Psychology 454: Latent Variable Modeling
Week 6: Regression + Reliability = SEM

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

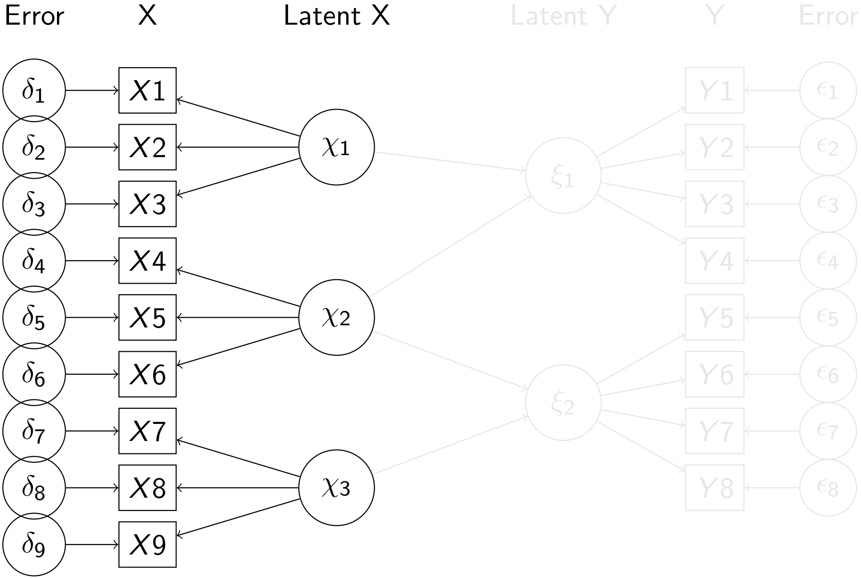
October, 2015

Part I

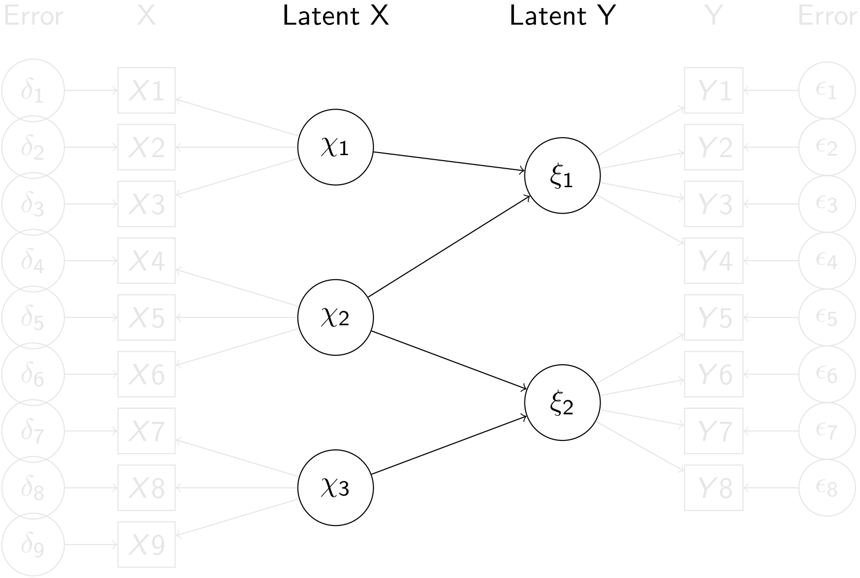
SEM=reliability + regression

Outline

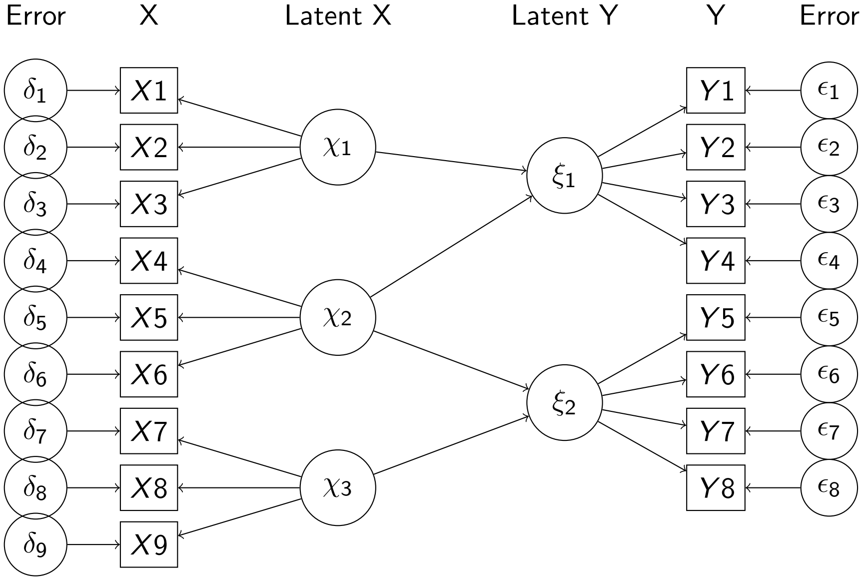
Measurement: A latent variable approach.



Theory


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

Psychometric Theory: A conceptual Syllabus



Two types of variables, three types of relationships

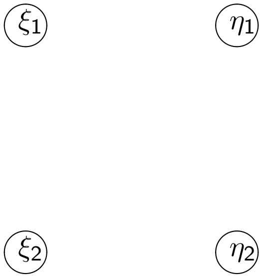
1. Variables
   1.1 Observed Variables (X, Y)
   1.2 Latent Variables (ξ η ϵ ζ)
2. Three kinds of variance/covariances
   2.1 Observed with Observed C_{xy} or σ_{xy}
   2.2 Observed with Latent λ
   2.3 Latent with Latent ϕ
3. Direction
   - Bidirectional (correlation)
   - Directional (regression)

Observed Variables

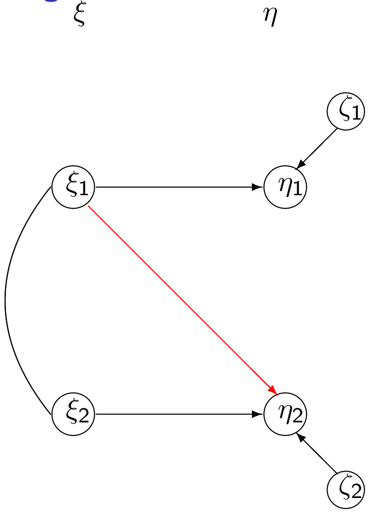
X	Y
$\boxed{X_1}$	$\boxed{Y_1}$
$\boxed{X_2}$	$\boxed{Y_2}$
$\boxed{X_3}$	$\boxed{Y_3}$
$\boxed{X_4}$	$\boxed{Y_4}$
$\boxed{X_5}$	$\boxed{Y_5}$
$\boxed{X_6}$	$\boxed{Y_6}$

Latent Variables

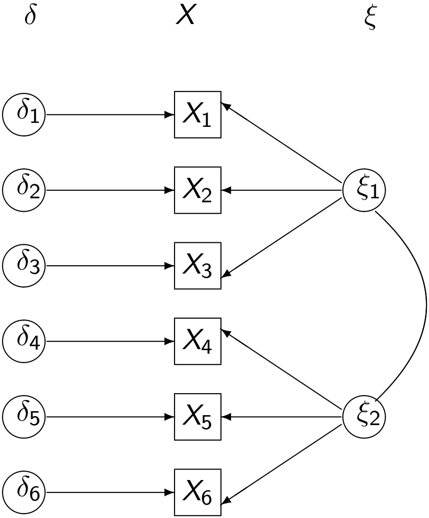
ξ η



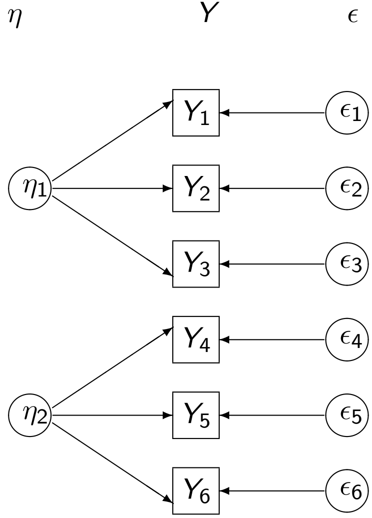
Theory: A regression model of latent variables



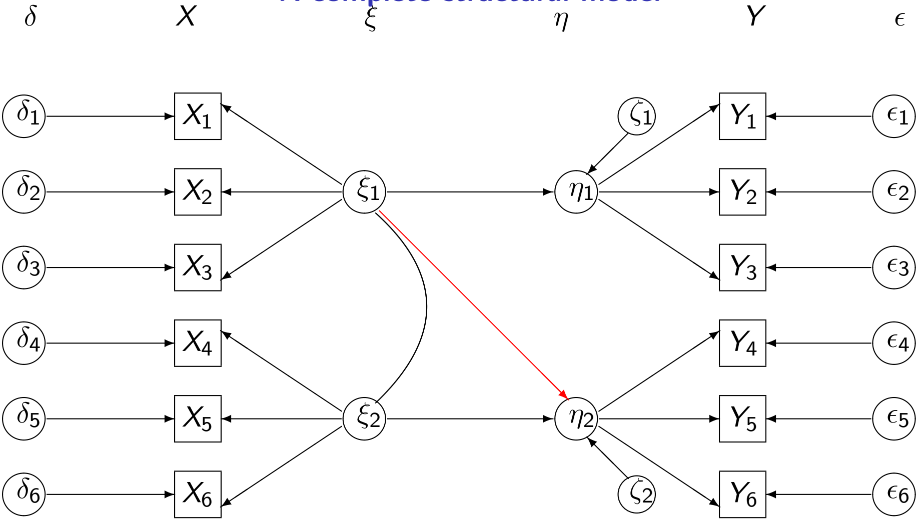
A measurement model for X



A measurement model for Y



A complete structural model



Latent Variable Modeling

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

Observed-observed

1. The Kerckhoff (1974) data set is used in the LISREL manual as an example of regression path models
   - "A sample of boys originally studied as ninth graders in 1969 was recontacted in 1974 to obtain information about high school performance and educational attainment."
   - 767 twelfth grade males
   - A study of background, aspiration and educational attainment
2. Variables
   - Intelligence
   - Number of siblings
   - Fathers Education
   - Fathers's occupation
   - Grades
   - Educational expectation
   - Occupational aspiration
3. Correlation matrix is available in the sem (Fox, Nie & Byrnes, 2013) package.

Getting the Kerchoff Correlation matrix

The data matrix is in the *sem* package as a moments matrix. We read this in, and then convert the lower diagonal matrix to a square matrix using the `lowerUpper` function.

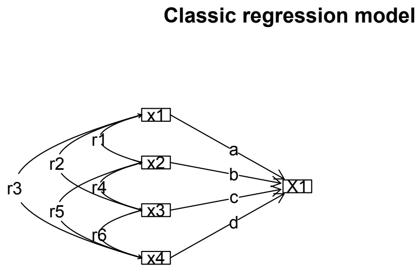
R code

```
library(sem)
R.kerch <- readMoments(diag=FALSE,
        names=c('Intelligence','Siblings',
       'FatherEd','FatherOcc','Grades','EducExp','OccupAsp'))
    -.100
     .277  -.152
     .250  -.108  .611
     .572  -.105  .294   .248
     .489  -.213  .446   .410   .597
     .335  -.153  .303   .331   .478   .651

 R.kerch <- lowerUpper(R.kerch)$lower
```

The Kerckhoff correlation matrix (N=767)

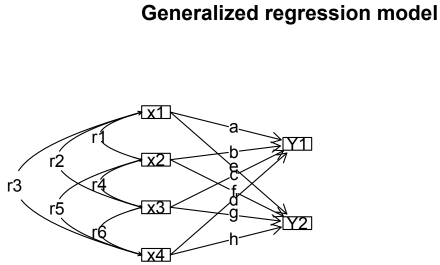
```
lowerMat(R.kerch)
             Intll Sblng FthrE FthrO Grads EdcEx Oc
Intelligence  1.00
Siblings     -0.10  1.00
FatherEd      0.28 -0.15  1.00
FatherOcc     0.25 -0.11  0.61  1.00
Grades        0.57 -0.10  0.29  0.25  1.00
EducExp       0.49 -0.21  0.45  0.41  0.60  1.00
OccupAsp      0.34 -0.15  0.30  0.33  0.48  0.65  1
```

The classic regression model



$$\hat{Y} = \beta_x x + \epsilon$$

The generalized regression model

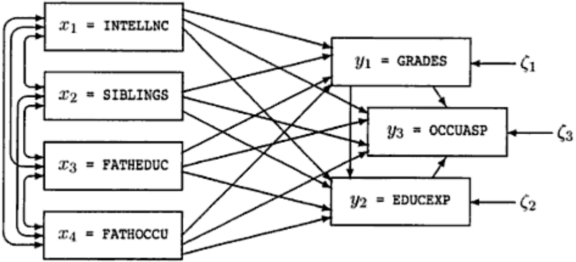


$$\hat{Y} = \beta_x x + \epsilon$$

Conceptual Kerckhoff model

1. Background variables
   - Intelligence
   - Number of siblings
   - Fathers Education
   - Fathers's occupation
2. Intermediate variables
   - Grades
   - Educational expectation
3. Final outcomes
   - Occupational aspiration

The model (From the LISREL manual)



Matrix regression

1. Most regression examples (and functions) use raw data
   - $\hat{Y} = X\beta + \epsilon$
   - $beta = (X'X)^{-1}X'Y$
   - lm(y ~x)
2. Regression is just solving the matrix equation
   - $\beta = R^{-1}r_{xy}$
   - mat.regress(R,x,y) (deprecated)
   - set.cor(y,x,R) (recommended)

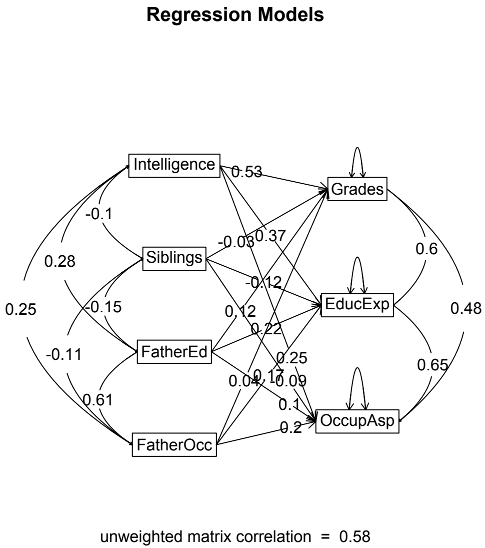
Simple regression 4 predictors, 3 criteria

R code

```
setCor(y=5:7,x=1:4,data=R.kerch, n.obs=737)
```

```
Call: setCor(y = 5:7, x = 1:4, data = R.kerch, n.obs = 737)

Multiple Regression from matrix input

Beta weights
             Grades EducExp OccupAsp
Intelligence   0.53    0.37     0.25
Siblings      -0.03   -0.12    -0.09
FatherEd       0.12    0.22     0.10
FatherOcc      0.04    0.17     0.20

Multiple R
  Grades  EducExp OccupAsp
    0.59     0.61     0.44
multiple R2
  Grades  EducExp OccupAsp
    0.35     0.38     0.19

Unweighted multiple R
  Grades  EducExp OccupAsp
    0.46     0.59     0.42
```

The simple Kerchoff regressions



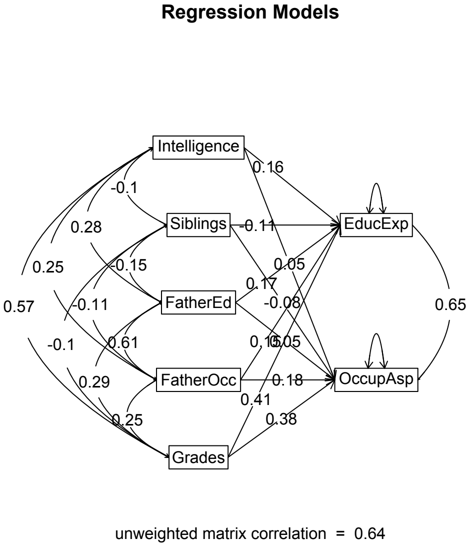
More complicated regression

R code

```
setCor(y=6:7,x=1:5,data=R.kerch)
```

```
Call: setCor(y = 6:7, x = 1:5, data = R.kerch)

Multiple Regression from matrix input

Beta weights
             EducExp OccupAsp
Intelligence    0.16     0.05
Siblings       -0.11    -0.08
FatherEd        0.17     0.05
FatherOcc       0.15     0.18
Grades          0.41     0.38

Multiple R
 EducExp OccupAsp
    0.70     0.54
multiple R2
 EducExp OccupAsp
    0.48     0.29

 Unweighted multiple R
 EducExp OccupAsp
    0.67     0.50
```

The simple Kerchoff regressions



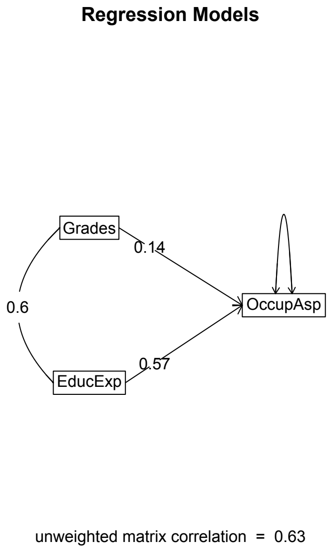
Predicting occupational aspirations from the intermediate set

R code

```
setCor(y=7,x=5:6,data=R.kerch)
```

```
Call: setCor(y = 7, x = 5:6, data = R.kerch)

Multiple Regression from matrix input

Beta weights
        OccupAsp
Grades      0.14
EducExp     0.57

Multiple R
         OccupAsp
OccupAsp     0.66
multiple R2
         OccupAsp
OccupAsp     0.44

 Unweighted multiple R
OccupAsp
    0.63
```

Predict Occupational Aspirations from the mediating variables



Predicting occupational aspirations from the entire set

R code

```
setCor(y=7,x=1:6,data=R.kerch)
```

```
Call: setCor(y = 7, x = 1:6, data = R.kerch)

Multiple Regression from matrix input

Beta weights
             OccupAsp
Intelligence    -0.04
Siblings        -0.02
FatherEd        -0.04
FatherOcc        0.10
Grades           0.16
EducExp          0.55

Multiple R
         OccupAsp
OccupAsp     0.67
multiple R2
         OccupAsp
OccupAsp     0.44

 Unweighted multiple R
OccupAsp
    0.57
 Unweighted multiple R2
OccupAsp
    0.32
```

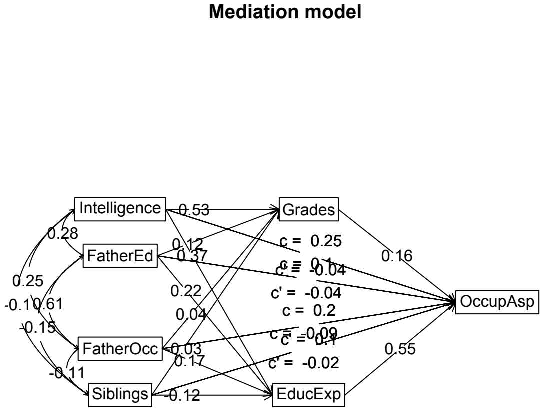
Try a mediation model

R code

```
mediate("OccupAsp",x=1:4,m=5:6,data=R.kerch,n.obs=737)
```

```
 'a'  paths
             Grades EducExp
Intelligence   0.53    0.37
Siblings      -0.03   -0.12
FatherEd       0.12    0.22
FatherOcc      0.04    0.17

'b' paths
        OccupAsp
Grades      0.16
EducExp     0.55

'ab' paths
             Grades EducExp
Intelligence   0.08    0.06
Siblings      -0.02   -0.07
FatherEd       0.02    0.03
FatherOcc      0.02    0.09
```

A 1 stage mediation model



Can use sem functions (in either sem or lavaan) to estimate the mediation model

1. Treat all variables as observed (fixed)
   - Specify a limited number of paths rather than the full regression model
2. *sem* commands
   - either in RAM (path) notation or
   - causal notation
3. *lavaan* commands similar to causal notation of sem

Kerckhoff-Kenny path analysis (modified from sem help page) to just predict the DVs)

```
model.kerch <- specifyModel()
    Intelligence -> Grades,        gam51
    Siblings -> Grades,            gam52
    FatherEd -> Grades,            gam53
    FatherOcc -> Grades,           gam54
    Intelligence -> EducExp,       gam61
    Siblings -> EducExp,           gam62
    FatherEd -> EducExp,           gam63
    FatherOcc -> EducExp,          gam64
     Intelligence -> OccupAsp,      gam71
    Siblings -> OccupAsp,          gam72
    FatherEd -> OccupAsp,          gam73
    FatherOcc -> OccupAsp,         gam74
  # Grades -> EducExp,             beta65
   # Grades -> OccupAsp,            beta75
 #   EducExp -> OccupAsp,           beta76

sem.kerch <- sem(model.kerch, R.kerch, 737, fixed.x=c('Intelligence','
    'FatherEd','FatherOcc'))
summary(sem.kerch)
```

Fixed path model – not modeling the DV correlations

```
Model Chisquare =  411.72   Df =  3 Pr(>Chisq) = 6.4133e-89
Chisquare (null model) =  1664.3   Df =  21
Goodness-of-fit index =  0.85747
Adjusted goodness-of-fit index =  -0.33031
RMSEA index =  0.43024   90% CI: (0.39572, 0.46581)
Bentler-Bonnett NFI =  0.75262
Tucker-Lewis NNFI =  -0.74103
Bentler CFI =  0.75128
SRMR =  0.099759
AIC =  441.72
AICc =  412.38
BIC =  510.76
CAIC =  388.91

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000   0.000   0.000   0.956   0.000  10.100

R-square for Endogenous Variables
 Grades   EducExp OccupAsp
 0.3490    0.3765   0.1930
```

With path coefficients of

```
Parameter Estimates
            Estimate  Std Error z value  Pr(>|z|)
gam51        0.525902 0.031182  16.86530 8.0987e-64 Grades <--- Intelligence
gam52       -0.029942 0.030149  -0.99314 3.2064e-01 Grades <--- Siblings
gam53        0.118966 0.038259   3.10951 1.8740e-03 Grades <--- FatherEd
gam54        0.040603 0.037785   1.07456 2.8257e-01 Grades <--- FatherOcc
gam61        0.373339 0.030517  12.23376 2.0521e-34 EducExp <--- Intelligence
gam62       -0.123910 0.029506  -4.19954 2.6745e-05 EducExp <--- Siblings
gam63        0.220918 0.037442   5.90022 3.6302e-09 EducExp <--- FatherEd
gam64        0.168302 0.036979   4.55125 5.3328e-06 EducExp <--- FatherOcc
gam71        0.248827 0.034718   7.16718 7.6561e-13 OccupAsp <--- Intelligence
gam72       -0.091653 0.033567  -2.73047 6.3245e-03 OccupAsp <--- Siblings
gam73        0.098869 0.042596   2.32109 2.0282e-02 OccupAsp <--- FatherEd
gam74        0.198486 0.042069   4.71809 2.3807e-06 OccupAsp <--- FatherOcc
V[Grades]    0.650995 0.033935  19.18333 5.1010e-82 Grades <--> Grades
V[EducExp]   0.623511 0.032503  19.18333 5.1010e-82 EducExp <--> EducExp
V[OccupAsp]  0.806964 0.042066  19.18333 5.1010e-82 OccupAsp <--> OccupAsp
```

Note, we are not modeling the DV correlations so the residuals will be large

Residuals from this model

```
> lowerMat(resid(sem.kerch))
             Intll Sblng FthrE FthrO Grads EdcEx OccpA
Intelligence 0.00
Siblings     0.00  0.00
FatherEd     0.00  0.00  0.00
FatherOcc    0.00  0.00  0.00  0.00
Grades       0.00  0.00  0.00  0.00  0.00
EducExp      0.00  0.00  0.00  0.00  0.26  0.00
OccupAsp     0.00  0.00  0.00  0.00  0.25  0.38  0.00
```

fixed sem = regression

Compare the coefficients from this sem with the regression β values

```
round(sem.kerch$coeff,3)
     gam51       gam52       gam53       gam54       gam61       gam62       gam63
      gam64       gam71       gam72       gam73     gam74
     0.526      -0.030       0.119       0.041       0.373      -0.124       0.221
     0.168       0.249      -0.092       0.099       0.198
    V[Grades]  V[EducExp] V[OccupAsp]
   0.651       0.624       0.807

 mr.kk <- mat.regress(data=R.kerch,x=c(1:4),y=c(5:7))

 > round(as.vector(mr.kk$beta),3)
[1]  0.526 -0.030  0.119  0.041  0.373 -0.124  0.221
 0.168  0.249 -0.092  0.099  0.198
```

More complicated regression

1. Able to model the intercorrelations of the Y variables
   - This treats the Ys as both predictors and predicted
2. Able to let some of the Ys be part of the regression model

Complete Kerckhoff-Kenny path analysis (taken from sem help page)

```
model.kerch1 <- specifyModel()
    Intelligence -> Grades,        gam51
    Siblings -> Grades,            gam52
    FatherEd -> Grades,            gam53
    FatherOcc -> Grades,           gam54
    Intelligence -> EducExp,       gam61
    Siblings -> EducExp,           gam62
    FatherEd -> EducExp,           gam63
    FatherOcc -> EducExp,          gam64
    Grades -> EducExp,             beta65
    Intelligence -> OccupAsp,      gam71
    Siblings -> OccupAsp,          gam72
    FatherEd -> OccupAsp,          gam73
    FatherOcc -> OccupAsp,         gam74
    Grades -> OccupAsp,            beta75
    EducExp -> OccupAsp,           beta76

sem.kerch1 <- sem(model.kerch1, R.kerch, 737, fixed.x=c('Intelligence'
    'FatherEd','FatherOcc'))
summary(sem.kerch1)
```

sem output

```
 Model Chisquare =  3.2685e-13   Df =  0 Pr(>Chisq) = NA
 Chisquare (null model) =  1664.3   Df =  21
 Goodness-of-fit index =  1
 AIC =  36
 AICc =  0.95265
 BIC =  118.85
 CAIC =  3.2685e-13
 Normalized Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-4.26e-15 -1.35e-15  0.00e+00 -4.17e-16  0.00e+00  1.49e-15
 Parameter Estimates
           Estimate  Std Error z value  Pr(>|z|)
gam51       0.525902 0.031182  16.86530 8.0987e-64 Grades <--- Intelligence
gam52      -0.029942 0.030149  -0.99314 3.2064e-01 Grades <--- Siblings
gam53       0.118966 0.038259   3.10951 1.8740e-03 Grades <--- FatherEd
gam54       0.040603 0.037785   1.07456 2.8257e-01 Grades <--- FatherOcc
gam61       0.160270 0.032710   4.89979 9.5940e-07 EducExp <--- Intelligence
gam62      -0.111779 0.026876  -4.15899 3.1966e-05 EducExp <--- Siblings
gam63       0.172719 0.034306   5.03461 4.7882e-07 EducExp <--- FatherEd
gam64       0.151852 0.033688   4.50758 6.5571e-06 EducExp <--- FatherOcc
beta65      0.405150 0.032838  12.33799 5.6552e-35 EducExp <--- Grades
gam71      -0.039405 0.034500  -1.14215 2.5339e-01 OccupAsp <--- Intelligence
gam72      -0.018825 0.028222  -0.66700 5.0477e-01 OccupAsp <--- Siblings
gam73      -0.041333 0.036216  -1.14126 2.5376e-01 OccupAsp <--- FatherEd
gam74       0.099577 0.035446   2.80924 4.9658e-03 OccupAsp <--- FatherOcc
beta75      0.157912 0.037443   4.21738 2.4716e-05 OccupAsp <--- Grades
beta76      0.549593 0.038260  14.36486 8.5976e-47 OccupAsp <--- EducExp
V[Grades]   0.650995 0.033935  19.18333 5.1010e-82 Grades <--> Grades
V[EducExp]  0.516652 0.026932  19.18333 5.1010e-82 EducExp <--> EducExp
V[OccupAsp] 0.556617 0.029016  19.18333 5.1010e-82 OccupAsp <--> OccupAsp
```

Note that these models give different path coefficients

```
> round(sem.kerch$coeff,3)
     gam51       gam52       gam53       gam54       gam61       gam62       gam63       gam6
      gam71       gam72       gam73
     0.526      -0.030       0.119       0.041       0.373      -0.124       0.221       0.1
     0.249      -0.092       0.099
     gam74   V[Grades]  V[EducExp] V[OccupAsp]
     0.198       0.651       0.624       0.807
> round(sem.kerch1$coeff,3)
     gam51       gam52       gam53       gam54       gam61       gam62       gam63       gam6
     beta65       gam71       gam72
     0.526      -0.030       0.119       0.041       0.160      -0.112       0.173       0.1
     0.405      -0.039      -0.019
     gam73       gam74      beta75      beta76   V[Grades]  V[EducExp] V[OccupAsp]
    -0.041       0.100       0.158       0.550       0.651       0.517       0.557

> round(mr.kk$beta,2)
             Grades EducExp OccupAsp
Intelligence   0.53    0.37     0.25
Siblings      -0.03   -0.12    -0.09
FatherEd       0.12    0.22     0.10
FatherOcc      0.04    0.17     0.20
```

A latent variable structural model

1. Taken from the LISREL User's reference guide
2. Data from, Caslyn and Kenny (1977)
   - Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement
3. Variables
   - self concept
   - parental evaluation
   - teacher evaluation
   - friend evaluation
   - educational aspiration
   - college plans

Caslyn and Kenny data

	self	parent	teacher	friend	edu_asp	college
self_concept	1.00	0.73	0.70	0.58	0.46	0.56
parental_eval	0.73	1.00	0.68	0.61	0.43	0.52
teacher_eval	0.70	0.68	1.00	0.57	0.40	0.48
friend_eval	0.58	0.61	0.57	1.00	0.37	0.41
edu_aspir	0.46	0.43	0.40	0.37	1.00	0.72
college_plans	0.56	0.52	0.48	0.41	0.72	1.00

Creating the model using `structure.diagam`

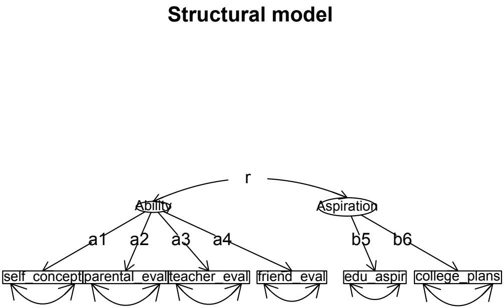
```
fx <- structure.list(6,list(c(1:4),c(5:6)),item.labels = rownames(abil
                     f.labels=c("Ability","Aspiration"))
mod.edu <- structure.diagram(fx,"r",title="Lisrel example 3.2",
                        errors=TRUE,lr=FALSE,cex=.8)
fx
```

```
fx
              Ability Aspiration
self_concept  "a1"    "0"
parental_eval "a2"    "0"
teacher_eval  "a3"    "0"
friend_eval   "a4"    "0"
edu_aspir     "0"     "b5"
college_plans "0"     "b6"
```

The sem commands are in the mod.edu object

```
mod.edu
      Path                                  Parameter Value
 [1,] "Ability->self_concept"               "a1"      NA
 [2,] "Ability->parental_eval"              "a2"      NA
 [3,] "Ability->teacher_eval"               "a3"      NA
 [4,] "Ability->friend_eval"                "a4"      NA
 [5,] "Aspiration->edu_aspir"               "b5"      NA
 [6,] "Aspiration->college_plans"           "b6"      NA
 [7,] "self_concept<->self_concept"         "x1e"     NA
 [8,] "parental_eval<->parental_eval"       "x2e"     NA
 [9,] "teacher_eval<->teacher_eval"         "x3e"     NA
[10,] "friend_eval<->friend_eval"           "x4e"     NA
[11,] "edu_aspir<->edu_aspir"               "x5e"     NA
[12,] "college_plans<->college_plans"       "x6e"     NA
[13,] "Aspiration<->Ability"                "rF2F1"   NA
[14,] "Ability<->Ability"                   NA        "1"
[15,] "Aspiration<->Aspiration"             NA        "1"
```

A model of the Caslyn-Kenny (1997) data set



```
ability <- as.matrix(ability) #sem requires matrix input
sem.edu <- sem(mod=mod.edu,S=ability,N=556)
summary(sem.edu)
```

```
Model Chisquare =  9.2557   Df =  8 Pr(>Chisq) = 0.32118
 Chisquare (null model) =  1832   Df =  15
 Goodness-of-fit index =  0.99443
 Adjusted goodness-of-fit index =  0.98537
 RMSEA index =  0.016817   90% CI: (NA, 0.054321)
 Bentler-Bonnett NFI =  0.99495
 Tucker-Lewis NNFI =  0.9987
 Bentler CFI =  0.99931
 SRMR =  0.012011
 AIC =  35.256
 AICc =  9.9273
 BIC =  91.426
 CAIC =  -49.31

 Normalized Residuals
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4410 -0.1870  0.0000 -0.0131  0.2110  0.5330

 R-square for Endogenous Variables
 self_concept parental_eval  teacher_eval   friend_eval     edu_aspir college_plans
       0.7451        0.7213        0.6482        0.4834        0.6008        0.8629
```

With parameter estimates

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1    0.86320  0.035145  24.5612 3.2846e-133 self_concept <--- Ability
a2    0.84932  0.035450  23.9582 7.5937e-127 parental_eval <--- Abilit
a3    0.80509  0.036405  22.1149 2.2725e-108 teacher_eval <--- Ability
a4    0.69527  0.038634  17.9964  2.0795e-72 friend_eval <--- Ability
b5    0.77508  0.040357  19.2058  3.3077e-82 edu_aspir <--- Aspiration
b6    0.92893  0.039410  23.5712 7.6153e-123 college_plans <--- Aspira
x1e   0.25488  0.023367  10.9075  1.0617e-27 self_concept <--> self_co
x2e   0.27865  0.024128  11.5491  7.4600e-31 parental_eval <--> parent
x3e   0.35184  0.026919  13.0703  4.8660e-39 teacher_eval <--> teacher
x4e   0.51660  0.034725  14.8768  4.6594e-50 friend_eval <--> friend_e
x5e   0.39924  0.038196  10.4525  1.4266e-25 edu_aspir <--> edu_aspir
x6e   0.13709  0.043505   3.1511  1.6264e-03 college_plans <--> colleg
rF2F1 0.66637  0.030954  21.5276 8.5783e-103 Ability <--> Aspiration
```

Three competing models

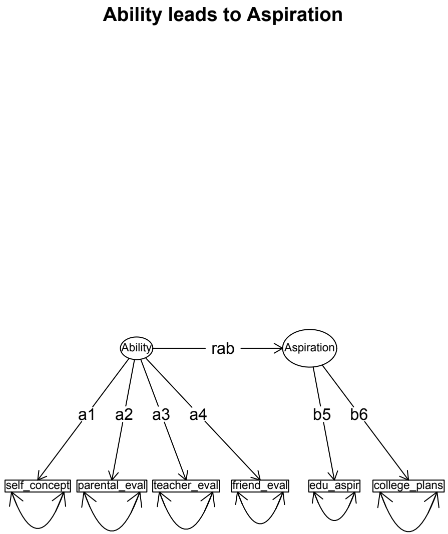
1. Ability and aspirations are correlated
    - r = .66
2. Ability causes aspirations
    - beta = .89
3. Aspirations cause ability
    - beta = .89

Create the model where ability lead to aspirations

```
phi <- phi.list(2,c(2))
phi
mod.edu <- structure.diagram(fx,phi,title="Aspiration leads to ability
                        errors=TRUE,lr=FALSE,cex=.7)
```

```
mod.edu
      Path                                Parameter Value
 [1,] "Ability->self_concept"             "a1"      NA
 [2,] "Ability->parental_eval"            "a2"      NA
 [3,] "Ability->teacher_eval"             "a3"      NA
 [4,] "Ability->friend_eval"              "a4"      NA
 [5,] "Aspiration->edu_aspir"             "b5"      NA
 [6,] "Aspiration->college_plans"         "b6"      NA
 [7,] "self_concept<->self_concept"       "x1e"     NA
 [8,] "parental_eval<->parental_eval"     "x2e"     NA
 [9,] "teacher_eval<->teacher_eval"       "x3e"     NA
[10,] "friend_eval<->friend_eval"         "x4e"     NA
[11,] "edu_aspir<->edu_aspir"             "x5e"     NA
[12,] "college_plans<->college_plans"     "x6e"     NA
[13,] "Ability ->Aspiration"              "rF1F2"   NA
[14,] "Ability<->Ability"                 NA        "1"
[15,] "Aspiration<->Aspiration"           NA        "1"
```

Ability causes aspiration



Fit statistics are identical

```
> sem.edu <- sem(mod=mod.edu,S=ability,N=556)
> summary(sem.edu)

summary(sem.edu)

 Model Chisquare =  9.2557   Df =  8 Pr(>Chisq) = 0.32118
 Chisquare (null model) =  1832   Df =  15
 Goodness-of-fit index =  0.99443
 Adjusted goodness-of-fit index =  0.98537
 RMSEA index =  0.016817   90% CI: (NA, 0.054321)
 Bentler-Bonnett NFI =  0.99495
 Tucker-Lewis NNFI =  0.9987
 Bentler CFI =  0.99931
 SRMR =  0.012011
 AIC =  35.256
 AICc =  9.9273
 BIC =  91.426
 CAIC =  -49.31

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4410 -0.1870  0.0000 -0.0131  0.2110  0.5330

 R-square for Endogenous Variables
 self_concept parental_eval  teacher_eval   friend_eval    Aspiration     edu_aspir college_pl
       0.7451        0.7213        0.6482        0.4834        0.4440        0.6008         0.
```

But the paths are different

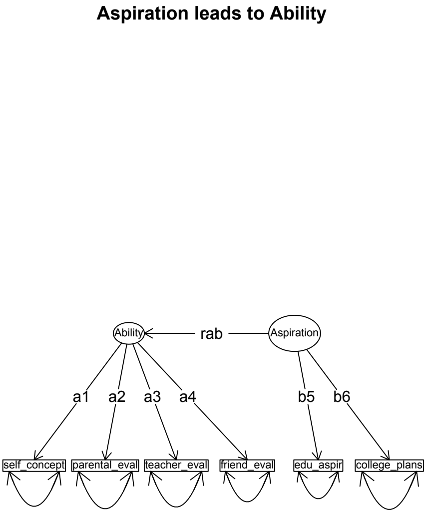
```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1     0.86320  0.035145  24.5612 3.2848e-133 self_concept <--- Ability
a2     0.84932  0.035450  23.9582 7.5919e-127 parental_eval <--- Abilit
a3     0.80509  0.036405  22.1149 2.2732e-108 teacher_eval <--- Ability
a4     0.69527  0.038634  17.9964  2.0794e-72 friend_eval <--- Ability
b5     0.57792  0.030630  18.8678  2.0977e-79 edu_aspir <--- Aspiration
b6     0.69263  0.037979  18.2370  2.6257e-74 college_plans <--- Aspira
x1e    0.25488  0.023367  10.9075  1.0617e-27 self_concept <--> self_co
x2e    0.27865  0.024128  11.5491  7.4610e-31 parental_eval <--> parent
x3e    0.35184  0.026919  13.0703  4.8654e-39 teacher_eval <--> teacher
x4e    0.51660  0.034725  14.8768  4.6595e-50 friend_eval <--> friend_e
x5e    0.39924  0.038196  10.4525  1.4266e-25 edu_aspir <--> edu_aspir
x6e    0.13709  0.043505   3.1511  1.6264e-03 college_plans <--> colleg
rF1F2 0.89371  0.074673  11.9683  5.2068e-33 Aspiration <--- Ability

 Iterations =  30
```

Let aspirations cause ability

```
> phi[1,2] <- phi[2,1]
> phi[2,1] <- "0"
> mod.edu <- structure.diagram(fx,phi,main="Aspiration leads to Abilit
>mod.edu
```

```
> mod.edu
      Path                                 Parameter Value
 [1,] "Ability->self_concept"              "a1"      NA
 [2,] "Ability->parental_eval"             "a2"      NA
 [3,] "Ability->teacher_eval"              "a3"      NA
 [4,] "Ability->friend_eval"               "a4"      NA
 [5,] "Aspiration->edu_aspir"              "b5"      NA
 [6,] "Aspiration->college_plans"          "b6"      NA
 [7,] "self_concept<->self_concept"        "x1e"     NA
 [8,] "parental_eval<->parental_eval"      "x2e"     NA
 [9,] "teacher_eval<->teacher_eval"        "x3e"     NA
[10,] "friend_eval<->friend_eval"          "x4e"     NA
[11,] "edu_aspir<->edu_aspir"              "x5e"     NA
[12,] "college_plans<->college_plans"      "x6e"     NA
[13,] "Aspiration<-Ability"                "rF2F1"   NA
[14,] "Ability<->Ability"                  NA        "1"
[15,] "Aspiration<->Aspiration"            NA        "1"
```

Aspiration leads to ability



Fits are still identical

```
> sem.edu <- sem(mod=mod.edu,S=ability,N=556)
> summary(sem.edu)
 Model Chisquare =  9.2557   Df =  8 Pr(>Chisq) = 0.32118
 Chisquare (null model) =  1832   Df =  15
 Goodness-of-fit index =  0.99443
 Adjusted goodness-of-fit index =  0.98537
 RMSEA index =  0.016817   90% CI: (NA, 0.054321)
 Bentler-Bonnett NFI =  0.99495
 Tucker-Lewis NNFI =  0.9987
 Bentler CFI =  0.99931
 SRMR =  0.012011
 AIC =  35.256
 AICc =  9.9273
 BIC =  91.426
 CAIC =  -49.31

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4410 -0.1870  0.0000 -0.0131  0.2110  0.5330

 R-square for Endogenous Variables
 self_concept parental_eval  teacher_eval   friend_eval    Aspiration
       0.7451        0.7213        0.6482        0.4834        0.4440
```

And the crucial coefficient is different

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1     0.86320  0.035145  24.5612 3.2848e-133 self_concept <--- Ability
a2     0.84932  0.035450  23.9582 7.5919e-127 parental_eval <--- Abilit
a3     0.80509  0.036405  22.1149 2.2732e-108 teacher_eval <--- Ability
a4     0.69527  0.038634  17.9964  2.0794e-72 friend_eval <--- Ability
b5     0.57792  0.030630  18.8678  2.0977e-79 edu_aspir <--- Aspiration
b6     0.69263  0.037979  18.2370  2.6257e-74 college_plans <--- Aspira
x1e    0.25488  0.023367  10.9075  1.0617e-27 self_concept <--> self_co
x2e    0.27865  0.024128  11.5491  7.4610e-31 parental_eval <--> parent
x3e    0.35184  0.026919  13.0703  4.8654e-39 teacher_eval <--> teacher
x4e    0.51660  0.034725  14.8768  4.6595e-50 friend_eval <--> friend_e
x5e    0.39924  0.038196  10.4525  1.4266e-25 edu_aspir <--> edu_aspir
x6e    0.13709  0.043505   3.1511  1.6264e-03 college_plans <--> colleg
rF2F1  0.89371  0.074673  11.9683  5.2068e-33 Aspiration <--- Ability
```

Compare the three models

	correlated	asp	abil
a1	0.86	0.64	0.86
a2	0.85	0.63	0.85
a3	0.81	0.60	0.81
a4	0.70	0.52	0.70
b5	0.78	0.78	0.58
b6	0.93	0.93	0.69
x1e	0.25	0.25	0.25
x2e	0.28	0.28	0.28
x3e	0.35	0.35	0.35
x4e	0.52	0.52	0.52
x5e	0.40	0.40	0.40
x6e	0.14	0.14	0.14
rF2F1	0.67	0.89	0.89

1. Although fits were identical
2. Paths differ as a function of presumed influence
3. Which solution is correct?
4. Is this even possible to answer?

Implications of arrows

1. Need to fit alternative models
   - Create alternative plausible models
   - Create alternative implausible models (they will fit also).
2. Need to consider alternative representations
   - Try reversing arrows
3. Are there external variables (e.g., time) that allow one to choose between models?
4. Confirmation that a model fits does not confirm theoretical adequacy of the model.

Two fundamentally different types of observed variables

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

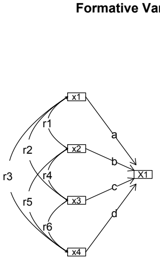
Formative indicators Bollen (2002)

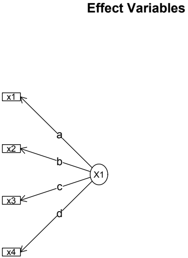
1. The correlational structure of formative indicators is independent of the loadings on a factor.
    - They are not locally independent
2. Examples of formative indicators Time spent in social interaction
    - Time spent with family, time spent with friends, time spent with coworkers.
    - These might in fact be negatively correlated even though total score is important.

Effect (reflective) indicators

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

Formative vs. Effect variables as regression versus factors





Measurement Invariance: Does a test measure the same thing

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

Consider the Holzinger Swineford data set

1. 9 ability measures from two schools
   - 145 from Grant-White
   - 156 from Pasteur
2. Are the factor structures the same across schools
   - Although lavaan does this in one call, lets do it part by part
   - over all factor structure
   - factors with schools
   - constrain factors to have the same loadings, etc.

First, some descriptive statistics

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

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

describeBy each group

```
> describeBy(HolzingerSwineford1939,group=HolzingerSwineford1939$schoo
group: Grant-White
         var   n   mean    sd median trimmed   mad   min    max  range
sex        2 145   1.50  0.50   2.00    1.50  0.00  1.00   2.00   1.00
ageyr      3 145  12.72  0.97  13.00   12.67  1.48 11.00  16.00   5.00
...
grade      6 144   7.45  0.50   7.00    7.44  0.00  7.00   8.00   1.00
x1         7 145   4.93  1.15   5.00    4.96  1.24  1.83   8.50   6.67
x2         8 145   6.20  1.11   6.25    6.14  1.11  2.25   9.25   7.00
...
x8        14 145   5.49  1.05   5.50    5.45  0.89  3.05  10.00   6.95
x9        15 145   5.33  1.03   5.31    5.33  1.15  3.11   9.25   6.14
----------------------------------------------------------------------
group: Pasteur
         var   n  mean    sd median trimmed   mad   min    max  range
sex        2 156  1.53  0.50   2.00    1.53  0.00  1.00   2.00   1.00 0
ageyr      3 156 13.25  1.06  13.00   13.15  1.48 12.00  16.00   4.00 0
...
grade      6 156  7.50  0.50   7.50    7.50  0.74  7.00   8.00   1.00 0
x1         7 156  4.94  1.19   5.00    4.97  1.24  0.67   7.50   6.83 0
x2         8 156  5.98  1.23   5.75    5.89  1.11  3.50   9.25   5.75 0
...
```

EFA for both groups

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

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

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

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

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

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

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

How similar are the solutions: factor congruence

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

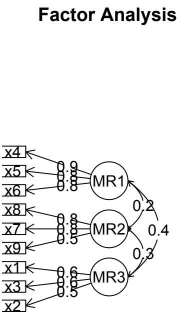
$$Congruence = (X'X)^{.5}Y'X(Y'Y)^{.5}$$

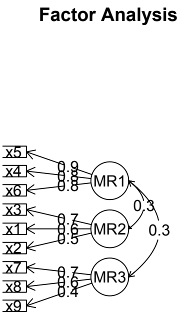
R code

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

```
     MR1  MR2  MR3
MR1 0.97 0.03 0.10
MR2 0.12 0.11 0.99
MR3 0.08 0.97 0.05
  cross <- t(y) %*% x
  sumsx <- sqrt(1/diag(t(x) %*% x))
   sumsy <- sqrt(1/diag(t(y) %*% y))
```

Do they look alike?





Test if the model fits the combined data

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

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

```
lavaan (0.4-14) converged normally after 41 iterations

  Number of observations                           301

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

Chi-square test baseline model:

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

Full model versus baseline model:

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

Loglikelihood and Information Criteria:

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

With values of

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

Root Mean Square Error of Approximation:

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

Standardized Root Mean Square Residual:

  SRMR                                           0.065

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

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

Now do it for both groups one analysis

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

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

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

Chi-square for each group:

  Pasteur                                        64.309
  Grant-White                                    51.542

Parameter estimates:

  Information                                  Expected
  Standard Errors                              Standard
```

Results for Pasteur

```
Group 1 [Pasteur]:

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

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

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

Results for Grant-White

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

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

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

Constrain the loadings to be the same

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

lavaan (0.4-14) converged normally after 30 iterations

  Number of observations per group
  Pasteur                                           156
  Grant-White                                       145

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

Chi-square for each group:

  Pasteur                                        71.064
  Grant-White                                    56.770

Parameter estimates:

  Information                                  Expected
  Standard Errors                              Standard
```

With parameter values of

```
Group 1 [Pasteur]:

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

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

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

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

```
> anova(fit2,fit2e)

Chi Square Difference Test

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

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

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

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

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

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

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

Test fit by comparing models

```
> anova(fit2,fit2e)

Chi Square Difference Test

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

Continue this logic, of successive tests with more constraints, but do it automatically: old lavaan way

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

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

Measurement invariance with lavaan and semTools

R code

```
library(semTools)  #make this additional package active
mi <- measurementInvariance(HS.model,
          data=HolzingerSwineford1939, group="school")
```

```
Measurement invariance models:

Model 1 : fit.configural
Model 2 : fit.loadings
Model 3 : fit.intercepts
Model 4 : fit.means

Chi Square Difference Test

               Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
fit.configural 48 7484.4 7706.8 115.85
fit.loadings   54 7480.6 7680.8 124.04      8.192       6     0.2244
fit.intercepts 60 7508.6 7686.6 164.10     40.059       6  4.435e-07 ***
fit.means      63 7543.1 7710.0 204.61     40.502       3  8.338e-09 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Fit measures:
                 cfi rmsea cfi.delta rmsea.delta
fit.configural 0.923 0.097        NA          NA
fit.loadings   0.921 0.093     0.002       0.004
fit.intercepts 0.882 0.107     0.038       0.015
fit.means      0.840 0.122     0.042       0.015
```

Using sim.structural to create models

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

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

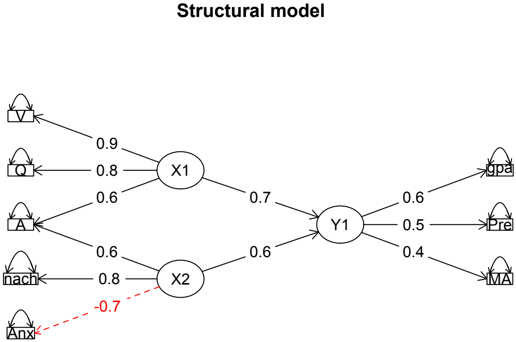
```
Call: sim.structural(fx = fx, Phi = Phi, fy = fy)

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

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


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


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

Create a figure (and write sem code)

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



Show the sem code

```
> mod4

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

Part II

Change

Part II: Detecting and measuring change

Measuring structure at two (or more) time points

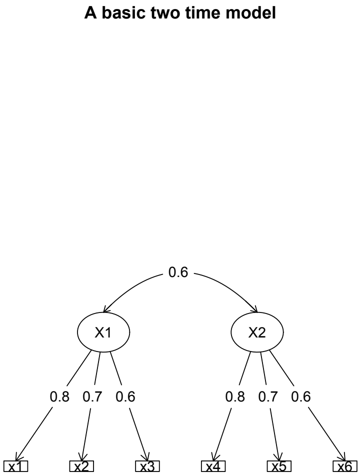
1. Is the structure the same
   - Structural Invariance (is the graph the same)
   - Measurement invariance (are the loadings the same)
   - Strong measurement invariance (are the item intercepts the same?)
   - Measuring change
2. Do the means change (is there growth)
   - This is the means of the latent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

Create some basic data and add in some change

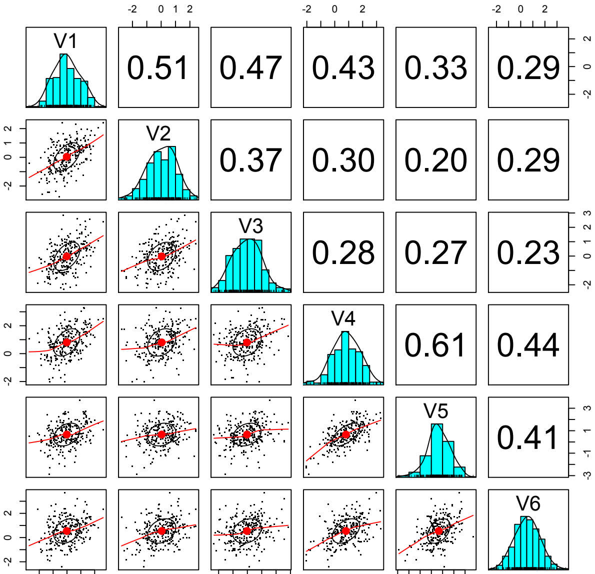
```
> set.seed(42)
 > fx <- matrix(c(.8,.7,.6,rep(0,6),.8,.7,.6),ncol=2)
> fx
     [,1] [,2]
[1,]  0.8  0.0
[2,]  0.7  0.0
[3,]  0.6  0.0
[4,]  0.0  0.8
[5,]  0.0  0.7
[6,]  0.0  0.6
> Phi <- matrix(c(1,.6,.6,1),ncol=2)
> Phi
     [,1] [,2]
[1,]  1.0  0.6
[2,]  0.6  1.0

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

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

   var   n  mean   sd median trimmed  mad   min  max range   se
V1   1 250 -0.01 0.94  -0.06   -0.02 1.00 -2.74 2.62  5.36 0.06
V2   2 250  0.04 0.98   0.11    0.06 0.99 -2.76 2.41  5.17 0.06
V3   3 250 -0.02 0.99  -0.03   -0.04 1.07 -2.32 2.84  5.17 0.06
V4   4 250  0.81 0.98   0.80    0.82 1.01 -2.03 3.29  5.32 0.06
V5   5 250  0.66 1.01   0.61    0.68 0.92 -2.89 3.72  6.61 0.06
V6   6 250  0.54 1.01   0.54    0.54 1.01 -2.42 3.71  6.13 0.06
```

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

A basic two occasion trait model



Splom of 6 basic variables



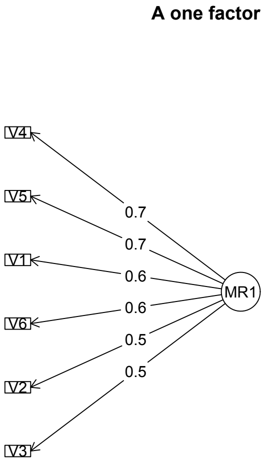
Multiple models

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

A one factor model

```
> fa(x)

Factor Analysis using method =  minres
Call: fa(r = x)
Standardized loadings (pattern matrix)
          based upon correlation matrix
    MR1   h2   u2
V1 0.64 0.41 0.59
V2 0.51 0.26 0.74
V3 0.50 0.25 0.75
V4 0.75 0.56 0.44
V5 0.66 0.43 0.57
V6 0.56 0.31 0.69

                MR1
SS loadings    2.22
Proportion Var 0.37

Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the null model are  1
     and the objective function was  1.58
       with Chi Square of  388.97
The degrees of freedom for the model are 9
          and the objective function was  0.3

The root mean square of the residuals (RMSR) is
The df corrected root mean square of the
                   residuals is  0.12
The number of observations was  250
       with Chi Square =  73.6  with prob <  3e-

Tucker Lewis Index of factoring reliability =  0
RMSEA index =  0.171  and the
           90 % confidence intervals are  0.135
BIC =  23.91
Fit based upon off diagonal values = 0.94
Measures of factor score adequacy
Correlation of scores with factors             0
Multiple R square of scores with factors       0
Minimum correlation of possible factor scores  0
```

Fit a one factor model to the data



Try a two factor solution

```
> fa(x,2)
```

```
Loading required package: GPArotation
Factor Analysis using method =  minres
Call: fa(r = x, nfactors = 2)
Standardized loadings (pattern matrix)
         based upon correlation matrix
     MR1   MR2   h2   u2
V1  0.05  0.76 0.62 0.38
V2 -0.06  0.69 0.43 0.57
V3  0.04  0.55 0.33 0.67
V4  0.75  0.08 0.64 0.36
V5  0.81 -0.07 0.59 0.41
V6  0.47  0.12 0.30 0.70

                       MR1  MR2
SS loadings           1.49 1.43
Proportion Var        0.25 0.24
Cumulative Var        0.25 0.49
Proportion Explained  0.51 0.49
Cumulative Proportion 0.51 1.00
 With factor correlations of
     MR1  MR2
MR1 1.00 0.57
MR2 0.57 1.00
Measures of factor score adequacy
```

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

The degrees of freedom for the null model are  1
  and the objective function was  1.58
   with Chi Square of  388.97
The degrees of freedom for the model are 4
            and the objective function was  0.02

The root mean square of the residuals (RMSR) is
The df corrected root mean square of
                 the residuals is  0.04
The number of observations was  250
      with Chi Square =  5.87  with prob <  0.21

Tucker Lewis Index of factoring reliability =  0
RMSEA index =  0.045  and the
         90 % confidence intervals are  NA 0.112
BIC =  -16.21
Fit based upon off diagonal values = 1
```

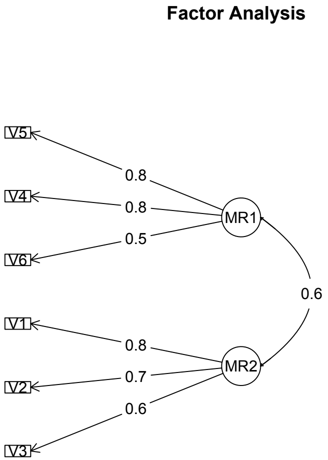
```
                                              MR1  MR2
Correlation of scores with factors           0.89 0.87
Multiple R square of scores with factors     0.80 0.77
Minimum correlation of possible factor scores 0.59 0.53
```

Compare the two solutions

```
Factor Analysis using method =  minres
Call: fa(r = x)
Standardized loadings (pattern matrix)
    based upon correlation matrix
    MR1   h2   u2
V1 0.64 0.41 0.59
V2 0.51 0.26 0.74
V3 0.50 0.25 0.75
V4 0.75 0.56 0.44
V5 0.66 0.43 0.57
V6 0.56 0.31 0.69

                MR1
SS loadings    2.22
Proportion Var 0.37
```

```
Factor Analysis using method =  minres
Call: fa(r = x, nfactors = 2)
Standardized loadings (pattern matrix)
      based upon correlation matrix
     MR1   MR2   h2   u2
V1  0.05  0.76 0.62 0.38
V2 -0.06  0.69 0.43 0.57
V3  0.04  0.55 0.33 0.67
V4  0.75  0.08 0.64 0.36
V5  0.81 -0.07 0.59 0.41
V6  0.47  0.12 0.30 0.70

                       MR1  MR2
SS loadings           1.49 1.43
Proportion Var        0.25 0.24
Cumulative Var        0.25 0.49
Proportion Explained  0.51 0.49
Cumulative Proportion 0.51 1.00

 With factor correlations of
     MR1  MR2
MR1 1.00 0.57
MR2 0.57 1.00
```

EFA two factor solution



Now try three different sems (using lavaan)

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

A simple sem with a standardized solution

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

```
   lhs op rhs est.std se  z pvalue
1   F1 =~  V1   0.819 NA NA     NA
2   F1 =~  V2   0.618 NA NA     NA
3   F1 =~  V3   0.579 NA NA     NA
4   F2 =~  V4   0.835 NA NA     NA
5   F2 =~  V5   0.720 NA NA     NA
6   F2 =~  V6   0.552 NA NA     NA
7   F1 ~~  F2   0.607 NA NA     NA
```

lavaan fit statistics

```
lavaan (0.4-14) converged normally after 16 iterations

  Number of observations                           250

  Estimator                                         ML
  Minimum Function Chi-square                    9.341
  Degrees of freedom                                 8
  P-value                                        0.314

Chi-square test baseline model:

  Minimum Function Chi-square                  395.026
  Degrees of freedom                                15
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.996
  Tucker-Lewis Index (TLI)                       0.993

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -1909.112
  Loglikelihood unrestricted model (H1)      -1904.445

  Number of free parameters                         13
  Akaike (AIC)                                3844.224
  Bayesian (BIC)                              3890.003
  Sample-size adjusted Bayesian (BIC)         3848.792
```

```
Root Mean Square Error of Approximation:

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

Standardized Root Mean Square Residual:

  SRMR

Parameter estimates:

  Information                                            Ex
  Standard Errors                                        St

                   Estimate  Std.err  Z-value  P
Latent variables:
  F1 =~
    V1                0.768    0.062   12.363
    V2                0.604    0.065    9.313
    V3                0.569    0.066    8.672
  F2 =~
    V4                0.818    0.062   13.296
    V5                0.727    0.064   11.346
    V6                0.558    0.066    8.472

Covariances:
  F1 ~~
    F2                0.607    0.062    9.861
```

Create two new models with some equality constraints

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

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

Equal across time

```
lavaan (0.4-14) converged normally after 15 iterations

  Number of observations                            250

  Estimator                                          ML
  Minimum Function Chi-square                    11.797
  Degrees of freedom                                 11
  P-value                                         0.379

Chi-square test baseline model:

  Minimum Function Chi-square                   395.026
  Degrees of freedom                                 15
  P-value                                         0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                     0.998
  Tucker-Lewis Index (TLI)                        0.997

  Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)               -1910.340
  Loglikelihood unrestricted model (H1)       -1904.441

  Number of free parameters                          10
  Akaike (AIC)                                 3840.679
  Bayesian (BIC)                               3875.894
  Sample-size adjusted Bayesian (BIC)          3844.193
```

```
Root Mean Square Error of Approximation:

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

Standardized Root Mean Square Residual:

  SRMR

Parameter estimates:

  Information                                            Ex
  Standard Errors                                        St

                   Estimate  Std.err  Z-value  P
Latent variables:
  F1 =~
    V1        (a)     0.793    0.046   17.162
    V2        (b)     0.669    0.047   14.116
    V3        (c)     0.564    0.048   11.819
  F2 =~
    V4        (a)     0.793    0.046   17.162
    V5        (b)     0.669    0.047   14.116
    V6        (c)     0.564    0.048   11.819

Covariances:
  F1 ~~
    F2                0.606    0.061    9.869
```

All loadings equal

```
> summary(fit2e,fit.measures=TRUE)
lavaan (0.4-14) converged normally after 12 iterations

  Number of observations                           250

  Estimator                                         ML
  Minimum Function Chi-square                   25.457
  Degrees of freedom                                13
  P-value                                        0.020

Chi-square test baseline model:

  Minimum Function Chi-square                  395.026
  Degrees of freedom                                15
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.967
  Tucker-Lewis Index (TLI)                       0.962

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -1917.169
  Loglikelihood unrestricted model (H1)      -1904.441

  Number of free parameters                          8
  Akaike (AIC)                                3850.339
  Bayesian (BIC)                              3878.510
  Sample-size adjusted Bayesian (BIC)         3853.150
```

```
Root Mean Square Error of Approximation:

  RMSEA
  90 Percent Confidence Interval          0.024
  P-value RMSEA <= 0.05

Standardized Root Mean Square Residual:

  SRMR

Parameter estimates:

  Information                                 Ex
  Standard Errors                             St

                   Estimate  Std.err  Z-value  P
Latent variables:
  F1 =~
    V1        (a)     0.686    0.033   20.987
    V2        (a)     0.686    0.033   20.987
    V3        (a)     0.686    0.033   20.987
  F2 =~
    V4        (a)     0.686    0.033   20.987
    V5        (a)     0.686    0.033   20.987
    V6        (a)     0.686    0.033   20.987

Covariances:
  F1 ~~
    F2                0.624    0.063    9.909
```

Create a data set with non-invariant factor loadings

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

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

One factor model

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

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

                MR1
SS loadings    2.18
Proportion Var 0.36

Test of the hypothesis that 1 factor is sufficient.

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

```
The root mean square of the residuals (RMSR) is
The df corrected root mean square of the residua
The number of observations was  250  with Chi Sq

Tucker Lewis Index of factoring reliability =  0
RMSEA index =  0.179  and the 90 % confidence in
BIC =  30.24
Fit based upon off diagonal values = 0.93
Measures of factor score adequacy

Correlation of scores with factors             0
Multiple R square of scores with factors       0
Minimum correlation of possible factor scores  0
```

Two factor model

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


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

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

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

Correlation of scores with factors              0.90 0.88
Multiple R square of scores with factors        0.80 0.77
Minimum correlation of possible factor scores   0.61 0.54
```

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

The degrees of freedom for the null model are  1
             and the objective function was  1.5
             with Chi Square of  389.85
The degrees of freedom for the model are 4
                    and the objective function

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

Tucker Lewis Index of factoring reliability =  0
RMSEA index =  0.028  and the 90 % confidence in
BIC =  -17.39
Fit based upon off diagonal values = 1
Measures of factor score adequacy
```

Compare the two solutions

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

                MR1
SS loadings    2.18
Proportion Var 0.36
```

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

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

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

CFA 2 correlated factors

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


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

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

Full model versus baseline model:

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

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -1952.265
  Loglikelihood unrestricted model (H1)      -1950.989

  Number of free parameters                         13
  Akaike (AIC)                                3930.530
  Bayesian (BIC)                              3976.309
  Sample-size adjusted Bayesian (BIC)         3935.09
```

```
Root Mean Square Error of Approximation:

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

Standardized Root Mean Square Residual:

  SRMR

Parameter estimates:

  Information                                      Ex
  Standard Errors                                  St

                   Estimate  Std.err  Z-value  P
Latent variables:
  F1 =~
    V1                0.750    0.063   11.880
    V2                0.725    0.065   11.160
    V3                0.621    0.066    9.410
  F2 =~
    V4                0.585    0.074    7.896
    V5                0.610    0.068    8.950
    V6                0.700    0.066   10.545

Covariances:
  F1 ~~
    F2                0.607    0.067    9.067

Variances:
```

Are the factors measurement invariant? Well, maybe.

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

lavaan (0.4-14) converged normally after 14 iterations

  Number of observations                            250

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

Chi-square test baseline model:

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

Full model versus baseline model:

  Comparative Fit Index (CFI)                     1.000
  Tucker-Lewis Index (TLI)                        1.012

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)               -1955.036
  Loglikelihood unrestricted model (H1)       -1950.989

  Number of free parameters                          10
  Akaike (AIC)                                 3930.072
  Bayesian (BIC)                               3965.286
```

```
Root Mean Square Error of Approximation:

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

Standardized Root Mean Square Residual:

  SRMR

Parameter estimates:

  Information                                       Ex
  Standard Errors                                   St

                    Estimate  Std.err  Z-value  P
Latent variables:
  F1 =~
    V1        (a)     0.686    0.050   13.774
    V2        (b)     0.675    0.049   13.834
    V3        (c)     0.657    0.048   13.677
  F2 =~
    V4        (a)     0.686    0.050   13.774
    V5        (b)     0.675    0.049   13.834
    V6        (c)     0.657    0.048   13.677

Covariances:
  F1 ~~
    F2                0.610    0.067    9.089

Variances:
```

Create the original data set again

```
> fx <- matrix(c(.8,.7,.6,rep(0,6),.8,.7,.6),ncol=2)
> x.model <- sim(fx=fx,Phi=Phi,mu=c(0,1),n=250)
> x <- x.model$observed
> x.df <- data.frame(x)

> describe(x,skew=FALSE)
   var   n mean   sd median trimmed  mad   min  max range   se
V1   1 250 0.02 0.99  -0.01   -0.01 1.04 -2.38 3.26  5.64 0.06
V2   2 250 0.13 1.07   0.03    0.11 1.08 -2.53 3.65  6.19 0.07
V3   3 250 0.03 1.03   0.05    0.02 0.97 -2.68 3.12  5.80 0.07
V4   4 250 0.87 1.04   0.90    0.85 1.03 -1.63 3.61  5.24 0.07
V5   5 250 0.73 0.97   0.76    0.75 0.90 -2.36 3.29  5.65 0.06
V6   6 250 0.65 1.00   0.67    0.64 1.03 -2.58 3.17  5.75 0.06
```

Model change in the items

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

lavaan: change over time?

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

  Number of observations                           250

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

Chi-square test baseline model:

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

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.998
  Tucker-Lewis Index (TLI)                       0.998

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -1952.674
  Loglikelihood unrestricted model (H1)      -1947.365

  Number of free parameters                         17
  Akaike (AIC)                                3939.347
  Bayesian (BIC)                              3999.212
  Sample-size adjusted Bayesian (BIC)         3945.320
```

```
Root Mean Square Error of Approximation:

  RMSEA
  90 Percent Confidence Interval          0
  P-value RMSEA <= 0.05

Standardized Root Mean Square Residual:

  SRMR

Parameter estimates:

  Information
  Standard Errors

                   Estimate  Std.err  Z-valu
Latent variables:
  F1 =~
    V1        (a)     1.000
    V2        (b)     0.925    0.076   12.2
    V3        (c)     0.816    0.072   11.3
  F2 =~
    V4        (a)     1.000
    V5        (b)     0.925    0.076   12.2
    V6        (c)     0.816    0.072   11.3

Regressions:
  F2 ~
    F1                0.644    0.075    8.5

Intercepts:
```

With the intercepts

Intercepts:

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

Variances:

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

Try fitting a moments model

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

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

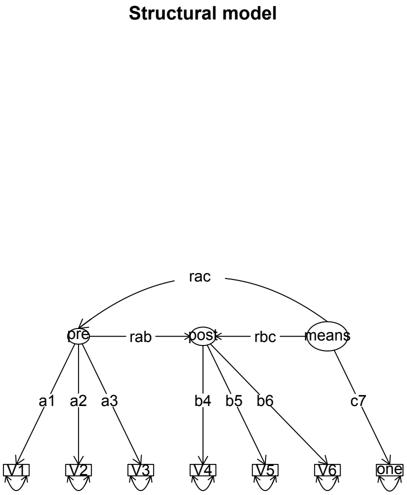
Modeling change using the moments matrix

1. McArdle (2009) Latent variable modeling of differences and changes with longitudinal data. Annual Review of Psychology, 60, 577-605
    - Use moments rather than covariances
2. Probably can be done in lavaan, but I don't know how. Can be done in sem package

Create the model to be fit in sem

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

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

Modeling the means in a moments matrix



the basic path model, with some editing

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

sem output

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

Model fit to raw moment matrix.

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

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

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1    0.830111 0.069823 11.8888  1.3536e-32 V1 <--- pre
a2    0.881776 0.076506 11.5256  9.7982e-31 V2 <--- pre
a3    0.698963 0.074824  9.3415  9.5005e-21 V3 <--- pre
b4    0.664838 0.062395 10.6553  1.6468e-26 V4 <--- post
b5    0.554498 0.053510 10.3626  3.6687e-25 V5 <--- post
b6    0.510270 0.051499  9.9084  3.8265e-23 V6 <--- post
c7    1.118034 0.050000 22.3607 9.5054e-111 one <--- mea
x1e   0.525613 0.078258  6.7164  1.8623e-11 V1 <--> V1
x2e   0.673405 0.093381  7.2114  5.5387e-13 V2 <--> V2
x3e   0.836504 0.090362  9.2572  2.0983e-20 V3 <--> V3
x4e   0.567286 0.081353  6.9732  3.0990e-12 V4 <--> V4
x5e   0.628041 0.073463  8.5490  1.2412e-17 V5 <--> V5
x6e   0.750522 0.080300  9.3465  9.0592e-21 V6 <--> V6
rF1F2 0.893184 0.142998  6.2461  4.2076e-10 post <--- pr
rF3F1 0.086583 0.073172  1.1833  2.3669e-01 pre <--- mea
rF3F2 1.375097 0.159247  8.6350  5.8723e-18 post <--- me

 Iterations =  21
```

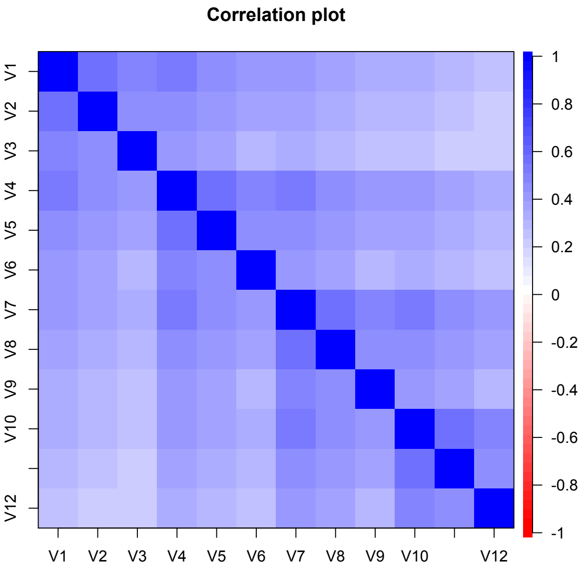
Traits and States and time

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

A factor simplex

```
simp <- sim()

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

A simplex


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

Factor structure of a simplex

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

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

                MR1
SS loadings    4.50
Proportion Var 0.38
```

Factors over time

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

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

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

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

44 ways to fool yourself with sem

44 ways to fool yourself with SEM

Adapted from Rex Kline; Principals and Practice of Structural Equation Modeling, 2005

1. Specification
2. Data
3. Analysis and Respecificaton
4. Interpretation

Specification errors

1. Specifying the model after the data are collected.
    - Particularly a problem when using archival data.
2. Are key variables omitted?
3. Is the model identifiable?
4. Omitting causes that are correlated with other variables in the structural model.
5. Failure to have sufficient number of indicators of latent variables.
    - "Two might be fine, three is better, four is best, anything more is gravy" (Kenny, 1979)
6. Failure to give careful consideration to directionality.
    - Path techniques are good for estimating strengths if we know the underlying model, but are not good for determining the model (Meehl and Walker, 2002)

Specification errors (continued)

7. Specifying feedback loops ("non recursive models") as a way to mask uncertainty
8. Overfit the model, ignoring parsimony
9. Add disturbances ("measurement error correlations" aka "correlated residuals") with substantive reason
10. Specifying indicators that are multivocal without substantive reason

Data Errors

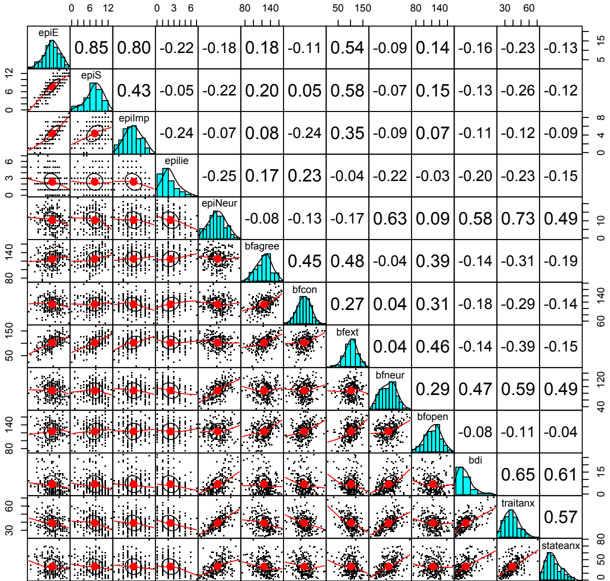
1. Failure to check the accuracy of data input or coding
   - Missing data codes (use a clear missing value)
   - Misytyped, mis-scanned data matrices
   - Improperly reversed items
     - Let the computer do it for you
     - Why reverse an item when a negative sign will do it for you?
2. Ignoring the pattern of missing data, is it random or systematic.
3. Failure to examine distributional characteristics
   - Weird data -> weird results
4. Failure to screen for outliers
   - Outliers due to mistakes
   - Outliers due to systematic differences

Describe the data

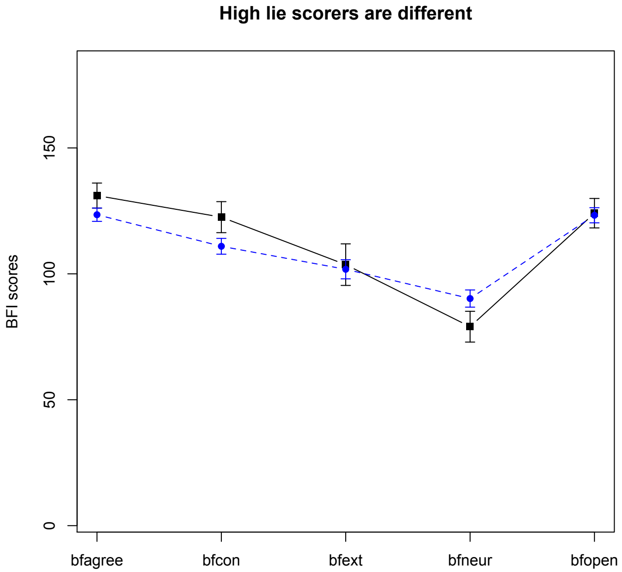
```
> describe(epi.bfi)

 pairs.panels(epi.bfi,pch=".",gap=0)   #mind the gap
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
epiE	1	231	13.33	4.14	14	13.49	4.45	1	22	21	-0.33	-0.06	0.27
epiS	2	231	7.58	2.69	8	7.77	2.97	0	13	13	-0.57	-0.02	0.18
epiImp	3	231	4.37	1.88	4	4.36	1.48	0	9	9	0.06	-0.62	0.12
epilie	4	231	2.38	1.50	2	2.27	1.48	0	7	7	0.66	0.24	0.10
epiNeur	5	231	10.41	4.90	10	10.39	4.45	0	23	23	0.06	-0.50	0.32
bfagree	6	231	125.00	18.14	126	125.26	17.79	74	167	93	-0.21	-0.27	1.19
bfcon	7	231	113.25	21.88	114	113.42	22.24	53	178	125	-0.02	0.23	1.44
bfext	8	231	102.18	26.45	104	102.99	22.24	8	168	160	-0.41	0.51	1.74
bfneur	9	231	87.97	23.34	90	87.70	23.72	34	152	118	0.07	-0.55	1.54
bfopen	10	231	123.43	20.51	125	123.78	20.76	73	173	100	-0.16	-0.16	1.35
bdi	11	231	6.78	5.78	6	5.97	4.45	0	27	27	1.29	1.50	0.38
traitanx	12	231	39.01	9.52	38	38.36	8.90	22	71	49	0.67	0.47	0.63
stateanx	13	231	39.85	11.48	38	38.92	10.38	21	79	58	0.72	-0.01	0.76

Graphic descriptions using SPLOMs



High lie score subjects seem different



Data errors (continued)

5. Assuming all relationships are linear without checking
    - graphical techniques are helpful for non-linearities
    - Simple graphical techniques do not help for interactions
6. Ignoring lack of independence among observations
    - Nesting of subjects within pairs, within classrooms, with managers
    - Can we model the nesting?

Errors of analysis and respecification

1. Failure to check the accuracy of computer syntax
   - Direction of effects
   - Error specifications
   - Omitted paths
2. Respecifying the model based entirely on statistical criteria
   - Just because it does not fit does not mean it should be fixed
3. Failure to check for admissible solutions
   - Are some of the paths impossible?
   - Do some of the variables have negative variances?
4. Reporting only standardized estimates
   - These are sample based estimates and reflect variances (errorful) and covariances (supposedly error free)
5. Analyzing a correlation matrix when the covariance matrix is more appropriate
   - Anything that has growth or change component must be done with covariances

Errors of Analysis and respecification (continued)

6. Analyzing a data set with extremely high correlations
   - solution will either be unstable or will not work if variables are too "colinear"
7. Not enough subjects for complexity of the data
   - This is ambiguous – what is enough?
   - Remember, the standard error of a correlation reflects sample size $se = \frac{r^2}{\sqrt{(1-r^2)(n-2)}}$

Errors of Analysis and respecification (continued)

8. Setting scales of latent variables inappropriately.
    - particularly a problem with multiple group comparisons
9. Ignoring the start values or giving bad ones.
    - Supplying reasonable start values helps a great deal
10. Do different start values lead to different solutions?
11. Failure to recognize empirical underidentification
    - for some data sets, the model is underidentified even though there are enough parameters
    - Failure to separate measurement from structural portion of model
        - Use the two or four step procedure

Errors of Analysis and respecification (continued)

12. Estimating means and intercepts without showing measurement invariance
13. Analyzing parcels without checking if parcels are in fact factorially homogeneous.
    - Factorial Homogeneous Item Domains (FHID)
    - Homogenous Item Composites (HIC)
    - (but consider contradictory advice on parcels)

Errors of Interpretation

1. Looking only at indexes of overall fit
    - need to examine the residuals to see where there is misfit, even though overall model is fine
2. Interpreting good fit as meaning model is "proved".
    - consider alternative models
    - better able to reject alternatives
3. Interpreting good fit as meaning that the endogenous variables are strongly predicted.
    - What is predicted is the covariance of the variables, not the variables
    - Are the residual covariances small, not whether the error variance is small
4. Relying solely on statistical criterion in model evaluation
    - What can the model not explain
    - What are alternative models
    - What constraints does the model imply

Errors of interpretation (continued)

5. Relying too much on statistical tests
   - significance of particular path coefficients does not imply effect size or importance
   - Overall statistical fit (χ^2) is sensitive to model misfit as f(N)
6. Misinterpreting the standardized solution in multiple group problems
7. Failure to consider equivalent models
   - Why is this model better than equivalent models?
8. Failure to consider non-equivalent models
   - Why is this model better than other, non-equivalent models?
9. Reifying the latent variables
   - Latent variables are just models of observed data
   - "Factors are fictions"
10. Believing that naming a factor means it is understood

Errors of interpretation (continued)

11. Believing that a strong analytical method like SEM can overcome poor theory or poor design.
12. Failure to report enough so that you can be replicated
13. Interpreting estimates of large effects as evidence for "causality"

Final Comments

1. Theory First
   - What are the alternative theories?
   - Are there specific differences in the theories that are testable?
2. Measurement Model
   - Comparison of measurement models?
   - How many latent variables? How do you know?
   - Measurement Invariance?
3. Structural Model
   - Comparison of multiple models?
   - What happens if the arrows are reversed?
4. Theory Last
   - What do we know now that we did not know before?
   - What do we have shown is not correct?

Bollen, K. A. (2002). *Latent variables in psychology and the social sciences*. US: Annual Reviews.

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

Kerckhoff, A. C. (1974). *Ambition and Attainment: A Study of Four Samples of American Boys.* Washington, D. C.: American Sociological Association.