Answers Session 3: Longitudinal SEMs

```r
library(lavaan)
```

```
## This is lavaan 0.6-1
```

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

```r
library(semPlot)
```

Additional exercise 2

(Similar to exercise 5.1 from the book.)

Here are the means and covariance matrix of six measurements in a sample called 'group 0':

```r
gr0.cov <- lav_matrix_lower2full(c(
  3.59,
  3.11, 3.10,
  2.91, 2.80, 2.82,
  3.22, 3.05, 2.86, 3.30,
  2.88, 2.63, 2.62, 2.82, 2.71
))
gr0.means <- c(11.97, 11.72, 12.03, 11.96, 12.10)

colnames(gr0.cov) <- rownames(gr0.cov) <-
  c("T1", "T2", "T3", "T4", "T5")
```

a) Fit the consecutive latent growth curve models to the data. Find the best-fitting model.

```r
growth.mod <- '
  i =~ 1*T1 + 1*T2 + 1*T3 + 1*T4 + 1*T5
  s =~ 0*T1 + 1*T2 + 2*T3 + 3*T4 + 4*T5
'
growth.fit <- growth(growth.mod, sample.cov = gr0.cov,
                     sample.mean = gr0.means, sample.nobs = 30)
summary(growth.fit, standardize = TRUE)
```

```
## lavaan (0.6-1) converged normally after  68 iterations
##
##   Number of observations                            30
##
##   Estimator                                         ML
##   Model Fit Test Statistic                      23.625
##   Degrees of freedom                                10
##   P-value (Chi-square)                           0.009
##
## 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
```

```
##   i =~
##     T1                1.000                                1.739    0.951
##     T2                1.000                                1.739    0.984
##     T3                1.000                                1.739    1.013
##     T4                1.000                                1.739    1.030
##     T5                1.000                                1.739    1.050
##   s =~
##     T1                0.000                                0.000    0.000
##     T2                1.000                                0.064    0.036
##     T3                2.000                                0.127    0.074
##     T4                3.000                                0.191    0.113
##     T5                4.000                                0.254    0.154
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   i ~~
##     s                -0.065    0.058   -1.107    0.268   -0.584   -0.584
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .T1                0.000                                0.000    0.000
##    .T2                0.000                                0.000    0.000
##    .T3                0.000                                0.000    0.000
##    .T4                0.000                                0.000    0.000
##    .T5                0.000                                0.000    0.000
##     i                11.836    0.325   36.366    0.000    6.808    6.808
##     s                 0.058    0.030    1.961    0.050    0.919    0.919
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .T1                0.317    0.110    2.886    0.004    0.317    0.095
##    .T2                0.223    0.074    3.030    0.002    0.223    0.072
##    .T3                0.167    0.057    2.912    0.004    0.167    0.057
##    .T4                0.176    0.059    2.971    0.003    0.176    0.062
##    .T5                0.172    0.071    2.417    0.016    0.172    0.063
##     i                 3.023    0.821    3.681    0.000    1.000    1.000
##     s                 0.004    0.009    0.465    0.642    1.000    1.000
```

```
indices <- c("chisq", "df", "pvalue", "cfi", "rmsea", "srmr", "aic")
fitmeasures(growth.fit, indices)
```

```
##   chisq      df  pvalue     cfi   rmsea    srmr     aic
##  23.625  10.000   0.009   0.952   0.213   0.056 338.907
```

We see that the mean of the latent intercept is significant, the mean of the latent slope is on the borderline of significance (because we have a very small sample size, so let's leave it in). The variance of the latent intercept is significant, but the variance of the latent slope is obviously not. The differences between residual variances over time are often smaller than their standard errors, so it may also be good to restrict those to equality:

```
growth.mod2 <- '
  i =~ 1*T1 + 1*T2 + 1*T3 + 1*T4 + 1*T5
  s =~ 0*T1 + 1*T2 + 2*T3 + 3*T4 + 4*T5
  # Zero variance and covariance for latent slope:
  s ~~ 0*s
  i ~~ 0*s
```

```r
  # Restrict residual variances to be equal across timepoints:
  T1 ~~ r*T1
  T2 ~~ r*T2
  T3 ~~ r*T3
  T4 ~~ r*T4
  T5 ~~ r*T5
'
growth.fit2 <- growth(growth.mod2, sample.cov = gr0.cov,
                      sample.mean = gr0.means, sample.nobs = 30)
fitmeasures(growth.fit2, indices)
```

```
##   chisq      df  pvalue     cfi   rmsea    srmr     aic
##  28.873  16.000   0.025   0.955   0.164   0.073 332.155
```

```r
lavTestLRT(growth.fit, growth.fit2)
```

```
## Chi Square Difference Test
##
##              Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## growth.fit  10 338.91 352.92 23.625
## growth.fit2 16 332.15 337.76 28.873      5.248       6     0.5124
```

CFI, RMSEA and AIC indicate better fit for the second model, SRMR is acceptable for both. The chi-square difference test indicates no significant deterioration of fit by removing some of the parameters. We take the second, more restricted model as the best fitting one.

b) What do the means and variances of the latent intercept and slope, and the standardized factor loadings tell you about inter- and intra-individual differences?

```r
summary(growth.fit2, standardize = TRUE)
```

```
## lavaan (0.6-1) converged normally after  32 iterations
##
##   Number of observations                            30
##
##   Estimator                                         ML
##   Model Fit Test Statistic                      28.873
##   Degrees of freedom                                16
##   P-value (Chi-square)                           0.025
##
## 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
##   i =~
##     T1                1.000                               1.671    0.963
##     T2                1.000                               1.671    0.963
##     T3                1.000                               1.671    0.963
##     T4                1.000                               1.671    0.963
##     T5                1.000                               1.671    0.963
##   s =~
##     T1                0.000                               0.000    0.000
```

```
##     T2                1.000                               0.000    0.000
##     T3                2.000                               0.000    0.000
##     T4                3.000                               0.000    0.000
##     T5                4.000                               0.000    0.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   i ~~
##     s                 0.000                                 NaN      NaN
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .T1                0.000                               0.000    0.000
##    .T2                0.000                               0.000    0.000
##    .T3                0.000                               0.000    0.000
##    .T4                0.000                               0.000    0.000
##    .T5                0.000                               0.000    0.000
##     i                11.856    0.312   37.979    0.000    7.097    7.097
##     s                 0.050    0.027    1.842    0.066      Inf      Inf
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##     s                 0.000                                 NaN      NaN
##    .T1         (r)    0.221    0.029    7.746    0.000    0.221    0.073
##    .T2         (r)    0.221    0.029    7.746    0.000    0.221    0.073
##    .T3         (r)    0.221    0.029    7.746    0.000    0.221    0.073
##    .T4         (r)    0.221    0.029    7.746    0.000    0.221    0.073
##    .T5         (r)    0.221    0.029    7.746    0.000    0.221    0.073
##     i                 2.791    0.732    3.812    0.000    1.000    1.000
```

The standardized factor loadings for the latent intercept indicate that most variance in observed variables can be explained by intercept (inter-individual) differences. The latent intercept has a significant non-zero variance of 2.79, so there seems to be substantial variation between observations. The intercept of the slope is not significant, but positive. This may indicate a small increase over time, but difficult to say with such small sample size, so there may as well be no change over time at all.

Additional exercise 3

Demo.growth is a dataset that is included in the lavaan package, consisting of 400 observations on the following variables:

$t1$ - $t4$: variable of interest, measured at four timepoints

$x1$ - $x2$: two time-invariant covariates

$c1$ - $c4$: a time-varying covariate

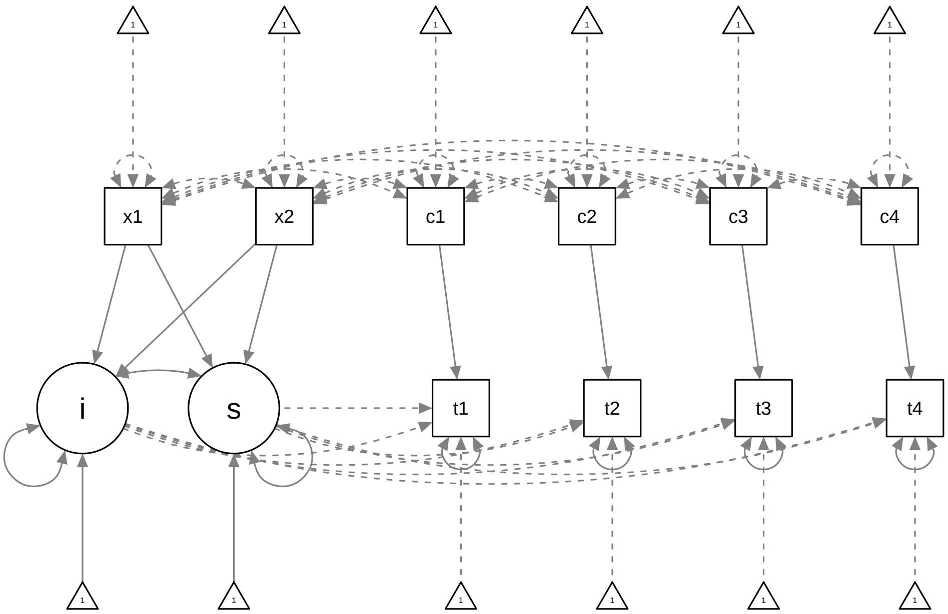
```r
data(Demo.growth)
```

```r
LGCM1 <- '
  # define latent intercept and slope:
  i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
  s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4

  # define regressions:
  i ~ x1 + x2
```

```r
  s ~ x1 + x2
  t1 ~ c1
  t2 ~ c2
  t3 ~ c3
  t4 ~ c4
'
fit1 <- growth(LGCM1, data = Demo.growth)
```

Fit an LGCM with $x1$ and $x2$ as time-constant predictors of the latent intercept and slope, and $c1$ trough $c4$ as time-varying predictors of the observed variables $t1$ through $t4$. Like in the following picture:



a) Test whether x1 and x2 are significant predictors of the slope and/or intercept.

```r
summary(fit1, standardized = TRUE)
```

```
## lavaan (0.6-1) converged normally after  31 iterations
##
##   Number of observations                           400
##
##   Estimator                                         ML
##   Model Fit Test Statistic                      26.059
##   Degrees of freedom                                21
##   P-value (Chi-square)                           0.204
##
## 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
##   i =~
##     t1                1.000                               1.386    0.875
##     t2                1.000                               1.386    0.660
```

```
##     t3                  1.000                                     1.386    0.507
##     t4                  1.000                                     1.386    0.412
##   s =~
##     t1                  0.000                                     0.000    0.000
##     t2                  1.000                                     0.768    0.366
##     t3                  2.000                                     1.536    0.562
##     t4                  3.000                                     2.304    0.684
##
## Regressions:
##                      Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   i ~
##     x1                  0.608    0.060   10.134    0.000    0.439    0.451
##     x2                  0.604    0.064    9.412    0.000    0.436    0.419
##   s ~
##     x1                  0.262    0.029    9.198    0.000    0.341    0.351
##     x2                  0.522    0.031   17.083    0.000    0.679    0.653
##   t1 ~
##     c1                  0.143    0.050    2.883    0.004    0.143    0.089
##   t2 ~
##     c2                  0.289    0.046    6.295    0.000    0.289    0.131
##   t3 ~
##     c3                  0.328    0.044    7.361    0.000    0.328    0.112
##   t4 ~
##     c4                  0.330    0.058    5.655    0.000    0.330    0.091
##
## Covariances:
##                      Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##  .i ~~
##    .s                   0.075    0.040    1.855    0.064    0.152    0.152
##
## Intercepts:
##                      Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .t1                  0.000                                     0.000    0.000
##    .t2                  0.000                                     0.000    0.000
##    .t3                  0.000                                     0.000    0.000
##    .t4                  0.000                                     0.000    0.000
##    .i                   0.580    0.062    9.368    0.000    0.419    0.419
##    .s                   0.958    0.029   32.552    0.000    1.247    1.247
##
## Variances:
##                      Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .t1                  0.580    0.080    7.230    0.000    0.580    0.231
##    .t2                  0.596    0.054   10.969    0.000    0.596    0.135
##    .t3                  0.481    0.055    8.745    0.000    0.481    0.064
##    .t4                  0.535    0.098    5.466    0.000    0.535    0.047
##    .i                   1.079    0.112    9.609    0.000    0.562    0.562
##    .s                   0.224    0.027    8.429    0.000    0.379    0.379
```

Indeed, $x1$ and $x2$ are significant predictors of the intercepts and slope. All associations are positive, indicating that higher values of $x1$ and/or $x2$ result in higher levels at the first timepoint, and stronger growth over time.

b) Test whether c has the same effect on t at each timepoint.

The estimated regression coefficients between c and x at each timepoint indicate that the effect of c grows stronger. Let's fit a model in which the regression coefficients are restricted to be equal across timepoints:

```
LGCM2 <- '
  # define latent intercept and slope:
  i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
  s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4

  # define regressions:
  i ~ x1 + x2
  s ~ x1 + x2
  t1 ~ c*c1
  t2 ~ c*c2
  t3 ~ c*c3
  t4 ~ c*c4
'
fit2 <- growth(LGCM2, data = Demo.growth)
lavTestLRT(fit1, fit2)
```

```
## Chi Square Difference Test
##
##      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit1 21 4996.2 5064.1 26.059
## fit2 24 4999.4 5055.3 35.264     9.2052       3    0.02668 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fitmeasures(fit1, indices)
```

```
##    chisq       df   pvalue      cfi    rmsea     srmr      aic
##   26.059   21.000    0.204    0.998    0.025    0.014 4996.244
```

```
fitmeasures(fit2, indices)
```

```
##    chisq       df   pvalue      cfi    rmsea     srmr      aic
##   35.264   24.000    0.065    0.995    0.034    0.018 4999.449
```

The model fit deteriorates significantly according to the chi-square value. According to AIC, the regression coefficients can also not be assumed equal. RMSEA, CFI and SRMR indicate good fit for both models, but better fit for the model with freely estimated regression coefficients. Conclusion: The effect of c on x is not equal across timepoints, but increases with time.

c) Test whether the residual variances are the same across timepoints.

```
LGCM3 <- '
  # define latent intercept and slope:
  i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
  s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4

  # define regressions:
  i ~ x1 + x2
  s ~ x1 + x2
  t1 ~ c1
  t2 ~ c2
  t3 ~ c3
  t4 ~ c4

  # set residual variances equal across timepoints:
  t1 ~~ r*t1
```

```r
  t2 ~~ r*t2
  t3 ~~ r*t3
  t4 ~~ r*t4
'
fit3 <- growth(LGCM3, data = Demo.growth)
lavTestLRT(fit1, fit3)
```

```
## Chi Square Difference Test
##
##      Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit1 21 4996.2 5064.1 26.059
## fit3 24 4992.9 5048.7 28.680     2.6213       3     0.4538
```

```r
fitmeasures(fit1, indices)
```

```
##    chisq       df   pvalue      cfi    rmsea     srmr      aic
##   26.059   21.000    0.204    0.998    0.025    0.014 4996.244
```

```r
fitmeasures(fit3, indices)
```

```
##    chisq       df   pvalue      cfi    rmsea     srmr      aic
##   28.680   24.000    0.232    0.998    0.022    0.015 4992.865
```

The chi-square difference test indicates that restricting the residual variances to be equal across timepoints does not significantly deteriorate model fit. According to AIC and RMSEA, the model with residual variances restricted to be equal fits best. CFI is the same for both models, SRMR is only marginally higher for the model with residual variances restricted to be equal. Conclusion: Residual variances can be assumed equal across timepoints.