27/28, great work!

Lab #3

Ahmad Saquib Sina and Bo Ra Joo

February 28, 2019

```
library(lavaan)

## This is lavaan 0.6-3

## lavaan is BETA software! Please report any bugs.
```

Question one

In this model, Family is an exogenous variable. This variable's variance is assumed to be caused entirely by variables not in the causal model. On the other hand, the rest of the variables, Ability, Motive, Course, and Achieve are endogenous variables because their variance is considered to be explained in part by other variables in the model. From the model, it is apparent that there is no unanalyzed effect because there is no bi-direction path between any two variables. From the conceptual model, we have noticed that Family has a direct causal effect on Achieve. Family has seven indirect effects on Achieve, and they are c*j, d*g, a*b*f*g, a*h, a*b*j, a*e*g, c*f*g. There is no spurious effect of Family on Achieve. Ability has a direct effect on Achieve. It has also indirect effects, and they are e*g and b*j. There is a spurious effect of Ability on Achieve because Family causes both Ability and Achieve. Course has a direct causal effect on Achieve. It does not have indirect effects on Achieve. However, there are spurious effects due to Ability, Family, and Motive. In the case of Motive, there is a direct causal effect on Achieve. Also, there is an indirect effect, f*g. There is spurious effects due to Family and Ability.

Question two

Family is an exogenous variable and the other reamining variables that are Ability, Motive, Course, and Achieve are endogenous variables.

Question three

Good

This is a just identified model. In a just-identified model, there is a direct path from each variable to each other variable. In the conceptual model, we can see that there is a direct path from each variable to each other variable. The fit between the data and the model looks good. The decomposed correlations for this model can be used to “reproduce" the original correlations perfectly. Therefore, it is a just-identified model. The values of RMSRA, CFI, TFI, and SRMR are 0, 1, 1, and 0, respectively, which represents a good global fit of the model.

-0.5 Not really: those measures are meaningless in a just-identified model because as you note, the are reproducing the original correlations perfectly and will always be that. We can't evaluate the global fit of the model because we have no residual variation to explain

Question four

There are seven indirect effects of Family on Achieve. These indirect effects are *cj*, *dg*, a*bfg*, *a*h, a*b*j, *aeg*, and *cfg*.

Question five

```
lower = '
    1
   .417       1
   .190     .205        1
   .372     .498      .375      1
   .417     .737      .255     .615     1 '

beth = getCov(lower, names=c("Family","Ability", "Motive",
                      "Course","Achieve"))
print(beth)
```

Nice name :-)

```
##          Family Ability Motive Course Achieve
## Family   1.000   0.417  0.190  0.372   0.417
## Ability  0.417   1.000  0.205  0.498   0.737
## Motive   0.190   0.205  1.000  0.375   0.255
## Course   0.372   0.498  0.375  1.000   0.615
## Achieve  0.417   0.737  0.255  0.615   1.000
```

Question six

```
mod <- '

# structural paths

Course ~ d*Family + e*Ability+f*Motive
Motive ~ c * Family + b* Ability
Ability ~ a * Family
Achieve ~ i * Family + h * Ability+ j*Motive + g*Course

# indirect effects
cj := c*j
dg := d*g
abfg := a*b*f*g
ah := a *h
abj := a*b*j
aeg := a*e*g
cfg := c*f*g
'

fit = sem(model = mod, sample.cov=beth, sample.nobs = 250)
summary(fit, fit.measures=TRUE,standardized=TRUE)

## lavaan 0.6-3 ended normally after 18 iterations
##
##   Optimization method                           NLMINB
##   Number of free parameters                         14
```

```
##
##   Number of observations                           250
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       0.000
##   Degrees of freedom                                0
##
## Model test baseline model:
##
##   Minimum Function Test Statistic              416.751
##   Degrees of freedom                               10
##   P-value                                        0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    1.000
##   Tucker-Lewis Index (TLI)                       1.000
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1562.793
##   Loglikelihood unrestricted model (H1)      -1562.793
##
##   Number of free parameters                         14
##   Akaike (AIC)                                3153.585
##   Bayesian (BIC)                              3202.885
##   Sample-size adjusted Bayesian (BIC)         3158.504
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.000
##   90 Percent Confidence Interval          0.000  0.000
##   P-value RMSEA <= 0.05                             NA
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.000
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   Course ~
##     Family     (d)    0.165    0.057    2.919    0.004    0.165    0.165
##     Ability    (e)    0.374    0.057    6.597    0.000    0.374    0.374
##     Motive     (f)    0.267    0.053    5.079    0.000    0.267    0.267
```

```
##   Motive ~
##     Family     (c)    0.127    0.068    1.871    0.061    0.127    0.127
##     Ability    (b)    0.152    0.068    2.251    0.024    0.152    0.152
##   Ability ~
##     Family     (a)    0.417    0.057    7.254    0.000    0.417    0.417
##   Achieve ~
##     Family     (i)    0.069    0.043    1.601    0.109    0.069    0.069
##     Ability    (h)    0.551    0.046   11.879    0.000    0.551    0.551
##     Motive     (j)    0.013    0.042    0.302    0.763    0.013    0.013
##     Course     (g)    0.310    0.048    6.498    0.000    0.310    0.310
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Course            0.649    0.058   11.180    0.000    0.649    0.652
##    .Motive            0.941    0.084   11.180    0.000    0.941    0.945
##    .Ability           0.823    0.074   11.180    0.000    0.823    0.826
##    .Achieve           0.370    0.033   11.180    0.000    0.370    0.371
##
## Defined Parameters:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##     cj                0.002    0.005    0.298    0.766    0.002    0.002
##     dg                0.051    0.019    2.663    0.008    0.051    0.051
##     abfg              0.005    0.003    1.894    0.058    0.005    0.005
##     ah                0.230    0.037    6.191    0.000    0.230    0.230
##     abj               0.001    0.003    0.299    0.765    0.001    0.001
##     aeg               0.048    0.012    3.902    0.000    0.048    0.048
##     cfg               0.010    0.006    1.695    0.090    0.010    0.010
```

```r
standardizedSolution(fit, ci = TRUE)
```

```
##        lhs op     rhs est.std    se      z pvalue ci.lower ci.upper
## 1   Course  ~  Family   0.165 0.056  2.960  0.003    0.056    0.275
## 2   Course  ~ Ability   0.374 0.054  6.939  0.000    0.269    0.480
## 3   Course  ~  Motive   0.267 0.051  5.201  0.000    0.166    0.367
## 4   Motive  ~  Family   0.127 0.067  1.889  0.059   -0.005    0.258
## 5   Motive  ~ Ability   0.152 0.067  2.273  0.023    0.021    0.284
## 6  Ability  ~  Family   0.417 0.050  8.353  0.000    0.319    0.515
## 7  Achieve  ~  Family   0.069 0.043  1.603  0.109   -0.015    0.154
## 8  Achieve  ~ Ability   0.551 0.042 12.969  0.000    0.468    0.634
## 9  Achieve  ~  Motive   0.013 0.042  0.302  0.763   -0.069    0.094
## 10 Achieve  ~  Course   0.310 0.047  6.565  0.000    0.217    0.403
## 11  Course ~~  Course   0.652 0.048 13.590  0.000    0.558    0.746
## 12  Motive ~~  Motive   0.945 0.028 33.834  0.000    0.890    0.999
## 13 Ability ~~ Ability   0.826 0.042 19.841  0.000    0.745    0.908
## 14 Achieve ~~ Achieve   0.371 0.037 10.090  0.000    0.299    0.443
## 15  Family ~~  Family   1.000 0.000     NA     NA    1.000    1.000
## 16      cj :=     c*j   0.002 0.005  0.298  0.765   -0.009    0.012
## 17      dg :=     d*g   0.051 0.019  2.684  0.007    0.014    0.089
## 18    abfg := a*b*f*g   0.005 0.003  1.926  0.054    0.000    0.011
## 19      ah :=     a*h   0.230 0.033  6.975  0.000    0.165    0.294
```

```
## 20     abj :=   a*b*j   0.001 0.003  0.299  0.765   -0.004    0.006
## 21     aeg :=   a*e*g   0.048 0.012  4.179  0.000    0.026    0.071
## 22     cfg :=   c*f*g   0.010 0.006  1.700  0.089   -0.002    0.023
```

Question seven

1) Interpretation for all the direct paths to Achieve

Each one standard deviation increase or difference in Family is associated with a 0.069 standard deviation predicted increase or difference in Achieve after controlling for the difference of Ability, Motive and Course. Also, each one standard deviation increase or difference in Ability is associated with a 0.551 standard deviation predicted increase or difference in Achieve after controlling for the difference of Family, Motive, and Course. Each one standard deviation increase or difference in Motive is associated with a 0.013 standard deviation predicted increase or difference in Achieve after controlling for the difference of Ability, Family and Course. Also, each one standard deviation increase or difference in Course is associated with a 0.310 standard deviation predicted increase or difference in Achieve after controlling for the difference of Ability, Motive and Family.

We can also put in this way; A direct causal effect of Family on Achieve is 0.069. A direct causal effect of Ability on Achieve is 0.551. A direct causal effect of Motive on Achieve is 0.013. A direct causal effect of Course on Achieve is 0.310.

2) Interpretation for the indirect paths from Family on Achieve

Family has an indirect causal effect on Achieve through Motive, and the effect is 0.002. Family has an indirect causal effect on Achieve through, Course and the effect is 0.051. Family has an indirect causal effect on Achieve through Motive, Ability, and Course and the effect is 0.005. Family has an indirect causal effect on Achieve through Ability, and the effect is 0.230. Family has an indirect causal effect on Achieve through Motive and Ability, and the effect is 0.001. Family has an indirect causal effect on Achieve through Ability and Course, and the effect is 0.048. Family has an indirect causal effect on Achieve through Motive and Course, and the effect is 0.010.

Question eight

No, there is no non-causal effect (neither spurious nor correlational) of family on achieve in the model. The reported value is zero.

Question nine

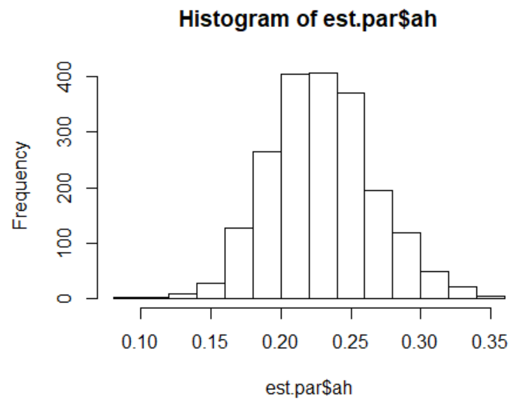
The bivariate correlation between Family and Achieve is 0.069+0.002+0.051+0.005+0.230+0.001+0.048+0.010= 0.416.

Question ten

```r
set.seed (123512)
# print the parameter estimates for fit
parameterEstimates(fit)[, ]

##        lhs op     rhs label   est    se      z pvalue ci.lower ci.upper
## 1   Course  ~  Family     d 0.165 0.057  2.919  0.004    0.054    0.276
```

```
## 2   Course  ~ Ability     e 0.374 0.057  6.597  0.000    0.263    0.486
## 3   Course  ~  Motive     f 0.267 0.053  5.079  0.000    0.164    0.370
## 4   Motive  ~  Family     c 0.127 0.068  1.871  0.061   -0.006    0.259
## 5   Motive  ~ Ability     b 0.152 0.068  2.251  0.024    0.020    0.285
## 6  Ability  ~  Family     a 0.417 0.057  7.254  0.000    0.304    0.530
## 7  Achieve  ~  Family     i 0.069 0.043  1.601  0.109   -0.016    0.155
## 8  Achieve  ~ Ability     h 0.551 0.046 11.879  0.000    0.460    0.642
## 9  Achieve  ~  Motive     j 0.013 0.042  0.302  0.763   -0.069    0.094
## 10 Achieve  ~  Course     g 0.310 0.048  6.498  0.000    0.216    0.404
## 11  Course ~~  Course       0.649 0.058 11.180  0.000    0.536    0.763
## 12  Motive ~~  Motive       0.941 0.084 11.180  0.000    0.776    1.106
## 13 Ability ~~ Ability       0.823 0.074 11.180  0.000    0.679    0.967
## 14 Achieve ~~ Achieve       0.370 0.033 11.180  0.000    0.305    0.434
## 15  Family ~~  Family       0.996 0.000     NA     NA    0.996    0.996
## 16      cj :=     c*j    cj 0.002 0.005  0.298  0.766   -0.009    0.012
## 17      dg :=     d*g    dg 0.051 0.019  2.663  0.008    0.014    0.089
## 18    abfg := a*b*f*g  abfg 0.005 0.003  1.894  0.058    0.000    0.011
## 19      ah :=     a*h    ah 0.230 0.037  6.191  0.000    0.157    0.303
## 20     abj :=   a*b*j   abj 0.001 0.003  0.299  0.765   -0.004    0.006
## 21     aeg :=   a*e*g   aeg 0.048 0.012  3.902  0.000    0.024    0.073
## 22     cfg :=   c*f*g   cfg 0.010 0.006  1.695  0.090   -0.002    0.023

# you are looking for the row numbers that correspond to the indirect
effects.
first <-    16
second <-   22
# We want to extract and save their estimates. The "est" part.

# number of bootstrap samples to take
k <- 2000
# sample size
n <- 250

# number of indirect effects you calculated above
effs <-   7

# this initializes a matrix that will save your bootstrapped samples (each
row)
# and the parameter estimates (each column).

est.par <- matrix(nrow = k,
                  ncol = effs)

for(i in 1:k){
  # generate a multivariate normal random sample
  tmp <- MASS::mvrnorm(n, mu = rep(0, ncol(beth)), Sigma = beth)

  # convert it to a data.frame
  tmp <- as.data.frame(tmp)
```

```
  # estimate the model
  # assuming the model you defined earlier is named mod
  fit.tmp <-sem(mod, data = tmp)

 # save the parameter estimates
  est.par[i, ] <- parameterEstimates(fit.tmp)[first:second, "est"]
}

# convert the matrix to a data.frame
est.par <- as.data.frame(est.par)

# name your parameters
colnames(est.par) <- parameterEstimates(fit.tmp)[first:second, "label"]

# print the first 6 rows
head(est.par)

##              cj         dg         abfg        ah           abj        aeg
## 1  0.0059437078 0.07473977  0.005631903 0.2616177  3.333021e-03 0.05354487
## 2 -0.0161970201 0.09694654  0.009175846 0.2273390 -8.552168e-03 0.07076285
## 3  0.0008302763 0.01753844  0.001679555 0.3134670  2.067139e-04 0.03468491
## 4  0.0021629328 0.03028502 -0.001305126 0.2730043 -3.735303e-04 0.05007310
## 5  0.0001390439 0.04126595  0.002150049 0.2808910  1.033229e-05 0.07198522
## 6  0.0036530233 0.03864671  0.007154541 0.2368057  4.018516e-03 0.04265737
##           cfg
## 1 0.010043256
## 2 0.017378208
## 3 0.006746012
## 4 0.007557349
## 5 0.028933676
## 6 0.006503820

# lets say one of your indirect effects is named ah
# Hint, one of your indirect effects should be ah :D
# plot it!
hist(est.par$ah)
```




```r
# extract the CIs
quantile(est.par$ah, probs = c(.025, .975))

##      2.5%     97.5%
## 0.1622315 0.3073054

# you may find this piece of code very helpful, too (HINT)
apply(est.par, 2, quantile, probs = c(.025, .975))

##                cj         dg         abfg        ah          abj
## 2.5%  -0.01083271 0.01601227 0.0006015997 0.1622315 -0.004994350
## 97.5%  0.01497490 0.09298865 0.0121747747 0.3073054  0.007415488
##              aeg           cfg
## 2.5%  0.02679996 -3.847253e-05
## 97.5% 0.07556610  2.394449e-02
```

Question 11

-0.5 Please also write out the confidence intervals rather than only the raw output, since we were asking for those in the question -- this helps provide evidence of you understanding the output

Based on the answer from Question 10, there is an evidence of mediation. For example, we have found that confidence intervals for dg, ah, and aeg all do not contain 0, which show the significance of mediation. In other words, Course, Ability, and Ability and Course play the mediation roles between Family and Achieve. However, cj, abfg, abj, and cfg are not statistically significant to see their confidence intervals contain 0.

Yes, the causal link between Family and Achieve goes through other variables, and they are Course, Ability and both Course and Ability. There is no direct effect of Family on Achieve since p value (0.109) is greater than a significance level, which means the impact is not statistically significant.

Question 12

In my view, the causal model seems reasonable. Family has indirect causal effects on Achieve; the causal relation of Family on Achieve goes through other variables such as Course and Ability and both Course and Ability. It is apparent that this model seems reasonable as we get the aforementioned indirect causal effects of family on achieve. However, one thing to point out is that we have found the complete mediation from this

Good point!

model since there is no significant direct effect of Family on Achieve. Also, we did not find the mediation effect of Motive on Achieve. Therefore, in my view, this model could include other influential exogenous variables based on literature reviews to analyze their direct effects on Achieve, which would make the model more reasonable.