

Psychology 454: Latent Variable Modeling

The effect of items on latent structures

Department of Psychology
Northwestern University
Evanston, Illinois USA



January, 2011

Outline

- 1 The problem of items
- 2 Simulated items
 - The pure model
 - 4 level items
 - Dichotomous items
 - Summary of model fits
- 3 Real data
 - Dichotomous items – IQ
 - Polytomous items– Mood

The problem with items

- Maximum Likelihood estimation is based upon continuous normal assumptions
- But items are discrete.
 - Correlations are reduced by forming categories. $\phi < \rho$
 - Correlations will differ as a function of item difficulty.
- Can demonstrate these problems with simulated items.
- Demonstration with dichotomous (e.g., IQ) or polytomous (e.g., mood) items

Simulate items with known structures

- Steps of simulation
 - Create a pure, continuous population model
 - Sample from population model
 - Distort with categorical items
 - Distort with dichotomous items
- Problem of categorical items most pronounced with differences in item difficulty.
 - Functionally, this leads to differences in skew.
 - Skew is a major problem when studying emotion.

Pure congeneric model

- Simulate 9 items with moderate loadings (.5 - .6)
 - Examine the covariances
 - Examine the structure
- Repeat with 4 categories
- Repeat with 2 categories

9 congeneric items - pure model

```
set.seed(42)
latent <- sim.congeneric(loads =sample(c(.4,.5,.6),9,replace=TRUE),N=200,short=F)
latent$model
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.36	0.24	0.36	0.30	0.30	0.36	0.24	0.30
V2	0.36	1.00	0.24	0.36	0.30	0.30	0.36	0.24	0.30
V3	0.24	0.24	1.00	0.24	0.20	0.20	0.24	0.16	0.20
V4	0.36	0.36	0.24	1.00	0.30	0.30	0.36	0.24	0.30
V5	0.30	0.30	0.20	0.30	1.00	0.25	0.30	0.20	0.25
V6	0.30	0.30	0.20	0.30	0.25	1.00	0.30	0.20	0.25
V7	0.36	0.36	0.24	0.36	0.30	0.30	1.00	0.24	0.30
V8	0.24	0.24	0.16	0.24	0.20	0.20	0.24	1.00	0.20
V9	0.30	0.30	0.20	0.30	0.25	0.25	0.30	0.20	1.00

Factor the model

```
fa(latent$model)
```

```
Call: fa(r = latent$model)
```

```
Standardized loadings based upon correlation matrix
```

	MR1	h2	u2
V1	0.6	0.36	0.64
V2	0.6	0.36	0.64
V3	0.4	0.16	0.84
V4	0.6	0.36	0.64
V5	0.5	0.25	0.75
V6	0.5	0.25	0.75
V7	0.6	0.36	0.64
V8	0.4	0.16	0.84
V9	0.5	0.25	0.75

	MR1
SS loadings	2.51
Proportion Var	0.28

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

```
The degrees of freedom for the null model are 36 and
the objective function was 1.46
```

```
The degrees of freedom for the model are 27 and
the objective function was 0
```

```
The root mean square of the residuals is 0
```

```
The df corrected root mean square of the residuals is 0
```

```
Fit based upon off diagonal values = 1
```

```
Measures of factor score adequacy
```

	MR1
Correlation of scores with factors	0.89
Multiple R square of scores with factors	0.78
Minimum correlation of possible factor scores	0.57

Factor the simulated data

```
> fa(latent$observed)
```

Test of the hypothesis that 1 factor is sufficient.

Factor Analysis using method

= minres The degrees of freedom for the null model are 36 and

Call: fa(r = latent\$observed) the objective function was 1.75 with Chi Square of

Standardized loadings based on the degrees of freedom for the model are 27 and

upon correlation matrix the objective function was 0.12

	MR1	h2	u2
V1	0.67	0.45	0.55
V2	0.60	0.36	0.64
V3	0.49	0.24	0.76
V4	0.61	0.38	0.62
V5	0.53	0.28	0.72
V6	0.47	0.23	0.77
V7	0.50	0.25	0.75
V8	0.41	0.17	0.83
V9	0.57	0.32	0.68

The root mean square of the residuals is 0.03

The df corrected root mean square of the residuals

is 0.05

The number of observations was 200

with Chi Square = 24.02 with prob < 0

Tucker Lewis Index of factoring reliability = 1.013

RMSEA index = 0 and the 90 % confidence intervals are

BIC = -119.04

Fit based upon off diagonal values = 0.98

Measures of factor score adequacy

	MR1
SS loadings	2.68
Proportion Var	0.30

	MR1
Correlation of scores with factors	0.89
Multiple R square of scores with factors	0.80
Minimum correlation of possible factor scores	0.60

Create the RAM commands to do this

```
> mod1 <- structure.diagram(fa(latent$observed),errors=TRUE)
> mod1
```

	Path	Parameter	Value
[1,]	"MR1->V1"	"F1V1"	NA
[2,]	"MR1->V2"	"F1V2"	NA
[3,]	"MR1->V3"	"F1V3"	NA
[4,]	"MR1->V4"	"F1V4"	NA
[5,]	"MR1->V5"	"F1V5"	NA
[6,]	"MR1->V6"	"F1V6"	NA
[7,]	"MR1->V7"	"F1V7"	NA
[8,]	"MR1->V8"	"F1V8"	NA
[9,]	"MR1->V9"	"F1V9"	NA
[10,]	"V1<->V1"	"x1e"	NA
[11,]	"V2<->V2"	"x2e"	NA
[12,]	"V3<->V3"	"x3e"	NA
[13,]	"V4<->V4"	"x4e"	NA
[14,]	"V5<->V5"	"x5e"	NA
[15,]	"V6<->V6"	"x6e"	NA
[16,]	"V7<->V7"	"x7e"	NA
[17,]	"V8<->V8"	"x8e"	NA
[18,]	"V9<->V9"	"x9e"	NA
[19,]	"MR1<->MR1"	NA	"1"
attr(,"class")			
[1]	"mod"		

Run sem

```
library(sem)
sem1 <- sem(mod1,S=cov(latent$observed),N=200)

Model Chisquare = 24.571 Df = 27 Pr(>Chisq) = 0.59848
Chisquare (null model) = 348.95 Df = 36
Goodness-of-fit index = 0.97387
Adjusted goodness-of-fit index = 0.95644
RMSEA index = 0 90% CI: (NA, 0.049115)
Bentler-Bonnett NFI = 0.92959
Tucker-Lewis NNFI = 1.0103
Bentler CFI = 1
SRMR = 0.03681
BIC = -118.48
```

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.32000	-0.32400	0.00004	-0.00704	0.19300	1.13000

Edit the model to fix all loadings to be the same

```
> mod2 <- edit(mod1)
> mod2
```

	Path	Parameter	Value
[1,]	"MR1->V1"	"a"	NA
[2,]	"MR1->V2"	"a"	NA
[3,]	"MR1->V3"	"a"	NA
[4,]	"MR1->V4"	"a"	NA
[5,]	"MR1->V5"	"a"	NA
[6,]	"MR1->V6"	"a"	NA
[7,]	"MR1->V7"	"a"	NA
[8,]	"MR1->V8"	"a"	NA
[9,]	"MR1->V9"	"a"	NA
[10,]	"V1<->V1"	"x1e"	NA
[11,]	"V2<->V2"	"x2e"	NA
[12,]	"V3<->V3"	"x3e"	NA
[13,]	"V4<->V4"	"x4e"	NA
[14,]	"V5<->V5"	"x5e"	NA
[15,]	"V6<->V6"	"x6e"	NA
[16,]	"V7<->V7"	"x7e"	NA
[17,]	"V8<->V8"	"x8e"	NA
[18,]	"V9<->V9"	"x9e"	NA
[19,]	"MR1<->MR1"	NA	"1"

Try the modified SEM model – better BIC, worse RMSEA

```
> sem2 <-sem(mod2,S=cov(latent$observed),N=200)
> summary(sem2)
```

```
Model Chisquare = 36.86   Df = 35 Pr(>Chisq) = 0.38289
Chisquare (null model) = 348.95   Df = 36
Goodness-of-fit index = 0.96218
Adjusted goodness-of-fit index = 0.95138
RMSEA index = 0.016341   90% CI: (NA, 0.05442)
Bentler-Bonnett NFI = 0.89437
Tucker-Lewis NNFI = 0.99389
Bentler CFI = 0.99406
SRMR = 0.069728
BIC = -148.58
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-2.2200	-0.6210	-0.0104	-0.0463	0.6660	2.1800

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
a	0.55728	0.035522	15.6880	0	V1 <--- MR1
x1e	0.64944	0.072581	8.9478	0	V1 <--> V1
x2e	0.70441	0.077944	9.0374	0	V2 <--> V2

Convert the continuous data to 4 categories

A slightly devious use of the scrub function

```
> cat7 <- round(latent$observed)
```

```
> table(cat7)
```

```
cat7
```

```
 -3  -2  -1   0   1   2   3  
11 119 420 694 421 123  12
```

```
#this does not work
```

```
> cat4 <- scrub(cat7,isvalue=c(-3,-2,-1,0,1,2,3),newvalue=c(1,1,1,2,3,4,4))
```

```
> table(cat4)cat4
```

```
  4  
1800
```

```
#but this does
```

```
> cat4 <- scrub(cat7,isvalue=c(-3,-2,-1,0,1,2,3),newvalue=c(4,4,4,5,6,7,7)) - 3
```

```
> table(cat4)
```

```
cat4
```

```
  1   2   3   4  
550 694 421 135
```

Compare the continuous to the categorical correlations

In general, the categorical correlations are smaller.

```
> r4 <- cor(cat4)
> round(r4,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.36	0.33	0.33	0.30	0.32	0.25	0.24	0.36
V2	0.36	1.00	0.31	0.39	0.29	0.28	0.25	0.17	0.28
V3	0.33	0.31	1.00	0.20	0.23	0.21	0.20	0.07	0.27
V4	0.33	0.39	0.20	1.00	0.30	0.25	0.25	0.24	0.25
V5	0.30	0.29	0.23	0.30	1.00	0.26	0.28	0.18	0.13
V6	0.32	0.28	0.21	0.25	0.26	1.00	0.18	0.17	0.22
V7	0.25	0.25	0.20	0.25	0.28	0.18	1.00	0.10	0.28
V8	0.24	0.17	0.07	0.24	0.18	0.17	0.10	1.00	0.25
V9	0.36	0.28	0.27	0.25	0.13	0.22	0.28	0.25	1.00

```
> round(latent$r - r4,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	0.00	-0.01	0.01	0.05	0.04	0.05	0.09	0.05	0.06
V2	-0.01	0.00	0.04	0.04	0.04	-0.06	0.05	0.05	0.06
V3	0.01	0.04	0.00	0.05	0.02	0.08	0.04	0.04	0.02
V4	0.05	0.04	0.05	0.00	0.03	0.05	0.02	0.05	0.10
V5	0.04	0.04	0.02	0.03	0.00	0.01	0.07	0.06	0.10
V6	0.05	-0.06	0.08	0.05	0.01	0.00	0.05	0.00	0.00
V7	0.09	0.05	0.04	0.02	0.07	0.05	0.00	0.07	0.01
V8	0.05	0.05	0.04	0.05	0.06	0.00	0.07	0.00	0.07
V9	0.06	0.06	0.02	0.10	0.10	0.00	0.01	0.07	0.00

What about the model fits?

EFA of 4 category data

```

> f4 <- fa(cat4)
> f4
Factor Analysis using
      method = minres
Call: fa(r = cat4)
Standardized loadings based
upon correlation matrix
      MR1    h2    u2
V1 0.64 0.41 0.59
V2 0.60 0.36 0.64
V3 0.46 0.21 0.79
V4 0.56 0.32 0.68
V5 0.48 0.23 0.77
V6 0.47 0.22 0.78
V7 0.44 0.19 0.81
V8 0.35 0.12 0.88
V9 0.50 0.25 0.75

```

Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the null model are 36
and the objective function was 1.39 with Chi Square of

The degrees of freedom for the model are 27
and the objective function was

The root mean square of the residuals is 0.03
The df corrected root mean square of the residuals is
The number of observations was 200 with
Chi Square = 20.99 with prob <

Tucker Lewis Index of factoring reliability = 1.034
RMSEA index = 0 and the 90 % confidence intervals are
BIC = -122.06
Fit based upon off diagonal values = 0.98
Measures of factor score adequacy

	MR1		
	MR1	Correlation of scores with factors	0.88
SS loadings	2.31	Multiple R square of scores with factors	0.77
Proportion Var	0.26	Minimum correlation of possible factor scores	0.53

What about the SEM model - It is actually better!

```
> sem4.1 <- sem(mod1,S=cov(cat4), N=200)
> summary(sem4.1)
```

```
Model Chisquare = 21.476   Df = 27 Pr(>Chisq) = 0.76349
Chisquare (null model) = 276.19   Df = 36
Goodness-of-fit index = 0.9777
Adjusted goodness-of-fit index = 0.96284
RMSEA index = 0   90% CI: (NA, 0.039538)
Bentler-Bonnett NFI = 0.92224
Tucker-Lewis NNFI = 1.0307
Bentler CFI = 1
SRMR = 0.036341
BIC = -121.58
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-1.56000	-0.28400	0.00000	-0.00475	0.40000	1.01000

Try the SEM model with equal parameters

```
> sem4.2 <- sem(mod2,S=cov(cat4),N= 200)
> summary(sem4.2)
```

```
Model Chisquare = 33.676   Df = 35 Pr(>Chisq) = 0.532
Chisquare (null model) = 276.19   Df = 36
Goodness-of-fit index = 0.96547
Adjusted goodness-of-fit index = 0.9556
RMSEA index = 0   90% CI: (NA, 0.048351)
Bentler-Bonnett NFI = 0.87807
Tucker-Lewis NNFI = 1.0057
Bentler CFI = 1
SRMR = 0.067984
BIC = -151.77
```

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.2200	-0.6170	-0.0172	-0.0485	0.4700	1.8800

Polychorics correct for categorical data

Compare polychoric correlations to the Pearson

```
> rpoly <- polychoric(cat4)
```

```
Loading required package: mvtnorm
```

```
> round(rpoly$rho,2)
```

```
> round(latent$r - rpoly$rho,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.42	0.36	0.39	0.36	0.36	0.29	0.28	0.40
V2	0.42	1.00	0.37	0.44	0.31	0.32	0.30	0.21	0.32
V3	0.36	0.37	1.00	0.22	0.28	0.23	0.23	0.07	0.28
V4	0.39	0.44	0.22	1.00	0.35	0.29	0.28	0.28	0.30
V5	0.36	0.31	0.28	0.35	1.00	0.29	0.33	0.21	0.14
V6	0.36	0.32	0.23	0.29	0.29	1.00	0.21	0.19	0.25
V7	0.29	0.30	0.23	0.28	0.33	0.21	1.00	0.12	0.32
V8	0.28	0.21	0.07	0.28	0.21	0.19	0.12	1.00	0.29
V9	0.40	0.32	0.28	0.30	0.14	0.25	0.32	0.29	1.00

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	0.00	-0.06	-0.02	0.00	-0.02	0.01	0.06	0.01	0.01
V2	-0.06	0.00	-0.02	-0.01	0.02	-0.10	0.01	0.01	0.01
V3	-0.02	-0.02	0.00	0.04	-0.03	0.05	0.01	0.04	0.00
V4	0.00	-0.01	0.04	0.00	-0.01	0.02	-0.02	0.01	0.05
V5	-0.02	0.02	-0.03	-0.01	0.00	-0.03	0.02	0.03	0.09
V6	0.01	-0.10	0.05	0.02	-0.03	0.00	0.02	-0.03	-0.02
V7	0.06	0.01	0.01	-0.02	0.02	0.02	0.00	0.05	-0.03
V8	0.01	0.01	0.04	0.01	0.03	-0.03	0.05	0.00	0.03
V9	0.01	0.01	0.00	0.05	0.09	-0.02	-0.03	0.03	0.00

The residuals are much smaller, but what about SEM?

SEM model 1 on polychoric correlation matrix

```
> semp.1 <- sem(mod1, rpoly$rho, N=200)
> summary(sem.1)

Model Chisquare = 33.745   Df = 27 Pr(>Chisq) = 0.17358
Chisquare (null model) = 354.99   Df = 36
Goodness-of-fit index = 0.96645
Adjusted goodness-of-fit index = 0.94408
RMSEA index = 0.035431   90% CI: (NA, 0.069201)
Bentler-Bonnett NFI = 0.90494
Tucker-Lewis NNFI = 0.9718
Bentler CFI = 0.97886
SRMR = 0.042953
BIC = -109.31
```

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.90000	-0.27200	0.00000	-0.00667	0.33000	1.18000

SEM on reduced model for polychoric correlations

```
> semp.2 <- sem(mod2, rpoly$rho, N=200)
> summary(semb.2)
```

```
Model Chisquare = 49.485   Df = 35 Pr(>Chisq) = 0.053178
Chisquare (null model) = 354.99   Df = 36
Goodness-of-fit index = 0.95117
Adjusted goodness-of-fit index = 0.93722
RMSEA index = 0.045603   90% CI: (NA, 0.073079)
Bentler-Bonnett NFI = 0.8606
Tucker-Lewis NNFI = 0.9533
Bentler CFI = 0.9546
SRMR = 0.080113
BIC = -135.96
```

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.780	-0.804	-0.119	-0.088	0.509	2.060

Create dichotomous items from the original items

A simple recoding

```
> cat2 <- latent$observed  
> cat2[cat2 < 0] <- 0  
> cat2[cat2 > 0] <- 1  
> table(cat2)
```

```
cat2  
  0   1  
895 905
```

Dichotomous items

Correlations of dichotomous variables are smaller yet

```
> r2 <- cor(cat2)
> round(r2,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.13	0.22	0.24	0.16	0.26	0.19	0.18	0.31
V2	0.13	1.00	0.19	0.13	0.14	0.23	0.14	0.14	0.17
V3	0.22	0.19	1.00	0.04	0.09	0.18	0.11	0.17	0.16
V4	0.24	0.13	0.04	1.00	0.15	0.14	0.19	0.03	0.18
V5	0.16	0.14	0.09	0.15	1.00	0.11	0.24	0.19	0.17
V6	0.26	0.23	0.18	0.14	0.11	1.00	0.15	0.13	0.14
V7	0.19	0.14	0.11	0.19	0.24	0.15	1.00	0.16	0.13
V8	0.18	0.14	0.17	0.03	0.19	0.13	0.16	1.00	0.15
V9	0.31	0.17	0.16	0.18	0.17	0.14	0.13	0.15	1.00

```
> round(latent$r - r2,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	0.00	0.23	0.12	0.15	0.17	0.11	0.15	0.11	0.11
V2	0.23	0.00	0.16	0.30	0.19	-0.01	0.17	0.08	0.16
V3	0.12	0.16	0.00	0.22	0.15	0.11	0.13	-0.07	0.12
V4	0.15	0.30	0.22	0.00	0.18	0.16	0.07	0.25	0.17
V5	0.17	0.19	0.15	0.18	0.00	0.15	0.11	0.04	0.06
V6	0.11	-0.01	0.11	0.16	0.15	0.00	0.08	0.04	0.09
V7	0.15	0.17	0.13	0.07	0.11	0.08	0.00	0.01	0.16
V8	0.11	0.08	-0.07	0.25	0.04	0.04	0.01	0.00	0.17
V9	0.11	0.16	0.12	0.17	0.06	0.09	0.16	0.17	0.00

EFA on dichotomous data

```
> fa(cat2)
```

```
Factor Analysis using
method = minres
Call: fa(r = cat2)
Standardized loadings
based upon correlation matrix
```

	MR1	h2	u2
V1	0.55	0.31	0.69
V2	0.37	0.14	0.86
V3	0.36	0.13	0.87
V4	0.35	0.13	0.87
V5	0.37	0.14	0.86
V6	0.42	0.18	0.82
V7	0.39	0.15	0.85
V8	0.35	0.12	0.88
V9	0.45	0.20	0.80

	MR1
SS loadings	1.50
Proportion Var	0.17

```
Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the null model are 36 and
the objective function was 0.71 with
Chi Square of 139.28
```

```
The degrees of freedom for the model are 27 and
the objective function was 0.1
```

```
The root mean square of the residuals is 0.03
The df corrected root mean square of the residuals is
The number of observations was 200 with
Chi Square = 19.27 with prob < 0.86
```

```
Tucker Lewis Index of factoring reliability = 1.1
RMSEA index = 0 and the 90 % confidence intervals are
BIC = -123.79
Fit based upon off diagonal values = 0.93
Measures of factor score adequacy
```

	MR1
Correlation of scores with factors	0.81
Multiple R square of scores with factors	0.65
Minimum correlation of possible factor scores	0.30

SEM of dichotomous data

```
> semd.1 <- sem(mod1,cov(cat2),N=200)
> summary(semd.1)
```

```
Model Chisquare = 19.714    Df = 27 Pr(>Chisq) = 0.84254
Chisquare (null model) = 142.01    Df = 36
Goodness-of-fit index = 0.9783
Adjusted goodness-of-fit index = 0.96384
RMSEA index = 0    90% CI: (NA, 0.032607)
Bentler-Bonnett NFI = 0.86118
Tucker-Lewis NNFI = 1.0916
Bentler CFI = 1
SRMR = 0.039298
BIC = -123.34
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-1.28000	-0.21800	0.00000	0.00376	0.30800	1.30000

Dichotomous items

With parameter estimates of (based upon covariances)

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)			
F1V1	0.27652	0.042768	6.4656	1.0090e-10	V1	<---	MR1
F1V2	0.18719	0.043634	4.2899	1.7873e-05	V2	<---	MR1
F1V3	0.18141	0.043373	4.1825	2.8834e-05	V3	<---	MR1
F1V4	0.17737	0.043468	4.0804	4.4950e-05	V4	<---	MR1
F1V5	0.18633	0.043381	4.2951	1.7460e-05	V5	<---	MR1
F1V6	0.21077	0.043083	4.8923	9.9686e-07	V6	<---	MR1
F1V7	0.19647	0.043405	4.5265	5.9956e-06	V7	<---	MR1
F1V8	0.17447	0.043338	4.0258	5.6782e-05	V8	<---	MR1
F1V9	0.22678	0.042843	5.2932	1.2019e-07	V9	<---	MR1
x1e	0.17175	0.023111	7.4316	1.0725e-13	V1	<-->	V1
x2e	0.21612	0.023930	9.0310	0.0000e+00	V2	<-->	V2
x3e	0.21812	0.023908	9.1234	0.0000e+00	V3	<-->	V3
x4e	0.21957	0.023968	9.1610	0.0000e+00	V4	<-->	V4
x5e	0.21403	0.023690	9.0346	0.0000e+00	V5	<-->	V5
x6e	0.20681	0.023582	8.7697	0.0000e+00	V6	<-->	V6
x7e	0.21255	0.023791	8.9342	0.0000e+00	V7	<-->	V7
x8e	0.21921	0.023857	9.1886	0.0000e+00	V8	<-->	V8
x9e	0.19920	0.023341	8.5342	0.0000e+00	V9	<-->	V9

These may be converted to standardized coefficients

```
> std.coef(semdata.1)
```

		Std. Estimate	
F1V1	F1V1	0.5550216	V1 <--- MR1
F1V2	F1V2	0.3735111	V2 <--- MR1
F1V3	F1V3	0.3620711	V3 <--- MR1
F1V4	F1V4	0.3540061	V4 <--- MR1
F1V5	F1V5	0.3735918	V5 <--- MR1
F1V6	F1V6	0.4205076	V6 <--- MR1
F1V7	F1V7	0.3920419	V7 <--- MR1
F1V8	F1V8	0.3491852	V8 <--- MR1
F1V9	F1V9	0.4529830	V9 <--- MR1
x1e	x1e	0.6919511	V1 <--> V1
x2e	x2e	0.8604895	V2 <--> V2
x3e	x3e	0.8689045	V3 <--> V3
x4e	x4e	0.8746797	V4 <--> V4
x5e	x5e	0.8604292	V5 <--> V5
x6e	x6e	0.8231734	V6 <--> V6
x7e	x7e	0.8463032	V7 <--> V7
x8e	x8e	0.8780697	V8 <--> V8
x9e	x9e	0.7948064	V9 <--> V9
		1.0000000	MR1 <--> MR1

Force them to be equal

```
> summary(sem2.2)
```

```
Model Chisquare = 24.246   Df = 35 Pr(>Chisq) = 0.9139
Chisquare (null model) = 142.01   Df = 36
Goodness-of-fit index = 0.97392
Adjusted goodness-of-fit index = 0.96646
RMSEA index = 0   90% CI: (NA, 0.019814)
Bentler-Bonnett NFI = 0.82927
Tucker-Lewis NNFI = 1.1043
Bentler CFI = 1
SRMR = 0.049417
BIC = -161.20
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-1.7800	-0.3270	-0.0650	-0.0126	0.3530	2.0700

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
a	0.20280	0.016085	12.6077	0	V1 <--- MR1
x1e	0.19161	0.020953	9.1449	0	V1 <--> V1
x2e	0.21167	0.022865	9.2577	0	V2 <--> V2
x3e	0.21559	0.023250	9.2724	0	V3 <--> V3

Standardized coefficients for all equal model

```
> std.coef(sem.d.2)
```

		Std. Estimate		
1	a	0.4203691	V1 <--- MR1	
2	a	0.4033421	V2 <--- MR1	
3	a	0.4002575	V3 <--- MR1	
4	a	0.3984961	V4 <--- MR1	
5	a	0.4048270	V5 <--- MR1	
6	a	0.4059531	V6 <--- MR1	
7	a	0.4047453	V7 <--- MR1	
8	a	0.4008939	V8 <--- MR1	
9	a	0.4086144	V9 <--- MR1	
10	x1e	0.8232898	V1 <--> V1	
11	x2e	0.8373151	V2 <--> V2	
12	x3e	0.8397939	V3 <--> V3	
13	x4e	0.8412009	V4 <--> V4	
14	x5e	0.8361151	V5 <--> V5	
15	x6e	0.8352021	V6 <--> V6	
16	x7e	0.8361812	V7 <--> V7	
17	x8e	0.8392841	V8 <--> V8	
18	x9e	0.8330343	V9 <--> V9	
19		1.0000000	MR1 <--> MR1	

Dichotomous items

Use the tetrachoric correlation to see how well it fits

```
> rtet <- tetrachoric(cat2)
> round(rtet$rho,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.20	0.34	0.37	0.26	0.40	0.30	0.28	0.47
V2	0.20	1.00	0.29	0.20	0.22	0.35	0.22	0.22	0.27
V3	0.34	0.29	1.00	0.06	0.15	0.28	0.17	0.27	0.25
V4	0.37	0.20	0.06	1.00	0.24	0.22	0.30	0.05	0.28
V5	0.26	0.22	0.15	0.24	1.00	0.17	0.37	0.30	0.26
V6	0.40	0.35	0.28	0.22	0.17	1.00	0.23	0.21	0.22
V7	0.30	0.22	0.17	0.30	0.37	0.23	1.00	0.25	0.20
V8	0.28	0.22	0.27	0.05	0.30	0.21	0.25	1.00	0.23
V9	0.47	0.27	0.25	0.28	0.26	0.22	0.20	0.23	1.00

```
> round(latent$r - rtet$rho,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	0.00	0.15	0.00	0.02	0.08	-0.03	0.04	0.01	-0.05
V2	0.15	0.00	0.06	0.23	0.11	-0.13	0.09	0.00	0.06
V3	0.00	0.06	0.00	0.20	0.10	0.01	0.07	-0.16	0.03
V4	0.02	0.23	0.20	0.00	0.09	0.08	-0.03	0.23	0.07
V5	0.08	0.11	0.10	0.09	0.00	0.09	-0.02	-0.06	-0.04
V6	-0.03	-0.13	0.01	0.08	0.09	0.00	0.00	-0.04	0.01
V7	0.04	0.09	0.07	-0.03	-0.02	0.00	0.00	-0.08	0.09
V8	0.01	0.00	-0.16	0.23	-0.06	-0.04	-0.08	0.00	0.09
V9	-0.05	0.06	0.03	0.07	-0.04	0.01	0.09	0.09	0.00

SEM of tetrachoric matrix

```
> semdt.1 <- sem(mod1,rtet$rho,N=200)
> summary(semdt.1)
```

```
Model Chisquare = 59.245   Df = 27 Pr(>Chisq) = 0.000331
Chisquare (null model) = 316.61   Df = 36
Goodness-of-fit index = 0.93886
Adjusted goodness-of-fit index = 0.8981
RMSEA index = 0.077468   90% CI: (0.050605, 0.10431)
Bentler-Bonnett NFI = 0.81288
Tucker-Lewis NNFI = 0.84679
Bentler CFI = 0.88509
SRMR = 0.06001
BIC = -83.81
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-1.98000	-0.31100	0.00001	0.01280	0.49200	1.99000

With standardized coefficients

```
> std.coef(sem dt.1)
```

		Std. Estimate	
F1V1	F1V1	0.6913730	V1 <--- MR1
F1V2	F1V2	0.4576086	V2 <--- MR1
F1V3	F1V3	0.4533130	V3 <--- MR1
F1V4	F1V4	0.4467387	V4 <--- MR1
F1V5	F1V5	0.4612190	V5 <--- MR1
F1V6	F1V6	0.5238008	V6 <--- MR1
F1V7	F1V7	0.4838795	V7 <--- MR1
F1V8	F1V8	0.4347005	V8 <--- MR1
F1V9	F1V9	0.5685366	V9 <--- MR1
x1e	x1e	0.5220034	V1 <--> V1
x2e	x2e	0.7905944	V2 <--> V2
x3e	x3e	0.7945073	V3 <--> V3
x4e	x4e	0.8004245	V4 <--> V4
x5e	x5e	0.7872770	V5 <--> V5
x6e	x6e	0.7256328	V6 <--> V6
x7e	x7e	0.7658606	V7 <--> V7
x8e	x8e	0.8110355	V8 <--> V8
x9e	x9e	0.6767661	V9 <--> V9
		1.0000000	MR1 <--> MR1

Equal loadings for tetrachoric correlations of dichotomized data

```
> semdt.2 <- sem(mod2,rtet$rho,N=200)
> summary(semdt.2)
```

```
Model Chisquare = 69.536   Df = 35 Pr(>Chisq) = 0.000456
Chisquare (null model) = 316.61   Df = 36
Goodness-of-fit index = 0.93068
Adjusted goodness-of-fit index = 0.91087
RMSEA index = 0.070416   90% CI: (0.045789, 0.094538)
Bentler-Bonnett NFI = 0.78038
Tucker-Lewis NNFI = 0.87341
Bentler CFI = 0.87693
SRMR = 0.076466
BIC = -115.91
```

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.6900	-0.5330	-0.1350	-0.0436	0.4750	3.0900

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
a	0.50758	0.033854	14.9932	0	V1 <--- MR1
x1e	0.62738	0.070403	8.9113	0	V1 <--> V1

Standardized coefficients for the tetrachoric solution - equal paths

```
> std.coef(semdt.2)
```

		Std. Estimate	
1	a	0.5395490	V1 <--- MR1
2	a	0.5033171	V2 <--- MR1
3	a	0.4968128	V3 <--- MR1
4	a	0.4931004	V4 <--- MR1
5	a	0.5030261	V5 <--- MR1
6	a	0.5101995	V6 <--- MR1
7	a	0.5067306	V7 <--- MR1
8	a	0.4958995	V8 <--- MR1
9	a	0.5157357	V9 <--- MR1
10	x1e	0.7088869	V1 <--> V1
11	x2e	0.7466719	V2 <--> V2
12	x3e	0.7531770	V3 <--> V3
13	x4e	0.7568520	V4 <--> V4
14	x5e	0.7469647	V5 <--> V5
15	x6e	0.7396964	V6 <--> V6
16	x7e	0.7432241	V7 <--> V7
17	x8e	0.7540837	V8 <--> V8
18	x9e	0.7340167	V9 <--> V9
19		1.0000000	MR1 <--> MR1

Comparing solutions

Table: Comparison of various models

Data	Method	χ^2 Null	χ^2 (model)	RMSEA	df	BIC
Continuous	EFA	342.23	24.02	0	27	-119.04
	CFA	348.95	24.57	0	27	-118.48
	CFA equal	348.95	36.86	0.016	35	-148.58
Categorical (4)	EFA	270.87	20.99	0	27	-122.06
	CFA	276.19	21.476	0	27	-121.58
	CFA equal	276.19	33.676	0	35	-151.77
Polychoric	EFA					
	CFA	354.99	33.745	0.035	27	-109.31
	CFA equal	354.99	49.485	0.045	35	-135.96
Dichotomous	EFA	139.28	19.27	0	27	-123.79
	CFA	142.01	19.714	0	27	-123.34
	CFA equal	142.01	24.246	0	35	-161.20
Tetrachoric	EFA					
	CFA	316.61	59.245	0.077	27	-83.81
	CFA equal	316.61	69.536	0.0704	35	-115.91

Real data sets with item level data

- Several examples in *psych* have item level data
 - `iqitems` has 14 iq items collected on the web (need to score multiple choice to convert to true/false)
 - `bfi` has 25 personality items supposedly representing 5 factors of personality (5 items per scale) with 6 alternatives.
 - `msq` has 72 emotion words (0-3 scale) aggregated from multiple studies. Structure can be found for 2-6 dimensions.
- Each data set may be analyzed for single factor data (one bfi scale, just arousal items from msq, g level of iq)

iq items

- Items collected as part of 'SAPA' project.
 - 14 items selected from 56 items.
 - Items were “home brewed” to reflect aspects of iq
 - Items are collected from the web with volunteer participants taking as long as they wanted (power condition) for each item
- Items can also be analyzed using Item Response Techniques

Motivational State Questionnaire (MSQ)

- Items taken from several popular measures of mood.
 - Positive Affect - Negative Affect Scales (PANAS) (Watson, Clark & Tellegen, 1988)
 - Activation-Deactivation Adjective Check List (ADACL) (Thayer, 1970)
 - Circumplex measures of affect (Larsen & Diener, 1992)
- Given as part of all studies conducted in the Personality-Motivation-Cognition lab over 10 years.
- Total $N = 3896$

- Larsen, R. J. & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Emotion* (pp. 25–59). Thousand Oaks, CA: Sage Publications, Inc.
- Thayer, R. E. (1970). Activation states as assessed by verbal report and four psychophysiological variables. *Psychophysiology*, 7(1), 86–94.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.