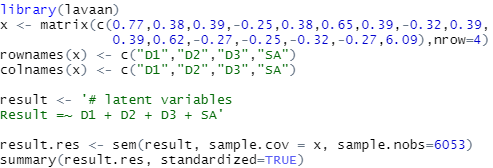
과제 4

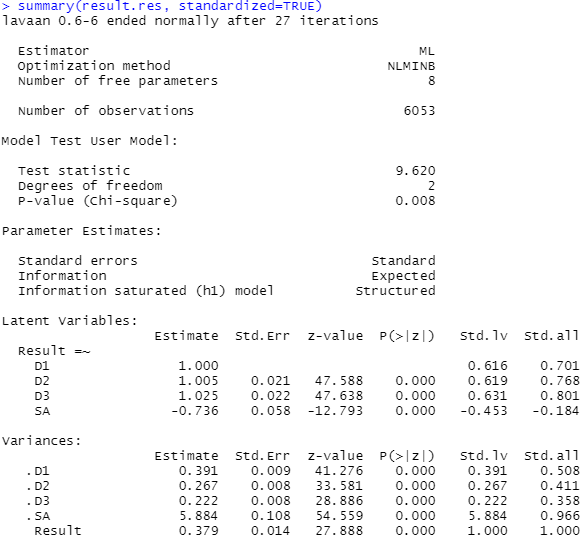
다음 1번문제는 다음 URL에 의해 볼 수 있는 “Latent Variable Modeling Using R – A Step-by-Step Guide”의 3장 pp 52-54의 연습문제들임.

([http://docshare.tips/stat-latent-variable-modeling-using-r-a-step-by-step-guide-2013-routledgepdf\_58825ee3b6d87fe8078b45d4.html](http://docshare.tips/stat-latent-variable-modeling-using-r-a-step-by-step-guide-2013-routledgepdf_58825ee3b6d87fe8078b45d4.html%20pp%2047-48))

1. Umstattd 등은(2013) 허약한 사회심리적인 건강상태(poor psychosocial health)를 우울증 질문으로부터의 3가지 항목과 사회활동의 측정값을 사용하여 일인자 모형(아래 그림 참조)으로 측정하였다. 공분산행렬은 표 3.4와 같다.

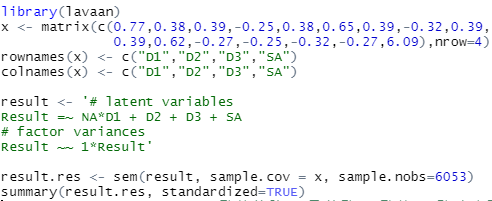
a) Depression1의 계수를 1로 하여 인자분석을 수행하시오.

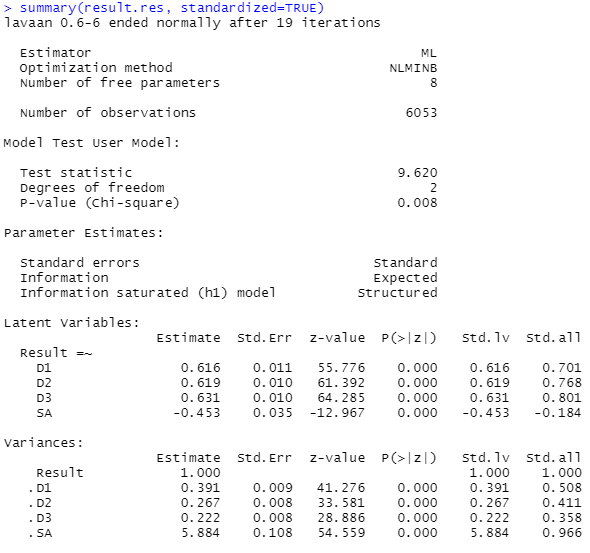


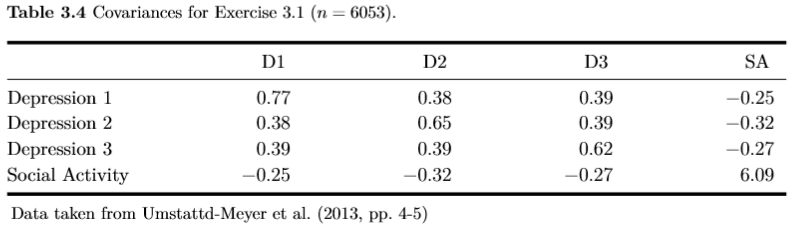


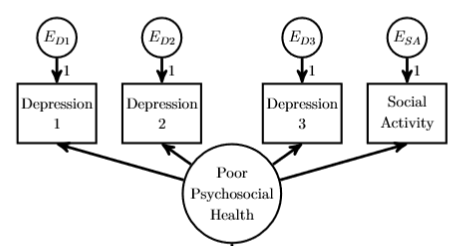
b) 인자의 분산을 1로 하여 인자분석을 수행하시오.

( a)와 b)의 과 자유도는 동일하여야 함. )









2. HolzingerSwineford1939 파일은 1939년에 Holzinger와 Swineford에 의해 수집된 심리학적인 시험 결과임. R에서 lavaan 패키지를 install하면 파일을 액세스할 수 있음

‘>HolzingerSwineford1939’ 라고 입력하면 됨.

>?HolzingerSwineford1939 라고 입력하면 설명이 나옴 이때 사용된 변수는 다음과 같음.

cf)

X1 = Visual Perception

X2 = Cubes

X3 = Lozenges

X4 = Paragraph comprehension

X5 = Sentence completion

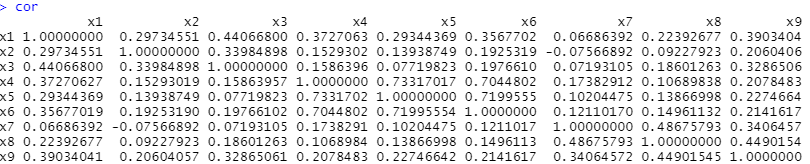
X6 = Word meaning

X7 = Addition

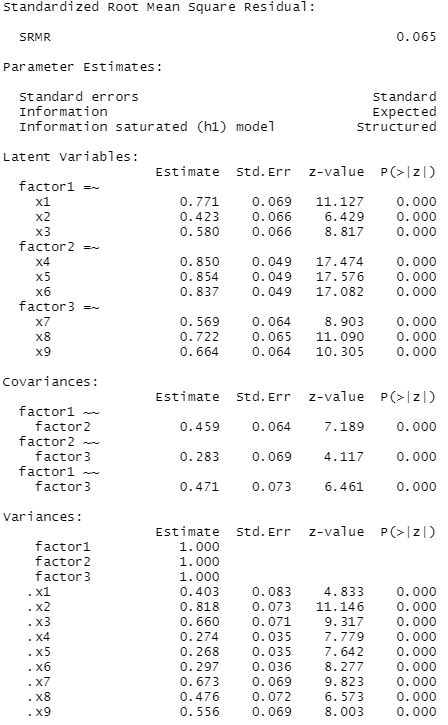
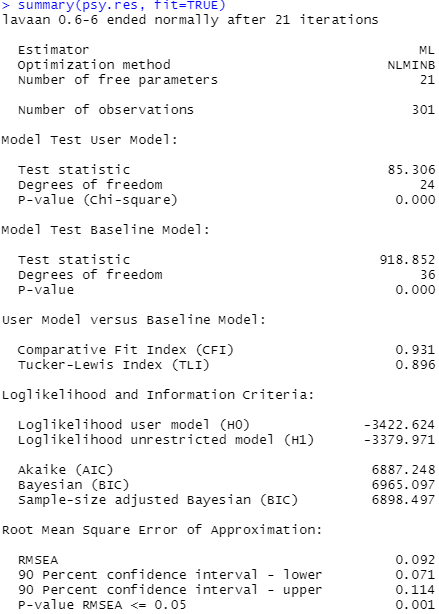
X8 = Counting dots

X9 = Straight-curved capitals

a) 9개 변수들의 상관행렬을 구하시오.



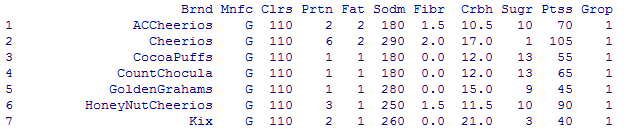
b) 9개의 시험이 3개의 공통 인자를 가진다는 가정을 확인적 인자분석에 의해 시험하시오. 이때 처음 3개의 시험은 visual perception을, 다음 3개의 시험은 verbal ability를, 마지막 3개의 시험은 speed라는 인자를 측정한다고 가정한다.



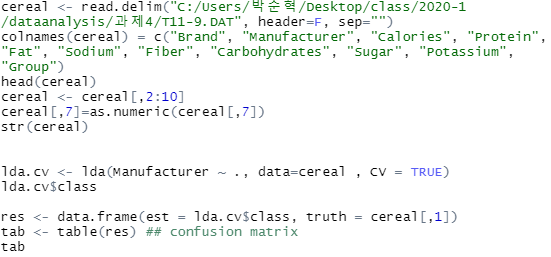
자유도가 24이고 카이제곱량은 85.306이며 p-value는 0.000이다. 이는 가정된 모형이 적절하지 않다는 것을 나타낸다.

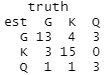
3. R. A. Johnson and D. W. Wichern의 책 “Applied Multivariate Statistical Analysis” (6th edition) 665 페이지의 다음 11.34 문제를 풀으시오.

Table 11.9는 3개의 다른 제조업체(General Mills(G), Kellogg(K), Quaker(Q))에서 만든 breakfast cereal들의 8개 변수의 측정값을 나타냄(T11-9.DAT에 있음).



a) 제조업체들의 공분산이 동일하다고 가정하고 CV(crossover validation) 방법에 의해 시리얼 브랜드들을 제조업체(두번째 열인 Manufacturer)에 대해 분류(판별분석)하시오. 단 판별을 위한 변수들로 위에서 세번째 열인 Calories부터 10번째 열인 Potassium들을 이용하시오.





Mass 라이브러리에 있는 lda함수를 사용하여 CV방법에 의해 제조업체에 대해 판별분석한 결과이다. 31개의 데이터가 예측값과 실제값이 같게 나왔고, 12개의 데이터가 예측값과 실제값이 다르게 나왔다.

b) CV(crossover validation) 방법을 사용했을 때의 에러율을 계산하시오.

27%의 에러율이 나왔다.

cf) Table 11.9는 R의 다음 명령문에 의해 읽을 수 있음.(T11-9.DAT 파일이 C\\Users\\ 디렉토리에 있다고 가정할 때)

* cereal <- read.delim("C:\\Users\\T11-9.DAT", header=F, sep="")
* colnames(cereal) = c("Brand", "Manufacturer", "Calories", "Protein", "Fat", "Sodium", "Fiber", "Carbohydrates", "Sugar", "Potassium", "Group")
* head(cereal)
* cereal <- cereal[,2:10]
* cereal[,7]=as.numeric(cereal[,7])
* str(cereal)

4. R. A. Johnson and D. W. Wichern의 책 “Applied Multivariate Statistical Analysis” (6th edition) 603 페이지의 다음 예제 11.8을 풀으시오.

연어들은 민물에서 태어나서 1년 또는 2년 후에 바다로 감. 그 후 몇 년간 바다에서 지낸후 알을 낳기 위해 그들의 탄생지로 돌아와서 알을 낳은 후 죽음. 전형적으로 알래스카 산 연어는 캐나다 산 연어보다 매년 성장하는 비늘의 링의 폭(diameter of ring of scale)이 적음. Table 11.2는 100배로 확장된 성장 링의 직경을 나타냄.

: 처음 1년간 민물에서 성장할 때의 링의 직경(1/100 인치)

: 처음 1년간 바다에서 성장할 때의 링의 직경(1/100 인치)

Country : 1=Alaskan, 2=Canadian

Gender : 1=암놈, 2=숫놈

Country Gender Freshwater Marine

1 1 2 108 368

2 1 1 131 355

3 1 1 105 469

4 1 2 86 506

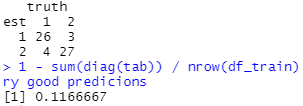
5 1 1 99 402

6 1 2 87 423

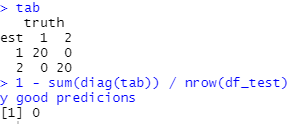
a) 알래스카 산 연어와 캐나다 산 연어의 공분산이 동일하다고 가정하고 1-30, 51-80번째 연어들을 training data, 31-50, 81-100 번째 연어들을 test data로 한 후 선형판별분석에 의해 알래스카 산 연어와 캐나다 산 연어들을 분류하고자 한다. 다음 경우에 대해 각각 training data와 test data에 대한 에러율을 계산하시오.

a-1) 판별을 위한 변수들로 위에서 세번째 열인 Freshwater와 네번째 열인 Marine을 이용하는 경우.

Training data 에러 : 11.66%

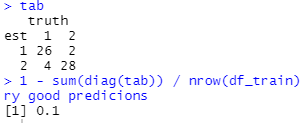


Test data 에러 : 0%

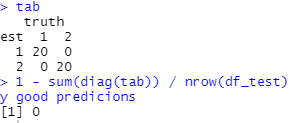


a-2) 판별을 위한 변수들로 위에서 두번째 열인 Gender, 세번째열인 Freshwater와 네번째 열인 Marine을 이용하는 경우.

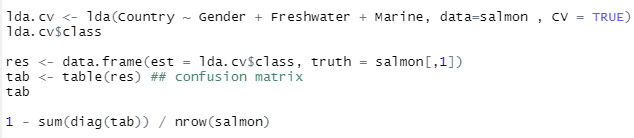
Training data 에러 : 10%

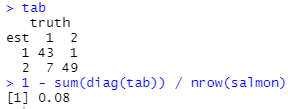


Test data 에러 : 0%



b) 선형판별분석에 의해 분석할 때 CV(crossover validation) 방법에 의해 알래스카 산 연어와 캐나다 산 연어들을 분류하고 에러율을 계산하시오. 단 판별을 위한 변수들로 위에서 두번째 열인 Gender, 세번째열인 Freshwater와 네번째 열인 Marine을 이용하시오.





에러율은 8%이다.

cf) Table 11.2는 R의 다음 명령문에 의해 읽을 수 있음.(T11-2.DAT 파일이 C\\Users\\ 디렉토리에 있다고 가정할 때)

salmon <- read.delim("C\\Users \\T11-2.DAT", header=F, sep="")

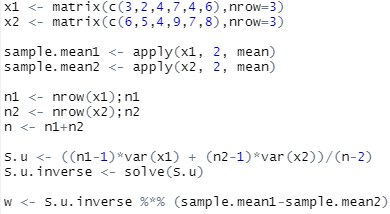
salmon

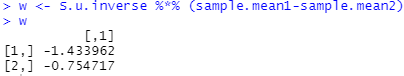
colnames(salmon) = c("Country", "Gender", "Freshwater", "Marine")

5. 다음 두개의 데이터 셑을 고려함.

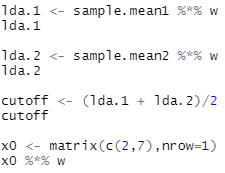
= and =

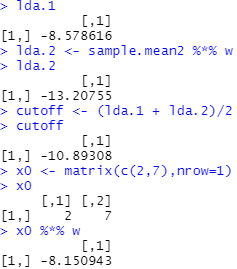
1. 선형 판별식을 a = W−1()을 이용하여 구하시오.





1. 을 판별법칙 |−| < | −| for all 를 사용하여 집단 1( 또는 집단 2(로 판단하시오..





답 : w0의 값이 cutoff값보다 크기 때문에 집단 1에 속한다.