다변량 통계 분석 과제 #2

2014150137 통계학과

박정진

6.1

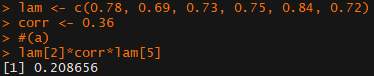
값을 통하여 reliability를 평가한다.



주어진 Parameter Estimates를 통하여 얻은 값은 0.6028이다.

6.2

(a)



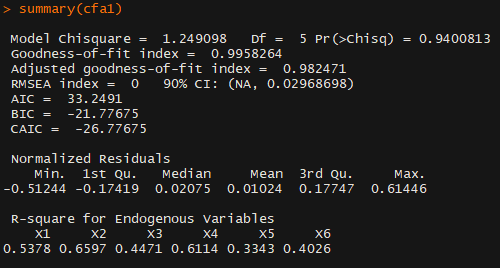
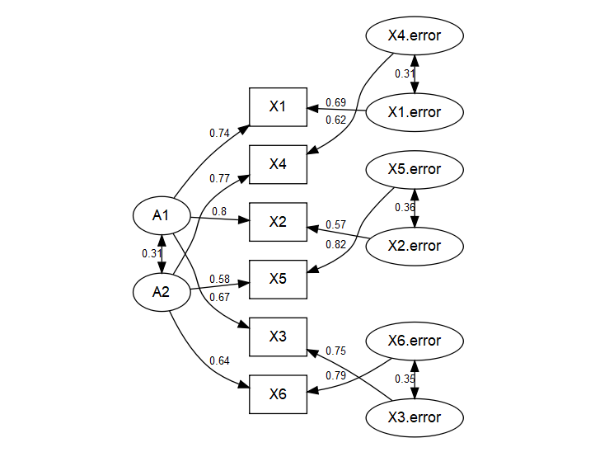
Corr(X2, X5) = Corr(+, +)== 0.69 \* 0.36 \* 0.84 = 0.208656 이다

(b)

Corr(X1, X2)가 새로 생겼다고 하여도 이 상관관계는 X2, X5와는 관계가 없다.

따라서, Corr(X2, X5) = 0.208656으로 변하지 않는다.

6.4



Model Fit :

X1=+

X2=+

X3=+

X4=+

X5=+

X6=+

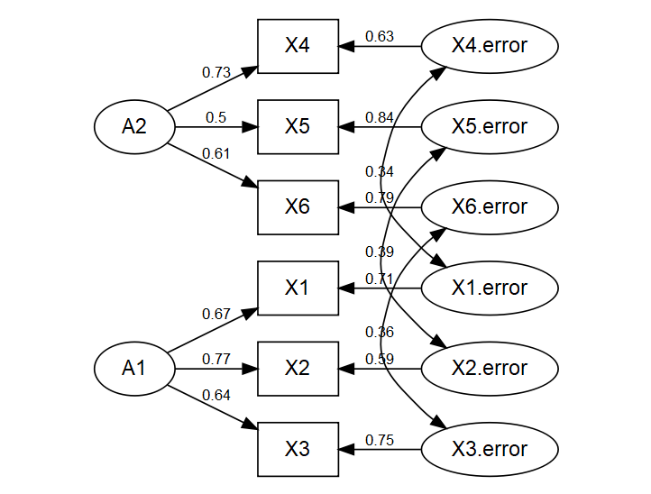
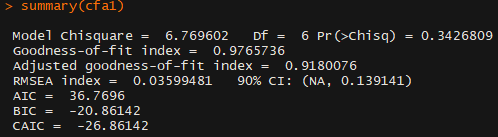
Assumptions.

Corr(X1, X4), Corr(X2, X5), Corr(X3, X6) 존재함.

모델 자체는 Chisq 통계량의 p-value가 커서 적합하다고 할 수 있고

Goodness of Fit 도 매우 높아 괜찮다고 할 수 있다.

(b)



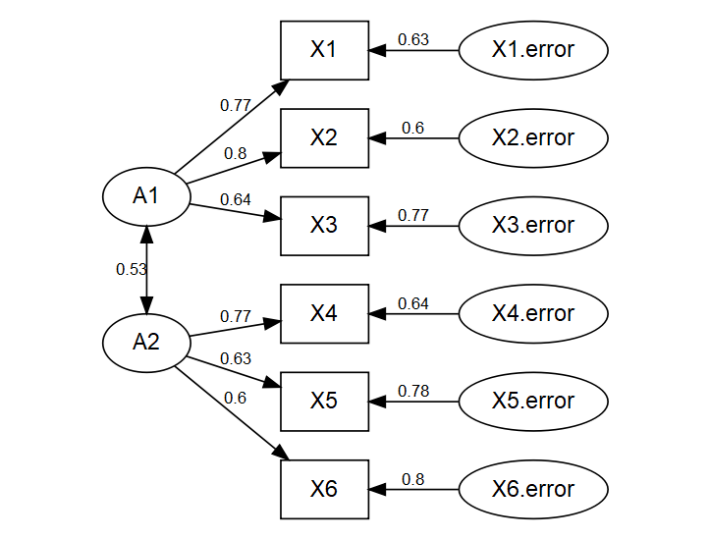
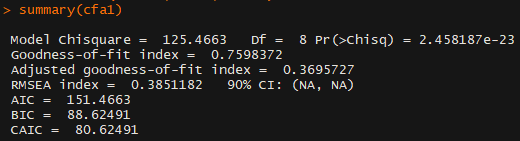
둘이 Separate 되어 있을 때의 통계량은 = 6.7696

연관성이 있을 때의 통계량은 = 1.2491

이므로, 두 deviance를 비교하는 통계량은 6.7696-1.2491=5.5206 > =3.91

따라서 유의수준 0.05 하에서 연관성이 있을 때의 모형이 더 낫다고 할 수 있다.

(c)



Method Variance를 없앤 모델이다.

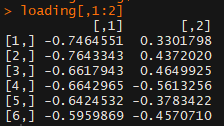
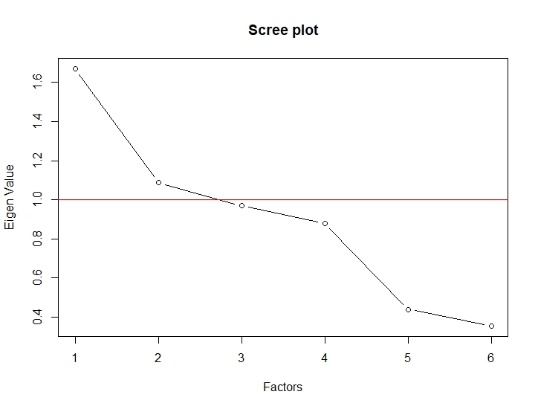
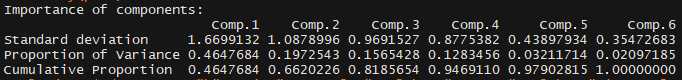
Chisq 통계량의 P-value가 0에 수렴하는 모습을 보이고 있어서 모델이 적합하지 않다고 할 수 있다. Goodness of fit 역시 좋지 않다.

첫 모델과 비교 검정을 해보자면,

125.4663 – 1.249098 = 124.217202 > = 7.81

이므로, 유의수준 0.05 하에서 첫 모델이 훨씬 적합함을 알 수 있다.

(d)



Variance가 1 이상인 2개의 Component를 사용하였을 것이다.

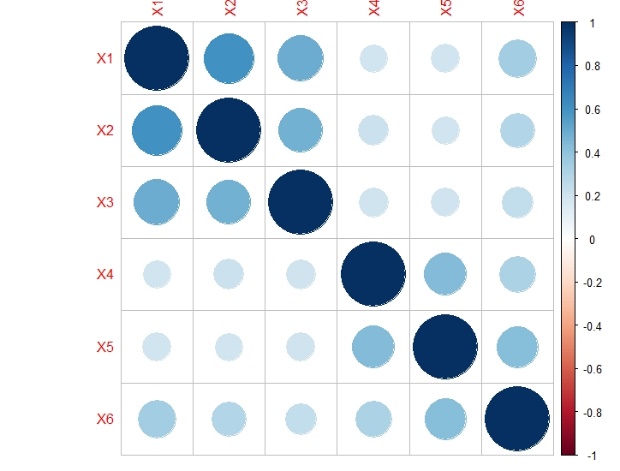
2개의 component 들은 66%가량의 설명력을 보이며,

PC1은 모든 변수와 같은 방향, PC2는 (X1,X2,X3), (X4,X5,X6)이 다른 방향으로 설명된다.

CFA와의 차이점은 CFA의 경우에는 사전에 가정한 모델을 통하여 각 변수들의 효과를 설명하고자 하는 경향이 있고, 각종 검정 방법들을 사용한다. 하지만, PCA의 경우에는 사전에 가정하는 모델이 없으며, 각종 검정 방법들을 사용하기보다는 loading matrix 등을 통하여 설명력을 보고자 한다.

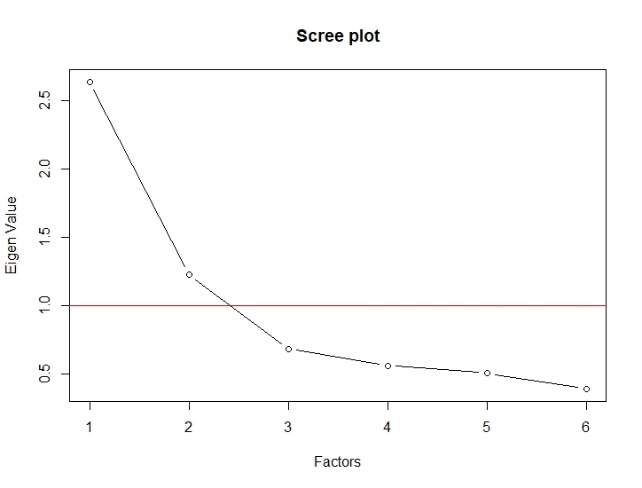
6.6

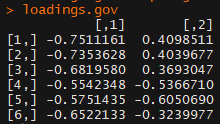
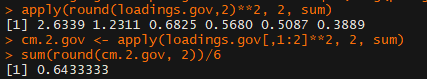
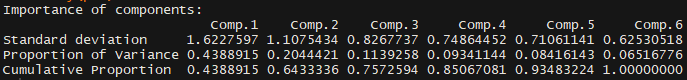
(a)



X1~3, X4~6간의 correlation이 보인다.

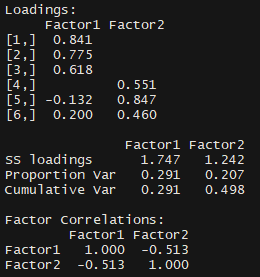
-pca





Variance가 1 이상인 2개의 PC를 사용하는 것이 좋아보인다. 두 개의 PC로 64%정도의 설명력을 보이며, PC1은 모든 변수가 같은 방향성을, PC2는 X1~3 / X4~6이 다른 방향성을 보여준다.

- EFA



PCA를 통하여 확인한대로 2개의 Factors들을 사용하며,

Factor1은 X1~X3와 큰 연관성을,

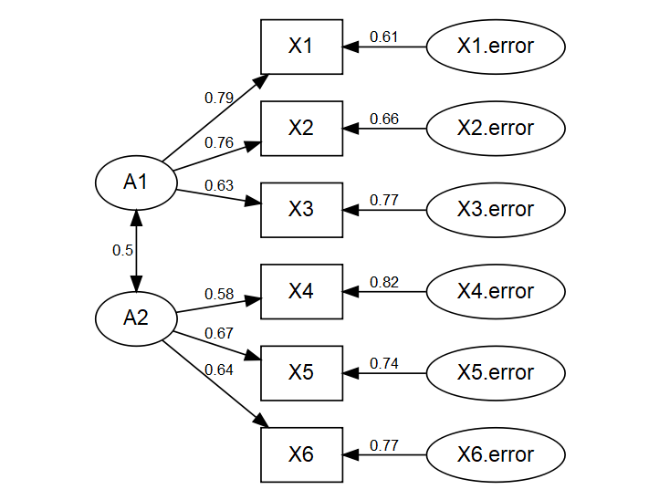
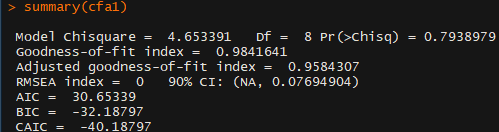
Factor2는 X4~X6와 큰 연관성을 보여준다.

이러한 연관성들을 이용하여 2-Factors CFA 모델에 적용할 수 있다. 또한 Factor간 연관성이 존재하기 때문에 이 점도 모델에 반영하도록 한다.

(b)

앞의 PCA, EFA를 통하여 2개의 Factors들을 사용하는 것이 괜찮다는 것을 보았다.

- CFA

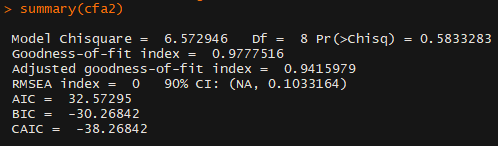


오른쪽 그림과 같은 모델로 CFA를 실시하였을 때에,

Chiq 검정 통계량의 p-value는 0.8정도로 유의수준 0.05 하에서 모델은 적합하다고 할 수 있다.

Goodness-of-fit index도 매우 높아서 이 모델은 괜찮다고 할 수 있다.

(c)



같은 모델에 대하여 동일한 방식으로 추출한 데이터를 대입하여 확인한 결과,

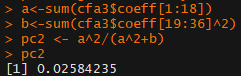
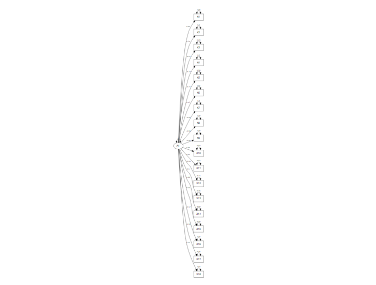
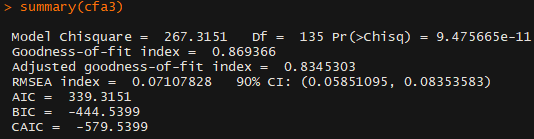
Chisq 통계량의 p-value가 0.5833으로 유의수준 0.05 하에서 모델이 적합하며,

또한 goodness-of-fit index 역시 동일하게 높아서

part b에서 가정한 모델이 유효함을 검증할 수 있다.

6.8

(a)



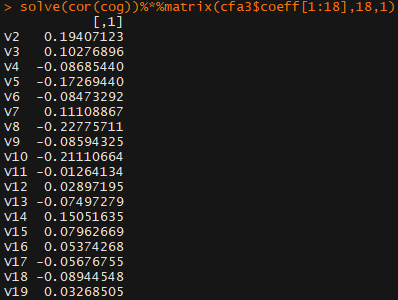
하나의 factor 아래에서 data를 적합시킨 결과,

Chisq 통계량이 0에 수렴하여 유의수준 0.05 하에서 모델이 적절하지 않다.

또한 값을 통하여 reliability를 평가해본 결과,

값이 0.02584로 매우 낮아서 reliable 하지 않다.

(b)



B = 을 구한 결과, 각 변수들의 Coeff 값들이 천차만별이다. 따라서 이들은 하나의 Factor 즉, equally weighted linear combination으로 나타내어질 수 없다.