SEM Review

```r
# Load the lavaan library
library(lavaan)
```

```
## This is lavaan 0.6-6
```

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

```r
# =~  to define latent variables
# ~~ to define covariance and correlation
# ~ to define direct prediction
```

```r
# Look at the dataset
data(HolzingerSwineford1939)
head(HolzingerSwineford1939[ , 7:15])
```

```
##         x1   x2    x3       x4   x5        x6       x7   x8       x9
## 1 3.333333 7.75 0.375 2.333333 5.75 1.2857143 3.391304 5.75 6.361111
## 2 5.333333 5.25 2.125 1.666667 3.00 1.2857143 3.782609 6.25 7.916667
## 3 4.500000 5.25 1.875 1.000000 1.75 0.4285714 3.260870 3.90 4.416667
## 4 5.333333 7.75 3.000 2.666667 4.50 2.4285714 3.000000 5.30 4.861111
## 5 4.833333 4.75 0.875 2.666667 4.00 2.5714286 3.695652 6.30 5.916667
## 6 5.333333 5.00 2.250 1.000000 3.00 0.8571429 4.347826 6.65 7.500000
```

Define your model specification

```r
text.model <- 'textspeed =~ x4 + x5 + x6 + x7 + x8 + x9'

#model name: 'text.model',
#latent variable : 'textspeed' (1 latent var) ,
#observed variables: x4 through x9 (6 observed var)
```

Analyze the model with cfa()

```r
text.fit <- cfa(model = text.model, data = HolzingerSwineford1939)
```

```r
#Summarize the model
summary(text.fit)
```

```
## lavaan 0.6-6 ended normally after 20 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         12
##
##   Number of observations                           301
##
## Model Test User Model:
##
##   Test statistic                               149.786
##   Degrees of freedom                                 9
##   P-value (Chi-square)                           0.000
```

```
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   textspeed =~
##     x4                1.000
##     x5                1.130    0.067   16.946    0.000
##     x6                0.925    0.056   16.424    0.000
##     x7                0.196    0.067    2.918    0.004
##     x8                0.186    0.062    2.984    0.003
##     x9                0.279    0.062    4.539    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .x4                0.383    0.048    7.903    0.000
##    .x5                0.424    0.059    7.251    0.000
##    .x6                0.368    0.044    8.419    0.000
##    .x7                1.146    0.094   12.217    0.000
##    .x8                0.988    0.081   12.215    0.000
##    .x9                0.940    0.077   12.142    0.000
##     textspeed         0.968    0.112    8.647    0.000
```

It is also important to examine model variances, which indicate the size of error in estimating manifest or latent variables.

You were able to view the coefficients for the model using the summary() function. However, the standardized coefficients in the Estimate column are often hard to interpret for how well they represent the latent variable.

```r
#standardized solution
summary(text.fit, standardized=TRUE)
```

```
## lavaan 0.6-6 ended normally after 20 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         12
##
##   Number of observations                           301
##
## Model Test User Model:
##
##   Test statistic                               149.786
##   Degrees of freedom                                 9
##   P-value (Chi-square)                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
```

```
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   textspeed =~
##     x4                1.000                               0.984    0.846
##     x5                1.130    0.067   16.946    0.000    1.112    0.863
##     x6                0.925    0.056   16.424    0.000    0.910    0.832
##     x7                0.196    0.067    2.918    0.004    0.193    0.177
##     x8                0.186    0.062    2.984    0.003    0.183    0.181
##     x9                0.279    0.062    4.539    0.000    0.275    0.273
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .x4                0.383    0.048    7.903    0.000    0.383    0.284
##    .x5                0.424    0.059    7.251    0.000    0.424    0.256
##    .x6                0.368    0.044    8.419    0.000    0.368    0.308
##    .x7                1.146    0.094   12.217    0.000    1.146    0.969
##    .x8                0.988    0.081   12.215    0.000    0.988    0.967
##    .x9                0.940    0.077   12.142    0.000    0.940    0.926
##     textspeed         0.968    0.112    8.647    0.000    1.000    1.000
```

Look at the Std.all column for the completely standardized solution to see which variables have a poor relationship to the text speed latent variable.

Looking at 'Latent Variables: Std.all', we can tell that variables 7, 8, and 9 do not measure text speed very well, as these loading coefficients are close to zero.(.177, .181, .273)

After reviewing the standardized loadings in the previous exercise, we found that several of the manifest variables may not represent our latent variable well.

As a second measure of our model, you can examine the fit indices to see if the model appropriately fits the data. You can look at both the goodness of fit and badness of fit statistics using the fit.measures argument within the summary() function.

```r
#goodness of fit and badness of fit statistics
summary(text.fit, fit.measures=TRUE )
```

```
## lavaan 0.6-6 ended normally after 20 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         12
##
##   Number of observations                           301
##
## Model Test User Model:
##
##   Test statistic                               149.786
##   Degrees of freedom                                 9
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                               681.336
##   Degrees of freedom                                15
##   P-value                                        0.000
##
## User Model versus Baseline Model:
```

```
##
##   Comparative Fit Index (CFI)                    0.789
##   Tucker-Lewis Index (TLI)                       0.648
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -2476.130
##   Loglikelihood unrestricted model (H1)      -2401.237
##
##   Akaike (AIC)                                4976.261
##   Bayesian (BIC)                              5020.746
##   Sample-size adjusted Bayesian (BIC)         4982.689
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.228
##   90 Percent confidence interval - lower         0.197
##   90 Percent confidence interval - upper         0.261
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.148
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   textspeed =~
##     x4                1.000
##     x5                1.130    0.067   16.946    0.000
##     x6                0.925    0.056   16.424    0.000
##     x7                0.196    0.067    2.918    0.004
##     x8                0.186    0.062    2.984    0.003
##     x9                0.279    0.062    4.539    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .x4                0.383    0.048    7.903    0.000
##    .x5                0.424    0.059    7.251    0.000
##    .x6                0.368    0.044    8.419    0.000
##    .x7                1.146    0.094   12.217    0.000
##    .x8                0.988    0.081   12.215    0.000
##    .x9                0.940    0.077   12.142    0.000
##     textspeed         0.968    0.112    8.647    0.000
```

Remember that goodness of fit statistics, like the CFI and TLI, should be large (over .90) and close to one, while badness of fit measures like the RMSEA and SRMR should be small (less than .10) and close to zero.

We can see that our fit indices are poor, with low CFI and TLI and high RMSEA and SRMR values. CFI=.789, TLI=.648, RMSE=..228, SRMR=.148

```r
#model with zero degrees of freedom
text.model1 <- 'text =~ x4 + x5 + x6'
text.fit1 <- cfa(model = text.model1, data = HolzingerSwineford1939)
summary(text.fit1)
```

```
## lavaan 0.6-6 ended normally after 15 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          6
##
##   Number of observations                           301
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   text =~
##     x4                1.000
##     x5                1.133    0.067   16.906    0.000
##     x6                0.924    0.056   16.391    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .x4                0.382    0.049    7.805    0.000
##    .x5                0.416    0.059    7.038    0.000
##    .x6                0.369    0.044    8.367    0.000
##     text              0.969    0.112    8.640    0.000
```

```r
#You should have at least one degree of freedom for any model.
```

If a model has zero degrees of freedom, it means we need to fix the model identification.

Update the model specification by setting two of the coefficient paths to 'a' to set them equal to each other.

```r
#fix model with zero degrees of freedom
text.model2 <- 'text =~ x4 + a*x5 + a*x6'
text.fit2 <- cfa(model = text.model2, data = HolzingerSwineford1939)
summary(text.fit2)
```

```
## lavaan 0.6-6 ended normally after 14 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          6
##   Number of equality constraints                     1
##
```

```
##   Number of observations                           301
##
## Model Test User Model:
##
##   Test statistic                                11.227
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.001
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   text =~
##     x4                1.000
##     x5         (a)    1.009    0.054   18.747    0.000
##     x6         (a)    1.009    0.054   18.747    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .x4                0.383    0.050    7.631    0.000
##    .x5                0.499    0.054    9.164    0.000
##    .x6                0.328    0.045    7.285    0.000
##     text              0.967    0.113    8.585    0.000
```

```
#two equal parameter estimates for x5 and x6
```

You have now created a two-factor model of the reading comprehension and speeded addition factors. Is that better than a one-factor model? Use the cfa() and summary() functions on your new two-factor model of the HolzingerSwineford1939 dataset to show the fit indices.

MULTIFACTOR MODEL:Two factor model

```
#two-factor model of text and speed variables
twofactor.model <- 'text =~ x4 + x5 + x6
speed =~ x7+ x8+x9'

#two-factor model of the reading comprehension and speeded addition factors
```

Is that better than a one-factor model? Use the cfa() and summary() functions to show the fit indices.

```
twofactor.fit <- cfa(model=twofactor.model, data=HolzingerSwineford1939)
summary(twofactor.fit, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan 0.6-6 ended normally after 24 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         13
##
##   Number of observations                           301
##
## Model Test User Model:
##
```

```
##   Test statistic                                14.354
##   Degrees of freedom                                 8
##   P-value (Chi-square)                           0.073
##
## Model Test Baseline Model:
##
##   Test statistic                               681.336
##   Degrees of freedom                                15
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.990
##   Tucker-Lewis Index (TLI)                       0.982
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -2408.414
##   Loglikelihood unrestricted model (H1)      -2401.237
##
##   Akaike (AIC)                                4842.828
##   Bayesian (BIC)                              4891.021
##   Sample-size adjusted Bayesian (BIC)         4849.792
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.051
##   90 Percent confidence interval - lower         0.000
##   90 Percent confidence interval - upper         0.093
##   P-value RMSEA <= 0.05                          0.425
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.039
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   text =~
##     x4                1.000                               0.984    0.847
##     x5                1.132    0.067   16.954    0.000    1.114    0.865
##     x6                0.925    0.056   16.438    0.000    0.911    0.833
##   speed =~
##     x7                1.000                               0.674    0.619
##     x8                1.150    0.165    6.990    0.000    0.775    0.766
##     x9                0.878    0.123    7.166    0.000    0.592    0.587
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
```

```
##   text ~~
##     speed             0.173    0.052    3.331    0.001    0.261    0.261
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .x4                0.382    0.049    7.854    0.000    0.382    0.283
##    .x5                0.418    0.059    7.113    0.000    0.418    0.252
##    .x6                0.367    0.044    8.374    0.000    0.367    0.307
##    .x7                0.729    0.084    8.731    0.000    0.729    0.616
##    .x8                0.422    0.084    5.039    0.000    0.422    0.413
##    .x9                0.665    0.071    9.383    0.000    0.665    0.655
##     text              0.969    0.112    8.647    0.000    1.000    1.000
##     speed             0.454    0.096    4.728    0.000    1.000    1.000
```

In comparing the one- and two-factor models, you should see that the fit indices are improved in the two-factor model.

Covariances: was added (amount by which 2 variables change together)

check covariance under std. all

R automatically correlates the latent variables, in order to:

Set covariance = 0, 'speed ~~ 0*visual'

Specify direct relationship between latent variables: 'speed ~ visual'

THREE-FACTOR MODEL:

Three-factor model of personality. This inventory includes 57 questions that measure extraversion, neuroticism, and lying.

Three factor model using the latent variables: extraversion, neuroticism, and lying with four manifest variables on each item.

Remember when you create multiple latent variables, these endogenous variables are automatically correlated. Set the correlation between the extraversion latent variable and neuroticism latent variable to zero, by using the ~~ in model specification code.

EPI

```r
# Load the library and data
library(psych)
```

```
##
## Attaching package: 'psych'

## The following object is masked from 'package:lavaan':
##
##     cor2cov
```

```r
epi <- read.csv("~/Downloads/epi.csv", row.names=1)

# Specify a three-factor model with correlation between extraversion and neuroticism set to zero
epi.model <- 'extraversion =~ V1 + V3 + V5 + V8
neuroticism =~ V2 + V4 + V7 + V9
lying =~ V6 + V12 + V18 + V24
extraversion ~~ 0*neuroticism'

# Run the model
epi.fit <- cfa(model = epi.model, data = epi)
```

```r
# Examine the output
summary(epi.fit, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan 0.6-6 ended normally after 118 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         26
##
##                                                   Used       Total
##   Number of observations                          3193        3570
##
## Model Test User Model:
##
##   Test statistic                               584.718
##   Degrees of freedom                                52
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              2196.019
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.750
##   Tucker-Lewis Index (TLI)                       0.683
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -23208.145
##   Loglikelihood unrestricted model (H1)     -22915.787
##
##   Akaike (AIC)                               46468.291
##   Bayesian (BIC)                             46626.077
##   Sample-size adjusted Bayesian (BIC)        46543.464
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.057
##   90 Percent confidence interval - lower         0.053
##   90 Percent confidence interval - upper         0.061
##   P-value RMSEA <= 0.05                          0.004
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.058
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
```

```
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion =~
##     V1                1.000                               0.052    0.115
##     V3                1.360    0.329    4.127    0.000    0.070    0.141
##     V5               -2.829    0.554   -5.109    0.000   -0.146   -0.391
##     V8                7.315    1.832    3.992    0.000    0.377    0.797
##   neuroticism =~
##     V2                1.000                               0.228    0.457
##     V4                0.424    0.053    8.004    0.000    0.097    0.196
##     V7                1.395    0.093   15.023    0.000    0.318    0.648
##     V9                1.205    0.078   15.506    0.000    0.275    0.553
##   lying =~
##     V6                1.000                               0.135    0.272
##     V12              -0.851    0.132   -6.435    0.000   -0.115   -0.291
##     V18              -0.785    0.122   -6.421    0.000   -0.106   -0.289
##     V24               1.086    0.161    6.734    0.000    0.147    0.339
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion ~~
##     neuroticism       0.000                               0.000    0.000
##     lying            -0.002    0.001   -3.313    0.001   -0.258   -0.258
##   neuroticism ~~
##     lying            -0.014    0.002   -6.867    0.000   -0.469   -0.469
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .V1                0.198    0.005   39.567    0.000    0.198    0.987
##    .V3                0.243    0.006   39.278    0.000    0.243    0.980
##    .V5                0.118    0.005   23.900    0.000    0.118    0.847
##    .V8                0.082    0.026    3.084    0.002    0.082    0.364
##    .V2                0.197    0.006   32.516    0.000    0.197    0.791
##    .V4                0.235    0.006   38.906    0.000    0.235    0.962
##    .V7                0.140    0.007   19.412    0.000    0.140    0.580
##    .V9                0.172    0.006   26.591    0.000    0.172    0.694
##    .V6                0.228    0.007   34.520    0.000    0.228    0.926
##    .V12               0.143    0.004   33.670    0.000    0.143    0.916
##    .V18               0.124    0.004   33.753    0.000    0.124    0.917
##    .V24               0.166    0.005   31.021    0.000    0.166    0.885
##     extraversion      0.003    0.001    2.480    0.013    1.000    1.000
##     neuroticism       0.052    0.005   10.010    0.000    1.000    1.000
##     lying             0.018    0.004    4.500    0.000    1.000    1.000
```

Create a DIRECT PATH

```r
#Edit the epi.model to include a direct regression path between lying and neuroticism.
#We might expect that a person's level of neuroticism would predict their level of lying.

epi.model1 <- 'extraversion =~ V1 + V3 + V5 + V8
neuroticism =~ V2 + V4 + V7 + V9
lying =~ V6 + V12 + V18 + V24
lying ~ neuroticism'                   #THIS LINE
```

```r
# Run the model
epi.fit1 <- cfa(model = epi.model1, data = epi)

# Examine the output
summary(epi.fit1, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan 0.6-6 ended normally after 120 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         26
##
##                                                   Used       Total
##   Number of observations                          3193        3570
##
## Model Test User Model:
##
##   Test statistic                               534.426
##   Degrees of freedom                                52
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              2196.019
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.774
##   Tucker-Lewis Index (TLI)                       0.713
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -23183.000
##   Loglikelihood unrestricted model (H1)     -22915.787
##
##   Akaike (AIC)                               46417.999
##   Bayesian (BIC)                             46575.786
##   Sample-size adjusted Bayesian (BIC)        46493.173
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.054
##   90 Percent confidence interval - lower         0.050
##   90 Percent confidence interval - upper         0.058
##   P-value RMSEA <= 0.05                          0.058
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.053
##
## Parameter Estimates:
```

```
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion =~
##     V1                1.000                               0.052    0.115
##     V3                1.135    0.268    4.230    0.000    0.059    0.118
##     V5               -2.497    0.443   -5.638    0.000   -0.129   -0.346
##     V8                8.223    2.008    4.096    0.000    0.425    0.898
##   neuroticism =~
##     V2                1.000                               0.223    0.447
##     V4                0.462    0.054    8.493    0.000    0.103    0.209
##     V7                1.435    0.093   15.368    0.000    0.320    0.652
##     V9                1.214    0.078   15.570    0.000    0.271    0.545
##   lying =~
##     V6                1.000                               0.125    0.252
##     V12              -0.943    0.150   -6.274    0.000   -0.118   -0.298
##     V18              -0.905    0.143   -6.339    0.000   -0.113   -0.308
##     V24               1.187    0.182    6.509    0.000    0.148    0.342
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   lying ~
##     neuroticism      -0.298    0.043   -6.943    0.000   -0.532   -0.532
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion ~~
##     neuroticism       0.003    0.001    3.761    0.000    0.240    0.240
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .V1                0.198    0.005   39.671    0.000    0.198    0.987
##    .V3                0.244    0.006   39.651    0.000    0.244    0.986
##    .V5                0.123    0.004   28.256    0.000    0.123    0.881
##    .V8                0.043    0.033    1.302    0.193    0.043    0.193
##    .V2                0.200    0.006   33.262    0.000    0.200    0.800
##    .V4                0.233    0.006   38.804    0.000    0.233    0.956
##    .V7                0.139    0.007   20.087    0.000    0.139    0.575
##    .V9                0.174    0.006   27.907    0.000    0.174    0.703
##    .V6                0.231    0.007   35.398    0.000    0.231    0.936
##    .V12               0.143    0.004   33.349    0.000    0.143    0.911
##    .V18               0.122    0.004   32.825    0.000    0.122    0.905
##    .V24               0.166    0.005   30.854    0.000    0.166    0.883
##     extraversion      0.003    0.001    2.643    0.008    1.000    1.000
##     neuroticism       0.050    0.005    9.947    0.000    1.000    1.000
##    .lying             0.011    0.003    3.970    0.000    0.717    0.717
```

UPDATING POOR MODELS:

if model has CFI and TLI below our criteria (.9) also if bad fit indices RMSEA and SRMS is higher tan criteria (.1)

```r
#CHECK MODEL VARIANCE

#In order to evaluate your three-factor model of the epi, you can examine the variance of the
#manifest(observable) variables to check for potential problems with the model. Very large variances
#can indicate potential issues; however, this value should be compared to the original scale of the dat
# Calculate the variance of V1

var(epi$V1) #0.2017972
```

```
## [1] NA
```

```r
#You can see that your variance from the model (0.199) is very similar to the real variance (0.201)
#which indicates our model does not have variance issues.
```

```r
#Examine MODIFICATION INDICES
#The fit indices for our epi.model are low (in the .70s) for CFI and TLI.
#You can use modification indices to find potential parameters (paths) to add to the model specificatio
#to improve model fit.

# Original model summary
summary(epi.fit, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan 0.6-6 ended normally after 118 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         26
##
##                                                   Used       Total
##   Number of observations                          3193        3570
##
## Model Test User Model:
##
##   Test statistic                               584.718
##   Degrees of freedom                                52
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              2196.019
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.750
##   Tucker-Lewis Index (TLI)                       0.683
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -23208.145
##   Loglikelihood unrestricted model (H1)     -22915.787
##
##   Akaike (AIC)                               46468.291
##   Bayesian (BIC)                             46626.077
```

```
##   Sample-size adjusted Bayesian (BIC)       46543.464
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.057
##   90 Percent confidence interval - lower         0.053
##   90 Percent confidence interval - upper         0.061
##   P-value RMSEA <= 0.05                          0.004
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.058
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion =~
##     V1                1.000                               0.052    0.115
##     V3                1.360    0.329    4.127    0.000    0.070    0.141
##     V5               -2.829    0.554   -5.109    0.000   -0.146   -0.391
##     V8                7.315    1.832    3.992    0.000    0.377    0.797
##   neuroticism =~
##     V2                1.000                               0.228    0.457
##     V4                0.424    0.053    8.004    0.000    0.097    0.196
##     V7                1.395    0.093   15.023    0.000    0.318    0.648
##     V9                1.205    0.078   15.506    0.000    0.275    0.553
##   lying =~
##     V6                1.000                               0.135    0.272
##     V12              -0.851    0.132   -6.435    0.000   -0.115   -0.291
##     V18              -0.785    0.122   -6.421    0.000   -0.106   -0.289
##     V24               1.086    0.161    6.734    0.000    0.147    0.339
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion ~~
##     neuroticism       0.000                               0.000    0.000
##     lying            -0.002    0.001   -3.313    0.001   -0.258   -0.258
##   neuroticism ~~
##     lying            -0.014    0.002   -6.867    0.000   -0.469   -0.469
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .V1                0.198    0.005   39.567    0.000    0.198    0.987
##    .V3                0.243    0.006   39.278    0.000    0.243    0.980
##    .V5                0.118    0.005   23.900    0.000    0.118    0.847
##    .V8                0.082    0.026    3.084    0.002    0.082    0.364
##    .V2                0.197    0.006   32.516    0.000    0.197    0.791
##    .V4                0.235    0.006   38.906    0.000    0.235    0.962
##    .V7                0.140    0.007   19.412    0.000    0.140    0.580
```

```
##    .V9                0.172    0.006   26.591    0.000    0.172    0.694
##    .V6                0.228    0.007   34.520    0.000    0.228    0.926
##    .V12               0.143    0.004   33.670    0.000    0.143    0.916
##    .V18               0.124    0.004   33.753    0.000    0.124    0.917
##    .V24               0.166    0.005   31.021    0.000    0.166    0.885
##     extraversion      0.003    0.001    2.480    0.013    1.000    1.000
##     neuroticism       0.052    0.005   10.010    0.000    1.000    1.000
##     lying             0.018    0.004    4.500    0.000    1.000    1.000
```

```r
# Examine the modification indices
modificationindices(epi.fit, sort=TRUE)
```

```
##              lhs op         rhs      mi    epc sepc.lv sepc.all sepc.nox
## 40   neuroticism =~          V3 152.701 -0.609  -0.139   -0.279   -0.279
## 39   neuroticism =~          V1 122.735  0.493   0.112    0.251    0.251
## 48         lying =~          V3 121.175  1.269   0.171    0.345    0.345
## 58            V1 ~~          V2  76.218  0.032   0.032    0.164    0.164
## 70            V3 ~~          V7  71.613 -0.033  -0.033   -0.178   -0.178
## 13  extraversion ~~ neuroticism  70.230  0.003   0.236    0.236    0.236
## 42   neuroticism =~          V8  68.905  0.372   0.085    0.179    0.179
## 47         lying =~          V1  62.368 -0.819  -0.111   -0.247   -0.247
## 50         lying =~          V8  56.929 -1.095  -0.148   -0.313   -0.313
## 87            V8 ~~          V7  38.504  0.022   0.022    0.203    0.203
## 33  extraversion =~          V7  30.415  1.034   0.053    0.109    0.109
## 59            V1 ~~          V4  28.442  0.021   0.021    0.095    0.095
## 32  extraversion =~          V4  27.525  1.079   0.056    0.113    0.113
## 75            V3 ~~         V24  20.299  0.017   0.017    0.084    0.084
## 52         lying =~          V4  18.610 -0.618  -0.084   -0.169   -0.169
## 103           V4 ~~         V12  17.780  0.014   0.014    0.078    0.078
## 86            V8 ~~          V4  15.339  0.015   0.015    0.109    0.109
## 113           V9 ~~         V18  15.043  0.012   0.012    0.081    0.081
## 53         lying =~          V7  13.292 -0.567  -0.077   -0.156   -0.156
## 35  extraversion =~          V6  10.893 -0.816  -0.042   -0.085   -0.085
## 76            V5 ~~          V8   9.434  0.103   0.103    1.046    1.046
## 116           V6 ~~         V18   9.357  0.011   0.011    0.067    0.067
## 45   neuroticism =~         V18   9.199  0.178   0.041    0.111    0.111
## 74            V3 ~~         V18   8.727 -0.009  -0.009   -0.054   -0.054
## 64            V1 ~~         V18   8.624  0.008   0.008    0.054    0.054
## 68            V3 ~~          V2   8.157 -0.012  -0.012   -0.054   -0.054
## 99            V2 ~~         V24   7.503  0.010   0.010    0.055    0.055
## 51         lying =~          V2   7.304  0.389   0.053    0.105    0.105
## 84            V5 ~~         V24   7.237  0.008   0.008    0.054    0.054
## 89            V8 ~~          V6   6.987 -0.011  -0.011   -0.084   -0.084
## 66            V3 ~~          V5   6.798 -0.010  -0.010   -0.060   -0.060
## 107           V7 ~~          V6   6.068 -0.010  -0.010   -0.057   -0.057
## 61            V1 ~~          V9   6.029  0.009   0.009    0.048    0.048
## 111           V9 ~~          V6   5.999  0.010   0.010    0.051    0.051
## 46   neuroticism =~         V24   5.729  0.180   0.041    0.095    0.095
## 71            V3 ~~          V9   5.614 -0.009  -0.009   -0.046   -0.046
## 54         lying =~          V9   5.263 -0.339  -0.046   -0.092   -0.092
## 56            V1 ~~          V5   5.014  0.007   0.007    0.047    0.047
## 57            V1 ~~          V8   4.821  0.017   0.017    0.136    0.136
## 60            V1 ~~          V7   4.784  0.008   0.008    0.046    0.046
## 117           V6 ~~         V24   4.689  0.010   0.010    0.051    0.051
## 34  extraversion =~          V9   4.329  0.401   0.021    0.042    0.042
```

```
## 69            V3 ~~          V4   3.827  0.008   0.008    0.035    0.035
## 37  extraversion =~         V18   3.057 -0.325  -0.017   -0.046   -0.046
## 106           V7 ~~          V9   2.624 -0.017  -0.017   -0.112   -0.112
## 83            V5 ~~         V18   2.479  0.004   0.004    0.031    0.031
## 96            V2 ~~          V6   2.361  0.006   0.006    0.030    0.030
## 88            V8 ~~          V9   2.253  0.005   0.005    0.046    0.046
## 94            V2 ~~          V7   2.142  0.012   0.012    0.071    0.071
## 92            V8 ~~         V24   2.050  0.006   0.006    0.049    0.049
## 55            V1 ~~          V3   1.617 -0.005  -0.005   -0.023   -0.023
## 43   neuroticism =~          V6   1.585  0.098   0.022    0.045    0.045
## 49         lying =~          V5   1.582  0.116   0.016    0.042    0.042
## 98            V2 ~~         V18   1.192 -0.003  -0.003   -0.022   -0.022
## 65            V1 ~~         V24   1.135  0.004   0.004    0.020    0.020
## 120          V18 ~~         V24   1.004 -0.003  -0.003   -0.024   -0.024
## 110           V7 ~~         V24   0.949  0.004   0.004    0.024    0.024
## 114           V9 ~~         V24   0.942 -0.004  -0.004   -0.021   -0.021
## 63            V1 ~~         V12   0.922  0.003   0.003    0.018    0.018
## 115           V6 ~~         V12   0.905  0.004   0.004    0.021    0.021
## 81            V5 ~~          V6   0.722  0.003   0.003    0.016    0.016
## 100           V4 ~~          V7   0.697 -0.004  -0.004   -0.022   -0.022
## 38  extraversion =~         V24   0.639  0.185   0.010    0.022    0.022
## 44   neuroticism =~         V12   0.585  0.049   0.011    0.028    0.028
## 62            V1 ~~          V6   0.573  0.003   0.003    0.014    0.014
## 80            V5 ~~          V9   0.511 -0.002  -0.002   -0.014   -0.014
## 119          V12 ~~         V24   0.501 -0.003  -0.003   -0.017   -0.017
## 95            V2 ~~          V9   0.439  0.004   0.004    0.024    0.024
## 101           V4 ~~          V9   0.432 -0.003  -0.003   -0.014   -0.014
## 93            V2 ~~          V4   0.420  0.003   0.003    0.013    0.013
## 41   neuroticism =~          V5   0.401 -0.022  -0.005   -0.014   -0.014
## 72            V3 ~~          V6   0.398 -0.003  -0.003   -0.012   -0.012
## 78            V5 ~~          V4   0.355  0.002   0.002    0.011    0.011
## 77            V5 ~~          V2   0.290  0.002   0.002    0.010    0.010
## 36  extraversion =~         V12   0.273 -0.105  -0.005   -0.014   -0.014
## 85            V8 ~~          V2   0.267  0.002   0.002    0.015    0.015
## 105           V4 ~~         V24   0.227  0.002   0.002    0.009    0.009
## 31  extraversion =~          V2   0.206  0.090   0.005    0.009    0.009
## 91            V8 ~~         V18   0.191 -0.001  -0.001   -0.014   -0.014
## 102           V4 ~~          V6   0.158 -0.002  -0.002   -0.007   -0.007
## 97            V2 ~~         V12   0.143 -0.001  -0.001   -0.007   -0.007
## 73            V3 ~~         V12   0.130 -0.001  -0.001   -0.007   -0.007
## 82            V5 ~~         V12   0.115  0.001   0.001    0.007    0.007
## 118          V12 ~~         V18   0.109 -0.001  -0.001   -0.007   -0.007
## 90            V8 ~~         V12   0.107 -0.001  -0.001   -0.011   -0.011
## 67            V3 ~~          V8   0.102 -0.003  -0.003   -0.022   -0.022
## 79            V5 ~~          V7   0.059 -0.001  -0.001   -0.005   -0.005
## 112           V9 ~~         V12   0.054 -0.001  -0.001   -0.005   -0.005
## 108           V7 ~~         V12   0.023 -0.001  -0.001   -0.004   -0.004
## 104           V4 ~~         V18   0.011  0.000   0.000   -0.002   -0.002
## 109           V7 ~~         V18   0.000  0.000   0.000    0.000    0.000
```

```
#Update the model specification code to include the largest mi value.
# Edit the model specification
epi.model2 <- 'extraversion =~ V1 + V3 + V5 + V8
neuroticism =~ V2 + V4 + V7 + V9
```

```
lying =~ V6 + V12 + V18 + V24
neuroticism =~ V3'
```

```
# Reanalyze the model
epi.fit2 <- cfa(model = epi.model2, data = epi)

# Summarize the updated model
summary(epi.fit2, standardized = TRUE, fit.measures = TRUE)
```

```
## lavaan 0.6-6 ended normally after 126 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         28
##
##                                                   Used       Total
##   Number of observations                          3193        3570
##
## Model Test User Model:
##
##   Test statistic                               332.891
##   Degrees of freedom                                50
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              2196.019
##   Degrees of freedom                                66
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.867
##   Tucker-Lewis Index (TLI)                       0.825
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -23082.232
##   Loglikelihood unrestricted model (H1)     -22915.787
##
##   Akaike (AIC)                               46220.465
##   Bayesian (BIC)                             46390.389
##   Sample-size adjusted Bayesian (BIC)        46301.421
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.042
##   90 Percent confidence interval - lower         0.038
##   90 Percent confidence interval - upper         0.046
##   P-value RMSEA <= 0.05                          0.999
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.040
```

```
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion =~
##     V1                1.000                               0.068    0.152
##     V3                1.798    0.325    5.532    0.000    0.123    0.246
##     V5               -2.268    0.360   -6.291    0.000   -0.155   -0.414
##     V8                5.077    0.887    5.725    0.000    0.346    0.732
##   neuroticism =~
##     V2                1.000                               0.222    0.445
##     V4                0.432    0.053    8.134    0.000    0.096    0.194
##     V7                1.493    0.093   16.025    0.000    0.331    0.675
##     V9                1.186    0.074   15.938    0.000    0.263    0.530
##   lying =~
##     V6                1.000                               0.135    0.272
##     V12              -0.851    0.127   -6.699    0.000   -0.115   -0.290
##     V18              -0.799    0.119   -6.728    0.000   -0.108   -0.294
##     V24               1.115    0.157    7.087    0.000    0.151    0.347
##   neuroticism =~
##     V3               -0.732    0.066  -11.074    0.000   -0.163   -0.327
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   extraversion ~~
##     neuroticism       0.004    0.001    4.953    0.000    0.283    0.283
##     lying            -0.003    0.001   -4.380    0.000   -0.346   -0.346
##   neuroticism ~~
##     lying            -0.016    0.002   -7.337    0.000   -0.521   -0.521
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .V1                0.196    0.005   39.250    0.000    0.196    0.977
##    .V3                0.217    0.006   34.642    0.000    0.217    0.878
##    .V5                0.116    0.004   29.066    0.000    0.116    0.828
##    .V8                0.104    0.014    7.603    0.000    0.104    0.465
##    .V2                0.200    0.006   33.875    0.000    0.200    0.802
##    .V4                0.235    0.006   39.046    0.000    0.235    0.962
##    .V7                0.131    0.007   19.577    0.000    0.131    0.544
##    .V9                0.178    0.006   29.830    0.000    0.178    0.720
##    .V6                0.228    0.007   34.969    0.000    0.228    0.926
##    .V12               0.144    0.004   34.186    0.000    0.144    0.916
##    .V18               0.123    0.004   34.035    0.000    0.123    0.914
##    .V24               0.166    0.005   31.188    0.000    0.166    0.879
##     extraversion      0.005    0.001    3.265    0.001    1.000    1.000
##     neuroticism       0.049    0.005   10.127    0.000    1.000    1.000
##     lying             0.018    0.004    4.651    0.000    1.000    1.000
```

Your fit indices should improve to the .80s by including this one extra parameter to the model.

(Now, CFI= 0.867, TLI=0.825)

(Before, CFI= 0.750, TLI =0.683)

#COMPARE TWO MODELS

The original model epi.model and the updated model with the modified path epi.model2 can now be compared using the anova() function to determine if the change in fit indices was a large change.

We can use the anova() function because these models are nested, which means they are the same manifest variables with different parameters.

```
# Analyze the original model
epi.fit <- cfa(model = epi.model, data =epi)

# Analyze the updated model
epi.fit2 <- cfa(model = epi.model2, data = epi)

# Compare those models
anova(epi.fit,epi.fit2)
```

```
## Chi-Squared Difference Test
##
##          Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## epi.fit2 50 46220 46390 332.89
## epi.fit  52 46468 46626 584.72     251.83       2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The updated model appears better, as the chi-square difference test is significant. (***)

```
#Select Specific Fit Indices
#You can also compare models by using the AIC or ECVI fit indices, rather than the anova() function.
#These fit indices are very useful if your models include different manifest variables.
#When comparing sets of AIC or ECVI values, the best model would have the smallest fit index.

# Find the fit indices for the original model
fitmeasures(epi.fit, c('aic', 'ecvi'))
```

```
##       aic      ecvi
## 46468.291     0.199
```

```
# Find the fit indices for the updated model
fitmeasures(epi.fit2, c('aic', 'ecvi'))
```

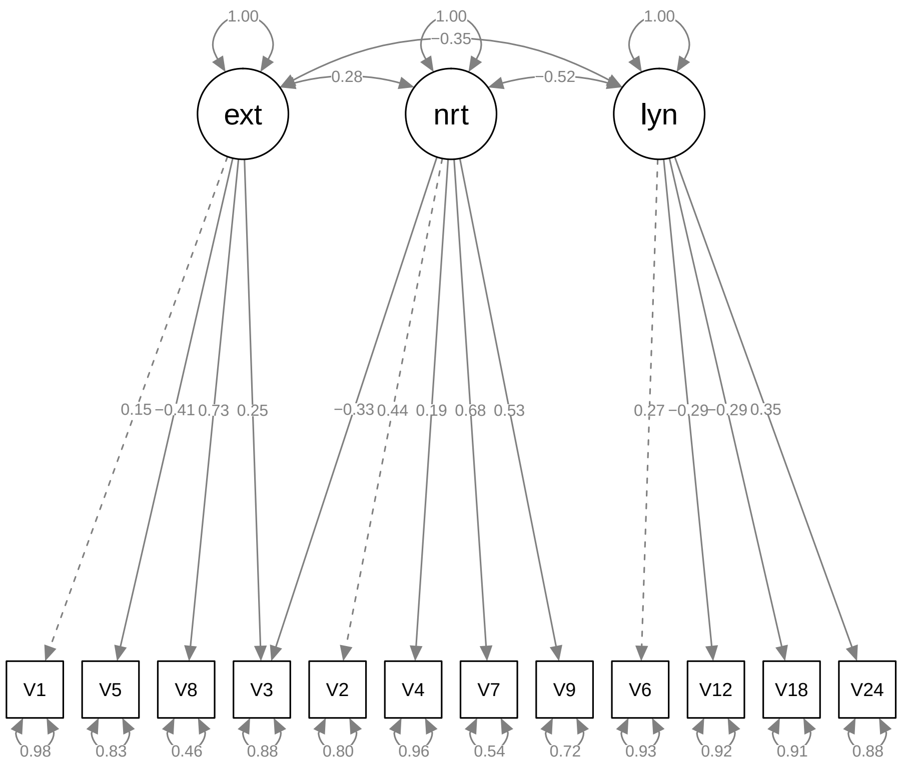
```
##       aic      ecvi
## 46220.465     0.122
```

For both AIC and ECVI, the updated model included the smaller fit indices and would be considered the better model.

```
library(semPlot)
```

```
## Registered S3 methods overwritten by 'huge':
##   method    from
##   plot.sim  BDgraph
##   print.sim BDgraph
```

```
semPaths(epi.fit2, whatLabels = 'std', rotation=1)
```


ext
nrt
lyn
1.00
1.00
1.00
−0.35
0.28
−0.52
0.15
−0.41
0.73
0.25
−0.33
0.44
0.19
0.68
0.53
0.27
−0.29
−0.29
0.35
V1
V5
V8
V3
V2
V4
V7
V9
V6
V12
V18
V24
0.98
0.83
0.46
0.88
0.80
0.96
0.54
0.72
0.93
0.92
0.91
0.88