

# Using the psych package to generate and test structural models

William Revelle

December 21, 2018

## Contents

<b>1</b>	<b>The psych package</b>	<b>3</b>
1.1	Preface . . . . .	3
1.2	Creating and modeling structural relations . . . . .	3
<b>2</b>	<b>Functions for generating correlational matrices with a particular structure</b>	<b>4</b>
2.1	sim.congeneric . . . . .	5
2.2	sim.hierarchical . . . . .	7
2.3	sim.item and sim.circ . . . . .	8
2.4	sim.structure . . . . .	8
2.4.1	$\mathbf{f}_x$ is a vector implies a congeneric model . . . . .	8
2.4.2	$\mathbf{f}_x$ is a matrix implies an independent factors model: . . . . .	11
2.4.3	$\mathbf{f}_x$ is a matrix and $\Phi \neq I$ is a correlated factors model . . . . .	11
2.4.4	$\mathbf{f}_x$ and $\mathbf{f}_y$ are matrices, and $\Phi \neq I$ represents their correlations . . . . .	15
2.4.5	A hierarchical structure among the latent predictors. . . . .	16
<b>3</b>	<b>Exploratory functions for analyzing structure</b>	<b>16</b>
3.1	Exploratory simple structure models . . . . .	19
3.2	Exploratory hierarchical models . . . . .	23
3.2.1	A bifactor solution . . . . .	23
3.2.2	A hierarchical solution . . . . .	23
<b>4</b>	<b>Exploratory Structural Equation Modeling (ESEM)</b>	<b>28</b>
<b>5</b>	<b>Confirmatory models</b>	<b>29</b>
5.1	Using psych as a front end for the sem package . . . . .	31
5.2	Testing a congeneric model versus a tau equivalent model . . . . .	31

5.3	Testing the dimensionality of a hierarchical data set by creating the model .	33
5.4	Testing the dimensionality based upon an exploratory analysis . . . . .	35
5.5	Specifying a three factor model . . . . .	35
5.6	Allowing for an oblique solution . . . . .	36
5.7	Extract a bifactor solution using <b>omega</b> and then test that model using <b>sem</b>	38
5.7.1	sem of Thurstone 9 variable problem . . . . .	38
5.8	Examining a hierarchical solution . . . . .	41
5.9	Estimating Omega using EFA followed by CFA . . . . .	44
<b>6</b>	<b>Summary and conclusion</b>	<b>48</b>

# 1 The psych package

## 1.1 Preface

The *psych* package (Revelle, 2014) has been developed to include those functions most useful for teaching and learning basic psychometrics and personality theory. Functions have been developed for many parts of the analysis of test data, including basic descriptive statistics (`describe` and `pairs.panels`), dimensionality analysis (`ICLUST`, `VSS`, `principal`, `factor.pa`), reliability analysis (`omega`, `guttman`) and eventual scale construction (`cluster.cor`, `score.items`). The use of these and other functions is described in more detail in the accompanying vignette ([overview.pdf](#)) as well as in the complete user's manual and the relevant help pages. (These vignettes are also available at <https://personality-project.org/r/overview.pdf>) and [https://personality-project.org/r/psych\\_for\\_sem.pdf](https://personality-project.org/r/psych_for_sem.pdf)).

This vignette is concerned with the problem of modeling structural data and using the *psych* package as a front end for the much more powerful *sem* package of John Fox (Fox, 2006, 2009; Fox et al., 2013). Future releases of this vignette will include examples for using the *lavaan* package of Yves Rosseel (Rosseel, 2012).

The first section discusses how to simulate particular latent variable structures. The second considers several Exploratory Factor Analysis (EFA) solutions to these problems. The third section considers how to do confirmatory factor analysis and structural equation modeling using the *sem* package but with the input prepared using functions in the *psych* package.

## 1.2 Creating and modeling structural relations

One common application of *psych* is the creation of simulated data matrices with particular structures to use as examples for principal components analysis, factor analysis, cluster analysis, and structural equation modeling. This vignette describes some of the functions used for creating, analyzing, and displaying such data sets. The examples use two other packages: *Rgraphviz* and *sem*. Although not required to use the *psych* package, *sem* is required for these examples. Although *Rgraphviz* had been used for the graphical displays, it has now been replaced with graphical functions within *psych*. The analyses themselves require only the *sem* package to do the structural modeling.

## 2 Functions for generating correlational matrices with a particular structure

The `sim` family of functions create data sets with particular structure. Most of these functions have default values that will produce useful examples. Although graphical summaries of these structures will be shown here, some of the options of the graphical displays will be discussed in a later section.

The `sim` functions include:

**`sim.structure`** A function to combine a measurement and structural model into one data matrix. Useful for understanding structural equation models. Combined with `structure.diagram` to see the proposed structure.

**`sim.congeneric`** A function to create congeneric items/tests for demonstrating classical test theory. This is just a special case of `sim.structure`.

**`sim.hierarchical`** A function to create data with a hierarchical (bifactor) structure.

**`sim.general`** A function to simulate a general factor and multiple group factors. This is done in a somewhat more obvious, although less general, method than `sim.hierarchical`.

**`sim.item`** A function to create items that either have a simple structure or a circumplex structure.

**`sim.circ`** Create data with a circumplex structure.

**`sim.dichot`** Create dichotomous item data with a simple or circumplex structure.

**`sim.minor`** Create a factor structure for `nvar` variables defined by `nfact` major factors and  $\frac{nvar}{2}$  “minor” factors for `n` observations.

**`sim.parallel`** Create a number of simulated data sets using `sim.minor` to show how parallel analysis works.

**`sim.rasch`** Create IRT data following a Rasch model.

**`sim.irt`** Create a two parameter IRT logistic (2PL) model.

**`sim.anova`** Simulate a 3 way balanced ANOVA or linear model, with or without repeated measures. Useful for teaching courses in research methods.

To make these examples replicable for readers, all simulations are prefaced by setting the random seed to a fixed (and for some, memorable) number ([Adams, 1980](#)). For normal use of the simulations, this is not necessary.

## 2.1 sim.congeneric

Classical test theory considers tests to be *tau* equivalent if they have the same covariance with a vector of latent true scores, but perhaps different error variances. Tests are considered *congeneric* if they each have the same true score component (perhaps to a different degree) and independent error components. The `sim.congeneric` function may be used to generate either structure.

The first example considers four tests with equal loadings on a latent factor (that is, a  $\tau$  equivalent model). If the number of subjects is not specified, a population correlation matrix will be generated. If N is specified, then the sample correlation matrix is returned. If the “short” option is FALSE, then the population matrix, sample matrix, and sample data are all returned as elements of a list.

```
> library(psych)
> set.seed(42)
> tau <- sim.congeneric(loads=c(.8,.8,.8,.8)) #population values
> tau.samp <- sim.congeneric(loads=c(.8,.8,.8,.8),N=100) # sample correlation matrix for 100 cases
> round(tau.samp,2)

      V1  V2  V3  V4
V1 1.00 0.68 0.72 0.66
V2 0.68 1.00 0.65 0.67
V3 0.72 0.65 1.00 0.76
V4 0.66 0.67 0.76 1.00

> tau.samp <- sim.congeneric(loads=c(.8,.8,.8,.8),N=100, short=FALSE)
> tau.samp

Call: NULL

$model (Population correlation matrix)
      V1  V2  V3  V4
V1 1.00 0.64 0.64 0.64
V2 0.64 1.00 0.64 0.64
V3 0.64 0.64 1.00 0.64
V4 0.64 0.64 0.64 1.00

$r (Sample correlation matrix for sample size = 100 )
      V1  V2  V3  V4
V1 1.00 0.70 0.62 0.58
V2 0.70 1.00 0.65 0.64
V3 0.62 0.65 1.00 0.59
V4 0.58 0.64 0.59 1.00

> dim(tau.samp$observed)

[1] 100  4
```

In this last case, the generated data are retrieved from `tau.samp$observed`. Congeneric data are created by specifying unequal loading values. The default values are loadings of `c(.8,.7,.6,.5)`. As seen in Figure 1, tau equivalence is the special case where all paths are equal.

```

> cong <- sim.congeneric(N=100)
> round(cong,2)

      V1  V2  V3  V4
V1 1.00 0.57 0.53 0.46
V2 0.57 1.00 0.35 0.41
V3 0.53 0.35 1.00 0.43
V4 0.46 0.41 0.43 1.00

> #plot.new()
> m1 <- structure.diagram(c("a","b","c","d"))

```

### Structural model

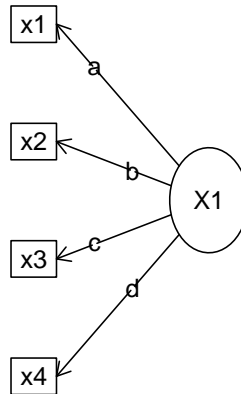


Figure 1: Tau equivalent tests are special cases of congeneric tests. Tau equivalence assumes  $a=b=c=d$

## 2.2 `sim.hierarchical`

The previous function, `sim.congeneric`, is used when one factor accounts for the pattern of correlations. A slightly more complicated model is when one broad factor and several narrower factors are observed. An example of this structure might be the structure of mental abilities, where there is a broad factor of general ability and several narrower factors (e.g., spatial ability, verbal ability, working memory capacity). Another example is in the measure of psychopathology where a broad general factor of neuroticism is seen along with more specific anxiety, depression, and aggression factors. This kind of structure may be simulated with `sim.hierarchical` specifying the loadings of each sub factor on a general factor (the g-loadings) as well as the loadings of individual items on the lower order factors (the f-loadings). An early paper describing a *bifactor* structure was by [Holzinger and Swineford \(1937\)](#). A helpful description of what makes a good general factor is that of [Jensen and Weng \(1994\)](#).

For those who prefer real data to simulated data, six data sets are included in the `bifactor` data set. One is the original 14 variable problem of [Holzinger and Swineford \(1937\)](#) (`holzinger`), a second is a nine variable problem adapted by [Bechtoldt \(1961\)](#) from [Thurstone and Thurstone \(1941\)](#) (the data set is used as an example in the SAS manual and discussed in great detail by [McDonald \(1999\)](#)), a third is from a recent paper by [Reise et al. \(2007\)](#) with 16 measures of patient reports of interactions with their health care provider.

```
> set.seed(42)
> gload=matrix(c(.9,.8,.7),nrow=3)
> fload <- matrix(c(.8,.7,.6,rep(0,9),.7,.6,.5,
+ rep(0,9),.7,.6,.4), ncol=3)
> fload #echo it to see the structureSw

      [,1] [,2] [,3]
[1,]  0.8  0.0  0.0
[2,]  0.7  0.0  0.0
[3,]  0.6  0.0  0.0
[4,]  0.0  0.7  0.0
[5,]  0.0  0.6  0.0
[6,]  0.0  0.5  0.0
[7,]  0.0  0.0  0.7
[8,]  0.0  0.0  0.6
[9,]  0.0  0.0  0.4

> bifact <- sim.hierarchical(gload=gload,fload=fload)
> round(bifact,2)

      V1  V2  V3  V4  V5  V6  V7  V8  V9
V1 1.00 0.56 0.48 0.40 0.35 0.29 0.35 0.30 0.20
V2 0.56 1.00 0.42 0.35 0.30 0.25 0.31 0.26 0.18
V3 0.48 0.42 1.00 0.30 0.26 0.22 0.26 0.23 0.15
V4 0.40 0.35 0.30 1.00 0.42 0.35 0.27 0.24 0.16
V5 0.35 0.30 0.26 0.42 1.00 0.30 0.24 0.20 0.13
V6 0.29 0.25 0.22 0.35 0.30 1.00 0.20 0.17 0.11
V7 0.35 0.31 0.26 0.27 0.24 0.20 1.00 0.42 0.28
```

```
V8 0.30 0.26 0.23 0.24 0.20 0.17 0.42 1.00 0.24
V9 0.20 0.18 0.15 0.16 0.13 0.11 0.28 0.24 1.00
```

These data can be represented as either a *bifactor* (Figure 2 panel A) or *hierarchical* (Figure 2 Panel B) factor solution. The analysis was done with the `omega` function.

### 2.3 `sim.item` and `sim.circ`

Many personality questionnaires are thought to represent multiple, independent factors. A particularly interesting case is when there are two factors and the items either have *simple structure* or *circumplex structure*. Examples of such items with a circumplex structure are measures of emotion (Rafaeli and Revelle, 2006) where many different emotion terms can be arranged in a two dimensional space, but where there is no obvious clustering of items. Typical personality scales are constructed to have simple structure, where items load on one and only one factor.

An additional challenge to measurement with emotion or personality items is that the items can be highly skewed and are assessed with a small number of discrete categories (do not agree, somewhat agree, strongly agree).

The more general `sim.item` function, and the more specific, `sim.circ` functions simulate items with a two dimensional structure, with or without skew, and varying the number of categories for the items. An example of a circumplex structure is shown in Figure 3

### 2.4 `sim.structure`

A more general case is to consider three matrices,  $\mathbf{f}_x, \phi_{xy}, \mathbf{f}_y$  which describe, in turn, a measurement model of x variables,  $\mathbf{f}_x$ , a measurement model of y variables,  $\mathbf{f}_y$ , and a covariance matrix between and within the two sets of factors. If  $\mathbf{f}_x$  is a vector and  $\mathbf{f}_y$  and  $\mathbf{phi}_{xy}$  are NULL, then this is just the congeneric model. If  $\mathbf{f}_x$  is a matrix of loadings with n rows and c columns, then this is a measurement model for n variables across c factors. If  $\mathbf{phi}_{xy}$  is not null, but  $\mathbf{f}_y$  is NULL, then the factors in  $\mathbf{f}_x$  are correlated. Finally, if all three matrices are not NULL, then the data show the standard linear structural relations (LISREL) structure.

Consider the following examples:

#### 2.4.1 $\mathbf{f}_x$ is a vector implies a congeneric model

```
> set.seed(42)
> fx <- c(.9, .8, .7, .6)
```



```

> op <- par(mfrow=c(1,2))
> m.bi <- omega(bifact,title="A bifactor model")
> m.hi <- omega(bifact,sl=FALSE,title="A hierarchical model")
> op <- par(mfrow = c(1,1))

```

**A bifactor model**

**A hierarchical model**

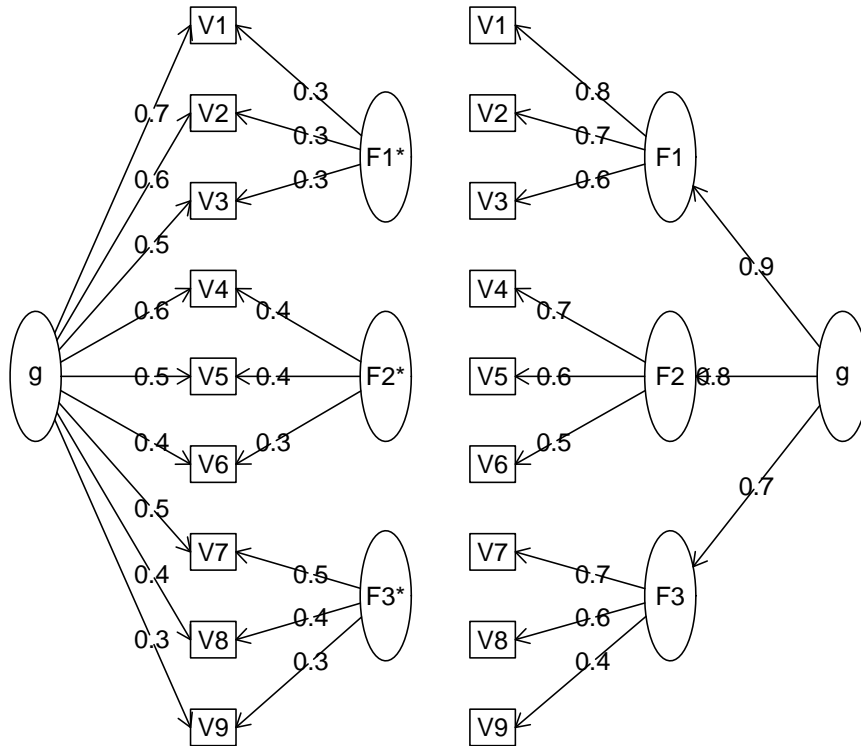


Figure 2: (Left panel) A bifactor solution represents each test in terms of a general factor and a residualized group factor. (Right Panel) A hierarchical factor solution has g as a second order factor accounting for the correlations between the first order factors

```

> circ <- sim.circ(16)
> f2 <- fa(circ,2)
> plot(f2,title="16 simulated variables in a circumplex pattern")

```

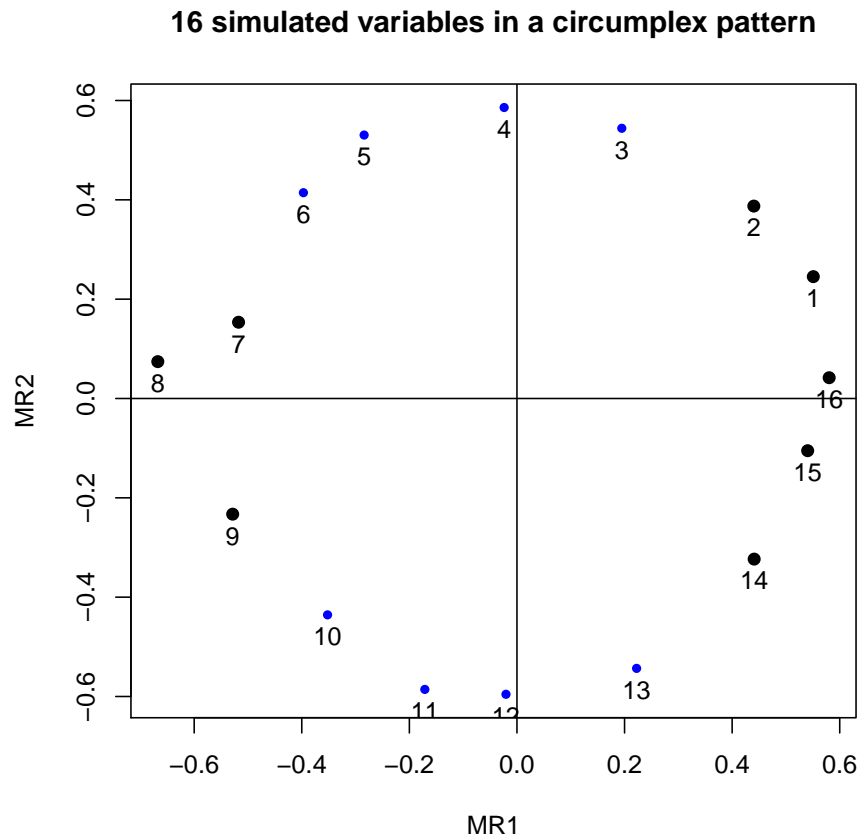


Figure 3: Emotion items or interpersonal items frequently show a circumplex structure. Data generated by `sim.circ` and factor loadings found by the principal axis algorithm using `factor.pa`.

```

> cong1 <- sim.structure(fx)
> cong1

Call: sim.structure(fx = fx)

$model (Population correlation matrix)
      V1  V2  V3  V4
V1 1.00 0.72 0.63 0.54
V2 0.72 1.00 0.56 0.48
V3 0.63 0.56 1.00 0.42
V4 0.54 0.48 0.42 1.00

$reliability (population reliability)
[1] 0.81 0.64 0.49 0.36

```

### 2.4.2 $\mathbf{f}_x$ is a matrix implies an independent factors model:

```

> set.seed(42)
> fx <- matrix(c(.9,.8,.7,rep(0,9),.7,.6,.5,rep(0,9),.6,.5,.4), ncol=3)
> three.fact <- sim.structure(fx)
> three.fact

Call: sim.structure(fx = fx)

$model (Population correlation matrix)
      V1  V2  V3  V4  V5  V6  V7  V8  V9
V1 1.00 0.72 0.63 0.00 0.00 0.00 0.00 0.0 0.00
V2 0.72 1.00 0.56 0.00 0.00 0.00 0.00 0.0 0.00
V3 0.63 0.56 1.00 0.00 0.00 0.00 0.00 0.0 0.00
V4 0.00 0.00 0.00 1.00 0.42 0.35 0.00 0.0 0.00
V5 0.00 0.00 0.00 0.42 1.00 0.30 0.00 0.0 0.00
V6 0.00 0.00 0.00 0.35 0.30 1.00 0.00 0.0 0.00
V7 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.3 0.24
V8 0.00 0.00 0.00 0.00 0.00 0.00 0.30 1.0 0.20
V9 0.00 0.00 0.00 0.00 0.00 0.00 0.24 0.2 1.00

$reliability (population reliability)
[1] 0.81 0.64 0.49 0.49 0.36 0.25 0.36 0.25 0.16

```

### 2.4.3 $\mathbf{f}_x$ is a matrix and $\Phi \neq I$ is a correlated factors model

```

> Phi = matrix(c(1,.5,.3,.5,1,.2,.3,.2,1), ncol=3)
> cor.f3 <- sim.structure(fx,Phi)
> fx

      [,1] [,2] [,3]
[1,] 0.9 0.0 0.0
[2,] 0.8 0.0 0.0
[3,] 0.7 0.0 0.0
[4,] 0.0 0.7 0.0
[5,] 0.0 0.6 0.0
[6,] 0.0 0.5 0.0
[7,] 0.0 0.0 0.6
[8,] 0.0 0.0 0.5
[9,] 0.0 0.0 0.4

```

### Structural model

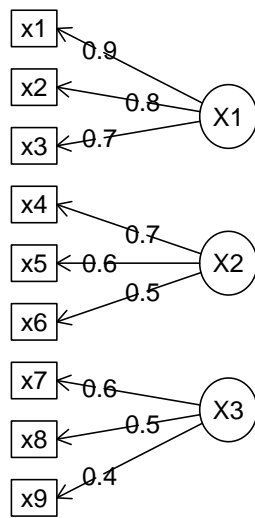


Figure 4: Three uncorrelated factors generated using the `sim.structure` function and drawn using `structure.diagram`.

```

> Phi

      [,1] [,2] [,3]
[1,]  1.0  0.5  0.3
[2,]  0.5  1.0  0.2
[3,]  0.3  0.2  1.0

> cor.f3

Call: sim.structure(fx = fx, Phi = Phi)

$model (Population correlation matrix)
      V1  V2  V3  V4  V5  V6  V7  V8  V9
V1 1.00 0.720 0.630 0.315 0.270 0.23 0.162 0.14 0.108
V2 0.72 1.000 0.560 0.280 0.240 0.20 0.144 0.12 0.096
V3 0.63 0.560 1.000 0.245 0.210 0.17 0.126 0.10 0.084
V4 0.32 0.280 0.245 1.000 0.420 0.35 0.084 0.07 0.056
V5 0.27 0.240 0.210 0.420 1.000 0.30 0.072 0.06 0.048
V6 0.23 0.200 0.175 0.350 0.300 1.00 0.060 0.05 0.040
V7 0.16 0.144 0.126 0.084 0.072 0.06 1.000 0.30 0.240
V8 0.14 0.120 0.105 0.070 0.060 0.05 0.300 1.00 0.200
V9 0.11 0.096 0.084 0.056 0.048 0.04 0.240 0.20 1.000

$reliability (population reliability)
[1] 0.81 0.64 0.49 0.49 0.36 0.25 0.36 0.25 0.16

```

**Using symbolic loadings and path coefficients** For some purposes, it is helpful not to specify particular values for the paths, but rather to think of them symbolically. This can be shown with symbolic loadings and path coefficients by using the `structure.list` and `phi.list` functions to create the `fx` and `Phi` matrices (Figure 5).

```

> fxs <- structure.list(9,list(F1=c(1,2,3),F2=c(4,5,6),F3=c(7,8,9)))
> Phis <- phi.list(3,list(F1=c(2,3),F2=c(1,3),F3=c(1,2)))
> fxs #show the matrix

      F1  F2  F3
[1,] "a1" "0"  "0"
[2,] "a2" "0"  "0"
[3,] "a3" "0"  "0"
[4,] "0"  "b4" "0"
[5,] "0"  "b5" "0"
[6,] "0"  "b6" "0"
[7,] "0"  "0"  "c7"
[8,] "0"  "0"  "c8"
[9,] "0"  "0"  "c9"

> Phis #show this one as well

      F1  F2  F3
F1 "1"  "rba" "rca"

```

```

F2 "rab" "1"    "rcb"
F3 "rac" "rbc" "1"

```

The `structure.list` and `phi.list` functions allow for creation of  $\mathbf{f}_x$ ,  $\mathbf{\Phi}$ , and  $\mathbf{f}_y$  matrices in a very compact form, just by specifying the relevant variables.

```

> #plot.new()
> corf3.mod <- structure.diagram(fxs,Phis)

```

### Structural model

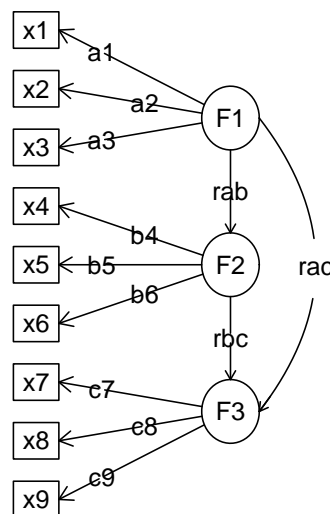


Figure 5: Three correlated factors with symbolic paths. Created using `structure.diagram` and `structure.list` and `phi.list` for ease of input.

**Drawing path models from Exploratory Factor Analysis solutions** Alternatively, this result can represent the estimated factor loadings and oblique correlations found using `factanal` (Maximum Likelihood factoring) or `fa` (Principal axis or minimum residual (minres) factoring) followed by a promax rotation using the `Promax` function (Figure 6.

Comparing this figure with the previous one (Figure 5), it will be seen that one path was dropped because it was less than the arbitrary “cut” value of .2.

```
> f3.p <- Promax(fa(cor.f3$model,3))
> #plot.new()
> mod.f3p <- structure.diagram(f3.p,cut=.2)
```

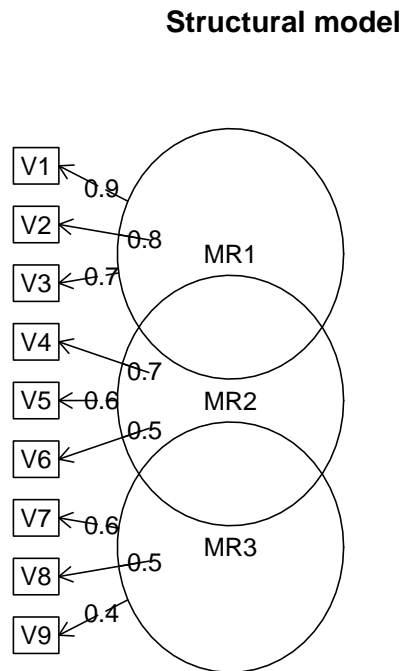


Figure 6: The empirically fitted structural model. Paths less than cut (.2 in this case, the default is .3) are not shown.

#### 2.4.4 $\mathbf{f}_x$ and $\mathbf{f}_y$ are matrices, and $\mathbf{\Phi} \neq \mathbf{I}$ represents their correlations

A more complicated model is when there is a  $\mathbf{f}_y$  vector or matrix representing a set of Y latent variables that are associated with a set of y variables. In this case, the  $\mathbf{\Phi}$  matrix is a set of correlations within the X set and between the X and Y set.

```

> set.seed(42)
> fx <- matrix(c(.9,.8,.7,rep(0,9),.7,.6,.5,rep(0,9),.6,.5,.4), ncol=3)
> fy <- c(.6,.5,.4)
> Phi <- matrix(c(1,.48,.32,.4,.48,1,.32,.3,.32,.32,1,.2,.4,.3,.2,1), ncol=4)
> twelveV <- sim.structure(fx,Phi, fy)$model
> colnames(twelveV) <- rownames(twelveV) <- c(paste("x",1:9,sep=""),paste("y",1:3,sep=""))
> round(twelveV,2)

```

	x1	x2	x3	x4	x5	x6	x7	x8	x9	y1	y2	y3
x1	1.00	0.72	0.63	0.30	0.26	0.22	0.17	0.14	0.12	0.22	0.18	0.14
x2	0.72	1.00	0.56	0.27	0.23	0.19	0.15	0.13	0.10	0.19	0.16	0.13
x3	0.63	0.56	1.00	0.24	0.20	0.17	0.13	0.11	0.09	0.17	0.14	0.11
x4	0.30	0.27	0.24	1.00	0.42	0.35	0.13	0.11	0.09	0.13	0.10	0.08
x5	0.26	0.23	0.20	0.42	1.00	0.30	0.12	0.10	0.08	0.11	0.09	0.07
x6	0.22	0.19	0.17	0.35	0.30	1.00	0.10	0.08	0.06	0.09	0.08	0.06
x7	0.17	0.15	0.13	0.13	0.12	0.10	1.00	0.30	0.24	0.07	0.06	0.05
x8	0.14	0.13	0.11	0.11	0.10	0.08	0.30	1.00	0.20	0.06	0.05	0.04
x9	0.12	0.10	0.09	0.09	0.08	0.06	0.24	0.20	1.00	0.05	0.04	0.03
y1	0.22	0.19	0.17	0.13	0.11	0.09	0.07	0.06	0.05	1.00	0.30	0.24
y2	0.18	0.16	0.14	0.10	0.09	0.08	0.06	0.05	0.04	0.30	1.00	0.20
y3	0.14	0.13	0.11	0.08	0.07	0.06	0.05	0.04	0.03	0.24	0.20	1.00

Data with this structure may be created using the `sim.structure` function, and shown either with the numeric values or symbolically using the `structure.diagram` function (Figure 7).

```

> fxs <- structure.list(9,list(X1=c(1,2,3), X2 =c(4,5,6),X3 = c(7,8,9)))
> phi <- phi.list(4,list(F1=c(4),F2=c(4),F3=c(4),F4=c(1,2,3)))
> fyx <- structure.list(3,list(Y=c(1,2,3)),"Y")

```

#### 2.4.5 A hierarchical structure among the latent predictors.

Measures of intelligence and psychopathology frequently have a general factor as well as multiple group factors. The general factor then is thought to predict some dependent latent variable. Compare this with the previous model (see Figure 7).

These two models can be compared using structural modeling procedures (see below).

### 3 Exploratory functions for analyzing structure

Given correlation matrices such as those seen above for congeneric or bifactor models, the question becomes how best to estimate the underlying structure. Because these data sets were generated from a known model, the question becomes how well does a particular model recover the underlying structure.



```

> #plot.new()
> sg3 <- structure.diagram(fxs,phi,fyx)

```

### Structural model

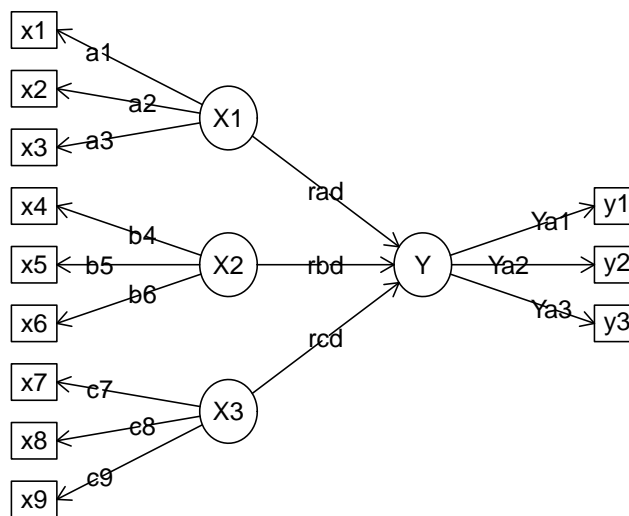


Figure 7: A symbolic structural model. Three independent latent variables are regressed on a latent Y.

```

> fxh <- structure.list(9,list(X1=c(1:3),X2=c(4:6),X3=c(7:9),g=NULL))
> fy <- structure.list(3,list(Y=c(1,2,3)))
> Phi <- diag(1,5,5)
> Phi[4,c(1:3)] <- letters[1:3]
> Phi[5,4] <- "r"
> #plot.new()
> hi.mod <-structure.diagram(fxh,Phi, fy)

```

### Structural model

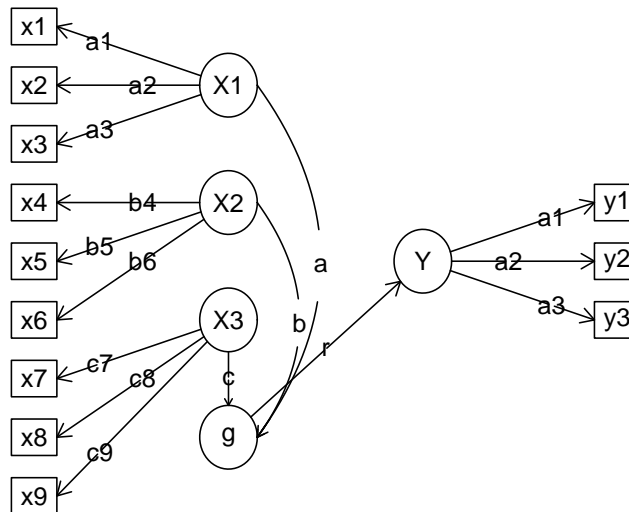


Figure 8: A symbolic structural model with a general factor and three group factors. The general factor is regressed on the latent Y variable.

### 3.1 Exploratory simple structure models

The technique of *principal components* provides a set of weighted linear composites that best approximates a particular correlation or covariance matrix. If these are then *rotated* to provide a more interpretable solution, the components are no longer the *principal* components. The `principal` function will extract the first `n` principal components (default value is 1) and if `n>1`, rotate to *simple structure* using a `varimax`, `quartimin`, or `Promax` criterion.

```
> principal(cong1$model)

Principal Components Analysis
Call: principal(r = cong1$model)
Standardized loadings (pattern matrix) based upon correlation matrix
      PC1    h2    u2 com
V1 0.89 0.80 0.20    1
V2 0.85 0.73 0.27    1
V3 0.80 0.64 0.36    1
V4 0.73 0.53 0.47    1

      PC1
SS loadings    2.69
Proportion Var 0.67

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

The root mean square of the residuals (RMSR) is 0.11

Fit based upon off diagonal values = 0.96

> fa(cong1$model)

Factor Analysis using method = minres
Call: fa(r = cong1$model)
Standardized loadings (pattern matrix) based upon correlation matrix
      MR1    h2    u2 com
V1 0.9 0.81 0.19    1
V2 0.8 0.64 0.36    1
V3 0.7 0.49 0.51    1
V4 0.6 0.36 0.64    1

      MR1
SS loadings    2.30
Proportion Var 0.57

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

The degrees of freedom for the null model are 6 and the objective function was 1.65
The degrees of freedom for the model are 2 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
```

Measures of factor score adequacy	
	MR1
Correlation of (regression) scores with factors	0.94
Multiple R square of scores with factors	0.88
Minimum correlation of possible factor scores	0.77

It is important to note that although the `principal` components function does not exactly reproduce the model parameters, the `factor.pa` function, implementing principal axes or minimum residual (minres) factor analysis, does.

Consider the case of three underlying factors as seen in the bifact example above. Because the number of observations is not specified, there is no associated  $\chi^2$  value. The `factor.congruence` function reports the cosine of the angle between the factors.

```
> pc3 <- principal(bifact,3)
> pa3 <- fa(bifact,3,fm="pa")
> ml3 <- fa(bifact,3,fm="ml")
> pc3
```

```
Principal Components Analysis
Call: principal(r = bifact, nfactors = 3)
Standardized loadings (pattern matrix) based upon correlation matrix
```

	RC1	RC3	RC2	h2	u2	com
V1	0.75	0.27	0.21	0.69	0.31	1.4
V2	0.76	0.21	0.16	0.64	0.36	1.2
V3	0.78	0.11	0.10	0.63	0.37	1.1
V4	0.29	0.69	0.15	0.59	0.41	1.5
V5	0.20	0.71	0.11	0.56	0.44	1.2
V6	0.07	0.76	0.08	0.59	0.41	1.0
V7	0.26	0.16	0.70	0.58	0.42	1.4
V8	0.20	0.11	0.71	0.55	0.45	1.2
V9	0.00	0.06	0.73	0.53	0.47	1.0

	RC1	RC3	RC2
SS loadings	1.99	1.73	1.64
Proportion Var	0.22	0.19	0.18
Cumulative Var	0.22	0.41	0.60
Proportion Explained	0.37	0.32	0.31
Cumulative Proportion	0.37	0.69	1.00

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

```
The root mean square of the residuals (RMSR) is 0.1
```

```
Fit based upon off diagonal values = 0.88
```

```
> pa3
```

```
Factor Analysis using method = pa
Call: fa(r = bifact, nfactors = 3, fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	PA1	PA3	PA2	h2	u2	com
V1	0.8	0.0	0.00	0.64	0.36	1
V2	0.7	0.0	0.00	0.49	0.51	1
V3	0.6	0.0	0.00	0.36	0.64	1
V4	0.0	0.7	0.00	0.49	0.51	1

```
V5 0.0 0.6 0.00 0.36 0.64 1
V6 0.0 0.5 0.00 0.25 0.75 1
V7 0.0 0.0 0.69 0.48 0.52 1
V8 0.0 0.0 0.61 0.36 0.64 1
V9 0.0 0.0 0.40 0.16 0.84 1
```

```
          PA1  PA3  PA2
SS loadings      1.49 1.10 1.01
Proportion Var    0.17 0.12 0.11
Cumulative Var    0.17 0.29 0.40
Proportion Explained 0.41 0.31 0.28
Cumulative Proportion 0.41 0.72 1.00
```

```
With factor correlations of
      PA1  PA3  PA2
PA1 1.00 0.72 0.63
PA3 0.72 1.00 0.56
PA2 0.63 0.56 1.00
```

```
Mean item complexity = 1
Test of the hypothesis that 3 factors are sufficient.
```

```
The degrees of freedom for the null model are 36 and the objective function was 1.88
The degrees of freedom for the model are 12 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
Measures of factor score adequacy
```

```
          PA1  PA3  PA2
Correlation of (regression) scores with factors 0.9 0.85 0.83
Multiple R square of scores with factors        0.8 0.72 0.69
Minimum correlation of possible factor scores    0.6 0.45 0.38
```

```
> ml3
```

```
Factor Analysis using method = ml
Call: fa(r = bifact, nfactors = 3, fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
      ML1 ML3 ML2  h2  u2 com
V1 0.8 0.0 0.0 0.64 0.36 1
V2 0.7 0.0 0.0 0.49 0.51 1
V3 0.6 0.0 0.0 0.36 0.64 1
V4 0.0 0.7 0.0 0.49 0.51 1
V5 0.0 0.6 0.0 0.36 0.64 1
V6 0.0 0.5 0.0 0.25 0.75 1
V7 0.0 0.0 0.7 0.49 0.51 1
V8 0.0 0.0 0.6 0.36 0.64 1
V9 0.0 0.0 0.4 0.16 0.84 1
```

```
      ML1  ML3  ML2
SS loadings      1.49 1.10 1.01
Proportion Var    0.17 0.12 0.11
Cumulative Var    0.17 0.29 0.40
Proportion Explained 0.41 0.31 0.28
Cumulative Proportion 0.41 0.72 1.00
```

```

With factor correlations of
      ML1 ML3 ML2
ML1 1.00 0.72 0.63
ML3 0.72 1.00 0.56
ML2 0.63 0.56 1.00

```

```

Mean item complexity = 1
Test of the hypothesis that 3 factors are sufficient.

```

```

The degrees of freedom for the null model are 36 and the objective function was 1.88
The degrees of freedom for the model are 12 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
Measures of factor score adequacy

```

	ML1	ML3	ML2
Correlation of (regression) scores with factors	0.90	0.85	0.83
Multiple R square of scores with factors	0.80	0.72	0.69
Minimum correlation of possible factor scores	0.61	0.45	0.38

```

> factor.congruence(list(pc3,pa3,ml3))

```

	RC1	RC3	RC2	PA1	PA3	PA2	ML1	ML3	ML2
RC1	1.00	0.49	0.42	0.93	0.24	0.21	0.93	0.24	0.21
RC3	0.49	1.00	0.35	0.27	0.94	0.15	0.27	0.93	0.15
RC2	0.42	0.35	1.00	0.22	0.16	0.94	0.22	0.16	0.94
PA1	0.93	0.27	0.22	1.00	0.00	0.00	1.00	0.00	0.00
PA3	0.24	0.94	0.16	0.00	1.00	0.00	0.00	1.00	0.00
PA2	0.21	0.15	0.94	0.00	0.00	1.00	0.00	0.00	1.00
ML1	0.93	0.27	0.22	1.00	0.00	0.00	1.00	0.00	0.00
ML3	0.24	0.93	0.16	0.00	1.00	0.00	0.00	1.00	0.00
ML2	0.21	0.15	0.94	0.00	0.00	1.00	0.00	0.00	1.00

By default, all three of these procedures use the varimax rotation criterion. Perhaps it is useful to apply an oblique transformation such as **Promax** or **oblimin** to the results. The **Promax** function in *psych* differs slightly from the standard **promax** in that it reports the factor intercorrelations.

```

> ml3p <- Promax(ml3)
> ml3p

```

```

Call: NULL

```

```

Standardized loadings (pattern matrix) based upon correlation matrix

```

	ML1	ML3	ML2	h2	u2
V1	0.8	0.0	0.0	0.64	0.36
V2	0.7	0.0	0.0	0.49	0.51
V3	0.6	0.0	0.0	0.36	0.64
V4	0.0	0.7	0.0	0.49	0.51
V5	0.0	0.6	0.0	0.36	0.64
V6	0.0	0.5	0.0	0.25	0.75
V7	0.0	0.0	0.7	0.49	0.51
V8	0.0	0.0	0.6	0.36	0.64
V9	0.0	0.0	0.4	0.16	0.84

	ML1	ML3	ML2
SS loadings	1.49	1.10	1.01
Proportion Var	0.17	0.12	0.11

Cumulative Var		0.17	0.29	0.40
Proportion Explained		0.41	0.31	0.28
Cumulative Proportion		0.41	0.72	1.00
	ML1	ML3	ML2	
ML1	1	0	0	
ML3	0	1	0	
ML2	0	0	1	

## 3.2 Exploratory hierarchical models

In addition to the conventional oblique factor model, an alternative model is to consider the correlations between the factors to represent a higher order factor. This can be shown either as a *bifactor* solution [Holzinger and Swineford \(1937\)](#); [Schmid and Leiman \(1957\)](#) with a general factor for all variables and a set of residualized group factors, or as a hierarchical structure. An exploratory hierarchical model can be applied to this kind of data structure using the `omega` function. Graphic options include drawing a Schmid - Leiman bifactor solution (Figure 9) or drawing a hierarchical factor solution (Figure 10).

### 3.2.1 A bifactor solution

The *bifactor* solution has a general factor loading for each variable as well as a set of residual group factors. This approach has been used extensively in the measurement of ability and has more recently been used in the measure of psychopathology ([Reise et al., 2007](#)). Data sets included in the `bifactor` data include the original ([Holzinger and Swineford, 1937](#)) data set (`holzinger`) as well as a set from [Reise et al. \(2007\)](#) (`reise`) and a nine variable problem from Thurstone.

### 3.2.2 A hierarchical solution

Both of these graphical representations are reflected in the output of the `omega` function. The first was done using a Schmid-Leiman transformation, the second was not. As will be seen later, the objects returned from these two analyses may be used as models for a `sem` analysis. It is also useful to examine the estimates of reliability reported by `omega`.

```
> om.bi
```

```
Omega
Call: omega(m = bifact)
Alpha:                0.78
G.6:                  0.78
Omega Hierarchical:    0.7
Omega H asymptotic:    0.85
```

```
> om.bi <- omega(bifact)
```

**Omega**

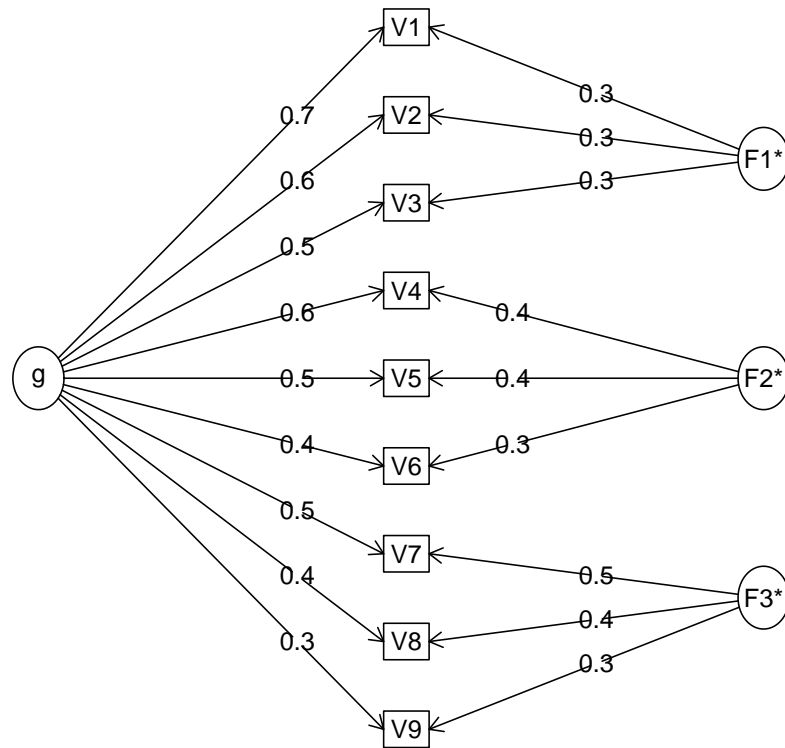


Figure 9: An exploratory bifactor solution to the nine variable problem



```
> om.hi <- omega(bifact,sl=FALSE)
```

**Omega**

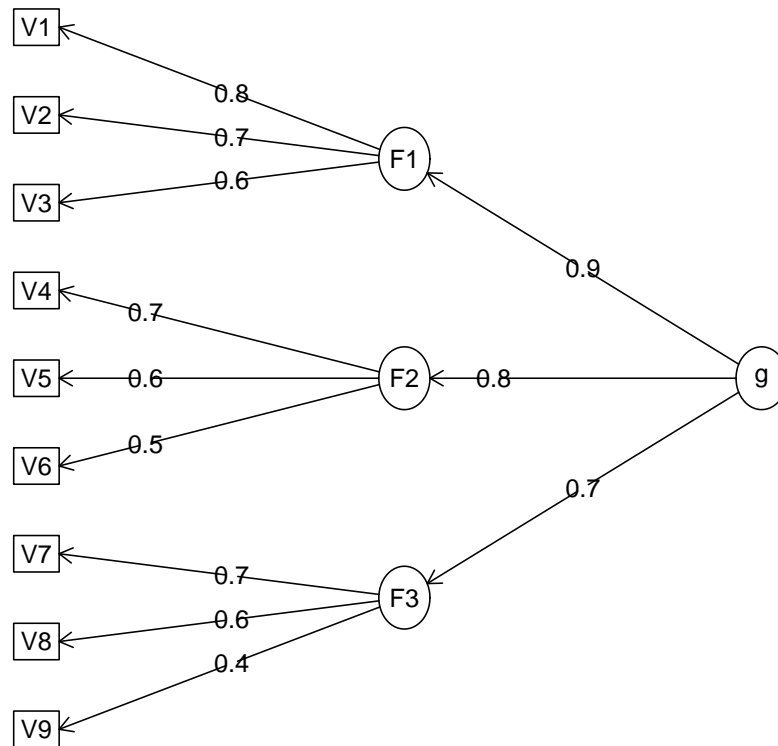


Figure 10: An exploratory hierarchical solution to the nine variable problem.

Omega Total 0.82

Schmid Leiman Factor loadings greater than 0.2

	g	F1*	F2*	F3*	h2	u2	p2
V1	0.72	0.35			0.64	0.36	0.81
V2	0.63	0.31			0.49	0.51	0.81
V3	0.54	0.26			0.36	0.64	0.81
V4	0.56		0.42		0.49	0.51	0.64
V5	0.48		0.36		0.36	0.64	0.64
V6	0.40		0.30		0.25	0.75	0.64
V7	0.49			0.50	0.49	0.51	0.49
V8	0.42			0.43	0.36	0.64	0.49
V9	0.28			0.29	0.16	0.84	0.49

With eigenvalues of:

	g	F1*	F2*	F3*
2.41	0.28	0.40	0.52	

general/max 4.67 max/min = 1.82

mean percent general = 0.65 with sd = 0.14 and cv of 0.21

Explained Common Variance of the general factor = 0.67

The degrees of freedom are 12 and the fit is 0

The root mean square of the residuals is 0

The df corrected root mean square of the residuals is 0

Compare this with the adequacy of just a general factor and no group factors

The degrees of freedom for just the general factor are 27 and the fit is 0.23

The root mean square of the residuals is 0.07

The df corrected root mean square of the residuals is 0.08

Measures of factor score adequacy

	g	F1*	F2*	F3*
Correlation of scores with factors	0.86	0.47	0.57	0.64
Multiple R square of scores with factors	0.74	0.22	0.33	0.41
Minimum correlation of factor score estimates	0.47	-0.56	-0.35	-0.18

Total, General and Subset omega for each subset

g	F1*	F2*	F3*
---	-----	-----	-----

Omega total for total scores and subscales	0.82	0.74	0.63	0.59
Omega general for total scores and subscales	0.70	0.60	0.40	0.29
Omega group for total scores and subscales	0.12	0.14	0.23	0.30

Yet one more way to treat the hierarchical structure of a data set is to consider hierarchical cluster analysis using the ICLUST algorithm (Figure 11). ICLUST is most appropriate for forming item composites.

### Hierarchical cluster analysis of bifact data

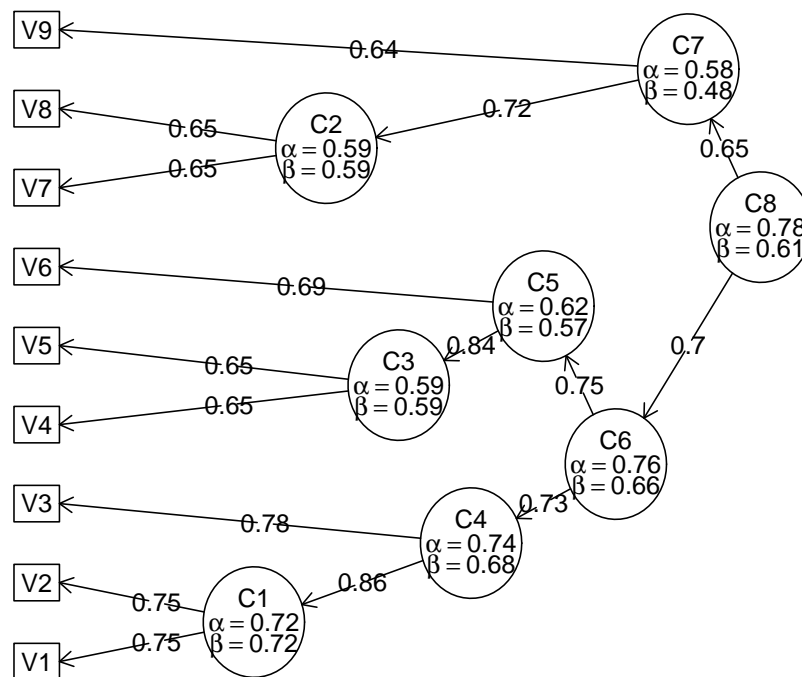


Figure 11: A hierarchical cluster analysis of the bifact data set using ICLUST

## 4 Exploratory Structural Equation Modeling (ESEM)

Traditional Exploratory Factor Analysis (EFA) examines how latent variables can account for the correlations within a data set. All loadings and cross loadings are found and rotation is done to some approximation of simple structure. Traditional Confirmatory Factor Analysis (CFA) tests such models by fitting just a limited number of loadings and typically does not allow any (or many) cross loadings. Structural Equation Modeling then applies two such measurement models, one to a set of X variables, another to a set of Y variables, and then tries to estimate the correlation between these two sets of latent variables. (Some SEM procedures estimate all the parameters from the same model, thus making the loadings in set Y affect those in set X.) It is possible to do a similar, exploratory modeling (ESEM) by conducting two Exploratory Factor Analyses, one in set X, one in set Y, and then finding the correlations of the X factors with the Y factors, as well as the correlations of the Y variables with the X factors and the X variables with the Y factors.

Consider the simulated data set of three ability variables, two motivational variables, and three outcome variables:

```
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 0.50 0.42 0.34
nach 0.00 0.00 0.48 1.00 -0.56 0.34 0.28 0.22
Anx  0.00 0.00 -0.42 -0.56 1.00 -0.29 -0.24 -0.20
gpa  0.38 0.34 0.50 0.34 -0.29 1.00 0.30 0.24
Pre  0.32 0.28 0.42 0.28 -0.24 0.30 1.00 0.20
MA   0.25 0.22 0.34 0.22 -0.20 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
```

We can fit this by using the `esem` function and then draw the solution (see Figure 12) using the `esem.diagram` function (which is normally called automatically by `esem`).

```
Exploratory Structural Equation Modeling Analysis using method = minres
Call: esim(r = gre.gpa$model, varsX = 1:5, varsY = 6:8, nfX = 2, nfY = 1,
  n.obs = 1000, plot = FALSE)
```

For the 'X' set:

	X1	X2
V	0.91	-0.06
Q	0.81	-0.05
A	0.53	0.57
nach	-0.10	0.81
Anx	0.08	-0.71

For the 'Y' set:

	Y1
gpa	0.6
Pre	0.5
MA	0.4

Correlations between the X and Y sets.

	X1	X2	Y1
X1	1.00	0.19	0.68
X2	0.19	1.00	0.67
Y1	0.68	0.67	1.00

The degrees of freedom for the null model are 56 and the empirical chi square function was 21.83  
The degrees of freedom for the model are 7 and the empirical chi square function was 21.83  
with prob < 0.0027

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

The df corrected root mean square of the residuals is 0.04

with the empirical chi square 21.83 with prob < 0.0027

The total number of observations was 1000 with fitted Chi Square = 2175.06 with prob < 0.0027

Empirical BIC = -26.53

ESABIC = -4.29

Fit based upon off diagonal values = 1

To see the item loadings for the X and Y sets combined, and the associated fa output, print

## 5 Confirmatory models

Although the exploratory models shown above do estimate the goodness of fit of the model and compare the residual matrix to a zero matrix using a  $\chi^2$  statistic, they estimate more

### Exploratory Structural Model

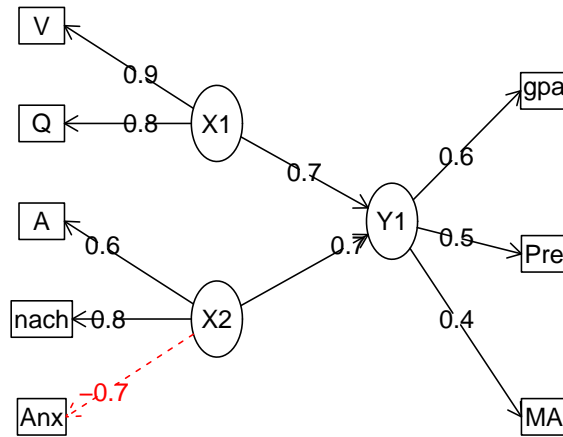


Figure 12: An example of a Exploratory Structure Equation Model.

parameters than are necessary if there is indeed a simple structure, and they do not allow for tests of competing models. The `sem` function in the *sem* package by John Fox allows for confirmatory tests. The interested reader is referred to the *sem* manual for more detail (Fox et al., 2013).

## 5.1 Using psych as a front end for the sem package

Because preparation of the `sem` commands is a bit tedious, several of the *psych* package functions have been designed to provide the appropriate commands. That is, the functions `structure.list`, `phi.list`, `structure.diagram`, `structure.sem`, and `omega.graph` may be used as a front end to `sem`. Usually with no modification, but sometimes with just slight modification, the model output from the `structure.diagram`, `structure.sem`, and `omega.graph` functions is meant to provide the appropriate commands for `sem`.

## 5.2 Testing a congeneric model versus a tau equivalent model

The congeneric model is a one factor model with possibly unequal factor loadings. The tau equivalent model model is one with equal factor loadings. Tests for these may be done by creating the appropriate structures. The `structure.graph` function which requires *Rgraphviz*, or `structure.diagram` or the `structure.sem` functions which do not may be used.

The following example tests the hypothesis (which is actually false) that the correlations found in the cong data set (see 2.1) are tau equivalent. Because the variable labels in that data set were V1 ... V4, we specify the labels to match those.

```
> library(sem)
> mod.tau <- structure.sem(c("a","a","a","a"),labels=paste("V",1:4,sep=""))
> mod.tau #show it
```

	Path	Parameter	Value
[1,]	"X1->V1"	"a"	NA
[2,]	"X1->V2"	"a"	NA
[3,]	"X1->V3"	"a"	NA
[4,]	"X1->V4"	"a"	NA
[5,]	"V1<->V1"	"x1e"	NA
[6,]	"V2<->V2"	"x2e"	NA
[7,]	"V3<->V3"	"x3e"	NA
[8,]	"V4<->V4"	"x4e"	NA
[9,]	"X1<->X1"	NA	"1"

```
attr("class")
[1] "mod"

> sem.tau <- sem(mod.tau,cong,100)
> summary(sem.tau,digits=2)
```

Model	Chisquare	Df	Pr(>Chisq)
	6.593496	5	0.2526696

```
AIC = 16.5935
```

```

BIC = -16.43236

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.03157 -0.44199 -0.25025 -0.07905  0.52702  0.88767

R-square for Endogenous Variables
  V1    V2    V3    V4
0.5245 0.4592 0.4500 0.4432

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a  0.6865481 0.06299180 10.899007 1.165221e-27 V1 <--- X1
x1e 0.4272839 0.08086561  5.283876 1.264786e-07 V1 <--> V1
x2e 0.5551772 0.09751222  5.693411 1.245260e-08 V2 <--> V2
x3e 0.5760999 0.10030974  5.743210 9.289853e-09 V3 <--> V3
x4e 0.5920607 0.10245375  5.778809 7.523134e-09 V4 <--> V4

Iterations = 11

Test whether the data are congeneric. That is, whether a one factor model fits. Compare
this to the prior model using the anova function.

> mod.cong <- structure.sem(c("a", "b", "c", "d"), labels=paste("V", 1:4, sep=""))
> mod.cong #show the model

      Path      Parameter Value
[1,] "X1->V1" "a"          NA
[2,] "X1->V2" "b"          NA
[3,] "X1->V3" "c"          NA
[4,] "X1->V4" "d"          NA
[5,] "V1<->V1" "x1e"       NA
[6,] "V2<->V2" "x2e"       NA
[7,] "V3<->V3" "x3e"       NA
[8,] "V4<->V4" "x4e"       NA
[9,] "X1<->X1" NA          "1"
attr(,"class")
[1] "mod"

> sem.cong <- sem(mod.cong, cong, 100)
> summary(sem.cong, digits=2)

Model Chisquare = 2.941678 Df = 2 Pr(>Chisq) = 0.2297327
AIC = 18.94168
BIC = -6.268663

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.5739 -0.0699  0.0339  0.0113  0.1605  0.5412

R-square for Endogenous Variables
  V1    V2    V3    V4
0.6880 0.4384 0.3942 0.3524

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a  0.8294562 0.09786772  8.475279 2.345174e-17 V1 <--- X1
b  0.6621164 0.10066777  6.577243 4.792500e-11 V2 <--- X1
c  0.6278767 0.10146860  6.187891 6.097433e-10 V3 <--- X1

```



```

d    0.5936695 0.10238816 5.798224 6.702094e-09 V4 <--- X1
x1e 0.3120026 0.10044870 3.106089 1.895798e-03 V1 <--> V1
x2e 0.5616018 0.10154893 5.530356 3.195810e-08 V2 <--> V2
x3e 0.6057707 0.10421285 5.812822 6.142832e-09 V3 <--> V3
x4e 0.6475566 0.10732995 6.033326 1.606191e-09 V4 <--> V4

Iterations = 12

> anova(sem.cong,sem.tau) #test the difference between the two models

LR Test for Difference Between Models

      Model Df Model Chisq Df LR Chisq Pr(>Chisq)
sem.cong      2      2.9417
sem.tau       5      6.5935  3    3.6518    0.3016

```

The `anova` comparison of the congeneric versus tau equivalent model shows that the change in  $\chi^2$  is significant given the change in degrees of freedom.

### 5.3 Testing the dimensionality of a hierarchical data set by creating the model

The bifactor correlation matrix was created to represent a hierarchical structure. Various confirmatory models can be applied to this matrix.

The first example creates the model directly, the next several create models based upon exploratory factor analyses. `mod.one` is a congeneric model of one factor accounting for the relationships between the nine variables. Although not correct, with 100 subjects, this model can not be rejected. However, an examination of the residuals suggests serious problems with the model.

```

> mod.one <- structure.sem(letters[1:9],labels=paste("V",1:9,sep=""))
> mod.one #show the model

      Path      Parameter Value
[1,] "X1->V1"    "a"      NA
[2,] "X1->V2"    "b"      NA
[3,] "X1->V3"    "c"      NA
[4,] "X1->V4"    "d"      NA
[5,] "X1->V5"    "e"      NA
[6,] "X1->V6"    "f"      NA
[7,] "X1->V7"    "g"      NA
[8,] "X1->V8"    "h"      NA
[9,] "X1->V9"    "i"      NA
[10,] "V1<->V1"  "x1e"     NA
[11,] "V2<->V2"  "x2e"     NA
[12,] "V3<->V3"  "x3e"     NA
[13,] "V4<->V4"  "x4e"     NA
[14,] "V5<->V5"  "x5e"     NA
[15,] "V6<->V6"  "x6e"     NA
[16,] "V7<->V7"  "x7e"     NA
[17,] "V8<->V8"  "x8e"     NA
[18,] "V9<->V9"  "x9e"     NA

```

```

[19,] "X1<->X1" NA      "1"
attr(,"class")
[1] "mod"

> sem.one <- sem(mod.one,bifactor,100)
> summary(sem.one,digits=2)

Model Chisquare = 21.16848 Df = 27 Pr(>Chisq) = 0.778334
AIC = 57.16848
BIC = -103.1711

Normalized Residuals
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-0.3336510 -0.2923555 -0.1940195  0.0369476  0.0000019  1.8875159

R-square for Endogenous Variables
      V1      V2      V3      V4      V5      V6      V7      V8      V9
0.5636 0.4524 0.3377 0.3292 0.2522 0.1798 0.2568 0.1980 0.0932

Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a  0.7507129 0.09512871 7.891549 2.984584e-15 V1 <--- X1
b  0.6726412 0.09807150 6.858682 6.949882e-12 V2 <--- X1
c  0.5811209 0.10137932 5.732145 9.916850e-09 V3 <--- X1
d  0.5737425 0.10163173 5.645309 1.648847e-08 V4 <--- X1
e  0.5021915 0.10392785 4.832116 1.350893e-06 V5 <--- X1
f  0.4239908 0.10609489 3.996335 6.433059e-05 V6 <--- X1
g  0.5067957 0.10378883 4.882950 1.045102e-06 V7 <--- X1
h  0.4450171 0.10554867 4.216227 2.484236e-05 V8 <--- X1
i  0.3052415 0.10867168 2.808841 4.972015e-03 V9 <--- X1
x1e 0.4364302 0.08884720 4.912143 9.008624e-07 V1 <--> V1
x2e 0.5475539 0.09653118 5.672300 1.408927e-08 V2 <--> V2
x3e 0.6622982 0.10678972 6.201891 5.578883e-10 V3 <--> V3
x4e 0.6708197 0.10761127 6.233731 4.554546e-10 V4 <--> V4
x5e 0.7478036 0.11527456 6.487153 8.747374e-11 V5 <--> V5
x6e 0.8202314 0.12278021 6.680485 2.381530e-11 V6 <--> V6
x7e 0.7431581 0.11480162 6.473411 9.581477e-11 V7 <--> V7
x8e 0.8019593 0.12086557 6.635134 3.242077e-11 V8 <--> V8
x9e 0.9068284 0.13200380 6.869714 6.433079e-12 V9 <--> V9

Iterations = 11

> round(residuals(sem.one),2)

      V1      V2      V3      V4      V5      V6      V7      V8      V9
V1  0.00  0.06  0.04 -0.03 -0.03 -0.03 -0.03 -0.03 -0.03
V2  0.06  0.00  0.03 -0.03 -0.04 -0.03 -0.03 -0.03 -0.03
V3  0.04  0.03  0.00 -0.03 -0.03 -0.03 -0.03 -0.03 -0.03
V4 -0.03 -0.03 -0.03  0.00  0.13  0.11 -0.02 -0.02 -0.02
V5 -0.03 -0.04 -0.03  0.13  0.00  0.09 -0.02 -0.02 -0.02
V6 -0.03 -0.03 -0.03  0.11  0.09  0.00 -0.02 -0.02 -0.02
V7 -0.03 -0.03 -0.03 -0.02 -0.02 -0.02  0.00  0.19  0.13
V8 -0.03 -0.03 -0.03 -0.02 -0.02 -0.02  0.19  0.00  0.10
V9 -0.03 -0.03 -0.03 -0.02 -0.02 -0.02  0.13  0.10  0.00

```

## 5.4 Testing the dimensionality based upon an exploratory analysis

Alternatively, the output from an exploratory factor analysis can be used as input to the `structure.sem` function.

```
> f1 <- fa(bifact)
> mod.f1 <- structure.sem(f1)
> sem.f1 <- sem(mod.f1,bifact,100)
> sem.f1

Model Chisquare = 21.16848 Df = 27

      V1      V2      V3      V4      V5      V6      V7      V8      V9      x1e
0.7507129 0.6726412 0.5811209 0.5737425 0.5021915 0.4239908 0.5067957 0.4450171 0.3052415 0.4364302
      x2e      x3e      x4e      x5e      x6e      x7e      x8e      x9e
0.5475539 0.6622982 0.6708197 0.7478036 0.8202314 0.7431581 0.8019593 0.9068284

Iterations = 11
```

The answers are, of course, identical.

## 5.5 Specifying a three factor model

An alternative model is to extract three factors and try this solution. The `fa` factor analysis function (using the *minimum residual* algorithm) is used to detect the structure. Alternatively, the `factanal` could have been used. Rather than use the default rotation of `oblimin`, we force an orthogonal solution (even though we know it will be a poor solution).

```
> f3 <-fa(bifact,3,rotate="varimax")
> mod.f3 <- structure.sem(f3)
> sem.f3 <- sem(mod.f3,bifact,100)
> summary(sem.f3,digits=2)

Model Chisquare = 53.86635 Df = 27 Pr(>Chisq) = 0.001579738
AIC = 89.86635
BIC = -70.47325

Normalized Residuals
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.000003 0.000000  1.950175  1.642171  2.632737  4.011789

R-square for Endogenous Variables
      V1  V2  V3  V4  V5  V6  V7  V8  V9
0.64 0.49 0.36 0.49 0.36 0.25 0.49 0.36 0.16

Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
F1V1 0.8000000 0.1114517 7.177994 7.074151e-13 V1 <--- MR1
F1V2 0.7000001 0.1089845 6.422931 1.336754e-10 V2 <--- MR1
F1V3 0.6000000 0.1068002 5.617968 1.932167e-08 V3 <--- MR1
F2V4 0.6999999 0.1427544 4.903527 9.413091e-07 V4 <--- MR3
F2V5 0.6000001 0.1328610 4.515998 6.301927e-06 V5 <--- MR3
F2V6 0.4999995 0.1238740 4.036354 5.428827e-05 V6 <--- MR3
```

```

F3V7 0.7000001 0.1680059 4.166522 3.092827e-05 V7 <--- MR2
F3V8 0.6000000 0.1530271 3.920873 8.822871e-05 V8 <--- MR2
F3V9 0.4000005 0.1265677 3.160368 1.575701e-03 V9 <--- MR2
x1e 0.3600000 0.1297434 2.774707 5.525146e-03 V1 <--> V1
x2e 0.5099999 0.1165643 4.375268 1.212834e-05 V2 <--> V2
x3e 0.6399999 0.1130156 5.662936 1.488043e-08 V3 <--> V3
x4e 0.5100000 0.1739239 2.932316 3.364440e-03 V4 <--> V4
x5e 0.6399998 0.1475345 4.337967 1.438068e-05 V5 <--> V5
x6e 0.7500007 0.1336788 5.610466 2.017821e-08 V6 <--> V6
x7e 0.5100000 0.2136118 2.387509 1.696298e-02 V7 <--> V7
x8e 0.6400001 0.1734024 3.690837 2.235172e-04 V8 <--> V8
x9e 0.8400000 0.1362332 6.165898 7.008420e-10 V9 <--> V9

```

```
Iterations = 24
```

```
> round(residuals(sem.f3),2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	0.00	0.00	0.00	0.40	0.35	0.29	0.35	0.30	0.20
V2	0.00	0.00	0.00	0.35	0.30	0.25	0.31	0.26	0.18
V3	0.00	0.00	0.00	0.30	0.26	0.22	0.26	0.23	0.15
V4	0.40	0.35	0.30	0.00	0.00	0.00	0.27	0.24	0.16
V5	0.35	0.30	0.26	0.00	0.00	0.00	0.24	0.20	0.13
V6	0.29	0.25	0.22	0.00	0.00	0.00	0.20	0.17	0.11
V7	0.35	0.31	0.26	0.27	0.24	0.20	0.00	0.00	0.00
V8	0.30	0.26	0.23	0.24	0.20	0.17	0.00	0.00	0.00
V9	0.20	0.18	0.15	0.16	0.13	0.11	0.00	0.00	0.00

The residuals show serious problems with this model. Although the residuals within each of the three factors are zero, the residuals between groups are much too large.

## 5.6 Allowing for an oblique solution

The previous solution is clearly very bad. What would happen if the exploratory solution were allowed to have correlated (oblique) factors?

```

> f3 <-fa(bifactor,3)      #extract three factors and do an oblique rotation
> mod.f3 <- structure.sem(f3) #create the sem model
> mod.f3      #show it

```

	Path	Parameter	Value
[1,]	"MR1->V1"	"F1V1"	NA
[2,]	"MR1->V2"	"F1V2"	NA
[3,]	"MR1->V3"	"F1V3"	NA
[4,]	"MR3->V4"	"F2V4"	NA
[5,]	"MR3->V5"	"F2V5"	NA
[6,]	"MR3->V6"	"F2V6"	NA
[7,]	"MR2->V7"	"F3V7"	NA
[8,]	"MR2->V8"	"F3V8"	NA
[9,]	"MR2->V9"	"F3V9"	NA
[10,]	"V1<->V1"	"x1e"	NA
[11,]	"V2<->V2"	"x2e"	NA
[12,]	"V3<->V3"	"x3e"	NA
[13,]	"V4<->V4"	"x4e"	NA
[14,]	"V5<->V5"	"x5e"	NA

```

[15,] "V6<->V6" "x6e" NA
[16,] "V7<->V7" "x7e" NA
[17,] "V8<->V8" "x8e" NA
[18,] "V9<->V9" "x9e" NA
[19,] "MR3<->MR1" "xF2F1" NA
[20,] "MR2<->MR1" "xF3F1" NA
[21,] "MR2<->MR3" "xF3F2" NA
[22,] "MR1<->MR1" NA "1"
[23,] "MR3<->MR3" NA "1"
[24,] "MR2<->MR2" NA "1"
attr(,"class")
[1] "mod"

```

The structure being tested may be seen using `structure.graph`

### Structural model

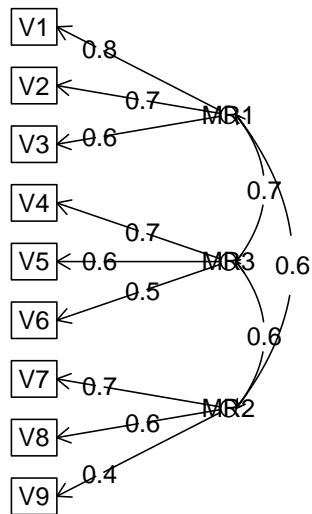


Figure 13: A three factor, oblique solution.

This makes much better sense, and in fact (as hoped) recovers the original structure.

## 5.7 Extract a bifactor solution using omega and then test that model using sem

A bifactor solution has previously been shown (Figure 9). The output from the `omega` function includes the sem commands for the analysis. As an example of doing this with real rather than simulated data, consider 9 variables from Thurstone. For completeness, the `stdCoef` from `sem` is used as well as the `summary` function.

### 5.7.1 sem of Thurstone 9 variable problem

The `sem` manual includes an example of a hierarchical solution to 9 mental abilities originally reported by Thurstone and used in the SAS manual for PROC CALIS and discussed in detail by McDonald (1999). The data matrix, as reported by Fox may be found in the Thurstone data set (which is “lazy loaded”). Using the commands just shown, it is possible to analyze this data set using a bifactor solution (Figure 14).

```
> sem.bi <- sem(om.th.bi$model$sem,Thurstone,213) #use the model created by omega
> summary(sem.bi,digits=2)
```

Model Chisquare = 24.2163 Df = 18 Pr(>Chisq) = 0.1480685  
AIC = 78.2163  
BIC = -72.28696

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.8211748	-0.3341044	-0.0000009	0.0281693	0.1561983	1.7968088

R-square for Endogenous Variables

Sentences	Vocabulary	Sent.Completion	First.Letters	Four.Letter.Words
0.8276	0.8302	0.7315	0.7472	0.6126
Suffixes	Letter.Series	Pedigrees	Letter.Group	
0.4824	0.8503	0.4996	0.4483	

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z )
Sentences	0.7678671	0.07059396	10.8772353	1.479833e-27
Vocabulary	0.7909248	0.06969232	11.3488087	7.518003e-30
Sent.Completion	0.7536211	0.07113218	10.5946585	3.154903e-26
First.Letters	0.6083814	0.07063841	8.6126138	7.141338e-18
Four.Letter.Words	0.5973349	0.07092937	8.4215455	3.715499e-17
Suffixes	0.5717903	0.07157752	7.9884057	1.366950e-15
Letter.Series	0.5668949	0.07249339	7.8199523	5.284337e-15
Pedigrees	0.6623314	0.07003035	9.4577757	3.145633e-21
Letter.Group	0.5299524	0.07332494	7.2274501	4.921470e-13
F1*Sentences	0.4878698	0.08141095	5.9926801	2.064107e-09
F1*Vocabulary	0.4523234	0.08353995	5.4144562	6.147524e-08
F1*Sent.Completion	0.4044507	0.08727334	4.6342988	3.581494e-06
F2*First.Letters	0.6140531	0.08471145	7.2487623	4.205973e-13
F2*Four.Letter.Words	0.5058063	0.08145488	6.2096500	5.310276e-10
F2*Suffixes	0.3943208	0.07805383	5.0519075	4.374195e-07
F3*Letter.Series	0.7272955	0.15844866	4.5901015	4.430304e-06
F3*Pedigrees	0.2468417	0.08677536	2.8446053	4.446649e-03

## Omega

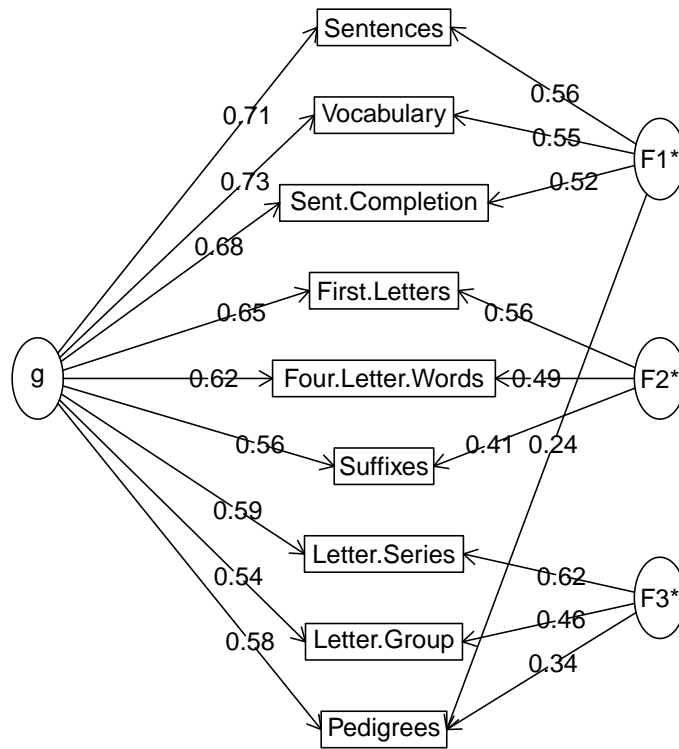


Figure 14: A bifactor solution to the Thurstone 9 variable problem. All items load on a general factor of ability, the residual factors account for the correlations between items within groups.

F3*Letter.Group	0.4091495	0.11352380	3.6040854	3.132541e-04
e1	0.1723633	0.03405646	5.0611045	4.168346e-07
e2	0.1698419	0.03001233	5.6590697	1.521958e-08
e3	0.2684749	0.03316228	8.0957909	5.689350e-16
e4	0.2528108	0.07942835	3.1828791	1.458185e-03
e5	0.3873510	0.06317399	6.1314949	8.705712e-10
e6	0.5175679	0.05955079	8.6912013	3.586269e-18
e7	0.1496709	0.21861502	0.6846325	4.935759e-01
e8	0.5003855	0.05956551	8.4005902	4.442400e-17
e9	0.5517474	0.08455914	6.5249884	6.800680e-11

Sentences	Sentences <--- g
Vocabulary	Vocabulary <--- g
Sent.Completion	Sent.Completion <--- g
First.Letters	First.Letters <--- g
Four.Letter.Words	Four.Letter.Words <--- g
Suffixes	Suffixes <--- g
Letter.Series	Letter.Series <--- g
Pedigrees	Pedigrees <--- g
Letter.Group	Letter.Group <--- g
F1*Sentences	Sentences <--- F1*
F1*Vocabulary	Vocabulary <--- F1*
F1*Sent.Completion	Sent.Completion <--- F1*
F2*First.Letters	First.Letters <--- F2*
F2*Four.Letter.Words	Four.Letter.Words <--- F2*
F2*Suffixes	Suffixes <--- F2*
F3*Letter.Series	Letter.Series <--- F3*
F3*Pedigrees	Pedigrees <--- F3*
F3*Letter.Group	Letter.Group <--- F3*
e1	Sentences <--> Sentences
e2	Vocabulary <--> Vocabulary
e3	Sent.Completion <--> Sent.Completion
e4	First.Letters <--> First.Letters
e5	Four.Letter.Words <--> Four.Letter.Words
e6	Suffixes <--> Suffixes
e7	Letter.Series <--> Letter.Series
e8	Pedigrees <--> Pedigrees
e9	Letter.Group <--> Letter.Group

Iterations = 72

> stdCoef(sem.bi,digits=2)

	Std. Estimate	
1 Sentences	0.7678671	Sentences <--- g
2 Vocabulary	0.7909246	Vocabulary <--- g
3 Sent.Completion	0.7536211	Sent.Completion <--- g
4 First.Letters	0.6083814	First.Letters <--- g
5 Four.Letter.Words	0.5973349	Four.Letter.Words <--- g
6 Suffixes	0.5717900	Suffixes <--- g
7 Letter.Series	0.5668950	Letter.Series <--- g
8 Pedigrees	0.6623317	Pedigrees <--- g
9 Letter.Group	0.5299523	Letter.Group <--- g
10 F1*Sentences	0.4878697	Sentences <--- F1*
11 F1*Vocabulary	0.4523233	Vocabulary <--- F1*
12 F1*Sent.Completion	0.4044507	Sent.Completion <--- F1*
13 F2*First.Letters	0.6140531	First.Letters <--- F2*
14 F2*Four.Letter.Words	0.5058063	Four.Letter.Words <--- F2*



```

15      F2*Suffixes      0.3943206      Suffixes <--- F2*
16      F3*Letter.Series 0.7272957      Letter.Series <--- F3*
17      F3*Pedigrees     0.2468418      Pedigrees <--- F3*
18      F3*Letter.Group  0.4091494      Letter.Group <--- F3*
19      e1               0.1723633      Sentences <--> Sentences
20      e2               0.1698418      Vocabulary <--> Vocabulary
21      e3               0.2684748      Sent.Completion <--> Sent.Completion
22      e4               0.2528108      First.Letters <--> First.Letters
23      e5               0.3873510      Four.Letter.Words <--> Four.Letter.Words
24      e6               0.5175675      Suffixes <--> Suffixes
25      e7               0.1496710      Letter.Series <--> Letter.Series
26      e8               0.5003859      Pedigrees <--> Pedigrees
27      e9               0.5517473      Letter.Group <--> Letter.Group
28      1.0000000        F1* <--> F1*
29      1.0000000        F2* <--> F2*
30      1.0000000        F3* <--> F3*
31      1.0000000        g <--> g

```

Compare this solution to the one reported below, and to the *sem* manual.

## 5.8 Examining a hierarchical solution

A hierarchical solution to this data set was previously found by the `omega` function (Figure 10). The output of that analysis can be used as a model for a `sem` analysis. Once again, the `stdCoef` function helps see the structure. Alternatively, using the `omega` function on the `Thurstone` data will create the model for this particular data set.

```

> sem.hi <- sem(om.hi$model$sem,Thurstone,213)
> summary(sem.hi,digits=2)

Model Chisquare = 38.1963   Df = 24 Pr(>Chisq) = 0.03310059
AIC = 80.1963
BIC = -90.47471

Normalized Residuals
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-0.9724643 -0.4164673 -0.0000001  0.0401007  0.0938588  1.6274666

R-square for Endogenous Variables
      F1      F2      F3      Sentences      Vocabulary
      0.6758      0.6112      0.6642      0.8185      0.8351
Sent.Completion First.Letters Four.Letter.Words Suffixes Letter.Series
      0.7329      0.6985      0.6355      0.4936      0.6097
Pedigrees Letter.Group
      0.5186      0.4949

Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
gF1      1.4438115 0.25653564 5.628113 1.821922e-08
gF2      1.2538296 0.21136562 5.932041 2.991910e-09
gF3      1.4065517 0.26890804 5.230605 1.689563e-07
F1Sentences 0.5151232 0.06292248 8.186632 2.686376e-16
F1Vocabulary 0.5203104 0.06338431 8.208820 2.233734e-16
F1Sent.Completion 0.4874316 0.06081528 8.014954 1.101786e-15

```

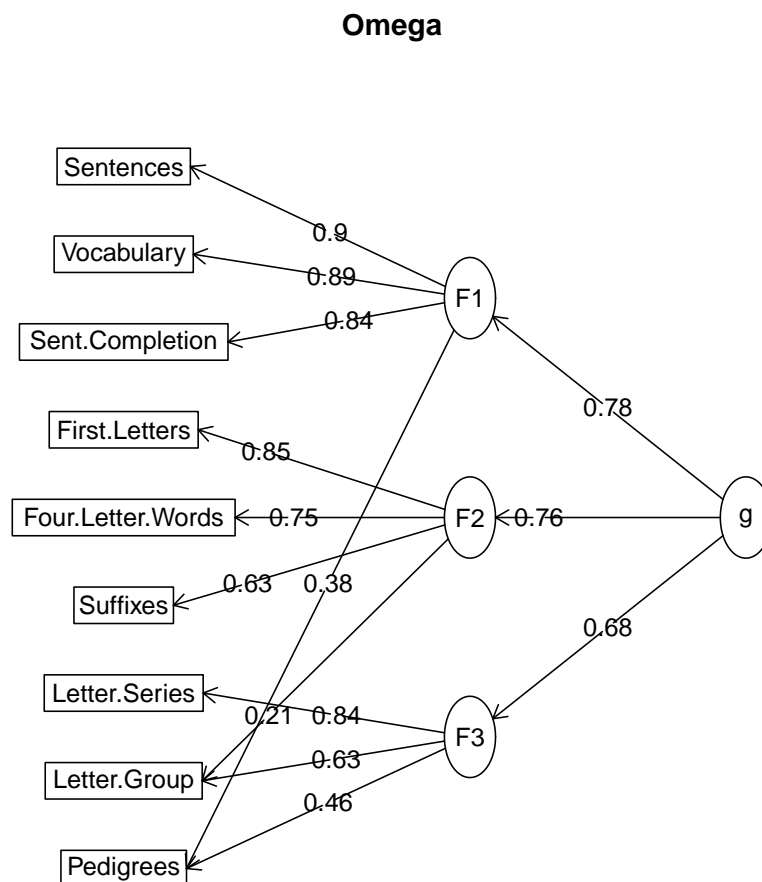


Figure 15: Hierarchical analysis of the Thurstone 9 variable problem using an exploratory algorithm can provide the appropriate sem code for analysis using the sem package.

```

F2First.Letters      0.5211221 0.06106205 8.534304 1.410015e-17
F2Four.Letter.Words 0.4970664 0.05902388 8.421446 3.718664e-17
F2Suffixes           0.4380644 0.05595794 7.828458 4.938915e-15
F3Letter.Series      0.4524352 0.06596903 6.858297 6.968649e-12
F3Pedigrees          0.4172903 0.06215816 6.713363 1.901887e-11
F3Letter.Group       0.4076312 0.06131399 6.648258 2.965820e-11
e1                   0.1814979 0.02847741 6.373397 1.848862e-10
e2                   0.1649304 0.02776938 5.939292 2.862558e-09
e3                   0.2671331 0.03336340 8.006771 1.177597e-15
e4                   0.3015024 0.05102191 5.909274 3.436179e-09
e5                   0.3645010 0.05263547 6.925008 4.359513e-12
e6                   0.5064150 0.05962608 8.493180 2.010593e-17
e7                   0.3903313 0.05933649 6.578268 4.759607e-11
e8                   0.4813697 0.06224844 7.733041 1.050075e-14
e9                   0.5051017 0.06332869 7.975875 1.513055e-15

```

```

gF1      F1 <--- g
gF2      F2 <--- g
gF3      F3 <--- g
F1Sentences      Sentences <--- F1
F1Vocabulary      Vocabulary <--- F1
F1Sent.Completion      Sent.Completion <--- F1
F2First.Letters      First.Letters <--- F2
F2Four.Letter.Words      Four.Letter.Words <--- F2
F2Suffixes      Suffixes <--- F2
F3Letter.Series      Letter.Series <--- F3
F3Pedigrees      Pedigrees <--- F3
F3Letter.Group      Letter.Group <--- F3
e1      Sentences <--> Sentences
e2      Vocabulary <--> Vocabulary
e3      Sent.Completion <--> Sent.Completion
e4      First.Letters <--> First.Letters
e5      Four.Letter.Words <--> Four.Letter.Words
e6      Suffixes <--> Suffixes
e7      Letter.Series <--> Letter.Series
e8      Pedigrees <--> Pedigrees
e9      Letter.Group <--> Letter.Group

```

Iterations = 54

```
> stdCoef(sem.hi,digits=2)
```

		Std. Estimate	
1	gF1	0.8220754	F1 <--- g
2	gF2	0.7817998	F2 <--- g
3	gF3	0.8150140	F3 <--- g
4	F1Sentences	0.9047111	Sentences <--- F1
5	F1Vocabulary	0.9138214	Vocabulary <--- F1
6	F1Sent.Completion	0.8560764	Sent.Completion <--- F1
7	F2First.Letters	0.8357617	First.Letters <--- F2
8	F2Four.Letter.Words	0.7971819	Four.Letter.Words <--- F2
9	F2Suffixes	0.7025560	Suffixes <--- F2
10	F3Letter.Series	0.7808129	Letter.Series <--- F3
11	F3Pedigrees	0.7201599	Pedigrees <--- F3
12	F3Letter.Group	0.7034902	Letter.Group <--- F3
13	e1	0.1814979	Sentences <--> Sentences
14	e2	0.1649304	Vocabulary <--> Vocabulary
15	e3	0.2671331	Sent.Completion <--> Sent.Completion

```

16          e4      0.3015024      First.Letters <--> First.Letters
17          e5      0.3645010 Four.Letter.Words <--> Four.Letter.Words
18          e6      0.5064151      Suffixes <--> Suffixes
19          e7      0.3903313      Letter.Series <--> Letter.Series
20          e8      0.4813697      Pedigrees <--> Pedigrees
21          e9      0.5051016      Letter.Group <--> Letter.Group
22          0.3241920      F1 <--> F1
23          0.3887891      F2 <--> F2
24          0.3357521      F3 <--> F3
25          1.0000000      g <--> g

```

```
> anova(sem.hi,sem.bi)
```

LR Test for Difference Between Models

	Model	Df	Model	Chisq	Df	LR	Chisq	Pr(>Chisq)
sem.hi		24		38.196				
sem.bi		18	24.216	6	13.98	0.02986	*	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Using the Thurstone data set, we see what happens when a hierarchical model is applied to real data. The exploratory structure derived from the `omega` function (Figure 15) provides estimates in close approximation to those found using `sem`. The model definition created by using `omega` is the same hierarchical model discussed in the `sem` help page. The *bifactor* model, with 6 more parameters does provide a better fit to the data than the hierarchical model.

Similar analyses can be done with other data that are organized hierarchically. Examples of these analyses are analyzing the 14 variables of `holzinger` and the 16 variables of `reise`. The output from the following analyses has been limited to just the comparison between the bifactor and hierarchical solutions.

```

> om.holz.bi <- omega(Holzinger,4)
> sem.holz.bi <- sem(om.holz.bi$model$sem,Holzinger,355)
> om.holz.hi <- omega(Holzinger,4,sl=FALSE)
> sem.holz.hi <- sem(om.holz.hi$model$sem,Holzinger,355)
> anova(sem.holz.bi,sem.holz.hi)

```

LR Test for Difference Between Models

	Model	Df	Model	Chisq	Df	LR	Chisq	Pr(>Chisq)
sem.holz.bi		63		147.66				
sem.holz.hi		73	178.79	10	31.129	0.0005587	***	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## 5.9 Estimating Omega using EFA followed by CFA

The function `omegaSem` combines both an exploratory factor analysis using `omega`, then calls the appropriate `sem` functions and organizes the results as in a standard `omega` analysis.

An example is found from the Thurstone data set of 9 cognitive variables:

```
> om.sem <- omegaSem(Thurstone,n.obs=213)

Call: omegaSem(m = Thurstone, n.obs = 213)
Omega
Call: omega(m = m, nfactors = nfactors, fm = fm, key = key, flip = flip,
  digits = digits, title = title, sl = sl, labels = labels,
  plot = plot, n.obs = n.obs, rotate = rotate, Phi = Phi, option = option)
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

general/max 3.73  max/min = 1.34
mean percent general = 0.6  with sd = 0.05 and cv of 0.09
Explained Common Variance of the general factor = 0.6

The degrees of freedom are 12 and the fit is 0.01
The number of observations was 213 with Chi Square = 2.98 with prob < 1
The root mean square of the residuals is 0.01
The df corrected root mean square of the residuals is 0.01
RMSEA index = 0 and the 10 % confidence intervals are 0 0
BIC = -61.36

Compare this with the adequacy of just a general factor and no group factors
The degrees of freedom for just the general factor are 27 and the fit is 1.48
The number of observations was 213 with Chi Square = 306.71 with prob < 3.3e-49
The root mean square of the residuals is 0.14
The df corrected root mean square of the residuals is 0.16

RMSEA index = 0.224 and the 10 % confidence intervals are 0.199 0.244
BIC = 161.96

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

The following analyses were done using the lavaan package

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 24.33052 with p = 0.1444947  
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

To get the standard sem fit statistics, ask for summary on the fitted object

Comparing the two models graphically (Figure 16 with Figure 14 shows that while not identical, they are very similar. The sem version is basically a forced simple structure. Notice that the values of  $\omega_h$  are not identical from the EFA and CFA models. The CFA solution yields higher values of  $\omega_h$  because, by forcing a pure cluster solution (no cross loadings), the correlations between the factors is forced to be through the  $g$  factor.

### Omega from SEM

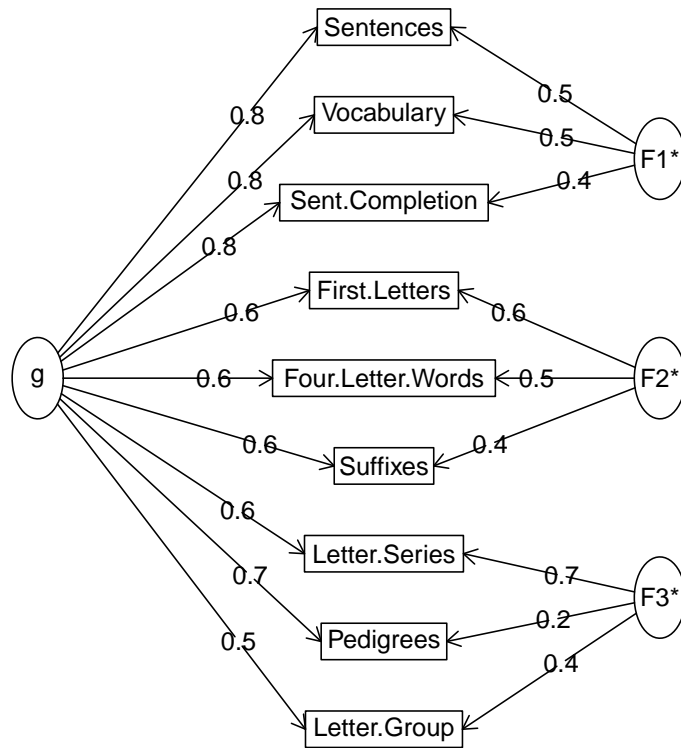


Figure 16: Confirmatory Omega structure using `omegaSem`

## 6 Summary and conclusion

The use of exploratory and confirmatory models for understanding real data structures is an important advance in psychological research. To understand these approaches it is helpful to try them first on “baby” data sets. To the extent that the models we use can be tested on simple, artificial examples, it is perhaps easier to practice their application. The *psych* tools for simulating structural models and for specifying models are a useful supplement to the power of packages such as *sem*. The techniques that can be used on simulated data set can also be applied to real data sets.



```

> sessionInfo()

R Under development (unstable) (2018-04-05 r74542)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS 10.14.2

Matrix products: default
BLAS: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib

locale:
[1] C

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods   base

other attached packages:
[1] sem_3.1-9          GPArotation_2014.11-1 psych_1.8.12

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.16      lattice_0.20-35   matrixcalc_1.0-3  MASS_7.3-49      grid_3.6.0
 [6] arm_1.9-3         nlme_3.1-136     stats4_3.6.0     coda_0.19-1     minqa_1.2.4
[11] mi_1.0            nloptr_1.0.4     Matrix_1.2-13    pbivnorm_0.6.0   boot_1.3-20
[16] splines_3.6.0     lme4_1.1-16      tools_3.6.0      foreign_0.8-70   abind_1.4-5
[21] parallel_3.6.0    compiler_3.6.0   mnormt_1.5-5     lavaan_0.6-3

```

## References

- Adams, D. (1980). *The hitchhiker's guide to the galaxy*. Harmony Books, New York, 1st American edition.
- Bechtoldt, H. (1961). An empirical study of the factor analysis stability hypothesis. *Psychometrika*, 26(4):405–432.
- Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modeling*, 13:465–486.
- Fox, J. (2009). *sem: Structural Equation Models*. R package version 0.9-15.
- . Fox, J. Nie, Zhenghua and Byrnes, J. (2013) *sem: Structural Equation Models*. R package version 3.1-3.
- Holzinger, K. and Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1):41–54.
- Jensen, A. R. and Weng, L.-J. (1994). What is a good g? *Intelligence*, 18(3):231–258.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. L. Erlbaum Associates, Mahwah, N.J.
- Rafaeli, E. and Revelle, W. (2006). A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*, 30(1):1–12.

- Reise, S., Morizot, J., and Hays, R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(0):19–31.
- Revelle, W. (2014). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston. R package version 1.4.1
- Rosseel, Y. (2012). *lavaan: An R Package for Structural Equation Modeling*. *Journal of Statistical Software*, 48(2):1-36. R package version 0.5-14.
- Schmid, J. J. and Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1):83–90.
- Thurstone, L. L. and Thurstone, T. G. (1941). *Factorial studies of intelligence*. The University of Chicago press, Chicago, Ill.

## Index

anova, 32, 33

bifactor, 7, 8, 23, 44

circumplex structure, 8

cluster.cor, 3

congeneric, 5

describe, 3

esem, 28

esem.diagram, 28

fa, 14, 35

factanal, 14, 35

factor.congruence, 20

factor.pa, 3, 20

guttman, 3

hierarchical, 8

holzinger, 23, 44

ICLUST, 3, 27

lavaan, 3

minimum residual, 35

oblimin, 22, 35

omega, 3, 8, 23, 38, 41, 44

omega.graph, 31

omegaSem, 44, 47

pairs.panels, 3

phi.list, 13, 14, 31

principal, 3, 19, 20

principal components, 19

Promax, 14, 19, 22

promax, 22

psych, 3, 22, 31, 48

quartimin, 19

R function

- anova, 32, 33
- bifactor, 7, 23
- cluster.cor, 3
- describe, 3
- esem, 28
- esem.diagram, 28
- fa, 14, 35
- factanal, 14, 35
- factor.congruence, 20
- factor.pa, 3, 20
- guttman, 3
- holzinger, 23, 44
- ICLUST, 3, 27
- oblimin, 22, 35
- omega, 3, 8, 23, 38, 41, 44
- omega.graph, 31
- omegaSem, 44, 47
- pairs.panels, 3
- phi.list, 13, 14, 31
- principal, 3, 19, 20
- Promax, 14, 19, 22
- promax, 22
- psych, 3

psych package

- bifactor, 7, 23
- cluster.cor, 3
- describe, 3
- esem, 28
- esem.diagram, 28
- fa, 14, 35
- factor.congruence, 20
- factor.pa, 3, 20
- guttman, 3
- holzinger, 23, 44
- ICLUST, 3, 27

- oblimin, 35
- omega, 3, 8, 23, 38, 41, 44
- omega.graph, 31
- omegaSem, 44, 47
- pairs.panels, 3
- phi.list, 13, 14, 31
- principal, 3, 19, 20
- Promax, 14, 19, 22
- psych, 3
- reise, 23, 44
- score.items, 3
- sim, 4
- sim.circ, 8
- sim.congeneric, 5, 7
- sim.hierarchical, 7
- sim.hierarcical, 4
- sim.item, 8
- sim.structure, 8, 16
- structure.diagram, 16, 31
- structure.graph, 31, 37
- structure.list, 13, 14, 31
- structure.sem, 31
- Thurstone, 38, 41
- VSS, 3
- quartimin, 19
- reise, 23, 44
- Rgraphviz, 31
- score.items, 3
- sem, 23, 31, 38, 41, 44
- sim, 4
- sim.circ, 8
- sim.congeneric, 5, 7
- sim.hierarchical, 7
- sim.hierarcical, 4
- sim.item, 8
- sim.structure, 8, 16
- stdCoef, 38, 41
- structure.diagram, 16, 31
- structure.graph, 31, 37
- structure.list, 13, 14, 31
- structure.sem, 31
- summary, 38
- tau, 5
- Thurstone, 38, 41
- varimax, 19
- VSS, 3
- R package
  - lavaan, 3
  - psych, 3, 22, 31, 48
  - Rgraphviz, 3
  - sem, 3, 31, 38, 41, 48
- reise, 23, 44
- Rgraphviz, 3, 31
- rotated, 19
- score.items, 3
- sem, 3, 23, 31, 38, 41, 44, 48
- sim, 4
- sim.circ, 8
- sim.congeneric, 5, 7
- sim.hierarchical, 7
- sim.hierarcical, 4
- sim.item, 8
- sim.structure, 8, 16
- simple structure, 8, 19
- stdCoef, 38, 41
- structure.diagram, 16, 31
- structure.graph, 31, 37
- structure.list, 13, 14, 31
- structure.sem, 31
- summary, 38