

Chapter 10

Reliability + Validity = Structural Equation Models

10.1 Generating simulated data structures

10.2 Measures of fit

As has been seen in the previous sections, the use of fit statistics does not guarantee meaningful models. If we do not specify the model correctly, either because we do not include the correct variables or because we fail to use the appropriate measurement model, we will lead to incorrect conclusions. [Widaman and Thompson \(2003\)](#) [MacCallum et al. \(2006\)](#) [Marsh et al. \(2005\)](#)

Even if we have a very good fit, we are unable to determine causal structure from the model, even if we bother to add time into the model.

10.2.1 χ^2

As we saw in the previous chapter, χ^2 is very sensitive to many sources of error in our model specification. χ^2 is sensitive to failures of our distributional assumptions (continuous, multivariate normal) as well as to our failures to correctly specify the structure.

10.2.2 GFI, NFI, ...

10.2.3 RMSEA

10.3 Reliability (Measurement) models

10.3.1 One factor — congeneric measurement model

10.3.1.1 Generating congeneric data structures

10.3.1.2 Testing for Tau equivalent and congeneric structures

10.3.2 Two (perhaps correlated) factors

10.3.2.1 Generating multiple factorial data

10.3.2.2 Confirmatory factoring using sem

10.3.3 Hierarchical measurement models

10.3.3.1 Generating the data for three correlated factors

10.3.3.2 Testing hierarchical models

10.4 Reliability + Validity = Structural Equation Models

10.4.1 Factorial invariance

10.4.2 Multiple group models

10.5 Evaluating goodness of fit

10.5.1 Model misspecification: Item quality

10.5.1.1 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 \(2006\)](#).

10.5.1.2 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 interrelated; 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 \(1994\)](#)([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 et al., 1992b](#)) 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.

10.5.2 Model misspecification: failure to include variables

10.5.3 Model misspecification: incorrect structure

10.6 What does it mean to fit a model

Chapter 11

Testing alternative models of data

In this chapter we consider how to test nested alternative models of some basic data types. Using the simulation tools introduced in the previous chapter, we generate a data set from a congeneric reliability model with unequal true score loadings and fit three alternative models to the data. Then we simulate a two factor data structure and consider a set of alternative models. Finally, we consider ways of representing (and modeling) hierarchical data structures.

For these examples, as well as the other ones, we need to load the `psych` and `sem` packages.

```
> library(sem)
> library(psych)
```

11.1 One factor — congeneric data model

The classic test theory structure of 4 observed variables V1 ... V4 all loading on a single factor, θ , (Figure 11.1 may be analyzed in multiple ways. The most restrictive model considers all the loadings to be fixed values (perhaps .7). A more reasonable model is to consider the four variables to be parallel, that is to say, that they have equal loadings on the latent variable and equal error variances. A less restrictive models would be tau equivalence where the paths are equal but the errors can be unequal, and then the least restrictive model is known as the “congeneric” model where all paths are allowed to vary.

We can generate data under a congeneric model and then test it with progressively more restricted models (i.e., start with the most unrestricted model, the congeneric model, fix some parameters for the tau equivalent model, add equality constraints for the parallel test model, and then fit arbitrarily fixed parameters). To do this, we first create a function, `sim.sem`, which we apply to make our data.

11.1.1 Generating the data

We create a function, `sim.sem`, to simulate data with a variety of possible structures. Although the function defaults to four variables with specific loadings on one factor, we can vary both the number of variables as well as the loadings and the number of factors. The

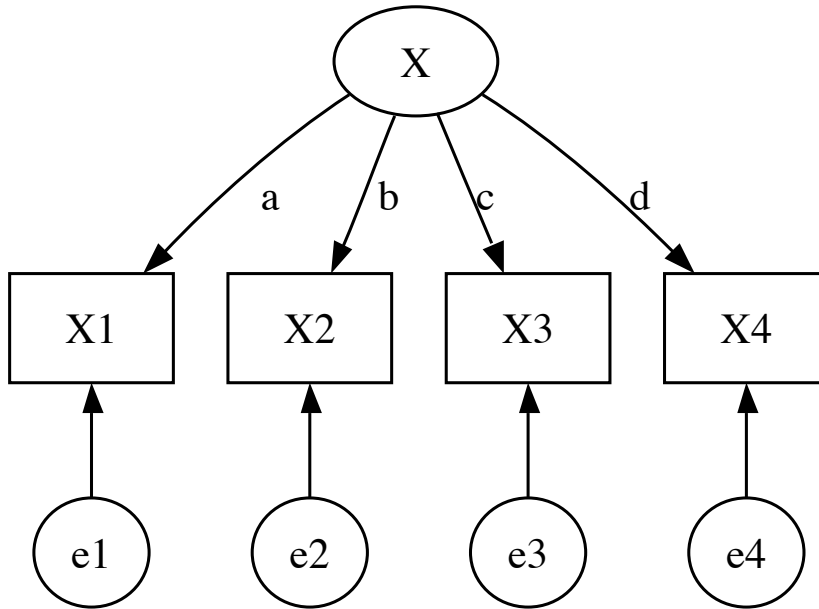


Fig. 11.1 The basic congeneric model is one latent (true score) factor accounting for the correlations of multiple observed scores. If there are at least 4 observed variables, the model is identified. For fewer variables, assumptions need to be made (i.e., for two parallel tests, the path coefficients are all equal.)

function returns the pattern matrix used to generate the data and the implied structure matrix, or just the simulated raw data.

```

> sim.sem <- function(N = 1000, loads = c(0.8, 0.7, 0.6, 0.5), phi = NULL,
+   obs = TRUE) {
+   if (!is.matrix(loads)) {
+     loading <- matrix(loads, ncol = 1)
+   }
+   else {
+     loading <- loads
+   }
+   nv <- dim(loading)[1]
+   nf <- dim(loading)[2]
+   error <- diag(1, nrow = nv)
+   diag(error) <- sqrt(1 - diag(loading %*% t(loading)))
+   if (is.null(phi))
+     phi <- diag(1, nrow = nf)
+   pattern <- cbind(loading, error)
+   colnames(pattern) <- c(paste("theta", seq(1:nf), sep = ""),
+     paste("e", seq(1:nv), sep = ""))
+   rownames(pattern) <- c(paste("V", seq(1:nv), sep = ""))
+   temp <- diag(1, nv + nf)
+   temp[1:nf, 1:nf] <- phi
  
```

```

+   phi <- temp
+   colnames(phi) <- c(paste("theta", seq(1:nf), sep = ""), paste("e",
+     seq(1:nv), sep = ""))
+   structure <- pattern %>% phi
+   latent <- matrix(rnorm(N * (nf + nv)), ncol = (nf + nv))
+   if (nf > 1) {
+     for (i in 1:nf) {
+       for (j in i + 1:nf) {
+         phi[i, j] <- 0
+       }
+     }
+   }
+   observed <- latent %>% t(pattern %>% phi)
+   if (obs) {
+     return(observed)
+   }
+   else {
+     ps <- list(pattern = pattern, structure = structure, phi)
+     return(ps)
+   }
+ }

```

Specifying a particular factor pattern, we can use the **sim.sem** function to show the extended pattern matrix, the implied population correlation matrix, and then take a sample of 1000 from that population. Note that even with 1000 simulated subjects the sample correlation matrix is not the same as the population matrix. As we develop our theory testing skills, it is useful to remember that we are trying to make inferences about the population based upon parameter estimates derived from the sample.

```

> N <- 1000
> sim <- sim.sem(obs = FALSE)
> round(sim$pattern, 2)
      theta1 e1  e2  e3  e4
V1    0.8 0.6 0.00 0.0 0.00
V2    0.7 0.0 0.71 0.0 0.00
V3    0.6 0.0 0.00 0.8 0.00
V4    0.5 0.0 0.00 0.0 0.87

> population <- (sim$pattern %>% t(sim$pattern))
> population
      V1  V2  V3  V4
V1 1.00 0.56 0.48 0.40
V2 0.56 1.00 0.42 0.35
V3 0.48 0.42 1.00 0.30
V4 0.40 0.35 0.30 1.00

> set.seed(42)
> data.f1 <- sim.sem()
> round(cor(data.f1), 2)

```

```

      V1  V2  V3  V4
V1 1.00 0.54 0.52 0.41
V2 0.54 1.00 0.41 0.32
V3 0.52 0.41 1.00 0.32
V4 0.41 0.32 0.32 1.00

```

11.1.2 Estimate a congeneric model

Using the simulated data generated above, we find the covariance matrix from the sample data and apply `sem` to the data. (The `sem` package needs to be loaded first.) Examine the statistics of fit as well as the residual matrix.

```

> S.congeneric <- cov(data.f1)
> model.congeneric <- matrix(c("theta -> V1", "a", NA, "theta -> V2",
+   "b", NA, "theta -> V3", "c", NA, "theta -> V4", "d", NA, "V1 <-> V1",
+   "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3", "w", NA, "V4 <-> V4",
+   "x", NA, "theta <-> theta", NA, 1), ncol = 3, byrow = TRUE)
> colnames(model.congeneric) <- c("path", "label", "initial estimate")
> model.congeneric

      path          label initial estimate
[1,] "theta -> V1"    "a"      NA
[2,] "theta -> V2"    "b"      NA
[3,] "theta -> V3"    "c"      NA
[4,] "theta -> V4"    "d"      NA
[5,] "V1 <-> V1"      "u"      NA
[6,] "V2 <-> V2"      "v"      NA
[7,] "V3 <-> V3"      "w"      NA
[8,] "V4 <-> V4"      "x"      NA
[9,] "theta <-> theta" NA      "1"

> sem.congeneric = sem(model.congeneric, S.congeneric, N)
> summary(sem.congeneric, digits = 3)

Model Chisquare = 0.46  Df = 2 Pr(>Chisq) = 0.795
Chisquare (null model) = 910  Df = 6
Goodness-of-fit index = 1
Adjusted goodness-of-fit index = 0.999
RMSEA index = 0  90% CI: (NA, 0.0398)
Bentler-Bonnett NFI = 1
Tucker-Lewis NNFI = 1.01
Bentler CFI = 1
BIC = -13.4

Normalized Residuals
      Min.  1st Qu.  Median  Mean  3rd Qu.  Max.
-0.177000 -0.032200 -0.000271 0.010600 0.017000 0.319000

```



```

Parameter Estimates
Estimate Std Error z value Pr(>|z|)
a 0.829    0.0320   25.90  0      V1 <--- theta
b 0.657    0.0325   20.23  0      V2 <--- theta
c 0.632    0.0325   19.43  0      V3 <--- theta
d 0.503    0.0340   14.80  0      V4 <--- theta
u 0.316    0.0346    9.12  0      V1 <--> V1
v 0.580    0.0334   17.35  0      V2 <--> V2
w 0.604    0.0337   17.94  0      V3 <--> V3
x 0.776    0.0382   20.31  0      V4 <--> V4

```

```
Iterations = 13
```

```
> round(residuals(sem.congeneric), 2)
```

```

      V1      V2      V3      V4
V1  0  0.00  0.00  0.00
V2  0  0.00 -0.01  0.00
V3  0 -0.01  0.00  0.01
V4  0  0.00  0.01  0.00

```

11.1.3 Estimate a tau equivalent model with equal true score and unequal error loadings

A more constrained model, “Tau equivalence”, assumes that the theta paths in Figure 11.1 are equal but allows the error variances to be unequal.

```

> S.congeneric <- cov(data.f1)
> model.tau <- matrix(c("theta -> V1", "a", NA, "theta -> V2", "a",
+   NA, "theta -> V3", "a", NA, "theta -> V4", "a", NA, "V1 <-> V1",
+   "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3", "w", NA, "V4 <-> V4",
+   "x", NA, "theta <-> theta", NA, 1), ncol = 3, byrow = TRUE)
> colnames(model.tau) <- c("path", "label", "initial estimate")
> model.tau

      path          label initial estimate
[1,] "theta -> V1"    "a"      NA
[2,] "theta -> V2"    "a"      NA
[3,] "theta -> V3"    "a"      NA
[4,] "theta -> V4"    "a"      NA
[5,] "V1 <-> V1"      "u"      NA
[6,] "V2 <-> V2"      "v"      NA
[7,] "V3 <-> V3"      "w"      NA
[8,] "V4 <-> V4"      "x"      NA
[9,] "theta <-> theta" NA      "1"

> sem.tau = sem(model.tau, S.congeneric, N)
> summary(sem.tau, digits = 3)

```

```

Model Chisquare = 56.1   Df = 5 Pr(>Chisq) = 7.64e-11
Chisquare (null model) = 910   Df = 6
Goodness-of-fit index = 0.974
Adjusted goodness-of-fit index = 0.947
RMSEA index = 0.101   90% CI: (0.0783, 0.126)
Bentler-Bonnett NFI = 0.938
Tucker-Lewis NNFI = 0.932
Bentler CFI = 0.943
BIC = 21.6

```

Normalized Residuals

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--------|---------|--------|--------|---------|-------|
| -3.160 | -2.890 | -0.967 | -0.418 | 2.290 | 3.000 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|---|----------|-----------|---------|----------|---------------|
| a | 0.668 | 0.0202 | 33.2 | 0 | V1 <--- theta |
| u | 0.448 | 0.0270 | 16.6 | 0 | V1 <--> V1 |
| v | 0.565 | 0.0315 | 18.0 | 0 | V2 <--> V2 |
| w | 0.576 | 0.0319 | 18.1 | 0 | V3 <--> V3 |
| x | 0.730 | 0.0386 | 18.9 | 0 | V4 <--> V4 |

Iterations = 10

```
> round(residuals(sem.tau), 2)
```

| | V1 | V2 | V3 | V4 |
|----|-------|-------|-------|-------|
| V1 | 0.11 | 0.10 | 0.08 | -0.03 |
| V2 | 0.10 | 0.00 | -0.04 | -0.12 |
| V3 | 0.08 | -0.04 | -0.02 | -0.12 |
| V4 | -0.03 | -0.12 | -0.12 | -0.15 |

Note that this model has a much worse fit (as it should), with a very large change in the χ^2 that far exceeds the benefit of greater parsimony (the change in degrees of freedom from 2 to 5). However, note that some traditional measures of fit (e.g., the GFI) seem reasonable. The RMSEA and NFI suggest a poor fit, as do the residuals.

11.1.4 Estimate a parallel test model with equal true score and equal error loadings

An even more unrealistic model would a model of parallel tests where the true score variances are the same for all tests, as are the error variances.

```

> model.parallel <- matrix(c("theta -> V1", "a", NA, "theta -> V2",
+   "a", NA, "theta -> V3", "a", NA, "theta -> V4", "a", NA, "V1 <-> V1",
+   "u", NA, "V2 <-> V2", "u", NA, "V3 <-> V3", "u", NA, "V4 <-> V4",
+   "u", NA, "theta <-> theta", NA, 1), ncol = 3, byrow = TRUE)

```

```
> colnames(model.parallel) <- c("path", "label", "initial estimate")
> model.parallel
```

| | path | label | initial estimate |
|------|-------------------|-------|------------------|
| [1,] | "theta -> V1" | "a" | NA |
| [2,] | "theta -> V2" | "a" | NA |
| [3,] | "theta -> V3" | "a" | NA |
| [4,] | "theta -> V4" | "a" | NA |
| [5,] | "V1 <-> V1" | "u" | NA |
| [6,] | "V2 <-> V2" | "u" | NA |
| [7,] | "V3 <-> V3" | "u" | NA |
| [8,] | "V4 <-> V4" | "u" | NA |
| [9,] | "theta <-> theta" | NA | "1" |

```
> sem.parallel = sem(model.parallel, S.congeneric, N)
```

```
> summary(sem.parallel, digits = 3)
```

```
Model Chisquare = 91.2   Df = 8 Pr(>Chisq) = 2.22e-16
Chisquare (null model) = 910   Df = 6
Goodness-of-fit index = 0.959
Adjusted goodness-of-fit index = 0.949
RMSEA index = 0.102   90% CI: (0.0838, 0.121)
Bentler-Bonnett NFI = 0.9
Tucker-Lewis NNFI = 0.931
Bentler CFI = 0.908
BIC = 36.0
```

```
Normalized Residuals
```

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|-----------|-----------|-----------|----------|----------|----------|
| | -2.78e+00 | -1.04e+00 | -2.74e-01 | 7.11e-07 | 9.99e-01 | 3.51e+00 |

```
Parameter Estimates
```

| | Estimate | Std Error | z value | Pr(> z) | |
|---|----------|-----------|---------|----------|---------------|
| a | 0.652 | 0.0198 | 32.9 | 0 | V1 <--- theta |
| u | 0.586 | 0.0152 | 38.7 | 0 | V1 <--> V1 |

```
Iterations = 11
```

```
> round(residuals(sem.parallel), 2)
```

| | V1 | V2 | V3 | V4 |
|----|-------|-------|-------|-------|
| V1 | -0.01 | 0.12 | 0.10 | -0.01 |
| V2 | 0.12 | 0.00 | -0.02 | -0.09 |
| V3 | 0.10 | -0.02 | -0.01 | -0.10 |
| V4 | -0.01 | -0.09 | -0.10 | 0.02 |

11.1.5 Estimate a parallel test model with fixed loadings

The most restrictive model estimates the fewest parameters and considers the case where all loadings are fixed at a particular value. (This is truly a stupid model). Notice how large the residuals are.

```
> model.fixed <- matrix(c("theta -> V1", NA, 0.6, "theta -> V2",
+   NA, 0.6, "theta -> V3", NA, 0.6, "theta -> V4", NA, 0.6, "V1 <-> V1",
+   "u", NA, "V2 <-> V2", "u", NA, "V3 <-> V3", "u", NA, "V4 <-> V4",
+   "u", NA, "theta <-> theta", NA, 1), ncol = 3, byrow = TRUE)
> colnames(model.fixed) <- c("path", "label", "initial estimate")
> model.fixed

      path          label initial estimate
[1,] "theta -> V1"    NA      "0.6"
[2,] "theta -> V2"    NA      "0.6"
[3,] "theta -> V3"    NA      "0.6"
[4,] "theta -> V4"    NA      "0.6"
[5,] "V1 <-> V1"      "u"     NA
[6,] "V2 <-> V2"      "u"     NA
[7,] "V3 <-> V3"      "u"     NA
[8,] "V4 <-> V4"      "u"     NA
[9,] "theta <-> theta" NA      "1"

> sem.fixed = sem(model.fixed, S.congeneric, N)
> summary(sem.fixed, digits = 3)

Model Chisquare = 98.6   Df = 9 Pr(>Chisq) = 0
Chisquare (null model) = 910   Df = 6
Goodness-of-fit index = 0.957
Adjusted goodness-of-fit index = 0.952
RMSEA index = 0.0998   90% CI: (0.0826, 0.118)
Bentler-Bonnett NFI = 0.892
Tucker-Lewis NNFI = 0.934
Bentler CFI = 0.9
BIC = 36.4

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.980  0.633   1.510   1.850   2.590   5.790

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
u 0.594    0.0153   38.9    0      V1 <-> V1

Iterations = 8

> round(residuals(sem.fixed), 2)

      V1    V2    V3    V4
V1 0.05  0.19  0.16  0.05
```

```
V2 0.19  0.06  0.05 -0.03
V3 0.16  0.05  0.05 -0.03
V4 0.05 -0.03 -0.03  0.08
```

11.1.6 Comparison of models

We can examine the degrees of freedom in each of previous analyses and compare them to the goodness of fit. The difference in χ^2 of a nested model is itself a χ^2 with degrees of freedom = the difference of the two different degrees of freedom. We form a list of the different analyses, and then show the summary statistics.

```
> summary.list <- list()
> summary.list[[1]] <- summary(sem.congeneric)[1:2]
> summary.list[[2]] <- summary(sem.tau)[1:2]
> summary.list[[3]] <- summary(sem.parallel)[1:2]
> summary.list[[4]] <- summary(sem.fixed)[1:2]
> summary.data <- matrix(unlist(summary.list), nrow = 4, byrow = TRUE)
> rownames(summary.data) <- c("congeneric", "tau", "parallel", "fixed")
> colnames(summary.data) <- c("chisq", "df")
> summary.data
```

| | chisq | df |
|------------|------------|----|
| congeneric | 0.4597646 | 2 |
| tau | 56.1290414 | 5 |
| parallel | 91.2250565 | 8 |
| fixed | 98.5734749 | 9 |

A simple conclusion is that although the congeneric model has estimated the most parameters, the χ^2 difference between congeneric and tau equivalence models justifies rejecting tau equivalence in favor of the less restrictive congeneric model.

11.2 Two (perhaps correlated) factors

We now consider more interesting problems. The case of two correlated factors sometimes appears as a classic prediction problem (multiple measures of X, multiple measures of Y, what is the correlation between the two latent constructs) and sometimes as a measurement problem (multiple subfactors of X). The generation structure is similar.

11.2.1 Generating the data

We use the `sim.sem` function from before, and specify a two factor, uncorrelated structure. Figure 11.2 shows the general case where the two factors could be correlated. By specifying the path between the two latent variables to be 0, we make them uncorrelated.

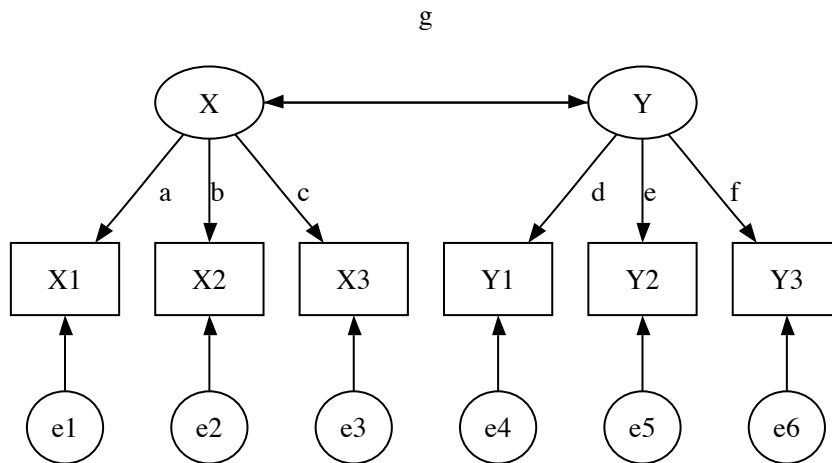


Fig. 11.2 Six variables with two factors. This notation shows the error of measurement in the observed and latent variables. If $g > 0$, then the two factors are correlated.

```
> set.seed(42)
> N <- 1000
> pattern <- matrix(c(0.9, 0, 0.8, 0, 0.7, 0, 0, 0.8, 0, 0.7, 0,
+ 0.6), ncol = 2, byrow = TRUE)
> phi <- matrix(c(1, 0, 0, 1), ncol = 2)
> population <- sim.sem(loads = pattern, phi = phi, obs = FALSE)
> round(population$pattern, 2)

  theta1 theta2 e1 e2 e3 e4 e5 e6
V1  0.9   0.0 0.44 0.0 0.00 0.0 0.00 0.0
V2  0.8   0.0 0.00 0.6 0.00 0.0 0.00 0.0
V3  0.7   0.0 0.00 0.0 0.71 0.0 0.00 0.0
V4  0.0   0.8 0.00 0.0 0.00 0.6 0.00 0.0
V5  0.0   0.7 0.00 0.0 0.00 0.0 0.71 0.0
V6  0.0   0.6 0.00 0.0 0.00 0.0 0.00 0.8

> pop.cor <- round(population$structure %*% t(population$pattern),
+ 2)
> pop.cor

  V1 V2 V3 V4 V5 V6
V1 1.00 0.72 0.63 0.00 0.00 0.00
V2 0.72 1.00 0.56 0.00 0.00 0.00
V3 0.63 0.56 1.00 0.00 0.00 0.00
V4 0.00 0.00 0.00 1.00 0.56 0.48
V5 0.00 0.00 0.00 0.56 1.00 0.42
V6 0.00 0.00 0.00 0.48 0.42 1.00

> data.f2 <- sim.sem(loads = pattern, phi = phi)

The SPLOM suggests two separate factors in the data.
```

```
> pairs.panels(data.f2)
```

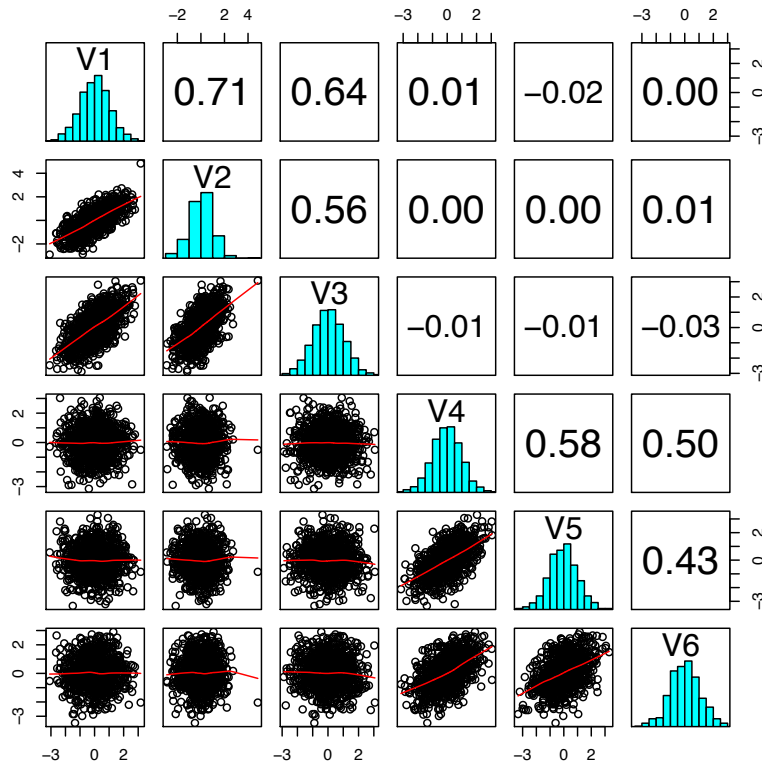


Fig. 11.3 A ScatterPlot Matrix, SPLOM, of the six variables.

11.2.2 Exploratory Factor analysis of the data

This structure may be analyzed in a variety of different ways, including exploratory factor analysis. A “scree” plot of the eigen values of the matrix suggests a two factor solution. Based upon this “prior” hypotheses, we extract two factors using the **factanal** function.

```
> f2 <- factanal(data.f2, 2)
> f2
```

Call:

```
factanal(x = data.f2, factors = 2)
```

Uniquenesses:

| V1 | V2 | V3 | V4 | V5 | V6 |
|-------|-------|-------|-------|-------|-------|
| 0.201 | 0.374 | 0.491 | 0.329 | 0.496 | 0.626 |

Loadings:

```
> VSS.scree(cor(data.f2))
```

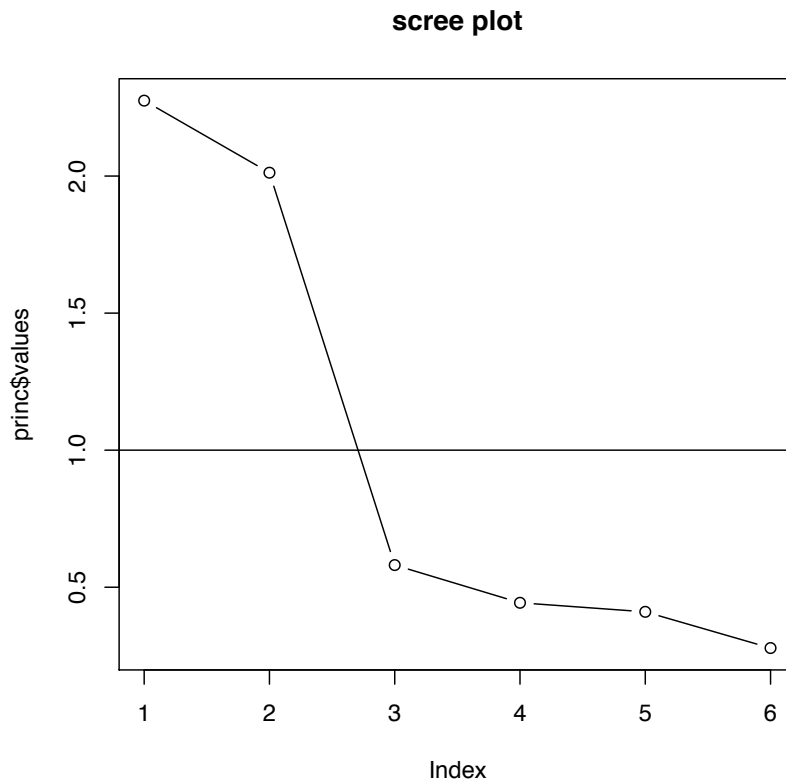


Fig. 11.4 A scree plot of the eigen values of the simulated data suggests that two factors are the best representation of the data. Compare this to the two correlated factor problem, Figure 11.6, and the three correlated factor problem, Figure 11.9

| | Factor1 | Factor2 |
|----|---------|---------|
| V1 | 0.894 | |
| V2 | 0.791 | |
| V3 | 0.713 | |
| V4 | | 0.819 |
| V5 | | 0.709 |
| V6 | | 0.611 |

| | Factor1 | Factor2 |
|----------------|---------|---------|
| SS loadings | 1.934 | 1.548 |
| Proportion Var | 0.322 | 0.258 |
| Cumulative Var | 0.322 | 0.580 |

Test of the hypothesis that 2 factors are sufficient.

The chi square statistic is 3.97 on 4 degrees of freedom.
The p-value is 0.41

The factor loadings nicely capture the population values specified in section [11.2.1](#).

11.2.3 Confirmatory analysis with a predicted structure

We can also analyze these data taking a confirmatory, proposing that the first 3 variables load on one factor, and the second 3 variables load on a second factor.

```
> S.f2 <- cov(data.f2)
> model.two <- matrix(c("theta1 -> V1", "a", NA, "theta1 -> V2",
+   "b", NA, "theta1 -> V3", "c", NA, "theta2 -> V4", "d", NA,
+   "theta2 -> V5", "e", NA, "theta2 -> V6", "f", NA, "V1 <-> V1",
+   "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3", "w", NA, "V4 <-> V4",
+   "x", NA, "V5 <-> V5", "y", NA, "V6 <-> V6", "z", NA, "theta1 <-> theta1",
+   NA, 1, "theta2 <-> theta2", NA, 1), ncol = 3, byrow = TRUE)
> colnames(model.two) <- c("path", "label", "initial estimate")
> model.two
```

| | path | label | initial estimate |
|-------|---------------------|-------|------------------|
| [1,] | "theta1 -> V1" | "a" | NA |
| [2,] | "theta1 -> V2" | "b" | NA |
| [3,] | "theta1 -> V3" | "c" | NA |
| [4,] | "theta2 -> V4" | "d" | NA |
| [5,] | "theta2 -> V5" | "e" | NA |
| [6,] | "theta2 -> V6" | "f" | NA |
| [7,] | "V1 <-> V1" | "u" | NA |
| [8,] | "V2 <-> V2" | "v" | NA |
| [9,] | "V3 <-> V3" | "w" | NA |
| [10,] | "V4 <-> V4" | "x" | NA |
| [11,] | "V5 <-> V5" | "y" | NA |
| [12,] | "V6 <-> V6" | "z" | NA |
| [13,] | "theta1 <-> theta1" | NA | "1" |
| [14,] | "theta2 <-> theta2" | NA | "1" |

```
> sem.two = sem(model.two, S.f2, N)
> summary(sem.two, digits = 3)
```

Model Chisquare = 4.91 Df = 9 Pr(>Chisq) = 0.842
Chisquare (null model) = 2004 Df = 15
Goodness-of-fit index = 0.998
Adjusted goodness-of-fit index = 0.996
RMSEA index = 0 90% CI: (NA, 0.0205)
Bentler-Bonnett NFI = 0.998
Tucker-Lewis NNFI = 1.00
Bentler CFI = 1
BIC = -57.3

```

Normalized Residuals
  Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
-8.02e-01 -4.43e-02 -7.84e-06 -8.03e-02  2.63e-05  3.51e-01

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a 0.898    0.0283    31.66  0.00e+00 V1 <--- theta1
b 0.770    0.0282    27.31  0.00e+00 V2 <--- theta1
c 0.727    0.0301    24.18  0.00e+00 V3 <--- theta1
d 0.817    0.0340    24.03  0.00e+00 V4 <--- theta2
e 0.731    0.0345    21.19  0.00e+00 V5 <--- theta2
f 0.644    0.0348    18.50  0.00e+00 V6 <--- theta2
u 0.204    0.0267     7.64  2.22e-14 V1 <--> V1
v 0.354    0.0244    14.53  0.00e+00 V2 <--> V2
w 0.510    0.0282    18.11  0.00e+00 V3 <--> V3
x 0.327    0.0391     8.36  0.00e+00 V4 <--> V4
y 0.527    0.0375    14.06  0.00e+00 V5 <--> V5
z 0.696    0.0385    18.09  0.00e+00 V6 <--> V6

```

```
Iterations = 22
```

```

> round(residuals(sem.two), 2)

      V1  V2  V3  V4  V5  V6
V1  0.00 0.00  0.00  0.01 -0.02  0.00
V2  0.00 0.00  0.00  0.00  0.00  0.01
V3  0.00 0.00  0.00 -0.01 -0.01 -0.03
V4  0.01 0.00 -0.01  0.00  0.00  0.00
V5 -0.02 0.00 -0.01  0.00  0.00  0.00
V6  0.00 0.01 -0.03  0.00  0.00  0.00

> std.coef(sem.two)

      Std. Estimate
a a 0.89320      V1 <--- theta1
b b 0.79150      V2 <--- theta1
c c 0.71341      V3 <--- theta1
d d 0.81904      V4 <--- theta2
e e 0.70975      V5 <--- theta2
f f 0.61113      V6 <--- theta2

```

It is useful to compare these “confirmatory” factor loadings with the factor loadings obtained by the exploratory factor analysis in section 11.2.2. Although the unstandardized loadings differ, the standardized loadings are identical to 3 decimals.

11.2.4 Confirmatory factor analysis with two independent factors with equal loadings within factors

The previous model allowed the factor loadings (and hence the quality of measurement of the variables) to differ. A more restrictive model (e.g., tau equivalence) forces the true score loadings to be equal within each factor. Note that although the χ^2 suggests that the model is not adequate, the more standard measures of adequacy of fit (e.g., RMSEA and the NFI) incorrectly show a good fit.

```
> model.twotau <- matrix(c("theta1 -> V1", "a", NA, "theta1 -> V2",
+   "a", NA, "theta1 -> V3", "a", NA, "theta2 -> V4", "d", NA,
+   "theta2 -> V5", "d", NA, "theta2 -> V6", "d", NA, "V1 <-> V1",
+   "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3", "w", NA, "V4 <-> V4",
+   "x", NA, "V5 <-> V5", "y", NA, "V6 <-> V6", "z", NA, "theta1 <-> theta1",
+   NA, 1, "theta2 <-> theta2", NA, 1), ncol = 3, byrow = TRUE)
> colnames(model.twotau) <- c("path", "label", "initial estimate")
> model.twotau
```

| | path | label | initial estimate |
|-------|---------------------|-------|------------------|
| [1,] | "theta1 -> V1" | "a" | NA |
| [2,] | "theta1 -> V2" | "a" | NA |
| [3,] | "theta1 -> V3" | "a" | NA |
| [4,] | "theta2 -> V4" | "d" | NA |
| [5,] | "theta2 -> V5" | "d" | NA |
| [6,] | "theta2 -> V6" | "d" | NA |
| [7,] | "V1 <-> V1" | "u" | NA |
| [8,] | "V2 <-> V2" | "v" | NA |
| [9,] | "V3 <-> V3" | "w" | NA |
| [10,] | "V4 <-> V4" | "x" | NA |
| [11,] | "V5 <-> V5" | "y" | NA |
| [12,] | "V6 <-> V6" | "z" | NA |
| [13,] | "theta1 <-> theta1" | NA | "1" |
| [14,] | "theta2 <-> theta2" | NA | "1" |

```
> sem.twotau = sem(model.twotau, S.f2, N)
> summary(sem.twotau, digits = 3)
```

```
Model Chisquare = 46.1   Df = 13 Pr(>Chisq) = 1.38e-05
Chisquare (null model) = 2004   Df = 15
Goodness-of-fit index = 0.985
Adjusted goodness-of-fit index = 0.976
RMSEA index = 0.0505   90% CI: (0.0352, 0.0666)
Bentler-Bonnett NFI = 0.977
Tucker-Lewis NNFI = 0.98
Bentler CFI = 0.983
BIC = -43.7
```

Normalized Residuals

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
-2.3600 -0.7210 -0.0444 -0.2320  0.2160  1.8500
```

Parameter Estimates

```
Estimate Std Error z value Pr(>|z|)
a 0.808    0.0215   37.5    0      V1 <--- theta1
d 0.739    0.0221   33.5    0      V4 <--- theta2
u 0.280    0.0198   14.2    0      V1 <--> V1
v 0.321    0.0212   15.1    0      V2 <--> V2
w 0.480    0.0275   17.5    0      V3 <--> V3
x 0.400    0.0270   14.8    0      V4 <--> V4
y 0.510    0.0309   16.5    0      V5 <--> V5
z 0.648    0.0366   17.7    0      V6 <--> V6
```

Iterations = 15

```
> round(residuals(sem.twotau), 2)
```

```
      V1  V2  V3  V4  V5  V6
V1  0.08  0.04  0.00  0.01 -0.02  0.00
V2  0.04 -0.03 -0.09  0.00  0.00  0.01
V3  0.00 -0.09 -0.09 -0.01 -0.01 -0.03
V4  0.01  0.00 -0.01  0.05  0.05 -0.02
V5 -0.02  0.00 -0.01  0.05  0.01 -0.08
V6  0.00  0.01 -0.03 -0.02 -0.08 -0.08
```

```
> std.coef(sem.twotau)
```

```
Std. Estimate
1 a 0.83640      V1 <--- theta1
2 a 0.81881      V2 <--- theta1
3 a 0.75897      V3 <--- theta1
4 d 0.75999      V4 <--- theta2
5 d 0.71919      V5 <--- theta2
6 d 0.67637      V6 <--- theta2
```

11.2.5 Structure invariance, part I— unequal loadings within factors - matched across factors

Are the two factors measured the same way? That is, are the loadings for the first factor the same as those for the second factor? We can test the model that the ordered loadings are the same across the two factors. We allow the errors to differ.

```
> model.two.invar <- matrix(c("theta1 -> V1", "a", NA, "theta1 -> V2",
+ "b", NA, "theta1 -> V3", "c", NA, "theta2 -> V4", "a", NA,
+ "theta2 -> V5", "b", NA, "theta2 -> V6", "c", NA, "V1 <-> V1",
+ "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3", "w", NA, "V4 <-> V4",
+ "x", NA, "V5 <-> V5", "y", NA, "V6 <-> V6", "z", NA, "theta1 <-> theta1",
+ NA, 1, "theta2 <-> theta2", NA, 1), ncol = 3, byrow = TRUE)
```

```
> colnames(model.two.invar) <- c("path", "label", "initial estimate")
> model.two.invar
```

| | path | label | initial estimate |
|-------|---------------------|-------|------------------|
| [1,] | "theta1 -> V1" | "a" | NA |
| [2,] | "theta1 -> V2" | "b" | NA |
| [3,] | "theta1 -> V3" | "c" | NA |
| [4,] | "theta2 -> V4" | "a" | NA |
| [5,] | "theta2 -> V5" | "b" | NA |
| [6,] | "theta2 -> V6" | "c" | NA |
| [7,] | "V1 <-> V1" | "u" | NA |
| [8,] | "V2 <-> V2" | "v" | NA |
| [9,] | "V3 <-> V3" | "w" | NA |
| [10,] | "V4 <-> V4" | "x" | NA |
| [11,] | "V5 <-> V5" | "y" | NA |
| [12,] | "V6 <-> V6" | "z" | NA |
| [13,] | "theta1 <-> theta1" | NA | "1" |
| [14,] | "theta2 <-> theta2" | NA | "1" |

```
> sem.two.invar = sem(model.two.invar, S.f2, N)
> summary(sem.two.invar, digits = 3)
```

```
Model Chisquare = 10.7   Df = 12 Pr(>Chisq) = 0.557
Chisquare (null model) = 2004   Df = 15
Goodness-of-fit index = 0.996
Adjusted goodness-of-fit index = 0.994
RMSEA index = 0   90% CI: (NA, 0.0293)
Bentler-Bonnett NFI = 0.995
Tucker-Lewis NNFI = 1
Bentler CFI = 1
BIC = -72.2
```

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|---------|---------|---------|---------|---------|--------|
| | -1.7400 | -0.8010 | -0.0444 | -0.0818 | 0.4770 | 1.6000 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|---|----------|-----------|---------|----------|----------------|
| a | 0.862 | 0.0214 | 40.26 | 0 | V1 <--- theta1 |
| b | 0.750 | 0.0215 | 34.91 | 0 | V2 <--- theta1 |
| c | 0.690 | 0.0225 | 30.67 | 0 | V3 <--- theta1 |
| u | 0.211 | 0.0249 | 8.48 | 0 | V1 <--> V1 |
| v | 0.350 | 0.0235 | 14.86 | 0 | V2 <--> V2 |
| w | 0.513 | 0.0277 | 18.53 | 0 | V3 <--> V3 |
| x | 0.312 | 0.0315 | 9.89 | 0 | V4 <--> V4 |
| y | 0.536 | 0.0330 | 16.24 | 0 | V5 <--> V5 |
| z | 0.692 | 0.0371 | 18.66 | 0 | V6 <--> V6 |

```
Iterations = 18
```

```

> round(residuals(sem.two.invar), 2)

      V1  V2  V3  V4  V5  V6
V1  0.06 0.04 0.06 0.01 -0.02 0.00
V2  0.04 0.03 0.04 0.00 0.00 0.01
V3  0.06 0.04 0.05 -0.01 -0.01 -0.03
V4  0.01 0.00 -0.01 -0.06 -0.05 -0.07
V5 -0.02 0.00 -0.01 -0.05 -0.04 -0.05
V6  0.00 0.01 -0.03 -0.07 -0.05 -0.06

> std.coef(sem.two.invar)

      Std. Estimate
1 a 0.88236      V1 <--- theta1
2 b 0.78527      V2 <--- theta1
3 c 0.69405      V3 <--- theta1
4 a 0.83927      V4 <--- theta2
5 b 0.71551      V5 <--- theta2
6 c 0.63875      V6 <--- theta2

```

What is both interesting and disappointing from this example is that although the true loadings (refer back to 11.2.1) are not matched across the two factors, estimating a model that they are equivalent across factors can not be rejected, even with 1000 subjects. In addition, the correct population values are not included in the normal confidence intervals of the estimated values of a,b, and c.

11.2.6 Estimate two correlated factors

This next example is a bit more subtle, in that we generate data with a particular causal structure. The matrix of intercorrelations of the two factors leads to correlations between the variables, but reflects the idea of a path coefficient from the first latent variable to the second one.¹

```

> set.seed(42)
> N <- 1000
> pattern <- matrix(c(0.9, 0, 0.8, 0, 0.7, 0, 0, 0.8, 0, 0.7, 0,
+ 0.6), ncol = 2, byrow = TRUE)
> phi <- matrix(c(1, 0.4, 0.4, 1), ncol = 2)
> population <- sim.sem(loads = pattern, phi = phi, obs = FALSE)
> round(population$pattern, 2)

      theta1 theta2  e1  e2  e3  e4  e5  e6
V1    0.9    0.0 0.44 0.0 0.00 0.0 0.00 0.0
V2    0.8    0.0 0.00 0.6 0.00 0.0 0.00 0.0
V3    0.7    0.0 0.00 0.0 0.71 0.0 0.00 0.0
V4    0.0    0.8 0.00 0.0 0.00 0.6 0.00 0.0

```

¹ Once again, we set the seed for the random number generator to a particular value in order to have reproducible results.

```
V5  0.0  0.7 0.00 0.0 0.00 0.0 0.71 0.0
V6  0.0  0.6 0.00 0.0 0.00 0.0 0.00 0.8
```

```
> round(population$structure, 2)
```

```
      theta1 theta2  e1 e2  e3 e4  e5 e6
V1  0.90  0.36 0.44 0.0 0.00 0.0 0.00 0.0
V2  0.80  0.32 0.00 0.6 0.00 0.0 0.00 0.0
V3  0.70  0.28 0.00 0.0 0.71 0.0 0.00 0.0
V4  0.32  0.80 0.00 0.0 0.00 0.6 0.00 0.0
V5  0.28  0.70 0.00 0.0 0.00 0.0 0.71 0.0
V6  0.24  0.60 0.00 0.0 0.00 0.0 0.00 0.8
```

```
> pop.cor <- population$structure %*% t(population$pattern)
```

```
> round(pop.cor, 2)
```

```
      V1  V2  V3  V4  V5  V6
V1  1.00 0.72 0.63 0.29 0.25 0.22
V2  0.72 1.00 0.56 0.26 0.22 0.19
V3  0.63 0.56 1.00 0.22 0.20 0.17
V4  0.29 0.26 0.22 1.00 0.56 0.48
V5  0.25 0.22 0.20 0.56 1.00 0.42
V6  0.22 0.19 0.17 0.48 0.42 1.00
```

```
> data.f2 <- sim.sem(loads = pattern, phi = phi)
```

The scree test for this problem also suggests two factors, although not as clearly as in example 11.2.1. We first conduct an exploratory factor analysis of the data. Rather than accepting the default value of a VARIMAX rotation, we examine the unrotated solution. For comparisons to a confirmatory factor analysis, we repeat the exploratory analysis with a VARIMAX rotation to simple structure.

```
> f2 <- factanal(data.f2, 2, rotation = "none")
```

```
> f2
```

```
Call:
```

```
factanal(x = data.f2, factors = 2, rotation = "none")
```

```
Uniquenesses:
```

```
      V1  V2  V3  V4  V5  V6
0.201 0.374 0.491 0.295 0.463 0.600
```

```
Loadings:
```

```
      Factor1 Factor2
V1  0.845  -0.291
V2  0.749  -0.256
V3  0.667  -0.254
V4  0.552   0.633
V5  0.463   0.568
V6  0.404   0.487
```

```
> pairs.panels(data.f2)
```

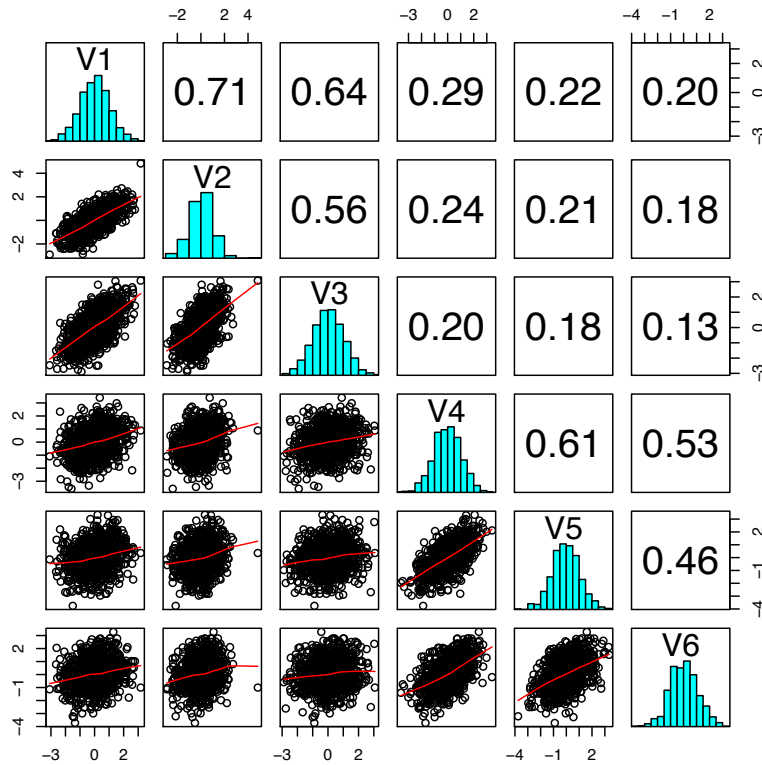


Fig. 11.5 Six variables loading on 2 correlated factors

| | Factor1 | Factor2 |
|----------------|---------|---------|
| SS loadings | 2.401 | 1.175 |
| Proportion Var | 0.400 | 0.196 |
| Cumulative Var | 0.400 | 0.596 |

Test of the hypothesis that 2 factors are sufficient.
 The chi square statistic is 3.99 on 4 degrees of freedom.
 The p-value is 0.407

```
> f2v <- factanal(data.f2, 2)
> f2v
```

```
Call:
factanal(x = data.f2, factors = 2)
```

```
Uniquenesses:
  V1  V2  V3  V4  V5  V6
```



```
> VSS.scree(cor(data.f2))
```

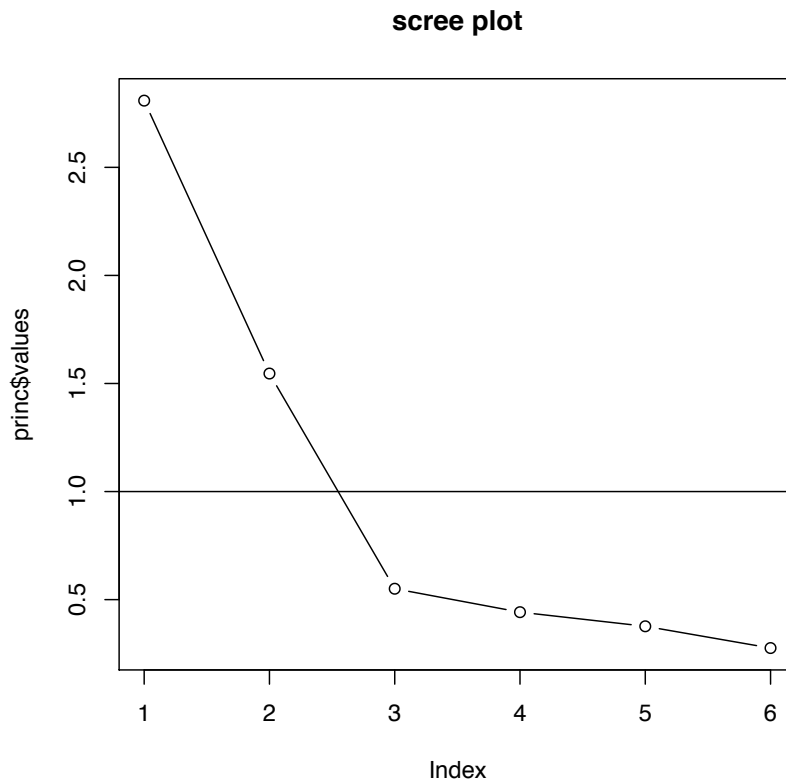


Fig. 11.6 Scree plot of two correlated factors. Compare to Figure 11.4

```
0.201 0.374 0.491 0.295 0.463 0.600
```

Loadings:

| | Factor1 | Factor2 |
|----|---------|---------|
| V1 | 0.888 | |
| V2 | 0.786 | |
| V3 | 0.711 | |
| V4 | 0.226 | 0.809 |
| V5 | 0.174 | 0.711 |
| V6 | 0.156 | 0.613 |

| | Factor1 | Factor2 |
|----------------|---------|---------|
| SS loadings | 2.018 | 1.558 |
| Proportion Var | 0.336 | 0.260 |
| Cumulative Var | 0.336 | 0.596 |

Test of the hypothesis that 2 factors are sufficient.
 The chi square statistic is 3.99 on 4 degrees of freedom.
 The p-value is 0.407

The sem for uncorrelated factors does not fit very well (as it should not!)

```
> S.f2 <- cov(data.f2)
> model.two <- matrix(c("theta1 -> V1", "a", NA, "theta1 -> V2",
+ "b", NA, "theta1 -> V3", "c", NA, "theta2 -> V4", "d", NA,
+ "theta2 -> V5", "e", NA, "theta2 -> V6", "f", NA, "V1 <-> V1",
+ "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3", "w", NA, "V4 <-> V4",
+ "x", NA, "V5 <-> V5", "y", NA, "V6 <-> V6", "z", NA, "theta1 <-> theta1",
+ NA, 1, "theta2 <-> theta2", NA, 1), ncol = 3, byrow = TRUE)
> colnames(model.two) <- c("path", "label", "initial estimate")
> model.two
```

| | path | label | initial estimate |
|-------|---------------------|-------|------------------|
| [1,] | "theta1 -> V1" | "a" | NA |
| [2,] | "theta1 -> V2" | "b" | NA |
| [3,] | "theta1 -> V3" | "c" | NA |
| [4,] | "theta2 -> V4" | "d" | NA |
| [5,] | "theta2 -> V5" | "e" | NA |
| [6,] | "theta2 -> V6" | "f" | NA |
| [7,] | "V1 <-> V1" | "u" | NA |
| [8,] | "V2 <-> V2" | "v" | NA |
| [9,] | "V3 <-> V3" | "w" | NA |
| [10,] | "V4 <-> V4" | "x" | NA |
| [11,] | "V5 <-> V5" | "y" | NA |
| [12,] | "V6 <-> V6" | "z" | NA |
| [13,] | "theta1 <-> theta1" | NA | "1" |
| [14,] | "theta2 <-> theta2" | NA | "1" |

```
> sem.two = sem(model.two, S.f2, N)
> summary(sem.two, digits = 3)
```

```
Model Chisquare = 101 Df = 9 Pr(>Chisq) = 0
Chisquare (null model) = 2206 Df = 15
Goodness-of-fit index = 0.969
Adjusted goodness-of-fit index = 0.927
RMSEA index = 0.101 90% CI: (0.0838, 0.119)
Bentler-Bonnett NFI = 0.954
Tucker-Lewis NNFI = 0.93
Bentler CFI = 0.958
BIC = 38.5
```

```
Normalized Residuals
  Min. 1st Qu. Median Mean 3rd Qu. Max.
-1.56e-05 -3.80e-06 2.10e+00 3.26e+00 6.43e+00 9.05e+00
```

```
Parameter Estimates
```

```

      Estimate Std Error z value Pr(>|z|)
a 0.898      0.0283    31.66  0.00e+00 V1 <--- theta1
b 0.770      0.0282    27.31  0.00e+00 V2 <--- theta1
c 0.727      0.0301    24.18  0.00e+00 V3 <--- theta1
d 0.883      0.0341    25.88  0.00e+00 V4 <--- theta2
e 0.782      0.0344    22.74  0.00e+00 V5 <--- theta2
f 0.682      0.0346    19.72  0.00e+00 V6 <--- theta2
u 0.204      0.0267     7.64  2.22e-14 V1 <--> V1
v 0.354      0.0244    14.53  0.00e+00 V2 <--> V2
w 0.510      0.0282    18.11  0.00e+00 V3 <--> V3
x 0.328      0.0401     8.18  2.22e-16 V4 <--> V4
y 0.527      0.0376    14.02  0.00e+00 V5 <--> V5
z 0.696      0.0383    18.19  0.00e+00 V6 <--> V6

```

```
Iterations = 22
```

```
> std.coef(sem.two)
```

```

      Std. Estimate
a a 0.89320      V1 <--- theta1
b b 0.79150      V2 <--- theta1
c c 0.71341      V3 <--- theta1
d d 0.83905      V4 <--- theta2
e e 0.73296      V5 <--- theta2
f f 0.63266      V6 <--- theta2

```

```
> round(residuals(sem.two), 2)
```

```

      V1  V2  V3  V4  V5  V6
V1 0.00 0.00 0.00 0.30 0.23 0.22
V2 0.00 0.00 0.00 0.25 0.22 0.19
V3 0.00 0.00 0.00 0.22 0.20 0.15
V4 0.30 0.25 0.22 0.00 0.00 0.00
V5 0.23 0.22 0.20 0.00 0.00 0.00
V6 0.22 0.19 0.15 0.00 0.00 0.00

```

and so we allow the two factors to be correlated.

```

> S.f2 <- cov(data.f2)
> model.two <- matrix(c("theta1 -> V1", "a", NA, "theta1 -> V2",
+   "b", NA, "theta1 -> V3", "c", NA, "theta2 -> V4", "d", NA,
+   "theta2 -> V5", "e", NA, "theta2 -> V6", "f", NA, "theta1 <-> theta2",
+   "g", NA, "V1 <-> V1", "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3",
+   "w", NA, "V4 <-> V4", "x", NA, "V5 <-> V5", "y", NA, "V6 <-> V6",
+   "z", NA, "theta1 <-> theta1", NA, 1, "theta2 <-> theta2", NA,
+   1), ncol = 3, byrow = TRUE)
> colnames(model.two) <- c("path", "label", "initial estimate")
> model.two

```

```

      path          label initial estimate
[1,] "theta1 -> V1"    "a"      NA

```

```

[2,] "theta1 -> V2"      "b"  NA
[3,] "theta1 -> V3"      "c"  NA
[4,] "theta2 -> V4"      "d"  NA
[5,] "theta2 -> V5"      "e"  NA
[6,] "theta2 -> V6"      "f"  NA
[7,] "theta1 <-> theta2" "g"  NA
[8,] "V1 <-> V1"         "u"  NA
[9,] "V2 <-> V2"         "v"  NA
[10,] "V3 <-> V3"        "w"  NA
[11,] "V4 <-> V4"        "x"  NA
[12,] "V5 <-> V5"        "y"  NA
[13,] "V6 <-> V6"        "z"  NA
[14,] "theta1 <-> theta1" NA   "1"
[15,] "theta2 <-> theta2" NA   "1"

```

```

> sem.two = sem(model.two, S.f2, N)
> summary(sem.two, digits = 3)

```

```

Model Chisquare = 5.39   Df = 8 Pr(>Chisq) = 0.715
Chisquare (null model) = 2206   Df = 15
Goodness-of-fit index = 0.998
Adjusted goodness-of-fit index = 0.995
RMSEA index = 0   90% CI: (NA, 0.0278)
Bentler-Bonnett NFI = 0.998
Tucker-Lewis NNFI = 1.00
Bentler CFI = 1
BIC = -49.9

```

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|-----------|-----------|-----------|-----------|----------|----------|
| | -0.888000 | -0.057500 | -0.000009 | -0.062200 | 0.086200 | 0.443000 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|---|----------|-----------|---------|----------|--------------------|
| a | 0.899 | 0.0280 | 32.10 | 0e+00 | V1 <--- theta1 |
| b | 0.770 | 0.0280 | 27.46 | 0e+00 | V2 <--- theta1 |
| c | 0.726 | 0.0300 | 24.20 | 0e+00 | V3 <--- theta1 |
| d | 0.890 | 0.0332 | 26.84 | 0e+00 | V4 <--- theta2 |
| e | 0.776 | 0.0338 | 22.98 | 0e+00 | V5 <--- theta2 |
| f | 0.679 | 0.0344 | 19.74 | 0e+00 | V6 <--- theta2 |
| g | 0.359 | 0.0337 | 10.65 | 0e+00 | theta2 <--> theta1 |
| u | 0.201 | 0.0256 | 7.86 | 4e-15 | V1 <--> V1 |
| v | 0.354 | 0.0239 | 14.83 | 0e+00 | V2 <--> V2 |
| w | 0.513 | 0.0280 | 18.34 | 0e+00 | V3 <--> V3 |
| x | 0.315 | 0.0378 | 8.34 | 0e+00 | V4 <--> V4 |
| y | 0.536 | 0.0361 | 14.82 | 0e+00 | V5 <--> V5 |
| z | 0.700 | 0.0380 | 18.42 | 0e+00 | V6 <--> V6 |

```
Iterations = 22
```

```

> std.coef(sem.two)

      Std. Estimate
a a 0.89472      V1 <--- theta1
b b 0.79114      V2 <--- theta1
c c 0.71163      V3 <--- theta1
d d 0.84578      V4 <--- theta2
e e 0.72766      V5 <--- theta2
f f 0.63007      V6 <--- theta2

> round(residuals(sem.two), 2)

      V1  V2  V3  V4  V5  V6
V1  0.00 0.00 0.00 0.02 -0.02 0.00
V2  0.00 0.00 0.00 0.00 0.00 0.01
V3  0.00 0.00 0.00 -0.01 -0.01 -0.03
V4  0.02 0.00 -0.01 0.00 0.00 0.00
V5 -0.02 0.00 -0.01 0.00 0.00 0.01
V6  0.00 0.01 -0.03 0.00 0.01 0.00

```

11.3 Hierarchical models

The two correlated factors of section 11.2.1 may be thought of as representing two lower level factors each of which loads on a higher level factor. With just two lower level factors, the loadings on the higher level factor are not unique (one correlation, r , between the two factors may be represented in an infinite number of ways as the product of loadings ga and gb).

There are several ways of representing hierarchical models, including correlated level one factors with a g factor and uncorrelated lower level factors with a g factor (a bifactor solution). The latter may be estimated directly from the data, or may be found by using the Schmid-Leiman transformation of the correlated factors.

11.3.1 Two Correlated factors with a g factor

The hierarchical model of a g factor is underidentified unless we specify one of the g paths. Here we set it to 1 and then estimate the rest of the model.

```

> S.g2 <- cov(data.f2)
> model.g2 <- matrix(c("theta1 -> V1", "a", NA, "theta1 -> V2",
+ "b", NA, "theta1 -> V3", "c", NA, "theta2 -> V4", "d", NA,
+ "theta2 -> V5", "e", NA, "theta2 -> V6", "f", NA, "g -> theta1",
+ NA, 1, "g -> theta2", "g2", NA, "V1 <-> V1", "u", NA, "V2 <-> V2",
+ "v", NA, "V3 <-> V3", "w", NA, "V4 <-> V4", "x", NA, "V5 <-> V5",
+ "y", NA, "V6 <-> V6", "z", NA, "theta1 <-> theta1", NA, 1,
+ "theta2 <-> theta2", NA, 1, "g <-> g", NA, 1), ncol = 3, byrow = TRUE)
> colnames(model.g2) <- c("path", "label", "initial estimate")
> model.g2

```

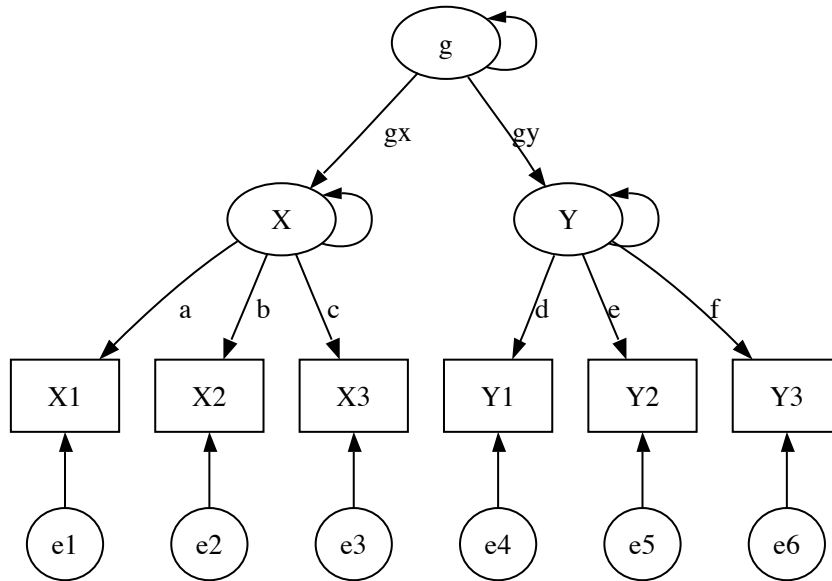


Fig. 11.7 The correlation between two factors may be modeled by a g , general, factor. This representation shows all the errors that need to be estimated.

| | path | label | initial | estimate |
|-------|---------------------|-------|---------|----------|
| [1,] | "theta1 -> V1" | "a" | NA | |
| [2,] | "theta1 -> V2" | "b" | NA | |
| [3,] | "theta1 -> V3" | "c" | NA | |
| [4,] | "theta2 -> V4" | "d" | NA | |
| [5,] | "theta2 -> V5" | "e" | NA | |
| [6,] | "theta2 -> V6" | "f" | NA | |
| [7,] | "g -> theta1" | NA | "1" | |
| [8,] | "g -> theta2" | "g2" | NA | |
| [9,] | "V1 <-> V1" | "u" | NA | |
| [10,] | "V2 <-> V2" | "v" | NA | |
| [11,] | "V3 <-> V3" | "w" | NA | |
| [12,] | "V4 <-> V4" | "x" | NA | |
| [13,] | "V5 <-> V5" | "y" | NA | |
| [14,] | "V6 <-> V6" | "z" | NA | |
| [15,] | "theta1 <-> theta1" | NA | "1" | |
| [16,] | "theta2 <-> theta2" | NA | "1" | |
| [17,] | "g <-> g" | NA | "1" | |

```
> sem.g2 = sem(model.g2, S.g2, N)
> summary(sem.g2, digits = 3)
```

```
Model Chisquare = 5.39  Df = 8  Pr(>Chisq) = 0.715
Chisquare (null model) = 2206  Df = 15
Goodness-of-fit index = 0.998
Adjusted goodness-of-fit index = 0.995
```

```

RMSEA index = 0 90% CI: (NA, 0.0278)
Bentler-Bonnett NFI = 0.998
Tucker-Lewis NNFI = 1.00
Bentler CFI = 1
BIC = -49.9

```

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|-----------|-----------|----------|-----------|----------|----------|
| | -8.88e-01 | -5.75e-02 | 6.72e-07 | -6.22e-02 | 8.61e-02 | 4.43e-01 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|----|----------|-----------|---------|----------|----------------|
| a | 0.636 | 0.0198 | 32.10 | 0.00e+00 | V1 <--- theta1 |
| b | 0.544 | 0.0198 | 27.46 | 0.00e+00 | V2 <--- theta1 |
| c | 0.513 | 0.0212 | 24.20 | 0.00e+00 | V3 <--- theta1 |
| d | 0.767 | 0.0361 | 21.21 | 0.00e+00 | V4 <--- theta2 |
| e | 0.669 | 0.0345 | 19.36 | 0.00e+00 | V5 <--- theta2 |
| f | 0.585 | 0.0337 | 17.33 | 0.00e+00 | V6 <--- theta2 |
| g2 | 0.590 | 0.0747 | 7.90 | 2.89e-15 | theta2 <--- g |
| u | 0.201 | 0.0256 | 7.86 | 4.00e-15 | V1 <--> V1 |
| v | 0.354 | 0.0239 | 14.83 | 0.00e+00 | V2 <--> V2 |
| w | 0.513 | 0.0280 | 18.34 | 0.00e+00 | V3 <--> V3 |
| x | 0.315 | 0.0378 | 8.34 | 0.00e+00 | V4 <--> V4 |
| y | 0.536 | 0.0361 | 14.82 | 0.00e+00 | V5 <--> V5 |
| z | 0.700 | 0.0380 | 18.42 | 0.00e+00 | V6 <--> V6 |

```
Iterations = 26
```

```
> std.coef(sem.g2)
```

| | Std. Estimate | |
|-------|---------------|----------------|
| a a | 0.89472 | V1 <--- theta1 |
| b b | 0.79114 | V2 <--- theta1 |
| c c | 0.71163 | V3 <--- theta1 |
| d d | 0.84578 | V4 <--- theta2 |
| e e | 0.72766 | V5 <--- theta2 |
| f f | 0.63007 | V6 <--- theta2 |
| | 0.70711 | theta1 <--- g |
| g2 g2 | 0.50800 | theta2 <--- g |

```
> round(residuals(sem.g2), 2)
```

| | V1 | V2 | V3 | V4 | V5 | V6 |
|----|-------|------|-------|-------|-------|-------|
| V1 | 0.00 | 0.00 | 0.00 | 0.02 | -0.02 | 0.00 |
| V2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| V3 | 0.00 | 0.00 | 0.00 | -0.01 | -0.01 | -0.03 |
| V4 | 0.02 | 0.00 | -0.01 | 0.00 | 0.00 | 0.00 |
| V5 | -0.02 | 0.00 | -0.01 | 0.00 | 0.00 | 0.01 |
| V6 | 0.00 | 0.01 | -0.03 | 0.00 | 0.01 | 0.00 |

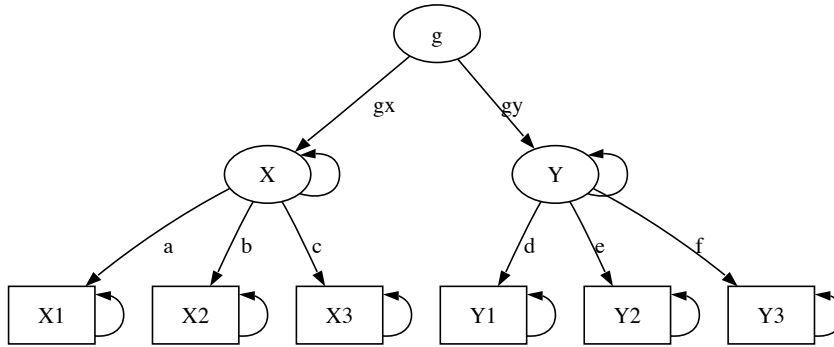


Fig. 11.8 The correlation between two factors may be modeled by a g , general, factor. This representation is somewhat more compact than the previous figure (11.7.)

11.3.2 Generating the data for 3 correlated factors

We have two demonstrations: the first is the two correlated factor data from section 11.2.1, the second is a three correlated factors. To create the later we use the `sim.sem` function with three latent variables.

```

> pattern <- matrix(c(0.9, 0.8, 0.7, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.8,
+ 0.7, 0.6, 0, 0, 0, 0, 0, 0, 0, 0, 0.6, 0.5, 0.4), ncol = 3)
> colnames(pattern) <- c("F1", "F2", "F3")
> rownames(pattern) <- paste("V", 1:dim(pattern)[1], sep = "")
> pattern
      F1 F2 F3
V1 0.9 0.0 0.0
V2 0.8 0.0 0.0
V3 0.7 0.0 0.0
V4 0.0 0.8 0.0
V5 0.0 0.7 0.0
V6 0.0 0.6 0.0
V7 0.0 0.0 0.6
V8 0.0 0.0 0.5
V9 0.0 0.0 0.4

> phi <- matrix(c(1, 0, 0, 0.5, 1, 0, 0.4, 0.4, 1), ncol = 3, byrow = TRUE)
> phi
      [,1] [,2] [,3]
[1,] 1.0 0.0 0
[2,] 0.5 1.0 0
[3,] 0.4 0.4 1

> data.f3 <- sim.sem(loads = pattern, phi = phi)

```



```
> VSS.scree(cor(data.f3))
```

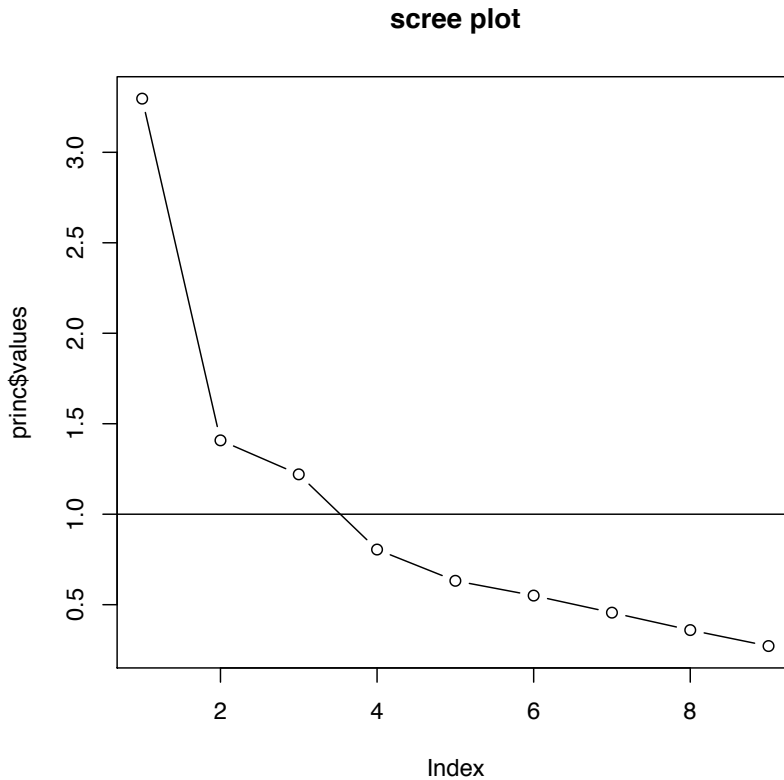


Fig. 11.9 Scree plot of three correlated factors. Compare to the two uncorrelated factors, Figure 11.4, and the two correlated factors, ??

11.3.3 Exploratory factor analysis with 3 factors

As a first approximation to these data, we can do a three factor exploratory analysis to try to understand the structure of the data.

```
> f3 <- factanal(data.f3, 3, rotation = "none")
> f3
```

Call:

```
factanal(x = data.f3, factors = 3, rotation = "none")
```

Uniquenesses:

| V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.203 | 0.357 | 0.516 | 0.319 | 0.412 | 0.622 | 0.485 | 0.746 | 0.839 |

Loadings:

| | Factor1 | Factor2 | Factor3 |
|----|---------|---------|---------|
| V1 | 0.835 | -0.316 | |
| V2 | 0.745 | -0.297 | |
| V3 | 0.659 | -0.221 | |
| V4 | 0.648 | 0.497 | -0.121 |
| V5 | 0.572 | 0.486 | -0.154 |
| V6 | 0.459 | 0.399 | |
| V7 | 0.322 | 0.188 | 0.613 |
| V8 | 0.226 | 0.196 | 0.406 |
| V9 | 0.196 | 0.148 | 0.317 |

| | Factor1 | Factor2 | Factor3 |
|----------------|---------|---------|---------|
| SS loadings | 2.837 | 0.974 | 0.688 |
| Proportion Var | 0.315 | 0.108 | 0.076 |
| Cumulative Var | 0.315 | 0.424 | 0.500 |

Test of the hypothesis that 3 factors are sufficient.
 The chi square statistic is 11.91 on 12 degrees of freedom.
 The p-value is 0.453

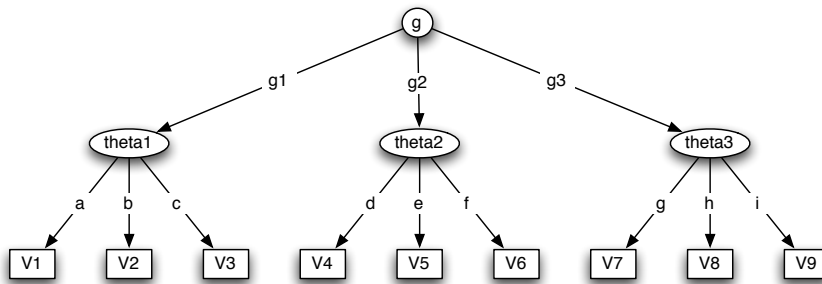


Fig. 11.10 The correlation between three factors may be modeled by a g , general, factor.

11.3.3.1 Orthogonal Rotation

The loadings from this factor analysis are not particularly easy to understand and can be rotated to a more somewhat more understandable structure using the VARIMAX rotation (which is actually the default for factanal). We use the **GPArotation** package.

```

> library(GPArotation)
> f3v <- Varimax(loadings(f3))
> round(loadings(f3v), 2)

```

| | Factor1 | Factor2 | Factor3 |
|----|---------|---------|---------|
| V1 | 0.87 | 0.20 | 0.07 |
| V2 | 0.78 | 0.17 | 0.05 |

| | | | |
|----|------|------|------|
| V3 | 0.67 | 0.17 | 0.09 |
| V4 | 0.26 | 0.77 | 0.12 |
| V5 | 0.20 | 0.74 | 0.07 |
| V6 | 0.16 | 0.59 | 0.09 |
| V7 | 0.14 | 0.14 | 0.69 |
| V8 | 0.06 | 0.16 | 0.48 |
| V9 | 0.07 | 0.13 | 0.37 |

The structure is more easy to understand than the original one, but still is somewhat hard to understand.

11.3.3.2 Oblique Rotation

By allowing the factors to be correlated, we are able to find a more simple representation of the factor pattern. However, we need to report both the factor loadings as well as the factor intercorrelations.

```
> f3o <- oblimin(loadings(f3))
> round(loadings(f3o), 2)
```

| | Factor1 | Factor2 | Factor3 |
|----|---------|---------|---------|
| V1 | 0.89 | 0.00 | 0.00 |
| V2 | 0.81 | -0.01 | -0.01 |
| V3 | 0.68 | 0.02 | 0.03 |
| V4 | 0.03 | 0.80 | 0.02 |
| V5 | -0.01 | 0.78 | -0.03 |
| V6 | -0.02 | 0.62 | 0.01 |
| V7 | 0.02 | -0.02 | 0.72 |
| V8 | -0.04 | 0.06 | 0.49 |
| V9 | -0.02 | 0.05 | 0.38 |

The alternatives to exploratory factor analysis is to apply a confirmatory model specifying the “expected” structure. We do this with both a hierarchical g factor model as well as a bifactor model.

11.3.4 Three correlated factors with a g factor

```
> S.g3 <- cov(data.f3)
> model.g3 <- matrix(c("theta1 -> V1", "a", NA, "theta1 -> V2", "b",
+ NA, "theta1 -> V3", "c", NA, "theta2 -> V4", "d", NA, "theta2 -> V5",
+ "e", NA, "theta2 -> V6", "f", NA, "theta3 -> V7", "g", NA,
+ "theta3 -> V8", "h", NA, "theta3 -> V9", "i", NA, "g -> theta1",
+ "g1", NA, "g -> theta2", "g2", NA, "g -> theta3", "g3", NA,
+ "V1 <-> V1", "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3", "w",
+ NA, "V4 <-> V4", "x", NA, "V5 <-> V5", "y", NA, "V6 <-> V6",
+ "z", NA, "V7 <-> V7", "s", NA, "V8 <-> V8", "t", NA, "V9 <-> V9",
+ "r", NA, "theta1 <-> theta1", NA, 1, "theta2 <-> theta2", NA,
```

```
+ 1, "theta3 <-> theta3", NA, 1, "g <-> g", NA, 1), ncol = 3,
+ byrow = TRUE)
> colnames(model.g3) <- c("path", "label", "initial estimate")
> model.g3
```

| | path | label | initial estimate |
|-------|---------------------|-------|------------------|
| [1,] | "theta1 -> V1" | "a" | NA |
| [2,] | "theta1 -> V2" | "b" | NA |
| [3,] | "theta1 -> V3" | "c" | NA |
| [4,] | "theta2 -> V4" | "d" | NA |
| [5,] | "theta2 -> V5" | "e" | NA |
| [6,] | "theta2 -> V6" | "f" | NA |
| [7,] | "theta3 -> V7" | "g" | NA |
| [8,] | "theta3 -> V8" | "h" | NA |
| [9,] | "theta3 -> V9" | "i" | NA |
| [10,] | "g -> theta1" | "g1" | NA |
| [11,] | "g -> theta2" | "g2" | NA |
| [12,] | "g -> theta3" | "g3" | NA |
| [13,] | "V1 <-> V1" | "u" | NA |
| [14,] | "V2 <-> V2" | "v" | NA |
| [15,] | "V3 <-> V3" | "w" | NA |
| [16,] | "V4 <-> V4" | "x" | NA |
| [17,] | "V5 <-> V5" | "y" | NA |
| [18,] | "V6 <-> V6" | "z" | NA |
| [19,] | "V7 <-> V7" | "s" | NA |
| [20,] | "V8 <-> V8" | "t" | NA |
| [21,] | "V9 <-> V9" | "r" | NA |
| [22,] | "theta1 <-> theta1" | NA | "1" |
| [23,] | "theta2 <-> theta2" | NA | "1" |
| [24,] | "theta3 <-> theta3" | NA | "1" |
| [25,] | "g <-> g" | NA | "1" |

```
> sem.g3 = sem(model.g3, S.g3, N)
> summary(sem.g3, digits = 3)
```

```
Model Chisquare = 20.5  Df = 24 Pr(>Chisq) = 0.665
Chisquare (null model) = 2647  Df = 36
Goodness-of-fit index = 0.995
Adjusted goodness-of-fit index = 0.991
RMSEA index = 0  90% CI: (NA, 0.0211)
Bentler-Bonnett NFI = 0.992
Tucker-Lewis NNFI = 1.00
Bentler CFI = 1
BIC = -145
```

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|-----------|-----------|----------|----------|----------|----------|
| | -1.27e+00 | -1.64e-01 | 1.63e-05 | 3.16e-02 | 3.43e-01 | 1.19e+00 |

```

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a  0.699   0.0380  18.40  0.00e+00 V1 <--- theta1
b  0.642   0.0361  17.77  0.00e+00 V2 <--- theta1
c  0.542   0.0327  16.55  0.00e+00 V3 <--- theta1
d  0.543   0.0744   7.29  3.02e-13 V4 <--- theta2
e  0.482   0.0664   7.26  3.79e-13 V5 <--- theta2
f  0.379   0.0535   7.09  1.34e-12 V6 <--- theta2
g  0.618   0.0485  12.72  0.00e+00 V7 <--- theta3
h  0.461   0.0392  11.77  0.00e+00 V8 <--- theta3
i  0.377   0.0379   9.95  0.00e+00 V9 <--- theta3
g1 0.788   0.0985   8.00  1.11e-15 theta1 <--- g
g2 1.370   0.2804   4.89  1.02e-06 theta2 <--- g
g3 0.583   0.0758   7.69  1.47e-14 theta3 <--- g
u  0.204   0.0238   8.54  0.00e+00 V1 <--> V1
v  0.375   0.0251  14.92  0.00e+00 V2 <--> V2
w  0.503   0.0268  18.78  0.00e+00 V3 <--> V3
x  0.342   0.0364   9.39  0.00e+00 V4 <--> V4
y  0.524   0.0356  14.71  0.00e+00 V5 <--> V5
z  0.702   0.0365  19.21  0.00e+00 V6 <--> V6
s  0.575   0.0609   9.43  0.00e+00 V7 <--> V7
t  0.781   0.0475  16.44  0.00e+00 V8 <--> V8
r  0.925   0.0480  19.26  0.00e+00 V9 <--> V9

```

```
Iterations = 35
```

```
> std.coef(sem.g3)
```

```

Std. Estimate
a a 0.89200 V1 <--- theta1
b b 0.80019 V2 <--- theta1
c c 0.69736 V3 <--- theta1
d d 0.84402 V4 <--- theta2
e e 0.74887 V5 <--- theta2
f f 0.60882 V6 <--- theta2
g g 0.68604 V7 <--- theta3
h h 0.51711 V8 <--- theta3
i i 0.41359 V9 <--- theta3
g1 g1 0.61912 theta1 <--- g
g2 g2 0.80777 theta2 <--- g
g3 g3 0.50360 theta3 <--- g

```

11.3.5 Bifactor solutions

An alternative to the correlated lower level factors and a g factor is a “bifactor” model where each item is represented by two factors, a lower level, group, factor and a higher level, g,

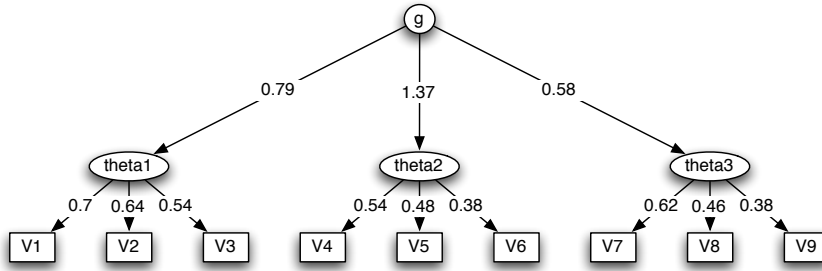


Fig. 11.11 A hierarchical solution to the three correlated factors problem.

factor. This may be found directly through sem - cfa, or may be done indirectly by using a Schmid-Leiman transformation of the correlated factors. We use the same three factor data set as in the two previous sections (11.3.2, 11.3.4)

| | path | label | initial | estimate |
|-------|---------------------|-------|---------|----------|
| [1,] | "theta1 -> V1" | "a" | NA | |
| [2,] | "theta1 -> V2" | "b" | NA | |
| [3,] | "theta1 -> V3" | "c" | NA | |
| [4,] | "theta2 -> V4" | "d" | NA | |
| [5,] | "theta2 -> V5" | "e" | NA | |
| [6,] | "theta2 -> V6" | "f" | NA | |
| [7,] | "theta3 -> V7" | "g" | NA | |
| [8,] | "theta3 -> V8" | "h" | NA | |
| [9,] | "theta3 -> V9" | "i" | NA | |
| [10,] | "g -> V1" | "g1" | NA | |
| [11,] | "g -> V2" | "g2" | NA | |
| [12,] | "g -> V3" | "g3" | NA | |
| [13,] | "g -> V4" | "g4" | NA | |
| [14,] | "g -> V5" | "g5" | NA | |
| [15,] | "g -> V6" | "g6" | NA | |
| [16,] | "g -> V7" | "g7" | NA | |
| [17,] | "g -> V8" | "g8" | NA | |
| [18,] | "g -> V9" | "g9" | NA | |
| [19,] | "V1 <-> V1" | "u" | NA | |
| [20,] | "V2 <-> V2" | "v" | NA | |
| [21,] | "V3 <-> V3" | "w" | NA | |
| [22,] | "V4 <-> V4" | "x" | NA | |
| [23,] | "V5 <-> V5" | "y" | NA | |
| [24,] | "V6 <-> V6" | "z" | NA | |
| [25,] | "V7 <-> V7" | "s" | NA | |
| [26,] | "V8 <-> V8" | "t" | NA | |
| [27,] | "V9 <-> V9" | "r" | NA | |
| [28,] | "theta1 <-> theta1" | NA | "1" | |
| [29,] | "theta2 <-> theta2" | NA | "1" | |
| [30,] | "theta3 <-> theta3" | NA | "1" | |
| [31,] | "g <-> g" | NA | "1" | |

Model Chisquare = 16.8 Df = 18 Pr(>Chisq) = 0.536
 Chisquare (null model) = 2647 Df = 36
 Goodness-of-fit index = 0.996
 Adjusted goodness-of-fit index = 0.99
 RMSEA index = 0 90% CI: (NA, 0.0263)
 Bentler-Bonnett NFI = 0.994
 Tucker-Lewis NNFI = 1
 Bentler CFI = 1
 BIC = -108

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|-----------|-----------|-----------|----------|----------|----------|
| | -9.92e-01 | -5.80e-02 | -1.59e-05 | 1.55e-02 | 1.22e-01 | 8.96e-01 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|----|----------|-----------|---------|----------|----------------|
| a | 0.697 | 0.0412 | 16.91 | 0.00e+00 | V1 <--- theta1 |
| b | 0.658 | 0.0406 | 16.21 | 0.00e+00 | V2 <--- theta1 |
| c | 0.517 | 0.0416 | 12.44 | 0.00e+00 | V3 <--- theta1 |
| d | 0.501 | 0.0842 | 5.95 | 2.63e-09 | V4 <--- theta2 |
| e | 0.533 | 0.0755 | 7.06 | 1.69e-12 | V5 <--- theta2 |
| f | 0.409 | 0.0710 | 5.77 | 8.14e-09 | V6 <--- theta2 |
| g | 0.642 | 0.0745 | 8.62 | 0.00e+00 | V7 <--- theta3 |
| h | 0.451 | 0.0579 | 7.78 | 7.33e-15 | V8 <--- theta3 |
| i | 0.355 | 0.0512 | 6.94 | 4.06e-12 | V9 <--- theta3 |
| g1 | 0.552 | 0.0477 | 11.57 | 0.00e+00 | V1 <--- g |
| g2 | 0.490 | 0.0474 | 10.34 | 0.00e+00 | V2 <--- g |
| g3 | 0.455 | 0.0466 | 9.77 | 0.00e+00 | V3 <--- g |
| g4 | 0.758 | 0.0603 | 12.56 | 0.00e+00 | V4 <--- g |
| g5 | 0.639 | 0.0591 | 10.81 | 0.00e+00 | V5 <--- g |
| g6 | 0.503 | 0.0560 | 8.98 | 0.00e+00 | V6 <--- g |
| g7 | 0.356 | 0.0426 | 8.37 | 0.00e+00 | V7 <--- g |
| g8 | 0.268 | 0.0411 | 6.53 | 6.37e-11 | V8 <--- g |
| g9 | 0.237 | 0.0423 | 5.60 | 2.15e-08 | V9 <--- g |
| u | 0.206 | 0.0272 | 7.57 | 3.84e-14 | V1 <--> V1 |
| v | 0.369 | 0.0283 | 13.06 | 0.00e+00 | V2 <--> V2 |
| w | 0.504 | 0.0267 | 18.87 | 0.00e+00 | V3 <--> V3 |
| x | 0.364 | 0.0370 | 9.84 | 0.00e+00 | V4 <--> V4 |
| y | 0.500 | 0.0431 | 11.60 | 0.00e+00 | V5 <--> V5 |
| z | 0.695 | 0.0378 | 18.37 | 0.00e+00 | V6 <--> V6 |
| s | 0.546 | 0.0898 | 6.09 | 1.15e-09 | V7 <--> V7 |
| t | 0.791 | 0.0553 | 14.30 | 0.00e+00 | V8 <--> V8 |
| r | 0.933 | 0.0494 | 18.88 | 0.00e+00 | V9 <--> V9 |

Iterations = 54

| | Std. Estimate | |
|-----|---------------|----------------|
| a a | 0.69823 | V1 <--- theta1 |

```

b b 0.64468      V2 <--- theta1
c c 0.52288      V3 <--- theta1
d d 0.45929      V4 <--- theta2
e e 0.48816      V5 <--- theta2
f f 0.38775      V6 <--- theta2
g g 0.61642      V7 <--- theta3
h h 0.43658      V8 <--- theta3
i i 0.33635      V9 <--- theta3
g1 g1 0.55298    V1 <--- g
g2 g2 0.47981    V2 <--- g
g3 g3 0.45993    V3 <--- g
g4 g4 0.69494    V4 <--- g
g5 g5 0.58531    V5 <--- g
g6 g6 0.47594    V6 <--- g
g7 g7 0.34180    V7 <--- g
g8 g8 0.25990    V8 <--- g
g9 g9 0.22427    V9 <--- g

```

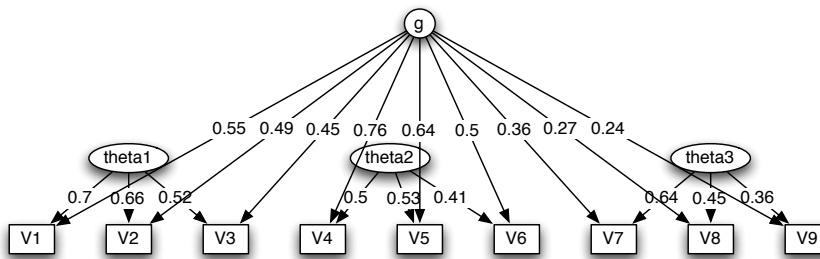


Fig. 11.12 A bifactor solution to the three correlated factors problem.

11.3.6 Schmid Leiman transformations to orthogonalize the factors

An alternative to a confirmatory hierarchical analysis or bifactor solution is to extract at least 3 factors from a correlation matrix, transform them obliquely to a simple pattern solution, and then extract the first factor from the correlations of these factors. From the resulting matrices, it is possible to find the g loading for each of the original variables (based upon the product of the g loadings of the factors and the loadings of the variables on these factors). Functionally, this is an alternative way of estimating a bifactor solution.

The `schmid` function found in the `psych` package finds the g factor and group factor loadings by doing a Schmid Leiman transformation. Here we show the Schmid Leiman analysis applied to the 3 correlated factors problem of section 11.3.4.

```

$s1
  g factor Factor1 Factor2 Factor3  h2  u2

```


| | | | | | | |
|----|------|-------|--------|--------|------|------|
| V1 | 0.56 | 0.692 | 0.0016 | 0.0031 | 0.80 | 0.20 |
| V2 | 0.50 | 0.629 | 0.0053 | 0.0122 | 0.66 | 0.34 |
| V3 | 0.46 | 0.525 | 0.0098 | 0.0290 | 0.46 | 0.54 |
| V4 | 0.66 | 0.022 | 0.5005 | 0.0161 | 0.65 | 0.35 |
| V5 | 0.59 | 0.011 | 0.4874 | 0.0226 | 0.61 | 0.39 |
| V6 | 0.48 | 0.017 | 0.3859 | 0.0129 | 0.38 | 0.62 |
| V7 | 0.34 | 0.014 | 0.0150 | 0.6298 | 0.52 | 0.48 |
| V8 | 0.26 | 0.034 | 0.0376 | 0.4296 | 0.25 | 0.75 |
| V9 | 0.22 | 0.012 | 0.0308 | 0.3361 | 0.15 | 0.85 |

\$orthog

| | Factor1 | Factor2 | Factor3 |
|----|---------|---------|---------|
| V1 | 0.892 | 0.0025 | -0.0035 |
| V2 | 0.810 | -0.0085 | -0.0139 |
| V3 | 0.677 | 0.0157 | 0.0332 |
| V4 | 0.028 | 0.8039 | 0.0185 |
| V5 | -0.014 | 0.7829 | -0.0258 |
| V6 | -0.022 | 0.6198 | 0.0148 |
| V7 | 0.018 | -0.0240 | 0.7207 |
| V8 | -0.044 | 0.0603 | 0.4916 |
| V9 | -0.016 | 0.0495 | 0.3846 |

\$fcor

| | [,1] | [,2] | [,3] |
|------|------|------|------|
| [1,] | 1.00 | 0.49 | 0.31 |
| [2,] | 0.49 | 1.00 | 0.38 |
| [3,] | 0.31 | 0.38 | 1.00 |

\$gloading

Loadings:

| | Factor1 |
|------|---------|
| [1,] | 0.631 |
| [2,] | 0.783 |
| [3,] | 0.486 |

| | Factor1 |
|----------------|---------|
| SS loadings | 1.247 |
| Proportion Var | 0.416 |

Although not identical to the results of the bifactor solution, the results agree to two decimal places. Why are these estimates not the same? Because in the case of the confirmatory model, the loadings of the variables on one factor on other factors are set to 0, while in the Schmid Leiman case, they are allowed to be non-zero.

The loadings on the general factor are used to calculate the ω_h coefficient discussed by Zinbarg et al. 2006.

11.3.7 Omega as an estimate of reliability

Many scales are assumed by their developers and users to be primarily a measure of one latent variable. When it is also assumed that the scale conforms to the effect indicator model of measurement (as is almost always the case in psychological assessment), it is important to support such an interpretation with evidence regarding the internal structure of that scale. In particular, it is important to examine two related properties pertaining to the internal structure of such a scale. The first property relates to whether all the indicators forming the scale measure a latent variable in common.

The second internal structural property pertains to the proportion of variance in the scale scores (derived from summing or averaging the indicators) accounted for by this latent variable that is common to all the indicators (Cronbach, 1951; McDonald, 1999; Revelle, 1979). That is, if an effect indicator scale is primarily a measure of one latent variable common to all the indicators forming the scale, then that latent variable should account for the majority of the variance in the scale scores. Put differently, this variance ratio provides important information about the sampling fluctuations when estimating individuals' standing on a latent variable common to all the indicators arising from the sampling of indicators (i.e., when dealing with either Type 2 or Type 12 sampling, to use the terminology of Lord, 1956). That is, this variance proportion can be interpreted as the square of the correlation between the scale score and the latent variable common to all the indicators in the infinite universe of indicators of which the scale indicators are a subset. Put yet another way, this variance ratio is important both as reliability and a validity coefficient. This is a reliability issue as the larger this variance ratio is, the more accurately one can predict an individual's relative standing on the latent variable common to all the scale's indicators based on his or her observed scale score. At the same time, this variance ratio also bears on the construct validity of the scale given that construct validity encompasses the internal structure of a scale." (Zinbarg, Yovel, Revelle, and McDonald, 2006). McDonald has proposed coefficient omega as an estimate of the general factor saturation of a test. [Zinbarg, Revelle, Yovel and Li \(2005\)](#) compare McDonald's Omega to Cronbach's alpha and Revelle's beta. They conclude that omega is the best estimate. (See also Zinbarg et al., 2006)

One way to find omega is to do a factor analysis of the original data set, rotate the factors obliquely, do a Schmid-Leiman (schmid) transformation, and then find omega. The **psych** package function **omega** does that.

Chapter 12

sem in R and in LISREL

There are many programs that allow one to analyze latent variable models. Almost all statistical packages will include the ability to do exploratory factor analysis and many allow for confirmatory analysis. Commercially available sem programs include AMOS, EQS, LISREL, MPlus, and SAS. Open source programs include R and Mx. The Loehlin text gives sample code for many problems in LISREL and EQS syntax, Raykov and Marcoulides (2006) give examples in EQS, LISREL and Mplus. In this chapter we compare the set up and output of the sem package in R with the unix version of LISREL for several problems.

12.1 Example data set 1: 9 cognitive variables (from Rakov and Marcoulides)

Tenko Raykov and George Marcoulides, in their textbook on SEM (Raykov and Marcoulides, 2006), present a data set based upon 220 high school students on 9 cognitive measures. They report three measures of Induction taken in the junior year, three measures of Figural Relations in the junior year, and three measures of figural relations in the senior year.

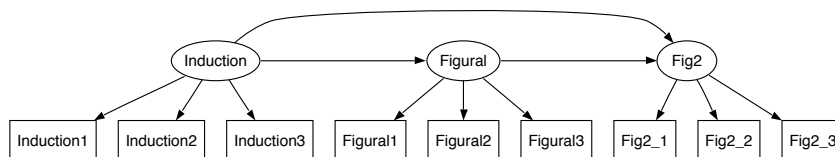


Fig. 12.1 9 cognitive variables (adapted from Raykov and Marcoulides, 2006)

They present the data set as a lower triangular covariance matrix which we can read this into R using the `scan` function embedded in a function to convert the data to a rectangular matrix. This function has been written so that it should work on both Macs and PCs:

```
> lower.triangle <- function(nrow = 2, data = NULL) {
+   if (is.null(data)) {
+     MAC <- Sys.info()[1] == "Darwin"
```

```

+       if (!MAC) {
+         data <- scan(file("clipboard"))
+       }
+       else data <- scan(pipe("pbpaste"))
+     }
+     mat <- diag(0, nrow)
+     k <- 1
+     for (i in 1:nrow) {
+       for (j in 1:i) {
+         mat[i, j] <- data[k]
+         k <- k + 1
+       }
+     }
+     mat <- mat + t(mat)
+     diag(mat) <- diag(mat)/2
+     return(mat)
+ }

```

```

p5.in <- scan("")
56.21
31.55 75.55
23.27 28.30 44.45
24.48 32.24 22.56 84.64
22.51 29.54 20.61 57.61 78.93
22.65 27.56 15.33 53.57 49.27 73.76
33.24 46.49 31.44 67.81 54.76 54.58 141.77
32.56 40.37 25.58 55.82 52.33 47.74 98.62 117.33
30.32 40.44 27.69 54.78 53.44 59.52 96.95 84.87 106.35

```

```
prob5 <- lower.triangle(9,p5.in)
```

```

> colnames(prob5) <- rownames(prob5) <- c("Induct1", "Induct2", "Induct3",
+     "Figural1", "Figural2", "Figural3", "Fig2.1", "Fig2.2", "Fig2.3")
> prob5

```

| | Induct1 | Induct2 | Induct3 | Figural1 | Figural2 | Figural3 | Fig2.1 | Fig2.2 | Fig2.3 |
|----------|---------|---------|---------|----------|----------|----------|--------|--------|--------|
| Induct1 | 56.21 | 31.55 | 23.27 | 24.48 | 22.51 | 22.65 | 33.24 | 32.56 | 30.32 |
| Induct2 | 31.55 | 75.55 | 28.30 | 32.24 | 29.54 | 27.56 | 46.49 | 40.37 | 40.44 |
| Induct3 | 23.27 | 28.30 | 44.45 | 22.56 | 20.61 | 15.33 | 31.44 | 25.58 | 27.69 |
| Figural1 | 24.48 | 32.24 | 22.56 | 84.64 | 57.61 | 53.57 | 67.81 | 55.82 | 54.78 |
| Figural2 | 22.51 | 29.54 | 20.61 | 57.61 | 78.93 | 49.27 | 54.76 | 52.33 | 53.44 |
| Figural3 | 22.65 | 27.56 | 15.33 | 53.57 | 49.27 | 73.76 | 54.58 | 47.74 | 59.52 |
| Fig2.1 | 33.24 | 46.49 | 31.44 | 67.81 | 54.76 | 54.58 | 141.77 | 98.62 | 96.95 |
| Fig2.2 | 32.56 | 40.37 | 25.58 | 55.82 | 52.33 | 47.74 | 98.62 | 117.33 | 84.87 |
| Fig2.3 | 30.32 | 40.44 | 27.69 | 54.78 | 53.44 | 59.52 | 96.95 | 84.87 | 106.35 |

12.2 Using R to analyze the data set

The model proposed for this is that Induction in year1 predicts Figural Ability in Year 1 and Year 2 and that Figural Ability in Year 1 predicts Figural Ability in Year 2.

12.2.1 An initial formulation is empirically underidentified

The R code for doing the basic analysis is straightforward:

```

      path                                label initial estimate
[1,] "Induction -> Induct1"              NA    "1"
[2,] "Induction -> Induct2"              "2"   NA
[3,] "Induction -> Induct3"              "3"   NA
[4,] "Figural -> Figural1"               NA    "1"
[5,] "Figural -> Figural2"              "5"   NA
[6,] "Figural -> Figural3"              "6"   NA
[7,] "Figural.time2 -> Fig2.1"          NA    "1"
[8,] "Figural.time2 -> Fig2.2"          "8"   NA
[9,] "Figural.time2 -> Fig2.3"          "9"   NA
[10,] "Induction -> Figural"             "i"   NA
[11,] "Induction -> Figural.time2"      "j"   NA
[12,] "Figural -> Figural.time2"        "k"   NA
[13,] "Induct1 <-> Induct1"              "u"   NA
[14,] "Induct2 <-> Induct2"              "v"   NA
[15,] "Induct3 <-> Induct3"              "w"   NA
[16,] "Figural1 <-> Figural1"            "x"   NA
[17,] "Figural2 <-> Figural2"            "y"   NA
[18,] "Figural3 <-> Figural3"            "z"   NA
[19,] "Fig2.1 <-> Fig2.1"                "q"   NA
[20,] "Fig2.2 <-> Fig2.2"                "r"   NA
[21,] "Fig2.3 <-> Fig2.3"                "s"   NA
[22,] "Induction <-> Induction"          "A"   "1"
[23,] "Figural <-> Figural"              "B"   "1"
[24,] "Figural.time2 <-> Figural.time2"  "C"   "1"

```

```

Model Chisquare = 124   Df = 24 Pr(>Chisq) = 2.1e-15
Chisquare (null model) = 1177   Df = 36
Goodness-of-fit index = 0.88
Adjusted goodness-of-fit index = 0.78
RMSEA index = 0.14   90% CI: (0.11, 0.16)
Bentler-Bonnett NFI = 0.9
Tucker-Lewis NNFI = 0.87
Bentler CFI = 0.91
BIC = -5.7

```

Normalized Residuals

```

      Min.  1st Qu.  Median    Mean  3rd Qu.    Max.

```

-1.6e+00 -4.7e-01 6.3e-05 1.4e-01 5.5e-01 3.2e+00

```

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
2  1.3e+00 0.118      10.6   0.0e+00 Induct2 <--- Induction
3  8.5e-01 0.100       8.5   0.0e+00 Induct3 <--- Induction
5  9.3e-01 0.027      34.6   0.0e+00 Figural2 <--- Figural
6  8.8e-01 0.022      40.6   0.0e+00 Figural3 <--- Figural
8  8.8e-01 0.039      22.3   0.0e+00 Fig2.2 <--- Figural.time2
9  8.8e-01 0.028      31.6   0.0e+00 Fig2.3 <--- Figural.time2
i  2.0e+00   NaN       NaN     NaN   Figural <--- Induction
j -2.0e+03   NaN       NaN     NaN   Figural.time2 <--- Induction
k  1.0e+03   NaN       NaN     NaN   Figural.time2 <--- Figural
u  4.2e+01 4.208       10.0   0.0e+00 Induct1 <--> Induct1
v  5.3e+01 5.347        9.9   0.0e+00 Induct2 <--> Induct2
w  3.4e+01 3.391       10.0   0.0e+00 Induct3 <--> Induct3
x  2.6e+01 3.045        8.5   0.0e+00 Figural1 <--> Figural1
y  2.9e+01 3.535        8.2   2.2e-16 Figural2 <--> Figural2
z  2.8e+01 3.382        8.3   0.0e+00 Figural3 <--> Figural3
q  3.2e+01 4.234        7.5   8.8e-14 Fig2.1 <--> Fig2.1
r  3.2e+01 4.058        8.0   1.8e-15 Fig2.2 <--> Fig2.2
s  2.0e+01 3.051        6.6   3.7e-11 Fig2.3 <--> Fig2.3
A  1.4e+01   NaN       NaN     NaN   Induction <--> Induction
B -7.0e-04   NaN       NaN     NaN   Figural <--> Figural
C  7.5e+02   NaN       NaN     NaN   Figural.time2 <--> Figural.time2

```

Iterations = 500

Aliased parameters: i j k A B C

12.2.2 Adjusting to model to converge

Unfortunately, the estimation in 12.2.1 does not converge and failed after 500 iterations. This is not an unusual problem in estimation. By specifying start values for the Induction -> Figural.time2 path, we can get a satisfactory solution:

```

      path                                label initial estimate
[1,] "Induction -> Induct1"              NA "1"
[2,] "Induction -> Induct2"              "2" NA
[3,] "Induction -> Induct3"              "3" NA
[4,] "Figural -> Figural1"               NA "1"
[5,] "Figural -> Figural2"              "5" NA
[6,] "Figural -> Figural3"              "6" NA
[7,] "Figural.time2 -> Fig2.1"           NA "1"
[8,] "Figural.time2 -> Fig2.2"          "8" NA
[9,] "Figural.time2 -> Fig2.3"          "9" NA

```

```

[10,] "Induction -> Figural"           "i" NA
[11,] "Induction -> Figural.time2"    "j" NA
[12,] "Figural -> Figural.time2"      "k" "0.75"
[13,] "Induct1 <-> Induct1"           "u" NA
[14,] "Induct2 <-> Induct2"           "v" NA
[15,] "Induct3 <-> Induct3"           "w" NA
[16,] "Figural1 <-> Figural1"         "x" NA
[17,] "Figural2 <-> Figural2"         "y" NA
[18,] "Figural3 <-> Figural3"         "z" NA
[19,] "Fig2.1 <-> Fig2.1"            "q" NA
[20,] "Fig2.2 <-> Fig2.2"            "r" NA
[21,] "Fig2.3 <-> Fig2.3"            "s" NA
[22,] "Induction <-> Induction"       "A" "1"
[23,] "Figural <-> Figural"           "B" "1"
[24,] "Figural.time2 <-> Figural.time2" "C" "1"

```

```

Model Chisquare = 52 Df = 24 Pr(>Chisq) = 0.00076
Chisquare (null model) = 1177 Df = 36
Goodness-of-fit index = 0.95
Adjusted goodness-of-fit index = 0.91
RMSEA index = 0.073 90% CI: (0.046, 0.1)
Bentler-Bonnett NFI = 0.96
Tucker-Lewis NNFI = 0.96
Bentler CFI = 0.98
BIC = -77

```

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|----------|----------|----------|----------|---------|---------|
| | -9.5e-01 | -8.9e-02 | -7.3e-05 | -1.2e-02 | 1.4e-01 | 1.3e+00 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|---|----------|-----------|---------|----------|------------------------------|
| 2 | 1.27 | 0.159 | 8.0 | 1.3e-15 | Induct2 <--- Induction |
| 3 | 0.89 | 0.114 | 7.8 | 7.1e-15 | Induct3 <--- Induction |
| 5 | 0.92 | 0.066 | 13.9 | 0.0e+00 | Figural2 <--- Figural |
| 6 | 0.88 | 0.066 | 13.4 | 0.0e+00 | Figural3 <--- Figural |
| 8 | 0.88 | 0.052 | 16.9 | 0.0e+00 | Fig2.2 <--- Figural.time2 |
| 9 | 0.88 | 0.048 | 18.2 | 0.0e+00 | Fig2.3 <--- Figural.time2 |
| i | 0.98 | 0.147 | 6.6 | 3.6e-11 | Figural <--- Induction |
| j | 0.60 | 0.178 | 3.4 | 7.2e-04 | Figural.time2 <--- Induction |
| k | 0.81 | 0.110 | 7.4 | 1.5e-13 | Figural.time2 <--- Figural |
| u | 30.90 | 3.891 | 7.9 | 2.0e-15 | Induct1 <--> Induct1 |
| v | 34.83 | 5.067 | 6.9 | 6.2e-12 | Induct2 <--> Induct2 |
| w | 24.49 | 3.075 | 8.0 | 1.8e-15 | Induct3 <--> Induct3 |
| x | 22.83 | 3.450 | 6.6 | 3.7e-11 | Figural1 <--> Figural1 |
| y | 26.87 | 3.459 | 7.8 | 8.0e-15 | Figural2 <--> Figural2 |
| z | 26.33 | 3.353 | 7.9 | 4.0e-15 | Figural3 <--> Figural3 |
| q | 31.31 | 4.451 | 7.0 | 2.0e-12 | Fig2.1 <--> Fig2.1 |

```

r 32.17    4.043    8.0    1.8e-15  Fig2.2 <--> Fig2.2
s 20.44    3.213    6.4    2.0e-10  Fig2.3 <--> Fig2.3
A 25.31    5.156    4.9    9.1e-07  Induction <--> Induction
B 37.70    6.085    6.2    5.8e-10  Figural <--> Figural
C 36.00    6.017    6.0    2.2e-09  Figural.time2 <--> Figural.time2

```

Iterations = 168

```

      Std. Estimate
1   0.67105      Induct1 <--- Induction
2  2 0.73412      Induct2 <--- Induction
3  3 0.67018      Induct3 <--- Induction
4   0.85457      Figural1 <--- Figural
5  5 0.81215      Figural2 <--- Figural
6  6 0.80191      Figural3 <--- Figural
7   0.88269      Fig2.1 <--- Figural.time2
8  8 0.85197      Fig2.2 <--- Figural.time2
9  9 0.89877      Fig2.3 <--- Figural.time2
10 i 0.62464      Figural <--- Induction
11 j 0.28886      Figural.time2 <--- Induction
12 k 0.60902      Figural.time2 <--- Figural

```

| | Induct1 | Induct2 | Induct3 | Figural1 | Figural2 | Figural3 | Fig2.1 | Fig2.2 | Fig2.3 |
|----------|---------|---------|---------|----------|----------|----------|--------|--------|--------|
| Induct1 | 0.00 | -0.55 | 0.79 | -0.23 | -0.17 | 1.01 | -2.15 | 1.49 | -0.89 |
| Induct2 | -0.55 | 0.00 | -0.21 | 0.90 | 0.78 | 0.11 | 1.61 | 0.96 | 0.86 |
| Induct3 | 0.79 | -0.21 | 0.00 | 0.62 | 0.47 | -3.89 | 0.01 | -2.02 | -0.03 |
| Figural1 | -0.23 | 0.90 | 0.62 | 0.00 | 0.88 | -0.58 | 2.58 | -1.46 | -2.75 |
| Figural2 | -0.17 | 0.78 | 0.47 | 0.88 | 0.00 | -0.42 | -5.11 | -0.24 | 0.64 |
| Figural3 | 1.01 | 0.11 | -3.89 | -0.58 | -0.42 | 0.00 | -2.56 | -2.43 | 9.13 |
| Fig2.1 | -2.15 | 1.61 | 0.01 | 2.58 | -5.11 | -2.56 | 0.00 | 1.63 | -0.46 |
| Fig2.2 | 1.49 | 0.96 | -2.02 | -1.46 | -0.24 | -2.43 | 1.63 | 0.00 | -0.67 |
| Fig2.3 | -0.89 | 0.86 | -0.03 | -2.75 | 0.64 | 9.13 | -0.46 | -0.67 | 0.00 |

12.2.3 Modifying the model to improve the fit

We see from the residuals (and Rakov and Marcoulides) that the fit is not very good and that we should allow for correlated errors for Figural3 in the junior year with Fig2.3 in the senior year. We adjust the model (and thus are no longer strictly doing a confirmatory analysis) to allow for these correlated errors.

| path | label | initial estimate |
|-----------------------------|-------|------------------|
| [1,] "Induction -> Induct1" | NA | "1" |
| [2,] "Induction -> Induct2" | "2" | NA |
| [3,] "Induction -> Induct3" | "3" | NA |
| [4,] "Figural -> Figural1" | NA | "1" |
| [5,] "Figural -> Figural2" | "5" | NA |
| [6,] "Figural -> Figural3" | "6" | NA |


```

[7,] "Figural.time2 -> Fig2.1"      NA    "1"
[8,] "Figural.time2 -> Fig2.2"      "8"   NA
[9,] "Figural.time2 -> Fig2.3"      "9"   NA
[10,] "Induction -> Figural"         "i"   NA
[11,] "Induction -> Figural.time2"  "j"   NA
[12,] "Figural -> Figural.time2"    "k"   NA
[13,] "Figural3 <-> Fig2.3"         "10"  NA
[14,] "Induct1 <-> Induct1"          "u"   NA
[15,] "Induct2 <-> Induct2"          "v"   NA
[16,] "Induct3 <-> Induct3"          "w"   NA
[17,] "Figural1 <-> Figural1"        "x"   NA
[18,] "Figural2 <-> Figural2"        "y"   NA
[19,] "Figural3 <-> Figural3"        "z"   NA
[20,] "Fig2.1 <-> Fig2.1"           "q"   NA
[21,] "Fig2.2 <-> Fig2.2"           "r"   NA
[22,] "Fig2.3 <-> Fig2.3"           "s"   NA
[23,] "Induction <-> Induction"      "A"   "1"
[24,] "Figural <-> Figural"          "B"   "1"
[25,] "Figural.time2 <-> Figural.time2" "C"   "1"

```

```

Model Chisquare = 21  Df = 23 Pr(>Chisq) = 0.61
Chisquare (null model) = 1177  Df = 36
Goodness-of-fit index = 0.98
Adjusted goodness-of-fit index = 0.96
RMSEA index = 0  90% CI: (NA, 0.049)
Bentler-Bonnett NFI = 0.98
Tucker-Lewis NNFI = 1
Bentler CFI = 1
BIC = -104

```

Normalized Residuals

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-8.3e-01 -8.4e-02  1.6e-04 -9.5e-05  1.5e-01  4.5e-01

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
2  1.27  0.159    8.0  1.1e-15 Induct2 <--- Induction
3  0.89  0.115    7.8  6.9e-15 Induct3 <--- Induction
5  0.89  0.064   13.8  0.0e+00 Figural2 <--- Figural
6  0.83  0.062   13.4  0.0e+00 Figural3 <--- Figural
8  0.87  0.051   17.2  0.0e+00 Fig2.2 <--- Figural.time2
9  0.86  0.047   18.3  0.0e+00 Fig2.3 <--- Figural.time2
i  1.00  0.150    6.7  2.6e-11 Figural <--- Induction
j  0.67  0.181    3.7  2.1e-04 Figural.time2 <--- Induction
k  0.75  0.106    7.1  1.5e-12 Figural.time2 <--- Figural
10 12.27  2.488    4.9  8.2e-07 Fig2.3 <--> Figural3
u  31.04  3.891    8.0  1.6e-15 Induct1 <--> Induct1
v  34.91  5.060    6.9  5.2e-12 Induct2 <--> Induct2

```

```

w 24.32 3.068 7.9 2.2e-15 Induct3 <--> Induct3
x 19.67 3.398 5.8 7.1e-09 Figural1 <--> Figural1
y 27.71 3.554 7.8 6.4e-15 Figural2 <--> Figural2
z 28.54 3.484 8.2 2.2e-16 Figural3 <--> Figural3
q 29.40 4.300 6.8 8.1e-12 Fig2.1 <--> Fig2.1
r 31.34 3.954 7.9 2.2e-15 Fig2.2 <--> Fig2.2
s 22.50 3.296 6.8 8.6e-12 Fig2.3 <--> Fig2.3
A 25.17 5.140 4.9 9.8e-07 Induction <--> Induction
B 39.88 6.286 6.3 2.2e-10 Figural <--> Figural
C 39.37 6.072 6.5 9.0e-11 Figural.time2 <--> Figural.time2

```

Iterations = 154

```

      Std. Estimate
1 0.66912 Induct1 <--- Induction
2 2 0.73342 Induct2 <--- Induction
3 3 0.67292 Induct3 <--- Induction
4 0.87615 Figural1 <--- Figural
5 5 0.80558 Figural2 <--- Figural
6 6 0.78263 Figural3 <--- Figural
7 0.89029 Fig2.1 <--- Figural.time2
8 8 0.85607 Fig2.2 <--- Figural.time2
9 9 0.88792 Fig2.3 <--- Figural.time2
10 i 0.62144 Figural <--- Induction
11 j 0.31746 Figural.time2 <--- Induction
12 k 0.56941 Figural.time2 <--- Figural

      Induct1 Induct2 Induct3 Figural1 Figural2 Figural3 Fig2.1 Fig2.2 Fig2.3
Induct1 0.00 -0.43 0.76 -0.65 0.20 1.71 -2.46 1.33 -0.52
Induct2 -0.43 0.00 -0.30 0.31 1.19 0.95 1.13 0.69 1.26
Induct3 0.76 -0.30 0.00 0.09 0.66 -3.40 -0.49 -2.35 0.11
Figural1 -0.65 0.31 0.09 0.00 -0.08 -0.57 2.30 -1.49 -1.81
Figural2 0.20 1.19 0.66 -0.08 0.00 1.20 -3.41 1.45 3.20
Figural3 1.71 0.95 -3.40 -0.57 1.20 0.10 -0.01 -0.01 0.10
Fig2.1 -2.46 1.13 -0.49 2.30 -3.41 -0.01 0.00 0.32 -0.11
Fig2.2 1.33 0.69 -2.35 -1.49 1.45 -0.01 0.32 0.00 -0.04
Fig2.3 -0.52 1.26 0.11 -1.81 3.20 0.10 -0.11 -0.04 0.01

```

12.2.4 Changing from a regression model to a correlation model

For theoretical reasons, the meaning of a regression model (X predicts Y or in the case of latent variables, latent X predicts latent Y) is very different than a simple correlation model. Both models fit the data equally well, but the path coefficients are very different. Compared the results from 12.2.3 with the results from a model that assumes just correlated latent variables:

```

      path          label initial estimate
[1,] "Induction -> Induct1"      NA      "1"

```

```

[2,] "Induction -> Induct2"      "2"  NA
[3,] "Induction -> Induct3"      "3"  NA
[4,] "Figural -> Figural1"      NA   "1"
[5,] "Figural -> Figural2"      "5"  NA
[6,] "Figural -> Figural3"      "6"  NA
[7,] "Figural.time2 -> Fig2.1"  NA   "1"
[8,] "Figural.time2 -> Fig2.2"  "8"  NA
[9,] "Figural.time2 -> Fig2.3"  "9"  NA
[10,] "Induction <-> Figural"    "i"  NA
[11,] "Induction <-> Figural.time2" "j"  NA
[12,] "Figural <-> Figural.time2" "k"  NA
[13,] "Figural3 <-> Fig2.3"     "10" NA
[14,] "Induct1 <-> Induct1"     "u"  NA
[15,] "Induct2 <-> Induct2"     "v"  NA
[16,] "Induct3 <-> Induct3"     "w"  NA
[17,] "Figural1 <-> Figural1"   "x"  NA
[18,] "Figural2 <-> Figural2"   "y"  NA
[19,] "Figural3 <-> Figural3"   "z"  NA
[20,] "Fig2.1 <-> Fig2.1"       "q"  NA
[21,] "Fig2.2 <-> Fig2.2"       "r"  NA
[22,] "Fig2.3 <-> Fig2.3"       "s"  NA
[23,] "Induction <-> Induction"  "A"  "1"
[24,] "Figural <-> Figural"     "B"  "1"
[25,] "Figural.time2 <-> Figural.time2" "C"  "1"

```

```

Model Chisquare = 21 Df = 23 Pr(>Chisq) = 0.61
Chisquare (null model) = 1177 Df = 36
Goodness-of-fit index = 0.98
Adjusted goodness-of-fit index = 0.96
RMSEA index = 0 90% CI: (NA, 0.049)
Bentler-Bonnett NFI = 0.98
Tucker-Lewis NNFI = 1
Bentler CFI = 1
BIC = -104

```

Normalized Residuals

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-8.3e-01 -8.4e-02  2.3e-04 -3.9e-05  1.6e-01  4.5e-01

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
2    1.27      0.159     8.0    1.3e-15 Induct2 <--- Induction
3    0.89      0.115     7.8    6.9e-15 Induct3 <--- Induction
5    0.89      0.064    13.8    0.0e+00 Figural2 <--- Figural
6    0.83      0.062    13.4    0.0e+00 Figural3 <--- Figural
8    0.87      0.051    17.2    0.0e+00 Fig2.2 <--- Figural.time2
9    0.86      0.047    18.3    0.0e+00 Fig2.3 <--- Figural.time2
i   25.13     4.361     5.8    8.3e-09 Figural <--> Induction

```

| | | | | | |
|----|--------|--------|-----|---------|----------------------------------|
| j | 35.70 | 5.801 | 6.2 | 7.6e-10 | Figural.time2 <--> Induction |
| k | 65.51 | 8.340 | 7.9 | 4.0e-15 | Figural.time2 <--> Figural |
| 10 | 12.26 | 2.488 | 4.9 | 8.2e-07 | Fig2.3 <--> Figural3 |
| u | 31.04 | 3.891 | 8.0 | 1.6e-15 | Induct1 <--> Induct1 |
| v | 34.91 | 5.060 | 6.9 | 5.2e-12 | Induct2 <--> Induct2 |
| w | 24.32 | 3.068 | 7.9 | 2.2e-15 | Induct3 <--> Induct3 |
| x | 19.67 | 3.398 | 5.8 | 7.1e-09 | Figural1 <--> Figural1 |
| y | 27.71 | 3.555 | 7.8 | 6.4e-15 | Figural2 <--> Figural2 |
| z | 28.54 | 3.483 | 8.2 | 2.2e-16 | Figural3 <--> Figural3 |
| q | 29.40 | 4.300 | 6.8 | 8.1e-12 | Fig2.1 <--> Fig2.1 |
| r | 31.34 | 3.954 | 7.9 | 2.2e-15 | Fig2.2 <--> Fig2.2 |
| s | 22.50 | 3.295 | 6.8 | 8.6e-12 | Fig2.3 <--> Fig2.3 |
| A | 25.16 | 5.143 | 4.9 | 9.9e-07 | Induction <--> Induction |
| B | 64.97 | 8.369 | 7.8 | 8.2e-15 | Figural <--> Figural |
| C | 112.37 | 13.662 | 8.2 | 2.2e-16 | Figural.time2 <--> Figural.time2 |

Iterations = 215

| | Std. Estimate | |
|---|---------------|---------------------------|
| 1 | 0.66911 | Induct1 <--- Induction |
| 2 | 0.73340 | Induct2 <--- Induction |
| 3 | 0.67292 | Induct3 <--- Induction |
| 4 | 0.87615 | Figural1 <--- Figural |
| 5 | 0.80555 | Figural2 <--- Figural |
| 6 | 0.78266 | Figural3 <--- Figural |
| 7 | 0.89029 | Fig2.1 <--- Figural.time2 |
| 8 | 0.85608 | Fig2.2 <--- Figural.time2 |
| 9 | 0.88793 | Fig2.3 <--- Figural.time2 |

| | Induct1 | Induct2 | Induct3 | Figural1 | Figural2 | Figural3 | Fig2.1 | Fig2.2 | Fig2.3 |
|----------|---------|---------|---------|----------|----------|----------|--------|--------|--------|
| Induct1 | 0.00 | -0.43 | 0.76 | -0.65 | 0.20 | 1.71 | -2.46 | 1.33 | -0.52 |
| Induct2 | -0.43 | 0.00 | -0.30 | 0.31 | 1.19 | 0.95 | 1.13 | 0.69 | 1.26 |
| Induct3 | 0.76 | -0.30 | 0.00 | 0.09 | 0.66 | -3.40 | -0.49 | -2.35 | 0.11 |
| Figural1 | -0.65 | 0.31 | 0.09 | 0.00 | -0.08 | -0.57 | 2.30 | -1.49 | -1.81 |
| Figural2 | 0.20 | 1.19 | 0.66 | -0.08 | 0.00 | 1.20 | -3.40 | 1.45 | 3.20 |
| Figural3 | 1.71 | 0.95 | -3.40 | -0.57 | 1.20 | 0.10 | -0.01 | -0.02 | 0.10 |
| Fig2.1 | -2.46 | 1.13 | -0.49 | 2.30 | -3.40 | -0.01 | 0.00 | 0.32 | -0.11 |
| Fig2.2 | 1.33 | 0.69 | -2.35 | -1.49 | 1.45 | -0.02 | 0.32 | 0.00 | -0.04 |
| Fig2.3 | -0.52 | 1.26 | 0.11 | -1.81 | 3.20 | 0.10 | -0.11 | -0.04 | 0.01 |

Note that the coefficients i,j, and k are now covariances rather than beta weights.

12.3 Using LISREL to analyze the data set

The commercial computer package LISREL, developed by Karl Joreskog, was the first commercial program to do Linear Structural RELations. Although seemingly complicated than other packages, LISREL uses a matrix formulation that clearly shows the difference between

observed and latent variables, the errors associated with each, and distinguishes between the predictor set of variables and the criterion set of variables.

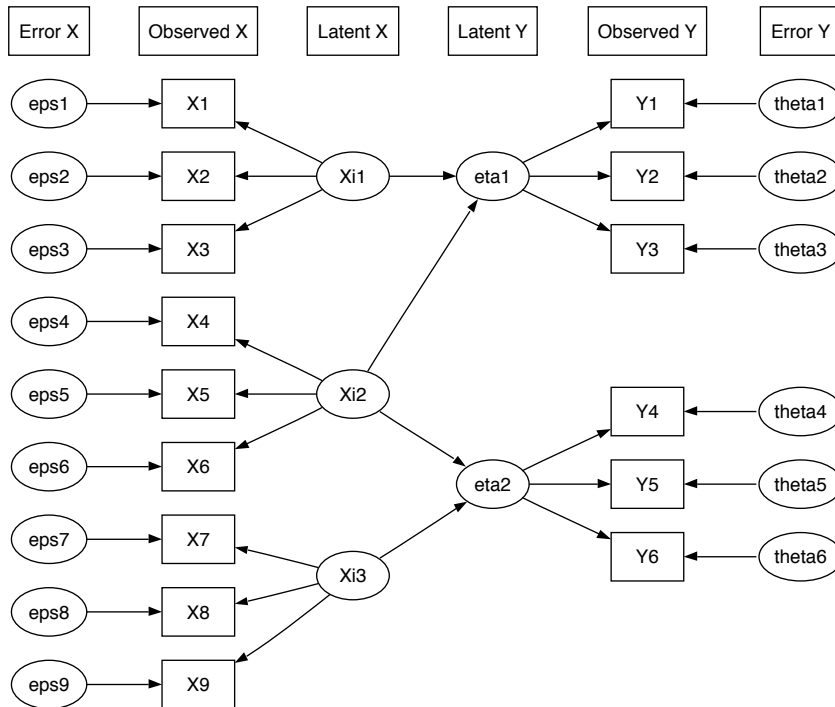


Fig. 12.2 The Linear Structural Relations (LISREL) model integrates two measurement models with one regression model. How well are the X's represented by the latent variables (factors) Xi, and how well are the Y variables represented by the factors etas.

The matrices are:

1. X variables (the observed variables)
2. Lambda X (LX:the factor loadings for the X variables on the eta factors)
3. Beta (BE:the beta weights linking the eta to the psi latent variables)
4. Lambda Y (LY: the factor loadings for the Y variables on the psi factors)
5. Psi (PS: the dependent latent factor variances and covariances)
6. Theta and Epsilon (TE: the error variances and covariances for the X and Y variables).

LISREL is available for PCs as an add on to SPSS, but is also available as a stand alone package at the Northwestern Social Science Computing Cluster. To use LISREL at the SSCC it is necessary to have [an account](#) and then to log in as a remote user.

12.3.1 Instructions for using the SSCC

1. Log on to the system using SSH (see the “how to” for doing this)
2. upload the appropriate batch command file using a [sftp connection](#).
The file we will submit is taken from Raykov and Marcoulides (2006):

```

STRUCTURAL REGRESSION MODEL
DA NI=9 NO=220
CM
56.21
31.55 75.55
23.27 28.30 44.45
24.48 32.24 22.56 84.64
22.51 29.54 20.61 57.61 78.93
22.65 27.56 15.33 53.57 49.27 73.76
33.24 46.49 31.44 67.81 54.76 54.58 141.77
32.56 40.37 25.58 55.82 52.33 47.74 98.62 117.33
30.32 40.44 27.69 54.78 53.44 59.52 96.95 84.87 106.35
LA
IND1 IND2 IND3 FR11 FR12 FR13 FR21 FR22 FR23
MO NY=9 NE=3 PS=SY,FI TE=DI,FR LY=FU,FI BE=FU,FI
LE
INDUCTN FIGREL1 FIGREL2
FR LY(2, 1) LY(3, 1)
FR LY(5, 2) LY(6, 2)
FR LY(8, 3) LY(9, 3)
VA 1 LY(1, 1) LY(4, 2) LY(7, 3)
FR BE(2, 1) BE(3, 1) BE(3, 2)
FR PS(1, 1) PS(2, 2) PS(3, 3)
OU

```

This file is created (or in this case copied) and saved on the Mac/PC with a meaningful name, rm5.txt, and then uploaded to the SSCC using a sftp operation. (From my Mac I use Interarchy as my sftp client.)

3. submit the lisrel job by invoking lisrel8:

```
[revelle@hardin ~]$ lisrel8 rm5.txt rm5.out
```

```

+-----+
|
|           L I S R E L  8.72
|
|           by
|
| Karl G. Joreskog and Dag Sorbom
| Available Workspace 16941056 bytes|
+-----+

```

This program is published exclusively by
Scientific Software International, Inc.

7383 N.Lincoln Avenue - Suite 100
 Lincolnwood, IL 60712-1704, U.S.A.
 Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
 Copyright by Scientific Software International, Inc., 1981-2005
 Use of this program is subject to the terms specified in the
 Universal Copyright Convention.
 Website: www.ssicentral.com
 Revision LISREL872_03/28/2005
 Input file [INPUT] :
 Input file [INPUT] :
 rm5.txt
 Output file [OUTPUT]:
 Output file [OUTPUT]:
 rm5.out
 Reading input from file rm5.txt

STRUCTURAL REGRESSION MODEL
 Computing Initial Estimates
 Computing Information Matrix
 Inverting Information Matrix
 Iteration 1 for LISREL Estimates
 Iteration 2 for LISREL Estimates
 Iteration 3 for LISREL Estimates
 Iteration 4 for LISREL Estimates
 Iteration 5 for LISREL Estimates
 Iteration 6 for LISREL Estimates
 Computing Information Matrix
 Inverting Information Matrix
 Computing Goodness of Fit Statistics

[revelle@hardin ~]\$

4. Transfer the output file (in this case "rm5.out") back to your host machine (using sftp).
5. Examine the output

DATE: 2/12/2007
 TIME: 11:14

L I S R E L 8.72

BY

Karl G. Jöreskog & Dag Sörbom

This program is published exclusively by
 Scientific Software International, Inc.

7383 N. Lincoln Avenue, Suite 100
 Lincolnwood, IL 60712, U.S.A.
 Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
 Copyright by Scientific Software International, Inc., 1981-2005
 Use of this program is subject to the terms specified in the
 Universal Copyright Convention.
 Website: www.ssicentral.com

The following lines were read from file rm5.txt:

```
STRUCTURAL REGRESSION MODEL
DA NI=9 NO=220
CM
56.21
31.55 75.55
23.27 28.30 44.45
24.48 32.24 22.56 84.64
22.51 29.54 20.61 57.61 78.93
22.65 27.56 15.33 53.57 49.27 73.76
33.24 46.49 31.44 67.81 54.76 54.58 141.77
32.56 40.37 25.58 55.82 52.33 47.74 98.62 117.33
30.32 40.44 27.69 54.78 53.44 59.52 96.95 84.87 106.35
LA
IND1 IND2 IND3 FR11 FR12 FR13 FR21 FR22 FR23
MO NY=9 NE=3 PS=SY,FI TE=DI,FR LY=FU,FI BE=FU,FI
LE
INDUCTN FIGREL1 FIGREL2
FR LY(2, 1) LY(3, 1)
FR LY(5, 2) LY(6, 2)
FR LY(8, 3) LY(9, 3)
VA 1 LY(1, 1) LY(4, 2) LY(7, 3)
FR BE(2, 1) BE(3, 1) BE(3, 2)
FR PS(1, 1) PS(2, 2) PS(3, 3)
OU
```

STRUCTURAL REGRESSION MODEL

```
Number of Input Variables 9
Number of Y - Variables 9
Number of X - Variables 0
Number of ETA - Variables 3
Number of KSI - Variables 0
Number of Observations 220
```

STRUCTURAL REGRESSION MODEL

Covariance Matrix

| | IND1 | IND2 | IND3 | FR11 | FR12 | FR13 |
|------|-------|-------|-------|-------|-------|-------|
| | ----- | ----- | ----- | ----- | ----- | ----- |
| IND1 | 56.21 | | | | | |
| IND2 | 31.55 | 75.55 | | | | |
| IND3 | 23.27 | 28.30 | 44.45 | | | |
| FR11 | 24.48 | 32.24 | 22.56 | 84.64 | | |
| FR12 | 22.51 | 29.54 | 20.61 | 57.61 | 78.93 | |
| FR13 | 22.65 | 27.56 | 15.33 | 53.57 | 49.27 | 73.76 |
| FR21 | 33.24 | 46.49 | 31.44 | 67.81 | 54.76 | 54.58 |
| FR22 | 32.56 | 40.37 | 25.58 | 55.82 | 52.33 | 47.74 |
| FR23 | 30.32 | 40.44 | 27.69 | 54.78 | 53.44 | 59.52 |

Covariance Matrix

| | FR21 | FR22 | FR23 |
|------|--------|--------|--------|
| | ----- | ----- | ----- |
| FR21 | 141.77 | | |
| FR22 | 98.62 | 117.33 | |
| FR23 | 96.95 | 84.87 | 106.35 |

STRUCTURAL REGRESSION MODEL

Parameter Specifications

LAMBDA-Y

| | INDUCTN | FIGREL1 | FIGREL2 |
|------|---------|---------|---------|
| | ----- | ----- | ----- |
| IND1 | 0 | 0 | 0 |
| IND2 | 1 | 0 | 0 |
| IND3 | 2 | 0 | 0 |
| FR11 | 0 | 0 | 0 |
| FR12 | 0 | 3 | 0 |
| FR13 | 0 | 4 | 0 |
| FR21 | 0 | 0 | 0 |
| FR22 | 0 | 0 | 5 |
| FR23 | 0 | 0 | 6 |

BETA

| | INDUCTN | FIGREL1 | FIGREL2 |
|---------|---------|---------|---------|
| | ----- | ----- | ----- |
| INDUCTN | 0 | 0 | 0 |
| FIGREL1 | 7 | 0 | 0 |
| FIGREL2 | 8 | 9 | 0 |

PSI

| INDUCTN | FIGREL1 | FIGREL2 | | | |
|-----------|---------|---------|-------|-------|-------|
| ----- | ----- | ----- | | | |
| 10 | 11 | 12 | | | |
| THETA-EPS | | | | | |
| IND1 | IND2 | IND3 | FR11 | FR12 | FR13 |
| ----- | ----- | ----- | ----- | ----- | ----- |
| 13 | 14 | 15 | 16 | 17 | 18 |
| THETA-EPS | | | | | |
| FR21 | FR22 | FR23 | | | |
| ----- | ----- | ----- | | | |
| 19 | 20 | 21 | | | |

STRUCTURAL REGRESSION MODEL

Number of Iterations = 5

LISREL Estimates (Maximum Likelihood)

| LAMBDA-Y | | | |
|----------|------------------------|-------------------------|---------|
| | INDUCTN | FIGREL1 | FIGREL2 |
| | ----- | ----- | ----- |
| IND1 | 1.00 | - - | - - |
| IND2 | 1.27 (0.16) 8.08 | - - | - - |
| IND3 | 0.89 (0.12) 7.70 | - - | - - |
| FR11 | - - | 1.00 | - - |
| FR12 | - - | 0.92 (0.07) 13.76 | - - |
| FR13 | - - | 0.88 (0.06) 13.54 | - - |

| | | | |
|------|-----|-----|-------------------------|
| FR21 | - - | - - | 1.00 |
| FR22 | - - | - - | 0.88 (0.05) 16.79 |
| FR23 | - - | - - | 0.88 (0.05) 18.39 |

BETA

| | INDUCTN | FIGREL1 | FIGREL2 |
|---------|------------------------|------------------------|---------|
| | ----- | ----- | ----- |
| INDUCTN | - - | - - | - - |
| FIGREL1 | 0.98 (0.15) 6.64 | - - | - - |
| FIGREL2 | 0.60 (0.18) 3.41 | 0.81 (0.11) 7.40 | - - |

Covariance Matrix of ETA

| | INDUCTN | FIGREL1 | FIGREL2 |
|---------|---------|---------|---------|
| | ----- | ----- | ----- |
| INDUCTN | 25.31 | | |
| FIGREL1 | 24.71 | 61.81 | |
| FIGREL2 | 35.39 | 65.23 | 110.46 |

PSI

Note: This matrix is diagonal.

| | INDUCTN | FIGREL1 | FIGREL2 |
|--|-------------------------|-------------------------|-------------------------|
| | ----- | ----- | ----- |
| | 25.31 (5.14) 4.92 | 37.69 (6.10) 6.18 | 36.00 (5.92) 6.08 |

Squared Multiple Correlations for Structural Equations

| INDUCTN | FIGREL1 | FIGREL2 |
|---------|---------|---------|
|---------|---------|---------|

```

-----
- -      0.39      0.67

```

THETA-EPS

```

      IND1      IND2      IND3      FR11      FR12      FR13
-----
30.90      34.84      24.49      22.83      26.87      26.33
(3.88)      (5.06)      (3.07)      (3.42)      (3.47)      (3.31)
7.97      6.89      7.98      6.67      7.75      7.95

```

THETA-EPS

```

      FR21      FR22      FR23
-----
31.31      32.17      20.44
(4.40)      (4.02)      (3.15)
7.12      7.99      6.50

```

Squared Multiple Correlations for Y - Variables

```

      IND1      IND2      IND3      FR11      FR12      FR13
-----
0.45      0.54      0.45      0.73      0.66      0.64

```

Squared Multiple Correlations for Y - Variables

```

      FR21      FR22      FR23
-----
0.78      0.73      0.81

```

Goodness of Fit Statistics

Degrees of Freedom = 24

Minimum Fit Function Chi-Square = 52.10 (P = 0.00076)

Normal Theory Weighted Least Squares Chi-Square = 48.28 (P = 0.0023)

Estimated Non-centrality Parameter (NCP) = 24.28

90 Percent Confidence Interval for NCP = (8.23 ; 48.09)

Minimum Fit Function Value = 0.24

Population Discrepancy Function Value (F0) = 0.11

90 Percent Confidence Interval for F0 = (0.038 ; 0.22)

Root Mean Square Error of Approximation (RMSEA) = 0.068

90 Percent Confidence Interval for RMSEA = (0.040 ; 0.096)

P-Value for Test of Close Fit (RMSEA < 0.05) = 0.14

Expected Cross-Validation Index (ECVI) = 0.41
 90 Percent Confidence Interval for ECVI = (0.34 ; 0.52)
 ECVI for Saturated Model = 0.41
 ECVI for Independence Model = 9.49

Chi-Square for Independence Model with 36 Degrees of Freedom = 2060.02
 Independence AIC = 2078.02
 Model AIC = 90.28
 Saturated AIC = 90.00
 Independence CAIC = 2117.56
 Model CAIC = 182.54
 Saturated CAIC = 287.71

Normed Fit Index (NFI) = 0.97
 Non-Normed Fit Index (NNFI) = 0.98
 Parsimony Normed Fit Index (PNFI) = 0.65
 Comparative Fit Index (CFI) = 0.99
 Incremental Fit Index (IFI) = 0.99
 Relative Fit Index (RFI) = 0.96

Critical N (CN) = 181.68

Root Mean Square Residual (RMR) = 1.99
 Standardized RMR = 0.023
 Goodness of Fit Index (GFI) = 0.95
 Adjusted Goodness of Fit Index (AGFI) = 0.91
 Parsimony Goodness of Fit Index (PGFI) = 0.51

12.3.2 Modify the model to allow for correlated errors

Just as we did for the sem using R, an examination of the residuals suggests that we need to modify the model to allow for correlated errors for the Figural3 at time 1 and time 2. This leads to the following LISREL commands:

```
STRUCTURAL REGRESSION MODEL
DA NI=9 NO=220
CM
56.21
31.55 75.55
23.27 28.30 44.45
24.48 32.24 22.56 84.64
22.51 29.54 20.61 57.61 78.93
22.65 27.56 15.33 53.57 49.27 73.76
```

```

33.24 46.49 31.44 67.81 54.76 54.58 141.77
32.56 40.37 25.58 55.82 52.33 47.74 98.62 117.33
30.32 40.44 27.69 54.78 53.44 59.52 96.95 84.87 106.35
LA
IND1 IND2 IND3 FR11 FR12 FR13 FR21 FR22 FR23
MO NY=9 NE=3 PS=SY,FI TE=SY,FI LY=FU,FI BE=FU,FI
LE
INDUCTN FIGREL1 FIGREL2
FR LY(2, 1) LY(3, 1)
FR LY(5, 2) LY(6, 2)
FR LY(8, 3) LY(9, 3)
VA 1 LY(1, 1) LY(4, 2) LY(7, 3)
FR BE(2, 1) BE(3, 1) BE(3, 2)
FR PS(1, 1) PS(2, 2) PS(3, 3)
FR TE(1,1) TE (2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6) TE(7,7) TE(8,8) TE(9,9) TE(9,6)
OU

```

Compare this set of commands to the previous set. What we have done is added a line to specify the errors in the “theta” matrix and specified that the 6th error correlates with the 9th error.

Uploading this revised command file to the SSCC and running it leads to the following output:

```

DATE: 2/12/2007
TIME: 11:37

```

L I S R E L 8.72

BY

Karl G. Jöreskog & Dag Sörbom

This program is published exclusively by
 Scientific Software International, Inc.
 7383 N. Lincoln Avenue, Suite 100
 Lincolnwood, IL 60712, U.S.A.
 Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
 Copyright by Scientific Software International, Inc., 1981-2005
 Use of this program is subject to the terms specified in the
 Universal Copyright Convention.
 Website: www.ssicentral.com

The following lines were read from file rm5a.txt:

STRUCTURAL REGRESSION MODEL

```

DA NI=9 NO=220
CM
56.21
31.55 75.55
23.27 28.30 44.45
24.48 32.24 22.56 84.64
22.51 29.54 20.61 57.61 78.93
22.65 27.56 15.33 53.57 49.27 73.76
33.24 46.49 31.44 67.81 54.76 54.58 141.77
32.56 40.37 25.58 55.82 52.33 47.74 98.62 117.33
30.32 40.44 27.69 54.78 53.44 59.52 96.95 84.87 106.35
LA
IND1 IND2 IND3 FR11 FR12 FR13 FR21 FR22 FR23
MO NY=9 NE=3 PS=SY,FI TE=SY,FI LY=FU,FI BE=FU,FI
LE
INDUCTN FIGREL1 FIGREL2
FR LY(2, 1) LY(3, 1)
FR LY(5, 2) LY(6, 2)
FR LY(8, 3) LY(9, 3)
VA 1 LY(1, 1) LY(4, 2) LY(7, 3)
FR BE(2, 1) BE(3, 1) BE(3, 2)
FR PS(1, 1) PS(2, 2) PS(3, 3)
FR TE(1,1) TE (2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6) TE(7,7) TE(8,8) TE(9,9) TE(9,6)
OU

```

STRUCTURAL REGRESSION MODEL

```

Number of Input Variables 9
Number of Y - Variables 9
Number of X - Variables 0
Number of ETA - Variables 3
Number of KSI - Variables 0
Number of Observations 220

```

STRUCTURAL REGRESSION MODEL

Covariance Matrix

| | IND1 | IND2 | IND3 | FR11 | FR12 | FR13 |
|------|-------|-------|-------|-------|-------|-------|
| IND1 | 56.21 | | | | | |
| IND2 | 31.55 | 75.55 | | | | |
| IND3 | 23.27 | 28.30 | 44.45 | | | |
| FR11 | 24.48 | 32.24 | 22.56 | 84.64 | | |
| FR12 | 22.51 | 29.54 | 20.61 | 57.61 | 78.93 | |
| FR13 | 22.65 | 27.56 | 15.33 | 53.57 | 49.27 | 73.76 |
| FR21 | 33.24 | 46.49 | 31.44 | 67.81 | 54.76 | 54.58 |
| FR22 | 32.56 | 40.37 | 25.58 | 55.82 | 52.33 | 47.74 |

| | | | | | | |
|------|-------|-------|-------|-------|-------|-------|
| FR23 | 30.32 | 40.44 | 27.69 | 54.78 | 53.44 | 59.52 |
|------|-------|-------|-------|-------|-------|-------|

Covariance Matrix

| | FR21 | FR22 | FR23 |
|------|--------|--------|--------|
| FR21 | 141.77 | | |
| FR22 | 98.62 | 117.33 | |
| FR23 | 96.95 | 84.87 | 106.35 |

STRUCTURAL REGRESSION MODEL

Parameter Specifications

LAMBDA-Y

| | INDUCTN | FIGREL1 | FIGREL2 |
|------|---------|---------|---------|
| IND1 | 0 | 0 | 0 |
| IND2 | 1 | 0 | 0 |
| IND3 | 2 | 0 | 0 |
| FR11 | 0 | 0 | 0 |
| FR12 | 0 | 3 | 0 |
| FR13 | 0 | 4 | 0 |
| FR21 | 0 | 0 | 0 |
| FR22 | 0 | 0 | 5 |
| FR23 | 0 | 0 | 6 |

BETA

| | INDUCTN | FIGREL1 | FIGREL2 |
|---------|---------|---------|---------|
| INDUCTN | 0 | 0 | 0 |
| FIGREL1 | 7 | 0 | 0 |
| FIGREL2 | 8 | 9 | 0 |

PSI

| | INDUCTN | FIGREL1 | FIGREL2 |
|--|---------|---------|---------|
| | 10 | 11 | 12 |

THETA-EPS

| | IND1 | IND2 | IND3 | FR11 | FR12 | FR13 |
|------|------|------|------|------|------|------|
| IND1 | 13 | | | | | |

| | | | | | | |
|------|---|----|----|----|----|----|
| IND2 | 0 | 14 | | | | |
| IND3 | 0 | 0 | 15 | | | |
| FR11 | 0 | 0 | 0 | 16 | | |
| FR12 | 0 | 0 | 0 | 0 | 17 | |
| FR13 | 0 | 0 | 0 | 0 | 0 | 18 |
| FR21 | 0 | 0 | 0 | 0 | 0 | 0 |
| FR22 | 0 | 0 | 0 | 0 | 0 | 0 |
| FR23 | 0 | 0 | 0 | 0 | 0 | 21 |

THETA-EPS

| | FR21 | FR22 | FR23 |
|------|-------|-------|-------|
| | ----- | ----- | ----- |
| FR21 | 19 | | |
| FR22 | 0 | 20 | |
| FR23 | 0 | 0 | 22 |

STRUCTURAL REGRESSION MODEL

Number of Iterations = 5

LISREL Estimates (Maximum Likelihood)

LAMBDA-Y

| | INDUCTN | FIGREL1 | FIGREL2 |
|------|------------------------|-------------------------|---------|
| | ----- | ----- | ----- |
| IND1 | 1.00 | - - | - - |
| IND2 | 1.27 (0.16) 8.07 | - - | - - |
| IND3 | 0.89 (0.12) 7.71 | - - | - - |
| FR11 | - - | 1.00 | - - |
| FR12 | - - | 0.89 (0.06) 13.89 | - - |
| FR13 | - - | 0.83 (0.06) 13.46 | - - |

| | | | |
|------|-----|-----|-------------------------|
| FR21 | - - | - - | 1.00 |
| FR22 | - - | - - | 0.87 (0.05) 17.20 |
| FR23 | - - | - - | 0.86 (0.05) 18.39 |

BETA

| | INDUCTN | FIGREL1 | FIGREL2 |
|---------|------------------------|------------------------|---------|
| | ----- | ----- | ----- |
| INDUCTN | - - | - - | - - |
| FIGREL1 | 1.00 (0.15) 6.68 | - - | - - |
| FIGREL2 | 0.67 (0.18) 3.74 | 0.75 (0.11) 7.10 | - - |

Covariance Matrix of ETA

| | INDUCTN | FIGREL1 | FIGREL2 |
|---------|---------|---------|---------|
| | ----- | ----- | ----- |
| INDUCTN | 25.17 | | |
| FIGREL1 | 25.13 | 64.97 | |
| FIGREL2 | 35.70 | 65.51 | 112.37 |

PSI

Note: This matrix is diagonal.

| | INDUCTN | FIGREL1 | FIGREL2 |
|--|-------------------------|-------------------------|-------------------------|
| | ----- | ----- | ----- |
| | 25.17 (5.13) 4.91 | 39.88 (6.26) 6.37 | 39.37 (6.05) 6.51 |

Squared Multiple Correlations for Structural Equations

| INDUCTN | FIGREL1 | FIGREL2 |
|---------|---------|---------|
|---------|---------|---------|

| | ----- | ----- | ----- | | | |
|-----------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | - - | 0.39 | 0.65 | | | |
| THETA-EPS | | | | | | |
| | IND1 | IND2 | IND3 | FR11 | FR12 | FR13 |
| | ----- | ----- | ----- | ----- | ----- | ----- |
| IND1 | 31.04 (3.88) 8.00 | | | | | |
| IND2 | - - | 34.91 (5.05) 6.91 | | | | |
| IND3 | - - | - - | 24.32 (3.06) 7.95 | | | |
| FR11 | - - | - - | - - | 19.67 (3.35) 5.87 | | |
| FR12 | - - | - - | - - | - - | 27.71 (3.53) 7.84 | |
| FR13 | - - | - - | - - | - - | - - | 28.54 (3.46) 8.24 |
| FR21 | - - | - - | - - | - - | - - | - - |
| FR22 | - - | - - | - - | - - | - - | - - |
| FR23 | - - | - - | - - | - - | - - | 12.26 (2.46) 4.99 |

| | | | |
|-----------|-------------------------|-------|-------|
| THETA-EPS | | | |
| | FR21 | FR22 | FR23 |
| | ----- | ----- | ----- |
| FR21 | 29.40 (4.28) 6.87 | | |

| | | |
|------|-----|-----------------------------|
| FR22 | - - | 31.34 (3.95) 7.93 |
| FR23 | - - | - - 22.50 (3.28) 6.86 |

Squared Multiple Correlations for Y - Variables

| IND1 | IND2 | IND3 | FR11 | FR12 | FR13 |
|-------|-------|-------|-------|-------|-------|
| ----- | ----- | ----- | ----- | ----- | ----- |
| 0.45 | 0.54 | 0.45 | 0.77 | 0.65 | 0.61 |

Squared Multiple Correlations for Y - Variables

| FR21 | FR22 | FR23 |
|-------|-------|-------|
| ----- | ----- | ----- |
| 0.79 | 0.73 | 0.79 |

Goodness of Fit Statistics

Degrees of Freedom = 23
 Minimum Fit Function Chi-Square = 20.55 (P = 0.61)
 Normal Theory Weighted Least Squares Chi-Square = 20.01 (P = 0.64)
 Estimated Non-centrality Parameter (NCP) = 0.0
 90 Percent Confidence Interval for NCP = (0.0 ; 11.11)

Minimum Fit Function Value = 0.094
 Population Discrepancy Function Value (F0) = 0.0
 90 Percent Confidence Interval for F0 = (0.0 ; 0.051)
 Root Mean Square Error of Approximation (RMSEA) = 0.0
 90 Percent Confidence Interval for RMSEA = (0.0 ; 0.047)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.96

Expected Cross-Validation Index (ECVI) = 0.31
 90 Percent Confidence Interval for ECVI = (0.31 ; 0.36)
 ECVI for Saturated Model = 0.41
 ECVI for Independence Model = 9.49

Chi-Square for Independence Model with 36 Degrees of Freedom = 2060.02
 Independence AIC = 2078.02
 Model AIC = 64.01
 Saturated AIC = 90.00
 Independence CAIC = 2117.56
 Model CAIC = 160.67

Saturated CAIC = 287.71

Normed Fit Index (NFI) = 0.99
 Non-Normed Fit Index (NNFI) = 1.00
 Parsimony Normed Fit Index (PNFI) = 0.63
 Comparative Fit Index (CFI) = 1.00
 Incremental Fit Index (IFI) = 1.00
 Relative Fit Index (RFI) = 0.98

Critical N (CN) = 444.69

Root Mean Square Residual (RMR) = 1.27
 Standardized RMR = 0.016
 Goodness of Fit Index (GFI) = 0.98
 Adjusted Goodness of Fit Index (AGFI) = 0.96
 Parsimony Goodness of Fit Index (PGFI) = 0.50

12.4 Comparing the R and LISREL output

Each sem author has his or her own preferences about how to organize the output. Compare the LISREL output [12.3.2](#) with the R output for the prediction model [12.2.3](#) and the correlation model [12.2.4](#).

As one would hope, the chi square values and df are equal between the two programs. LISREL gives far more goodness of fit statistics and also has a more detailed output than sem.

12.5 Testing for factorial invariance

The models tested above measured Figural Relations in the Junior and Senior year. Were these tests measuring the same concept? If they were, then we would expect the factor loadings to be the same in both years. We can test this by constraining the equivalent loadings to be identical and comparing the differences in χ^2 for the two models. (The first model is discussed in section [12.2.3](#))

| | path | label | initial | estimate |
|------|---------------------------|-------|---------|----------|
| [1,] | "Induction -> Induct1" | NA | "1" | |
| [2,] | "Induction -> Induct2" | "2" | NA | |
| [3,] | "Induction -> Induct3" | "3" | NA | |
| [4,] | "Figural -> Figural1" | NA | "1" | |
| [5,] | "Figural -> Figural2" | "5" | NA | |
| [6,] | "Figural -> Figural3" | "6" | NA | |
| [7,] | "Figural.time2 -> Fig2.1" | NA | "1" | |

```

[8,] "Figural.time2 -> Fig2.2"      "5"  NA
[9,] "Figural.time2 -> Fig2.3"      "6"  NA
[10,] "Induction -> Figural"         "i"  NA
[11,] "Induction -> Figural.time2"  "j"  NA
[12,] "Figural -> Figural.time2"    "k"  NA
[13,] "Figural3 <-> Fig2.3"         "10" NA
[14,] "Induct1 <-> Induct1"         "u"  NA
[15,] "Induct2 <-> Induct2"         "v"  NA
[16,] "Induct3 <-> Induct3"         "w"  NA
[17,] "Figural1 <-> Figural1"       "x"  NA
[18,] "Figural2 <-> Figural2"       "y"  NA
[19,] "Figural3 <-> Figural3"       "z"  NA
[20,] "Fig2.1 <-> Fig2.1"           "q"  NA
[21,] "Fig2.2 <-> Fig2.2"           "r"  NA
[22,] "Fig2.3 <-> Fig2.3"           "s"  NA
[23,] "Induction <-> Induction"     "A"  "1"
[24,] "Figural <-> Figural"         "B"  "1"
[25,] "Figural.time2 <-> Figural.time2" "C"  "1"

```

```

Model Chisquare = 21 Df = 25 Pr(>Chisq) = 0.7
Chisquare (null model) = 1177 Df = 36
Goodness-of-fit index = 0.98
Adjusted goodness-of-fit index = 0.96
RMSEA index = 0 90% CI: (NA, 0.043)
Bentler-Bonnett NFI = 0.98
Tucker-Lewis NNFI = 1
Bentler CFI = 1
BIC = -114

```

Normalized Residuals

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-9.1e-01 -1.1e-01  4.9e-05 -1.1e-02  1.7e-01  6.0e-01

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
2  1.27      0.159      8.0    1.1e-15 Induct2 <--- Induction
3  0.89      0.115      7.8    6.9e-15 Induct3 <--- Induction
5  0.88      0.040     21.8    0.0e+00 Figural2 <--- Figural
6  0.86      0.042     20.5    0.0e+00 Figural3 <--- Figural
i  0.99      0.147      6.8    1.3e-11 Figural <--- Induction
j  0.67      0.181      3.7    2.0e-04 Figural.time2 <--- Induction
k  0.76      0.100      7.6    2.7e-14 Figural.time2 <--- Figural
10 12.23     2.481      4.9    8.3e-07 Fig2.3 <--> Figural3
u 31.03     3.891      8.0    1.6e-15 Induct1 <--> Induct1
v 34.90     5.060      6.9    5.3e-12 Induct2 <--> Induct2
w 24.34     3.069      7.9    2.2e-15 Induct3 <--> Induct3
x 19.85     3.234      6.1    8.3e-10 Figural1 <--> Figural1
y 28.00     3.461      8.1    6.7e-16 Figural2 <--> Figural2

```

| | | | | | |
|---|-------|-------|-----|---------|----------------------------------|
| z | 28.19 | 3.417 | 8.2 | 2.2e-16 | Figural3 <--> Figural3 |
| q | 29.21 | 4.243 | 6.9 | 5.9e-12 | Fig2.1 <--> Fig2.1 |
| r | 31.14 | 3.917 | 8.0 | 1.8e-15 | Fig2.2 <--> Fig2.2 |
| s | 22.74 | 3.275 | 6.9 | 3.8e-12 | Fig2.3 <--> Fig2.3 |
| A | 25.18 | 5.141 | 4.9 | 9.7e-07 | Induction <--> Induction |
| B | 39.43 | 5.833 | 6.8 | 1.4e-11 | Figural <--> Figural |
| C | 39.62 | 5.962 | 6.6 | 3.0e-11 | Figural.time2 <--> Figural.time2 |

Iterations = 162

| | Std. | Estimate | | | | | | | | |
|----|-----------|----------|---------------|------|---------------|--|--|--|--|--|
| 1 | 0.66928 | | Induct1 | <--- | Induction | | | | | |
| 2 | 2 0.73353 | | Induct2 | <--- | Induction | | | | | |
| 3 | 3 0.67263 | | Induct3 | <--- | Induction | | | | | |
| 4 | 0.87392 | | Figural1 | <--- | Figural | | | | | |
| 5 | 5 0.79947 | | Figural2 | <--- | Figural | | | | | |
| 6 | 6 0.79051 | | Figural3 | <--- | Figural | | | | | |
| 7 | 0.89178 | | Fig2.1 | <--- | Figural.time2 | | | | | |
| 8 | 5 0.85900 | | Fig2.2 | <--- | Figural.time2 | | | | | |
| 9 | 6 0.88599 | | Fig2.3 | <--- | Figural.time2 | | | | | |
| 10 | i 0.62103 | | Figural | <--- | Induction | | | | | |
| 11 | j 0.31730 | | Figural.time2 | <--- | Induction | | | | | |
| 12 | k 0.57035 | | Figural.time2 | <--- | Figural | | | | | |

| | Induct1 | Induct2 | Induct3 | Figural1 | Figural2 | Figural3 | Fig2.1 | Fig2.2 | Fig2.3 |
|----------|---------|---------|---------|----------|----------|----------|--------|--------|--------|
| Induct1 | 0.00 | -0.44 | 0.77 | -0.48 | 0.56 | 1.29 | -2.65 | 1.01 | -0.38 |
| Induct2 | -0.44 | 0.00 | -0.29 | 0.52 | 1.66 | 0.42 | 0.89 | 0.28 | 1.43 |
| Induct3 | 0.77 | -0.29 | 0.00 | 0.25 | 1.00 | -3.76 | -0.64 | -2.62 | 0.25 |
| Figural1 | -0.48 | 0.52 | 0.25 | 0.61 | 1.19 | -1.33 | 2.33 | -1.74 | -1.24 |
| Figural2 | 0.56 | 1.66 | 1.00 | 1.19 | 1.33 | 1.01 | -2.80 | 1.73 | 4.20 |
| Figural3 | 1.29 | 0.42 | -3.76 | -1.33 | 1.01 | -1.40 | -1.44 | -1.50 | -0.62 |
| Fig2.1 | -2.65 | 0.89 | -0.64 | 2.33 | -2.80 | -1.44 | -0.89 | -1.11 | -0.10 |
| Fig2.2 | 1.01 | 0.28 | -2.62 | -1.74 | 1.73 | -1.50 | -1.11 | -1.49 | -0.45 |
| Fig2.3 | -0.38 | 1.43 | 0.25 | -1.24 | 4.20 | -0.62 | -0.10 | -0.45 | 0.58 |

The difference in χ^2 is trivial and we have gained two degrees of freedom. This suggests that the two measures are factorially equivalent.

12.5.1 Testing for factorial equivalence in multiple groups

Not shown in this chapter is how to test for equivalence of measurement across different groups. This involves best fitting the model for multiple groups simultaneously and will be discussed in the next section (as yet unwritten).

Chapter 13

Further issues: Item quality

13.1 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 \(2006\)](#).

13.2 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 interrelated; 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)(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 et al., 1992b) 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.

13.3 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. The `circ.sim` and `item.sim` functions 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.

13.4 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 13.4 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 `%%` to convert `i` to `i modulo 2`.) Note that the confirmatory model has more degrees of freedom than the exploratory, because it is forcing the small loadings to be exactly zero.

```
> 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
```

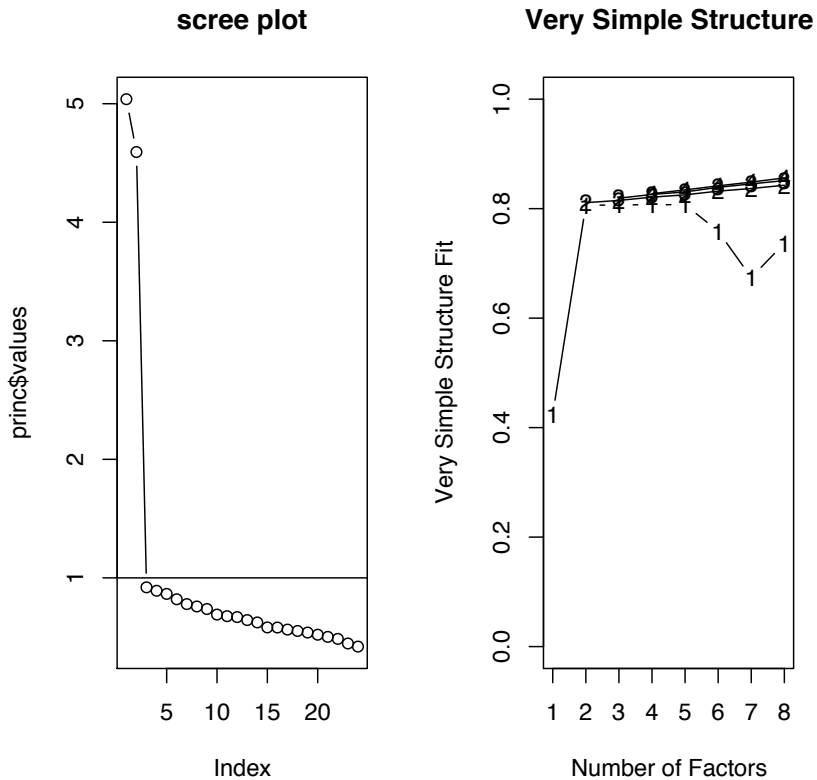


Fig. 13.1 Determining the number of factors to extract from 24 variables generated with a simple structure. The left hand panel shows the scree plot, the right hand panel a VSS plot. Notice the inflection at two factors, suggesting a two factor solution

```
+ }
+ 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)
```

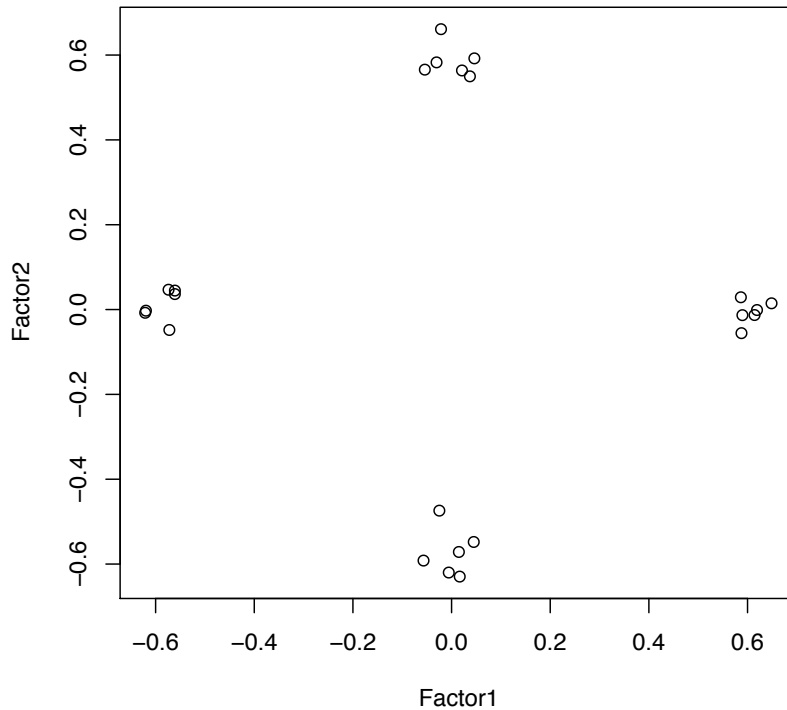


Fig. 13.2 Factor loadings for 24 items on two dimensions.

```

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

13.4.1 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 |
| 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

13.4.2 3 categories of responses

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

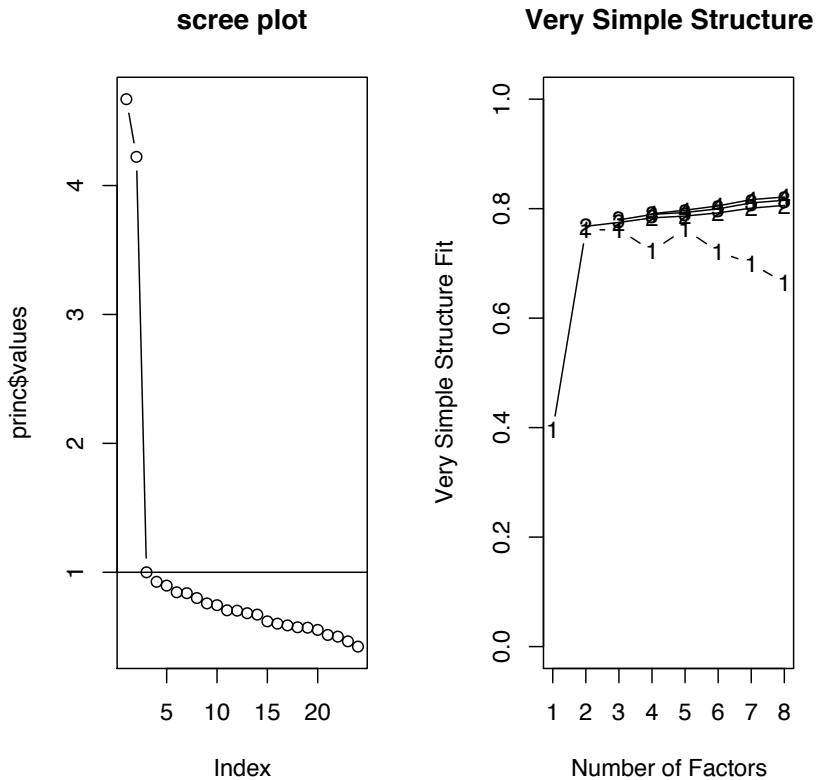


Fig. 13.3 Determining the number of factors to extract from 24 variables generated with a simple structure with 5-point items. The left hand panel shows the scree plot, the right hand panel a VSS plot. Compare with Figure 13.4

```
> 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 | | | | | | |

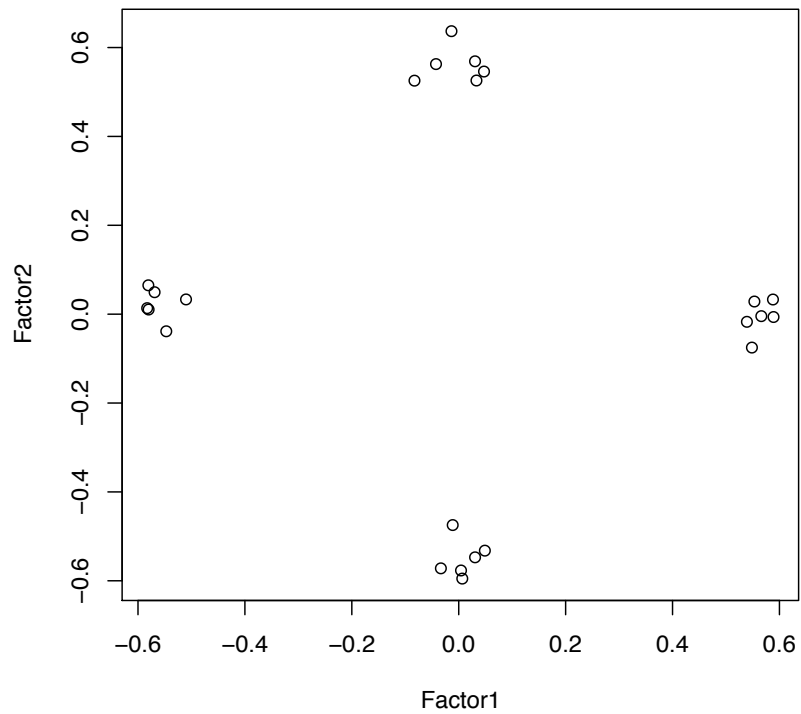


Fig. 13.4 24 variables loading on two factors for categorical items. Compare with Figure ??

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

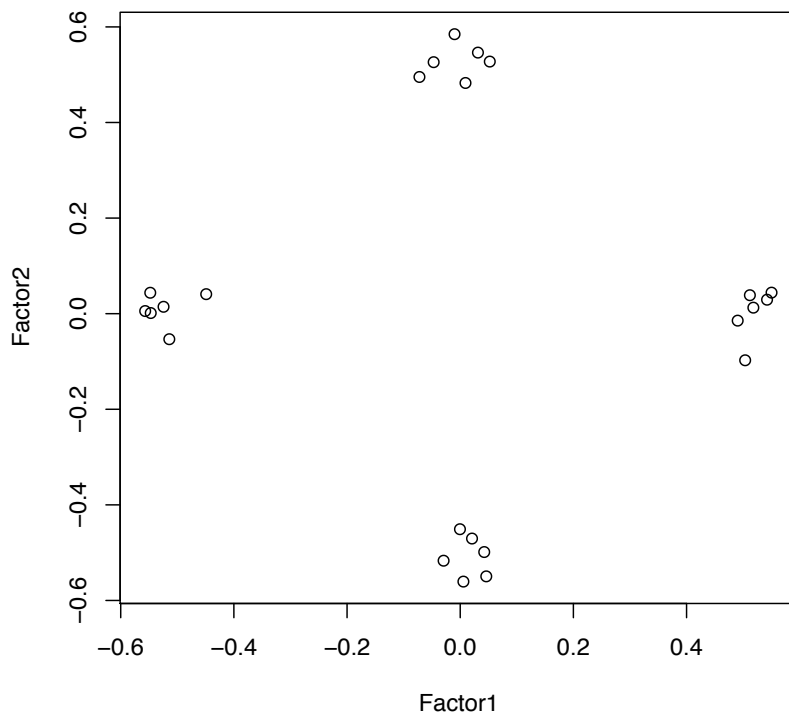


Fig. 13.5 24 variables, simple structure. Items are constrained to have 3 categories

13.4.3 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:

| | 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

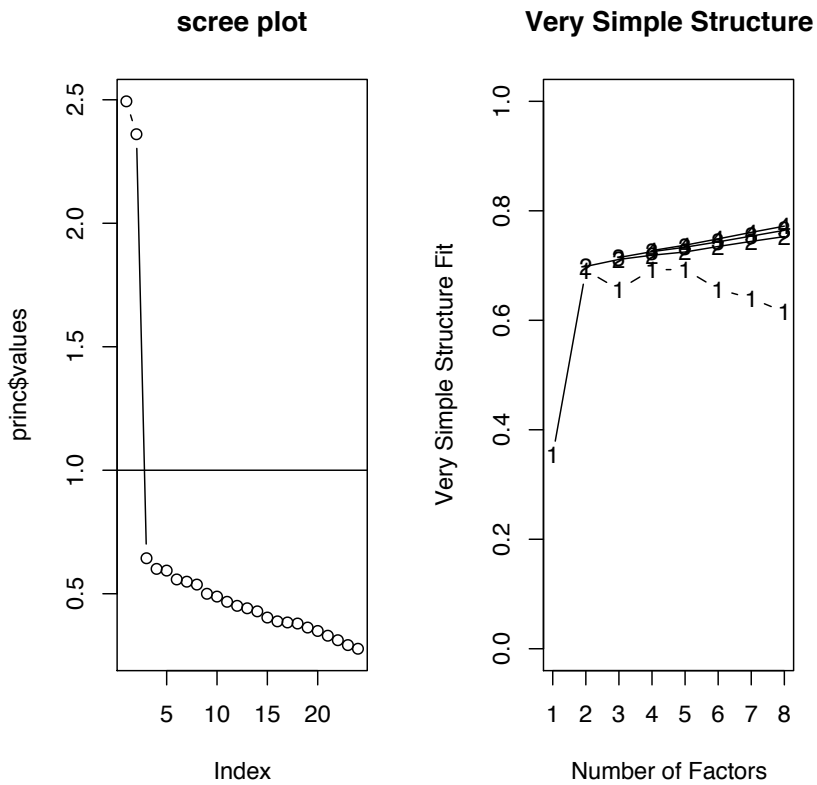


Fig. 13.6 Determining the number of factors to extract from 24 variables generated with a simple structure for dichotomous items. The left hand panel shows the scree plot, the right hand panel a VSS plot. Compare with Figures 13.4 and 13.4.1

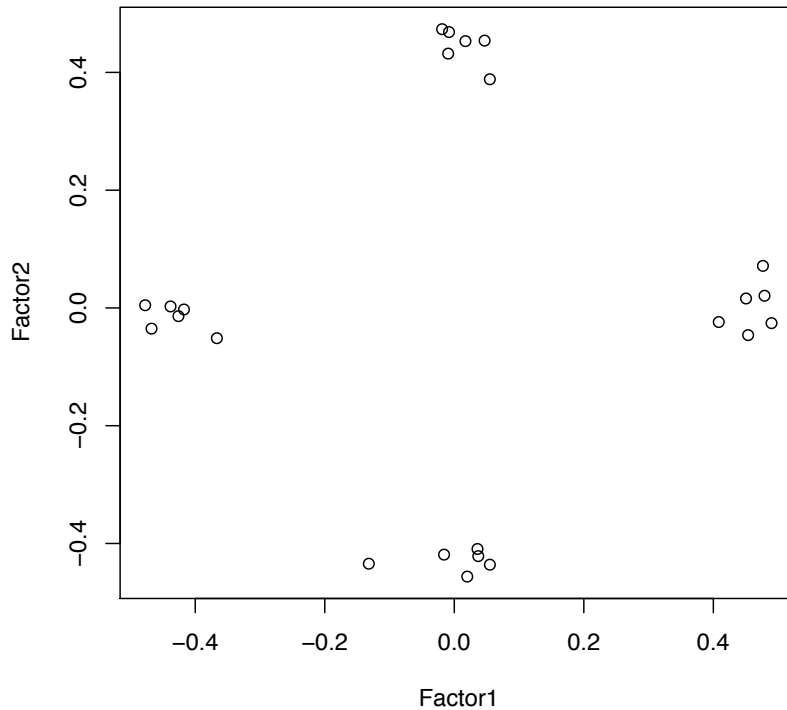


Fig. 13.7 24 variables, simple structure. Items are constrained to be dichotomous.

13.5 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 |

| | 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

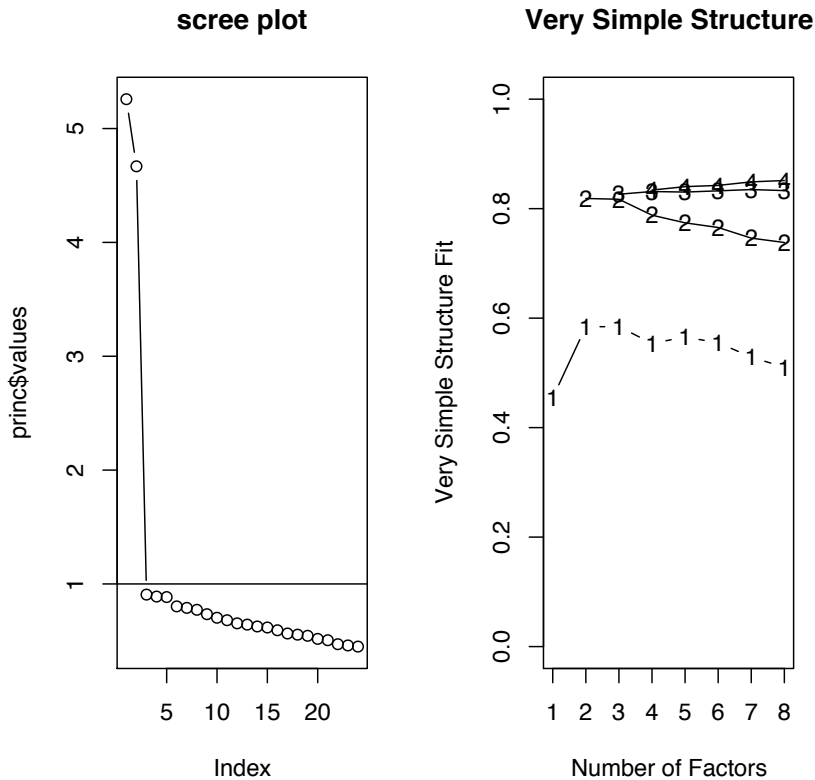


Fig. 13.8 Determining the number of factors to extract from 24 variables generated with a circumplex structure. The left hand panel shows the scree plot, the right hand panel a VSS plot. Notice the inflection at two factors, suggesting a two factor solution

13.5.1 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 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 Chi-square = 2297   Df = 252   Pr(>ChiSq) = 0
Chi-square (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)
```

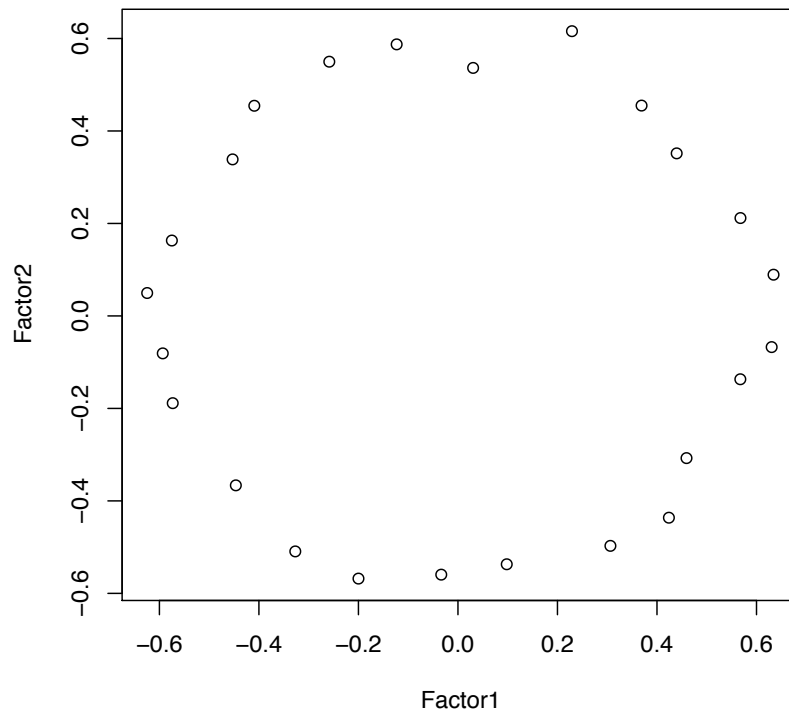


Fig. 13.9 Factor loadings for 24 items on two dimensions. Given that the data were generated to reflect uniform locations around a two dimensional space, the circular ordering of loadings is not surprising.

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

13.5.2 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 [13.5.1](#) 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.

13.6 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.

13.6.1 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 13.6.1](#)). Although the factor analysis loadings recover the structure very well ([Figure 13.6.1](#)

```
> 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 13.4.1) 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 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.

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

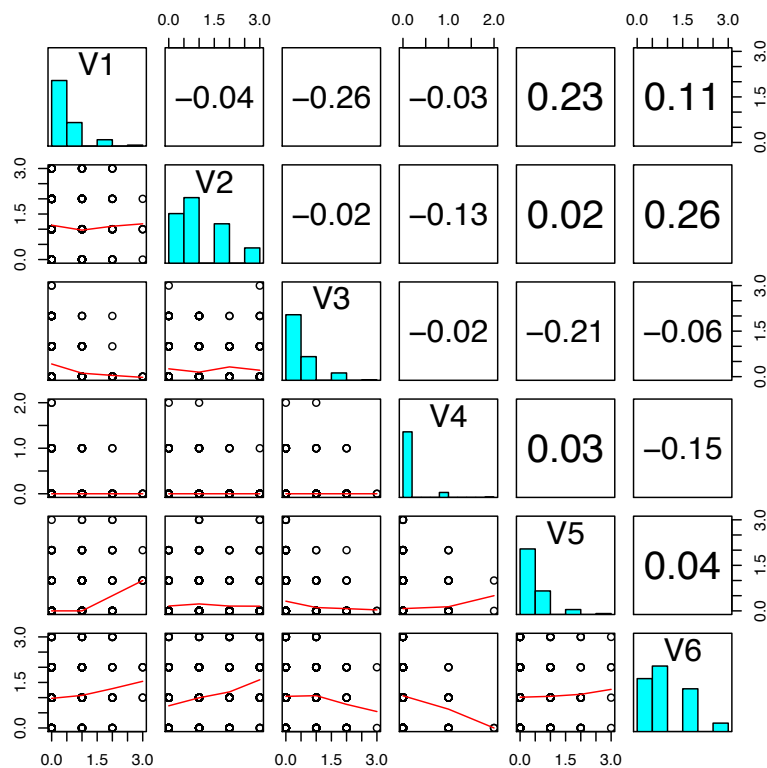


Fig. 13.10 SPLOM of the first 6 variables showing the effect of skew. Note how the correlations of items with opposite skew are very attenuated.

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

Model Chi-square = 703   Df = 254 Pr(>ChiSq) = 0
Chi-square (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.
```

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

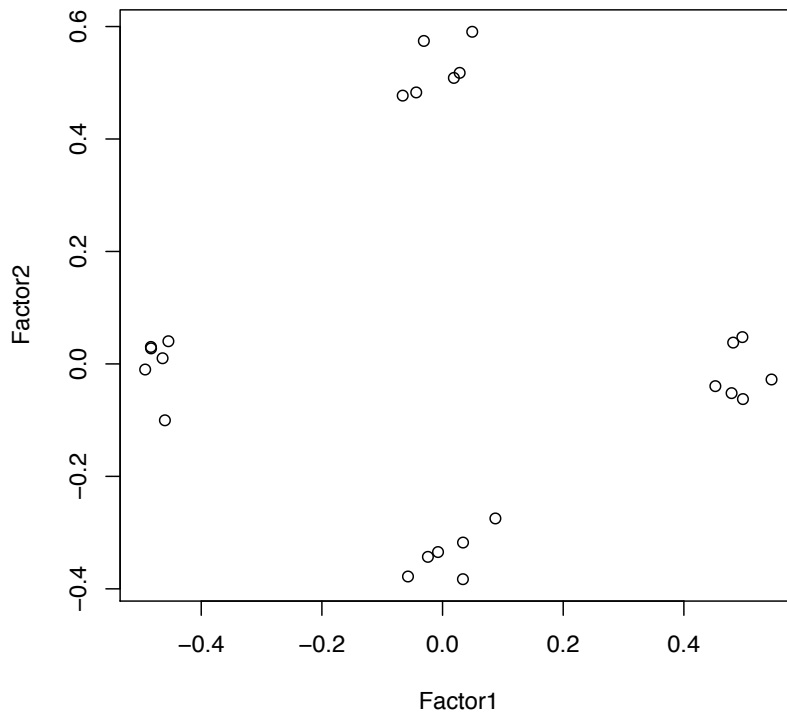


Fig. 13.11 The factor structure of very skewed items recovers the space quite well, at least in terms of angular location. The loadings are less than they should be given the data generation algorithm.

```
-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 |
| 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



Fig. 13.12 A two dimensional solution does not fit very well.

13.6.2 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

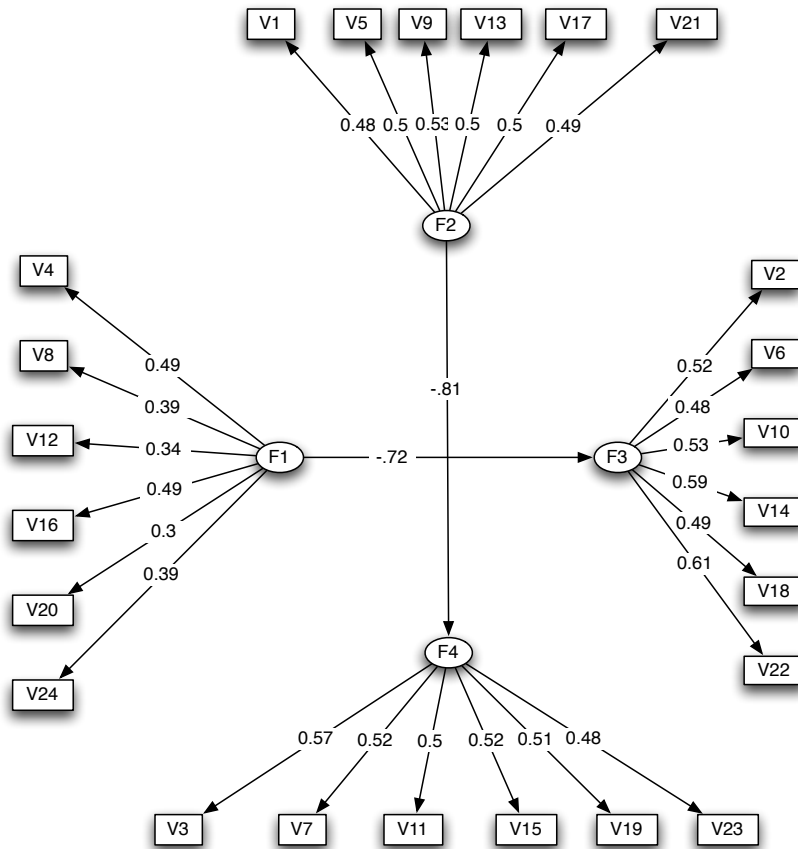


Fig. 13.13 A two dimensional solution does not fit very well, but a 4 factor model in two space matches the generating function very well.

13.7 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)

      [,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 13.6.2. 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 |
| 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

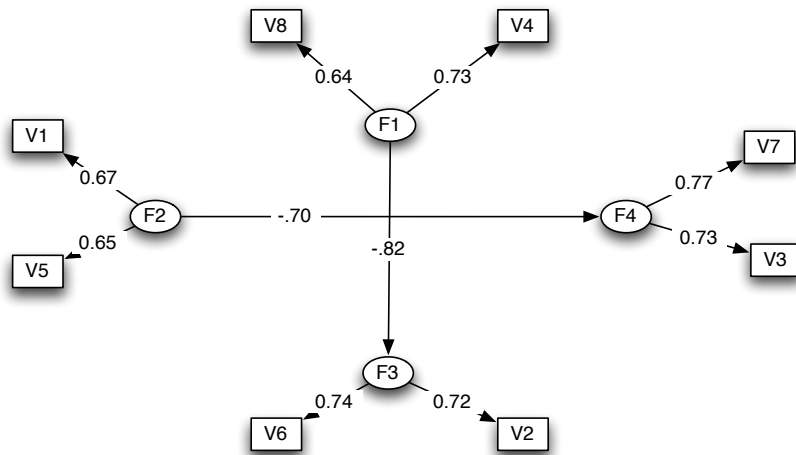


Fig. 13.14 An alternative solution is to group the variables into “testlets” or “homogeneous item composites” (HICs) and then to examine the structure of the HICs.

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.

Chapter 14

Evaluating models

Evaluating model fit in SEM may be done by examining the various indices of fit and misfit supplied by the programs used. The R `sem` package provides 8 indices which we will discuss in this chapter. LISREL, EQS, and Mplus provide even more.

However, before considering the various indices of fit, it is important to consider the reasons that models do not fit.

1. Errors in theory
 - a. Failure to include the appropriate variables
 - b. Failure to model the appropriate relationships
2. Errors in assumptions
 - a. Problems in distributions
 - b. Methods or correlated error factors

14.1 Model misspecification: failure to include variables

A classic problem in statements of causal structure is the failure to include appropriate variables. Such model misspecification is the bane of using correlations to infer anything about causality, for there is always the lurking third variable that could explain the relationship.

In an attempt to demonstrate this effect, consider the correlation between three variables at time 1 as predictors of an important outcome at time 2. The measured variables at time 1 are Yellow Fingers, Yellow Teeth and Bad Breath. The outcome variable is probability of Lung Cancer (rescored with a logistic transformation to be a continuous variable ranging from -3 to 3.)¹

For the purposes of this demonstration, we create an artificial correlation matrix of these four variables by defining a latent variable, θ , with factor loadings θ . The product of $\theta\theta^T$ is the observed correlation matrix:

```
> theta <- matrix(c(0.8, 0.7, 0.6, 0.5), nrow = 4)
> observed <- theta %*% t(theta)
```

¹ As I hope is obvious, this is an artificial example. It was inspired, in part, by the webpage on causal and statistical reasoning at Carnegie Mellon University (www.cmu.edu/CSR/index.html)

```

> diag(observed) <- 1
> rownames(observed) <- colnames(observed) <- c("breath", "teeth",
+       "fingers", "cancer")
> observed

      breath teeth fingers cancer
breath  1.00  0.56   0.48  0.40
teeth   0.56  1.00   0.42  0.35
fingers 0.48  0.42   1.00  0.30
cancer  0.40  0.35   0.30  1.00

```

14.1.1 Misspecified Linear Regression

Using classical linear regression, we can predict cancer risk given 1, 2, or 3 predictors. To do this from the observed correlation matrix, we can use the `solve` function in base R, or alternatively the `mat.regress` function in the `psych` package. This latter function will take a correlation matrix and then find the beta weights for a set of X predictors of Y variables. We do this multiple times, first to regress smoking on yellow fingers, then upon yellow teeth and yellow fingers, and then finally, on breath, yellow teeth and yellow fingers. Finally, compare the `mat.regress` output with using the `solve` function.

Remember to load the `psych` package before running this analysis.

```

> library(psych)
> mat.regress(observed, 3, 4)

$beta
fingers
  0.3

$R2
cancer
  0.09

> mat.regress(observed, c(2, 3), 4)

$beta
  teeth fingers
  0.27   0.19

$R2
cancer
  0.15

> mat.regress(observed, c(1:3), 4)

$beta
breath  teeth fingers
  0.26   0.16   0.11

```

```

$R2
cancer
  0.19

> beta <- solve(observed[1:3, 1:3], observed[4, 1:3])
> round(beta, 2)

breath  teeth fingers
  0.26   0.16   0.11

```

Note how the beta weight for yellow fingers decreases as we add more variables into the model. The final model with all three predictors may be summarized as in Figure 14.1.

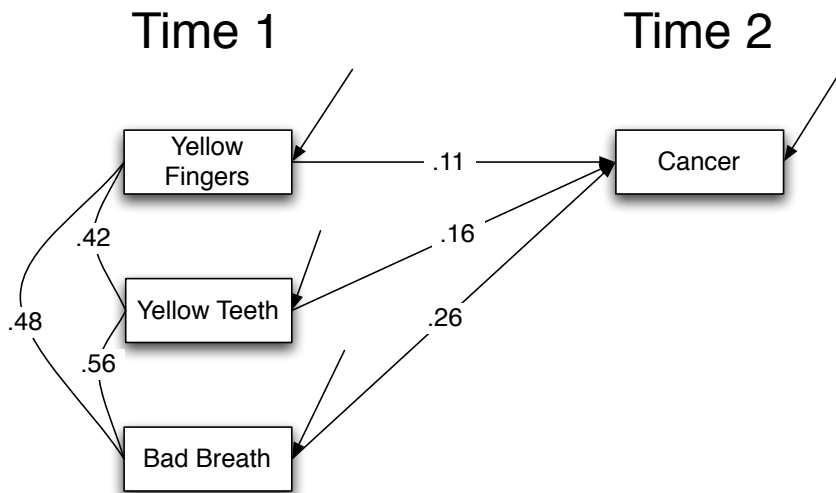


Fig. 14.1 The direct and indirect effect of three predictors upon a criterion variable. The “real”, causal variable is missing from the model.

14.1.2 Regression with the correct variables included

We can restate the θ term in the generating model (14.1) as “smoking” and generate the correlation matrix again, as well as the regressions. This time we add the “smoking” variable with a loading of 1.0 on the latent variable.

```

> theta <- matrix(c(1, 0.8, 0.7, 0.6, 0.5), nrow = 5)
> observed <- theta %*% t(theta)
> diag(observed) <- 1
> rownames(observed) <- colnames(observed) <- c("smoking", "breath",
+       "teeth", "fingers", "cancer")
> observed

```

| | smoking | breath | teeth | fingers | cancer |
|---------|---------|--------|-------|---------|--------|
| smoking | 1.0 | 0.80 | 0.70 | 0.60 | 0.50 |
| breath | 0.8 | 1.00 | 0.56 | 0.48 | 0.40 |
| teeth | 0.7 | 0.56 | 1.00 | 0.42 | 0.35 |
| fingers | 0.6 | 0.48 | 0.42 | 1.00 | 0.30 |
| cancer | 0.5 | 0.40 | 0.35 | 0.30 | 1.00 |

```
> mat.regress(observed, 4, 5)
```

```
$beta
fingers
  0.3
```

```
$R2
cancer
  0.09
```

```
> mat.regress(observed, c(3, 4), 5)
```

```
$beta
  teeth fingers
  0.27   0.19
```

```
$R2
cancer
  0.15
```

```
> mat.regress(observed, c(2:4), 5)
```

```
$beta
  breath  teeth fingers
  0.26   0.16   0.11
```

```
$R2
cancer
  0.19
```

```
> mat.regress(observed, c(1:4), 5)
```

```
$beta
smoking  breath  teeth fingers
  0.5     0.0     0.0     0.0
```

```
$R2
cancer
  0.25
```

Notice how if the model is correctly specified (i.e., the causal variable, smoking, is introduced), the beta weights for the non-causal variables go to zero. This is understandable if we consider the beta weights in the two predictor case:

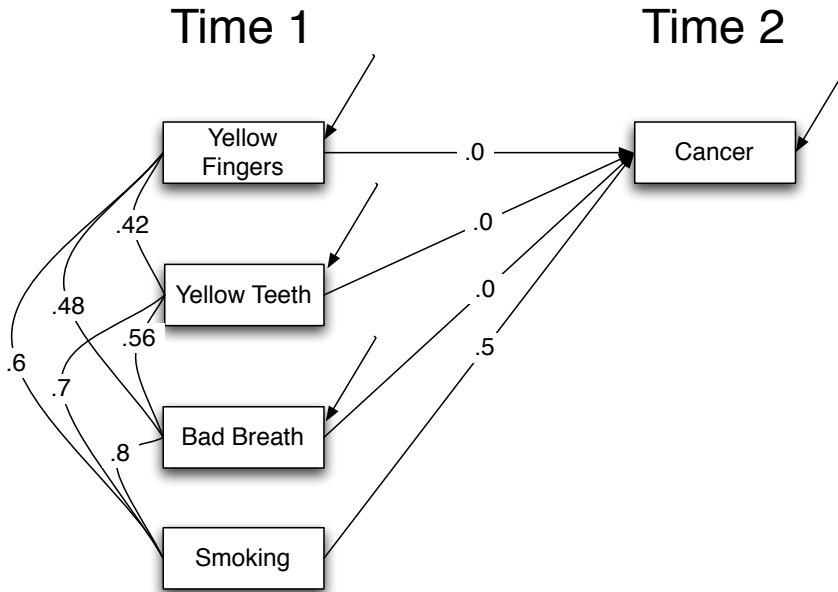


Fig. 14.2 . The direct and indirect effect of four predictors upon a criterion variable If the “correct” causal variable is specified, the β weights of the remaining variables are reduced to zero. Compare the β weights with those in Figure 14.1

$$\left\{ \begin{aligned} \beta_1 &= (r_{x1y}r_{x2x2} - r_{x1x2}r_{x2y}) / (r_{x1x1}r_{x2x2} - r_{x1x2}^2) \\ \beta_2 &= (r_{x2y}r_{x1x1} - r_{x1x2}r_{x1y}) / (r_{x1x1}r_{x2x2} - r_{x1x2}^2) \end{aligned} \right\}$$

In the more general case,

$$\beta R = r_{xy} \tag{14.1}$$

and we can solve E.16 for β by multiplying both sides by the inverse of R.

$$\beta = \beta R R^{-1} = r_{xy} R^{-1}$$

In the two variable case (see Appendix 2), finding the inverse of a two by two matrix is discussed and is shown to be

$$R^{-1} = \begin{pmatrix} \frac{r_{22}}{r_{11}r_{22}-r_{12}^2} & -\frac{r_{12}}{r_{11}r_{22}-r_{12}^2} \\ -\frac{r_{12}}{r_{11}r_{22}-r_{12}^2} & \frac{r_{11}}{r_{11}r_{22}-r_{12}^2} \end{pmatrix} \tag{14.2}$$

14.1.3 Misspecified Structural Equation Models

The regression models in 14.1 are misspecified in that the “real” causal variable is not included in the model. This same problem can arise in structural equations. That is, we can fit the data very well with a model which is, in fact, incorrect. In parallel with the misspecification of the linear regression, compare a series of structural equation models. The first one is fully

saturated (has no degrees of freedom), and models the effect of yellow fingers as leading to cancer. Note how we are using a subset of the correlation matrix. Remember to load the **sem** package before running this analysis.

14.1.3.1 one predictor

```

      path                parameter initial value
[1,] "fingers -> cancer"  "1"          NA
[2,] "fingers <-> fingers" "5"          NA
[3,] "cancer <-> cancer"  "8"          NA

Model Chisquare = -9.6e-15  Df = 0 Pr(>Chisq) = NA
Chisquare (null model) = 9.3  Df = 1
Goodness-of-fit index = 1
BIC = -9.6e-15

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0      0      0      0      0      0

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.30     0.096     3.1    1.8e-03 cancer <--- fingers
5 1.00     0.142     7.0    2.0e-12 fingers <--> fingers
8 0.91     0.129     7.0    2.0e-12 cancer <--> cancer

Iterations = 0

```

Note how the path coefficient for fingers -> cancer is identical to the beta weight found in the regression model for one predictor variable (and is, in the one predictor case equal, of course, to the zero order correlation). Also note that the unexplained variance of cancer is equal to $1 - r^2$.

Compare this result to the model that just models the correlation between yellow fingers and cancer:

```

      path                parameter initial value
[1,] "fingers <-> cancer"  "1"          NA
[2,] "fingers <-> fingers" "5"          NA
[3,] "cancer <-> cancer"  "8"          NA

Model Chisquare = -9.6e-15  Df = 0 Pr(>Chisq) = NA
Chisquare (null model) = 9.3  Df = 1
Goodness-of-fit index = 1
BIC = -9.6e-15

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0      0      0      0      0      0

```



```

Parameter Estimates
Estimate Std Error z value Pr(>|z|)
1 0.3      0.10      2.9      4.2e-03  cancer <--> fingers
5 1.0      0.14      7.0      2.0e-12  fingers <--> fingers
8 1.0      0.14      7.0      2.0e-12  cancer <--> cancer

Iterations = 0

```

14.1.3.2 Two predictors, don't model the correlation

A slightly more complicated model adds the effects of having yellow teeth.

```

      path                parameter initial value
[1,] "fingers -> cancer"    "1"          NA
[2,] "teeth -> cancer"     "2"          NA
[3,] "fingers <-> fingers" "5"          NA
[4,] "teeth <-> teeth"     "6"          NA
[5,] "cancer <-> cancer"   "8"          NA

Model Chisquare = 19  Df = 1 Pr(>Chisq) = 1.2e-05
Chisquare (null model) = 35  Df = 3
Goodness-of-fit index = 0.9
Adjusted goodness-of-fit index = 0.37
RMSEA index = 0.43  90% CI: (0.28, 0.6)
Bentler-Bonnett NFI = 0.46
Tucker-Lewis NNFI = -0.69
Bentler CFI = 0.44
BIC = 15

```

```

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00  0.31   0.76   1.39   1.14   4.18

```

```

Parameter Estimates
Estimate Std Error z value Pr(>|z|)
1 0.19    0.10     1.8    6.9e-02  cancer <--- fingers
2 0.27    0.10     2.7    7.7e-03  cancer <--- teeth
5 1.00    0.14     7.0    2.0e-12  fingers <--> fingers
6 1.00    0.14     7.0    2.0e-12  teeth <--> teeth
8 0.85    0.12     7.0    2.0e-12  cancer <--> cancer

Iterations = 0

```

```

      teeth fingers cancer
teeth 0.000    0.42 0.078
fingers 0.420    0.00 0.114
cancer 0.078    0.11 0.042

```

Even with 100 subjects, the model does not fit in terms of χ^2 or any of the conventional fit statistics. Although the path coefficients predicting cancer exactly match the regression betas, the failure to fit is due to the failure to model the correlations between the predictor variables. That is, our measurement model is faulty (because we are not actually trying to measure it.)

14.1.3.3 Two predictors, model the correlation

```

      path                parameter initial value
[1,] "fingers -> cancer"    "1"          NA
[2,] "teeth -> cancer"     "2"          NA
[3,] "fingers <-> fingers" "5"          NA
[4,] "teeth <-> teeth"    "6"          NA
[5,] "fingers <-> teeth"  "7"          NA
[6,] "cancer <-> cancer"  "8"          NA

```

```

Model Chisquare = 5.5e-15  Df = 0 Pr(>Chisq) = NA
Chisquare (null model) = 35  Df = 3
Goodness-of-fit index = 1
BIC = 5.5e-15

```

Normalized Residuals

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0         0         0         0         0         0

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
1 0.19      0.10      1.8    6.9e-02 cancer <--- fingers
2 0.27      0.10      2.7    7.7e-03 cancer <--- teeth
5 1.00      0.14      7.0    2.0e-12 fingers <--> fingers
6 1.00      0.14      7.0    2.0e-12 teeth <--> teeth
7 0.42      0.11      3.9    1.2e-04 teeth <--> fingers
8 0.85      0.12      7.0    2.0e-12 cancer <--> cancer

```

```
Iterations = 0
```

```

      teeth fingers cancer
teeth      0         0         0
fingers    0         0         0
cancer     0         0         0

```

Fitting the correlation between fingers and teeth produces a fully saturated model (with no degrees of freedom). The paths are the correct beta weights.

14.1.4 Three predictors - alternative models

There are a variety of ways to model the effect of three predictors on the outcome variable. The model that is logically the equivalent of the regression model is to consider the three predictors as independent. Alternatives to this consider various ways in which the predictors could be related.

14.1.4.1 Three predictors, don't model the correlations

```

      path                parameter initial value
[1,] "fingers -> cancer"  "1"          NA
[2,] "teeth -> cancer"    "2"          NA
[3,] "breath -> cancer"   "3"          NA
[4,] "fingers <-> fingers" "5"          NA
[5,] "teeth <-> teeth"    "6"          NA
[6,] "breath <-> breath"  "7"          NA
[7,] "cancer <-> cancer"  "8"          NA

Model Chi-square = 68  Df = 3 Pr(>ChiSq) = 1.4e-14
Chi-square (null model) = 89  Df = 6
Goodness-of-fit index = 0.74
Adjusted goodness-of-fit index = 0.12
RMSEA index = 0.47  90% CI: (0.37, 0.57)
Bentler-Bonnett NFI = 0.24
Tucker-Lewis NNFI = -0.56
Bentler CFI = 0.22
BIC = 54

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.0    1.2    2.0    2.5    4.3    5.6

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.11     0.11     1.0    3.0e-01 cancer <--- fingers
2 0.16     0.11     1.4    1.5e-01 cancer <--- teeth
3 0.26     0.12     2.2    2.5e-02 cancer <--- breath
5 1.00     0.14     7.0    2.0e-12 fingers <--> fingers
6 1.00     0.14     7.0    2.0e-12 teeth <--> teeth
7 1.00     0.14     7.0    2.0e-12 breath <--> breath
8 0.81     0.11     7.0    2.0e-12 cancer <--> cancer

Iterations = 0

      breath teeth fingers cancer
breath  0.00  0.56   0.48  0.142
teeth   0.56  0.00   0.42  0.190

```

```
fingers  0.48  0.42   0.00  0.191
cancer   0.14  0.19   0.19  0.088
```

As we saw before, although the prediction paths from the predictors to the criterion match the beta weights, the model does not fit, because this model fails to model the correlation between the predictors. Once again, our failure to have a measurement model is at fault.

We can fix the variance of the predictors to increase the degrees of freedom, but we are still not modeling the covariances.

```
      path                parameter initial value
[1,] "fingers -> cancer"   "1"          NA
[2,] "teeth -> cancer"     "2"          NA
[3,] "breath -> cancer"    "3"          NA
[4,] "fingers <-> fingers" NA          "1"
[5,] "teeth <-> teeth"     NA          "1"
[6,] "breath <-> breath"   NA          "1"
[7,] "cancer <-> cancer"   "8"          NA

> summary(sem.4a, digits = 2)

Model Chisquare = 68  Df = 6 Pr(>Chisq) = 1.3e-12
Chisquare (null model) = 89  Df = 6
Goodness-of-fit index = 0.74
Adjusted goodness-of-fit index = 0.56
RMSEA index = 0.32  90% CI: (0.26, 0.39)
Bentler-Bonnett NFI = 0.24
Tucker-Lewis NNFI = 0.26
Bentler CFI = 0.26
BIC = 40

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.0    1.2    2.0    2.5    4.3    5.6

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.11    0.11    1.0    3.0e-01 cancer <--- fingers
2 0.16    0.11    1.4    1.5e-01 cancer <--- teeth
3 0.26    0.12    2.2    2.5e-02 cancer <--- breath
8 0.81    0.11    7.0    2.0e-12 cancer <--> cancer

Iterations = 0

> print(standardized.residuals(sem.4a), digits = 2)

      breath teeth fingers cancer
breath  0.00  0.56   0.48  0.142
teeth   0.56  0.00   0.42  0.190
fingers 0.48  0.42   0.00  0.191
cancer  0.14  0.19   0.19  0.088
```

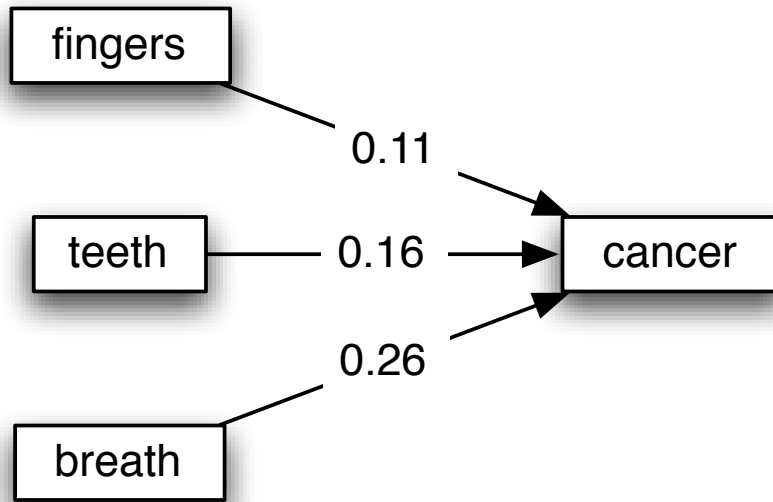


Fig. 14.3 The direct effect of three predictors upon a criterion variable. The “real”, causal variable is missing from the model. Although the direct paths are correct (match the beta weights), the model has a poor fit because the predictors are modeled as uncorrelated. That is, it is not possible to see the indirect effects from this model.

14.1.4.2 Three predictors, model the correlations, case 1

Revise the previous model to include a “yellow” latent variable. That is, we notice from the residuals that yellow teeth and fingers seem to go together. Perhaps, with a bit of creativity, we can explain this as due to the influence of yellowing agents which need to be controlled.

| | path | parameter | initial | value |
|-------|-----------------------|-----------|---------|-------|
| [1,] | "fingers -> cancer" | "1" | NA | |
| [2,] | "teeth -> cancer" | "2" | NA | |
| [3,] | "breath -> cancer" | "3" | NA | |
| [4,] | "fingers <-> fingers" | "5" | NA | |
| [5,] | "teeth <-> teeth" | "6" | NA | |
| [6,] | "breath <-> breath" | "7" | NA | |
| [7,] | "cancer <-> cancer" | "8" | NA | |
| [8,] | "yellow <-> yellow" | NA | "1" | |
| [9,] | "yellow -> fingers" | "10" | NA | |
| [10,] | "yellow -> teeth" | NA | "1" | |

Model Chisquare = 48 Df = 2 Pr(>Chisq) = 3.2e-11
 Chisquare (null model) = 89 Df = 6
 Goodness-of-fit index = 0.84
 Adjusted goodness-of-fit index = 0.19
 RMSEA index = 0.48 90% CI: (0.37, 0.61)
 Bentler-Bonnett NFI = 0.45
 Tucker-Lewis NNFI = -0.68

Bentler CFI = 0.44

BIC = 39

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|---------|---------|---------|---------|---------|---------|
| | 0.0e+00 | 7.1e-06 | 1.3e+00 | 1.9e+00 | 2.3e+00 | 5.6e+00 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|----|----------|-----------|----------|----------|----------------------|
| 1 | 1.1e-01 | 0.11 | 1.0e+00 | 3.0e-01 | cancer <--- fingers |
| 2 | 1.6e-01 | 0.11 | 1.4e+00 | 1.5e-01 | cancer <--- teeth |
| 3 | 2.6e-01 | 0.12 | 2.2e+00 | 2.5e-02 | cancer <--- breath |
| 5 | 8.2e-01 | 0.12 | 6.9e+00 | 6.1e-12 | fingers <--> fingers |
| 6 | -6.5e-07 | 0.14 | -4.6e-06 | 1.0e+00 | teeth <--> teeth |
| 7 | 1.0e+00 | 0.14 | 7.0e+00 | 2.0e-12 | breath <--> breath |
| 8 | 8.1e-01 | 0.11 | 7.0e+00 | 2.0e-12 | cancer <--> cancer |
| 10 | 4.2e-01 | 0.11 | 3.9e+00 | 1.2e-04 | fingers <--- yellow |

Iterations = 14

| | breath | teeth | fingers | cancer |
|---------|--------|---------|---------|--------|
| breath | 0.00 | 5.6e-01 | 4.8e-01 | 0.142 |
| teeth | 0.56 | 6.5e-07 | 7.8e-07 | 0.145 |
| fingers | 0.48 | 7.8e-07 | 4.0e-07 | 0.124 |
| cancer | 0.14 | 1.4e-01 | 1.2e-01 | 0.073 |

This model is significant improvement over the previous model, (examine the change in χ^2 for the one degree of freedom used), but still does not fit very well.

14.1.4.3 Three predictors, model the correlations, case 2

Looking at the residuals suggests perhaps we should model a latent mouth variable as well. Perhaps the yellowing of the teeth have an additional component related to being in the mouth.

| | path | parameter | initial value |
|-------|-----------------------|-----------|---------------|
| [1,] | "fingers -> cancer" | "1" | NA |
| [2,] | "teeth -> cancer" | "2" | NA |
| [3,] | "breath -> cancer" | "3" | NA |
| [4,] | "fingers <-> fingers" | "5" | NA |
| [5,] | "teeth <-> teeth" | "6" | NA |
| [6,] | "breath <-> breath" | "7" | NA |
| [7,] | "cancer <-> cancer" | "8" | NA |
| [8,] | "yellow <-> yellow" | NA | "1" |
| [9,] | "yellow -> fingers" | "10" | NA |
| [10,] | "yellow -> teeth" | NA | "1" |
| [11,] | "mouth -> teeth" | NA | "1" |
| [12,] | "mouth -> breath" | "11" | NA |
| [13,] | "mouth <-> mouth" | NA | "1" |

```

Model Chisquare = 26   Df = 1 Pr(>Chisq) = 3.5e-07
Chisquare (null model) = 89   Df = 6
Goodness-of-fit index = 0.9
Adjusted goodness-of-fit index = -0.033
RMSEA index = 0.5   90% CI: (0.35, 0.68)
Bentler-Bonnett NFI = 0.71
Tucker-Lewis NNFI = -0.81
Bentler CFI = 0.7
BIC = 21

```

Normalized Residuals

```

      Min.  1st Qu.  Median    Mean  3rd Qu.    Max.
-8.9e-07  6.4e-01  7.8e-01  1.4e+00  1.8e+00  4.8e+00

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
1    0.11    0.105     1.0   3.0e-01  cancer <--- fingers
2    0.16    0.111     1.4   1.5e-01  cancer <--- teeth
3    0.26    0.115     2.2   2.5e-02  cancer <--- breath
5    0.96    0.136     7.1   1.7e-12  fingers <--> fingers
6   -1.09    0.124    -8.7   0.0e+00  teeth <--> teeth
7    0.78    0.118     6.6   3.4e-11  breath <--> breath
8    0.81    0.115     7.0   2.0e-12  cancer <--> cancer
10   0.20    0.097     2.0   4.3e-02  fingers <--- yellow
11   0.47    0.114     4.1   4.4e-05  breath <--- mouth

```

```
Iterations = 21
```

```

      breath teeth fingers cancer
breath 1.3e-07 0.094 4.8e-01 0.067
teeth  9.4e-02 0.088 2.2e-01 0.063
fingers 4.8e-01 0.224 -1.3e-07 0.160
cancer  6.7e-02 0.063 1.6e-01 0.045

```

This is a great improvement (once again, look at the change in χ^2 for the 1 degree of freedom more complex model), but the model still does not fit at all well.

14.1.4.4 Three predictors, model the correlations, case 3

Alternatively, we could just allow all the predictors to correlate:

```

      path                parameter initial value
[1,] "fingers -> cancer"  "1"         NA
[2,] "teeth -> cancer"   "2"         NA
[3,] "breath -> cancer"  "3"         NA
[4,] "fingers <-> fingers" "5"         NA
[5,] "teeth <-> teeth"   "6"         NA
[6,] "breath <-> breath" "7"         NA

```

```
[7,] "cancer <-> cancer"   "8"      NA
[8,] "teeth <-> breath"    "9"      NA
[9,] "teeth <-> fingers"  "10"     NA
[10,] "fingers <-> breath" "11"     NA
```

```
Model Chisquare = 2.2e-14 Df = 0 Pr(>Chisq) = NA
Chisquare (null model) = 89 Df = 6
Goodness-of-fit index = 1
BIC = 2.2e-14
```

Normalized Residuals

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0e+00 0.0e+00 0.0e+00 1.3e-16 1.3e-16 5.3e-16
```

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|----|----------|-----------|---------|----------|----------------------|
| 1 | 0.11 | 0.11 | 1.0 | 3.0e-01 | cancer <--- fingers |
| 2 | 0.16 | 0.11 | 1.4 | 1.5e-01 | cancer <--- teeth |
| 3 | 0.26 | 0.12 | 2.2 | 2.5e-02 | cancer <--- breath |
| 5 | 1.00 | 0.14 | 7.0 | 2.0e-12 | fingers <--> fingers |
| 6 | 1.00 | 0.14 | 7.0 | 2.0e-12 | teeth <--> teeth |
| 7 | 1.00 | 0.14 | 7.0 | 2.0e-12 | breath <--> breath |
| 8 | 0.81 | 0.11 | 7.0 | 2.0e-12 | cancer <--> cancer |
| 9 | 0.56 | 0.12 | 4.9 | 1.2e-06 | breath <--> teeth |
| 10 | 0.42 | 0.11 | 3.9 | 1.2e-04 | fingers <--> teeth |
| 11 | 0.48 | 0.11 | 4.3 | 1.7e-05 | breath <--> fingers |

```
Iterations = 0
```

| | breath | teeth | fingers | cancer |
|---------|---------|-------|---------|---------|
| breath | 0.0e+00 | 0 | 0.0e+00 | 5.6e-17 |
| teeth | 0.0e+00 | 0 | 0.0e+00 | 0.0e+00 |
| fingers | 0.0e+00 | 0 | 0.0e+00 | 5.6e-17 |
| cancer | 5.6e-17 | 0 | 5.6e-17 | 0.0e+00 |

This model is fully saturated, and thus the χ^2 statistic is meaningless. The β weights match the regression model, and the modeled correlations match the data.

However, if we fix the variances of the three predictors to be 1, then we have gained 3 degrees of freedom and now the model looks great!

| | path | parameter | initial value |
|------|-----------------------|-----------|---------------|
| [1,] | "fingers -> cancer" | "1" | NA |
| [2,] | "teeth -> cancer" | "2" | NA |
| [3,] | "breath -> cancer" | "3" | NA |
| [4,] | "fingers <-> fingers" | NA | "1" |
| [5,] | "teeth <-> teeth" | NA | "1" |
| [6,] | "breath <-> breath" | NA | "1" |
| [7,] | "cancer <-> cancer" | "8" | NA |
| [8,] | "teeth <-> breath" | "9" | NA |


```

[9,] "teeth <-> fingers"  "10"      NA
[10,] "fingers <-> breath" "11"      NA

Model Chisquare = 2.2e-14 Df = 3 Pr(>Chisq) = 1
Chisquare (null model) = 89 Df = 6
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 = 1.1
Bentler CFI = 1
BIC = -14

Normalized Residuals
  Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0e+00 0.0e+00 0.0e+00 1.3e-16 1.3e-16 5.3e-16

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.11 0.105 1.0 3.0e-01 cancer <--- fingers
2 0.16 0.111 1.4 1.5e-01 cancer <--- teeth
3 0.26 0.115 2.2 2.5e-02 cancer <--- breath
8 0.81 0.115 7.0 2.0e-12 cancer <--> cancer
9 0.56 0.060 9.3 0.0e+00 breath <--> teeth
10 0.42 0.075 5.6 2.6e-08 fingers <--> teeth
11 0.48 0.069 6.9 4.6e-12 breath <--> fingers

Iterations = 0

      breath teeth fingers cancer
breath 0.0e+00 0 0.0e+00 5.6e-17
teeth 0.0e+00 0 0.0e+00 0.0e+00
fingers 0.0e+00 0 0.0e+00 5.6e-17
cancer 5.6e-17 0 5.6e-17 0.0e+00

```

14.1.5 Three predictors, model the correlations with one latent variable

An alternative model to that in Figure 14.4 is to note that the three predictors correlate and to consider that perhaps they reflect an unknown, latent variable. Perhaps it is this latent variable which leads to cancer.

```

      path                parameter initial value
[1,] "latent -> cancer"  "1"          NA
[2,] "latent -> breath"  "2"          NA
[3,] "latent -> fingers" "3"          NA

```

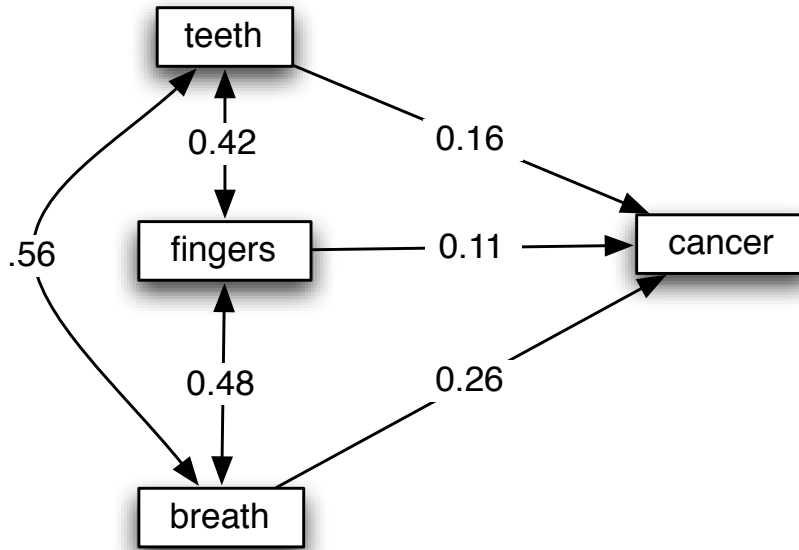


Fig. 14.4 The direct and indirect effect of three predictors upon a criterion variable using sem. The “real”, causal variable is missing from the model. The direct paths are correct (match the beta weights), and the model has an excellent fit because the correlations between the predictors are modeled.

```

[4,] "latent -> teeth"      "4"      NA
[5,] "fingers <-> fingers" "5"      NA
[6,] "teeth <-> teeth"    "6"      NA
[7,] "breath <-> breath"  "7"      NA
[8,] "cancer <-> cancer"  "8"      NA
[9,] "latent <-> latent"   NA       "1"
  
```

```

Model Chisquare = 1.9e-10 Df = 2 Pr(>Chisq) = 1
Chisquare (null model) = 89 Df = 6
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 = 1.1
Bentler CFI = 1
BIC = -9.2
  
```

Normalized Residuals

```

Min. 1st Qu. Median Mean 3rd Qu. Max.
9.4e-07 1.7e-06 3.0e-06 3.5e-06 4.7e-06 1.2e-05
  
```

Parameter Estimates

```

Estimate Std Error z value Pr(>|z|)
1 0.50 0.11 4.7 2.8e-06 cancer <--- latent
  
```

```

2 0.80    0.10    7.8    5.1e-15  breath <--- latent
3 0.60    0.10    5.8    8.2e-09  fingers <--- latent
4 0.70    0.10    6.8    9.8e-12  teeth <--- latent
5 0.64    0.11    5.9    4.8e-09  fingers <--> fingers
6 0.51    0.10    4.9    1.2e-06  teeth <--> teeth
7 0.36    0.11    3.3    9.1e-04  breath <--> breath
8 0.75    0.12    6.4    2.0e-10  cancer <--> cancer

```

Iterations = 13

```

          breath  teeth fingers  cancer
breath  6.1e-07  5.4e-07  3.5e-07  5.2e-07
teeth   5.4e-07  2.8e-07  1.1e-07  2.9e-07
fingers 3.5e-07  1.1e-07  1.3e-07  1.8e-07
cancer  5.2e-07  2.9e-07  1.8e-07  1.7e-06

```

Ah, that did it. We now understand the “causal” structure (although our inference about what is common between bad breath, yellow teeth and yellow fingers will probably ignore the real cause). The secret to solving lung cancer is to use mouth freshners, visit your dentist, and wear latex gloves! (See Figure 14.5)

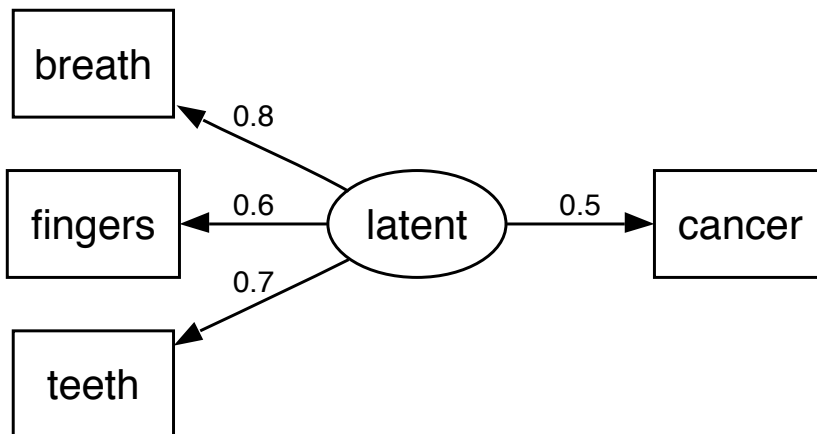


Fig. 14.5 Faulty inference can be the result of a misspecified model. Whatever is common to bad breath, yellow teeth, and yellow hands seems to lead to lung cancer. Thus, one should use mouth freshners, visit your dentist, and wear latex gloves.

14.1.6 Three predictors with shared “error”

An alternative model is to consider the latent variable as accounting for the correlations between the three observed variables, and then to have direct paths from yellow teeth, yellow fingers, and bad breath to cancer. (This is functionally equivalent to the regression model.)

Note how this model is conceptually very different from the previous one in which the latent variable was seen as common to all four variables.

```

      path                parameter initial value
[1,] "breath -> cancer"    "A"          NA
[2,] "teeth -> cancer"     "B"          NA
[3,] "fingers -> cancer"  "C"          NA
[4,] "latent -> breath"   "2"          NA
[5,] "latent -> fingers"  "3"          NA
[6,] "latent -> teeth"    "4"          NA
[7,] "fingers <-> fingers" "5"          NA
[8,] "teeth <-> teeth"    "6"          NA
[9,] "breath <-> breath"  "7"          NA
[10,] "cancer <-> cancer" "8"          NA
[11,] "latent <-> latent" NA          "1"

Model Chisquare = 1.5e-10  Df = 0 Pr(>Chisq) = NA
Chisquare (null model) = 89  Df = 6
Goodness-of-fit index = 1
BIC = 1.5e-10

```

```

Normalized Residuals
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-4.8e-06 -1.4e-06 -7.0e-07 -1.0e-06 -1.6e-07  1.5e-06

```

```

Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
A 0.26      0.12      2.2    2.5e-02 cancer <--- breath
B 0.16      0.11      1.4    1.5e-01 cancer <--- teeth
C 0.11      0.11      1.0    3.0e-01 cancer <--- fingers
2 0.80      0.11      7.2    7.1e-13 breath <--- latent
3 0.60      0.11      5.6    1.9e-08 fingers <--- latent
4 0.70      0.11      6.4    1.3e-10 teeth <--- latent
5 0.64      0.11      5.7    1.5e-08 fingers <--> fingers
6 0.51      0.12      4.4    1.2e-05 teeth <--> teeth
7 0.36      0.13      2.8    5.5e-03 breath <--> breath
8 0.81      0.11      7.0    2.0e-12 cancer <--> cancer

```

```
Iterations = 15
```

```

      breath  teeth  fingers  cancer
breath  1.1e-07 -5.5e-07 -2.6e-07 -8.8e-08
teeth  -5.5e-07  2.1e-07 -1.7e-08 -1.1e-07
fingers -2.6e-07 -1.7e-08  8.5e-08 -6.1e-08
cancer  -8.8e-08 -1.1e-07 -6.1e-08 -4.7e-08

```

But the previous model is fully saturated. We can revise the model somewhat by forcing all three paths with the latent variable to be equal. This frees up two degrees of freedom and results in a very good fit.

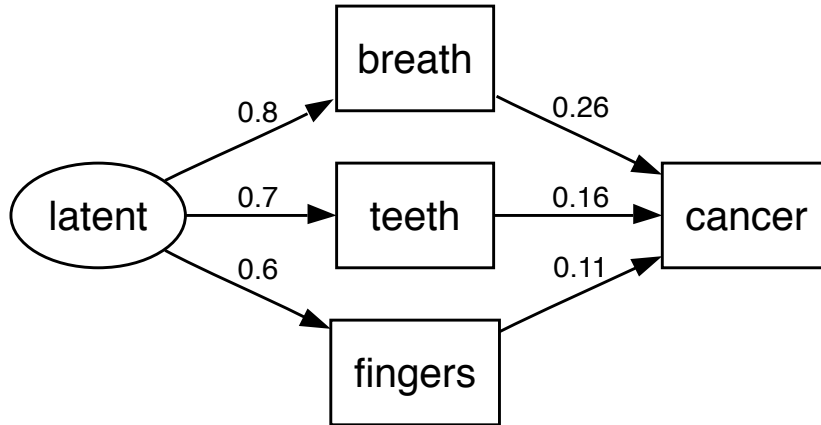


Fig. 14.6 Correlated errors not associated with the criterion. Alternatively, there is something in common to yellow teeth, yellow fingers, and bad breath, but whatever it is that they do not share leads to cancer.

```

> model.8b <- matrix(c("breath -> cancer", "A", NA, "teeth -> cancer",
+ "B", NA, "fingers -> cancer", "C", NA, "latent -> breath",
+ 2, NA, "latent -> fingers", 2, NA, "latent -> teeth", 2, NA,
+ "fingers <-> fingers", 5, NA, "teeth <-> teeth", 6, NA, "breath <-> breath",
+ 7, NA, "cancer <-> cancer", 8, NA, "latent <-> latent", NA,
+ 1), byrow = TRUE, ncol = 3)
> model.8b

      [,1]      [,2] [,3]
[1,] "breath -> cancer" "A" NA
[2,] "teeth -> cancer" "B" NA
[3,] "fingers -> cancer" "C" NA
[4,] "latent -> breath" "2" NA
[5,] "latent -> fingers" "2" NA
[6,] "latent -> teeth" "2" NA
[7,] "fingers <-> fingers" "5" NA
[8,] "teeth <-> teeth" "6" NA
[9,] "breath <-> breath" "7" NA
[10,] "cancer <-> cancer" "8" NA
[11,] "latent <-> latent" NA "1"

> sem.8b <- sem(model.8b, observed[2:5, 2:5], 100)
> summary(sem.8b, digits = 2)

Model Chi-square = 1.9 Df = 2 Pr(>ChiSq) = 0.39
Chi-square (null model) = 89 Df = 6
Goodness-of-fit index = 1
Adjusted goodness-of-fit index = 0.95
RMSEA index = 0 90% CI: (NA, 0.20)
Bentler-Bonnett NFI = 0.98

```

```
Tucker-Lewis NNFI = 1
Bentler CFI = 1
BIC = -7.3
```

Normalized Residuals

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--------|---------|--------|--------|---------|-------|
| -0.660 | -0.233 | 0.039 | -0.020 | 0.223 | 0.583 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|---|----------|-----------|---------|----------|----------------------|
| A | 0.26 | 0.115 | 2.2 | 2.5e-02 | cancer <--- breath |
| B | 0.16 | 0.111 | 1.4 | 1.5e-01 | cancer <--- teeth |
| C | 0.11 | 0.105 | 1.0 | 3.0e-01 | cancer <--- fingers |
| 2 | 0.70 | 0.069 | 10.3 | 0.0e+00 | breath <--- latent |
| 5 | 0.59 | 0.108 | 5.4 | 5.3e-08 | fingers <--> fingers |
| 6 | 0.50 | 0.096 | 5.2 | 2.4e-07 | teeth <--> teeth |
| 7 | 0.45 | 0.091 | 4.9 | 8.2e-07 | breath <--> breath |
| 8 | 0.81 | 0.115 | 7.0 | 2.0e-12 | cancer <--> cancer |

Iterations = 10

```
> print(standardized.residuals(sem.8b), digits = 2)
```

| | breath | teeth | fingers | cancer |
|---------|--------|---------|---------|---------|
| breath | 0.058 | 0.0637 | -0.016 | 0.0232 |
| teeth | 0.064 | 0.0065 | -0.076 | 0.0092 |
| fingers | -0.016 | -0.0763 | -0.083 | -0.0255 |
| cancer | 0.023 | 0.0092 | -0.025 | 0.0047 |

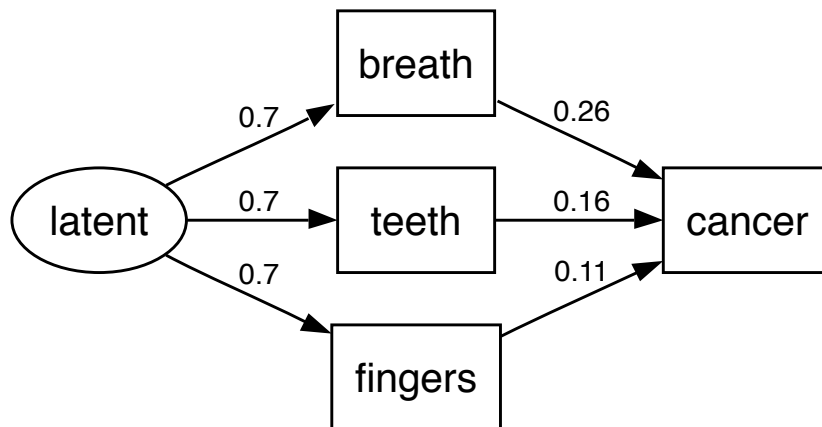


Fig. 14.7 Fixing the correlated errors paths. There is something in common to yellow teeth, yellow fingers, and bad breath, but whatever it is that they do not share leads to cancer. By fixing the paths to the latent variable to be the same the model has gained two degrees of freedom. The fit is very good.

14.1.7 Reverse the causal paths

Yet another alternative model is to think of cancer as the cause of bad breath, yellow teeth and yellow fingers. (That is to say, to reverse the causal arrows from the previous model.) The paths from cancer to the former “predictors” are now longer β weights, but have become the zero order correlation coefficients.

```
> model.8c <- matrix(c("breath <- cancer", "A", NA, "teeth <- cancer",
+   "B", NA, "fingers <- cancer", "C", NA, "latent -> breath",
+   1, NA, "latent -> fingers", 2, NA, "latent -> teeth", 3, NA,
+   "fingers <-> fingers", 5, NA, "teeth <-> teeth", 6, NA, "breath <-> breath",
+   7, NA, "cancer <-> cancer", NA, 1, "latent <-> latent", NA,
+   1), byrow = TRUE, ncol = 3)
> model.8c
```

```
      [,1]      [,2] [,3]
[1,] "breath <- cancer"      "A" NA
[2,] "teeth <- cancer"      "B" NA
[3,] "fingers <- cancer"    "C" NA
[4,] "latent -> breath"     "1" NA
[5,] "latent -> fingers"    "2" NA
[6,] "latent -> teeth"      "3" NA
[7,] "fingers <-> fingers"  "5" NA
[8,] "teeth <-> teeth"      "6" NA
[9,] "breath <-> breath"    "7" NA
[10,] "cancer <-> cancer"   NA  "1"
[11,] "latent <-> latent"   NA  "1"
```

```
> sem.8c <- sem(model.8c, observed[2:5, 2:5], 100)
> summary(sem.8c, digits = 2)
```

```
Model Chi-square = 1.0e-10 Df = 1 Pr(>ChiSq) = 1
Chi-square (null model) = 89 Df = 6
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 = 1.1
Bentler CFI = 1
BIC = -4.6
```

Normalized Residuals

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|----------|----------|----------|----------|----------|---------|
| | -6.5e-06 | -4.7e-06 | -4.0e-06 | -3.2e-06 | -1.5e-06 | 4.8e-08 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|---|----------|-----------|---------|----------|--------------------|
| A | 0.40 | 0.092 | 4.3 | 1.4e-05 | breath <--- cancer |
| B | 0.35 | 0.094 | 3.7 | 2.0e-04 | teeth <--- cancer |

```

C 0.30    0.096    3.1    1.8e-03  fingers <--- cancer
1 0.69    0.116    6.0    2.4e-09  breath <--- latent
2 0.52    0.110    4.7    2.1e-06  fingers <--- latent
3 0.61    0.113    5.4    7.3e-08  teeth <--- latent
5 0.64    0.113    5.7    1.5e-08  fingers <--> fingers
6 0.51    0.117    4.4    1.2e-05  teeth <--> teeth
7 0.36    0.130    2.8    5.5e-03  breath <--> breath

```

```
Iterations = 16
```

```
> print(standardized.residuals(sem.8c), digits = 2)
```

```

          breath  teeth  fingers  cancer
breath -6.1e-08 -5.5e-07 -2.1e-07  5.2e-09
teeth  -5.5e-07 -6.6e-07 -4.4e-07 -6.9e-07
fingers -2.1e-07 -4.4e-07 -5.7e-07 -4.5e-07
cancer  5.2e-09 -6.9e-07 -4.5e-07  0.0e+00

```

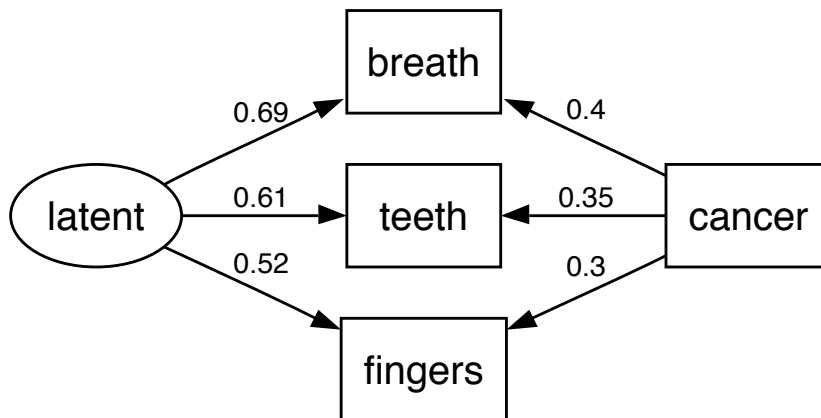


Fig. 14.8 Changing the direction of causation. Perhaps cancer is the causal agent and breath, teeth, and fingers are merely signs of the underlying disease. In addition, perhaps they have some shared error.

14.2 Including the correct variables, but misspecifying the models

Based upon the previous model fitting in section 14.1 we have concluded that there is some latent variable that ties our four variables together. We now examine what happens when we add yet another variable to the mix.

We use the correlation matrix from section 14.1. Note that the correlation matrix is identical for the previous four variables, and that the smoking variable is equivalent to the latent factor that generated the data.

14.2.1 Including the correct variables in linear regression

Remember that if we include smoking into the linear regression, the effect of the other variables vanishes (Figure 14.2)

```
> mat.regress(observed, c(2:4), 5)
```

```
$beta
  breath  teeth fingers
    0.26   0.16   0.11
```

```
$R2
cancer
  0.19
```

```
> mat.regress(observed, c(1:4), 5)
```

```
$beta
smoking  breath  teeth fingers
    0.5     0.0     0.0     0.0
```

```
$R2
cancer
  0.25
```

If, however, we were to make smoking an unreliable measure and thus not perfectly correlated with the latent factor, the other variables still seem to have an effect. We show this by making the latent path from θ to smoking less than one. In the first case, we make the path .9. This is the same as making the reliability of smoking .81. Call this new correlation matrix `observed1`.

```
> theta <- matrix(c(0.9, 0.8, 0.7, 0.6, 0.5), nrow = 5)
> observed1 <- theta %*% t(theta)
> diag(observed1) <- 1
> rownames(observed1) <- colnames(observed1) <- c("smoking", "breath",
+ "teeth", "fingers", "cancer")
> observed1
```

```
      smoking breath teeth fingers cancer
smoking  1.00  0.72  0.63  0.54  0.45
breath   0.72  1.00  0.56  0.48  0.40
teeth    0.63  0.56  1.00  0.42  0.35
fingers  0.54  0.48  0.42  1.00  0.30
cancer   0.45  0.40  0.35  0.30  1.00
```

```
> mat.regress(observed1, c(2:4), 5)
```

```
$beta
  breath  teeth fingers
    0.26   0.16   0.11
```

```

$R2
cancer
  0.19
> mat.regress(observed1, c(1:4), 5)
$beta
smoking  breath  teeth fingers
  0.28    0.13    0.08    0.05
$R2
cancer
  0.22

```

Compare the regression weights for the two data sets (observed and observed1). Note how the other variables still contribute to the regression unless smoking is measured perfectly reliably.

To show it even more clearly, make the θ to smoking path = .5 (This is the equivalent of having a reliability of smoking of .25)

```

> theta <- matrix(c(0.5, 0.8, 0.7, 0.6, 0.5), nrow = 5)
> observed2 <- theta %*% t(theta)
> diag(observed2) <- 1
> rownames(observed2) <- colnames(observed2) <- c("smoking", "breath",
+ "teeth", "fingers", "cancer")
> observed2

      smoking breath teeth fingers cancer
smoking  1.00  0.40  0.35  0.30  0.25
breath   0.40  1.00  0.56  0.48  0.40
teeth    0.35  0.56  1.00  0.42  0.35
fingers  0.30  0.48  0.42  1.00  0.30
cancer   0.25  0.40  0.35  0.30  1.00
> mat.regress(observed2, c(2:4), 5)
$beta
breath  teeth fingers
  0.26   0.16   0.11
$R2
cancer
  0.19
> mat.regress(observed2, c(1:4), 5)
$beta
smoking  breath  teeth fingers
  0.07    0.24    0.15    0.10
$R2
cancer
  0.2

```

Note that in this case, we completely over estimate the contribution of the other variables and underestimate the contribution of smoking. In regression, there is no way to correct for this, but structural equation modeling does allow for various ways of correcting this problem.

14.2.2 Including the correct variables in the Structural Equation

Here we apply the identical model to our three different correlation matrices.

```
> model.9 <- matrix(c("latent -> cancer", 1, NA, "latent -> breath",
+ 2, NA, "latent -> fingers", 3, NA, "latent -> teeth", 4, NA,
+ "latent -> smoking", 9, NA, "fingers <-> fingers", 5, NA, "teeth <-> teeth",
+ 6, NA, "breath <-> breath", 7, NA, "cancer <-> cancer", 8,
+ NA, "smoking <-> smoking", 10, NA, "latent <-> latent", NA,
+ 1), byrow = TRUE, ncol = 3)
> model.9

      [,1]      [,2] [,3]
[1,] "latent -> cancer" "1" NA
[2,] "latent -> breath" "2" NA
[3,] "latent -> fingers" "3" NA
[4,] "latent -> teeth" "4" NA
[5,] "latent -> smoking" "9" NA
[6,] "fingers <-> fingers" "5" NA
[7,] "teeth <-> teeth" "6" NA
[8,] "breath <-> breath" "7" NA
[9,] "cancer <-> cancer" "8" NA
[10,] "smoking <-> smoking" "10" NA
[11,] "latent <-> latent" NA "1"

> sem.9 <- sem(model.9, observed, 100)
> summary(sem.9, digits = 2)

Model Chisquare = 1.8e-11 Df = 5 Pr(>Chisq) = 1
Chisquare (null model) = 240 Df = 10
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 = -23

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-8.2e-07 -3.6e-07 5.1e-08 2.3e-07 8.5e-07 1.2e-06

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
```

```

1  5.0e-01 0.095      5.3e+00 1.3e-07  cancer <--- latent
2  8.0e-01 0.085      9.4e+00 0.0e+00  breath <--- latent
3  6.0e-01 0.092      6.5e+00 7.5e-11  fingers <--- latent
4  7.0e-01 0.089      7.9e+00 3.6e-15  teeth <--- latent
9  1.0e+00 0.075      1.3e+01 0.0e+00  smoking <--- latent
5  6.4e-01 0.093      6.9e+00 4.9e-12  fingers <--> fingers
6  5.1e-01 0.076      6.7e+00 2.3e-11  teeth <--> teeth
7  3.6e-01 0.060      6.0e+00 1.7e-09  breath <--> breath
8  7.5e-01 0.107      7.0e+00 2.8e-12  cancer <--> cancer
10 -7.1e-08 0.048      -1.5e-06 1.0e+00  smoking <--> smoking

```

Iterations = 15

```
> print(standardized.residuals(sem.9), digits = 2)
```

```

          smoking  breath  teeth  fingers  cancer
smoking -1.9e-08 -3.1e-08 -1.0e-07 -2.4e-08 1.2e-07
breath  -3.1e-08  9.0e-08 -5.1e-08  5.7e-09 1.2e-07
teeth   -1.0e-07 -5.1e-08 -6.0e-08 -3.9e-08 6.7e-08
fingers -2.4e-08  5.7e-09 -3.9e-08  1.7e-07 8.9e-08
cancer  1.2e-07  1.2e-07  6.7e-08  8.9e-08 1.6e-07

```

Note that with the perfect data set, the estimate for the error variance of smoking is appropriately very small.

Repeat this analysis with the less than perfect reliability of smoking of the observed1 data set:

```
> sem.10 <- sem(model.9, observed1, 100)
> summary(sem.10, digits = 2)
```

```

Model Chisquare = 1.1e-10  Df = 5  Pr(>Chisq) = 1
Chisquare (null model) = 188  Df = 10
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 = 1.1
Bentler CFI = 1
BIC = -23

```

Normalized Residuals

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-3.8e-06 -2.5e-06  3.1e-07  2.7e-07  2.7e-06  5.7e-06

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
1  0.50      0.099      5.0    5.0e-07  cancer <--- latent
2  0.80      0.088      9.1    0.0e+00  breath <--- latent
3  0.60      0.096      6.2    4.5e-10  fingers <--- latent
4  0.70      0.092      7.6    3.3e-14  teeth <--- latent

```

```

9  0.90    0.084    10.7    0.0e+00  smoking <--- latent
5  0.64    0.098     6.5    7.7e-11  fingers <--> fingers
6  0.51    0.084     6.1    1.1e-09  teeth <--> teeth
7  0.36    0.070     5.1    3.0e-07  breath <--> breath
8  0.75    0.111     6.7    1.7e-11  cancer <--> cancer
10 0.19    0.064     3.0    2.8e-03  smoking <--> smoking

```

Iterations = 12

```
> print(standardized.residuals(sem.10), digits = 2)
```

```

          smoking  breath  teeth fingers  cancer
smoking  1.5e-07  8.2e-08 -1.8e-07 6.5e-07 -2.7e-07
breath   8.2e-08  4.4e-08 -2.2e-07 5.3e-07 -2.8e-07
teeth   -1.8e-07 -2.2e-07  2.9e-09 3.0e-07 -3.9e-07
fingers  6.5e-07  5.3e-07  3.0e-07 5.1e-07  1.2e-07
cancer  -2.7e-07 -2.8e-07 -3.9e-07 1.2e-07 -5.4e-07

```

Repeat this analysis with the even less reliability of smoking of the observed2 data set:

```
> sem.11 <- sem(model.9, observed2, 100)
```

```
> summary(sem.11, digits = 2)
```

```

Model Chisquare = 4.2e-10  Df = 5 Pr(>Chisq) = 1
Chisquare (null model) = 110  Df = 10
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 = 1.1
Bentler CFI = 1
BIC = -23

```

Normalized Residuals

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.0e-05 -2.4e-06 -1.1e-06 -5.4e-07 2.2e-06 5.5e-06

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
1  0.50    0.105    4.7    2.1e-06  cancer <--- latent
2  0.80    0.098    8.2    2.2e-16  breath <--- latent
3  0.60    0.102    5.9    4.6e-09  fingers <--- latent
4  0.70    0.100    7.0    2.3e-12  teeth <--- latent
9  0.50    0.105    4.7    2.1e-06  smoking <--- latent
5  0.64    0.107    6.0    2.2e-09  fingers <--> fingers
6  0.51    0.099    5.1    2.7e-07  teeth <--> teeth
7  0.36    0.097    3.7    2.1e-04  breath <--> breath
8  0.75    0.117    6.4    1.3e-10  cancer <--> cancer
10 0.75    0.117    6.4    1.3e-10  smoking <--> smoking

```

Iterations = 11

```
> print(standardized.residuals(sem.11), digits = 2)

           smoking  breath  teeth  fingers  cancer
smoking  7.7e-07 -2.4e-07  5.2e-07 -2.6e-07  2.3e-07
breath   -2.4e-07 -6.2e-07 -1.4e-08 -1.1e-06 -2.4e-07
teeth    5.2e-07 -1.4e-08 -2.7e-07 -1.2e-07  5.2e-07
fingers  -2.6e-07 -1.1e-06 -1.2e-07 -1.1e-07 -2.6e-07
cancer   2.3e-07 -2.4e-07  5.2e-07 -2.6e-07  7.7e-07
```

We now see the real power of the SEM approach. For by modeling the correlations between the X predictor set, we are able to estimate how unreliable each variable is (the path from a variable to itself reflects the unreliability) and see the structure of the data. But, the conclusion is still wrong, because now we are forced to interpret that whatever it is that is common to smoking, bad breath, yellow fingers and yellow teeth lead to cancer. Although our latent modeling approach has helped and is able to reproduce the data perfectly, it has not led to the correct conclusion as to causality. (See Figure ??).

14.2.3 Direct the causal path

What happens if we make smoking a causal variable that leads to the latent variable?

```
> model.12 <- matrix(c("latent -> cancer", 1, NA, "latent -> breath",
+ 2, NA, "latent -> fingers", 3, NA, "latent -> teeth", 4, NA,
+ "smoking -> latent", NA, 1, "fingers <-> fingers", 5, NA, "teeth <-> teeth",
+ 6, NA, "breath <-> breath", 7, NA, "cancer <-> cancer", 8,
+ NA, "smoking <-> smoking", NA, 1, "latent <-> latent", 12,
+ NA), byrow = TRUE, ncol = 3)
> sem.12 <- sem(model.12, observed, 100)
> summary(sem.12, digits = 2)

Model Chisquare = 4.9e-12  Df = 6 Pr(>Chisq) = 1
Chisquare (null model) = 240  Df = 10
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 = -28

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-5.8e-07 -1.0e-07  1.7e-07  3.1e-07  4.4e-07  1.2e-06

Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1  5.0e-01  0.087    5.7e+00 9.2e-09 cancer <--- latent
2  8.0e-01  0.060    1.3e+01 0.0e+00 breath <--- latent
```

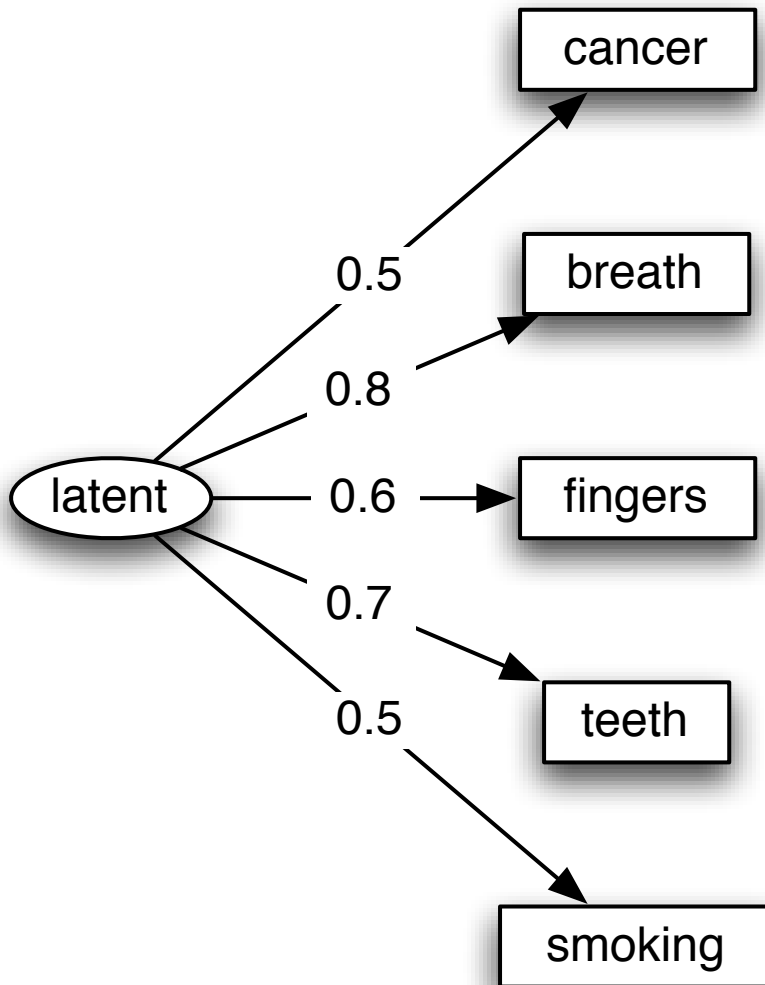


Fig. 14.9 Good fit does not imply “causality”—the problem of incorrect inference. Whatever is common to smoking, bad breath, yellow teeth, and yellow hands also seems to lead to lung cancer. Thus, one should use mouth fresheners, visit your dentist, and wear latex gloves. It is unclear why the latent variable leads to smoking.

```

3  6.0e-01  0.080      7.5e+00  8.5e-14  fingers <--- latent
4  7.0e-01  0.072      9.8e+00  0.0e+00  teeth <--- latent
5  6.4e-01  0.093      6.9e+00  4.9e-12  fingers <--> fingers
6  5.1e-01  0.076      6.7e+00  2.3e-11  teeth <--> teeth
7  3.6e-01  0.060      6.0e+00  1.7e-09  breath <--> breath
8  7.5e-01  0.107      7.0e+00  2.8e-12  cancer <--> cancer
12 9.7e-09  0.048      2.0e-07  1.0e+00  latent <--> latent
  
```

Iterations = 15

```
> print(standardized.residuals(sem.12), digits = 2)
```

```

      smoking  breath  teeth  fingers  cancer
smoking  0.0e+00  5.5e-08  1.0e-07 -1.4e-08 -1.2e-08
breath   5.5e-08  1.4e-07  1.2e-07  1.7e-08  1.5e-08
teeth    1.0e-07  1.2e-07  1.7e-07  4.8e-08  4.0e-08
fingers -1.4e-08  1.7e-08  4.8e-08  2.5e-08 -1.7e-08
cancer  -1.2e-08  1.5e-08  4.0e-08 -1.7e-08 -8.2e-08

```

Repeat this analysis with noisy data from `observed2`. (Remember that in this case, smoking is not measured reliably).

```
> sem.13 <- sem(model.12, observed2, 100)
```

```
> summary(sem.13, digits = 2)
```

```

Model Chisquare = 7.4e-11  Df = 6  Pr(>Chisq) = 1
Chisquare (null model) = 110  Df = 10
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 = 1.1
Bentler CFI = 1
BIC = -28

```

Normalized Residuals

```

      Min.  1st Qu.  Median    Mean  3rd Qu.    Max.
-5.1e-06 -2.7e-06 -2.4e-06 -2.3e-06 -1.1e-06  6.4e-07

```

Parameter Estimates

```

      Estimate Std Error z value Pr(>|z|)
1  0.25      0.070    3.6    3.6e-04  cancer <--- latent
2  0.40      0.087    4.6    4.0e-06  breath <--- latent
3  0.30      0.075    4.0    7.1e-05  fingers <--- latent
4  0.35      0.082    4.3    1.8e-05  teeth <--- latent
5  0.64      0.107    6.0    2.2e-09  fingers <--> fingers
6  0.51      0.099    5.1    2.7e-07  teeth <--> teeth
7  0.36      0.097    3.7    2.1e-04  breath <--> breath
8  0.75      0.117    6.4    1.3e-10  cancer <--> cancer
12 3.00      1.364    2.2    2.8e-02  latent <--> latent

```

Iterations = 18

```
> print(standardized.residuals(sem.13), digits = 2)
```

```

      smoking  breath  teeth  fingers  cancer
smoking  0.0e+00 -3.0e-07 -1.2e-07 -1.2e-07 -2.4e-07
breath   -3.0e-07 -7.3e-07 -3.1e-07 -2.9e-07 -4.8e-07
teeth    -1.2e-07 -3.1e-07  9.1e-08 -8.8e-08 -2.8e-07
fingers  -1.2e-07 -2.9e-07 -8.8e-08 -7.1e-07 -2.5e-07
cancer   -2.4e-07 -4.8e-07 -2.8e-07 -2.5e-07 -2.8e-07

```


We can also model smoking as a noisy variable, and then fix one path (in this case, the latent to cancer) to estimate the model for pure, moderate, and very noisy smoking.

```
> model.14 <- matrix(c("latent -> cancer", NA, 1, "latent -> breath",
+ 2, NA, "latent -> fingers", 3, NA, "latent -> teeth", 4, NA,
+ "smoking -> latent", 11, NA, "fingers <-> fingers", 5, NA,
+ "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA, "cancer <-> cancer",
+ 8, NA, "smoking <-> smoking", 10, NA, "latent <-> latent",
+ NA, 1), byrow = TRUE, ncol = 3)
> sem.14 <- sem(model.14, observed, 100)
> summary(sem.14, digits = 2)
```

```
Model Chisquare = 111  Df = 6 Pr(>Chisq) = 0
Chisquare (null model) = 240  Df = 10
Goodness-of-fit index = 0.76
Adjusted goodness-of-fit index = 0.39
RMSEA index = 0.42  90% CI: (0.35, 0.49)
Bentler-Bonnett NFI = 0.54
Tucker-Lewis NNFI = 0.24
Bentler CFI = 0.55
BIC = 83
```

Normalized Residuals

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|------|
| -4.63 | -3.38 | -1.81 | -1.68 | 0.00 | 0.81 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|----|----------|-----------|---------|----------|----------------------|
| 2 | 0.73 | 0.082 | 8.9 | 0.0e+00 | breath <--- latent |
| 3 | 0.54 | 0.090 | 6.0 | 2.2e-09 | fingers <--- latent |
| 4 | 0.63 | 0.087 | 7.2 | 6.0e-13 | teeth <--- latent |
| 11 | 0.95 | 0.133 | 7.2 | 8.5e-13 | latent <--- smoking |
| 5 | 0.63 | 0.102 | 6.2 | 5.4e-10 | fingers <--> fingers |
| 6 | 0.50 | 0.091 | 5.5 | 3.5e-08 | teeth <--> teeth |
| 7 | 0.32 | 0.082 | 3.9 | 8.5e-05 | breath <--> breath |
| 8 | 1.02 | 0.208 | 4.9 | 9.7e-07 | cancer <--> cancer |
| 10 | 1.00 | 0.142 | 7.0 | 2.0e-12 | smoking <--> smoking |

Iterations = 13

```
> print(standardized.residuals(sem.14), digits = 2)
```

| | smoking | breath | teeth | fingers | cancer |
|---------|---------|--------|-------|---------|--------|
| smoking | 0.00 | 0.10 | 0.10 | 0.09 | -0.45 |
| breath | 0.10 | -0.34 | -0.31 | -0.27 | -0.99 |
| teeth | 0.10 | -0.31 | -0.25 | -0.22 | -0.84 |
| fingers | 0.09 | -0.27 | -0.22 | -0.18 | -0.72 |
| cancer | -0.45 | -0.99 | -0.84 | -0.72 | -1.92 |

```
> sem.15 <- sem(model.14, observed1, 100)
> summary(sem.15, digits = 2)
```

```

Model Chisquare = 71   Df = 6 Pr(>Chisq) = 2.1e-13
Chisquare (null model) = 188   Df = 10
Goodness-of-fit index = 0.81
Adjusted goodness-of-fit index = 0.53
RMSEA index = 0.33   90% CI: (0.27, 0.4)
Bentler-Bonnett NFI = 0.62
Tucker-Lewis NNFI = 0.39
Bentler CFI = 0.63
BIC = 44

```

Normalized Residuals

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|------|
| -4.31 | -3.26 | -1.73 | -1.61 | 0.00 | 0.72 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|----|----------|-----------|---------|----------|----------------------|
| 2 | 0.77 | 0.088 | 8.7 | 0.0e+00 | breath <--- latent |
| 3 | 0.57 | 0.096 | 6.0 | 2.2e-09 | fingers <--- latent |
| 4 | 0.67 | 0.093 | 7.2 | 5.1e-13 | teeth <--- latent |
| 11 | 0.81 | 0.125 | 6.5 | 9.1e-11 | latent <--- smoking |
| 5 | 0.64 | 0.104 | 6.2 | 7.6e-10 | fingers <--> fingers |
| 6 | 0.51 | 0.093 | 5.5 | 4.6e-08 | teeth <--> teeth |
| 7 | 0.35 | 0.086 | 4.0 | 5.3e-05 | breath <--> breath |
| 8 | 0.92 | 0.181 | 5.1 | 3.6e-07 | cancer <--> cancer |
| 10 | 1.00 | 0.142 | 7.0 | 2.0e-12 | smoking <--> smoking |

Iterations = 13

```
> print(standardized.residuals(sem.15), digits = 2)
```

| | smoking | breath | teeth | fingers | cancer |
|---------|---------|--------|--------|---------|--------|
| smoking | 0.000 | 0.095 | 0.087 | 0.075 | -0.36 |
| breath | 0.095 | -0.330 | -0.295 | -0.252 | -0.88 |
| teeth | 0.087 | -0.295 | -0.249 | -0.216 | -0.76 |
| fingers | 0.075 | -0.252 | -0.216 | -0.183 | -0.65 |
| cancer | -0.362 | -0.878 | -0.760 | -0.651 | -1.58 |

```
> sem.16 <- sem(model.14, observed2, 100)
```

```
> summary(sem.16, digits = 2)
```

```

Model Chisquare = 26   Df = 6 Pr(>Chisq) = 0.00021
Chisquare (null model) = 110   Df = 10
Goodness-of-fit index = 0.92
Adjusted goodness-of-fit index = 0.8
RMSEA index = 0.18   90% CI: (0.12, 0.26)
Bentler-Bonnett NFI = 0.76
Tucker-Lewis NNFI = 0.67
Bentler CFI = 0.8
BIC = -1.5

```

Normalized Residuals

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|---------|--------|------|---------|------|
| -3.4 | -2.8 | -1.4 | -1.4 | 0.0 | 0.4 |

Parameter Estimates

| | Estimate | Std Error | z value | Pr(> z) | |
|----|----------|-----------|---------|----------|----------------------|
| 2 | 0.88 | 0.108 | 8.1 | 6.7e-16 | breath <--- latent |
| 3 | 0.67 | 0.114 | 5.9 | 4.4e-09 | fingers <--- latent |
| 4 | 0.78 | 0.110 | 7.1 | 1.6e-12 | teeth <--- latent |
| 11 | 0.40 | 0.113 | 3.6 | 3.7e-04 | latent <--- smoking |
| 5 | 0.64 | 0.107 | 6.0 | 1.9e-09 | fingers <--> fingers |
| 6 | 0.52 | 0.098 | 5.3 | 1.4e-07 | teeth <--> teeth |
| 7 | 0.39 | 0.096 | 4.1 | 4.5e-05 | breath <--> breath |
| 8 | 0.77 | 0.144 | 5.4 | 8.3e-08 | cancer <--> cancer |
| 10 | 1.00 | 0.142 | 7.0 | 2.0e-12 | smoking <--> smoking |

Iterations = 14

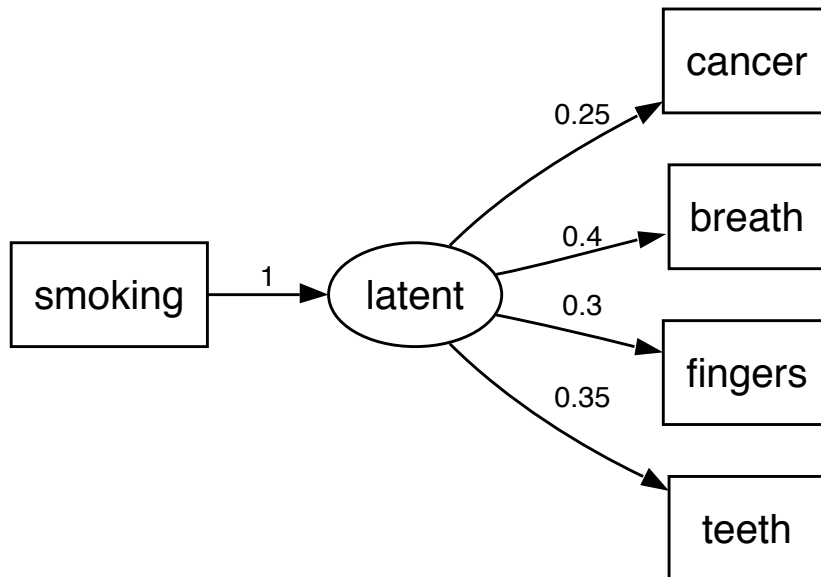


Fig. 14.10 The correct model does not necessarily fit better. Smoking seems to affect something that leads to bad breath, yellow teeth, and yellow hands as well as lung cancer.

14.3 Measures of fit

As has been seen in the previous sections, the use of fit statistics does not guarantee meaningful models. If we do not specify the model correctly, either because we do not include the correct variables or because we fail to use the appropriate measurement model, we will lead to incorrect conclusions.

Even if we have a very good fit, we are unable to determine causal structure from the model, even if we bother to add time into the model.

14.3.1 χ^2

As we saw in the previous chapter, χ^2 is very sensitive to many sources of error in our model specification. χ^2 is sensitive to failures of our distributional assumptions (continuous, multivariate normal) as well as to our failures to correctly specify the structure.

14.3.2 *GFI, NFI, ...*

14.3.3 *RMSEA*

14.4 What does it mean to fit a model

What should we do when the model does not fit? This is a recurring controversy, discussed, for instance in the March, 2007 issue of *Personality and Individual Differences*. It is also a continuing source of debate on the SEM-net list serve. There are those who treat fit statistics (particularly χ^2) as the definitive test and evidence for model adequacy. There are others who do not take such an all or none approach, and are concerned with comparisons of models to alternative models.

Chapter 15

Multidimensional Scaling and Multi-Mode Methods

15.1 Basic models of Multidimensional Scaling

15.1.1 Metric models

15.1.2 Non-metric models

15.2 Measuring Individual Differences in MDS

15.2.1 INDSCAL and ALSCAL

The [Carroll and Chang \(1970\)](#) algorithm has been implemented in the **SensoMineR** package.

Part IV
**The construction of tests and the analysis of
data**

