Principal Components Example

Below is a simple example of a principal components analysis (PCA) to illustrate a few of the concepts. **I do not recommend principal components analysis** for several important reasons: 1) the method assumes components are uncorrelated, which is likely an unreasonable and unnecessary assumption, b) PCA assumes no measurement error in the measured variables and is not a true factor analysis method, and c) PCA has biased estimates of loadings for small sample sizes (Snook & Gorsuch, 1989). Moreover, my example below analyzes the correlation matrix to help illustrate the meaning of eigenvalues and exploratory factor analysis. I usually recommend the principal axis factoring method with an oblique rotation, such as promax, in which the factors can be correlated, if an exploratory factor analysis method is desired.¹ For additional information, see the “Principal Components Analysis” handout for this class as well as the handouts “A Quick Primer on Exploratory Factor Analysis” and “Exploratory Factor Analysis Example: SPSS and R” from my structural equation modeling class, [http://web.pdx.edu/~newsomj/semclass](http://web.pdx.edu/~newsomj/semclass). See also Preacher & MacCallum (2003) for a good primer on exploratory factor analysis considerations and differences between exploratory factor analysis approaches and PCA.

```
factor var=rfelpos rnotprdr ramable ramfailr rnumqal rnotworr
    /method=correlation
    /analysis=rfelpos rnotprdr ramable ramfailr rnumqal rnotworr
    /print=initial extraction rotation correlation sig
    /plot=eigen
    /criteria=factor(6).
```

Note that SPSS automatically uses the Kaiser-Guttman rule of selecting components with eigenvalues over 1.0, which is known to over- or under-estimated the number of components or factors (e.g., Cattell & Vogelmann, 1977; Gorsuch, 1983; Zwick & Velicer, 1982). Usually a scree plot is examined and then the number of factors are selected. I “extracted” all six possible components in this case using `/criteria=factor(6)` in order to get all of the six eigenvalues.

¹ I analyze the correlation matrix in this example for the sake of simplicity and for maximizing correspondence across packages, but the usual recommendation for some factor analyses is to use the covariance matrix so that variances are preserved (Cudeck, 1989; Morrison, 1967, p. 222). If the items have similar variances the two methods will lead to similar conclusions.
### Total Variance Explained

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sums of Squared Loadings</th>
<th>Rotation Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% of Variance</td>
<td>Cumulative %</td>
</tr>
<tr>
<td>1</td>
<td>2.344</td>
<td>39.072</td>
<td>39.072</td>
</tr>
<tr>
<td>2</td>
<td>1.021</td>
<td>17.015</td>
<td>56.088</td>
</tr>
<tr>
<td>3</td>
<td>0.948</td>
<td>15.807</td>
<td>71.895</td>
</tr>
<tr>
<td>4</td>
<td>0.858</td>
<td>14.307</td>
<td>86.201</td>
</tr>
<tr>
<td>5</td>
<td>0.465</td>
<td>7.755</td>
<td>93.956</td>
</tr>
<tr>
<td>6</td>
<td>0.363</td>
<td>6.044</td>
<td>100.000</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

### Scree Plot

![Scree Plot Image]

### Component Matrix

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>relpos</td>
<td>.454</td>
<td>.005</td>
<td>-.402</td>
<td>.792</td>
<td>.068</td>
<td>.022</td>
</tr>
<tr>
<td>molprdr</td>
<td>.844</td>
<td>.020</td>
<td>-.125</td>
<td>-.209</td>
<td>.040</td>
<td>-.476</td>
</tr>
<tr>
<td>ramable</td>
<td>.226</td>
<td>.718</td>
<td>.614</td>
<td>217</td>
<td>.096</td>
<td>.023</td>
</tr>
<tr>
<td>ramfair</td>
<td>.795</td>
<td>-.143</td>
<td>.218</td>
<td>.006</td>
<td>-.525</td>
<td>.158</td>
</tr>
<tr>
<td>rumqal</td>
<td>.605</td>
<td>.437</td>
<td>-.438</td>
<td>-.374</td>
<td>.151</td>
<td>.296</td>
</tr>
<tr>
<td>rootworr</td>
<td>.614</td>
<td>-.542</td>
<td>.394</td>
<td>-.017</td>
<td>.390</td>
<td>.149</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

* 6 components extracted.
> library(psych)
> library(stats)
> #perform pca analysis uses correlation matrix here just to illustrate eigens
> #this code gives the eigenvalues
> #eigen values are in standard deviation form so should be squared for eigen values
> pca1=princomp(mydata, cor = TRUE)
> summary(pca1)
> #prints square roots of eigen values
> values(pca1)

R

> library(psych)
> library(stats)
> #perform pca analysis uses correlation matrix here just to illustrate eigens
> #this code gives the eigenvalues
> #eigen values are in standard deviation form so should be squared for eigen values
> pca1=princomp(mydata, cor = TRUE)
> summary(pca1)
> #prints square roots of eigen values
> values(pca1)

6 variables and 118 observations.
> eigs = pca1$sdev^2
> eigs

2.3443376 1.0209221 0.9484331 0.8583942 0.4653001 0.3626130

> #produces scree plot
> plot(pca1)

> this set of commands replicates the loadings found with SPSS (rotated component matrix)
> #use Varimax (capital V) with eps=1e-7 for Kaiser normalization
> pca2=principal(mydata,nfactors = 6,covar = FALSE, rotate="Varimax",eps=1e-7)
> #prints loadings - cutoff needed so it will print all loadings
> loads = loadings(pca2, digits = 3)
> print(loads, cutoff = 0.0001)

Loadings:
    RC3  RC6  RC4  RC2  RC5  RC1
rfelpos  0.070  0.049  0.987  0.013  0.097  0.095

Rotated Component Matrix a

<table>
<thead>
<tr>
<th></th>
<th>Comp.1</th>
<th>Comp.2</th>
<th>Comp.3</th>
<th>Comp.4</th>
<th>Comp.5</th>
<th>Comp.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>rfelpos</td>
<td>0.070</td>
<td>0.049</td>
<td>0.987</td>
<td>0.013</td>
<td>0.097</td>
<td>0.095</td>
</tr>
<tr>
<td>motprdr</td>
<td>0.287</td>
<td>0.204</td>
<td>0.122</td>
<td>0.039</td>
<td>0.255</td>
<td>0.892</td>
</tr>
<tr>
<td>ramable</td>
<td>0.045</td>
<td>0.002</td>
<td>0.012</td>
<td>0.996</td>
<td>0.066</td>
<td>0.030</td>
</tr>
<tr>
<td>rainfail</td>
<td>0.125</td>
<td>0.238</td>
<td>0.115</td>
<td>0.082</td>
<td>0.924</td>
<td>0.231</td>
</tr>
<tr>
<td>munqal</td>
<td>0.964</td>
<td>0.003</td>
<td>0.074</td>
<td>0.050</td>
<td>0.111</td>
<td>0.227</td>
</tr>
<tr>
<td>motworr</td>
<td>0.004</td>
<td>0.962</td>
<td>0.051</td>
<td>0.000</td>
<td>0.213</td>
<td>0.166</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.
rnotprdr 0.287 0.204 0.122 0.039 0.255 0.892
ramable 0.045 0.002 0.012 0.996 0.066 0.030
ramfailr 0.125 0.238 0.115 0.082 0.924 0.231
rnumqal 0.964 0.003 0.074 0.050 0.111 0.227
rnotworr 0.004 0.962 0.051 0.000 0.212 0.166

<table>
<thead>
<tr>
<th></th>
<th>RC3</th>
<th>RC6</th>
<th>RC4</th>
<th>RC2</th>
<th>RC5</th>
<th>RC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS loadings</td>
<td>1.033</td>
<td>1.026</td>
<td>1.010</td>
<td>1.003</td>
<td>0.991</td>
<td>0.937</td>
</tr>
<tr>
<td>Proportion var</td>
<td>0.172</td>
<td>0.171</td>
<td>0.168</td>
<td>0.167</td>
<td>0.165</td>
<td>0.156</td>
</tr>
<tr>
<td>Cumulative var</td>
<td>0.172</td>
<td>0.343</td>
<td>0.512</td>
<td>0.679</td>
<td>0.844</td>
<td>1.000</td>
</tr>
</tbody>
</table>

References