FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS
MASTER OF STATISTICS AND DATA SCIENCE
STRUCTURAL EQUATION MODELING

KU LEUVEN

Assignment

Van den Broeck Sebastiaan (r0902562)

13/03/2023

1 Introduction

Harm avoidance and self-directedness have been linked to depression. A behaviour can be classified under harm avoidance if it is done to avoid novelty and punishment. Self-directedness, on the other hand, is a form of self-determination and ability to regulate behaviour to suit goals and values. It has been proposed that harm avoidance and self-directedness are indirectly linked to depression through social functioning (Tse et al., 2011). In this work I will test this hypothesis on a new dataset, which will be discussed next. A structural equation model has been used to test the hypothesis and will be discussed next. Lastly, the results and implications thereof will be considered.

2 Data

The data treated in the report is the Midlife in the United States (MIDUS) series. Currently, there are three waves in the study, which were collected via phone interviews, surveys and by bringing participants into clinical settings to facility collecting biological data. All three waves cover the contiguous United States in its entirety. The first wave was collected in 1995 and 1996, while the second wave was collected in 2004 and 2005. The most recent wave was collected in 2013 and 2014. The second and third wave have been combined to create a bigger dataset. It was not possible to incorporate the first dataset, since a lot of variables changed between the first and second and third waves (Radler, 2014).

An important reason for choosing this dataset is that it contains a lot of documentation for which variables form certain latent constructs such as depression or social anxiety. Since I am not familiar with the field of psychology this would save me a lot of time. Depression is the most important latent variable in this work. It has been measures through seven questions during which the respondent reflects over the last two weeks. For example, the questions include losing interest, becoming tired, having trouble falling asleep or thinking about death. Each variable which measures this latent construct has been coded such that a 1 reflects a yes answer. As could be expected, a 0 then means a respondent has answered no.

Construct	**Code**	**Question**
Depression	PA63	During those two weeks, did you lose interest in most things?
	P164	Thinking about these same two weeks, did you feel more tired out or low on energy?
	PA65	During those same two weeks, did you lose appetite?
	PA66	Did you have more trouble falling asleep than you usually do during those two weeks?
	PA67	During that same two week period, did you have a lot more trouble concentrating than usual?
	PA68	People sometimes feel down on themselves, no good, or worthless. During that two-week period, did you feel this way?
	PA69	Did you think a lot about death - either your own, someone else's or death in general - during those two weeks?

Construct	**Code**	**Count**	
		0	1
Depression	PA63	156	633
	P164	61	726
	PA65	338	445
	PA66	223	565
	PA67	111	675
	PA68	283	507
	PA69	304	485

Another important aspect in this report is harm avoidance. Since it cannot be measured directly, four questions were asked to get an idea about this variable. First, interviewees were asked whether they would enjoy experiencing an earthquake or learning to walk the tightrope. These two variables were reverse recoded such that a 4 reflects not agreeing with the statement at all (harm avoidance), while a 1 indicates fully agreeing (no avoidance). Second, interviewees were presented with two scenario's

twice. For each question, one scenario corresponds to a harmful situation, while the other scenario's is harmless. Again, there was a recoding such that a higher score on these two variables indicates avoiding harm.

Construct	Code	Question
Harm avoidance	SE7D	It might be fun and exciting to be in an earthquake.
	SE7V	It might be fun learning to walk a tightrope.
	SE8	Of these two situations, I would dislike more: Situation 1: Riding a long stretch of rapids in a canoe; Situation 2: Waiting for someone who's late.
	SE9	Of these two situations, I would dislike more: Situation 1: Being at the circus when two lions suddenly get loose down in the ring; Situation 2: Bringing my whole family to the circus and then not being able to get in because a clerk sold me tickets for the wrong night.

Construct	Code	Count			
		1 (harm)	2	3	4 (no harm)
Harm avoidance	SE7D	274	875	838	4889
	SE7V	367	1222	1163	4238

Construct	Code	Count	
		0 (harm)	1 (no harm)
Harm avoidance	SE8	3803	3089
	SE7V	2994	3898

We should not forget about self-directedness, which has been measured through three variables. Making plans for the future, knowing what to want out of life and setting goals are important for this dimension. Again, the variables were reverse coded such that a higher score reflects agreeing more with the statement. The data indicates that most participants agree somewhat or fully what the three statements.

Construct	Code	Question
Self-directedness	SE14O	I like to make plans for the future.
	SE14R	I know what I want out of life.
	SE14P	I find it helpful to set goals for the near future.

Construct	Code	Count			
		1	2	3	4
Self-directedness	SE14O	247	1303	2754	2590
	SE14R	251	1089	2929	2604
	SE14P	318	1320	3071	2184

Lastly, the latent variable social functioning has been used in the analysis. Seven questions related to this dimension were asked. The variables SE1BB, SE1D, SE1I and SE1V were reverse coded such that a higher score indicates a higher degree of social functioning. The dataset counts 601 observations after deleting rows with missing values.

Construct	Code	Question
Social functioning	SE1BB	People would describe me as a giving person, willing to share my time with others.
	SE1D	Most people see me as loving and affectionate.
	SE1HH	I have not experienced many warm and trusting relationships with others.
	SE1J	Maintaining close relationships has been difficult and frustrating for me.
	SE1I	I think it is important to have new experiences that challenge how you think about yourself and the world.
	SE1P	I often feel lonely because I have few close friends with whom to share my concerns.
	SE1V	I enjoy personal and mutual conversations with family members and friends.

Construct	Code	Count						
		1	2	3	4	5	6	7
Social functioning	SE1BB	28	76	127	472	799	2382	3055
	SE1D	43	126	219	807	815	2598	230
	SE1HH	291	501	575	441	497	1382	3242
	SE1J	236	575	818	711	512	1524	2550
	SE1I	113	136	142	709	1154	2190	2483
	SE1P	242	447	744	655	479	1321	3036
	SE1V	63	64	105	211	589	1879	4024

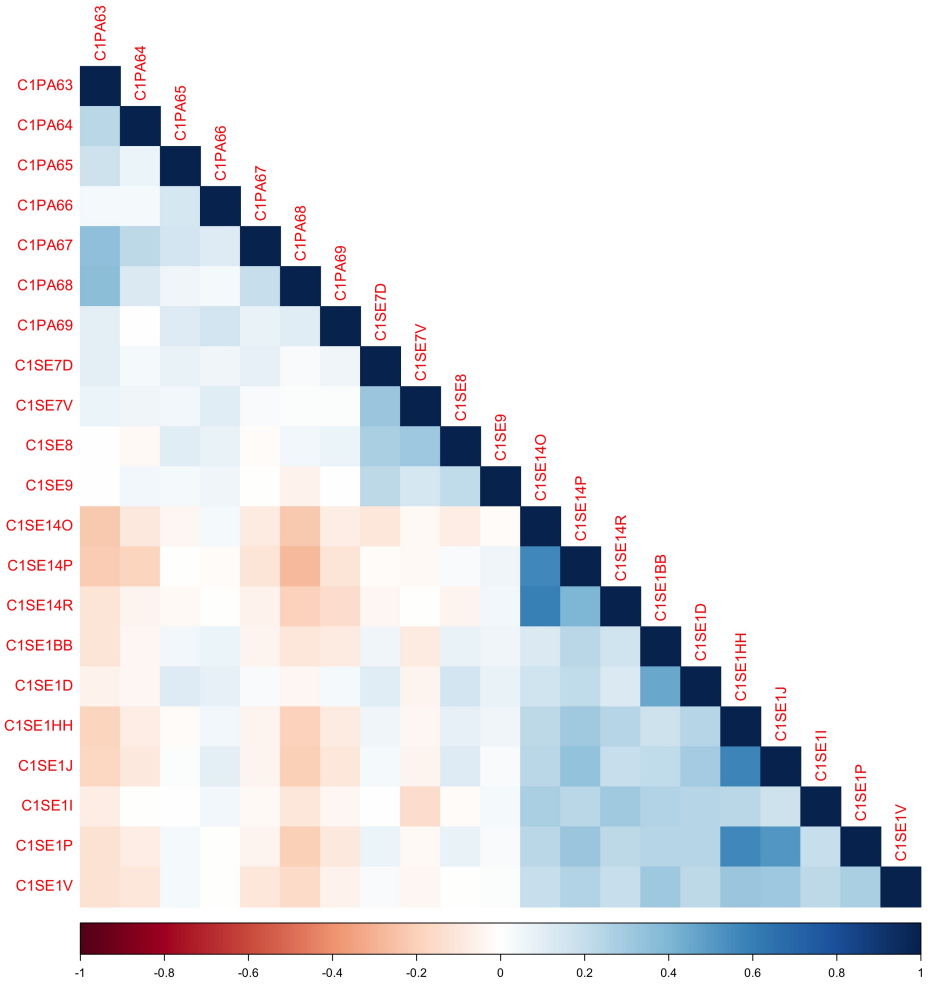

Figure 1: Correlation plot

3 The base model

The base model has four latent variables which are related to each other through a structural model. Social functioning has been expressed linearly through harm avoidance and self directedness. Depression is estimated to have a linear relationship with social functioning. Hence, there is no direct effect of the exogenous variables harm avoidance and self directedness on depression.

$$
\begin{cases}
\text{PA63} & = \lambda_{11}\text{depression} + \delta_{11} \\
\text{PA64} & = \lambda_{12}\text{depression} + \delta_{12} \\
\text{PA65} & = \lambda_{13}\text{depression} + \delta_{13} \\
\text{PA66} & = \lambda_{14}\text{depression} + \delta_{14} \\
\text{PA67} & = \lambda_{15}\text{depression} + \delta_{15} \\
\text{PA68} & = \lambda_{16}\text{depression} + \delta_{16} \\
\text{PA69} & = \lambda_{17}\text{depression} + \delta_{17} \\
\text{SE7V} & = \lambda_{21}\text{harm avoidance} + \delta_{21} \\
\text{SE7D} & = \lambda_{22}\text{harm avoidance} + \delta_{22} \\
\text{SE8} & = \lambda_{23}\text{harm avoidance} + \delta_{23} \\
\text{SE9} & = \lambda_{24}\text{harm avoidance} + \delta_{24} \\
\text{SE14O} & = \lambda_{31}\text{self directedness} + \delta_{31} \\
\text{SE14P} & = \lambda_{32}\text{self directedness} + \delta_{32} \\
\text{SE14R} & = \lambda_{33}\text{self directedness} + \delta_{33} \\
\text{SE1BB} & = \lambda_{41}\text{social functioning} + \delta_{41} \\
\text{SE1D} & = \lambda_{42}\text{social functioning} + \delta_{42} \\
\text{SE1HH} & = \lambda_{43}\text{social functioning} + \delta_{43} \\
\text{SE1J} & = \lambda_{44}\text{social functioning} + \delta_{44} \\
\text{SE1I} & = \lambda_{45}\text{social functioning} + \delta_{45} \\
\text{SE1P} & = \lambda_{46}\text{social functioning} + \delta_{46} \\
\text{SE1V} & = \lambda_{47}\text{social functioning} + \delta_{47} \\
\text{social functioning} & = \beta_1\text{harm avoidance} + \beta_2\text{self directedness} + \delta_1 \\
\text{depression} & = \beta_3\text{social functioning} + \delta_2
\end{cases}
\tag{1}
$$

The measurement errors δ are supposed to have an expected value of 0. It is assumed that they have constant variance across observations and are mutually uncorrelated. There should be a covariance of zero between these errors and the latent variables. The first loading of each latent variable has been fixed to one in order to identify the model.

There is a deviation from the multivariate normality assumption on the residuals δ, since all variables are ordinal in nature. A solution can then be obtained using thresholds, which represent the values of a normal distribution that reproduce the proportions of the indicators. Maximum likelihood is traditionally used in SEM to find estimates of the unknown parameters. However, ordinary ML should not be used when at least one factor indicator is categorical, since this approach can lead to incorrect estimates (Brown, 2015). In this case, robust maximum likelihood or diagonally weighted least squares can be used. The latter approach has been specifically developed for ordinal data and has been shown to yield better results when the sample size is not small (Li, 2016). It has therefore been applied here.

First and foremost, we will take a closer look at the measurement model, which indicates how the variables relate to their latent constructs. By squaring the standardized loading one obtains the communality, which indicates the proportion of the variance in the indicator that is explained by the latent variable. The residual variance then indicates the proportion of the variance that is not explained by the latent factor. Although there are no hard rules, a popular cut-off value for the communality appears to be 0.5 (**hair2010**). The standardized loading should then be larger than 0.7, which means that the indicator does a good job at reflecting the latent construct.

Inspecting Table 1, it is evident to see that some indicators have a low communality. The variables PA63, PA64 and PA69 have a high residual variance and load on the latent variable depression. The problematic indicators PA65, PA66 and PA69 assess a loss of appetite, trouble falling asleep and thinking about death. A simple way to improve the model fit may be to reduce the number of variables that load on depression. However, this action would lead to a decline of the theoretical support and validity of the model as well (**hair2010**).

Next, regarding harm avoidance the variables SE7V and SE9 have a communality of 0.379 and 0.201. On the one hand, SE7V evaluates whether the participant believes it could be fun to experience an earthquake. On the other hand, through SE9 the respondent has to choose between a harmful and

Table 1: Measurement model

Variable	Loading	Standard error	z-value	p-value	Stand. loading	Communality	Residual var.
PA63	1				0.851	0.724	0.275
PA64	0.637	0.109	5.844	<0.001	0.542	0.294	0.706
PA65	0.177	0.083	2.138	0.033	0.151	0.023	0.977
PA66	0.050	0.086	0.588	0.556	0.043	0.002	0.998
PA67	0.655	0.098	6.712	<0.001	0.557	0.310	0.689
PA68	0.894	0.110	8.090	<0.001	0.761	0.579	0.421
PA69	0.344	0.088	3.894	<0.001	0.293	0.086	0.914
SE7V	1				0.616	0.379	0.620
SE7D	1.140	0.147	7.744	<0.001	0.703	0.494	0.506
SE8	1.173	0.153	7.650	<0.001	0.723	0.523	0.477
SE9	0.727	0.113	6.405	<0.001	0.448	0.201	0.799
SE14O	1				0.819	0.671	0.330
SE14P	0.981	0.049	20.185	<0.001	0.803	0.645	0.355
SE14R	0.872	0.047	18.615	<0.001	0.714	0.510	0.490
SE1BB	1				0.556	0.309	0.691
SE1D	0.947	0.072	13.178	<0.001	0.526	0.277	0.723
SE1HH	1.391	0.099	14.080	<0.001	0.773	0.598	0.403
SE1J	1.337	0.093	14.310	<0.001	0.743	0.552	0.448
SE1I	0.848	0.080	10.592	<0.001	0.471	0.222	0.778
SE1P	1.310	0.087	15.104	<0.001	0.728	0.530	0.470
SE1V	1.126	0.093	12.170	<0.001	0.626	0.392	0.609

a safe situation. A simple way to get a more parsimonious model would be to delete these variables. However, the aspiration to achieve a good model fit should never compromise the theory being tested (**hair2010**). Since these variables have been specifically designed by the authors of the dataset I will not be deleting them from the model.

The indicators that load on self-directedness don't have a problem regarding a low communality.

Lastly, the variables SE1BB, SE1D, SE1I and SE1V also have a low communality.

Second, more attention will be paid to the structural model, which has been summarized in Table 2. The results indicate a significant effect ($p < 0.001$) of self directedness on social functioning with a standardized magnitude of 0.610. Hence, it is estimated by the model that an increase of one standard deviation in self-directedness will lead to an increase of 0.610 standardized units in social functioning. However, we are not able to make a conclusion regarding the effect of harm avoidance on social functioning since its p-value is not low enough. In other words, it is plausible that this coefficient is zero on the population level, although there a small effect was detected in the sample. Next, social functioning appears to have a strong and significant ($p < 0.001$) negative effect on depression. Specifically, the standardized coefficient indicates a decrease of 0.454 in depression for every increase of one standard deviation in social functioning.

Strongly related to the structural model is the notion of discriminant validity. Discriminant validity gives an indication of the notion that theoretically different constructs should not be highly intercorrelated. In other words, if two latent variables are highly correlated they could represent the same construct and they could be merged into one latent variable to obtain a more parsimonious solution (Brown, 2015). The low and insignificant ($p=0.258$) covariance of -0.035 between harm avoidance and self-directedness indicates that there is no problem with discriminant validity here.

Table 2: Structural model

		Coefficient	Standard error	z-value	p-value	Stand. coefficient
social functioning	harm avoidance	0.090	0.051	1.760	0.078	0.100
social functioning	self directedness	0.414	0.038	11.050	<0.001	0.610
depression	social functioning	-0.696	0.098	-7.131	<0.001	-0.454

Table 3: Modification indices of base model

Left hand side	Operation	Right hand side	Modification index	Expected parameter change	Stand. expected parameter change
1SE1BB	covariance	1SE1D	94.016	0.329	0.466
depression	regression	self_directedness	66.730	-0.467	-0.449
depression	covariance	self_directedness	61.695	-0.306	-0.492
self_directedness	regression	depression	61.694	-0.532	-0.553
depression	loading	1SE14P	55.819	-0.422	-0.359
social_functioning	regression	depression	53.335	0.351	0.538
depression	covariance	social_functioning	53.333	0.202	0.606
self_directedness	loading	1PA68	45.156	-0.418	-0.343
self_directedness	loading	1SE1I	44.925	0.413	0.338
social_functioning	loading	1SE14P	35.233	0.647	0.359
1SE14O	covariance	1SE14R	32.091	0.266	0.661
1SE1HH	covariance	1SE1J	31.628	0.202	0.475
1SE1BB	covariance	1SE1HH	29.529	-0.264	-0.501
social_functioning	loading	1PA65	29.419	0.454	0.252
depression	regression	harm_avoidance	26.757	0.294	0.213
1SE1HH	covariance	1SE1P	26.553	0.187	0.429
depression	loading	1SE1D	25.581	0.334	0.284
social_functioning	loading	1SE7V	25.571	-0.281	-0.156
social_functioning	loading	1SE14O	23.932	-0.586	-0.326
harm_avoidance	regression	depression	23.132	0.182	0.251

Third, the goodness of fit of the model will be evaluated.

Table 4: Test statistics

Statistic	Value	Target
χ^2	658	217.74
CFI	0.938	0.9
TLI	0.930	0.9
RMSEA	0.065	0.05
SRMR	0.087	0.08

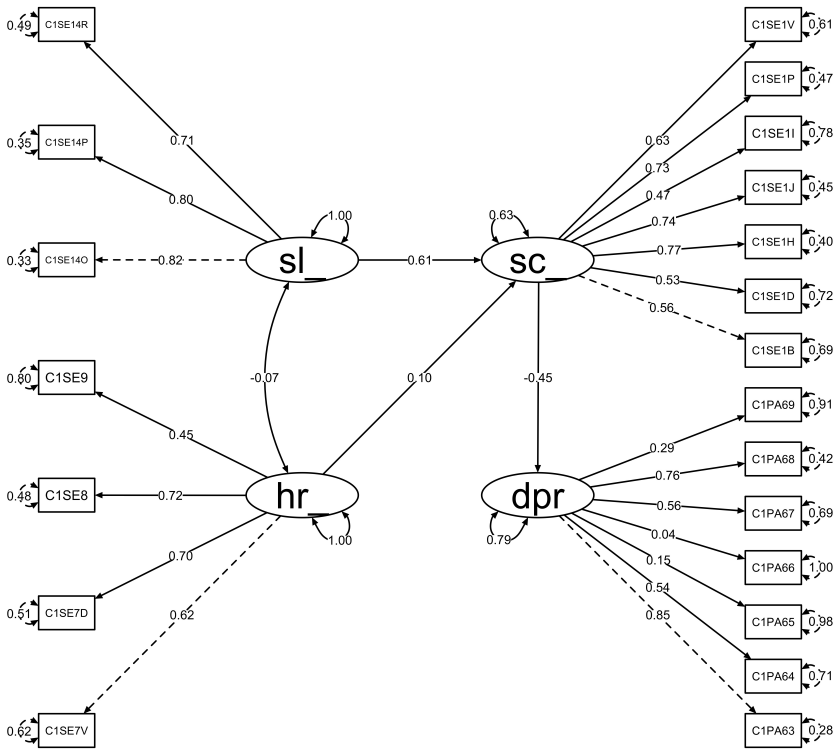

Figure 2: Summary of base model

4 Expanding the base model

5 Conclusion