Exploratory Factor Analysis

Today . . .

- The common factor model
- The communality problem
- Rotation

Factor analysis and principal components analysis are often used for the same purposes but they have different underlying models that sometimes make one more appropriate than the other for certain statistical goals.

- Principal components analysis seeks linear combinations that best capture the variation in the original variables.
- Factor analysis seeks linear combinations that best capture the correlations among the original variables.

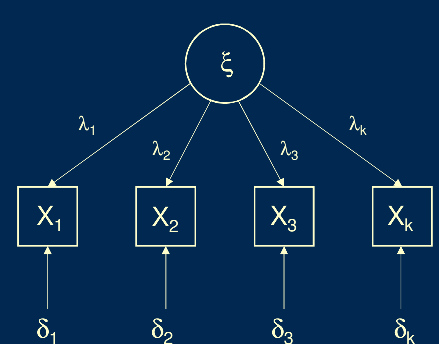
In factor analysis, the observed variance in each measure is assumed to be attributable to:

- Common factors that influence more than one measure
- A specific factor that is idiosyncratic to each measure.

The specific factor explicitly acknowledges that measures are faulty and have a part of their variance that is unrelated to other measures (and so, effectively, random).

In factor analysis we assume an explicit measurement model that specifies the causes of variation in observed measurements.

- Some causes are unobservable (latent) variables that affect more than one measure.
- Other causes are unobservable (latent) variables that are specific (unique) to each measure.
- The specific factors are assumed to be uncorrelated with each other, and, uncorrelated with the common factors.



$X_{i,1} = \lambda_1 \xi_i + \delta_{i,1}$

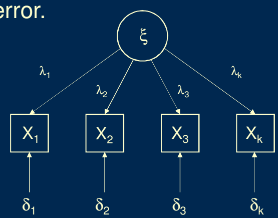
$X_{i,2} = \lambda_2 \xi_i + \delta_{i,2}$

$X_{i,3} = \lambda_3 \xi_i + \delta_{i,3}$

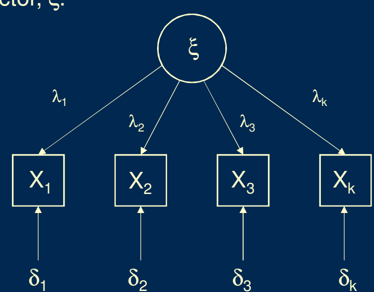
$X_{i,k} = \lambda_k \xi_i + \delta_{i,k}$

In a simple one-factor model, each measure is assumed to be a simple linear combination of the common factor and a specific factor. The specific factor is assumed to include both unique systematic influences and random error.

For each X, λ represents the extent to which each measure reflects the underlying common factor, ξ.



This model assumes that all covariation among the observed variables, X, is due to the underlying common factor, ξ.



In the simple one-factor model, the variance in a measure is assumed to be captured completely by variance due to the common factor and variance due to the specific factor. The two sources are assumed to be uncorrelated and so their variances are additive.

$$\sigma^2_{X_k} = \sigma^2_{\lambda_k \xi + \delta_k} = \lambda_k^2 + \sigma^2_{\delta_k}$$

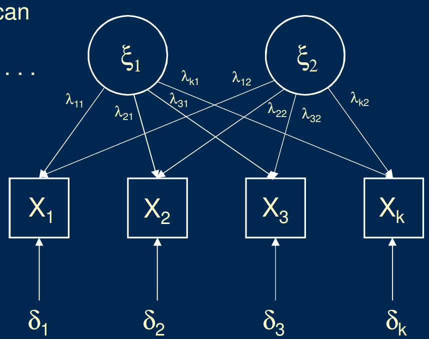
If the variables are standardized, then λ is a correlation coefficient and λ^2 is the proportion of variance in X due to the common factor. This is the communality of X. The remaining variance is assumed to be due to random sources.

The communality in X can be generally defined as the proportion of variance in X that is due to the common factors, no matter how many their number. This means that the communality can also be defined as:

$$1 - \theta_{kk}^2$$

where θ^2 is used to indicate the specific factor variance.

Measurement models can be more complex . . .



Adding more common factors requires expanding the linear combinations for X to accommodate the additional common sources of variance.

$$X_{i,1} = \lambda_{1,1}\xi_{i,1} + \lambda_{1,2}\xi_{i,2} + \delta_{i,1}$$

$$X_{i,2} = \lambda_{2,1}\xi_{i,1} + \lambda_{2,2}\xi_{i,2} + \delta_{i,2}$$

$$X_{i,3} = \lambda_{3,1}\xi_{i,1} + \lambda_{3,2}\xi_{i,2} + \delta_{i,3}$$

$$X_{i,k} = \lambda_{k,1}\xi_{i,1} + \lambda_{k,2}\xi_{i,2} + \delta_{i,k}$$

Provided the additional common factors are assumed to be independent, each λ is still a correlation coefficient and the variance of X can be viewed as the sum of independent proportions of variance:

$$\sigma^2_{X_k} = \sum_{j=1}^{f} \lambda^2_{k,j} + \theta^2_k = 1$$

and the communality is still:

$$h^2_{kk} = 1 - \theta^2_{kk}$$

Because factor analysis only seeks to identify the common factors that influence the correlations among measures, it is not a correlation matrix that is analyzed.

The correlation matrix contains ones on the main diagonal, implying an attempt to account for all of the variance in X—the goal of principal components analysis.

Instead, in factor analysis the main diagonal is replaced by the communalities—the variances in X that are due only to the common factors.

But, the communalities are not known in advance of the factor analysis, giving rise to the communality problem and the need to solve for the common factors iteratively.

The analysis begins with initial and perhaps crude estimates of the communalities on the main diagonal of the correlation matrix. These are used to derive initial and perhaps crude estimates of the common factors. The initial estimates of the common factors are used to generate improved estimates of the communalities, which can then be substituted into the main diagonal of the correlation matrix.

The process of substituting better communality estimates to derive better approximations to the common factors continues until little change occurs from one iteration to the next.

The procedure converges on the best estimates for the common factors and the communalities.

One potentially good starting value for the communality of any given measure is the squared multiple correlation of that measure with all of the other measures in the **X** matrix.

These are lower bound estimates of the true communality because each measure is assumed to have an error component that attenuates the relations between measures (which are presumed to reflect the common factors).

In factor analysis, the linear combinations merely reflect the minimum number of common variance sources necessary to capture the correlations among the measures.

Their “location” or orientation or meaning relative to the original reference system (the variables) has no particular privileged status. This creates what is known as the rotational indeterminacy of common factors and motivates the search for orientations or rotations that have some optimal “meaning”.

The solution to the common factor model begins in same way that principal components analysis began:

$$X = Z_s D^{\frac{1}{2}} U'$$

The correlation matrix among the measures in X can be defined in matrix form as:

$$R = \frac{1}{N-1} X'X$$

We can substitute in the definition of X in terms of its component matrices:

$$R = \frac{1}{N-1} (Z_s D^{\frac{1}{2}} U')' (Z_s D^{\frac{1}{2}} U')$$

$$R = \frac{1}{N-1} (Z_s D^{\frac{1}{2}} U')' (Z_s D^{\frac{1}{2}} U')$$

We can rearrange terms because (AB)` = (B`A`) and $(D^{½})$` = $D^{½}$. We can further simplify because the correlation matrix for Z_s is an identity matrix:

$$R = U D^{\frac{1}{2}} D^{\frac{1}{2}} U'$$

This form is a reminder that the correlation matrix among the original measures, which is a variance-covariance matrix for standardized scores, can be obtained by applying the weights (U = eigenvectors) for creating the linear combinations to the variance-covariance matrix for those linear combinations (D, with eigenvalues on the diagonal).

$$R = UDU'$$

Another way to represent R is in terms of the factor loadings, which are just rescaled eigenvectors:

$$R = UD^{\frac{1}{2}}(UD^{\frac{1}{2}})'$$

$$R = FF'$$

Factor analysis also tries to approximate the correlation matrix, **R**, but does not attempt to account for all variability in **X**. Instead, only the common factor variance is of interest. This means that rather than substituting $\mathbf{Z_s D^{1/2} U}$` for **X** into the formula for **R**, we must instead substitute **X** as defined by the common factor model.

$$X_{i,1} = \lambda_{1,1}\xi_{i,1} + \lambda_{1,2}\xi_{i,2} + \ldots + \lambda_{1,f}\xi_{i,f} + \delta_{i,1}$$

$$X_{i,2} = \lambda_{2,1}\xi_{i,1} + \lambda_{2,2}\xi_{i,2} + \ldots + \lambda_{2,f}\xi_{i,f} + \delta_{i,2}$$

$$X_{i,3} = \lambda_{3,1}\xi_{i,1} + \lambda_{3,2}\xi_{i,2} + \ldots + \lambda_{3,f}\xi_{i,f} + \delta_{i,3}$$

$$X_{i,k} = \lambda_{k,1}\xi_{i,1} + \lambda_{k,2}\xi_{i,2} + \ldots + \lambda_{k,f}\xi_{i,f} + \delta_{i,k}$$

In matrix form, the common factor model can be represented as:

$$\mathbf{X} = \Xi \Lambda' + \Delta$$

An N x V matrix of standardized scores on the original variables.

An N x F matrix of common factor scores.

An F x V matrix of factor loadings.

An N x V matrix of specific factor scores.

In the common factor model, we make the following assumptions:

$$\frac{1}{N-1}\Xi'\Xi = I$$

The common factors are uncorrelated.

$$\Theta = \frac{1}{N-1}\Delta'\Delta$$

The specific factors have a diagonal covariance matrix.

$$\Xi'\Delta = 0$$

The common factors and specific factors are uncorrelated.

Substituting the common factor definitions for X:

$$R = \frac{1}{N-1}(\Xi\Lambda' + \Delta)'(\Xi\Lambda' + \Delta)$$

$$R = \boxed{\frac{1}{N-1}}(\Lambda\boxed{\Xi'\Xi}\Lambda' + \boxed{\Delta'\Xi\Lambda'} + \boxed{\Lambda\Xi'\Delta} + \Delta'\Delta)$$

An identity matrix

Expected to be zero

$$R = \Lambda\Lambda' + \Theta$$

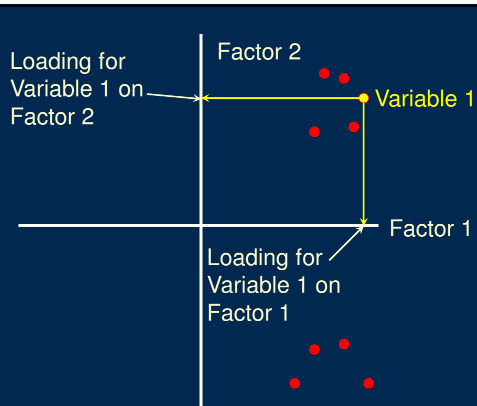
Expected to be zero

$$R - \Theta = \Lambda\Lambda'$$

The matrix that is reproduced by factor analysis is the correlation matrix less the specific factor variances.

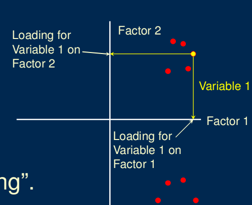
The factor loadings (Λ) can be thought of as the coordinates for the variables in multidimensional space with the factors as the reference axes. By noting which variables load highly on particular factors, we can discern the meaning of those factors.

The original orientation of the reference system, however, may not be optimal from an interpretation standpoint. The factors can be "rotated" to a new position that might make interpretation easier.



In factor analysis, the linear combinations merely reflect the minimum number of common variance sources necessary to capture the correlations among the measures. Their "location" or orientation relative to the original reference system has no particular privileged status.

This creates what is known as the *rotational indeterminacy* of common factors and motivates the search for orientations that have some optimal "meaning".



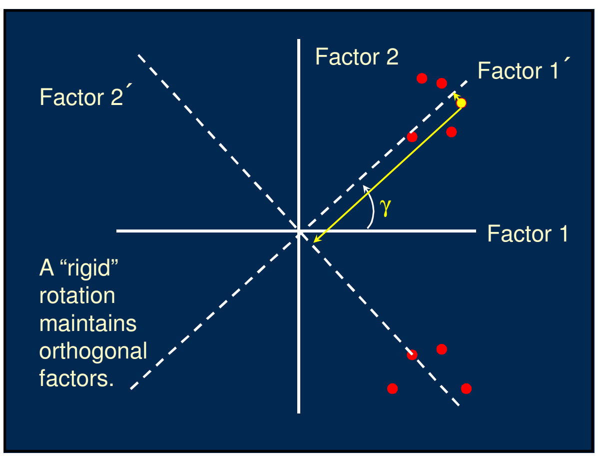
The purpose of factor rotation is to provide an orientation that achieves an easier interpretation of the underlying common factors.

This is often referred to as the search for simple structure—reflecting the fact that the factor loadings (in the factor structure matrix) are used to infer the meaning of the factors.

The loadings provide the coordinates for the location of the variables in a reference system defined by the factors. Rotation shifts the reference system so that the pattern of loadings may enhance interpretation.

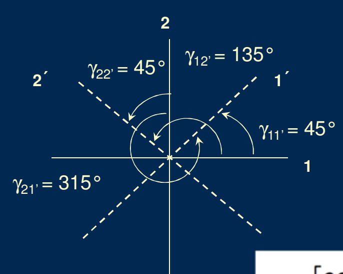
Simple structure occurs when:

- Most of the loadings on any given factor are small and a few loadings are large in absolute value
- Most of the loadings for any given variable are small, with ideally only one loading being large in absolute value.
- Any pair of factors have dissimilar patterns of loadings.



$$\Lambda^* = \Lambda T$$

T is the transformation matrix of cosines that indicate the location of the new axes compared to the original axes.



$$T = \begin{bmatrix} cos(\gamma_{11'}) & cos(\gamma_{12'}) \\ cos(\gamma_{21'}) & cos(\gamma_{22'}) \end{bmatrix} = \begin{bmatrix} .7071 & -.7071 \\ .7071 & .7071 \end{bmatrix}$$

Simple structure occurs when:

- Most of the loadings on any given factor are small and a few loadings are large in absolute value
- Most of the loadings for any given variable are small, with ideally only one loading being large in absolute value.
- Any pair of factors have dissimilar patterns of loadings.

The location of the new axes is determined by simple structure, which ideally might look like this for the factor loadings:

	F_1	F_2	F_3
X_1	1	0	0
X_2	1	0	0
X_3	1	0	0
X_4	0	1	0
X_5	0	1	0
X_6	0	1	0
X_7	0	0	1
X_8	0	0	1
X_9	0	0	1

One way to approach this ideal pattern is to find the rotation that maximizes the variance of the loadings in the columns of the factor structure matrix. This approach was originally suggested by Kaiser and is called *varimax* rotation.

	F_1	F_2	F_3
X_1	1	0	0
X_2	1	0	0
X_3	1	0	0
X_4	0	1	0
X_5	0	1	0
X_6	0	1	0
X_7	0	0	1
X_8	0	0	1
X_9	0	0	1

A second way to approach this ideal pattern is to find the rotation that maximizes the variance of the loadings in the rows of the factor structure matrix. This approach is called *quartimax* rotation.

	F_1	F_2	F_3
X_1	1	0	0
X_2	1	0	0
X_3	1	0	0
X_4	0	1	0
X_5	0	1	0
X_6	0	1	0
X_7	0	0	1
X_8	0	0	1
X_9	0	0	1

For those who want the best of both worlds, *equimax* rotation attempts to satisfy both goals. Varimax is the most commonly used and the three rarely produce results that are very discrepant.

	F_1	F_2	F_3
X_1	1	0	0
X_2	1	0	0
X_3	1	0	0
X_4	0	1	0
X_5	0	1	0
X_6	0	1	0
X_7	0	0	1
X_8	0	0	1
X_9	0	0	1

Next time . . .

- Example
- Orthogonal versus oblique rotation