

Übungsaufgabe CFA mit lavaan

Musterlösung

Stand: 27.07.2023

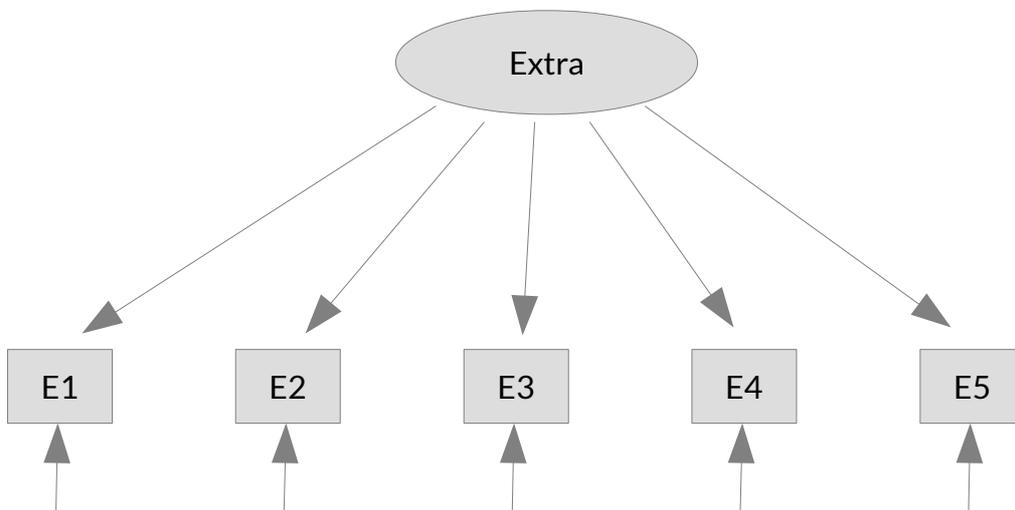
Testen Sie bitte auf Basis des Datensatzes *bfi* aus dem *psych* Modul von R, ob die fünf Items auf der Persönlichkeitsdimension Extraversion ein eindimensionales Konstrukt darstellen.

Starten Sie bitte damit, die umgekehrt gepolten Items zu rekodieren. Nutzen Sie dafür z.B. die Funktion *item.reverse* aus dem Modul *misty* (Sie können aber gerne auch eine andere Funktion zum Rekodieren nutzen).

Anschließend durchlaufen Sie bitte alle sechs Schritte einer normalen CFA-Modellierung:

- Model specification
- Model identification
- Model estimation
- Model evaluation
- Model respecification (wenn nötig)
- Model interpretation

Theoretisches Modell:



Anmerkung. Die Abbildung berücksichtigt keine eventuell invers kodierten Items.

Musterlösung für die Aufgabe zur CFA:

```
# laden der Module
# vorher ggf. einmalig installieren mit install.packages("...")
library(lavaan)
library(psych)
library(semPower)
library(misty)

head(bfi)
?bfi
```

Beim Betrachten der Items sehen wir, dass die Items 1 und 2 umgekehrt kodiert sind („Don't talk a lot“, „Find it difficult to approach others“), da hier hohe Werte für die Antworten einer geringen Extraversion entsprechen. Daher kodieren wir diese beiden Items um und verwenden die umkodierten Items in unserer Modellspezifikation.

```
# Rekodierung inverser Items
# mit misty package
bfi$E1r <- item.reverse(bfi$E1, min=1, max=6)
bfi$E2r <- item.reverse(bfi$E2, min=1, max=6)
```

1. Model Specification

```
extra_model_1 <- '
      # Factor loadings
      extrav =~ E1r + E2r + E3 + E4 + E5
      '
```

2. Model Identification

```
# Mit semPower package
semPower.getDf(extra_model_1)
```

```
[1] 5
```

Unser Modell hat fünf Freiheitsgrade (diesen Schritt würde ich normalerweise durchführen, bevor ich überhaupt die Daten erhebe), ist also global overidentified. Und außerdem hat jeder Faktor (hier gibt es ja nur einen) drei oder mehr Indikatoren. Damit sollten wir ein identifiziertes Modell haben, also ein Modell, was theoretisch schätzbar ist.

3. Model Estimation

Für eine CFA möchte ich die Faktorvarianzen auf 1 setzen, indem ich die Option *std.lv = TRUE* verwende. Aber es wäre auch korrekt, die voreingestellte Standardmethode (erste Ladung = 1) zu verwenden.

```
extra_fit_1 <- cfa(model = extra_model_1, data = bfi,
      std.lv = TRUE)
```

4. Model Evaluation

```
summary(extra_fit_1, fit.measures = TRUE, standardized=TRUE)
```

lavaan 0.6.15 ended normally after 18 iterations

Estimator	ML	
Optimization method	NLMINB	
Number of model parameters	10	
	Used	Total
Number of observations	2713	2800

Model Test User Model:

Test statistic	86.601
Degrees of freedom	5
P-value (Chi-square)	0.000

Model Test Baseline Model:

Test statistic	3015.292
Degrees of freedom	10
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.973
Tucker-Lewis Index (TLI)	0.946

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-23033.411
Loglikelihood unrestricted model (H1)	-22990.110
Akaike (AIC)	46086.822
Bayesian (BIC)	46145.880
Sample-size adjusted Bayesian (SABIC)	46114.107

Root Mean Square Error of Approximation:

RMSEA	0.078
90 Percent confidence interval - lower	0.064
90 Percent confidence interval - upper	0.092
P-value H ₀ : RMSEA ≤ 0.050	0.001
P-value H ₀ : RMSEA ≥ 0.080	0.410

Standardized Root Mean Square Residual:

SRMR	0.030
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
extrav =~						
E1r	0.992	0.032	30.693	0.000	0.992	0.608
E2r	1.173	0.031	38.021	0.000	1.173	0.730
E3	0.774	0.027	28.603	0.000	0.774	0.572
E4	1.012	0.028	35.755	0.000	1.012	0.692
E5	0.692	0.027	25.494	0.000	0.692	0.518

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.E1r	1.680	0.055	30.461	0.000	1.680	0.631
.E2r	1.206	0.050	24.285	0.000	1.206	0.467
.E3	1.230	0.039	31.569	0.000	1.230	0.673
.E4	1.111	0.042	26.653	0.000	1.111	0.521
.E5	1.307	0.040	32.899	0.000	1.307	0.732
extrav	1.000				1.000	1.000

Zunächst konnte der Algorithmus eine Lösung finden ("ended normally"). Der Chi-Quadrat-Test für die Modellgüte war signifikant (P -Wert $< .05$), daher haben wir keine perfekte Modellanpassung (das haben wir jedoch fast nie).

Der Fit-Index CFI ist gut (.973, höher als .95), ebenso der SRMR (.030, niedriger als .08). Der RMSEA ist immer noch zu hoch (.078, höher als .06). Da wir jedoch nur wenige Freiheitsgrade (5) haben, ist dies nicht allzu problematisch.

Alle fünf Items zeigen signifikante Ladungen auf den einen Faktor. Die standardisierten Ladungen liegen meistens zwischen .50 und .70. Ladungen über .70 wären besser (da ungefähr dann mehr als 50% der Varianz der Items durch den gemeinsamen Faktor erklärt würden). Da wir jedoch ein sehr breites Konstrukt betrachten und die Items verschiedene Facetten dieses Konstrukts ansprechen, denke ich, dass die Ergebnisse hier in Ordnung sind.

Und wir haben keine negativen Varianzen (Heywood-Fälle).

Basierend auf diesen Ergebnissen ist die globale Modellanpassung (global fit) in Ordnung. Aber wir sollten uns auch die lokale Anpassung (local fit) ansehen. Hierfür ziehe ich die Modifikationsindizes vor gegenüber der Residualmatrix.

```
modindices(extra_fit_1, minimum.value = 10, sort=TRUE)
```

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
20	E3	~~	E5	58.251	0.220	0.220	0.174	0.174
16	E2r	~~	E3	33.472	-0.209	-0.209	-0.172	-0.172
21	E4	~~	E5	23.613	-0.152	-0.152	-0.126	-0.126
12	E1r	~~	E2r	14.620	0.172	0.172	0.121	0.121

Die mögliche Modifikation mit dem höchsten Modifikationsindex (mi) befindet sich zwischen den Items:

E3 Know how to captivate people.

E5 Take charge.

Da die Fähigkeit, Menschen zu fesseln, beispielsweise die Fähigkeit zur Übernahme der Führung erleichtern kann, ist diese Modifikation für mich theoretisch sinnvoll. Angesichts des immer noch hoch signifikanten Modelltests und des hohen RMSEA (aber mit einer eher geringen Anzahl von Freiheitsgraden) werde ich diese Modifikation zulassen.

Allerdings, da in diesem Fall die Forschungsfrage lautet, ob wir eine eindimensionale latente Variable haben, halte ich es nicht für unbedingt notwendig, diese Modifikation in diesem Modell mit gutem CFI und SRMR hinzuzufügen. Wenn Sie sich entschieden hätten, Ihr Modell nicht weiter zu modifizieren, wäre das auch in Ordnung gewesen.

5. Model Respecification

Hier habe ich eine Fehlerkovarianz zwischen den Items E3 und E5 hinzugefügt.

```
extra_model_2 <- '  
  # Factor loadings  
  extrav =~ E1r + E2r + E3 + E4 + E5  
  
  #Error covariance  
  E3 ~~ E5  
'  
  
extra_fit_2 <- cfa(model = extra_model_2, data = bfi,  
  std.lv = TRUE)  
  
summary(extra_fit_2, fit.measures = TRUE, standardized=TRUE)
```

lavaan 0.6.15 ended normally after 22 iterations

Estimator	ML		
Optimization method	NLMINB		
Number of model parameters	11		
	Used	Total	
Number of observations	2713	2800	

Model Test User Model:

Test statistic	29.925
Degrees of freedom	4
P-value (Chi-square)	0.000

Model Test Baseline Model:

Test statistic	3015.292
Degrees of freedom	10
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.991
Tucker-Lewis Index (TLI)	0.978

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-23005.073
Loglikelihood unrestricted model (H1)	-22990.110
Akaike (AIC)	46032.146
Bayesian (BIC)	46097.110
Sample-size adjusted Bayesian (SABIC)	46062.160

Root Mean Square Error of Approximation:

RMSEA	0.049
90 Percent confidence interval - lower	0.033
90 Percent confidence interval - upper	0.066
P-value H ₀ : RMSEA ≤ 0.050	0.508
P-value H ₀ : RMSEA ≥ 0.080	0.001

Standardized Root Mean Square Residual:

SRMR	0.015
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
extrav =~						
E1r	1.002	0.032	30.851	0.000	1.002	0.614
E2r	1.192	0.031	38.206	0.000	1.192	0.742
E3	0.733	0.028	26.465	0.000	0.733	0.542
E4	1.020	0.029	35.708	0.000	1.020	0.698
E5	0.647	0.028	23.200	0.000	0.647	0.484

Covariances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.E3 ~~						
.E5	0.218	0.030	7.256	0.000	0.218	0.164

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.E1r	1.660	0.055	29.994	0.000	1.660	0.623
.E2r	1.161	0.051	22.747	0.000	1.161	0.450
.E3	1.291	0.041	31.863	0.000	1.291	0.706
.E4	1.095	0.042	25.818	0.000	1.095	0.513
.E5	1.367	0.041	33.067	0.000	1.367	0.765
extrav	1.000				1.000	1.000

Bei der Auswertung des umstrukturierten Modells sehen wir, dass die Modellanpassung verbessert wurde. CFI und SRMR sind sehr gut und auch im Modell mit nur 4 Freiheitsgraden ist der RMSEA ebenfalls gut.

Als Nächstes vergleichen wir dieses Modell mit dem vorherigen Modell mithilfe des Chi-Quadrat-Differenztests (Likelihood-Quotienten-Test, likelihood ratio test, LR-Test).

```
lavTestLRT(extra_fit_1, extra_fit_2)
```

Chi-Squared Difference Test

```
      Df   AIC   BIC  Chisq  Chisq diff  RMSEA Df diff Pr(>Chisq)
extra_fit_2  4 46032 46097 29.925
extra_fit_1  5 46087 46146 86.601      56.676  0.14325      1 5.139e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Der LR-Test zeigt einen signifikant besseren Modellfit ($p < .05$).

Zur Überprüfung des local fit fordere ich noch Modifikationsindizes an.

```
modindices(extra_fit_2, minimum.value = 10, sort=TRUE)
```

```
  lhs op rhs   mi   epc sepc.lv sepc.all sepc.nox
20  E3 ~~ E4 27.973  0.167  0.167  0.141  0.141
17 E2r ~~ E3 15.856 -0.143 -0.143 -0.117 -0.117
21  E4 ~~ E5 10.632 -0.101 -0.101 -0.083 -0.083
```

Es gibt immer noch mögliche Modifikationen. Der höchste verbleibende Modifikationsvorschlag befindet sich zwischen den Items:

E3 Know how to captivate people.

E4 Make friends easily.

Man könnte argumentieren, dass dies eine theoretisch sinnvolle Modifikation ist. Allerdings, da in diesem Fall die Forschungsfrage lautet, ob wir eine eindimensionale latente Variable haben, halte ich es nicht für strikt notwendig, diese Modifikation hinzuzufügen, da wir bereits eine sehr gute Modellanpassung aufgrund aller drei Fit-Indizes erreicht haben.

Wenn Sie sich aber dafür entschieden haben, Ihr Modell weiter zu modifizieren und eine Fehlerkovarianz zwischen E3 und E4 zuzulassen, wäre das auch in Ordnung.

6. Model Interpretation

Basierend auf dem guten Modellfit des modifizierten Modells und den standardisierten Ladungen kann man schlussfolgern, dass dieses Messinstrument für Extraversion im Wesentlichen eindimensional ist.