1ZM31 Multivariate Statistics (2017/2018)

Group Assignment 3

Report

Multivariate Statistics

Dr. Jan de Jonge

Structural Equation Modeling (SEM)

Group	23

A.W.H (Sander) Berkers	0658745
Bhoomica Mysore Nataraja	1282832
D.F (Diego Fernando) Barreto Trujillo	1280341
Nitish Singh	1283901

Date	<19/10/2017>

PART C

1. Search for the maximum likelihood (ML) ratio chi-square, write it down as well as its degrees of freedom. Further, use multiple key fit indices of differing types as suggested by Hair et al. (2014) and SEM lecture 2, and write them all down in a nice table (i.e. the models tested in the columns, and all necessary fit indices in the rows). Could you say something about the fit of Model 1 (linear), making use of all necessary fit indices?

The analyzed fit indicators and results of the model are shown in Table 1:

	Model 1
Maximum Likelihood Ratio Chi-Square (ML)	83.36
Degrees of Freedom (df)	15
P-Value	! 0.00
Root Mean Square Error of Approximation (RMSEA)	! 0.06
Non-Normed Fit Index (NNFI)	! 0.83
Parsimony Normed Fit Index (PNFI)	0.26
Comparative Fit Index (CFI)	✓ 0.95

Table 1. Fit indicators for model 1

According to table 1, the model does not have a good fitting because from the 4 fit indicators (not taking into account the PNFI because it is a relative indicator), only one, the CFI, supports for a good fit of the model (CNFI>=0.95); this indicator is not conservative so this model is not the most representative of a good fit.

The difference is significant between the observed covariance matrix and the estimated covariance matrix because the p-value is <0.05 (H_o = observed and estimated matrices are equal, rejected), the RMSEA is >0.05 and the NNFI (the most conservative fit indicator) is less than 0.95.

2. The output is showing a warning. Could you please explain this warning, and argument whether or not it could be ignored.

The warning output of the first model run is:

```
WARNING: The Condition Number indicates severe multicollinearity.
         One or more variables may be redundant.
```

Figure 1. Warning of the otuput in LISREL for the model 1

This warning pops-up in LISREL because one or more variables are a linear or almost-linear combination of other variables (Garson, 2007). This warning can be ignored since:

1) SEM is a robust method that compared to Multiple Regression includes more flexible assumptions (allowing interpretation although in presence of multicollinearity) (Garson, 2007).
2) Multicollinearity impacts causality interpretation (Hair, et.al, 2014) but what matters in this part is the quality of the model fitting.

3. Run a second model (Model 2) by adding the non-linear paths from the squared terms to the outcomes (suggestion: add additional lines in the syntax, starting with FR). Have a look at the output of Model 2. Again, write down all required fit indices in the table, Chi-square inclusive. What about model fit of Model 2?

The analyzed fit indicators and results of Model 2 are shown in Table 2:

	Model 2
Maximum Likelihood Ratio Chi-Square (ML)	42.19
Degrees of Freedom (df)	6
P-Value	0.000
Root Mean Square Error of Approximation (RMSEA)	0.067
Non-Normed Fit Index (NNFI)	0.771
Parsimony Normed Fit Index (PNFI)	0.106
Comparative Fit Index (CFI)	0.975

Table 2. Fit Indicators for model 1 and 2

By adding the non-linear paths from the squared terms to the outcomes (LISREL code is shown in APPENDIX 1), the model fit is poor because 3 out of 4 indicators are not meeting the goodness-of-fit criteria, leading to a significant difference between the observed and estimated covariance matrix:

- P-value is <0.05.
- RMSEA>0.05
- NNFI<0.95

Although, the values of RMSEA and NNFI show that model fit of Model 2 is not better, it is important to review the delta on ML ratio Chi-Squares and in degrees of freedom to make a choice on the model.

4. Compute the difference in the ML ratio Chi-squares between Model 1 and Model 2. Do the same for their degrees of freedom. This difference between chi-squares has also a Chi-square distribution itself with a number of degrees of freedom equal to their difference (see common Chi-square tables at statistical text books or at the internet, and use a p-value of .05). What could be concluded from this computation? Which of the two models is the best fitting structural model?

The differences in the ML, df, are computed and shown in the next table:

	Model 1 - Model 2
Maximum Likelihood Ratio Chi-Square (ML) (DELTA)	41.17
Degrees of Freedom (df) (DELTA)	9
Chi Square reference	16.92

Table 3. Differences on Maximum Likelihood and Degrees of Freedom between model 1 and 2

Chi-Square with 9 degrees of freedom and p=0.05 is 16.92, lesser than the delta of chi-square of the two models (41.17) which means that the model 2 has a significant reduction of Chi-Square points with slight decrease in degrees of freedom; thus, model 2 with these criteria has a better goodness of fit.

Although from question 3 it seemed that model 2 was poor-fit, it is better than model 1 because the ML delta is significant and good for model fitting; also, model 1 is a linear model while model 2 has non-linear variables which makes model 2 a better choice since it is analogous to the non-linearity of vitamin model.

5. LISREL is providing some modification indices for potential model improvement. Use the best fitting model, search for the largest modification index in its output file, and write down its Chi-square value and expected parameter change. Based upon this modification index, run a third model (only) by relaxation or fixing of the corresponding parameter. Have a look at the output of this model and write down all required fit indices in the table, Chi-square inclusive. What about model fit?

LISREL outputs of Model 2 was used to check the modification indices as shown in the table below.

Modifications Indices for BETA						
	SAT	EXH	ANX	CAUT	CDEM	CSUP
SAT	- -	- -	- -	- -	- -	- -
EXH	- -	- -	- -	- -	- -	- -
ANX	- -	- -	- -	- -	- -	- -
CAUT	- -	- -	- -	- -	- -	- -
CDEM	- -	- -	- -	- -	- -	- -
CSUP	- -	- -	- -	- -	- -	- -
AUT2	23.106	7.222	1.097	10.328	12.775	20.403
DEM2	1.975	4.874	5.934	4.236	3.900	2.450
SUP2	0.000	0.089	0.071	0.008	0.093	0.000

Expected Changes for BETA						
	SAT	EXH	ANX	CAUT	CDEM	CSUP
SAT	- -	- -	- -	- -	- -	- -
EXH	- -	- -	- -	- -	- -	- -
ANX	- -	- -	- -	- -	- -	- -
CAUT	- -	- -	- -	- -	- -	- -
CDEM	- -	- -	- -	- -	- -	- -
CSUP	- -	- -	- -	- -	- -	- -
AUT2	-0.583	-1.567	-1.106	0.428	0.831	-1.264
DEM2	0.204	-1.544	-3.086	-0.329	0.551	0.525
SUP2	-0.001	-0.064	-0.103	-0.004	0.026	0.000

Modifications Indices for GAMMA		
	SEX	AGE
SAT	- -	- -
EXH	- -	- -
ANX	- -	- -
CAUT	- -	- -
CDEM	- -	- -
CSUP	- -	- -
AUT2	3.987	32.910
DEM2	5.371	0.312
SUP2	0.024	0.058

Expected Changes for GAMMA		
	SEX	AGE
SAT	- -	- -
EXH	- -	- -
ANX	- -	- -
CAUT	- -	- -
CDEM	- -	- -
CSUP	- -	- -
AUT2	-0.069	0.009
DEM2	0.096	0.001
SUP2	0.002	0.000

Figure 2. Output of LISREL for BETA and GAMMA (Modification Indices)

The highest modification index is 32.91 with an expected parameter change of 0.009 that connects AGE (Age) and AUT2 (The squared term of Job Autonomy) in the GAMMA Matrix. To have a model 3, this path is relaxed in LISREL (the code is in APPENDIX 1). The results are presented in Table 4.

	Model 1	Model 2	Model 3
Maximum Likelihood Ratio Chi-Square (ML)	83.36	42.19	8.80
Degrees of Freedom (df)	15	6	5
P-Value	! 0.000	! 0.000	✓ 0.117
Root Mean Square Error of Approximation (RMSEA)	! 0.059	! 0.067	✓ 0.024
Non-Normed Fit Index (NNFI)	! 0.827	! 0.771	✓ 0.971
Parsimony Normed Fit Index (PNFI)	0.258	0.106	0.090
Comparative Fit Index (CFI)	✓ 0.953	✓ 0.975	✓ 0.997

Table 4. Comparison of three models on fitting indicators

Model 3 is the only model that has good fit indicators because:

- p-value>0.05
- RMSEA<0.05
- NNFI>0.95
- CFI>0.95

Also, it is important to check the differences of ML Chi-Square and degrees of freedom among models. This differences and value of Chi Square for de delta of df are shown on Table 5.

	Delta Chi Square (degrees of freedom)		
	Model 1	Model 2	Model 3
Model 1		41.17 (9)*	74.56 (10)*
Model 2			33.39 (1)*
Model 3			
*** Difference is significantly better**			

Table 5. Differences on Maximum Likelihood and Degrees of Freedom between model 1 and 2 and 3

Also, according to Table 5, Model 3 is better than model 1 and model 2 with respect to ML Chi-square because there is a great reduction in Chi-Square points with few decreasing degrees of freedom (being a significant reduction compared with the table of Chi-Square values).

Altogether, not only the delta of ML Chi-Square but also all the important fit indicators support that model 3 is the only model where the observed and estimated covariance matrix are statistically equal.

6. How many significant non-linear relations can be discovered in the output of Model 3? Could you please describe these non-linear relations in more detail? For example, are they U-shaped or inverted U-shaped, and what is the precise relation between the variables of interest? Are the effects found according to theoretical and/or logical assumptions? Suggestion: create and calculate a simple, standardized, regression equation (y=c+x1+(x1)2+e) of each non-linear effect and draw the regression line (c and e could be ignored for practical reasons).

Looking at Figure 33, the output of model 3 shows that in beta matrix, there are 4 significant relations between squared terms and health/well-being (the absolute value of the third value of each shaded square is greater than 1.96, showing significant relation)

	AUT2	DEM2	SUP2
SAT	0.063 (0.048) 1.313	-0.063 (0.040) -1.589	-0.553 (0.132) -4.179
EXH	-0.112 (0.045) -2.465	0.024 (0.037) 0.654	0.300 (0.124) 2.415
ANX	-0.008 (0.026) -0.298	0.068 (0.021) 3.219	0.107 (0.071) 1.515

Figure 3. Maximum Likelihood Estimates for non-linear relations of model 3

It is important to describe each significant relation:

- ***Squared term of Job Autonomy relates to Exhaustion.***

  As the ML Estimate is negative, it is an inverted U-shaped relation showing that low or high autonomy is leading to lesser exhaustion and in moderate autonomy, exhaustion is the highest. Expectations were to have a U-shaped model where low and high autonomy would lead to emotional exhaustion; however, exhaustion is not only emotional but physical, so this could explain somehow the result.

$$y = -0.112x^2 + 0.092x \text{ where } x \text{ is Autonomy and } y \text{ is Exhaustion}$$

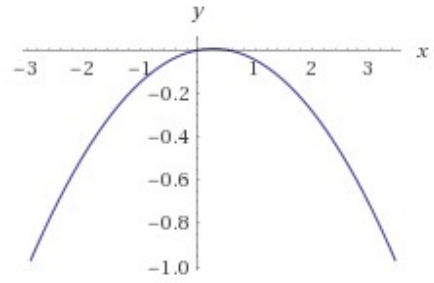

Figure 4. Exhaustion vs. Autonomy[1]

- ***Squared term of Job Demands relates to Job Related Anxiety.***

  As the ML Estimate is positive for this relation, there is a U-shaped relation between job demands and anxiety as shown in Figure 5.

[1] Wolframalpha software output used to plot equations; just for reference of the u shape because c and error are not considered

$$y = 0.068x^2 + 0.114x \; where \; x \; is \; job \; demands \; and \; y \; is \; anxiety$$

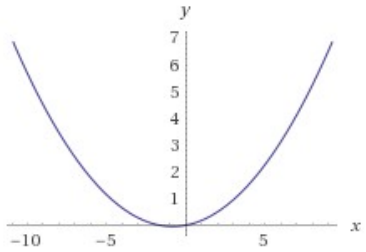

Figure 5. Anxiety vs. Job Demands[2]

This relation suggests that low and high job demands will lead to high anxiety while moderate job demands will lower the level of anxiety. This is logical because at low job demands there is uncertainty (and thus anxiety) if employee will be fired and if there are too many job demands there will be nervousness and anxiety due to time pressure and responsibilities.

- ***Squared term of Social Support relates to Job Satisfaction.***

  As the ML Estimate is positive, it is an inverted U-shaped relation between workplace social support and job satisfaction as shown in Figure 56.

$$y = -0.553x^2 + 0.729x \; where \; x \; is \; Social \; Support \; and \; y \; is \; Job \; Satisfaction$$

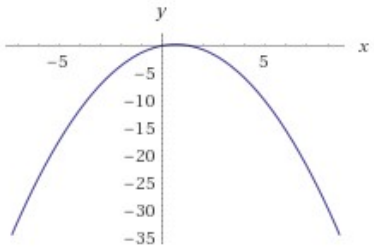

Figure 6. Job Satisfaction vs. Social Support[3]

This relation suggests that low and high social support will lead to lower job satisfaction while moderate support from the co-workers will increase the level of job satisfaction. It is a logical relation because an individual in need of social support would feel alone without guidance and this will lead to less job satisfaction; in the other way, if there is too much social support, this can lead to some type of conflict (difficulty in decision-making) that will reduce job satisfaction.

[2] Wolframalpha software output used to plot equations; just for reference of the u shape because c and error are not considered

[3] Wolframalpha software output used to plot equations; just for reference of the u shape because c and error are not considered

- ***Squared term of Social Support relates to Exhaustion.***

  As the ML Estimate is positive, it is a U-shaped relation between workplace social support and Exhaustion as shown in Figure 57.

$$y = 0.300x^2 - 0.567x \; where \; x \; is \; Social \; Support \; and \; y \; is \; Exhaustion$$

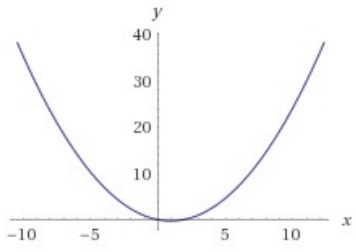

Figure 7. Exhaustion vs. Social Support[4]

This relation suggests that low and high social support will lead to high level of exhaustion while moderate social support will lower exhaustion. It is a logical relation because with lack of social support, all the job demands lie on the individual so more exhaustion; on the other hand, too much social support, as in the past relation, may lead to more conflict and in turn more exhaustion because the individual has to deal and manage conflict, spending unnecessary effort.

7. Please write down the explained variance for each of the outcome variables.

The explained variance (R2) for outcomes in best fitting model are shown in Figure 8. Looking at the 3 dependent variables regarding health and well-being (SAT, EXH, ANX), burnout/exhaustion is the dependent variable which variance is better explained by the model (27.6%, which is significantly good). Job satisfaction (SAT) and job-related anxiety (ANX) have a variance explained of 16.6% and 8.0% respectively by the model. Anxiety (ANX) is not much explained.

Squared Multiple Correlations for Structural Equations								
SAT	EXH	ANX	CAUT	CDEM	CSUP	AUT2	DEM2	SUP2
0.166	0.276	0.080	0.030	0.009	0.016	0.026	- -	- -

Figure 8. R2 Outcomes of model 3

[4] Wolframalpha software output used to plot equations; just for reference of the u shape because c and error are not considered

PART D

What could be said about the confounding role of both gender and age in the best fitting model?

SEX has a significant effect on Satisfaction (+), Autonomy (-), Demands (+), and Social Support (+) based on LISREL estimates. Social Support is also shown to have a significant impact on satisfaction. It is hence possible, that SEX acts as a confounder and gives rise to spurious covariance among these estimates.

Similarly, Age has a significant effect on Exhaustion (-), Demands (+), Social Support (-), and Autonomy (+). Autonomy(AUT2) is shown to have a significant effect on Exhaustion. It is possible that Age acts as a confounder and gives rise to spurious covariance among these estimates (as an increase in Age leads to an increase in autonomy and decrease in exhaustion).

	SEX	AGE		SEX	AGE
	--------	--------		--------	--------
SAT	0.18	-0.004	CDEM	0.096	0.006
	-0.062	-0.003		-0.045	-0.002
	2.915	-1.651		2.14	2.945
EXH	-0.038	-0.007	CSUP	0.098	-0.003
	-0.058	-0.003		-0.026	-0.001
	-0.664	-2.813		3.824	-2.312
ANX	-0.018	-0.002	AUT2	-	0.009
	-0.033	-0.001			0.001
	-0.535	-1.388			5.814
CAUT	-0.244	0.003	DEM2	-	-
	-0.042	-0.002			
	-5.755	1.758	SUP2	-	-

Table 6: Effect of X on Y (X: Age and SEX)

PART E

What could be mentioned about causality? To put differently, are all causality criteria met?

For causality, it is important to look for 4 types of evidence (Hair, et.al, 2014):

1. Systematic covariation
2. Temporal Sequence
3. Nonspurious Covariance
4. Theoretical Support

If one of these evidences is not fulfilled, then no causality can be found.

1. Systematic covariation

If there are some correlation between paths, there is systematic covariation. According to Table 7, there is systematic covariation since there Is some correlation among the example paths.

	Significant relation	CAUT--> SAT	CDEM--> SAT	CSUP--> SAT
Significant relation	Path	BE (1,4)	BE (1,5)	BE (1,6)
CAUT--> SAT	BE (1,4)	1.000	0.238	-0.135
CDEM--> SAT	BE (1,5)		1.000	0.130
CSUP--> SAT	BE (1,6)			1.000

Table 7. Correlation among some example Beta paths

2. Temporal Sequence

Looking to all significant paths of the model, it can be established that there is temporal sequence because the independent variable logically must happen first than the depend variables; for instance, it is important that autonomy, social support or job demands occur first so that exhaustion or satisfaction are affected or reflected. Thus, there is temporal sequence.

3. Nonspurious Covariance

In this model, there is spurious covariance because there are some variables that can affect some variables separately and not necessarily because they are strictly related; for example, in this model, there is a significant relation between autonomy and satisfaction, however, autonomy and satisfaction can be affected independently by leadership style. Leadership style can lead to have more autonomy because of empowerment leadership style (Wang, Demerouti, & Bakker, 2017) and independently satisfaction can be achieved when a leadership style shows intelligent and model-to-follow leader (Wang, Demerouti, & Bakker, 2017).

Since there is spurious covariance existing, one of the evidences for causality (out of the possible 4 shown below) is not met, hence, there is no causality.

	Fulfill Causality?
1. Systematic covariation	YES
2. Temporal Sequence	YES
3. Nonspurious Covariance	NO
4. Theoretical Support	YES

Table 8. 4 Evidences of Causality

PART F

What is your final conclusion with regard to the models tested? In other words, what about empirical support for the Vitamin Model? Please be concise (1-3 sentences at maximum)!

The non-linear model has a better fit than the linear model. Differing from our expectation, job autonomy and exhaustion followed an inverted U-shaped relationship while the other curvilinear relationships are as per our logical inference: U-shaped relationship between job demands and anxiety; U-shaped patterns between workplace social support and job satisfaction; inverted U-shaped pattern between Social support and job satisfaction. The LISREL output data supports that the Model 3 (with non-linear relationships between job characteristics and employee well-being) closely follows Warr's Vitamin Model: job characteristics act like psychological "vitamins" for health and well-being.

APPENDIX 1 (LISREL CODES)

Model 2 LISREL Code

```
SEM TEST VITAMIN MODEL
MODEL 2 (NON-LINEAR)
DA NI=12 NO=1332
CM FI=DATA2017.COV
LA
SAT CHA EXH ANX CAUT CDEM CSUP AUT2 DEM2 SUP2 SEX AGE
SE
1 3 4 5 6 7 8 9 10 11 12/
MO NY=9 NX=2 BE=FU,FI GA=FU,FI PH=SY,FR PS=SY,FI
FR BE(1,4) BE(2,4) BE(3,4)
FR BE(1,5) BE(2,5) BE(3,5)
FR BE(1,6) BE(2,6) BE(3,6)

FR BE(1,7) BE(2,7) BE(3,7)
FR BE(1,8) BE(2,8) BE(3,8)
FR BE(1,9) BE(2,9) BE(3,9)

FR GA(1,1) GA(2,1) GA(3,1) GA(4,1) GA(5,1) GA(6,1)
FR GA(1,2) GA(2,2) GA(3,2) GA(4,2) GA(5,2) GA(6,2)
FR PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(5,5) PS(6,6)
FR PS(7,7) PS(8,8) PS(9,9)
FR PS(4,5) PS(4,6) PS(4,7) PS(4,8) PS(4,9)
FR PS(5,6) PS(5,7) PS(5,8) PS(5,9)
FR PS(6,7) PS(6,8) PS(6,9)
FR PS(7,8) PS(7,9) PS(8,9)
FR PS(1,2) PS(1,3) PS(2,3)
OU ALL
```

Model 3 LISREL Code

```
SEM TEST VITAMIN MODEL
MODEL 3 (NON-LINEAR RELAXING AGE AND AUT2 PATH)
DA NI=12 NO=1332
CM FI=DATA2017.COV
LA
SAT CHA EXH ANX CAUT CDEM CSUP AUT2 DEM2 SUP2 SEX AGE
SE
1 3 4 5 6 7 8 9 10 11 12/
MO NY=9 NX=2 BE=FU,FI GA=FU,FI PH=SY,FR PS=SY,FI
FR BE(1,4) BE(2,4) BE(3,4)
FR BE(1,5) BE(2,5) BE(3,5)
FR BE(1,6) BE(2,6) BE(3,6)

FR BE(1,7) BE(2,7) BE(3,7)
FR BE(1,8) BE(2,8) BE(3,8)
FR BE(1,9) BE(2,9) BE(3,9)

FR GA(7,2)

FR GA(1,1) GA(2,1) GA(3,1) GA(4,1) GA(5,1) GA(6,1)
FR GA(1,2) GA(2,2) GA(3,2) GA(4,2) GA(5,2) GA(6,2)
FR PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(5,5) PS(6,6)
FR PS(7,7) PS(8,8) PS(9,9)
FR PS(4,5) PS(4,6) PS(4,7) PS(4,8) PS(4,9)
FR PS(5,6) PS(5,7) PS(5,8) PS(5,9)
FR PS(6,7) PS(6,8) PS(6,9)
FR PS(7,8) PS(7,9) PS(8,9)
FR PS(1,2) PS(1,3) PS(2,3)
OU ALL
```

References

Garson, D. (2007). *Structural Equation Modeling.* ResearchGate.

Hair, et.al. (2014). *Multivariate Data Analysis.* United States: Pearson.

Landy, F., & Conte, J. (2013). *Work in the 21st Century.* Philadelphia: Wiley.