Confirmatory Factor Analysis

Statsomat.com

23 December 2020

Warning: The automatic computation and interpretation delivered by the Statsomat should not completely replace the classical, made by humans graphical exploratory data analysis and statistical analysis. There may be data cases for which the Statsomat does not deliver the most optimal solution or output interpretation.

Basic Information

Automatic statistics for the file:

File
case2.csv

Your selection for the encoding: Auto
Your selection for the decimal character: Auto
Observations (rows with at least one non-missing value): 500
Variables (columns with at least one non-missing value): 9
Variables considered continuous: 7

Variables considered continuous
SUBJID
ACTIV
SOMA
SOCF
VITAL
GENHLTH
AGE

Numerical variables considered binary or ordinal: 2

Numerical variables considered binary or ordinal
PAIN
MENTH

Character variables considered binary: 0
Character variables considered nominal and transformed to binary: 0

Model Syntax

The following table describes the applied model equations in lavaan model syntax, either as entered by you in the text area (denoted by User=1) or established internally (User=0). The last column numbers the free parameters which are estimated.

Left hand side	Operator	Right hand side	User	Free parameter
PHYSF	=~	ACTIV	1	0
PHYSF	=~	SOMA	1	1
PHYSF	=~	PAIN	1	2
MENTF	=~	MENTH	1	0
MENTF	=~	SOCF	1	3
MENTF	=~	VITAL	1	4
GWB	=~	GENHLTH	1	0
AGEF	=~	AGE	1	0
GENHLTH	~~	GENHLTH	1	0
AGE	~~	AGE	1	0
ACTIV	~~	SOMA	1	5
ACTIV	~~	ACTIV	0	6
SOMA	~~	SOMA	0	7
PAIN	~~	PAIN	0	8
MENTH	~~	MENTH	0	9
SOCF	~~	SOCF	0	10
VITAL	~~	VITAL	0	11
PHYSF	~~	PHYSF	0	12
MENTF	~~	MENTF	0	13
GWB	~~	GWB	0	14
AGEF	~~	AGEF	0	15
PHYSF	~~	MENTF	0	16
PHYSF	~~	GWB	0	17
PHYSF	~~	AGEF	0	18
MENTF	~~	GWB	0	19
MENTF	~~	AGEF	0	20
GWB	~~	AGEF	0	21
ACTIV	~1		0	22
SOMA	~1		0	23
PAIN	~1		0	24
MENTH	~1		0	25
SOCF	~1		0	26
VITAL	~1		0	27
GENHLTH	~1		0	28
AGE	~1		0	29
PHYSF	~1		0	0
MENTF	~1		0	0
GWB	~1		0	0
AGEF	~1		0	0

Assumptions

Open issue

Model Settings

Outputs

Model Fit Summary → in appendix

```
lavaan 0.6-7 ended normally after 234 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of free parameters                         29

  Number of observations                           500
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                                45.009
  Degrees of freedom                                15
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              1942.397
  Degrees of freedom                                28
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.971

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -13787.363
  Loglikelihood unrestricted model (H1)     -13764.859

  Akaike (AIC)                               27632.726
  Bayesian (BIC)                             27754.950
  Sample-size adjusted Bayesian (BIC)        27662.902
```

Root Mean Square Error of Approximation:

RMSEA	0.063
90 Percent confidence interval - lower	0.043
90 Percent confidence interval - upper	0.085
P-value RMSEA <= 0.05	0.137

Standardized Root Mean Square Residual:

SRMR	0.026

Parameter Estimates:

Standard errors	Standard
Information	Observed
Observed information based on	Hessian

Latent Variables:

	Estimate	Std.Err	z-value	P(>|z|)
PHYSF =~				
ACTIV	1.000			
SOMA	0.868	0.050	17.527	0.000
PAIN	0.150	0.018	8.480	0.000
MENTF =~				
MENTH	1.000			
SOCF	1.850	0.074	25.077	0.000
VITAL	2.357	0.109	21.659	0.000
GWB =~				
GENHLTH	1.000			
AGEF =~				
AGE	1.000			

Covariances:

	Estimate	Std.Err	z-value	P(>|z|)
.ACTIV ~~				
.SOMA	87.987	17.805	4.942	0.000
PHYSF ~~				
MENTF	28.370	4.307	6.586	0.000
GWB	69.057	7.958	8.677	0.000
AGEF	-30.305	9.003	-3.366	0.001
MENTF ~~				
GWB	25.179	2.280	11.044	0.000
AGEF	-5.290	2.627	-2.014	0.044
GWB ~~				
AGEF	-9.664	4.502	-2.147	0.032

Intercepts:

	Estimate	Std.Err	z-value	P(>|z|)

```
   .ACTIV            34.268    0.892   38.426    0.000
   .SOMA             21.936    0.708   30.975    0.000
   .PAIN              8.212    0.146   56.202    0.000
   .MENTH             7.060    0.248   28.497    0.000
   .SOCF             14.750    0.418   35.249    0.000
   .VITAL            16.016    0.618   25.906    0.000
   .GENHLTH          28.322    0.378   74.915    0.000
   .AGE              33.210    0.530   62.651    0.000
    PHYSF             0.000
    MENTF             0.000
    GWB               0.000
    AGEF              0.000

Variances:
                  Estimate  Std.Err  z-value  P(>|z|)
   .GENHLTH          7.861
   .AGE              0.000
   .ACTIV          215.497   24.583    8.766    0.000
   .SOMA           113.524   15.800    7.185    0.000
   .PAIN             6.567    0.583   11.260    0.000
   .MENTH            8.548    0.782   10.935    0.000
   .SOCF            11.779    1.949    6.045    0.000
   .VITAL           68.148    5.272   12.926    0.000
    PHYSF          182.151   29.416    6.192    0.000
    MENTF           22.140    1.948   11.367    0.000
    GWB             63.601    4.520   14.072    0.000
    AGEF           140.490    8.885   15.811    0.000
```

Completely Standardized Parameter Estimates → in appendix

```
Latent Variables:
                   est.std  Std.Err  z-value  P(>|z|) ci.lower ci.upper
  PHYSF =~
    ACTIV            0.677    0.045   15.014    0.000    0.588    0.765
    SOMA             0.740    0.043   17.204    0.000    0.655    0.824
    PAIN             0.620    0.041   14.990    0.000    0.539    0.701
  MENTF =~
    MENTH            0.849    0.017   51.456    0.000    0.817    0.882
    SOCF             0.930    0.013   73.920    0.000    0.906    0.955
    VITAL            0.802    0.019   42.728    0.000    0.765    0.839
  GWB =~
    GENHLTH             NA       NA                        NA       NA
  AGEF =~
    AGE              1.000                              1.000    1.000

Covariances:
                   est.std  Std.Err  z-value  P(>|z|) ci.lower ci.upper
```

.ACTIV ~~						
.SOMA	0.563	0.054	10.461	0.000	0.457	0.668
PHYSF ~~						
MENTF	0.447	0.051	8.734	0.000	0.346	0.547
GWB	0.231	0.020	11.559	0.000	0.192	0.270
AGEF	-0.189	0.054	-3.509	0.000	-0.295	-0.084
MENTF ~~						
GWB	0.241	0.017	14.482	0.000	0.209	0.274
AGEF	-0.095	0.046	-2.044	0.041	-0.186	-0.004
GWB ~~						
AGEF	-0.037	0.017	-2.159	0.031	-0.070	-0.003

Intercepts:

	est.std	Std.Err	z-value	P(>|z|)	ci.lower	ci.upper
.ACTIV	-0.000	0.045	-0.000	1.000	-0.088	0.088
.SOMA	-0.000	0.045	-0.000	1.000	-0.088	0.088
.PAIN	0.000	0.045	0.000	1.000	-0.088	0.088
.MENTH	0.000	0.045	0.000	1.000	-0.088	0.088
.SOCF	-0.000	0.045	-0.000	1.000	-0.088	0.088
.VITAL	0.000	0.045	0.000	1.000	-0.088	0.088
.GENHLTH	0.000	0.045	0.000	1.000	-0.088	0.088
.AGE	-0.000	0.045	-0.000	1.000	-0.088	0.088
PHYSF	0.000				0.000	0.000
MENTF	0.000				0.000	0.000
GWB	NA	NA			NA	NA
AGEF	0.000				0.000	0.000

Variances:

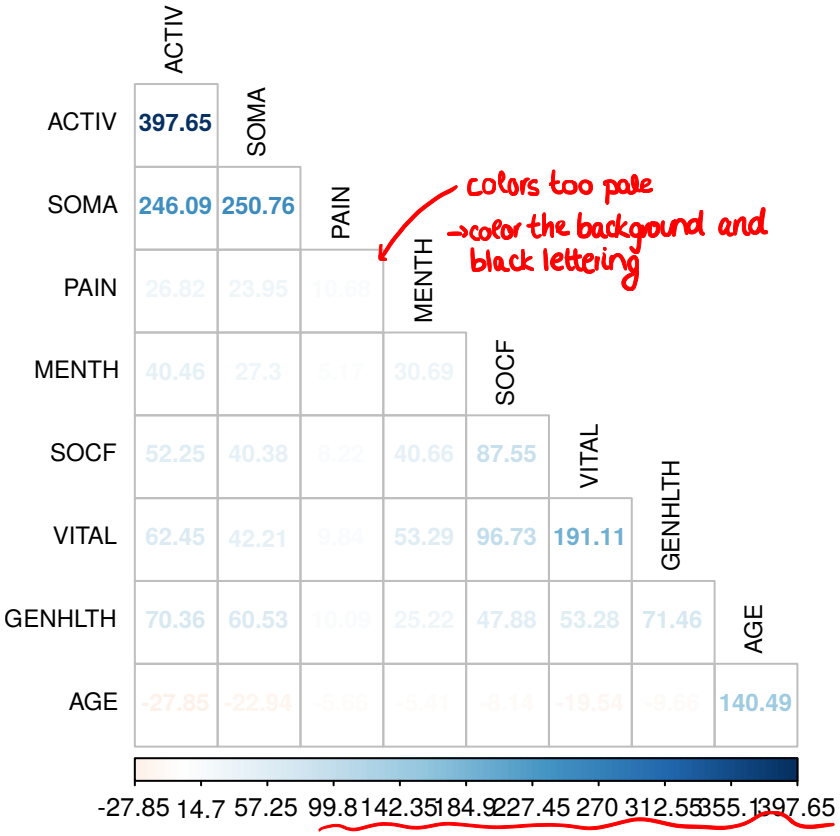
	est.std	Std.Err	z-value	P(>|z|)	ci.lower	ci.upper
.GENHLTH	7.877	0.498	15.811	0.000	6.900	8.853
.AGE	0.000				0.000	0.000
.ACTIV	0.542	0.061	8.881	0.000	0.422	0.662
.SOMA	0.453	0.064	7.116	0.000	0.328	0.577
.PAIN	0.615	0.051	11.981	0.000	0.515	0.716
.MENTH	0.279	0.028	9.933	0.000	0.224	0.334
.SOCF	0.135	0.023	5.746	0.000	0.089	0.180
.VITAL	0.357	0.030	11.840	0.000	0.298	0.416
PHYSF	1.000				1.000	1.000
MENTF	1.000				1.000	1.000
GWB	NA	NA			NA	NA
AGEF	1.000				1.000	1.000

Communality

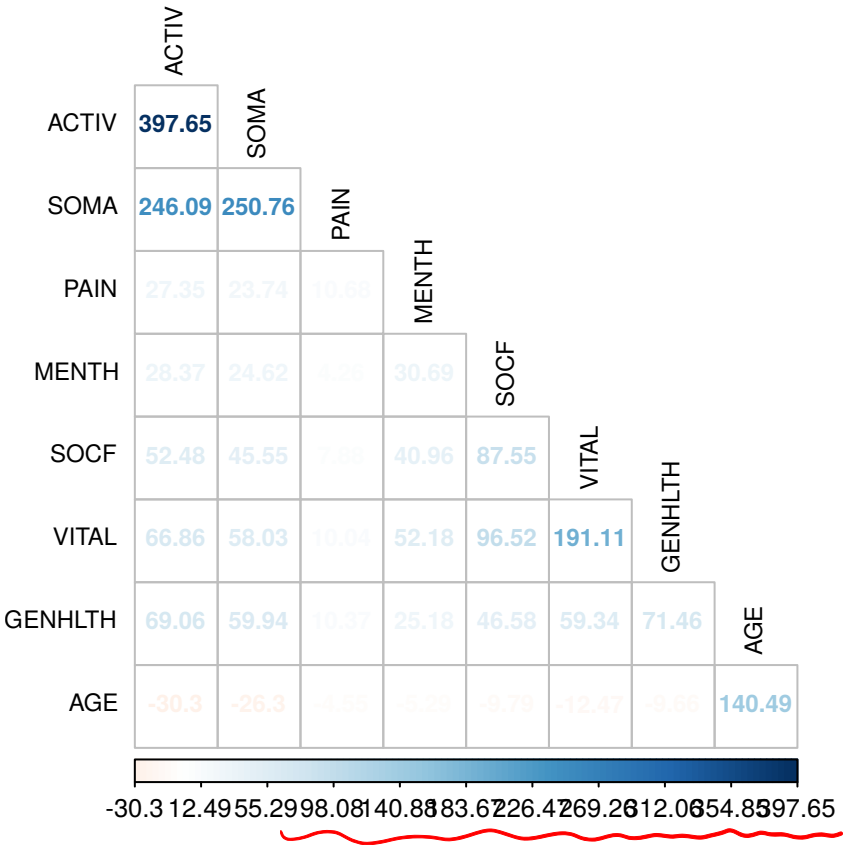
Table 1: Communality

Variable	Communality
GENHLTH	0.89
AGE	1.00
ACTIV	0.46
SOMA	0.55
PAIN	0.38
MENTH	0.72
SOCF	0.87
VITAL	0.64

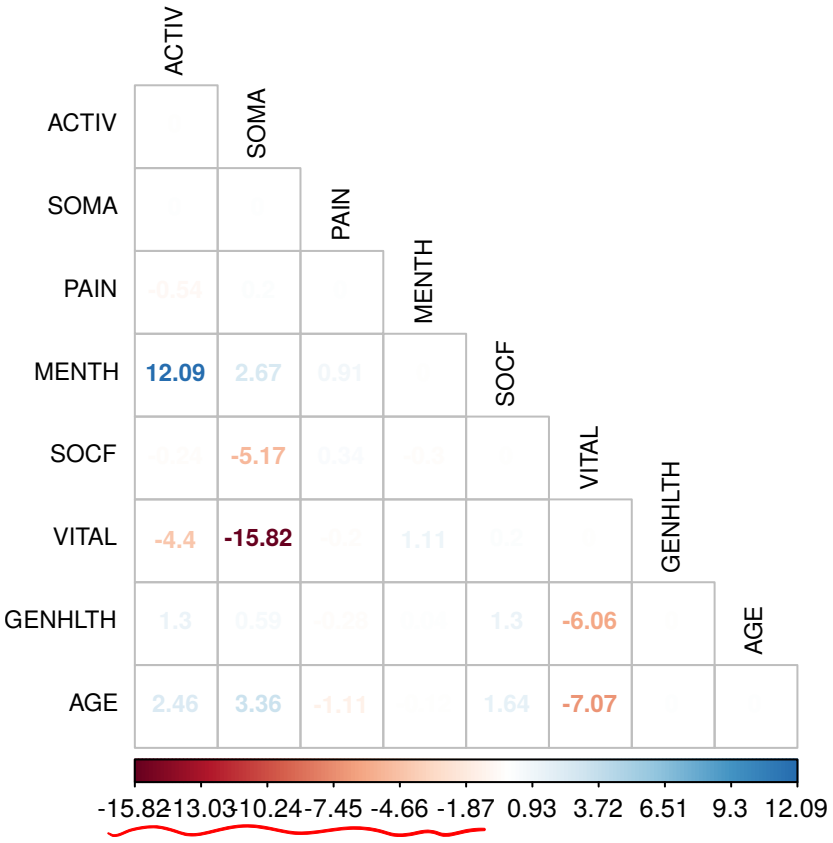
Observed Covariance Matrix



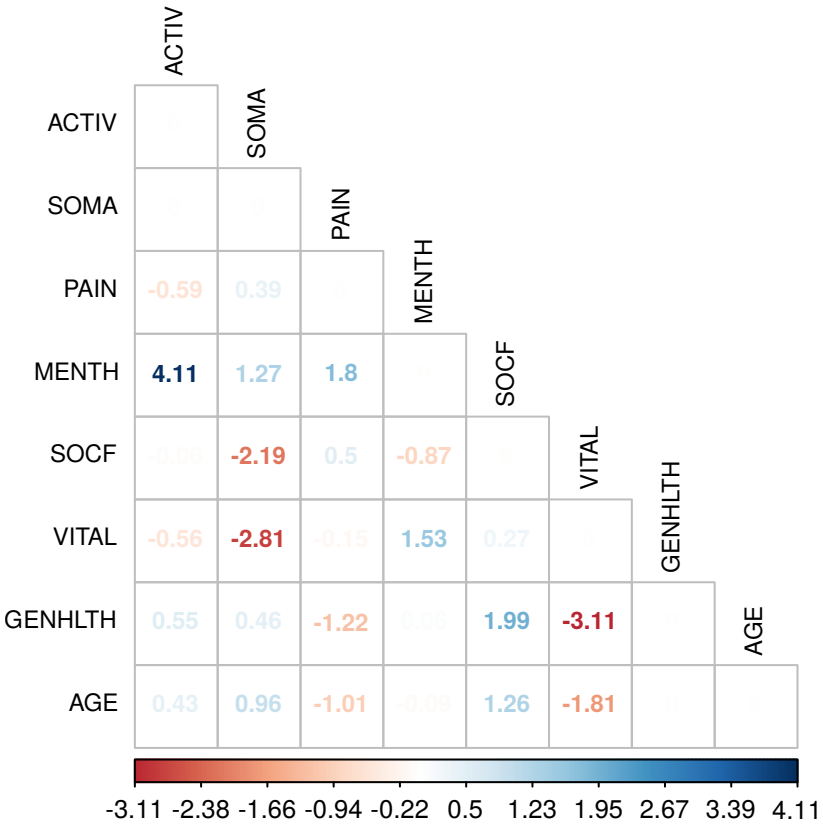
Fitted Covariance Matrix



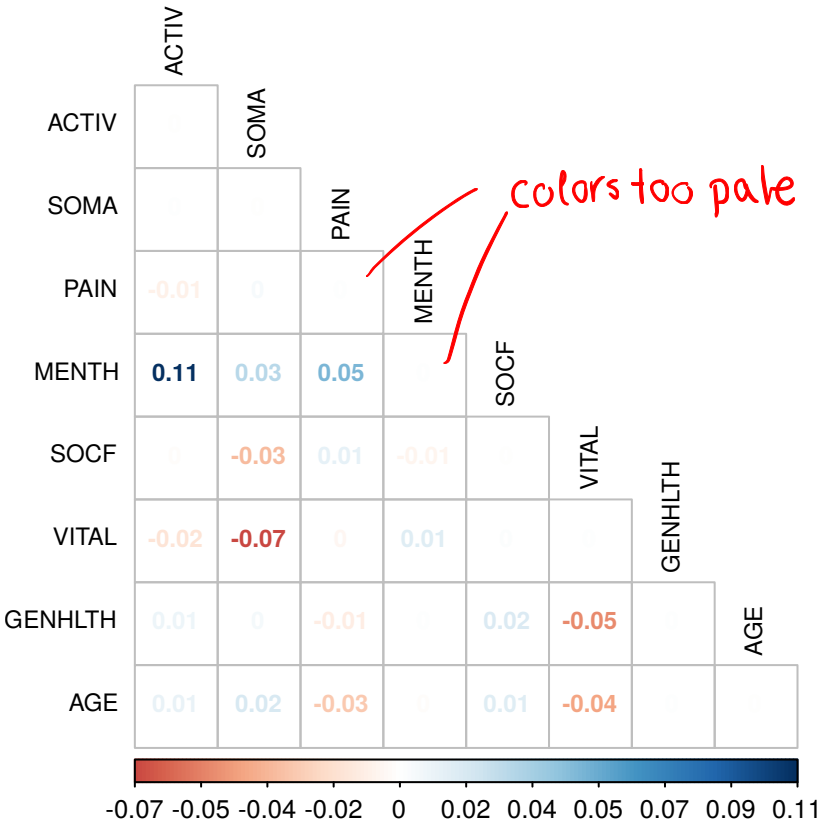
Residual Covariance Matrix



Standardized Residual Matrix



Residual Correlation Matrix



Modification Indices

add list of abbreviations

Table 2: Modification Indices With Respect To (Residual) Correlation

Left	Operator	Right	Modification Index	Expected Parameter Change	Delta	Power	Decision
ACTIV	~~	MENTH	14.761	6.982	0.1	0.050	**(m)**
SOCF	~~	GENHLTH	9.929	5.308	0.1	0.050	**(m)**
MENTH	~~	SOCF	8.687	-5.760	0.1	0.050	**(m)**
VITAL	~~	GENHLTH	5.425	-6.083	0.1	0.050	**(m)**
ACTIV	~~	SOCF	4.522	-5.811	0.1	0.050	**(m)**
VITAL	~~	AGE	4.178	-9.674	0.1	0.050	**(m)**
MENTH	~~	VITAL	4.025	4.230	0.1	0.050	**(m)**
SOMA	~~	VITAL	2.279	-5.726	0.1	0.050	(i)
SOMA	~~	GENHLTH	2.053	4.753	0.1	0.050	(i)
PAIN	~~	GENHLTH	1.880	-1.758	0.1	0.051	(i)
MENTH	~~	GENHLTH	1.845	-1.358	0.1	0.051	(i)
SOCF	~~	AGE	1.566	3.530	0.1	0.050	(i)
SOMA	~~	MENTH	1.137	-1.490	0.1	0.051	(i)
PAIN	~~	AGE	1.010	-1.680	0.1	0.050	(i)
ACTIV	~~	GENHLTH	0.728	-3.400	0.1	0.050	(i)
SOMA	~~	PAIN	0.551	1.330	0.1	0.050	(i)
ACTIV	~~	PAIN	0.551	-1.532	0.1	0.050	(i)
PAIN	~~	MENTH	0.366	0.251	0.1	0.057	(i)
SOMA	~~	AGE	0.297	2.944	0.1	0.050	(i)
SOCF	~~	VITAL	0.279	2.218	0.1	0.050	(i)
ACTIV	~~	VITAL	0.171	2.040	0.1	0.050	(i)
SOMA	~~	SOCF	0.167	0.865	0.1	0.050	(i)
PAIN	~~	VITAL	0.134	0.413	0.1	0.051	(i)
MENTH	~~	AGE	0.119	0.606	0.1	0.050	(i)
PAIN	~~	SOCF	0.008	-0.056	0.1	0.053	(i)
ACTIV	~~	AGE	0.000	0.125	0.1	0.050	(i)

Table 3: Modification Indices With Respect To Factor Loadings

Left	Operator	Right	Modification Index	Expected Parameter Change	Delta	Power	Decision
PHYSF	=~	MENTH	9.314	0.049	0.4	1.000	epc:nm
GWB	=~	VITAL	9.278	-0.248	0.4	0.998	epc:nm
MENTF	=~	SOMA	5.317	-0.301	0.4	0.864	epc:nm
PHYSF	=~	VITAL	5.292	-0.099	0.4	1.000	epc:nm
GWB	=~	SOCF	4.805	0.118	0.4	1.000	epc:nm
MENTF	=~	ACTIV	3.843	0.311	0.4	0.712	**(m)**
AGEF	=~	VITAL	3.166	-0.061	0.4	1.000	(nm)
AGEF	=~	SOCF	2.041	0.029	0.4	1.000	(nm)
GWB	=~	PAIN	1.376	-0.107	0.4	0.992	(nm)
AGEF	=~	PAIN	1.078	-0.013	0.4	1.000	(nm)
PHYSF	=~	SOCF	0.804	-0.024	0.4	1.000	(nm)
MENTF	=~	PAIN	0.793	0.040	0.4	1.000	(nm)
GWB	=~	ACTIV	0.405	0.089	0.4	0.818	(nm)
AGEF	=~	SOMA	0.366	0.024	0.4	1.000	(nm)
GWB	=~	MENTH	0.173	0.013	0.4	1.000	(nm)
GWB	=~	SOMA	0.132	-0.044	0.4	0.916	(nm)
AGEF	=~	MENTH	0.008	-0.001	0.4	1.000	(nm)
AGEF	=~	ACTIV	0.001	-0.001	0.4	1.000	(nm)

Interpretation

Goodness of Fit Indices

We consider some of the model fit indices from the Model Fit Summary section to check the goodness-of-fit of the model. To decide for an acceptable or non-acceptable model, we apply thresholds considered in the References: [@brown], [@kline].

Model Test User Model

The degrees of freedom are calculated as the number of known parameters minus the number of free parameters: 44 - 29 = 15 . The 15 degrees of freedom indicate an over-identified model, fact which basically enables further analysis and interpretation.

The test statistic with the value 45.009 is called the ~~Chi-square~~ χ^2- model fit index and represents the difference between summaries of the model-implied covariance matrix and the observed covariance matrix which is hypothesized and desirable to be zero. In general, if the p-value is larger than 0.01 then the test is not statistically significant at 1 % error, the hypothesis cannot be rejected, which would be in favour of the model.

what is meant with „summaries"?

In our case, the p-value is <0.001 suggesting that the model may not be acceptable for the data. The Chi-square model fit index is based on a very stringent statistical hypothesis which may have no practical relevance. We will consider it only in connection with other model fit indices.

Model Test Baseline Model

The test statistic with the value 1942.397 represents the difference between summaries of the baseline model (an alternative model-implied covariance matrix having zero covariances, i.e. a worst fitting model assuming independent variables) and the observed covariance matrix. The p-value of the test of a zero difference is <0.001 suggesting that

the baseline model does not fit good to the data. This result is used indirectly in the construction of other model fit indices.

Root Mean Square Error of Approximation:
The Root Mean Square Error of Approximation (RMSEA) is a fit index based on the chi-square test statistic, which corrects for parsimony, i.e. overly complex models are penalized. RMSEA can be greater or equal than zero, with values close to zero suggesting an acceptable model fit.
In our case, the RMSEA is 0.063. The upper bound of the 90% confidence interval of the RMSEA is 0.085 and smaller than the threshold value 0.1, suggesting an acceptable model fit.

Standardized Root Mean Square Residual:
The Standardized Root Mean Square Residual (SRMR) is a fit index derived from the residual correlation matrix with a range between zero and one with values close to zero suggesting an acceptable model fit.
In our case, the SRMR is 0.03 which is smaller than the threshold value 0.06 suggesting an acceptable model fit.

User Model versus Baseline Model
The Comparative Fit Index (CFI), evaluates the fit of the the model in relation to the worst-fitting baseline model described above. It ranges between zero and one, with values close to one suggesting good models (in the sense of departure from the baseline model).
In our case, the CFI is 0.984 which is greater or equal than the threshold value 0.95, suggesting a good model fit.

Similarly to the CFI, the Tucker-Lewis Index (TLI) evaluates the fit of the model in relation to the worst-fitting baseline model described above. Moreover, overly complex models are penalized. Values can range outside zero and one but the index is interpreted similarly to the CFI.
In our case, the TLI is 0.971 which is greater or equal than the threshold value 0.95, suggesting a good model fit.

Summary of the Goodness of Fit Indices

All the indices besides the Chi-square model fit index suggest an acceptable model fit. The Chi-square model fit index is based on a very stringent statistical hypothesis. Therefore, we ignorit and tentatively assume an acceptable model fit. We verify this assertion by considering further metrics.

Residuals

We analyze the residual matrices from the Outputs chapter. The residual covariance matrix represents the difference between the observed covariance matrix and the fitted model-implied covariance matrix. Large absolute values indicate local areas of misfit. However, the residuals are affected by the raw metric and are difficult to interpret more precisely.
A better interpretation allows the standardized residual matrix (residuals divided by their estimated asymptotic standard error) and the residual correlation matrix.
Following variable pairs have standardized residuals which are larger or equal than the considered threshold 2.58 [@brown] or correlation residuals which are larger or equal than the considered threshold 0.1 [@kline]. In these cases, the zero-order covariance relationship between the involved variables is probably underestimated:

Table 4: Pairs with Underestimated Covariance

	Pair(s)	
Pair 1	ACTIV	MENTH

Following variable pairs have standardized residuals which are smaller or equal than the considered threshold -2.58 [@brown] or correlation residuals which are smaller or equal than the considered threshold -0.1 [@kline]. In these cases, the zero-order covariance relationship between the involved variables is probably overestimated:

Table 5: Pairs with Overestimated Covariance

	Pair(s)	
Pair 1	GENHLTH	VITAL
Pair 2	VITAL	GENHLTH

Depending on the sample size, the misspecification detected by the analysis of the residual covariance resp. correlation matrices can be relevant or could be in practice neglected. This is matter of subject in the next section.

Modification Indices

In the interpretation of the modification indices we rely mostly on [@brown] and [@mi]. We cite from [@brown]: "The modification index reflects an approximation of how much the overall model Chi² will decrease if the fixed or constrained parameter is freely estimated." In other words, if adding a line with a high modification index to the model, i.e. if adding a parameter, the overall goodness-of-fit may be improved. Nevertheless, this should be done only under certain conditions, described in the sequel.
We consider only modification indices greater or equal than 3.84 (which are statistically significant at 5% type I error). Next, we search only for modification indices which achieve a power of minimum 75% in detecting a (relevant) misspecification of at least 0.1 for (residual) correlations, respectively 0.4 for factor loadings. These are characterized in the decision column by the label "epc:m".
We remark that these conditions are not fulfilled for modification indices with respect to (residual) correlations. Therefore, there exist no significant and relevant modification indices with respect to (residual) correlations.
We remark that these conditions are not fulfilled for modification indices with respect to factor loadings. Therefore, there exist no significant and relevant modification indices with respect to factor loadings.

Parameter Estimates

Factor Loadings

We remark that the completely standardized factor loadings (section "Completely Standardized Parameter Estimates") are all statistically significant at 5% type I error. Moreover, in absolute value they are all greater than 0.4. This cutoff-value is considered in some CFA research areas a magnitude that is substantively meaningful [@brown]. Please consider also cutoff-values from your particular research area when interpreting the factor loadings. We summarize the interpretation of the completely standardized factor loadings in the next table:

Table 6: Check Completely Standardized Factor Loadings

Latent Variable	Observed Variable	Factor Loading[1]	P-Value	Significant?[2]	Relevance[3]	Direction	Check
PHYSF	ACTIV	0.68	<0.001	Yes	**	Ok	Ok
PHYSF	SOMA	0.74	<0.001	Yes	**	Ok	Ok
PHYSF	PAIN	0.62	<0.001	Yes	**	Ok	Ok
MENTF	MENTH	0.85	<0.001	Yes	***	Ok	Ok

MENTF	SOCF	0.93	<0.001	Yes	***	Ok	Ok
MENTF	VITAL	0.80	<0.001	Yes	***	Ok	Ok
GWB	GENHLTH					Ok	Ok
AGEF	AGE	1.00			***	Ok	Ok

[1] The completely standardized factor loading can be interpreted as the correlation with the factor.

[2] 5% type I error.

[3] Stars correspond to factor loadings cutoff-values: 0.4, 0.6, 0.8.

[4] The observed variable is probably not related to latent factor.

[5] Uncertain. The evidence is insufficient or the model is misspecified.

[6] Uncertain. Significant but small(er) effect size.

Next, we inspect the unstandardized factor loadings (section “Model Fit Summary”). We remark that the unstandardized factor loadings are all statistically significant at 5% type I error. Therefore, the significance test results for standardized and unstandardized factor loadings coincide (for non-marker variables). We summarize the interpretation of the unstandardized factor loadings in the next table(s):

Table 7: Check Unstandardized Factor Loadings

Latent Variable	Observed Variable	Factor Loading	P-Value	Significant?[1]	Direction
PHYSF	ACTIV	1.00			Ok
PHYSF	SOMA	0.87	<0.001	Yes	Ok
PHYSF	PAIN	0.15	<0.001	Yes	Ok
MENTF	MENTH	1.00			Ok
MENTF	SOCF	1.85	<0.001	Yes	Ok
MENTF	VITAL	2.36	<0.001	Yes	Ok
GWB	GENHLTH	1.00			Ok
AGEF	AGE	1.00			Ok

[1] 5% type I error.

Table 8: Interpretation of Unstandardized Factor Loadings

Interpretation of Unstandardized Factor Loadings
ACTIV is marker variable for PHYSF
A 1-unit increase in PHYSF leads to a 0.87 -unit increase in the SOMA
A 1-unit increase in PHYSF leads to a 0.15 -unit increase in the PAIN
MENTH is marker variable for MENTF
A 1-unit increase in MENTF leads to a 1.85 -unit increase in the SOCF
A 1-unit increase in MENTF leads to a 2.36 -unit increase in the VITAL
GENHLTH is marker variable for GWB
AGE is marker variable for AGEF

Factor Correlations

As noted by [@brown], “the interpretability of the size and statistical significance of factor intercorrelations depends on the specific research context.”Though, the largest estimated factor intercorrelation within the section “Completely Standardized Parameter Estimates” is 0.45 which we regard as a proof of a reasonable discriminant validity.

Factor Reliability

Error Variances

Table 9: Completely Standardized Error Variances and Communality

Observed Variable	Error Variance[1]	Communality[2,3]	P-Value	Significant Error Variance?[4]
ACTIV	0.54	0.46	<0.001	Yes
AGE	0.00	1.00		
GENHLTH	7.88	0.89	<0.001	Yes
MENTH	0.28	0.72	<0.001	Yes
PAIN	0.62	0.38	<0.001	Yes
SOCF	0.13	0.87	<0.001	Yes
SOMA	0.45	0.55	<0.001	Yes
VITAL	0.36	0.64	<0.001	Yes

[1] Can be interpreted as proportion of unexplained variance by the latent factor(s) (%).

[2] Corresponds to the squared factor loading.

[3] Can be interpreted as proportion of explained variance by the latent factor(s) (%).

[4] 5% type I error. Typically significant since a large portion of variance is not explained by the latent variable.

Error Correlations

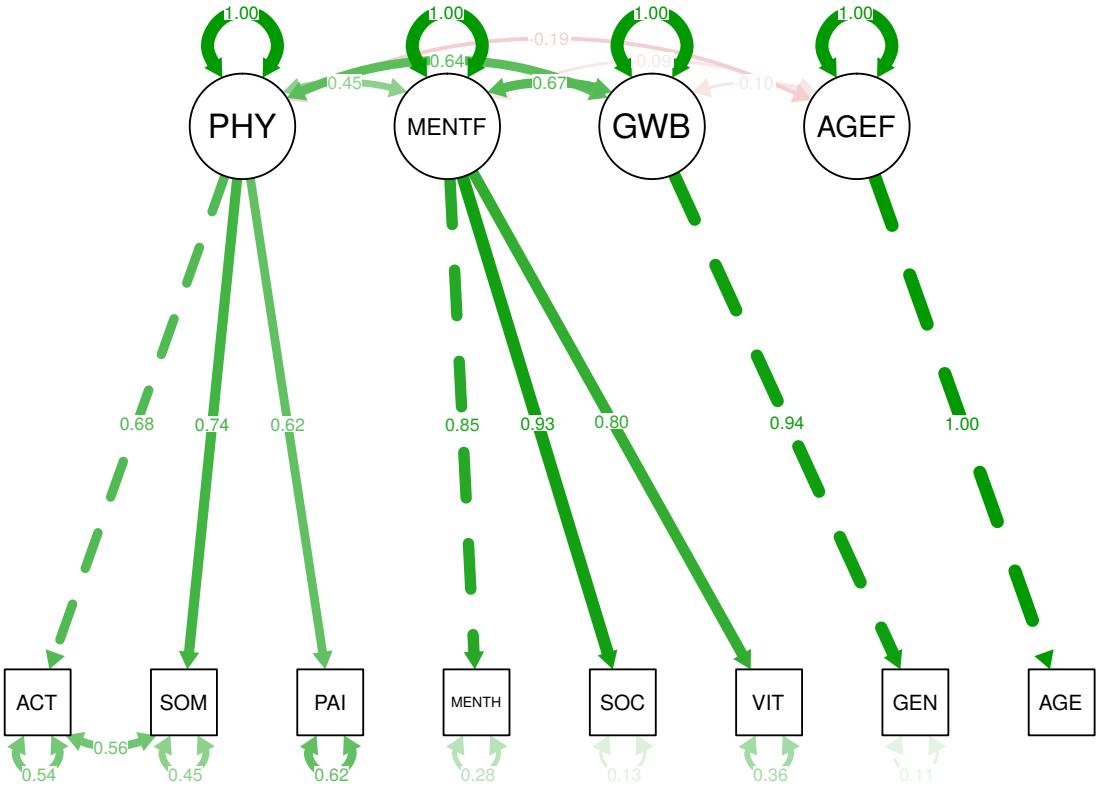
We remark, that all error covariances have completely standardized parameter estimates above 0.1 (in absolute value) and which are statistically significant at 5% type I error. Therefore, we assume that they are well defined in the model.

Intercepts

In case of missing values and estimation via FIML, a meanstructure i.e. the intercepts of the observed variables are added to the model. The means of the latent factors are fixed to zero. Therefore, the estimated intercepts within the section “Model Fit Summary” are just the means of the observed variables.

+Summary?

Path Diagram



References