Structural Equation Models 2019

Week 2

Exercise 2.1

The null hypothesis is that self-concept (SC) is a multidimensional construct composed of four factors:

- General SC (GSC)
- Academic SC (ASC)
- English SC (ESC)
- Mathematics SC (MSC)

Alternative hypothesis is that self-concept is not a multidimensional construct composed of four factors.

Let us bring the data in R and prepare it for analysis:

```
library(lavaan)
```

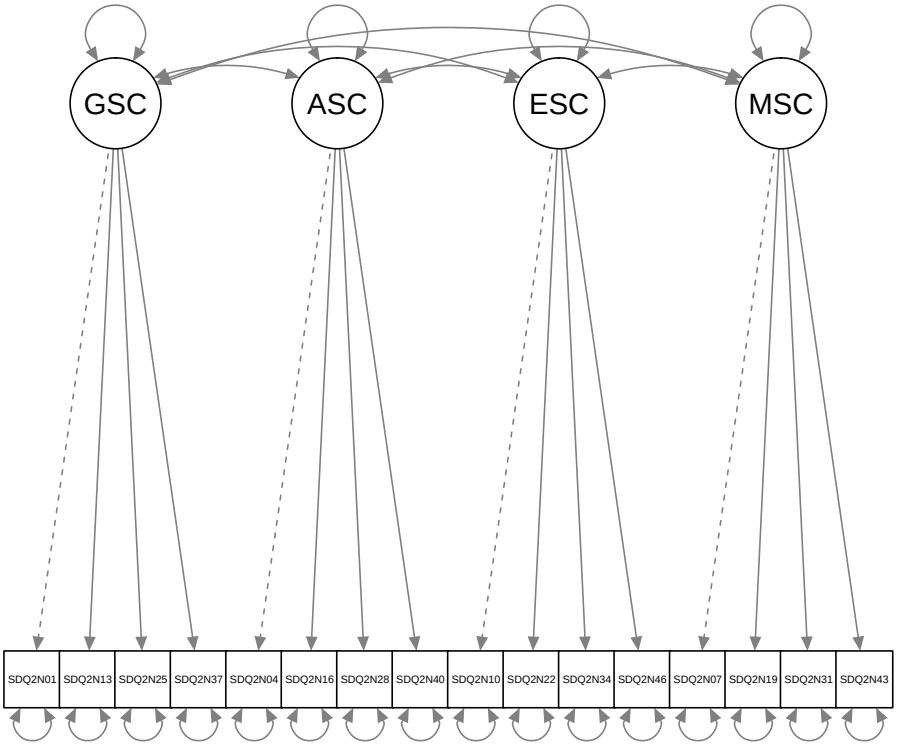
```
## This is lavaan 0.6-3
```

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

```
library(semPlot)

ex2.1 <- read.fortran(file = "ASC7INDM.DAT", c("40F1.0","X","6F2.0"))
colnames(ex2.1) <- c("SPPCN08", "SPPCN18", "SPPCN28", "SPPCN38", "SPPCN48",
                     "SPPCN58", "SPPCN01", "SPPCN11" ,"SPPCN21", "SPPCN31",
                     "SPPCN41", "SPPCN51", "SPPCN06", "SPPCN16", "SPPCN26",
                     "SPPCN36", "SPPCN46", "SPPCN56", "SPPCN03", "SPPCN13",
                     "SPPCN23", "SPPCN33", "SPPCN43", "SPPCN53", "SDQ2N01",
                     "SDQ2N13", "SDQ2N25", "SDQ2N37", "SDQ2N04", "SDQ2N16",
                     "SDQ2N28", "SDQ2N40", "SDQ2N10", "SDQ2N22", "SDQ2N34",
                     "SDQ2N46", "SDQ2N07", "SDQ2N19", "SDQ2N31", "SDQ2N43",
                     "MASTENG1", "MASTMAT1", "TENG1", "TMAT1", "SENG1", "SMAT1")
```

Then let us specify the model, visualize the model structure:

```
modelF4 <- "
GSC =~ SDQ2N01 + SDQ2N13 + SDQ2N25 + SDQ2N37
ASC =~ SDQ2N04 + SDQ2N16 + SDQ2N28 + SDQ2N40
ESC =~ SDQ2N10 + SDQ2N22 + SDQ2N34 + SDQ2N46
MSC =~ SDQ2N07 + SDQ2N19 + SDQ2N31 + SDQ2N43
"

fitF4 <- cfa(modelF4, data = ex2.1)

semPaths(fitF4, layout='tree2')
```




```
summary(fitF4, fit.measures = T)
```

```
## lavaan 0.6-3 ended normally after 49 iterations
##
##   Optimization method                           NLMINB
##   Number of free parameters                         38
##
##   Number of observations                           265
##
##   Estimator                                         ML
##   Model Fit Test Statistic                     159.112
##   Degrees of freedom                                98
##   P-value (Chi-square)                           0.000
##
## Model test baseline model:
##
##   Minimum Function Test Statistic            1703.155
##   Degrees of freedom                              120
##   P-value                                        0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    0.961
##   Tucker-Lewis Index (TLI)                       0.953
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -6562.678
##   Loglikelihood unrestricted model (H1)      -6483.122
##
```

```
##   Number of free parameters                         38
##   Akaike (AIC)                                13201.356
##   Bayesian (BIC)                              13337.386
##   Sample-size adjusted Bayesian (BIC)         13216.905
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.049
##   90 Percent Confidence Interval          0.034  0.062
##   P-value RMSEA <= 0.05                          0.556
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.048
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   GSC =~
##     SDQ2N01           1.000
##     SDQ2N13           1.083    0.154    7.044    0.000
##     SDQ2N25           0.851    0.132    6.455    0.000
##     SDQ2N37           0.934    0.131    7.131    0.000
##   ASC =~
##     SDQ2N04           1.000
##     SDQ2N16           1.279    0.150    8.520    0.000
##     SDQ2N28           1.247    0.154    8.097    0.000
##     SDQ2N40           1.259    0.156    8.048    0.000
##   ESC =~
##     SDQ2N10           1.000
##     SDQ2N22           0.889    0.103    8.658    0.000
##     SDQ2N34           0.670    0.148    4.539    0.000
##     SDQ2N46           0.843    0.117    7.225    0.000
##   MSC =~
##     SDQ2N07           1.000
##     SDQ2N19           0.841    0.058   14.495    0.000
##     SDQ2N31           0.952    0.049   19.516    0.000
##     SDQ2N43           0.655    0.049   13.298    0.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   GSC ~~
##     ASC               0.415    0.078    5.292    0.000
##     ESC               0.355    0.072    4.947    0.000
##     MSC               0.635    0.118    5.387    0.000
##   ASC ~~
##     ESC               0.464    0.078    5.921    0.000
##     MSC               0.873    0.134    6.519    0.000
##   ESC ~~
```

```
##     MSC               0.331    0.100    3.309    0.001
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .SDQ2N01           1.198    0.126    9.537    0.000
##    .SDQ2N13           1.119    0.124    9.019    0.000
##    .SDQ2N25           1.056    0.107    9.897    0.000
##    .SDQ2N37           0.771    0.087    8.821    0.000
##    .SDQ2N04           1.394    0.128   10.900    0.000
##    .SDQ2N16           0.616    0.068    9.020    0.000
##    .SDQ2N28           0.896    0.090    9.959    0.000
##    .SDQ2N40           0.952    0.095   10.029    0.000
##    .SDQ2N10           0.653    0.082    7.941    0.000
##    .SDQ2N22           0.657    0.075    8.735    0.000
##    .SDQ2N34           2.590    0.233   11.128    0.000
##    .SDQ2N46           1.201    0.118   10.183    0.000
##    .SDQ2N07           0.854    0.100    8.551    0.000
##    .SDQ2N19           1.228    0.121   10.153    0.000
##    .SDQ2N31           0.365    0.065    5.649    0.000
##    .SDQ2N43           0.964    0.092   10.473    0.000
##     GSC               0.613    0.137    4.464    0.000
##     ASC               0.561    0.126    4.453    0.000
##     ESC               0.668    0.116    5.749    0.000
##     MSC               2.307    0.273    8.460    0.000
```

The hypothesis that SC has four factors is not supported by the results (chi square statistic = 159.112, p = 0.000), which suggest that the fit of the data to the model is not adequate and null hypothesis should be rejected. However, the indices CFI (0.961), TLI (0.953) and RMSEA (0.049) support the null hypothesis.

Exercise 2.2

a) Hypothesis 2:

The null hypothesis is that SC has two factors:

- General SC (GSC)
- Academic SC (ASC)

Alternative hypothesis is that SC does not have two factors. Let us form the model:

```
modelF2 <- "
GSC =~ SDQ2N01 + SDQ2N13 + SDQ2N25 + SDQ2N37
OSC =~ SDQ2N04 + SDQ2N16 + SDQ2N28 + SDQ2N40
+ SDQ2N10 + SDQ2N22 + SDQ2N34 + SDQ2N46
+ SDQ2N07 + SDQ2N19 + SDQ2N31 + SDQ2N43
"

fitF2 <- cfa(modelF2, data = ex2.1)
summary(fitF2, fit.measures = T)
```

```
## lavaan 0.6-3 ended normally after 38 iterations
##
##   Optimization method                           NLMINB
##   Number of free parameters                         33
##
```

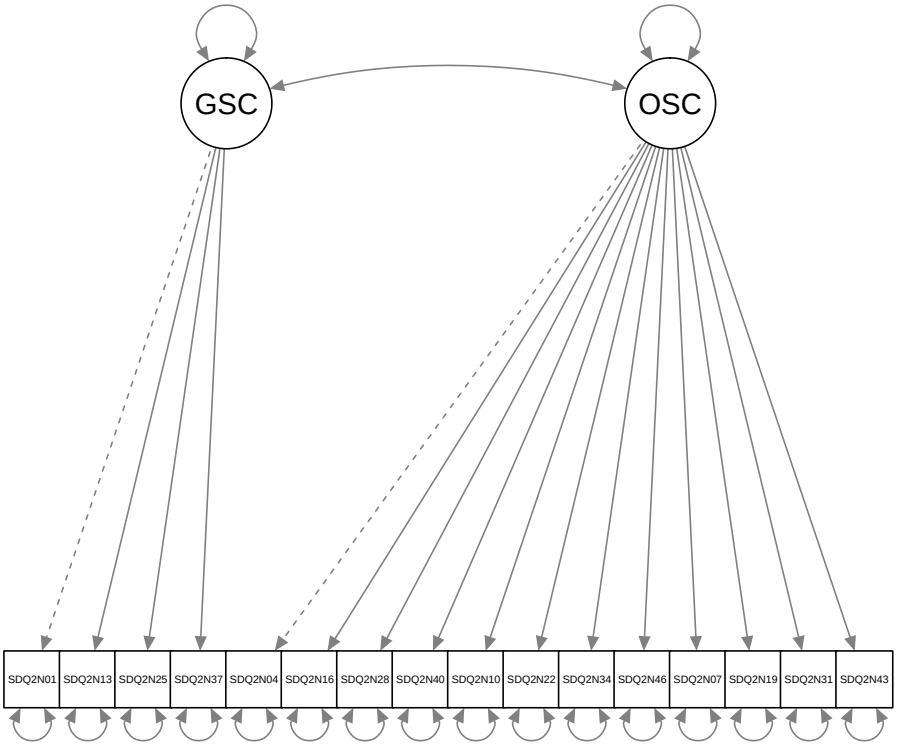
```
##   Number of observations                           265
##
##   Estimator                                         ML
##   Model Fit Test Statistic                     457.653
##   Degrees of freedom                               103
##   P-value (Chi-square)                           0.000
##
## Model test baseline model:
##
##   Minimum Function Test Statistic             1703.155
##   Degrees of freedom                               120
##   P-value                                        0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    0.776
##   Tucker-Lewis Index (TLI)                       0.739
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -6711.949
##   Loglikelihood unrestricted model (H1)      -6483.122
##
##   Number of free parameters                         33
##   Akaike (AIC)                               13489.897
##   Bayesian (BIC)                             13608.028
##   Sample-size adjusted Bayesian (BIC)        13503.401
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.114
##   90 Percent Confidence Interval          0.103  0.125
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.101
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   GSC =~
##     SDQ2N01           1.000
##     SDQ2N13           1.048    0.151    6.930    0.000
##     SDQ2N25           0.860    0.131    6.542    0.000
##     SDQ2N37           0.890    0.128    6.957    0.000
##   OSC =~
##     SDQ2N04           1.000
##     SDQ2N16           1.263    0.170    7.440    0.000
```

```
##     SDQ2N28           1.276    0.177    7.221    0.000
##     SDQ2N40           1.235    0.176    7.026    0.000
##     SDQ2N10           0.581    0.123    4.736    0.000
##     SDQ2N22           0.558    0.117    4.786    0.000
##     SDQ2N34           0.065    0.161    0.406    0.685
##     SDQ2N46           0.514    0.132    3.885    0.000
##     SDQ2N07           2.069    0.262    7.885    0.000
##     SDQ2N19           1.871    0.242    7.721    0.000
##     SDQ2N31           2.021    0.247    8.192    0.000
##     SDQ2N43           1.442    0.193    7.481    0.000
##
## Covariances:
##                   Estimate  Std.Err  z-value  P(>|z|)
##   GSC ~~
##     OSC               0.340    0.068    4.975    0.000
##
## Variances:
##                   Estimate  Std.Err  z-value  P(>|z|)
##    .SDQ2N01           1.170    0.127    9.216    0.000
##    .SDQ2N13           1.134    0.127    8.906    0.000
##    .SDQ2N25           1.026    0.107    9.582    0.000
##    .SDQ2N37           0.799    0.090    8.842    0.000
##    .SDQ2N04           1.495    0.134   11.171    0.000
##    .SDQ2N16           0.799    0.076   10.490    0.000
##    .SDQ2N28           1.018    0.095   10.695    0.000
##    .SDQ2N40           1.138    0.105   10.828    0.000
##    .SDQ2N10           1.166    0.103   11.364    0.000
##    .SDQ2N22           1.043    0.092   11.360    0.000
##    .SDQ2N34           2.888    0.251   11.510    0.000
##    .SDQ2N46           1.554    0.136   11.425    0.000
##    .SDQ2N07           1.191    0.123    9.654    0.000
##    .SDQ2N19           1.247    0.124   10.067    0.000
##    .SDQ2N31           0.575    0.073    7.852    0.000
##    .SDQ2N43           0.996    0.095   10.442    0.000
##     GSC               0.641    0.142    4.508    0.000
##     OSC               0.461    0.114    4.034    0.000
```

The hypothesis that SC has two factors is not supported by the results (chi square statistic = 457.653, p = 0.000), which suggest that the fit of the data to the model is not adequate and null hypothesis should be rejected. In addition the indices CFI (0.776), TLI (0.739) and RMSEA (0.144) are not supporting the null hypothesis.

Let us visualize the model structure:

```
library(semPlot)
semPaths(fitF2, layout='tree2')
```



b) Hypothesis 3:

The null hypothesis is that SC is unidimensional (only one SC factor).

Alternative hypothesis is that SC is not unidimensional.

```
modelF1 <- "
SC =~ SDQ2N01 + SDQ2N13 + SDQ2N25 + SDQ2N37
+ SDQ2N04 + SDQ2N16 + SDQ2N28 + SDQ2N40
+ SDQ2N10 + SDQ2N22 + SDQ2N34 + SDQ2N46
+ SDQ2N07 + SDQ2N19 + SDQ2N31 + SDQ2N43
"

fitF1 <- cfa(modelF1, data = ex2.1)
summary(fitF1, fit.measures = T)
```

```
## lavaan 0.6-3 ended normally after 43 iterations
##
##   Optimization method                           NLMINB
##   Number of free parameters                         32
##
##   Number of observations                           265
##
##   Estimator                                         ML
##   Model Fit Test Statistic                     531.918
##   Degrees of freedom                               104
##   P-value (Chi-square)                           0.000
##
## Model test baseline model:
##
##   Minimum Function Test Statistic             1703.155
```

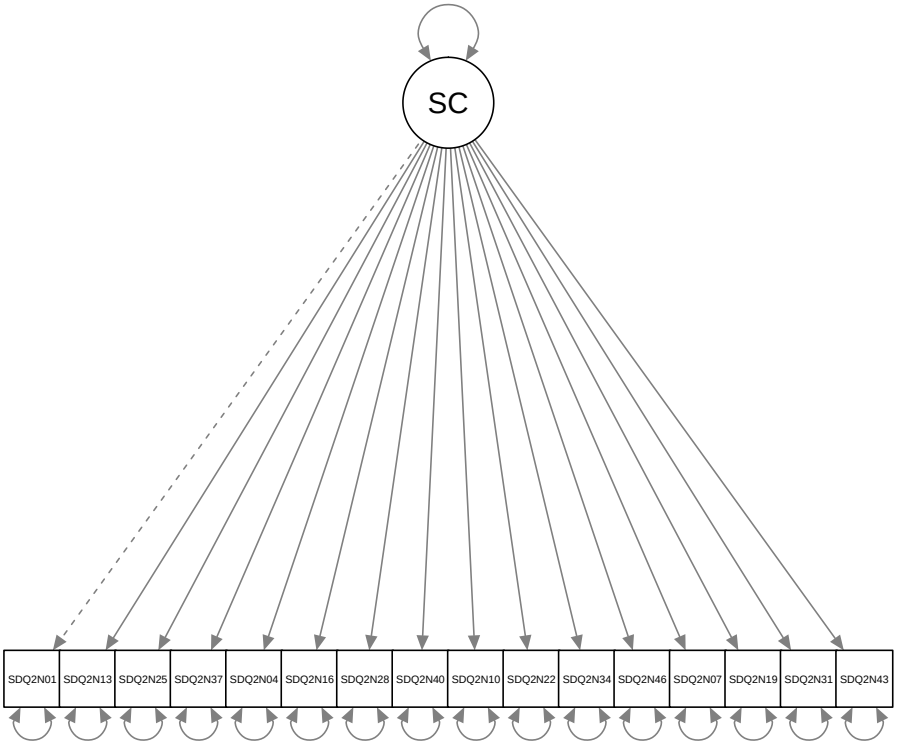
```
##   Degrees of freedom                                120
##   P-value                                         0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                     0.730
##   Tucker-Lewis Index (TLI)                        0.688
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -6749.081
##   Loglikelihood unrestricted model (H1)      -6483.122
##
##   Number of free parameters                         32
##   Akaike (AIC)                               13562.162
##   Bayesian (BIC)                             13676.713
##   Sample-size adjusted Bayesian (BIC)        13575.256
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                           0.125
##   90 Percent Confidence Interval          0.114  0.135
##   P-value RMSEA <= 0.05                           0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                            0.104
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   SC =~
##     SDQ2N01           1.000
##     SDQ2N13           1.158    0.247    4.690    0.000
##     SDQ2N25           0.903    0.209    4.330    0.000
##     SDQ2N37           1.126    0.224    5.018    0.000
##     SDQ2N04           1.407    0.278    5.063    0.000
##     SDQ2N16           1.772    0.310    5.716    0.000
##     SDQ2N28           1.775    0.317    5.605    0.000
##     SDQ2N40           1.744    0.315    5.541    0.000
##     SDQ2N10           0.859    0.197    4.362    0.000
##     SDQ2N22           0.816    0.187    4.371    0.000
##     SDQ2N34           0.181    0.222    0.815    0.415
##     SDQ2N46           0.756    0.202    3.732    0.000
##     SDQ2N07           2.743    0.471    5.826    0.000
##     SDQ2N19           2.505    0.434    5.768    0.000
##     SDQ2N31           2.711    0.454    5.970    0.000
##     SDQ2N43           1.929    0.341    5.659    0.000
##
```

```
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .SDQ2N01           1.565    0.138   11.335    0.000
##    .SDQ2N13           1.508    0.134   11.266    0.000
##    .SDQ2N25           1.299    0.115   11.338    0.000
##    .SDQ2N37           0.994    0.089   11.160    0.000
##    .SDQ2N04           1.469    0.132   11.140    0.000
##    .SDQ2N16           0.762    0.073   10.368    0.000
##    .SDQ2N28           0.994    0.093   10.633    0.000
##    .SDQ2N40           1.093    0.102   10.742    0.000
##    .SDQ2N10           1.140    0.101   11.333    0.000
##    .SDQ2N22           1.022    0.090   11.332    0.000
##    .SDQ2N34           2.882    0.250   11.508    0.000
##    .SDQ2N46           1.535    0.135   11.409    0.000
##    .SDQ2N07           1.311    0.132    9.913    0.000
##    .SDQ2N19           1.316    0.129   10.186    0.000
##    .SDQ2N31           0.650    0.078    8.367    0.000
##    .SDQ2N43           1.040    0.099   10.520    0.000
##     SC                0.246    0.083    2.972    0.003
```

The hypothesis that SC is unidimensional is not supported by the results (chi square statistic = 531.918, p = 0.000), which suggest that the fit of the data to the model is not adequate and null hypothesis should be rejected. In addition the indices CFI (0.730), TLI (0.688) and RMSEA (0.125) are not supporting the null hypothesis.

Let us visualize the model structure:

```
library(semPlot)
semPaths(fitF1, layout='tree2')
```



Overall the best fit of these three models is the first one with four factors since it is the only model that had at least some support for the model fit.