

Chapter 5: Further issues: Item quality

William Revelle

Northwestern University

Prepared as part of course on latent variable analysis ([Psychology 454](#))
and as a supplement to the [Short Guide to R for psychologists](#)

February 21, 2007

5.1	Continuous, ordinal, and dichotomous data	1
5.2	Simple structure versus circumplex structure	1
5.3	Data generation using the circ.sim function	2
5.4	Simple structure - normal items	3
5.4.1	5 categories of responses	8
5.4.2	3 categories of responses	13
5.4.3	dichotomous items	16
5.5	Circumplex structure - normal items	22
5.5.1	Fitting a circumplex data set with a simple structure model	23
5.5.2	An alternative model	26
5.6	Simple Structure - categorical and skewed items	29
5.6.1	Two dimensions with 4 point scales, differing in skew	29
5.6.2	An alternative model of two bipolar dimensions	34
5.7	Forming clusters or homogeneous item composites	38
5.8	References	44

5.1 Continuous, ordinal, and dichotomous data

Most advice on the use of latent variable models discusses the assumption of multivariate normality in the data. Further discussions include the need for continuous measures of the observed variables. But how does this relate to the frequent use of SEM techniques in analysis of personality or social psychological items rather than scales? In this chapter we consider typical problems in personality where we are interested in the structure of self reports of personality, emotion, or attitude. Using simulation techniques, we consider the effects of normally distributed items, ordinal items with 6 or 4 or 2 levels, and then the effect of skew on these results. We use simulations to show the results more clearly. For a discussion of real data with some of these problems, see [Rafaeli and Revelle \(2006\)](#).

5.2 Simple structure versus circumplex structure

Most personality scales are created to have “simple structure” where items load on one and only one factor (Revelle and Rocklin, 1979; Thurstone, 1947). The conventional estimate for the reliability and general factor saturation of such a test is Cronbach’s coefficient α (Cronbach, 1951). Variations of this model include hierarchical structures where all items load on a general factor, g , and then groups of items load on separate, group, factors (Carroll, 1993; Jensen and Weng, 1994). Estimates of the amount of general factor saturation for such hierarchical structures may be found using the ω coefficient discussed by McDonald (1999) and Zinbarg et al. (2005).

An alternative structure, particularly popular in the study of affect as well as studies of interpersonal behavior is a “circumplex structure” where items are thought to be more complex and to load on at most two factors.

“A number of elementary requirements can be teased out of the idea of circumplex structure. First, circumplex structure implies minimally that variables are inter-related; random noise does not a circumplex make. Second, circumplex structure implies that the domain in question is optimally represented by two and only two dimensions. Third, circumplex structure implies that variables do not group or clump along the two axes, as in simple structure, but rather that there are always interstitial variables between any orthogonal pair of axes (Saucier, 1992). In the ideal case, this quality will be reflected in equal spacing of variables along the circumference of the circle (Gurtman, 1994; Wiggins, Steiger, & Gaelick, 1981). Fourth, circumplex structure implies that variables have a constant radius from the center of the circle, which implies that all variables have equal communality on the two circumplex dimensions (Fisher, 1997; Gurtman, 1994). Fifth, circumplex structure implies that all rotations are equally good representations of the domain (Conte & Plutchik, 1981; Larsen & Diener, 1992).” (Acton and Revelle, 2004).

Variations of this model in personality assessment include the case where items load on two factors but the entire space is made up of more factors. The Abridged Big Five Circumplex Structure (AB5C) of Hofstee, de Raad and Golberg (1992) is an example of such a structure. That is, the AB5C items are of complexity one or two but are embedded in a five dimensional space.

5.3 Data generation using the `circ.sim` function

In investigations of circumplex versus simple structure, it is convenient to be able to generate artificial data sets. Here is a simple function (`circ.sim`) adapted from [Rafaeli and Revelle, 2006](#)). This function will generate either simple structure or circumplex structured items and can divide a continuously distributed item into a categorical scale. In addition the function can generate a higher order, g , factor and introduce skew into the items.

```
> circ.sim <- function(nvar = 72, nsub = 500, circum = TRUE,  
+   xloading = 0.6, yloading = 0.6, 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))
+   g <- rnorm(nsub)
+   trueex <- rnorm(nsub) * xloading + xbias
+   truey <- rnorm(nsub) * yloading + ybias
+   if (circum) {
+     radia <- seq(0, 2 * pi, len = nvar + 1)
+     rad <- radia[which(radia < 2 * pi)]
+   }
+   else rad <- rep(seq(0, 3 * pi/2, len = 4), nvar/4)
+   error <- matrix(rnorm(nsub * (nvar)), nsub)
+   trueitem <- outer(trueex, cos(rad)) + outer(truey, sin(rad))
+   item <- gloading * g + trueitem + errorweight * error
+   if (categorical) {
+     item = round(item)
+     item[(item <= low)] <- low
+     item[(item > high)] <- high
+   }
+   if (truncate) {
+     item[item < cutpoint] <- 0
+   }
+   return(item)
+ }

```

5.4 Simple structure - normal items

The first simulation is to generate 24 items with a two dimensional simple structure. Items are assumed to be continuous. To allow for replicability of the simulation, we set the random number seed to a memorable value (Adams, 1979). As can be seen in the loadings matrix as well as Figure 5.4 the solution is clearly a simple structure. For the purpose of this first simulation, we simulate 500 subjects.

```

> 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)

```

```

Call:
factanal(x = ss.items, factors = 2)

```

```

Uniquenesses:

```

```

  V1  V2  V3  V4  V5  V6  V7  V8  V9  V10  V11  V12  V13  V14  V15
0.61 0.70 0.66 0.65 0.62 0.68 0.65 0.77 0.67 0.66 0.65 0.62 0.68 0.56 0.62
  V16  V17  V18  V19  V20  V21  V22  V23  V24
0.70 0.68 0.65 0.58 0.60 0.67 0.68 0.62 0.67

```

Loadings:

```

      Factor1 Factor2
V1  -0.62   -0.01
V2   0.04    0.55
V3   0.59    0.03
V4  -0.06   -0.59
V5  -0.62    0.00
V6  -0.05    0.57
V7   0.59   -0.06
V8  -0.02   -0.47
V9  -0.57    0.05
V10 -0.03    0.58
V11  0.59   -0.01
V12 -0.01   -0.62
V13 -0.56    0.04
V14 -0.02    0.66
V15  0.62    0.00
V16  0.04   -0.55
V17 -0.56    0.04
V18  0.05    0.59
V19  0.65    0.01
V20  0.02   -0.63
V21 -0.57   -0.05
V22  0.02    0.56
V23  0.61   -0.01
V24  0.01   -0.57

```

```

      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

We can compare the results of this exploratory factor analysis with a confirmatory factor analysis using the sem package. To simplify the generation of our model matrix, we make a small function, **modelmat** to do it for us, and then do use the sem program to test the model. (**modelmat** uses the modulo operator

```

> 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
+   }
+ }

```

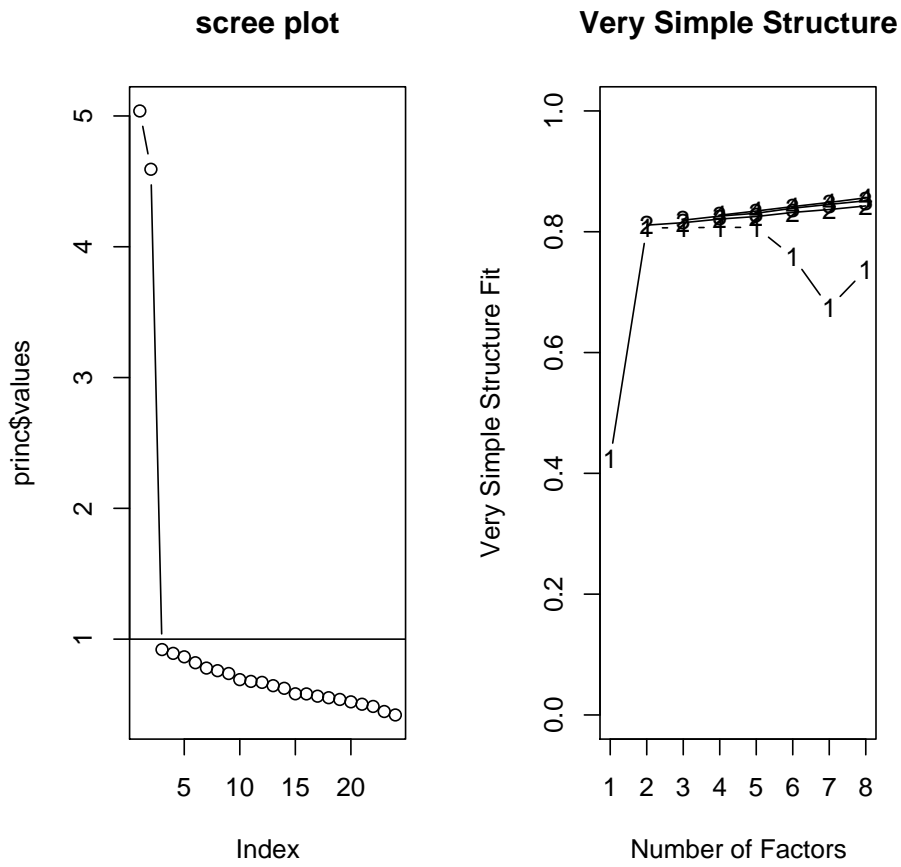


Figure 5.1: Determining the number of factors to extract from 24 variables generated with a simple structure. The left hand panel shows the scree plot, the right hand panel a VSS plot. Notice the inflection at two factors, suggesting a two factor solution

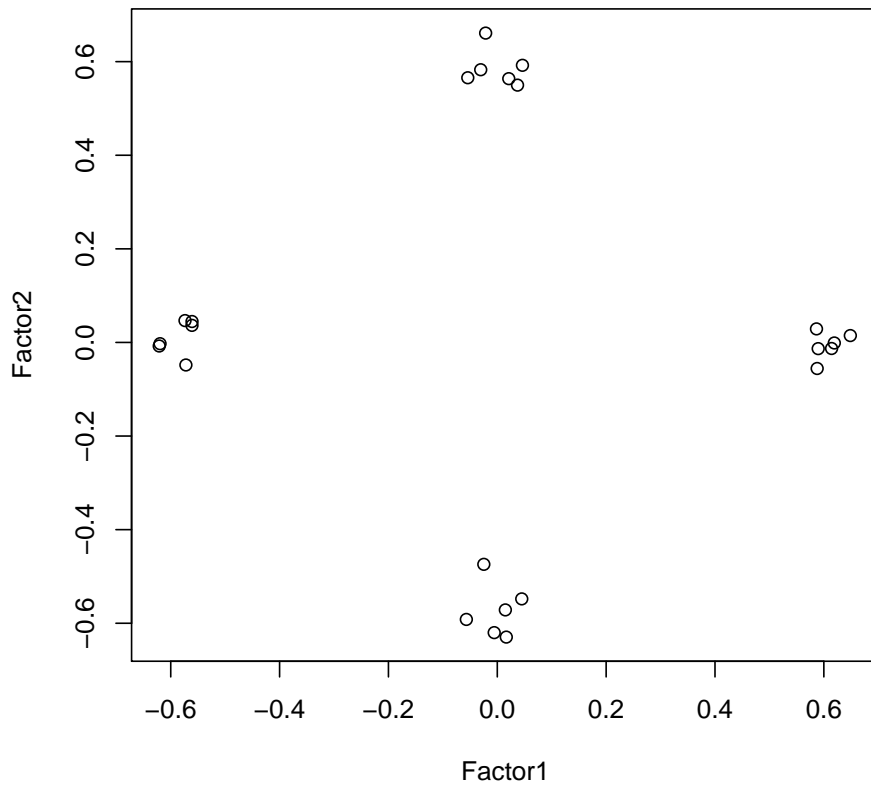


Figure 5.2: Factor loadings for 24 items on two dimensions.

```

+   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)
+ }
> model.ss <- modelmat(24)
> ss.cov <- cov(ss.items)
> sem.ss <- sem(model.ss, ss.cov, nsub)
> summary(sem.ss, digits = 2)

```

```

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)	
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
7	0.57	0.042	14	0	V7 <--- F1
8	-0.45	0.043	-10	0	V8 <--- F2
9	-0.60	0.045	-13	0	V9 <--- F1
10	0.56	0.042	13	0	V10 <--- F2
11	0.58	0.043	14	0	V11 <--- F1
12	-0.60	0.042	-14	0	V12 <--- F2
13	-0.56	0.044	-13	0	V13 <--- F1
14	0.67	0.043	16	0	V14 <--- F2
15	0.64	0.044	14	0	V15 <--- F1
16	-0.53	0.043	-12	0	V16 <--- F2
17	-0.55	0.043	-13	0	V17 <--- F1
18	0.60	0.044	14	0	V18 <--- F2
19	0.67	0.044	15	0	V19 <--- F1

20	-0.62	0.042	-15	0	V20 <--- F2
21	-0.58	0.044	-13	0	V21 <--- F1
22	0.56	0.044	13	0	V22 <--- F2
23	0.63	0.044	14	0	V23 <--- F1
24	-0.56	0.043	-13	0	V24 <--- F2
25	0.62	0.043	14	0	V1 <--> V1
26	0.62	0.043	15	0	V2 <--> V2
27	0.73	0.050	15	0	V3 <--> V3
28	0.64	0.044	14	0	V4 <--> V4
29	0.60	0.042	14	0	V5 <--> V5
30	0.73	0.050	15	0	V6 <--> V6
31	0.62	0.042	15	0	V7 <--> V7
32	0.71	0.047	15	0	V8 <--> V8
33	0.72	0.049	15	0	V9 <--> V9
34	0.61	0.042	14	0	V10 <--> V10
35	0.63	0.043	15	0	V11 <--> V11
36	0.58	0.041	14	0	V12 <--> V12
37	0.69	0.047	15	0	V13 <--> V13
38	0.58	0.042	14	0	V14 <--> V14
39	0.65	0.045	14	0	V15 <--> V15
40	0.67	0.045	15	0	V16 <--> V16
41	0.66	0.045	15	0	V17 <--> V17
42	0.67	0.047	14	0	V18 <--> V18
43	0.61	0.044	14	0	V19 <--> V19
44	0.59	0.042	14	0	V20 <--> V20
45	0.69	0.047	15	0	V21 <--> V21
46	0.67	0.046	15	0	V22 <--> V22
47	0.65	0.045	14	0	V23 <--> V23
48	0.65	0.045	15	0	V24 <--> V24

Iterations = 18

5.4.1 5 categories of responses

Unfortunately, although we like to think of our items as continuous measures of the underlying traits, items typically have 2-6 categories of response. What is the effect of this on our structural measures? Here we use the `circ.sim` function to break the continuous items down to a five category items (-2, -1, 0, 1, 2). We reset the seed to 42 so that our simulation produces the same items as before.

We do an exploratory factor analysis of the data. The `sem` package converges only if we specify two factor loadings to be one.

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

Call:
factanal(x = ss.items, factors = 2)

Uniquenesses:

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
0.66	0.72	0.69	0.67	0.66	0.72	0.69	0.77	0.66	0.68	0.71	0.67	0.74	0.59	0.65
V16	V17	V18	V19	V20	V21	V22	V23	V24						
0.71	0.67	0.70	0.65	0.65	0.70	0.68	0.68	0.70						

Loadings:

	Factor1	Factor2
V1	-0.58	0.01
V2	0.03	0.53
V3	0.55	0.03
V4	-0.03	-0.57
V5	-0.58	0.01
V6	-0.08	0.53
V7	0.55	-0.08
V8	-0.01	-0.47
V9	-0.58	0.06
V10	-0.04	0.56
V11	0.54	-0.02
V12	0.00	-0.58
V13	-0.51	0.03
V14	-0.01	0.64
V15	0.59	-0.01
V16	0.05	-0.53
V17	-0.57	0.05
V18	0.05	0.55
V19	0.59	0.03
V20	0.01	-0.60
V21	-0.55	-0.04
V22	0.03	0.57
V23	0.57	0.00
V24	0.03	-0.55

	Factor1	Factor2
SS loadings	3.82	3.74
Proportion Var	0.16	0.16
Cumulative Var	0.16	0.32

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 256.83 on 229 degrees of freedom.
The p-value is 0.0999

```
> 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.ss5 <- sem(model.ss, ss.cov, nsub)
> summary(sem.ss5, digits = 2)

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

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
7	-0.63	0.050	-12	0	V7 <--- F1
8	-0.55	0.053	-10	0	V8 <--- F2
9	0.70	0.053	13	0	V9 <--- F1
10	0.68	0.053	13	0	V10 <--- F2
11	-0.63	0.052	-12	0	V11 <--- F1
12	-0.67	0.051	-13	0	V12 <--- F2
13	0.60	0.053	11	0	V13 <--- F1
14	0.78	0.052	15	0	V14 <--- F2
15	-0.71	0.052	-14	0	V15 <--- F1
16	-0.63	0.054	-12	0	V16 <--- F2
17	0.67	0.051	13	0	V17 <--- F1
18	0.64	0.052	12	0	V18 <--- F2
19	-0.71	0.053	-14	0	V19 <--- F1
20	-0.73	0.053	-14	0	V20 <--- F2
21	0.65	0.052	12	0	V21 <--- F1
22	0.68	0.053	13	0	V22 <--- F2
23	-0.69	0.053	-13	0	V23 <--- F1
24	-0.66	0.053	-12	0	V24 <--- F2
25	0.68	0.051	13	0	V1 <--> V1
26	0.69	0.052	13	0	V2 <--> V2
27	0.78	0.054	15	0	V3 <--> V3
28	0.64	0.045	14	0	V4 <--> V4
29	0.63	0.044	14	0	V5 <--> V5
30	0.74	0.050	15	0	V6 <--> V6
31	0.67	0.046	15	0	V7 <--> V7
32	0.71	0.047	15	0	V8 <--> V8
33	0.71	0.049	14	0	V9 <--> V9
34	0.68	0.047	15	0	V10 <--> V10
35	0.70	0.048	15	0	V11 <--> V11
36	0.62	0.043	14	0	V12 <--> V12

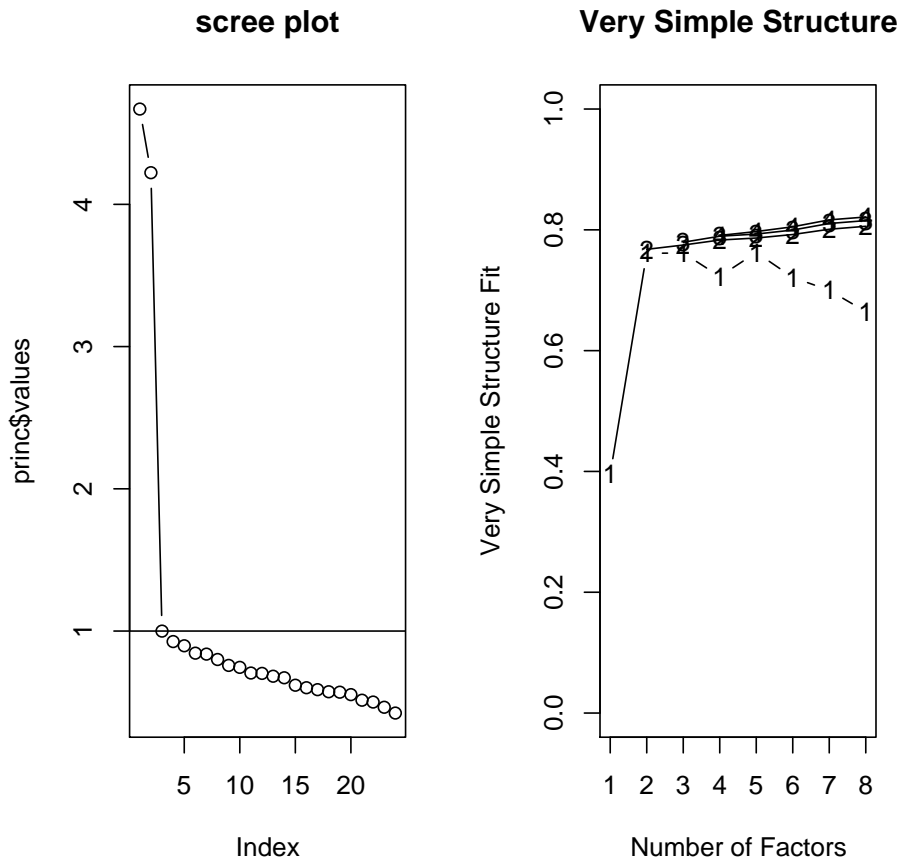


Figure 5.3: Determining the number of factors to extract from 24 variables generated with a simple structure with 5-point items. The left hand panel shows the scree plot, the right hand panel a VSS plot. Compare with Figure 5.4

37	0.75	0.051	15	0	V13 <--> V13
38	0.59	0.043	14	0	V14 <--> V14
39	0.69	0.048	14	0	V15 <--> V15
40	0.70	0.048	15	0	V16 <--> V16
41	0.68	0.047	14	0	V17 <--> V17
42	0.67	0.046	15	0	V18 <--> V18
43	0.71	0.049	14	0	V19 <--> V19
44	0.66	0.046	14	0	V20 <--> V20
45	0.72	0.049	15	0	V21 <--> V21
46	0.65	0.045	14	0	V22 <--> V22
47	0.72	0.050	14	0	V23 <--> V23
48	0.69	0.047	15	0	V24 <--> V24

Iterations = 16

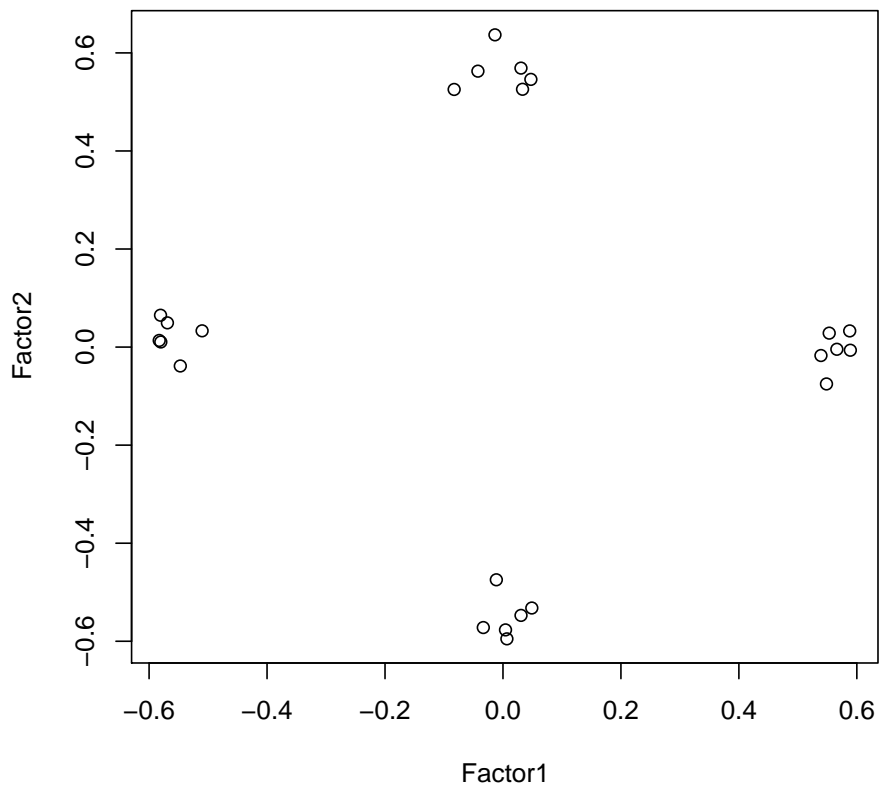


Figure 5.4: 24 variables loading on two factors for categorical items. Compare with Figure ??

5.4.2 3 categories of responses

Try this for 3 categories of response. Help the solution along by giving it appropriate start values.

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

Call:

```
factanal(x = ss.items, factors = 2)
```

Uniquenesses:

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
0.69	0.77	0.74	0.73	0.70	0.75	0.74	0.80	0.72	0.72	0.76	0.69	0.80	0.66	0.73
V16	V17	V18	V19	V20	V21	V22	V23	V24						
0.75	0.70	0.72	0.71	0.70	0.73	0.70	0.70	0.78						

Loadings:

	Factor1	Factor2
V1	-0.56	0.01
V2	0.01	0.48
V3	0.51	0.04
V4	-0.03	-0.52
V5	-0.55	0.00
V6	-0.07	0.50
V7	0.50	-0.10
V8	0.00	-0.45
V9	-0.52	0.01
V10	-0.05	0.53
V11	0.49	-0.01
V12	0.01	-0.56
V13	-0.45	0.04
V14	-0.01	0.58
V15	0.52	0.01
V16	0.04	-0.50
V17	-0.55	0.04
V18	0.05	0.53
V19	0.54	0.03
V20	0.05	-0.55
V21	-0.51	-0.05
V22	0.03	0.55
V23	0.55	0.04
V24	0.02	-0.47

	Factor1	Factor2
SS loadings	3.29	3.25
Proportion Var	0.14	0.14

Cumulative Var 0.14 0.27

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 277.76 on 229 degrees of freedom.
The p-value is 0.0152

```
> ss.cov <- cov(ss.items)
> ss.cor <- cor(ss.items)
> print(model.ss, digits = 2)
```

	path	label	initial	estimate
[1,]	"F1-> V1"	NA	"1"	
[2,]	"F2-> V2"	NA	"1"	
[3,]	"F1-> V3"	"3"	NA	
[4,]	"F2-> V4"	"4"	NA	
[5,]	"F1-> V5"	"5"	NA	
[6,]	"F2-> V6"	"6"	NA	
[7,]	"F1-> V7"	"7"	NA	
[8,]	"F2-> V8"	"8"	NA	
[9,]	"F1-> V9"	"9"	NA	
[10,]	"F2-> V10"	"10"	NA	
[11,]	"F1-> V11"	"11"	NA	
[12,]	"F2-> V12"	"12"	NA	
[13,]	"F1-> V13"	"13"	NA	
[14,]	"F2-> V14"	"14"	NA	
[15,]	"F1-> V15"	"15"	NA	
[16,]	"F2-> V16"	"16"	NA	
[17,]	"F1-> V17"	"17"	NA	
[18,]	"F2-> V18"	"18"	NA	
[19,]	"F1-> V19"	"19"	NA	
[20,]	"F2-> V20"	"20"	NA	
[21,]	"F1-> V21"	"21"	NA	
[22,]	"F2-> V22"	"22"	NA	
[23,]	"F1-> V23"	"23"	NA	
[24,]	"F2-> V24"	"24"	NA	
[25,]	"V1<-> V1"	"25"	NA	
[26,]	"V2<-> V2"	"26"	NA	
[27,]	"V3<-> V3"	"27"	NA	
[28,]	"V4<-> V4"	"28"	NA	
[29,]	"V5<-> V5"	"29"	NA	
[30,]	"V6<-> V6"	"30"	NA	
[31,]	"V7<-> V7"	"31"	NA	
[32,]	"V8<-> V8"	"32"	NA	
[33,]	"V9<-> V9"	"33"	NA	
[34,]	"V10<-> V10"	"34"	NA	
[35,]	"V11<-> V11"	"35"	NA	
[36,]	"V12<-> V12"	"36"	NA	
[37,]	"V13<-> V13"	"37"	NA	
[38,]	"V14<-> V14"	"38"	NA	
[39,]	"V15<-> V15"	"39"	NA	
[40,]	"V16<-> V16"	"40"	NA	
[41,]	"V17<-> V17"	"41"	NA	
[42,]	"V18<-> V18"	"42"	NA	

```
[43,] "V19<-> V19" "43" NA
[44,] "V20<-> V20" "44" NA
[45,] "V21<-> V21" "45" NA
[46,] "V22<-> V22" "46" NA
[47,] "V23<-> V23" "47" NA
[48,] "V24<-> V24" "48" NA
[49,] "F1 <-> F1" NA "1"
[50,] "F2 <-> F2" NA "1"
```

```
> sem.ss3 <- sem(model.ss, ss.cor, nsub)
> summary(sem.ss3, digits = 2)
```

```
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
```

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-7.020	-1.320	-0.081	0.022	1.280	5.850

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
3	-0.59	0.054	-11.0	0	V3 <--- F1
4	-0.62	0.055	-11.3	0	V4 <--- F2
5	0.65	0.053	12.4	0	V5 <--- F1
6	0.60	0.055	10.9	0	V6 <--- F2
7	-0.59	0.054	-10.9	0	V7 <--- F1
8	-0.54	0.056	-9.7	0	V8 <--- F2
9	0.61	0.054	11.3	0	V9 <--- F1
10	0.63	0.055	11.4	0	V10 <--- F2
11	-0.58	0.054	-10.7	0	V11 <--- F1
12	-0.66	0.054	-12.3	0	V12 <--- F2
13	0.53	0.055	9.6	0	V13 <--- F1
14	0.70	0.053	13.1	0	V14 <--- F2
15	-0.61	0.054	-11.4	0	V15 <--- F1
16	-0.59	0.055	-10.7	0	V16 <--- F2
17	0.65	0.053	12.2	0	V17 <--- F1
18	0.62	0.055	11.3	0	V18 <--- F2
19	-0.64	0.053	-12.0	0	V19 <--- F1
20	-0.66	0.054	-12.1	0	V20 <--- F2
21	0.61	0.054	11.3	0	V21 <--- F1
22	0.66	0.054	12.2	0	V22 <--- F2
23	-0.65	0.053	-12.3	0	V23 <--- F1
24	-0.57	0.055	-10.3	0	V24 <--- F2
25	0.69	0.053	12.8	0	V1 <--> V1
26	0.80	0.060	13.2	0	V2 <--> V2
27	0.75	0.051	14.7	0	V3 <--> V3

28	0.73	0.050	14.5	0	V4 <--> V4
29	0.70	0.049	14.3	0	V5 <--> V5
30	0.75	0.051	14.6	0	V6 <--> V6
31	0.75	0.051	14.7	0	V7 <--> V7
32	0.80	0.053	14.9	0	V8 <--> V8
33	0.74	0.051	14.6	0	V9 <--> V9
34	0.73	0.050	14.5	0	V10 <--> V10
35	0.76	0.052	14.7	0	V11 <--> V11
36	0.70	0.049	14.3	0	V12 <--> V12
37	0.80	0.054	15.0	0	V13 <--> V13
38	0.66	0.047	14.0	0	V14 <--> V14
39	0.74	0.051	14.6	0	V15 <--> V15
40	0.76	0.052	14.7	0	V16 <--> V16
41	0.70	0.049	14.4	0	V17 <--> V17
42	0.74	0.051	14.5	0	V18 <--> V18
43	0.71	0.049	14.4	0	V19 <--> V19
44	0.70	0.049	14.3	0	V20 <--> V20
45	0.74	0.051	14.6	0	V21 <--> V21
46	0.70	0.049	14.3	0	V22 <--> V22
47	0.70	0.049	14.3	0	V23 <--> V23
48	0.78	0.052	14.8	0	V24 <--> V24

Iterations = 13

5.4.3 dichotomous items

This is the worst case scenario, in which items are scored as either yes or no. I can not get the sem of the covariance matrix to work, but I can for the correlation matrix.

```
> set.seed(42)
> nsub = 500
> model.ss[1, 2] <- NA
> model.ss[1, 3] <- 1
> model.ss[2, 2] <- NA
> model.ss[2, 3] <- 1
> 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)
```

Call:

```
factanal(x = ss.items, factors = 2)
```

Uniquenesses:

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
0.78	0.80	0.81	0.81	0.79	0.77	0.79	0.86	0.78	0.77	0.83	0.78	0.85	0.76	0.82
V16	V17	V18	V19	V20	V21	V22	V23	V24						
0.83	0.79	0.83	0.79	0.77	0.81	0.79	0.82	0.82						

Loadings:

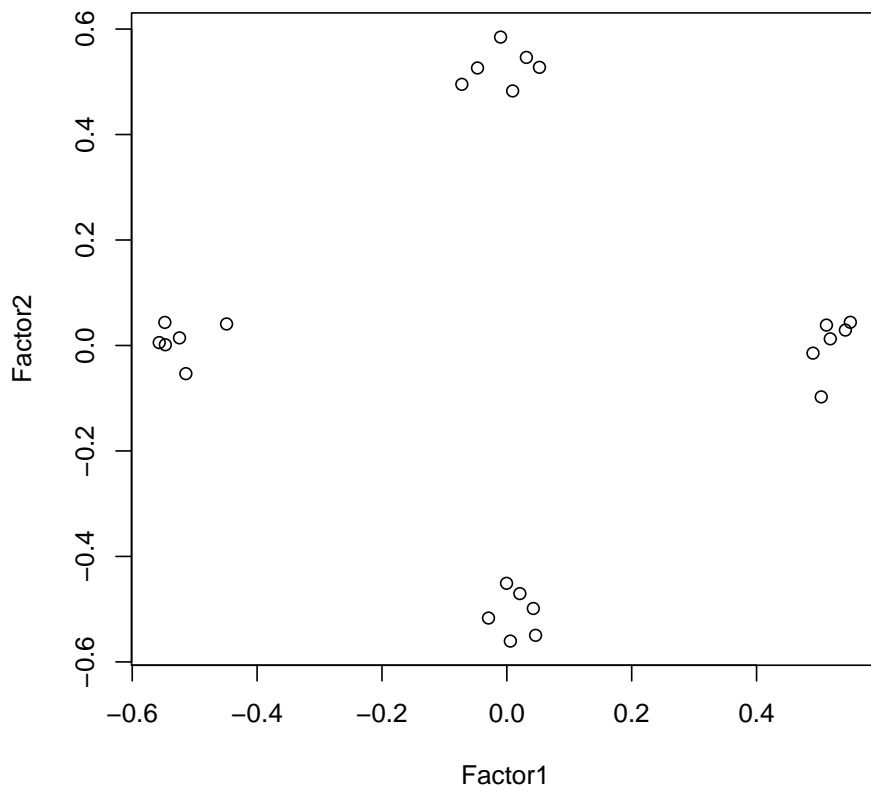


Figure 5.5: 24 variables, simple structure. Items are constrained to have 3 categories

	Factor1	Factor2
V1	-0.02	0.47
V2	0.45	0.02
V3	0.05	-0.44
V4	-0.44	0.00
V5	0.02	0.45
V6	0.48	0.07
V7	-0.13	-0.43
V8	-0.37	-0.05
V9	-0.01	0.47
V10	0.48	0.02
V11	0.04	-0.41
V12	-0.47	-0.04
V13	0.05	0.39
V14	0.49	-0.03
V15	-0.02	-0.42
V16	-0.42	0.00
V17	0.05	0.45
V18	0.41	-0.02
V19	0.02	-0.46
V20	-0.48	0.00
V21	-0.01	0.43
V22	0.45	-0.05
V23	0.04	-0.42
V24	-0.43	-0.01

	Factor1	Factor2
SS loadings	2.43	2.31
Proportion Var	0.10	0.10
Cumulative Var	0.10	0.20

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 245.65 on 229 degrees of freedom.
The p-value is 0.214

```

> ss.cor <- cor(ss.items)
> sem.ss2 <- sem(model.ss, ss.cor, nsub)
> summary(sem.ss2, digits = 2)

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

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
3	-0.50	0.059	-8.4	0.0e+00	V3 <--- F1
4	-0.53	0.058	-9.0	0.0e+00	V4 <--- F2
5	0.54	0.058	9.2	0.0e+00	V5 <--- F1
6	0.56	0.058	9.7	0.0e+00	V6 <--- F2
7	-0.52	0.059	-8.8	0.0e+00	V7 <--- F1
8	-0.45	0.059	-7.5	6.0e-14	V8 <--- F2
9	0.54	0.059	9.3	0.0e+00	V9 <--- F1
10	0.58	0.058	10.0	0.0e+00	V10 <--- F2
11	-0.47	0.059	-7.9	3.3e-15	V11 <--- F1
12	-0.55	0.058	-9.4	0.0e+00	V12 <--- F2
13	0.46	0.059	7.8	4.9e-15	V13 <--- F1
14	0.58	0.058	10.0	0.0e+00	V14 <--- F2
15	-0.49	0.059	-8.3	0.0e+00	V15 <--- F1
16	-0.48	0.059	-8.0	8.9e-16	V16 <--- F2
17	0.54	0.059	9.2	0.0e+00	V17 <--- F1
18	0.48	0.059	8.0	8.9e-16	V18 <--- F2
19	-0.53	0.059	-9.0	0.0e+00	V19 <--- F1
20	-0.56	0.058	-9.5	0.0e+00	V20 <--- F2
21	0.52	0.059	8.8	0.0e+00	V21 <--- F1
22	0.54	0.058	9.3	0.0e+00	V22 <--- F2
23	-0.49	0.059	-8.3	0.0e+00	V23 <--- F1
24	-0.50	0.059	-8.5	0.0e+00	V24 <--- F2
25	0.76	0.063	12.1	0.0e+00	V1 <--> V1
26	0.80	0.064	12.4	0.0e+00	V2 <--> V2
27	0.83	0.056	14.6	0.0e+00	V3 <--> V3
28	0.81	0.055	14.6	0.0e+00	V4 <--> V4
29	0.80	0.055	14.5	0.0e+00	V5 <--> V5
30	0.78	0.054	14.4	0.0e+00	V6 <--> V6
31	0.81	0.056	14.6	0.0e+00	V7 <--> V7
32	0.86	0.058	15.0	0.0e+00	V8 <--> V8
33	0.79	0.055	14.4	0.0e+00	V9 <--> V9
34	0.77	0.054	14.3	0.0e+00	V10 <--> V10
35	0.85	0.057	14.8	0.0e+00	V11 <--> V11
36	0.79	0.055	14.5	0.0e+00	V12 <--> V12
37	0.85	0.057	14.9	0.0e+00	V13 <--> V13
38	0.77	0.054	14.3	0.0e+00	V14 <--> V14
39	0.83	0.056	14.7	0.0e+00	V15 <--> V15
40	0.84	0.057	14.8	0.0e+00	V16 <--> V16
41	0.80	0.055	14.5	0.0e+00	V17 <--> V17
42	0.84	0.057	14.9	0.0e+00	V18 <--> V18
43	0.80	0.056	14.5	0.0e+00	V19 <--> V19
44	0.79	0.055	14.4	0.0e+00	V20 <--> V20
45	0.81	0.056	14.6	0.0e+00	V21 <--> V21
46	0.80	0.055	14.5	0.0e+00	V22 <--> V22
47	0.83	0.057	14.7	0.0e+00	V23 <--> V23
48	0.83	0.056	14.7	0.0e+00	V24 <--> V24

Iterations = 10

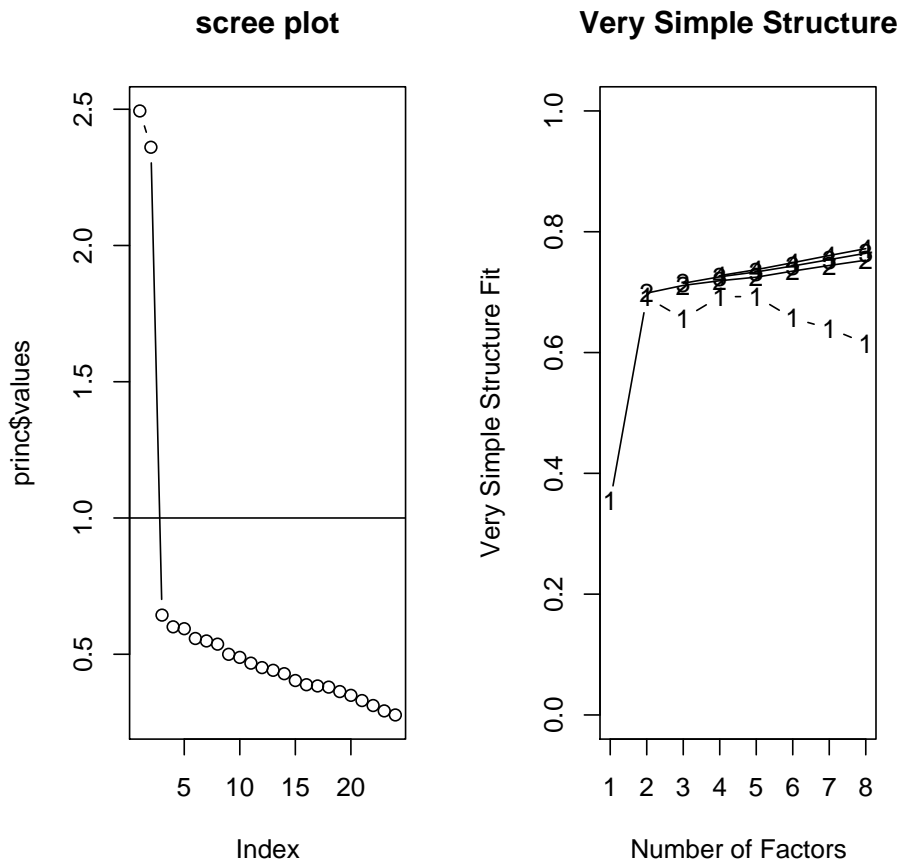


Figure 5.6: Determining the number of factors to extract from 24 variables generated with a simple structure for dichotomous items. The left hand panel shows the scree plot, the right hand panel a VSS plot. Compare with Figures 5.4 and 5.4.1

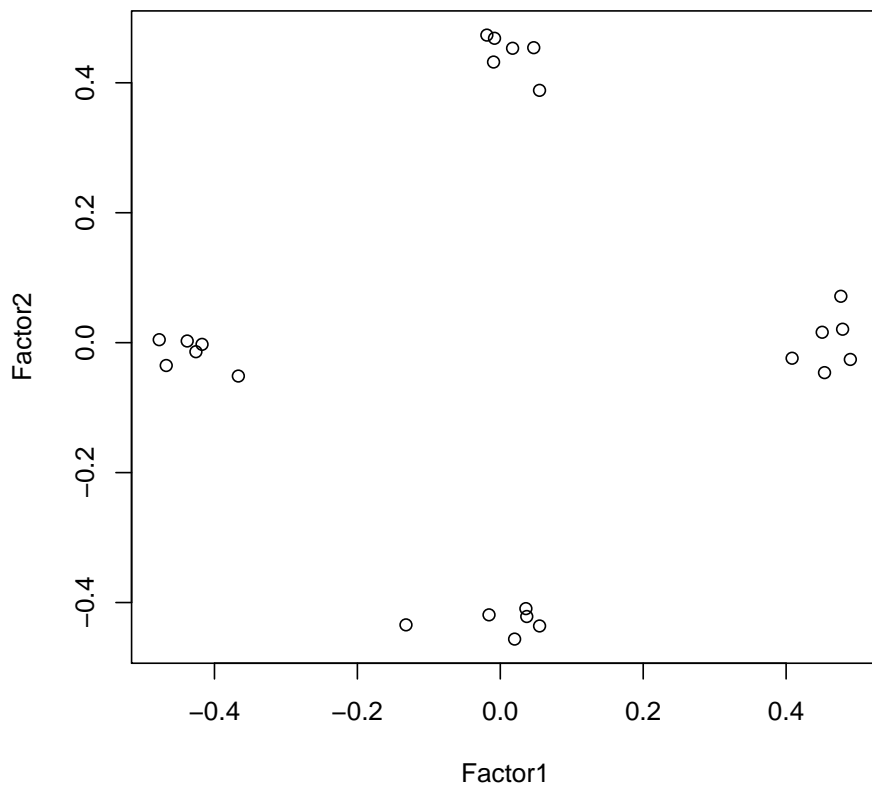


Figure 5.7: 24 variables, simple structure. Items are constrained to be dichotomous.

5.5 Circumplex structure - normal items

We now repeat the data generation, EFA and CFA for circumplex data. Exploratory Factor Analysis correctly suggests that we have a two dimensional structure and identifies the item loadings quite well. As is discussed by Acton and Revelle (2004), a circumplex structure will be relatively insensitive to rotation, e.g., the varimax criterion will not change as we rotate. In fact, this is one of the tests for circumplex structure versus simple structure suggested by Acton and Revelle.

```
> 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)
```

Call:

```
factanal(x = circ.items, factors = 2)
```

Uniquenesses:

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
0.61	0.64	0.68	0.63	0.63	0.64	0.71	0.57	0.66	0.68	0.63	0.59	0.60	0.66	0.69
V16	V17	V18	V19	V20	V21	V22	V23	V24						
0.63	0.66	0.70	0.69	0.64	0.63	0.67	0.64	0.64						

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
V9	0.37	0.45
V10	0.44	0.35
V11	0.57	0.21
V12	0.63	0.09
V13	0.63	-0.07
V14	0.57	-0.14
V15	0.46	-0.31
V16	0.42	-0.44
V17	0.31	-0.50
V18	0.10	-0.54
V19	-0.03	-0.56
V20	-0.20	-0.57
V21	-0.33	-0.51
V22	-0.45	-0.37
V23	-0.57	-0.19
V24	-0.59	-0.08

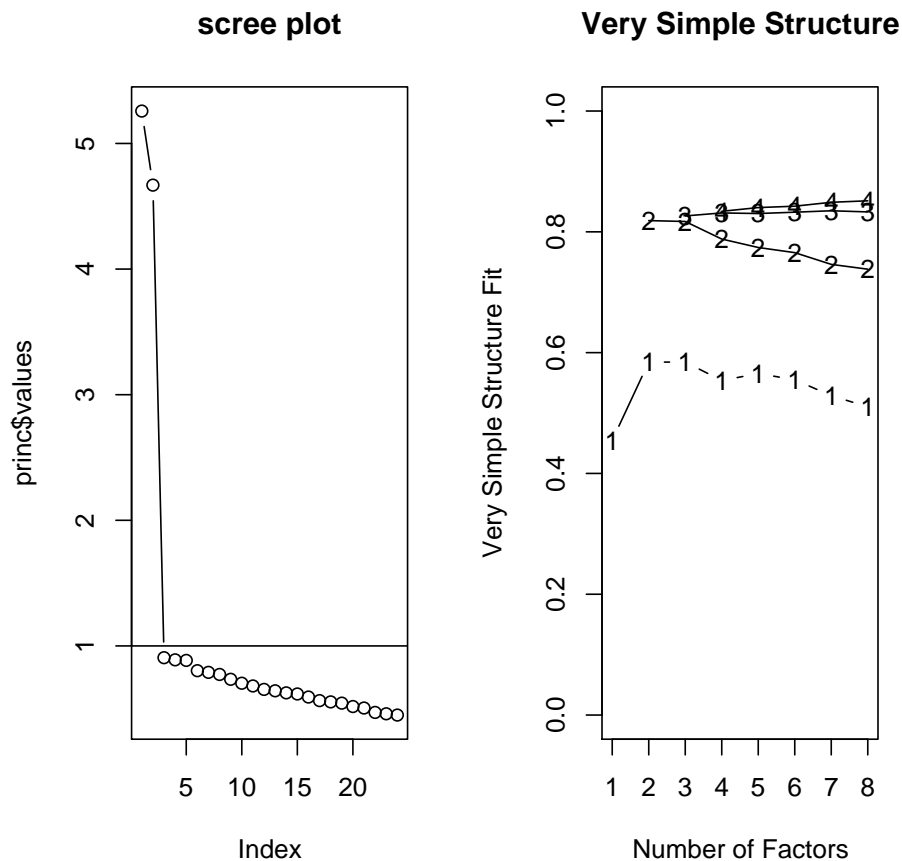


Figure 5.8: Determining the number of factors to extract from 24 variables generated with a circumplex structure. The left hand panel shows the scree plot, the right hand panel a VSS plot. Notice the inflection at two factors, suggesting a two factor solution

	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

5.5.1 Fitting a circumplex data set with a simple structure model

We can compare the results of this exploratory factor analysis with a confirmatory factor analysis using the sem package. As can be seen below, the model we used in the previous examples fits very poorly and should be revised. What is particularly interesting is that all of

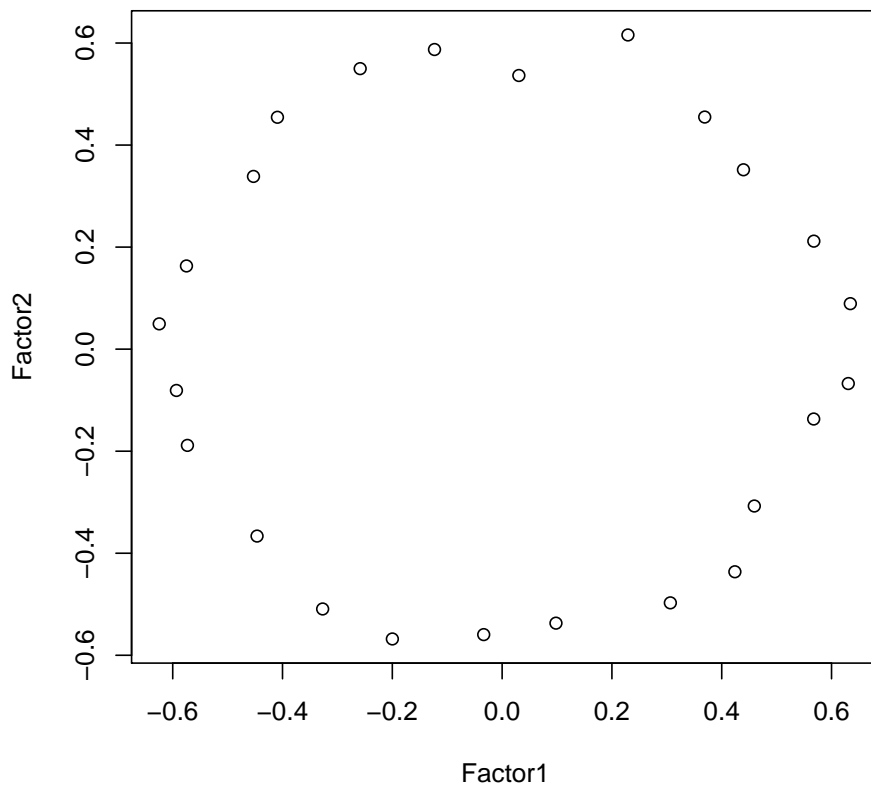


Figure 5.9: Factor loadings for 24 items on two dimensions. Given that the data were generated to reflect uniform locations around a two dimensional space, the circular ordering of loadings is not surprising.

the paths are very large, even though the model is terrible.

```
> model.cs <- modelmat(24)
> cs.cov <- cov(circ.items)
> sem.cs <- sem(model.cs, cs.cov, nsub)
> summary(sem.cs, digits = 2)

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

Parameter Estimates

	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
12	0.654	0.048	13.53	0.0e+00	V12 <--- F2
13	0.681	0.050	13.58	0.0e+00	V13 <--- F1
14	0.539	0.048	11.17	0.0e+00	V14 <--- F2
15	0.494	0.052	9.48	0.0e+00	V15 <--- F1
16	0.407	0.060	6.82	8.9e-12	V16 <--- F2
17	0.321	0.057	5.66	1.5e-08	V17 <--- F1
18	0.062	0.063	0.99	3.2e-01	V18 <--- F2
19	0.026	0.058	0.44	6.6e-01	V19 <--- F1
20	-0.186	0.061	-3.06	2.2e-03	V20 <--- F2
21	-0.270	0.061	-4.42	1.0e-05	V21 <--- F1
22	-0.418	0.057	-7.33	2.3e-13	V22 <--- F2
23	-0.520	0.052	-9.99	0.0e+00	V23 <--- F1
24	-0.602	0.047	-12.70	0.0e+00	V24 <--- F2
25	0.590	0.049	11.93	0.0e+00	V1 <--> V1
26	0.686	0.053	12.87	0.0e+00	V2 <--> V2
27	0.884	0.064	13.70	0.0e+00	V3 <--> V3
28	0.878	0.065	13.60	0.0e+00	V4 <--> V4
29	0.841	0.057	14.64	0.0e+00	V5 <--> V5
30	1.085	0.069	15.61	0.0e+00	V6 <--> V6

```

31 0.905 0.057 15.78 0.0e+00 V7 <--> V7
32 1.076 0.073 14.85 0.0e+00 V8 <--> V8
33 0.942 0.065 14.59 0.0e+00 V9 <--> V9
34 0.745 0.055 13.64 0.0e+00 V10 <--> V10
35 0.692 0.054 12.81 0.0e+00 V11 <--> V11
36 0.597 0.051 11.61 0.0e+00 V12 <--> V12
37 0.682 0.056 12.18 0.0e+00 V13 <--> V13
38 0.645 0.050 12.80 0.0e+00 V14 <--> V14
39 0.778 0.058 13.52 0.0e+00 V15 <--> V15
40 0.878 0.065 13.60 0.0e+00 V16 <--> V16
41 0.891 0.061 14.58 0.0e+00 V17 <--> V17
42 0.995 0.063 15.73 0.0e+00 V18 <--> V18
43 0.932 0.059 15.78 0.0e+00 V19 <--> V19
44 0.933 0.061 15.27 0.0e+00 V20 <--> V20
45 1.009 0.068 14.93 0.0e+00 V21 <--> V21
46 0.859 0.062 13.75 0.0e+00 V22 <--> V22
47 0.739 0.057 13.02 0.0e+00 V23 <--> V23
48 0.628 0.050 12.59 0.0e+00 V24 <--> V24

```

```
Iterations = 34
```

5.5.2 An alternative model

An examination of the exploratory factor analysis suggests that a two factor model might work, but with a very different pattern of loadings than seen before. It seems as if the items can be grouped into four sets of 6, best represented by two dimensions: Such an alternative model can be formed by creating a simple function, `modelcirc`, to save us the time in writing out all 48 equations. but still does not provide an answer unless we specify one loading for each factor to be 1.

```

> modelcirc <- function(n = 24) {
+   mat = matrix(rep(NA, 3 * (n * 2 + 2)), ncol = 3)
+   for (i in 1:24) {
+     mat[i, 1] <- paste("F", 1 + trunc(i/6)%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)
+ }
> model.circ <- modelcirc(24)

```

```
> print(model.circ)
```

	path	label	initial	estimate
[1,]	"F1-> V1"	"1"	NA	
[2,]	"F1-> V2"	"2"	NA	
[3,]	"F1-> V3"	"3"	NA	
[4,]	"F1-> V4"	"4"	NA	
[5,]	"F1-> V5"	"5"	NA	
[6,]	"F2-> V6"	"6"	NA	
[7,]	"F2-> V7"	"7"	NA	
[8,]	"F2-> V8"	"8"	NA	
[9,]	"F2-> V9"	"9"	NA	
[10,]	"F2-> V10"	"10"	NA	
[11,]	"F2-> V11"	"11"	NA	
[12,]	"F1-> V12"	"12"	NA	
[13,]	"F1-> V13"	"13"	NA	
[14,]	"F1-> V14"	"14"	NA	
[15,]	"F1-> V15"	"15"	NA	
[16,]	"F1-> V16"	"16"	NA	
[17,]	"F1-> V17"	"17"	NA	
[18,]	"F2-> V18"	"18"	NA	
[19,]	"F2-> V19"	"19"	NA	
[20,]	"F2-> V20"	"20"	NA	
[21,]	"F2-> V21"	"21"	NA	
[22,]	"F2-> V22"	"22"	NA	
[23,]	"F2-> V23"	"23"	NA	
[24,]	"F1-> V24"	"24"	NA	
[25,]	"V1<-> V1"	"25"	NA	
[26,]	"V2<-> V2"	"26"	NA	
[27,]	"V3<-> V3"	"27"	NA	
[28,]	"V4<-> V4"	"28"	NA	
[29,]	"V5<-> V5"	"29"	NA	
[30,]	"V6<-> V6"	"30"	NA	
[31,]	"V7<-> V7"	"31"	NA	
[32,]	"V8<-> V8"	"32"	NA	
[33,]	"V9<-> V9"	"33"	NA	
[34,]	"V10<-> V10"	"34"	NA	
[35,]	"V11<-> V11"	"35"	NA	
[36,]	"V12<-> V12"	"36"	NA	
[37,]	"V13<-> V13"	"37"	NA	
[38,]	"V14<-> V14"	"38"	NA	
[39,]	"V15<-> V15"	"39"	NA	
[40,]	"V16<-> V16"	"40"	NA	
[41,]	"V17<-> V17"	"41"	NA	
[42,]	"V18<-> V18"	"42"	NA	
[43,]	"V19<-> V19"	"43"	NA	
[44,]	"V20<-> V20"	"44"	NA	
[45,]	"V21<-> V21"	"45"	NA	
[46,]	"V22<-> V22"	"46"	NA	
[47,]	"V23<-> V23"	"47"	NA	
[48,]	"V24<-> V24"	"48"	NA	
[49,]	"F1 <-> F1"	NA	"1"	
[50,]	"F2 <-> F2"	NA	"1"	

```

> model.circ[1, 2] <- NA
> model.circ[1, 3] <- 1
> model.circ[7, 2] <- NA
> model.circ[7, 3] <- 1
> cs.cov <- cov(circ.items)
> sem.circ <- sem(model.circ, cs.cov, nsub)
> summary(sem.circ, digits = 2)

```

```

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

```

Normalized Residuals

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-8.160	-2.650	-0.248	-0.021	2.720	8.070

Parameter Estimates

	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
5	0.49	0.052	9.5	0	V5 <--- 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
10	0.58	0.054	10.7	0	V10 <--- F2
11	0.50	0.057	8.9	0	V11 <--- 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
16	-0.67	0.053	-12.7	0	V16 <--- F1
17	-0.54	0.054	-10.0	0	V17 <--- F1
18	-0.53	0.057	-9.1	0	V18 <--- F2
19	-0.60	0.054	-11.2	0	V19 <--- F2
20	-0.73	0.053	-13.8	0	V20 <--- F2
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
25	0.66	0.051	13.0	0	V1 <--> V1
26	0.66	0.046	14.3	0	V2 <--> V2
27	0.76	0.053	14.4	0	V3 <--> V3
28	0.73	0.050	14.5	0	V4 <--> V4
29	0.75	0.050	15.0	0	V5 <--> V5
30	0.87	0.059	14.8	0	V6 <--> V6

31	0.74	0.056	13.1	0	V7 <--> V7
32	0.66	0.049	13.4	0	V8 <--> V8
33	0.71	0.050	14.2	0	V9 <--> V9
34	0.70	0.047	14.7	0	V10 <--> V10
35	0.79	0.052	15.0	0	V11 <--> V11
36	0.73	0.050	14.5	0	V12 <--> V12
37	0.73	0.052	14.0	0	V13 <--> V13
38	0.61	0.043	14.2	0	V14 <--> V14
39	0.73	0.050	14.6	0	V15 <--> V15
40	0.72	0.050	14.4	0	V16 <--> V16
41	0.78	0.052	14.9	0	V17 <--> V17
42	0.82	0.054	15.0	0	V18 <--> V18
43	0.69	0.047	14.6	0	V19 <--> V19
44	0.61	0.044	13.9	0	V20 <--> V20
45	0.69	0.050	13.9	0	V21 <--> V21
46	0.75	0.052	14.5	0	V22 <--> V22
47	0.84	0.056	15.1	0	V23 <--> V23
48	0.74	0.050	14.7	0	V24 <--> V24

Iterations = 15

As would be expected, this is still not a very good fit, although it is much better than the fit in 5.5.1 for we are fitting a simple structure model to a circumplex data set. Although we are modeling each item as of complexity one, in reality some of the items are of complexity two. One way to model this additional complexity is to allow for correlated errors between those variables at the 45 and 135 degree locations.

5.6 Simple Structure - categorical and skewed items

A recurring debate in the emotion literature is the proper structure of affect and whether valence is indeed bipolar. Part of the controversy arises from the way affect is measured, with some using unipolar scales (not at all happy, somewhat happy, very happy) whereas others use bipolar (Very sad, somewhat sad, somewhat happy, very happy.) It has been claimed that by using unipolar scales we are introducing skew since any person who is feeling very sad, or somewhat sad will give a 0 on the happiness scale. The example is measuring temperature with a bipolar versus a unipolar scale.

This issue has been addressed very thoroughly by Rafaeli and Revelle (2006) who suggest that happiness and sadness are not bipolar opposites. In particular, Rafaeli and Revelle examine the effect of skew. Here we use the circ.sim simulation function again, to introduce serious skew into our data.

5.6.1 Two dimensions with 4 point scales, differing in skew

circ.sim is used with four point scales, with any values less than 0 being cut to 0. This leads to substantial skew for these items. (See Figure 5.6.1). Although the factor analysis loadings recover the structure very well (Figure 5.6.1

```

> 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)

```

Call:

```
factanal(x = skew.items, factors = 2)
```

Uniquenesses:

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
0.79	0.74	0.70	0.85	0.77	0.77	0.77	0.89	0.76	0.73	0.77	0.90	0.77	0.67	0.75
V16	V17	V18	V19	V20	V21	V22	V23	V24						
0.85	0.78	0.76	0.75	0.92	0.78	0.65	0.79	0.88						

Loadings:

	Factor1	Factor2
V1	-0.45	0.04
V2	0.02	0.51
V3	0.55	-0.03
V4	-0.06	-0.38
V5	-0.48	0.03
V6	-0.07	0.48
V7	0.48	0.04
V8	-0.01	-0.33
V9	-0.49	-0.01
V10	0.03	0.52
V11	0.48	-0.05
V12	0.03	-0.32
V13	-0.48	0.03
V14	-0.03	0.57
V15	0.50	-0.06
V16	0.03	-0.38
V17	-0.46	0.01
V18	-0.04	0.48
V19	0.50	0.05
V20	0.09	-0.27
V21	-0.46	-0.10
V22	0.05	0.59
V23	0.45	-0.04
V24	-0.02	-0.34

	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

Using our simple structure model (from section 5.4.1) on the covariance matrix shows the structure as well. We find that the χ^2 value for the null model is smaller than for the

```
> pairs.panels(skew.items[, 1:6])
```

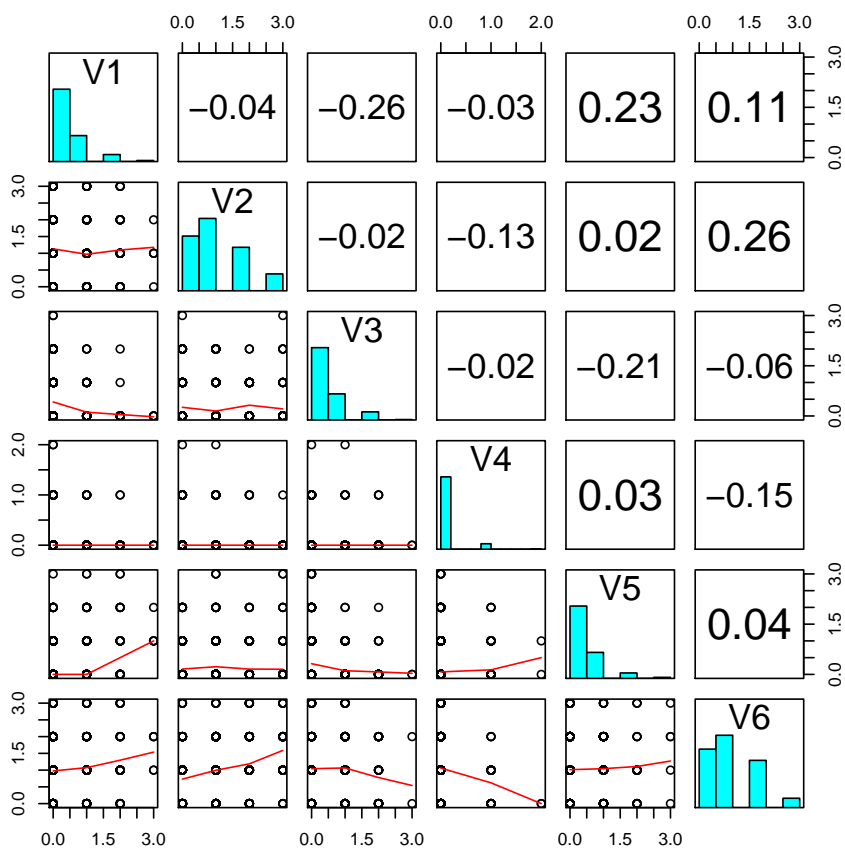


Figure 5.10: SPLOM of the first 6 variables showing the effect of skew. Note how the correlations of items with opposite skew are very attenuated.

```
> plot(fcs$loadings)
```

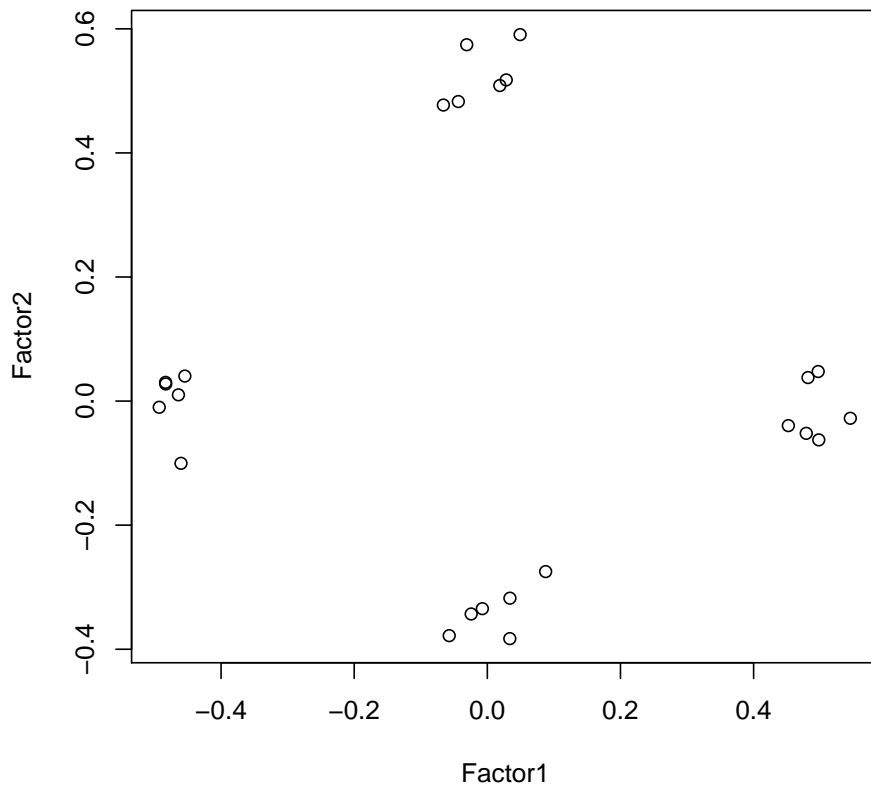


Figure 5.11: The factor structure of very skewed items recovers the space quite well, at least in terms of angular location. The loadings are less than they should be given the data generation algorithm.

non-skewed data and the fit is not nearly as good. The problem is that the differences in skew between the positive and negatively keyed items is creating the functional equivalent of method or group factors. That is, items loading on the latent factors with the same sign are much more highly correlated than those with an opposite sign.

```
> skew.cov <- cov(skew.items)
> sem.skew <- sem(model.ss, skew.cov, nsub)
> summary(sem.skew, digits = 2)

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
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-11.000	-1.010	0.098	0.170	1.440	8.710

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
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
10	0.571	0.0523	10.9	0.0e+00	V10 <--- F2
11	-0.430	0.0440	-9.8	0.0e+00	V11 <--- F1
12	-0.106	0.0165	-6.4	1.3e-10	V12 <--- F2
13	0.411	0.0404	10.2	0.0e+00	V13 <--- F1
14	0.662	0.0547	12.1	0.0e+00	V14 <--- F2
15	-0.463	0.0447	-10.4	0.0e+00	V15 <--- F1
16	-0.134	0.0179	-7.5	7.1e-14	V16 <--- F2
17	0.420	0.0428	9.8	0.0e+00	V17 <--- F1
18	0.523	0.0534	9.8	0.0e+00	V18 <--- F2
19	-0.463	0.0438	-10.6	0.0e+00	V19 <--- F1
20	-0.078	0.0149	-5.2	1.8e-07	V20 <--- F2
21	0.416	0.0433	9.6	0.0e+00	V21 <--- 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
25	0.419	0.0357	11.7	0.0e+00	V1 <--> V1
26	0.639	0.0536	11.9	0.0e+00	V2 <--> V2
27	0.296	0.0211	14.1	0.0e+00	V3 <--> V3
28	0.070	0.0046	15.0	0.0e+00	V4 <--> V4
29	0.309	0.0212	14.6	0.0e+00	V5 <--> V5
30	0.595	0.0412	14.4	0.0e+00	V6 <--> V6

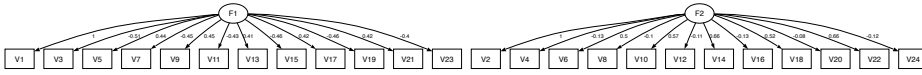


Figure 5.12: A two dimensional solution does not fit very well.

```

31  0.329   0.0225   14.6   0.0e+00  V7 <--> V7
32  0.056   0.0037   15.2   0.0e+00  V8 <--> V8
33  0.295   0.0204   14.5   0.0e+00  V9 <--> V9
34  0.599   0.0425   14.1   0.0e+00  V10 <--> V10
35  0.313   0.0214   14.6   0.0e+00  V11 <--> V11
36  0.066   0.0044   15.2   0.0e+00  V12 <--> V12
37  0.261   0.0180   14.5   0.0e+00  V13 <--> V13
38  0.626   0.0461   13.6   0.0e+00  V14 <--> V14
39  0.316   0.0218   14.5   0.0e+00  V15 <--> V15
40  0.075   0.0050   15.0   0.0e+00  V16 <--> V16
41  0.296   0.0202   14.6   0.0e+00  V17 <--> V17
42  0.640   0.0443   14.4   0.0e+00  V18 <--> V18
43  0.303   0.0210   14.4   0.0e+00  V19 <--> V19
44  0.055   0.0036   15.4   0.0e+00  V20 <--> V20
45  0.305   0.0207   14.7   0.0e+00  V21 <--> V21
46  0.550   0.0414   13.3   0.0e+00  V22 <--> V22
47  0.316   0.0213   14.8   0.0e+00  V23 <--> V23
48  0.074   0.0049   15.1   0.0e+00  V24 <--> V24

```

Iterations = 74

5.6.2 An alternative model of two bipolar dimensions

We can revise the model to take into account the bipolar nature of the data by modeling it in terms of four factors grouped into two sets of two highly correlated factors. This solution is very good (in terms of χ^2) and RMSEA). The loadings, however, look very small until we realize that modeling covariances produces smaller path coefficients than modeling the correlations. Standardizing the loadings makes this point clearer.

```

> modelmat4 <- function(n = 24) {
+   mat = matrix(rep(NA, 3 * (n * 2 + 6)), ncol = 3)
+   for (i in 1:n) {
+     mat[i, 1] <- paste("F", i%4 + 1, "-> 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 + 3, 1] <- "F3 <-> F3"

```

```

+   mat[n * 2 + 4, 1] <- "F4 <-> F4"
+   mat[n * 2 + 5, 1] <- "F1 <-> F3"
+   mat[n * 2 + 6, 1] <- "F2 <-> F4"
+   mat[n * 2 + 1, 3] <- mat[n * 2 + 2, 3] <- mat[n * 2 + 3,
+       3] <- mat[n * 2 + 4, 3] <- 1
+   mat[n * 2 + 5, 2] <- 2 * n + 1
+   mat[n * 2 + 6, 2] <- 2 * n + 2
+   return(mat)
+ }

```

```

> model.4 <- modelmat4()
> sem.skew4 <- sem(model.4, skew.cov, nsub)
> summary(sem.skew4, digits = 2)

```

```

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

```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-2.8e+00	-6.1e-01	-1.3e-06	-2.7e-02	5.4e-01	3.3e+00

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
1	0.314	0.0325	9.7	0.0e+00	V1 <--- F2
2	0.488	0.0454	10.8	0.0e+00	V2 <--- F3
3	0.370	0.0309	12.0	0.0e+00	V3 <--- F4
4	0.138	0.0157	8.8	0.0e+00	V4 <--- F1
5	0.327	0.0310	10.5	0.0e+00	V5 <--- F2
6	0.416	0.0430	9.7	0.0e+00	V6 <--- F3
7	0.335	0.0316	10.6	0.0e+00	V7 <--- F4
8	0.098	0.0137	7.1	9.9e-13	V8 <--- F1
9	0.331	0.0304	10.9	0.0e+00	V9 <--- F2
10	0.474	0.0439	10.8	0.0e+00	V10 <--- F3
11	0.317	0.0307	10.3	0.0e+00	V11 <--- F4
12	0.093	0.0149	6.2	4.7e-10	V12 <--- F1
13	0.292	0.0286	10.2	0.0e+00	V13 <--- F2
14	0.565	0.0458	12.3	0.0e+00	V14 <--- F3
15	0.337	0.0313	10.8	0.0e+00	V15 <--- F4
16	0.146	0.0163	8.9	0.0e+00	V16 <--- F1
17	0.308	0.0302	10.2	0.0e+00	V17 <--- F2
18	0.444	0.0444	10.0	0.0e+00	V18 <--- F3
19	0.325	0.0308	10.5	0.0e+00	V19 <--- F4
20	0.073	0.0134	5.4	6.0e-08	V20 <--- F1
21	0.301	0.0306	9.8	0.0e+00	V21 <--- F2
22	0.557	0.0436	12.8	0.0e+00	V22 <--- F3
23	0.299	0.0306	9.8	0.0e+00	V23 <--- F4

24	0.113	0.0158	7.2	8.1e-13	V24 <--- F1
25	0.335	0.0237	14.1	0.0e+00	V1 <--> V1
26	0.633	0.0462	13.7	0.0e+00	V2 <--> V2
27	0.282	0.0214	13.2	0.0e+00	V3 <--> V3
28	0.061	0.0048	12.7	0.0e+00	V4 <--> V4
29	0.294	0.0214	13.7	0.0e+00	V5 <--> V5
30	0.591	0.0417	14.2	0.0e+00	V6 <--> V6
31	0.310	0.0224	13.8	0.0e+00	V7 <--> V7
32	0.053	0.0038	14.1	0.0e+00	V8 <--> V8
33	0.277	0.0206	13.5	0.0e+00	V9 <--> V9
34	0.590	0.0431	13.7	0.0e+00	V10 <--> V10
35	0.298	0.0213	14.0	0.0e+00	V11 <--> V11
36	0.065	0.0045	14.6	0.0e+00	V12 <--> V12
37	0.254	0.0183	13.9	0.0e+00	V13 <--> V13
38	0.598	0.0467	12.8	0.0e+00	V14 <--> V14
39	0.302	0.0220	13.7	0.0e+00	V15 <--> V15
40	0.066	0.0052	12.6	0.0e+00	V16 <--> V16
41	0.283	0.0204	13.9	0.0e+00	V17 <--> V17
42	0.624	0.0445	14.0	0.0e+00	V18 <--> V18
43	0.297	0.0214	13.8	0.0e+00	V19 <--> V19
44	0.053	0.0036	14.9	0.0e+00	V20 <--> V20
45	0.294	0.0210	14.0	0.0e+00	V21 <--> V21
46	0.529	0.0422	12.5	0.0e+00	V22 <--> V22
47	0.301	0.0212	14.2	0.0e+00	V23 <--> V23
48	0.071	0.0050	14.2	0.0e+00	V24 <--> V24
49	-0.722	0.0550	-13.1	0.0e+00	F3 <--> F1
50	-0.810	0.0413	-19.6	0.0e+00	F4 <--> F2

Iterations = 93

> *std.coef(sem.skew4)*

	Std. Estimate	
1 1	0.47714	V1 <--- F2
2 2	0.52321	V2 <--- F3
3 3	0.57183	V3 <--- F4
4 4	0.48879	V4 <--- F1
5 5	0.51584	V5 <--- F2
6 6	0.47547	V6 <--- F3
7 7	0.51551	V7 <--- F4
8 8	0.39005	V8 <--- F1
9 9	0.53238	V9 <--- F2
10 10	0.52563	V10 <--- F3
11 11	0.50162	V11 <--- F4
12 12	0.34165	V12 <--- F1
13 13	0.50106	V13 <--- F2
14 14	0.59007	V14 <--- F3
15 15	0.52281	V15 <--- F4
16 16	0.49488	V16 <--- F1
17 17	0.50054	V17 <--- F2
18 18	0.48950	V18 <--- F3
19 19	0.51244	V19 <--- F4
20 20	0.29937	V20 <--- F1

```

21 21 0.48595      V21 <--- F2
22 22 0.60821      V22 <--- F3
23 23 0.47872      V23 <--- F4
24 24 0.39074      V24 <--- F1

```

Alternatively, we can repeat this analysis, modeling the correlations rather than the covariances. Examine how the goodness of fit for the four (correlated) factor model is identical for the covariances or the correlations, but in either case, the fits are far better than the two factor model

```

> skew.cor <- cor(skew.items)
> sem.skew4 <- sem(model.4, skew.cor, nsub)
> summary(sem.skew4, digits = 2)

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

```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-2.8e+00	-6.1e-01	-1.7e-05	-2.7e-02	5.4e-01	3.3e+00

Parameter Estimates

	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
10	0.53	0.049	10.8	0.0e+00	V10 <--- F3
11	0.50	0.049	10.3	0.0e+00	V11 <--- F4
12	0.34	0.055	6.2	4.7e-10	V12 <--- F1
13	0.50	0.049	10.2	0.0e+00	V13 <--- F2
14	0.59	0.048	12.3	0.0e+00	V14 <--- F3
15	0.52	0.049	10.8	0.0e+00	V15 <--- F4
16	0.49	0.055	8.9	0.0e+00	V16 <--- F1
17	0.50	0.049	10.2	0.0e+00	V17 <--- F2
18	0.49	0.049	10.0	0.0e+00	V18 <--- F3
19	0.51	0.049	10.5	0.0e+00	V19 <--- F4
20	0.30	0.055	5.4	6.0e-08	V20 <--- F1
21	0.49	0.049	9.8	0.0e+00	V21 <--- F2
22	0.61	0.048	12.8	0.0e+00	V22 <--- F3
23	0.48	0.049	9.8	0.0e+00	V23 <--- F4

24	0.39	0.055	7.2	8.1e-13	V24 <---> F1
25	0.77	0.055	14.1	0.0e+00	V1 <--> V1
26	0.73	0.053	13.7	0.0e+00	V2 <--> V2
27	0.67	0.051	13.2	0.0e+00	V3 <--> V3
28	0.76	0.060	12.7	0.0e+00	V4 <--> V4
29	0.73	0.053	13.7	0.0e+00	V5 <--> V5
30	0.77	0.055	14.2	0.0e+00	V6 <--> V6
31	0.73	0.053	13.8	0.0e+00	V7 <--> V7
32	0.85	0.060	14.2	0.0e+00	V8 <--> V8
33	0.72	0.053	13.5	0.0e+00	V9 <--> V9
34	0.72	0.053	13.7	0.0e+00	V10 <--> V10
35	0.75	0.054	14.0	0.0e+00	V11 <--> V11
36	0.88	0.060	14.6	0.0e+00	V12 <--> V12
37	0.75	0.054	13.9	0.0e+00	V13 <--> V13
38	0.65	0.051	12.8	0.0e+00	V14 <--> V14
39	0.73	0.053	13.7	0.0e+00	V15 <--> V15
40	0.76	0.060	12.7	0.0e+00	V16 <--> V16
41	0.75	0.054	13.9	0.0e+00	V17 <--> V17
42	0.76	0.054	14.0	0.0e+00	V18 <--> V18
43	0.74	0.053	13.8	0.0e+00	V19 <--> V19
44	0.91	0.061	14.9	0.0e+00	V20 <--> V20
45	0.76	0.055	14.0	0.0e+00	V21 <--> V21
46	0.63	0.050	12.5	0.0e+00	V22 <--> V22
47	0.77	0.054	14.2	0.0e+00	V23 <--> V23
48	0.85	0.060	14.2	0.0e+00	V24 <--> V24
49	-0.72	0.055	-13.1	0.0e+00	F3 <--> F1
50	-0.81	0.041	-19.6	0.0e+00	F4 <--> F2

Iterations = 11

5.7 Forming clusters or homogeneous item composites

An alternative treatment for the non-continuous nature of items is to group them into “testlets” or “homogeneous item composites”, (HICs). This can be done by a set of transformations, or by recognizing that forming such scales is the equivalent of multiplying a “keys” matrix times the original data matrix. The **psych** package includes two functions, **cluster.cor** and **cluster.loadings** that do this and finds the resulting correlation of the scales.

The function requires us to first form a “keys” matrix composed of item weights of -1, 0, and 1:

```
> 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)
```

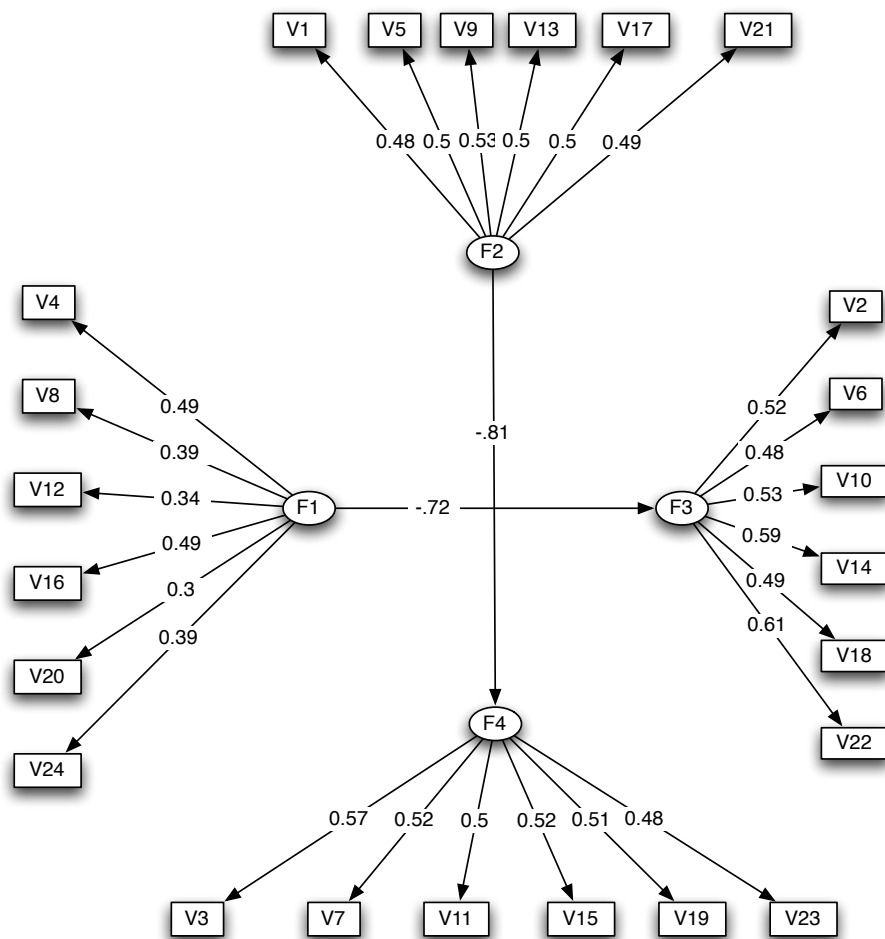


Figure 5.13: A two dimensional solution does not fit very well, but a 4 factor model in two space matches the generating function very well.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	0	1	0	0	0	0	0	0
[2,]	0	0	1	0	0	0	0	0
[3,]	0	0	0	1	0	0	0	0
[4,]	0	0	0	0	1	0	0	0
[5,]	0	0	0	0	0	1	0	0
[6,]	0	0	0	0	0	0	1	0
[7,]	0	0	0	0	0	0	0	1
[8,]	1	0	0	0	0	0	0	0
[9,]	0	1	0	0	0	0	0	0
[10,]	0	0	1	0	0	0	0	0
[11,]	0	0	0	1	0	0	0	0
[12,]	0	0	0	0	1	0	0	0
[13,]	0	0	0	0	0	1	0	0
[14,]	0	0	0	0	0	0	1	0
[15,]	0	0	0	0	0	0	0	1
[16,]	1	0	0	0	0	0	0	0
[17,]	0	1	0	0	0	0	0	0
[18,]	0	0	1	0	0	0	0	0
[19,]	0	0	0	1	0	0	0	0
[20,]	0	0	0	0	1	0	0	0
[21,]	0	0	0	0	0	1	0	0
[22,]	0	0	0	0	0	0	1	0
[23,]	0	0	0	0	0	0	0	1
[24,]	1	0	0	0	0	0	0	0

```
> clusters <- cluster.loadings(keys, skew.cor)
> print(clusters, digits = 2)
```

```
$loadings
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
V1 -0.05 0.33 0.00 -0.32 -0.02 0.34 0.06 -0.25
V2 -0.22 -0.03 0.37 0.00 -0.23 -0.01 0.41 -0.02
V3 0.06 -0.34 -0.03 0.43 0.04 -0.33 0.00 0.37
V4 0.37 0.01 -0.23 -0.03 0.18 0.06 -0.24 -0.05
V5 -0.01 0.37 0.02 -0.30 -0.03 0.37 0.04 -0.29
V6 -0.25 0.05 0.34 -0.10 -0.26 0.03 0.38 -0.03
V7 0.01 -0.27 0.04 0.36 0.03 -0.29 0.04 0.40
V8 0.20 -0.01 -0.22 0.01 0.25 0.02 -0.23 0.00
V9 0.03 0.36 0.03 -0.29 0.03 0.41 0.01 -0.29
V10 -0.27 -0.01 0.38 -0.01 -0.23 -0.05 0.39 -0.01
V11 -0.01 -0.28 -0.04 0.38 0.08 -0.30 -0.06 0.35
V12 0.22 -0.03 -0.24 0.04 0.16 -0.03 -0.20 0.02
V13 0.05 0.35 0.01 -0.32 -0.07 0.34 0.05 -0.29
V14 -0.26 0.04 0.43 0.02 -0.25 0.04 0.44 -0.06
V15 0.02 -0.28 -0.03 0.35 0.07 -0.32 -0.06 0.41
V16 0.27 -0.05 -0.23 0.02 0.34 0.00 -0.24 0.03
V17 0.00 0.35 -0.01 -0.30 -0.04 0.37 0.02 -0.27
V18 -0.24 0.06 0.33 -0.02 -0.24 0.02 0.38 -0.04
V19 0.00 -0.32 0.04 0.40 -0.06 -0.34 0.01 0.31
V20 0.22 0.00 -0.15 0.06 0.11 -0.10 -0.22 0.09
V21 0.00 0.39 -0.08 -0.31 0.02 0.29 -0.08 -0.25
V22 -0.29 0.01 0.45 0.03 -0.26 -0.06 0.43 0.02
```

```
V23 -0.02 -0.28 -0.08 0.32 -0.03 -0.24 -0.05 0.39
V24 0.22 0.04 -0.22 0.02 0.24 0.02 -0.25 -0.02
```

```
$cor
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] 1.00 -0.01 -0.34 0.03 0.43 0.02 -0.36 0.01
[2,] -0.01 1.00 0.01 -0.43 -0.01 0.53 0.04 -0.38
[3,] -0.34 0.01 1.00 -0.01 -0.33 -0.02 0.56 -0.03
[4,] 0.03 -0.43 -0.01 1.00 0.03 -0.44 -0.02 0.47
[5,] 0.43 -0.01 -0.33 0.03 1.00 -0.04 -0.35 0.03
[6,] 0.02 0.53 -0.02 -0.44 -0.04 1.00 0.00 -0.39
[7,] -0.36 0.04 0.56 -0.02 -0.35 0.00 1.00 -0.03
[8,] 0.01 -0.38 -0.03 0.47 0.03 -0.39 -0.03 1.00
```

```
$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
```

```
$sd
```

```
[1] 2.0 2.1 2.1 2.2 1.9 2.1 2.2 2.2
```

```
$alpha
```

```
[1] 0.35 0.49 0.51 0.55 0.24 0.48 0.57 0.55
```

```
$size
```

```
[1] 3 3 3 3 3 3 3 3
```

The function returns the item by cluster correlation (roughly equivalent to a factor loading), the raw correlation matrix, and the correlation matrix corrected for unreliability. For our purposes, we want to examine the raw correlation matrix of the composite scales. We create a new structural model similar to the one created in section 5.6.2. Note how the fit is very good and is very similar to the results from the more extensive analysis using all 24 variables.

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

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
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-8.9e-01	-6.1e-02	-5.8e-06	9.6e-06	5.6e-03	8.9e-01

Parameter Estimates

	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
9	0.56	0.062	9.0	0.0e+00	V1 <--> V1
10	0.48	0.051	9.5	0.0e+00	V2 <--> V2
11	0.47	0.058	8.2	2.2e-16	V3 <--> V3
12	0.47	0.055	8.5	0.0e+00	V4 <--> V4
13	0.58	0.060	9.6	0.0e+00	V5 <--> V5
14	0.46	0.052	8.9	0.0e+00	V6 <--> V6
15	0.41	0.061	6.7	2.8e-11	V7 <--> V7
16	0.58	0.052	11.2	0.0e+00	V8 <--> V8
17	-0.82	0.046	-18.0	0.0e+00	F3 <--> F1
18	-0.70	0.052	-13.5	0.0e+00	F4 <--> F2

Iterations = 25

This last example has shown that there are multiple alternative methods for representing sets of items. Forming “testlets” or “HICS” is one way for compensating for problems at the item level. Another way of organizing the eight testlets is in terms of two orthogonal factors:

```
> m4 <- modelmat(8)
> sem4 <- sem(m4, clusters$cor, nsub)
> summary(sem4, digits = 2)

Model Chisquare = 53   Df = 20 Pr(>Chisq) = 9e-05
Chisquare (null model) = 884   Df = 28
Goodness-of-fit index = 0.97
Adjusted goodness-of-fit index = 0.95
RMSEA index = 0.057   90% CI: (0.039, 0.076)
Bentler-Bonnett NFI = 0.94
Tucker-Lewis NNFI = 0.95
Bentler CFI = 0.96
BIC = -72
```

Normalized Residuals

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-0.894	-0.056	0.477	0.370	0.689	3.300

Parameter Estimates

	Estimate	Std Error	z value	Pr(> z)	
1	-0.53	0.050	-11	0	V1 <--- F1

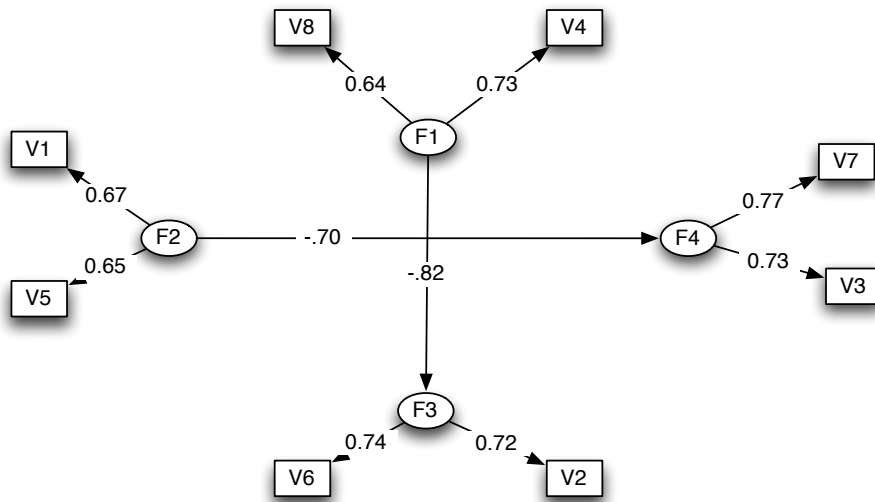


Figure 5.14: An alternative solution is to group the variables into “testlets” or “homogeneous item composites” (HICs) and then to examine the structure of the HICs.

2	0.69	0.046	15	0	V2 <--- F2
3	0.71	0.047	15	0	V3 <--- F1
4	-0.66	0.047	-14	0	V4 <--- F2
5	-0.52	0.050	-10	0	V5 <--- F1
6	0.70	0.046	15	0	V6 <--- F2
7	0.74	0.047	16	0	V7 <--- F1
8	-0.60	0.048	-13	0	V8 <--- F2
9	0.72	0.054	13	0	V1 <--> V1
10	0.52	0.048	11	0	V2 <--> V2
11	0.50	0.049	10	0	V3 <--> V3
12	0.56	0.049	11	0	V4 <--> V4
13	0.73	0.055	13	0	V5 <--> V5
14	0.50	0.048	11	0	V6 <--> V6
15	0.45	0.050	9	0	V7 <--> V7
16	0.64	0.051	13	0	V8 <--> V8

Iterations = 24

All of these techniques are meant to deal with the problem of real items that tend to be categorical, of low reliability, and faced with problems of skew.

5.8 References

- Acton, G. S. , & Revelle, W. (2002). Interpersonal personality measures show circumplex structure based on new psychometric criteria. *Journal of Personality Assessment*, 79, 456-481.
- Acton, G. S., & Revelle, W. (2004). Evaluation of ten psychometric criteria for circumplex structure. *Methods of Psychological Research*. <http://www.dgps.de/fachgruppen/methoden/mpr-online/> 9, 1-27. (pdf)
- Adams, D. (1979) *Hitchhikers Guide to the Galaxy*, London: Pan Books
- Browne, M. W. (1992). Circumplex models for correlation matrices. *Psychometrika*, 57, 469-497.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Gurtman, M. B. (1993). Constructing personality tests to meet a structural criterion: Application of the interpersonal circumplex. *Journal of Personality*, 61, 237-263.
- Gurtman, M. B. (1994). The circumplex as a tool for studying normal and abnormal personality: A methodological primer. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 243-263). New York: Springer.
- Hofstee, W. K. B., de Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63, 146-163.
- Jensen, A. R., & Weng, L. (1994). What is a good g? *Intelligence*, 18, 231-258.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Rafaeli, E. & Revelle, W. (2006). A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*. (pdf from Springer).
- Revelle, W., & Rocklin, T. (1979). Very Simple Structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14, 403-414. (pdf)
- Saucier, G. (1992). Benchmarks: Integrating affective and interpersonal circles with the Big-Five personality factors. *Journal of Personality and Social Psychology*, 62, 1025-1035.
- Saucier, G., Ostendorf, F., & Peabody, D. (2001). The non-evaluative circumplex of personality adjectives. *Journal of Personality*, 69, 537-582.
- Thurstone, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.
- Zinbarg, R.E., Revelle, W., Yovel, I., & Li. W. (2005). Cronbach's Alpha, Revelle's Beta,

McDonald's Omega: Their relations with each and two alternative conceptualizations of reliability. *Psychometrika*. 70, 123-133. [\(pdf\)](#)