인자분석을 위한 보충 R 코드와 결과

이재용, 임요한

서울대학교 통계학과

September 20, 2017

Contents

1	셋업	1
	기대수명 자료 분석	2
	2.1 자료	2
3	마약사용 자료 분석	8

1 셋업

```
library("MVA")

## Loading required package: HSAUR2

## Loading required package: tools

set.seed(280875)

library("lattice")

##

## Attaching package: 'lattice'

##

## The following object is masked from 'package:boot':

##

## melanoma
```

```
lattice.options(default.theme = function() standard.theme("pdf", color = FALSE))

if (file.exists("deparse.R")) {
   if (!file.exists("figs")) dir.create("figs")
   source("deparse.R")
   options(prompt = "R", continue = "+ ", width = 64,
        digits = 4, show.signif.stars = FALSE, useFancyQuotes = FALSE)

options(SweaveHooks = list(onefig = function() {par(mfrow = c(1,1))},
        twofig = function() {par(mfrow = c(1,2))},
        figtwo = function() {par(mfrow = c(2,1))},
        threefig = function() {par(mfrow = c(3,1))},
        figthree = function() {par(mfrow = c(2,2))},
        sixfig = function() {par(mfrow = c(3,2))},
        nomar = function() par("mai" = c(0, 0, 0, 0))))
}
```

참고 lattice.options 함수는 lattice 패키지를 유지 보수하는데 쓰이는 함수이다. 여기에 꼭 필요한 것인지 모르겠다.

2 기대수명 자료 분석

2.1 자료

```
"life" <-
structure(.Data = list(c(63., 34., 38., 59., 56., 62., 50., 65., 56., 69., 65., 64., 56., 60.,
61., 49., 59., 63., 59., 65., 65., 64., 64., 67., 61., 68., 67., 65., 59., 58., 57.)
, c(51., 29., 30., 42., 38., 44., 39., 44., 46., 47., 48., 50., 44., 44., 45., 40., 42., 44.,
44., 48., 48., 63., 43., 45., 40., 46., 45., 46., 43., 44., 46.)
, c(30., 13., 17., 20., 18., 24., 20., 22., 24., 24., 26., 28., 25., 22., 22., 22., 22., 23.,
24., 28., 26., 21., 21., 23., 21., 23., 23., 24., 23., 24., 28.)
, c(13., 5., 7., 6., 7., 7., 7., 7., 11., 8., 9., 11., 10., 6., 8., 9., 6., 8., 8., 14., 9.,
7., 6., 8., 10., 8., 8., 9., 10., 9., 9.)
, c(67., 38., 38., 64., 62., 69., 55., 72., 63., 75., 68., 66., 61., 65., 65., 51., 61., 67.,
63., 68., 67., 68., 68., 74., 67., 75., 74., 71., 66., 62., 60.)
```

life 자료를 구축했다.

```
toLatex(HSAURtable(life), pcol = 1, rownames = TRUE,
caption = "Life expectancies for different countries by age and gender.", label = "ch:EFA:life:tab")
## \index{life data@\Robject{life} data}
## \begin{center}
## \begin{longtable}{l rrrrrrrr }
## \caption{\Robject{life} data. Life expectancies for different countries by age and gender. \label{ch
## \\
## \hline
     & \Robject{m0} & \Robject{m25} & \Robject{m50} & \Robject{m75} & \Robject{w0} & \Robject{w25} & \lambda
## \endfirsthead
## \caption[]{\Robject{life} data (continued).} \\
## \hline
     & \Robject{m0} & \Robject{m25} & \Robject{m50} & \Robject{m75} & \Robject{w0} & \Robject{w25} & \lambda
## \endhead
## Algeria
             & 63 & 51 & 30 & 13 & 67 & 54 & 34 & 15 \\
```

Cameroon & 34 & 29 & 13 & 5 & 38 & 32 & 17 & 6 \\

```
& 38 & 30 & 17 & 7 & 38 & 34 & 20 & 7 \\
## Madagascar
## Mauritius
            & 59 & 42 & 20 & 6 & 64 & 46 & 25 & 8 \\
## Reunion
           & 56 & 38 & 18 & 7 & 62 & 46 & 25 & 10 \\
## Seychelles
             & 62 & 44 & 24 & 7 & 69 & 50 & 28 & 14 \\
## South Africa (C) & 50 & 39 & 20 & 7 & 55 & 43 & 23 & 8 \\
## South Africa (W)
                  & 65 & 44 & 22 & 7 & 72 & 50 & 27 & 9 \\
## Tunisia
         & 56 & 46 & 24 & 11 & 63 & 54 & 33 & 19 \\
## Canada
          & 69 & 47 & 24 & 8 & 75 & 53 & 29 & 10 \\
## Costa Rica & 65 & 48 & 26 & 9 & 68 & 50 & 27 & 10 \\
## El Salvador & 56 & 44 & 25 & 10 & 61 & 48 & 27 & 12 \\
## Greenland & 60 & 44 & 22 & 6 & 65 & 45 & 25 & 9 \\
          & 61 & 45 & 22 & 8 & 65 & 49 & 27 & 10 \\
## Grenada
## Guatemala
            & 49 & 40 & 22 & 9 & 51 & 41 & 23 & 8 \\
## Honduras & 59 & 42 & 22 & 6 & 61 & 43 & 22 & 7 \\
         & 63 & 44 & 23 & 8 & 67 & 48 & 26 & 9 \\
## Jamaica
          & 59 & 44 & 24 & 8 & 63 & 46 & 25 & 8 \\
## Mexico
## Nicaragua
            & 65 & 48 & 28 & 14 & 68 & 51 & 29 & 13 \\
          & 65 & 48 & 26 & 9 & 67 & 49 & 27 & 10 \\
## Panama
## Trinidad (62)
               & 64 & 63 & 21 & 7 & 68 & 47 & 25 & 9 \\
## Trinidad (67)
                & 64 & 43 & 21 & 6 & 68 & 47 & 24 & 8 \\
                   & 67 & 45 & 23 & 8 & 74 & 51 & 28 & 10 \\
## United States (66)
                    & 61 & 40 & 21 & 10 & 67 & 46 & 25 & 11 \\
## United States (NW66)
## Argentina
            & 65 & 46 & 24 & 9 & 71 & 51 & 28 & 10 \\
## Chile
        & 59 & 43 & 23 & 10 & 66 & 49 & 27 & 12 \\
## Colombia & 58 & 44 & 24 & 9 & 62 & 47 & 25 & 10 \
## Ecuador & 57 & 46 & 28 & 9 & 60 & 49 & 28 & 11 \\
## \hline
## \end{longtable}
## \end{center}
```

toLatex는 R 객체를 라텍으로 만들어주는 함수이다. HSAURtable은 latex의 longtable 라텍 environment 를 만들어주는 함수이다.

```
sapply(1:3, function(f) factanal(life, factors = f)$PVAL)

## objective objective
## 1.880e-24 1.912e-05 4.578e-01
```

sapply는 벡터나 리스트의 원소에 함수를 적용하는 함수이다. 여기서 함수는 인자의 개수를 인자(argument)로 받아들여 그 개수의 인자모형을 적합해서 P 값을 구하는 것이다. factanal은 인자분석을 수해하는 명령어이다. method는 언제나 mle라고 한다. 결과는 각 인자의 개수가 1, 2, 3인 모형 3개를 적합하고 각모형의 p 값을 출력하는 명령이다. 여기서 1:3을 넣어서 돌아갔는데, 4이상이 들어가면 에러가 발생한다. 이를 기반으로 factor의 개수가 3이라고 결론을 낸다.

```
factanal(life, factors = 3, method ="mle", rotation="varimax")
##
## Call:
## factanal(x = life, factors = 3, rotation = "varimax", method = "mle")
##
## Uniquenesses:
##
     m0 m25 m50 m75
                            w0 w25
                                      w50
                                          พ75
## 0.005 0.362 0.066 0.288 0.005 0.011 0.020 0.146
##
## Loadings:
      Factor1 Factor2 Factor3
## m0 0.964
            0.122 0.226
## m25 0.646 0.169 0.438
## m50 0.430
            0.354
                    0.790
## m75
             0.525
                     0.656
## w0 0.970
             0.217
## w25 0.764
             0.556
                     0.310
## w50 0.536
             0.729
                    0.401
## w75 0.156
                     0.280
             0.867
##
##
                Factor1 Factor2 Factor3
## SS loadings
                  3.375 2.082 1.640
## Proportion Var
                  0.422 0.260
                                 0.205
## Cumulative Var 0.422 0.682
                                 0.887
##
```

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 6.73 on 7 degrees of freedom.
The p-value is 0.458

다음의 모형을 적합했다.

$$X = \Lambda f + u, \ u \sim N(0, \Psi).$$

여기서 X는 k=8차원 랜덤벡터이고, Λ 는 8×3 인자적재(factor loading)이다. 인자는 $f=(f_1,f_2,f_3)$ 로 나타낸다. 결과에서 uniqueness는 특정분산으로 u의 분산 대각행렬 Ψ 의 대각원소이다. Loadings는 Λ 행렬을 나타낸다. 빈칸이 있는데 0을 의미하는 것 같다.

SS loadings는 각 인자의 공통성 h_i 이다. cummulative var는 공통성과 유일성을 합했을 때 공통성의 비율이다. 즉,

$$h_1 + h_2 + h_3 + \sum_{j=1}^{k} \psi_j$$

를 100으로 했을 때 h_i 들의 비율이다.

마지막 부분은 가설검정인데

 H_0 : 적합한 모형이 옳다.

에 대한 가설이다. p-value가 작으면 이 모형이 적당하지 않다는 강한 증거이다. 여기서는 0.458이므로 이 모형이 틀리다는 뚜렷한 증거는 없는 것이므로, 이 모형이 적당하다는 뜻이다.

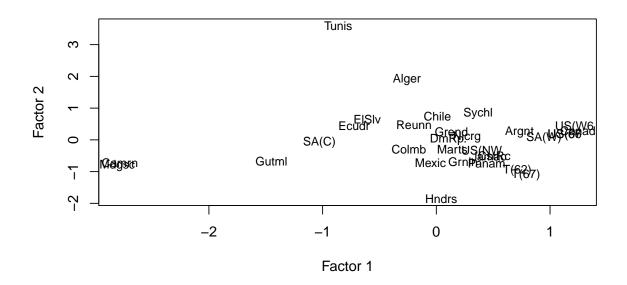
```
scores <- factanal(life, factors = 3, method = "mle", scores = "regression")$scores</pre>
```

인자점수(factor scores)를 구한다.

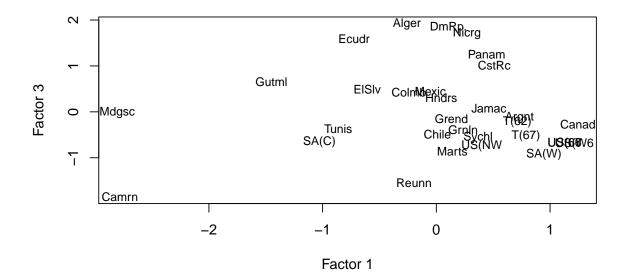
$$X = Lf + u$$

에서 \hat{f} 을 구한 것이다. 따라서, 각 관측치마다 한 개의 인자점수 벡터가 있다. 이는 각 관측치를 간단하게 요약한 통계량이라 볼 수 있다.

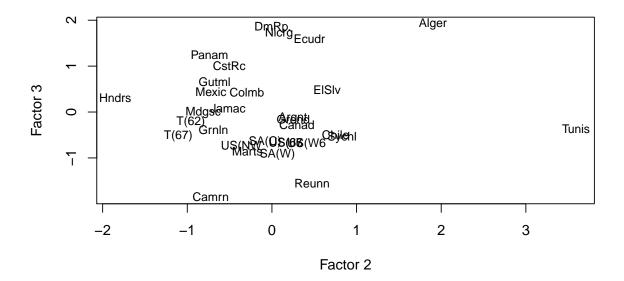
```
cex <- 0.8
plot(scores[,1], scores[,2], type = "n", xlab = "Factor 1", ylab = "Factor 2")
text(scores[,1], scores[,2], abbreviate(rownames(life), 5), cex = cex)</pre>
```



```
plot(scores[,1], scores[,3], type = "n", xlab = "Factor 1", ylab = "Factor 3")
text(scores[,1], scores[,3], abbreviate(rownames(life), 5), cex = cex)
```



```
plot(scores[,2], scores[,3], type = "n", xlab = "Factor 2", ylab = "Factor 3")
text(scores[,2], scores[,3], abbreviate(rownames(life), 5), cex = cex)
```



인자점수를 이용하여 관측치들을 인자들의 좌표에 산점도를 그린 것이다.

3 마약사용 자료 분석

참고 코드를 자세히 보지 않았다.

```
d <-
c(0.447,
    0.422, 0.619,
    0.435, 0.604, 0.583,
    0.114, 0.068, 0.053, 0.115,
    0.203, 0.146, 0.139, 0.258, 0.349,
    0.091, 0.103, 0.110, 0.122, 0.209, 0.221,
    0.082, 0.063, 0.066, 0.097, 0.321, 0.355, 0.201,
    0.513, 0.445, 0.365, 0.482, 0.186, 0.315, 0.150, 0.154,
    0.304, 0.318, 0.240, 0.368, 0.303, 0.377, 0.163, 0.219, 0.534,
    0.245, 0.203, 0.183, 0.255, 0.272, 0.323, 0.310, 0.288, 0.301, 0.302,
    0.101, 0.088, 0.074, 0.139, 0.279, 0.367, 0.232, 0.320, 0.204, 0.368, 0.340,</pre>
```

```
0.245, 0.199, 0.184, 0.293, 0.278, 0.545, 0.232, 0.314, 0.394, 0.467, 0.392, 0.511)
```

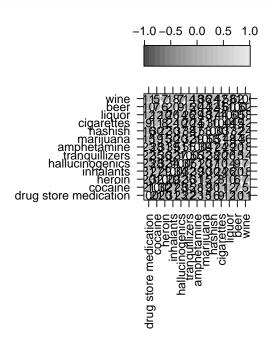
길이가 78인 벡터이다. 어떤 값인지는 모르겠다.

```
druguse <- diag(13) / 2
druguse[upper.tri(druguse)] <- d</pre>
```

upper.tri 함수는 druguse 행렬과 동일한 크기의 논리행렬으로서 상삼각 원소는 TRUE 나머지는 FALSE로 값이 들어간다. 여기서 상삼각원소라는 것은 j j i 를 만족하는 (i,j)를 말한다. druguse의 상삼각부분에 벡터 d의 원소를 대입한다.

druguse를 대칭이 되도록 만들었다. 행과 열의 이름을 대입했다.

```
### code chunk number 8: ch:EFA:druguse:plot
ord <- order.dendrogram(as.dendrogram(hclust(dist(druguse))))</pre>
panel.corrgram <-</pre>
     function(x, y, z, subscripts, at,
             level = 0.9, label = FALSE, ...)
 {
     require("ellipse", quietly = TRUE)
     x <- as.numeric(x)[subscripts]</pre>
     y <- as.numeric(y)[subscripts]</pre>
     z <- as.numeric(z)[subscripts]</pre>
     zcol <- level.colors(z, at = at, col.regions = grey.colors, ...)</pre>
     for (i in seq(along = z)) {
         ell <- ellipse(z[i], level = level, npoints = 50,
                       scale = c(.2, .2), centre = c(x[i], y[i]))
         panel.polygon(ell, col = zcol[i], border = zcol[i], ...)
     if (label)
         panel.text(x = x, y = y, lab = 100 * round(z, 2), cex = 0.8,
```



```
(factanal(covmat = druguse, factors = 6,
                     method = "mle", n.obs = 1634))
##
## Call:
## factanal(factors = 6, covmat = druguse, n.obs = 1634, method = "mle")
## Uniquenesses:
##
             cigarettes
                                         beer
                                                               wine
##
                  0.563
                                        0.368
                                                              0.374
                                                     tranquillizers
##
                 liquor
                                      cocaine
                  0.412
                                                              0.522
##
                                        0.681
## drug store medication
                                       heroin
                                                          marijuana
                  0.785
                                        0.669
##
                                                              0.318
##
                hashish
                                    inhalants
                                                    hallucinogenics
                  0.005
                                        0.541
                                                              0.620
##
##
            amphetamine
                  0.005
##
##
## Loadings:
                        Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
##
## cigarettes
                         0.494
                                                         0.407
                                                                 0.110
                        0.776
## beer
                                                         0.112
## wine
                         0.786
## liquor
                         0.720
                                 0.121
                                         0.103 0.115
                                                         0.160
## cocaine
                                 0.519
                                                 0.132
                                                                 0.158
## tranquillizers
                         0.130
                                 0.564
                                         0.321
                                                 0.105
                                                         0.143
                                                                 0.372
## drug store medication
                                 0.255
## heroin
                                 0.532
                                         0.101
                                                                 0.190
## marijuana
                                0.158
                                         0.152
                                                 0.259
                                                         0.609 0.110
                         0.429
## hashish
                         0.244
                                 0.276
                                         0.186
                                                 0.881
                                                         0.194
                                                                 0.100
## inhalants
                         0.166
                                 0.308
                                         0.150
                                                         0.140
                                                                 0.537
## hallucinogenics
                                 0.387
                                         0.335
                                                 0.186
                                                                 0.288
## amphetamine
                         0.151
                                 0.336
                                         0.886
                                                 0.145
                                                         0.137
                                                                 0.187
##
```

```
Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
##
## SS loadings
                   2.301 1.415 1.116
                                         0.964
                                                  0.676
                                                          0.666
## Proportion Var
                   0.177
                         0.109
                                  0.086
                                         0.074
                                                  0.052
                                                          0.051
## Cumulative Var
                   0.177
                          0.286
                                  0.372
                                         0.446
                                                  0.498
                                                          0.549
##
## Test of the hypothesis that 6 factors are sufficient.
## The chi square statistic is 22.41 on 15 degrees of freedom.
## The p-value is 0.0975
```

```
### code chunk number 11: ch:EFA:druqdiff
pfun <- function(nf) {</pre>
     fa <- factanal(covmat = druguse, factors = nf,</pre>
                  method = "mle", n.obs = 1634)
     est <- tcrossprod(fa$loadings) + diag(fa$uniquenesses)</pre>
     ret <- round(druguse - est, 3)
     colnames(ret) <- rownames(ret) <-</pre>
         abbreviate(rownames(ret), 3)
     ret
pfun(6)
##
         cgr
               ber
                     win
                           lqr
                                 ccn
                                        trn
                                              dsm
                                                     hrn
                                                           mrj hsh
## cgr 0.000 -0.001 0.014 -0.018 0.010 0.001 -0.020 -0.004 0.001
## ber -0.001 0.000 -0.002 0.004 0.004 -0.011 -0.001 0.007
## win 0.014 -0.002 0.000 -0.001 -0.001 -0.005 0.008 0.008 -0.004
## lqr -0.018 0.004 -0.001 0.000 -0.008 0.021 -0.006 -0.018 0.003
## ccn 0.010 0.004 -0.001 -0.008 0.000 0.000 0.008 0.004 -0.004
## trn 0.001 -0.011 -0.005 0.021 0.000 0.000 0.006 -0.004 -0.004
## dsm -0.020 -0.001 0.008 -0.006 0.008 0.006 0.000 -0.015 0.008
## hrn -0.004 0.007 0.008 -0.018 0.004 -0.004 -0.015 0.000
                                                         0.006
## mrj 0.001 0.002 -0.004 0.003 -0.004 -0.004 0.008 0.006
                                                         0.000
## hsh 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## inh 0.010 -0.004 -0.007 0.012 -0.003 0.002 0.004 -0.002 -0.006
```

```
## hll -0.005 0.005 -0.001 -0.005 -0.008 -0.008 -0.002 0.020 0.003 0
## amp 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##
       inh hll amp
## cgr 0.010 -0.005 0
## ber -0.004 0.005 0
## win -0.007 -0.001
## lqr 0.012 -0.005 0
## ccn -0.003 -0.008 0
## trn 0.002 -0.008 0
## dsm 0.004 -0.002 0
## hrn -0.002 0.020 0
## mrj -0.006 0.003 0
## hsh 0.000 0.000 0
## inh 0.000 -0.002 0
## hll -0.002 0.000 0
## amp 0.000 0.000
```

```
ber -0.001 0.000 -0.002 0.002 0.002 -0.014 0.000 0.005
##
##
                win 0.009 -0.002 0.000 0.000 -0.002 -0.004 0.012 0.013
##
                lgr -0.013 0.002 0.000 0.000 -0.008 0.024 -0.017 -0.020
##
                   0.011 0.002 -0.002 -0.008 0.000 0.031 0.038 0.082
                    0.009 -0.014 -0.004 0.024 0.031
                                                     0.000 -0.021
##
                dsm -0.011 0.000 0.012 -0.017 0.038 -0.021
##
                                                            0.000
                                                                  0.021
               hrn -0.004 0.005 0.013 -0.020 0.082 0.026
##
                                                            0.021
                                                                  0.000
                    0.003 -0.001 0.001 -0.001 -0.002 -0.002 0.007 0.006
##
               mrj
                           ##
               hsh -0.027
##
                    0.039 -0.002 -0.007 -0.002 0.023 -0.038
                                                           0.113 0.031
                inh
               hll -0.017 0.009 0.004 -0.015 -0.030 -0.058 0.000 -0.005
##
##
                    0.002 -0.007
                                 0.002 0.006 -0.075 0.044 -0.038 -0.049
                amp
##
                      mrj
                             hsh
                                   inh
                                          hll
                                                 amp
                cgr 0.003 -0.027 0.039 -0.017 0.002
##
##
                ber -0.001 0.019 -0.002 0.009 -0.007
                win 0.001 -0.017 -0.007 0.004 0.002
##
##
               lqr -0.001 0.014 -0.002 -0.015 0.006
                ccn -0.002 0.041 0.023 -0.030 -0.075
##
##
                trn -0.002 -0.016 -0.038 -0.058 0.044
##
                    0.007 -0.040
                                 0.113 0.000 -0.038
                dsm
               hrn 0.006 -0.035 0.031 -0.005 -0.049
##
                    0.000 0.001 0.003 -0.002 -0.002
##
               mrj
                    0.001 0.000 -0.035 0.034 0.010
##
               hsh
                    0.003 -0.035 0.000 0.007 -0.015
##
                inh
##
               hll -0.002 0.034
                                 0.007
                                        0.000
                amp -0.002 0.010 -0.015 0.041
##
pfun(4)
##
                             ber
                                          lqr
                                                              dsm
                      cgr
                                   win
                                                 ccn
                                                       trn
                                                                     hrn
                cgr 0.000 -0.001 0.008 -0.012 0.009 0.008 -0.015 -0.007
##
##
               ber -0.001
                           0.000 -0.001 0.001 0.000 -0.016 -0.002 0.003
##
                   0.008 -0.001 0.000 0.000 -0.001 -0.005 0.012 0.014
                lqr -0.012 0.001
                                0.000 0.000 -0.004 0.029 -0.015 -0.015
##
                           0.000 -0.001 -0.004 0.000 0.024 -0.014 0.007
##
                    0.009
##
                   0.008 -0.016 -0.005 0.029 0.024 0.000 -0.020 0.027
                trn
```

```
##
              hrn -0.007 0.003 0.014 -0.015 0.007 0.027 -0.018 0.000
##
##
              mrj 0.001 -0.001 0.001 -0.001 -0.003 -0.001 0.003 0.003
##
              ##
              inh 0.037 -0.005 -0.008 0.001 -0.022 -0.032 0.090 -0.001
              hll -0.020 0.006 0.001 -0.010 -0.028 -0.028 0.008 0.005
##
              amp 0.000 0.000 0.000 -0.001 0.000 0.001 0.000 0.000
##
##
                    mrj
                         hsh
                               inh
                                     hll
                                           amp
              cgr 0.001 -0.023 0.037 -0.020 0.000
##
##
              ber -0.001 0.018 -0.005 0.006 0.000
##
              win 0.001 -0.020 -0.008 0.001 0.000
##
              lgr -0.001 0.018 0.001 -0.010 -0.001
##
              ccn -0.003 0.035 -0.022 -0.028 0.000
              trn -0.001 0.001 -0.032 -0.028 0.001
##
##
              dsm 0.003 -0.042 0.090 0.008 0.000
##
              hrn 0.003 -0.037 -0.001 0.005 0.000
              mrj 0.000 0.000 0.001 -0.002 0.000
##
##
              hsh 0.000 0.000 -0.031 0.055 -0.001
##
              inh 0.001 -0.031 0.000 0.021 0.000
##
              hll -0.002 0.055
                              0.021
                                    0.000
##
              amp 0.000 -0.001 0.000 0.000 0.000
```