Analysis of Unstructured Covariance Matrices with lavaan*

The Longitudinal IQ data are taken from *The Statistical Sleuth* by F. Ramsey and D. Schafer. The source is Marie Skodak and Harold M. Skeels, "A final follow-up study of one hundred adopted children," *Journal of Genetic Psychology*, 1949, 75, 85-125. Data are used here without permission. The article by Skodak and Skeels includes the complete data set, including variables not in the data set that appears in Ramsey and Schafer's book.

The variables used here (from Ramsey and Schafer) are

- Adoptive mother's education
- Birth mother's IQ
- Child's IQ at age 2
- Child's IQ at age 4
- Child's IQ at age 8
- Child's IQ at age 13

```
> # install.packages("lavaan", dependencies = TRUE) # Only need to do this once
> library(lavaan)
This is lavaan 0.6-15
lavaan is FREE software! Please report any bugs.
> longIQ =
read.table("https://www.utstat.toronto.edu/brunner/data/illegal/origIQ.data.txt")
> colnames(longIQ) = c("amEduc", "bmIQ", "IQ2", "IQ4", "IQ8", "IQ13")
> head(longIQ)
  amEduc bmIQ IQ2 IQ4 IQ8 IQ13
1     10  100 120 115 109  106
2     10   71 131 109 113   95
3     14   89 126 115 113   90
4      7   73 120 102 111  121
5     14   64 126 125 114   96
6      8   64 125 109  96   87
>
> dim(longIQ)
[1] 62  6
> r = round(cor(longIQ), 3); r
       amEduc  bmIQ    IQ2   IQ4   IQ8   IQ13
amEduc  1.000 0.070 -0.095 0.014 0.023 -0.004
bmIQ    0.070 1.000  0.042 0.272 0.362  0.380
IQ2    -0.095 0.042  1.000 0.509 0.476  0.300
IQ4     0.014 0.272  0.509 1.000 0.646  0.539
IQ8     0.023 0.362  0.476 0.646 1.000  0.774
IQ13   -0.004 0.380  0.300 0.539 0.774  1.000
> # Under H0: rho = 0, t = r * sqrt(n-2) / sqrt(1-r^2) ~ t(n-2)
> n = dim(longIQ)[1]; n
[1] 62
> critval = qt(0.975,n-2); critval
[1] 2.000298
> tstat = r * sqrt(n-2) / sqrt(1-r^2); round(tstat,3)
       amEduc  bmIQ    IQ2   IQ4   IQ8   IQ13
amEduc    Inf 0.544 -0.739 0.108 0.178 -0.031
bmIQ    0.544   Inf  0.326 2.189 3.008  3.182
IQ2    -0.739 0.326    Inf 4.580 4.193  2.436
IQ4     0.108 2.189  4.580   Inf 6.555  4.957
IQ8     0.178 3.008  4.193 6.555   Inf  9.469
IQ13   -0.031 3.182  2.436 4.957 9.469    Inf
```

```
> # In this model, the only parameters are variances and covariances
> # of the observable variables.
> covmod0 = "amEduc ~~ amEduc + bmIQ  + IQ2 +  IQ4 +  IQ8 + IQ13
+                      bmIQ ~~ bmIQ  + IQ2 +  IQ4 +  IQ8 + IQ13
+                              IQ2 ~~ IQ2 +  IQ4 +  IQ8 + IQ13
+                                     IQ4 ~~ IQ4 +  IQ8 + IQ13
+                                            IQ8 ~~ IQ8 + IQ13
+                                                   IQ13 ~~ IQ13"
> fit0 = lavaan(covmod0,data=longIQ)
>
> summary(fit0)
lavaan 0.6.13 ended normally after 222 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        21

  Number of observations                            62

Model Test User Model:

  Test statistic                                 0.000
  Degrees of freedom                                 0

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  amEduc ~~
    bmIQ              3.209    5.816    0.552    0.581
    IQ2              -3.542    4.755   -0.745    0.456
    IQ4               0.539    4.726    0.114    0.909
    IQ8               0.892    4.960    0.180    0.857
    IQ13             -0.157    5.582   -0.028    0.978
  bmIQ ~~
    IQ2               8.625   26.031    0.331    0.740
    IQ4              55.577   26.905    2.066    0.039
    IQ8              77.660   28.970    2.681    0.007
    IQ13             91.703   32.802    2.796    0.005
  IQ2 ~~
    IQ4              84.854   23.769    3.570    0.000
    IQ8              83.373   24.621    3.386    0.001
    IQ13             59.065   26.123    2.261    0.024
  IQ4 ~~
    IQ8             112.841   26.415    4.272    0.000
    IQ13            106.089   28.382    3.738    0.000
  IQ8 ~~
    IQ13            159.709   33.140    4.819    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    amEduc            8.315    1.493    5.568    0.000
    bmIQ            250.958   45.073    5.568    0.000
    IQ2             167.104   30.013    5.568    0.000
    IQ4             166.532   29.910    5.568    0.000
    IQ8             183.314   32.924    5.568    0.000
    IQ13            232.310   41.724    5.568    0.000


> # Okay, that's kind of reasonable. Try to make them correlations.
```

```
> fit0.1 = lavaan(covmod0, data=longIQ, std.ov=TRUE)
> # summary(fit0.1)

> fitted(fit0.1) # Sigma(thetahat)
$cov
       amEduc   bmIQ    IQ2    IQ4    IQ8   IQ13
amEduc  0.984
bmIQ    0.069  0.984
IQ2    -0.093  0.041  0.984
IQ4     0.014  0.267  0.500  0.984
IQ8     0.022  0.356  0.469  0.635  0.984
IQ13   -0.004  0.374  0.295  0.531  0.761  0.984

> r # For comparison
       amEduc  bmIQ    IQ2   IQ4   IQ8   IQ13
amEduc  1.000 0.070 -0.095 0.014 0.023 -0.004
bmIQ    0.070 1.000  0.042 0.272 0.362  0.380
IQ2    -0.095 0.042  1.000 0.509 0.476  0.300
IQ4     0.014 0.272  0.509 1.000 0.646  0.539
IQ8     0.023 0.362  0.476 0.646 1.000  0.774
IQ13   -0.004 0.380  0.300 0.539 0.774  1.000

> fitted(fit0.1)$cov * n/(n-1)

       amEduc   bmIQ    IQ2    IQ4    IQ8   IQ13
amEduc  1.000
bmIQ    0.070  1.000
IQ2    -0.095  0.042  1.000
IQ4     0.014  0.272  0.509  1.000
IQ8     0.023  0.362  0.476  0.646  1.000
IQ13   -0.004  0.380  0.300  0.539  0.774  1.000

> # So asymptotically it's fine, but I want the MLEs to be sample correlations.
> # Standardize the variables myself.
>
> standIQ = scale(longIQ) * sqrt(n/(n-1))
> fit0.2 = lavaan(covmod0, data=standIQ)
> fitted(fit0.2); r # Sigma(thetahat) should be correlation matrix
$cov
       amEduc   bmIQ    IQ2    IQ4    IQ8   IQ13
amEduc  1.000
bmIQ    0.070  1.000
IQ2    -0.095  0.042  1.000
IQ4     0.014  0.272  0.509  1.000
IQ8     0.023  0.362  0.476  0.646  1.000
IQ13   -0.004  0.380  0.300  0.539  0.774  1.000

       amEduc  bmIQ    IQ2   IQ4   IQ8   IQ13
amEduc  1.000 0.070 -0.095 0.014 0.023 -0.004
bmIQ    0.070 1.000  0.042 0.272 0.362  0.380
IQ2    -0.095 0.042  1.000 0.509 0.476  0.300
IQ4     0.014 0.272  0.509 1.000 0.646  0.539
IQ8     0.023 0.362  0.476 0.646 1.000  0.774
IQ13   -0.004 0.380  0.300 0.539 0.774  1.000
```

```
> # Now want names for parameters so I can test differences, etc.
>
> covmod1 =
+ "amEduc ~~ amEduc + rho12*bmIQ + rho13*IQ2 + rho14*IQ4 + rho15*IQ8 + rho16*IQ13
+                   bmIQ ~~ bmIQ + rho23*IQ2 + rho24*IQ4 + rho25*IQ8 + rho26*IQ13
+                                  IQ2 ~~ IQ2 + rho34*IQ4 + rho35*IQ8 + rho36*IQ13
+                                               IQ4 ~~ IQ4 + rho45*IQ8 + rho46*IQ13
+                                                            IQ8 ~~ IQ8 + rho56*IQ13
+                                                                         IQ13 ~~ IQ13"
> fit1 = lavaan(covmod1, data=standIQ) # This will be the full model.
> parameterEstimates(fit1)
      lhs op    rhs label    est    se      z pvalue ci.lower ci.upper
1  amEduc ~~ amEduc        1.000 0.180  5.568  0.000    0.648    1.352
2  amEduc ~~   bmIQ rho12  0.070 0.127  0.552  0.581   -0.179    0.320
3  amEduc ~~    IQ2 rho13 -0.095 0.128 -0.745  0.456   -0.345    0.155
4  amEduc ~~    IQ4 rho14  0.014 0.127  0.114  0.909   -0.234    0.263
5  amEduc ~~    IQ8 rho15  0.023 0.127  0.180  0.857   -0.226    0.272
6  amEduc ~~   IQ13 rho16 -0.004 0.127 -0.028  0.978   -0.252    0.245
7    bmIQ ~~   bmIQ        1.000 0.180  5.568  0.000    0.648    1.352
8    bmIQ ~~    IQ2 rho23  0.042 0.127  0.331  0.740   -0.207    0.291
9    bmIQ ~~    IQ4 rho24  0.272 0.132  2.066  0.039    0.014    0.530
10   bmIQ ~~    IQ8 rho25  0.362 0.135  2.681  0.007    0.097    0.627
11   bmIQ ~~   IQ13 rho26  0.380 0.136  2.796  0.005    0.114    0.646
12    IQ2 ~~    IQ2        1.000 0.180  5.568  0.000    0.648    1.352
13    IQ2 ~~    IQ4 rho34  0.509 0.142  3.570  0.000    0.229    0.788
14    IQ2 ~~    IQ8 rho35  0.476 0.141  3.386  0.001    0.201    0.752
15    IQ2 ~~   IQ13 rho36  0.300 0.133  2.261  0.024    0.040    0.560
16    IQ4 ~~    IQ4        1.000 0.180  5.568  0.000    0.648    1.352
17    IQ4 ~~    IQ8 rho45  0.646 0.151  4.272  0.000    0.350    0.942
18    IQ4 ~~   IQ13 rho46  0.539 0.144  3.738  0.000    0.257    0.822
19    IQ8 ~~    IQ8        1.000 0.180  5.568  0.000    0.648    1.352
20    IQ8 ~~   IQ13 rho56  0.774 0.161  4.819  0.000    0.459    1.089
21   IQ13 ~~   IQ13        1.000 0.180  5.568  0.000    0.648    1.352
> # fitted(fit1); r # Checks

> # Likelihood ratio test of relationship between adoptive mom's education
> # and kid's IQ at any age
> noAM = lavaan(covmod1, data=standIQ, constraints = "rho13==0
+                                                     rho14==0
+                                                     rho15==0
+                                                     rho16==0")
> anova(noAM,fit1)
Chi-Squared Difference Test

     Df    AIC     BIC  Chisq Chisq diff RMSEA Df diff Pr(>Chisq)
fit1  0 970.22 1014.89 0.0000
noAM  4 963.43  999.59 1.2108     1.2108     0       4     0.8763
>
> # Likelihood ratio test of relationship between birth mom's IQ
> # and kid's IQ at any age
>
> noBM = lavaan(covmod1, data=standIQ, constraints = "rho23==0
+                                                     rho24==0
+                                                     rho25==0
+                                                     rho26==0")
> anova(noBM,fit1)

Chi-Squared Difference Test

     Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
fit1  0 970.22 1014.9  0.000
noBM  4 974.47 1010.6 12.256     12.256 0.18246       4    0.01555 *
---
Signif. Codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> # Multivariate regression yields p = 0.022
```

```
> # Likelihood ratio test: Is birth mother's IQ correlated with child's IQ
> # differently at different ages?
>
> BMsame = lavaan(covmod1, data=standIQ, constraints = "rho23==rho24
+                                                       rho24==rho25
+                                                       rho25==rho26")
> anova(BMsame,fit1)

Chi-Squared Difference Test

       Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
fit1    0 970.22 1014.9 0.0000
BMsame  3 970.81 1009.1 6.5956     6.5956 0.13904       3    0.08597 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> # This is surprising. Try a Wald test
> # For Wald tests: Wtest = function(L,Tn,Vn,h=0) # H0: L theta = h
> source("https://www.utstat.toronto.edu/brunner/openSEM/fun/Wtest.txt")
>
> Vhat = vcov(fit1) # Asymptotic covariance matrix of the estimates
> rhohat = coef(fit1); rhohat
amEduc~~amEduc          rho12          rho13          rho14          rho15          rho16
         1.000          0.070         -0.095          0.014          0.023         -0.004
    bmIQ~~bmIQ          rho23          rho24          rho25          rho26       IQ2~~IQ2
         1.000          0.042          0.272          0.362          0.380          1.000
         rho34          rho35          rho36       IQ4~~IQ4          rho45          rho46
         0.509          0.476          0.300          1.000          0.646          0.539
      IQ8~~IQ8          rho56     IQ13~~IQ13
         1.000          0.774          1.000
> length(rhohat)
[1] 21
> rbind(names(rhohat))
     [,1]           [,2]    [,3]    [,4]    [,5]    [,6]    [,7]          [,8]    [,9]
[1,] "amEduc~~amEduc" "rho12" "rho13" "rho14" "rho15" "rho16" "bmIQ~~bmIQ" "rho23" "rho24"
     [,10]   [,11]   [,12]      [,13]   [,14]   [,15]   [,16]      [,17]   [,18]   [,19]
[1,] "rho25" "rho26" "IQ2~~IQ2" "rho34" "rho35" "rho36" "IQ4~~IQ4" "rho45" "rho46"
"IQ8~~IQ8"
     [,20]   [,21]
[1,] "rho56" "IQ13~~IQ13"
>
> L0 = matrix(0,3,21)
> L0[1,8]  = 1;  L0[1,9]  = -1 # rho23==rho24
> L0[2,9]  = 1;  L0[2,10] = -1 # rho24==rho25
> L0[3,10] = 1;  L0[3,11] = -1 # rho25==rho26
>
> Wtest(L0,rhohat,Vhat)
       W       df  p-value
5.683357 3.000000 0.128074

> # Just test 2 years old versus 13
>
> Two_vs_13 = matrix(0,1,21); Two_vs_13[1,8] = 1; Two_vs_13[1,11] = -1
> Wtest(Two_vs_13,rhohat,Vhat)
         W         df    p-value
4.66802402 1.00000000 0.03072926

>
> # However, Bonferroni correcting for 6 pairwise tests, need p < ...
> 0.05/6
[1] 0.008333333

> # One could bootstrap, but IQ scores are well known to be normally distributed.
```