Analyzing scales

Examples of simulation

Steps in simulation

- Create pure, continuous, population model
- Sample from population model
- Distort sample with categorical items
- Distort sample with dichotomous item
- Distort sample with dichotomous items varying in difficulty

The pure model

```
> latent <- sim.congeneric(rep(.6,9),N=200,short=FALSE)
> latent

$model (Population correlation matrix)
     V1   V2   V3   V4   V5   V6   V7   V8   V9
V1 1.00 0.36 0.36 0.36 0.36 0.36 0.36 0.36 0.36
V2 0.36 1.00 0.36 0.36 0.36 0.36 0.36 0.36 0.36
V3 0.36 0.36 1.00 0.36 0.36 0.36 0.36 0.36 0.36
V4 0.36 0.36 0.36 1.00 0.36 0.36 0.36 0.36 0.36
V5 0.36 0.36 0.36 0.36 1.00 0.36 0.36 0.36 0.36
V6 0.36 0.36 0.36 0.36 0.36 1.00 0.36 0.36 0.36
V7 0.36 0.36 0.36 0.36 0.36 0.36 1.00 0.36 0.36
V8 0.36 0.36 0.36 0.36 0.36 0.36 0.36 1.00 0.36
V9 0.36 0.36 0.36 0.36 0.36 0.36 0.36 0.36 1.00
```

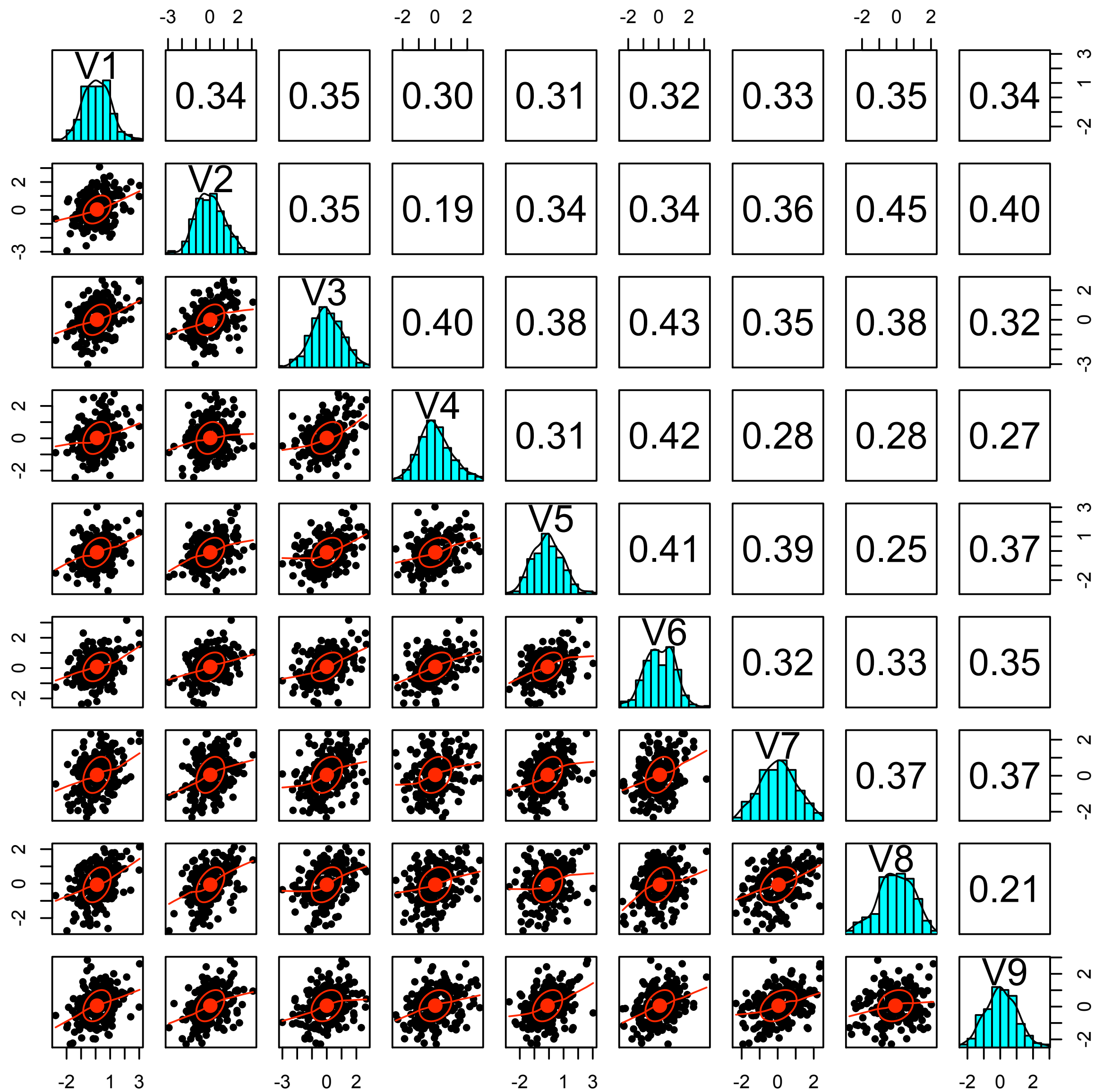
Sample of model

$r (Sample correlation matrix for sample size = 200)

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.34	0.35	0.30	0.31	0.32	0.33	0.35	0.34
V2	0.34	1.00	0.35	0.19	0.34	0.34	0.36	0.45	0.40
V3	0.35	0.35	1.00	0.40	0.38	0.43	0.35	0.38	0.32
V4	0.30	0.19	0.40	1.00	0.31	0.42	0.28	0.28	0.27
V5	0.31	0.34	0.38	0.31	1.00	0.41	0.39	0.25	0.37
V6	0.32	0.34	0.43	0.42	0.41	1.00	0.32	0.33	0.35
V7	0.33	0.36	0.35	0.28	0.39	0.32	1.00	0.37	0.37
V8	0.35	0.45	0.38	0.28	0.25	0.33	0.37	1.00	0.21
V9	0.34	0.40	0.32	0.27	0.37	0.35	0.37	0.21	1.00

```
pairs.panels(latent$observed)
```

SPLOM

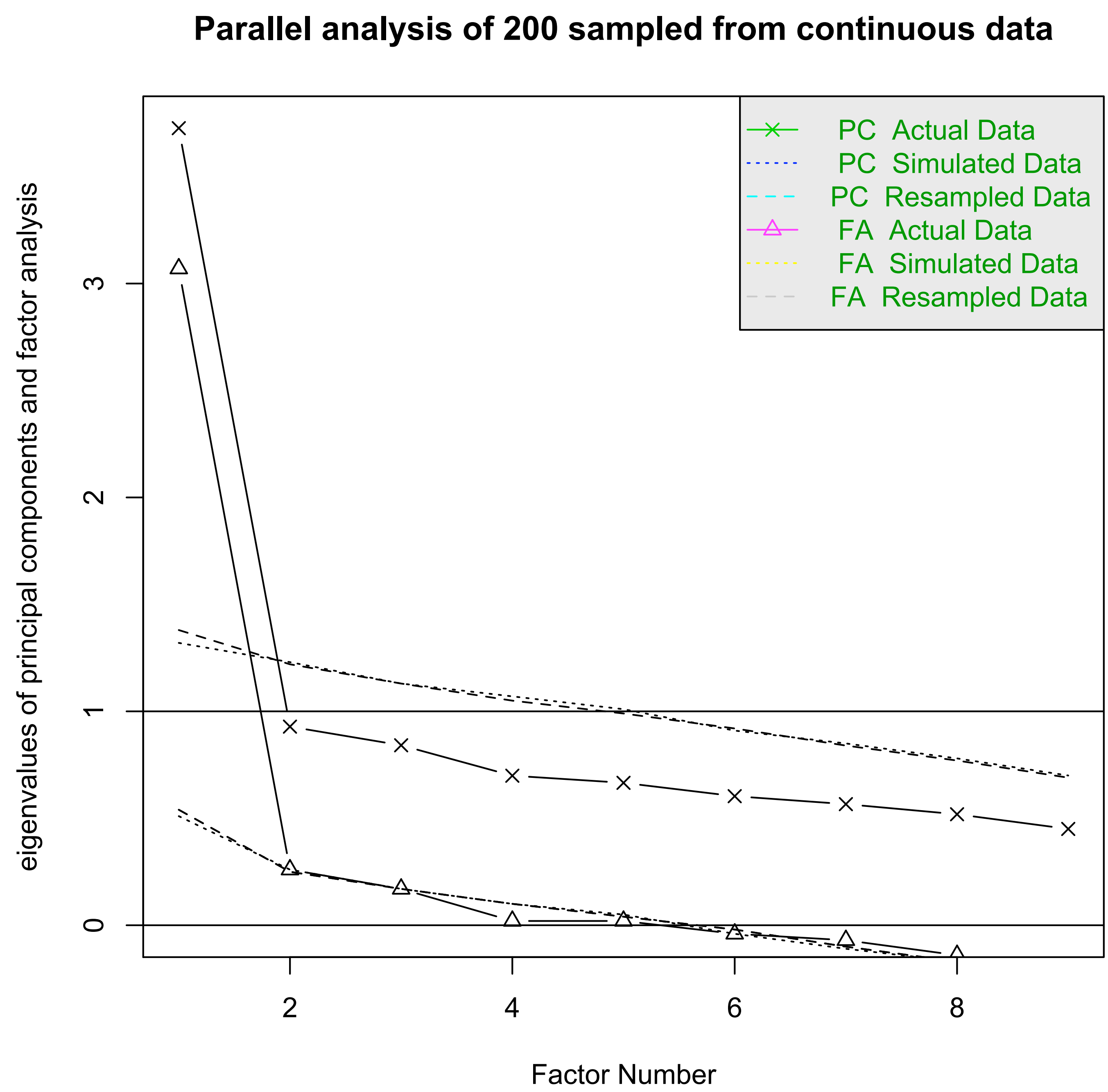


Number of factors?

```
> vss <- VSS(latent$r,SMC=FALSE,n.obs=200)
> vss
Very Simple Structure
VSS complexity 1 achieves a maximimum of 0.77  with  1
factors
VSS complexity 2 achieves a maximimum of 0.8  with  2
factors

The Velicer MAP criterion achieves a minimum of 1  with  1
factors

Velicer MAP
[1] 0.02 0.04 0.07 0.12 0.17 0.28 0.48 1.00

Very Simple Structure Complexity 1
[1] 0.77 0.48 0.45 0.44 0.43 0.44 0.44 0.48
Very Simple Structure Complexity 2
[1] 0.00 0.80 0.75 0.74 0.69 0.67 0.65 0.66
```

Chi^2

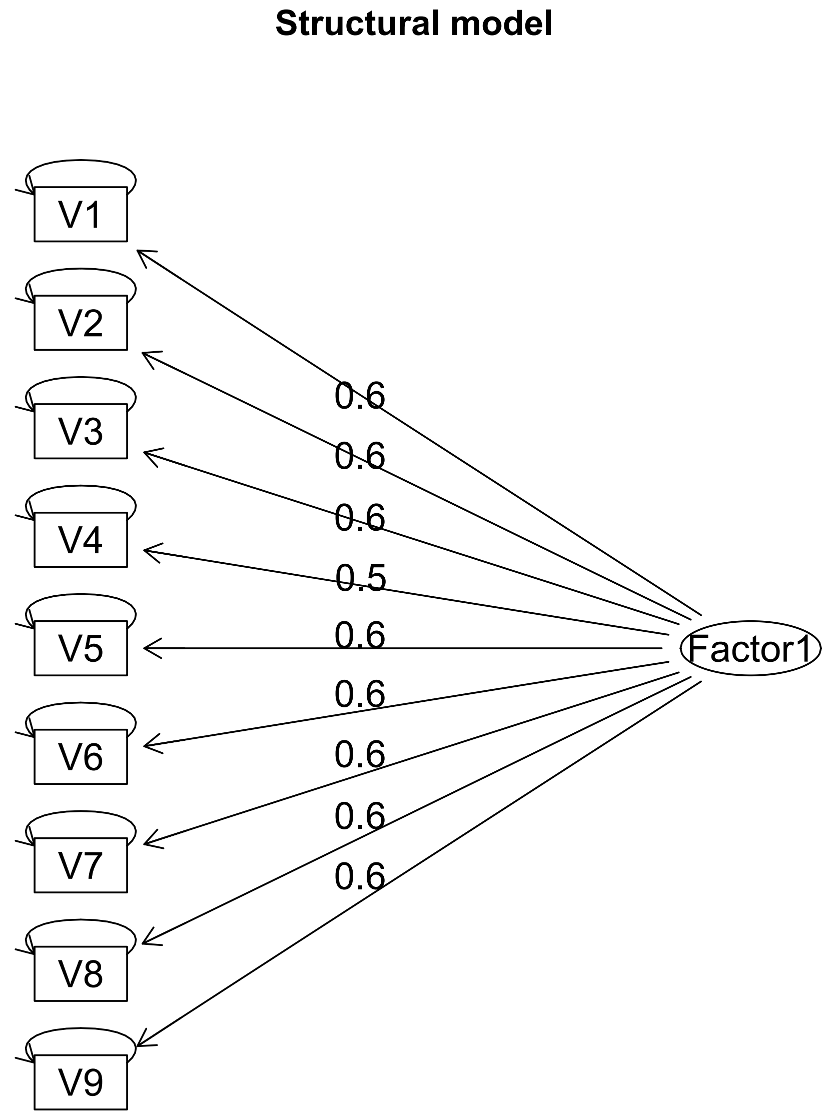
	dof	chisq	prob
1	27	3.757936e+01	0.08475115
2	19	1.880069e+01	0.46968761
3	12	4.713637e+00	0.96686629
4	6	3.818237e+00	0.70125805
5	1	6.672074e-01	0.41402694
6	-3	2.398666e-01	NA
7	-6	6.638422e-02	NA
8	-8	1.026952e-04	NA


Parallel analysis of 200 sampled from continuous data
PC Actual Data
PC Simulated Data
PC Resampled Data
FA Actual Data
FA Simulated Data
FA Resampled Data
eigenvalues of principal components and factor analysis
0
1
2
3
2
4
6
8
Factor Number

Factor analysis

```
> f1 <- factanal(latent$observed,1)
factanal(x = latent$observed, factors = 1)
Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8    V9 
0.688 0.646 0.596 0.726 0.644 0.608 0.653 0.684 0.684 
Loadings:
   Factor1
V1 0.559  
V2 0.595  
V3 0.635  
V4 0.524  
V5 0.597  
V6 0.626  
V7 0.589  
V8 0.562  
V9 0.563  
               Factor1
SS loadings      3.072
Proportion Var   0.341
Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 37.57 on 27 degrees of freedom.
The p-value is 0.0849
```

Structural model




```
mod.1 <- structure.graph(f1)   mod.1 <- structure.sem(f1)
```

```
> mod.1
      Path                     Parameter Value
 [1,] "Factor1->V1"            "V1"      NA
 [2,] "Factor1->V2"            "V2"      NA
 [3,] "Factor1->V3"            "V3"      NA
 [4,] "Factor1->V4"            "V4"      NA
 [5,] "Factor1->V5"            "V5"      NA
 [6,] "Factor1->V6"            "V6"      NA
 [7,] "Factor1->V7"            "V7"      NA
 [8,] "Factor1->V8"            "V8"      NA
 [9,] "Factor1->V9"            "V9"      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,] "Factor1<->Factor1" NA            "1"
```

sem model

sem summary

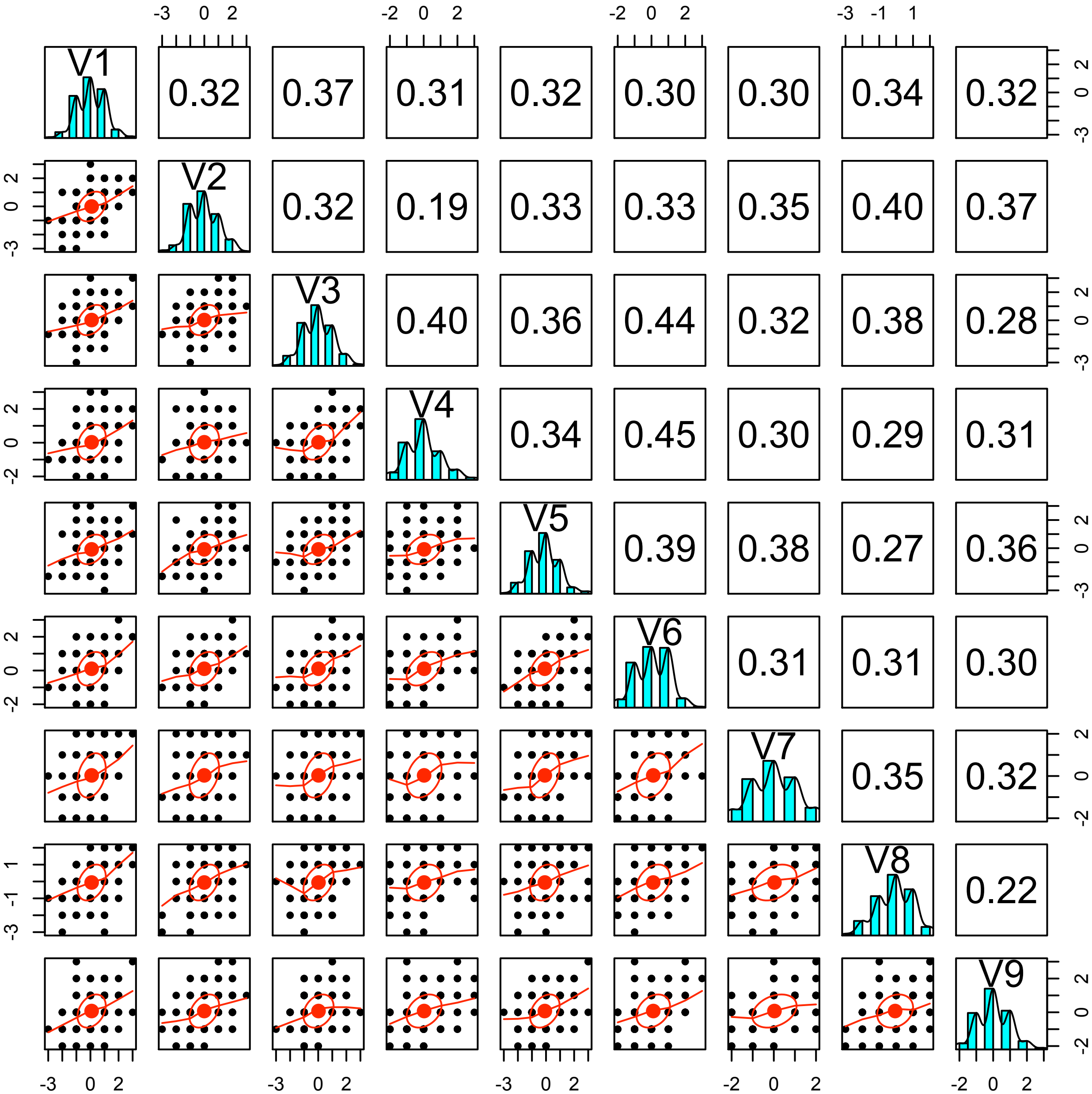
```
> summary(sem.1)

 Model Chisquare =  38.441   Df =  27 Pr(>Chisq) = 0.071153
 Chisquare (null model) =  442.57   Df =  36
 Goodness-of-fit index =  0.95875
 Adjusted goodness-of-fit index =  0.93125
 RMSEA index =  0.046145   90% CI: (NA, 0.077156)
 Bentler-Bonnett NFI =  0.91314
 Tucker-Lewis NNFI =  0.96248
 Bentler CFI =  0.97186
 SRMR =  0.043651
 BIC =  -104.61
```

Parameters - 1 factor

```
 Parameter Estimates
    Estimate Std Error z value Pr(>|z|)
V1  0.55887  0.071501  7.8162  5.5511e-15 V1 <--- Factor1
V2  0.59536  0.070960  8.3901  0.0000e+00 V2 <--- Factor1
V3  0.63541  0.069606  9.1286  0.0000e+00 V3 <--- Factor1
V4  0.52357  0.072652  7.2066  5.7376e-13 V4 <--- Factor1
V5  0.59670  0.070646  8.4463  0.0000e+00 V5 <--- Factor1
V6  0.62646  0.069951  8.9556  0.0000e+00 V6 <--- Factor1
V7  0.58882  0.070844  8.3115  0.0000e+00 V7 <--- Factor1
V8  0.56170  0.071800  7.8232  5.1070e-15 V8 <--- Factor1
V9  0.56253  0.071583  7.8584  3.9968e-15 V9 <--- Factor1
x1e 0.68767  0.076447  8.9954  0.0000e+00 V1 <--> V1
x2e 0.64554  0.073940  8.7307  0.0000e+00 V2 <--> V2
x3e 0.59625  0.070155  8.4991  0.0000e+00 V3 <--> V3
x4e 0.72588  0.079565  9.1231  0.0000e+00 V4 <--> V4
x5e 0.64395  0.073447  8.7675  0.0000e+00 V5 <--> V5
x6e 0.60755  0.071076  8.5479  0.0000e+00 V6 <--> V6
x7e 0.65329  0.074098  8.8166  0.0000e+00 V7 <--> V7
x8e 0.68449  0.076652  8.9299  0.0000e+00 V8 <--> V8
x9e 0.68356  0.076353  8.9527  0.0000e+00 V9 <--> V9
```

Distort by making categorical

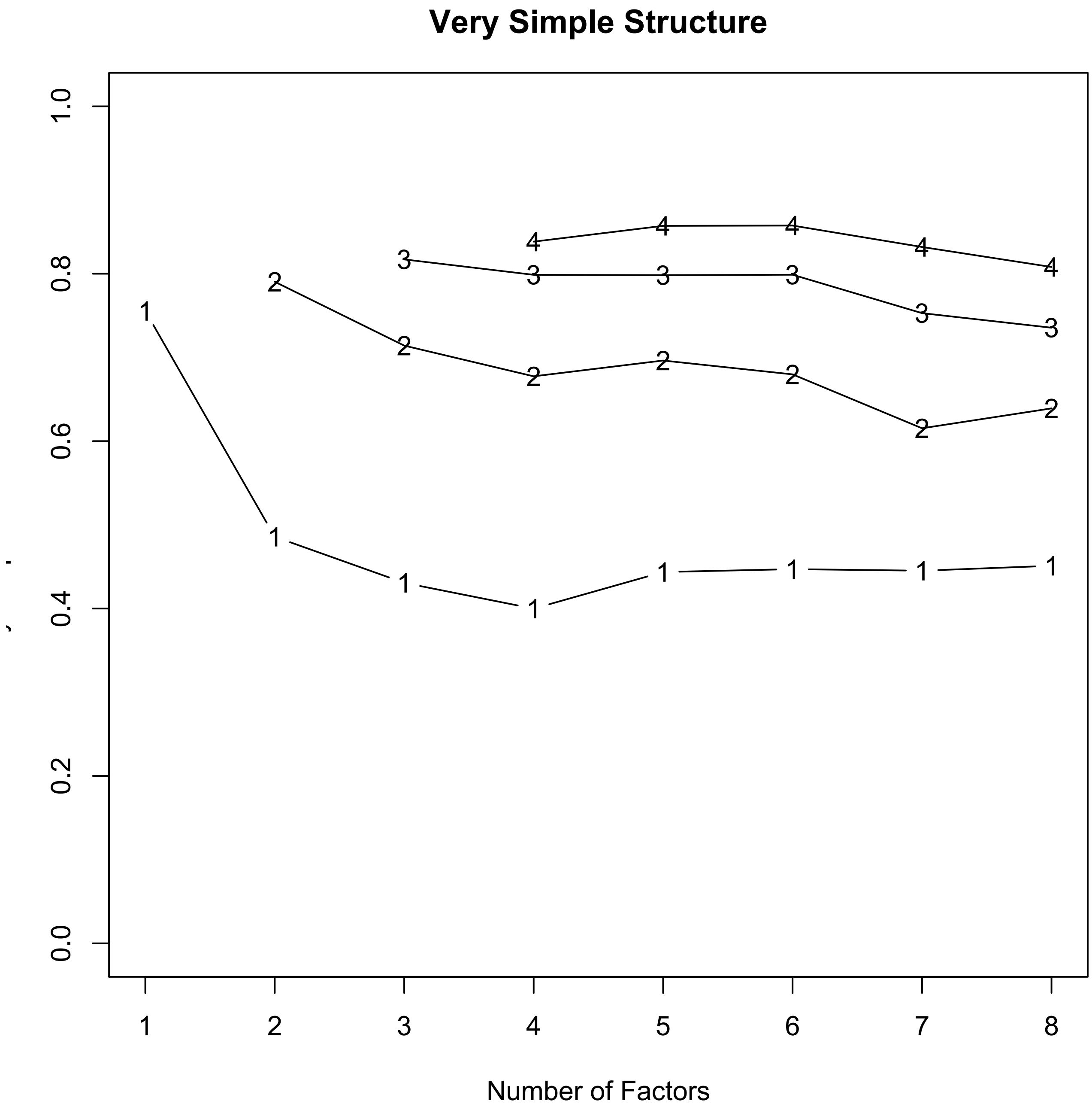
```
> cate <- round(latent$observed)
> pairs.panels(cate)
```



V1
0.32
0.37
0.31
0.32
0.30
0.30
0.34
0.32
V2
0.32
0.19
0.33
0.33
0.35
0.40
0.37
V3
0.40
0.36
0.44
0.32
0.38
0.28
V4
0.34
0.45
0.30
0.29
0.31
V5
0.39
0.38
0.27
0.36
V6
0.31
0.31
0.30
V7
0.35
0.32
V8
0.22
V9

Differences in r by rounding

```
> round(cor(latent$observed) - cor(cate),2)
      V1   V2    V3    V4    V5    V6    V7    V8    V9
V1  0.00 0.02 -0.02 -0.01 -0.01  0.01  0.03  0.01  0.02
V2  0.02 0.00  0.03  0.00  0.01  0.01  0.01  0.06  0.03
V3 -0.02 0.03  0.00  0.00  0.02 -0.01  0.03  0.00  0.04
V4 -0.01 0.00  0.00  0.00 -0.02 -0.03 -0.02 -0.01 -0.04
V5 -0.01 0.01  0.02 -0.02  0.00  0.02  0.01 -0.01  0.02
V6  0.01 0.01 -0.01 -0.03  0.02  0.00  0.00  0.02  0.05
V7  0.03 0.01  0.03 -0.02  0.01  0.00  0.00  0.02  0.05
V8  0.01 0.06  0.00 -0.01 -0.01  0.02  0.02  0.00  0.00
V9  0.02 0.03  0.04 -0.04  0.02  0.05  0.05  0.00  0.00
```



VSS

```
> vss.c

Very Simple Structure
VSS complexity 1 achieves a maximimum of 0.76  with  1
factors
VSS complexity 2 achieves a maximimum of 0.79  with  2
factors
The Velicer MAP criterion achieves a minimum of 1  with  1
factors
 Velicer MAP
[1] 0.02 0.04 0.07 0.11 0.17 0.27 0.44 1.00

Very Simple Structure Complexity 1
[1] 0.76 0.49 0.43 0.40 0.44 0.45 0.45 0.45
Very Simple Structure Complexity 2
[1] 0.00 0.79 0.71 0.68 0.70 0.68 0.62 0.64
```

Chi^2

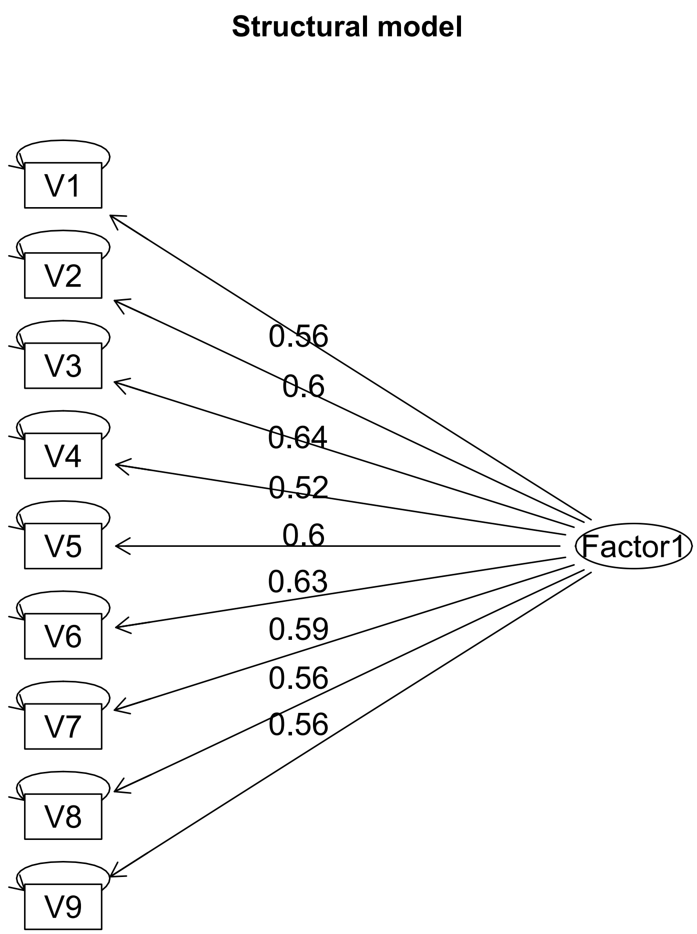
```
> vss.c$vss.stats[,1:3]
  dof        chisq      prob
1  27 3.198998e+01 0.2324556
2  19 1.490603e+01 0.7285507
3  12 6.314856e+00 0.8993859
4   6 4.911160e+00 0.5552571
5   1 1.806164e+00 0.1789690
6  -3 1.377620e+00        NA
7  -6 8.126361e-03        NA
8  -8 2.158884e-04        NA
```

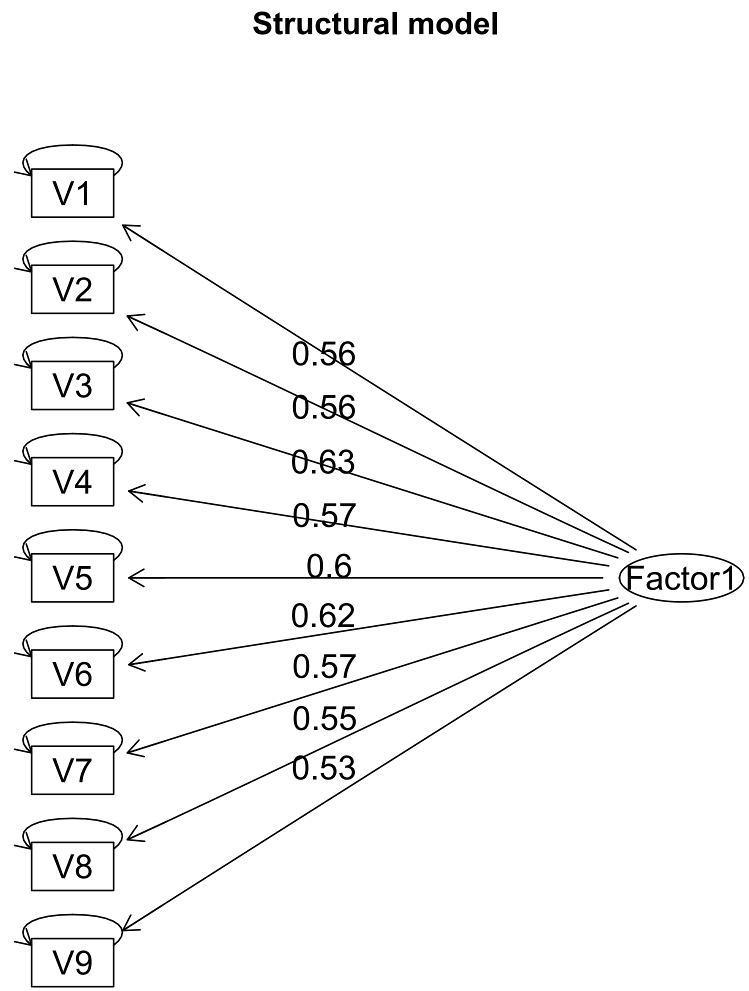
factanal

```
Call:
factanal(x = cate, factors = 1)
Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8    V9
0.691 0.687 0.603 0.680 0.643 0.611 0.679 0.696 0.721
Loadings:
   Factor1
V1 0.556
V2 0.559
V3 0.630
V4 0.565
V5 0.598
V6 0.623
V7 0.567
V8 0.551
V9 0.528
               Factor1
SS loadings      2.989
Proportion Var   0.332
Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 31.97 on 27 degrees of freedom
The p-value is 0.233
```

Compare the structures

```
> mod.f1 <- structure.graph(f1,digits=2)
```






```
mod.c1 <- structure.graph(c1,digits=2)
```

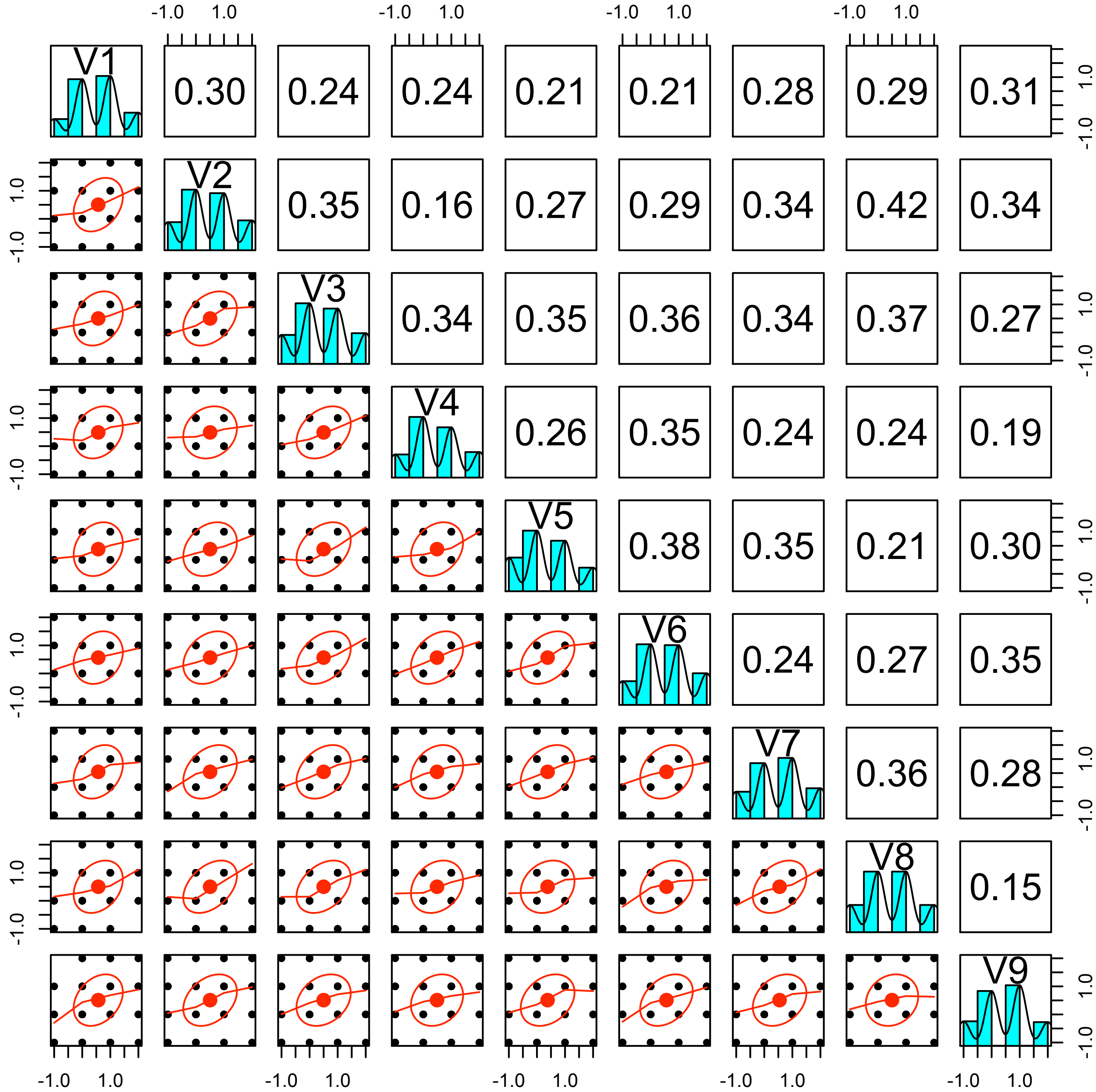
Sem of categorical data

```
> sem.c1 <- sem(mod.c1,cor(cate),200)
> summary(sem.c1)

 Model Chisquare =  32.711   Df =  27 Pr(>Chisq) = 0.20678
 Chisquare (null model) =  418.48   Df =  36
 Goodness-of-fit index =  0.9643
 Adjusted goodness-of-fit index =  0.94049
 RMSEA index =  0.032602   90% CI: (NA, 0.067294)
 Bentler-Bonnett NFI =  0.92183
 Tucker-Lewis NNFI =  0.9801
 Bentler CFI =  0.98507
 SRMR =  0.040899
 BIC =  -110.34
```

Make 4 categories

```
> cat4 <- latent$observed
> cat4[cat4< -1  ] <- -1
> cat4[(cat4<0)&cat4>-1] <- 0
> cat4[cat4>1] <- 2
> cat4[(cat4< 1 ) & (cat4>0)] <- 1
```

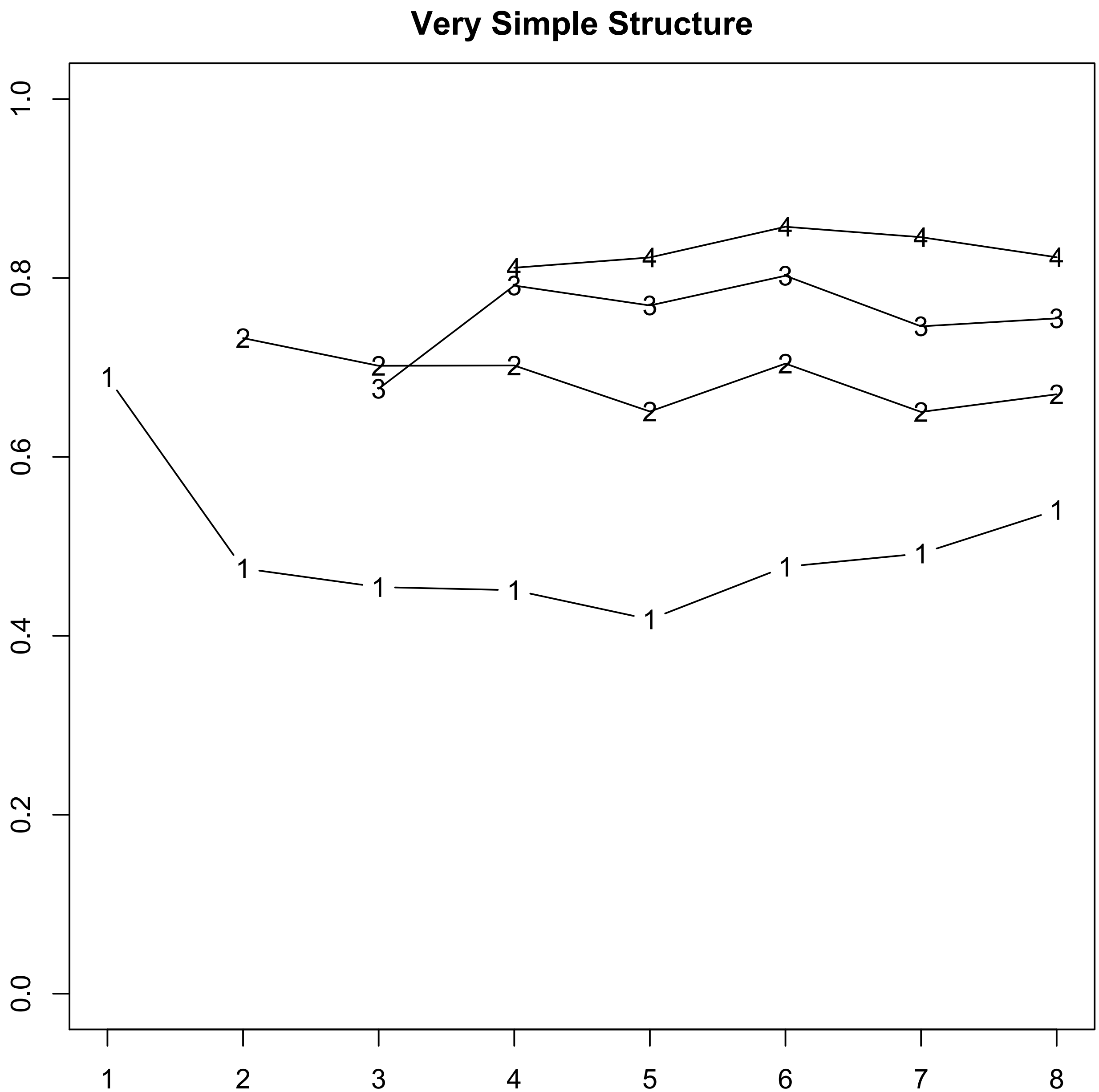
Splom



V1
V2
V3
V4
V5
V6
V7
V8
V9
0.30 0.24 0.24 0.21 0.21 0.28 0.29 0.31
0.35 0.16 0.27 0.29 0.34 0.42 0.34
0.34 0.35 0.36 0.34 0.37 0.27
0.26 0.35 0.24 0.24 0.19
0.38 0.35 0.21 0.30
0.24 0.27 0.35
0.36 0.28
0.15
-1.0
1.0

4 categories reduce r

```
> round(cor(latent$observed) - cor(cat4),2)
     V1   V2   V3   V4   V5   V6   V7   V8   V9
V1 0.00 0.04 0.10 0.07 0.09 0.11 0.06 0.06 0.03
V2 0.04 0.00 0.00 0.03 0.07 0.05 0.02 0.04 0.06
V3 0.10 0.00 0.00 0.06 0.03 0.07 0.01 0.01 0.04
V4 0.07 0.03 0.06 0.00 0.06 0.07 0.04 0.04 0.08
V5 0.09 0.07 0.03 0.06 0.00 0.03 0.04 0.05 0.07
V6 0.11 0.05 0.07 0.07 0.03 0.00 0.08 0.06 0.00
V7 0.06 0.02 0.01 0.04 0.04 0.08 0.00 0.01 0.09
V8 0.06 0.04 0.01 0.04 0.05 0.06 0.01 0.00 0.06
V9 0.03 0.06 0.04 0.08 0.07 0.00 0.09 0.06 0.00
```



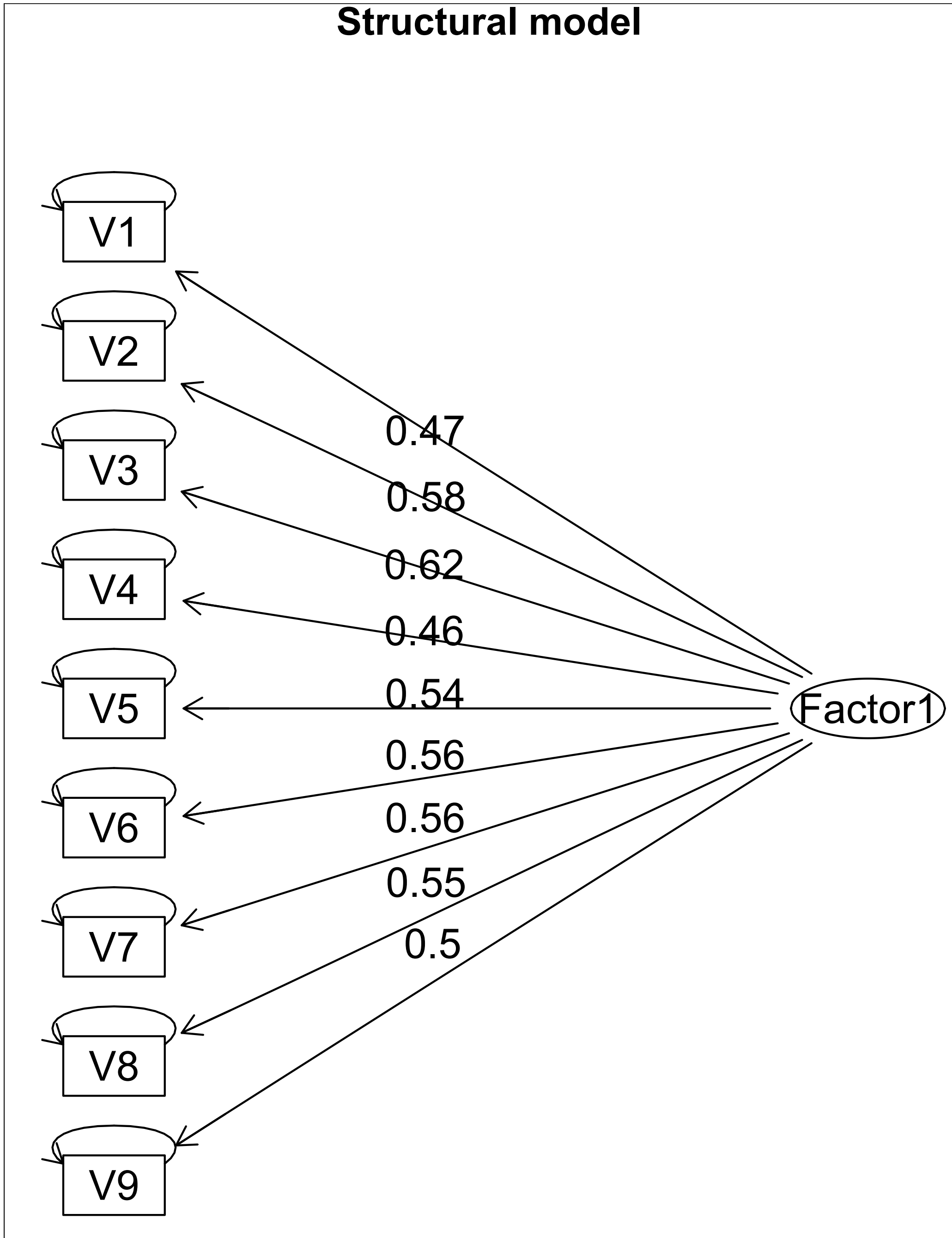
VSS

```
> vss.4
Very Simple Structure
VSS complexity 1 achieves a maximimum of 0.69  with  1
factors
VSS complexity 2 achieves a maximimum of 0.73  with  2
factors

The Velicer MAP criterion achieves a minimum of 1  with  1
factors
 Velicer MAP
[1] 0.02 0.04 0.07 0.11 0.17 0.26 0.50 1.00

Very Simple Structure Complexity 1
[1] 0.69 0.48 0.45 0.45 0.42 0.48 0.49 0.54
Very Simple Structure Complexity 2
[1] 0.00 0.73 0.70 0.70 0.65 0.70 0.65 0.67
```

```
> vss$vss.stats[,1:3]
  dof        chisq       prob
1  27 3.757936e+01 0.08475115
2  19 1.880069e+01 0.46968761
3  12 4.713637e+00 0.96686629
4   6 3.818237e+00 0.70125805
5   1 6.672074e-01 0.41402694
6  -3 2.398666e-01         NA
7  -6 6.638422e-02         NA
8  -8 1.026952e-04         NA
```

factanal

```
> f4 <- factanal(cat4,1)
factanal(x = cat4, factors = 1)
Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8    V9
0.780 0.662 0.616 0.787 0.705 0.684 0.682 0.703 0.748
Loadings:
   Factor1
V1 0.469
V2 0.581
V3 0.619
V4 0.461
V5 0.543
V6 0.562
V7 0.564
V8 0.545
V9 0.502
               Factor1
SS loadings      2.632
Proportion Var   0.292
Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 40.94 on 27 degrees of freedom.
The p-value is 0.0418
```



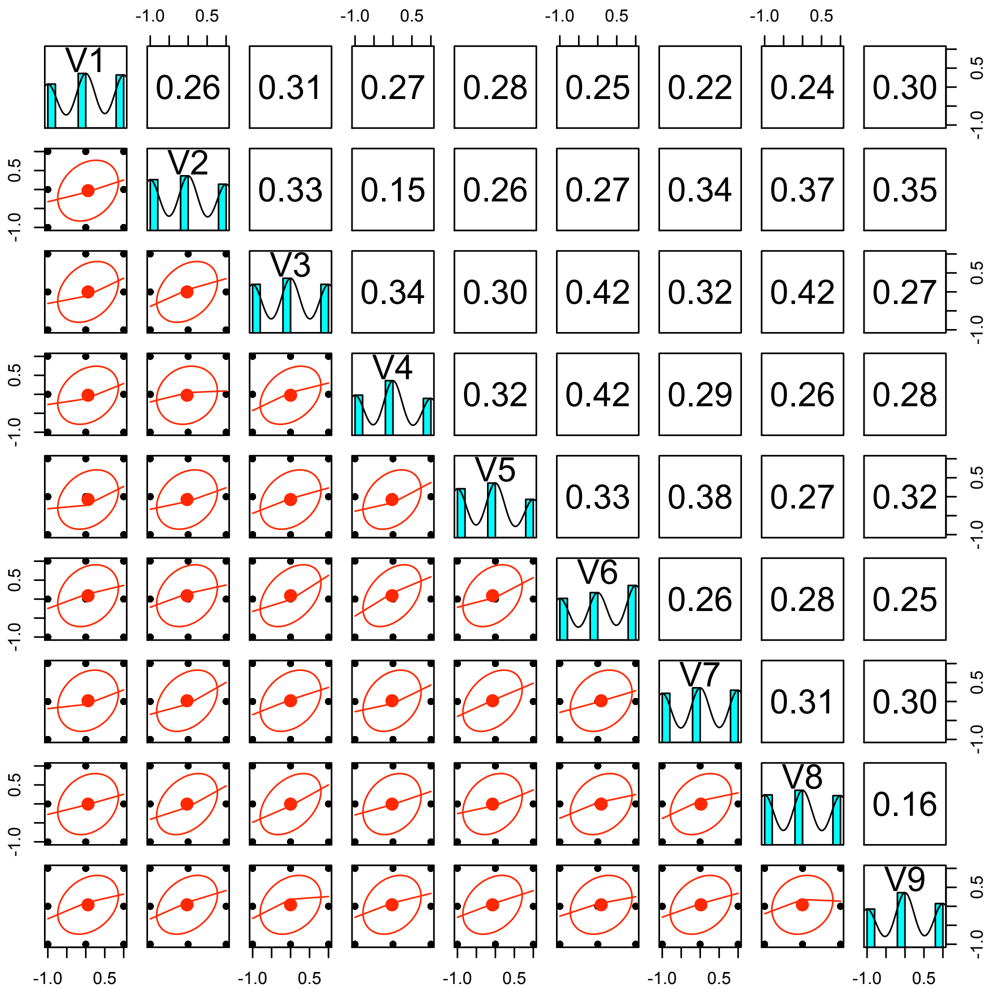
loadings are lower

sem fit

```
> sem.4 <- sem(mod.4,cor(cat4),200)
> summary(sem.4)

 Model Chisquare =  41.884   Df =  27 Pr(>Chisq) = 0.033789
 Chisquare (null model) =  354.65   Df =  36
 Goodness-of-fit index =  0.95555
 Adjusted goodness-of-fit index =  0.9259
 RMSEA index =  0.052633   90% CI: (0.014957, 0.082409)
 Bentler-Bonnett NFI =  0.8819
 Tucker-Lewis NNFI =  0.93772
 Bentler CFI =  0.95329
 SRMR =  0.048461
 BIC =  -101.17

 Normalized Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.65e+00 -5.12e-01 -1.13e-06  9.86e-04  4.06e-01  1.37e+00
```

3 level items

```
> cat3 <- latent$observed
> cat3[cat3< -.5] <- -1
> cat3[cat3 > .5 ] <- 1
> cat3[abs(cat3) < 1] <- 0
> pairs.panels(cat3)
```

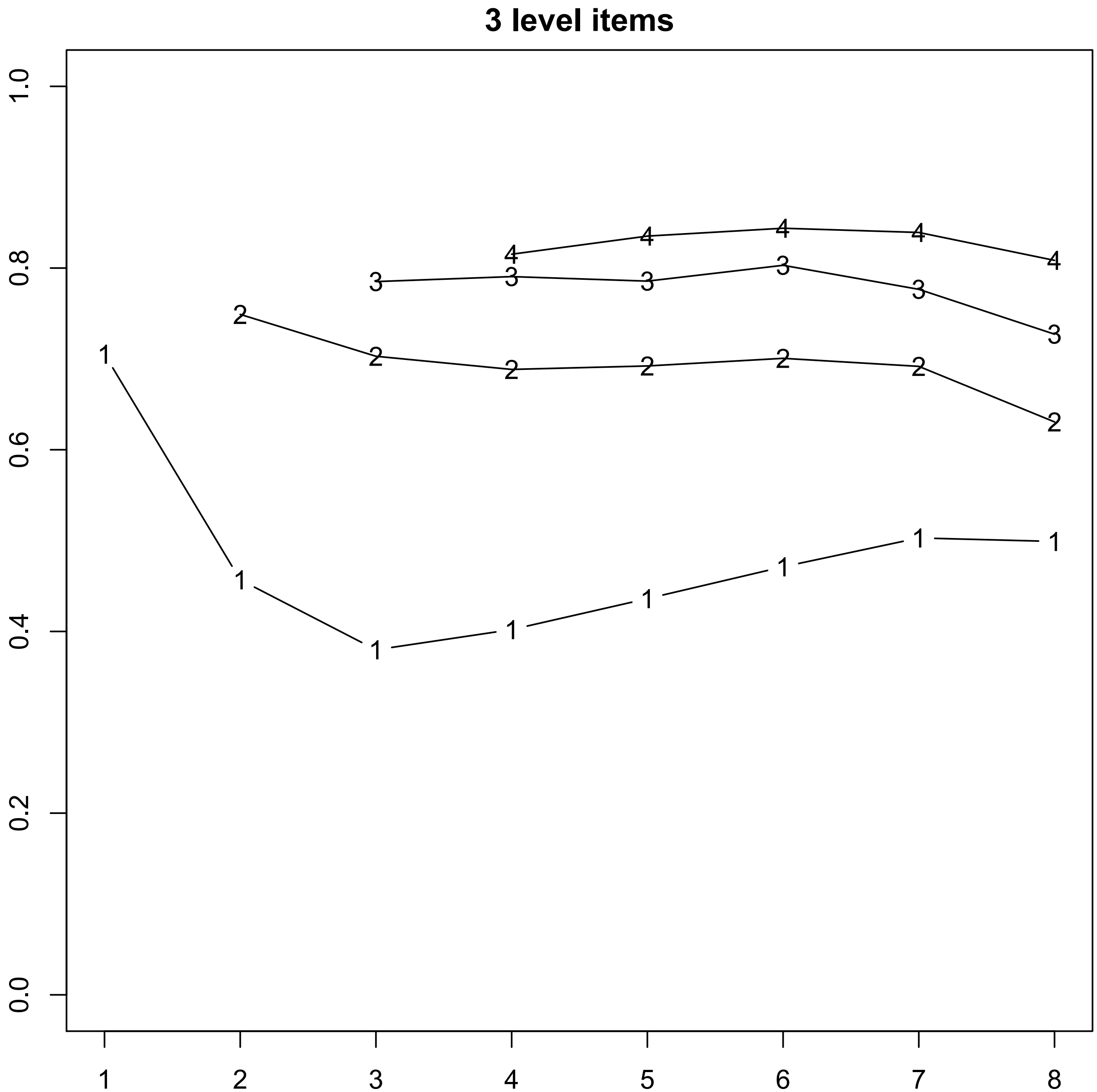


Correlations are reduced

Most correlations are smaller

```
> round(cor(latent$observed) - cor(cat3),2)
     V1   V2    V3    V4    V5   V6    V7    V8    V9
V1 0.00 0.08  0.04  0.03  0.03 0.06  0.11  0.11  0.04
V2 0.08 0.00  0.02  0.04  0.08 0.07  0.02  0.08  0.05
V3 0.04 0.02  0.00  0.05  0.07 0.00  0.03 -0.04  0.04
V4 0.03 0.04  0.05  0.00 -0.01 0.00 -0.02  0.02 -0.01
V5 0.03 0.08  0.07 -0.01  0.00 0.08  0.01 -0.02  0.05
V6 0.06 0.07  0.00  0.00  0.08 0.00  0.05  0.04  0.10
V7 0.11 0.02  0.03 -0.02  0.01 0.05  0.00  0.06  0.07
V8 0.11 0.08 -0.04  0.02 -0.02 0.04  0.06  0.00  0.05
V9 0.04 0.05  0.04 -0.01  0.05 0.10  0.07  0.05  0.00
```

VSS



VSS fits

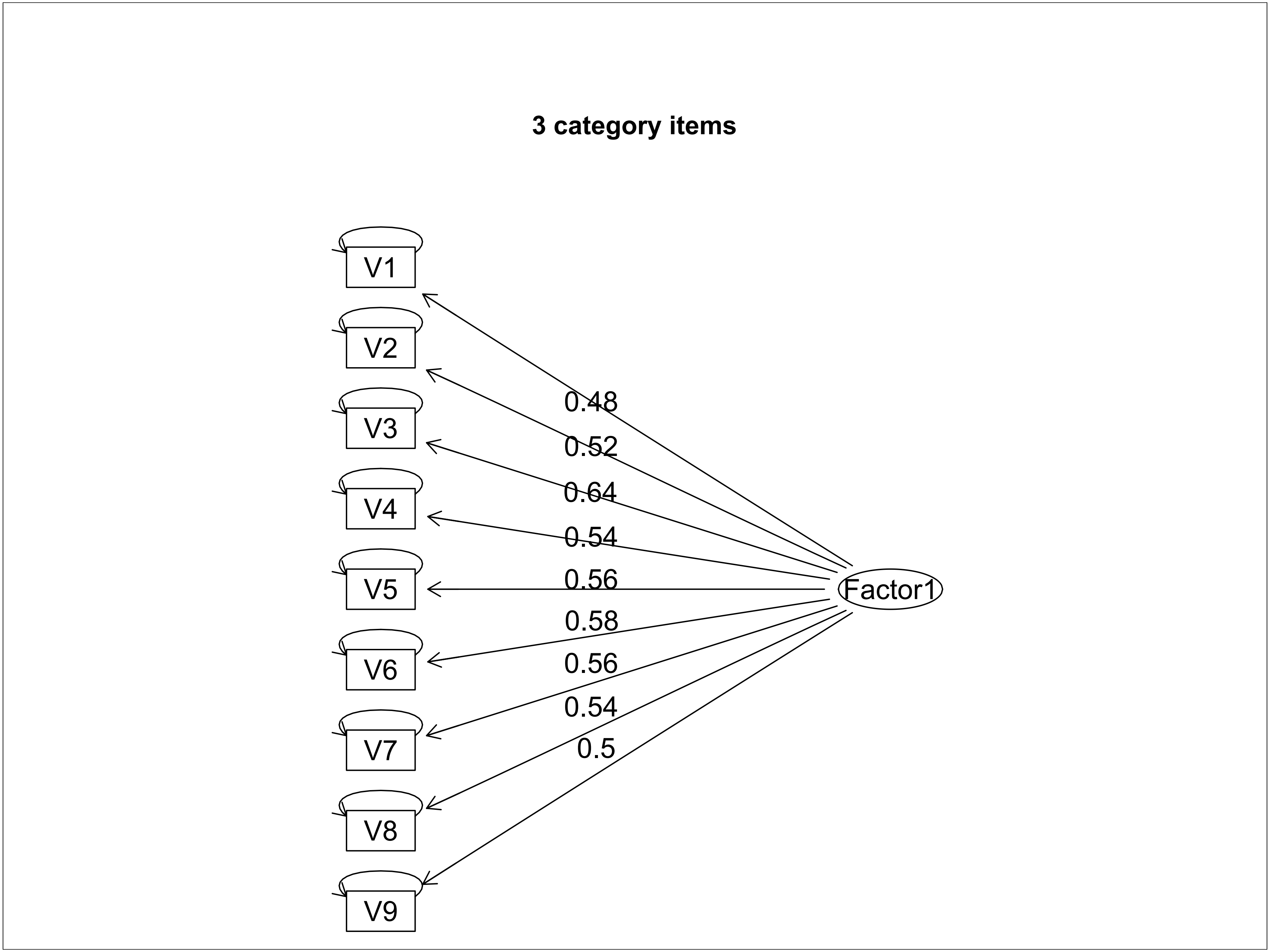
```
>  vss.3
Very Simple Structure of  3 level items
VSS complexity 1 achieves a maximimum of 0.71  with  1
factors
VSS complexity 2 achieves a maximimum of 0.75  with  2
factors

The Velicer MAP criterion achieves a minimum of 1  with  1
factors
 Velicer MAP
[1] 0.02 0.04 0.07 0.11 0.17 0.30 0.46 1.00

Very Simple Structure Complexity 1
[1] 0.71 0.46 0.38 0.40 0.44 0.47 0.50 0.50
Very Simple Structure Complexity 2
[1] 0.00 0.75 0.70 0.69 0.69 0.70 0.69 0.63
```

How many factors?

```
> vss.3$vss.stats[,1:3]
  dof        chisq       prob
1  27 4.011865e+01 0.04994193
2  19 2.175496e+01 0.29662056
3  12 8.841514e+00 0.71640433
4   6 7.578853e+00 0.27060879
5   1 3.089800e+00 0.07878449
6  -3 2.628320e+00         NA
7  -6 1.882974e-02         NA
8  -8 4.302505e-05         NA
```

factanal

```
> f3 <- factanal(cat3,1)
factanal(x = cat3, factors = 1)
Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8    V9
0.770 0.725 0.595 0.709 0.682 0.661 0.692 0.710 0.751
Loadings:
   Factor1
V1 0.480
V2 0.525
V3 0.636
V4 0.540
V5 0.564
V6 0.582
V7 0.555
V8 0.538
V9 0.499
               Factor1
SS loadings      2.705
Proportion Var   0.301
Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 40.09 on 27 degrees of freedom.
The p-value is 0.0502
```


3 category items
V1
V2
V3
V4
V5
V6
V7
V8
V9
Factor1
0.48
0.52
0.64
0.54
0.56
0.58
0.56
0.54
0.5

```
>  sem3 <- sem(mod3,cor(cat3),200)
> summary(sem3)

 Model Chisquare =  41.018   Df =  27 Pr(>Chisq) = 0.041018
 Chisquare (null model) =  368.11   Df =  36
 Goodness-of-fit index =  0.95617
 Adjusted goodness-of-fit index =  0.92695
 RMSEA index =  0.051079   90% CI: (0.010671, 0.081127)
 Bentler-Bonnett NFI =  0.88857
 Tucker-Lewis NNFI =  0.94372
 Bentler CFI =  0.95779
 SRMR =  0.047473
 BIC =  -102.04

 Normalized Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.86e+00 -4.43e-01 -7.19e-08  2.19e-03  2.64e-01  1.45e+00
```

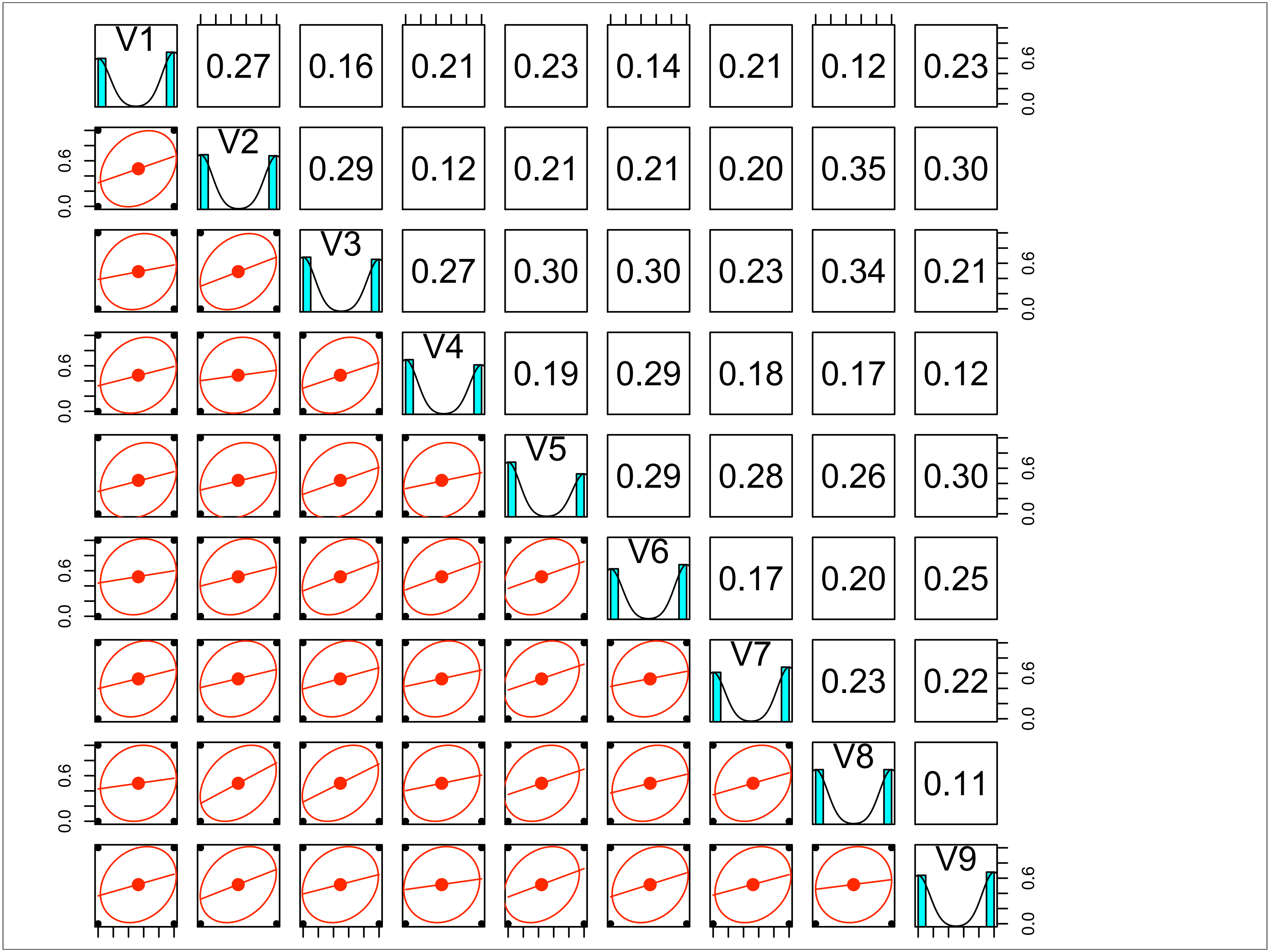
3 category parameters

```
 Parameter Estimates
    Estimate Std Error z value Pr(>|z|)
V1  0.47990  0.074676  6.4264  1.3063e-10 V1 <--- Factor1
V2  0.52484  0.074248  7.0687  1.5641e-12 V2 <--- Factor1
V3  0.63603  0.071347  8.9146  0.0000e+00 V3 <--- Factor1
V4  0.53957  0.073819  7.3093  2.6845e-13 V4 <--- Factor1
V5  0.56357  0.073055  7.7143  1.2212e-14 V5 <--- Factor1
V6  0.58192  0.072733  8.0007  1.3323e-15 V6 <--- Factor1
V7  0.55507  0.073273  7.5754  3.5749e-14 V7 <--- Factor1
V8  0.53832  0.073777  7.2966  2.9510e-13 V8 <--- Factor1
V9  0.49946  0.074634  6.6922  2.1983e-11 V9 <--- Factor1
x1e 0.76970  0.083653  9.2011  0.0000e+00 V1 <--> V1
x2e 0.72455  0.081107  8.9332  0.0000e+00 V2 <--> V2
x3e 0.59547  0.072907  8.1675  2.2204e-16 V3 <--> V3
x4e 0.70887  0.079989  8.8620  0.0000e+00 V4 <--> V4
x5e 0.68239  0.078053  8.7427  0.0000e+00 V5 <--> V5
x6e 0.66137  0.076873  8.6034  0.0000e+00 V6 <--> V6
x7e 0.69189  0.078684  8.7933  0.0000e+00 V7 <--> V7
x8e 0.71021  0.080001  8.8775  0.0000e+00 V8 <--> V8
x9e 0.75054  0.082664  9.0794  0.0000e+00 V9 <--> V9
```

Dichotomous items

```
> dichot <- (latent$observed>0)+0
> pairs.panels(dichot)
```

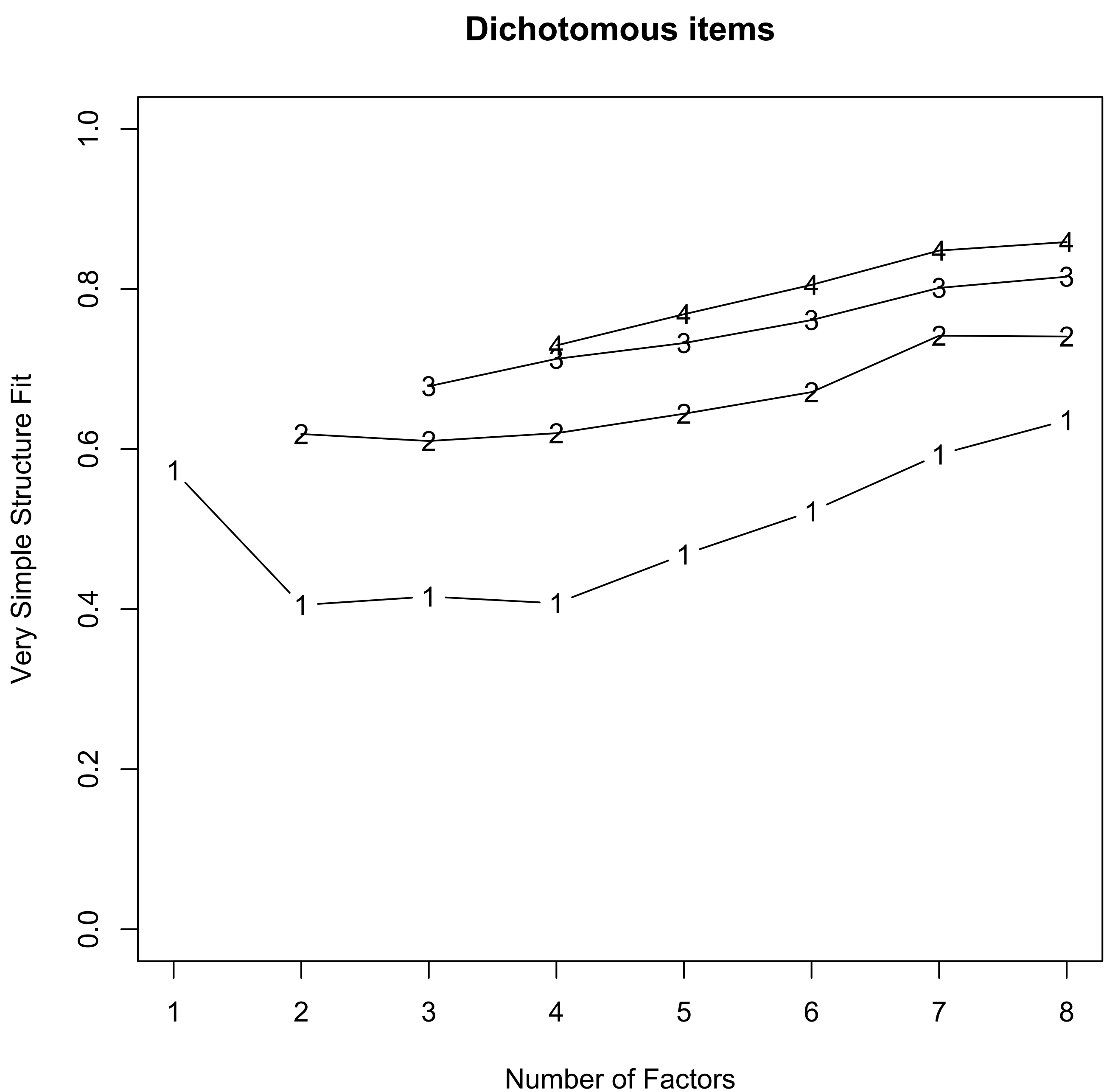
But note that these are all of equal difficulty


V1
V2
V3
V4
V5
V6
V7
V8
V9
0.27
0.16
0.21
0.23
0.14
0.21
0.12
0.23
0.29
0.12
0.21
0.21
0.20
0.35
0.30
0.27
0.30
0.30
0.23
0.34
0.21
0.19
0.29
0.18
0.17
0.12
0.29
0.28
0.26
0.30
0.17
0.20
0.25
0.23
0.22
0.11
0.0
0.6

Most correlations are reduced

```
> round(cor(latent$observed) - cor(dichot),2)
     V1   V2   V3   V4    V5   V6   V7    V8   V9
V1 0.00 0.07 0.18 0.09  0.08 0.18 0.12  0.23 0.11
V2 0.07 0.00 0.06 0.07  0.13 0.13 0.16  0.10 0.10
V3 0.18 0.06 0.00 0.13  0.08 0.12 0.11  0.04 0.10
V4 0.09 0.07 0.13 0.00  0.13 0.13 0.09  0.11 0.15
V5 0.08 0.13 0.08 0.13  0.00 0.12 0.11 -0.01 0.08
V6 0.18 0.13 0.12 0.13  0.12 0.00 0.15  0.13 0.10
V7 0.12 0.16 0.11 0.09  0.11 0.15 0.00  0.14 0.15
V8 0.23 0.10 0.04 0.11 -0.01 0.13 0.14  0.00 0.10
V9 0.11 0.10 0.10 0.15  0.08 0.10 0.15  0.10 0.00
```

VSS



Dichotomous items
Very Simple Structure Fit
Number of Factors
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
8

VSS stats

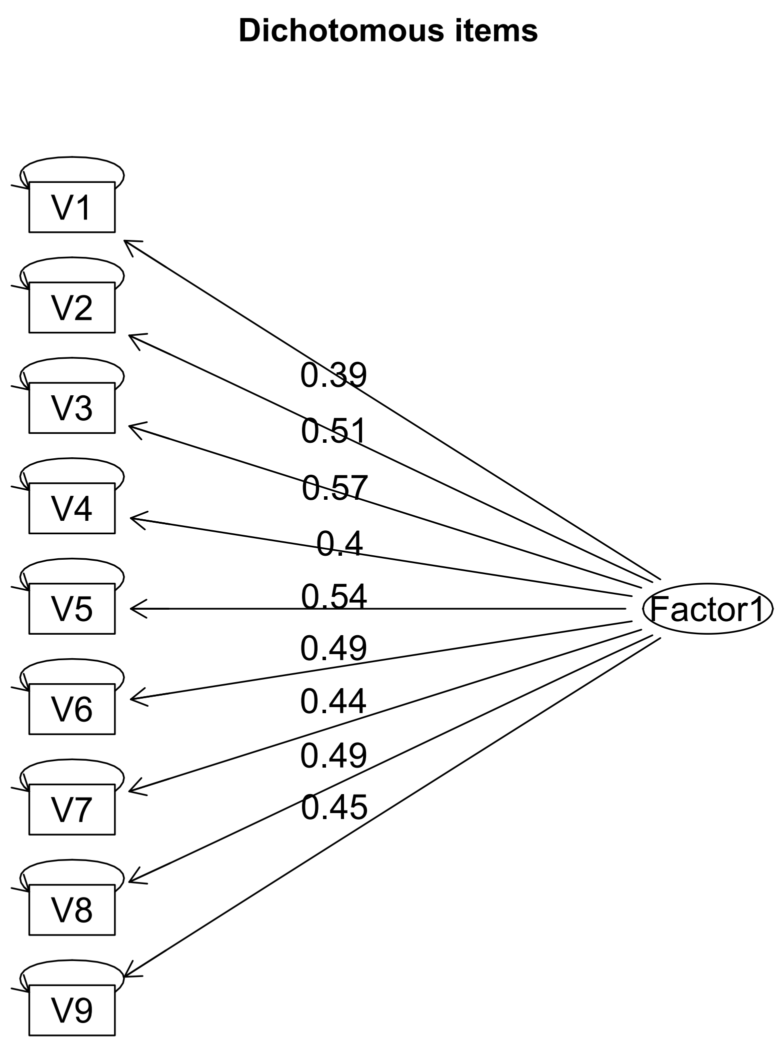
```
> vss.d
Very Simple Structure of  Dichotomous items
VSS complexity 1 achieves a maximimum of 0.64  with  8
factors
VSS complexity 2 achieves a maximimum of 0.74  with  7
factors

The Velicer MAP criterion achieves a minimum of 1  with  1
factors
Velicer MAP
[1] 0.02 0.04 0.07 0.11 0.17 0.29 0.47 1.00

Very Simple Structure Complexity 1
[1] 0.57 0.41 0.42 0.41 0.47 0.52 0.59 0.64
Very Simple Structure Complexity 2
[1] 0.00 0.62 0.61 0.62 0.64 0.67 0.74 0.74
```

VSS chi square uses factor.pa rather than mle

	dof	chisq	prob
1	27	3.310306e+01	0.1936903
2	19	2.273101e+01	0.2493977
3	12	1.038772e+01	0.5819866
4	6	4.544930e+00	0.6033534
5	1	2.419505e+00	0.1198328
6	-3	1.891661e-01	NA
7	-6	1.902598e-02	NA
8	-8	4.720899e-05	NA

factanal

```
> fd <- factanal(dichot,1)
factanal(x = dichot, factors = 1)
Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8    V9
0.844 0.736 0.676 0.842 0.705 0.764 0.803 0.764 0.795
Loadings:
   Factor1
V1 0.394
V2 0.514
V3 0.569
V4 0.397
V5 0.543
V6 0.486
V7 0.444
V8 0.486
V9 0.453
               Factor1
SS loadings      2.071
Proportion Var   0.230
Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 33.08 on 27 degrees of freedom.
The p-value is 0.194
```



Once again, loadings are lower

```
> summary(sem.d)

 Model Chisquare =  33.85   Df =  27 Pr(>Chisq) = 0.17044
 Chisquare (null model) =  243.65   Df =  36
 Goodness-of-fit index =  0.9649
 Adjusted goodness-of-fit index =  0.9415
 RMSEA index =  0.035706   90% CI: (NA, 0.069392)
 Bentler-Bonnett NFI =  0.86107
 Tucker-Lewis NNFI =  0.95601
 Bentler CFI =  0.96701
 SRMR =  0.046655
 BIC =  -109.20

 Normalized Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-1.52e+00 -4.97e-01 -9.91e-06  2.16e-03  4.53e-01  1.38e+00
```

Manipulating dichotomous items

- Correlations of dichotomous items can be used to estimate continuous correlations using tetrachoric or polychoric correlations
- The hetcor function in the polychor package will do this

Using hetcor

- Items need to be in the form of `factors' rather than integers or reals
- Returned matrix is lower diagonal

```
> dichot.fact <- matrix(factor(dichot),ncol=9)
> head(dichot.fact)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,] "1"  "1"  "1"  "1"  "0"  "1"  "0"  "1"  "0"
[2,] "1"  "1"  "1"  "1"  "1"  "1"  "1"  "1"  "1"
[3,] "0"  "1"  "1"  "0"  "1"  "1"  "1"  "1"  "1"
[4,] "1"  "0"  "1"  "0"  "1"  "0"  "0"  "1"  "0"
[5,] "0"  "0"  "1"  "0"  "0"  "0"  "1"  "0"  "0"
[6,] "1"  "1"  "0"  "0"  "1"  "0"  "0"  "0"  "0"
> poly.r <- hetcor(dichot.fact)
> poly.r
```

```
> poly.r

Two-Step Estimates

Correlations/Type of Correlation:
            dichot.fact         X2         X3         X4         X5         X6         X7         X8         X9
dichot.fact           1 Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric
X2               0.4137          1 Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric
X3               0.2513     0.4398          1 Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric
X4               0.3302      0.187     0.4109          1 Polychoric Polychoric Polychoric Polychoric Polychoric
X5               0.3551     0.3257     0.4555     0.2885          1 Polychoric Polychoric Polychoric Polychoric
X6               0.2152      0.325      0.456     0.4445     0.4389          1 Polychoric Polychoric Polychoric
X7               0.3207     0.3104     0.3559      0.284     0.4266     0.2615          1 Polychoric Polychoric
X8               0.1879     0.5225     0.5092     0.2643     0.4017     0.3094     0.3542          1 Polychoric
X9               0.3519     0.4548     0.3252     0.1902      0.451     0.3815     0.3373      0.172          1


Standard Errors:
            dichot.fact      X2      X3      X4      X5      X6     X7     X8
dichot.fact
X2              0.09744
X3               0.1062 0.09548
X4               0.1026  0.1084 0.09759
X5               0.1017  0.1031 0.09469   0.105
X6               0.1076  0.1027 0.09433 0.09527 0.09608
X7               0.1031  0.1035  0.1011  0.1049   0.097  0.1058
X8               0.1084 0.08871 0.08991  0.1056 0.09856  0.1035 0.1012
X9               0.1013 0.09438  0.1027  0.1083 0.09513 0.09949 0.1021 0.1088

n = 200
```

How to get the correlations? Examine the structure

```
> str(poly.r)
List of 7
 $ correlations: num [1:9, 1:9] 1 0.414 0.251 0.33 0.355 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:9] "dichot.fact" "X2" "X3" "X4" ...
  .. ..$ : chr [1:9] "dichot.fact" "X2" "X3" "X4" ...
 $ type        : chr [1:9, 1:9] "" "Polychoric" "Polychoric"
"Polychoric" ...
 $ NA.method   : chr "complete.obs"
 $ ML          : logi FALSE
 $ std.errors  : num [1:9, 1:9] 0 0.0974 0.1062 0.1026 0.1017 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:9] "dichot.fact" "X2" "X3" "X4" ...
  .. ..$ : chr [1:9] "dichot.fact" "X2" "X3" "X4" ...
 $ n           : int 200
 $ tests       : num [1:9, 1:9] 0 NA NA NA NA NA NA NA NA 0 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:9] "dichot.fact" "X2" "X3" "X4" ...
  .. ..$ : chr [1:9] "dichot.fact" "X2" "X3" "X4" ...
 - attr(*, "class")= chr "hetcor"
```

The polychoric correlations

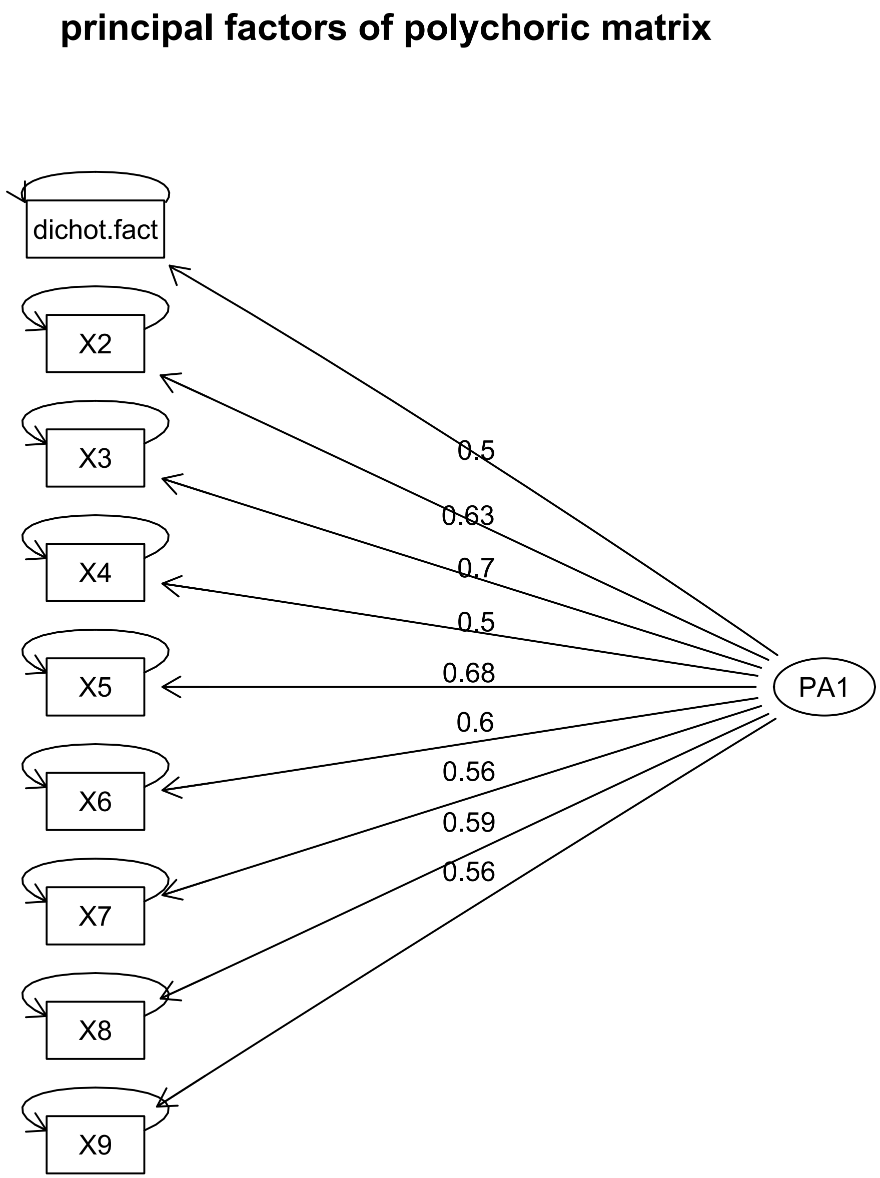
```
       > round(poly.r$correlation,2)
    dichot.fact   X2   X3   X4   X5   X6   X7   X8   X9
t.fact      1.00 0.41 0.25 0.33 0.36 0.22 0.32 0.19 0.35
            0.41 1.00 0.44 0.19 0.33 0.32 0.31 0.52 0.45
            0.25 0.44 1.00 0.41 0.46 0.46 0.36 0.51 0.33
            0.33 0.19 0.41 1.00 0.29 0.44 0.28 0.26 0.19
            0.36 0.33 0.46 0.29 1.00 0.44 0.43 0.40 0.45
            0.22 0.32 0.46 0.44 0.44 1.00 0.26 0.31 0.38
            0.32 0.31 0.36 0.28 0.43 0.26 1.00 0.35 0.34
            0.19 0.52 0.51 0.26 0.40 0.31 0.35 1.00 0.17
            0.35 0.45 0.33 0.19 0.45 0.38 0.34 0.17 1.00
```

Polychor recovers the correlations (somewhat)

```
> round(latent$r - poly.r$correlations,2)
      V1    V2    V3    V4    V5    V6    V7    V8    V9
V1  0.00 -0.07  0.09 -0.03 -0.05  0.10  0.01  0.16 -0.01
V2 -0.07  0.00 -0.09  0.00  0.01  0.01  0.05 -0.07 -0.06
V3  0.09 -0.09  0.00 -0.01 -0.08 -0.03 -0.01 -0.13 -0.01
V4 -0.03  0.00 -0.01  0.00  0.03 -0.02 -0.01  0.01  0.08
V5 -0.05  0.01 -0.08  0.03  0.00 -0.03 -0.03 -0.15 -0.08
V6  0.10  0.01 -0.03 -0.02 -0.03  0.00  0.05  0.02 -0.03
V7  0.01  0.05 -0.01 -0.01 -0.03  0.05  0.00  0.01  0.03
V8  0.16 -0.07 -0.13  0.01 -0.15  0.02  0.01  0.00  0.04
V9 -0.01 -0.06 -0.01  0.08 -0.08 -0.03  0.03  0.04  0.00
```

But, the fits are worse

```
> fpc <- factor.pa(poly.r$correlations,1,n.obs=200)
> fpc
  V  PA1
1 1 0.50
2 2 0.63
3 3 0.70
4 4 0.50
5 5 0.68
6 6 0.60
7 7 0.56
8 8 0.59
9 9 0.56
               PA1
SS loadings    3.18
Proportion Var 0.35
Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the model is 27  and the fit was 
0.57
The number of observations was  200  with Chi Square = 
111.24  with prob <  3.4e-12
```



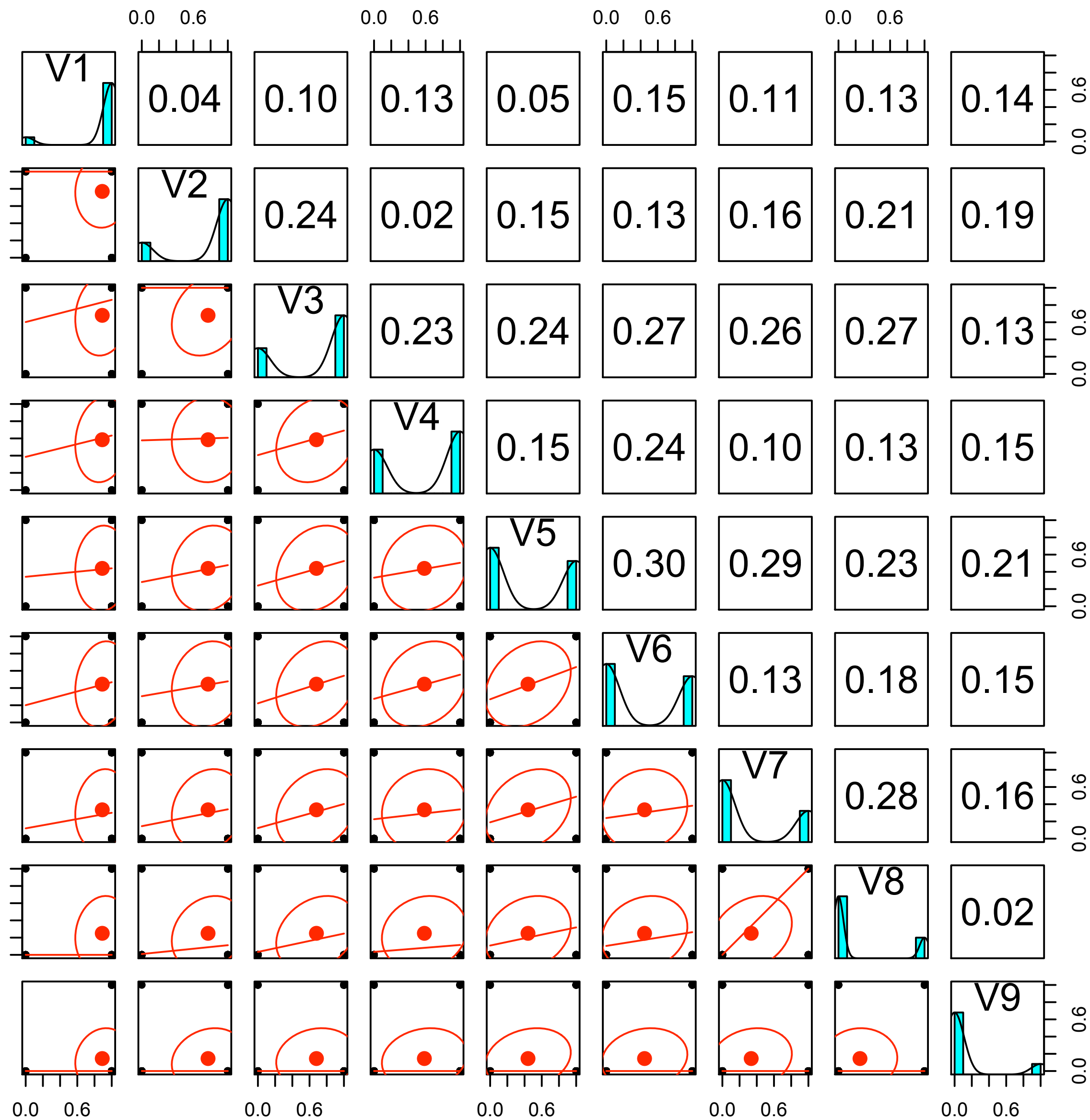
Structure of loadings

sem with polychorics

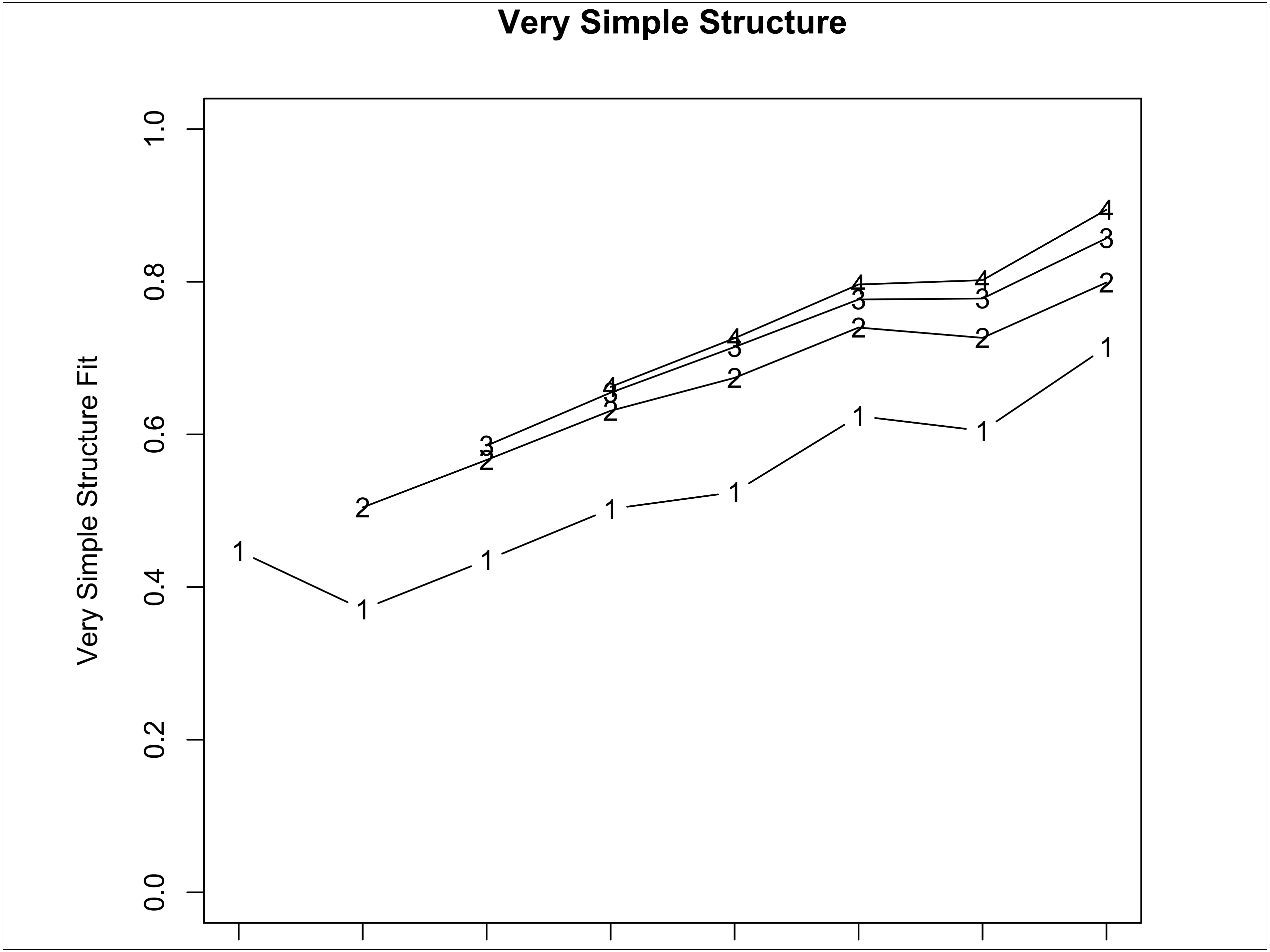
```
> sem.pc <- sem(mod.pc,poly.r$correlations,200)
> summary(sem.pc)

 Model Chisquare =  113.67   Df =  27 Pr(>Chisq) =
1.3231e-12
 Chisquare (null model) =  545.37   Df =  36
 Goodness-of-fit index =  0.8996
 Adjusted goodness-of-fit index =  0.83266
 RMSEA index =  0.12701   90% CI: (0.10340, 0.15158)
 Bentler-Bonnett NFI =  0.79157
 Tucker-Lewis NNFI =  0.77313
 Bentler CFI =  0.82985
 SRMR =  0.068918
 BIC =  -29.384
 Normalized Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-2.23e+00 -7.24e-01  9.55e-08  4.73e-03  6.85e-01  1.97e+00
```

Items differ in difficulty

```
> diff <- seq(-1,1,.25)
> diff
[1] -1.00 -0.75 -0.50 -0.25  0.00  0.25  0.50  0.75  1.00

diff.dichot <-t((t(latent$observed)>diff) +0)

> describe(diff.dichot)

   var   n mean   sd median trimmed mad min max range  skew kurtosis   se
V1   1 200 0.89 0.31      1    0.99   0   0   1     1 -2.47     4.14 0.02
V2   2 200 0.77 0.42      1    0.84   0   0   1     1 -1.27    -0.38 0.03
V3   3 200 0.68 0.47      1    0.72   0   0   1     1 -0.77    -1.42 0.03
V4   4 200 0.58 0.49      1    0.61   0   0   1     1 -0.34    -1.89 0.03
V5   5 200 0.44 0.50      0    0.42   0   0   1     1  0.24    -1.95 0.04
V6   6 200 0.44 0.50      0    0.43   0   0   1     1  0.22    -1.96 0.04
V7   7 200 0.34 0.47      0    0.29   0   0   1     1  0.69    -1.53 0.03
V8   8 200 0.25 0.43      0    0.19   0   0   1     1  1.15    -0.69 0.03
V9   9 200 0.14 0.35      0    0.06   0   0   1     1  2.00     2.02 0.02
>
```



The effect of difficulty on r


Very Simple Structure
Very Simple Structure Fit
0.0
0.2
0.4
0.6
0.8
1.0

MAP does a good job

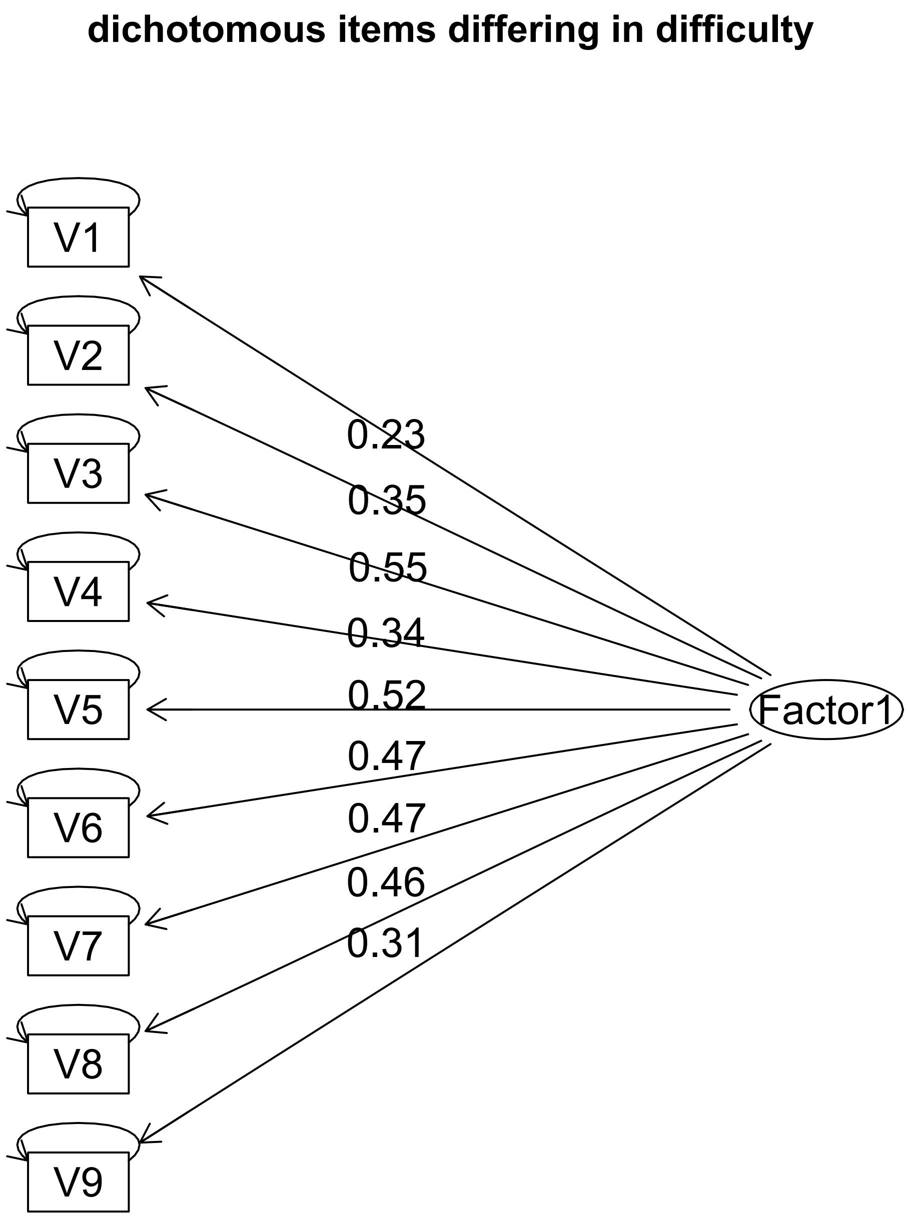
```
> vss.dd
Very Simple Structure
VSS complexity 1 achieves a maximimum of 0.71  with  8
factors
VSS complexity 2 achieves a maximimum of 0.8  with  8
factors

The Velicer MAP criterion achieves a minimum of 1  with  1
factors
Velicer MAP
[1] 0.02 0.04 0.07 0.12 0.18 0.26 0.45 1.00

Very Simple Structure Complexity 1
[1] 0.45 0.37 0.44 0.50 0.52 0.62 0.60 0.71
Very Simple Structure Complexity 2
[1] 0.00 0.50 0.57 0.63 0.67 0.74 0.73 0.80
```

Chi square

```
  > vss.dd$vss.stats[,1:3]
  dof        chisq      prob
1  27 2.716383e+01 0.4549829
2  19 1.808306e+01 0.5168968
3  12 9.767687e+00 0.6363331
4   6 7.077246e+00 0.3137626
5   1 2.690807e+00 0.1009287
6  -3 3.601937e-01        NA
7  -6 1.347089e-02        NA
8  -8 2.712477e-05        NA
```

```
factanal(x = diff.dichot, factors = 1)
Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8    V9
0.946 0.880 0.698 0.882 0.730 0.776 0.782 0.790 0.902
Loadings:
   Factor1
V1 0.233
V2 0.347
V3 0.549
V4 0.344
V5 0.519
V6 0.473
V7 0.467
V8 0.458
V9 0.314
               Factor1
SS loadings      1.615
Proportion Var   0.179

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 27.15 on 27 degrees of freedom.
The p-value is 0.456
```

The effect of difficulty



SEM of dichotomous items differing in difficulty

```
> summary(sem.dd)

 Model Chisquare =  27.777   Df =  27 Pr(>Chisq) = 0.42255
 Chisquare (null model) =  167.01   Df =  36
 Goodness-of-fit index =  0.97115
 Adjusted goodness-of-fit index =  0.95192
 RMSEA index =  0.012024   90% CI: (NA, 0.057117)
 Bentler-Bonnett NFI =  0.83368
 Tucker-Lewis NNFI =  0.9921
 Bentler CFI =  0.99407
 SRMR =  0.045246
 BIC =  -115.28
```

Try with polychorics

```
> dd.fact <- matrix(factor(diff.dichot),ncol=9)
> poly.dd <- hetcor(dd.fact)
Warning message:
In hetcor.data.frame(dframe, ML = ML, std.err = std.err, bins =
bins,  :
  the correlation matrix has been adjusted to make it positive-
definite
> poly.dd

                                  Two-Step Estimates

                             Correlations/Type of Correlation:
        dd.fact         X2         X3         X4         X5         X6         X7         X8         X9
 dd.fact      1 Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric
 X2      0.1212          1 Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric
 X3      0.2133     0.3975          1 Polychoric Polychoric Polychoric Polychoric Polychoric Polychoric
 X4      0.2657    0.04403     0.3618          1 Polychoric Polychoric Polychoric Polychoric Polychoric
 X5      0.1402      0.279     0.3955     0.2467          1 Polychoric Polychoric Polychoric Polychoric
 X6      0.3228     0.2216     0.4363     0.3788     0.4518          1 Polychoric Polychoric Polychoric
 X7      0.2649     0.3024     0.4521     0.1696     0.4492     0.2124          1 Polychoric Polychoric
 X8      0.3055      0.413     0.5225     0.2294     0.3699     0.3096     0.4474          1 Polychoric
 X9      0.8024     0.4893     0.2776     0.2835     0.3618     0.2926     0.2996     0.0982          1
```

Some polychors are bad

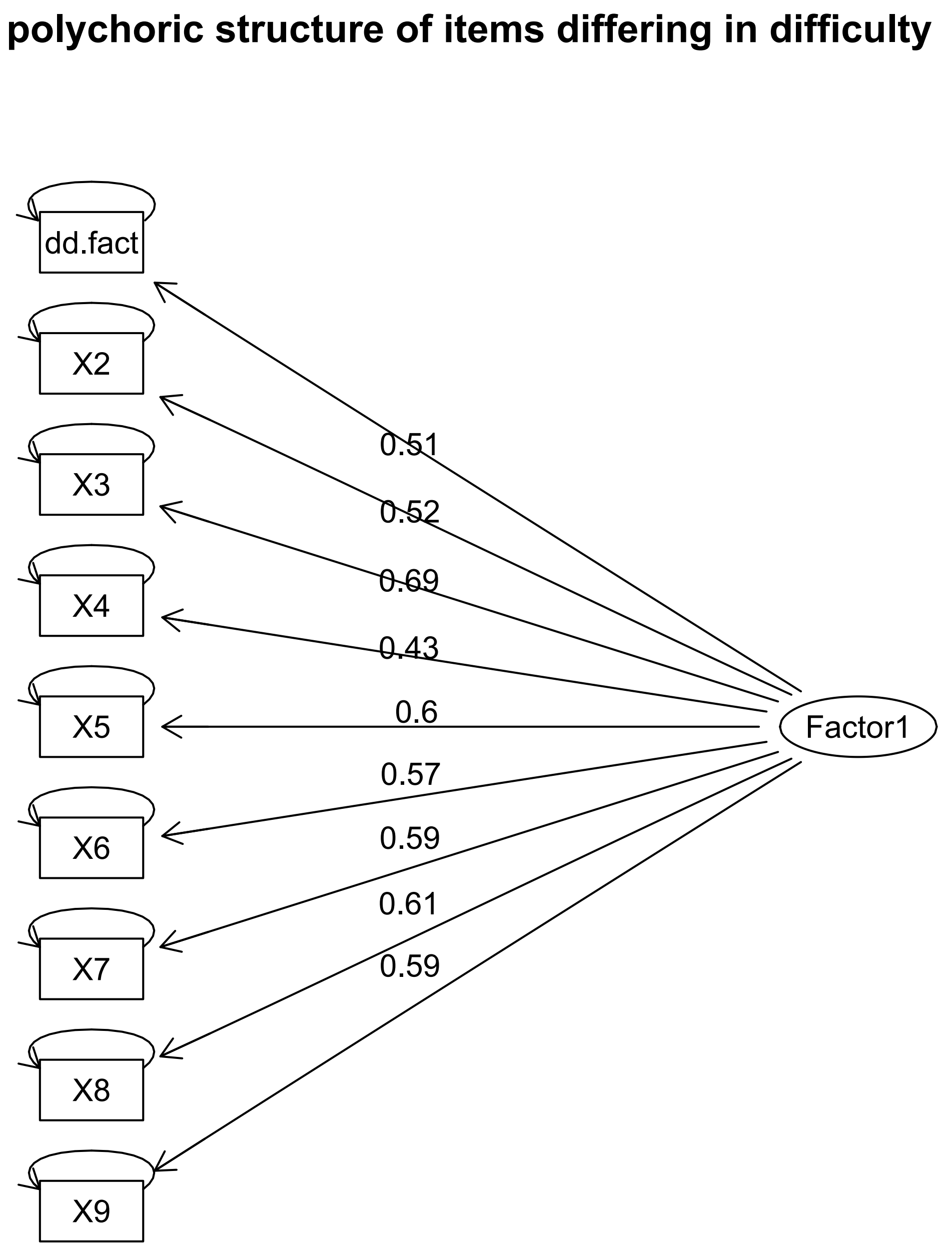
```
> round(poly.dd$correlations,2)
         dd.fact   X2   X3   X4   X5   X6   X7   X8   X9
dd.fact     1.00 0.12 0.21 0.27 0.14 0.32 0.26 0.31 0.80
X2          0.12 1.00 0.40 0.04 0.28 0.22 0.30 0.41 0.49
X3          0.21 0.40 1.00 0.36 0.40 0.44 0.45 0.52 0.28
X4          0.27 0.04 0.36 1.00 0.25 0.38 0.17 0.23 0.28
X5          0.14 0.28 0.40 0.25 1.00 0.45 0.45 0.37 0.36
X6          0.32 0.22 0.44 0.38 0.45 1.00 0.21 0.31 0.29
X7          0.26 0.30 0.45 0.17 0.45 0.21 1.00 0.45 0.30
X8          0.31 0.41 0.52 0.23 0.37 0.31 0.45 1.00 0.10
X9          0.80 0.49 0.28 0.28 0.36 0.29 0.30 0.10 1.00
```

Compare to original sample

```
> round(latent$r-poly.dd$correlations,2)
      V1    V2    V3    V4    V5    V6    V7    V8    V9
V1  0.00  0.22  0.13  0.04  0.17 -0.01  0.07  0.05 -0.46
V2  0.22  0.00 -0.05  0.15  0.06  0.12  0.06  0.04 -0.09
V3  0.13 -0.05  0.00  0.04 -0.02 -0.01 -0.11 -0.15  0.04
V4  0.04  0.15  0.04  0.00  0.07  0.04  0.11  0.05 -0.01
V5  0.17  0.06 -0.02  0.07  0.00 -0.04 -0.06 -0.12  0.01
V6 -0.01  0.12 -0.01  0.04 -0.04  0.00  0.10  0.02  0.06
V7  0.07  0.06 -0.11  0.11 -0.06  0.10  0.00 -0.08  0.07
V8  0.05  0.04 -0.15  0.05 -0.12  0.02 -0.08  0.00  0.12
V9 -0.46 -0.09  0.04 -0.01  0.01  0.06  0.07  0.12  0.00
```

factanal

```
Call:
factanal(factors = 1, covmat = poly.dd$correlations, n.obs =
dd.fact       X2       X3       X4       X5       X6       X7
X9
  0.735   0.728   0.524   0.814   0.635   0.673   0.650   0.
0.653
Loadings:
        Factor1
dd.fact 0.514
X2      0.522
X3      0.690
X4      0.432
X5      0.604
X6      0.572
X7      0.592
X8      0.607
X9      0.589
               Factor1
SS loadings      2.958
Proportion Var   0.329
Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 3178.95 on 27 degrees of freedom
The p-value is 0
```



Factors recover structure

But fit is terrible!

```
Model Chisquare =  3252.5   Df =  27 Pr(>Chisq) = 0
Chisquare (null model) =  3635.7   Df =  36
Goodness-of-fit index =  0.7789
Adjusted goodness-of-fit index =  0.63149
RMSEA index =  0.7748   90% CI: (NA, NA)
Bentler-Bonnett NFI =  0.10539
Tucker-Lewis NNFI =  -0.19474
Bentler CFI =  0.10395
SRMR =  0.11425
BIC =  3109.4
```

Parameters are fine!

```
Parameter Estimates
         Estimate Std Error z value Pr(>|z|)
dd.fact 0.51449  0.078937  6.5177  7.1372e-11 dd.fact <--- Factor1
X2      0.52201  0.073263  7.1251  1.0396e-12 X2 <--- Factor1
X3      0.68994  0.070252  9.8209  0.0000e+00 X3 <--- Factor1
X4      0.43166  0.074963  5.7584  8.4930e-09 X4 <--- Factor1
X5      0.60444  0.071100  8.5012  0.0000e+00 X5 <--- Factor1
X6      0.57186  0.071986  7.9441  1.9984e-15 X6 <--- Factor1
X7      0.59192  0.071588  8.2683  2.2204e-16 X7 <--- Factor1
X8      0.60726  0.072851  8.3356  0.0000e+00 X8 <--- Factor1
X9      0.58886  0.077977  7.5518  4.2855e-14 X9 <--- Factor1
x1e     0.73530  0.086127  8.5374  0.0000e+00 dd.fact <--> dd.fact
x2e     0.72751  0.080257  9.0647  0.0000e+00 X2 <--> X2
x3e     0.52398  0.069762  7.5109  5.8620e-14 X3 <--> X3
x4e     0.81367  0.086306  9.4277  0.0000e+00 X4 <--> X4
x5e     0.63465  0.073724  8.6085  0.0000e+00 X5 <--> X5
x6e     0.67297  0.076378  8.8111  0.0000e+00 X6 <--> X6
x7e     0.64963  0.074946  8.6681  0.0000e+00 X7 <--> X7
x8e     0.63123  0.076089  8.2960  0.0000e+00 X8 <--> X8
x9e     0.65324  0.083448  7.8281  4.8850e-15 X9 <--> X9
```

Conclusions about items

- No easy answers
- Need to examine factors, correlations, scatter plots, fits, residuals, etc.
- Consider different ways of analyzing the data