LUIS FELIPE VILLOTA

LUND UNIVERSITY
QDA USING R - SIMM61

LAB ASSIGNMENT 4 - SEM: HOLZINGER AND SWINEFORD DATASET

Introduction

Our objective in the present report is to to evaluate the fit of a theoretical factor structure of intelligence-test indicators and their relationships between different test scores based on the Holzinger and Swineford dataset through the structural equation modeling statistical technique. It is a theory-based confirmatory research which, in this exercise, covers (3) established latent and exogenous factors of *Visual Perception Ability*, *Processing Speed*, and *Verbal Ability* underlying and causally determining the scores of participants in 10 different tests (endogenous and manifest variables) (below described).

Data and Methods

Data cleaning. In the full database we have 301 observations and 33 variables. We had 312 missing values in total in variables "*t25_frmbord2*" and "*t26_flags*". We didn't remove any *NA*s, because it would've reduced the observations considerably and those were not variables of interest here. We renamed variable "*female*" into "*sex*".

Model A specification. We built Model A according to the theoretical assumptions in which the measurement model was composed of three (3) exogenous variables (here latent factors) and 10 endogenous variables (manifest variables) in the following manner: a.) *Visual perception* causally determines test scores in "*t01_visperc*", "*t02_cubes*", "*t03_frmbord*", and "*t04_lozenges*"; b.) *Verbal ability* causally determines test scores in "*t06_paracomp*", "*t07_sentcomp*", and "*t09_wordmean*"; c.) *Processing speed* causally determines test scores in "*t10_addition*", "*t12_countdot*" and "*t13_sccaps"*. Finally, these three exogenous variables (latent factors) are correlated to one another.

Checking assumptions. All multivariate normality tests (*Henze-Zirkler*, *MVN based on kurtosis*, *MVN based on skewness*) show significant p values (< 0.05), so the assumption of MVN is violated. For our model, it means that the model fit indices are not reliable. In this line, we used the Maximum likelihood estimator with robust SE and test statistics to find the most likely solution for Model A. Although this estimator assumes MVN, it allows for unstandardized observed variables and it is scale-free (the test scores have different scales).

We proceeded with the normality-adjusted robust variants of the statistics: The model Chi-squared = 0.00, Satorra-Bentler (corrects Chi-squared fit) = 1.047, degrees of freedom = 32, p-value = 0.00, TLI = 0.917, CFI = 0.941 and RMSEA =0.075, with 90 Percent confidence interval in the lower = 0.056, and the upper= 0.095. We find the model is acceptable: is identified df >1; and according to the predominant literature of Hu and Bentler's (1999), the model shows acceptable goodness of fit indices: TLI < 0.95, CFI < 0.95, SRMR = 0.062 < 0.08.

Model B specification. We included in our previous model a suspected correlation between "*t10_addition*" and "*t12_countdot*" (not determined by *Processing Speed*). This was due to the realization that these two variables had a similar structure (distinct from the one of "*t13_sccaps*") so it was deduced that after accounting for the influence of *Processing Speed,* some possible correlation would remain between the mentioned two variables above.

Results

Comparison of Model A and Model B. Model A (AIC= 8296.856, BIC= 8382.120, SABIC= 8309.177) and Model B (AIC= 8267.652, BIC= 8356.623, SABIC = 8280.508). There is a significant difference between the model fits, the AIC of both have a difference >2. The accepted criteria regarding this indicator tells us to prefer the model with the smallest AIC, which in this case is Model B. Regarding BIC and SABIC, and following a similar logic of comparison, we have again Model B as a better fit for the data.

Influence of the Visual Perception Ability factor in Model B. For this final model, we interpret the standardized solution estimates. We found that "*t02_cubes" (Cubes, Simplification of Brighams Spatial Relations Test score)* is the manifest variable that was the least influenced by the Visual Perception Ability factor (latent variable). This is based on its factor loading = 0.442 and its R^2 = 0.195 (only ~20% of variance of the Visual Perception Ability factor explained). In Figure i, we observe the full path diagram of Model B with standardized estimates. The rest of the manifest variables factor loadings regarding Visual perception ability are: "*t01_visperc*" = 0.751, "*t04_lozenges*" = 0.596 and "*t03_frmbord"* = 0.495.

Discussion of Task 3 Path Diagram

We reproduced the path diagram proposed through the creation of **Model C**, in which "*t13_sccaps*" is a manifest and endogenous variable, determined by the endogneous variable "*t01_visperc*" mediated by "*t12_countdot*" (a seen in *Figures ii-iii*). If we look to the unstandardized estimates of the modeled paths, we see that if "*t01_visperc*" is increased by 1 unit, it is foreseen a total change of 0.394 in "*t13_sccaps*". The reasoning behind this claim is based on the model relationships, in which "*t01_visperc"* has a direct effect (labeled "*c"* = 0.308) on the dependent variable. It has also indirect effect on it mediated by "*t12_countdot*" (labeled and in the form of "*a*b"* = 0.379 * 0.227). So, 1-unit increase of "*t01_visperc*" would suppose a total influence *c+(a*b)* = 0.394.

R code: https://github.com/FelipeVillota/SIMM61_QDA-with-R/blob/main/mental.R

Figure i. Plot of Path Diagram of Model B (with standardized estimates)

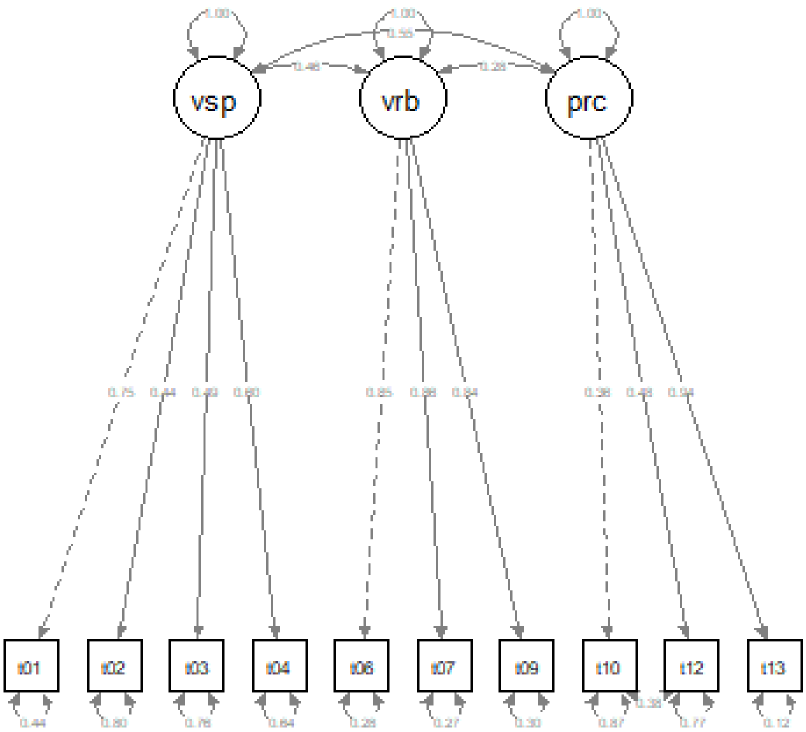

Figure ii. Plot of Path Diagram of Model C (with unstandardized estimates)

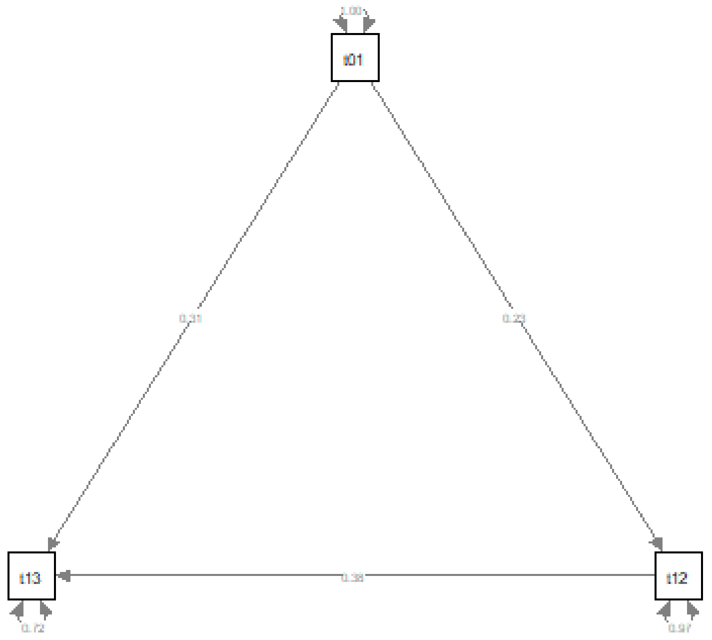

Figure iii. Plot of Path Diagram of Model C (with labels)