

8장 인자분석

덕성여자대학교 정보통계학과

김 재 희



《예제 8.3》 신체 측정 데이터에 대해 상관행렬을 이용하여 인자 분석을 하여 인자적재값, 공통성과 특수성을 구하여 [표 8.1]에 정리하였다.

공통인자인 인자1은 길이부분에 더 많이 내재하는 인자로 ‘길이 인자’로 볼 수 있으며 인자2는 무게와 둘레에 기여하는 부분이 큰 ‘무게 인자’로 볼 수 있다.

공통인자에 의해 설명되는 분산인 공통성은 각 변수에 대해 2개 인자의 인자적재값제곱의 합으로 계산되고, 특수인자에 의해 설명되는 분산은 $\psi_i = Var(X_i) - h_i^2 = 1 - h_i^2$ 으로 계산된다. 여기서 상관행렬을 이용했으므로 $Var(X_i) = 1$ 이다.

[표 8.1] 신체 측정 자료에 대한 인자분석 결과

| 변수 | 주성분법 인자1 인자2 | | 공통성 h_i^2 | 특수성 ψ_i |
|---------------|----------------------------------|-------|----------------|-----------------|
| X_1 (신장) | 1.00 | 0.00 | 1.00 | 0.00 |
| X_2 (앞은키) | 0.90 | -0.12 | 0.82 | 0.18 |
| X_3 (가슴둘레) | 0.31 | 0.68 | 0.56 | 0.44 |
| X_4 (흉골길이) | 0.55 | 0.12 | 0.32 | 0.68 |
| X_5 (머리둘레) | 0.11 | 0.18 | 0.04 | 0.96 |
| X_6 (엉덩이둘레) | 0.34 | 0.62 | 0.50 | 0.50 |
| X_7 (몸무게) | 0.50 | 0.86 | 0.99 | 0.01 |

《예제 8.4》 100명 고등학생들의 성적에 대한 데이터의 상관행렬을 이용하여 인자분석을 하여 인자적재값, 공통성과 특수성을 구하여 [표 8.2]에 정리하였다.

[표 8.2] 과목성적에 대한 인자분석 결과

| 변수 | 주성분법 | | 공통성 | 특수성 |
|------------|------|------|---------|----------|
| | 인자1 | 인자2 | h_i^2 | ψ_i |
| X_1 (국어) | 0.90 | 0.10 | 0.82 | 0.18 |
| X_2 (영어) | 0.73 | 0.12 | 0.55 | 0.45 |
| X_3 (사회) | 0.55 | 0.23 | 0.36 | 0.54 |
| X_4 (물리) | 0.23 | 0.87 | 0.32 | 0.68 |
| X_5 (수학) | 0.34 | 0.93 | 0.81 | 0.19 |
| X_6 (생물) | 0.36 | 0.50 | 0.38 | 0.62 |
| X_7 (국사) | 0.60 | 0.30 | 0.45 | 0.55 |

인자1은 국어, 영어, 사회, 국사 과목에 비중이 큰 인자로 ‘인문학 학습능력 인자’
인자2는 수학, 물리, 생물의 과목에서의 비중이 높은 것으로 보아 ‘자연과학 학습능력 인자’

《예제 8.5》 엔지니어 견습생 20명과 비행조종사(pilot) 20명의 여섯 가지의 테스트 결과 비교 자료가 [표 8.3]에 정리 되어있다. 인자분석을 통해 변수에 내재하는 공통인자들을 찾고, 여러 가지 인자적재 추정방법을 사용해 비교하고자 한다.

X_1 = 지능(intelligence)

X_2 = 상황설명능력(form relations)

X_3 = 동력측정검력계(dynamometer)

X_4 = 상세 표시 능력(dotting)

X_5 = 지각 기구 좌표화 능력(sensory motor coordination)

X_6 = 인내력(perseveration)

[표 8.3] 엔지니어와 비행조종사 검사 자료

| 엔지니어 수습생 | | | | | | 파일럿 | | | | | |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
| 121 | 22 | 74 | 223 | 54 | 254 | 132 | 17 | 77 | 232 | 50 | 249 |
| 108 | 30 | 80 | 175 | 40 | 300 | 123 | 32 | 79 | 192 | 64 | 315 |
| 122 | 49 | 87 | 266 | 41 | 223 | 129 | 31 | 96 | 250 | 55 | 319 |
| 77 | 37 | 66 | 178 | 80 | 209 | 131 | 23 | 67 | 291 | 48 | 310 |
| 140 | 35 | 71 | 175 | 38 | 261 | 110 | 24 | 96 | 239 | 42 | 268 |
| 108 | 37 | 57 | 241 | 59 | 245 | 47 | 22 | 87 | 231 | 40 | 217 |
| 124 | 39 | 52 | 194 | 72 | 242 | 125 | 32 | 87 | 227 | 30 | 324 |
| 130 | 34 | 89 | 200 | 85 | 242 | 129 | 29 | 102 | 234 | 58 | 300 |
| 149 | 55 | 91 | 198 | 50 | 277 | 130 | 26 | 104 | 256 | 58 | 270 |
| 129 | 38 | 72 | 162 | 47 | 268 | 147 | 47 | 82 | 240 | 30 | 322 |
| 154 | 37 | 87 | 170 | 60 | 244 | 159 | 37 | 80 | 227 | 58 | 317 |
| 145 | 33 | 88 | 208 | 51 | 228 | 135 | 41 | 83 | 216 | 39 | 306 |
| 112 | 40 | 60 | 232 | 29 | 279 | 100 | 35 | 83 | 183 | 57 | 242 |
| 120 | 39 | 73 | 159 | 39 | 233 | 149 | 37 | 94 | 227 | 30 | 240 |
| 118 | 21 | 83 | 152 | 88 | 233 | 149 | 38 | 78 | 258 | 42 | 271 |
| 141 | 42 | 80 | 195 | 36 | 241 | 153 | 27 | 89 | 283 | 66 | 291 |
| 135 | 49 | 73 | 152 | 42 | 249 | 136 | 31 | 83 | 257 | 31 | 311 |
| 151 | 37 | 76 | 223 | 74 | 268 | 97 | 36 | 100 | 252 | 30 | 225 |
| 97 | 46 | 83 | 164 | 31 | 243 | 141 | 37 | 105 | 250 | 27 | 243 |
| 109 | 42 | 82 | 188 | 57 | 267 | 164 | 32 | 76 | 187 | 30 | 264 |

[표 8.4] 인자분석 결과 인자적재값 비교

| 변수 | 주성분법 | | 주성분법 (Varimax) | | 주성분법 (Promax) | |
|-----------|---------|---------|----------------|---------|---------------|---------|
| | 인자1 | 인자2 | 인자1 | 인자2 | 인자1 | 인자2 |
| X_1 | 0.5369 | 0.4614 | -0.0634 | 0.8344 | 0.1311 | 0.6922 |
| X_2 | -0.1294 | 0.8696 | -0.3578 | 0.1005 | -0.6438 | 0.6126 |
| X_3 | 0.5135 | -0.2538 | 0.7237 | -0.0262 | 0.5598 | 0.1099 |
| X_4 | 0.7239 | -0.3659 | 0.7391 | 0.2947 | 0.7942 | 0.1486 |
| X_5 | -0.4155 | -0.4142 | -0.4833 | -0.0126 | -0.0663 | -0.5815 |
| X_6 | 0.7145 | 0.1236 | 0.2389 | 0.8001 | 0.4813 | 0.5321 |
| 분산 설명량 | 1.7751 | 1.3544 | 1.4930 | 1.4344 | 1.6113 | 1.5094 |

| 변수 | 주 축 인 자 법 (SMC) | | 최대우도법 | | 최대우도법 (Varimax) | |
|-----------|-----------------|---------|---------|---------|-----------------|---------|
| | 인자1 | 인자2 | 인자1 | 인자2 | 인자1 | 인자2 |
| X_1 | 0.4029 | 0.3120 | 0.1886 | 0.4790 | 0.3540 | 0.3737 |
| X_2 | -0.1062 | 0.5686 | 1.0000 | 0.0000 | -0.4239 | 0.9059 |
| X_3 | 0.3426 | -0.1385 | -0.1209 | 0.2697 | 0.2956 | 0.0046 |
| X_4 | 0.5586 | -0.2467 | -0.2664 | 0.5089 | 0.5738 | -0.0258 |
| X_5 | -0.2864 | -0.2464 | -0.2445 | -0.2914 | -0.1604 | -0.3449 |
| X_6 | 0.5557 | 0.0889 | -0.0673 | 0.6513 | 0.6185 | 0.2148 |
| 분산 설명량 | 0.9939 | 0.5694 | 1.4930 | 1.4344 | 1.1297 | 1.1261 |

8.8 R을 이용한 인자분석

[프로그램 8.1] 비행조종사 자료에 대한 인자분석

```
pilot=read.csv("C:/data/pilot_f[REDACTED]header=T)
pilot ; attach(pilot)
x=pilot[,2:7]
m=mean(x) ; m
S=cov(x) ; S
R=cor(x) ; R
fact1=factanal(x, factors=2, rotation="none")          # no rotation
fact2=factanal(x, factors=2, scores = "regression")    # varimax is the default
fact3=factanal(x, factors=2, rotation="promax")       # promax rotation
fact1 ; fact2 ; fact3
```

```

# scree plot
library(graphics)
prin=princomp(x)
screeplot(prin,npcs=6, type="lines",main="scree plot") # 그림 8.7

# plot of factor pattern
namevar=names(fact2$loadings)=c(
plot(fact2$loadings[,1],fact2$loadings[,2],
      xlab="factor1",ylab="factor2", main="factor pattern") # 그림 8.8
      text( x=fact2$loadings[,1], y=fact2$loadings[,2], labels=namevar, adj=0)
      abline(v=0,h=0)

# plot of factor scores
plot(fact2$scores[,1], fact2$scores[,2], pch="*",
      xlab="factor1",ylab="factor2", main="factor scores") # 그림 8.9

```


● R에서는 factanal() 함수를 이용하여 상관행렬을 이용한 인자분석을 할 수 있다. default는 상관행렬로 최대우도법을 이용하고 varimax 회전을 하여 주성분분석을 한다. 주성분법을 이용한 인자분석은 princomp()를 활용한다.

인자분석시 여러 방법으로 인자분석 수행 후 출력결과 중 특히 인자패턴을 비교하며 인자의 의미를 파악하도록 해야 한다.

[결과 8.1]에서 각 방법에 따른 인자적재값을 비교할 수 있다.

$$H_0 : \text{인자 개수 2 개다} \text{ vs } H_0 : \text{not } H_0$$

varimax 회전 인자분석의 카이제곱검정 결과

$$p\text{-값} = 0.333 > 0.05 = \alpha \text{ 이므로}$$

인자 2 개인 모형이 적합하다.

[결과 8.1] 인자분석 결과

```
> fact1=factanal(x, factors=2, rotation="none", scores = "Bartlett") #no rotation
> fact1
factanal(x = x, factors = 2, scores = "Bartlett", rotation = "none")
Uniquenesses:
      x1      x2      x3      x4      x5      x6
0.734 0.005 0.913 0.671 0.856 0.570
Loadings:
      Factor1 Factor2
x1  0.189    0.480
x2  0.997
x3 -0.121    0.269
x4 -0.267    0.508
x5 -0.245   -0.291
x6          0.652
```

| | Factor1 | Factor2 |
|----------------|---------|---------|
| SS loadings | 1.181 | 1.071 |
| Proportion Var | 0.197 | 0.178 |
| Cumulative Var | 0.197 | 0.375 |

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 4.59 on 4 degrees of freedom.
The p-value is 0.333

```
> fact2=factanal(x, factors=2, scores = "regression") # varimax is the default
> fact2
factanal(x = x, factors = 2, scores = "regression")
Uniquenesses:
      x1      x2      x3      x4      x5      x6
0.734 0.005 0.913 0.671 0.856 0.570
```

Loadings:

| | Factor1 | Factor2 |
|----|---------|---------|
| x1 | 0.354 | 0.375 |
| x2 | -0.425 | 0.903 |
| x3 | 0.295 | |
| x4 | 0.573 | |
| x5 | -0.159 | -0.345 |
| x6 | 0.619 | 0.217 |

| | Factor1 | Factor2 |
|----------------|---------|---------|
| SS loadings | 1.130 | 1.122 |
| Proportion Var | 0.188 | 0.187 |
| Cumulative Var | 0.188 | 0.375 |

Test of the hypothesis that 2 factors are sufficient.

The chi square statistic is 4.59 on 4 degrees of freedom.

The p-value is 0.333

```
> fact3=factanal(x, factors=2, rotation="promax")  
# promax rotation
```

```
> fact3
```

Call:

```
factanal(x = x, factors = 2, rotation = "promax")
```

Uniquenesses:

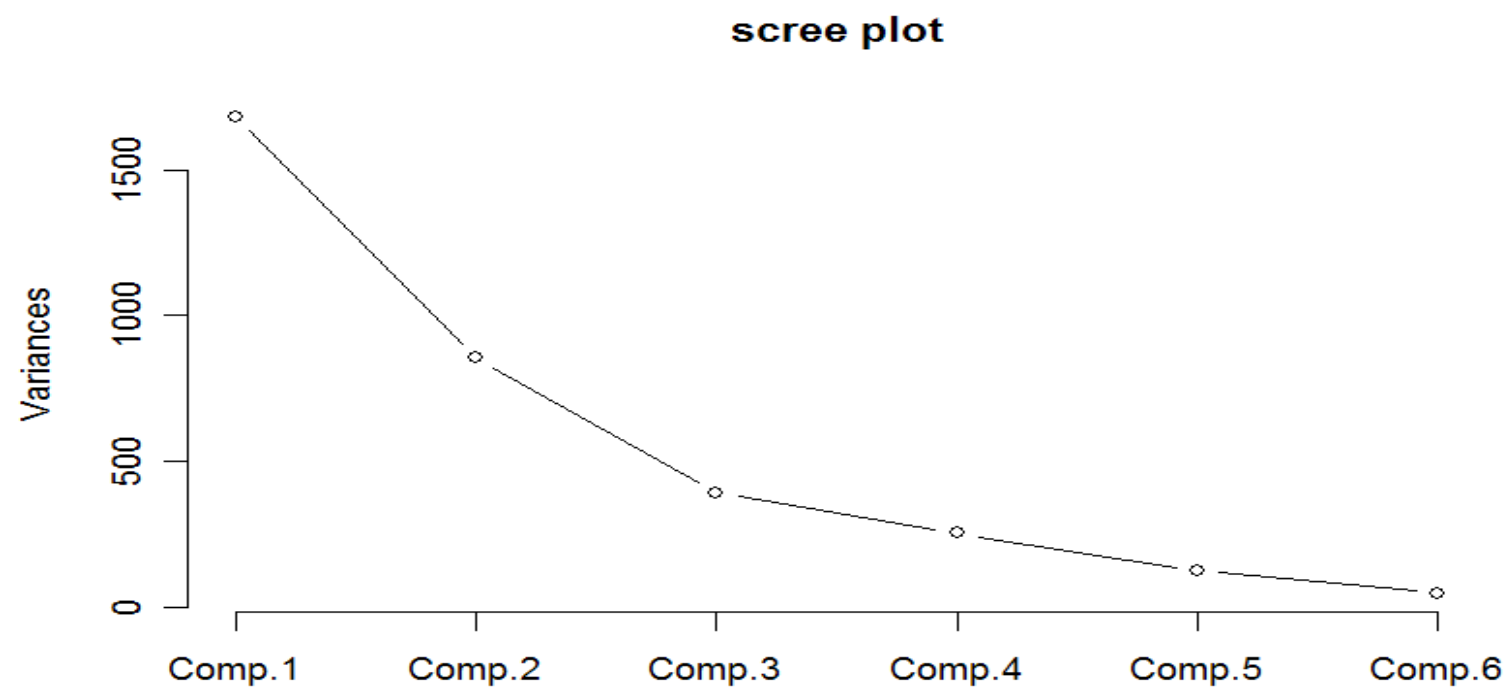
| | x1 | x2 | x3 | x4 | x5 | x6 |
|--|-------|-------|-------|-------|-------|-------|
| | 0.734 | 0.005 | 0.913 | 0.671 | 0.856 | 0.570 |

Loadings:

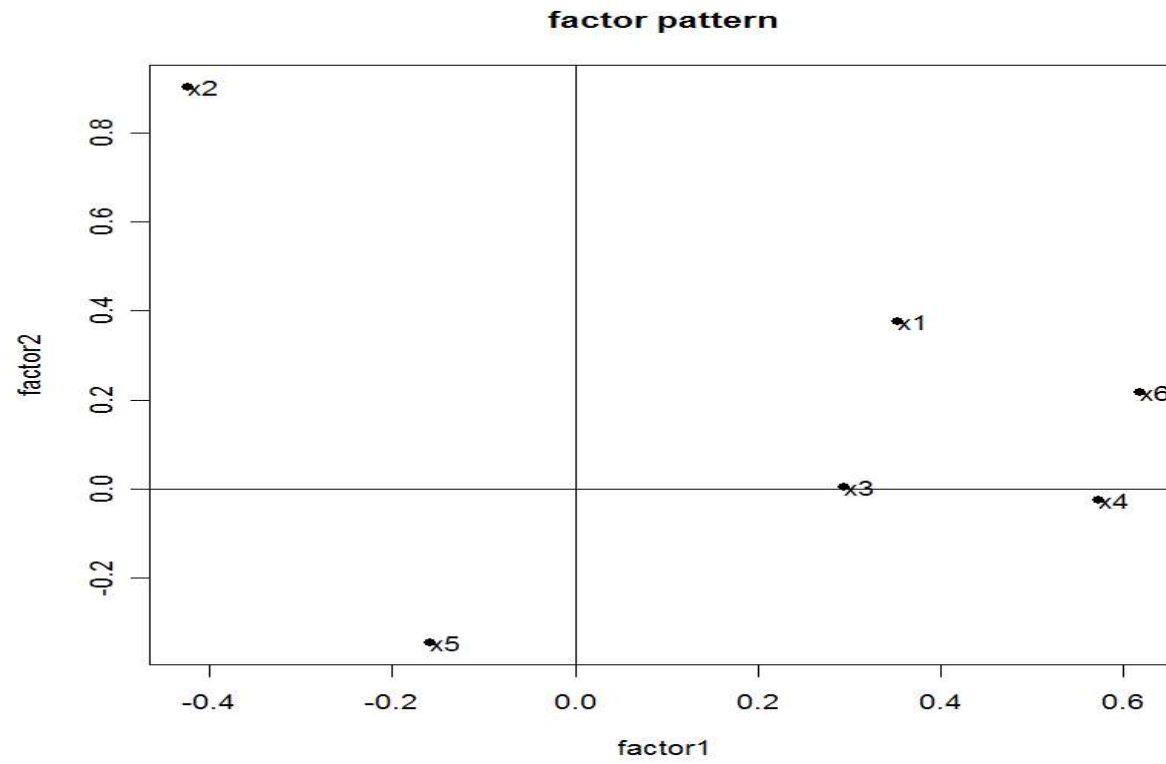
| | Factor1 | Factor2 |
|----|---------|---------|
| x1 | 0.231 | 0.404 |
| x2 | 1.032 | -0.345 |
| x3 | -0.105 | 0.305 |
| x4 | -0.239 | 0.589 |
| x5 | -0.275 | -0.200 |
| x6 | | 0.661 |

The p-value is 0.333

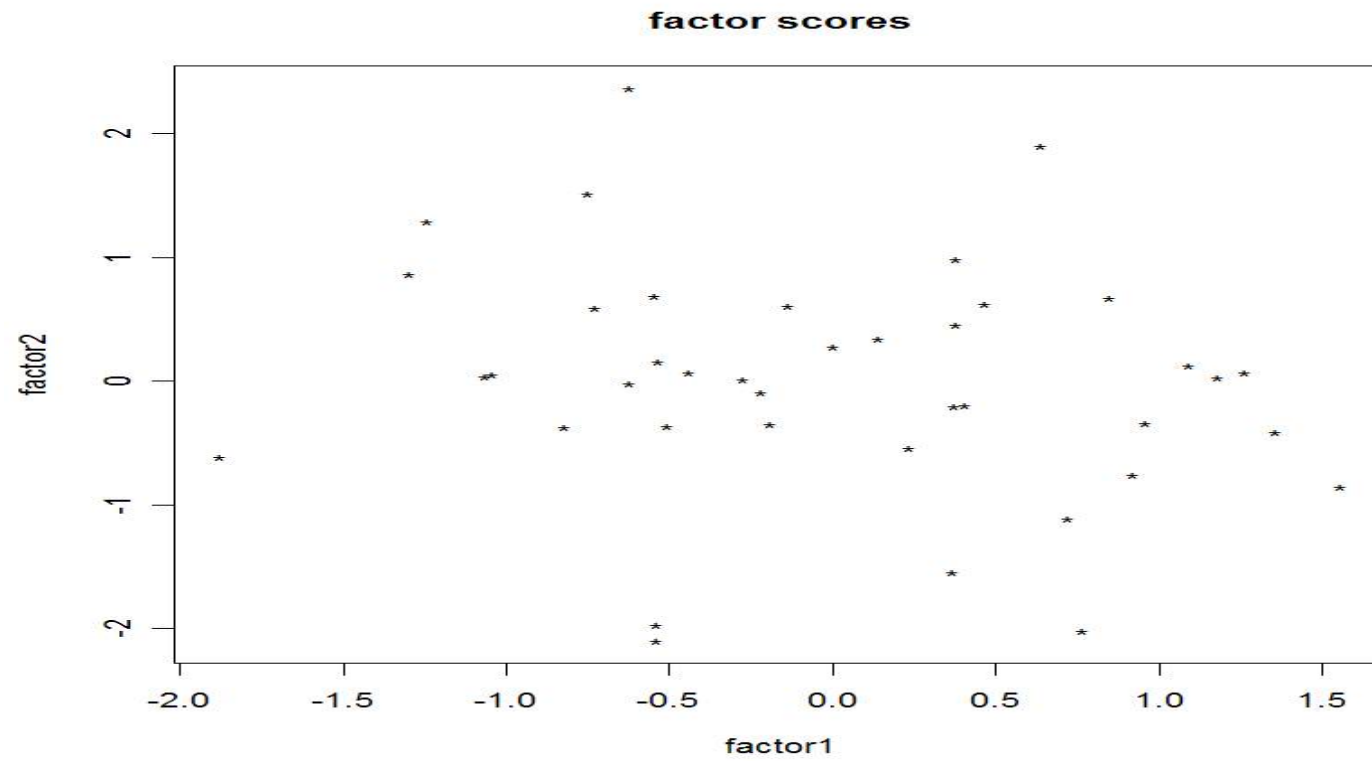
| | Factor1 | Factor2 |
|---|---------|---------|
| SS loadings | 1.262 | 1.199 |
| Proportion Var | 0.210 | 0.200 |
| Cumulative Var | 0.210 | 0.410 |
| Test of the hypothesis that 2 factors are sufficient. | | |
| The chi square statistic is 4.59 on 4 degrees of freedom. | | |



[그림 8.7] 인자 개수에 대한 스크리 그래프



[그림 8.8] Varimax 회전된 인자 패턴



[그림 8.9] Varimax 회전된 인자 점수 그래프