

Forschungspraktikum

Gruppenbezogene Menschenfeindlichkeit

18. Juni 2007:

Konfirmatorische Faktorenanalyse zur Prüfung eines Messmodells

Konfirmatorische Faktorenanalyse

Das zentrale Problem der explorativen Faktorenanalyse ist, dass es bei Modellen mit mehr als einem Faktor beliebig viele gleichwertige Lösungen gibt.

Eine "Lösung" ist dabei die spezifische Beziehungsstruktur unter den Faktoren sowie zwischen den Faktoren und den Indikatoren, d.h. sie ist gekennzeichnet durch die Korrelation unter den Faktoren, durch die Regressionsgewichte (Ladungen), die die Effekte der Faktoren auf die Indikatoren erfassen, und die durch die Varianzen der Residuen, also die Uniqueness.

Ursache ist, dass sich die reproduzierten Korrelationen nach der Formel:

$$\hat{r}_{ij} = \sum_{n=1}^K \sum_{m=1}^K \lambda_{in} \cdot \lambda_{jm} \cdot r(F_n, F_m)$$

berechnen, und es beliebig viele Kombinationen der Ladungen λ_{bm} und Faktorkorrelationen $r(F_n, F_m)$ gibt, die zu identischen Korrelationen führen.

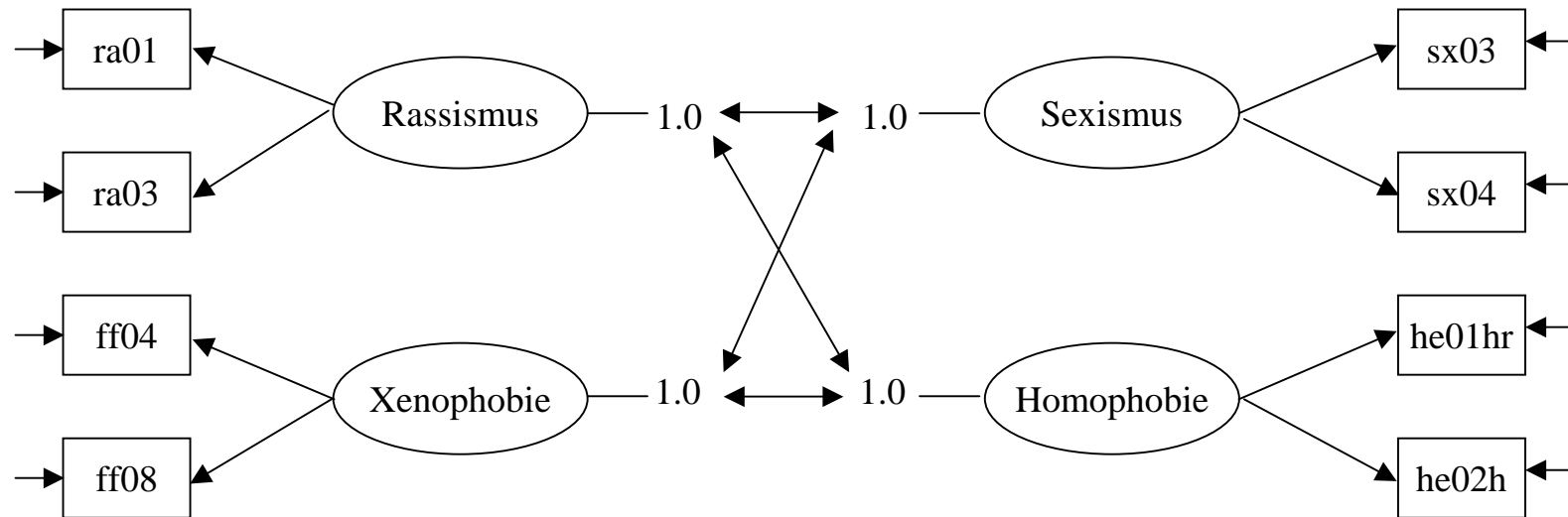
Eine eindeutige Lösung gibt es nur dann, wenn bestimmte Ladungen a priori auf den Wert "0" festgesetzt sind, weil es dadurch unmöglich wird, Lineartransformationen der Faktoren durchzuführen, die einerseits zu den gleichen reproduzierten Korrelationen führen, andererseits aber die auf "0" festgesetzten Werte unverändert lassen.

Inhaltlich bedeutet diese Festsetzung auf "0", dass aus theoretischen Gründen unterstellt wird, dass bestimmte Indikatoren nicht direkt, sondern höchstens indirekt durch einen Faktor beeinflusst werden. Vorausgesetzt wird also eine Vorstellung über die inhaltliche Bedeutung der Faktoren und deren Einfluss auf die Indikatoren.

Konfirmatorische Faktorenanalyse

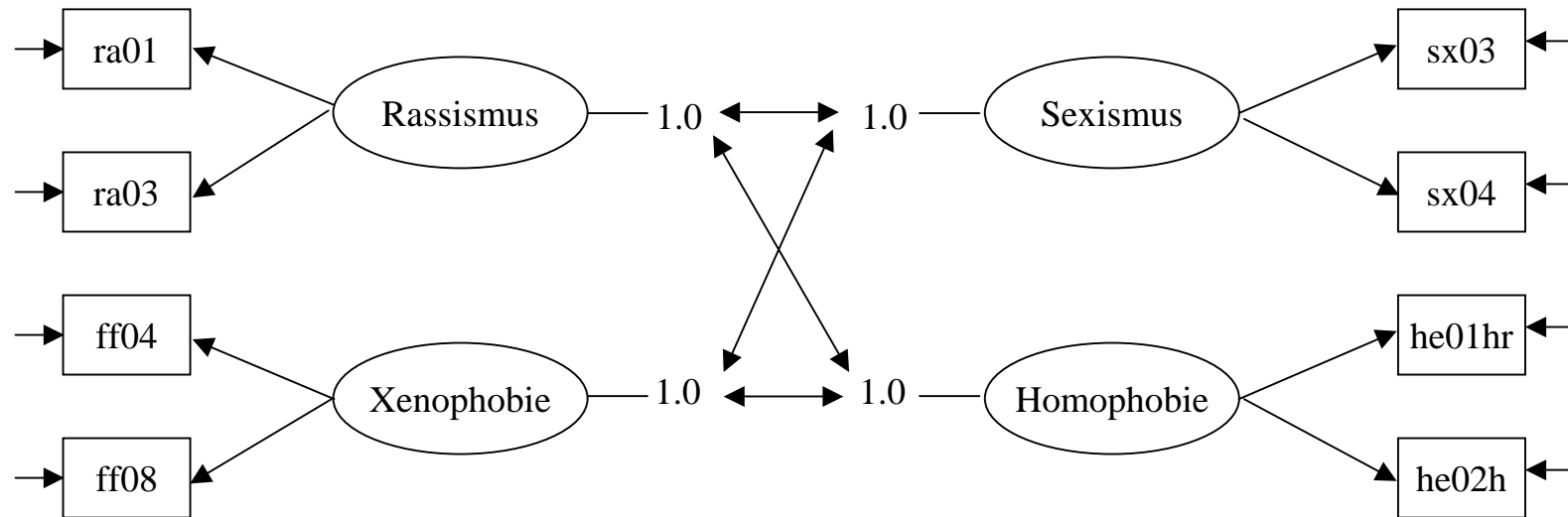
Am Beispiel der Indikatoren von Rassismus, Fremdenfeindlichkeit, Sexismus und Homophobie soll dies verdeutlicht werden.

Wenn die jeweils zwei Indikatoren dieser GMF-Dimensionen tatsächlich nur auf jeweils einen Faktor laden, ergibt sich folgendes Pfaddiagramm:



Im Pfaddiagramm stehen die Indikatoren in Rechtecken und die Faktoren in Oliven. Gerichtete Pfeile stehen für kausale Beziehungen, ungerichtete Pfeile (mit 2 Spitzen) für Korrelationen oder Kovarianzen. Die Residualvariablen (bzw. Uniqueness) ist nicht explizit aufgeführt, sondern nur durch Pfeile auf die Indikatoren gekennzeichnet. Um auszudrücken, dass die Faktoren standardisiert sind, sind sie mit Linien verbunden, hinter denen jeweils der Wert 1.0 für die Faktorvarianz steht.

Konfirmatorische Faktorenanalyse



Aus dem Pfaddiagramm geht hervor, dass die beiden Indikatoren für Rassismus ausschließlich auf den "Rassismus" genannten Faktor laden und entsprechendes für die Indikatoren der anderen drei GMF-Dimensionen gilt. Außerdem sind die Faktoren untereinander korreliert.

Diese Struktur wird theoretisch erwartet. Anhand der GMF-Daten soll geprüft werden, ob sie zutrifft. Wegen dieser Prüfung einer theoretisch erwarteten Beziehungsstruktur spricht man von konfirmatorischer Faktorenanalyse.

Zur Prüfung der Modellstruktur werden die GFM-Daten verwendet. Im Unterschied zur klassischen explorativen Faktorenanalyse, die Korrelationen analysiert, sind in der konfirmatorischen Faktorenanalyse i.a. Varianzen und Kovarianzen Ausgangsdaten.

Konfirmatorische Faktorenanalyse

Im ersten Schritt werden diese berechnet. Dies kann etwa über die SPSS-Prozedur Regression erfolgen:

```
compute he01hr=5-he01h.
regression var=ra01 ra03 ff04d ff08d sx03 sx04 he01hr he02h
/dep ra01 /ent ra03.
```

Korrelationen

		ra01 Aussiedler besser stellen	ra03 Weißen sind zurecht führend in der Welt	ff04d Es leben zu viele Ausländer in D.	ff08d Ausländer nach Hause schicken, wenn Arbeitsplätze knapp	sx03 Rolle der Ehefrau u Mutter für Frauen	sx04 Als Karrierehilfe für den Mann	he01hr	he02h Ekelhaft, wenn Homosexuell e sich küssen
Kovarianz	ra01 Aussiedler besser stellen	.699	.245	.247	.268	.191	.204	.199	.188
	ra03 Weißen sind zurecht führend in der Welt	.245	.614	.251	.257	.192	.222	.187	.178
	ff04d Es leben zu viele Ausländer in D.	.247	.251	1.016	.568	.214	.215	.283	.317
	ff08d Ausländer nach Hause schicken, wenn Arbeitsplätze knapp	.268	.257	.568	.908	.207	.223	.264	.282
	sx03 Rolle der Ehefrau u Mutter für Frauen	.191	.192	.214	.207	.777	.434	.348	.280
	sx04 Als Karrierehilfe für den Mann	.204	.222	.215	.223	.434	.682	.299	.290
	he01hr	.199	.187	.283	.264	.348	.299	1.296	.646
	he02h Ekelhaft, wenn Homosexuelle sich küssen	.188	.178	.317	.282	.280	.290	.646	1.154

Konfirmatorische Faktorenanalyse

Deskriptive Statistiken

	Mittelwert	N
ra01 Aussiedler besser stellen	3.19	9012
ra03 Weißen sind zurecht führend in der Welt	3.48	9012
ff04d Es leben zu viele Ausländer in D.	2.38	9012
ff08d Ausländer nach Hause schicken, wenn Arbeitsplätze knapp	2.87	9012
sx03 Rolle der Ehefrau u Mutter für Frauen	3.09	9012
sx04 Als Karrierehilfe für den Mann	3.24	9012
he01hr	2.9658	9012
he02h Ekelhaft, wenn Homosexuelle sich küssen	2.92	9012

Neben den Varianzen und Kovarianzen sind auch Mittelwerte angefordert. Dann wird automatisch auch die Fallzahl mit ausgegeben.

Für die konfirmatorische Faktorenanalyse gibt es in SPSS keine Prozedur. Hierfür wird im folgenden das Programm LISREL verwendet, das mit der sog. SIMPLIS-Syntax eine sehr einfache Möglichkeit besitzt, Modelle für konfirmatorische Faktorenanalyse einfach zu spezifizieren und zu schätzen.

In der SIMPLIS-Syntax werden Befehle in Anlehnung an die englische Sprache formuliert. Notwendige Informationen sind Variablennamen für die Indikatoren, Variablennamen für die Faktoren, als Daten entweder Rohdaten oder Varianzen und Kovarianzen plus der Fallzahl und die Angabe der Beziehung zwischen den Modellvariablen.

Die folgende Folie enthält alle Befehle einschließlich der Daten für die Durchführung einer konfirmatorischen Faktorenanalyse.

Die Befehle werden anschließend im einzelnen vorgestellt.

Konfirmatorische Faktorenanalyse

```
Faktorenmodell fuer Rassism, Xenophobie, Sexismus u. Homophobie
Observed Variables
  ra01 ra03 ff04d ff08d sx03 sx04 he01hr he02h
Latent Variables
  Rassism Xenophob Sexismus Homophob
Covariances
  0.699
  0.245 0.614
  0.247 0.251 1.016
  0.268 0.257 0.568 0.908
  0.191 0.192 0.214 0.207 0.777
  0.204 0.222 0.215 0.223 0.434 0.682
  0.199 0.187 0.283 0.264 0.348 0.299 1.296
  0.188 0.178 0.317 0.282 0.280 0.290 0.646 1.154
Sample Size 9012
Relationships
  ra01    ra03    = Rassism
  ff04d   ff08d   = Xenophob
  sx03    sx04    = Sexismus
  he01hr  he02h   = Homophob
Options ND=3 SC RS MI
Path diagram
End of problem
```

Konfirmatorische Faktorenanalyse

```
Faktorenmodell fuer Rassism, Xenophobie, Sexismus u. Homophobie  
Observed Variables  
ra01 ra03 ff04d ff08d sx03 sx04 he01hr he02h  
Latent Variables  
Rassism Xenophob Sexismus Homophob
```

Die erste Zeile ist die nicht notwendige Titelzeile, die zur Identifikation eines Modell hilfreich ist.

Anschließend werden mit den Befehlen "Observed Variables" und "Latent Variables" die Variablennamen der Indikatoren und der Faktoren (latenten Variablen) spezifiziert. Direkt hinter den Befehlen folgen die Variablennamen.

Ein Befehl muss am Beginn einer Zeile stehen. Allerdings können mehrere logische Zeilen durch Semikolon getrennt in eine physikalische Zeile der Befehlsdatei geschrieben werden.

Das LISREL-Programm scannt die logischen Zeilen der Befehlsdatei nach den Befehlnamen durch. Bei Befehlen ist es egal ob sie in Groß- oder Kleinbuchstaben geschrieben werden.

Bei den Variablennamen wird dagegen zwischen Groß- und Kleinschreibung unterschieden. In der SIMPLIS-Syntax wären "ra01", "RA01" "rA01" und "Ra01" also verschiedene Variablen.

Wenn man sich bei den Variablennamen verschreibt, läuft das Programm sofort auf einen Fehler.

Bei den Variablennamen werden nur die ersten 8 Zeichen gelesen. Längere Namen werden nach 8 Zeichen abgeschnitten. Es ist sinnvoll zu versuchen, mit den max. 8 Zeichen Namen zu geben, die die Bedeutung der Variablen erschließen lassen.

Konfirmatorische Faktorenanalyse

```
Covariances
ra01      0.699
ra03      0.245 0.614
ff04d     0.247 0.251 1.016
ff08d     0.268 0.257 0.568 0.908
sx03      0.191 0.192 0.214 0.207 0.777
sx04      0.204 0.222 0.215 0.223 0.434 0.682
he01hr    0.199 0.187 0.283 0.264 0.348 0.299 1.296
he02h     0.188 0.178 0.317 0.282 0.280 0.290 0.646 1.154
Sample Size 9012
```

Nach der Angaben der Indikatoren (Observed Variables) und Faktoren (Latent Variables) werden die Eingabedaten spezifiziert. Im Beispiel werden nach dem Befehl "Covariances" die Varianzen und Kovarianzen der Indikatoren eingelesen. Während in der SPSS-Ausgabe die gesamte quadratische Kovarianzmatrix ausgegeben wird, benötigt SIMPLIS nur die untere Hälfte einschließlich der Diagonalelemente. Die einzelnen Zahlen werden hintereinander durch jeweils eine Leerstelle getrennt eingegeben.

Sie können theoretisch auch in einer Zeile stehen. Durch die Dreiecksform ist aber leichter erkennbar, wofür welche Zahl steht: Die erste Zahl ist die Varianz der ersten beobachteten Variable, also die Varianz von "ra01". Die zweite Zahl ist die Kovarianz zwischen der ersten und der zweiten Variable, also die Kovarianz zwischen "ra01" und "ra03". Dann folgt die Varianz der zweiten Variable ("ra03"), dann wieder Kovarianzen usw..

Konfirmatorische Faktorenanalyse

```
Covariances
ra01      0.699
ra03      0.245 0.614
ff04d     0.247 0.251 1.016
ff08d     0.268 0.257 0.568 0.908
sx03      0.191 0.192 0.214 0.207 0.777
sx04      0.204 0.222 0.215 0.223 0.434 0.682
he01hr    0.199 0.187 0.283 0.264 0.348 0.299 1.296
he02h     0.188 0.178 0.317 0.282 0.280 0.290 0.646 1.154
Sample Size 9012
```

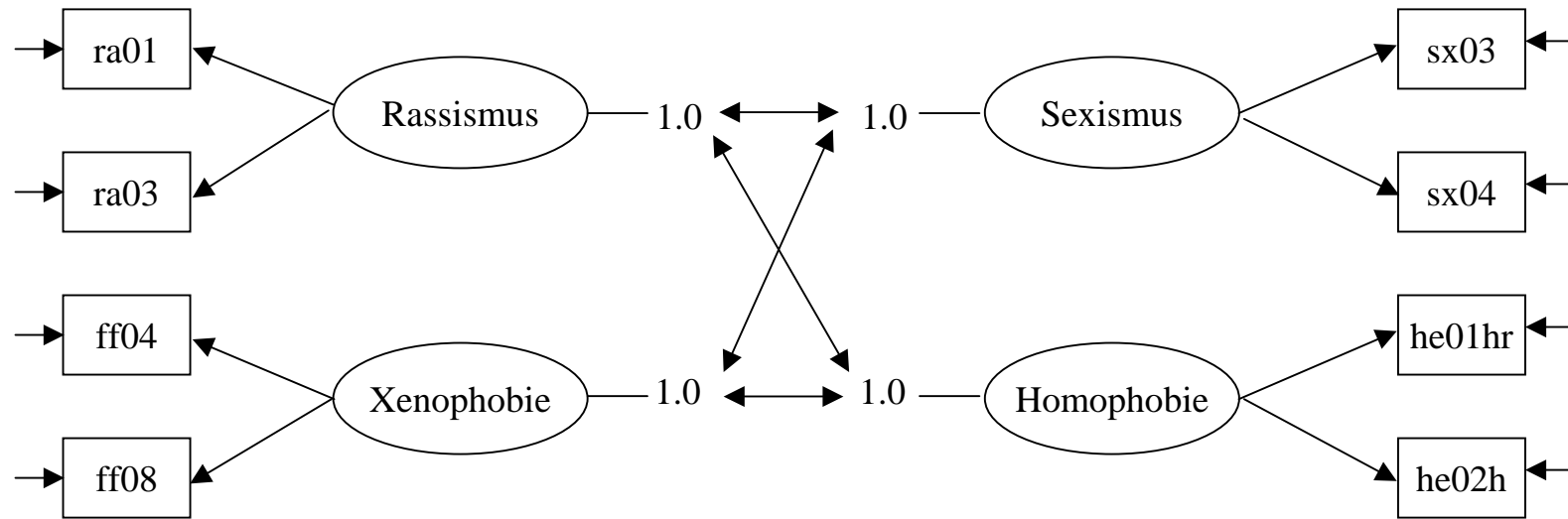
Die Varianzen und Kovarianzen können auch in einer Textdatei stehen, aus der sie eingelesen werden. Der Befehl würde dann lauten:

```
Covariances from file "c:\...\dateiname"
```

Wenn die Datei im selben Verzeichnis steht wie die Datei mit den SIMPLIS-Befehlen, kann auf die Angaben des Pfades und auf die Anführungsstriche verzichtet werden

Nach den Varianzen und Kovarianzen werden die Fallzahlen angegeben, auf denen die Berechnung der Varianzen und Kovarianzen beruht. Dazu wird hinter dem SIMPLIS-Befehl "Sample Size" die Fallzahl angegeben.

Konfirmatorische Faktorenanalyse



Relationships

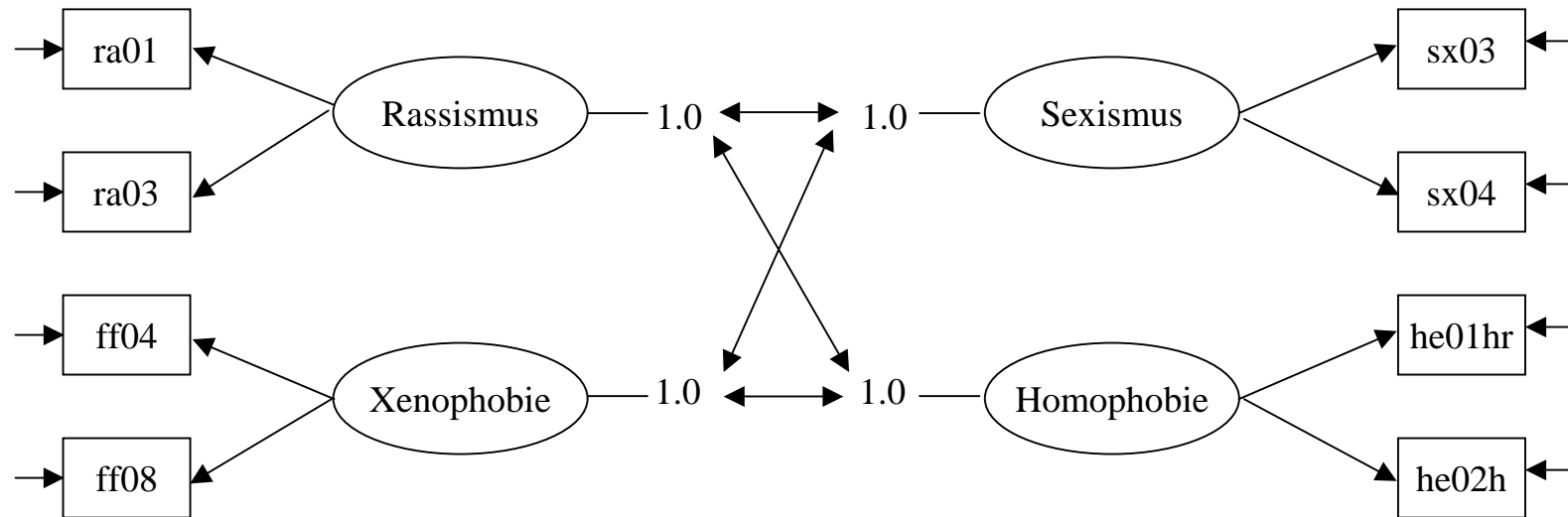
```
ra01    ra03    = Rassism
ff04d   ff08d   = Xenophob
sx03    sx04    = Sexismus
he01hr  he02h   = Homophob
```

Hinter dem Schlüsselwort "Relationships" folgt die Modellspezifikation, wie sie sich aus dem Pfaddiagramm ergibt.

Links steht eine oder mehrere abhängige Variablen, dann folgt ein Gleichheitszeichen, hinter dem die erklärende(n) Variablen folgen.

Aus der Spezifikation ist erkennbar, dass "ra01" und "ra03" durch "Rassism" erklärt wird, "ff04d" und "ff08d" durch "Xenophob", "sx03" und "sx04" durch "Sexismus" und "he01hr" und "he02h" durch "Homophob".

Konfirmatorische Faktorenanalyse



Relationships

ra01	ra03	=	Rassism
ff04d	ff08d	=	Xenophob
sx03	sx04	=	Sexismus
he01hr	he02h	=	Homophob

Es ist nicht nötig, die Residualvariablen anzugeben. Die Voreinstellungen von SIMPLIS sind so, dass für jede abhängige Variable automatisch eine Residualvariable existiert, deren Varianz geschätzt wird, und die nicht mit anderen Residualvariablen und den erklärenden Variablen korreliert ist.

Es ist auch nicht notwendig, die Korrelationen zwischen den Faktoren zu spezifizieren. Als Voreinstellung gilt in SIMPLIS, dass die Faktoren eine Varianz von "1" haben und alle unabhängigen Faktoren untereinander korrelieren.

Konfirmatorische Faktorenanalyse

```
Options ND=3 SC RS MI  
Path diagram  
End of problem
```

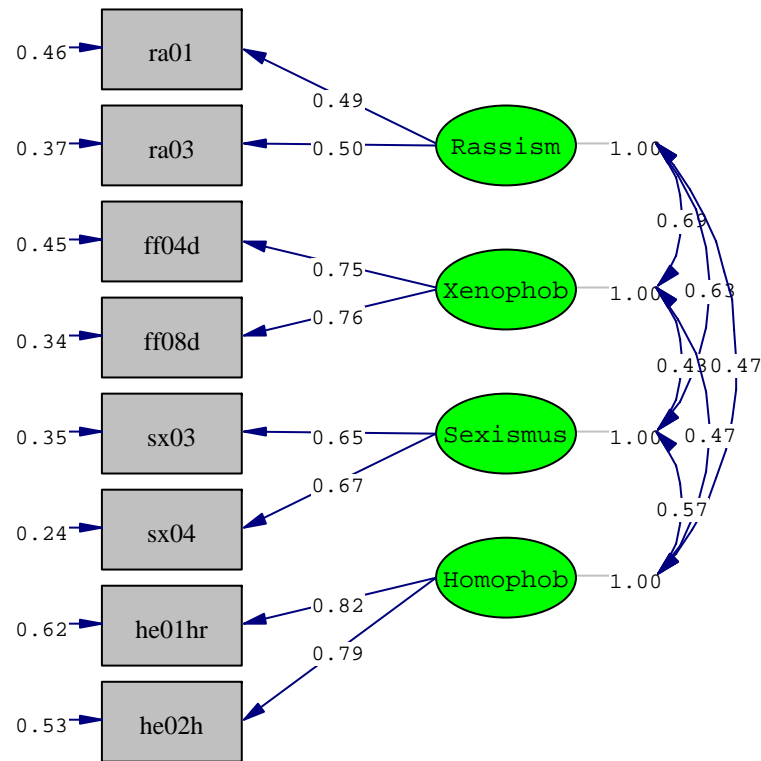
Die Modellspezifikation endet mit der Angabe von Optionen und dem Befehl "End of problem", der das Ende der Modellspezifikation signalisiert.

In den Optionen wird mit "ND=3" die Anzahl der signifikanten Stellen bei den Parametern auf 3 festgesetzt, voreingestellt sind 2 Stellen. Mit "SC" werden zusätzlich zu den unstandardisierten Koeffizienten auch die standardisierten Koeffizienten ausgegeben. Mit "RS" werden die reproduzierten Varianzen und Kovarianzen, sowie die Residuen, das sind die Differenzen zwischen den empirischen und den reproduzierten Varianzen und Kovarianzen, angefordert. Mit "MI" werden sog. "Modifikationsindizes" berechnet, die helfen, ein Modell zu verbessern, wenn es nicht zu den Daten passt.

Mit "Path diagram" kann schließlich ein Pfaddiagramm angefordert werden, in dem die Schätzungen der Koeffizienten eingetragen sind.

Für das Beispiel ist auf der nächsten Folie das von LISREL produzierte Pfaddiagramm wiedergeben.

Konfirmatorische Faktorenanalyse



Chi-Square=144.58, df=14, P-value=0.00000, RMSEA=0.032

Die Zahlen ganz links geben die geschätzten Residualvarianzen an, die Zahlen bei den blauen Pfeilen die unstandardisierten Ladungen, also die Effekte der Faktoren auf die Indikatoren. Rechts stehen in den kurvenförmigen Pfeilen mit Doppelspitzen die Korrelationen zwischen den Faktoren.

Konfirmatorische Faktorenanalyse

```
Goodness of Fit Statistics
Degrees of Freedom = 14
Minimum Fit Function Chi-Square = 144.821 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 144.580 (P = 0.0)
...
Root Mean Square Error of Approximation (RMSEA) = 0.0322
90 Percent Confidence Interval for RMSEA = (0.0275 ; 0.0370)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
...
```

Bevor ein Modell inhaltlich interpretiert wird, sollte zunächst überprüft werden, ob das Modell mit den Daten vereinbar ist, d.h. ob die reproduzierten Varianzen und Kovarianzen so gut mit den empirischen Varianzen und Kovarianzen übereinstimmen, dass davon ausgegangen werden kann, dass Abweichung vermutlich auf zufälligen Stichprobengegebenheiten basieren und jedenfalls keine inhaltliche bedeutsamen Abweichungen bestehen.

Zur Beurteilung findet sich in der Ausgabe eine Vielzahl von Kenngrößen, die die Übereinstimmung von Modell und Daten erfassen sollen. Hier sollen nur die beiden wichtigsten vorgestellt werden.

Mit einem Chi-Quadrat-Anpassungstest wird die Nullhypothese geprüft, dass die reproduzierten (in LISREL: modellimplizierten oder "fitted") Varianzen und Kovarianzen nicht signifikant von den empirischen Varianzen und Kovarianzen abweichen.

Konfirmatorische Faktorenanalyse

```
Goodness of Fit Statistics
Degrees of Freedom = 14
Minimum Fit Function Chi-Square = 144.821 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 144.580 (P = 0.0)
...
Root Mean Square Error of Approximation (RMSEA) = 0.0322
90 Percent Confidence Interval for RMSEA = (0.0275 ; 0.0370)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
...
```

Sinnvoll ist das nur, wenn ein Modell *überidentifiziert* ist. Dann werden weniger Koeffizienten geschätzt als Informationen vorliegen. Im Beispiel basiert das Modell auf 36 Varianzen und Kovarianzen. Geschätzt werden aber nur 8 Ladungen, 8 Residualvarianzen und 6 Faktorkorrelationen, also 22 Koeffizienten. Die Differenz zwischen 36 und 22 ist 14 und gleich der Zahl der Freiheitsgrade des Chiquadrattests.

Die Teststatistik wird auf zwei unterschiedliche Weisen berechnet. Im Pfaddiagramm ist die Statistik ausgegeben, die auf einer generalisierten Kleinstquadratschätzung beruht. Normalerweise wird allerdings die Statistik verwendet, die auf der sog. Minimierungs- oder Fitfunktion beruht und im Beispiel den Wert 144.821 aufweist.

Der zugehörige Signifikanzwert ist 0.0, d.h. die Wahrscheinlichkeit ist praktisch null, dass die Abweichungen der modellimplizierten Varianzen und Kovarianzen von den empirischen Varianzen und Kovarianzen nur auf Zufallsschwankungen der Stichprobe beruht.

Konfirmatorische Faktorenanalyse

```
Goodness of Fit Statistics
Degrees of Freedom = 14
Minimum Fit Function Chi-Square = 144.821 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 144.580 (P = 0.0)
...
Root Mean Square Error of Approximation (RMSEA) = 0.0322
90 Percent Confidence Interval for RMSEA = (0.0275 ; 0.0370)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
...
```

Bei sehr großen Fallzahlen führen allerdings bereits sehr kleine und vermutlich inhaltlich unbedeutende Abweichungen zu einem signifikanten Ergebnis, also der Ablehnung des spezifizierten Modells.

Daher wird neben dem exakten Übereinstimmungstest auch ein Test auf annähernde Übereinstimmung berechnet, der auf dem sog. "root mean square error of approximation" beruht. Aus Erfahrungswerten ist die Faustregel gebildet, dass der Wert des Koeffizienten RMSEA kleiner als 0.05 sein soll, wenn die Modellanpassung noch als halbwegs befriedigend gelten soll. Die Wahrscheinlichkeit, dass diese Bedingung erfüllt wird, ist im Beispiel praktisch 1.00, sodass nach diesem Kriterium das Modell mit den Daten vereinbar ist.

Neben dem Paddiagramm werden in der Ausgabedatei auch die Schätzungen der Modellgleichungen aufgeführt.

Konfirmatorische Faktorenanalyse

```
LISREL Estimates (Maximum Likelihood)
  Measurement Equations
  ra01 = 0.493*Rassism, Errorvar.= 0.456    , R2 = 0.348
        (0.0103)                          (0.00931)
        47.928                             48.969
  ra03 = 0.497*Rassism, Errorvar.= 0.367    , R2 = 0.402
        (0.00984)                          (0.00847)
        50.504                             43.321
  ff04d = 0.752*Xenophob, Errorvar.= 0.451  , R2 = 0.556
        (0.0115)                            (0.0126)
        65.438                             35.709
  ff08d = 0.756*Xenophob, Errorvar.= 0.337  , R2 = 0.629
        (0.0110)                            (0.0119)
        68.924                             28.299
  sx03 = 0.651*Sexismus, Errorvar.= 0.353   , R2 = 0.545
        (0.00988)                          (0.00932)
        65.880                             37.881
  sx04 = 0.667*Sexismus, Errorvar.= 0.238   , R2 = 0.652
        (0.00936)                          (0.00881)
        71.200                             26.960
  he01hr = 0.820*Homophob, Errorvar.= 0.623 , R2 = 0.519
        (0.0139)                            (0.0179)
        59.021                             34.854
  he02h = 0.787*Homophob, Errorvar.= 0.534 , R2 = 0.537
        (0.0132)                            (0.0162)
        59.775                             33.018
```

Konfirmatorische Faktorenanalyse

```
ra01 = 0.493*Rassism, Errorvar.= 0.456    , R2 = 0.348  
      (0.0103)                          (0.00931)  
      47.928                             48.969
```

Die Gleichungen geben die unstandardisierte Lösung wieder. Allerdings gibt es keine Interzepte (Regressionskonstanten), da das Modell davon ausgeht, dass alle Variablen zentriert sind, also in Abweichungen vom Mittelwert gemessen werden. Das ändert die Werte der unstandardisierten Regressionsgewichte nicht.

Im Beispiel ist erkennbar, dass der (standardisierte) Faktor Rassismus mit einem Regressionsgewicht von 0.493 auf den Indikator "ra01" wirkt. Steigt Rassismus um 1 Einheit, also 1 Standardabweichung, an erhöht sich der Mittelwert von Rassismus im Durchschnitt um 0.493 Einheiten, also eine knappe halbe Einheit auf der vierstufigen Ausgangsskala.

Unter dem geschätzten Wert ist in Klammern der Standardfehler und danach der T-Wert, das ist der Quotient aus Schätzwert und Standardfehler aufgeführt. Die T-Statistik ist asymptotisch standardnormalverteilt, wenn der Koeffizient in der Grundgesamtheit 0.0 ist. Werte größer oder gleich 2 oder kleiner oder gleich -2 weisen darauf hin, dass der Koeffizient bei einer Irrtumswahrscheinlichkeit von etwa 5% signifikant von null verschieden ist.

Die Messfehlervarianz beträgt 0.046, die Erklärungskraft 34.8%, d.h. 34.8% der Varianz des Indikators lässt sich auf den Faktor zurückführen. Dies ist gleichzeitig die Reliabilität oder Kommunalität des einzelnen Indikators.

Konfirmatorische Faktorenanalyse

Correlation Matrix of Independent Variables

	Rassism	Xenophob	Sexismus	Homophob
Rassism	1.000			
Xenophob	0.687 (0.013) 51.970	1.000		
Sexismus	0.627 (0.014) 46.024	0.433 (0.012) 35.249	1.000	
Homophob	0.471 (0.016) 30.319	0.471 (0.013) 37.244	0.568 (0.012) 48.495	1.000

Nach den unstandardisierten Regressionsgewichten (Ladungen) des Messmodells wird die Kovarianzmatrix bzw. wegen der Standardisierung Korrelationsmatrix der Faktoren ausgegeben. Wie gut zu erkennen ist, korrelieren die vier Faktoren sehr hoch miteinander.

Konfirmatorische Faktorenanalyse

Completely Standardized Solution

	LAMBDA-X			
	Rassism	Xenophob	Sexismus	Homophob
	-----	-----	-----	-----
ra01	0.590	- -	- -	- -
ra03	0.634	- -	- -	- -
ff04d	- -	0.746	- -	- -
ff08d	- -	0.793	- -	- -
sx03	- -	- -	0.739	- -
sx04	- -	- -	0.807	- -
he01hr	- -	- -	- -	0.721
he02h	- -	- -	- -	0.733

Die mit der Option "SC" angeforderte "Completely Standardized Solution" gibt die Werte der Modellparameter an, wenn alle Indikatoren und Faktoren standardisiert sind.

Diese Werte entsprechen also den Ladungen in der Mustermatrix einer explorativen Faktorenanalyse bei einer schiefwinkligen Rotation bzw. der Faktormatrix bei einer orthogonalen Lösung. Gut erkennbar ist, dass jeder Indikator immer nur auf einen Faktor lädt, wodurch die Eindeutigkeit der Lösung garantiert wird.

Die vollständig standardisierten Ladungen sollten möglichst Werte von mindestens ± 0.5 aufweisen. Ist dies nicht der Fall, bedeutet das, dass die Faktoren die Indikatoren nicht sehr stark beeinflussen und dann umgekehrt die Schätzungen der Korrelationen unter den Faktoren unsicherer werden.

Konfirmatorische Faktorenanalyse

Xenoph Sexism Homoph Rassism
 Mustermatrix^a

Completely Standardized Solution

LAMBDA-X

	Rassism	Xenophob	Sexismus	Homophob
ra01	0.590	- -	- -	- -
ra03	0.634	- -	- -	- -
ff04d	- -	0.746	- -	- -
ff08d	- -	0.793	- -	- -
sx03	- -	- -	0.739	- -
sx04	- -	- -	0.807	- -
he01hr	- -	- -	- -	0.721
he02h	- -	- -	- -	0.733

	Faktor			
	1	2	3	4
ra01 Aussiedler besser stellen	.016	-.018	.023	.585
ra03 Weißen sind zurecht führend in der Welt	.022	.031	-.013	.603
ff04d Es leben zu viele Ausländer in D.	.774	.004	.023	-.035
ff08d Ausländer nach Hause schicken, wenn Arbeitsplätze knapp	.744	.001	-.020	.063
sx03 Rolle der Ehefrau u Mutter für Frauen	.013	.780	.021	-.058
sx04 Als Karrierehilfe für den Mann	-.010	.748	-.009	.074
he01hr	-.041	.030	.725	.014
he02h Ekelhaft, wenn Homosexuelle sich küssen	.047	-.017	.716	-.006

Extraktionsmethode: Hauptachsen-Faktorenanalyse.
 Rotationsmethode: Promax mit Kaiser-Normalisierung.
 a. Die Rotation ist in 6 Iterationen konvergiert.

Der Vergleich der vollstandardisierten Lösung mit der Mustermatrix bei einer explorativen Hauptachsen-Faktorenanalyse mit 4 vorgegebenen Faktoren und Promax-Rotation zeigt, dass die Ergebnisse relativ ähnlich sind.

Dies ist nur dann der Fall, wenn ein Modell recht gut auf die Daten passt. Dann sind auch die meisten Ladungen der explorativen Faktorenanalyse nahe null sind und nur die Ladungen haben große Werte, die auch in der konfirmatorischen Faktorenanalyse spezifiziert sind.

Konfirmatorische Faktorenanalyse

Xenoph Sexism Homoph Rassism

Korrelationsmatrix für Faktor

Faktor	1	2	3	4
1	1.000	.415	.460	.655
2	.415	1.000	.557	.603
3	.460	.557	1.000	.448
4	.655	.603	.448	1.000

Extraktionsmethode: Hauptachsen-Faktorenanalyse.
Rotationsmethode: Promax mit Kaiser-Normalisierung.

	Rassism	Xenophob	Sexismus	Homophob
Rassism	1.000			
Xenophob	0.687	1.000		
Sexismus	0.627	0.433	1.000	
Homophob	0.471	0.471	0.568	1.000

Die Korrelation unter den Faktoren sind bei der konfirmatorischen Faktorenanalyse höher als bei der explorativen Faktorenanalyse.

Kommunalitäten

	Anfänglich	Extraktion
ra01 Aussiedler besser stellen	.218	.354
ra03 Weißen sind zurecht führend in der Welt	.245	.397
ff04d Es leben zu viele Ausländer in D.	.385	.583
ff08d Ausländer nach Hause schicken, wenn Arbeitsplätze knapp	.402	.605
sx03 Rolle der Ehefrau u Mutter für Frauen	.391	.584
sx04 Als Karrierehilfe für den Mann	.416	.616
he01hr	.328	.533
he02h Ekelhaft, wenn Homosexuelle sich küssen	.330	.528

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

$R^2 = 0.348$
$R^2 = 0.402$
$R^2 = 0.556$
$R^2 = 0.629$
$R^2 = 0.545$
$R^2 = 0.652$
$R^2 = 0.519$
$R^2 = 0.537$

Die erklärten Varianzen der konfirmatorischen Faktorenanalyse entsprechen den Kommunalitäten der explorativen Faktorenanalyse.

Es gibt Abweichungen, aber das Muster ist doch ähnlich.

Das Beispiel verdeutlicht, dass eine explorative Faktorenanalyse durchaus zu vergleichbaren Ergebnissen wie eine konfirmatorische Faktorenanalyse kommt, wenn es eine klare und relativ eindeutige Struktur in den Daten gibt.

Konfirmatorische Faktorenanalyse

Fitted Covariance Matrix

	ra01	ra03	ff04d	ff08d	sx03	sx04
ra01	0.699					
ra03	0.245	0.614				
ff04d	0.255	0.257	1.016			
ff08d	0.256	0.258	0.568	0.908		
sx03	0.201	0.203	0.212	0.213	0.777	
sx04	0.206	0.208	0.217	0.218	0.434	0.682
he01hr	0.190	0.192	0.291	0.292	0.303	0.311
he02h	0.183	0.184	0.279	0.280	0.291	0.298
	he01hr	he02h				
he01hr	1.296					
he02h	0.646	1.154				

Über die Option "RS" wird die durch das Modell reproduzierte Varianz-Kovarianzmatrix angefordert.

Von größeren Interesse ist allerdings die Matrix der Residuen, die bei dieser Option ebenfalls ausgegeben wird und Hinweise darauf gibt, wann eine Varianz bzw. Kovarianz überschätzt (negative Werte) bzw. unterschätzt (positive Werte) wird,

Konfirmatorische Faktorenanalyse

Fitted Residuals						
	ra01	ra03	ff04d	ff08d	sx03	sx04
	-----	-----	-----	-----	-----	-----
ra01	0.000					
ra03	0.000	0.000				
ff04d	-0.008	-0.006	0.000			
ff08d	0.012	-0.001	0.000	0.000		
sx03	-0.010	-0.011	0.002	-0.006	0.000	
sx04	-0.002	0.014	-0.002	0.005	0.000	0.000
he01hr	0.009	-0.005	-0.008	-0.028	0.045	-0.012
he02h	0.005	-0.006	0.038	0.002	-0.011	-0.008
	he01hr	he02h				
	-----	-----				
he01hr	0.000					
he02h	0.000	0.000				

Im Beispiel sind die Werte sehr klein, am größten ist noch die Unterschätzung der Kovarianz zwischen "sx03" und "he01hr".

Allerdings muss auch die Skalierung berücksichtigt werden.

Informativer sind daher die sog. *standardisierten Residuen*, die dich ergeben, wenn die Residuen durch ihre jeweiligen geschätzten Standardfehler geteilt werden. Wenn das Modell korrekt spezifiziert ist, sind die Residuen asymptotisch normalverteilt, allerdings nicht statistisch unabhängig voneinander. Trotzdem weisen sehr große oder sehr kleine Werte darauf hin, dass das Modell die betreffende Varianz oder Kovarianz möglicherweise nicht sehr gut wiedergibt.

Konfirmatorische Faktorenanalyse

Standardized Residuals						
	ra01	ra03	ff04d	ff08d	sx03	sx04
	-----	-----	-----	-----	-----	-----
ra01	- -					
ra03	- -	- -				
ff04d	-1.870	-1.655	- -			
ff08d	3.499	-0.334	- -	- -		
sx03	-2.523	-3.128	0.374	-1.394	- -	
sx04	-0.632	5.110	-0.523	1.385	- -	- -
he01hr	1.412	-0.914	-1.232	-5.433	9.760	-3.399
he02h	0.932	-1.260	6.769	0.355	-2.683	-2.625
	he01hr	he02h				
	-----	-----				
he01hr	- -					
he02h	- -	- -				

Insgesamt sind vier Werte außergewöhnlich weit von null verschieden, was darauf hinweist, dass das Modell nicht ganz optimal ist.

Die sog. **Modifikationsindizes** geben dann Hinweise, wie das Modell verbessert werden kann. Ein Modifikationsindex gibt eine Schätzung an, um welchen Wert sich der Chi-Quadratwert des Anpassungstests vermutlich verringert, wenn ein fixierter Parameter frei geschätzt würde. Bei einer solchen Modifikation verliert das Modell 1 Freiheitsgrad. Es kann aber sein, dass die Verringerung beim Chi-Quadrat dies deutlich kompensiert.

Konfirmatorische Faktorenanalyse

	Modification Indices for LAMBDA-X				Expected Change for LAMBDA-X			
	Rassism	Xenophob	Sexismus	Homophob	Rassism	Xenophob	Sexismus	Homophob
ra01	- -	3.115	3.090	2.138	- -	0.045	-0.036	0.024
ra03	- -	3.114	3.090	2.138	- -	-0.045	0.036	-0.024
ff04d	17.561	- -	0.018	18.253	-0.212	- -	0.002	0.076
ff08d	17.561	- -	0.018	18.253	0.213	- -	-0.002	-0.076
sx03	25.241	1.395	- -	29.106	-0.123	-0.017	- -	0.109
sx04	25.241	1.395	- -	29.106	0.126	0.017	- -	-0.111
he01hr	0.878	26.987	27.046	- -	-0.029	-0.116	0.188	- -
he02h	0.878	26.987	27.047	- -	0.028	0.111	-0.180	- -

Tatsächlich sind einige Modifikationsindizes recht groß. Setzt man sie frei, würde sich der Modellfit deutlich verbessern.

Neben den Modifikationsindizes wird auch eine Schätzung angegeben, welche Werte sich vermutlich ergeben würden, wenn ein vorgegebener Parameterwert zum Schätzen freigegeben wird.

Durch die Freigabe von Ladungen nähert sich das Modell immer mehr einer explorativen Faktorenanalyse an. Wenn das Modell nicht mehr eindeutig ist, gibt es eine Warnung, dass das Modell möglicherweise unteridentifiziert ist. Dann können weder Standardfehler noch Chiquadrat-tests berechnet werden.

Generell gilt beim Modifizieren die Regel, dass immer nur ein Parameter verändert wird und dass möglichst nur solche Parameter verändert werden, die auch inhaltlich plausible sind.

Konfirmatorische Faktorenanalyse

Modification Indices for THETA-DELTA								
	ra01	ra03	ff04d	ff08d	sx03	sx04	he01hr	he02h
ra01	- -							
ra03	- -	- -						
ff04d	6.442	0.298	- -					
ff08d	10.089	0.032	- -	- -				
sx03	1.284	12.583	1.915	2.166	- -			
sx04	0.699	26.254	3.545	3.779	- -	- -		
he01hr	2.125	0.052	2.003	8.366	82.763	31.541		
he02h	0.048	1.735	27.459	0.644	22.005	2.285	- -	- -

Im Unterschied zur explorativen Faktorenanalyse ist es bei der konfirmatorischen Faktorenanalyse möglich, Kovarianzen zwischen den Residualvariablen zu erlauben. Inhaltlich bedeutet eine solche Kovarianz, dass zwischen zwei Indikatoren eine Beziehung besteht, die nicht vollständig durch die Faktorenstruktur wiedergegeben wird.

Im Beispiel scheint es eine Beziehung zwischen sx03 (Frauen sollen sich auf ihre Rolle als Ehefrau und Mutter beschränken) und he01h (Ehen zwischen zwei Männern oder Frauen sollten möglich sein) zu geben, die nicht hinreichend durch die Faktorenkorrelation zwischen "Sexismus" und "Homopobie" erklärt werden kann.

Die erwartete Änderung des Wertes bei einer Schätzung ergibt 0.069, also einen positiven Wert. Möglicherweise ist es der Stimulus "Ehe", der in beiden Itemformulierungen vorkommt und eine zusätzliche positive Assoziation schafft.

Konfirmatorische Faktorenanalyse

Mit dem SIMPLIS-Befehl "Set ... free" kann ein gefixter Parameterwert freigegeben werden. Hier wird die Residualkovarianz zwischen sx03 und he01hr freigegeben:

```
Modif. Modell: Residualkovarianz zwischen sx03 und he01hr
... Relationships
ra01    ra03    = Rassism
ff04d   ff08d   = Xenophob
sx03    sx04    = Sexismus
he01hr  he02h   = Homophob
Set error covariance of sx03 and he01hr free
...
```

Die Freigabe verbessert den Modellfit erwartungsgemäß sehr deutlich:

Ausgangsmodell:

```
Goodness of Fit Statistics
Degrees of Freedom = 14
Minimum Fit Function Chi-Square = 144.821 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 144.580 (P = 0.0)
```

Modifiziertes Modell:

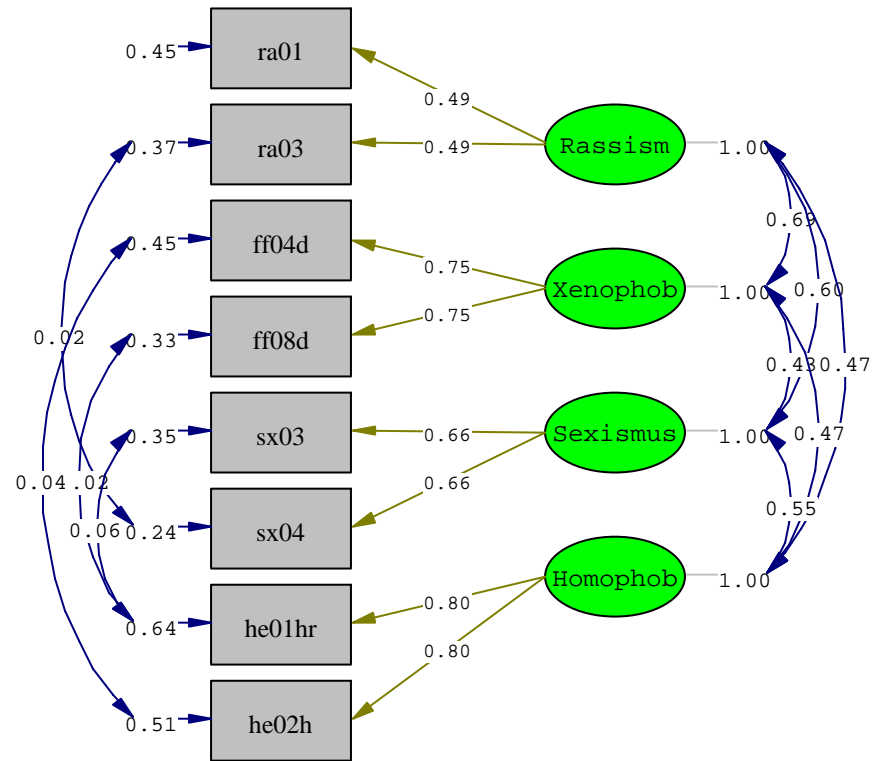
```
Goodness of Fit Statistics
Degrees of Freedom = 13
Minimum Fit Function Chi-Square = 62.455 (P = 0.000)
Normal Theory Weighted Least Squares Chi-Square = 62.797 (P = .000)
```

Konfirmatorische Faktorenanalyse

Über weitere Modifikationen kommt man schließlich zu einem Modell, das sehr gut auf die Daten passt. Neben zusätzlichen freien Residualkovarianzen sind in diesem Modell jeweils die Ladungen der Indikatoren eines Faktors mit dem Simplis Befehl "Set ... and ... equal" gleichgesetzt. Das bedeutet, dass die geschätzten unstandardisierten Ladungen den gleichen Wert aufweisen.

```
Modif. Modell: Residualkovarianz zwischen sx03 und he01hr
... Relationships
ra01    ra03    = Rassism
ff04d   ff08d   = Xenophob
sx03    sx04    = Sexismus
he01hr  he02h   = Homophob
Set error covariance of sx03 and he01hr free
set error covariance of ff04d and he02h free
set error covariance of ff08d and he01hr free
set Rassism -> ra01 and Rassism -> ra03 equal
set Homophob -> he01hr and Homophob -> he02h equal
set Xenophob -> ff04d and Xenophob -> ff08d equal
set Sexismus -> sx03 and Sexismus -> sx04 equal
set error covariance of ra03 and sx04 free
...
```

Konfirmatorische Faktorenanalyse



Chi-Square=21.00, df=14, P-value=0.10159, RMSEA=0.007

Der Chi-Quadratwert der Minimierungsfunktion beträgt 20.734 bei $df=14$, was zu einer empirischen Signifikanz von 0.109 führt. Mit einer Wahrscheinlichkeit von immerhin knapp 11% weichen die modellimplizierten und die empirischen Kovarianzen wie in diesem Modell oder noch stärker ab, wenn das spezifizierte Modell in der Population zutrifft.

Konfirmatorische Faktorenanalyse

Ausgangsmodell:

Completely Standardized Solution

LAMBDA-X

Rassism Xenophob Sexismus Homophob

	Rassism	Xenophob	Sexismus	Homophob
ra01	0.590	- -	- -	- -
ra03	0.634	- -	- -	- -
ff04d	- -	0.746	- -	- -
ff08d	- -	0.793	- -	- -
sx03	- -	- -	0.739	- -
sx04	- -	- -	0.807	- -
he01hr	- -	- -	- -	0.721
he02h	- -	- -	- -	0.733

PHI

Rassism Xenophob Sexismus Homophob

	Rassism	Xenophob	Sexismus	Homophob
Rassism	1.000			
Xenophob	0.687	1.000		
Sexismus	0.627	0.433	1.000	
Homophob	0.471	0.471	0.568	1.000

Gefittetes Modell:

Completely Standardized Solution

LAMBDA-X

Rassism Xenophob Sexismus Homophob

	Rassism	Xenophob	Sexismus	Homophob
	0.592	- -	- -	- -
	0.630	- -	- -	- -
	- -	0.746	- -	- -
	- -	0.794	- -	- -
	- -	- -	0.745	- -
	- -	- -	0.801	- -
	- -	- -	- -	0.708
	- -	- -	- -	0.747

PHI

Rassism Xenophob Sexismus Homophob

	Rassism	Xenophob	Sexismus	Homophob
	1.000			
	0.688	1.000		
	0.603	0.434	1.000	
	0.472	0.466	0.548	1.000

Der Vergleich der Resultate zeigt, dass sich das an die Daten angepasste Modell inhaltlich kaum von dem Ausgangsmodell unterscheidet, was für eine sehr robuste Lösung spricht.

Konfirmatorische Faktorenanalyse

Ausgangsmodell:

Completely Standardized Solution

THETA-DELTA

	ra01	ra03	ff04d	ff08d	sx03	sx04	he01hr	he02h
	0.652	0.598	0.444	0.371	0.455	0.348	0.481	0.463

Gefittetes Modell:

Completely Standardized Solution

THETA-DELTA

	ra01	ra03	ff04d	ff08d	sx03	sx04	he01hr	he02h
ra01	0.649							
ra03	- -	0.604						
ff04d	- -	- -	0.444					
ff08d	- -	- -	- -	0.369				
sx03	- -	- -	- -	- -	0.445			
sx04	- -	0.037	- -	- -	- -	0.358		
he01hr	- -	- -	- -	-0.021	0.058	- -	0.499	
he02h	- -	- -	0.036	- -	- -	- -	- -	0.442

Allerdings gibt es inhaltliche Unterschiede durch die zusätzliche Residualkovarianzen.

Konfirmatorische Faktorenanalyse

Wenn mit der SIMPLIS-Syntax Rohdaten eingelesen werden, müssen diese zunächst erzeugt werden. Dies kann etwa mit SPSS erfolgen:

```
count nm=ra01 ra03 ff04d ff08d sx03 sx04 he01hr he02h (missing).  
temporary.  
select if (nm=0).  
write outfile='C:\Anwender\Lehre\2007SS\ForschungspraktikumGMF\07-06-18\cfa.dat'  
  / ra01 ra03 ff04d ff08d sx03 sx04 he01hr he02h (8f2.0).  
exec.
```

Die Rohdaten können dann direkt in LISREL mit dem Befehl "Raw data from file" eingelesen werden.

Auf der nächsten Folie ist die gesamte Modellspezifikation angegeben.

Bis auf Rundungsfehler, die sich durch die nur 3 Nachkommastellen bei der Analyse von Varianzen und Kovarianzen anstelle der Rohdaten ergeben, ergibt sich die gleiche Lösung.

Konfirmatorische Faktorenanalyse

```
Gefittetes Modell mit Rohdateneingabe.  
Observed Variables  
  ra01 ra03 ff04d ff08d sx03 sx04 he01hr he02h  
Latent Variables  
  Rassism Xenophob Sexismus Homophob  
Raw data from file cfa.dat  
Relationships  
  ra01    ra03    = Rassism  
  ff04d   ff08d   = Xenophob  
  sx03    sx04    = Sexismus  
  he01hr  he02h   = Homophob  
Set error covariance of sx03 and he01hr free  
set error covariance of ff04d and he02h free  
set error covariance of ff08d and he01hr free  
set Rassism -> ra01 and Rassism -> ra03 equal  
set Homophob -> he01hr and Homophob -> he02h equal  
set Xenophob -> ff04d and Xenophob -> ff08d equal  
set Sexismus -> sx03 and Sexismus -> sx04 equal  
set error covariance of ra03 and sx04 free  
Options ND=3 SC RS MI  
Path diagram  
End of problem
```

Konfirmatorische Faktorenanalyse

Anwendungsvoraussetzungen

Die inferenzstatistischen Aussagen basieren darauf, dass die Fälle der Stichprobe als identisch verteilte voneinander unabhängige Realisationen einer multivariaten Normalverteilung aufgefasst werden können.

Diese Annahme ist nicht sehr realistisch, was bedeutet, dass Testergebnisse möglicherweise in die Irre führen.

Es kann jedoch gezeigt werden, dass die Normalverteilungsannahme nicht notwendig ist, wenn das spezifizierte Modell die Population korrekt beschreibt und die Residualvariablen nicht nur unkorreliert, sondern statistisch unabhängig von den substantiellen erklärenden Variablen, hier also den Faktoren sind. Dann sind die Schätzungen als Folge einer Version des zentralen Grenzwertsatzes asymptotisch normalverteilt und die Inferenzstatistik ist korrekt.

Wenn dies nicht der Fall ist, können alternative Schätzmethoden angewendet werden, die keine Normalverteilung voraussetzen, aber aufwendiger sind.

In der Realität sind schon als Folge der Stichprobenziehung die Fälle möglicherweise nicht unabhängig und identisch verteilt. Inzwischen ist es auch möglich, komplexe geschichtete und mehrstufige Stichprobenpläne zu berücksichtigen.