Answers to exercises LVM session 1

Exercise 2.1

a) Load packages and data:

```
library("lavaan")
Math.data <- read.table("MathHmwk.txt", header = TRUE)
```

Fit a linear regression model:

```
Math.lm <- lm(MathAchievement ~ MathHomework, data = Math.data)
Math.lm
```

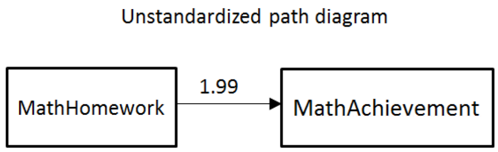
```
##
## Call:
## lm(formula = MathAchievement ~ MathHomework, data = Math.data)
##
## Coefficients:
##  (Intercept)  MathHomework
##        47.03          1.99
```

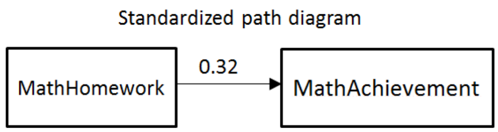
The correlation between `MathHomework` and `MathAchievement`:

```
obs_cor <- coef(Math.lm)[2] * sd(Math.data$MathHomework) / sd(Math.data$MathAchievement)
obs_cor
```

```
## MathHomework
##    0.3199936
```

b) The standardized and unstandardized path diagrams of the regression look as follows:





The path diagrams would be even more complete, if a variance for Math Homework and a residual variance for Math Achievement would have been included.

c) Fit the SEM model:

```r
Math.model <- '
  MathAchievement ~ MathHomework
'
Math.fit <- sem(Math.model, data = Math.data)
summary(Math.fit, standardized = TRUE)
```

```
## lavaan 0.6-18 ended normally after 1 iteration
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         2
##
##   Number of observations                           100
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   MathAchievement ~
##     MathHomework       1.990    0.589    3.378    0.001    1.990    0.320
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .MathAchievemnt  113.190   16.007    7.071    0.000  113.190    0.898
```

d) The results in a) and d) give the same standardized and unstandardized parameter estimates.

Exercise 2.2

a) The exogenous variables is Race. The endogenous variables are SES, Cognitive Abilities, School Type and Academic Achievement.

b) Supply the correlation matrix:

```r
HighSchool.cor <- lav_matrix_lower2full(c(
  1.000,
  0.178, 1.000,
  0.230, 0.327, 1.000,
  0.106, 0.245, 0.183, 1.000,
  0.195, 0.356, 0.721, 0.178, 1.000
```

```r
))
rownames(HighSchool.cor) <- colnames(HighSchool.cor) <-
  c("Race", "SES", "CognAb", "SchoolTy", "AcAch")
```

Specify the model:

```r
HighSchool.mod <- '
  CognAb ~ Race + SES
  SES ~ Race
  SchoolTy ~ SES + Race + CognAb
  AcAch ~ Race + SES + SchoolTy + CognAb
'
HighSchool.fit <- sem(HighSchool.mod, sample.cov = HighSchool.cor,
                      sample.nobs = 18058)
summary(HighSchool.fit, standardized = TRUE)
```

```
## lavaan 0.6-18 ended normally after 1 iteration
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        14
##
##   Number of observations                         18058
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   CognAb ~
##     Race              0.177    0.007   25.260    0.000    0.177    0.177
##     SES               0.295    0.007   42.061    0.000    0.295    0.295
##   SES ~
##     Race              0.178    0.007   24.308    0.000    0.178    0.178
##   SchoolTy ~
##     SES               0.202    0.008   26.499    0.000    0.202    0.202
##     Race              0.046    0.007    6.151    0.000    0.046    0.046
##     CognAb            0.106    0.008   13.804    0.000    0.106    0.106
##   AcAch ~
##     Race              0.015    0.005    2.935    0.003    0.015    0.015
##     SES               0.128    0.005   23.336    0.000    0.128    0.128
##     SchoolTy          0.022    0.005    4.191    0.000    0.022    0.022
##     CognAb            0.671    0.005  122.482    0.000    0.671    0.671
##
## Variances:
```

```
##                     Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .CognAb             0.863    0.009   95.021    0.000    0.863    0.863
##    .SES                0.968    0.010   95.021    0.000    0.968    0.968
##    .SchoolTy           0.926    0.010   95.021    0.000    0.926    0.926
##    .AcAch              0.463    0.005   95.021    0.000    0.463    0.463
```

c) Schooltype has a (very) small positive effect on Academic Achievement, but it is statistically significant (because the sample size is very large).

d) The zero-order, bivariate correlation between Race and Academic Achievement was .195.

e) The standardized direct effect of Race on Academic Achievement in the model we fitted is .015 (but with $> 18,000$ observations, still significant).

f) The SEM model 'explains' the observed correlation between Race and Academic Achievement through indirect effects via SES, Cognitive Abilities and School Type.

g) To assume that Race 'causes' Academic Achievement or Cognitive Ability is probably ridiculous. There are probably additional social, economic or other factors at play, that can explain this association, but that are not included in the model. Thus, our model does not provide a complete explanation of reality, but does provide insight into how these variables are associated in the real world, and how some associations can be in large part explained by other variables.

Exercise 2.3

a) The endogenous variables are Social Climate, Material Covered and Student Achievement. The exogenous variable in the model is Teacher Expectations.

We read in the covariance matrix:

```
MacKinnon.cov <- lav_matrix_lower2full(c(
  84.85,
  71.28, 140.34,
  18.83,  -6.25, 72.92,
  60.05,  84.54, 37.18, 139.48
))
rownames(MacKinnon.cov) <- colnames(MacKinnon.cov) <-
  c("TeachExp", "SocCli", "MatCov", "StudAch")
```

b) The original bivariate correlation between Teacher Expectations and Student Achievement is 0.552:

```
cov2cor(MacKinnon.cov)
```

```
##           TeachExp      SocCli      MatCov   StudAch
## TeachExp 1.0000000  0.65320730  0.23938728 0.5519899
## SocCli   0.6532073  1.00000000 -0.06178257 0.6042484
## MatCov   0.2393873 -0.06178257  1.00000000 0.3686635
## StudAch  0.5519899  0.60424844  0.36866345 1.0000000
```

c) Read in the covariance matrix:

```
MacKinnon.cov <- lav_matrix_lower2full(c(
  84.85,
  71.28, 140.34,
  18.83,  -6.25, 72.92,
  60.05,  84.54, 37.18, 139.48
))
rownames(MacKinnon.cov) <- colnames(MacKinnon.cov) <-
  c("TeachExp", "SocCli", "MatCov", "StudAch")
```

Write the syntax:

```
MacKinnon.mod <- '
  SocCli ~ a1*TeachExp
  MatCov ~ a2*TeachExp
  StudAch ~ TeachExp + b1*SocCli + b2*MatCov
  a1b1 := a1*b1
  a2b2 := a2*b2
'
```

Fit the model to the data:

```r
MacKinnon.fit <- sem(MacKinnon.mod, sample.cov = MacKinnon.cov,
                     sample.nobs = 40, fixed.x = FALSE)
summary(MacKinnon.fit, standardized = TRUE)
```

```
## lavaan 0.6-18 ended normally after 2 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         9
##
##   Number of observations                            40
##
## Model Test User Model:
##
##   Test statistic                                 3.687
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.055
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   SocCli ~
##     TeachExp  (a1)    0.840    0.154    5.456    0.000    0.840    0.653
##   MatCov ~
##     TeachExp  (a2)    0.222    0.142    1.559    0.119    0.222    0.239
##   StudAch ~
##     TeachExp          0.112    0.186    0.603    0.546    0.112    0.084
##     SocCli    (b1)    0.569    0.142    4.006    0.000    0.569    0.545
##     MatCov    (b2)    0.530    0.154    3.446    0.001    0.530    0.366
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .SocCli           78.448   17.542    4.472    0.000   78.448    0.573
##    .MatCov           67.023   14.987    4.472    0.000   67.023    0.943
##    .StudAch          63.323   14.159    4.472    0.000   63.323    0.425
##     TeachExp         82.729   18.499    4.472    0.000   82.729    1.000
##
## Defined Parameters:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##     a1b1              0.478    0.148    3.229    0.001    0.478    0.356
##     a2b2              0.118    0.083    1.421    0.155    0.118    0.088
```

The direct effect of Teacher Expectations on Student Achievement is $\hat{b} = 0.112$, the standardized value is 0.084, and it is not significant. The original bivariate association is mostly explained by the indirect effect through Social Climate. There thus seems to be mediation.

Note that `lavaan` by default does not include (co)variances of exogenous variables in the output. They are assumed equal to the sample values, but not explicitly included as model parameters. To explicitly include (co)variances of exogenous variables as parameters in our model, we need to override the default specification, by adding `fixed.x = FALSE` (`fixed.x = TRUE` is the default). This will not affect the parameter estimates or the fit of our model, but merely yields more parameter estimates in the output.

d) We request the $\hat{\beta}$ and $\hat{\psi}$ matrices as follows:

```
beta <- inspect(MacKinnon.fit, "coef")$beta
beta
```

```
##          SocCli MatCov StdAch TchExp
## SocCli    0.000   0.00      0  0.840
## MatCov    0.000   0.00      0  0.222
## StudAch   0.569   0.53      0  0.112
## TeachExp  0.000   0.00      0  0.000
```

```
psi <- inspect(MacKinnon.fit, "coef")$psi
psi
```

```
##          SocCli MatCov StdAch TchExp
## SocCli   78.448
## MatCov    0.000 67.023
## StudAch   0.000  0.000 63.323
## TeachExp  0.000  0.000  0.000 82.729
```