Item characteristics the effect on model fit

A comparison of item distributions

Standard scales and tests

I. Tests based upon normal theory

II. Scales are thought to

   A. Be normally distributed

   B. Have continuous responses

   C. Joint distributions of scales are multivariate normal distributed

III. In addition, scales are hoped to be reliable measures of one construct

But, we work with items

I. Items are

   A. discrete

   B. skewed

   C. unreliable

   D. possibly multi-vocal

Three broad classes of item structures

I. Simple structured

   A. items are of complexity one (univocal)

   B. Perhaps multiple dimensions -- task is to determine dimensionality

II. Hierarchical structure

   A. correlated simple structures

III. Circumplex (or beyond) structured

   A. Items are of complexity two (rarely more)

   B. At least two, probably more dimensions

Multiple forms of item response

I. Dichotomous (yes/no or pass/fail)

II. 3 level responses (Yes/Neutral/No)

III. 4 -6 levels

IV. Many levels (>6)

Treatment of correlations

I. As is (phi/Pearson)

II. With assumptions

   A. phi -> tetrachoric

   B. 3 or 4 alternative items -> polychoric

Examples through simulation

I. Simple structured items - 2 dimensions

   A. continuous

   B. 5/3/2 category response

II. Circumplex items -- problems of fitting with CFA

III. Skewed simple structure items

Item generation

```
circ.sim <- function (nvar = 72 ,nsub = 500,
   circum = TRUE, xloading =.6, yloading = .6,
   gloading=0, xbias=0,  ybias = 0,categorical=FALSE,
   low=-3,high=3,truncate=FALSE,cutpoint=0)
     {
     avloading <- (xloading+yloading)/2
     errorweight <- sqrt(1-(avloading^2  + gloading^2))  #squared errors and true score
weights add to 1
   g <- rnorm(nsub)
     truex <- rnorm(nsub)* xloading  +xbias #generate normal true scores for x + xbias
     truey <- rnorm(nsub) * yloading + ybias #generate normal true scores for y + ybias
     if (circum)  #make a vector of radians (the whole way around the circle) if circumplex
     {radia <- seq(0,2*pi,len=nvar+1)
     rad <- radia[which(radia<2*pi)]        #get rid of the last one
    } else rad <- rep(seq(0,3*pi/2,len=4),nvar/4) #simple structure
     error<- matrix(rnorm(nsub*(nvar)),nsub)    #create normal error scores
     #true score matrix for each item reflects structure in radians
     trueitem <- outer(truex, cos(rad)) + outer(truey,sin(rad))
     item<- gloading * g +  trueitem  + errorweight*error   #observed item = true score + error
score
   if (categorical) {
     item = round(item)       #round all items to nearest integer value
          item[(item<= low)] <- low
          item[(item>high) ] <- high   }
     if (truncate) {item[item < cutpoint] <- 0  }
     return (item)     }
```

Item generation

I. Create a new function, circ.sim

II. Give flexibility for doing alternative types of simulations

    A. simple structure, circumplex, hierarchical

III. Define basic defaults that make sense

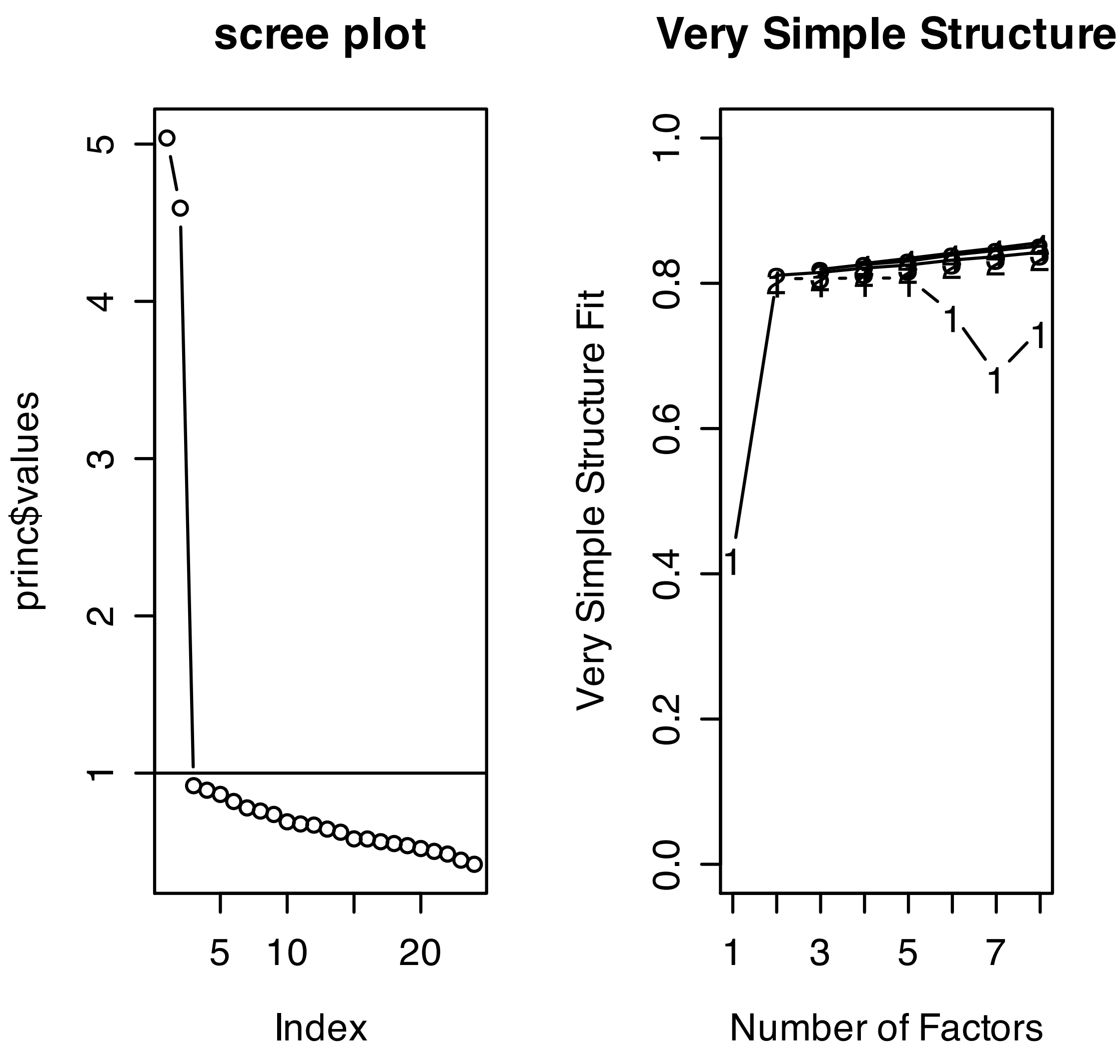
    A. nsub=500, items=72, item loadings=.6

    B. truncate=FALSE, g=0, circum=TRUE

24 items, simple structure

```
library(sem)
library(psych)
set.seed(42)
nsub=500
ss.items <- circ.sim(nvar=24,circum=FALSE,nsub)
colnames(ss.items) <- paste("V",seq(1:24),sep="")
ss.cov <- cov(ss.items)
fss <- factanal(ss.items,2)
print(fss,digits=2,cutoff=0)
```

Number of factors



Exploratory Factor Analysis

	Factor1	Factor2
V1	−0.62	−0.01
V2	0.04	0.55
V3	0.59	0.03
V4	−0.06	−0.59
...		
V22	0.02	0.56
V23	0.61	−0.01
V24	0.01	−0.57

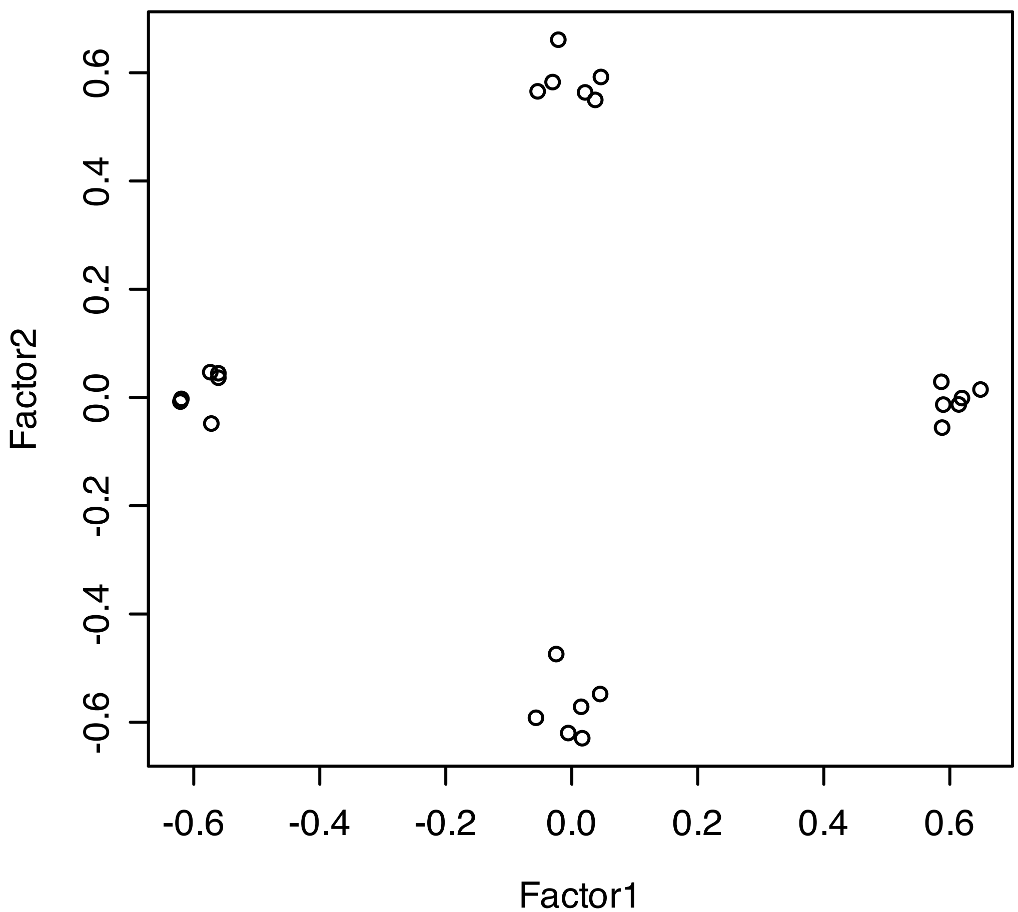
Uniquenesses:

V1	V2	V3	V4	V5	V6
0.61	0.70	0.66	0.65	0.62	

Loadings:

	Factor1	Factor2
SS loadings	4.29	4.06
Proportion Var	0.18	0.17
Cumulative Var	0.18	0.35

Items show simple structure



EFA fit statistics

	Factor1	Factor2
SS loadings	4.29	4.06
Proportion Var	0.18	0.17
Cumulative Var	0.18	0.35

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 235.15 on 229 degrees of freedom.
The p-value is 0.376

Confirmatory FA

I. Specify the model

II. Run the analysis

III. Examine fit statistics

IV. Examine parameter estimates

V. Compare with alternative models

Creating the model

I. Specify a model matrix of path coefficients

II. Can simplify the procedure by programming

   A. modelmat: a function to generate sem model matrix

   B. uses the modulo command

Model creation

```
modelmat <- function(n=24) {
  mat = matrix(rep(NA, 3*(n*2+2)),ncol=3)
  for (i in 1:n) {
    mat[i,1] <- paste('F', 2- i%%2 ,'-> V',i,sep='')
    mat[i,2] <- i}
    for (i in 1:n) {
      mat[i+n,1]  <- paste('V',i, '<-> V',i,sep ='')
      mat[i+n,2]  <- n + i  }
    colnames(mat) <- c("path","label","initial
estimate")
    mat[n*2+1,1] <- 'F1 <-> F1'
    mat[n*2+2,1] <- 'F2 <-> F2'
    mat[n*2+1,3] <- 1
    mat[n*2+2,3] <- 1
    return(mat) }
```

Simple Structure SEM

```
set.seed(42)
nsub=500
ss.items <- circ.sim
(nvar=24,circum=FALSE,nsub)
colnames(ss.items) <- paste("V",seq
(1:24),sep="")
ss.cov <- cov(ss.items)
model.ss <- modelmat(24)
ss.cov <- cov(ss.items)
sem.ss <- sem(model.ss,ss.cov,nsub)
summary(sem.ss,digits=2)
```

SEM fits: Simple Structure

```
Model Chisquare =  257   Df =  252 Pr(>Chisq) = 0.4
 Chisquare (null model) =  3380   Df =  276
 Goodness-of-fit index =  0.96
 Adjusted goodness-of-fit index =  0.95
 RMSEA index =  0.0064   90% CI: (NA, 0.019)
 Bentler-Bonnett NFI =  0.92
 Tucker-Lewis NNFI =  1
 Bentler CFI =  1
 BIC =  -1309

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.2e+00 -4.3e-01  5.8e-05  8.7e-03  4.7e-01  2.1e+00
```

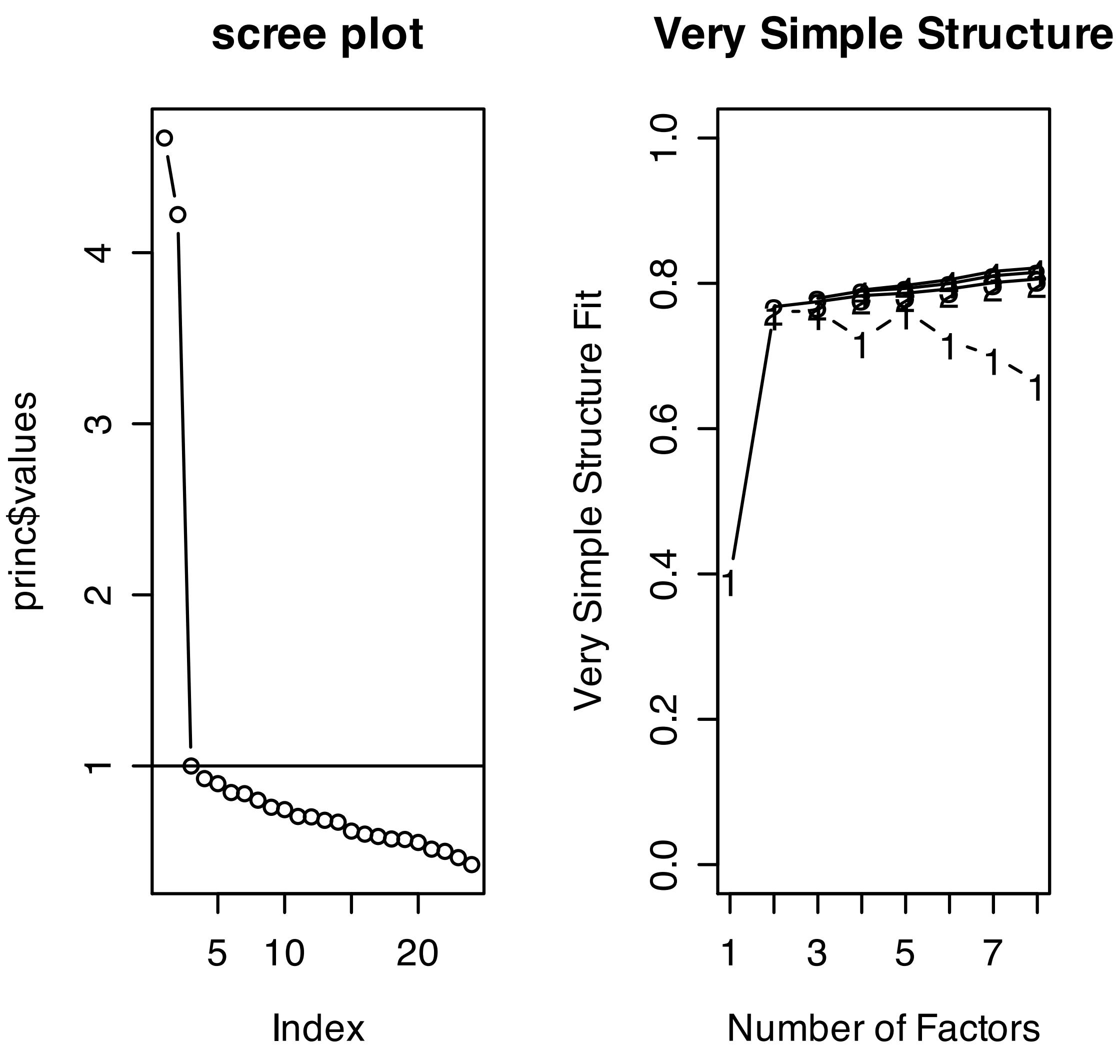
SEM parameters: SS

```
 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1  -0.63    0.043     -15      0           V1 <--- F1
2   0.52    0.042      12      0           V2 <--- F2
3   0.62    0.046      13      0           V3 <--- F1
4  -0.59    0.043     -14      0           V4 <--- F2
5  -0.61    0.042     -14      0           V5 <--- F1
6   0.59    0.046      13      0           V6 <--- F2
...
47  0.65    0.045      14      0           V23 <--> V23
48  0.65    0.045      15      0           V24 <--> V24

 Iterations =  18
```

24 items, 5 categories, SS

```
set.seed(42)
nsub=500
ss.items <- circ.sim
(nvar=24,circum=FALSE,nsub=nsub,low=-2,high=2,categoric
al=TRUE)
colnames(ss.items) <- paste("V",seq(1:24),sep="")
fss <- factanal(ss.items,2)
print(fss,digits=2,cutoff=0)
ss.cov <- cov(ss.items)
model.ss <- modelmat(24)
model.ss[1,2] <- NA
model.ss[1,3] <- 1
model.ss[2,2] <- NA
model.ss[2,3] <- 1
ss.cov <- cov(ss.items)
sem.ss <- sem(model.ss,ss.cov,nsub)
summary(sem.ss,digits=2)
```

Number of factors?



CFA: Fit statistics - 5 point

```
 Model Chisquare =  451   Df =  254 Pr(>Chisq)
= 3.5e-13
 Chisquare (null model) =  2932   Df =  276
 Goodness-of-fit index =  0.94
 Adjusted goodness-of-fit index =  0.92
 RMSEA index =  0.039   90% CI: (0.033, 0.045)
 Bentler-Bonnett NFI =  0.85
 Tucker-Lewis NNFI =  0.92
 Bentler CFI =  0.93
 BIC =  -1128
 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -7.380  -1.580  -0.082   0.037   1.460   6.190
```

Compare with continuous

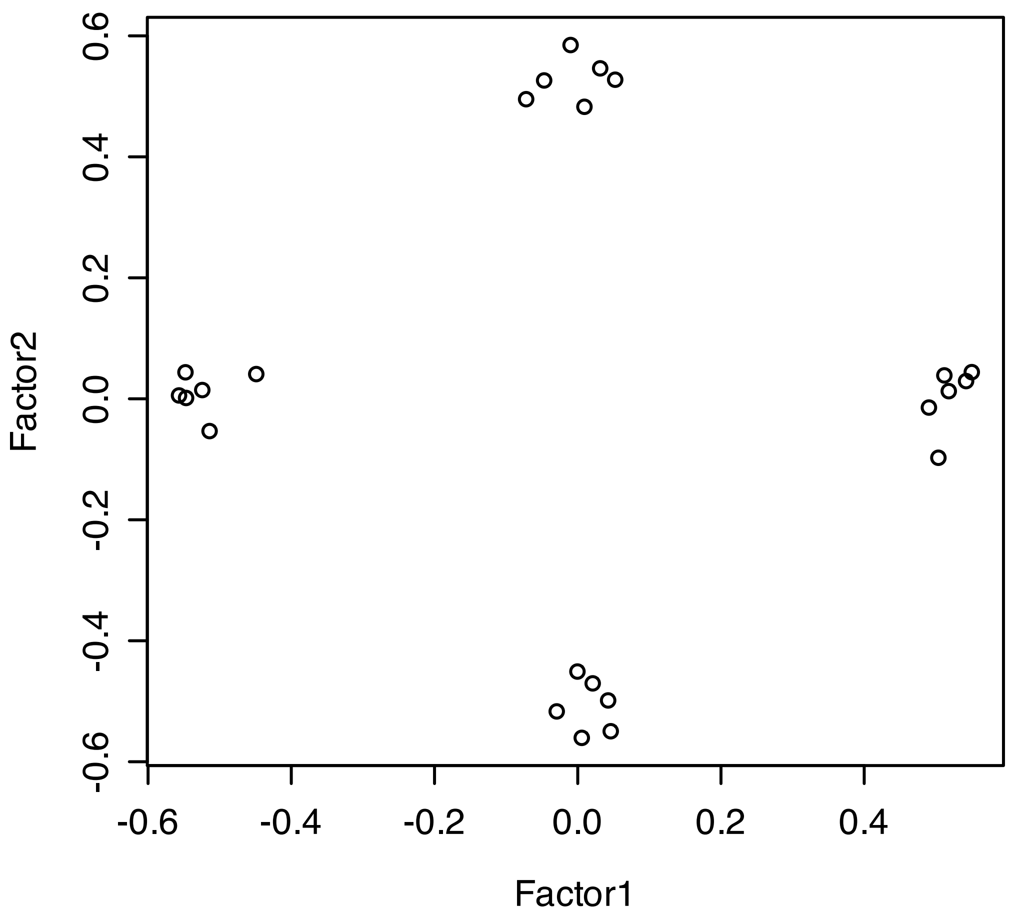
```
Model Chisquare =  257   Df =  252 Pr(>Chisq) = 0.4
 Chisquare (null model) =  3380   Df =  276
 Goodness-of-fit index =  0.96
 Adjusted goodness-of-fit index =  0.95
 RMSEA index =  0.0064   90% CI: (NA, 0.019)
 Bentler-Bonnett NFI =  0.92
 Tucker-Lewis NNFI =  1
 Bentler CFI =  1
 BIC =  -1309

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.2e+00 -4.3e-01  5.8e-05  8.7e-03  4.7e-01  2.1e+00
```

Parameter estimates

```
         Estimate Std Error z value Pr(>|z|)
 3   -0.68     0.055      -12      0          V3 <--- F1
 4   -0.68     0.052      -13      0          V4 <--- F2
 5    0.67     0.049       14      0          V5 <--- F1
 6    0.65     0.055       12      0          V6 <--- F2

47   0.72     0.050       14      0          V23 <--> V23
48   0.69     0.047       15      0          V24 <--> V24

              Iterations =  16
```

24 items, SS, 3 point scales

```
set.seed(42)
nsub=500
ss.items <- circ.sim
(nvar=24,circum=FALSE,nsub=nsub,low=-1,high=1,categoric
al=TRUE)
colnames(ss.items) <- paste("V",seq(1:24),sep="")
fss <- factanal(ss.items,2)
print(fss,digits=2,cutoff=0)
ss.cov <- cov(ss.items)
ss.cor <- cor(ss.items)

print(model.ss,digits=2)
sem.ss <- sem(model.ss,ss.cor,nsub)
summary(sem.ss,digits=2)
```



Factor2
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.6
-0.4
-0.2
0.0
0.2
0.4
Factor1

SS24 - 3 point fits

```
 Model Chisquare =  474   Df =  254 Pr(>Chisq) =
1.9e-15
 Chisquare (null model) =  2400   Df =  276
 Goodness-of-fit index =  0.93
 Adjusted goodness-of-fit index =  0.92
 RMSEA index =  0.042   90% CI: (0.036, 0.047)
 Bentler-Bonnett NFI =  0.8
 Tucker-Lewis NNFI =  0.89
 Bentler CFI =  0.9
 BIC =  -1105
```

SS 24 3 point: parameters

Estimate	Std Error	z value	Pr(>|z|)	
-0.59	0.054	-11.0	0	V3 <--- F1
-0.62	0.055	-11.3	0	V4 <--- F2
0.65	0.053	12.4	0	V5 <--- F1
0.60	0.055	10.9	0	V6 <--- F2
-0.59	0.054	-10.9	0	V7 <--- F1
-0.54	0.056	-9.7	0	V8 <--- F2
0.70	0.049	14.3	0	V23 <--> V23
0.78	0.052	14.8	0	V24 <--> V24

Simple Structure - dichotomous

```
set.seed(42)
nsub=500   #note this is 1000 in chapter
ss.items <- circ.sim
(nvar=24,circum=FALSE,nsub=nsub,low=0,high=
1,categorical=TRUE)
colnames(ss.items) <- paste("V",seq(1:24),sep="")
fss <- factanal(ss.items,2)
print(fss,digits=2,cutoff=0)
ss.cor <- cor(ss.items)
sem.ss <- sem(model.ss,ss.cor,nsub)
summary(sem.ss,digits=2)
```

Dichotomous Fits

Model Chisquare = 463 Df = 254 Pr(>Chisq) = 2.5e-14
Chisquare (null model) = 1519 Df = 276
Goodness-of-fit index = 0.93
Adjusted goodness-of-fit index = 0.92
RMSEA index = 0.041 90% CI: (0.035, 0.046)
Bentler-Bonnett NFI = 0.7
Tucker-Lewis NNFI = 0.82
Bentler CFI = 0.83
BIC = -1116
Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-7.010	-0.902	-0.098	0.113	1.110	5.590

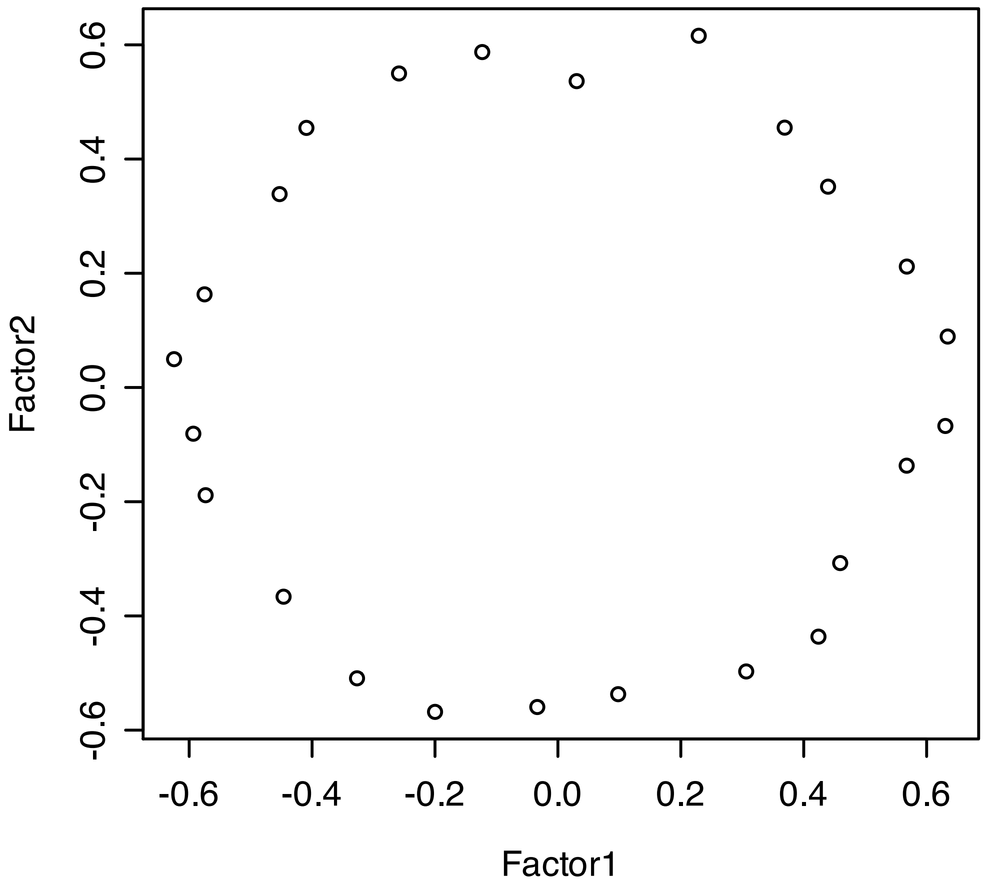
Compare fits

	df	chi^2	RMSEA	NFI	CFI
simple	252	257	.006	.92	.998
5	254	450	.04	.84	.926
3	254	473	.04	.80	.897
2	254	462	.04	.69	.832

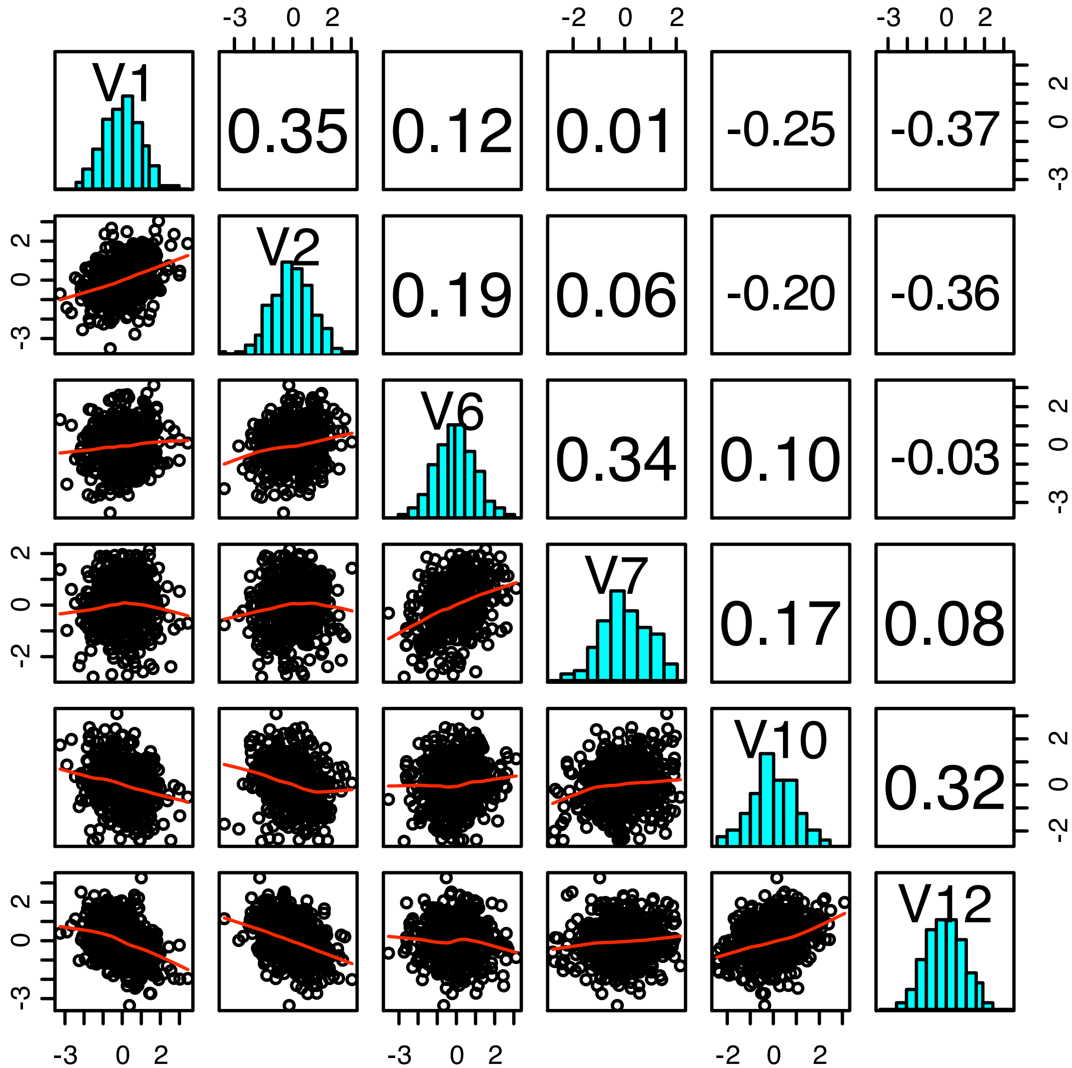
Circumplex Structures

I. Items are multi-vocal (complexity 2 or more)
   A. Typical in emotion data
   B. Interpersonal circumplex (Leary, Wiggins, etc.)
II. Simple structure models do not work

Simulated circumplex



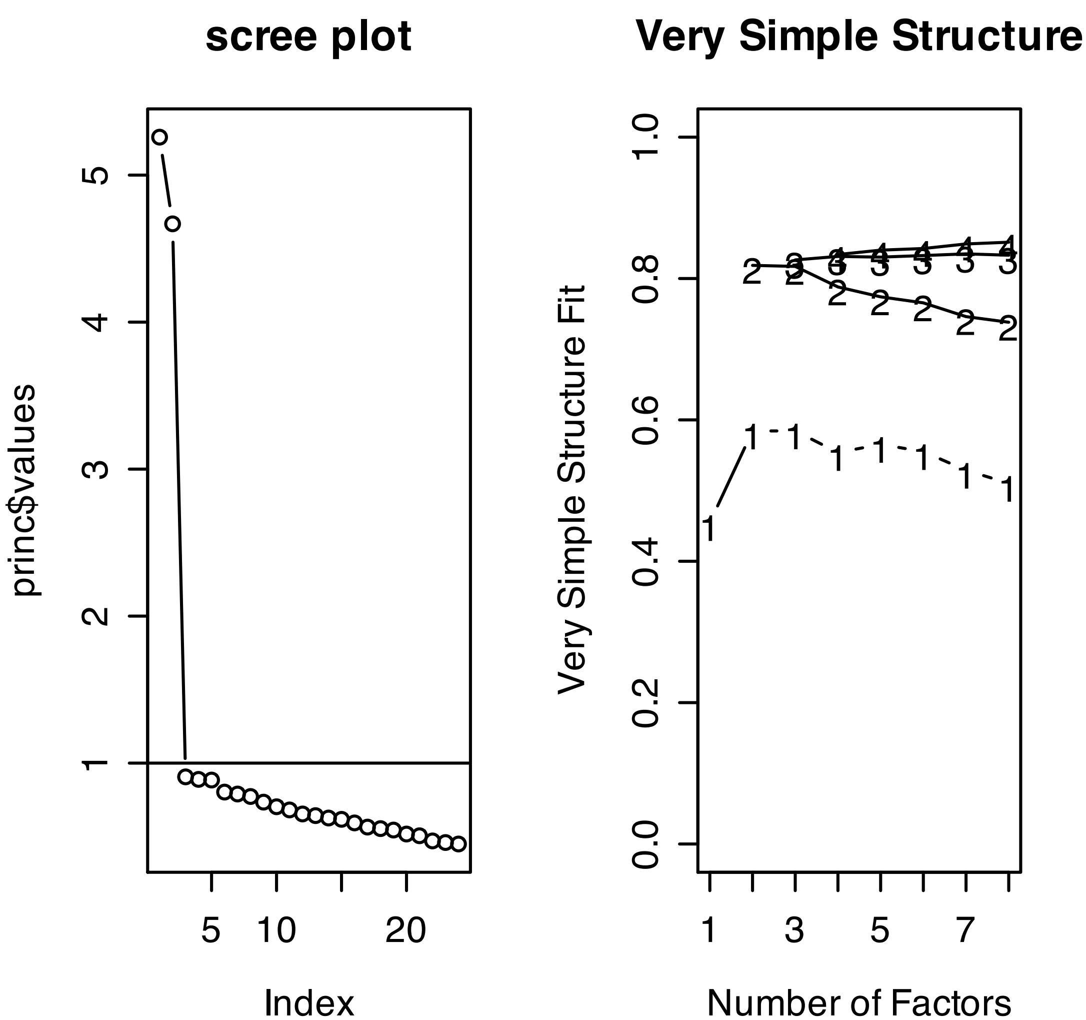
SPLOM of circumplex items



Simulation:

```
set.seed(42)
nsub=500
circ.items <- circ.sim
(nvar=24,circum=TRUE,nsub)
colnames(circ.items) <- paste("V",seq
(1:24),sep="")
fcs <- factanal(circ.items,2)
print(fcs,digits=2,cutoff=0)
```

Number of factors?



Factor loadings

Loadings:

	Factor1	Factor2
V1	-0.62	0.05
V2	-0.57	0.16
V3	-0.45	0.34
V4	-0.41	0.45
V5	-0.26	0.55
V6	-0.12	0.59
V7	0.03	0.54
V8	0.23	0.62
	...	
V23	-0.57	-0.19
V24	-0.59	-0.08

EFA goodness of fit

	Factor1	Factor2
SS loadings	4.52	3.96
Proportion Var	0.19	0.17
Cumulative Var	0.19	0.35

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 224.9 on 229 degrees of freedom.
The p-value is 0.564

Fitting the wrong CFA

```
Model Chisquare =  2297   Df =  252 Pr(>Chisq) = 0
 Chisquare (null model) =  3449   Df =  276
 Goodness-of-fit index =  0.55
 Adjusted goodness-of-fit index =  0.47
 RMSEA index =  0.13   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  0.33
 Tucker-Lewis NNFI =  0.29
 Bentler CFI =  0.36
 BIC =  731

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -8.95   -3.42   -0.10   -0.15    3.57    9.50
```

Parameters from wrong CFA

```
   Estimate Std Error z value Pr(>|z|)
1  -0.650   0.047     -13.78  0.0e+00  V1 <--- F1
2  -0.544   0.050     -10.92  0.0e+00  V2 <--- F2
3  -0.477   0.056      -8.53  0.0e+00  V3 <--- F1
4  -0.387   0.061      -6.36  2.1e-10  V4 <--- F2
5  -0.290   0.056      -5.16  2.4e-07  V5 <--- F1
6  -0.111   0.067      -1.66  9.7e-02  V6 <--- F2
7  -0.027   0.057      -0.47  6.4e-01  V7 <--- F1
8   0.247   0.069       3.59  3.3e-04  V8 <--- F2
9   0.319   0.059       5.40  6.6e-08  V9 <--- F1
10  0.415   0.053       7.89  3.1e-15  V10 <--- F2
11  0.512   0.051      10.05  0.0e+00  V11 <--- F1

 Iterations =  34
>
```

Fitting a better model (but still wrong)

Model Chisquare = 1339 Df = 254 Pr(>Chisq) = 0
Chisquare (null model) = 3449 Df = 276
Goodness-of-fit index = 0.8
Adjusted goodness-of-fit index = 0.77
RMSEA index = 0.093 90% CI: (NA, NA)
Bentler-Bonnett NFI = 0.61
Tucker-Lewis NNFI = 0.63
Bentler CFI = 0.66
BIC = -240

Circumplex structure ≠ Simple structure

	Estimate	Std Error	z value	Pr(>|z|)	
2	0.67	0.051	13.2	0	V2 <--- F1
3	0.70	0.054	12.9	0	V3 <--- F1
4	0.65	0.053	12.2	0	V4 <--- F1
6	0.59	0.060	9.8	0	V6 <--- F2
8	0.85	0.056	15.0	0	V8 <--- F2
9	0.71	0.056	12.7	0	V9 <--- F2
12	-0.64	0.053	-12.0	0	V12 <--- F1
13	-0.76	0.054	-14.1	0	V13 <--- F1
14	-0.67	0.049	-13.6	0	V14 <--- F1
15	-0.64	0.053	-12.0	0	V15 <--- F1
21	-0.77	0.056	-13.8	0	V21 <--- F2
22	-0.66	0.057	-11.6	0	V22 <--- F2
23	-0.50	0.059	-8.5	0	V23 <--- F2
24	0.59	0.053	11.3	0	V24 <--- F1

Now consider skew

I. Although scales are likely (but not guaranteed) to be normal, items are rarely so.

II. Scales with skew include Beck Depression, Eysenck Psychoticism, performance in 405

III. Items with skew reflect anger, depression, strong negative affect

Effect of skew

I. Will reduce the correlation if skews are in opposite directions but will not increase the correlation if in the same direction
II. A strange sort of method variance that only hurts relationships

The challenge of skew

I. Affects the interpretation of bipolarity in affect

A. If affect is bipolar (happy to sad), then a bipolar item will show this, but two unipolar items will not because their correlation is limited

B. Very Sad - somewhat sad, somewhat happy, very happy

C. Not at all sad, somewhat sad, very sad

D. Not at all happy, somewhat happy, very happy

The analogy to temperature

I. How hot is it today (not at all, somewhat, very)

II. How cold is it today (not at all, somewhat, very)

III. What is the temperature today?

   A. Very cold, somewhat cold, somewhat hot, very hot

Effect of skew in SEM

```
skew.items  <- circ.sim
(nvar=24,circum=FALSE,nsub=nsub,truncate=
TRUE,ybias=1,categorical=TRUE)
colnames(skew.items) <- paste("V",seq
(1:24),sep="")
fcs <- factanal(skew.items,2)
print(fcs,digits=2,cutoff=0)
```

EFA - 2 factors

	Factor1	Factor2
SS loadings	2.83	2.39
Proportion Var	0.12	0.10
Cumulative Var	0.12	0.22

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 288.48 on 229 degrees of freedom.
The p-value is 0.00466

Loadings are smaller

	Factor1	Factor2
V1	-0.45	0.04
V2	0.02	0.51
V3	0.55	-0.03
V6	-0.07	0.48
V7	0.48	0.04
V8	-0.01	-0.33
V11	0.48	-0.05
V12	0.03	-0.32
V22	0.05	0.59
V23	0.45	-0.04
V24	-0.02	-0.34

But CFA does badly

```
 Model Chisquare =  703   Df =  254 Pr(>Chisq)
= 0
 Chisquare (null model) =  1786   Df =  276
 Goodness-of-fit index =  0.91
 Adjusted goodness-of-fit index =  0.9
 RMSEA index =  0.06   90% CI: (0.054, 0.065)
 Bentler-Bonnett NFI =  0.61
 Tucker-Lewis NNFI =  0.68
 Bentler CFI =  0.7
 BIC =  -876
```

CFA parameters

3	-0.513	0.0441	-11.6	0.0e+00	V3 <--- F1
4	-0.126	0.0172	-7.3	2.6e-13	V4 <--- F2
5	0.444	0.0439	10.1	0.0e+00	V5 <--- F1
6	0.505	0.0514	9.8	0.0e+00	V6 <--- F2
7	-0.447	0.0452	-9.9	0.0e+00	V7 <--- F1
8	-0.100	0.0153	-6.5	6.1e-11	V8 <--- F2
9	0.446	0.0431	10.3	0.0e+00	V9 <--- F1
22	0.660	0.0518	12.7	0.0e+00	V22 <--- F2
23	-0.401	0.0438	-9.1	0.0e+00	V23 <--- F1
24	-0.120	0.0176	-6.8	8.5e-12	V24 <--- F2

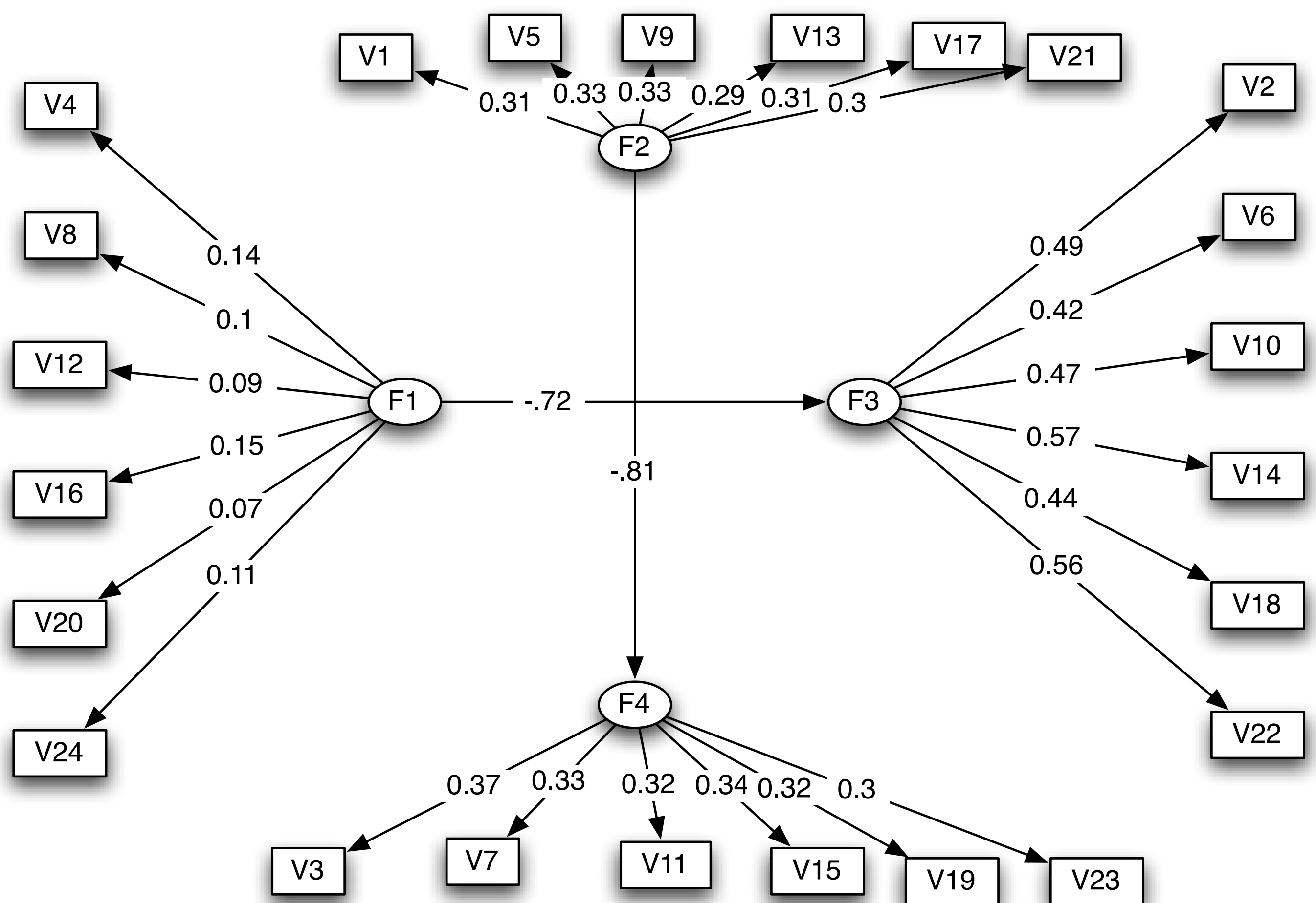
But, if we allow for 4 correlated factors, it works

Model Chisquare = 261 Df = 250 Pr(>Chisq) = 0.30
Chisquare (null model) = 1786 Df = 276
Goodness-of-fit index = 0.96
Adjusted goodness-of-fit index = 0.95
RMSEA index = 0.0094 90% CI: (NA, 0.021)
Bentler-Bonnett NFI = 0.85
Tucker-Lewis NNFI = 1
Bentler CFI = 1
BIC = -1293

2 sets of correlated factors

```
   Estimate Std Error z value Pr(>|z|)
1    0.48     0.049        9.7   0.0e+00  V1 <--- F2
2    0.52     0.049       10.8   0.0e+00  V2 <--- F3
3    0.57     0.048       12.0   0.0e+00  V3 <--- F4
4    0.49     0.055        8.8   0.0e+00  V4 <--- F1
5    0.52     0.049       10.5   0.0e+00  V5 <--- F2
6    0.48     0.049        9.7   0.0e+00  V6 <--- F3
7    0.52     0.049       10.6   0.0e+00  V7 <--- F4
8    0.39     0.055        7.1   9.9e-13  V8 <--- F1
9    0.53     0.049       10.9   0.0e+00  V9 <--- F2
...
49 -0.72     0.055      -13.1   0.0e+00  F3 <--> F1
50 -0.81     0.041      -19.6   0.0e+00  F4 <--> F2
```

Graphical representation



Can also form HICs

```
make.keys <- function(nvar=24,scales=8) {
  keys <- matrix(rep(0,scales * nvar),ncol=scales)
 for (i in 1:nvar) {
  keys[i,i %% scales + 1] <- 1}
  return(keys)
 }
keys <- make.keys()
print(keys)
clusters <- cluster.loadings(keys,skew.cor)
print(clusters,digits=2)
```

Correlation matrix of Homogeneous item Composites, HICs

```
$corrected
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]
[1,]  0.35 -0.02 -0.81  0.06  1.48  0.05 -0.82  0.01
[2,] -0.01  0.49  0.02 -0.82 -0.04  1.09  0.08 -0.73
[3,] -0.34  0.01  0.51 -0.02 -0.94 -0.05  1.04 -0.06
[4,]  0.03 -0.43 -0.01  0.55  0.09 -0.86 -0.04  0.85
[5,]  0.43 -0.01 -0.33  0.03  0.24 -0.11 -0.95  0.08
[6,]  0.02  0.53 -0.02 -0.44 -0.04  0.48  0.01 -0.77
[7,] -0.36  0.04  0.56 -0.02 -0.35  0.00  0.57 -0.05
[8,]  0.01 -0.38 -0.03  0.47  0.03 -0.39 -0.03  0.55
```

alpha on diagonal, corrected above diagonal

Fit the HIC correlations

```
m8 <- modelmat4(8)
sem8 <- sem(m8,clusters$cor,nsub)
summary(sem8,digits=2)
```

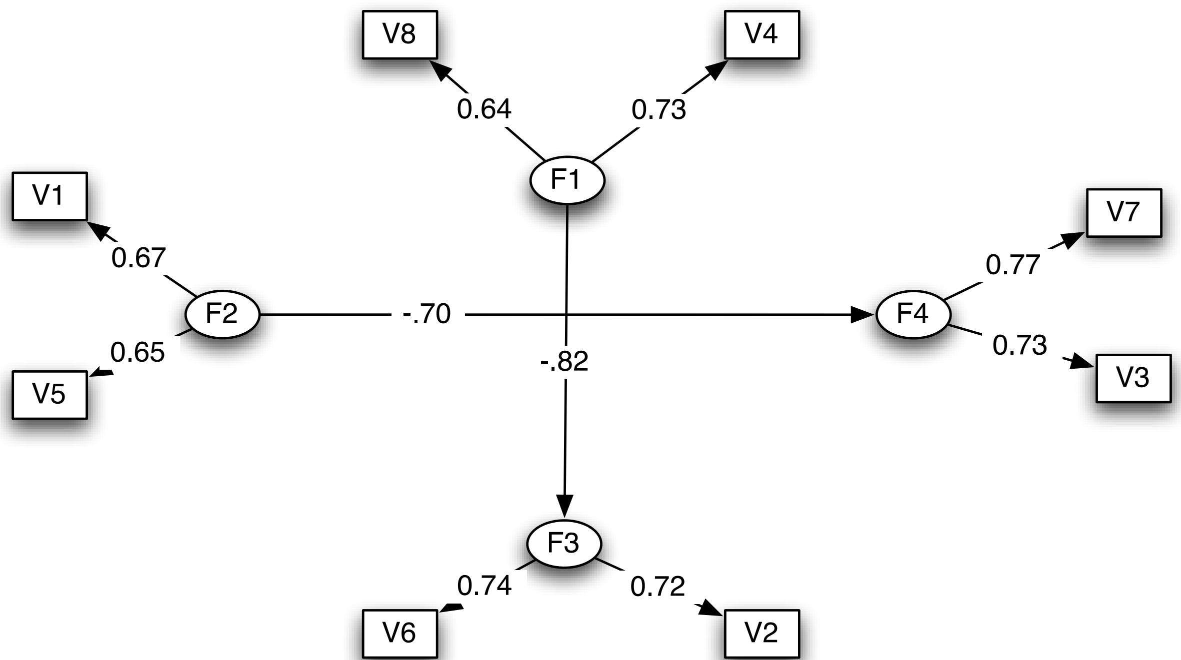
SEM Fits

Model Chisquare = 5.2 Df = 18 Pr(>Chisq) = 1
Chisquare (null model) = 884 Df = 28
Goodness-of-fit index = 1
Adjusted goodness-of-fit index = 1
RMSEA index = 0 90% CI: (NA, NA)
Bentler-Bonnett NFI = 1
Tucker-Lewis NNFI = 1.0
Bentler CFI = 1
BIC = -107

SEM of HICs, parameters

```
     Estimate Std Error z value Pr(>|z|)
 1   0.67    0.055      12.1   0.0e+00  V1 <--- F2
 2   0.72    0.048      15.0   0.0e+00  V2 <--- F3
 3   0.73    0.051      14.2   0.0e+00  V3 <--- F4
 4   0.73    0.050      14.6   0.0e+00  V4 <--- F1
 5   0.65    0.054      11.9   0.0e+00  V5 <--- F2
 6   0.74    0.048      15.4   0.0e+00  V6 <--- F3
 7   0.77    0.052      14.8   0.0e+00  V7 <--- F4
 8   0.64    0.049      13.2   0.0e+00  V8 <--- F1
                          ...
17 -0.82     0.046     -18.0   0.0e+00  F3 <--> F1
18 -0.70     0.052     -13.5   0.0e+00  F4 <--> F2
```

Structural model of HICS



Simple structure - skewed