Psychology 454: Latent Variable Modeling

Using the lavaan package for latent variable modeling

Department of Psychology
Northwestern University
Evanston, Illinois USA

January, 2011

Outline

3 major structural modeling programs in R

- sem (by John Fox)
  - Uses ram notation for parameters
  - *psych* will work as a front end for developing parameters
  - Development work seems to have switched to OpenMx
  - Will not do multiple groups
- lavaan (by Yves Rosseel)
  - Uses a more compact notation than sem
  - Will work on multiple groups
  - Still under development
- OpenMx (by Michael Neal, Steve Boker and the OpenMx group)
  - Very powerful structural equation package
  - Based upon Mx (developed for behavioral geneticists)
  - Somewhat idiosyncratic syntax

Getting lavaan

- Beta version (0.4-5) may be downloaded from lavaan website.
  - Will handle covariance matrices – objects to but will run correlations matrices

```
install.packages("lavaan", repos="http://www.da.ugent.be", type="source")
library(lavaan)
```

- R version on CRAN is 0.3-3.

```
install.packages("lavaan")
library(lavaan)
```

  - Will not handle covariance or correlation matrices
- Documentation is also available at http://users.ugent.be/~yrosseel/lavaan/lavaan_usersguide_0.3-1.pdf
- For more information about *lavaan* go to http://lavaan.ugent.be/

Using lavaan

- Confirmatory factoring models with cfa
  - Single group
  - Multiple group (factor invariance issues)
- Structural Equation Models with sem
  - Single group models
    - Regression models
    - Complex regression models
    - latent variable models

Confirmatory models for a Thurstone data set – Bechtoldt.1 and then Becholdt.2

- Bechtoldt (1961) split a data set from Thurstone & Thurstone (1941) into two equal parts (N=212, 213) to examine factor stability.
  - One set has become known as the "Thurstone" data set in SAS and in McDonald (1999).
  - Both are available in the *psych* package and can be analyzed using `cfa` in *lavaan*
- The following script forms two subsets (b2 is equivalent to "Thurstone") and then does a `cfa`

```
data(bifactor)}
b1 <- Bechtoldt.1[c(3:8,15:17),c(3:8,15:17)]
b2 <- Bechtoldt.2[c(3:8,15:17),c(3:8,15:17)]
Thurstone.mod <- ' F1 =~ Sentences  +        Vocabulary   +     Completion
                   F2 =~ First_Letters  +    Four_letter_words + Suffixes
                   F3 =~  Letter_Series +    Pedigrees     +     Letter_Grouping'
t.cfa.2 <- cfa(Thurstone.mod,sample.cov=b2,sample.nobs=213,std.lv=TRUE)
summary(t.cfa.2)
```

lavaan output for a cfa – first a warning

```
> t.cfa.2 <- cfa(Thurstone.mod,sample.cov=b2,
                    sample.nobs=213,std.lv=TRUE)
Warning message:
In Sample(data = data, group = group, sample.cov = sample.cov,
                     sample.mean = sample.mean,  :

sample covariance matrix looks like a correlation matrix!

  lavaan currently does not support the analysis of correlation
  matrices; the standard errors in the summary output will be most
  likely wrong; see the following reference:

  Cudeck, R. (1989). Analysis of correlation matrices using covariance
  structure models. Psychological Bulletin, 105, 317-327.
```

Limited output unless requested

```
> summary(t.cfa.2)
Lavaan (0.4-5) converged normally after 28 iterations

  Number of observations                          213

  Estimator                                        ML
  Minimum Function Chi-square                  38.376
  Degrees of freedom                               24
  P-value                                       0.032
```

More complete output

```
> summary(t.cfa.2,fit.measures=TRUE)

Lavaan (0.4-5) converged normally after 28 iterations

  Number of observations                           213

  Estimator                                         ML
  Minimum Function Chi-square                   38.376
  Degrees of freedom                                24
  P-value                                        0.032

Chi-square test baseline model:

  Minimum Function Chi-square                 1107.090
  Degrees of freedom                                36
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.987
  Tucker-Lewis Index (TLI)                       0.980

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2181.238
  Loglikelihood unrestricted model (H1)      -2162.050

  Number of free parameters                         21
  Akaike (AIC)                                4404.476
  Bayesian (BIC)                              4475.063
  Sample-size adjusted Bayesian (BIC)         4408.520

Root Mean Square Error of Approximation:

  RMSEA                                          0.053
  90 Percent Confidence Interval          0.016  0.083
  P-value RMSEA <= 0.05                          0.404

Standardized Root Mean Square Residual:

  SRMR                                           0.044
```

With parameter estimates - notice that we fixed latent variances to 1

```
Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  F1 =~
    Sentences         0.903    0.054   16.727    0.000
    Vocabulary        0.912    0.054   17.005    0.000
    Completion        0.854    0.056   15.317    0.000
  F2 =~
    First_Letters     0.834    0.060   13.783    0.000
    Four_letter_w     0.795    0.061   12.937    0.000
    Suffixes          0.701    0.064   10.960    0.000
  F3 =~
    Letter_Series     0.779    0.064   12.173    0.000
    Pedigrees         0.718    0.065   10.998    0.000
    Letter_Groupi     0.702    0.066   10.679    0.000

Covariances:
  F1 ~~
    F2                0.643    0.050   12.755    0.000
    F3                0.670    0.051   13.153    0.000
  F2 ~~
    F3                0.637    0.058   10.900    0.000

Variances:
    Sentences         0.181    0.028    6.388    0.000
    Vocabulary        0.164    0.028    5.953    0.000
    Completion        0.266    0.033    8.026    0.000
    First_Letters     0.300    0.051    5.923    0.000
    Four_letter_w     0.363    0.052    6.941    0.000
    Suffixes          0.504    0.059    8.513    0.000
    Letter_Series     0.388    0.059    6.594    0.000
    Pedigrees         0.479    0.062    7.751    0.000
    Letter_Groupi     0.503    0.063    7.995    0.000
    F1                1.000
    F2                1.000
    F3                1.000
```

Alternative parameterization one variable path per latent set to 1

summary(t.cfa.2,fit.measures=TRUE)

```
                  Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  F1 =~
    Sentences         1.000
    Vocabulary        1.010    0.051   19.938    0.000
    Completion        0.946    0.054   17.644    0.000
  F2 =~
    First_Letters     1.000
    Four_letter_w     0.954    0.082   11.668    0.000
    Suffixes          0.841    0.081   10.326    0.000
  F3 =~
    Letter_Series     1.000
    Pedigrees         0.922    0.097    9.469    0.000
    Letter_Groupi     0.901    0.097    9.288    0.000

Covariances:
  F1 ~~
    F2                0.484    0.072    6.751    0.000
    F3                0.471    0.071    6.653    0.000
  F2 ~~
    F3                0.414    0.068    6.118    0.000

Variances:
    Sentences             0.181    0.028    6.388    0.000
    Vocabulary            0.164    0.028    5.953    0.000
    Completion            0.266    0.033    8.026    0.000
    First_Letters         0.300    0.051    5.923    0.000
    Four_letter_w         0.363    0.052    6.941    0.000
    Suffixes              0.504    0.059    8.513    0.000
    Letter_Series         0.388    0.059    6.594    0.000
    Pedigrees             0.479    0.062    7.751    0.000
    Letter_Groupi         0.503    0.063    7.995    0.000
    F1                    0.815    0.097    8.363    0.000
    F2                    0.695    0.101    6.891    0.000
    F3                    0.607    0.100    6.087    0.000
```

Compare to the efa from psych and `sem` from sem

- This data set has been discussed before (many times, see e.g., Week 4)
  - We compared methods of factor extraction (minres and mle) and rotation (varimax and oblimin)
  - We compared EFA and SEM solutions
- Now compare those solutions to the *lavaan* solutions
- Both in ease of set up and in statistical modeling

create the sem commands by using psych

```
f3 <- fa(Thurstone,3,fm='mle')
mod3 <- structure.diagram(f3,cut=.45,errors=TRUE)
mod3
```

```
      Path          Parameter Value
 [1,] "ML1->V1"    "F1V1"    NA
 [2,] "ML1->V2"    "F1V2"    NA
 [3,] "ML1->V3"    "F1V3"    NA
 [4,] "ML2->V4"    "F2V4"    NA
 [5,] "ML2->V5"    "F2V5"    NA
 [6,] "ML2->V6"    "F2V6"    NA
 [7,] "ML3->V7"    "F3V7"    NA
 [8,] "ML3->V8"    "F3V8"    NA
 [9,] "ML3->V9"    "F3V9"    NA
[10,] "V1<->V1"    "x1e"     NA
[11,] "V2<->V2"    "x2e"     NA
...
[18,] "V9<->V9"    "x9e"     NA
[19,] "ML2<->ML1" "rF2F1"    NA
[20,] "ML3<->ML1" "rF3F1"    NA
[21,] "ML3<->ML2" "rF3F2"    NA
[22,] "ML1<->ML1" NA         "1"
[23,] "ML2<->ML2" NA         "1"
[24,] "ML3<->ML3" NA         "1"
```

Running sem

```
> rownames(Thurstone) <- colnames(Thurstone) #to get the names to match the modl
> sem3 <- sem(mod3,Thurstone,N=213)
> summary(sem3,digits=2)
```

```
 Model Chisquare =  38   Df =  24 Pr(>Chisq) = 0.033
 Chisquare (null model) =  1102   Df =  36
 Goodness-of-fit index =  0.96
 Adjusted goodness-of-fit index =  0.92
 RMSEA index =  0.053   90% CI: (0.015, 0.083)
 Bentler-Bonnett NFI =  0.97
 Tucker-Lewis NNFI =  0.98
 Bentler CFI =  0.99
 SRMR =  0.044
 BIC =  -90

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -0.97   -0.42    0.00    0.04    0.09    1.63
```

With parameter estimates

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
F1V1  0.90     0.054     16.7    0.0e+00  V1 <--- ML1
F1V2  0.91     0.054     17.0    0.0e+00  V2 <--- ML1
F1V3  0.86     0.056     15.3    0.0e+00  V3 <--- ML1
F2V4  0.84     0.061     13.8    0.0e+00  V4 <--- ML2
F2V5  0.80     0.062     12.9    0.0e+00  V5 <--- ML2
F2V6  0.70     0.064     10.9    0.0e+00  V6 <--- ML2
F3V7  0.78     0.065     12.0    0.0e+00  V7 <--- ML3
F3V8  0.72     0.067     10.7    0.0e+00  V8 <--- ML3
F3V9  0.70     0.067     10.5    0.0e+00  V9 <--- ML3
x1e   0.18     0.028      6.4    1.7e-10  V1 <--> V1
x2e   0.16     0.028      5.9    3.0e-09  V2 <--> V2
x3e   0.27     0.033      8.0    1.6e-15  V3 <--> V3
x4e   0.30     0.051      5.9    2.7e-09  V4 <--> V4
x5e   0.36     0.052      7.0    3.4e-12  V5 <--> V5
x6e   0.51     0.060      8.4    0.0e+00  V6 <--> V6
x7e   0.39     0.062      6.3    2.3e-10  V7 <--> V7
x8e   0.48     0.065      7.4    1.8e-13  V8 <--> V8
x9e   0.51     0.065      7.7    9.5e-15  V9 <--> V9
rF2F1 0.64     0.051     12.6    0.0e+00  ML1 <--> ML2
rF3F1 0.67     0.054     12.5    0.0e+00  ML1 <--> ML3
rF3F2 0.64     0.059     10.7    0.0e+00  ML2 <--> ML3
```

A direct comparison of statistical estimates

```
Model Chisquare =  38   Df =  24 Pr(>Chisq) = 0.033
Chisquare (null model) =  1102   Df =  36
Goodness-of-fit index =  0.96
Adjusted goodness-of-fit index =  0.92
RMSEA index =  0.053   90% CI: (0.015, 0.083)
Bentler-Bonnett NFI =  0.97
Tucker-Lewis NNFI =  0.98
Bentler CFI =  0.99
SRMR =  0.044
BIC =  -90

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -0.97   -0.42    0.00    0.04    0.09    1.63
```

```
  Number of observations                           213

  Estimator                                         ML
  Minimum Function Chi-square                   38.376
  Degrees of freedom                                24
  P-value                                        0.032

Chi-square test baseline model:

  Minimum Function Chi-square                 1107.090
  Degrees of freedom                                36
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.987
  Tucker-Lewis Index (TLI)                       0.980
Root Mean Square Error of Approximation:

  RMSEA                                          0.053
  90 Percent Confidence Interval          0.016  0.083
  P-value RMSEA <= 0.05                          0.404

Standardized Root Mean Square Residual:

  SRMR                                           0.044
```

A direct comparison of parameter estimates

sem

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
F1V1  0.90     0.054     16.7    0.0e+00  V1 <--- ML1
F1V2  0.91     0.054     17.0    0.0e+00  V2 <--- ML1
F1V3  0.86     0.056     15.3    0.0e+00  V3 <--- ML1
F2V4  0.84     0.061     13.8    0.0e+00  V4 <--- ML2
F2V5  0.80     0.062     12.9    0.0e+00  V5 <--- ML2
F2V6  0.70     0.064     10.9    0.0e+00  V6 <--- ML2
F3V7  0.78     0.065     12.0    0.0e+00  V7 <--- ML3
F3V8  0.72     0.067     10.7    0.0e+00  V8 <--- ML3
F3V9  0.70     0.067     10.5    0.0e+00  V9 <--- ML3
x1e   0.18     0.028      6.4    1.7e-10  V1 <--> V1
x2e   0.16     0.028      5.9    3.0e-09  V2 <--> V2
x3e   0.27     0.033      8.0    1.6e-15  V3 <--> V3
x4e   0.30     0.051      5.9    2.7e-09  V4 <--> V4
x5e   0.36     0.052      7.0    3.4e-12  V5 <--> V5
x6e   0.51     0.060      8.4    0.0e+00  V6 <--> V6
x7e   0.39     0.062      6.3    2.3e-10  V7 <--> V7
x8e   0.48     0.065      7.4    1.8e-13  V8 <--> V8
x9e   0.51     0.065      7.7    9.5e-15  V9 <--> V9
rF2F1 0.64     0.051     12.6    0.0e+00  ML1 <--> ML2
rF3F1 0.67     0.054     12.5    0.0e+00  ML1 <--> ML3
rF3F2 0.64     0.059     10.7    0.0e+00  ML2 <--> ML3
```

lavaan

```
Latent variables:
  F1 =~
    Sentences          0.903    0.054    16.727    0.000
    Vocabulary         0.912    0.054    17.005    0.000
    Completion         0.854    0.056    15.317    0.000
  F2 =~
    First_Letters      0.834    0.060    13.783    0.000
    Four_letter_w      0.795    0.061    12.937    0.000
    Suffixes           0.701    0.064    10.960    0.000
  F3 =~
    Letter_Series      0.779    0.064    12.173    0.000
    Pedigrees          0.718    0.065    10.998    0.000
    Letter_Groupi      0.702    0.066    10.679    0.000

Covariances:
  F1 ~~
    F2                 0.643    0.050    12.755    0.000
    F3                 0.670    0.051    13.153    0.000
  F2 ~~
    F3                 0.637    0.058    10.900    0.000
```

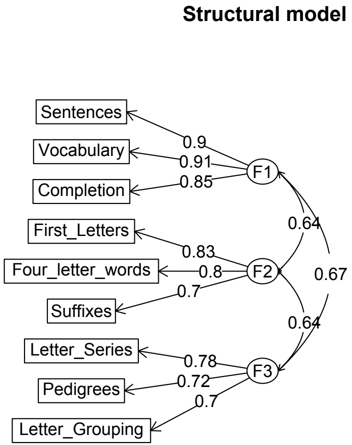
lavaan.diagram

Currently, *lavaan* does not draw structural diagrams. But, it is not hard to form a simple function to draw lavaan diagrams from lavaan output using tools from *psych*.

```
"lavaan.diagram" <-
function(fit,model="cfa",...) {
#if (is.null(fit@Model@GLIST[[1]]$beta)) {model <- "cfa"} else {model <- "sem"}
if(model=="cfa") {fx=fit@Model@GLIST$lambda
                  colnames(fx) <- fit@Model@dimNames$lambda[[2]]
                  Phi <-  fit@Model@GLIST$psi
                  Rx <- fit@Model@GLIST$theta
                  v.labels <- fit@Model@dimNames$lambda[[1]]
structure.diagram(fx=fx,Phi=Phi,Rx=Rx,labels=v.labels,...)}
else {structure.diagram(fx=fit@Model@GLIST$lambda,Phi=fit@Model@GLIST$beta,
                           Rx=fit@Model@GLIST$theta,...) }
```

This function is not ready for prime time because it does not yet draw sem (just cfa) diagrams.

lavaan diagam for the Thurstone (Bechtoldt.2) data set



Confirmatory factor structures across groups

- When comparing measures across age or across genders, it is important to make sure that the factor structures are in fact the same.
  - When measuring change, we want to make sure that our measure is the same for different ages.
  - When comparing ethnic groups, gender, genetic relationships, want to make sure that the measures are invariant across the groups
- This can be done by doing multiple group cfa.
- Possible to do in OpenMx and lavaan, but not in sem

Comparing Bechtoldt1 and Bechtoldt2

```
two.mod <- cfa(Thurstone.mod,sample.cov=list(b1,b2),
                    sample.nobs=list(212,213),std.lv=TRUE)
> summary(two.mod,fit.measures=TRUE)
Model converged normally after 26 iterations using ML

  Minimum Function Chi-square                       74.045
  Degrees of freedom                                    48
  P-value                                           0.0093

Chi-square for each group:

  Group 1                                           35.669
  Group 2                                           38.376

Chi-square test baseline model:

  Minimum Function Chi-square                     2205.154
  Degrees of freedom                                    63
  P-value                                           0.0000
```

Does not seem to work with lavaan beta– need to use the old

Loadings for two groups

```
Group 1 [Group 1]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  F1 =~
    Sentences         0.907    0.054   16.800    0.000
    Vocabulary        0.913    0.054   16.992    0.000
    Completion        0.840    0.056   14.890    0.000
  F2 =~
    First_Letters     0.829    0.064   12.939    0.000
    Four_letter_words     0.731    0.066   11.126    0.000
    Suffixes          0.650    0.067    9.668    0.000
  F3 =~
    Letter_Series     0.847    0.060   14.206    0.000
    Pedigrees         0.788    0.061   12.872    0.000
    Letter_Grouping     0.711    0.063   11.202    0.000

Latent covariances:
  F1 ~~
    F2                0.565    0.058    9.668    0.000
    F3                0.700    0.045   15.528    0.000
  F2 ~~
    F3                0.570    0.062    9.137    0.000

Latent variances:
    F1                1.000
    F2                1.000
    F3                1.000

Residual variances:
    Sentences         0.173    0.028    6.137    0.000
```

```
Group 2 [Group 2]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  F1 =~
    Sentences         0.903    0.054   16.727    0.000
    Vocabulary        0.912    0.054   17.005    0.000
    Completion        0.854    0.056   15.317    0.000
  F2 =~
    First_Letters     0.834    0.060   13.783    0.000
    Four_letter_words     0.795    0.061   12.937    0.000
    Suffixes          0.701    0.064   10.960    0.000
  F3 =~
    Letter_Series     0.779    0.064   12.173    0.000
    Pedigrees         0.718    0.065   10.998    0.000
    Letter_Grouping     0.702    0.066   10.679    0.000

Latent covariances:
  F1 ~~
    F2                0.643    0.050   12.755    0.000
    F3                0.670    0.051   13.153    0.000
  F2 ~~
    F3                0.637    0.058   10.900    0.000

Latent variances:
    F1                1.000
    F2                1.000
    F3                1.000

Residual variances:
```

Constrain the two groups to be equal

```
 two.mod <- cfa(Thurstone.mod,sample.cov=list(b1,b2),
                    sample.nobs=list(212,213),std.lv=TRUE,
                    group.constraints=c("loadings"))
summary(two.mod,fit.measures=TRUE)
```

```
Model converged normally after 25 iterations using ML

  Minimum Function Chi-square                   76.128
  Degrees of freedom                                57
  P-value                                       0.0461

Chi-square for each group:

  Group 1                                       36.700
  Group 2                                       39.428

Chi-square test baseline model:

  Minimum Function Chi-square                 2205.154
  Degrees of freedom                                63
  P-value                                       0.0000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.991
  Tucker-Lewis Index (TLI)                       0.990
```

Parameter estimates

```
Model estimates:

Group 1 [Group 1]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  F1 =~
    Sentences         0.903    0.038   23.705    0.000
    Vocabulary        0.911    0.038   24.046    0.000
    Completion        0.846    0.040   21.362    0.000
  F2 =~
    First_Letters     0.831    0.044   18.943    0.000
    Four_letter_words    0.767    0.045   17.120    0.000
    Suffixes          0.679    0.046   14.674    0.000
  F3 =~
    Letter_Series     0.816    0.044   18.752    0.000
    Pedigrees         0.756    0.045   16.976    0.000
    Letter_Grouping    0.705    0.046   15.465    0.000

Latent covariances:
  F1 ~~
    F2                0.565    0.056   10.044    0.000
    F3                0.697    0.044   15.746    0.000
  F2 ~~
    F3                0.569    0.061    9.304    0.000

Latent variances:
    F1                1.000
    F2                1.000
    F3                1.000
```

```
Group 2 [Group 2]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  F1 =~
    Sentences         0.903
    Vocabulary        0.911
    Completion        0.846
  F2 =~
    First_Letters     0.831
    Four_letter_words    0.767
    Suffixes          0.679
  F3 =~
    Letter_Series     0.816
    Pedigrees         0.756
    Letter_Grouping    0.705

Latent covariances:
  F1 ~~
    F2                0.641    0.049   13.108    0.000
    F3                0.672    0.048   13.939    0.000
  F2 ~~
    F3                0.633    0.057   11.145    0.000

Latent variances:
    F1                1.000
    F2                1.000
```

Compare goodness of fits

Because the models are in fact samples from the same data, they should agree.

```
Model converged normally after 26 iterations using ML

  Minimum Function Chi-square                   74.045
  Degrees of freedom                                48
  P-value                                       0.0093

Chi-square for each group:

  Group 1                                       35.669
  Group 2                                       38.376

Chi-square test baseline model:

  Minimum Function Chi-square                 2205.154
  Degrees of freedom                                63
  P-value                                       0.0000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.988
  Tucker-Lewis Index (TLI)                       0.984
```

```
Model converged normally after 25 iterations using ML

  Minimum Function Chi-square                   76.128
  Degrees of freedom                                57
  P-value                                       0.0461

Chi-square for each group:

  Group 1                                       36.700
  Group 2                                       39.428

Chi-square test baseline model:

  Minimum Function Chi-square                 2205.154
  Degrees of freedom                                63
  P-value                                       0.0000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.991
  Tucker-Lewis Index (TLI)                       0.990
```

Descriptive statistics of their data set

```
> describe(HolzingerSwineford1939)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
id	1	301	176.55	105.94	163.00	176.78	140.85	1.00	351.00	350.00	-0.01	-1.35	6.11
sex	2	301	1.51	0.50	2.00	1.52	0.00	1.00	2.00	1.00	-0.06	-2.01	0.03
ageyr	3	301	13.00	1.05	13.00	12.89	1.48	11.00	16.00	5.00	0.69	0.25	0.06
agemo	4	301	5.38	3.45	5.00	5.32	4.45	0.00	11.00	11.00	0.09	-1.21	0.20
school*	5	301	1.52	0.50	2.00	1.52	0.00	1.00	2.00	1.00	-0.07	-2.01	0.03
grade	6	300	7.48	0.50	7.00	7.47	0.00	7.00	8.00	1.00	0.09	-2.00	0.03
x1	7	301	4.94	1.17	5.00	4.96	1.24	0.67	8.50	7.83	-0.25	0.36	0.07
x2	8	301	6.09	1.18	6.00	6.02	1.11	2.25	9.25	7.00	0.47	0.38	0.07
x3	9	301	2.25	1.13	2.12	2.20	1.30	0.25	4.50	4.25	0.38	-0.89	0.07
x4	10	301	3.06	1.16	3.00	3.02	0.99	0.00	6.33	6.33	0.27	0.12	0.07
x5	11	301	4.34	1.29	4.50	4.40	1.48	1.00	7.00	6.00	-0.35	-0.53	0.07
x6	12	301	2.19	1.10	2.00	2.09	1.06	0.14	6.14	6.00	0.86	0.88	0.06
x7	13	301	4.19	1.09	4.09	4.16	1.10	1.30	7.43	6.13	0.25	-0.27	0.06
x8	14	301	5.53	1.01	5.50	5.49	0.96	3.05	10.00	6.95	0.53	1.24	0.06
x9	15	301	5.37	1.01	5.42	5.37	0.99	2.78	9.25	6.47	0.20	0.34	0.06

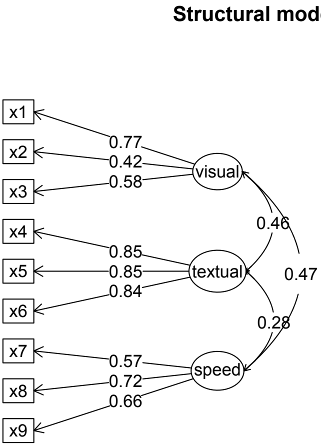
cfa syntax

Because we are using a covariance analysis, we need to standardize the observed variables to express the loadings as correlations.

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

 fit <- cfa(HS.model, data = HolzingerSwineford1939,std.lv=TRUE,std.ov=TRUE)
summary(fit)
lavaan.diagram(fit,cut=.2,digits=2)
```

Lavaan diagram of Holzinger-Swineford 1939 cfa



Now do multiple groups

```
fit.2g  <- cfa(HS.model, data=HolzingerSwineford1939, group="school",
                         std.lv=TRUE,std.ov=TRUE)
summary(fit.2g)
```

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

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

Chi-square for each group:

  Pasteur                                       64.309
  Grant-White                                   51.542

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard
```

Compare the parameters for both schools

```
Group 1 [Pasteur]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                0.884    0.111    7.934    0.000
    x2                0.335    0.089    3.753    0.000
    x3                0.513    0.093    5.525    0.000
  textual =~
    x4                0.821    0.069   11.927    0.000
    x5                0.854    0.068   12.604    0.000
    x6                0.836    0.068   12.230    0.000
  speed =~
    x7                0.545    0.098    5.557    0.000
    x8                0.679    0.104    6.531    0.000
    x9                0.550    0.098    5.596    0.000

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

Variances:
    x1                0.212    0.165    1.286    0.198
    x2                0.881    0.104    8.464    0.000
    x3                0.731    0.100    7.271    0.000
    x4                0.320    0.052    6.138    0.000
    x5                0.265    0.050    5.292    0.000
```

```
Group 2 [Grant-White]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                0.674    0.090    7.525    0.000
    x2                0.515    0.091    5.642    0.000
    x3                0.691    0.090    7.711    0.000
  textual =~
    x4                0.863    0.070   12.355    0.000
    x5                0.826    0.071   11.630    0.000
    x6                0.823    0.071   11.572    0.000
  speed =~
    x7                0.657    0.084    7.819    0.000
    x8                0.793    0.083    9.568    0.000
    x9                0.698    0.084    8.357    0.000

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

Variances:
    x1                0.538    0.095    5.675    0.000
    x2                0.728    0.099    7.339    0.000
    x3                0.515    0.095    5.409    0.000
    x4                0.249    0.051    4.870    0.000
    x5                0.310    0.053    5.812    0.000
```

Constrain the two schools to have equal loadings

(This works on lavaan 0.3.3 but not the beta version 0.4-5)

```
fit.2g  <- cfa(HS.model, data=HolzingerSwineford1939, group="school",
                        std.lv=TRUE,std.ov=TRUE,group.constraints=c("loadings"))
 summary(fit.2g)
```

```
Model converged normally after 27 iterations using ML

  Minimum Function Chi-square                  122.862
  Degrees of freedom                                57
  P-value                                       0.0000

Chi-square for each group:

  Grant-White                                   54.264
  Pasteur                                       68.598
```

Show more fit statistics

```
> summary(fit.2g,fit.measures=TRUE)

Model converged normally after 27 iterations using ML

  Minimum Function Chi-square                  122.862
  Degrees of freedom                                57
  P-value                                       0.0000

Chi-square for each group:

  Grant-White                                   54.264
  Pasteur                                       68.598

Chi-square test baseline model:

  Minimum Function Chi-square                  957.769
  Degrees of freedom                                63
  P-value                                       0.0000
```

```
Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.926
  Tucker-Lewis Index (TLI)                       0.919

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -3417.421
  Loglikelihood unrestricted model (H1)      -3355.990

  Akaike (AIC)                                6900.841
  Bayesian (BIC)                              7023.176

Root Mean Square Error of Approximation:

  RMSEA                                          0.088
  90 Percent Confidence Interval          0.066  0.109

Standardized Root Mean Square Residual:

  SRMR                                           0.084
```

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

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

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