Psychology 454: Psychological Measurement
Problem set # 2

William Revelle
Swift 315
email: revelle@northwestern.edu

January 13, 2009

The following problems assume that you have installed R on your machine and have also installed the **psych** and **sem** packages.

1. Create a data set of raw scores with the `sim.congeneric` function and a factor structure defined by loadings of .6, .5, .4, and .3 for 200 subjects. Draw the scatter plot matrix of correlations for this data set using the `pairs.panels` function.

   Find the descriptive statistics using the `describe` function.

2. For the following figure (Fig 1) , what are the implied correlations?

3. Simulate a congeneric correlation matrix using the default values using the `sim.congeneric` function

4. Find the factor structure for this using the `factanal` function.

5. Do a confirmatory factor analysis using the `sem` function.

6. Repeat the previous three items for sample sizes of 100, 200 and 400 cases.

7. What are the implied correlations in Figure 2

```
> F <- c(0.9, 0.8, 0.7, 0.6)
> structure.graph(F)
```

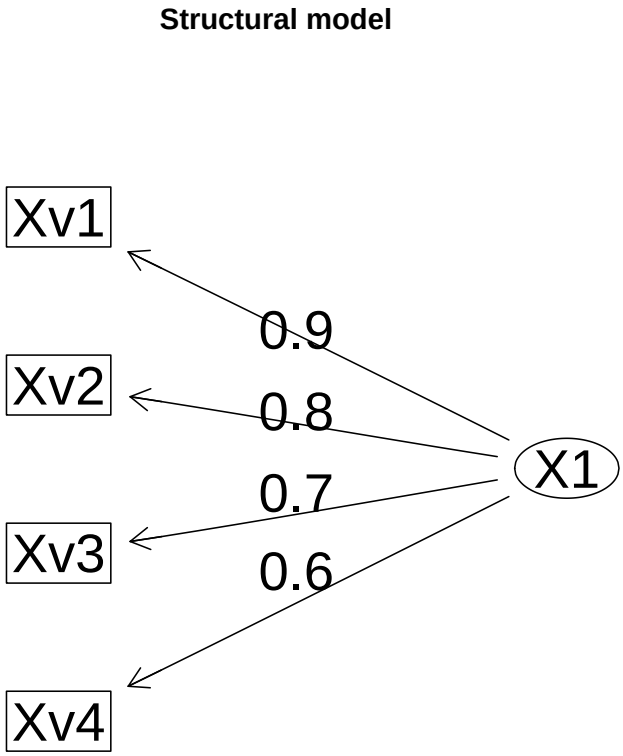

Figure 1: A congeneric model drawn using the `structure.graph` function.

```
> F <- matrix(c("a", "b", rep(0, 4), "c", "d"), ncol = 2)
> Phi <- matrix(c(1, "e", "e", 1), ncol = 2)
> structure.graph(F, Phi = Phi)
```

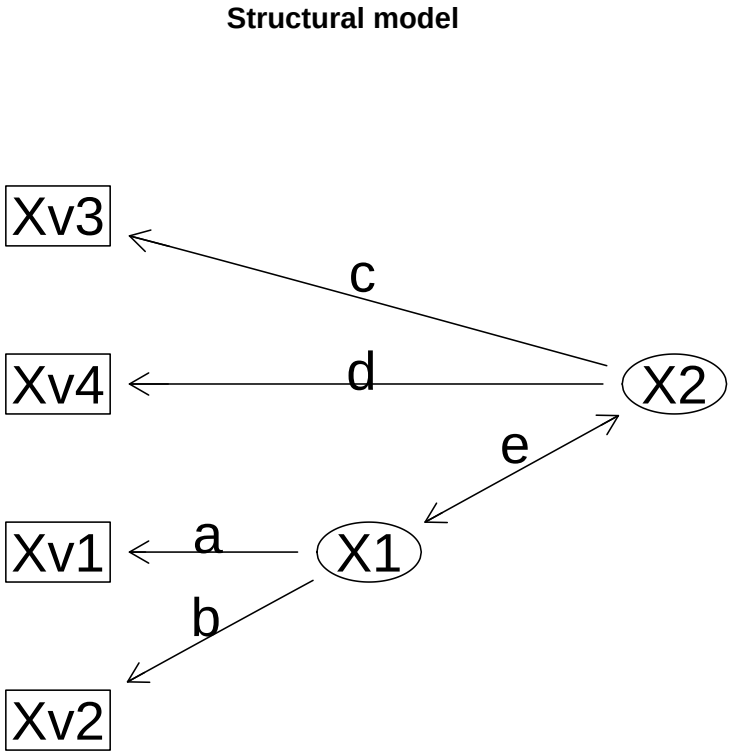

Figure 2: Two correlated factors. The structure was drawn using the `structure.graph` function.

1 Answers to problem set #2

1. ```
   > X <- sim.congeneric(N = 200, loads = c(0.6, 0.5, 0.4, 0.3), short = FALSE)
   > dim(X$observed)
   ```

   ```
   [1] 200   4
   ```

   ```
   > describe(X$observed)
   ```

   ```
      var   n mean   sd median trimmed  mad   min  max range  skew kurtosis   se
   V1   1 200 0.03 0.96   0.01    0.01 0.95 -2.18 2.82  5.00  0.20     0.01 0.07
   V2   2 200 0.09 1.00   0.07    0.08 1.03 -2.33 3.12  5.44  0.10    -0.36 0.07
   V3   3 200 0.04 0.96  -0.07    0.01 0.88 -3.33 3.07  6.40  0.30     0.69 0.07
   V4   4 200 0.10 0.96   0.10    0.11 0.94 -2.58 2.76  5.34 -0.02     0.04 0.07
   ```

```
> pairs.panels(X$observed)
```

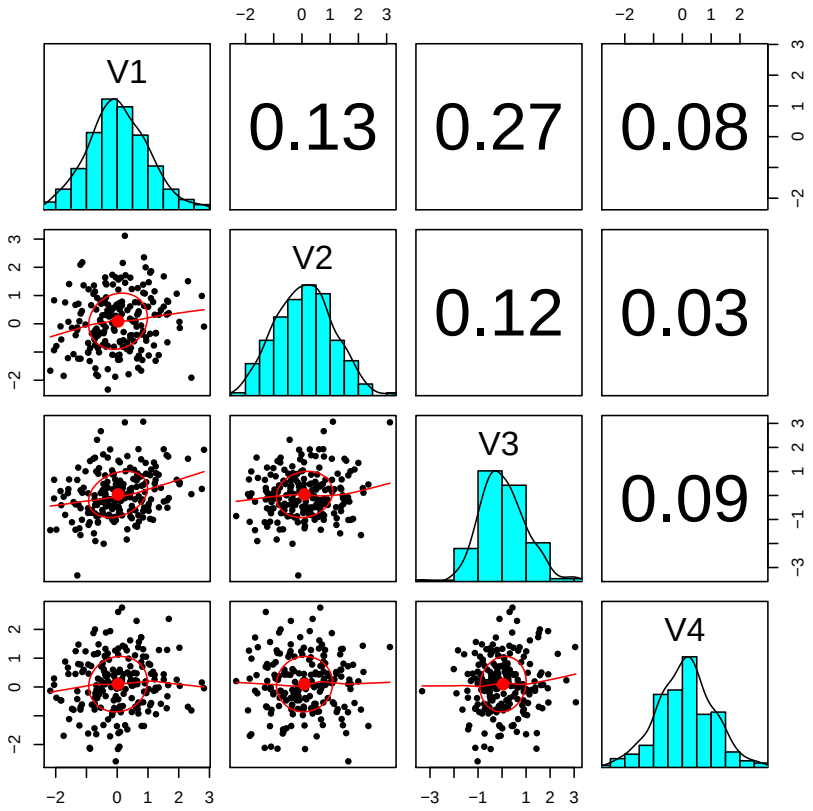

Figure 3: A scatter plot matrix (SPLOM) will show the pairwise correlations, their pairwise scatter plots (below the diagonal) as well as a local smooth

2. Implied correlations may be found by the factor model: $R = F\vec{F}'$

   ```
   > F <- c(0.9, 0.8, 0.7, 0.6)
   > R <- F %*% t(F)
   > R
   ```

   ```
        [,1] [,2] [,3] [,4]
   [1,] 0.81 0.72 0.63 0.54
   ```
   ```

```
[2,] 0.72 0.64 0.56 0.48
[3,] 0.63 0.56 0.49 0.42
[4,] 0.54 0.48 0.42 0.36
```

3.
```
> R <- sim.congeneric()
> R
```

```
     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
```

4.
```
> factanal(covmat = R, factors = 1)
```

```
Call:
factanal(factors = 1, covmat = R)

Uniquenesses:
  V1   V2   V3   V4
0.36 0.51 0.64 0.75

Loadings:
   Factor1
V1 0.8
V2 0.7
V3 0.6
V4 0.5

               Factor1
SS loadings      1.740
Proportion Var   0.435

The degrees of freedom for the model is 2 and the fit was 0
```

5. The structural equation functions are found in the `sem` pacakge

```
> library(sem)
> model = matrix(c("Latent ->  V1", "a", NA, "Latent -> V2", "b", NA, "Latent -> V3", "c",
+     NA, "Latent -> V4", "d", NA, "V1 <-> V1", "u", NA, "V2 <-> V2", "v", NA, "V3 <-> V3",
+     "w", NA, "V4 <-> V4", "x", NA, "Latent <-> Latent", NA, 1), ncol = 3, byrow = TRUE)
> model
```

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

```
> sem4 = sem(model, R, 1000)
> summary(sem4)
```

```
 Model Chisquare =  1.8894e-09   Df =  2 Pr(>Chisq) = 1
```

```
 Chisquare (null model) =  894.56   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 =  1.0068
 Bentler CFI =  1
 SRMR =  6.4327e-07
 BIC =  -13.816

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
3.00e-06 5.52e-06 9.38e-06 1.10e-05 1.48e-05 3.77e-05

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a 0.80     0.032181  24.860  0        V1 <--- Latent
b 0.70     0.032359  21.632  0        V2 <--- Latent
c 0.60     0.032771  18.309  0        V3 <--- Latent
d 0.50     0.033583  14.889  0        V4 <--- Latent
u 0.36     0.034180  10.532  0        V1 <--> V1
v 0.51     0.033047  15.433  0        V2 <--> V2
w 0.64     0.034422  18.593  0        V3 <--> V3
x 0.75     0.037129  20.200  0        V4 <--> V4

 Iterations =  13
```

6. Redoing the analysis for different sample sizes partly shows sample variability in the parameters, but also shows that the χ^2 statistic varies directly as a function of sample size.

```
> R100 <- sim.congeneric(N = 100)
> R200 <- sim.congeneric(N = 200)
> R400 <- sim.congeneric(N = 200)
> factanal(covmat = R100, factors = 1, n.obs = 100)

Call:
factanal(factors = 1, covmat = R100, n.obs = 100)

Uniquenesses:
  V1   V2   V3   V4
0.36 0.51 0.64 0.75

Loadings:
   Factor1
V1 0.8
V2 0.7
V3 0.6
V4 0.5

               Factor1
SS loadings      1.740
Proportion Var   0.435

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 0 on 2 degrees of freedom.
```

```
The p-value is 1
```

```
> factanal(covmat = R200, factors = 1, n.obs = 200)
```

```
Call:
factanal(factors = 1, covmat = R200, n.obs = 200)

Uniquenesses:
  V1   V2   V3   V4
0.36 0.51 0.64 0.75

Loadings:
   Factor1
V1 0.8
V2 0.7
V3 0.6
V4 0.5

               Factor1
SS loadings      1.740
Proportion Var   0.435

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 0 on 2 degrees of freedom.
The p-value is 1
```

```
> factanal(covmat = R400, factors = 1, n.obs = 400)
```

```
Call:
factanal(factors = 1, covmat = R400, n.obs = 400)

Uniquenesses:
  V1   V2   V3   V4
0.36 0.51 0.64 0.75

Loadings:
   Factor1
V1 0.8
V2 0.7
V3 0.6
V4 0.5

               Factor1
SS loadings      1.740
Proportion Var   0.435

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 0 on 2 degrees of freedom.
The p-value is 1
```

```
> sem1 = sem(model, R100, 100)
> sem2 = sem(model, R200, 200)
> sem4 = sem(model, R400, 400)
> summary(sem1)
```

```
 Model Chisquare =  1.8724e-10   Df =  2 Pr(>Chisq) = 1
 Chisquare (null model) =  88.65   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 =  1.0726
 Bentler CFI =  1
 SRMR =  6.4327e-07
 BIC =  -9.2103

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
9.45e-07 1.74e-06 2.95e-06 3.46e-06 4.67e-06 1.19e-05

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a 0.80     0.10223   7.8258  5.1070e-15 V1 <--- Latent
b 0.70     0.10279   6.8098  9.7764e-12 V2 <--- Latent
c 0.60     0.10410   5.7637  8.2292e-09 V3 <--- Latent
d 0.50     0.10668   4.6869  2.7731e-06 V4 <--- Latent
u 0.36     0.10858   3.3156  9.1445e-04 V1 <--> V1
v 0.51     0.10498   4.8582  1.1846e-06 V2 <--> V2
w 0.64     0.10935   5.8530  4.8284e-09 V3 <--> V3
x 0.75     0.11795   6.3589  2.0324e-10 V4 <--> V4

 Iterations =  13
```

```
> summary(sem2)
```

```
 Model Chisquare =  3.7636e-10   Df =  2 Pr(>Chisq) = 1
 Chisquare (null model) =  178.20   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 =  1.0348
 Bentler CFI =  1
 SRMR =  6.4327e-07
 BIC =  -10.597

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
1.34e-06 2.46e-06 4.19e-06 4.91e-06 6.62e-06 1.68e-05

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a 0.80     0.072102  11.0953 0.0000e+00 V1 <--- Latent
b 0.70     0.072503   9.6547 0.0000e+00 V2 <--- Latent
c 0.60     0.073425   8.1717 2.2204e-16 V3 <--- Latent
d 0.50     0.075244   6.6451 3.0309e-11 V4 <--- Latent
u 0.36     0.076583   4.7008 2.5915e-06 V1 <--> V1
v 0.51     0.074043   6.8879 5.6641e-12 V2 <--> V2
w 0.64     0.077125   8.2982 0.0000e+00 V3 <--> V3
x 0.75     0.083190   9.0155 0.0000e+00 V4 <--> V4

 Iterations =  13
```

```
> summary(sem4)

 Model Chisquare =  7.5461e-10   Df =  2 Pr(>Chisq) = 1
 Chisquare (null model) =  357.29   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 =  1.0171
 Bentler CFI =  1
 SRMR =  6.4327e-07
 BIC =  -11.983

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
1.90e-06 3.49e-06 5.93e-06 6.95e-06 9.38e-06 2.38e-05

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
a 0.80     0.050920  15.7109 0.0000e+00 V1 <--- Latent
b 0.70     0.051203  13.6710 0.0000e+00 V2 <--- Latent
c 0.60     0.051854  11.5710 0.0000e+00 V3 <--- Latent
d 0.50     0.053139   9.4093 0.0000e+00 V4 <--- Latent
u 0.36     0.054084   6.6563 2.8084e-11 V1 <--> V1
v 0.51     0.052291   9.7531 0.0000e+00 V2 <--> V2
w 0.64     0.054467  11.7502 0.0000e+00 V3 <--> V3
x 0.75     0.058751  12.7658 0.0000e+00 V4 <--> V4

 Iterations =  13
```

7. The implied correlation matrix from Figure 2 is found by path tracing.

	X1	X2	X3	X4
X1	1			
X2	ab	1		
X3	aec	bec	1	
X4	aed	bed	cd	1