Exercises Week 2 Latent Variable Modeling

```
library("lavaan")
```

Exercise 3.1

```
Health.cov <- lav_matrix_lower2full(c(
  0.77,
  0.38, 0.65,
  0.39, 0.39, 0.62,
 -0.25,-0.32,-0.27, 6.09,
  0.31, 0.29, 0.26,-0.36, 7.67,
  0.24, 0.25, 0.19,-0.18, 0.51, 1.69,
 -3.16,-3.56,-2.63, 6.09,-3.12,-4.58,204.79,
 -0.92,-0.88,-0.72, 0.88,-1.49,-1.41, 16.53, 7.24
))
rownames(Health.cov) <- colnames(Health.cov) <-
  c("Dep1","Dep2","Dep3", "SocAct", "Falls", "ChronCond",
    "PhysAct", "PersMob")

marker.mod <- '
  F =~ 1*Dep1 + Dep2 + Dep3 + SocAct
'
marker.fit <- cfa(marker.mod, sample.cov = Health.cov, sample.nobs = 6053)
summary(marker.fit, standardized = TRUE)
```

```
## lavaan (0.6-1) converged normally after  27 iterations
##
##   Number of observations                          6053
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       9.620
##   Degrees of freedom                                 2
##   P-value (Chi-square)                           0.008
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   F =~
##     Dep1              1.000                               0.616    0.701
##     Dep2              1.005    0.021   47.588    0.000    0.619    0.768
##     Dep3              1.025    0.022   47.638    0.000    0.631    0.801
##     SocAct           -0.736    0.058  -12.793    0.000   -0.453   -0.184
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Dep1              0.391    0.009   41.276    0.000    0.391    0.508
```

```
##    .Dep2              0.267    0.008   33.581    0.000    0.267    0.411
##    .Dep3              0.222    0.008   28.886    0.000    0.222    0.358
##    .SocAct            5.884    0.108   54.559    0.000    5.884    0.966
##     F                 0.379    0.014   27.888    0.000    1.000    1.000
```

```r
stdLV.mod <- '
  F =~ NA*Dep1 + Dep2 + Dep3 + SocAct
  F ~~ 1*F
'
stdLV.fit <- cfa(stdLV.mod, sample.cov = Health.cov, sample.nobs = 6053)
summary(stdLV.fit, standardized = TRUE)
```

```
## lavaan (0.6-1) converged normally after  19 iterations
##
##   Number of observations                          6053
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       9.620
##   Degrees of freedom                                 2
##   P-value (Chi-square)                           0.008
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   F =~
##     Dep1              0.616    0.011   55.776    0.000    0.616    0.701
##     Dep2              0.619    0.010   61.392    0.000    0.619    0.768
##     Dep3              0.631    0.010   64.285    0.000    0.631    0.801
##     SocAct           -0.453    0.035  -12.967    0.000   -0.453   -0.184
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##     F                 1.000                               1.000    1.000
##    .Dep1              0.391    0.009   41.276    0.000    0.391    0.508
##    .Dep2              0.267    0.008   33.581    0.000    0.267    0.411
##    .Dep3              0.222    0.008   28.886    0.000    0.222    0.358
##    .SocAct            5.884    0.108   54.559    0.000    5.884    0.966
```

```r
effects.mod <- '
  F =~ NA*Dep1 + a*Dep1 + b*Dep2 + c*Dep3 + d*SocAct
  a + b + c + d == 4
'
effects.fit <- cfa(effects.mod, sample.cov = Health.cov, sample.nobs = 6053)
summary(effects.fit, standardized = TRUE)
```

```
## lavaan (0.6-1) converged normally after  30 iterations
##
##   Number of observations                          6053
##
##   Estimator                                         ML
```

```
##   Model Fit Test Statistic                       9.620
##   Degrees of freedom                                 2
##   P-value (Chi-square)                           0.008
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   F =~
##     Dep1       (a)    1.744    0.048   36.449    0.000    0.616    0.701
##     Dep2       (b)    1.753    0.048   36.367    0.000    0.619    0.768
##     Dep3       (c)    1.787    0.049   36.377    0.000    0.631    0.801
##     SocAct     (d)   -1.284    0.131   -9.832    0.000   -0.453   -0.184
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Dep1              0.391    0.009   41.276    0.000    0.391    0.508
##    .Dep2              0.267    0.008   33.581    0.000    0.267    0.411
##    .Dep3              0.222    0.008   28.886    0.000    0.222    0.358
##    .SocAct            5.884    0.108   54.559    0.000    5.884    0.966
##     F                 0.125    0.007   17.914    0.000    1.000    1.000
##
## Constraints:
##                                                |Slack|
##     a+b+c+d - (4)                                0.000
```

All identification methods give the exact same chi-square value.

Additional question:

The paper authors have decided to measure poor psychosocial health by means of three depression items and a measure of social activities. However, it seems that social activities is not strongly associated with depressive symptoms, so one may wonder whether this is a valid model for measuring psychosocial health.

```r
fitmeasures(marker.fit)
```

```
##                npar                fmin               chisq
##               8.000               0.001               9.620
##                  df              pvalue      baseline.chisq
##               2.000               0.008            5907.581
##         baseline.df     baseline.pvalue                 cfi
##               6.000               0.000               0.999
##                 tli                nnfi                 rfi
##               0.996               0.996               0.995
##                 nfi                pnfi                 ifi
##               0.998               0.333               0.999
##                 rni                logl   unrestricted.logl
##               0.999          -33330.619          -33325.809
##                 aic                 bic              ntotal
##           66677.238           66730.905            6053.000
##                bic2               rmsea      rmsea.ci.lower
```

```
##           66705.483              0.025              0.011
##      rmsea.ci.upper       rmsea.pvalue                rmr
##               0.042              0.994              0.016
##          rmr_nomean               srmr       srmr_bentler
##               0.016              0.008              0.008
## srmr_bentler_nomean        srmr_bollen  srmr_bollen_nomean
##               0.008              0.008              0.008
##          srmr_mplus  srmr_mplus_nomean              cn_05
##               0.008              0.008           3770.918
##               cn_01                gfi               agfi
##            5796.282              0.999              0.996
##                pgfi                mfi               ecvi
##               0.200              0.999              0.004
```

Exercise 3.2

```
Health.mod <- '
  PPsyHealth =~ Dep1 + Dep2 + Dep3 + SocAct
  PPhysHealth =~ ChronCond + PhysAct + Falls
  PersMob ~ PPsyHealth + PPhysHealth
'
Health.fit <- sem(Health.mod, sample.cov = Health.cov, sample.nobs = 6053)
summary(Health.fit, fit.measures = TRUE, standardized = TRUE)
```

```
## lavaan (0.6-1) converged normally after  62 iterations
##
##   Number of observations                          6053
##
##   Estimator                                         ML
##   Model Fit Test Statistic                     254.865
##   Degrees of freedom                                18
##   P-value (Chi-square)                           0.000
##
## Model test baseline model:
##
##   Minimum Function Test Statistic            10290.938
##   Degrees of freedom                                28
##   P-value                                        0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    0.977
##   Tucker-Lewis Index (TLI)                       0.964
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -95467.244
##   Loglikelihood unrestricted model (H1)     -95339.812
##
##   Number of free parameters                         18
##   Akaike (AIC)                              190970.488
##   Bayesian (BIC)                            191091.238
##   Sample-size adjusted Bayesian (BIC)       191034.038
```

```
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.047
##   90 Percent Confidence Interval          0.042  0.052
##   P-value RMSEA <= 0.05                          0.856
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.027
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   PPsyHealth =~
##     Dep1              1.000                               0.624    0.712
##     Dep2              1.015    0.020   49.994    0.000    0.634    0.786
##     Dep3              0.972    0.020   49.608    0.000    0.607    0.771
##     SocAct           -0.771    0.056  -13.683    0.000   -0.481   -0.195
##   PPhysHealth =~
##     ChronCond         1.000                               0.610    0.469
##     PhysAct         -12.005    0.442  -27.173    0.000   -7.319   -0.511
##     Falls             1.073    0.070   15.337    0.000    0.654    0.236
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   PersMob ~
##     PPsyHealth        0.433    0.196    2.215    0.027    0.270    0.101
##     PPhysHealth      -4.034    0.286  -14.107    0.000   -2.459   -0.914
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   PPsyHealth ~~
##     PPhysHealth       0.250    0.011   22.634    0.000    0.656    0.656
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Dep1              0.380    0.009   41.810    0.000    0.380    0.494
##    .Dep2              0.248    0.007   33.798    0.000    0.248    0.382
##    .Dep3              0.251    0.007   35.721    0.000    0.251    0.405
##    .SocAct            5.858    0.107   54.534    0.000    5.858    0.962
##    .ChronCond         1.318    0.028   46.833    0.000    1.318    0.780
##    .PhysAct         151.194    3.462   43.667    0.000  151.194    0.738
##    .Falls             7.241    0.134   54.019    0.000    7.241    0.944
##    .PersMob           1.990    0.239    8.318    0.000    1.990    0.275
##     PPsyHealth        0.390    0.014   28.811    0.000    1.000    1.000
##     PPhysHealth       0.372    0.024   15.334    0.000    1.000    1.000
```

Poor physical health seems to be a much stronger predictor of personal mobility than poor psychosocial health.

Additional question:

The χ^2 value is significant, but the sample size is very large. CFI and TLI are both $> .95$, the RMSEA $< .06$ and SRMR is $< .05$, all indicating good model fit. Let's also inspect the residual (co)variances:

```r
fitMeasures(Health.fit, c("chisq", "df", "pvalue", "cfi", "rmsea", "srmr"))
```

```
##   chisq      df  pvalue     cfi   rmsea    srmr
## 254.865  18.000   0.000   0.977   0.047   0.027
```

```r
modificationIndices(Health.fit, sort = TRUE)[1:10,]
```

```
##            lhs op     rhs      mi    epc sepc.lv sepc.all sepc.nox
## 26 PPhysHealth =~    Dep3 100.581 -0.242  -0.148   -0.188   -0.188
## 28        Dep1 ~~    Dep2  90.131 -0.089  -0.089   -0.288   -0.288
## 47      SocAct ~~ PhysAct  68.471  3.297   3.297    0.111    0.111
## 29        Dep1 ~~    Dep3  37.601  0.054   0.054    0.176    0.176
## 45        Dep3 ~~ PersMob  29.876  0.098   0.098    0.139    0.139
## 34        Dep1 ~~ PersMob  28.808 -0.108  -0.108   -0.124   -0.124
## 24 PPhysHealth =~    Dep1  25.971  0.135   0.082    0.094    0.094
## 38        Dep2 ~~ PhysAct  24.638 -0.503  -0.503   -0.082   -0.082
## 27 PPhysHealth =~  SocAct  22.579 -0.409  -0.249   -0.101   -0.101
## 35        Dep2 ~~    Dep3  22.032  0.045   0.045    0.179    0.179
```

```r
residuals(Health.fit, type = "cor")
```

```
## $type
## [1] "cor.bollen"
##
## $cor
##           Dep1   Dep2   Dep3   SocAct ChrnCn PhysAc Falls  PersMb
## Dep1       0.000
## Dep2      -0.022  0.000
## Dep3       0.016  0.008  0.000
## SocAct     0.023 -0.008  0.011  0.000
## ChronCond -0.008 -0.003 -0.051  0.004  0.000
## PhysAct   -0.013 -0.045  0.025  0.107 -0.006  0.000
## Falls      0.017  0.008  0.000 -0.022  0.031  0.042  0.000
## PersMob   -0.035 -0.013  0.045  0.035 -0.005 -0.005  0.000  0.000
##
## $mean
##      Dep1      Dep2      Dep3    SocAct ChronCond   PhysAct     Falls
##         0         0         0         0         0         0         0
##   PersMob
##         0
```

If we look at the standardised parameter estimates though, the Falls variable is not well explained by the Physical Health factor. And, as noted earlier, Poor Psychosocial Health is not a strong predictor of Personal Mobility. However, residuals do not suggest this is problematic.

We could consider adding a correlated error between Physical Activities and Social Activities. These show the largest residual error ($> .10$) and it makes sense from a substantial poin of view: Both measure activity levels. In addition, if both were assessed with the same method (self-report? interview?), measurement errors are likely correlated.

Additional exercise 1

```
data(HolzingerSwineford1939)
summary(HolzingerSwineford1939)
```

```
##        id             sex            ageyr        agemo
##  Min.   :  1.0   Min.   :1.000   Min.   :11   Min.   : 0.000
##  1st Qu.: 82.0   1st Qu.:1.000   1st Qu.:12   1st Qu.: 2.000
##  Median :163.0   Median :2.000   Median :13   Median : 5.000
##  Mean   :176.6   Mean   :1.515   Mean   :13   Mean   : 5.375
##  3rd Qu.:272.0   3rd Qu.:2.000   3rd Qu.:14   3rd Qu.: 8.000
##  Max.   :351.0   Max.   :2.000   Max.   :16   Max.   :11.000
##
##          school        grade             x1               x2
##  Grant-White:145   Min.   :7.000   Min.   :0.6667   Min.   :2.250
##  Pasteur    :156   1st Qu.:7.000   1st Qu.:4.1667   1st Qu.:5.250
##                    Median :7.000   Median :5.0000   Median :6.000
##                    Mean   :7.477   Mean   :4.9358   Mean   :6.088
##                    3rd Qu.:8.000   3rd Qu.:5.6667   3rd Qu.:6.750
##                    Max.   :8.000   Max.   :8.5000   Max.   :9.250
##                    NA's   :1
##        x3              x4              x5              x6
##  Min.   :0.250   Min.   :0.000   Min.   :1.000   Min.   :0.1429
##  1st Qu.:1.375   1st Qu.:2.333   1st Qu.:3.500   1st Qu.:1.4286
##  Median :2.125   Median :3.000   Median :4.500   Median :2.0000
##  Mean   :2.250   Mean   :3.061   Mean   :4.341   Mean   :2.1856
##  3rd Qu.:3.125   3rd Qu.:3.667   3rd Qu.:5.250   3rd Qu.:2.7143
##  Max.   :4.500   Max.   :6.333   Max.   :7.000   Max.   :6.1429
##
##        x7              x8               x9
##  Min.   :1.304   Min.   : 3.050   Min.   :2.778
##  1st Qu.:3.478   1st Qu.: 4.850   1st Qu.:4.750
##  Median :4.087   Median : 5.500   Median :5.417
##  Mean   :4.186   Mean   : 5.527   Mean   :5.374
##  3rd Qu.:4.913   3rd Qu.: 6.100   3rd Qu.:6.083
##  Max.   :7.435   Max.   :10.000   Max.   :9.250
##
```

```
HS.mod.1 <- '
  IQ =~ x1 + x2 + x3 + x4 + x5 + x6
'
HS.fit.1 <- cfa(HS.mod.1, data = HolzingerSwineford1939, estimator = "MLR")
summary(HS.fit.1, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan (0.6-1) converged normally after  34 iterations
##
##   Number of observations                           301
##
##   Estimator                                         ML      Robust
##   Model Fit Test Statistic                     103.230     100.487
##   Degrees of freedom                                 9           9
##   P-value (Chi-square)                           0.000       0.000
##   Scaling correction factor                                  1.027
##     for the Yuan-Bentler correction (Mplus variant)
```

```
##
## Model test baseline model:
##
##   Minimum Function Test Statistic             668.643     605.920
##   Degrees of freedom                               15          15
##   P-value                                       0.000       0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                   0.856       0.845
##   Tucker-Lewis Index (TLI)                      0.760       0.742
##
##   Robust Comparative Fit Index (CFI)                           NA
##   Robust Tucker-Lewis Index (TLI)                              NA
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -2559.686   -2559.686
##   Loglikelihood unrestricted model (H1)     -2508.071   -2508.071
##
##   Number of free parameters                        12          12
##   Akaike (AIC)                               5143.372    5143.372
##   Bayesian (BIC)                             5187.857    5187.857
##   Sample-size adjusted Bayesian (BIC)        5149.800    5149.800
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                         0.187       0.184
##   90 Percent Confidence Interval         0.155  0.220       0.153  0.217
##   P-value RMSEA <= 0.05                         0.000       0.000
##
##   Robust RMSEA                                                 NA
##   90 Percent Confidence Interval                               NA     NA
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                          0.114       0.114
##
## Parameter Estimates:
##
##   Information                                Observed
##   Observed information based on               Hessian
##   Standard Errors                   Robust.huber.white
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   IQ =~
##     x1                1.000                               0.488    0.419
##     x2                0.511    0.152    3.354    0.001    0.250    0.212
##     x3                0.468    0.128    3.657    0.000    0.229    0.203
##     x4                2.028    0.322    6.303    0.000    0.990    0.852
##     x5                2.234    0.374    5.974    0.000    1.091    0.847
##     x6                1.882    0.295    6.375    0.000    0.919    0.840
##
```

```
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .x1               1.120    0.109   10.321    0.000    1.120    0.824
##    .x2               1.319    0.128   10.272    0.000    1.319    0.955
##    .x3               1.223    0.078   15.715    0.000    1.223    0.959
##    .x4               0.370    0.050    7.341    0.000    0.370    0.274
##    .x5               0.470    0.058    8.132    0.000    0.470    0.283
##    .x6               0.351    0.046    7.635    0.000    0.351    0.294
##     IQ               0.238    0.077    3.113    0.002    1.000    1.000
```

a) The model does not fit very well. Also, the standardized loadings of X1, X2 and X3 are much smaller than those of X4, X5 and X6, so do not seem well explained by the model.

b) The robust ML χ^2 value is lower than the ML, but the difference is small (103.230 vs. 100.487).

c) Let's see whether and how we can improve model fit:

```
residuals(HS.fit.1, type = "cor")
```

```
## $type
## [1] "cor.bollen"
##
## $cor
##    x1     x2     x3     x4     x5     x6
## x1  0.000
## x2  0.208  0.000
## x3  0.356  0.297  0.000
## x4  0.016 -0.028 -0.014  0.000
## x5 -0.061 -0.040 -0.094  0.012  0.000
## x6  0.005  0.014  0.027 -0.012  0.008  0.000
##
## $mean
## x1 x2 x3 x4 x5 x6
##  0  0  0  0  0  0
```

```
modificationIndices(HS.fit.1, sort. = TRUE)
```

```
##    lhs op rhs     mi    epc sepc.lv sepc.all sepc.nox
## 15  x1 ~~  x3 49.835  0.484   0.484    0.414    0.414
## 19  x2 ~~  x3 29.298  0.399   0.399    0.314    0.314
## 14  x1 ~~  x2 17.170  0.295   0.295    0.243    0.243
## 24  x3 ~~  x5 14.774 -0.206  -0.206   -0.271   -0.271
## 17  x1 ~~  x5  7.829 -0.149  -0.149   -0.205   -0.205
## 26  x4 ~~  x5  7.807  0.260   0.260    0.623    0.623
## 27  x4 ~~  x6  7.107 -0.207  -0.207   -0.573   -0.573
## 28  x5 ~~  x6  3.327  0.154   0.154    0.380    0.380
## 21  x2 ~~  x5  2.742 -0.092  -0.092   -0.117   -0.117
## 20  x2 ~~  x4  1.399 -0.059  -0.059   -0.085   -0.085
## 25  x3 ~~  x6  1.179  0.050   0.050    0.076    0.076
## 16  x1 ~~  x4  0.540  0.035   0.035    0.055    0.055
## 23  x3 ~~  x4  0.342 -0.028  -0.028   -0.042   -0.042
## 22  x2 ~~  x6  0.308  0.026   0.026    0.039    0.039
## 18  x1 ~~  x6  0.042  0.009   0.009    0.015    0.015
```

Residuals among X1, X2 and X3 are largest. Highest modification indices are for correlations between X1, X2 and X3. This matches what we already expected based on the standardized loadings, that the model can maybe be improved by adding a separate factor for X1, X2 and X3:

```r
HS.mod.2 <- '
  IQ1 =~ x1 + x2 + x3
  IQ2 =~ x4 + x5 + x6
'
HS.fit.2 <- cfa(HS.mod.2, data = HolzingerSwineford1939, estimator = "MLR")
summary(HS.fit.2, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan (0.6-1) converged normally after  28 iterations
##
##   Number of observations                           301
##
##   Estimator                                         ML      Robust
##   Model Fit Test Statistic                      24.361      24.373
##   Degrees of freedom                                 8           8
##   P-value (Chi-square)                           0.002       0.002
##   Scaling correction factor                                  1.000
##     for the Yuan-Bentler correction (Mplus variant)
##
## Model test baseline model:
##
##   Minimum Function Test Statistic              668.643     605.920
##   Degrees of freedom                                15          15
##   P-value                                        0.000       0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    0.975       0.972
##   Tucker-Lewis Index (TLI)                       0.953       0.948
##
##   Robust Comparative Fit Index (CFI)                            NA
##   Robust Tucker-Lewis Index (TLI)                               NA
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -2520.252   -2520.252
##   Loglikelihood unrestricted model (H1)      -2508.071   -2508.071
##
##   Number of free parameters                         13          13
##   Akaike (AIC)                                5066.503    5066.503
##   Bayesian (BIC)                              5114.696    5114.696
##   Sample-size adjusted Bayesian (BIC)         5073.467    5073.467
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.082       0.082
##   90 Percent Confidence Interval          0.046  0.121       0.046  0.121
##   P-value RMSEA <= 0.05                          0.067       0.067
##
##   Robust RMSEA                                                  NA
##   90 Percent Confidence Interval                                NA     NA
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.047       0.047
```

```
##
## Parameter Estimates:
##
##   Information                                 Observed
##   Observed information based on                Hessian
##   Standard Errors                   Robust.huber.white
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   IQ1 =~
##     x1                1.000                               0.907    0.778
##     x2                0.559    0.163    3.436    0.001    0.507    0.431
##     x3                0.708    0.162    4.369    0.000    0.642    0.568
##   IQ2 =~
##     x4                1.000                               0.991    0.852
##     x5                1.111    0.066   16.910    0.000    1.101    0.854
##     x6                0.925    0.062   14.966    0.000    0.917    0.838
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   IQ1 ~~
##     IQ2               0.414    0.106    3.889    0.000    0.461    0.461
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .x1                0.536    0.194    2.766    0.006    0.536    0.395
##    .x2                1.125    0.120    9.401    0.000    1.125    0.814
##    .x3                0.863    0.110    7.832    0.000    0.863    0.677
##    .x4                0.369    0.051    7.311    0.000    0.369    0.274
##    .x5                0.449    0.057    7.830    0.000    0.449    0.270
##    .x6                0.356    0.047    7.639    0.000    0.356    0.298
##     IQ1               0.822    0.215    3.831    0.000    1.000    1.000
##     IQ2               0.981    0.122    8.053    0.000    1.000    1.000
```

d) Looking at the estimated parameters, the standardized factor loadings of X1, X2 and X3 have substantially increased. Looking at model fit, that definitely improved. Chi-square value is much smaller (although still significant, but that is to be expected with $N = 300$). RMSEA is nearly adequate. CFI and SRMR indicate good model fit. AIC for the 2D model is lower than that for the 1D model. Furthermore, robust CFI and TLI are now $> .95$. Although the value of RMSEA $> .08$, the p-value for the test of the RMSEA being $< .05$ is $> .05$, and the confidence interval for the RMSEA includes .05, which seems acceptable. Finally, looking at the residuals:

```r
residuals(HS.fit.2, type="cor")
```

```
## $type
## [1] "cor.bollen"
##
## $cor
##        x1     x2     x3     x4     x5     x6
## x1  0.000
## x2 -0.038  0.000
## x3 -0.002  0.095  0.000
## x4  0.067 -0.016 -0.065  0.000
## x5 -0.013 -0.030 -0.146  0.005  0.000
## x6  0.056  0.026 -0.022 -0.010  0.004  0.000
```

```
##
## $mean
## x1 x2 x3 x4 x5 x6
##  0  0  0  0  0  0
```

We do see that there is one residual correlation > .1 (between X3 and X5). All other correlations seem well explained by the model.

Additional question 2

a) Fit the depicted model to the data:

```
## Input covariances:
cormat <- lav_matrix_lower2full(c(
  1.000,
  0.700,  1.000,
  0.713,  0.636,  1.000,
  0.079,  0.066,  0.076,  1.000,
  0.088,  0.058,  0.070,  0.681,  1.000,
  0.084,  0.056,  0.074,  0.712,  0.633,  1.000,
  0.279,  0.248,  0.240,  0.177,  0.155,  0.170,  1.000,
  0.250,  0.214,  0.222,  0.157,  0.143,  0.152,  0.373,  1.000,
  0.280,  0.236,  0.251,  0.173,  0.178,  0.171,  0.448,  0.344,  1.000
))

## Input standard deviations:
sds = c(2.5, 2.1, 3.0, 4.1, 3.9, 4.4, 1.2, 1.0, 1.2)

## Reconstruct covariance matrix from correlations and sds:
covmat <- diag(sds) %*% cormat %*% diag(sds)

## Assign row and column names:
rownames(covmat) <- colnames(covmat) <- c("Y1", "Y2", "Y3", "Y4", "Y5", "Y6",
                                          "X1", "X2", "X3")

## Define formative model:
form.mod <- '
  SATISFACTION =~ Y1 + Y2 + Y3
  OPTIMISM =~ Y4 + Y5 + Y6
  STRESS <~ 1*X1 + X2 +X3
  SATISFACTION ~ STRESS
  OPTIMISM ~ STRESS
'

## Fit model:
form.fit <- cfa(form.mod, sample.cov=covmat, sample.nobs = 500)
summary(form.fit, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan (0.6-1) converged normally after  67 iterations
##
##   Number of observations                           500
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       2.166
##   Degrees of freedom                                22
```

```
##   P-value (Chi-square)                           1.000
##
## Model test baseline model:
##
##   Minimum Function Test Statistic             1542.629
##   Degrees of freedom                                33
##   P-value                                        0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    1.000
##   Tucker-Lewis Index (TLI)                       1.020
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -9192.919
##   Loglikelihood unrestricted model (H1)      -9191.836
##
##   Number of free parameters                         17
##   Akaike (AIC)                               18419.837
##   Bayesian (BIC)                             18491.486
##   Sample-size adjusted Bayesian (BIC)        18437.526
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.000
##   90 Percent Confidence Interval          0.000  0.000
##   P-value RMSEA <= 0.05                          1.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.005
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   SATISFACTION =~
##     Y1                1.000                               2.217    0.888
##     Y2                0.746    0.038   19.570    0.000    1.655    0.789
##     Y3                1.086    0.055   19.930    0.000    2.409    0.804
##   OPTIMISM =~
##     Y4                1.000                               3.579    0.874
##     Y5                0.848    0.045   18.733    0.000    3.035    0.779
##     Y6                1.000    0.051   19.441    0.000    3.579    0.814
##
## Composites:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   STRESS <~
##     X1                1.000                               0.366    0.439
```

```
##     X2                1.053    0.445    2.369    0.018    0.386    0.385
##     X3                1.073    0.434    2.469    0.014    0.393    0.471
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   SATISFACTION ~
##     STRESS            0.317    0.083    3.806    0.000    0.390    0.390
##   OPTIMISM ~
##     STRESS            0.338    0.101    3.358    0.001    0.258    0.258
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##  .SATISFACTION ~~
##    .OPTIMISM          0.052    0.367    0.142    0.887    0.007    0.007
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Y1                1.321    0.186    7.083    0.000    1.321    0.212
##    .Y2                1.662    0.142   11.735    0.000    1.662    0.378
##    .Y3                3.181    0.284   11.213    0.000    3.181    0.354
##    .Y4                3.964    0.528    7.509    0.000    3.964    0.236
##    .Y5                5.971    0.506   11.795    0.000    5.971    0.393
##    .Y6                6.510    0.622   10.459    0.000    6.510    0.337
##    .SATISFACTION      4.169    0.364   11.458    0.000    0.848    0.848
##    .OPTIMISM         11.960    1.064   11.238    0.000    0.933    0.933
##     STRESS            0.000                               0.000    0.000
```

b) Fit a model with stress as a reflective LV:

```r
refl.mod <- '
  SATISFACTION =~ Y1 + Y2 + Y3
  OPTIMISM =~ Y4 + Y5 + Y6
  STRESS =~ 1*X1 + X2 +X3
  SATISFACTION ~ STRESS
  OPTIMISM ~ STRESS
'
refl.fit <- cfa(refl.mod, sample.cov = covmat, sample.nobs = 500)
summary(refl.fit, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan (0.6-1) converged normally after  65 iterations
##
##   Number of observations                           500
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       3.010
##   Degrees of freedom                                24
##   P-value (Chi-square)                           1.000
##
## Model test baseline model:
##
##   Minimum Function Test Statistic             1752.818
##   Degrees of freedom                                36
##   P-value                                        0.000
##
## User model versus baseline model:
```

```
##
##   Comparative Fit Index (CFI)                    1.000
##   Tucker-Lewis Index (TLI)                       1.018
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -9193.341
##   Loglikelihood unrestricted model (H1)      -9191.836
##
##   Number of free parameters                         21
##   Akaike (AIC)                               18428.681
##   Bayesian (BIC)                             18517.188
##   Sample-size adjusted Bayesian (BIC)        18450.533
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.000
##   90 Percent Confidence Interval          0.000  0.000
##   P-value RMSEA <= 0.05                          1.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.008
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Latent Variables:
##                     Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   SATISFACTION =~
##     Y1                1.000                                2.217    0.888
##     Y2                0.747    0.038   19.570    0.000    1.655    0.789
##     Y3                1.086    0.055   19.926    0.000    2.408    0.804
##   OPTIMISM =~
##     Y4                1.000                                3.580    0.874
##     Y5                0.848    0.045   18.731    0.000    3.034    0.779
##     Y6                1.000    0.051   19.440    0.000    3.579    0.814
##   STRESS =~
##     X1                1.000                                0.812    0.677
##     X2                0.675    0.078    8.696    0.000    0.548    0.548
##     X3                0.962    0.103    9.314    0.000    0.781    0.652
##
## Regressions:
##                     Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   SATISFACTION ~
##     STRESS            1.299    0.177    7.345    0.000    0.476    0.476
##   OPTIMISM ~
##     STRESS            1.388    0.272    5.110    0.000    0.315    0.315
##
## Covariances:
##                     Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
```

```
## .SATISFACTION ~~
##    .OPTIMISM         -0.337    0.378   -0.892    0.372   -0.051   -0.051
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Y1                1.321    0.187    7.082    0.000    1.321    0.212
##    .Y2                1.661    0.142   11.731    0.000    1.661    0.377
##    .Y3                3.182    0.284   11.214    0.000    3.182    0.354
##    .Y4                3.963    0.528    7.506    0.000    3.963    0.236
##    .Y5                5.972    0.506   11.796    0.000    5.972    0.393
##    .Y6                6.510    0.622   10.459    0.000    6.510    0.337
##    .X1                0.778    0.080    9.758    0.000    0.778    0.542
##    .X2                0.698    0.054   12.810    0.000    0.698    0.700
##    .X3                0.827    0.079   10.497    0.000    0.827    0.576
##    .SATISFACTION      3.804    0.367   10.353    0.000    0.774    0.774
##    .OPTIMISM         11.545    1.065   10.837    0.000    0.901    0.901
##     STRESS            0.659    0.099    6.662    0.000    1.000    1.000
```

c) For both models, fit indices indicate excellent model fit. The formative model fits slightly better than the reflective model according to the SRMR, but the difference is very small.