Psychology 454: Latent Variable Modeling

Week 2

Department of Psychology
Northwestern University
Evanston, Illinois USA

September, 2016

Outline

Required functions

- R
- library(psych)
- library(sem)
- library(lavaan)
    - sem is version 3.1-7, lavaan is 0.5-20, psych is 1.6.9
- A bit of confidence

See http://personality-project.org/revelle/syllabi/405/405-efa.pdf for a more detailed discussion of EFA/CFA/

Create an artificial data set

Create a data set of raw scores with the `sim.congeneric` function and a factor structure defined by loadings of .6, .5, .4, and .3 for 200 subjects.
Show the scatter plot matrix of correlations for this data set using the `pairs.panels` function.

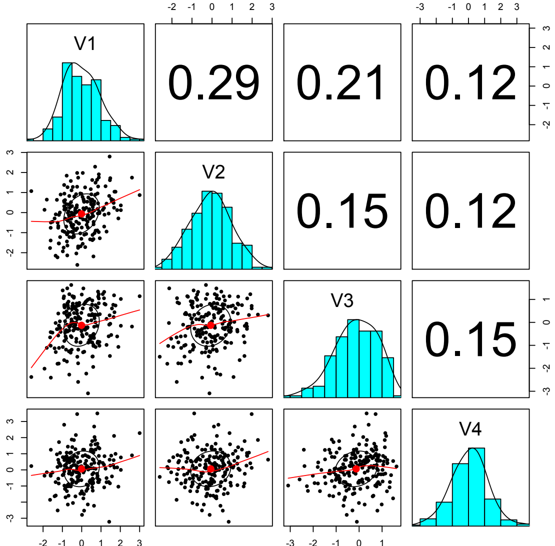
R code

```
set.seed(42)
 X <- sim.congeneric(N = 200, loads = c(0.6, 0.5, 0.4, 0.3),
           short = FALSE)
dim(X$observed)
describe(X$observed)
pairs.panels(X$observed)
```

```
[1] 200   4

   var   n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
V1   1 200 -0.01 0.92  -0.10   -0.04 0.96 -2.61 3.00  5.61  0.34     0.14 0.06
V2   2 200 -0.06 1.00  -0.03   -0.07 0.99 -2.61 2.80  5.40  0.03    -0.21 0.07
V3   3 200 -0.13 0.91  -0.11   -0.08 1.04 -3.09 1.63  4.72 -0.48    -0.03 0.06
V4   4 200  0.06 1.11   0.13    0.06 1.07 -3.22 3.50  6.72  0.03     0.60 0.08
```

A SPLOM: scatter plot matrix using pairs.panels

pairs.panels(X$observed)



Another congeneric correlation matrix

```
set.seed(42)
Cong <- sim.congeneric()  #use the default values
f1 <- fa(Cong)
```

```
Factor Analysis using method =  minres
Call: fa(r = Cong)
Standardized loadings based upon correlation matrix
   MR1   h2   u2
V1 0.8 0.64 0.36
V2 0.7 0.49 0.51
V3 0.6 0.36 0.64
V4 0.5 0.25 0.75
                MR1
SS loadings    1.74
Proportion Var 0.44
Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the null model are  6  and the objective fu
The degrees of freedom for the model are 2  and the objective function

The root mean square of the residuals 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 scores with factors              0.89
```

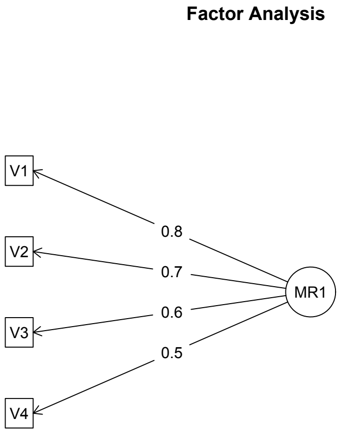
Do it with the factanal function

```
f <- factanal(covmat=Cong,n.obs=1000,factors=1)
```

```
> f <- factanal(covmat=Cong,n.obs=1000,factors=1)
> f
Call:
factanal(factors = 1, covmat = Cong, n.obs = 1000)
Uniquenesses:
  V1   V2   V3   V4
0.36 0.51 0.64 0.75
Loadings:
   Factor1
V1 0.8
V2 0.7
V3 0.6
V4 0.5
               Factor1
SS loadings      1.740
Proportion Var   0.435
Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 0 on 2 degrees of freedom.
The p-value is 1
```

diagram the result

```
diagram(f1) #or fa.diagram(f1)
```



Use the sem function from sem

```
model = matrix(c("Latent ->  V1", "a", NA, "Latent -> V2", "b", NA
 "Latent -> V3", "c",  NA, "Latent -> V4", "d", NA,
 "V1 <-> V1", "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3",
    "w", NA, "V4 <-> V4", "x", NA,
 "Latent <-> Latent", NA, 1), ncol = 3, byrow = TRUE)
model
sem4 = sem(model, Cong, 1000)
summary(sem4)
```

```
>  model
      [,1]                [,2] [,3]
 [1,] "Latent ->  V1"     "a"  NA
 [2,] "Latent -> V2"      "b"  NA
 [3,] "Latent -> V3"      "c"  NA
 [4,] "Latent -> V4"      "d"  NA
 [5,] "V1 <-> V1"         "u"  NA
 [6,] "V2 <-> V2"         "v"  NA
 [7,] "V3 <-> V3"         "w"  NA
 [8,] "V4 <-> V4"         "x"  NA
 [9,] "Latent <-> Latent" NA   "1"
```

sem continued

```
>  sem4 = sem(model, Cong, 1000)
> summary(sem4)
```

```
Model Chisquare =  1.8885e-09   Df =  2 Pr(>Chisq) = 1
 Chisquare (null model) =  894.56   Df =  6
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.0068
 Bentler CFI =  1
 SRMR =  6.4327e-07
 BIC =  -13.816

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
3.00e-06 5.52e-06 9.38e-06 1.10e-05 1.48e-05 3.77e-05

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a 0.80     0.032181  24.860  0        V1 <--- Latent
b 0.70     0.032359  21.632  0        V2 <--- Latent
c 0.60     0.032771  18.309  0        V3 <--- Latent
d 0.50     0.033583  14.889  0        V4 <--- Latent
u 0.36     0.034180  10.532  0        V1 <--> V1
```

Do it with 400 cases and “real” data

```
set.seed(17)  #only for teaching purposes
Cong <- sim.congeneric(loads=c(.8,.7,.6,.5),N=400,short=FALSE)
Cong$r
```

```
          V1        V2        V3        V4
V1 1.0000000 0.6024138 0.4340634 0.4081029
V2 0.6024138 1.0000000 0.4065795 0.4048413
V3 0.4340634 0.4065795 1.0000000 0.2749936
V4 0.4081029 0.4048413 0.2749936 1.0000000
```

Exploratory factor analysis

```
f1 <- fa(Cong$observed)
```

```
> f1
Factor Analysis using method =  minres
Call: fa(r = Cong$observed)
Standardized loadings based upon correlation matrix
    MR1   h2   u2
V1 0.79 0.63 0.37
V2 0.76 0.58 0.42
V3 0.54 0.29 0.71
V4 0.52 0.27 0.73
                MR1
SS loadings    1.77
Proportion Var 0.44
Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the null model are  6  and the objective function was  0.94
             with Chi Square of  372.34
The degrees of freedom for the model are 2  and the objective function was  0
The root mean square of the residuals is  0
The df corrected root mean square of the residuals is  0.01
The number of observations was  400  with Chi Square =  0.23  with prob <  0.89
Tucker Lewis Index of factoring reliability =  1.014
RMSEA index =  0  and the 90 % confidence intervals are  0 0.038
BIC =  -11.75
Fit based upon off diagonal values = 1
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  0.59
```

sem needs a covariance or correlation matrix

```
sem4 = sem(model, Cong$r,N=400)
> summary(sem4)
```

```
 Model Chisquare =  0.23618   Df =  2 Pr(>Chisq) = 0.88861
 Chisquare (null model) =  374.38   Df =  6
 Goodness-of-fit index =  0.9997
 Adjusted goodness-of-fit index =  0.9985
 RMSEA index =  0   90% CI: (NA, 0.046149)
 Bentler-Bonnett NFI =  0.99937
 Tucker-Lewis NNFI =  1.0144
 Bentler CFI =  1
 SRMR =  0.0043459
 BIC =  -11.747

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.12300 -0.08460 -0.00774 -0.00601  0.02920  0.14700

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a 0.79190  0.050413  15.7083 0.0000e+00 V1 <--- Latent
b 0.76186  0.050522  15.0798 0.0000e+00 V2 <--- Latent
c 0.54011  0.052069  10.3729 0.0000e+00 V3 <--- Latent
d 0.52099  0.052401   9.9425 0.0000e+00 V4 <--- Latent
u 0.37289  0.052508   7.1015 1.2341e-12 V1 <--> V1
v 0.41957  0.051757   8.1065 4.4409e-16 V2 <--> V2
w 0.70828  0.056405  12.5571 0.0000e+00 V3 <--> V3
x 0.72857  0.057370  12.6993 0.0000e+00 V4 <--> V4
```

But lavaan prefers a raw data matrix

```
model <- 'latent =~ V1 + V2 + V3 + V4'
 fit <- cfa(model,data=Cong$observed,std.lv=TRUE,std.ov=TRUE)
```

```
Lavaan (0.4-3) converged normally after 14 iterations
   Number of observations                            400
  Estimator                                           ML
  Minimum Function Chi-square                      0.237
  Degrees of freedom                                   2
  P-value                                          0.888
Model fitting results:
  Information                                   Expected
  Standard Errors                               Standard
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  latent =~
    V1                0.791    0.050   15.754    0.000
    V2                0.761    0.050   15.127    0.000
    V3                0.539    0.052   10.380    0.000
    V4                0.520    0.052    9.961    0.000
Latent variances:
    latent            1.000
Residual variances:
    V1                0.372    0.052    7.138    0.000
    V2                0.419    0.051    8.151    0.000
    V3                0.707    0.056   12.569    0.000
    V4                0.727    0.057   12.725    0.000
```

Mood data

```
data(msq)  #get all the data
 msq9 <- msq[,c("alert","wide.awake","energetic","sleepy","tired",
 "drowsy","tense","jittery","nervous","relaxed","calm","at.ease",
 "pleased","satisfied","content","distressed","grouchy","sad")]
```

```
describe(msq9)
           var    n mean   sd median trimmed  mad min max range  skew kurtosis   se
alert        1 3885 1.15 0.91      1    1.09 1.48   0   3     3  0.33    -0.76 0.01
wide-awake   2 3884 0.94 0.95      1    0.83 1.48   0   3     3  0.65    -0.63 0.02
energetic    3 3890 0.84 0.89      1    0.74 1.48   0   3     3  0.74    -0.42 0.01
sleepy       4 3880 1.25 1.05      1    1.18 1.48   0   3     3  0.40    -1.04 0.02
tired        5 3886 1.39 1.04      1    1.36 1.48   0   3     3  0.22    -1.10 0.02
drowsy       6 3884 1.16 1.03      1    1.08 1.48   0   3     3  0.46    -0.93 0.02
tense        7 3886 0.52 0.77      0    0.37 0.00   0   3     3  1.41     1.35 0.01
jittery      8 3890 0.59 0.80      0    0.45 0.00   0   3     3  1.24     0.82 0.01
nervous      9 3879 0.35 0.65      0    0.22 0.00   0   3     3  1.93     3.48 0.01
relaxed     10 3889 1.68 0.88      2    1.72 1.48   0   3     3 -0.17    -0.68 0.01
calm        11 3814 1.55 0.92      2    1.56 1.48   0   3     3 -0.01    -0.83 0.01
at-ease     12 3879 1.59 0.92      2    1.61 1.48   0   3     3 -0.09    -0.83 0.01
pleased     13 3883 0.90 0.87      1    0.82 1.48   0   3     3  0.57    -0.59 0.01
satisfied   14 3889 1.34 0.89      1    1.32 1.48   0   3     3  0.00    -0.83 0.01
content     15 3874 1.45 0.93      1    1.43 1.48   0   3     3  0.02    -0.85 0.01
distressed  16 3888 0.40 0.71      0    0.23 0.00   0   3     3  1.87     2.99 0.01
grouchy     17 3891 0.47 0.73      0    0.31 0.00   0   3     3  1.52     1.72 0.01
sad         18 3886 0.34 0.63      0    0.22 0.00   0   3     3  1.96     3.80 0.01
```

How many factors – no right answer, one wrong answer

Henry Kaiser once said (personal communication) that a solution to the number-of-factors problem in factor analysis is easy, that he used to make one up every morning before breakfast. (p 283)

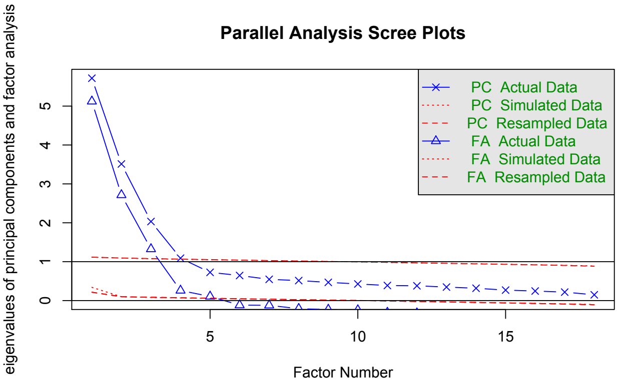
1. Statistical
   - Extracting factors until the χ^2 of the residual matrix is not significant.
   - Extracting factors until the change in χ^2 from factor n to factor n+1 is not significant.
2. Rules of Thumb
   - Parallel Extracting factors until the eigenvalues of the real data are less than the corresponding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

Exploratory factor analysis- how many factors

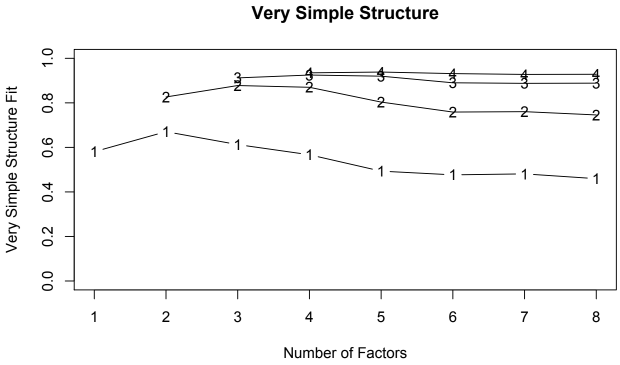
```
fa.parallel(msq9)
 VSS(msq9)
 nfactors(msq9)
```

```
Parallel analysis suggests that the number of factors =  5
                           and the number of components =  4

Very Simple Structure
Call: VSS(x = msq9)
VSS complexity 1 achieves a maximimum of 0.67  with  2  factors
VSS complexity 2 achieves a maximimum of 0.88  with  3  factors

The Velicer MAP criterion achieves a minimum of 0.02  with  4  factors

Velicer MAP
[1] 0.08 0.05 0.03 0.02 0.03 0.04 0.05 0.06

Very Simple Structure Complexity 1
[1] 0.58 0.67 0.61 0.57 0.49 0.48 0.48 0.46

Very Simple Structure Complexity 2
[1] 0.00 0.83 0.88 0.87 0.80 0.76 0.76 0.75
```

Parallel analysis – comparison with random data



Very Simple Structure – comparison with alternative solutions

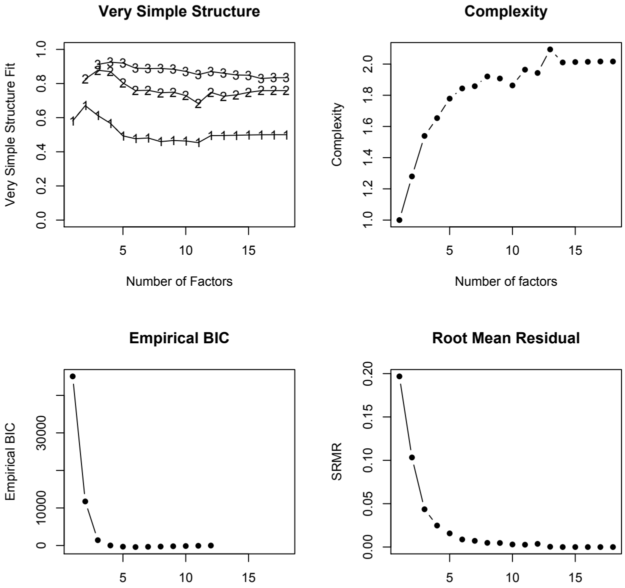


The number of factors problem – "it will break your heart"

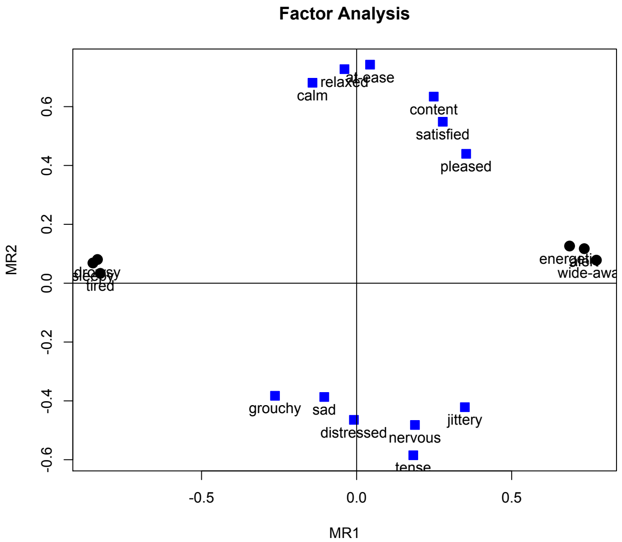
```
nfactors(msq9)
```

```
Number of factors
Call: vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,
    n.obs = n.obs, plot = FALSE, title = title, use = use, cor = cor)
VSS complexity 1 achieves a maximimum of 0.67  with  2  factors
VSS complexity 2 achieves a maximimum of 0.88  with  3  factors
The Velicer MAP achieves a minimum of 0.02  with  4  factors
Empirical BIC achieves a minimum of  -405.06  with  6  factors
Sample Size adjusted BIC achieves a minimum of  -93.48  with  8  factors

Statistics by number of factors
   vss1 vss2   map dof   chisq     prob sqresid  fit RMSEA   BIC SABIC complex  eChisq     SRMR
1  0.58 0.00 0.081 135 2.2e+04  0.0e+00    22.2 0.58 0.203 20601 21030     1.0 4.6e+04 2.0e-0
2  0.67 0.83 0.049 118 1.1e+04  0.0e+00     9.2 0.83 0.157 10382 10757     1.3 1.3e+04 1.0e-0
3  0.61 0.88 0.026 102 3.4e+03  0.0e+00     4.7 0.91 0.092  2588  2912     1.5 2.3e+03 4.4e-0
4  0.57 0.87 0.024  87 1.7e+03 2.5e-290     3.5 0.93 0.068   947  1223     1.7 7.3e+02 2.5e-0
5  0.49 0.80 0.029  73 7.0e+02 2.2e-102     3.0 0.94 0.047    92   324     1.8 3.0e+02 1.6e-0
6  0.48 0.76 0.036  60 2.8e+02  1.1e-30     2.6 0.95 0.031  -213   -22     1.8 9.1e+01 8.7e-0
7  0.48 0.76 0.046  48 1.9e+02  1.1e-18     2.5 0.95 0.028  -208   -55     1.9 6.1e+01 7.1e-0
8  0.46 0.75 0.061  37 9.5e+01  5.6e-07     2.3 0.96 0.020  -211   -93     1.9 2.8e+01 4.8e-0
9  0.47 0.75 0.076  27 7.6e+01  1.6e-06     2.3 0.96 0.022  -147   -62     1.9 2.6e+01 4.7e-0
10 0.46 0.73 0.109  18 3.4e+01  1.2e-02     1.9 0.96 0.015  -115   -58     1.9 1.1e+01 3.0e-0
11 0.45 0.68 0.153  10 2.3e+01  9.6e-03     1.8 0.97 0.019   -59   -28     2.0 8.7e+00 2.7e-0
12 0.49 0.75 0.172   3 3.8e+01  3.2e-08     2.1 0.96 0.055    13    22     1.9 1.6e+01 3.7e-0
13 0.50 0.73 0.232  -3 5.0e-01       NA     1.8 0.97    NA    NA    NA     2.1 1.3e-01 3.4e-0
14 0.50 0.73 0.314  -8 3.4e-05       NA     1.9 0.96    NA    NA    NA     2.0 1.0e-05 2.9e-0
15 0.50 0.75 0.368 -12 5.0e-06       NA     1.9 0.96    NA    NA    NA     2.0 1.7e-06 1.2e-0
16 0.50 0.76 0.569 -15 1.4e-11       NA     1.9 0.96    NA    NA    NA     2.0 6.2e-12 2.3e-0
17 0.50 0.76 1.000 -17 0.0e+00       NA     1.8 0.97    NA    NA    NA     2.0 4.4e-16 1.9e-1
18 0.50 0.76    NA -18 0.0e+00       NA     1.8 0.97    NA    NA    NA     2.0 4.4e-16 1.9e-1
```

Alternative estimates of the number of factors



2 dimensional solution

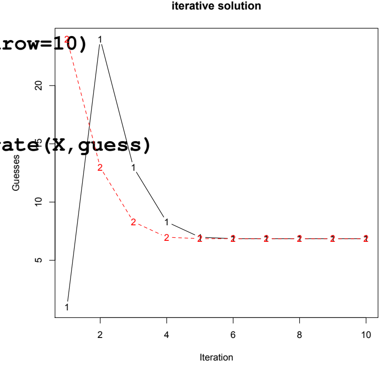


The basic concept of fitting

- Given some fit statistic, try to find the optimal values for that fit
  - Consider the case of the square root of 47
    - Guess X
    - find $47/Guess$
    - try a new $Guess = 47/Guess$
    - do it again
- The next example is a baby function to do this
- It is the concept, not the programming that is important

Iterative fitting to find a square root

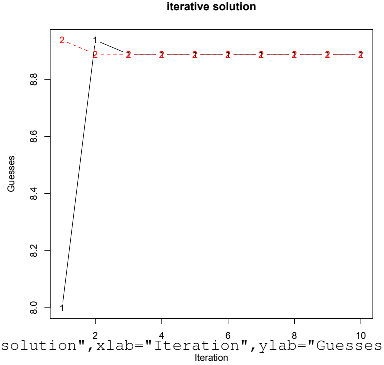
```
 iterative.sqrt <- function(X,guess) {
if(missing(guess)) guess <- 1  #a dumb guess
iterate <- function(X,guess) {
    low <- guess
    high <- X/guess
    guess <-((high+low)/2)
    guess}
Iter <- matrix(NA,ncol=2,nrow=10)
for (i in 1:10) {

Iter[i,1] <- guess
Iter[i,2] <- guess <- iterate(X,guess)
}
Iter}
```




```
> X <- 47
> iter <- iterative.sqrt(47)
> iter
           [,1]      [,2]
 [1,]  1.000000 24.000000
 [2,] 24.000000 12.979167
 [3,] 12.979167  8.300177
 [4,]  8.300177  6.981353
 [5,]  6.981353  6.856786
 [6,]  6.856786  6.855655
 [7,]  6.855655  6.855655
 [8,]  6.855655  6.855655
 [9,]  6.855655  6.855655
```

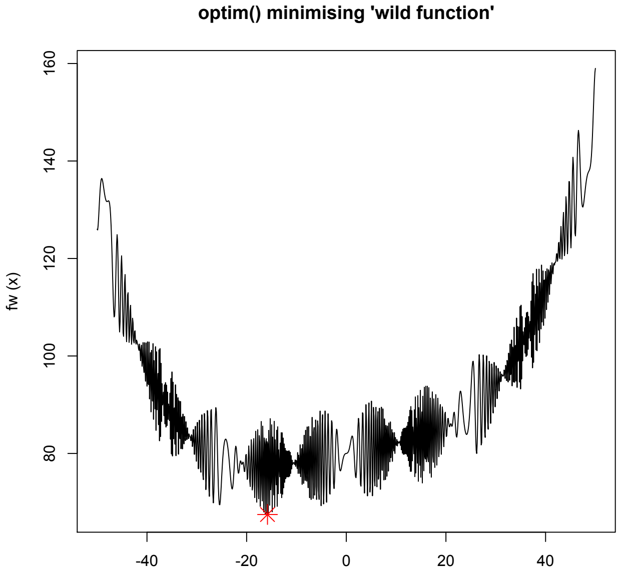
Iterative fitting to find a square root – a better guess

```
> X <- 79
> iter <- iterative.sqrt(X,8)
> iter
          [,1]     [,2]
 [1,] 8.000000 8.937500
 [2,] 8.937500 8.888330
 [3,] 8.888330 8.888194
 [4,] 8.888194 8.888194
 [5,] 8.888194 8.888194
 [6,] 8.888194 8.888194
 [7,] 8.888194 8.888194
 [8,] 8.888194 8.888194
 [9,] 8.888194 8.888194
[10,] 8.888194 8.888194
```




```
> matplot(iter,typ="b",main="iterative solution",xlab="Iteration",ylab="Guesses")
```

The optim function is one way to fit complex models



Iterative fitting has several steps

- Specifying a model and loss function
  - Ideally supplying a first derivative to the loss function
  - Can be done empirically
- Some way to evaluate the quality of the fit
- Minimization of function, but then how good is this minimum?
- The problem of local minima
- Can be resolved by multiple restarts starting at random locations

The basic equations

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

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

- where the *trace*, `tr`, of a matrix is the sum of the diagonal elements and the division by two reflects the symmetry of the **S** matrix. As discussed by Loehlin (2004), Equation 1 can be generalized to weight the residuals $(\mathbf{S} - \Sigma)$ by the inverse of the sample matrix, **S**, and thus to minimize

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

Maximum likelihood and its friends

$$E = \frac{1}{2} tr((\mathbf{S} - \Sigma)\mathbf{S}^{-1})^2 = \frac{1}{2} tr(\mathbf{I} - \Sigma\mathbf{S}^{\mathbf{-1}})^2. \tag{3}$$

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

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

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

$$E = \frac{1}{2}\left(tr(\Sigma^{-1}\mathbf{S}) - ln\left|\Sigma^{-1}\mathbf{S}\right| - p\right) \tag{5}$$

Perhaps a helpful intuitive explanation of Equation 5 is that if the model is correct, then $\Sigma = \mathbf{S}$ and thus $\Sigma^{-1}\mathbf{S} = \mathbf{I}$. The trace of an identity matrix of rank p is p, and the logarithm of $|\mathbf{I}|$ is 0. Thus, the value of E if the model has perfect fit is 0. With the assumption of multivariate normality of the residuals, and for large samples, a χ^2 statistic can be estimated for a model with p variables and f factors (Bartlett, 1951; Jöreskog, 1978; Lawley & Maxwell, 1962):

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

This χ^2 has degrees of freedom:

$$df = p * (p-1)/2 - p * f + f * (f-1)/2. \tag{7}$$

That is, the number of lower off-diagonal correlations - the number of unconstrained loadings (Lawley & Maxwell, 1962).

Fit statistics - it all starts with optimization

```
"fit" <- function(S,nf,fm,covar) {
          S.smc <- smc(S,covar)
           if((fm=="wls") | (fm =="gls") ) {S.inv <- solve(S)} else {
                                                               S.inv <- NULL
           if(!covar &&(sum(S.smc) == nf)) {start <- rep(.5,nf)}  else
                                                             start <- diag(S
                    #initial communality estimates are variance - smc
           if(fm=="ml")  {res <- optim(start, FAfn, FAgr, method =
                     "L-BFGS-B", lower = .005, upper = 1,
                        control = c(list(fnscale=1,
                                         parscale = rep(0.01, length(s
                                         nf = nf, S = S)
                 } else {
                                         res <- optim(start, fit.resid
                                          "L-BFGS-B", lower = .005, up
                                           control = c(list(fnscale =
                                          length(start))), nf= nf, S=
                                 }
   if((fm=="wls") | (fm=="gls") ) {Lambda <- FAout.wls(res$par, S, nf)
                                             Lambda <- FAout(res$par, S
    result <- list(loadings=Lambda,res=res,S=S)
    }
```

Maximum Likelihood criterion

```
FAfn <- function(Psi, S, nf)
   {
       sc <- diag(1/sqrt(Psi))
       Sstar <- sc %*% S %*% sc
       E <- eigen(Sstar, symmetric = TRUE, only.values = TRUE)
       e <- E$values[-(1:nf)]
       e <- sum(log(e) - e) - nf + nrow(S)
      -e
   }
   FAgr <- function(Psi, S, nf)  #the first derivatives
   {
       sc <- diag(1/sqrt(Psi))
       Sstar <- sc %*% S %*% sc
       E <- eigen(Sstar, symmetric = TRUE)
       L <- E$vectors[, 1:nf, drop = FALSE]
       load <- L %*% diag(sqrt(pmax(E$values[1:nf] - 1, 0)), nf)
       load <- diag(sqrt(Psi)) %*% load
       g <- load %*% t(load) + diag(Psi) - S      # g <- model - data
       diag(g)/Psi^2                              #normalized
   }
```

Goodness of Fit: a first approximation

- $\chi^2$and its relatives
- Root Mean Square Error of Approximation (RMSEA)
  - $\sqrt{\frac{\frac{\chi^2}{df}-1}{N-1}}$
- Akaike Information Criterion (AIC)
  - $\chi^2 + k(k-1) - 2df$
  - where k is number of variables
- Baysian Information Criterion (BIC)
  - $\chi^2 + (k(k-1) - 2df)ln(N)$
- Normed Fit Index (Bentler Bonnet Index)
  - $\frac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$
- Tucker Lews Index (non-normed fit index)
  - $\frac{\chi^2_{null}/df - \chi^2_{model}/df}{\chi^2_{null}/df}$

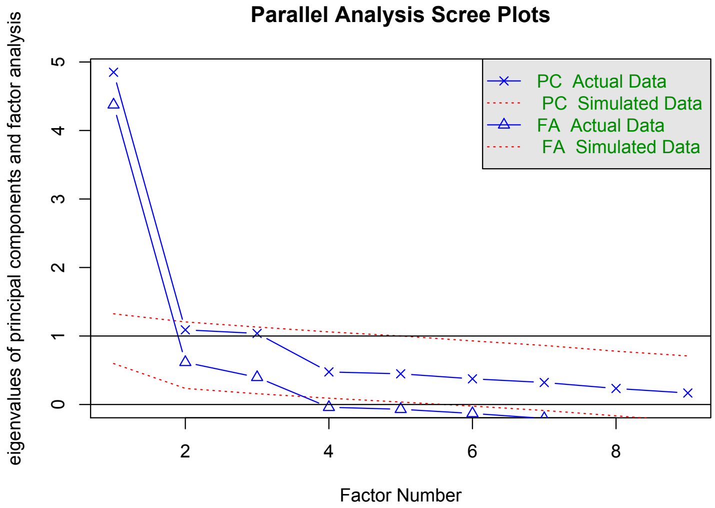
- A classic data set in factor analysis is from Thurstone & Thurstone (1941).
  - This data set has been used by Bechtoldt (1961) and others for demonstrations of factor analysis.
  - 213 subjects (taken as one of two data sets from the complete set of 425 subjects) were given 9 ability tests.
  - The variables are "Sentences" "Vocabulary" "Sent.Completion" "First.Letters" "4.Letter.Words" "Suffixes" "Letter.Series" "Pedigrees" "Letter.Group"
- Found in the `bifactor` data set.
- Thought to form three lower level factors and one higher level factor.
- First, lets use this data set to explore EFA and CFA

Exploratory analysis of Thurstone data set

```
> data(bifactor)
>  colnames(Thurstone) <- c("Sent","voc","Sent.com","first","4.let"
   "l.series","pedi","lett")
> Thurstone
```

	Sent	voc	Sent.com	first	4.let	Suff	l.series	pedi	lett
Sentences	1.00	0.83	0.78	0.44	0.43	0.45	0.45	0.54	0.38
Vocabulary	0.83	1.00	0.78	0.49	0.46	0.49	0.43	0.54	0.36
Sent.Completion	0.78	0.78	1.00	0.46	0.42	0.44	0.40	0.53	0.36
First.Letters	0.44	0.49	0.46	1.00	0.67	0.59	0.38	0.35	0.42
4.Letter.Words	0.43	0.46	0.42	0.67	1.00	0.54	0.40	0.37	0.45
Suffixes	0.45	0.49	0.44	0.59	0.54	1.00	0.29	0.32	0.32
Letter.Series	0.45	0.43	0.40	0.38	0.40	0.29	1.00	0.56	0.60
Pedigrees	0.54	0.54	0.53	0.35	0.37	0.32	0.56	1.00	0.45
Letter.Group	0.38	0.36	0.36	0.42	0.45	0.32	0.60	0.45	1.00

How many factors?



Extract 1 factor

```
> f1 <- fa(Thurstone,1,213)
> f1
```

```
Factor Analysis using method =  minres
Call: fa(r = Thurstone, nfactors = 1, n.obs = 213)
Standardized loadings based upon correlation matrix
          MR1   h2   u2
Sent      0.87 0.75 0.25
voc       0.88 0.77 0.23
Sent.com 0.83 0.70 0.30
first     0.62 0.39 0.61
4.let     0.61 0.37 0.63
Suff      0.59 0.34 0.66
l.series 0.57 0.32 0.68
pedi      0.64 0.41 0.59
lett      0.52 0.27 0.73

                MR1
SS loadings    4.32
Proportion Var 0.48
```

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

The degrees of freedom for the null model are  36  and the objective f
      was  5.2 with Chi Square of  1081.97
The degrees of freedom for the model are 27  and the objective
           function was  1.12

The root mean square of the residuals is  0.08
The df corrected root mean square of the residuals is  0.13
The number of observations was  213
          with Chi Square =  231.55  with prob <  2.1e-34

Tucker Lewis Index of factoring reliability =  0.738
RMSEA index =  0.191  and the 90 % confidence intervals are  0.19 0.19
BIC =  86.79
Fit based upon off diagonal values = 0.95
Measures of factor score adequacy
                                                 MR1
Correlation of scores with factors              0.96
Multiple R square of scores with factors        0.92
Minimum correlation of possible factor scores  0.85
>
```

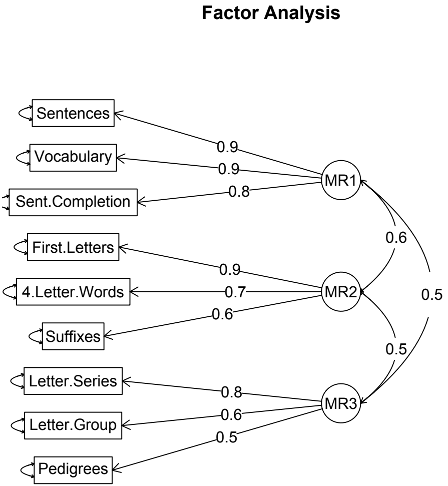
Extract 3 factors

```
> f3 <- fa(Thurstone,3,213)
Factor Analysis using method =  minres
Call: fa(r = Thurstone, nfactors = 3, n.obs = 213)
Standardized loadings based upon correlation matrix
           MR1   MR2   MR3   h2   u2
Sent       0.91 -0.04  0.04 0.82 0.18
voc        0.89  0.06 -0.03 0.84 0.16
Sent.com   0.83  0.04  0.00 0.73 0.27
first      0.00  0.86  0.00 0.73 0.27
4.let     -0.01  0.74  0.10 0.63 0.37
Suff       0.18  0.63 -0.08 0.50 0.50
l.series   0.03 -0.01  0.84 0.72 0.28
pedi       0.37 -0.05  0.47 0.50 0.50
lett      -0.06  0.21  0.64 0.53 0.47
                MR1  MR2  MR3
SS loadings    2.64 1.86 1.50
Proportion Var 0.29 0.21 0.17
Cumulative Var 0.29 0.50 0.67
 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.59 0.54
MR2 0.59 1.00 0.52
MR3 0.54 0.52 1.00
```

And fit statistics

```
Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are  36  and the objective f
        5.2 with Chi Square of  1081.97
The degrees of freedom for the model are 12  and the
          objective function was  0.01

The root mean square of the residuals is  0
The df corrected root mean square of the residuals is  0.01
The number of observations was  213  with Chi Square =  2.82  with pro

Tucker Lewis Index of factoring reliability =  1.027
RMSEA index =  0  and the 90 % confidence intervals are  0 0.023
BIC =  -61.51
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                 MR1  MR2  MR3
Correlation of 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
```

Examine the structure



Prepare for a sem run

```
> mod3 <- structure.diagram(f3,errors=TRUE)
> mod3
      Path                                  Parameter            Value
 [1,] "MR1->Sentences"                      "F1Sentences"        NA
 [2,] "MR1->Vocabulary"                     "F1Vocabulary"       NA
 [3,] "MR1->Sent.Completion"                "F1Sent.Completion"  NA
 [4,] "MR2->First.Letters"                  "F2First.Letters"    NA
 [5,] "MR2->4.Letter.Words"                 "F24.Letter.Words"   NA
 [6,] "MR2->Suffixes"                       "F2Suffixes"         NA
 [7,] "MR3->Letter.Series"                  "F3Letter.Series"    NA
 [8,] "MR1->Pedigrees"                      "F1Pedigrees"        NA
 [9,] "MR3->Pedigrees"                      "F3Pedigrees"        NA
[10,] "MR3->Letter.Group"                   "F3Letter.Group"     NA
[11,] "Sentences<->Sentences"               "x1e"                NA
[12,] "Vocabulary<->Vocabulary"             "x2e"                NA
[13,] "Sent.Completion<->Sent.Completion"   "x3e"                NA
[14,] "First.Letters<->First.Letters"       "x4e"                NA
[15,] "4.Letter.Words<->4.Letter.Words"     "x5e"                NA
[16,] "Suffixes<->Suffixes"                 "x6e"                NA
[17,] "Letter.Series<->Letter.Series"       "x7e"                NA
[18,] "Pedigrees<->Pedigrees"               "x8e"                NA
[19,] "Letter.Group<->Letter.Group"         "x9e"                NA
[20,] "MR2<->MR1"                           "rF2F1"              NA
[21,] "MR3<->MR1"                           "rF3F1"              NA
[22,] "MR3<->MR2"                           "rF3F2"              NA
[23,] "MR1<->MR1"                           NA                   "1"
[24,] "MR2<->MR2"                           NA                   "1"
[25,] "MR3<->MR3"                           NA                   "1"
attr(,"class")
[1] "mod"
```

Run sem (from the sem package)

```
> library(sem)
Attaching package: 'sem'
The following object(s) are masked from 'package:lavaan':
    sem
> sem3 <- sem(mod3,Thurstone,n.obs=213)
> sem3 <- sem(mod3,Thurstone,N=213)
> sem3
```

```
 Model Chisquare =  20.59086   Df =  23

     F1Sentences       F1Vocabulary F1Sent.Completion   F2First.Letter
       0.9039722          0.9138628         0.8575614         0.835950
     F1Pedigrees        F3Pedigrees    F3Letter.Group               x1
       0.3365901          0.4532355         0.7343129         0.182834
             x5e                x6e               x7e               x8
       0.3637161          0.5076057         0.3251154         0.505941
           rF3F2
       0.6015164

 Iterations =  23
>
```

Summary sem

```
> summary(sem3)

 Model Chisquare =  20.591   Df =  23 Pr(>Chisq) = 0.60606
 Chisquare (null model) =  1101.9   Df =  36
 Goodness-of-fit index =  0.97882
 Adjusted goodness-of-fit index =  0.95856
 RMSEA index =  0   90% CI: (NA, 0.049561)
 Bentler-Bonnett NFI =  0.98131
 Tucker-Lewis NNFI =  1.0035
 Bentler CFI =  1
 SRMR =  0.029241
 BIC =  -102.72

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.80800 -0.17100 -0.00001  0.01950  0.08700  1.29000
```

Path loadings

```
 Parameter Estimates
                   Estimate Std Error z value Pr(>|z|)
F1Sentences         0.90397  0.054203 16.6777 0.0000e+00 Sentences <--- MR1
F1Vocabulary        0.91386  0.053844 16.9724 0.0000e+00 Vocabulary <--- MR1
F1Sent.Completion   0.85756  0.055995 15.3151 0.0000e+00 Sent.Completion <--- MR1
F2First.Letters     0.83595  0.060664 13.7799 0.0000e+00 First.Letters <--- MR2
F24.Letter.Words    0.79767  0.061653 12.9381 0.0000e+00 4.Letter.Words <--- MR2
F2Suffixes          0.70171  0.064411 10.8943 0.0000e+00 Suffixes <--- MR2
F3Letter.Series     0.82151  0.067105 12.2422 0.0000e+00 Letter.Series <--- MR3
F1Pedigrees         0.33659  0.074302  4.5301 5.8969e-06 Pedigrees <--- MR1
F3Pedigrees         0.45324  0.079230  5.7205 1.0619e-08 Pedigrees <--- MR3
F3Letter.Group      0.73431  0.068128 10.7785 0.0000e+00 Letter.Group <--- MR3
x1e                 0.18283  0.028163  6.4920 8.4707e-11 Sentences <--> Sentences
x2e                 0.16486  0.027510  5.9927 2.0642e-09 Vocabulary <--> Vocabulary
x3e                 0.26459  0.033242  7.9596 1.7764e-15 Sent.Completion <--> Sent.Completion
x4e                 0.30119  0.050626  5.9492 2.6938e-09 First.Letters <--> First.Letters
x5e                 0.36372  0.052301  6.9543 3.5429e-12 4.Letter.Words <--> 4.Letter.Words
x6e                 0.50761  0.059994  8.4610 0.0000e+00 Suffixes <--> Suffixes
x7e                 0.32512  0.068680  4.7338 2.2041e-06 Letter.Series <--> Letter.Series
x8e                 0.50594  0.055984  9.0372 0.0000e+00 Pedigrees <--> Pedigrees
x9e                 0.46079  0.067704  6.8059 1.0042e-11 Letter.Group <--> Letter.Group
rF2F1               0.64002  0.051105 12.5237 0.0000e+00 MR1 <--> MR2
rF3F1               0.57469  0.060284  9.5331 0.0000e+00 MR1 <--> MR3
rF3F2               0.60152  0.063809  9.4268 0.0000e+00 MR2 <--> MR3
```

Modify the model to get rid of cross loading

```
> mod3.r <- edit(mod3)
Warning: class(es) of 'name' will be discarded
> mod3.r
      Path                                    Parameter             Value
 [1,] "MR1->Sentences"                        "F1Sentences"         NA
 [2,] "MR1->Vocabulary"                       "F1Vocabulary"        NA
 [3,] "MR1->Sent.Completion"                  "F1Sent.Completion"   NA
 [4,] "MR2->First.Letters"                    "F2First.Letters"     NA
 [5,] "MR2->4.Letter.Words"                   "F24.Letter.Words"    NA
 [6,] "MR2->Suffixes"                         "F2Suffixes"          NA
 [7,] "MR3->Letter.Series"                    "F3Letter.Series"     NA
 [8,] "MR3->Pedigrees"                        "F3Pedigrees"         NA
 [9,] "MR3->Letter.Group"                     "F3Letter.Group"      NA
[10,] "Sentences<->Sentences"                 "x1e"                 NA
[11,] "Vocabulary<->Vocabulary"               "x2e"                 NA
[12,] "Sent.Completion<->Sent.Completion"     "x3e"                 NA
[13,] "First.Letters<->First.Letters"         "x4e"                 NA
[14,] "4.Letter.Words<->4.Letter.Words"       "x5e"                 NA
[15,] "Suffixes<->Suffixes"                   "x6e"                 NA
[16,] "Letter.Series<->Letter.Series"         "x7e"                 NA
[17,] "Pedigrees<->Pedigrees"                 "x8e"                 NA
[18,] "Letter.Group<->Letter.Group"           "x9e"                 NA
[19,] "MR2<->MR1"                             "rF2F1"               NA
[20,] "MR3<->MR1"                             "rF3F1"               NA
```

Fit the modified model

```
> sem.3 <- sem(mod3.r,Thurstone,N=213)
> summary(sem.3)
```

```
Model Chisquare =  38.196   Df =  24 Pr(>Chisq) = 0.033101
Chisquare (null model) =  1101.9   Df =  36
Goodness-of-fit index =  0.95957
Adjusted goodness-of-fit index =  0.9242
RMSEA index =  0.052822   90% CI: (0.015262, 0.083067)
Bentler-Bonnett NFI =  0.96534
Tucker-Lewis NNFI =  0.98002
Bentler CFI =  0.98668
SRMR =  0.043595
BIC =  -90.475

Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.9720 -0.4160  0.0000  0.0401  0.0939  1.6300
```

Show the loadings

```
Parameter Estimates
                  Estimate Std Error z value Pr(>|z|)
F1Sentences       0.90471  0.054221  16.6857 0.0000e+00 Sentences <--- MR1
F1Vocabulary      0.91382  0.053888  16.9578 0.0000e+00 Vocabulary <--- MR1
F1Sent.Completion 0.85608  0.056059  15.2711 0.0000e+00 Sent.Completion <--- MR1
F2First.Letters   0.83576  0.060686  13.7720 0.0000e+00 First.Letters <--- MR2
F24.Letter.Words  0.79718  0.061676  12.9254 0.0000e+00 4.Letter.Words <--- MR2
F2Suffixes        0.70256  0.064407  10.9080 0.0000e+00 Suffixes <--- MR2
F3Letter.Series   0.78081  0.065156  11.9837 0.0000e+00 Letter.Series <--- MR3
F3Pedigrees       0.72016  0.067081  10.7357 0.0000e+00 Pedigrees <--- MR3
F3Letter.Group    0.70349  0.066950  10.5078 0.0000e+00 Letter.Group <--- MR3
x1e               0.18150  0.028401   6.3905 1.6540e-10 Sentences <--> Sentences
x2e               0.16493  0.027798   5.9332 2.9712e-09 Vocabulary <--> Vocabulary
x3e               0.26713  0.033469   7.9815 1.5543e-15 Sent.Completion <--> Sent.Completion
x4e               0.30150  0.050687   5.9483 2.7091e-09 First.Letters <--> First.Letters
x5e               0.36450  0.052358   6.9617 3.3609e-12 4.Letter.Words <--> 4.Letter.Words
x6e               0.50641  0.059962   8.4456 0.0000e+00 Suffixes <--> Suffixes
x7e               0.39033  0.061591   6.3375 2.3352e-10 Letter.Series <--> Letter.Series
x8e               0.48137  0.065376   7.3631 1.7986e-13 Pedigrees <--> Pedigrees
x9e               0.50510  0.065219   7.7447 9.5479e-15 Letter.Group <--> Letter.Group
rF2F1             0.64270  0.050945  12.6156 0.0000e+00 MR1 <--> MR2
rF3F1             0.67000  0.053807  12.4519 0.0000e+00 MR1 <--> MR3
rF3F2             0.63718  0.059368  10.7327 0.0000e+00 MR2 <--> MR3

 Iterations =  36
```

Exploratory versus confirmatory

- The real power of confirmatory analysis is the ability to test paths and models.
- EFA fits $\mathbf{R} = \mathbf{FF'} + \mathbf{U^2}$
    - for k variables and nf factors
    - The number of parameters to estimate is $k * nf - (nf * (nf - 1)/2)$
    - The number of observed correlations is $k * (k - 1)/2$
    - degrees of freedom $= k * (k-1)/2 - k * nf - (nf * (nf - 1)/2)$
    - all parameters are allowed to vary
- Confirmatory analysis fixes some parameters to zero, doesn't estimate others
    - This makes a more constrained model, with more degrees of freedom
- Compare the EFA and CFA of the Thurstone data sets (12 versus 24 df)
- Fixing parameters to specified values is also possible

Fix loadings on each factor to be equal, but different across factors

```
> mod.3f <- edit(mod.3r)
> mod.3f
      Path                                   Parameter Value
 [1,] "MR1->Sentences"                       "F1"      NA
 [2,] "MR1->Vocabulary"                      "F1"      NA
 [3,] "MR1->Sent.Completion"                 "F1"      NA
 [4,] "MR2->First.Letters"                   "F2"      NA
 [5,] "MR2->4.Letter.Words"                  "F2"      NA
 [6,] "MR2->Suffixes"                        "F2"      NA
 [7,] "MR3->Letter.Series"                   "F3"      NA
 [8,] "MR3->Pedigrees"                       "F3"      NA
 [9,] "MR3->Letter.Group"                    "F3"      NA
[10,] "Sentences<->Sentences"                "x1e"     NA
[11,] "Vocabulary<->Vocabulary"              "x2e"     NA
[12,] "Sent.Completion<->Sent.Completion"    "x3e"     NA
[13,] "First.Letters<->First.Letters"        "x4e"     NA
[14,] "4.Letter.Words<->4.Letter.Words"      "x5e"     NA
[15,] "Suffixes<->Suffixes"                  "x6e"     NA
[16,] "Letter.Series<->Letter.Series"        "x7e"     NA
[17,] "Pedigrees<->Pedigrees"                "x8e"     NA
[18,] "Letter.Group<->Letter.Group"          "x9e"     NA
[19,] "MR2<->MR1"                            "rF2F1"   NA
[20,] "MR3<->MR1"                            "rF3F1"   NA
[21,] "MR3<->MR2"                            "rF3F2"   NA
[22,] "MR1<->MR1"                            NA        "1"
[23,] "MR2<->MR2"                            NA        "1"
[24,] "MR3<->MR3"                            NA        "1"
```

sem with parameters fixed to be equal

```
> sem.3f <- sem(mod.3f,Thurstone,N=213)
> summary(sem.3f)
 Model Chisquare =  44.287   Df =  30 Pr(>Chisq) = 0.044911
 Chisquare (null model) =  1101.9   Df =  36
 Goodness-of-fit index =  0.95336
 Adjusted goodness-of-fit index =  0.93004
 RMSEA index =  0.047396   90% CI: (0.0074497, 0.075428)
 Bentler-Bonnett NFI =  0.9598
 Tucker-Lewis NNFI =  0.98392
 Bentler CFI =  0.9866
 SRMR =  0.05096
 BIC =  -116.55
 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.1500 -0.4100 -0.1220 -0.0464  0.4190  1.2800
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
F1    0.89541  0.047217  18.9636 0.0000e+00 Sentences <--- MR1
F2    0.78422  0.046718  16.7865 0.0000e+00 First.Letters <--- MR2
F3    0.73500  0.046620  15.7656 0.0000e+00 Letter.Series <--- MR3
x1e   0.18337  0.026644   6.8824 5.8857e-12 Sentences <--> Sentences
x2e   0.17162  0.025846   6.6402 3.1332e-11 Vocabulary <--> Vocabulary
x3e   0.25956  0.033050   7.8535 3.9968e-15 Sent.Completion <--> Sent.Completion
x4e   0.33085  0.045964   7.1980 6.1107e-13 First.Letters <--> First.Letters
x5e   0.37081  0.049050   7.5599 4.0412e-14 4.Letter.Words <--> 4.Letter.Words
x6e   0.48237  0.058982   8.1783 2.2204e-16 Suffixes <--> Suffixes
x7e   0.41554  0.055363   7.5058 6.1062e-14 Letter.Series <--> Letter.Series
x8e   0.47323  0.059526   7.9500 1.7764e-15 Pedigrees <--> Pedigrees
x9e   0.49870  0.062254   8.0108 1.1102e-15 Letter.Group <--> Letter.Group
rF2F1 0.65013  0.050468  12.8821 0.0000e+00 MR1 <--> MR2
rF3F1 0.67629  0.051829  13.0485 0.0000e+00 MR1 <--> MR3
rF3F2 0.64416  0.058881  10.9400 0.0000e+00 MR2 <--> MR3
```

Do these models differ?

- Do the models differ?
- compare the two χ^2 compared to degrees of freedom

```
> anova(sem.3,sem.3f)
LR Test for Difference Between Models

        Model Df Model Chisq Df LR Chisq Pr(>Chisq)
Model 1       24      38.196
Model 2       30      44.287  6   6.0906     0.4131
```

Hierarchical solutions for ability and personality

- Most ability tests, many personality tests are said to have a hierarchical structure
  - Lower level factors
  - Higher level factors to account for the correlation of lower level factors
- Can bested using EFA (omega function) or CFA (sem)

Omega for Thurstone

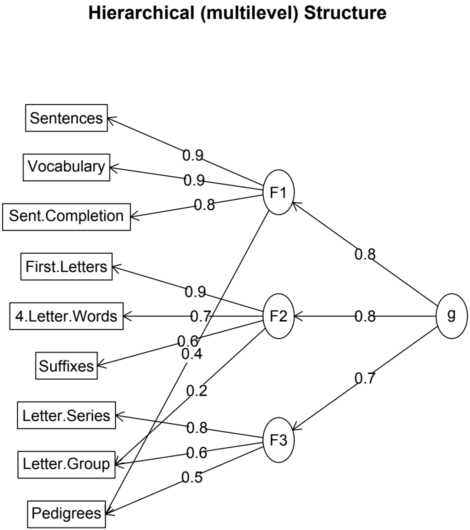
```
> om.t <- omega(Thurstone,n.obs=213)
> om.t
> diagram(om.t,sl=FALSE)
```

```
Omega
Call: omega(m = Thurstone, n.obs = 213)
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  0.57              0.82 0.18 0.61
Vocabulary      0.73  0.55              0.84 0.16 0.63
Sent.Completion 0.68  0.52              0.73 0.27 0.63
First.Letters   0.65        0.56        0.73 0.27 0.57
4.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              0.61 0.72 0.28 0.48
Pedigrees       0.58  0.23        0.34 0.50 0.50 0.66
Letter.Group    0.54              0.46 0.53 0.47 0.56
With eigenvalues of:
   g  F1*  F2*  F3*
3.58 0.96 0.74 0.71
```

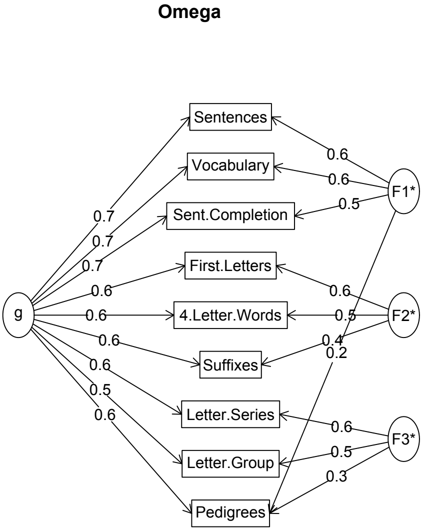
Omega output (continued)

```
general/max  3.71   max/min =   1.35
mean percent general =  0.6    with sd =  0.05 and cv of  0.09

The degrees of freedom are 12  and the fit is  0.01
The number of observations was  213  with Chi Square =  2.82  with prob <  1
The root mean square of the residuals is  0
The df corrected root mean square of the residuals is  0.01
RMSEA index =  0  and the 90 % confidence intervals are  0 0.023
BIC =  -61.51

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 =  307.1  with prob <  2.8e-49
The root mean square of the residuals is  0.1
The df corrected root mean square of the residuals is  0.16

RMSEA index =  0.224  and the 90 % confidence intervals are  0.223 0.226
BIC =  162.35

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 0.52 0.56
Minimum correlation of factor score estimates 0.49 0.08 0.03 0.11
```

A diagram is worth a 1000 words



Schmid Leiman transformation



Confirmatory using omega

- A confirmatory analysis will have more degrees of freedom because some paths are set to 0.
- Can do the run directly in sem or can use `omegaSem` to do it for you
  - `omegaSem` does an exploratory analysis and then specifies the sem paths for largest paths
  - Then calls sem to do the hierarchical analysis.
  - Reports both exploratory and confirmatory statistics
- Could just do the analysis directly!

omegaSem of Thurstone data set

```
> mod.om <- omegaSem(Thurstone,n.obs=213)
> mod.om
Call: omegaSem(m = Thurstone, n.obs = 213)
... (all the regular omega output )
  Then
 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
Sentences       0.77 0.49            0.83 0.17
Vocabulary      0.79 0.45            0.83 0.17
Sent.Completion 0.75 0.40            0.73 0.27
First.Letters   0.61      0.61       0.75 0.25
4.Letter.Words  0.60      0.51       0.61 0.39
Suffixes        0.57      0.39       0.48 0.52
Letter.Series   0.57           0.73 0.85 0.15
Pedigrees       0.66           0.25 0.50 0.50
Letter.Group    0.53           0.41 0.45 0.55

With eigenvalues of:
   g  F1*  F2*  F3*
3.88 0.61 0.79 0.76
>
```

sem output is available as part of omegaSem

```
> summary(mod.om$sem)  #note the use of a sub object
#use str(mod.om) to see what is available
```

```
Model Chisquare =  24.216   Df =  18 Pr(>Chisq) = 0.14807
Chisquare (null model) =  1101.9   Df =  36
Goodness-of-fit index =  0.97578
Adjusted goodness-of-fit index =  0.93944
RMSEA index =  0.040361   90% CI: (NA, 0.077994)
Bentler-Bonnett NFI =  0.97802
Tucker-Lewis NNFI =  0.98834
Bentler CFI =  0.99417
SRMR =  0.034895
BIC =  -72.287

Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.8210 -0.3340  0.0000  0.0282  0.1560  1.8000
```

more sem output

```
Parameter Estimates
                   Estimate Std Error z value  Pr(>|z|)
Sentences          0.76787  0.072626  10.57291 0.0000e+00 Sentences
Vocabulary         0.79092  0.072418  10.92170 0.0000e+00 Vocabular
Sent.Completion    0.75362  0.073402  10.26709 0.0000e+00 Sent.Comp
First.Letters      0.60838  0.072201   8.42617 0.0000e+00 First.Let
4.Letter.Words     0.59733  0.073851   8.08843 6.6613e-16 4.Letter.
Suffixes           0.57179  0.071492   7.99792 1.3323e-15 Suffixes
Letter.Series      0.56689  0.074271   7.63282 2.2871e-14 Letter.Se
Pedigrees          0.66233  0.069321   9.55455 0.0000e+00 Pedigrees
Letter.Group       0.52995  0.078985   6.70955 1.9522e-11 Letter.Gr
F1*Sentences       0.48787  0.085457   5.70898 1.1366e-08 Sentences
F1*Vocabulary      0.45232  0.090422   5.00233 5.6640e-07 Vocabular
F1*Sent.Completion 0.40445  0.093402   4.33024 1.4895e-05 Sent.Comp
F2*First.Letters   0.61405  0.085794   7.15733 8.2268e-13 First.Let
F2*4.Letter.Words  0.50581  0.084848   5.96130 2.5024e-09 4.Letter.
F2*Suffixes        0.39432  0.078289   5.03671 4.7359e-07 Suffixes
F3*Letter.Series   0.72730  0.159499   4.55988 5.1184e-06 Letter.Se
F3*Pedigrees       0.24684  0.089011   2.77317 5.5513e-03 Pedigrees
F3*Letter.Group    0.40915  0.122180   3.34875 8.1177e-04 Letter.Gr
e1                 0.17236  0.034113   5.05265 4.3571e-07 Sentences
e2                 0.16984  0.030037   5.65438 1.5641e-08 Vocabular
e3                 0.26847  0.033188   8.08958 6.6613e-16 Sent.Comp
e4                 0.25281  0.079472   3.18115 1.4669e-03 First.Let
```

The path model is there if you want

```
> mod.om$omegaSem$model
      Path                                  Parameter            Initial
 [1,] "g->Sentences"                        "Sentences"          NA
 [2,] "g->Vocabulary"                       "Vocabulary"         NA
 [...

  [7,] "g->Letter.Series"                    "Letter.Series"      NA
 [8,] "g->Pedigrees"                        "Pedigrees"          NA
 [9,] "g->Letter.Group"                     "Letter.Group"       NA
[10,] "F1*->Sentences"                      "F1*Sentences"       NA
[11,] "F1*->Vocabulary"                     "F1*Vocabulary"      NA
[12,] "F1*->Sent.Completion"                "F1*Sent.Completion" NA
[13,] "F2*->First.Letters"                  "F2*First.Letters"   NA
[14,] "F2*->4.Letter.Words"                 "F2*4.Letter.Words"  NA
[15,] "F2*->Suffixes"                       "F2*Suffixes"        NA
[16,] "F3*->Letter.Series"                  "F3*Letter.Series"   NA
[17,] "F3*->Pedigrees"                      "F3*Pedigrees"       NA
[18,] "F3*->Letter.Group"                   "F3*Letter.Group"    NA
[19,] "Sentences<->Sentences"               "e1"                 NA
[20,] "Vocabulary<->Vocabulary"             "e2"                 NA
[...
[27,] "Letter.Group<->Letter.Group"         "e9"                 NA
[28,] "F1*<->F1*"                           NA                   "1"
```

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

Bechtoldt, H. (1961). An empirical study of the factor analysis stability hypothesis. *Psychometrika*, *26*(4), 405–432.

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

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

Loehlin, J. C. (2004). *Latent variable models: an introduction to factor, path, and structural equation analysis* (4th ed.). Mahwah, N.J.: L. Erlbaum Associates.

Thurstone, L. L. & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago, Ill.: The University of Chicago press.