STAT 530/J530
September 29th, 2005

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Principal Components Factor Analysis

To get the matrix of factor loadings Λ:

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 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 Λ:

1. 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 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:

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

# Factors	2	3	4	5	6
# Variables Required	5	7	8	9	11

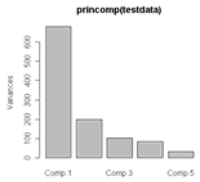
How Many Factors?

Kaiser's Crieterion: 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

Statisitcal Significance Rule of Thumb:

n	50	100	200	300	600	1000
λ	0.722	0.512	0.384	0.298	0.210	0.162

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 ≥ 300

Quote

"At the present time, factor analysis still maintains the flavor of an art, and no single strategy should yet be 'chiseled into stone'".

-Johnson and Wichern, 2002