Project 8: Confirmatory factor analysis

DA 410

Marjorie Blanco

Find an example on which a CFA analysis was performed using different library than Lavaan. Your task is to replicate the results, using the built-in dataset HolzingerSwineford1939 and the model which shows in Confirmatory Factor Analysis in R.pdf.

Write a short report of the previous analysis on how it defines its model and what the conclusion it gets (200 – 400 words). And show your replicated results with an interpretation.

The Holzinger and Swineford (1939) dataset consists of mental ability test scores of and grade children from two different schools (Pasteur and Grant-White). This dataset contains 9 variables:

x1: Visual perception

x2: Cubes

x3: Lozenges

x4: Paragraph comprehension

x5: Sentence completion

x6: Word meaning

x7: Speeded addition

x8: Speeded counting of dots

x9: Speeded discrimination straight and curved capitals

This figure is a graphical representation of a measurement model for the Holzinger and Swineford (1939) dataset consisting of scores of 301 school children on nine different tests. In this model the performance on the nine different tests is explained by three latent constructs: speed, textual, and visual.

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

There are three latent factors referring to students’ mental ability: visual, textual, and speed. The latent factors themselves are not directly measured but the assumption is that the nine observed variables are indicators of the latent factors. The visual latent factor is measured by x1, x2 and x3. The textual latent factor is measured by x4, x5, and x6. The speed latent factor is measured by x7, x8, and x9.

Table 1: Descriptive Statistics for Observed Variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Mean | SD | min | max | % Missing |
| x1 | 4.96 | 1.14 | 0.67 | 8.50 | 3.32 |
| x2 | 6.10 | 1.18 | 2.25 | 9.25 | 3.32 |
| x3 | 2.27 | 1.15 | 0.38 | 4.50 | 9.30 |
| x4 | 3.11 | 1.15 | 0.33 | 6.33 | 6.64 |
| x5 | 4.31 | 1.30 | 1.00 | 7.00 | 6.64 |
| x6 | 2.19 | 1.10 | 0.14 | 6.14 | 0.00 |
| x7 | 4.18 | 1.10 | 1.30 | 7.43 | 3.32 |
| x8 | 5.54 | 1.02 | 3.05 | 10.00 | 6.64 |
| x9 | 5.37 | 1.01 | 2.78 | 9.25 | 0.00 |

#load data  
data(HolzingerSwineford1939)  
#extract relevant variables x1 to x9  
data <- HolzingerSwineford1939[,7:15]

## Confirmatory factor analysis, using the sem() function

HS.S <- stats::cov(HolzingerSwineford1939[,7:15], use = "pairwise")  
  
HS.nobs <- nrow(HS.S)  
cfa.model <- specifyModel("CFA.r")  
cfa.model

## Path Parameter StartValue  
## 1 Visual -> x1 theta01   
## 2 Visual -> x2 theta02   
## 3 Visual -> x3 theta03   
## 4 Textual -> x4 theta04   
## 5 Textual -> x5 theta05   
## 6 Textual -> x6 theta06   
## 7 Speed -> x7 theta07   
## 8 Speed -> x8 theta08   
## 9 Speed -> x9 theta09   
## 10 x1 <-> x1 theta10   
## 11 x2 <-> x2 theta11   
## 12 x3 <-> x3 theta12   
## 13 x4 <-> x4 theta13   
## 14 x5 <-> x5 theta14   
## 15 x6 <-> x6 theta15   
## 16 x7 <-> x7 theta16   
## 17 x8 <-> x8 theta17   
## 18 x9 <-> x9 theta18   
## 19 Visual <-> Visual <fixed> 1   
## 20 Textual <-> Textual <fixed> 1   
## 21 Speed <-> Speed <fixed> 1

sem.fit <- sem::sem(cfa.model, HS.S, HS.nobs)  
sem.fit

##   
## Model Chisquare = 4.080454 Df = 27   
##   
## theta01 theta02 theta03 theta04 theta05 theta06 theta07   
## 0.7248950 0.5638459 0.8026435 0.9860081 1.1170427 0.9112517 0.6620916   
## theta08 theta09 theta10 theta11 theta12 theta13 theta14   
## 0.8111108 0.5656946 0.8374250 1.0684671 0.6348779 0.3829556 0.4175353   
## theta15 theta16 theta17 theta18   
## 0.3699675 0.7487185 0.3674891 0.6983773   
##   
## Iterations = 27

### Parameter Estimates

Here is the complete list of all of the parameters in the model.

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

##   
## Model Chisquare = 4.080454 Df = 27 Pr(>Chisq) = 0.9999999  
## AIC = 40.08045  
## BIC = -55.24461  
##   
## Normalized Residuals  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.2140240 -0.0000001 0.3023541 0.3403381 0.5590696 1.1040491   
##   
## R-square for Endogenous Variables  
## x1 x2 x3 x4 x5 x6 x7 x8 x9   
## 0.3856 0.2293 0.5037 0.7174 0.7493 0.6918 0.3693 0.6416 0.3142   
##   
## Parameter Estimates  
## Estimate Std Error z value Pr(>|z|)   
## theta01 0.7248950 0.5528521 1.3111917 0.189793023 x1 <--- Visual   
## theta02 0.5638459 0.5051216 1.1162577 0.264311861 x2 <--- Visual   
## theta03 0.8026435 0.5707436 1.4063118 0.159631561 x3 <--- Visual   
## theta04 0.9860081 0.3499981 2.8171811 0.004844720 x4 <--- Textual  
## theta05 1.1170427 0.3846748 2.9038622 0.003685904 x5 <--- Textual  
## theta06 0.9112517 0.3316441 2.7476797 0.006001860 x6 <--- Textual  
## theta07 0.6620916 0.4492414 1.4737993 0.140535640 x7 <--- Speed   
## theta08 0.8111108 0.4564704 1.7769188 0.075581605 x8 <--- Speed   
## theta09 0.5656946 0.4078045 1.3871711 0.165389599 x9 <--- Speed   
## theta10 0.8374250 0.7271184 1.1517038 0.249442810 x1 <--> x1   
## theta11 1.0684671 0.6440193 1.6590608 0.097103550 x2 <--> x2   
## theta12 0.6348779 0.7949449 0.7986439 0.424496903 x3 <--> x3   
## theta13 0.3829556 0.3009759 1.2723798 0.203238196 x4 <--> x4   
## theta14 0.4175353 0.3638786 1.1474577 0.251192514 x5 <--> x5   
## theta15 0.3699675 0.2712124 1.3641244 0.172528413 x6 <--> x6   
## theta16 0.7487185 0.5309213 1.4102252 0.158473211 x7 <--> x7   
## theta17 0.3674891 0.5941402 0.6185225 0.536230980 x8 <--> x8   
## theta18 0.6983773 0.4443689 1.5716159 0.116039649 x9 <--> x9   
##   
## Iterations = 27

sem.fit$criterion

## [1] 0.5100568

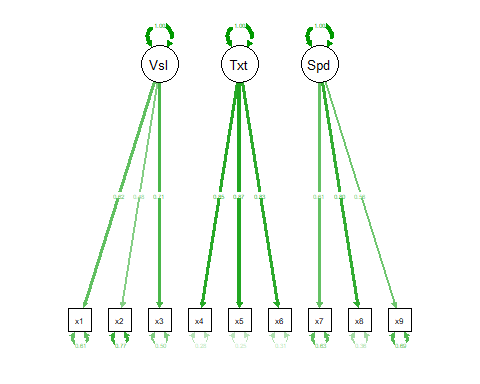
#modification indices  
modIndices(sem.fit)

##   
## 5 largest modification indices, A matrix (regression coefficients):  
## Visual<-x9 Textual<-x1 x9<-Visual x9<-x1 x1<-x4   
## 1.3816572 1.1226552 1.0642987 0.9032045 0.8392840   
##   
## 5 largest modification indices, P matrix (variances/covariances):  
## Visual<->x9 Textual<->x1 Textual<->Visual Speed<->Visual   
## 1.0642987 0.8365412 0.7056969 0.6015507   
## x9<->x1   
## 0.3826238

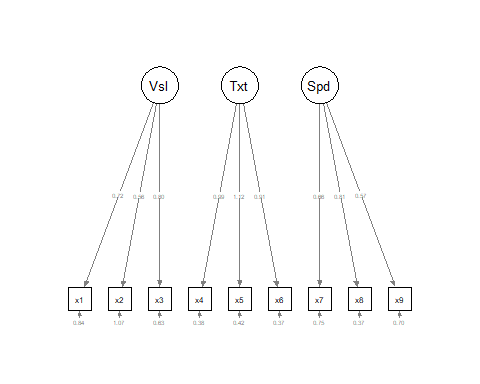
The chi-squared actual value > 0.05 indicating this is not a good fit.

Figure 1: Measurement model for the Holzinger and Swinford Data including parameter estimates

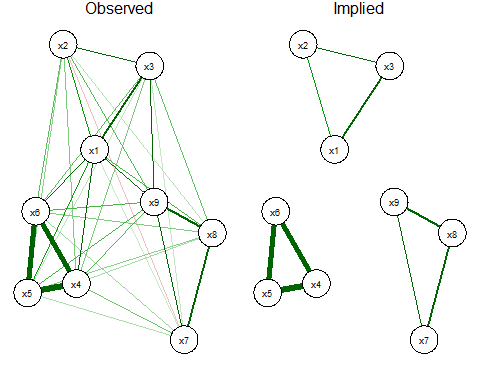
semPaths(sem.fit, what="std")



semPaths(sem.fit, "model", "est", style = "lisrel")



semCors(sem.fit, layout = "spring", cut = 0.3, esize = 20, titles = TRUE)



base.model <- specify.model("CFA-Full.r")  
base.model

## Path Parameter StartValue  
## 1 Visual -> x1 theta01   
## 2 Visual -> x2 theta02   
## 3 Visual -> x3 theta03   
## 4 Textual -> x4 theta04   
## 5 Textual -> x5 theta05   
## 6 Textual -> x6 theta06   
## 7 Speed -> x7 theta07   
## 8 Speed -> x8 theta08   
## 9 Speed -> x9 theta09   
## 10 x1 <-> x1 theta10   
## 11 x2 <-> x2 theta11   
## 12 x3 <-> x3 theta12   
## 13 x4 <-> x4 theta13   
## 14 x5 <-> x5 theta14   
## 15 x6 <-> x6 theta15   
## 16 x7 <-> x7 theta16   
## 17 x8 <-> x8 theta17   
## 18 x9 <-> x9 theta18   
## 19 Visual <-> Visual <fixed> 1   
## 20 Textual <-> Textual <fixed> 1   
## 21 Speed <-> Speed <fixed> 1   
## 22 Visual <-> Textual theta19   
## 23 Visual <-> Speed theta20   
## 24 Textual <-> Speed theta21

base.model.fit <- sem::sem(base.model, HS.S, HS.nobs)  
base.model.fit

##   
## Model Chisquare = 2.267256 Df = 24   
##   
## theta01 theta02 theta03 theta04 theta05 theta06 theta07   
## 0.9011177 0.4987686 0.6572487 0.9913406 1.1034382 0.9181265 0.6205055   
## theta08 theta09 theta10 theta11 theta12 theta13 theta14   
## 0.7321656 0.6710955 0.5508844 1.1376195 0.8471387 0.3724103 0.4477425   
## theta15 theta16 theta17 theta18 theta19 theta20 theta21   
## 0.3573900 0.8020564 0.4893231 0.5680181 0.4585089 0.4705347 0.2829845   
##   
## Iterations = 28

### Parameter Estimates

Here is the complete list of all of the parameters in the model.

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

##   
## Model Chisquare = 2.267256 Df = 24 Pr(>Chisq) = 1  
## AIC = 44.26726  
## BIC = -50.46613  
##   
## Normalized Residuals  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.53168 -0.10020 -0.02108 -0.01290 0.10252 0.40914   
##   
## R-square for Endogenous Variables  
## x1 x2 x3 x4 x5 x6 x7 x8 x9   
## 0.5958 0.1794 0.3377 0.7252 0.7311 0.7023 0.3243 0.5228 0.4422   
##   
## Parameter Estimates  
## Estimate Std Error z value Pr(>|z|)   
## theta01 0.9011177 0.4967327 1.8140898 0.069663898 x1 <--- Visual   
## theta02 0.4987686 0.4758935 1.0480676 0.294607464 x2 <--- Visual   
## theta03 0.6572487 0.4572545 1.4373806 0.150609887 x3 <--- Visual   
## theta04 0.9913406 0.3479837 2.8488138 0.004388256 x4 <--- Textual   
## theta05 1.1034382 0.3850883 2.8654161 0.004164617 x5 <--- Textual   
## theta06 0.9181265 0.3296850 2.7848596 0.005355091 x6 <--- Textual   
## theta07 0.6205055 0.4275251 1.4513897 0.146671383 x7 <--- Speed   
## theta08 0.7321656 0.4049562 1.8080121 0.070604620 x8 <--- Speed   
## theta09 0.6710955 0.3994734 1.6799505 0.092966956 x9 <--- Speed   
## theta10 0.5508844 0.6991422 0.7879433 0.430729868 x1 <--> x1   
## theta11 1.1376195 0.6260435 1.8171572 0.069193049 x2 <--> x2   
## theta12 0.8471387 0.5577288 1.5189079 0.128785687 x3 <--> x3   
## theta13 0.3724103 0.2936729 1.2681127 0.204757707 x4 <--> x4   
## theta14 0.4477425 0.3593706 1.2459072 0.212798473 x5 <--> x5   
## theta15 0.3573900 0.2648531 1.3493897 0.177211815 x6 <--> x6   
## theta16 0.8020564 0.5008522 1.6013833 0.109292043 x7 <--> x7   
## theta17 0.4893231 0.4566177 1.0716254 0.283888336 x8 <--> x8   
## theta18 0.5680181 0.4353412 1.3047653 0.191972781 x9 <--> x9   
## theta19 0.4585089 0.3912178 1.1720044 0.241195301 Textual <--> Visual  
## theta20 0.4705347 0.4467133 1.0533258 0.292191685 Speed <--> Visual   
## theta21 0.2829845 0.4215801 0.6712474 0.502062924 Speed <--> Textual   
##   
## Iterations = 28

base.model.fit$criterion

## [1] 0.283407

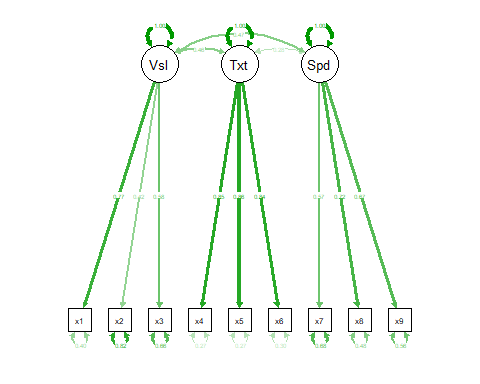
#modification indices  
modIndices(base.model.fit)

##   
## 5 largest modification indices, A matrix (regression coefficients):  
## x9<-Visual Speed<-x9 x8<-x7 x7<-x8 Visual<-x9   
## 0.9677343 0.9075159 0.9075108 0.9075107 0.7677746   
##   
## 5 largest modification indices, P matrix (variances/covariances):  
## Speed<->x9 x8<->x7 Visual<->x9 Visual<->x7 x9<->x8   
## 0.9075127 0.9075117 0.7677743 0.5672536 0.3972465

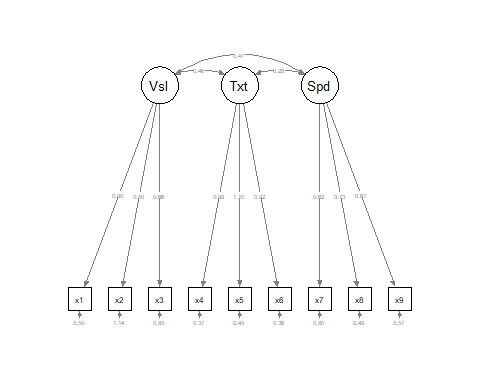
The chi-squared actual value > 0.05 indicating this is not a good fit.

Figure 3: Measurement model for the Holzinger and Swinford Data including parameter estimates

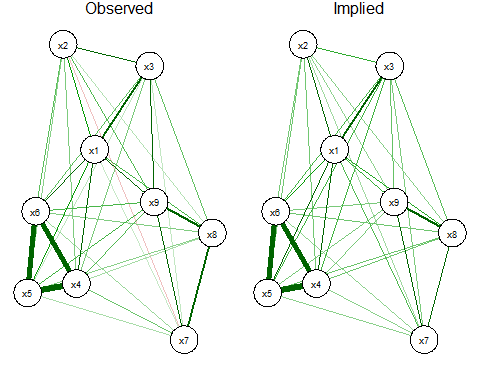
semPaths(base.model.fit, what="std")



semPaths(base.model.fit, "model", "est", style = "lisrel")



semCors(base.model.fit, layout = "spring", cut = 0.3, esize = 20, titles = TRUE)



## Confirmatory factor analysis, using the cfa() function

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

### Parameter Estimates

Here is the complete list of all of the parameters in the model.

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

The output shows that this model has very weak fit to the data ( = 85.306; DF = 24; CFI = .93; gammaHat = .96; RMSEA = .092; SRMR = 0.065).

CFI (Comparative fit index) measures whether the model fits the data better than a more restricted baseline model. The CFI = 0.931 which is acceptable.

TLI (Tucker-Lewis index) which penalizes complex models measures if the model fits the data better than a more restricted baseline model. The TLI = 0.896 which is acceptable.

Figure 2: Measurement model for the Holzinger and Swinford Data including parameter estimates

semPaths(cfa.fit, what="std")

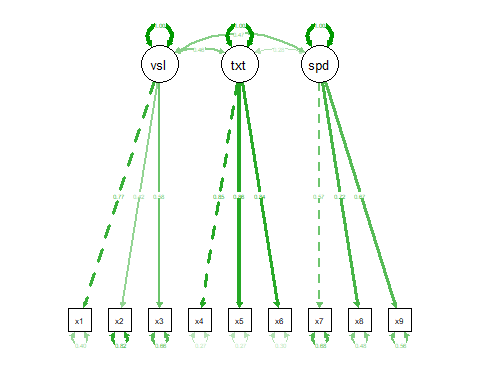


Table 2: Factor Loadings

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Latent Factor | Indicator | B | SE | Z | Beta | sig |
| visual | x1 | 1.000 | 0.000 |  | 0.772 |  |
| visual | x2 | 0.554 | 0.100 | 5.554 | 0.424 | \*\*\* |
| visual | x3 | 0.729 | 0.109 | 6.685 | 0.581 | \*\*\* |
| textual | x4 | 1.000 | 0.000 |  | 0.852 |  |
| textual | x5 | 1.113 | 0.065 | 17.014 | 0.855 | \*\*\* |
| textual | x6 | 0.926 | 0.055 | 16.703 | 0.838 | \*\*\* |
| speed | x7 | 1.000 | 0.000 |  | 0.570 |  |
| speed | x8 | 1.180 | 0.165 | 7.152 | 0.723 | \*\*\* |
| speed | x9 | 1.082 | 0.151 | 7.155 | 0.665 | \*\*\* |

Table 3: Latent Factor Correlations

|  |  |  |  |
| --- | --- | --- | --- |
| Factor 1 | Factor 2 | Correlation | sig |
| visual | visual | 0.809 | \*\*\* |
| textual | textual | 0.979 | \*\*\* |
| speed | speed | 0.384 | \*\*\* |
| visual | textual | 0.408 | \*\*\* |
| visual | speed | 0.262 | \*\*\* |
| textual | speed | 0.173 | \*\*\* |

moreFitIndices(cfa.fit)

## gammaHat adjGammaHat baseline.rmsea aic.smallN bic.priorN   
## 0.9566992 0.9188110 0.2854364 7520.8016812 7595.4088206   
## hqc sic   
## 7548.6417914 7588.1835282

References:

Holzinger, K., and Swineford, F. (1939). A study in factor analysis: The stability of a bifactor solution. Supplementary Educational Monograph, no. 48. Chicago: University of Chicago Press.

Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-202.