A (crude) example of factor analysis

William Revelle
Northwestern University

Abstract

The following was done in class to show what happens if we iteratively do an eigen value decomposition of a correlation matrix.

Create the original correlation matrix using a factor model $R = FF' + U^2$ where we do this by creating the correlations, and then making the diagonal 1.

```
> f <- c(.9,.8,.7,.6,.5,.4)
> R <- f %*% t(f)
> R
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.81 0.72 0.63 0.54 0.45 0.36
[2,] 0.72 0.64 0.56 0.48 0.40 0.32
[3,] 0.63 0.56 0.49 0.42 0.35 0.28
[4,] 0.54 0.48 0.42 0.36 0.30 0.24
[5,] 0.45 0.40 0.35 0.30 0.25 0.20
[6,] 0.36 0.32 0.28 0.24 0.20 0.16
> diag(R) <- 1
> R
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.00 0.72 0.63 0.54 0.45 0.36
[2,] 0.72 1.00 0.56 0.48 0.40 0.32
[3,] 0.63 0.56 1.00 0.42 0.35 0.28
[4,] 0.54 0.48 0.42 1.00 0.30 0.24
[5,] 0.45 0.40 0.35 0.30 1.00 0.20
[6,] 0.36 0.32 0.28 0.24 0.20 1.00
> e <- eigen(R)
> e
$values
[1] 3.1647347 0.8216365 0.7185017 0.5920839 0.4410344 0.2620089

$vectors
           [,1]        [,2]        [,3]       [,4]        [,5]        [,6]
[1,] -0.4964630 -0.06108095  0.09230163  0.1390722  0.23845173  0.81552052
[2,] -0.4680485 -0.07428828  0.12095171  0.2141704  0.65661667 -0.53269894
[3,] -0.4326827 -0.09628978  0.18197078  0.5298520 -0.67508185 -0.18417914
[4,] -0.3899672 -0.14160502  0.41427181 -0.7780083 -0.20056923 -0.10357339
[5,] -0.3397764 -0.29920322 -0.86039187 -0.1967243 -0.10763437 -0.06685542
[6,] -0.2823444  0.93375805 -0.17844447 -0.1002466 -0.06668308 -0.04515620
```

```
> round(e)
Error in round(e) : non-numeric argument to mathematical function
> round(e$vectors,2)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,] -0.50 -0.06  0.09  0.14  0.24  0.82
[2,] -0.47 -0.07  0.12  0.21  0.66 -0.53
[3,] -0.43 -0.10  0.18  0.53 -0.68 -0.18
[4,] -0.39 -0.14  0.41 -0.78 -0.20 -0.10
[5,] -0.34 -0.30 -0.86 -0.20 -0.11 -0.07
[6,] -0.28  0.93 -0.18 -0.10 -0.07 -0.05
> C1 <- e$vectors[,1] %*%  sqrt(e$values[1] )
Error in e$vectors[, 1] %*% sqrt(e$values[1]) : non-conformable arguments
> C1 <- e$vectors[,1]* sqrt(e$values[1] )
> C1
[1] -0.8831928 -0.8326442 -0.7697296 -0.6937401 -0.6044520 -0.5022823
> p1 <- principal(R,1)
> p1
Principal Components Analysis
Call: principal(r = R, nfactors = 1)
Standardized loadings (pattern matrix) based upon correlation matrix
   PC1   h2   u2
1 0.88 0.78 0.22
2 0.83 0.69 0.31
3 0.77 0.59 0.41
4 0.69 0.48 0.52
5 0.60 0.37 0.63
6 0.50 0.25 0.75

                PC1
SS loadings    3.16
Proportion Var 0.53

Test of the hypothesis that 1 component is sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 9  and the objective function was  0.1

Fit based upon off diagonal values = 0.95
> round(C1,2)
[1] -0.88 -0.83 -0.77 -0.69 -0.60 -0.50
> R1 <- C1 %*% t(C1)
> R1
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.7800296 0.7353854 0.6798197 0.6127063 0.5338477 0.4436121
[2,] 0.7353854 0.6932964 0.6409109 0.5776386 0.5032935 0.4182224
[3,] 0.6798197 0.6409109 0.5924836 0.5339923 0.4652646 0.3866215
[4,] 0.6127063 0.5776386 0.5339923 0.4812753 0.4193326 0.3484533
[5,] 0.5338477 0.5032935 0.4652646 0.4193326 0.3653623 0.3036056
[6,] 0.4436121 0.4182224 0.3866215 0.3484533 0.3036056 0.2522875
> round(R1,2)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.78 0.74 0.68 0.61 0.53 0.44
[2,] 0.74 0.69 0.64 0.58 0.50 0.42
[3,] 0.68 0.64 0.59 0.53 0.47 0.39
[4,] 0.61 0.58 0.53 0.48 0.42 0.35
[5,] 0.53 0.50 0.47 0.42 0.37 0.30
[6,] 0.44 0.42 0.39 0.35 0.30 0.25
```

```
> newR <- R
> diag(newR) <- diag(R1)
> newR
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.7800296 0.7200000 0.6300000 0.5400000 0.4500000 0.3600000
[2,] 0.7200000 0.6932964 0.5600000 0.4800000 0.4000000 0.3200000
[3,] 0.6300000 0.5600000 0.5924836 0.4200000 0.3500000 0.2800000
[4,] 0.5400000 0.4800000 0.4200000 0.4812753 0.3000000 0.2400000
[5,] 0.4500000 0.4000000 0.3500000 0.3000000 0.3653623 0.2000000
[6,] 0.3600000 0.3200000 0.2800000 0.2400000 0.2000000 0.2522875
> round(newR,2)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.78 0.72 0.63 0.54 0.45 0.36
[2,] 0.72 0.69 0.56 0.48 0.40 0.32
[3,] 0.63 0.56 0.59 0.42 0.35 0.28
[4,] 0.54 0.48 0.42 0.48 0.30 0.24
[5,] 0.45 0.40 0.35 0.30 0.37 0.20
[6,] 0.36 0.32 0.28 0.24 0.20 0.25
> e1 <- eigen(newR)
> e1
$values
[1] 2.765663532 0.118512431 0.110547014 0.094664707 0.073358001 0.001988977

$vectors
           [,1]        [,2]        [,3]        [,4]       [,5]       [,6]
[1,] -0.5298385 -0.02217985  0.04302330  0.03711866  0.1593246  0.8307624
[2,] -0.4854257 -0.04488776  0.09386445  0.09940594  0.7294311 -0.4599839
[3,] -0.4325923 -0.15980480  0.58314005 -0.46019350 -0.4396253 -0.2054886
[4,] -0.3734284  0.79467345 -0.37567619 -0.11590107 -0.2290442 -0.1483862
[5,] -0.3104960 -0.58080509 -0.69749569 -0.12417692 -0.2177392 -0.1301044
[6,] -0.2462528 -0.05581334  0.14715061  0.86493332 -0.3865281 -0.1306809

> C2 <- e1$vectors[,1] * sqrt(e1$values[1])
> round(C2,2)
[1] -0.88 -0.81 -0.72 -0.62 -0.52 -0.41
> round(C1,2)
[1] -0.88 -0.83 -0.77 -0.69 -0.60 -0.50
> R2 <- C1 %*% t(C1)
> round(R2,2)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.78 0.74 0.68 0.61 0.53 0.44
[2,] 0.74 0.69 0.64 0.58 0.50 0.42
[3,] 0.68 0.64 0.59 0.53 0.47 0.39
[4,] 0.61 0.58 0.53 0.48 0.42 0.35
[5,] 0.53 0.50 0.47 0.42 0.37 0.30
[6,] 0.44 0.42 0.39 0.35 0.30 0.25
> R2 <- C2 %*% t(C2)
> round(R2,2)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.78 0.71 0.63 0.55 0.45 0.36
[2,] 0.71 0.65 0.58 0.50 0.42 0.33
[3,] 0.63 0.58 0.52 0.45 0.37 0.29
[4,] 0.55 0.50 0.45 0.39 0.32 0.25
[5,] 0.45 0.42 0.37 0.32 0.27 0.21
[6,] 0.36 0.33 0.29 0.25 0.21 0.17
> newR <- R
> diag(newR) <- diag(R2)
> newR
```

```
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.7764014 0.7200000 0.6300000 0.5400000 0.4500000 0.3600000
[2,] 0.7200000 0.6516958 0.5600000 0.4800000 0.4000000 0.3200000
[3,] 0.6300000 0.5600000 0.5175555 0.4200000 0.3500000 0.2800000
[4,] 0.5400000 0.4800000 0.4200000 0.3856683 0.3000000 0.2400000
[5,] 0.4500000 0.4000000 0.3500000 0.3000000 0.2666315 0.2000000
[6,] 0.3600000 0.3200000 0.2800000 0.2400000 0.2000000 0.1677111
> e2 <- eigen(newR)
> e2
$values
[1]  2.713326569  0.026560749  0.021000548  0.015032428  0.008352109 -0.018608872

$vectors
           [,1]        [,2]        [,3]        [,4]        [,5]       [,6]
[1,] -0.5393397  0.01532272  0.07401932 -0.04557508  0.03186623  0.8368432
[2,] -0.4874507  0.05512168  0.38607617 -0.59044290 -0.35538143 -0.3679400
[3,] -0.4290380 -0.72071688 -0.47952904  0.13765266 -0.05414361 -0.2113426
[4,] -0.3674886  0.68859386 -0.57720818  0.13892389 -0.05146676 -0.1888720
[5,] -0.3052143  0.05157614  0.51388843  0.77003221 -0.08970178 -0.1977544
[6,] -0.2433664  0.02173466  0.13515783 -0.13454508  0.92685588 -0.2118220

> C3 <- e2$vectors[,1] * sqrt(e2$values[1])
> round(C3,2)
[1] -0.89 -0.80 -0.71 -0.61 -0.50 -0.40
> R3 <- C3 %*% t(C3)
> round(R3,2)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.79 0.71 0.63 0.54 0.45 0.36
[2,] 0.71 0.64 0.57 0.49 0.40 0.32
[3,] 0.63 0.57 0.50 0.43 0.36 0.28
[4,] 0.54 0.49 0.43 0.37 0.30 0.24
[5,] 0.45 0.40 0.36 0.30 0.25 0.20
[6,] 0.36 0.32 0.28 0.24 0.20 0.16
> round(R3-R,2)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,] -0.21 -0.01  0.00  0.00  0.00  0.00
[2,] -0.01 -0.36  0.01  0.01  0.00  0.00
[3,]  0.00  0.01 -0.50  0.01  0.01  0.00
[4,]  0.00  0.01  0.01 -0.63  0.00  0.00
[5,]  0.00  0.00  0.01  0.00 -0.75  0.00
[6,]  0.00  0.00  0.00  0.00  0.00 -0.84
> newR <- R
> diag(newR) <- diag(R3)
> e3 <- eigen(newR)
> C4 <- e3$vectors[,1] * sqrt(e3$values[1])
> round(C4,2)
[1] -0.89 -0.80 -0.70 -0.60 -0.50 -0.40
> p1 <- principal(R,1)
> p1
Principal Components Analysis
Call: principal(r = R, nfactors = 1)
Standardized loadings (pattern matrix) based upon correlation matrix
   PC1   h2   u2
1 0.88 0.78 0.22
2 0.83 0.69 0.31
3 0.77 0.59 0.41
4 0.69 0.48 0.52
5 0.60 0.37 0.63
```

```
6 0.50 0.25 0.75

                PC1
SS loadings    3.16
Proportion Var 0.53

Test of the hypothesis that 1 component is sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 9  and the objective function was  0.1

Fit based upon off diagonal values = 0.95
> p2 <- principal(R,2)
> p2
Principal Components Analysis
Call: principal(r = R, nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation matrix
   RC1   RC2   h2    u2
1 0.84  0.27 0.78 0.217
2 0.80  0.24 0.70 0.302
3 0.75  0.20 0.60 0.400
4 0.69  0.13 0.50 0.502
5 0.66 -0.04 0.44 0.561
6 0.16  0.97 0.97 0.031

                       RC1  RC2
SS loadings           2.86 1.12
Proportion Var        0.48 0.19
Cumulative Var        0.48 0.66
Proportion Explained  0.72 0.28
Cumulative Proportion 0.72 1.00

Test of the hypothesis that 2 components are sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 4  and the objective function was  0.2

Fit based upon off diagonal values = 0.95
> p2 <- principal(R,2,rotate="none")
> p2
Principal Components Analysis
Call: principal(r = R, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
   PC1   PC2   h2    u2
1 0.88 -0.06 0.78 0.217
2 0.83 -0.07 0.70 0.302
3 0.77 -0.09 0.60 0.400
4 0.69 -0.13 0.50 0.502
5 0.60 -0.27 0.44 0.561
6 0.50  0.85 0.97 0.031

                       PC1  PC2
SS loadings           3.16 0.82
Proportion Var        0.53 0.14
Cumulative Var        0.53 0.66
Proportion Explained  0.79 0.21
Cumulative Proportion 0.79 1.00
```

```
Test of the hypothesis that 2 components are sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 4  and the objective function was  0.2

Fit based upon off diagonal values = 0.95
> p3 <- principal(R,3,rotate="none")
> p3
Principal Components Analysis
Call: principal(r = R, nfactors = 3, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
   PC1   PC2   PC3   h2     u2
1 0.88 -0.06 -0.08 0.79 0.2108
2 0.83 -0.07 -0.10 0.71 0.2917
3 0.77 -0.09 -0.15 0.62 0.3761
4 0.69 -0.13 -0.35 0.62 0.3789
5 0.60 -0.27  0.73 0.97 0.0292
6 0.50  0.85  0.15 0.99 0.0084

                       PC1  PC2  PC3
SS loadings           3.16 0.82 0.72
Proportion Var        0.53 0.14 0.12
Cumulative Var        0.53 0.66 0.78
Proportion Explained  0.67 0.17 0.15
Cumulative Proportion 0.67 0.85 1.00

Test of the hypothesis that 3 components are sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 0  and the objective function was  0.31

Fit based upon off diagonal values = 0.97
> f1 <- fa(R,1)
Loading required package: MASS
> f1
Factor Analysis using method =  minres
Call: fa(r = R, nfactors = 1)
Standardized loadings (pattern matrix) based upon correlation matrix
  MR1   h2   u2 com
1 0.9 0.81 0.19   1
2 0.8 0.64 0.36   1
3 0.7 0.49 0.51   1
4 0.6 0.36 0.64   1
5 0.5 0.25 0.75   1
6 0.4 0.16 0.84   1

                MR1
SS loadings    2.71
Proportion Var 0.45

Mean item complexity =  1
Test of the hypothesis that 1 factor is sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 9  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  0
```

```
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                MR1
Correlation of scores with factors             0.94
Multiple R square of scores with factors       0.89
Minimum correlation of possible factor scores  0.78
> f2 <- fa(R,2,rotate="none")
> f2
Factor Analysis using method =  minres
Call: fa(r = R, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
   MR1   MR2   h2    u2 com
1 0.98 -0.05 0.96 0.035 1.0
2 0.75  0.29 0.64 0.360 1.3
3 0.65  0.25 0.49 0.510 1.3
4 0.56  0.21 0.36 0.640 1.3
5 0.47  0.18 0.25 0.750 1.3
6 0.37  0.14 0.16 0.840 1.3

                       MR1  MR2
SS loadings           2.62 0.24
Proportion Var        0.44 0.04
Cumulative Var        0.44 0.48
Proportion Explained  0.91 0.09
Cumulative Proportion 0.91 1.00

Mean item complexity =  1.2
Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 4  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  0

Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                MR1   MR2
Correlation of scores with factors             0.98  0.59
Multiple R square of scores with factors       0.97  0.35
Minimum correlation of possible factor scores  0.94 -0.29
> f2 <- fa(R,2))
Error: unexpected ')' in "f2 <- fa(R,2))"
> f2 <- fa(R,2)
> f2
Factor Analysis using method =  minres
Call: fa(r = R, nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation matrix
   MR1   MR2   h2    u2 com
1 0.97 -0.19 0.96 0.035 1.1
2 0.77  0.18 0.64 0.360 1.1
3 0.68  0.15 0.49 0.510 1.1
4 0.58  0.13 0.36 0.640 1.1
5 0.48  0.11 0.25 0.750 1.1
6 0.39  0.09 0.16 0.840 1.1

                       MR1  MR2
```

```
SS loadings           2.73 0.14
Proportion Var        0.45 0.02
Cumulative Var        0.45 0.48
Proportion Explained  0.95 0.05
Cumulative Proportion 0.95 1.00

 With factor correlations of
     MR1  MR2
MR1 1.00 0.04
MR2 0.04 1.00

Mean item complexity =  1.1
Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are  15  and the objective function was  2.06
The degrees of freedom for the model are 4  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  0

Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                MR1   MR2
Correlation of scores with factors             0.98  0.60
Multiple R square of scores with factors       0.96  0.36
Minimum correlation of possible factor scores  0.91 -0.28
```