da410\_project08

jason grahn

3/9/2019

head(HolzingerSwineford1939,5)

## id sex ageyr agemo school grade x1 x2 x3 x4 x5  
## 1 1 1 13 1 Pasteur 7 3.333333 7.75 0.375 2.333333 5.75  
## 2 2 2 13 7 Pasteur 7 5.333333 5.25 2.125 1.666667 3.00  
## 3 3 2 13 1 Pasteur 7 4.500000 5.25 1.875 1.000000 1.75  
## 4 4 1 13 2 Pasteur 7 5.333333 7.75 3.000 2.666667 4.50  
## 5 5 2 12 2 Pasteur 7 4.833333 4.75 0.875 2.666667 4.00  
## x6 x7 x8 x9  
## 1 1.2857143 3.391304 5.75 6.361111  
## 2 1.2857143 3.782609 6.25 7.916667  
## 3 0.4285714 3.260870 3.90 4.416667  
## 4 2.4285714 3.000000 5.30 4.861111  
## 5 2.5714286 3.695652 6.30 5.916667

HS.model <- ' visual =~ x1 + x2 + x3  
 textual =~ x4 + x5 + x6  
 speed =~ x7 + x8 + x9 '  
HS.model

## [1] " visual =~ x1 + x2 + x3\n textual =~ x4 + x5 + x6\n speed =~ x7 + x8 + x9 "

This code block builds the structured model that will be used by the lavaan technique. The Visual latent variable is measured by the x1, x2, and x3 variables; Textual is measured by x4, x5, and x6; and Speed is measured by x7, x8, and x9. For this example, we aren’t working with any covariances.

fit <- sem(HS.model,   
 data = HolzingerSwineford1939)  
fit

## lavaan 0.6-3 ended normally after 35 iterations  
##   
## Optimization method NLMINB  
## Number of free parameters 21  
##   
## Number of observations 301  
##   
## Estimator ML  
## Model Fit Test Statistic 85.306  
## Degrees of freedom 24  
## P-value (Chi-square) 0.000

SEM stands for Structural Equation Modeling and allows us to apply variable heirarchies. Because we built the structure in the previous code block, in this one we apply the SEM method *to* that model. This output tells us the model fit statististic (85.306) the degrees of freedom being used to test it; and the chi-square p-value of that fit. We have a brilliant fit with this as we see in the 0 p-value.

summary(fit,   
 standardized = TRUE,  
 fit.measures = TRUE)

## lavaan 0.6-3 ended normally after 35 iterations  
##   
## Optimization method NLMINB  
## Number of free parameters 21  
##   
## Number of observations 301  
##   
## Estimator ML  
## Model Fit Test Statistic 85.306  
## Degrees of freedom 24  
## P-value (Chi-square) 0.000  
##   
## Model test baseline model:  
##   
## Minimum Function Test Statistic 918.852  
## Degrees of freedom 36  
## P-value 0.000  
##   
## User model versus baseline model:  
##   
## Comparative Fit Index (CFI) 0.931  
## Tucker-Lewis Index (TLI) 0.896  
##   
## Loglikelihood and Information Criteria:  
##   
## Loglikelihood user model (H0) -3737.745  
## Loglikelihood unrestricted model (H1) -3695.092  
##   
## Number of free parameters 21  
## Akaike (AIC) 7517.490  
## Bayesian (BIC) 7595.339  
## Sample-size adjusted Bayesian (BIC) 7528.739  
##   
## Root Mean Square Error of Approximation:  
##   
## RMSEA 0.092  
## 90 Percent Confidence Interval 0.071 0.114  
## P-value RMSEA <= 0.05 0.001  
##   
## Standardized Root Mean Square Residual:  
##   
## SRMR 0.065  
##   
## 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  
## visual =~   
## x1 1.000 0.900 0.772  
## x2 0.554 0.100 5.554 0.000 0.498 0.424  
## x3 0.729 0.109 6.685 0.000 0.656 0.581  
## textual =~   
## x4 1.000 0.990 0.852  
## x5 1.113 0.065 17.014 0.000 1.102 0.855  
## x6 0.926 0.055 16.703 0.000 0.917 0.838  
## speed =~   
## x7 1.000 0.619 0.570  
## x8 1.180 0.165 7.152 0.000 0.731 0.723  
## x9 1.082 0.151 7.155 0.000 0.670 0.665  
##   
## Covariances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## visual ~~   
## textual 0.408 0.074 5.552 0.000 0.459 0.459  
## speed 0.262 0.056 4.660 0.000 0.471 0.471  
## textual ~~   
## speed 0.173 0.049 3.518 0.000 0.283 0.283  
##   
## Variances:  
## Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
## .x1 0.549 0.114 4.833 0.000 0.549 0.404  
## .x2 1.134 0.102 11.146 0.000 1.134 0.821  
## .x3 0.844 0.091 9.317 0.000 0.844 0.662  
## .x4 0.371 0.048 7.779 0.000 0.371 0.275  
## .x5 0.446 0.058 7.642 0.000 0.446 0.269  
## .x6 0.356 0.043 8.277 0.000 0.356 0.298  
## .x7 0.799 0.081 9.823 0.000 0.799 0.676  
## .x8 0.488 0.074 6.573 0.000 0.488 0.477  
## .x9 0.566 0.071 8.003 0.000 0.566 0.558  
## visual 0.809 0.145 5.564 0.000 1.000 1.000  
## textual 0.979 0.112 8.737 0.000 1.000 1.000  
## speed 0.384 0.086 4.451 0.000 1.000 1.000

There’s a LOT of output to this. The latent variables have been standardized and each of their estimates have been provided, as well as the measurements that comprise them. Down the line, we see statistically significant p-values, which is great. The Covariances are included, showing us that we have low estimates for each of those.

CFI = 0.931 TLI = 0.896 RMSEA = 0.092 SRMR = 0.065

While CFI is not greater than 0.95, it IS fairly close; unfortuntely, TLI also falls a little short. RMSEA is a more important measure though, and we want this less than 0.05. We can see that we are at 0.092. This is greater than 0.05. We’d also like to see SRMR at less than 0.06. We are unfortunately higher than that as well.

Since all of these are at inappropriate levels, we can determine this is not an ideal model; that the model fit is *not* good, despite passing the chi-square test.

Reference: <https://personality-project.org/r/tutorials/summerschool.14/rosseel_sem_intro.pdf>