An introduction to Psychometric Theory with applications in R
Exploratory vs. confirmatory analysis

William Revelle

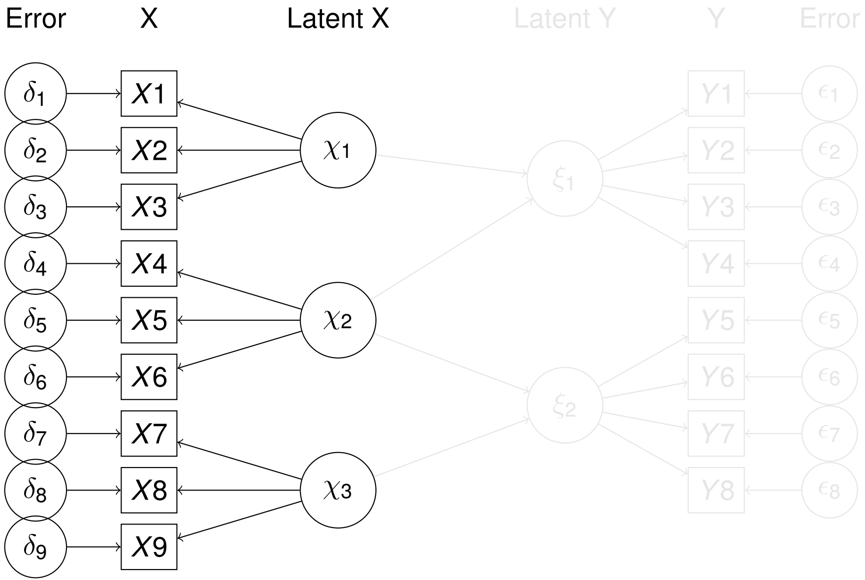
Department of Psychology
Northwestern University
Evanston, Illinois USA

NORTHWESTERN UNIVERSITY

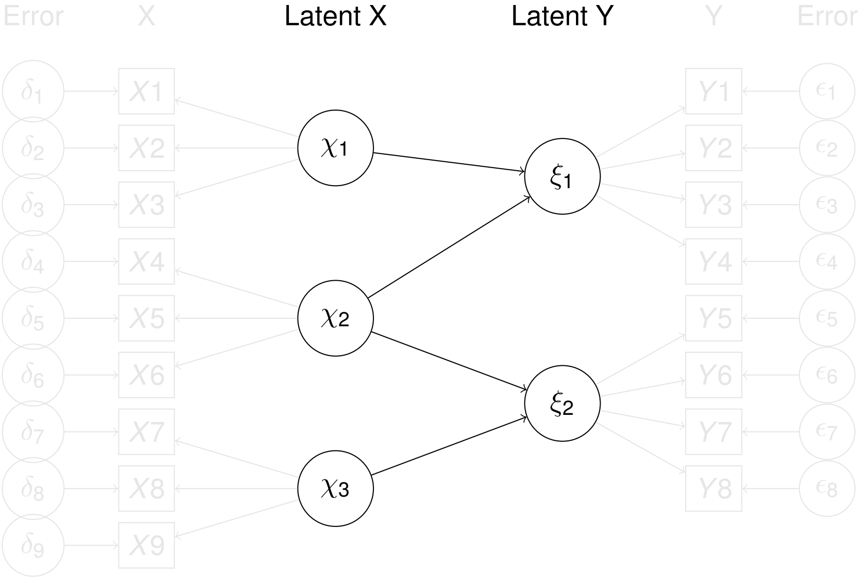
May, 2025

Outline

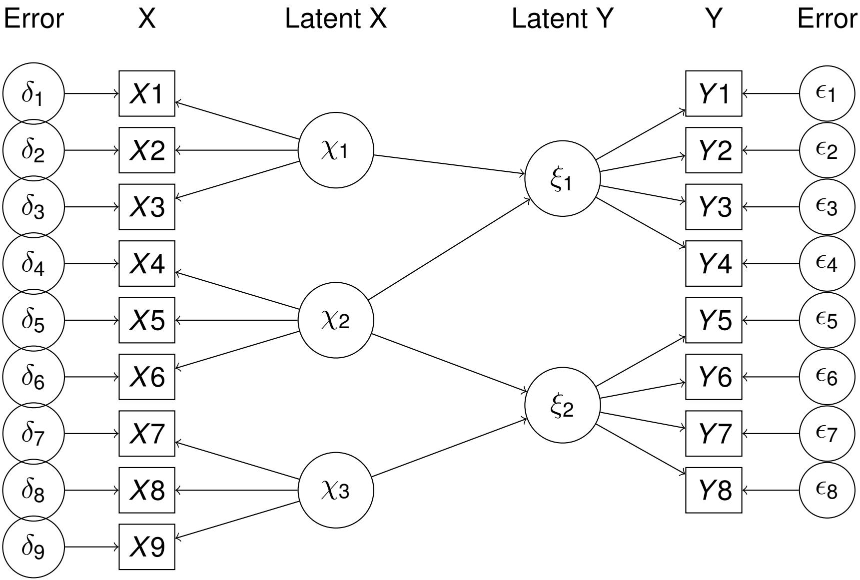
Measurement: A latent variable approach.



Theory


Error
X
Latent X
Latent Y
Y
Error
δ_1
δ_2
δ_3
δ_4
δ_5
δ_6
δ_7
δ_8
δ_9
$X1$
$X2$
$X3$
$X4$
$X5$
$X6$
$X7$
$X8$
$X9$
χ_1
χ_2
χ_3
ξ_1
ξ_2
$Y1$
$Y2$
$Y3$
$Y4$
$Y5$
$Y6$
$Y7$
$Y8$
ϵ_1
ϵ_2
ϵ_3
ϵ_4
ϵ_5
ϵ_6
ϵ_7
ϵ_8

Psychometric Theory: A conceptual Syllabus



Two types of variables, three types of relationships

1. Variables
   1.1 Observed Variables (X, Y)
   1.2 Latent Variables (ξ η ϵ ζ)
2. Three kinds of variance/covariances
   2.1 Observed with Observed C_{xy} or σ_{xy}
   2.2 Observed with Latent λ
   2.3 Latent with Latent ϕ
3. Direction
   - Bidirectional (correlation)
   - Directional (regression)

Observed Variables

X Y

X_1 Y_1

X_2 Y_2

X_3 Y_3

X_4 Y_4

X_5 Y_5

X_6 Y_6

Latent Variables

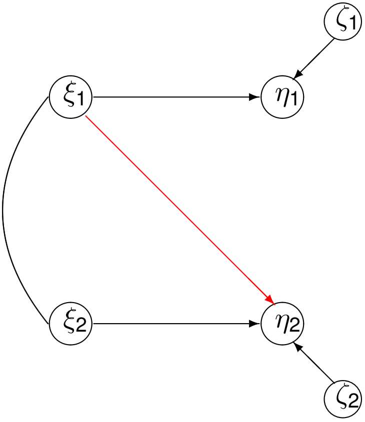
ξ η

ξ_1 η_1

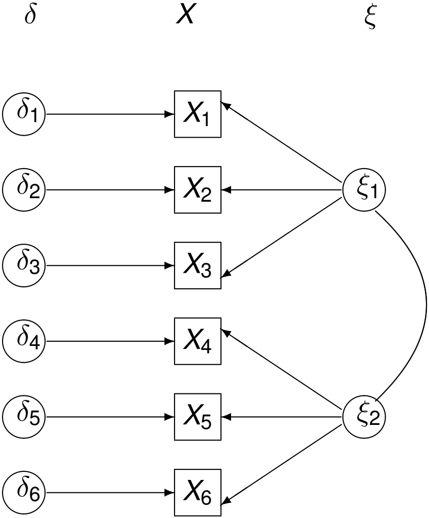
ξ_2 η_2

Theory: A regression model of latent variables

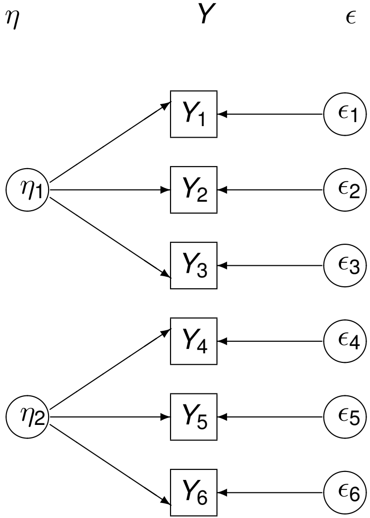
ξ η



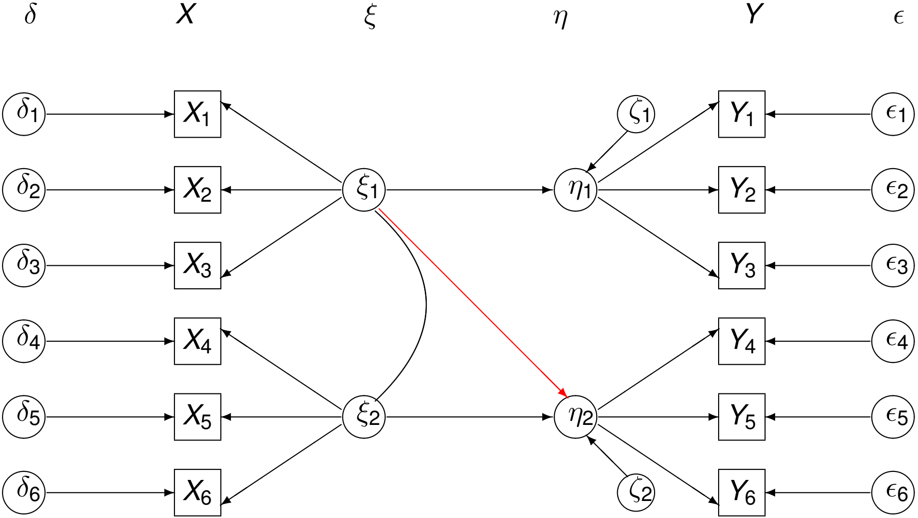
A measurement model for X



A measurement model for Y



A complete structural model



Latent Variable Modeling

1. Requires measuring observed variables
   - Requires defining what is relevant and irrelevant to our theory.
   - Issues in quality of scale information, levels of measurement.
2. Formulating a measurement model of the data: estimating latent constructs
   - Perhaps based upon exploratory and then confirmatory factor analysis, definitely based upon theory.
   - Includes understanding the reliability of the measures.
3. Modeling the structure of the constructs
   - This is a combination of theory and fitting. Do the data fit the theory.
   - Comparison of models. Does one model fit better than alternative models?

Two fundamentally different types of observed variables

1. Observed variables can be "reflective" of the latent variables. They are "effect indicators".
   - Variables are caused by the latent variables.
   - Covariation between the variables are explained by the latent variables
2. Observed variables can be "causal indicators" or formative indicators that can directly effect the latent variable
   - Variables cause the "latent" variable
   - Covariation of the the observed variables is not modeled

See Loevinger (1957) for a thoughtful introduction to the concept items as signs or samples.

Formative indicators Bollen (2002)

1. The correlational structure of formative indicators is independent of the loadings on a factor.
    - They are not locally independent
    - (Correlations remain even when removing factor structure.)
2. Examples of formative indicators
    - Time spent in social interaction
        - Time spent with family, time spent with friends, time spent with coworkers.
        - These might in fact be negatively correlated even though total score is important.
    - Socio-economic status
        - Education
        - Occupation
        - Income
        - Sum of variables is the index, but separate elements may be independent

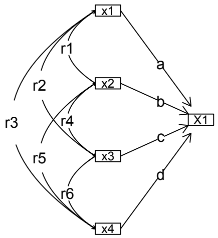
See also Loevinger (1957) for this distinction between signs and samples.

Effect (reflective) indicators

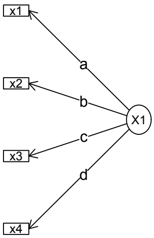
1. Test scores on various quantitative tests as effect indicators of trait
   - Feelings of self worth as effect indicators of self esteem
   - Reports of anxiety and depression as indicators of Neuroticism
   - Ability items as indicators of ability
2. Correlational structure is a function of the path coefficients with latent variables
3. Variables are locally independent
   - (uncorrelated with each other when latent variable is partialed out)

Formative vs. Effect variables as regression versus factors

Formative Variables



Effect Variables



Confirmatory Factor Analysis as a way of testing factor models

1. Exploratory factor analysis (EFA) is a way of summarizing the relationships between variables
   - May have goodness of fit of factor model to the covariances but this is unusual but becoming more common
   - All loadings are modeled
   - For n variables and p factors, the df = n*(n-1)/2 - n*p + p*(p-1)/2
   - Rotations/Transformations do not affect fit
2. Confirmatory factor analysis (CFA) also models the data
   - But limits paths to k specified paths, sets other to zero
   - Rotation/Transformations not done.
   - df = n*(n-1)/2 - k
   - Goodness of fit statistics are always produced
3. Blending CFA with regression is a full Structural Equation Model

CFA allows us to test specific measurement models

1. EFA fits correlations or covariances with loadings for every item on every factor and thus
   - for an n x n correlation matrix with p factors
   - total df = n * (n-1)/2 (the number of correlations)
   - df used in model = n * p - (p*(p-1)/2)
   - df is thus n * (n-1)/2 - n * p + (p*(p-1)/2)
2. CFA typically is a simple structure model and estimate just n parameters (one per item)
   - df = n * (n-1)/2 - n
3. The question is how much better fit do you get by forcing a “cluster structure" onto the items?
4. Lets compare EFA and CFA for some classic data sets

The Thurstone data set

R code

```
lowerMat(Thurstone)
```

```
                Sntnc Vcblr Snt.C Frs.L F.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
Four.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
```

Thurstone: 3 oblique factors using Maximum Likelihood

R code

```
f3ml <- fa(Thurstone,3,fm="mle",n.obs=213)
```

```
Factor Analysis using method =  ml
Call: fa(r = Thurstone, nfactors = 3, n.obs = 213, fm = "mle")
Standardized loadings (pattern matrix) based upon correlation matrix
                   ML1   ML2   ML3   h2   u2 com
Sentences         0.91 -0.04  0.04 0.83 0.17 1.0
Vocabulary        0.89  0.06 -0.03 0.84 0.16 1.0
Sent.Completion   0.83  0.04  0.00 0.73 0.27 1.0
First.Letters     0.00  0.86  0.01 0.73 0.27 1.0
Four.Letter.Words -0.01  0.74  0.10 0.63 0.37 1.0
Suffixes          0.18  0.63 -0.08 0.50 0.50 1.2
Letter.Series     0.03 -0.01  0.84 0.72 0.28 1.0
Pedigrees         0.37 -0.05  0.47 0.50 0.50 1.9
Letter.Group     -0.06  0.21  0.64 0.53 0.47 1.2

                       ML1  ML2  ML3
SS loadings           2.64 1.86 1.49
Proportion Var        0.29 0.21 0.17
Cumulative Var        0.29 0.50 0.67
Proportion Explained  0.44 0.31 0.25
Cumulative Proportion 0.44 0.75 1.00

 With factor correlations of
     ML1  ML2  ML3
ML1 1.00 0.59 0.54
ML2 0.59 1.00 0.52
ML3 0.54 0.52 1.00
```

MLE continued

R code

```
f3ml <- fa(Thurstone,3,fm="mle",n.obs=213)
```

```
The degrees of freedom for the null model are  36  and the objective function was  5.1976 wit
The degrees of freedom for the model are 12  and the objective function was  0.0137

The root mean square of the residuals (RMSR) is  0.0061
The df corrected root mean square of the residuals is  0.0106

The harmonic number of observations is  213 with the empirical chi square  0.5756  with prob
The total number of observations was  213  with Likelihood Chi Square =  2.8233  with prob <

Tucker Lewis Index of factoring reliability =  1.02658
RMSEA index =  0  and the 90 % confidence intervals are  0 0
BIC =  -61.5122
Fit based upon off diagonal values = 0.9998
Measures of factor score adequacy
                                                  ML1    ML2    ML3
Correlation of (regression) scores with factors   0.9640 0.9235 0.9022
Multiple R square of scores with factors          0.9293 0.8528 0.8140
Minimum correlation of possible factor scores     0.8586 0.7056 0.6281
```

Comparing multiple approaches: minres

R code

```
f3 <- fa(Thurstone,3, n.obs=231)
```

```
Factor Analysis using method =  minres
Call: fa(r = Thurstone, nfactors = 3, n.obs = 213)
Standardized loadings (pattern matrix) based upon correlation matrix
                  MR1   MR2   MR3   h2   u2 com
Sentences         0.90 -0.03  0.04 0.82 0.18 1.0
Vocabulary        0.89  0.06 -0.03 0.84 0.16 1.0
Sent.Completion   0.84  0.03  0.00 0.74 0.26 1.0
First.Letters     0.00  0.85  0.00 0.73 0.27 1.0
Four.Letter.Words -0.02  0.75  0.10 0.63 0.37 1.0
Suffixes          0.18  0.63 -0.08 0.50 0.50 1.2
Letter.Series     0.03 -0.01  0.84 0.73 0.27 1.0
Pedigrees         0.38 -0.05  0.46 0.51 0.49 2.0
Letter.Group     -0.06  0.21  0.63 0.52 0.48 1.2

                       MR1  MR2  MR3
SS loadings           2.65 1.87 1.49
Proportion Var        0.29 0.21 0.17
Cumulative Var        0.29 0.50 0.67
Proportion Explained  0.44 0.31 0.25
Cumulative Proportion 0.44 0.75 1.00

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.59 0.53
MR2 0.59 1.00 0.52
MR3 0.53 0.52 1.00

Mean item complexity =  1.2
Test of the hypothesis that 3 factors are sufficient.
```

Thurstone, minres continued

R code

```
f3 <- fa(Thurstone,3, n.obs=231)
```

```
...

Mean item complexity =  1.2
Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are  36  and the objective function was  5.1976 wit
The degrees of freedom for the model are 12  and the objective function was  0.0144

The root mean square of the residuals (RMSR) is  0.0058
The df corrected root mean square of the residuals is  0.0101

The harmonic number of observations is  213 with the empirical chi square  0.5233  with prob 
The total number of observations was  213  with Likelihood Chi Square =  2.9772  with prob <

Tucker Lewis Index of factoring reliability =  1.02614
RMSEA index =  0  and the 90 % confidence intervals are  0 0
BIC =  -61.3583
Fit based upon off diagonal values = 0.9999
Measures of factor score adequacy
                                                  MR1    MR2    MR3
Correlation of (regression) scores with factors   0.9635 0.9233 0.9032
Multiple R square of scores with factors          0.9283 0.8526 0.8157
Minimum correlation of possible factor scores     0.8566 0.7051 0.6314
```

Comparing solutions

Both minres and mle give χ^2 statistics. Although the dfs are identical we can still compare them.

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

This χ^2 has degrees of freedom:

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

R code

```
anova(f3,f3.ml)   #to compare the solutions
```

```
# compare the two solutions
Model 1 = fa(r = Thurstone, nfactors = 3, n.obs = 213)
Model 2 = fa(r = Thurstone, nfactors = 3, n.obs = 213, fm = "mle")
  df d.df chiSq d.chiSq PR test empirical d.empirical test.echi    BIC d.BIC
1 12   NA  2.98      NA NA   NA      0.52          NA        NA -61.36    NA
2 12    0  2.82    0.15  0 -Inf      0.58       -0.05       Inf -61.51 -0.15
```

Force orthogonality for efa

R code

```
f3 <- fa(Thurstone,3,n.obs=213,rotate="varimax",fm="mle")
```

```
Factor Analysis using method =  ml
Call: fa(r = Thurstone, nfactors = 3, n.obs = 213, rotate = "varimax",
    fm = "mle")
Standardized loadings (pattern matrix) based upon correlation matrix
                   ML1  ML2  ML3   h2   u2 com
Sentences         0.83 0.24 0.26 0.83 0.17 1.4
Vocabulary        0.83 0.32 0.22 0.84 0.16 1.4
Sent.Completion   0.77 0.28 0.23 0.73 0.27 1.5
First.Letters     0.23 0.79 0.23 0.73 0.27 1.3
Four.Letter.Words 0.21 0.71 0.29 0.63 0.37 1.5
Suffixes          0.31 0.62 0.13 0.50 0.50 1.6
Letter.Series     0.23 0.18 0.80 0.72 0.28 1.3
Pedigrees         0.45 0.17 0.53 0.50 0.50 2.2
Letter.Group      0.15 0.31 0.64 0.53 0.47 1.6

                       ML1  ML2  ML3
SS loadings           2.45 1.90 1.64
Proportion Var        0.27 0.21 0.18
Cumulative Var        0.27 0.48 0.67
Proportion Explained  0.41 0.32 0.27
Cumulative Proportion 0.41 0.73 1.00

The degrees of freedom for the null model are  36  and the objective function was  5.2
      with Chi Square of  1081.97

The total number of observations was  213  with Likelihood Chi Square =  2.82  with prob <  1
```

Varimax rotation of 3 factors (continued)

R code

```
f3 <- fa(Thurstone,3,n.obs=213,rotate="varimax",fm="mle")
```

```
Tucker Lewis Index of factoring reliability =  1.027
RMSEA index =  0  and the 90 % confidence intervals are  0 0
BIC =  -61.51
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                 ML1  ML2  ML3
Correlation of (regression) scores with factors   0.93 0.87 0.86
Multiple R square of scores with factors          0.86 0.76 0.73
Minimum correlation of possible factor scores     0.73 0.52 0.46

#compare with oblique
RMSEA index =  0  and the 90 % confidence intervals are  0 0
BIC =  -61.51
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                 ML1  ML2  ML3
Correlation of (regression) scores with factors   0.96 0.92 0.90
Multiple R square of scores with factors          0.93 0.85 0.81
Minimum correlation of possible factor scores     0.86 0.71 0.63
```

Confirmatory approaches

1. EFA estimates all model parameters
2. CFA estimates a reduced set of parameters (“Theory driven") to see how well this reduced set fits.
3. Original program was LISREL (Joreskog & Sorbom, 1993)
4. Proliferation of programs (Amos, SAS PROC CALIS, R packages sem, lavaan, OpenMx, LISREL, EQS, and Mplus) (Narayanan, 2012)
5. Open source packages in R (R Core Team, 2022) are sem (Fox, Nie & Byrnes, 2013), lavaan (Rosseel, 2012) and OpenMx (Boker, Neale, Maes, Wilde, Spiegel, Brick, Spies, Estabrook, Kenny, Bates, Mehta & Fox, 2011)

lavaan model

R code

```
model <-  'L1 =~ Sentences + Vocabulary + Sent.Completion
           L2 =~ First.Letters + Four.Letter.Words + Suffixes
           L3 =~ Letter.Series + Pedigrees + Letter.Group
library(lavaan)
cfa.thu <-cfa(model=model,sample.cov=Thurstone,sample.nobs=231,std.lv=
summary(cfa.thu)
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  L1 =~
    Sentences         0.903    0.052   17.419    0.000
    Vocabulary        0.912    0.051   17.709    0.000
    Sent.Completin    0.854    0.054   15.951    0.000
  L2 =~
    First.Letters     0.834    0.058   14.353    0.000
    Four.Lttr.Wrds    0.795    0.059   13.473    0.000
    Suffixes          0.701    0.061   11.414    0.000
  L3 =~
    Letter.Series     0.779    0.061   12.677    0.000
    Pedigrees         0.719    0.063   11.453    0.000
    Letter.Group      0.702    0.063   11.121    0.000
Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  L1 ~~
    L2                0.643    0.048   13.283    0.000
    L3                0.670    0.049   13.698    0.000
  L2 ~~
    L3                0.637    0.056   11.351    0.000
```

lavaan forced orthogonal

R code

```
cfa.thu.orth <-cfa(model=model,sample.cov=Thurstone,sample.nobs=231,
        orthogonal=TRUE,std.lv=TRUE)
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  L1 =~
    Sentences         0.906    0.052   17.416    0.000
    Vocabulary        0.910    0.052   17.523    0.000
    Sent.Completin    0.853    0.054   15.855    0.000
  L2 =~
    First.Letters     0.855    0.061   13.944    0.000
    Four.Lttr.Wrds    0.784    0.062   12.641    0.000
    Suffixes          0.687    0.063   10.914    0.000
  L3 =~
    Letter.Series     0.855    0.068   12.636    0.000
    Pedigrees         0.646    0.067    9.673    0.000
    Letter.Group      0.696    0.067   10.394    0.000

 anova(cfa.thu,cfa.thu.orth)
Chi-Squared Difference Test

             Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
cfa.thu      24 4773.9 4846.2  41.62
cfa.thu.orth 27 4962.1 5024.1 235.82      194.2       3  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Count the df = 9*8/2 - 9 = 27 vs 9*8/2 - 9 -3 = 24

Comparing cfa and efa

1. efa fits by using more parameters
2. efa fit statistic does not vary by rotation (although correlation with factors does)
3. cfa fits using fewer parameters
4. cfa fit statistic does vary by solution
5. cfa allows for a comparison of model fits

Compare exploratory omega with confirmatory omega

1. the `omega` function is an exploratory approach
2. the `omegaSem` first finds the exploratory solution and then develops a cfa model based upon that
3. call *lavaan* to do the cfa and the organizes it for omega output
4. These give somewhat different solutions

Exploratory omega

R code

```
om <- omega(Thurstone)
```

```
Alpha:                 0.89
G.6:                   0.91
Omega Hierarchical:    0.74
Omega H asymptotic:    0.79
Omega Total            0.93
Schmid Leiman Factor loadings greater than  0.2
                    g   F1*   F2*   F3*   h2   u2   p2
Sentences         0.71  0.56              0.82 0.18 0.61
Vocabulary        0.73  0.55              0.84 0.16 0.63
Sent.Completion   0.68  0.52              0.74 0.26 0.63
First.Letters     0.65        0.56        0.73 0.27 0.57
Four.Letter.Words 0.62        0.49        0.63 0.37 0.61
Suffixes          0.56        0.41        0.50 0.50 0.63
Letter.Series     0.59              0.62  0.73 0.27 0.48
Pedigrees         0.58  0.24        0.34  0.51 0.49 0.66
Letter.Group      0.54              0.46  0.52 0.48 0.56
With eigenvalues of:
   g  F1*  F2*  F3*
3.58 0.96 0.74 0.72
Measures of factor score adequacy
                                                g  F1*  F2*  F3*
Correlation of scores with factors           0.86 0.73 0.72 0.75
Multiple R square of scores with factors     0.74 0.54 0.51 0.57
Minimum correlation of factor score estimates 0.49 0.07 0.03 0.13
 Total, General and Subset omega for each subset
                                                g  F1*  F2*  F3*
Omega total for total scores and subscales    0.93 0.92 0.83 0.79
Omega general for total scores and subscales  0.74 0.58 0.50 0.47
Omega group for total scores and subscales    0.16 0.34 0.32 0.32
```

omegaSem calls omega and then call lavaan

R code

```
om.sem <- omegaSem(Thurstone,n.obs=231)
```

```
Omega Hierarchical from a confirmatory model using sem =  0.79
 Omega Total  from a confirmatory model using sem =  0.93
With loadings of
                     g  F1*  F2*  F3*   h2   u2   p2
Sentences         0.77 0.49           0.82 0.18 0.72
Vocabulary        0.79 0.45           0.83 0.17 0.75
Sent.Completion   0.75 0.40           0.73 0.27 0.77
First.Letters     0.61      0.61      0.74 0.26 0.50
Four.Letter.Words 0.60      0.50      0.61 0.39 0.59
Suffixes          0.57      0.39      0.48 0.52 0.68
Letter.Series     0.57           0.73 0.85 0.15 0.38
Pedigrees         0.66           0.25 0.50 0.50 0.87
Letter.Group      0.53           0.41 0.45 0.55 0.62
With eigenvalues of:
   g  F1*  F2*  F3*
3.86 0.60 0.78 0.75
The degrees of freedom of the confirmatory model are  18  and the fit is  26.38663  with p =
general/max  4.92   max/min =   1.3
mean percent general =  0.65    with sd =  0.15 and cv of  0.23
Explained Common Variance of the general factor =  0.64
Measures of factor score adequacy
                                                 g   F1*  F2*  F3*
Correlation of scores with factors            0.90  0.68 0.80 0.85
Multiple R square of scores with factors      0.81  0.46 0.63 0.73
Minimum correlation of factor score estimates 0.61 -0.08 0.27 0.45
 Total, General and Subset omega for each subset
                                                 g  F1*  F2*  F3*
Omega total for total scores and subscales    0.93 0.92 0.82 0.80
Omega general for total scores and subscales  0.79 0.69 0.48 0.50
Omega group for total scores and subscales    0.14 0.23 0.35 0.31
```

Why the difference?

1. Exploratory omega allows for cross loadings
2. Confirmatory omega eliminates cross loadings and thus forces the variance onto the general factor

R code
```
print(om,cut=0)
```

```
Schmid Leiman Factor loadings greater than  0
                     g   F1*   F2*   F3*   h2   u2   p2
Sentences         0.71  0.56 -0.02  0.03 0.82 0.18 0.61
Vocabulary        0.73  0.55  0.04 -0.02 0.84 0.16 0.63
Sent.Completion   0.68  0.52  0.02  0.00 0.74 0.26 0.63
First.Letters     0.65  0.00  0.56  0.00 0.73 0.27 0.57
Four.Letter.Words 0.62 -0.01  0.49  0.08 0.63 0.37 0.61
Suffixes          0.56  0.11  0.41 -0.06 0.50 0.50 0.63
Letter.Series     0.59  0.02 -0.01  0.62 0.73 0.27 0.48
Pedigrees         0.58  0.24 -0.03  0.34 0.51 0.49 0.66
Letter.Group      0.54 -0.04  0.14  0.46 0.52 0.48 0.56
```

Show the cross loadings from exploratory omega

R code

```
print(om,cut=0); print(om.sem,cut=0)
```

```
Schmid Leiman Factor loadings greater than  0
                     g   F1*   F2*   F3*   h2   u2   p2
Sentences         0.71  0.56 -0.02  0.03 0.82 0.18 0.61
Vocabulary        0.73  0.55  0.04 -0.02 0.84 0.16 0.63
Sent.Completion   0.68  0.52  0.02  0.00 0.74 0.26 0.63
First.Letters     0.65  0.00  0.56  0.00 0.73 0.27 0.57
Four.Letter.Words 0.62 -0.01  0.49  0.08 0.63 0.37 0.61
Suffixes          0.56  0.11  0.41 -0.06 0.50 0.50 0.63
Letter.Series     0.59  0.02 -0.01  0.62 0.73 0.27 0.48
Pedigrees         0.58  0.24 -0.03  0.34 0.51 0.49 0.66
Letter.Group      0.54 -0.04  0.14  0.46 0.52 0.48 0.56

Omega Hierarchical from a confirmatory model using sem =  0.79
 Omega Total  from a confirmatory model using sem =  0.93
With loadings of
                     g  F1*  F2*  F3*   h2   u2   p2
Sentences         0.77 0.49 0.00 0.00 0.82 0.18 0.72
Vocabulary        0.79 0.45 0.00 0.00 0.83 0.17 0.75
Sent.Completion   0.75 0.40 0.00 0.00 0.73 0.27 0.77
First.Letters     0.61 0.00 0.61 0.00 0.74 0.26 0.50
Four.Letter.Words 0.60 0.00 0.50 0.00 0.61 0.39 0.59
Suffixes          0.57 0.00 0.39 0.00 0.48 0.52 0.68
Letter.Series     0.57 0.00 0.00 0.73 0.85 0.15 0.38
Pedigrees         0.66 0.00 0.00 0.25 0.50 0.50 0.87
Letter.Group      0.53 0.00 0.00 0.41 0.45 0.55 0.62
```

Simulating structures to see how models work

1. It is always useful to create "toy" data sets with known structures to demonstrate alternative procedures.
2. There are many functions in *psych* to generate such structures.

Using sim.structural to create models

R code

```
fx <-matrix(c( .9,.8,.6,rep(0,4),.6,.8,-.7),ncol=2)
fy <- matrix(c(.6,.5,.4),ncol=1)
rownames(fx) <- c("V","Q","A","nach","Anx")
rownames(fy)<- c("gpa","Pre","MA")
Phi <-matrix( c(1,0,.7,.0,1,.6,.7,.6,1),ncol=3)
gre.gpa <- sim.structural(fx,Phi,fy,n=1000)
gre.gpa
```

```
Call: sim.structural(fx = fx, Phi = Phi, fy = fy)

 $model (Population correlation matrix)
        V    Q     A  nach   Anx   gpa   Pre    MA
V    1.00 0.72  0.54  0.00  0.00  0.38  0.32  0.25
Q    0.72 1.00  0.48  0.00  0.00  0.34  0.28  0.22
A    0.54 0.48  1.00  0.48 -0.42  0.47  0.39  0.31
nach 0.00 0.00  0.48  1.00 -0.56  0.29  0.24  0.19
Anx  0.00 0.00 -0.42 -0.56  1.00 -0.25 -0.21 -0.17
gpa  0.38 0.34  0.47  0.29 -0.25  1.00  0.30  0.24
Pre  0.32 0.28  0.39  0.24 -0.21  0.30  1.00  0.20
MA   0.25 0.22  0.31  0.19 -0.17  0.24  0.20  1.00

$reliability (population reliability)
   V    Q    A nach  Anx  gpa  Pre   MA
0.81 0.64 0.72 0.64 0.49 0.36 0.25 0.16

$reliability (population reliability)
   V    Q    A nach  Anx  gpa  Pre   MA
0.81 0.64 0.72 0.64 0.49 0.36 0.25 0.16
```

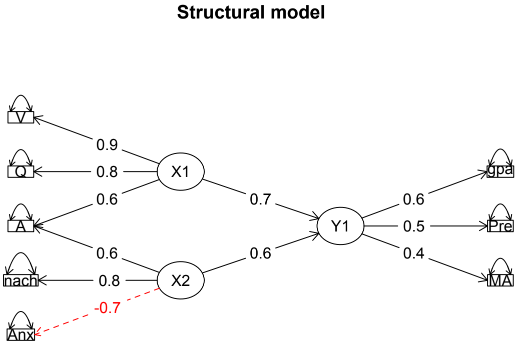
R code

```
> fx
> fy
> Phi
```

```
>  fx
     [,1] [,2]
V     0.9  0.0
Q     0.8  0.0
A     0.6  0.6
nach  0.0  0.8
Anx   0.0 -0.7
> Phi
     [,1] [,2] [,3]
[1,]  1.0  0.0  0.7
[2,]  0.0  1.0  0.6
[3,]  0.7  0.6  1.0
> fy
    [,1]
gpa  0.6
Pre  0.5
MA   0.4
```

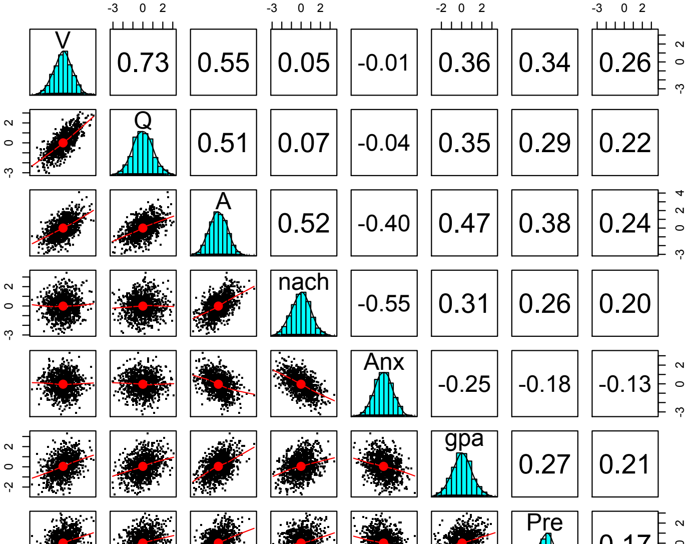
Create a figure (and write sem code for the sem package)

```
mod4 <- structure.diagram(fx,Phi,fy,errors=TRUE,e.size=.3)
```



Show the SPLOM

```
pairs.panels(gre.gpa$observed,pch=".")
```



two alternative sem/cfa packages

1. *sem* Fox et al. (2013)
   - Uses a complete "RAMpath" notation (Zhang, McArdle, Hamagami & Grimm, 2016):
2. *lavaan* (Rosseel, 2012)
   - Has become the standard R implementation of CFA/SEM

Show the sem code – This uses RAM path notation for all arrows

```
> mod4

       Path             Parameter  Value
 [1,]  "X1->V"          "F1V"      NA
 [2,]  "X1->Q"          "F1Q"      NA
 [3,]  "X1->A"          "F1A"      NA
 [4,]  "X2->A"          "F2A"      NA
 [5,]  "X2->nach"       "F2nach"   NA
 [6,]  "X2->Anx"        "F2Anx"    NA
 [7,]  "V<->V"          "x1e"      NA
 [8,]  "Q<->Q"          "x2e"      NA
 [9,]  "A<->A"          "x3e"      NA
[10,]  "nach<->nach"    "x4e"      NA
[11,]  "Anx<->Anx"      "x5e"      NA
[12,]  "Y1->gpa"        "Fygpa"    NA
[13,]  "Y1->Pre"        "FyPre"    NA
[14,]  "Y1->MA"         "FyMA"     NA
[15,]  "gpa<->gpa"      "y1e"      NA
[16,]  "Pre<->Pre"      "y2e"      NA
[17,]  "MA<->MA"        "y3e"      NA
[18,]  "X2<->X1"        "rF2F1"    NA
[19,]  "X1->Y1"         "rX1Y1"    NA
[20,]  "X2->Y1"         "rX2Y1"    NA
[21,]  "X1<->X1"        NA         "1"
[22,]  "X2<->X2"        NA         "1"
[23,]  "Y1<->Y1"        NA         "1"
attr(,"class")
[1] "mod"
```

Can we recover the structure using a confirmatory factor model?

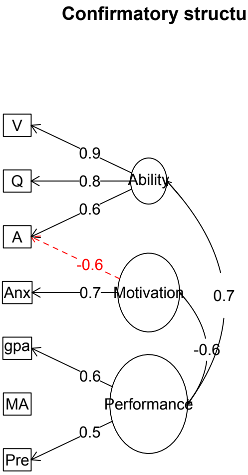
R code

```
model <- 'Ability =~ V + Q + A
              Motivation =~ A + Anx
              Performance =~ gpa + MA + Pre'
  fit <- sem(model,gre.gpa$observed,std.lv=TRUE)
  summary(fit)
  lavaan.diagram(fit)
```

lavaan output

```
lavaan (0.5-17) converged normally after  28 iterations

  Number of observations                          1000

  Estimator                                         ML
  Minimum Function Test Statistic               12.116
  Degrees of freedom                                10
  P-value (Chi-square)                           0.277

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  Ability =~
    V                 0.924    0.029   32.165    0.000
    Q                 0.830    0.029   28.263    0.000
    A                 0.614    0.033   18.540    0.000
  Motivation =~
    A                -0.593    0.038  -15.642    0.000
    Anx               0.660    0.044   14.990    0.000
  Performance =~
    gpa               0.570    0.036   16.001    0.000
    MA                0.341    0.035    9.637    0.000
    Pre               0.483    0.035   13.678    0.000

Covariances:
  Ability ~~
    Motivation       -0.033    0.051   -0.654    0.513
    Performance       0.748    0.039   19.386    0.000
  Motivation ~~
```

Create a figure from the lavaan output)



But the data can also be modeled as full SEM model

A SEM model is a CFA with regressions

R code

```
model <- 'Ability =~ V + Q + A
              Motivation =~ A + Anx
              Performance =~ gpa + MA + Pre
              Performance ~ Ability + Motivation'
  fit <- sem(model,gre.gpa$observed,std.lv=TRUE)
  summary(fit)
  standardizedSolution(fit)
  lavaan.diagram(fit)
```

lavaan solution for the regression model

```
standardizedSolution(fit)
           lhs op         rhs est.std    se       z pvalue
1      Ability =~           V   0.899 0.028  32.165  0.000
2      Ability =~           Q   0.810 0.029  28.263  0.000
3      Ability =~           A   0.593 0.032  18.540  0.000
4   Motivation =~           A  -0.573 0.037 -15.642  0.000
5   Motivation =~         Anx   0.672 0.045  14.990  0.000
6  Performance =~         gpa   0.579 0.198   2.932  0.003
7  Performance =~          MA   0.340 0.117   2.891  0.004
8  Performance =~         Pre   0.484 0.165   2.941  0.003
9  Performance  ~     Ability   0.729 0.252   2.889  0.004
10 Performance  ~  Motivation  -0.575 0.209  -2.749  0.006
11           V ~~           V   0.193 0.026   7.444  0.000
12           Q ~~           Q   0.344 0.025  13.704  0.000
13           A ~~           A   0.297 0.034   8.625  0.000
14         Anx ~~         Anx   0.548 0.053  10.301  0.000
15         gpa ~~         gpa   0.664 0.039  17.001  0.000
16          MA ~~          MA   0.885 0.041  21.435  0.000
17         Pre ~~         Pre   0.766 0.039  19.660  0.000
18     Ability ~~     Ability   1.000    NA      NA     NA
19  Motivation ~~  Motivation   1.000    NA      NA     NA
20 Performance ~~ Performance   0.110    NA      NA     NA
21     Ability ~~  Motivation  -0.033 0.051  -0.654  0.513
```

Measuring structure at two (or more) time points

1. Is the structure the same
   - Structural Invariance (is the graph the same)
   - Measurement invariance (are the loadings the same)
   - Strong measurement invariance (are the item intercepts the same?)
   - Measuring change
2. Do the means change (is there growth)
   - This is the means of the latent trait, not the means of the items
3. Do the latent traits correlate across two or more occasions?
   - Just two occasions, can not separate trait from state effects
   - With > 2 occasions, can examine trait and state effects
4. Compare several different simulations

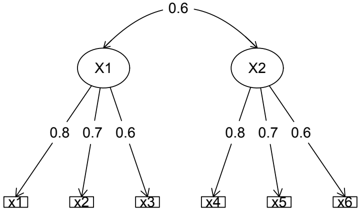
Create some basic data and add in some change

```
set.seed(42)
  fx <- matrix(c(.8,.7,.6,rep(0,6),.8,.7,.6),ncol=2)
 fx
     [,1] [,2]
[1,]  0.8  0.0
[2,]  0.7  0.0
[3,]  0.6  0.0
[4,]  0.0  0.8
[5,]  0.0  0.7
[6,]  0.0  0.6
Phi <- matrix(c(1,.6,.6,1),ncol=2)
 Phi
     [,1] [,2]
[1,]  1.0  0.6
[2,]  0.6  1.0

 x.model <- sim(fx=fx,Phi=Phi,mu=c(0,1),n=250)
 x <- x.model$observed
structure.diagram(fx,Phi,lr=FALSE,e.size=.3,main="A basic two time model")
 describe(x,skew=FALSE)

>  pairs.panels(x,pch=".")


  vars    n  mean    sd median trimmed   mad    min
max range   skew kurtosis    se
X1     1 250 -0.03 0.99  -0.07
-0.05 1.03 -2.54 2.62   5.17  0.17     -0.39 0.06
X2     2 250  0.04 0.96   0.02
0.03 0.90 -2.52 2.79   5.31  0.16     0.19 0.06
X3     3 250 -0.03 0.98  -0.04
-0.03 0.89 -2.52 2.52   5.05  0.02     -0.27 0.06
X4     4 250  0.76 1.03   0.81
```

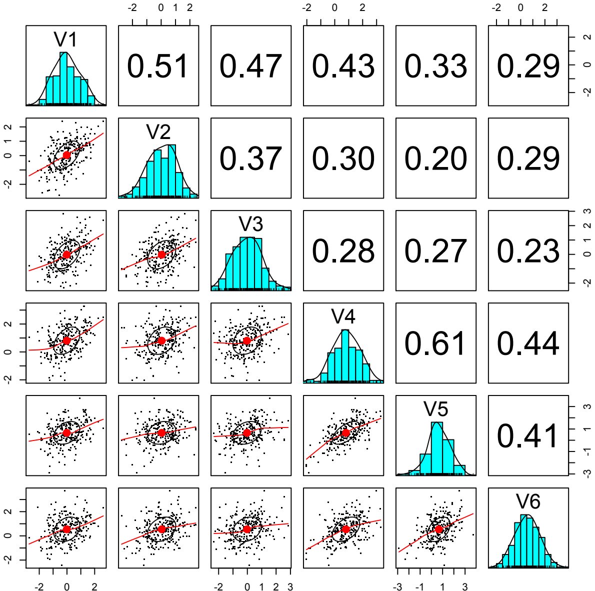
1. Set the random seed
2. Create a one factor structure
3. Put in some change
4. Show the structural model
5. Describe it
6. Show the splom (with small pch)

A basic two occasion trait model

A basic two time model



Splom of 6 basic variables



Multiple models

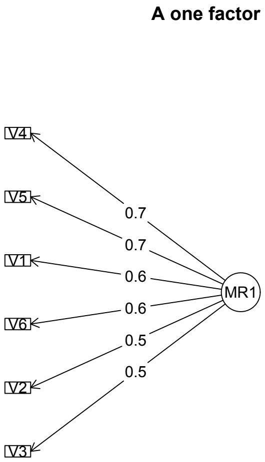
1. Ignore time, are the data congeneric?
   - Are they all measures of the same thing?
   - This is a one factor model
2. Include time, do we recover a correlation across time?
   - Try a two factor model
   - Plot the resulting structure

A one factor model

```
> fa(x)

Factor Analysis using method =  minres
Standardized loadings (pattern matrix) based upon correlation matrix
   MR1   h2   u2 com
1 0.65 0.42 0.58   1
2 0.62 0.38 0.62   1
3 0.46 0.21 0.79   1
4 0.74 0.54 0.46   1
5 0.60 0.36 0.64   1
6 0.59 0.35 0.65   1

                MR1
SS loadings    2.26
Proportion Var 0.38

Mean item complexity =  1
Test of the hypothesis that 1 factor is sufficient.
```

```
Test of the hypothesis that 1 factor is sufficient.

The degrees of freedom for the null model are
15  and the objective function was
                   1.6 with Chi Square of
393.33
The degrees of freedom for the model are 9
and the objective function was  0.27

The root mean square of the residuals (RMSR) is
0.09
The df corrected root mean square of the residuals
0.12

The harmonic number of observations is
250 with the empirical chi square  61.05
with prob <  8.4e-10
The total number of observations was
250  with Likelihood Chi Square =  66.6
with prob <  7e-11

Tucker Lewis Index of factoring reliability =
0.746
RMSEA index =  0.16
and the 90 % confidence intervals are
0.126 0.197
BIC =  16.9
Fit based upon off diagonal values = 0.95
Measures of factor score adequacy

Correlation of (regression) scores with factors
0.89
```

Fit a one factor model to the data fa.diagram(fa(x),sort=FALSE)



Try a two factor solution

```
> fa(x,2)

Factor Analysis using method =  minres
Factor Analysis using method =  minres
Call: fa(r = x, nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation matrix
    MR1   MR2   h2   u2 com
1  0.01  0.76 0.59 0.41   1
2  0.05  0.66 0.48 0.52   1
3 -0.06  0.59 0.31 0.69   1
4  0.75  0.08 0.63 0.37   1
5  0.77 -0.08 0.52 0.48   1
6  0.59  0.06 0.39 0.61   1

                       MR1  MR2
SS loadings           1.52 1.39
Proportion Var        0.25 0.23
Cumulative Var        0.25 0.49
Proportion Explained  0.52 0.48
Cumulative Proportion 0.52 1.00

 With factor correlations of
     MR1  MR2
MR1 1.00 0.61
MR2 0.61 1.00

Mean item complexity =  1
Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are
```

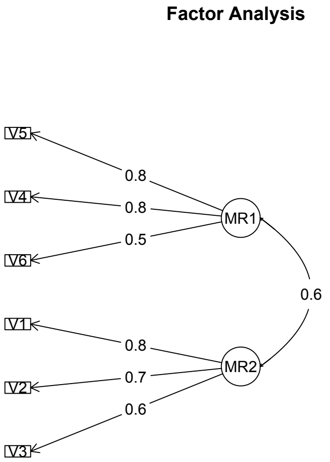
```
Test of the hypothesis that 2 factors are sufficien

The degrees of freedom for the null model are
15  and the objective function was
1.6 with Chi Square of  393.33
The degrees of freedom for the model are 4
and the objective function was  0.03

The root mean square of the residuals (RMSR) is
0.02
The df corrected root mean square of the residuals
0.04

The harmonic number of observations is
250 with the empirical chi square  3.83
with prob <  0.43
The total number of observations was
250  with Likelihood Chi Square =  7.28
with prob <  0.12

Tucker Lewis Index of factoring reliability =
0.967
RMSEA index =  0.057
and the 90 % confidence intervals are
0 0.123
BIC =  -14.81
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                MR1  MR2
Correlation of (regression) scores with factors
0.89 0.87
```

Compare the two solutions

```
Factor Analysis using method =  minres
Call: fa(r = x)
Factor Analysis using method =  minres
Call: fa(r = x)
Standardized loadings (pattern matrix) based upon correlation matrix
   MR1   h2   u2 com
1 0.76 0.57 0.43   1
2 0.65 0.42 0.58   1
3 0.56 0.31 0.69   1
4 0.64 0.41 0.59   1
5 0.50 0.25 0.75   1
6 0.50 0.25 0.75   1

                MR1
SS loadings    2.21
Proportion Var 0.37
```

```
Factor Analysis using method =  minres
Factor Analysis using method =  minres
Call: fa(r = x, nfactors = 2)
Standardized loadings (pattern matrix) based upon c
    MR1   MR2   h2   u2 com
1  0.88 -0.01 0.76 0.24 1.0
2  0.66  0.02 0.45 0.55 1.0
3  0.58  0.00 0.33 0.67 1.0
4 -0.01  0.87 0.75 0.25 1.0
5 -0.01  0.60 0.36 0.64 1.0
6  0.10  0.48 0.29 0.71 1.1

                       MR1  MR2
SS loadings           1.58 1.37
Proportion Var        0.26 0.23
Cumulative Var        0.26 0.49
Proportion Explained  0.53 0.47
Cumulative Proportion 0.53 1.00

 With factor correlations of
     MR1  MR2
MR1 1.00 0.58
MR2 0.58 1.00
```

EFA two factor solution



Now try three different sems (using lavaan)

1. Two correlated factors, free loadings
2. Two correlated factors, equal loadings across occasions
3. Two correlated factors, all loadings equal

A simple sem with a standardized solution using lavaan

R code

```
#factor model
mod2f <- 'F1 =~ V1 + V2 + V3
              F2 =~ V4 + V5 + V6
                 #correlation between factors
                F1 ~~F2'# now fit it and summarize it
fit <- sem(mod2f, data=x,std.lv=TRUE)
summary(fit, fit.measures=TRUE)
standardizedSolution(fit)
```

```
   lhs op rhs est.std    se      z pvalue ci.lower ci.upper
1   F1 =~  V1   0.773 0.047 16.573      0    0.682    0.865
2   F1 =~  V2   0.693 0.049 14.156      0    0.597    0.788
3   F1 =~  V3   0.537 0.056  9.568      0    0.427    0.647
4   F2 =~  V4   0.822 0.041 20.211      0    0.742    0.901
5   F2 =~  V5   0.687 0.046 15.072      0    0.598    0.777
6   F2 =~  V6   0.622 0.049 12.678      0    0.526    0.718
7   F1 ~~  F2   0.639 0.060 10.563      0    0.520    0.757
8   V1 ~~  V1   0.402 0.072  5.570      0    0.261    0.543
9   V2 ~~  V2   0.520 0.068  7.679      0    0.388    0.653
10  V3 ~~  V3   0.712 0.060 11.809      0    0.594    0.830
11  V4 ~~  V4   0.325 0.067  4.866      0    0.194    0.456
12  V5 ~~  V5   0.528 0.063  8.417      0    0.405    0.650
13  V6 ~~  V6   0.614 0.061 10.065      0    0.494    0.733
14  F1 ~~  F1   1.000 0.000     NA     NA    1.000    1.000
15  F2 ~~  F2   1.000 0.000     NA     NA    1.000    1.000
```

lavaan fit statistics

```
lavaan 0.6-9 ended normally after 16 iterations

  Estimator
ML
  Optimization method
NLMINB
  Number of model parameters
13

  Number of observations
250

Model Test User Model:

  Test statistic
9.546
  Degrees of freedom
8
  P-value (Chi-square)
0.298

Model Test Baseline Model:

  Test statistic
399.453
  Degrees of freedom
15
  P-value
0.000

User Model versus Baseline Model:
```

```
Root Mean Square Error of Approximation:

  RMSEA
0.000
  90 Percent Confidence Interval
0.000  0.040
  P-value RMSEA <= 0.05
0.972

Standardized Root Mean Square Residual:

  SRMR
0.017

Parameter estimates:

  Information
Expected
  Standard Errors
Standard
```

Create two new models with some equality constraints

The first sets the loadings for 1-3 equal to those of 4-6, the second says they are all equal

R code

```
mod2fa <- 'F1 =~ a*V1 + b*V2 + c*V3
                     F2 =~ a*V4 + b*V5 + c*V6
                     F1 ~~F2'
 fit2a <- sem(mod2fa,data=x.df,std.lv=TRUE)
 summary(fit2a,fit.measures=TRUE)

mod2fe <- 'F1 =~ a*V1 + a*V2 + a*V3
                     F2 =~ a*V4 + a*V5 + a*V6
                     F1 ~~F2'
 fit2e <- sem(mod2fe,data=x.df,std.lv=TRUE)
 summary(fit2e,fit.measures=TRUE)
```

Equal across time

```
lavaan (0.5-17) converged normally after
15 iterations

  Number of observations
250

  Estimator
ML
  Minimum Function Test Statistic
5.338
  Degrees of freedom
11
  P-value (Chi-square)
0.914

Model test baseline model:

  Minimum Function Test Statistic
393.670
  Degrees of freedom
15
  P-value
0.000

User model versus baseline model:

  Comparative Fit Index (CFI)
1.000
  Tucker-Lewis Index (TLI)
1.020

Loglikelihood and Information Criteria:
```

```
Root Mean Square Error of Approximation:

  RMSEA
0.000
  90 Percent Confidence Interval
0.000   0.025
  P-value RMSEA <= 0.05
0.991

Standardized Root Mean Square Residual:

  SRMR
0.035

Parameter estimates:

  Information
Expected
  Standard Errors
Standard


                        Estimate   Std.err
Z-value   P(>|z|)
Latent variables:
  F1 =~
    V1        (a)      0.817     0.046
17.711     0.000
    V2        (b)      0.662     0.049
13.630     0.000
    V3        (c)      0.549     0.046
11.949     0.000
```

All loadings equal

```
> summary(fit2e, fit.measures=TRUE)
lavaan (0.5-17) converged normally after
12 iterations

  Number of observations
250

  Estimator
ML
  Minimum Function Test Statistic
26.140
  Degrees of freedom
13
  P-value (Chi-square)
0.016

Model test baseline model:

  Minimum Function Test Statistic
393.670
  Degrees of freedom
15
  P-value
0.000

User model versus baseline model:

  Comparative Fit Index (CFI)
0.965
  Tucker-Lewis Index (TLI)
0.960

Root Mean Square Error of Approximation:

  RMSEA
0.064
  90 Percent Confidence Interval
0.026  0.099
  P-value RMSEA <= 0.05
0.235

Standardized Root Mean Square Residual:

  SRMR
0.084

Parameter estimates:

  Information
Expected
  Standard Errors
Standard

                    Estimate  Std.err
Z-value  P(>|z|)
Latent variables:
  F1 =~
    V1        (a)     0.686    0.033
20.964    0.000
    V2        (a)     0.686    0.033
20.964    0.000
    V3        (a)     0.686    0.033
```

Do the models differ?

These are nested models, and we can compare their χ^2 values.

R code
```
anova(fit,fit2a)
anova(fit2a,2e)
```

```
Chi Square Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit    8 3838.9 3884.6 3.9861                              
fit2a 11 3834.2 3869.4 5.3380     1.3519       3     0.7169

Chi Square Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit2a 11 3834.2 3869.4  5.338                              
fit2e 13 3851.0 3879.2 26.140     20.802       2   3.04e-05 ***
```

Measurement Invariance: Does a test measure the same thing

1. Across groups
   - Different schools
   - Different groups (e.g., ethnicity, age, gender)
2. Across time
   - Is todays' measure the same as next year's measure?
3. Types of invariance
   - Configural: Are the arrows the same
   - Weak invariance: Are the loadings the same across groups
   - Strong invariance: Loadings and intercepts are equal across groups
   - Super strong: Loadings, intercepts and means are equal across groups

Consider the Holzinger Swineford data set

1. 9 ability measures from two schools
   - 145 from Grant-White
   - 156 from Pasteur
2. Are the factor structures the same across schools
   - Although lavaan does this in one call, lets do it part by part
   - over all factor structure
   - factors with schools
   - constrain factors to have the same loadings, etc.
3. A subset of the data are available in lavaan as `HolzingerSwineford1939`
4. Complete data are available in *psychTools* as `holzinger.swineford` courtesy of Keith Widaman
5. Original data from Holzinger & Swineford (1939) and used in many different examples for LISREL, lavaan, etc.

The holzinger.swineford data description

The following commentary was provided by Keith Widaman:
“The Holzinger and Swineford (1939) data have been used as a model data set by many investigators. For example, Harman (1976) used the “24 Psychological Variables" example prominently in his authoritative text on multiple factor analysis, and the data presented under this rubric consisted of 24 of the variables from the Grant-White school (N = 145). Meredith (1964a, 1964b) used several variables from the Holzinger and Swineford study in his work on factorial invariance under selection. Joreskog (1971) based his work on multiple-group confirmatory factor analysis using the Holzinger and Swineford data, subsetting the data into four groups.
Rosseel, who developed the ‘lavaan’ package for R , included 9 of the manifest variables from Holzinger and Swineford (1939) as a “resident" data set when one downloads the ‘lavaan’ package. Several background variables are included in this “resident" data set in addition to 9 of the psychological tests (which are named x1 – x9 in the data set). When analyzing these data, I found the distributions of the variables (means, SDs) did not match the sample statistics from the original article. For example, in the “resident" data set in ‘lavaan’, scores on all manifest variables ranged between 0 and 10, sample means varied between 3 and 6, and sample SDs varied between 1.0 and 1.5. In the original data set, scores ranges were rather different across tests, with some variables having scores that ranged between 0 and 20, but other manifest variables having scores ranging from 50 to over 300 – with obvious attendant differences in sample means and SDs.
After a bit of snooping (i.e., data analysis), I discovered that the 9 variables in the “resident" data set in ‘lavaan’ had been rescored through ratio transformations. The ratio transformations involved dividing the raw score for each person on a given test by a particular constant for that test that transformed scores on the test to have the desired range.

I decided to perform transformations of all 26 variables so that two data sets could be available to interested researchers:"

First, some descriptive statistics

```
describe(HolzingerSwineford1939,skew=FALSE)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	se
id	1	301	176.55	105.94	163.00	176.78	140.85	1.00	351.00	350.00	6.11
sex	2	301	1.51	0.50	2.00	1.52	0.00	1.00	2.00	1.00	0.03
ageyr	3	301	13.00	1.05	13.00	12.89	1.48	11.00	16.00	5.00	0.06
agemo	4	301	5.38	3.45	5.00	5.32	4.45	0.00	11.00	11.00	0.20
school*	5	301	1.52	0.50	2.00	1.52	0.00	1.00	2.00	1.00	0.03
grade	6	300	7.48	0.50	7.00	7.47	0.00	7.00	8.00	1.00	0.03
x1	7	301	4.94	1.17	5.00	4.96	1.24	0.67	8.50	7.83	0.07
x2	8	301	6.09	1.18	6.00	6.02	1.11	2.25	9.25	7.00	0.07
x3	9	301	2.25	1.13	2.12	2.20	1.30	0.25	4.50	4.25	0.07
x4	10	301	3.06	1.16	3.00	3.02	0.99	0.00	6.33	6.33	0.07
x5	11	301	4.34	1.29	4.50	4.40	1.48	1.00	7.00	6.00	0.07
x6	12	301	2.19	1.10	2.00	2.09	1.06	0.14	6.14	6.00	0.06
x7	13	301	4.19	1.09	4.09	4.16	1.10	1.30	7.43	6.13	0.06
x8	14	301	5.53	1.01	5.50	5.49	0.96	3.05	10.00	6.95	0.06
x9	15	301	5.37	1.01	5.42	5.37	0.99	2.78	9.25	6.47	0.06

```
colnames(holzinger.swineford)
 [1] "case"         "school"       "grade"        "female"       "ageyr"
 [6] "mo"           "agemo"        "t01_visperc"  "t02_cubes"    "t03_frmbord"
[11] "t04_lozenges" "t05_geninfo"  "t06_paracomp" "t07_sentcomp" "t08_wordclas"
[16] "t09_wordmean" "t10_addition" "t11_code"     "t12_countdot" "t13_sccaps"
[21] "t14_wordrecg" "t15_numbrecg" "t16_figrrecg" "t17_objnumb"  "t18_numbfig"
[26] "t19_figword"  "t20_deduction" "t21_numbpuzz" "t22_probreas" "t23_series"
[31] "t24_woody"    "t25_frmbord2" "t26_flags"
```

describeBy each group

```
> describeBy(HolzingerSwineford1939,group=HolzingerSwineford1939$school,sk

group: Grant-White
        var   n  mean   sd median trimmed  mad   min   max range
se
sex       2 145  1.50 0.50   2.00    1.50 0.00  1.00  2.00  1.00 0.0
ageyr     3 145 12.72 0.97  13.00   12.67 1.48 11.00 16.00  5.00 0.0
...
grade     6 144  7.45 0.50   7.00    7.44 0.00  7.00  8.00  1.00 0.0
x1        7 145  4.93 1.15   5.00    4.96 1.24  1.83  8.50  6.67 0.1
x2        8 145  6.20 1.11   6.25    6.14 1.11  2.25  9.25  7.00 0.0
...
x8       14 145  5.49 1.05   5.50    5.45 0.89  3.05 10.00  6.95 0.0
x9       15 145  5.33 1.03   5.31    5.33 1.15  3.11  9.25  6.14 0.0
------------------------------------------------------------------------
group: Pasteur
        var   n  mean   sd median trimmed  mad   min   max range
se
sex       2 156  1.53 0.50   2.00    1.53 0.00  1.00  2.00  1.00 0.04
ageyr     3 156 13.25 1.06  13.00   13.15 1.48 12.00 16.00  4.00 0.09
...
grade     6 156  7.50 0.50   7.50    7.50 0.74  7.00  8.00  1.00 0.04
x1        7 156  4.94 1.19   5.00    4.97 1.24  0.67  7.50  6.83 0.09
x2        8 156  5.98 1.23   5.75    5.89 1.11  3.50  9.25  5.75 0.10
```

EFA for both groups

by(HolzingerSwineford1939[,7:15],HolzingerSwineford1939[,5],fa,nfactors

```
HolzingerSwineford1939[, 5]: Grant-White
Factor Analysis using method =  minres
Call: FUN(r = data[x, , drop = FALSE], nfactors = 3
Standardized loadings (pattern matrix) based upon c
     MR1   MR2   MR3   h2   u2
x1  0.09  0.06  0.64 0.50 0.50
x2  0.02 -0.03  0.51 0.26 0.74
x3  0.11 -0.03  0.64 0.47 0.53
x4  0.86 -0.04  0.04 0.76 0.24
x5  0.82  0.09 -0.03 0.70 0.30
x6  0.81 -0.04  0.05 0.68 0.32
x7  0.14  0.78 -0.19 0.60 0.40
x8 -0.11  0.79  0.18 0.69 0.31
x9  0.08  0.46  0.40 0.54 0.46

                       MR1  MR2  MR3
SS loadings           2.23 1.53 1.44
Proportion Var        0.25 0.17 0.16
Cumulative Var        0.25 0.42 0.58
Proportion Explained  0.43 0.29 0.28
Cumulative Proportion 0.43 0.72 1.00

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.25 0.41
MR2 0.25 1.00 0.31
MR3 0.41 0.31 1.00
```

```
HolzingerSwineford1939[, 5]: Pasteur
Factor Analysis using method =  minres
Call: FUN(r = data[x, , drop = FALSE], nfactors = 3
Standardized loadings (pattern matrix) based upon c
     MR1   MR2   MR3   h2   u2
x1  0.27  0.59  0.00 0.51 0.49
x2  0.03  0.49 -0.16 0.25 0.75
x3 -0.08  0.73  0.01 0.50 0.50
x4  0.80  0.02  0.06 0.68 0.32
x5  0.92 -0.07 -0.06 0.79 0.21
x6  0.78  0.13  0.06 0.70 0.30
x7  0.06 -0.14  0.71 0.52 0.48
x8 -0.02  0.13  0.60 0.39 0.61
x9 -0.02  0.37  0.40 0.34 0.66

                       MR1  MR2  MR3
SS loadings           2.22 1.36 1.10
Proportion Var        0.25 0.15 0.12
Cumulative Var        0.25 0.40 0.52
Proportion Explained  0.48 0.29 0.23
Cumulative Proportion 0.48 0.77 1.00

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.27 0.26
MR2 0.27 1.00 0.14
MR3 0.26 0.14 1.00
```

How similar are the solutions: factor congruence

Factor congruence is the cosine of the angle between two vectors:

$$Congruence = (diag(X'X))^{-.5} Y'X(diag((Y'Y))^{-.5}$$

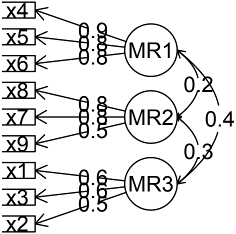
```
> f3.pasteur <- fa(HolzingerSwineford1939[1:156,7:15],3)
> f3.grant <- fa(HolzingerSwineford1939[157:301,7:15],3)
>factor.congruence(f3.pasteur,f3.grant)
```

	MR1	MR2	MR3
MR1	0.97	0.03	0.10
MR2	0.12	0.11	0.99
MR3	0.08	0.97	0.05

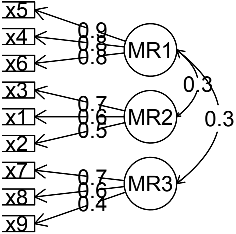
```
cross <- t(y) %*% x
sumsx <- sqrt(1/diag(t(x) %*% x))
 sumsy <- sqrt(1/diag(t(y) %*% y))
```

Do they look alike?

Factor Analysis



Factor Analysis



Test if the model fits the combined data

```
HS.model <- ' visual  =~ x1 + x2 + x3
              textual =~ x4 + x5 + x6
              speed   =~ x7 + x8 + x9 '

 fit <- cfa(HS.model, data=HolzingerSwineford1939, std.lv=TRUE)
summary(fit, fit.measures=TRUE)

lavaan (0.4-14) converged normally after 41 iterations

  Number of observations                           301

  Estimator                                         ML
  Minimum Function Chi-square                   85.306
  Degrees of freedom                                24
  P-value                                        0.000

Chi-square test baseline model:

  Minimum Function Chi-square                  918.852
  Degrees of freedom                                36
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.931
  Tucker-Lewis Index (TLI)                       0.896

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -3737.745
  Loglikelihood unrestricted model (H1)      -3695.092
```

With values of

```
Number of free parameters                          21
  Akaike (AIC)                                7517.490
  Bayesian (BIC)                              7595.339
  Sample-size adjusted Bayesian (BIC)         7528.739

Root Mean Square Error of Approximation:

  RMSEA                                          0.092
  90 Percent Confidence Interval        0.071   0.114
  P-value RMSEA <= 0.05                          0.001

Standardized Root Mean Square Residual:

  SRMR                                           0.065

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                0.900    0.081   11.127    0.000
    x2                0.498    0.077    6.429    0.000
    x3                0.656    0.074    8.817    0.000
  textual =~
    x4                0.990    0.057   17.474    0.000
    x5                1.102    0.063   17.576    0.000
    x6                0.917    0.054   17.082    0.000
  speed =~
    x7                0.619    0.070    8.903    0.000
    x8                0.731    0.066   11.090    0.000
```

Now do it for both groups one analysis

```
> fit2 <- cfa(HW.model, data=HolzingerSwineford1939, group="school",std.lv=TRUE)
> summary(fit2)


  Number of observations per group
  Pasteur                                           156
  Grant-White                                       145

  Estimator                                          ML
  Minimum Function Chi-square                   115.851
  Degrees of freedom                                 48
  P-value                                         0.000

Chi-square for each group:

  Pasteur                                        64.309
  Grant-White                                    51.542

Parameter estimates:

  Information                                  Expected
  Standard Errors                              Standard
```

Results for Pasteur

```
Group 1 [Pasteur]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                1.047    0.132    7.934    0.000
    x2                0.412    0.110    3.753    0.000
    x3                0.597    0.108    5.525    0.000
  textual =~
    x4                0.946    0.079   11.927    0.000
    x5                1.119    0.089   12.604    0.000
    x6                0.827    0.068   12.230    0.000
  speed =~
    x7                0.591    0.106    5.557    0.000
    x8                0.665    0.102    6.531    0.000
    x9                0.545    0.097    5.596    0.000

Covariances:
  visual ~~
    textual           0.484    0.086    5.600    0.000
    speed             0.299    0.109    2.755    0.006
  textual ~~
    speed             0.325    0.100    3.256    0.001

Intercepts:
    x1                4.941    0.095   52.249    0.000
    x2                5.984    0.098   60.949    0.000
    x3                2.487    0.093   26.778    0.000
    x4                2.823    0.092   30.689    0.000
    x5                3.995    0.105   38.183    0.000
    x6                1.922    0.079   24.321    0.000
    x7                4.432    0.087   51.181    0.000
    x8                5.563    0.078   71.214    0.000
```

Results for Grant-White

```
Latent variables:
  visual =~
    x1               0.777    0.103    7.525    0.000
    x2               0.572    0.101    5.642    0.000
    x3               0.719    0.093    7.711    0.000
  textual =~
    x4               0.971    0.079   12.355    0.000
    x5               0.961    0.083   11.630    0.000
    x6               0.935    0.081   11.572    0.000
  speed =~
    x7               0.679    0.087    7.819    0.000
    x8               0.833    0.087    9.568    0.000
    x9               0.719    0.086    8.357    0.000

Covariances:
  visual ~~
    textual          0.541    0.085    6.355    0.000
    speed            0.523    0.094    5.562    0.000
  textual ~~
    speed            0.336    0.091    3.674    0.000

Intercepts:
    x1               4.930    0.095   51.696    0.000
    x2               6.200    0.092   67.416    0.000
    x3               1.996    0.086   23.195    0.000
    x4               3.317    0.093   35.625    0.000
    x5               4.712    0.096   48.986    0.000
    x6               2.469    0.094   26.277    0.000
    x7               3.921    0.086   45.819    0.000
    x8               5.488    0.087   63.174    0.000
    x9               5.327    0.085   62.571    0.000
    visual           0.000
    textual          0.000
```

Constrain the loadings to be the same

```
> fit2e <- cfa(HW.model, data=HolzingerSwineford1939, group="school",std.lv=TRUE,group.equal="load
> summary(fit2e)

lavaan (0.4-14) converged normally after 30 iterations

  Number of observations per group
  Pasteur                                           156
  Grant-White                                       145

  Estimator                                          ML
  Minimum Function Chi-square                   127.834
  Degrees of freedom                                 57
  P-value                                         0.000

Chi-square for each group:

  Pasteur                                        71.064
  Grant-White                                    56.770

Parameter estimates:

  Information                                  Expected
  Standard Errors                              Standard
```

With parameter values of

```
Group 1 [Pasteur]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                0.866    0.078   11.149    0.000
    x2                0.523    0.076    6.916    0.000
    x3                0.683    0.071    9.689    0.000
  textual =~
    x4                0.954    0.056   17.002    0.000
    x5                1.033    0.061   17.012    0.000
    x6                0.870    0.052   16.750    0.000
  speed =~
    x7                0.630    0.066    9.500    0.000
    x8                0.752    0.065   11.586    0.000
    x9                0.650    0.064   10.205    0.000

Covariances:
  visual ~~
    textual           0.485    0.087    5.555    0.000
    speed             0.341    0.109    3.126    0.002
  textual ~~
    speed             0.336    0.094    3.590    0.000

Intercepts:
    x1                4.941    0.092   53.661    0.000
    x2                5.984    0.099   60.420    0.000
    x3                2.487    0.093   26.734    0.000
    x4                2.823    0.093   30.400    0.000
    x5                3.995    0.100   39.756    0.000
    x6                1.922    0.081   23.732    0.000
    x7                4.432    0.089   49.903    0.000
```

Now, compare the fit of the two group model to the one with equal parameters

```
> anova(fit2, fit2e)

Chi Square Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit2  48 7484.4 7706.8 115.85
fit2e 57 7478.4 7667.4 127.83     11.982       9     0.2143
```

But, another way to specify the fit2e model – without making the latent variables standardized

```
fit2e <- cfa(HW.model, data=HolzingerSwineford1939, group="school",group.equal=c("loadings"))
```

```
 Number of observations per group
  Pasteur                                          156
  Grant-White                                      145

  Estimator                                         ML
  Minimum Function Chi-square                  124.044
  Degrees of freedom                                54
  P-value                                        0.000

Chi-square for each group:
  Pasteur                                       68.825
  Grant-White                                   55.219
Parameter estimates:
  Information                                 Expected
  Standard Errors                             Standard

Group 1 [Pasteur]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                1.000
    x2                0.599    0.100    5.979    0.000
    x3                0.784    0.108    7.267    0.000
  textual =~
    x4                1.000
    x5                1.083    0.067   16.049    0.000
    x6                0.912    0.058   15.785    0.000
  speed =~
    x7                1.000
```

Test fit by comparing models

```
> anova(fit2, fit2e)

Chi Square Difference Test

      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit2  48 7484.4 7706.8 115.85
fit2e 54 7480.6 7680.8 124.04     8.1922       6     0.2244
```

Continue this logic, of successive tests with more constraints, but do it automatically

```
mi <- measurementInvariance(HW.model, data=HolzingerSwineford1939, group="school")
```

```
Measurement invariance tests:
Model 1: configural invariance:
   chisq       df    pvalue      cfi     rmsea       bic
 115.851   48.000    0.000     0.923     0.097  7706.822
Model 2: weak invariance (equal loadings):
   chisq       df    pvalue      cfi     rmsea       bic
 124.044   54.000    0.000     0.921     0.093  7680.771
[Model 1 versus model 2]
  delta.chisq       delta.df  delta.p.value      delta.cfi
        8.192          6.000          0.224          0.002
Model 3: strong invariance (equal loadings + intercepts):
   chisq       df    pvalue      cfi     rmsea       bic
 164.103   60.000    0.000     0.882     0.107  7686.588
[Model 1 versus model 3]
  delta.chisq       delta.df  delta.p.value      delta.cfi
       48.251         12.000          0.000          0.041
[Model 2 versus model 3]
  delta.chisq       delta.df  delta.p.value      delta.cfi
       40.059          6.000          0.000          0.038
Model 4: equal loadings + intercepts + means:
   chisq       df    pvalue      cfi     rmsea       bic
 204.605   63.000    0.000     0.840     0.122  7709.969
[Model 1 versus model 4]
  delta.chisq       delta.df  delta.p.value      delta.cfi
       88.754         15.000          0.000          0.083
[Model 3 versus model 4]
  delta.chisq       delta.df  delta.p.value      delta.cfi
       40.502          3.000          0.000          0.042
>
```

Create a data set with non-invariant factor loadings

```
> set.seed(42)
 > fx <- matrix(c(.8,.7,.6,rep(0,6),.6,.7,.8),ncol=2)
> fx
     [,1] [,2]
[1,]  0.8  0.0
[2,]  0.7  0.0
[3,]  0.6  0.0
[4,]  0.0  0.6
[5,]  0.0  0.7
[6,]  0.0  0.8
> Phi <- matrix(c(1,.6,.6,1),ncol=2)
> Phi
> set.seed(42)
> x.model <- sim(fx=fx,Phi=Phi,mu=c(0,1),n=250)
>  x <- x.model$observed
>  structure.diagram(fx,Phi,lr=FALSE,e.size=.3,main="A basic two time mod
>  describe(x,skew=FALSE)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	se
V1	1	250	0.02	0.99	0.01	0.03	1.02	−2.64	2.76	5.40	0.06
V2	2	250	−0.02	0.96	−0.01	−0.03	0.94	−2.66	3.12	5.78	0.06
V3	3	250	0.02	0.97	−0.06	0.01	0.95	−2.74	2.41	5.15	0.06
V4	4	250	0.61	0.99	0.59	0.65	0.99	−3.26	3.15	6.41	0.06

One factor model

```
> f1n <- fa(x)
> f1n
```

```
Factor Analysis using method =  minres
Call: fa(r = x)
Standardized loadings (pattern matrix) based upon correlation matrix
    MR1   h2   u2
V1 0.59 0.35 0.65
V2 0.56 0.31 0.69
V3 0.48 0.23 0.77
V4 0.54 0.29 0.71
V5 0.69 0.48 0.52
V6 0.72 0.52 0.48

                MR1
SS loadings    2.18
Proportion Var 0.36

Test of the hypothesis that 1 factor is sufficient.

The degrees of freedom for the null model are
15
        and the objective function was
1.58
         with Chi Square of  389.85
The degrees of freedom for the model are 9
         and the objective function was
0.33
```

```
The root mean square of the residuals (RMSR) is
0.07
The df corrected root mean square of the residuals
0.13
The number of observations was  250
with Chi Square =  79.93  with prob <
1.7e-13

Tucker Lewis Index of factoring reliability =
0.684
RMSEA index =  0.179
and the 90 % confidence intervals are
0.143 0.214
BIC =  30.24
Fit based upon off diagonal values = 0.93
Measures of factor score adequacy
                                                    MR1
Correlation of scores with factors
0.89
Multiple R square of scores with factors
0.79
Minimum correlation of possible factor scores
0.57
```

Two factor model

```
> f2n <- fa(x,2)
> f2n

Factor Analysis using method =  minres
Call: fa(r = x, nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation matrix
     MR1   MR2   h2   u2
V1 -0.01  0.79 0.62 0.38
V2 -0.01  0.73 0.53 0.47
V3  0.15  0.40 0.25 0.75
V4  0.51  0.07 0.30 0.70
V5  0.79 -0.03 0.60 0.40
V6  0.80  0.02 0.65 0.35

                       MR1  MR2
SS loadings           1.58 1.37
Proportion Var        0.26 0.23
Cumulative Var        0.26 0.49
Proportion Explained  0.53 0.47
Cumulative Proportion 0.53 1.00

 With factor correlations of
     MR1  MR2
MR1 1.00 0.54
MR2 0.54 1.00

Correlation of scores with factors
0.90 0.88
Multiple R square of scores with factors
0.80 0.77
Minimum correlation of possible factor scores

Test of the hypothesis that 2 factors are sufficier

The degrees of freedom for the null model are
15
                              and the objective function was
1.58
                with Chi Square of  389.85
The degrees of freedom for the model are 4
                     and the objective function was
0.02

The root mean square of the residuals (RMSR) is
0.01
The df corrected root mean square of the residuals
0.04
The number of observations was  250  with
                   Chi Square =  4.7
with prob <  0.32

Tucker Lewis Index of factoring reliability =
0.993
RMSEA index =  0.028
and the 90 % confidence intervals are
NA 0.102
BIC =  -17.39
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                  MR
MR2
```

Compare the two solutions

```
Factor Analysis using method =  minres
Call: fa(r = x)
Standardized loadings (pattern matrix)
  based upon correlation matrix
    MR1   h2   u2
V1 0.59 0.35 0.65
V2 0.56 0.31 0.69
V3 0.48 0.23 0.77
V4 0.54 0.29 0.71
V5 0.69 0.48 0.52
V6 0.72 0.52 0.48

                MR1
SS loadings    2.18
Proportion Var 0.36
```

```
Factor Analysis using method =  minres
Call: fa(r = x, nfactors = 2)
Standardized loadings (pattern matrix)
           based upon correlation matrix
     MR1   MR2   h2   u2
V1 -0.01  0.79 0.62 0.38
V2 -0.01  0.73 0.53 0.47
V3  0.15  0.40 0.25 0.75
V4  0.51  0.07 0.30 0.70
V5  0.79 -0.03 0.60 0.40
V6  0.80  0.02 0.65 0.35

                       MR1  MR2
SS loadings           1.58 1.37
Proportion Var        0.26 0.23
Cumulative Var        0.26 0.49
Proportion Explained  0.53 0.47
Cumulative Proportion 0.53 1.00

 With factor correlations of
     MR1  MR2
MR1 1.00 0.54
MR2 0.54 1.00
```

CFA 2 correlated factors

```
> y.df <- data.frame(y)
> fitn <- sem(mod2f, data=y.df, std.lv=TRUE)
> summary(fitn , fit.measures=TRUE)


lavaan (0.4-14) converged normally after 17 iterations
  Number of observations
250
  Estimator
ML
  Minimum Function Chi-square
2.552
  Degrees of freedom
8
  P-value
0.959
Chi-square test baseline model:

  Minimum Function Chi-square
341.309
  Degrees of freedom
15
  P-value
0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)
1.000
  Tucker-Lewis Index (TLI)
1.031

Loglikelihood and Information Criteria:
```

```
Root Mean Square Error of Approximation:

  RMSEA
0.000
  90 Percent Confidence Interval
0.000  0.000
  P-value RMSEA <= 0.05
0.994

Standardized Root Mean Square Residual:

  SRMR
0.014

Parameter estimates:

  Information
Expected
  Standard Errors
Standard

                          Estimate  Std.err
Z-value  P(>|z|)
Latent variables:
  F1 =~
    V1                       0.750    0.063
11.880     0.000
    V2                       0.725    0.065
11.160     0.000
    V3                       0.621    0.066
9.410     0.000
  F2 =~
    V1                       0.505    0.074
```

Are the factors measurement invariant? Well, maybe.

```
> summary(fitn2a, fit.measures=TRUE)

lavaan (0.4-14) converged normally after 14 iterations

  Number of observations
250

  Estimator
ML
  Minimum Function Chi-square
8.093
  Degrees of freedom
11
  P-value
0.705

Chi-square test baseline model:

  Minimum Function Chi-square
341.309
  Degrees of freedom
15
  P-value
0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)
1.000
  Tucker-Lewis Index (TLI)
```

```
Root Mean Square Error of Approximation:

  RMSEA
0.000
  90 Percent Confidence Interval
0.000   0.051
  P-value RMSEA <= 0.05
0.946

Standardized Root Mean Square Residual:

  SRMR
0.049

Parameter estimates:

  Information
Expected
  Standard Errors
Standard

                  Estimate  Std.err
Z-value  P(>|z|)
Latent variables:
  F1 =~
    V1        (a)     0.686    0.050
13.774    0.000
    V2        (b)     0.675    0.049
13.834    0.000
    V3        (c)     0.657    0.048
13.677    0.000
  F2 =~
```

Create the original data set again

```
set.seed(42)
 fx <- matrix(c(.8,.7,.6,rep(0,6),.8,.7,.6),ncol=2)
 x.model <- sim(fx=fx,Phi=Phi,mu=c(0,1),n=250)
 x <- x.model$observed
x.df <- data.frame(x)
colnames(x.df) <- paste0("V",1:6)
```

```
> describe(x,skew=FALSE)
   vars   n  mean   sd   min  max range   se
V1    1 250 -0.03 0.99 -2.54 2.62  5.17 0.06
V2    2 250  0.04 0.96 -2.52 2.79  5.31 0.06
V3    3 250 -0.03 0.98 -2.52 2.52  5.05 0.06
V4    4 250  0.76 1.03 -1.77 3.37  5.13 0.07
V5    5 250  0.69 0.94 -2.18 2.74  4.92 0.06
V6    6 250  0.54 0.98 -2.71 3.40  6.10 0.06
```

Model change in the items

```
mod2fc <- 'F1 =~ a*V1 + b* V2 + c*V3
                 F2 =~ a* V4 + b*V5 +c* V6
                     #correlation between factors
                    F2 ~  F1 '      #the regression
fit2c <- sem(mod2fc,data=x.df,meanstructure=TRUE)
summary(fit2c, fit.measures=TRUE)
```

Correlated factors

```
lavaan (0.4-14) converged normally after 23 iterations

  Number of observations
250

  Estimator
ML
  Minimum Function Chi-square
10.616
  Degrees of freedom
10
  P-value
0.388

Chi-square test baseline model:

  Minimum Function Chi-square
406.795
  Degrees of freedom
15
  P-value
0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)
```

```
Root Mean Square Error of Approximation:

  RMSEA
0.016
  90 Percent Confidence Interval
0.000  0.071
  P-value RMSEA <= 0.05
0.793

Standardized Root Mean Square Residual:

  SRMR
0.032

Parameter estimates:

  Information
Expected
  Standard Errors
Standard

                    Estimate  Std.err
Z-value  P(>|z|)
Latent variables:
  F1 =~
    V1         (a)     1.000
    V2         (b)     0.925    0.076
12.243    0.000
    V3         (c)     0.816    0.072
11.386    0.000
  F2 =~
    V4         (a)     1.000
```

With the intercepts

```
Intercepts:
    V1              0.021    0.063
0.332    0.740
    V2              0.135    0.066
2.033    0.042
    V3              0.032    0.066
0.482    0.630
    V4              0.865    0.065
13.294    0.000
    V5              0.729    0.062
11.696    0.000
    V6              0.649    0.063
10.369    0.000
    F1              0.000
    F2              0.000
```

```
Variances:
    V1              0.383    0.060
    V2              0.576    0.068
    V3              0.670    0.071
    V4              0.486    0.065
    V5              0.481    0.061
    V6              0.596    0.065
    F1              0.613    0.088
    F2              0.319    0.063
```

Try fitting a moments model

```
mod2fc <- 'F1 =~ a*V1 + b* V2 + c*V3
                 F2 =~ a* V4 + b*V5 +c* V6
                 means = ~ F1 + F2 + 1*V1 +1*V2+1*V3+1*V4+1*V5+1*V6
                    #correlation between factors

                 F2 ~ F1 '      #the regression
fit2c <- sem(mod2fc, data=x.df, meanstructure=TRUE)
summary(fit2c , fit.measures=TRUE)
```

Using the moments matrix

Modeling change using the moments matrix

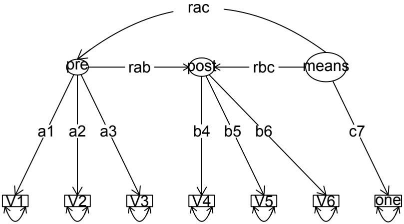
1. McArdle (2009) Latent variable modeling of differences and changes with longitudinal data. Annual Review of Psychology, 60, 577-605
   - Use moments rather than covariances

Create the model to be fit in sem

```
fxg
    pre  post  means
V1  "a1" "0"   "0"
V2  "a2" "0"   "0"
V3  "a3" "0"   "0"
V4  "0"  "b4"  "0"
V5  "0"  "b5"  "0"
V6  "0"  "b6"  "0"
one "0"  "0"   "c7"
> phi
    F1    F2  F3
F1  "1"   "0" "rac"
F2  "rab" "1" "rbc"
F3  "0"   "0" "1"

mod.mom1 <- structure.diagram(fog, phi, errors=TRUE)
```

Modeling the means in a moments matrix

Structural model



the basic path model, with some editing

```
mod.mom1
        Path               Parameter  Value
 [1,] "pre->V1"           "a1"       NA
 [2,] "pre->V2"           "a2"       NA
 [3,] "pre->V3"           "a3"       NA
 [4,] "post->V4"          "b4"       NA
 [5,] "post->V5"          "b5"       NA
 [6,] "post->V6"          "b6"       NA
 [7,] "means->one"        "c7"       NA
 [8,] "V1<->V1"           "x1e"      NA
 [9,] "V2<->V2"           "x2e"      NA
[10,] "V3<->V3"           "x3e"      NA
[11,] "V4<->V4"           "x4e"      NA
[12,] "V5<->V5"           "x5e"      NA
[13,] "V6<->V6"           "x6e"      NA
[14,] "one<->one"         NA     "1"         <-- edited
[15,] "pre␣->post"        "rF1F2"    NA
[16,] "means->pre"        "rF3F1"    NA       <- edited
[17,] "means->post"       "rF3F2"    NA     <- edited
[18,] "pre<->pre"         NA         "1"
[19,] "post<->post"       NA         "1"
[20,] "means<->means"     NA         "1"
attr(,"class")
[1] "mod"
```

sem output

```
> sem.mom1 <- sem(mod.mom, MomMat, N=250, raw=TRUE)
> summary(sem.mom1)


Model fit to raw moment matrix.

 Model Chisquare =   12.077    Df =
12
                        Pr(>Chisq) = 0.43954
 AIC =   44.077
 AICc =   14.411
 BIC =   100.42
 CAIC =   -66.181

 Normalized Residuals
   Min.  1st Qu.   Median
Mean  3rd Qu.     Max.
-1.2500  -0.1150  0.0000  -0.0103
0.0972   0.9730
```

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1     0.830111 0.069823  11.8888
1.3536e-32 V1 <--- pre
a2     0.881776 0.076506  11.5256
9.7982e-31 V2 <--- pre
a3     0.698963 0.074824   9.3415
9.5005e-21 V3 <--- pre
b4     0.664838 0.062395  10.6553
1.6468e-26 V4 <--- post
b5     0.554498 0.053510  10.3626
3.6687e-25 V5 <--- post
b6     0.510270 0.051499   9.9084
3.8265e-23 V6 <--- post
c7     1.118034 0.050000
22.3607 9.5054e-111 one <--- means
x1e    0.525613 0.078258   6.7164
1.8623e-11 V1 <--> V1
x2e    0.673405 0.093381   7.2114
5.5387e-13 V2 <--> V2
x3e    0.836504 0.090362   9.2572
2.0983e-20 V3 <--> V3
x4e    0.567286 0.081353   6.9732
3.0990e-12 V4 <--> V4
x5e    0.628041 0.073463   8.5490
1.2412e-17 V5 <--> V5
x6e    0.750522 0.080300   9.3465
9.0592e-21 V6 <--> V6
rF1F2  0.893184 0.142998   6.2461
4.2076e-10 post <--- pre
rF3F1  0.086583 0.073172   1.1833
2.3669e-01 pre <--- means
```

Traits and States and time

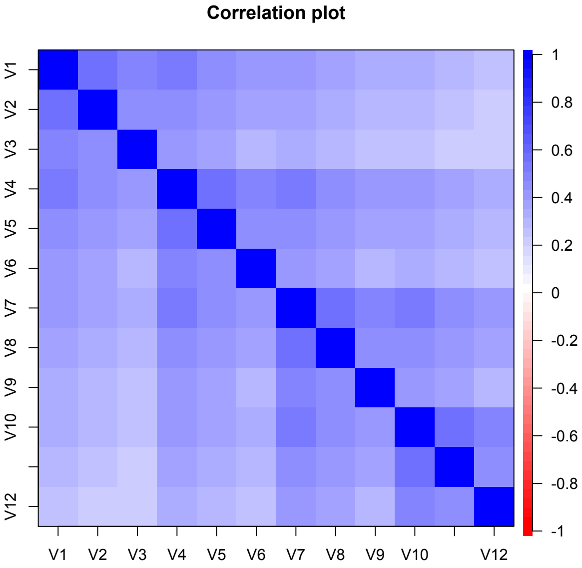
1. With just two time points, traits and states are confounded
   - Is the correlation a trait like stability
   - or does the state dissipate slowly?
2. With > 2 time points we can distinguish states and traits
   - States should have an autocorrelation component
   - Traits should be consistent across time
3. Consider the simplex structure of 4 time points
   - Clean within time factor structure
   - Simplex across time points

A factor simplex

```
simp <- sim()
```

```
 $model (Population correlation matrix)
      V1   V2   V3   V4   V5   V6   V7   V8   V9  V10  V11  V12
V1  1.00 0.56 0.48 0.51 0.45 0.38 0.41 0.36 0.31 0.33 0.29 0.25
V2  0.56 1.00 0.42 0.45 0.39 0.34 0.36 0.31 0.27 0.29 0.25 0.22
V3  0.48 0.42 1.00 0.38 0.34 0.29 0.31 0.27 0.23 0.25 0.22 0.18
V4  0.51 0.45 0.38 1.00 0.56 0.48 0.51 0.45 0.38 0.41 0.36 0.31
V5  0.45 0.39 0.34 0.56 1.00 0.42 0.45 0.39 0.34 0.36 0.31 0.27
V6  0.38 0.34 0.29 0.48 0.42 1.00 0.38 0.34 0.29 0.31 0.27 0.23
V7  0.41 0.36 0.31 0.51 0.45 0.38 1.00 0.56 0.48 0.51 0.45 0.38
V8  0.36 0.31 0.27 0.45 0.39 0.34 0.56 1.00 0.42 0.45 0.39 0.34
V9  0.31 0.27 0.23 0.38 0.34 0.29 0.48 0.42 1.00 0.38 0.34 0.29
V10 0.33 0.29 0.25 0.41 0.36 0.31 0.51 0.45 0.38 1.00 0.56 0.48
V11 0.29 0.25 0.22 0.36 0.31 0.27 0.45 0.39 0.34 0.56 1.00 0.42
V12 0.25 0.22 0.18 0.31 0.27 0.23 0.38 0.34 0.29 0.48 0.42 1.00
```

A simplex


Correlation plot
V1
V2
V3
V4
V5
V6
V7
V8
V9
V10
V12
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Factor structure of a simplex

```
> fsimp <- fa(simp$model)
> fsimp
```

```
Factor Analysis using method =  minres
Call: fa(r = simp$model)
Standardized loadings (pattern matrix) based upon correlation matrix
     MR1  h2   u2
V1   0.64 0.41 0.59
V2   0.57 0.33 0.67
V3   0.49 0.24 0.76
V4   0.73 0.54 0.46
V5   0.65 0.42 0.58
V6   0.56 0.31 0.69
V7   0.73 0.54 0.46
V8   0.65 0.42 0.58
V9   0.56 0.31 0.69
V10  0.64 0.41 0.59
V11  0.57 0.33 0.67
V12  0.49 0.24 0.76

                MR1
SS loadings    4.50
Proportion Var 0.38
```

Factors over time

```
> fsimp4 <- fa(simp$model,4)
> fsimp4
```

```
Factor Analysis using method =  minres
Call: fa(r = simp$model, nfactors = 4)
Standardized loadings (pattern matrix)
             based upon correlation matrix
    MR3 MR1 MR2 MR4   h2   u2
V1  0.8 0.0 0.0 0.0 0.64 0.36
V2  0.7 0.0 0.0 0.0 0.49 0.51
V3  0.6 0.0 0.0 0.0 0.36 0.64
V4  0.0 0.0 0.0 0.8 0.64 0.36
V5  0.0 0.0 0.0 0.7 0.49 0.51
V6  0.0 0.0 0.0 0.6 0.36 0.64
V7  0.0 0.8 0.0 0.0 0.64 0.36
V8  0.0 0.7 0.0 0.0 0.49 0.51
V9  0.0 0.6 0.0 0.0 0.36 0.64
V10 0.0 0.0 0.8 0.0 0.64 0.36
V11 0.0 0.0 0.7 0.0 0.49 0.51
V12 0.0 0.0 0.6 0.0 0.36 0.64
```

```
                       MR3  MR1  MR2  MR4
SS loadings           1.49 1.49 1.49 1.49
Proportion Var        0.12 0.12 0.12 0.12
Cumulative Var        0.12 0.25 0.37 0.50
Proportion Explained  0.25 0.25 0.25 0.25
Cumulative Proportion 0.25 0.50 0.75 1.00

 With factor correlations of
     MR3  MR1  MR2  MR4
MR3 1.00 0.64 0.51 0.80
MR1 0.64 1.00 0.80 0.80
MR2 0.51 0.80 1.00 0.64
MR4 0.80 0.80 0.64 1.00
```

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., & Fox, J. (2011). Openmx: An open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306–317.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–634.

Fox, J., Nie, Z., & Byrnes, J. (2013). *sem: Structural Equation Models*. R package version 3.1-3.

Holzinger, K. & Swineford, F. (1939). A study in factor analysis: the stability of a bi-factor solution. Department of Education, University of Chicago, Chicago, Il.

Joreskog, K. G. & Sorbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Lisrel 8: Lawrence Erlbaum Associates, Inc.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports Monograph Supplement 9*, *3*, 635–694.

Narayanan, A. (2012). A review of eight software packages for structural equation modeling. *The American Statistician*, *66*(2), 129–138.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Zhang, Z., McArdle, J., Hamagami, A., & Grimm, . K. (2016). *RAMpath: Structural Equation Modeling Using the Reticular Action Model (RAM) Notation*. R package version 0.4.