Statistics 5401
22. Factor Extraction (addendum)

Gary W. Oehlert
School of Statistics
313B Ford Hall
612-625-1557
gary@stat.umn.edu

OK, here is a "good" algorithm for doing factor extraction. It's called `facanal()` and it is in the file `facanal.mac` or `facanal.mac.txt` on the course web page. It's fast.

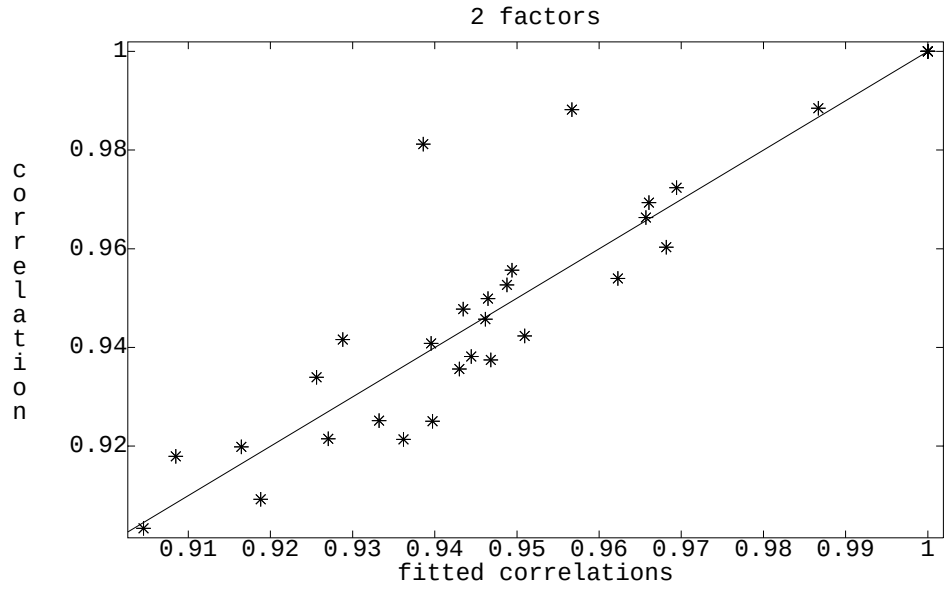
```
Cmd> readdata("",x1,x2,x3,x4,d)
Read from file "~/5401/JW5data/T4-3.DAT"
Column 1 saved as REAL vector x1
Column 2 saved as REAL vector x2
Column 3 saved as REAL vector x3
Column 4 saved as REAL vector x4
Column 5 saved as REAL vector d

Cmd> stiff <- hconcat(x1,x2,x3,x4)

Cmd> readdata("",x1,x2,x3,x4,gender)
Read from file "~/5401/JW5data/T6-12.DAT"
Column 1 saved as REAL vector x1
Column 2 saved as REAL vector x2
Column 3 saved as REAL vector x3
Column 4 saved as REAL vector x4
Column 5 saved as factor gender

Cmd> o2 <- hconcat(x1,x2,x3,x4)

Cmd> readdata("",x1,x2,x3,x4,x5,x6,x7,x8,nation)
Read from file "~/5401/JW5data/T8-6.DAT"
Column 1 saved as REAL vector x1
Column 2 saved as REAL vector x2
Column 3 saved as REAL vector x3
Column 4 saved as REAL vector x4
Column 5 saved as REAL vector x5
Column 6 saved as REAL vector x6
Column 7 saved as REAL vector x7
Column 8 saved as REAL vector x8
Column 9 saved as factor nation

Cmd> X <- hconcat(x1,x2,x3,x4,x5,x6,x7,x8)

Cmd> cars <- matrix(vecread("cars"),22)'
```

```
Read from file "cars"

Cmd> R <- cor(stiff)

Cmd> facanal(R,1)
Convergence in 11 iterations by criterion 2
estimated uniquenesses:
(1)  0.029287   0.16735    0.1751   0.15814
unrotated estimated loadings:
(1,1)      0.98525
(2,1)       0.9125
(3,1)      0.90824
(4,1)      0.91753
minimized ml criterion:
(1)      0.49985

Cmd> facanal(R,1,method:"gls")
Convergence in 10 iterations by criterion 2
estimated uniquenesses:
(1)  0.033067   0.11957  0.096379  0.072421
unrotated estimated loadings:
(1,1)      0.97858
(2,1)      0.90492
(3,1)      0.92847
(4,1)      0.93911
minimized gls criterion:
(1)      0.23904

Cmd> facanal(R,2)
WARNING: With m = 2 and p = 4, m >
(2*p + 1 - sqrt(8*p+1))/2
Convergence in 10 iterations by criterion 2
estimated uniquenesses:
(1)   0.04093   0.08901   0.10437  0.045864
unrotated estimated loadings:
(1,1)      0.97186     -0.12064
(2,1)      0.90119      -0.3144
(3,1)      0.93199      0.16439
(4,1)      0.95153      0.22075
minimized ml criterion:
(1)   1.1209e-11

Cmd> R <- cor(X)

Cmd> facanal(R,1)
```

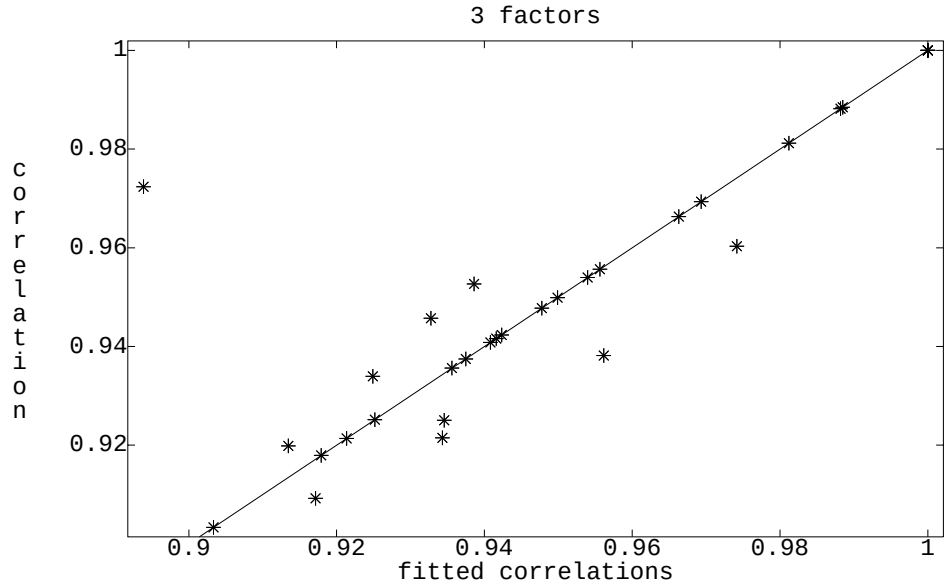
```
Convergence in 10 iterations by criterion 2
estimated uniquenesses:
(1)   0.55595   0.46087    0.3317   0.19564
(5)  0.093515  0.036322  0.024539   0.11755
unrotated estimated loadings:
(1,1)      0.66638
(2,1)      0.73426
(3,1)       0.8175
(4,1)      0.89686
(5,1)       0.9521
(6,1)      0.98167
(7,1)      0.98765
(8,1)      0.93939
minimized ml criterion:
(1)       3.1378

Cmd> facanal(R,2)
Convergence in 11 iterations by criterion 2
estimated uniquenesses:
(1)  0.080991  0.075815   0.15143   0.13533
(5)  0.081718  0.033793  0.017957  0.085993
unrotated estimated loadings:
(1,1)      0.73082     -0.62041
(2,1)      0.79165     -0.54542
(3,1)      0.85494     -0.34299
(4,1)      0.91585     -0.16087
(5,1)      0.95793    -0.025686
(6,1)      0.97239      0.14375
(7,1)      0.98062      0.14294
(8,1)      0.92291       0.2495
minimized ml criterion:
(1)      0.33273

Cmd> dim(X)
(1)           55            8

Cmd> (54-(2*8-5)/6-2*2/3)*.3327
(1)       16.912

Cmd> ((8-2)^2-8-2)/2
(1)           13

Cmd> 1-cumchi(16.9,13)
(1)      0.20392

Cmd> R <- cor(cars)
```

```
Cmd> dim(cars)
(1)           91           22

Cmd> for(i,run(4,10)) {
   out <- facanal(R,i,silent:T)
   mult <- (90 - (2*22-5)/6 - 2*i/3)
   df <- ((22-i)^2-22-i)/2
   pval <- 1-cumchi(mult*out$criterion,df)
   vector(out$criterion,df,pval)
}
(1)       6.1626          149            0
(1)        4.408          131            0
(1)       3.2526          114   3.0775e-13
(1)       2.2515           98   1.5934e-06
(1)       1.5937           83    0.0021496
(1)       1.1698           69     0.041313
(1)      0.90159           56      0.10959

Cmd> out$psihat
 (1)  1.08e-05  2.559e-6  7.074e-6  8.531e-8
 (5)  0.062139   0.51553   0.44655   0.17161
( 9)  4.872e-5  0.001189   0.13368   0.24403
(13)   0.27527   0.13032   0.10497   0.10601
(17)  0.090196  0.085037   0.21065  2.412e-5
(21)  0.016782   0.30618

Cmd> print(out$loadings,format:"f5.2")
MATRIX:
 (1,1)  0.63 -0.73 -0.09  0.20 -0.11  0.00 -0.00 -0.00  0.00 -0.00
 (2,1)  0.60 -0.80  0.01  0.01 -0.01  0.00  0.00  0.00 -0.00  0.00
 (3,1)  0.56 -0.81  0.08 -0.15  0.07  0.00 -0.00 -0.00  0.00 -0.00
 (4,1) -1.00 -0.03 -0.00  0.00  0.00 -0.00 -0.00 -0.00 -0.00  0.00
 (5,1) -0.94 -0.02 -0.01  0.06  0.05  0.03  0.09  0.18 -0.01 -0.03
 (6,1)  0.34 -0.46  0.04  0.09  0.11 -0.02 -0.04  0.33  0.01  0.14
 (7,1) -0.12  0.31 -0.03 -0.33 -0.16 -0.23 -0.18 -0.38  0.24  0.16
 (8,1)  0.69 -0.28 -0.18  0.35  0.30 -0.08 -0.04 -0.06  0.07  0.05
 (9,1)  0.75 -0.20 -0.28  0.36  0.44  0.01  0.00 -0.00 -0.00 -0.00
(10,1)  0.69 -0.45  0.03  0.21  0.22 -0.48  0.00  0.00  0.00  0.00
(11,1) -0.39 -0.26  0.28 -0.17 -0.37 -0.56  0.09  0.02  0.07 -0.27
(12,1) -0.70 -0.01  0.17 -0.26 -0.39 -0.03 -0.02 -0.12 -0.09 -0.05
(13,1) -0.47  0.11  0.34 -0.25 -0.30 -0.11  0.20  0.01 -0.41  0.07
(14,1)  0.81 -0.15 -0.21  0.02  0.21 -0.09 -0.23 -0.10 -0.14 -0.09
(15,1)  0.54  0.16 -0.53 -0.20  0.13  0.11 -0.34 -0.23  0.25 -0.08
(16,1)  0.70 -0.16 -0.33  0.23  0.23  0.01 -0.16  0.37  0.03 -0.10
(17,1)  0.72 -0.18 -0.42  0.03  0.26  0.03 -0.32  0.12  0.01 -0.06
(18,1)  0.72 -0.02 -0.25  0.22  0.44 -0.04 -0.26  0.15  0.02  0.01
(19,1)  0.67  0.02 -0.26  0.16  0.34 -0.05 -0.17  0.20  0.12  0.19
(20,1)  0.39 -0.11 -0.90 -0.14 -0.02 -0.00  0.00  0.00 -0.00  0.00
```

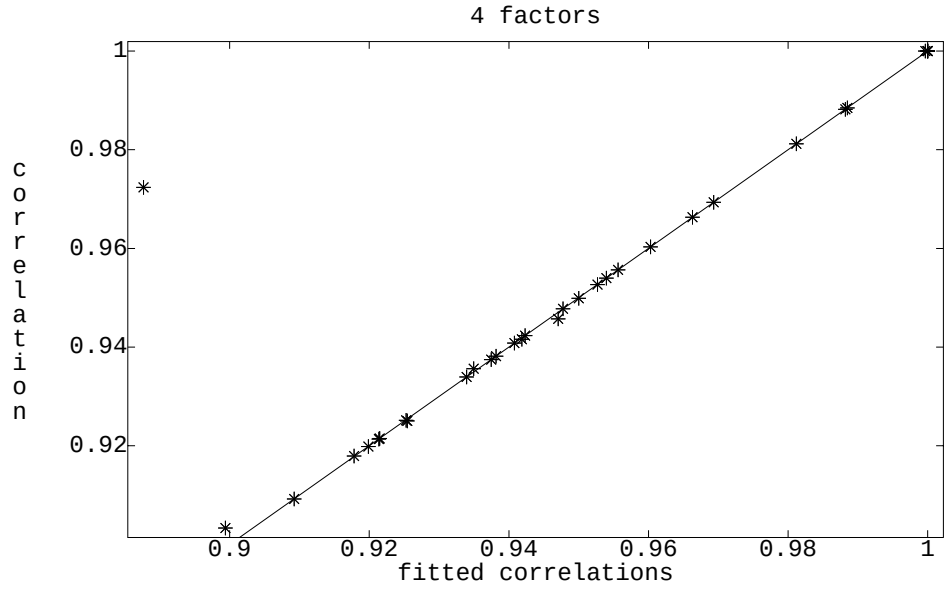
```
(21,1)  0.86 -0.19 -0.21  0.09  0.24 -0.11 -0.29 -0.01 -0.06  0.02
(22,1)  0.26  0.33 -0.08  0.18  0.37  0.16  0.08  0.24  0.46  0.20

Cmd> aaup <- matrix(vecread("aaup"),13)'
Read from file "aaup"

Cmd> aaup <- aaup[,run(8)]

Cmd> laaup <- log(aaup)

Cmd> R <- cor(laaup)

Cmd> for(i,run(2,6)) {
out <- facanal(R,i,silent:T)
L <- out$loadings
Psi <- dmat(out$psihat)
Rhat <- L%*%L'+Psi
plot(vector(Rhat),vector(R),\
xlab:"fitted correlations",ylab:"correlation",\
title:paste(i,"factors"),show:F)
addlines(vector(0,1),vector(0,1),show:F)
out$criterion
showplot()
}

(1)       7.8664
```



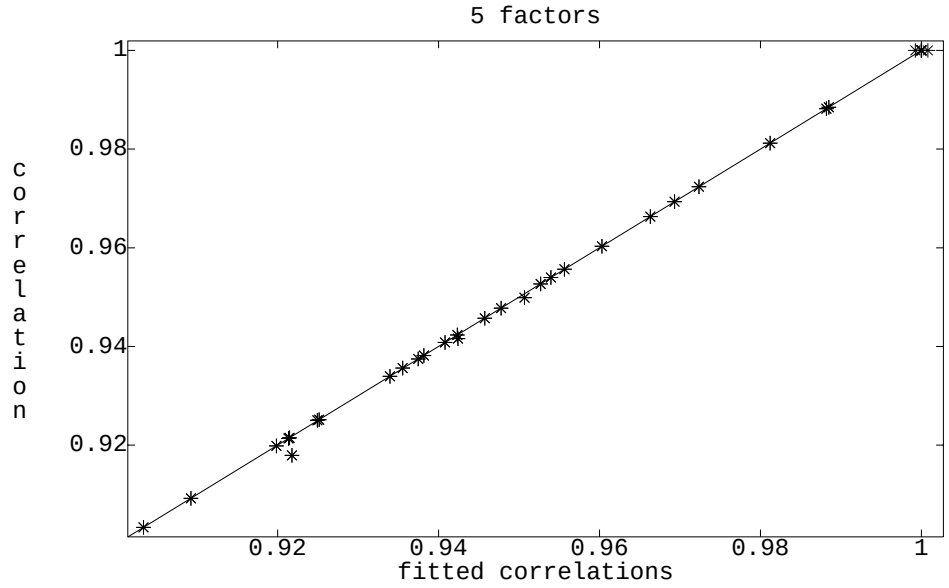

```
(1)       6.3239
```



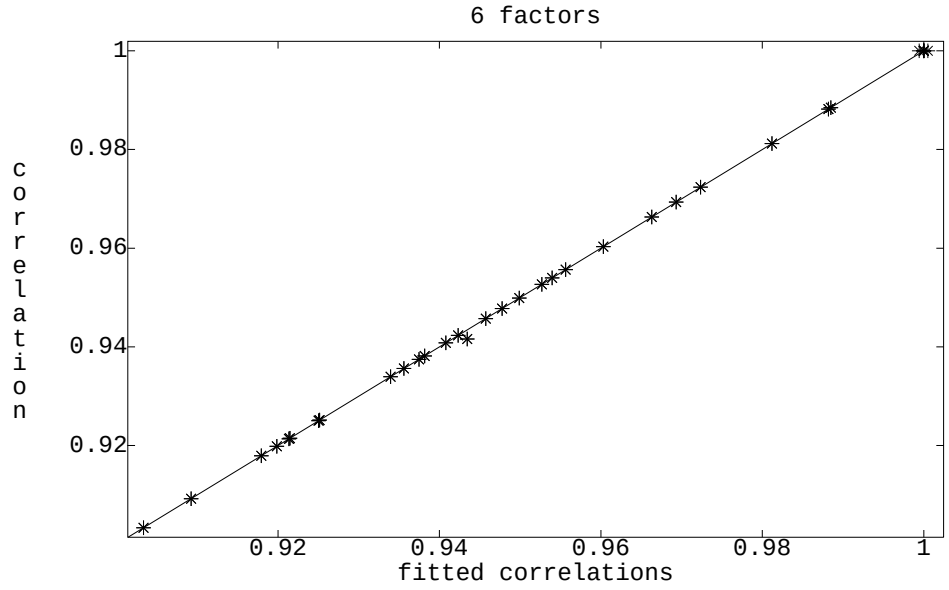
(1) 2.8523



(1) 0.61073




```
(1)      0.12113
```




```
Cmd> dim(aaup)
(1)         1074            8

Cmd> (1073-(2*8+5)/6-2*5/3)*.61073
(1)       651.14

Cmd> out <- facanal(R,5,quiet:T)
WARNING: With m = 5 and p = 8, m > (2*p + 1 - sqrt(8*p+1))/2

Cmd> print(out$psihat,format:"f6.3")
```

```
(1)  0.003  0.004  0.012  0.000
(5)  0.000  0.000  0.000  0.000

Cmd> print(out$loadings,format:"f6.3")
(1,1)  0.963 -0.083 -0.042 -0.247 -0.003
(2,1)  0.945 -0.070 -0.151 -0.113 -0.249
(3,1)  0.930 -0.085 -0.339 -0.020 -0.016
(4,1)  0.997 -0.080 -0.000  0.000 -0.000
(5,1)  0.968  0.056 -0.026 -0.244  0.000
(6,1)  0.950  0.111 -0.133 -0.097 -0.241
(7,1)  0.934  0.123 -0.336  0.000  0.000
(8,1)  0.997  0.074  0.000  0.000  0.000
```