Principal Component Analysis & Factor Analysis

Psych 818
DeShon

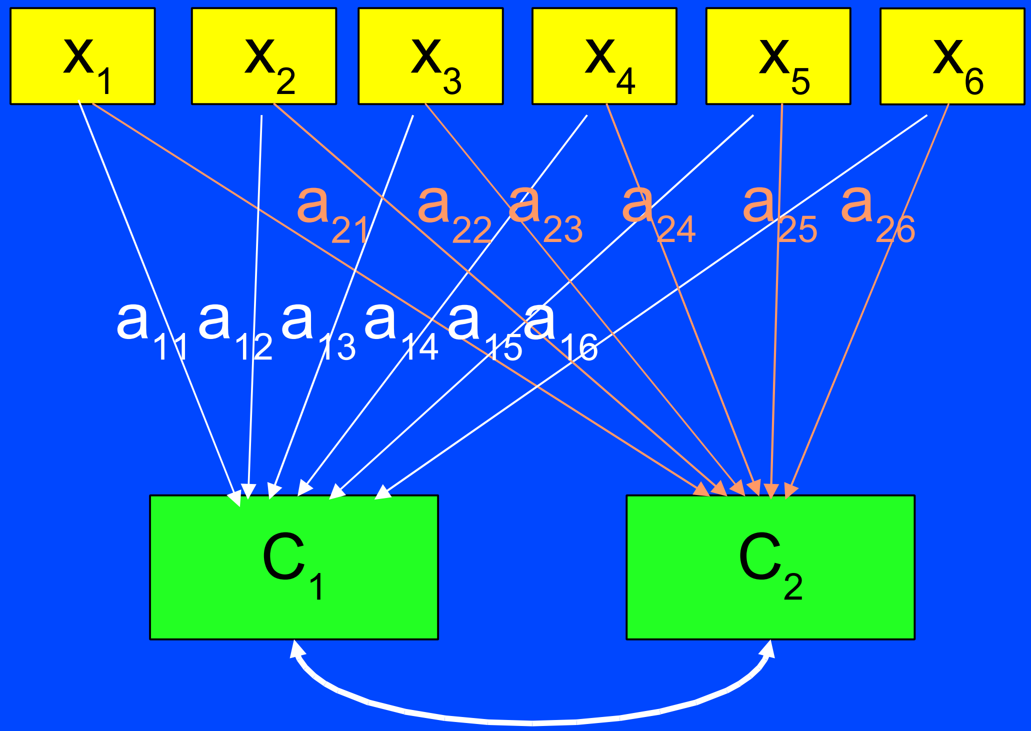
Purpose

- Both are used to reduce the dimensionality of correlated measurements
  - Can be used in a purely exploratory fashion to investigate dimensionality
  - Or, can be used in a quasi-confirmatory fashion to investigate whether the empirical dimensionality is consistent with the expected or theoretical dimensionality
- Conceptually, very different analyses
- Mathematically, there is substantial overlap

Principal Component Analysis

- Principal component analysis is conceptually and mathematically less complex
  - So, start here...
- First rule...
  - Don't interpret components as factors or latent variables.
  - Components are simply weighted composite variables
  - They should be interpreted and called components or composites

Principal Component Analysis



r = 0.0?

Principal Component Analysis

- Key Questions
  - How do you determine the weights?
  - How many composites do you need to reasonably reproduce the observed correlations among the measured variables?
    - Fewer is better!
  - Can the resulting components be transformed/rotated to yield more interpretable components?
  - How do you compute a person's score on the composite variable?

Conceptually...

From k original variables: $x_1, x_2, \ldots, x_k$:

Produce k new variables: $C_1, C_2, \ldots, C_k$:

$$C_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k$$

$$C_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2k}x_k$$

$$\ldots$$

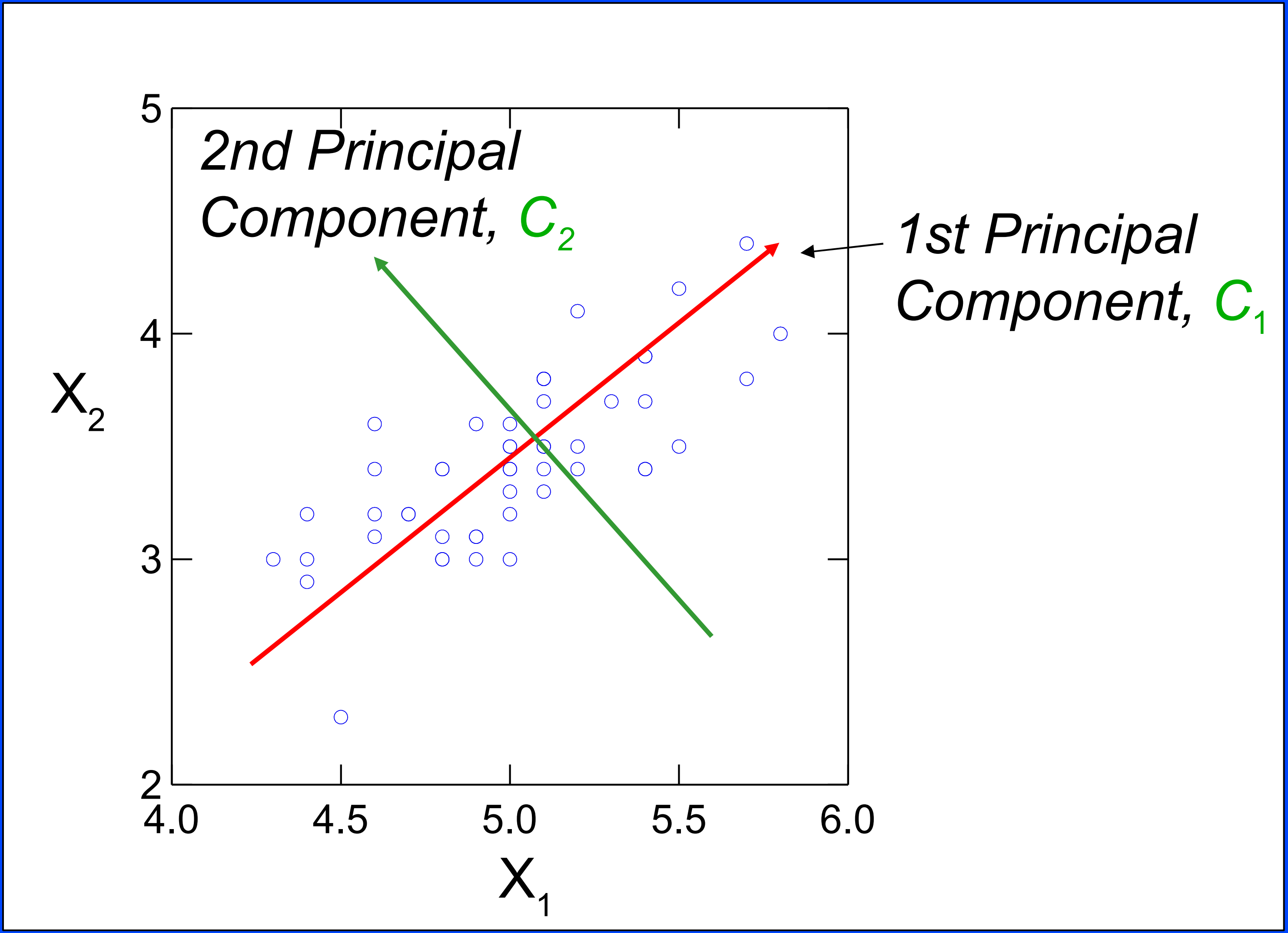
$$C_k = a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kk}x_k$$

Notice that there are as many components as there are original variables

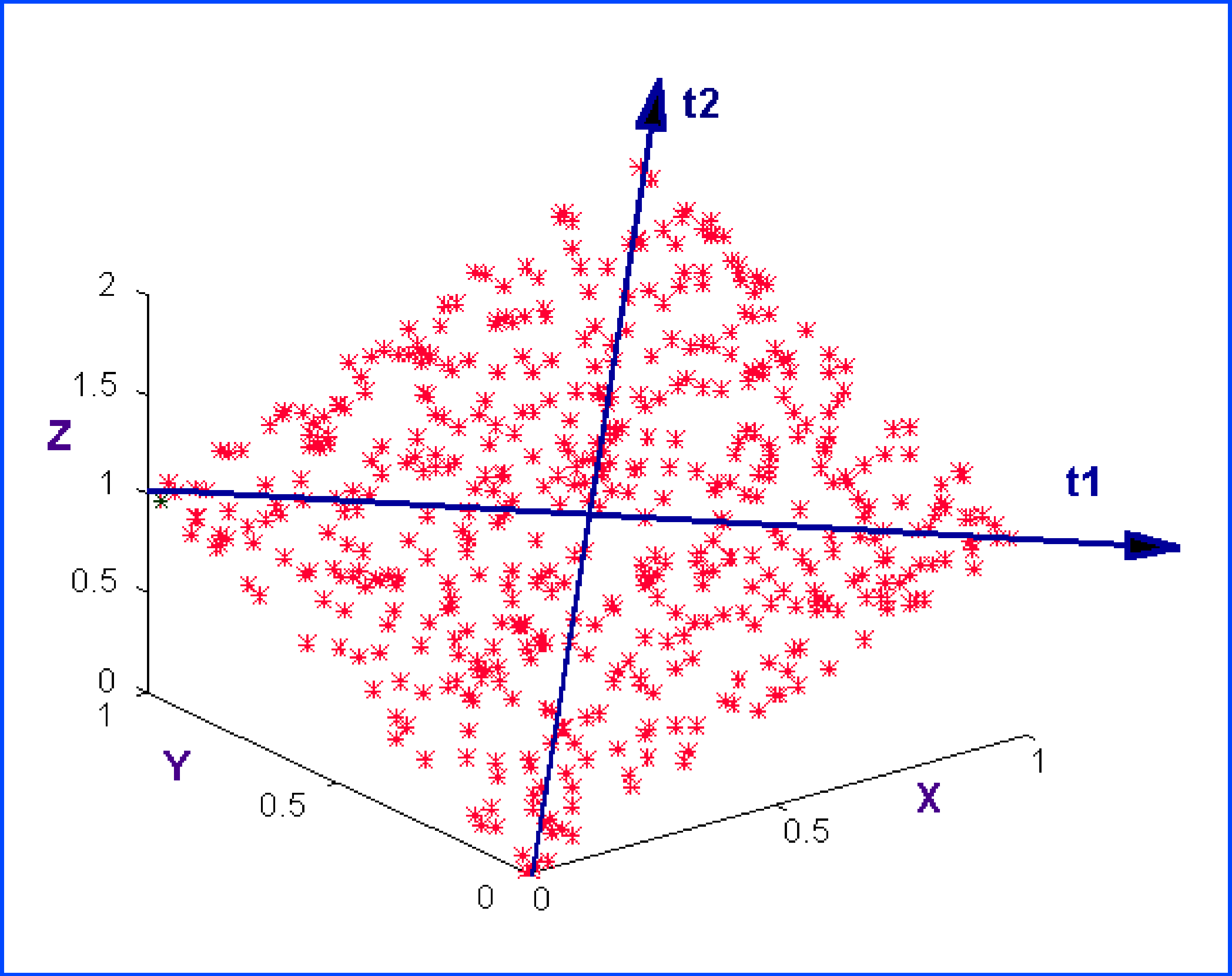
Conceptually...

- Find the weights such that
  - Composite variables are orthogonal/uncorrelated
  - C_1 explains as much variance as possible
    - maximum variance criterion
  - C_2 explains as much of the remaining variance as possible
  - etc...

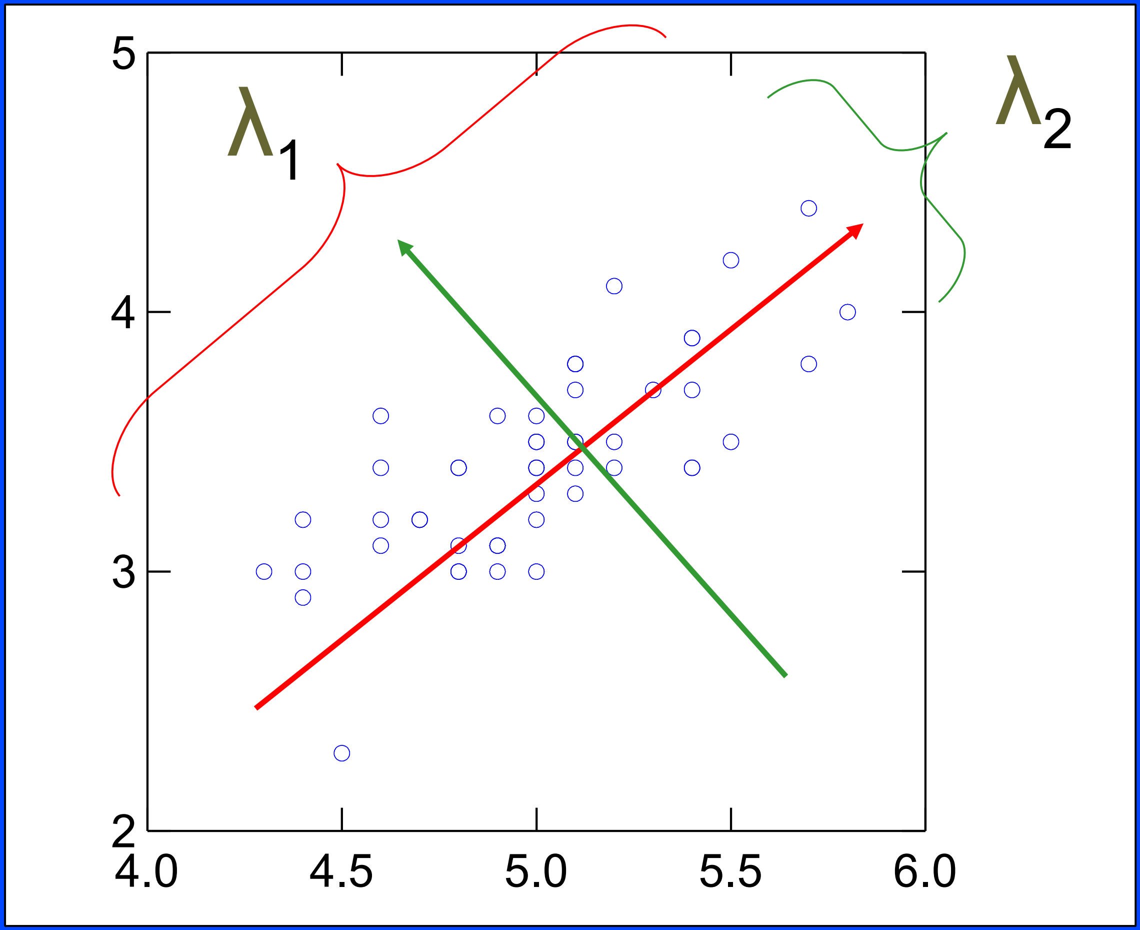
Conceptually...



In 3 dimensions...



Conceptually...



The variance of the resulting composite variables equal to the eigenvalues associated with the correlation or covariance matrix

Eigen..what?

- Determining the weights that maximize the variance of the components turns out to be a special case of Matrix Eigenvalues and Eigenvectors

$$Criterion: \Sigma\ s_{ij} a_i a_j = a^T S\ a \rightarrow max$$

- Problem:
  - Can make this quantity as large as desired by simply increasing the weights to infinity
  - So, put a constraint on the weights...

$$Constraint: \Sigma\ a_i a_j = a^T\ a = 1$$

Eigen..what?

- Lagrange Multipliers (λ) are frequently used when maximizing functions subject to constraints.

$$\phi_1 = a_1^T S a_i - \lambda_1 (a_1^T a_1 - 1)$$

- The partial derivative (used to find the maximum) is:

$$\frac{\partial \phi_1}{\partial a_1} 2 S a_1 - 2 \lambda_1 a_1$$

Eigen..what?

$$\phi_1 = a_1^T S a_i - \lambda_1 (a_1^T a_1 - 1)$$

- Set equal to zero, divide out the constant, and factor yields:

$$(S - \lambda_1 I) a_1 = 0$$

- So, we need to solve for both the eigenvalue (λ) and the weights (eigenvector)

Eigen..what?

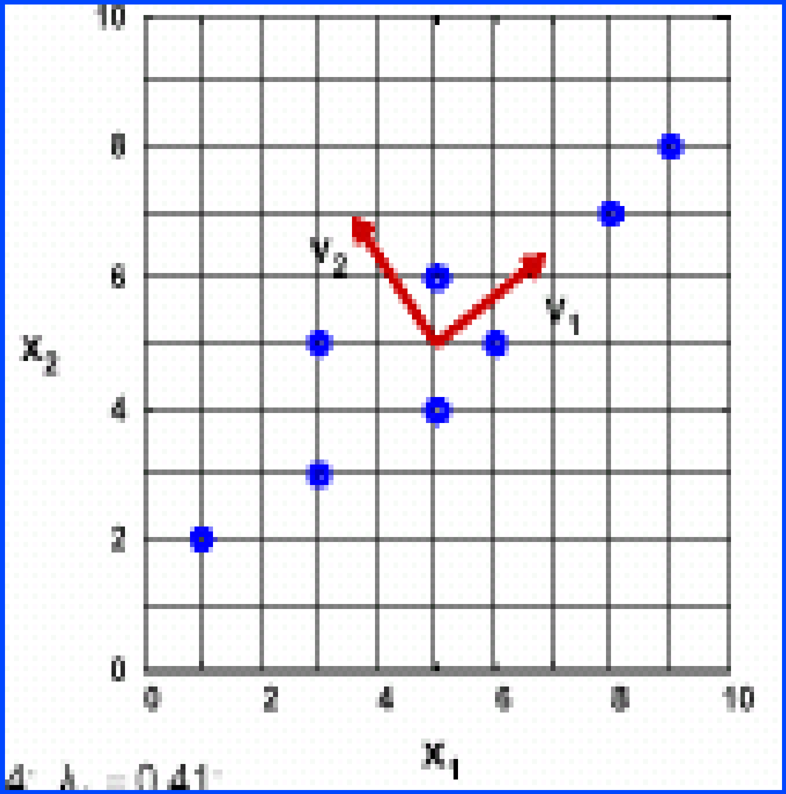
- Solving for the eigenvalues
  - Characteristic Equation...solving for the determinant

$$
|\mathbf{R} - \lambda \mathbf{I}| = \begin{vmatrix} r_{11} - \lambda & & \cdots & r_{1p} \\ r_{21} & r_{22} - \lambda & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} - \lambda \end{vmatrix} = 0
$$

- Once you have the eigenvalues, plug them back into the equation to solve for the eigenvectors

Example (by Hand...)

X1	X2
1	2
3	3
3	5
5	4
5	6
6	5
8	7
9	8

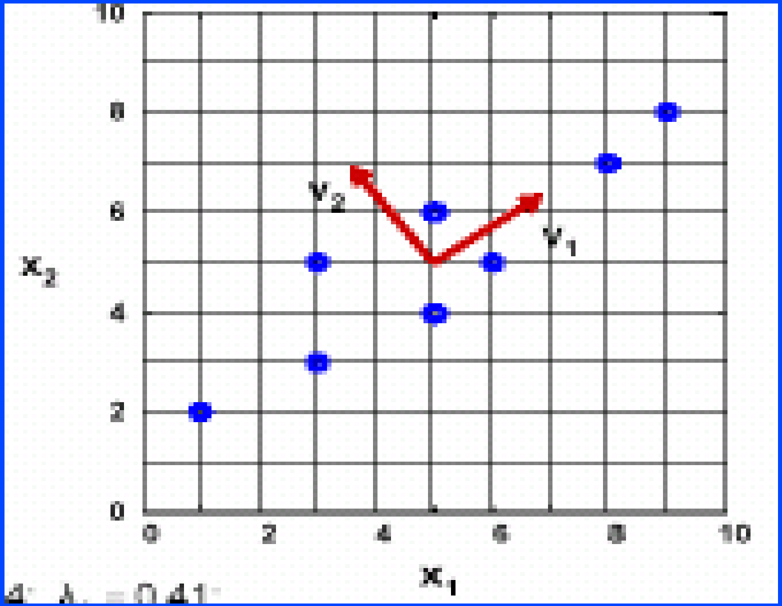


$$S = \begin{bmatrix} 6.25 & 4.25 \\ 4.25 & 3.5 \end{bmatrix}$$

Eigenvalues

$$Sa = \lambda a \rightarrow (S - \lambda I)\, a = 0 \rightarrow \begin{vmatrix} 6.25-\lambda & 4.25 \\ 4.25 & 3.5-\lambda \end{vmatrix} = 0$$

Example (by Hand...)



$$S = \begin{bmatrix} 6.25 & 4.25 \\ 4.25 & 3.5 \end{bmatrix}$$

$$\begin{vmatrix} 6.25-\lambda & 4.25 \\ 4.25 & 3.5-\lambda \end{vmatrix} = 0\,; \rightarrow \begin{matrix} \lambda_1 = 9.34 \\ \lambda_2 = 0.41 \end{matrix}$$

Eigenvectors

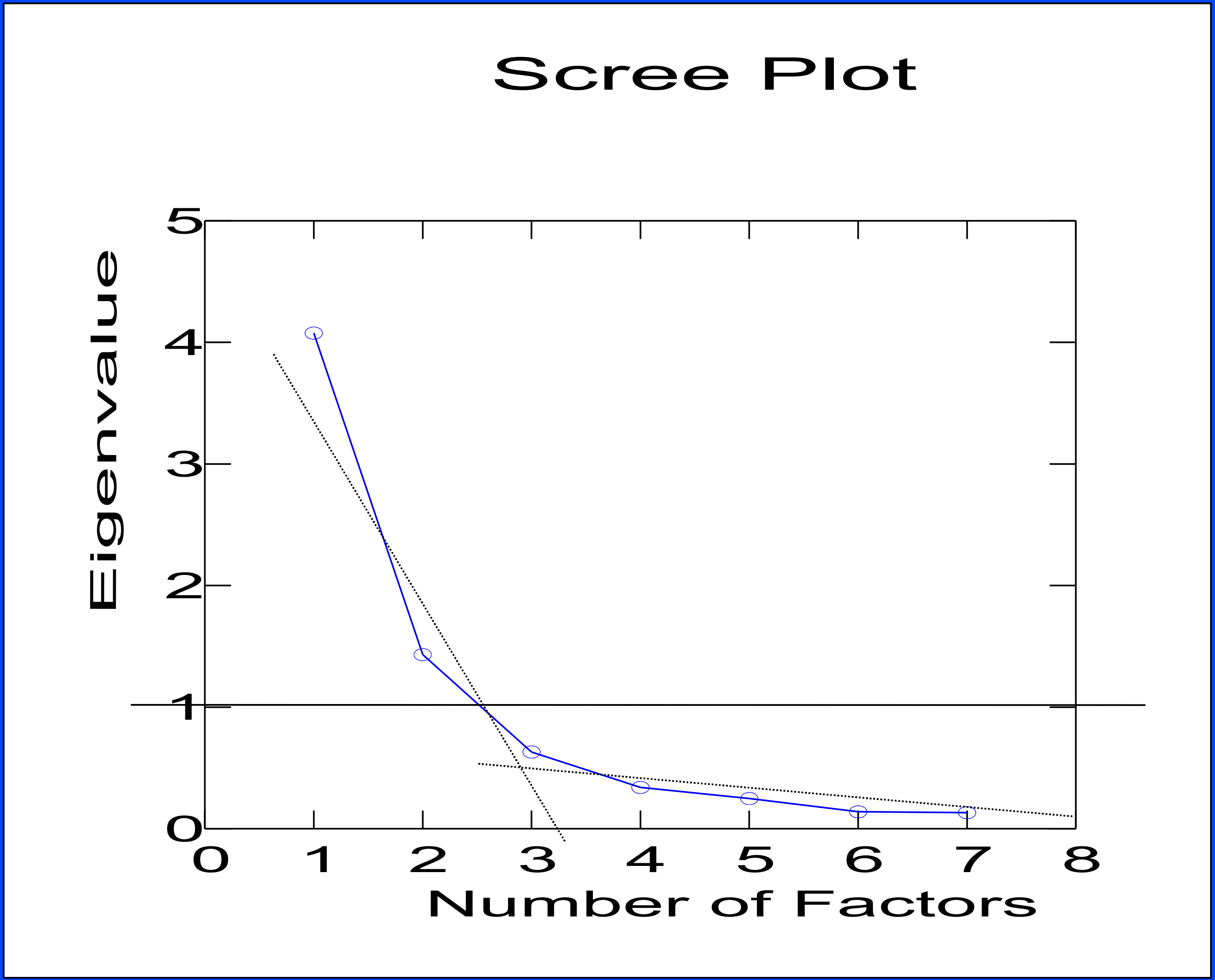
$$\begin{bmatrix} 6.25 & 4.25 \\ 4.25 & 3.5 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 a_{11} \\ \lambda_1 a_{12} \end{bmatrix} \rightarrow \begin{matrix} a_{11} = 0.81 \\ a_{12} = 0.59 \end{matrix}$$

$$\begin{bmatrix} 6.25 & 4.25 \\ 4.25 & 3.5 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} \lambda_2 a_{21} \\ \lambda_2 a_{22} \end{bmatrix} \rightarrow \begin{matrix} a_{21} = -.59 \\ a_{22} = 0.81 \end{matrix}$$

Stopping Rules

- Problem: It requires *k* principal components to perfectly reproduce an observed covariance matrix among *k* measured variables
- But, this doesn't simplify the dimensionality
- Instead, how many principal components do you need to reproduce the observed covariance matrix *reasonably well?*
  - *Kaiser's Criterion*
    - *If* $\lambda_j < 1$ *then component explains less variance than original variable (correlation matrix)*
  - *Cattell's Scree Criterion*

Scree Plot



Component Rotation

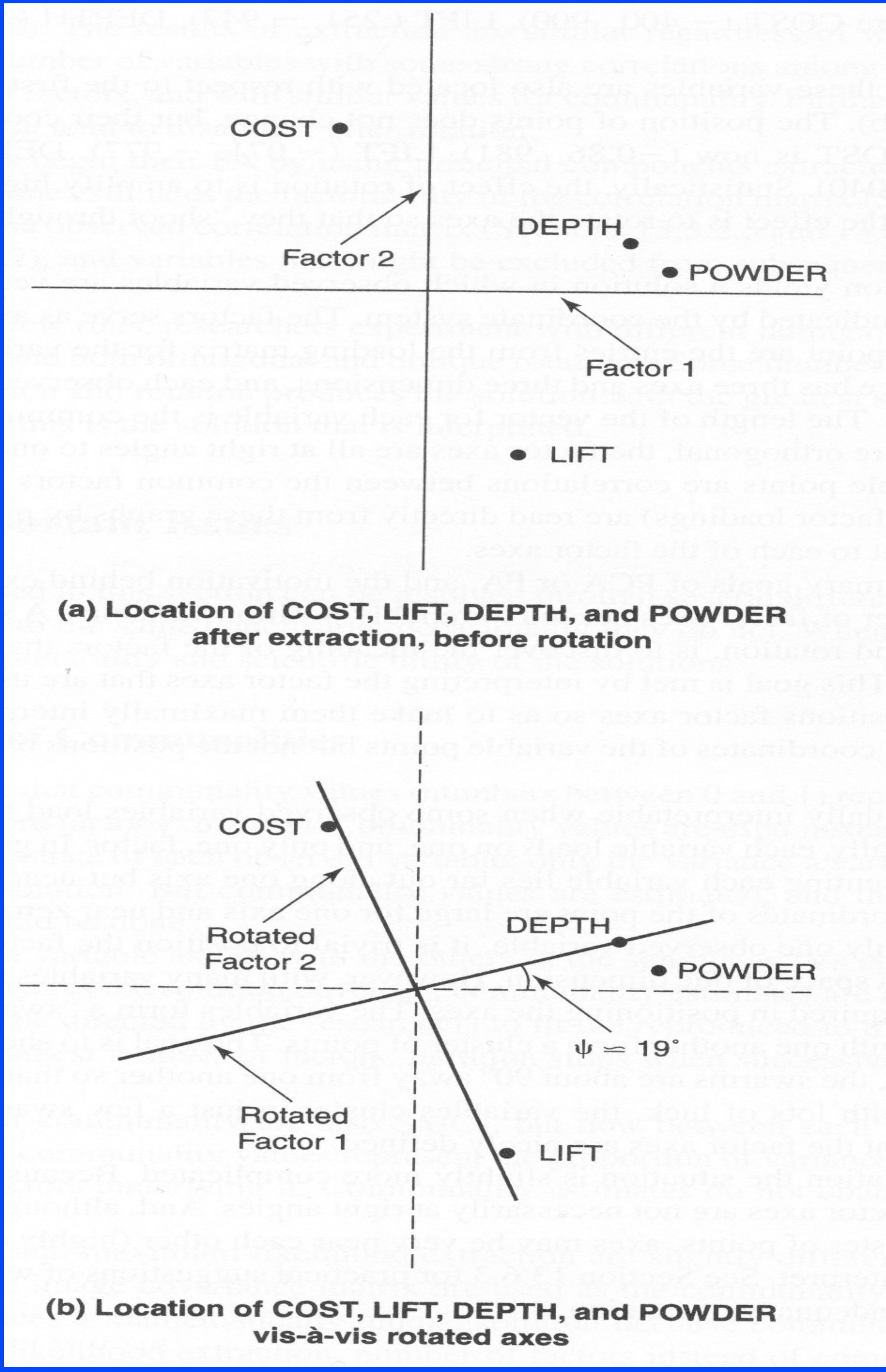
- The components have been achieved using a maximal variance criterion.
  - Good for prediction using the fewest possible composites
  - Bad for understanding
- So, once the number of desired components has been determined, rotate them to a more understandable pattern/criterion
  - Simple Structure!

Simple Structure

- Thurstone, 1944
  - Each variable has at least one zero loading
  - Each factor in a factor matrix with k columns should have k zero loadings
  - Each pair of columns in a factor matrix should have several variables loading on one factor but not the other
  - Each pair of columns should have a large proportion of variables with zero loadings in both columns
  - Each pair of columns should only have a small proportion of variables with non zero loadings in both columns

Component Rotation

- Geometric Version
- Factor loadings are found by dropping a line from the variable coordinates to the factor at a right angle
- Repositioning the axes changes the loadings on the factor but keeps the relative positioning of the points the same



Simple Structure Rotations

- Orthogonal vs. Oblique
  - Orthogonal rotation keeps factors un-correlated while increasing the meaning of the factors
  - Oblique rotation allows the factors to correlate leading to a conceptually clearer picture but a nightmare for explanation

Orthogonal Rotations

- Varimax – most popular
  - Simple structure by maximizing variance of loadings within factors across variables
  - Makes large loading larger and small loadings smaller
  - Spreads the variance from first (largest) factor to other smaller factors
- Quartimax - Not used as often
  - Opposite of Varimax
  - minimizes the number of factors needed to explain each variable
  - often generates a general factor on which most variables are loaded to a high or medium degree.

Orthogonal Rotations

- Equamax – Not popular
  - hybrid of the earlier two that tries to simultaneously simplify factors and variables
  - compromise between Varimax and Quartimax criteria.

Oblique Rotations

- Direct Oblimin – Most common oblique
  - Begins with an unrotated solution
  - Has a parameter (gamma in SPSS) that allows the user to define the amount of correlation acceptable
  - gamma values near -4 -> orthogonal, 0 leads to mild correlations (also direct quartimin) and 1 highly correlated
- Promax – more efficient
  - Solution is rotated maximally with an orthogonal rotation
  - Followed by oblique rotation - Easy and quick method
  - Orthogonal loadings are raised to powers in order to drive down small loadings - Simple structure

Component Loadings

- Component loadings are the correlations between the variables (rows) and components (columns).
- Most would say should be higher than .3
  - accounts for 10% of variance in composite
- The squared factor loading is the percent of variance in that variable explained by the component
- In oblique rotation, one gets both a pattern matrix and a structure matrix

Component Loadings

- Structure matrix
  - factor loading matrix as in orthogonal rotation
  - Correlation of the variable with the component
  - Contains both unique and common variance
- Pattern matrix
  - coefficients represent partial correlations with components.
  - Like regression weights
  - The more factors, the lower the pattern coefficients because there will be more common contributions to variance explained

Component Loadings

- For oblique rotation, you should look at both the structure and pattern coefficients when attributing a label to a factor
  - Pattern matrices often appear to give simpler structure.
  - Many authors argue that this “apparent” simple structure is misleading because it ignores the correlation among the components.

Pattern vs. Structure matrices

)elta = -.8

Pattern Matrix[a]

	Component		
	1	2	3
physical aggression	.249	.787	-2.49E-02
property damage	3.221E-02	.847	-.123
theft	-.231	.774	7.099E-02
extreme verbal abuse	.611	.511	-4.61E-02
sad	.852	-.154	-9.73E-02
anxious	.888	3.827E-02	1.194E-02
self-confidence	6.259E-02	6.719E-02	.939
compliance	-.116	-.120	.872

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

a. Rotation converged in 16 iterations.

Structure Matrix

	Component		
	1	2	3
physical aggression	.370	.826	-.128
property damage	.179	.861	-.192
theft	-.129	.735	.053
extreme verbal abuse	.694	.605	-.191
sad	.846	-.021	-.234
anxious	.892	.168	-.146
self-confidence	-.091	.006	.923
compliance	-.286	-.203	.901

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

Component Scores

- A person's score on a composite is simply the weighted sum of the variable scores
- A component score is a person's score on that composite variable -- when their variable values are applied as:

$$PC_1 = a_{11}X_1 + a_{21}X_2 + \ldots + a_{k1}X_k$$

  - The weights are the eigenvectors.
- These scores can be used as variables in further analyses (e.g., regression)

Covariance or Correlation Matrix?

- *Covariance Matrix:*
  - Variables must be in same units
  - Emphasizes variables with most variance
  - Mean eigenvalue $\neq 1.0$
- *Correlation Matrix:*
  - Variables are standardized (mean 0.0, SD 1.0)
  - Variables can be in different units
  - All variables have same impact on analysis
  - Mean eigenvalue = 1.0

Example: US crime statistics

- Variables
  - Murder
  - Rape
  - Robbery
  - Assault
  - Burglary
  - Larceny
  - Autotheft

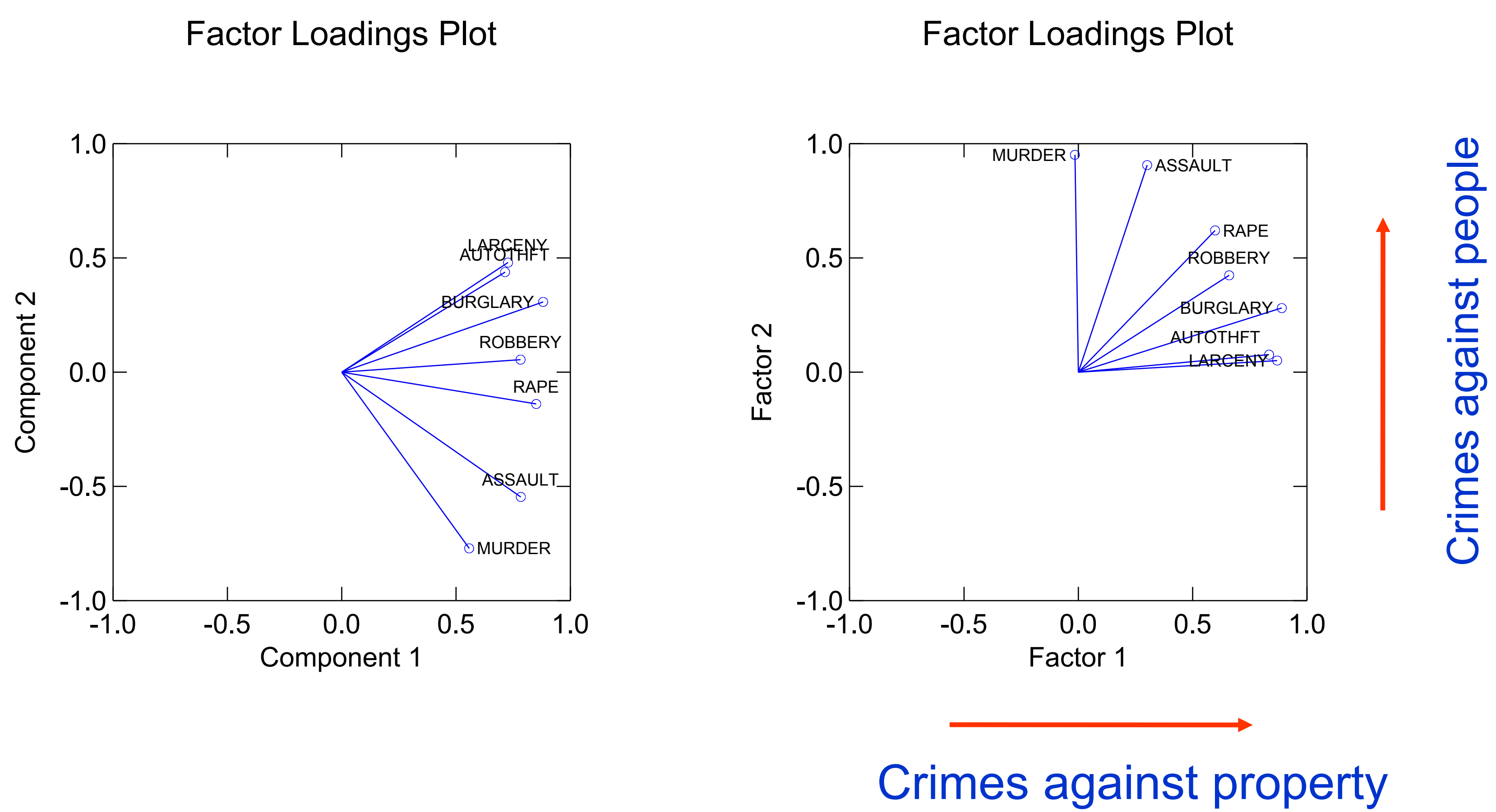
  - Data: Frequency by state

Component loadings

	1	2
MURDER	0.557	-0.771
RAPE	0.851	-0.139
ROBBERY	0.782	0.055
ASSAULT	0.784	-0.546
BURGLARY	0.881	0.308
LARCENY	0.728	0.480
AUTOTHFT	0.714	0.438

Example: Component Loadings

After Varimax Rotation:



PC process summary

- Decide whether to use correlation or covariance matrix
- Find eigenvectors (**components**) and eigenvalues (**variance accounted for**)
- Decide how many components to use by examining eigenvalues (perhaps using **scree diagram**)
- Rotate subset of components to simple structure
- Examine **loadings** (perhaps vector loading plot)
- Plot **scores**

PCA Terminology & Relations

- *j*th **principal component** is *j*th eigenvector of correlation/covariance matrix
- **scores** are values of units on components (produced using coefficients)
- **amount of variance accounted for** by component is given by eigenvalue, λ_j
- **proportion of variance accounted for** by component is given by $\lambda_j / \Sigma \lambda_j$
- **loading** of *k*th original variable on *j*th component is given by $a_{jk}\sqrt{\lambda_j}$ --correlation between variable and component

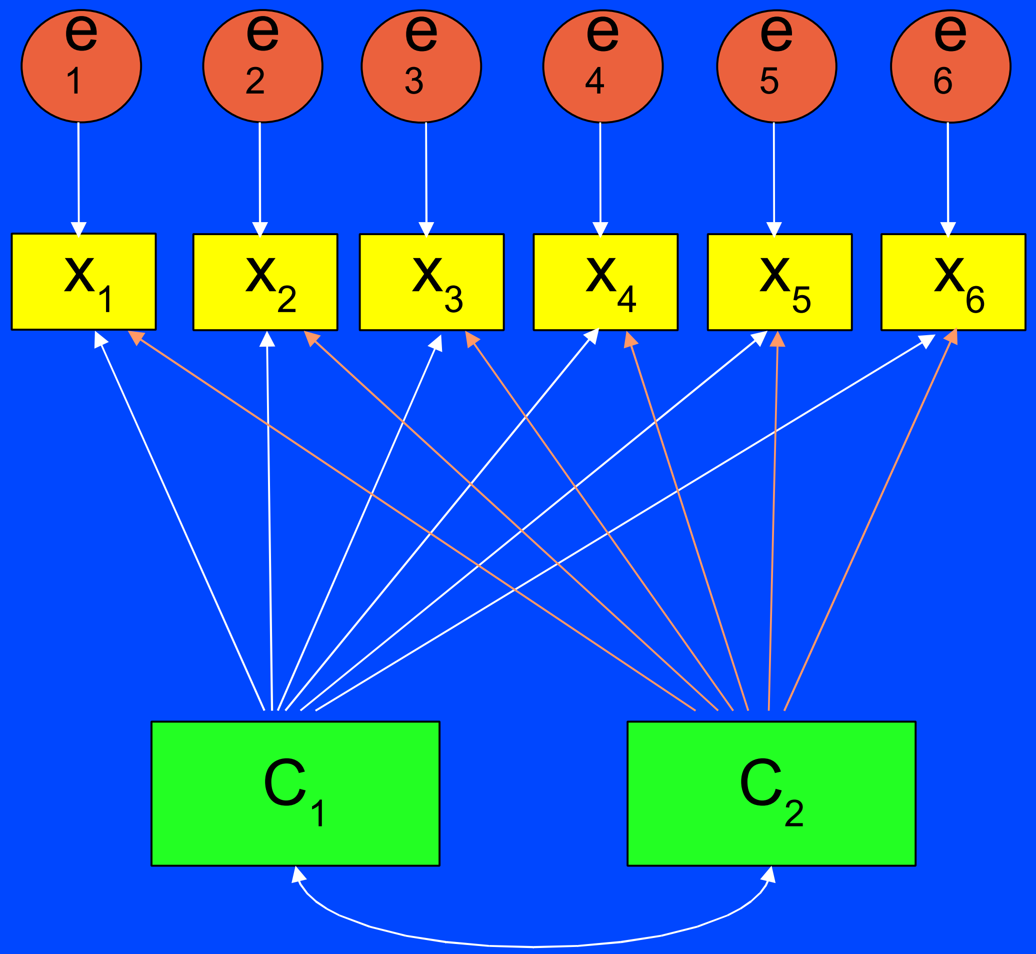
PCA Relations

- Sum of eigenvalues = p
  - if the input matrix was a correlation matrix
- Sum of eigenvalues = sum of input variances
  - if the input matrix was a covariance matrix
- Proportion of variance explained = eigenvalue / sum of eigenvalues
- Sum of squared factor loadings for jth principal component = eigenvalue_j

PCA Relations

- Sum of squared factor loadings for variable i
  - = variance explained in variable i
  - = Cii (diagonal entry i in matrix C)
  - = communality_i in common factor analysis
  - = variance of variable i if m = p
- Sum of crossproducts between columns i and j of factor loading matrix = Cij (entry ij in matrix C)
- The relations in #4, #5 and #6 are still true after rotation.

Factor Analysis Model



r = 0.0?

Factor Analysis

- Latent variables are thought to cause the manifest variables
- The relationship isn't perfect and so each measured variable is due, in part, to the latent variables and the residual variance is treated as random error

$$x_1 = a_{11} f_1 + a_{12} f_2 + \ldots + a_{1k} f_k + \mathrm{e}_1$$

$$x_2 = a_{21} f_1 + a_{22} f_2 + \ldots + a_{2k} f_k + \mathrm{e}_2$$

$$\ldots$$

$$x_p = a_{p1} f_1 + a_{p2} f_2 + \ldots + a_{pk} f_k + \mathrm{e}_3$$

Is the Factor model identified?

- Look back at the factor model...
- 6 measured variables
  - 6*7/2=21 free parameters
- How many parameters estimated in the factor analysis model?
  - 6 error variances
  - 12 path coefficients
  - 1 factor correlation
  - =19; okay....

Is the Factor model identified?

- What if you try to extract 3 factors?
  - 6 error variances
  - 18 loadings
  - 3 correlations
- 27 parameters being estimated
- Uh-ohhhh... 27>21

- There are many sources of mathematical indeterminacy in the factor analysis model

A useful dodge...

- The factor analysis method gets around the identification problem by estimating the loadings and the errors separately
- Mathematically, the main difference between FA and PCA is that FA uses a reduced correlation matrix
  - Based on communality estimates
- Factor analysis finds the eigenvalues and eigenvectors of the correlation matrix with the squared multiple correlations each variable with other variables on the main diagonal

Estimating Communality

- Many ways to estimate communality and the many varieties of FA differ with respect to how communality is estimated
  - Principle Factors aka Principle Axis
  - Iterated Principle Factors
- The most common estimate of communality(h^2) is the squared multiple correlation (SMC)
- In other words, regress each variable on all other variables and get the multiple R.

$$x_{i2} = b_o + b_1 x_{i1} + b_2 x_{i3} + \ldots + b_p x_{ip}$$

Reduce the correlation matrix

1.0	.72	.63	.54	.45
.72	1.0	.56	.48	.40
.63	.56	1.0	.42	.35
.54	.48	.42	1.0	.30
.45	.40	.35	.30	1.0

.81	.72	.63	.54	.45
.72	**.64**	.56	.48	.40
.63	.56	**.49**	.42	.35
.54	.48	.42	**.36**	.30
.45	.40	.35	.30	**.25**

FA Analysis

- Now, just perform a PCA on the reduced correlation matrix
- Re-estimate communalities based on the factor solution

Common problems in FA

- The communality estimates are just that...estimates.
- These estimates can often result in impossible results.
  - Communality estimates greater than 1.0
  - Error variance estimates less than 0.0
- Collectively referred to as “Heywood Cases”
- When encountered, the model does not fit.
  - Simplify the model or reduce the number of variables being analyzed.

Factor Scores

- Unlike PCA, a person's score on the latent variable is indeterminant
  - Two unknowns (latent true score and error) but only one observed score for each person
- Can't compute the factor score as you can in PCA.
- Instead you have to estimate the person's factor score.

Differences between PCA and FA

- Unless you have lots of error (very low communalities) you will get virtually identical results when you perform these two analyses
- I always do both
- I've only seen a discrepancy one or two times
  - Change FA model (number of factors extracted) or estimate communality differently or reduce the number of variables being factored

Some Guidelines

- Factors need at least three variables with high loadings or should not be interpreted
  - Since the vars won't perform as expected you should probably start out with 6 to 10 variables per factor.
- If the loadings are low, you will need more variables, 10 or 20 per factor may be required.
- The larger the n, the larger the number of vars per factor, and the larger the loadings, the better
  - Strength in one of these areas can compensate for weakness in another
  - Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. Psychological Methods, 3, 231-251.

Some Guidelines

- Large N, high h^2, and high **overdetermination** (each factor having at least three or four high loadings and simple structure) increase your chances of reproducing the population factor pattern
- When communalities are high (> .6), you should be in good shape even with N well below 100
- With communalities moderate (about .5) and the factors well-determined, you should have 100 to 200 subjects

Some Guidelines

- With communalities low (< .5) but high overdetermination of factors (not many factors, each with 6 or 7 high loadings), you probably need well over 100 subjects.
- With low communalities and only 3 or 4 high loadings on each, you probably need over 300 subjects.
- With low communalities and poorly determined factors, you will need well over 500 subjects.
  - MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4, 84-99.

Example...