Psychology 405: Psychometric Theory Constructs, Components, and Factor Models

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2012

Outline

Models of data

(MacCallum, 2004) "A factor analysis model is not an exact representation of real-world phenomena.
Always wrong to some degree, even in population.
At best, model is an approximation of real world."
Box (1979): "Models, of course, are never true, but fortunately it is only necessary that they be useful. For this it is usually needful only that they not be grossly wrong."
Tukey (1961): "In a single sentence, the moral is: Admit that complexity always increases, first from the model you fit to the data, thence to the model you use to think and plan about the experiment and its analysis, and thence to the true situation."

(From MacCallum, 2004); http://www.fa100.info/maccallum2.pdf

Observed Variables

X

Y

X_1

X_2

X_3

X_4

X_5

X_6

Y_1

Y_2

Y_3

Y_4

Y_5

Y_6

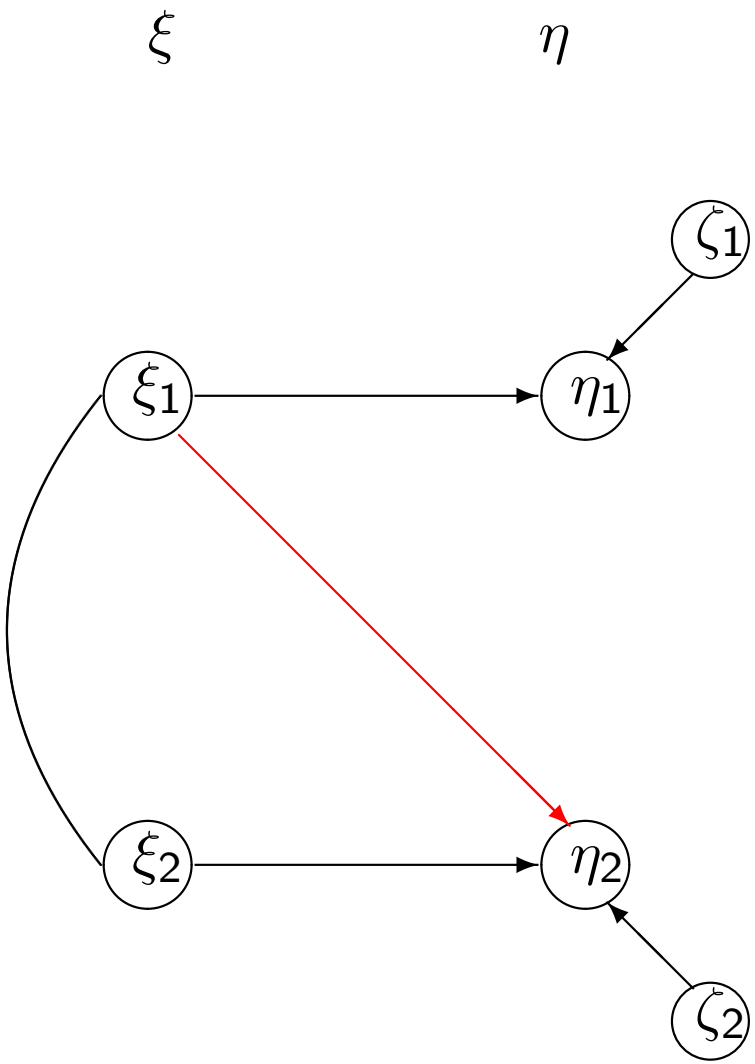
Latent Variables

ξ η

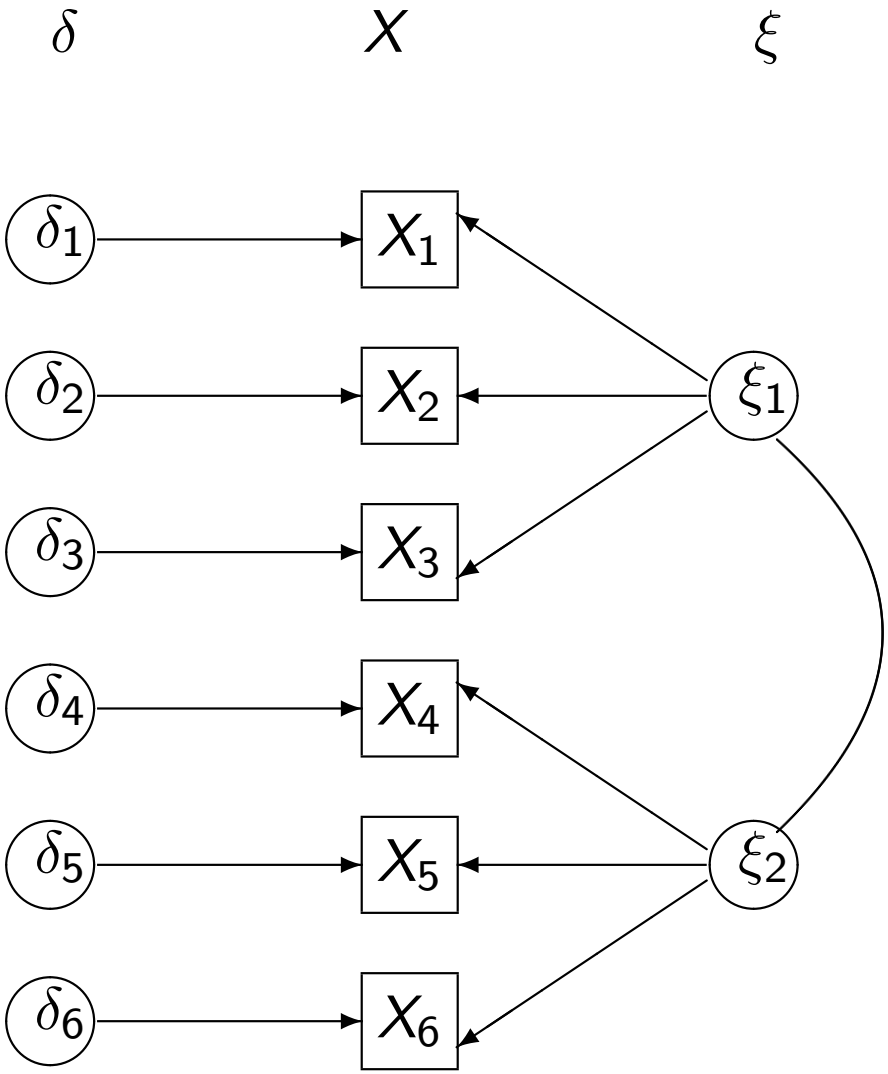
ξ_1 η_1

ξ_2 η_2

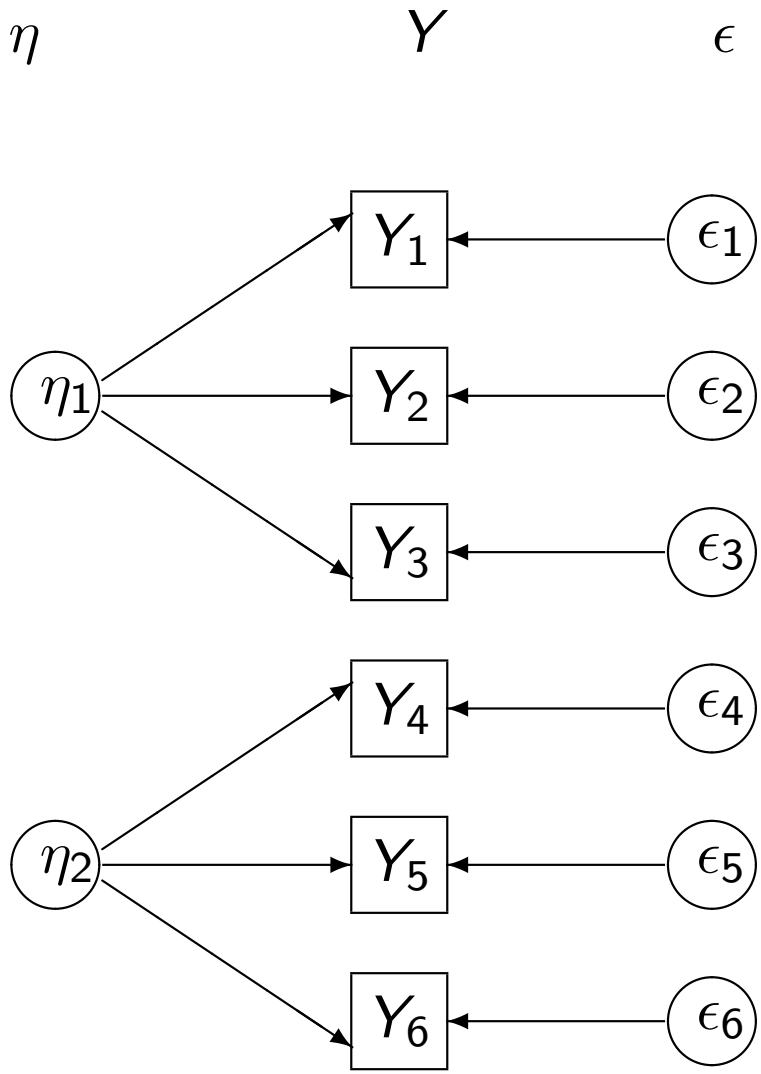
Theory: A regression model of latent variables



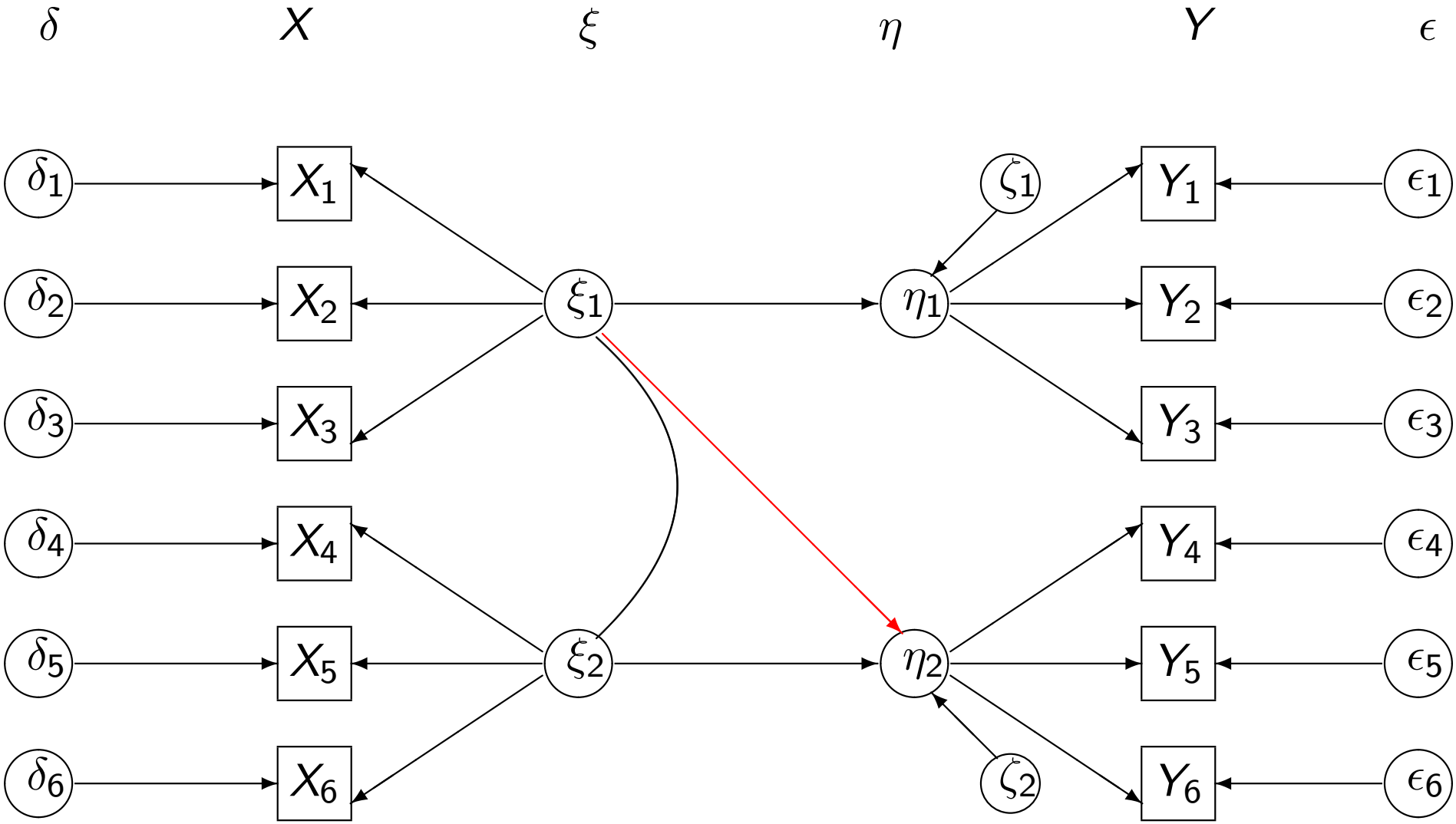
A measurement model for X



A measurement model for Y



A complete structural model



Various measurement models

1. Observed variables models
   - Singular Value Decompostion
   - Eigen Value – Eigen Vector decomposition
   - Principal Components
   - First k principal components as an approximation
2. Latent variable models
   - Factor analysis
3. Interpretation of models
   - Choosing the appropriate number of components/factors
   - Transforming/rotating towards interpretable structures

Singular Value Decomposition of the data matrix

Consider the matrix $\mathbf{X}$ of n deviation scores for N subjects, where each element, x_{ij}, represents the responses of the i^{th} individual to the j^{th} item or test. For simplicity, let the x_{ij} scores in each column be deviations from the mean for that column (i.e., they are column centered, perhaps by using `scale`). Let the number of variables be n. Then the `svd` function will find the *Singular Value Decomposition* of $\mathbf{X}$ which allows us to express $\mathbf{X}$ as the product of three orthogonal matrices:

$$_{\mathbf{N}}\mathbf{X}_{\mathbf{n}} = {}_{\mathbf{N}}\mathbf{U}_{\mathbf{nn}}\mathbf{D}_{\mathbf{nn}}\mathbf{V}'_{\mathbf{n}}$$

where $\mathbf{D}$ is a diagonal matrix of the *singular values* and the $\mathbf{U}$ and $\mathbf{V}$ matrices are matrices of the *singular vectors*. Although descriptive of the data, what is *meaning* of these vectors?

Decomposition (models) of Correlation and Covariance matrices

With **X** defined as before, the covariance matrix, **Cov**, is

$$\mathbf{Cov} = N^{-1}\mathbf{X}\mathbf{X}'$$

and the standard deviations are

$$\mathbf{sd} = \sqrt{diag(\mathbf{Cov})}.$$

Let the matrix $\mathbf{I_{sd}}$ be a diagonal matrix with elements $= \frac{1}{sd_i}$, then the correlation matrix **R** is

$$\mathbf{R} = \mathbf{I_{sd}}\mathbf{Cov}\mathbf{I_{sd}}.$$

The problem is how to approximate the matrix, **R** of rank n, with a matrix of lower rank? The solution to this problem may be seen if we think about how to create a model matrix to approximate **R**.

An example correlation matrix

Consider the following correlation matrix

	V1	V2	V3	V4	V5	V6
V1	1.00	0.72	0.63	0.54	0.45	0.36
V2	0.72	1.00	0.56	0.48	0.40	0.32
V3	0.63	0.56	1.00	0.42	0.35	0.28
V4	0.54	0.48	0.42	1.00	0.30	0.24
V5	0.45	0.40	0.35	0.30	1.00	0.20
V6	0.36	0.32	0.28	0.24	0.20	1.00

Is it possible to model these 36 correlations and variances with fewer terms? Yes, of course. The diagonal elements are all 1 and the off diagonal elements are symmetric. Thus, we have $n * (n - 1)$ correlations we want to model.

Eigen vector decomposition

Given a *nxn* matrix $\mathbf{R}$, each eigenvector, $\mathbf{x_i}$, solves the equation

$$\mathbf{x_i R} = \lambda_i \mathbf{x_i}$$

and the set of n eigenvectors are solutions to the equation

$$\mathbf{XR} = \boldsymbol{\lambda} \mathbf{X}$$

where $\mathbf{X}$ is a matrix of orthogonal eigenvectors and $\boldsymbol{\lambda}$ is a diagonal matrix of the the eigenvalues, λ_i. Then

$$\mathbf{x_i R} - \lambda_i \mathbf{XI} = 0 <=> \mathbf{x_i}(\mathbf{R} - \lambda_i \mathbf{I}) = 0$$

Finding the eigenvectors and eigenvalues is computationally tedious, but may be done using the `eigen` function. That the vectors making up $\mathbf{X}$ are orthogonal means that

$$\mathbf{X X'} = \mathbf{I}$$

and because they form the *basis space* for $\mathbf{R}$ that

$$\mathbf{R} = \mathbf{X} \boldsymbol{\lambda} \mathbf{X'}.$$

Consider the eigen value solution for the example correlation matrix.

```
> e <- eigen(R)
> print(e,digits=2)
$values
[1] 3.16 0.82 0.72 0.59 0.44 0.26

$vectors
      [,1]   [,2]   [,3]  [,4]   [,5]   [,6]
[1,] -0.50 -0.061  0.092  0.14  0.238  0.816
[2,] -0.47 -0.074  0.121  0.21  0.657 -0.533
[3,] -0.43 -0.096  0.182  0.53 -0.675 -0.184
[4,] -0.39 -0.142  0.414 -0.78 -0.201 -0.104
[5,] -0.34 -0.299 -0.860 -0.20 -0.108 -0.067
[6,] -0.28  0.934 -0.178 -0.10 -0.067 -0.045

> round(e$vectors %*% t(e$vectors),2)   #the eigen vectors are orthogonal
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    0    0    0    0
[2,]    0    1    0    0    0    0
[3,]    0    0    1    0    0    0
[4,]    0    0    0    1    0    0
[5,]    0    0    0    0    1    0
[6,]    0    0    0    0    0    1
```

Eigen Value decomposition and recreation of the original matrix

Find the eigen values (λ) and eigen vectors ($\mathbf{V_i}$).

```
> e <- eigen(R)
> print(e,digits=2)

$values
[1] 3.16 0.82 0.72 0.59 0.44 0.26

$vectors
      [,1]   [,2]   [,3]  [,4]   [,5]   [,6]
[1,] -0.50 -0.061  0.092  0.14  0.238  0.816
[2,] -0.47 -0.074  0.121  0.21  0.657 -0.533
[3,] -0.43 -0.096  0.182  0.53 -0.675 -0.184
[4,] -0.39 -0.142  0.414 -0.78 -0.201 -0.104
[5,] -0.34 -0.299 -0.860 -0.20 -0.108 -0.067
[6,] -0.28  0.934 -0.178 -0.10 -0.067 -0.045

#the eigen vectors are orthogonal

> round(e$vectors %*% t(e$vectors),2)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    0    0    0    0
[2,]    0    1    0    0    0    0
[3,]    0    0    1    0    0    0
[4,]    0    0    0    1    0    0
[5,]    0    0    0    0    1    0
[6,]    0    0    0    0    0    1
```

The eigen vectors and values recreate the observed correlations.

$$\mathbf{R} = \mathbf{V}\lambda\mathbf{V}'.$$

```
> round(e$vectors %*% diag(e$values) %*% t(e$vectors),2)

     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.00 0.72 0.63 0.54 0.45 0.36
[2,] 0.72 1.00 0.56 0.48 0.40 0.32
[3,] 0.63 0.56 1.00 0.42 0.35 0.28
[4,] 0.54 0.48 0.42 1.00 0.30 0.24
[5,] 0.45 0.40 0.35 0.30 1.00 0.20
[6,] 0.36 0.32 0.28 0.24 0.20 1.00
```

The eigen values reflect the scale, the vectors the structure

Consider the original data and solution

```
> R
> e <- eigen(R)
> print(e,digits=2)
```

```
     V1   V2   V3   V4   V5   V6
V1 1.00 0.72 0.63 0.54 0.45 0.36
V2 0.72 1.00 0.56 0.48 0.40 0.32
V3 0.63 0.56 1.00 0.42 0.35 0.28
V4 0.54 0.48 0.42 1.00 0.30 0.24
V5 0.45 0.40 0.35 0.30 1.00 0.20
V6 0.36 0.32 0.28 0.24 0.20 1.00

$values
[1] 3.16 0.82 0.72 0.59 0.44 0.26

$vectors
      [,1]   [,2]   [,3]  [,4]   [,5]   [,6]
[1,] -0.50 -0.061  0.092  0.14  0.238  0.816
[2,] -0.47 -0.074  0.121  0.21  0.657 -0.533
[3,] -0.43 -0.096  0.182  0.53 -0.675 -0.184
[4,] -0.39 -0.142  0.414 -0.78 -0.201 -0.104
[5,] -0.34 -0.299 -0.860 -0.20 -0.108 -0.067
[6,] -0.28  0.934 -0.178 -0.10 -0.067 -0.045
```

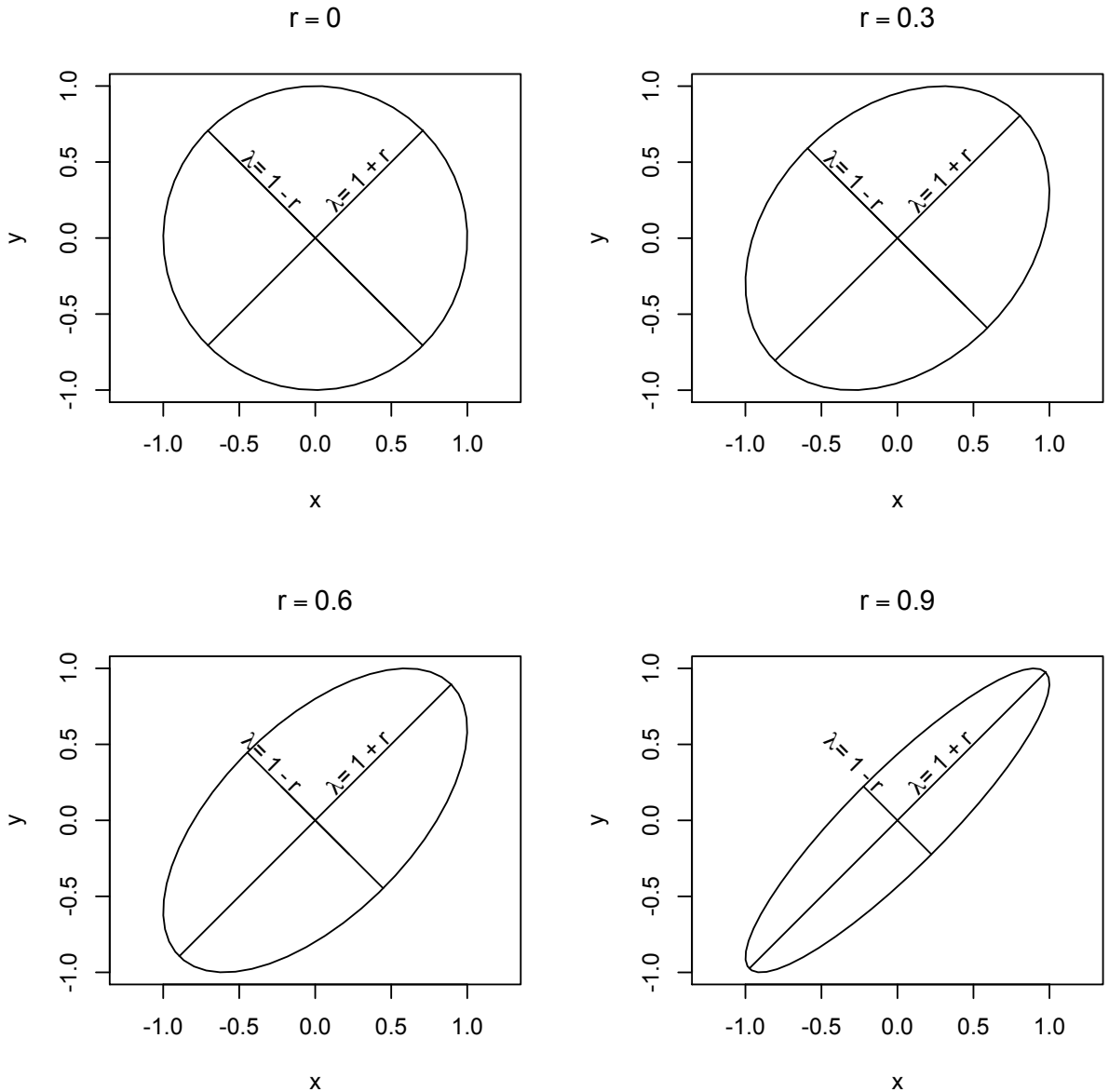
Note that the signs are arbitrary.

Consider if all the correlations are divided by 2.

```
> R.5 <- as.matrix(R/2)
> diag(R.5) <- 1
> R.5
> e.5 <- eigen(R.5)
> print(e.5,2)
```

```
      V1   V2    V3   V4    V5   V6
V1 1.000 0.36 0.315 0.27 0.225 0.18
V2 0.360 1.00 0.280 0.24 0.200 0.16
V3 0.315 0.28 1.000 0.21 0.175 0.14
V4 0.270 0.24 0.210 1.00 0.150 0.12
V5 0.225 0.20 0.175 0.15 1.000 0.10
V6 0.180 0.16 0.140 0.12 0.100 1.00

$values
[1] 2.08 0.91 0.86 0.80 0.72 0.63

$vectors
     [,1]   [,2]   [,3]  [,4]   [,5]   [,6]
[1,] 0.50 -0.061  0.092  0.14  0.238  0.816
[2,] 0.47 -0.074  0.121  0.21  0.657 -0.533
[3,] 0.43 -0.096  0.182  0.53 -0.675 -0.184
[4,] 0.39 -0.142  0.414 -0.78 -0.201 -0.104
[5,] 0.34 -0.299 -0.860 -0.20 -0.108 -0.067
[6,] 0.28  0.934 -0.178 -0.10 -0.067 -0.045
```

Eigen vectors of a 2 x 2 correlation matrix



Although the length (eigen values) of the axes differ, their orientation (eigen vectors) are the same.

```
> r2 <- matrix(c(1,.6,.6,1),2,2)
> print(eigen(r2),2)

$values
[1] 1.6 0.4

$vectors
     [,1]  [,2]
[1,] 0.71 -0.71
[2,] 0.71  0.71
```

From eigen vectors to Principal Components

1. For n variables, there are n eigen vectors
   - There is no parsimony in thinking of the eigen vectors
   - Except that the vectors provide the orthogonal basis for the variables
2. Principal components are formed from the eigen vectors and eigen values
   - $\mathbf{R} = \mathbf{V}\lambda\mathbf{V}' = \mathbf{C}\mathbf{C}'$
   - $\mathbf{C} = \mathbf{V}\sqrt{\lambda}$
3. But there will still be as many Principal Components as variables, so what is the point?
4. Take just the first k Principal Components and see how well this reduced model fits the data.

The first principal component.

```
> pc1 <- principal(R,1)
> pc1

Uniquenesses:
   V1    V2    V3    V4    V5    V6
0.220 0.307 0.408 0.519 0.635 0.748
Loadings:
   PC1
V1 0.88
V2 0.83
V3 0.77
V4 0.69
V5 0.60
V6 0.50

                 PC1
SS loadings    3.142
Proportion Var 0.524
```

```
 #show  the model
> round(pc1$loadings %*% t(pc1$loadings),2)

    V1   V2   V3   V4   V5   V6
V1 0.77 0.73 0.68 0.61 0.53 0.44
V2 0.73 0.69 0.64 0.57 0.50 0.42
V3 0.68 0.64 0.59 0.53 0.46 0.38
V4 0.61 0.57 0.53 0.48 0.41 0.34
V5 0.53 0.50 0.46 0.41 0.36 0.30
V6 0.44 0.42 0.38 0.34 0.30 0.25

#find the residuals
> Rresid <-  R - pc1$loadings %*% t(pc1$loadings)
> round(Rresid,2)

      V1    V2    V3    V4    V5    V6
V1  0.23 -0.01 -0.05 -0.07 -0.08 -0.08
V2 -0.01  0.31 -0.08 -0.09 -0.10 -0.09
V3 -0.05 -0.08  0.41 -0.11 -0.11 -0.10
V4 -0.07 -0.09 -0.11  0.52 -0.11 -0.10
V5 -0.08 -0.10 -0.11 -0.11  0.64 -0.10
V6 -0.08 -0.09 -0.10 -0.10 -0.10  0.75
```

The model fits pretty well, except that the diagonal is underestimated and the other correlations are over estimated.

Try 2 and 3 principal components

```
> p2 <- principal(R,2,rotate="none")
> p2
> resid(p2)

Principal Components Analysis
Call: principal(r = R, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix)
    PC1   PC2   h2    u2
V1 0.88 -0.06 0.78 0.217
V2 0.83 -0.07 0.70 0.302
V3 0.77 -0.09 0.60 0.400
V4 0.69 -0.13 0.50 0.502
V5 0.60 -0.27 0.44 0.561
V6 0.50  0.85 0.97 0.031

                 PC1  PC2
SS loadings     3.16 0.82
Proportion Var 0.53 0.14
Cumulative Var 0.53 0.66


Fit based upon off diagonal values = 0.95

   V1    V2    V3    V4    V5    V6
V1  0.22
V2 -0.02  0.30
V3 -0.05 -0.09  0.40
V4 -0.08 -0.11 -0.13  0.50
V5 -0.10 -0.12 -0.14 -0.15  0.56
V6 -0.04 -0.04 -0.03  0.00  0.13  0.03
```

```
> p3 <- principal(R,3,rotate="none")
> resid(p3)

Principal Components Analysis
Call: principal(r = R, nfactors = 3, rotate = "none")
Standardized loadings (pattern matrix)
    PC1   PC2   PC3   h2     u2
V1 0.88 -0.06 -0.08 0.79 0.2108
V2 0.83 -0.07 -0.10 0.71 0.2917
V3 0.77 -0.09 -0.15 0.62 0.3761
V4 0.69 -0.13 -0.35 0.62 0.3789
V5 0.60 -0.27  0.73 0.97 0.0292
V6 0.50  0.85  0.15 0.99 0.0084

                 PC1  PC2  PC3
SS loadings     3.16 0.82 0.72
Proportion Var 0.53 0.14 0.12
Cumulative Var 0.53 0.66 0.78


Fit based upon off diagonal values = 0.97

   V1    V2    V3    V4    V5    V6
V1  0.21
V2 -0.03  0.29
V3 -0.07 -0.10  0.38
V4 -0.11 -0.14 -0.18  0.38
V5 -0.04 -0.05 -0.03  0.10  0.03
V6 -0.02 -0.03 -0.01  0.05  0.02  0.01
```

Factor and component models

What if we did not model the diagonal, just the off diagonal elements?
The component model is that

$$\mathbf{R} = \mathbf{C}\mathbf{C}' \tag{1}$$

while the factor model is

$$\mathbf{R} = \mathbf{F}\mathbf{F}' + \mathbf{U}^2 \tag{2}$$

where U^2 is a diagonal matrix of uniquenesses. Functionally, this is modeling the off diagonals.

Factors vs. components

Originally developed by Spearman (1904) for the case of one common factor, and then later generalized by Thurstone (1947) and others to the case of multiple factors, factor analysis is probably the most frequently used and sometimes the most controversial psychometric procedure. The factor model, although seemingly very similar to the components model, is in fact very different. For rather than having components as linear sums of variables, in the factor model the variables are themselves linear sums of the unknown factors. That is, while components can be solved for by doing an *eigenvalue* or *singular value decomposition*, factors are estimated as best fitting solutions (Eckart & Young, 1936; Householder & Young, 1938), normally through iterative methods (Jöreskog, 1978; Lawley & Maxwell, 1963). Cattell (1965) referred to components analysis as a closed model and factor analysis as an open model, in that by explaining just the common variance, there was still more variance to explain.

Iterative principal axes factor analysis

Principal components represents a $n * n$ matrix in terms of the first k components. It attempts to reproduce all of the **R** matrix. *Factor analysis* on the other hand, attempts to model just the common part of the matrix, which means all of the off-diagonal elements and the common part of the diagonal (the *communalities*). The non-common part, the *uniquenesses*, are simply that which is left over. An easy to understand procedure is *principal axes* factor analysis. This is similar to principal components, except that it is done with a reduced matrix where the diagonals are the communalities. The communalities can either be specified a priori, estimated by such procedures as multiple linear regression, or found by iteratively doing an eigenvalue decomposition and repeatedly replacing the original 1s on the diagonal with the the value of 1 - u^2 where

$$\mathbf{U}^2 = diag(\mathbf{R} - \mathbf{F}\mathbf{F}').$$

Principal axes as eigen values of a reduced matrix

That is, starting with the original correlation or covariance matrix, $\mathbf{R}$, find the k largest principal components, reproduce the matrix using those principal components. Find the resulting residual matrix, $\mathbf{R}^*$ and uniqueness matrix, $\mathbf{U}^2$ by

$$\mathbf{R}^* = \mathbf{R} - \mathbf{F}\mathbf{F}' \tag{3}$$

$$\mathbf{U}^2 = diag(\mathbf{R}^*)$$

and then, for iteration i, find $\mathbf{R}_i$ by replacing the diagonal of the original $\mathbf{R}$ matrix with 1 - diag($\mathbf{U}^2$) found on the previous step. Repeat this process until the change from one iteration to the next is arbitrarily small.

Comparing 1 with 5 iterations

```
> f1 <- fa(R,1,fm='pa',max.iter=1)
> f1
> resid(f1)
```

```
Factor Analysis using method =  pa
Call: fa(r = R, nfactors = 1, max.iter = 1, fm = "pa")
Standardized loadings (pattern matrix)
    PA1   h2   u2
V1 0.86 0.74 0.26
V2 0.79 0.62 0.38
V3 0.70 0.48 0.52
V4 0.60 0.36 0.64
V5 0.50 0.25 0.75
V6 0.40 0.16 0.84

                PA1
SS loadings    2.62
Proportion Var 0.44

   V1   V2   V3   V4   V5   V6
V1 0.26
V2 0.04 0.38
V3 0.03 0.01 0.52
V4 0.02 0.01 0.00 0.64
V5 0.02 0.00 0.00 0.00 0.75
V6 0.01 0.00 0.00 0.00 0.00 0.84
```

```
> f1 <- fa(R,1,fm='pa',max.iter=5)
> f1
> resid(f1)
```

```
Factor Analysis using method =  pa
Call: fa(r = R, nfactors = 1, max.iter = 5, fm = "pa")
Standardized loadings (pattern matrix)
   PA1   h2   u2
V1 0.9 0.81 0.19
V2 0.8 0.64 0.36
V3 0.7 0.49 0.51
V4 0.6 0.36 0.64
V5 0.5 0.25 0.75
V6 0.4 0.16 0.84

                PA1
SS loadings    2.71
Proportion Var 0.45

   V1   V2   V3   V4   V5   V6
V1 0.19
V2 0.00 0.36
V3 0.00 0.00 0.51
V4 0.00 0.00 0.00 0.64
V5 0.00 0.00 0.00 0.00 0.75
V6 0.00 0.00 0.00 0.00 0.00 0.84
```

SMCs as initial communality estimates

Rather than starting with initial communality estimates of 1, the process can be started with other estimates of the communality. A conventional starting point is the lower bound estimate of the communalities, the *squared multiple correlation* or *SMC* (Roff, 1936).

The concept here is that a variable's communality must be at least as great as the amount of its variance that can be predicted by all of the other variables. The squared multiple correlations of each variable with the remaining variables are the diagonal elements of

$$\mathbf{I} - (diag(\mathbf{R}^{-1})^{-1}$$

and thus a starting estimate for $\mathbf{R_0}$ would be $\mathbf{R} - (diag(\mathbf{R}^{-1})^{-1}$.

Goodness of fit–simple estimates

At least three indices of goodness of fit of the principal factors model can be considered: One compares the sum of squared residuals to the sum of the squares of the original values:

$$GF_{total} = 1 - \frac{\mathbf{1R^{*2}1'}}{\mathbf{1R^{2}1'}}$$

The second does the same, but does not consider the diagonal of **R**

$$GF_{offdiagonal} = 1 - \frac{\sum_{i \neq j} r_{ij}^{*2}}{\sum_{i \neq j} r_{ij}^{*2}} = 1 - \frac{\mathbf{1R^{*2}1'} - tr(\mathbf{1R^{*2}1'}}{\mathbf{1R^{2}1'} - tr(\mathbf{1R^{2}1'})}$$

Finally, a χ^2 test of the size of the residuals simply sums all the squared residuals and multiplies by the number of observations:

$$\chi^2 = \sum_{i<j} r*_{ij}^2 (N-1)$$

with p * (p-1)/2 degrees of freedom.

OLS

The fundamental factor equation (Equation **??**) may be viewed as set of simultaneous equations which may be solved several different ways: *ordinary least squares*, *generalized least squares*, and *maximum likelihood*. Ordinary least squares (*OLS*) or *unweighted least squares* (*ULS*) minimizes the sum of the squared residuals when modeling the sample correlation or covariance matrix, $\mathbf{S}$, with $\Sigma = \mathbf{FF'} + \mathbf{U^2}$

$$E = \frac{1}{2} tr(\mathbf{S} - \Sigma)^2 \tag{4}$$

where the *trace*, `tr`, of a matrix is the sum of the diagonal elements and the division by two reflects the symmetry of the $\mathbf{S}$ matrix.

MLE

Equation 4 can be generalized to weight the residuals $(\mathbf{S} - \Sigma)$ by the inverse of the sample matrix, $\mathbf{S}$, and thus to minimize

$$E = \frac{1}{2} tr((\mathbf{S} - \Sigma)\mathbf{S}^{-1})^2 = \frac{1}{2} tr(\mathbf{I} - \Sigma\mathbf{S}^{-1})^2. \quad (5)$$

This is known as *generalized least squares* (*GLS*) or *weighted least squares* (*WLS*). Similarly, if the residuals are weighted by the inverse of the model, Σ, minimizing

$$E = \frac{1}{2} tr((\mathbf{S} - \Sigma)\Sigma^{-1})^2 = \frac{1}{2} tr(\mathbf{S}\Sigma^{-1} - I)^2 \quad (6)$$

will result in a model that maximizes the likelihood of the data. This procedure, *maximum likelihood estimation* (*MLE*) is also seen as finding the minimum of

$$E = \frac{1}{2} \left(tr(\Sigma^{-1}\mathbf{S}) - ln \left| \Sigma^{-1}\mathbf{S} \right| - p \right) \quad (7)$$

where p is the number of variables (Jöreskog, 1978). Perhaps a helpful intuitive explanation of Equation 10 is that if the model is

Maximum Likelihood Estimation

(*MLE*) is also seen as finding the minimum of

$$E = \frac{1}{2}\left(tr(\Sigma^{-1}\mathbf{S}) - ln\left|\Sigma^{-1}\mathbf{S}\right| - p\right) \tag{10}$$

where p is the number of variables (Jöreskog, 1978). Perhaps a helpful intuitive explanation of Equation 10 is that if the model is correct, then $\Sigma = \mathbf{S}$ and thus $\Sigma^{-1}\mathbf{S} = \mathbf{I}$. The trace of an identity matrix of rank p is p, and the logarithm of $|\mathbf{I}|$ is 0. Thus, the value of E if the model has perfect fit is 0. With the assumption of multivariate normality of the residuals, and for large samples, a χ^2 statistic can be estimated for a model with p variables and f factors:

$$\chi^2 = \left(tr(\Sigma^{-1}\mathbf{S}) - ln\left|\Sigma^{-1}\mathbf{S}\right| - p\right)(N - 1 - (2p + 5)/6 - (2f)/3). \tag{11}$$

This χ^2 has degrees of freedom:

$$df = p * (p - 1)/2 - p * f + f * (f - 1)/2. \tag{12}$$

Minimum Residual Factor Analysis

The previous factor analysis procedures attempt to optimize the fit of the model matrix (Σ) to the correlation or covariance matrix ($\mathbf{S}$). The diagonal of the matrix is treated as mixture of common variance and unique variance and the problem becomes one of estimating the common variance (the *communality* of each variable). An alternative is to ignore the diagonal and to find that model which minimizes the squared residuals of the off diagonal elements. This is done in the `fa` function using the "minres" option by finding the solution that minimizes

$$\frac{1}{2}\mathbf{1}\left((\mathbf{S}-\mathbf{I})-(\Sigma-tr(\Sigma)\right)^2\mathbf{1}'. \tag{13}$$

The advantage of the *minres* solution is that it does not require finding the inverse of either the original correlation matrix (as do *GLS* and *WLS*) nor of the model matrix (as does *MLE*, and thus can be performed on non-positive definite matrices or matrices that are not invertible.

Solutions with more than 1 factor or component

Nothing in the previous algebra restricted the dimensionality of the **F** matrix or **C** matrix to be one column. That is, why limit ourselves to a one dimensional solution? Consider the following correlation matrix (constructed by creating a factor matrix and then finding its inner product).

```
>F  <-  matrix(c(.9,.8,.7,rep(0,6),.8,.7,.6),ncol=2) #the model
> rownames(F) <- paste("V",seq(1:6),sep="")  #add labels
> colnames(F) <- c("F1", "F2")
> R <- F %*% t(F) #create the correlation matrix
> diag(R) <- 1  #adjust the diagonal of the matrix
> R
```

```
     V1   V2   V3   V4   V5   V6
V1 1.00 0.72 0.63 0.00 0.00 0.00
V2 0.72 1.00 0.56 0.00 0.00 0.00
V3 0.63 0.56 1.00 0.00 0.00 0.00
V4 0.00 0.00 0.00 1.00 0.56 0.48
V5 0.00 0.00 0.00 0.56 1.00 0.42
V6 0.00 0.00 0.00 0.48 0.42 1.00
```

Try one principal component to this model.

```
> pc1 <- principal(R)
> pc1
```

```
Principal Components Analysis
Call: principal(r = R)
Standardized loadings (pattern matrix) based upon correlation matrix
    PC1   h2   u2
V1 0.90 0.82 0.18
V2 0.88 0.77 0.23
V3 0.83 0.69 0.31
V4 0.00 0.00 1.00
V5 0.00 0.00 1.00
V6 0.00 0.00 1.00

                PC1
SS loadings    2.28
Proportion Var 0.38

Test of the hypothesis that 1 component is sufficient.

The degrees of freedom for the null model are  15  and the objective function was  1.96
The degrees of freedom for the model are 9  and the objective function was  0.87

Fit based upon off diagonal values = 0.61
```

The residuals are large for the second set of variables.

```
> resid(pc1)
```

```
   V1    V2    V3    V4    V5    V6
V1  0.18
V2 -0.07  0.23
V3 -0.12 -0.17  0.31
V4  0.00  0.00  0.00  1.00
V5  0.00  0.00  0.00  0.56  1.00
V6  0.00  0.00  0.00  0.48  0.42  1.00
```

Two Principal Components

```
> pc2 <- principal(R,2)
> pc2

Uniquenesses:
   V1    V2    V3    V4    V5    V6
0.182 0.234 0.309 0.282 0.332 0.409
Loadings:
   PC1  PC2
V1 0.90
V2 0.88
V3 0.83
V4      0.85
V5      0.82
V6      0.77
                 PC1   PC2
SS loadings    2.273 1.988
Proportion Var 0.379 0.331
Cumulative Var 0.379 0.710

> round(pc2$loadings %*% t(pc2$loadings),2)

   V1   V2   V3   V4   V5   V6
V1 0.81 0.79 0.75 0.00 0.00 0.00
V2 0.79 0.77 0.73 0.00 0.00 0.00
V3 0.75 0.73 0.69 0.00 0.00 0.00
V4 0.00 0.00 0.00 0.72 0.70 0.65
V5 0.00 0.00 0.00 0.70 0.67 0.63
V6 0.00 0.00 0.00 0.65 0.63 0.59
```

```
> Rresid <- R - pc2$loadings %*% t(pc2$loadings)
> round(Rresid,2)

      V1    V2    V3    V4    V5    V6
V1  0.19 -0.07 -0.12  0.00  0.00  0.00
V2 -0.07  0.23 -0.17  0.00  0.00  0.00
V3 -0.12 -0.17  0.31  0.00  0.00  0.00
V4  0.00  0.00  0.00  0.28 -0.14 -0.17
V5  0.00  0.00  0.00 -0.14  0.33 -0.21
V6  0.00  0.00  0.00 -0.17 -0.21  0.41

> resid(pc2)

   V1    V2    V3    V4    V5    V6
V1  0.18
V2 -0.07  0.23
V3 -0.12 -0.17  0.31
V4  0.00  0.00  0.00  0.28
V5  0.00  0.00  0.00 -0.13  0.33
V6  0.00  0.00  0.00 -0.17 -0.21  0.41
```

Try two factors

```
> f2 <- fa(R,2,rotate="none")
> f2


Factor Analysis using method =  minres
Call: fa(r = R, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
   MR1 MR2   h2   u2
V1 0.9 0.0 0.81 0.19
V2 0.8 0.0 0.64 0.36
V3 0.7 0.0 0.49 0.51
V4 0.0 0.8 0.64 0.36
V5 0.0 0.7 0.49 0.51
V6 0.0 0.6 0.36 0.64

                MR1  MR2
SS loadings    1.94 1.49
Proportion Var 0.32 0.25
Cumulative Var 0.32 0.57

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are  15  and the objective function was  1.96
The degrees of freedom for the model are 4  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  0

Fit based upon off diagonal values = 1
```

```
> resid(f2)


   V1   V2   V3   V4   V5   V6
V1 0.19
V2 0.00 0.36
V3 0.00 0.00 0.51
V4 0.00 0.00 0.00 0.36
V5 0.00 0.00 0.00 0.00 0.51
V6 0.00 0.00 0.00 0.00 0.00 0.64
```

Add two more variables (with a factor model)

```
#the model
> f  <-  matrix(c(.9,.8,.7,rep(0,3),.7,rep(0,4),.8,.7,.6,0,.5),ncol=2)
> rownames(f) <- paste("V",seq(1:8),sep="")  #add labels
>  colnames(f) <- c("F1", "F2")
>  R <- f %*% t(f) #create the correlation matrix
>  diag(R) <- 1  #adjust the diagonal of the matrix
>  R
```

```
     V1   V2   V3   V4   V5   V6   V7   V8
V1 1.00 0.72 0.63 0.00 0.00 0.00 0.63 0.00
V2 0.72 1.00 0.56 0.00 0.00 0.00 0.56 0.00
V3 0.63 0.56 1.00 0.00 0.00 0.00 0.49 0.00
V4 0.00 0.00 0.00 1.00 0.56 0.48 0.00 0.40
V5 0.00 0.00 0.00 0.56 1.00 0.42 0.00 0.35
V6 0.00 0.00 0.00 0.48 0.42 1.00 0.00 0.30
V7 0.63 0.56 0.49 0.00 0.00 0.00 1.00 0.00
V8 0.00 0.00 0.00 0.40 0.35 0.30 0.00 1.00
```

Factors loadings do not change, component loadings do

```
> R

     V1   V2   V3   V4   V5   V6   V7   V8
V1 1.00 0.72 0.63 0.00 0.00 0.00 0.63 0.00
V2 0.72 1.00 0.56 0.00 0.00 0.00 0.56 0.00
V3 0.63 0.56 1.00 0.00 0.00 0.00 0.49 0.00
V4 0.00 0.00 0.00 1.00 0.56 0.48 0.00 0.40
V5 0.00 0.00 0.00 0.56 1.00 0.42 0.00 0.35
V6 0.00 0.00 0.00 0.48 0.42 1.00 0.00 0.30
V7 0.63 0.56 0.49 0.00 0.00 0.00 1.00 0.00
V8 0.00 0.00 0.00 0.40 0.35 0.30 0.00 1.00

> f2 <- factanal(covmat=R,factors=2)
> f2

Call:
factanal(factors = 2, covmat = R)
Uniquenesses:
  V1   V2   V3   V4   V5   V6   V7   V8
0.19 0.36 0.51 0.36 0.51 0.64 0.51 0.75
Loadings:
   Factor1 Factor2
V1 0.9
V2 0.8
V3 0.7
V4         0.8
V5         0.7
V6         0.6
V7 0.7
V8         0.5
               Factor1 Factor2
SS loadings      2.430   1.740
```

```
pc2 <- principal(R,2)
pc2

Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8
0.194 0.271 0.367 0.311 0.379 0.468 0.367 0.575
Loadings:
   PC1  PC2
V1 0.90
V2 0.85
V3 0.80
V4      0.83
V5      0.79
V6      0.73
V7 0.80
V8      0.65
                 PC1   PC2
SS loadings    2.812 2.268
Proportion Var 0.352 0.284
Cumulative Var 0.352 0.635
```

Simple Structure

The original solution of a principal components or principal axes factor analysis is a set of vectors that best account for the observed covariance or correlation matrix, and where the components or factors account for progressively less and less variance. But such a solution, although maximally efficient in describing the data, is rarely easy to interpret. But what makes a structure easy to interpret? Thurstone's answer, *simple structure*, consists of five rules (Thurstone, 1947, p 335):

> *(1) Each row of the oblique factor matrix* $\mathbf{V}$ *should have at least one zero.*
> *(2) For each column p of the factor matrix* $\mathbf{V}$ *there should be a distinct set of r linearly independent tests whose factor loadings* v_{ip} *are zero.*
> *(3) For every pair of columns of* $\mathbf{V}$ *there should be several tests whose entries* v_{ip} *vanish in one column but not in the other.*
> *(4) For every pair of columns of* $\mathbf{V}$*, a large proportion of the tests should have zero entries in both columns. This applies to factor problems with four or five or more common factors.*
> *(5) For every pair of columns there should preferably be only a small number of tests with non-vanishing entries in both columns.*

Thurstone proposed to rotate the original solution to achieve simple structure.

Simple structure

A matrix is said to be *rotated* if it is multiplied by a matrix of orthogonal vectors that preserves the communalities of each variable. Just as the original matrix was orthogonal, so is the rotated solution. For two factors, the *rotation* matrix **T** will rotate the two factors θ radians in a counterclockwise direction.

$$T = \begin{pmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{pmatrix} \tag{14}$$

Generalizing equation 14 to larger matrices is straight forward:

$$T = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & cos(\theta) & \dots & sin(\theta) & \dots & 0 \\ \dots & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & -sin(\theta) & \dots & cos(\theta) & \dots & 0 \\ \dots & \dots & 0 & \dots & 0 & \dots & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix}. \tag{15}$$

Rotating to simple structure

When **F** is post-multiplied by **T**, **T** will rotate the i^{th} and j^{th} columns of **F** by θ radians in a counterclockwise direction.

$$\mathbf{F}_r = \mathbf{FT} \tag{16}$$

The `factor.rotate` function from the **psych** package will do this rotation for arbitrary angles (in degrees) for any pairs of factors. This is useful if there is a particular rotation that is desired. An entire package devoted to rotations is the *GPArotation* by Robert Jennrich (Jennrich, 2004).

Analytic Simple Structure

As pointed out by Carroll (1953) when discussing Thurstone's (1947) simple structure as a rotational criterion "it is obvious that there could hardly be any single mathematical expression which could embody all these characteristics." (p 24). Carroll's solution to this was to minimize the sum of the inner products of the squared (rotated) loading matrix. An alternative, discussed by Ferguson (1954) is to consider the *parsimony* of a group of n tests with r factors to be defined as the average parsimony of the individual tests (l_j) where

$$l_j = \sum_{m}^{r} a_{jm}^4 \tag{17}$$

(the squared communality) and thus the average parsimony is

$$l_. = n^{-1} \sum_{j}^{n} \sum_{m}^{r} a_{jm}^4$$

Rotation to parsimony

Parsimony as defined in equation 17 is a function of the variance as well as the mean of the squared loadings of a particular test on all the factors. For fixed communality h^2, it will be maximized if all but one loading is zero; a variable's parsimony will be maximal if one loading is 1.0 and the rest are zero. In path notation, parsimony is maximized if one and only one arrow is associated with a variable. This criterion, as well as the criterion of maximum variance taken over factors has been operationalized as the *quartimax* criterion by Neuhaus & Wrigley (1954). As pointed out by Kaiser (1958), the criterion can rotate towards a solution with one general factor, ignoring other, smaller factors.

Varimax and alternatives

If a general factor is not desired, an alternative measure of the parsimony of a factor, similar to equation 17 is to maximize the variance of the squared loadings taken over items instead of over factors. This, the *varimax* criterion was developed by Kaiser (1958) to avoid the tendency to yield a general factor. Both of these standard rotations as well as many others are available in the **GPArotation** package of rotations and transformations which uses the *Gradient Projection Algorithms* developed by Jennrich (2001, 2002, 2004).

Harmon 8 physical measures

```
> data(Harman23.cor)
> lower.mat(Harman23.cor$cov)
```

	heght	arm.s	forrm	lwr.l	weght	btr.d	chst.g	chst.w
height	1.00							
arm.span	0.85	1.00						
forearm	0.80	0.88	1.00					
lower.leg	0.86	0.83	0.80	1.00				
weight	0.47	0.38	0.38	0.44	1.00			
bitro.diameter	0.40	0.33	0.32	0.33	0.76	1.00		
chest.girth	0.30	0.28	0.24	0.33	0.73	0.58	1.00	
chest.width	0.38	0.42	0.34	0.36	0.63	0.58	0.54	1.00

Two solutions – loadings change, goodness of fits do not

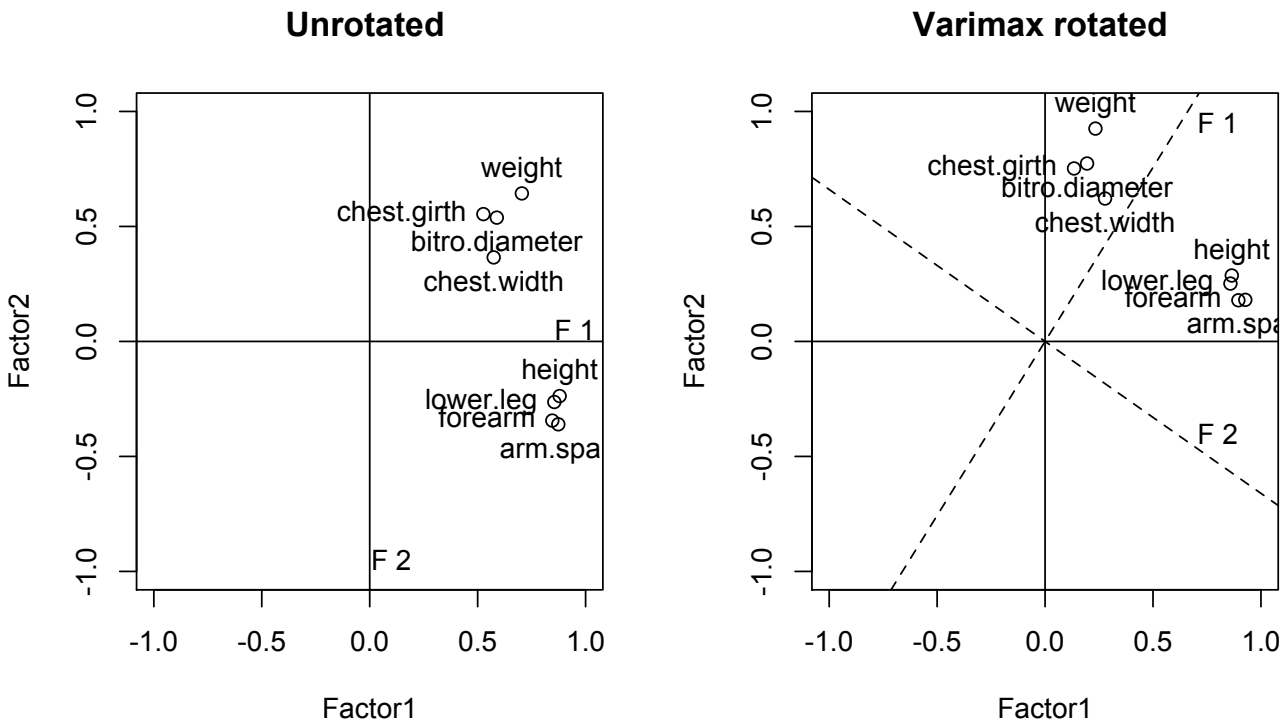
```
> f2 <- fa(Harman23.cor$cov,2,rotate="none")
> f2
```

```
Factor Analysis using method =  minres
Call: fa(r = Harman23.cor$cov, nfactors = 2,
    rotate = "none")
Standardized loadings (pattern matrix)
                MR1   MR2   h2   u2
height         0.89 -0.19 0.83 0.17
arm.span       0.89 -0.31 0.89 0.11
forearm        0.86 -0.30 0.83 0.17
lower.leg      0.87 -0.22 0.80 0.20
weight         0.67  0.67 0.89 0.11
bitro.diameter 0.56  0.58 0.65 0.35
chest.girth    0.50  0.59 0.59 0.41
chest.width    0.56  0.40 0.47 0.53

                MR1  MR2
SS loadings    4.40 1.56
Proportion Var 0.55 0.19
Cumulative Var 0.55 0.74
Test of the hypothesis that 2 factors are sufficient.
The root mean square of the residuals (RMSR) is  0.02
The df corrected root mean square of the residuals
             is  0.03
Fit based upon off diagonal values = 1
```

```
> f2 <- fa(Harman23.cor$cov,2,rotate="varimax")
> f2
```

```
Factor Analysis using method =  minres
Call: fa(r = Harman23.cor$cov, nfactors = 2,
     rotate = "varimax")
Standardized loadings (pattern matrix)
                MR1  MR2   h2   u2
height         0.86 0.30 0.83 0.17
arm.span       0.92 0.20 0.89 0.11
forearm        0.89 0.19 0.83 0.17
lower.leg      0.86 0.26 0.80 0.20
weight         0.22 0.92 0.89 0.11
bitro.diameter 0.18 0.78 0.65 0.35
chest.girth    0.12 0.76 0.59 0.41
chest.width    0.27 0.63 0.47 0.53

                MR1  MR2
SS loadings    3.30 2.66
Proportion Var 0.41 0.33
Cumulative Var 0.41 0.74

The root mean square of the residuals (RMSR)
     is   0.02
The df corrected root mean square of the residuals
     is  0.03
```

Alternative rotations



Oblique transfomations

Many of those who use factor analysis use it to identify theoretically meaningful constructs which they have no reason to believe are orthogonal. This has lead to the use of *oblique transformations* which allow the factors to be correlated. Although the term rotation is sometimes used for both *orthogonal* and *oblique* solutions, in the oblique case the factor matrix is not rotated so much as *transformed*.

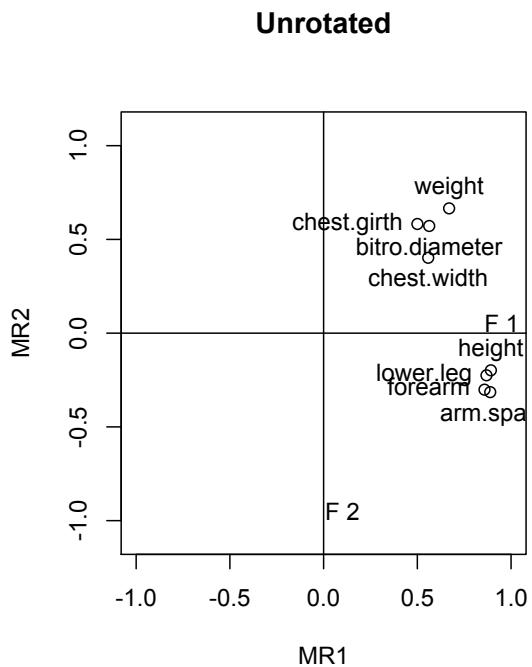
Oblique transformations lead to the distinction between the *factor pattern* and *factor structure* matrices. The *factor pattern* matrix is the set of *regression weights* (loadings) from the latent factors to the observed variables. The *factor structure* matrix is the matrix of *correlations* between the factors and the observed variables. If the factors are uncorrelated, structure and pattern are identical. But, if the factors are correlated, the structure matrix ($\mathbf{S}$) is the pattern matrix ($\mathbf{F}$) times the factor intercorrelations ϕ

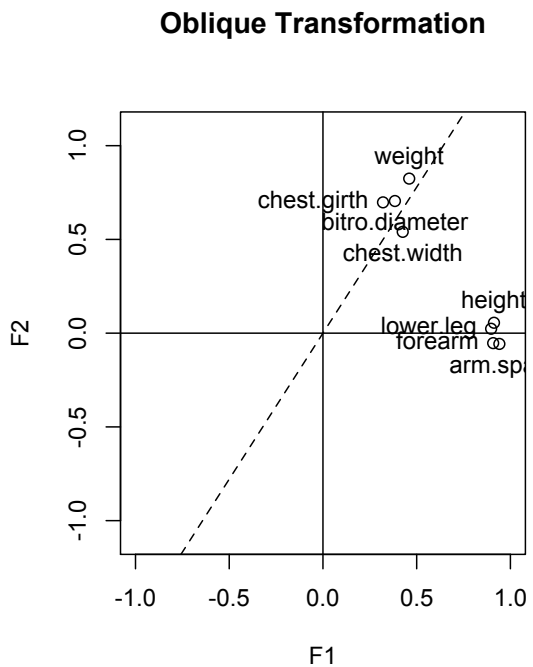
$\mathbf{S} = \mathbf{F}\phi <=> \mathbf{F} = \mathbf{S}\phi^{-1}$:

An oblique transformation of the Harman 8 physical variables

```
> f2t <- fa(Harman23.cor$cov,2,rotate="oblimin",n.obs=305)
> print(f2t)
Factor Analysis using method =  minres
Call: fa(r = Harman23.cor$cov, nfactors = 2, rotate = "oblimin", n.obs = 305)
                item   MR1   MR2   h2   u2
height             1  0.87  0.08 0.84 0.16
arm.span           2  0.96 -0.05 0.89 0.11
forearm            3  0.93 -0.04 0.83 0.17
lower.leg          4  0.88  0.04 0.81 0.19
weight             5  0.01  0.94 0.89 0.11
bitro.diameter     6  0.00  0.80 0.64 0.36
chest.girth        7 -0.06  0.79 0.59 0.41
chest.width        8  0.13  0.62 0.47 0.53

                 MR1  MR2
SS loadings     3.37 2.58
Proportion Var  0.42 0.32
Cumulative Var  0.42 0.74

 With factor correlations of
     MR1  MR2
MR1 1.00 0.46
MR2 0.46 1.00
```

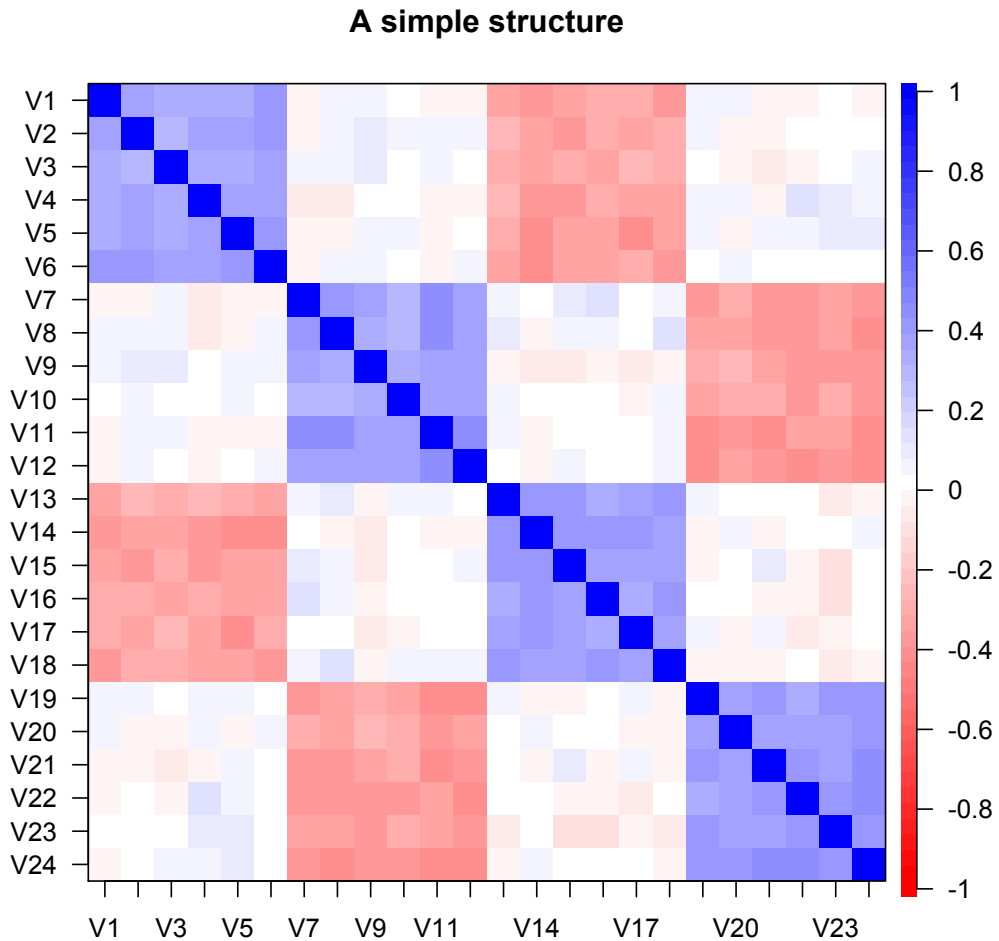
Oblique Transformations



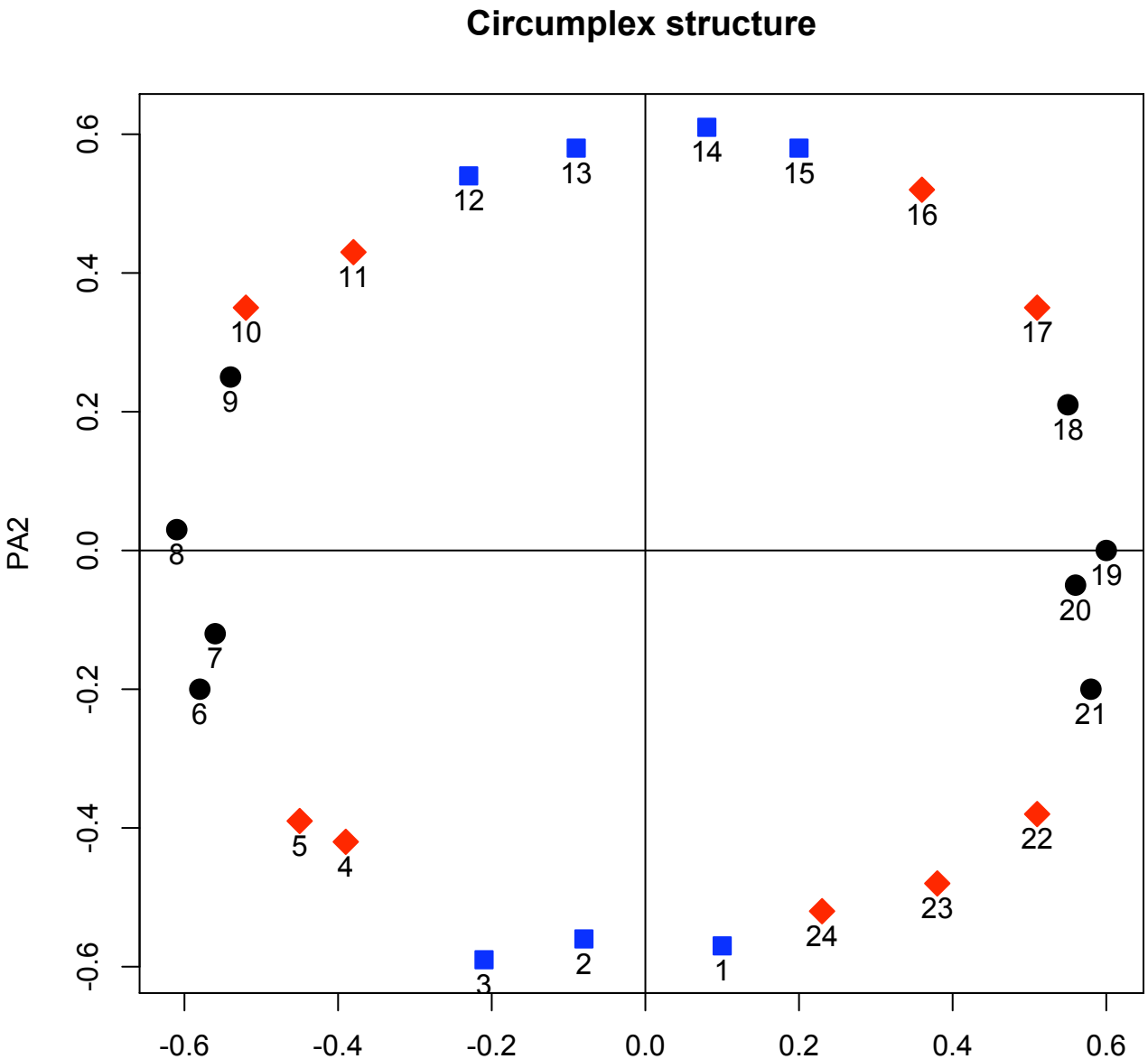


Another way to show simple structure

```
> simp24 <- sim.item(24,circum=FALSE)
> cor.plot(cor(simp24),main="A simple structure")
```

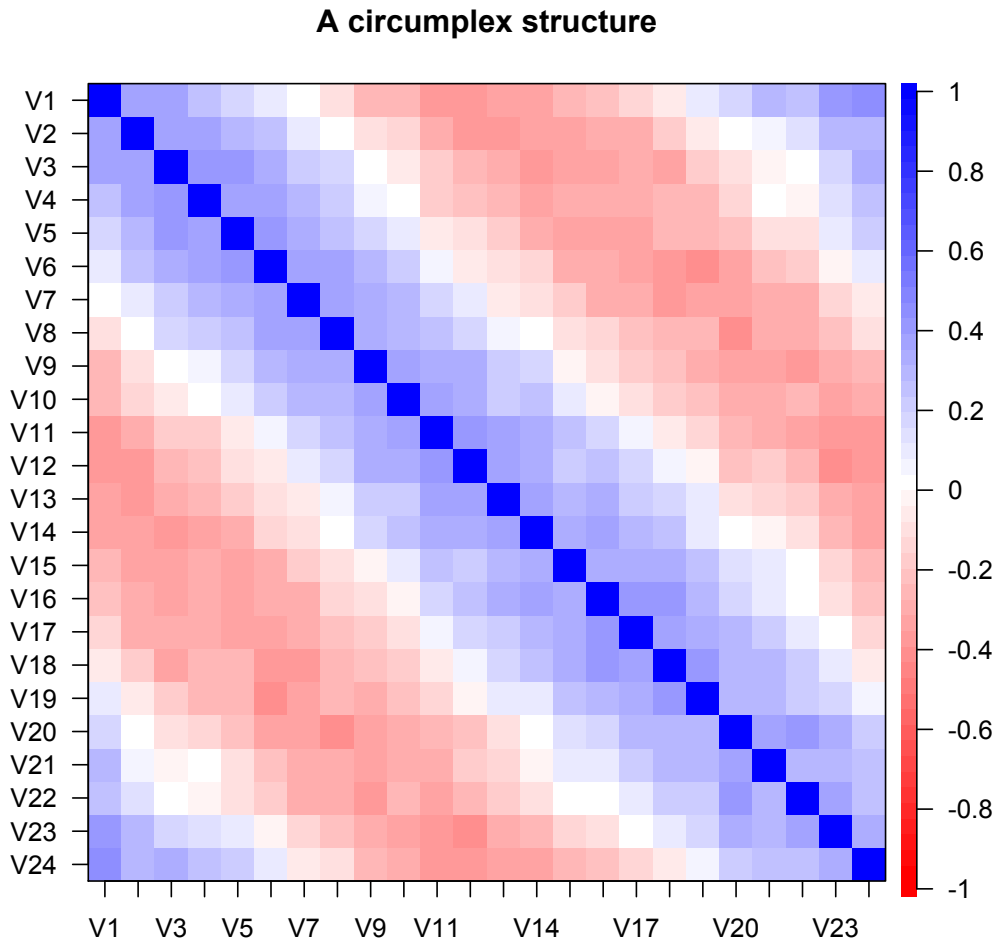


A circumplex is one alternative to simple structure



Another way of showing a circumplex – cor.plot

```
> circ24 <- sim.item(24,circum=TRUE)
> cor.plot(cor(circ24),main="A circumplex structure")
```



The Thurstone 9 variable problem

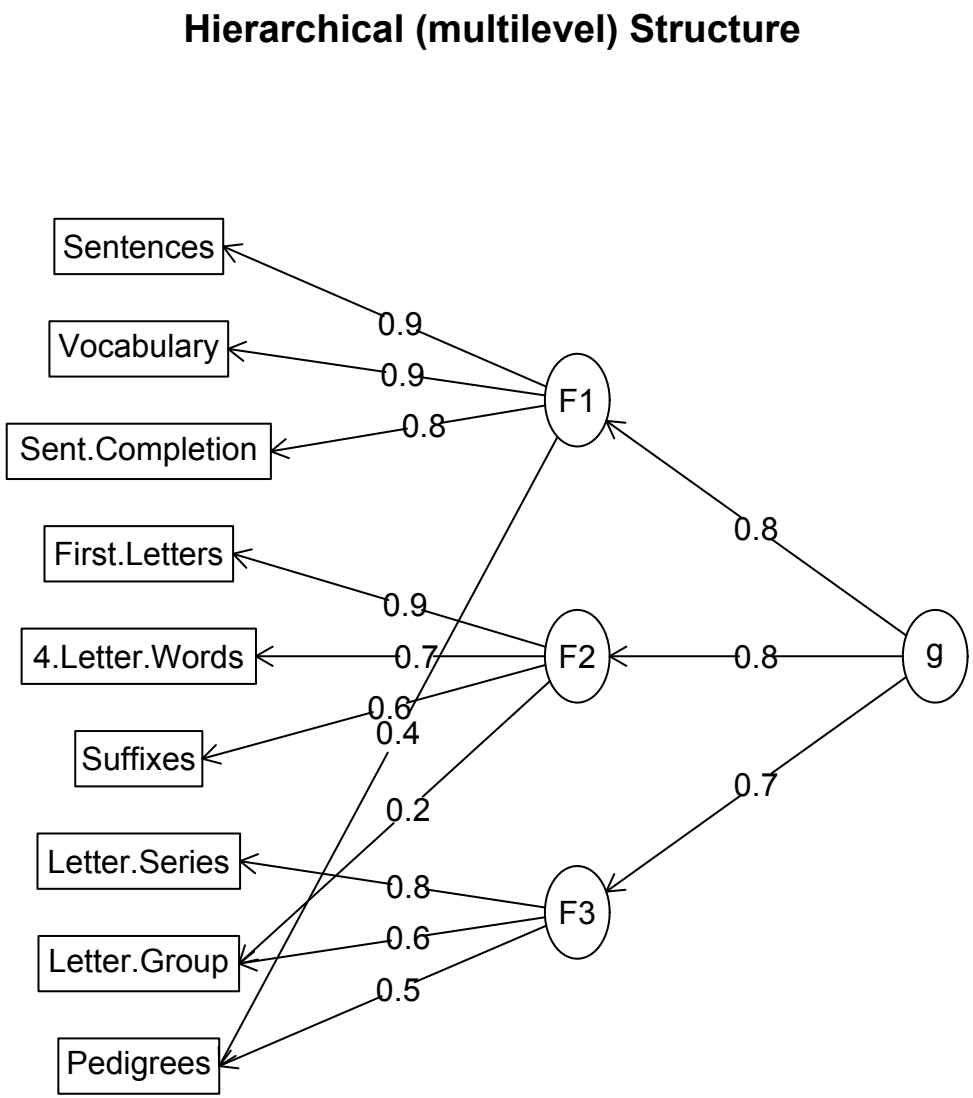
```
> lower.mat(Thurstone)
```

	Sntnc	Vcblr	Snt.C	Frs.L	4.L.W	Sffxs	Ltt.S	Pdgrs	Ltt.G
Sentences	1.00								
Vocabulary	0.83	1.00							
Sent.Completion	0.78	0.78	1.00						
First.Letters	0.44	0.49	0.46	1.00					
4.Letter.Words	0.43	0.46	0.42	0.67	1.00				
Suffixes	0.45	0.49	0.44	0.59	0.54	1.00			
Letter.Series	0.45	0.43	0.40	0.38	0.40	0.29	1.00		
Pedigrees	0.54	0.54	0.53	0.35	0.37	0.32	0.56	1.00	
Letter.Group	0.38	0.36	0.36	0.42	0.45	0.32	0.60	0.45	1.00

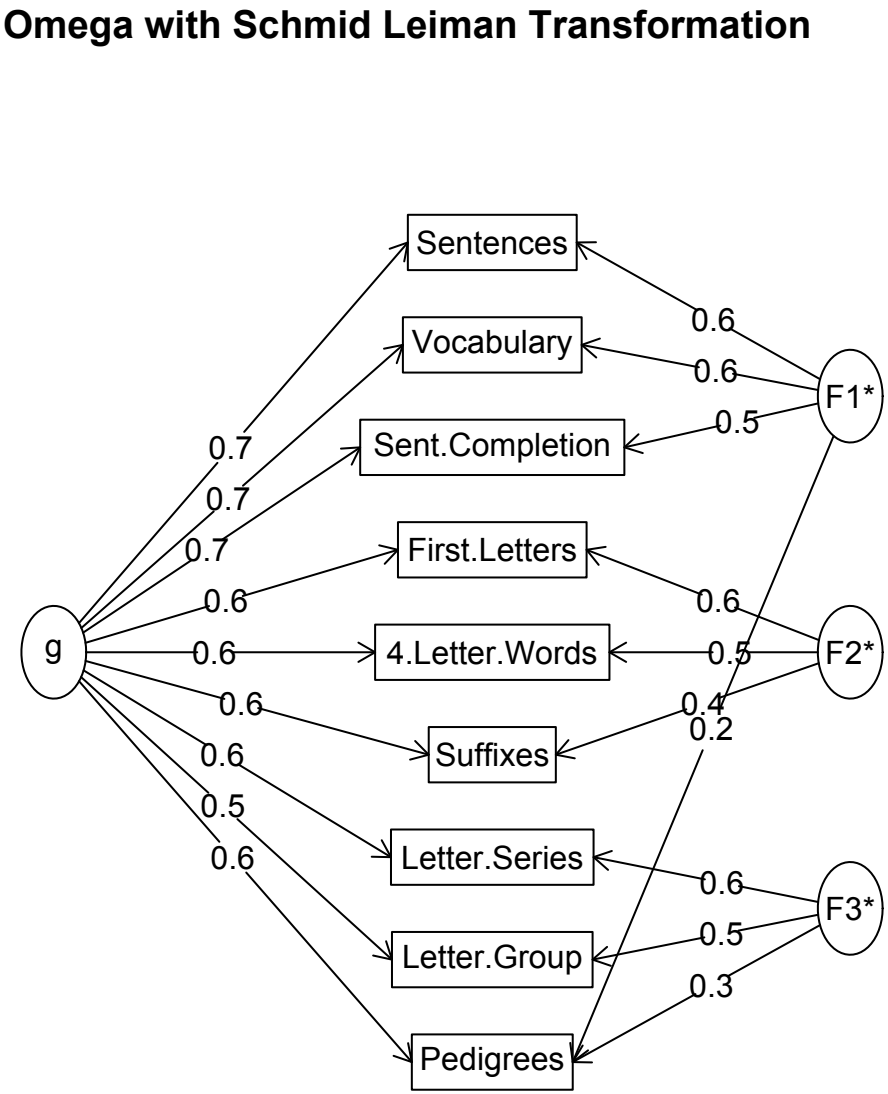
Three factors from Thurstone 9 variables

```
> f3 <- fa(Thurstone,3)
> f3


Factor Analysis using method =  minres
Call: fa(r = Thurstone, nfactors = 3)
Standardized loadings (pattern matrix) based upon correlation matrix
                   MR1   MR2   MR3   h2   u2
Sentences         0.91 -0.04  0.04 0.82 0.18
Vocabulary        0.89  0.06 -0.03 0.84 0.16
Sent.Completion   0.83  0.04  0.00 0.73 0.27
First.Letters     0.00  0.86  0.00 0.73 0.27
4.Letter.Words   -0.01  0.74  0.10 0.63 0.37
Suffixes          0.18  0.63 -0.08 0.50 0.50
Letter.Series     0.03 -0.01  0.84 0.72 0.28
Pedigrees         0.37 -0.05  0.47 0.50 0.50
Letter.Group     -0.06  0.21  0.64 0.53 0.47

                 MR1  MR2  MR3
SS loadings    2.64 1.86 1.50
Proportion Var 0.29 0.21 0.17
Cumulative Var 0.29 0.50 0.67

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.59 0.54
MR2 0.59 1.00 0.52
MR3 0.54 0.52 1.00
```

A hierarchical/multilevel solution to the Thurstone 9 variables



A bifactor solution using the Schmid Leiman transformation



How many factors – no right answer, one wrong answer

1. Statistical
   - Extracting factors until the χ^2 of the residual matrix is not significant.
   - Extracting factors until the change in χ^2 from factor n to factor n+1 is not significant.
2. Rules of Thumb
   - Parallel Extracting factors until the eigenvalues of the real data are less than the corresponding eigenvalues of a random data set of the same size (*parallel analysis*)
   - Plotting the magnitude of the successive eigenvalues and applying the *scree test*.
3. Interpretability
   - Extracting factors as long as they are interpretable.
   - Using the *Very Simple Structure* Criterion (`VSS`)
   - Using the Minimum Average Partial criterion (MAP).
4. Eigen Value of 1 rule

Bartlett, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika*, *38*(3/4), 337–344.

Carroll, J. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, *18*(1), 23–38.

Cattell, R. B. (1965). A biometrics invited paper. factor analysis: An introduction to essentials i. the purpose and underlying models. *Biometrics*, *21*(1), 190–215.

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*(3), 211–218.

Ferguson, G. (1954). The concept of parsimony in factor analysis. *Psychometrika*, *19*(4), 281–290.

Householder, A. S. & Young, G. (1938). Matrix approximation and latent roots. *The American Mathematical Monthly*, *45*(3), 165–171.

Jennrich, R. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, *66*(2), 289–306.

Jennrich, R. (2002). A simple general method for oblique rotation. *Psychometrika*, *67*(1), 7–19.

Jennrich, R. (2004). Derivative free gradient projection algorithms for rotation. *Psychometrika*, *69*(3), 475–480.

Jöreskog, K. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4), 443–477.

Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200.

Lawley, D. N. & Maxwell, A. E. (1962). Factor analysis as a statistical method. *The Statistician*, *12*(3), 209–229.

Lawley, D. N. & Maxwell, A. E. (1963). *Factor analysis as a statistical method*. London: Butterworths.

Neuhaus, J. & Wrigley, C. (1954). The quartimax method: an analytical approach to orthogonal simple structure. *British Journal of Statistical Psychology*, *7*, 81–91.

Roff, M. (1936). Some properties of the communality in multiple factor theory. *Psychometrika*, *1*(2), 1–6.

Spearman, C. (1904). "general intelligence," objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–292.

Thurstone, L. L. (1947). *Multiple-factor analysis: a development and expansion of The vectors of the mind*. Chicago, Ill.: The University of Chicago Press.