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. 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).
```

Correlation Matrix

		rfelpos	rnotprdr	ramable	ramfailr	rnumqal	rnotworr
Correlation	rfelpos	1.000	.260	.037	.246	.174	.134
	rnotprdr	.260	1.000	.097	.543	.518	.406
	ramable	.037	.097	1.000	.158	.108	.022
	ramfailr	.246	.543	.158	1.000	.289	.470
	rnumqal	.174	.518	.108	.289	1.000	.071
	rnotworr	.134	.406	.022	.470	.071	1.000
Sig. (1-tailed)	rfelpos		.002	.344	.004	.030	.073
	rnotprdr	.002		.147	.000	.000	.000
	ramable	.344	.147		.044	.122	.407
	ramfailr	.004	.000	.044		.001	.000
	rnumqal	.030	.000	.122	.001		.222
	rnotworr	.073	.000	.407	.000	.222	

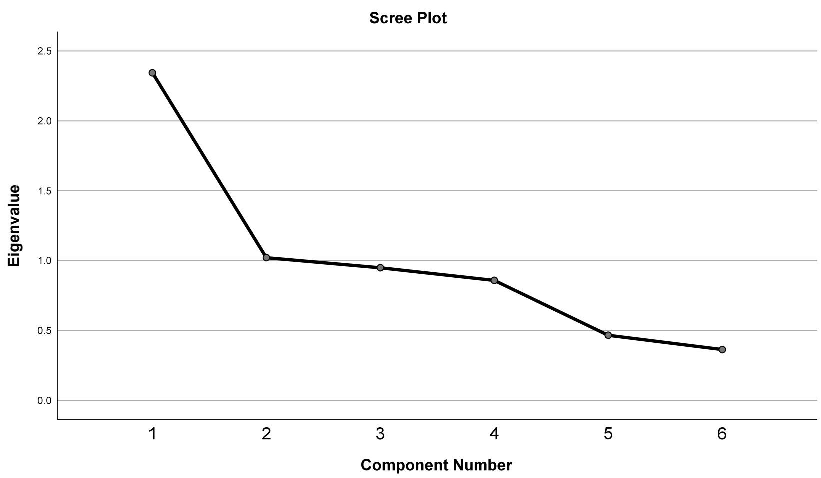
Note that SPSS automatically uses the Kaiser-Guttman rule of selecting components with eigenvalues over 1.0, which is known to over- or under-estimated the number of components or factors (e.g., Cattell & Vogelmann, 1977; Gorsuch, 1983; Zwick & Velicer, 1982). Usually a scree plot is examined and then 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

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.344	39.072	39.072	2.344	39.072	39.072	1.033	17.221	17.221
2	1.021	17.015	56.088	1.021	17.015	56.088	1.026	17.094	34.314
3	.948	15.807	71.895	.948	15.807	71.895	1.010	16.841	51.156
4	.858	14.307	86.201	.858	14.307	86.201	1.003	16.722	67.878
5	.465	7.755	93.956	.465	7.755	93.956	.991	16.513	84.390
6	.363	6.044	100.000	.363	6.044	100.000	.937	15.610	100.000

Extraction Method: Principal Component Analysis.



Component Matrix [a]

	Component					
	1	2	3	4	5	6
rfelpos	.454	.005	-.402	.792	.068	.022
rnotprdr	.844	.020	-.125	-.209	.040	-.476
ramable	.226	.718	.614	.217	.096	-.023
ramfailr	.795	-.143	.218	.006	-.525	.158
rnumqal	.605	.437	-.438	-.374	.151	.296
rnotworr	.614	-.542	.394	-.017	.390	.149

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

Rotated Component Matrix [a]

	Component					
	1	2	3	4	5	6
rfelpos	.070	.049	.987	.013	.097	.095
rnotprdr	.287	.204	.122	.039	.255	.892
ramable	.045	.002	.012	.996	.066	.030
ramfailr	.125	.238	.115	.082	.924	.231
rnumqal	.964	.003	.074	.050	.111	.227
rnotworr	.004	.962	.051	.000	.213	.166

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

a. Rotation converged in 5 iterations.

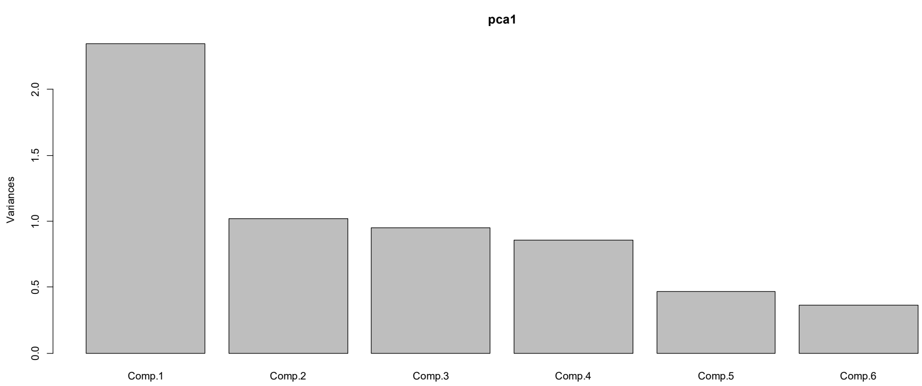
R

```
> 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)
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4     Comp.5     Comp.6
Standard deviation     1.5311230 1.0104069 0.9738753 0.9264957 0.68212908 0.60217352
Proportion of Variance 0.3907229 0.1701537 0.1580722 0.1430657 0.07755001 0.06043549
Cumulative Proportion  0.3907229 0.5608766 0.7189488 0.8620145 0.93956451 1.00000000
> #prints square roots of eigen values
> print(pca1)
Call:
princomp(x = mydata, cor = TRUE)

Standard deviations:
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6 
1.5311230 1.0104069 0.9738753 0.9264957 0.6821291 0.6021735 

 6  variables and  118 observations.
> eigens = pca1$sdev*pca1$sdev
> eigens
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6 
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
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

                 RC3   RC6   RC4   RC2   RC5   RC1
SS loadings    1.033 1.026 1.010 1.003 0.991 0.937
Proportion Var 0.172 0.171 0.168 0.167 0.165 0.156
Cumulative Var 0.172 0.343 0.512 0.679 0.844 1.000
```

References

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research*, *12*(3), 289-325.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*(2), 317.

Gorsuch, R. L. (1983). *Factor analysis (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Morrison, D.Â€F. (1976). *Multivariate statistical methods.* New York, NY: McGraw-Hill.

Preacher, K.J., & MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*, 13-43.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate behavioral research*, *17*(2), 253-269.