	PC1	PC2	PC3
이력서	-0.149	0.371	0.200
외모	-0.132	-0.029	0.042
대학_학점	-0.030	0.102	-0.131
친밀감	-0.203	-0.093	0.620
자신감	-0.231	-0.236	-0.189
명석	-0.337	-0.196	-0.125
진실성	-0.120	-0.301	0.447
경험	-0.379	-0.090	-0.282
추진력	-0.164	0.636	0.025
비전	-0.316	0.012	-0.113
장래성	-0.312	-0.122	-0.245
화합	-0.339	-0.074	-0.050
업무적합성	-0.357	-0.025	0.041
마케팅	-0.226	-0.045	0.385
업무_파악	-0.274	0.471	0.017

주성분분석
공분산 행렬 사용

고유값 동일
즉,요인/주성분 원변수 설명 정도 동일

요인 loading : 원변수 그룹에 사용
X=L(loading)F
주성분 계수: 주성분 이름 부여에 사용
Y=L(linear coefficient)X

$$f_{ij} = \sqrt{\lambda_i} l_{ij}$$

요인분석
요인추출: 주성분분석 방법 이용
공분산 행렬 사용

인자 분석: 이력서, 외모, 대학_

공분산 행렬에 대한 주성분 인자 분석

비회전 인자 적재 및 공통성 f_{ij}

변수	요인1	요인2	요인3
이력서	-1.216	1.584	0.652
외모	-1.079	-0.125	0.136
대학_학점	-0.242	0.434	-0.426
친밀감	-1.657	-0.397	2.017
자신감	-1.888	-1.005	-0.616
명석	-2.748	-0.836	-0.406
진실성	-0.981	-1.282	1.455
경험	-3.092	-0.384	-0.916
추진력	-1.338	2.713	0.081
비전	-2.578	0.053	-0.369
장래성	-2.546	-0.521	-0.796
화합	-2.763	-0.317	-0.164
업무적합성	-2.913	-0.106	0.134
마케팅	-1.844	-0.191	1.254
업무_파악	-2.239	2.008	0.055
분산	66.536	18.181	10.591
% Var	0.543	0.148	0.086

모형 (페이지 123)

- 모형

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

  - x_i는 i번째 원변수 ▷ x: 원변수 벡터
  - f_i는 i번째 공통요인 (common factor) ▷ f: 요인 벡터
  - l_{ij}는 i변수에 대한 j요인부하 (factor loading) ▷ L: 부하 행렬 (factor loading matrix)
  - η_i는 i번째 원변수에 대한 오차항(공통인자가 설명하지 못하는 원인), 특정요인(specific factor)라 함

- 가정
  - f_i는 서로 독립이고 평균 0, 분산 1인 동일분포를 따른다.

$$\underline{f} \sim iid(\underline{0}, I)$$

  - η_i는 서로 독립이고 평균 0, 분산 w_i인 동일분포를 따른다.

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

  - f_i와 η_i는 서로 독립이다. $Cov(\underline{\eta}, \underline{f}) = 0$

- 요인
  - 공통요인은 원변수 변동을 설명한다.

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

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

  - i번째 원변수의 변동 (분산)
    - 공통성(communality)+특정분산(specific variance)

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

    - 공통성은 요인들이 설명하는 원변수 변동
  - 원변수 xi, 원변수 xj의 공분산 $\text{cov}(x_i, x_j) = \sum_{k=1}^{m} l_{ik} l_{jk}$
  - 원변수 xi, 요인변수 fj의 공분산 $\text{cov}(x_i, f_k) = l_{ik}$
  - 요인 개수 m<p: 차원 축소 (페이지 129)
    - 변수 그룹화, 일부 요인점수 사용 (고유치 1 이상, 80% 규칙)
  - 요인은 서로 독립이다.
  - 만약 m=p인 경우에는 특정분산=0이다.

요인 부하 L 구하기

- 부하 L 얻는 방법 (페이지 124)
  - principal factoring w/ or w/o iteration
  - Rao's canonical factoring, alpha factoring, image factoring
  - Maximum likelihood, un-weighted least square factor analysis, Harris factoring

- 주성분 방법
  - 원변수 공분산행렬(상관계수 행렬)로부터 고유치 고유벡터를 구한다. (주성분과 동일)

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

  - i번째 공통요인에 의해 원변수의 설명 변동 크기는 λ_i이다.
  - i번째 공통요인의 원변수에 대한 설명 능력 다음과 같다.

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

- 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.

- 최대우도 추정방법 $\max_{L,\Psi} L(\underline{\mu}, \Sigma \mid \underline{x}) = L(\underline{\mu}, LL' + \Psi \mid \underline{x})$
  - 원변수가 다변량 정규분포 가정(강한 가정: 비율, 발생 회수, 소득, 자동차 가격 등 예외) 하에서
  - 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)

- L 부하의 의미 (페이지 134)
  - 각 요인이 원변수에 미치는 영향 정도

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

L 부하 loading 및 요인개수

- 부하(l_{ij}, loading) 의미
  - 각 요인(원변수 내재된 관계의 공통 부분)이 원변수에 미치는 영향력 크기
  - 각 요인에서 부하 절대값이 큰 것들만 선택하여 변수들을 그룹화 한다.
    - 그룹화: 그룹 내 변수들의 평균 값 계산
    - 부하 값이 음인 경우는 반대 개념: 반대 부호는 (-)를 붙여준 후 평균 계산
  - 요인의 개수는 변수 그룹 개수를 의미한다.
  - 공통성 (communality)
    - 요인들이 설명하는 원변수 변동 부분
    - 원변수 분산=공통성+특정분산

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

factor1	factor2 ...	factor M	Error
Common factors			

- 요인 개수 구하기 (페이지 144)
  - (1)trivial한 요인은 제외하자. 원 변수 1-2개에만 부하 값이 큰 요인은 제외하자. 이 요인에 의해 묶을 수 있는 변수는 1-2개이므로 그룹의 의미가 없기 때문이다.
  - (2)Kaiser 판단(가장 많이 이용): 변수들의 상관 관계가 0이면 (즉, 상관계수 행렬 R은 항등 행렬 I) 원 변수의 개수와 주성분의 개수가 같아지고 주성분의 분산은 모두 1이고 각 주성분이 가지는 분산 평균도 1이다. 이를 이용하여 상관계수 행렬로부터 구한 고유치가 평균인 1이상인 되어야 한다는 판단 하에 고유치가 1 이상인 것만으로 요인의 개수 결정
  - Scree Plot 사용: 주성분과 동일, 80% 규칙 사용 가능 (분산 변동 설명비율)
  - 요인 개수 강제 결정: 그룹의 개수를 미리 결정하거나, 원변수를 어느 한 그룹에 강제로 속하게 할 때 사용, 이런 경우 부하 값이 달라진다.
- 분류된 변수 사용
  - 그룹 내 변수들의 평균 ▷ 동일 속성 동일 개념
  - 요인점수를 주성분 점수와 동일하게 사용

요인 회전 rotation (페이지131)

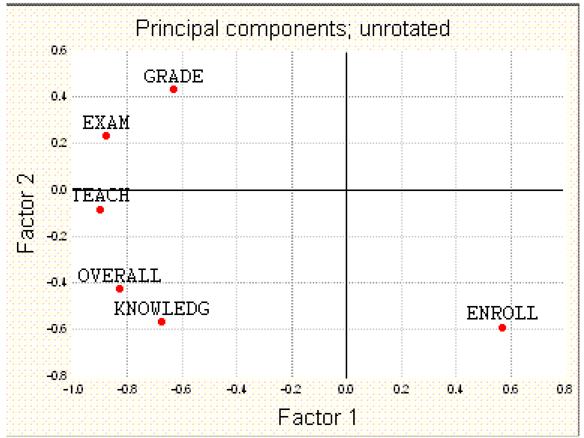
▪개념

- 요인의 개수 m<p인 경우 부하 행렬의 해는 많은 해가 존재.
- 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.

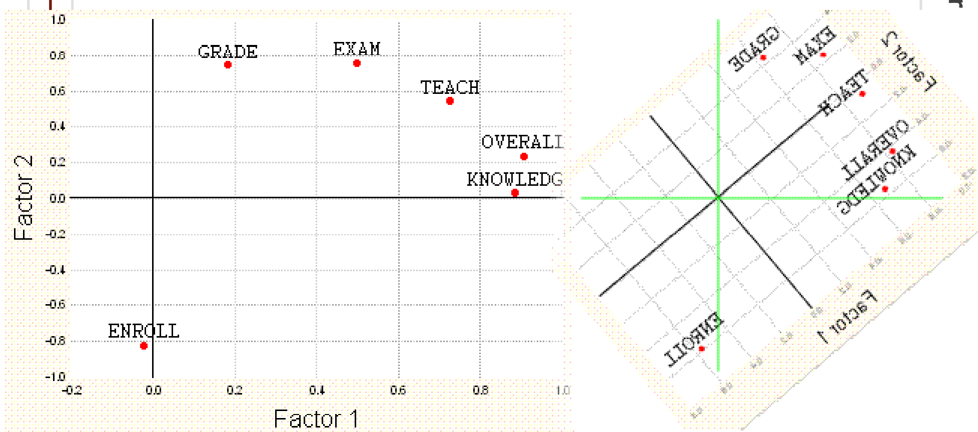
▪방법

- 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

▪예제



- VARIMAX 방법



요인 점수 score (페이지 141)

- 요인점수 (factor score) F 의미
  - 원변수에 내재된 공통성 측정
  - 요인들은 서로 독립이다.
  - 제일 요인이 원변수 구조를 가장 많이 설명, 제이 요인...
  - 주성분 점수와 비교
    - 주성분은 원변수의 선형결합, 원변수에 의해 만들어짐
    - 요인점수는 원변수의 내재된 공통 개념

- 부하 L의 의미
  - 주성분 (Y=LX)
    - 원변수가 주성분에 미치는 영향, 주성분 이름 부여하기 위하여 사용
  - 요인(X=LF)
    - 요인이 원변수에 미치는 영향, 원변수 그룹화에 활용, 역으로 요인점수 이름 부여에 사용하기도 한다.
    - 요인 F를 구하는 관측치로 활용

- 요인점수 구하기 ((X=LF+η)
  - 관측치 X와 주성분 분석 방법에 의해 계산된 L을 이용하여 F 값 계산 (unknown η)
    - Bartlett's Method (Weighted Least Square Method)

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

    - Thompson's Method (Regression Method)

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

    - SPSS는 Anderson Rubin 방법 제공: 추정 요인의 직교성을 보장하는 Bartlett 방법을 수정한 방법

- 요인 점수 활용 (페이지 144)
  - 주성분 점수와 동일하게
    - 가능하면 변수의 속성이 동일한 경우 활용하자. 그래서 리커트 척도에 가장 많이 적용 ▷ 일반적으로 그룹 내 변수들의 평균 점수 사용
    - 주성분 점수와 비해 장점: 요인 회전으로 인하여 다소 다른 개념의 요인 점수들을 얻을 수 있다.

요인분석 With Applicant Data (페이지 127)

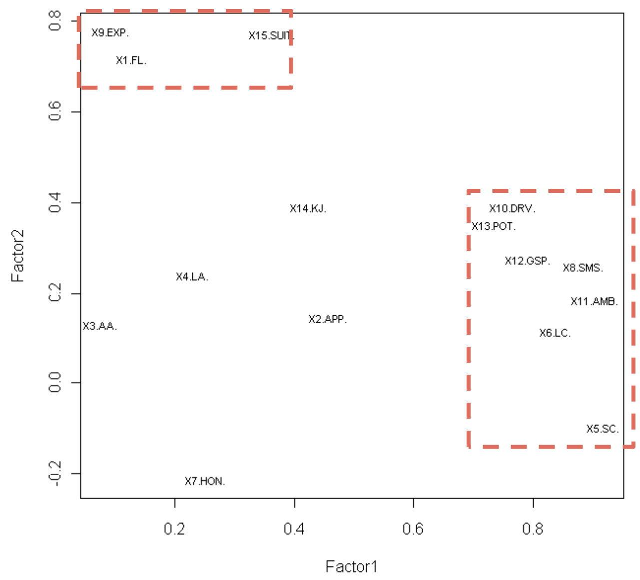
- 요인 부하 및 개수

```
app=read.table("Applicant.txt",header=T)
app_s=app[,2:16]
fa=factanal(app_s,4,rotation="varimax")
print(fa,digits=2,sort=T)
names(fa)
```

```
Loadings:
            Factor1 Factor2 Factor3 Factor4
X5.SC.       0.92            0.14
X6.LC.       0.84    0.11    0.29
X8.SMS.      0.88    0.26
X10.DRV.     0.77    0.39    0.17
X11.AMB.     0.90    0.18
X12.GSP.     0.79    0.28    0.35    0.15
X13.POT.     0.74    0.35    0.43    0.25
X1.FL.       0.13    0.72    0.11   -0.12
X9.EXP.              0.78            0.17
X15.SUIT.    0.36    0.77            0.14
X4.LA.       0.23    0.24    0.84
X7.HON.      0.25   -0.22    0.74
X3.AA.               0.13            0.68
X14.KJ.      0.42    0.39    0.55   -0.60
X2.APP.      0.46    0.14    0.24    0.16

               Factor1 Factor2 Factor3 Factor4
SS loadings       5.57    2.47    2.10    1.01
Proportion Var    0.37    0.16    0.14    0.07
Cumulative Var    0.37    0.54    0.68    0.74

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 84 on 51 degrees of freedom.
The p-value is 0.00247
```

- 상관계수 행렬, 공분산 행렬 설정 불필요? Surely
- promax
  - 부하 값 산점도

```
load=fa$loadings
plot(load,type="n") # plot factor 1 by 2
text(load,labels=names(app_s),cex=.7) # add variable names
```



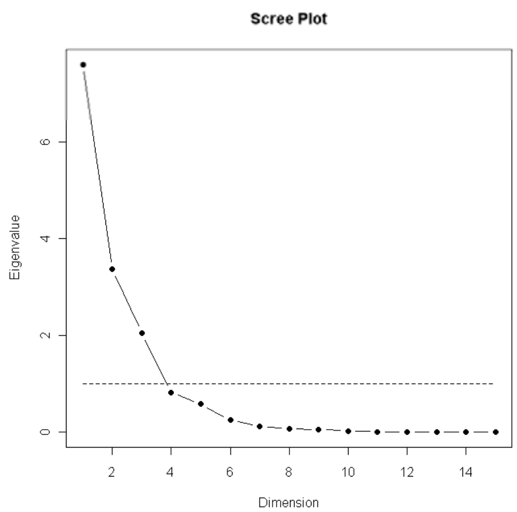
←요인의 개수가 4개이면 충분한가에 대한 가설 검정 결과 귀무가설이 기각되어 4개보다 많아야 된다.

요인분석 With Applicant Data 2 (페이지 139, 141)

- Scree Plot

```
#Scree Plot
library(psy)
scree.plot(fa$correlation)
```

- Kaiser 규칙에 의해 고유치 1이 기준이 된다. 앞의 결과와는 달리 고유치 기준이면 3개 이면 충분



- 슬라이드 52 에서 4대신 6을 넣으면 귀무가설이 채택되므로 6개가 적절하다.

```
Test of the hypothesis that 6 factors are sufficient.
The chi square statistic is 41.53 on 30 degrees of freedom.
The p-value is 0.0784
```

- 최적 요인 개수
  - 하나의 그룹으로 묶이는 변수의 개수와 요인의 수를 고려할 때 4개가 가장 적절

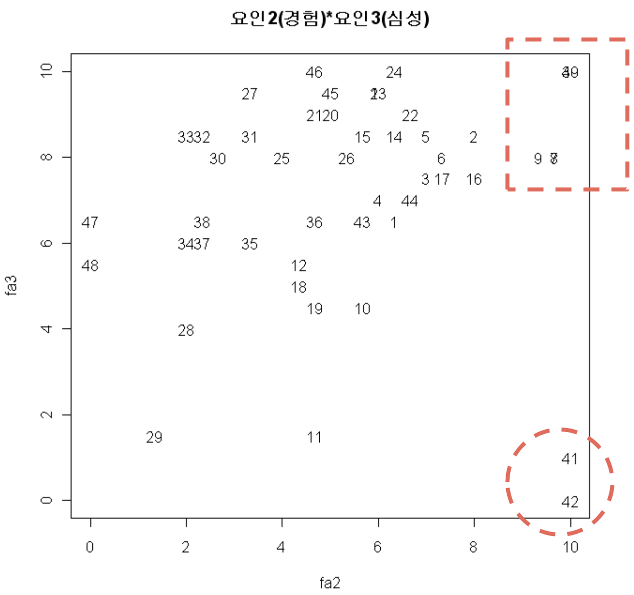
- 요인 분석 결과 활용
- 변수 그룹 사용
  - 요인 내 부하 값이 큰 변수를 동일한 그룹으로 묶음
  - KJ (keenness to join 화합)은 어느 변수 군으로? 요인1 혹은 요인4
  - 변수를 4개의 그룹으로 나눈다.
    - 단 APP(외모)는 어느 변수 군에도 속하지 않는다.
    - 변수 KJ는 요인4의 변수 군으로 나눈다. (AA, KJ)
- 요인 점수 활용
  - 주성분 점수 활용과 동일
  - 리커트 척도와 같이 변수의 속성과 단위가 동일한 경우

요인분석 With Applicant Data 3 (페이지 149)

▪그룹 내 변수 평균

- why 평균? 단위 일치
- 요인2(경험 그룹), 요인3(심성) 그룹 산점도: 이런 주성분 분석의 이름 부여와 동일하네? 당근

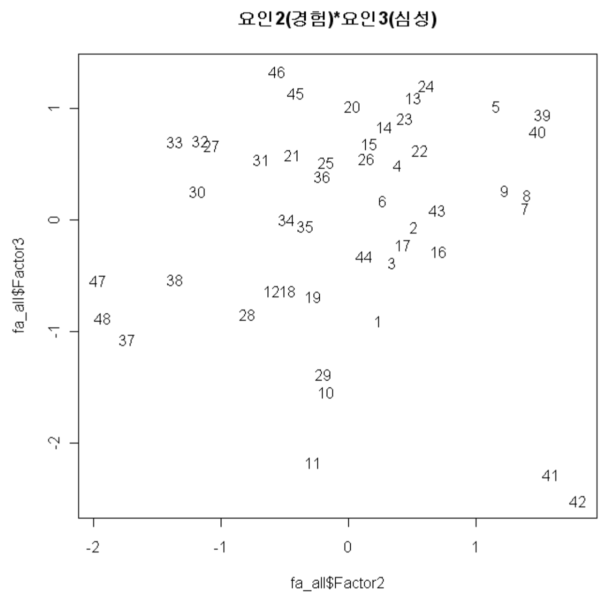
```
fa2=(app$X1.FL.+app$X9.EXP.+app$X15.SUIT.)/3
fa3=(app$X4.LA.+app$X7.HON.)/2

ds2=data.frame(cbind(app[,1:1],fa2,fa3))
attach(ds2)
plot(fa2,fa3,main="요인2(경험)*요인3(심성)",type="n")
text(fa2,fa3,labels=as.character(V1))
```



- 네모 박스 안: (40, 39, 7, 8, 9)
- 원 안: (41, 42)

▪요인 점수 factor score

```
> names(fa)
 [1] "converged"     "loadings"      "uniquenesses" "correla
 [6] "factors"       "dof"           "method"       "scores"
[11] "PVAL"          "n.obs"         "call"
> id=app[,1:1]
> fa_all=data.frame(cbind(id,fa$scores))
> plot(fa_all$Factor2,fa_all$Factor3,
+ main="요인2(경험)*요인3(심성)",type="n")
> text(fa_all$Factor2,fa_all$Factor3,
+ labels=as.character(id))
```



Chapter 5. Factor Analysis

Internal Consistency (페이지 153)

- 내적 일치도 (신뢰도 계수 reliability coefficient)
  - 리커트 척도 문항만 가능
  - 그룹을 형성하는 변수들간의 일치하는 정도
  - 측정 방법: Cronbach α
    - N: 문항 개수, σ_X: 총합 점수의 분산, σ_{Y_i}: 개별 문항 점수 분산

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N} \sigma_{Y_i}^2}{\sigma_X^2}\right)$$

```
> library(multilevel)
> ds4=rbind(app[,2:2],app[,10:10],app[,16:16])
> cronbach(ds4)
$Alpha
[1] 0.922462

$N
[1] 3
```

  - Guide Line
    - 관측치 200, 문항 개수 5개 ➔ 0.7
    - 요인분석 문항 그룹 결과와 일치
    - 요인분석 결과 부하 값이 큰 문항은 변환하여 사용
      - 1번 변수가 부하 값이 1인 경우, 10점 면접 경우 (X1=11-X1)

- 결과 제시
  - 부하값 표(고유치, 누적 기여율 %) 아래 행에 내적 일치도를 제공한다.
  - 다음은 PROMAX 회전 방법에 의해 구해진 부하 값 Varimax 보다는 더 명확하게 구별되어진다.

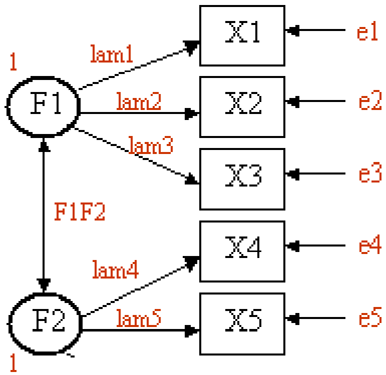
Loadings:

	Factor1	Factor2	Factor3	Factor4
X5.SC.	1.11	-0.31	-0.11	
X6.LC.	0.85		0.11	
X8.SMS.	0.98	0.10	-0.17	
X10.DRV.	0.75	0.26		
X11.AMB.	1.02		-0.17	
X12.GSP.	0.70	0.12	0.20	0.12
X13.POT.	0.56	0.21	0.32	0.21
X1.FL.	-0.11	0.76		-0.18
X9.EXP.	-0.13	0.87	-0.10	0.10
X15.SUIT.	0.16	0.79		
X4.LA.	-0.20	0.12	0.97	
X7.HON.		-0.38	0.86	
X3.AA.		0.16		0.67
X14.KJ.	0.20	0.25	0.50	-0.64
X2.APP.	0.38		0.17	0.15

	Factor1	Factor2	Factor3	Factor4
SS loadings	5.62	2.45	2.20	1.02
~~Proportion Var~~	~~0.37~~	~~0.16~~	~~0.15~~	~~0.07~~
Cumulative Var	0.37	0.54	0.68	0.75
신뢰도 계수 Cronbachα	?	0.92	?	?

Confirmatory Factor Analysis (페이지 159)

- SEM(Structural Equation Model) 구조방정식 모형
  - 인과관계 분석 회귀분석+변수의 공분산구조 설명 기법인 요인분석의 결합
  - 관측변수(네모) vs. 잠재변수(원, latent, factor)



  - 귀무가설: 모집단 공분산 행렬은 연구 공분산 행렬과 동일
  - 대립가설: 모집단 공분산 행렬은 제약이 없다.

- SEM in R

```
library(sem)
ds.cov=cov(app_s)
model.ds=specify.model()
F1 -> X1.FL., lam1, NA
F1 -> X9.EXP., lam2, NA
F1 -> X15.SUIT., lam3, NA
F2 -> X4.LA., lam4, NA
F2 -> X7.HON., lam5, NA
X1.FL. <-> X1.FL., e1,   NA
X9.EXP. <-> X9.EXP., e2,   NA
X15.SUIT. <-> X15.SUIT., e3,   NA
X4.LA. <-> X4.LA., e4,   NA
X7.HON. <-> X7.HON., e5,   NA
F1 <-> F1, NA,    1
F2 <-> F2, NA,    1
F1 <-> F2, F1F2, NA

ds.sem=sem(model.ds, ds.cov, nrow(app_s))
summary(ds.sem)
std.coef(ds.sem)
```

```
Model Chisquare =  6.3852   Df =  4 Pr(>Chisq) = 0.17217
Chisquare (null model) =  96.328   Df =  10
Goodness-of-fit index =  0.94992
Adjusted goodness-of-fit index =  0.81222
RMSEA index =  0.11264   90% CI: (NA, 0.26782)
Bentler-Bonnett NFI =  0.93371
Tucker-Lewis NNFI =  0.93093
Bentler CFI =  0.97237
SRMR =  0.067519
BIC =  -9.0996
```

```
Parameter Estimates
     Estimate  Std Error z value  Pr(>|z|)
lam1  1.89748  0.372697  5.0912 3.5577e-07 X1.FL. <--- F1
lam2  2.59395  0.437045  5.9352 2.9349e-09 X9.EXP. <--- F1
lam3  2.83242  0.429297  6.5978 4.1728e-11 X15.SUIT. <--- F1
lam4 26.56995  0.523975 50.7084 0.0000e+00 X4.LA. <--- F2
lam5  0.18040  0.045824  3.9368 8.2590e-05 X7.HON. <--- F2
e1    3.54845  0.947298  3.7459 1.7977e-04 X1.FL. <--> X1.FL.
```

요인분석 With Police Data (페이지 132)

데이터 description

- 경찰 지원자 신체적 특성 15개 항목
- 변수속성: 체격 및 체력 조건
  - 변수속성 군이 다르고 단위도 상이하다.
  - 그러므로 묶여진(그룹 내) 변수들의 평균은 의미 없음, 요인점수를 이용 (주성분 점수와 유사)
  - 리커트 척도와 상이하므로 내적일치도는 의미 없음

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	

```
ds=read.table("POLICE.txt",header=T)
ds0=ds[,2:16]

fa=factanal(ds0,4,rotation="varimax",scores="regression")
print(fa,digits=2,sort=T)
```

부하 값

Loadings:

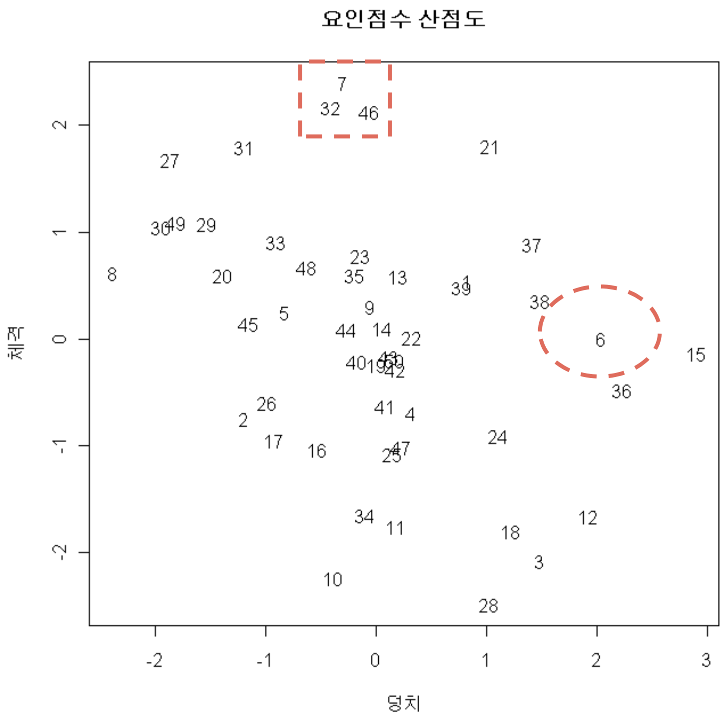
	Factor1	Factor2	Factor3	Factor4
W	0.67	0.60	-0.21	0.33
TW	0.89		0.17	-0.16
CH	-0.70	-0.18		
FAT	0.96	0.21		0.14
H	0.21	0.83	-0.17	-0.11
SW	0.16	0.79		0.18
PW	0.27	0.60	-0.26	0.12
BR	0.19	0.57		
PL		-0.14	0.52	
RECVR			0.99	
CW	0.57	0.42	-0.18	0.67
Re	0.20		-0.12	-0.25
DI		-0.17	0.15	
SP	-0.23	0.20	-0.42	-0.28
EN	-0.36	-0.20		

- 요인점수1: 덩치 (CH: 턱걸이 반대 개념) +
- 요인점수2: 체격 (폐활양?) +
- 요인점수3: 심장지구력 (맥박) -

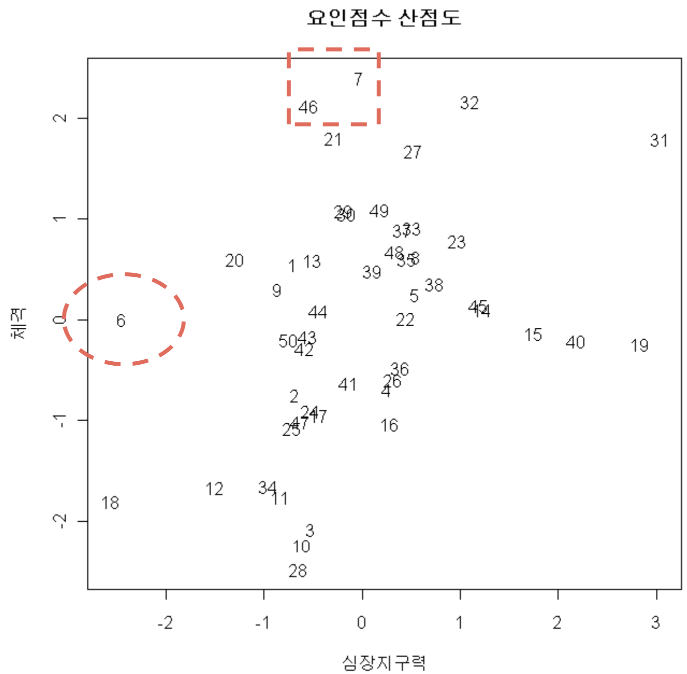
요인분석 With Police Data 2

- 요인점수 산점도
  - 덩치와 체격 요인점수에 의한 산점도

```
id=ds[,1:1]
fa_all=data.frame(cbind(id,fa$scores))
plot(fa_all$Factor1,fa_all$Factor2,
main="요인점수 산점도",type="n",
xlab="덩치",ylab="체격")
text(fa_all$Factor1,fa_all$Factor2,
labels=as.character(id))
```



- 체격과 심장지구력 산점도



- (7, 46) 체격(매우 좋음), 심장지구력(중간), 덩치(중간)
- (6) 체격 중간,심장지구력 매우좋음,덩치 중간

- 주성분 분석과 동일