Principal Components Factor Analysis

To get the matrix of factor loadings $\Lambda$:

1. Perform principal components analysis using the correlation matrix (works because $A^{-1}=A^T$)

2. Multiply each column of the coefficient matrix by the corresponding standard deviation (so that $\text{Var}(F)=1$)

Why Not PC-FA?

PC-FA method ignores the error structure, this means:

- The communality estimates will be inflated
- The correlation structure is not accurately captured

You are not actually fitting the underlying model!
Principal Factor Factor Analysis
To get the matrix of factor loadings $\Lambda$:

- Adjust the correlation matrix to have only the communalities.
  
  A. Replace the ones on the diagonal with the highest correlation for each variable.
  
  or
  
  B. Replace the ones on the diagonal with the $R^2$ for predicting that variable from the others.

2. Perform principal components analysis using the reduced correlation matrix (only modeling the common part of the correlation, not the errors)

3. Multiply each column of the coefficient matrix by the corresponding standard deviation (so that $\text{Var}(F)=1$)

Can we do even better?
What “new information” do we have after performing principal factor factor analysis?
How Many Factors?

**Degrees of Freedom Limit:** Define $s$ as the difference between the number of unique values in the correlation matrix and the number of parameters in the factor analysis model:

$$s = \frac{1}{2}(q-k)^2 - \frac{1}{2}(q+k)$$

<table>
<thead>
<tr>
<th># Factors</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td># Variables Required</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

How Many Factors?

**Kaiser's Criterion:** Take as many factors as there are eigenvalues $> 1$.

Good for around 20-50 variables but tends to choose too few if there are fewer variables and too many if there are too many variables. Works better with larger communalities and sample sizes.
How Many Factors?

Scree Plot: Take the number of factors corresponding to the last eigenvalue before they start to level off.

Tends to keep more than Kaiser.

How Many Factors?

Fixed % of Variance

A priori number of factors

Model Fits

Several significant loadings for each factor.

Significant Loadings?

Judging loadings:
±0.3 Minimal
±0.4 More Important
±0.5 Practically Significant

Statistical Significance Rule of Thumb:

<table>
<thead>
<tr>
<th>n</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>600</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>0.722</td>
<td>0.512</td>
<td>0.384</td>
<td>0.298</td>
<td>0.210</td>
<td>0.162</td>
</tr>
</tbody>
</table>
Several?

Rule of Thumb:

A factor is reliable if it has
3 or more loadings of 0.8
4 or more of 0.6
10 or more of 0.4 if \( n \geq 150 \)
Fewer loadings require \( n \geq 300 \)

Rotations

Orthogonal rotations are those that keep the factors orthogonal (perpendicular)

Varimax – Maximizes the sum of the variances of the squared loadings within columns. This tends to force each variable to load highly on as few factors as possible.
> beardat<-
+ read.table("http://www.stat.sc.edu/~habing/courses/data/bears.txt",head=T)
> bears<-beardat[,3:7]
> source("http://www.stat.sc.edu/~habing/courses/530/fact.txt")

> summary(princomp(bears,cor=T))
Importance of components:

          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Standard deviation     2.0777385 0.61455477 0.38299289 0.30941739 0.25080408
Proportion of Variance 0.8633994 0.07553551 0.02933671 0.01914782 0.01258054
Cumulative Proportion  0.8633994 0.93893493 0.96827164 0.98741946 1.00000000

> print(loadings(princomp(bears,cor=T)),cutoff=0)
Loadings:

          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Head.L    0.455 -0.290 -0.515  0.531  0.403
Head.W    0.405  0.860 -0.254 -0.170  0.060
Neck.G    0.463  0.055  0.449  0.504 -0.572
Length    0.452 -0.394 -0.326 -0.583 -0.441
Chest.G   0.459 -0.139  0.602 -0.309  0.558

> fact(bears,method="pca",maxfactors=5)
$eigen.values
[1]  4.317  0.378  0.147  0.096  0.063

$method
[1] "principal components - no rotation"

$loadings

          Factor1 Factor2 Factor3 Factor4 Factor5
Head.L    0.945  0.178  0.197  -0.164  0.101
Head.W    0.842 -0.528  0.097  0.053  0.015
Neck.G    0.961 -0.034 -0.172  -0.156 -0.144
Length    0.939  0.242  0.125  0.180  0.111
Chest.G   0.954  0.086  0.231  0.096  0.140

$communalities

1       1       1       1       1

$importance

          Factor1 Factor2 Factor3 Factor4 Factor5
variance explained  4.317  0.378  0.147  0.096  0.063
percent explained    0.863  0.076  0.029  0.019  0.013

$residuals

       Min.  1st Qu.  Median    Mean  3rd Qu.     Max.
0       0        0         0       0        0
> fact(bears, method="pca", maxfactors=2)

$eigen.values
[1] 4.317 0.378 0.147 0.096 0.063

$method
[1] "principal components - no rotation"

$loadings
  Factor1  Factor2
Head.L  0.945  0.178
Head.W  0.842 -0.528
Neck.G  0.961 -0.034
Length  0.939  0.242
Chest.G 0.954  0.086

$communalities
  0.924  0.988  0.926  0.940  0.918

$importance
  variance explained  4.317  0.378
  percent explained   0.863  0.076

$residuals
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -0.047  -0.027  -0.020  -0.015   0.000   0.020
> fact(bears, method="pf", maxfactors=2)

$eigen.values
[1] 4.317 0.378 0.147 0.096 0.063

$method
[1] "principal factor - no rotation"

$loadings

           Factor1 Factor2
Head.L      0.937  0.151
Head.W      0.812 -0.367
Neck.G      0.955 -0.084
Length      0.931  0.215
Chest.G     0.948  0.037

$communalities

0.901   0.793   0.919   0.914   0.900

$importance

                        Factor1 Factor2
variance explained      4.215 0.212
percent explained       0.843 0.042

$residuals

          Min.  1st Qu. Median   Mean  3rd Qu.    Max.
1st Qu.  -0.024  -0.009  -0.002  -0.002   0.007  0.017  
Median   -0.002  -0.002  -0.002  -0.002   0.007  0.017  
Mean     -0.002  -0.002  -0.002  -0.002   0.007  0.017  
3rd Qu.  0.007   0.007   0.007   0.007  0.017  0.017  
Max.     0.017   0.017   0.017   0.017  0.017  0.017  

r-squared= 0.014

mean= -0.002  s.d.= 0.012
```r
> fact(bears, method="iter", maxfactors=2)

$eigen.values
[1]  4.317  0.378  0.147  0.096  0.063

$method
[1] "iterated principal factor - no rotation"

$loadings
     Factor1 Factor2
Head.L   0.927  -0.078
Head.W   0.784   0.204
Neck.G   0.968   0.202
Length   0.957  -0.330
Chest.G  0.941   0.035

$communalities
Head.L  0.865   0.656  0.978    1.026   0.887
Head.W  0.656   0.865  0.978     1.026   0.887
Neck.G  0.978   0.978  0.865    0.656  0.865
Length  1.026   1.026  1.026     0.865  0.865
Chest.G 0.887   0.887  0.887    0.887  0.887

$importance
          Factor1 Factor2
variance explained   4.213   0.199
percent explained    0.843   0.040

$residuals
     Min. 1st Qu. Median    Mean 3rd Qu.    Max.
-0.002  -0.001   0.000    0.000  0.001    0.002

Warning message:
Communality greater than one, 36 was last legal iteration.
```
> fact(bears, method="iter", maxfactors=2, niter=36)
$eigen.values
[1] 4.317 0.378 0.147 0.096 0.063

$method
[1] "iterated principal factor - no rotation"

$loadings
Factor1 Factor2
Head.L  0.929 -0.093
Head.W  0.785  0.205
Neck.G  0.968  0.199
Length  0.952 -0.307
Chest.G 0.942  0.026

$communalities
     0.871    0.658    0.977   1.000    0.887

$importance
variance explained  4.209  0.185
percent explained    0.842  0.037

$residuals
          Min.  1st Qu.   Median      Mean  3rd Qu.    Max.
-0.003    -0.001     0.000      0.000    0.001    0.003

Warning message:  
Convergence not achieved, difference of 0.00113 after 36 iterations.
> fact(bears, method="iter", maxfactors=1)
$eigen.values
[1] 4.317 0.378 0.147 0.096 0.063

$method
[1] "iterated principal factor - no rotation"

$loadings
   Factor1
Head.L 0.934
Head.W 0.780
Neck.G 0.962
Length 0.924
Chest.G 0.950

$communalities
     0.872    0.608    0.925    0.855    0.903

$importance
   Factor1
variance explained  4.163
percent explained    0.833

$residuals
         Min. 1st Qu.    Median      Mean 3rd Qu.    Max.  
-0.039     -0.018    -0.007       0.000    0.010    0.050

![Graphs showing variances, distribution of predictions, and residual correlations.](image-url)
> fact(bears, method="iter", rotation="varimax", maxfactors=2, niter=36)

$eigen.values
[1]  4.317  0.378  0.147  0.096  0.063

$method
[1] "iterated principal factor - varimax rotation"

$loadings

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head.L</td>
<td>0.610</td>
<td>0.706</td>
</tr>
<tr>
<td>Head.W</td>
<td>0.711</td>
<td>0.391</td>
</tr>
<tr>
<td>Neck.G</td>
<td>0.840</td>
<td>0.521</td>
</tr>
<tr>
<td>Length</td>
<td>0.480</td>
<td>0.877</td>
</tr>
<tr>
<td>Chest.G</td>
<td>0.702</td>
<td>0.628</td>
</tr>
</tbody>
</table>

$communalities

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.871</td>
<td>0.658</td>
<td>0.977</td>
<td>1.000</td>
<td>0.887</td>
</tr>
</tbody>
</table>

$importance

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
</tr>
</thead>
<tbody>
<tr>
<td>variance explained</td>
<td>2.307</td>
<td>2.087</td>
</tr>
<tr>
<td>percent explained</td>
<td>0.461</td>
<td>0.417</td>
</tr>
</tbody>
</table>

$residuals

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.003</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Warning message:
Convergence not achieved, difference of 0.00113 after 36 iterations.
Using the data on page 73 of text...

```r
> subject<-c(1,0.44,0.41,0.29,0.33,0.25,1,0.35,0.35,0.32,0.33,1,0.16,0.19,
+ 0.18,1,0.59,0.47,1,0.46,1)
> subnames<-c("French","English","History","Arithmetic","Algebra","Geometry")
> subcor<-matrix(0,ncol=6,nrow=6)
> subcor[lower.tri(subcor,diag=T)]<-subject
> subcor<-subcor+t(subcor)
> diag(subcor)<-rep(1,6)
> subcor

[1,] 1.00 0.44 0.41 0.29 0.33 0.25
[2,] 0.44 1.00 0.35 0.35 0.32 0.33
[3,] 0.41 0.35 1.00 0.16 0.19 0.18
[4,] 0.29 0.35 0.16 1.00 0.59 0.47
[5,] 0.33 0.32 0.19 0.59 1.00 0.46
[6,] 0.25 0.33 0.18 0.47 0.46 1.00

> fact(r=subcor)
$eigen.values
[1] 2.729 1.129 0.615 0.603 0.523 0.402
$method
[1] "iterated principal factor - no rotation"
$loadings
   Factor1 Factor2
[1,]  0.591  -0.376
[2,]  0.594  -0.234
[3,]  0.430  -0.412
[4,]  0.710   0.341
[5,]  0.696   0.271
[6,]  0.584   0.184
$communalities
[1] 0.490 0.408 0.355 0.621 0.558 0.375
$importance
   Factor1 Factor2
variance explained  2.216  0.590
percent explained    0.369  0.098
$residuals
     Min.  1st Qu.   Median     Mean  3rd Qu.    Max.
-0.030  -0.004    0.001     0.000   0.004  0.026
```
> fact(r=subcor, rotation="varimax")
$eigen.values
[1] 2.729 1.129 0.615 0.603 0.523 0.402

-method
[1] "iterated principal factor - varimax rotation"

-loadings
Factor1 Factor2
[1,] 0.231 0.661
[2,] 0.321 0.552
[3,] 0.082 0.590
[4,] 0.769 0.172
[5,] 0.714 0.218
[6,] 0.572 0.217

-communalities
[1] 0.490 0.408 0.355 0.621 0.558 0.375

-importance
variance explained 1.592 1.214
percent explained 0.265 0.202

-residuals
 Min. 1st Qu.  Median     Mean 3rd Qu.     Max.
-0.030 -0.004  0.001  0.000  0.004  0.026