Exploratorische Faktorenanalyse

- Der Begriff Faktorenanalyse umfasst eine Gruppe multivariater Analyseverfahren, mit denen zugrundeliegende gemeinsame Dimensionen von Variablenmengen (z.B. Fragebogenitems) untersucht werden.
- Die Faktorenanalyse führt zu einer Datenreduktion: Die Variation in einer Vielzahl von Variablen wird auf eine geringere Zahl von gemeinsamen Faktoren zurückgeführt.

Exploratorische Faktorenanalyse

- Bei einer exploratorischen Faktorenanalyse existieren im „Idealfall“ keine Annahmen darüber, wie viele Faktoren den Variablen zugrunde liegen und welche Struktur die Zusammenhänge zwischen den Variablen haben.
- In der Praxis ist das natürlich fast nie der Fall; z.B. vermuten wir vorab, dass unseren Fragebögen nur eine einzige latente Dimension zugrunde liegt.

Exploratorische Faktorenanalyse

- Im Rahmen der exploratorischen Faktorenanalyse gehen wir, der Natur der Methode entsprechend, trotzdem „naiv“ vor. Die Fragen, welche am Ende der Analyse beantwortet sein sollen, sind:
  - **Wie viele** Faktoren sind angemessen, um die Zusammenhänge zwischen den untersuchten Variablen zu erklären?
  - **Welche** der Faktoren beeinflussen **welche** meiner Variablen?

Fundamentaltheorem der Faktorenanalyse

- Die Ausprägungen x_{im} von Individuen i in beobachteten Variablen m lassen sich als gewichtete Summe von $j = 1,...,q$ latenten Faktoren f_j ausdrücken:
  $$x_{im} = f_{i1} \cdot a_{m1} + ... + f_{ij} \cdot a_{mj} + ... + f_{iq} \cdot a_{mq} + e_{im}$$
  $$= \sum_{j=1}^{q} f_{ij} \cdot a_{mj} + e_{im}$$
  x_{im} = der Wert der Person i in der beobachteten Variablen m;
  f_{ij} = Faktorwert der Person i in Faktor j;
  a_{mj} = Ladung der Variablen m auf Faktor j;
  e_{im} = durch die Faktoren nicht erklärte Fehlerkomponente;
  q = Anzahl der Faktoren

Faktorenextraktion und Ladungsmatrix

- Ausgangspunkt einer Faktorenanalyse ist die Korrelationsmatrix der untersuchten Variablen.
- Auf Basis der Korrelationsmatrix erfolgt die Faktorenextraktion, d.h. die Bestimmung des Faktorenraumes durch die Schätzung der Ladungsmatrix, welche die Ladungen aller Variablen auf allen Faktoren umfasst.
- Sind die Faktoren definiert, bilden sie wiederum ein Koordinatensystem, innerhalb dessen jede Variable anhand ihrer Faktorladungen verortet werden kann.

Ladungsmatrix

- Die Ladungsmatrix A stellt das wichtigste Ergebnis der Faktorenanalyse dar. In ihr sind die Ladungen a_{mj} der Variablen m auf den Faktoren j zusammengefasst.
- Mithilfe der Ladungsmatrix A soll sich die Korrelationsmatrix der beobachteten Variablen reproduzieren lassen:
  $R = A'A$

Ladungen, Kommunalitäten, Eigenwerte

- Eine einzelne Faktorladung a_{mj} lässt sich interpretieren wie die Korrelation der beobachteten Variablen m mit dem Faktor j.
- Die Summe der quadrierten Ladungen *einer* Variablen auf *allen* Faktoren ergibt die Varianz dieser Variablen, die durch die Faktoren gemeinsam erklärt wird.
- Diese Größe wird als Kommunalität (h^2_m) einer Variablen bezeichnet.

Ladungen, Kommunalitäten, Eigenwerte

- Die Summe der quadrierten Ladungen *eines* Faktors auf *allen* Variablen ergibt die Varianz dieses Faktors, die er an der Gesamtvarianz der beobachteten Variablen erklärt.
- Diese Größe wird als Eigenwert (λ_j) eines Faktors bezeichnet.

Ladungen, Kommunalitäten, Eigenwerte

Schematische Veranschaulichung der Beziehungen von Faktorladungen, Kommunalitäten und Eigenwerten anhand des Ladungsmusters von vier Variablen auf zwei Faktoren.

Ladungsmatrix A

Variablen	Ladungen auf Faktor 1	Ladungen auf Faktor 2	Kommunalitäten
x_1	a_{11}	a_{12}	$\sum_{j=1}^{q} a_{1j}^2 = a_{11}^2 + a_{12}^2 = h_1^2$
x_2	a_{21}	a_{22}	$\sum_{j=1}^{q} a_{2j}^2 = a_{21}^2 + a_{22}^2 = h_2^2$
x_3	a_{31}	a_{32}	$\sum_{j=1}^{q} a_{3j}^2 = a_{31}^2 + a_{32}^2 = h_3^2$
x_4	a_{41}	a_{42}	$\sum_{j=1}^{q} a_{4j}^2 = a_{41}^2 + a_{42}^2 = h_4^2$
Eigenwerte	$\sum_{m=1}^{k} a_{m1}^2 = \lambda_1$	$\sum_{m=1}^{k} a_{m2}^2 = \lambda_2$	$\sum_{j=1}^{q} \lambda_j = \sum_{m=1}^{k} h_m^2$
% erklärte Varianz	$\frac{\lambda_1}{k} \cdot 100$	$\frac{\lambda_2}{k} \cdot 100$	$\frac{\lambda_1 + \lambda_2}{k} \cdot 100$ = Durch beide Faktoren insgesamt erklärte Varianz

k = 4 = Anzahl der Variablen; q = 2 = Anzahl der Faktoren.

Faktorenextraktion

- Die Faktorenextraktion erfolgt nach dem Kriterium, dass jeder Faktor sukzessive so viel Varianz in den untersuchten Variablen erklärt wie möglich.
- Jeder neue Faktor erklärt dabei nur Varianz, die von den zuvor extrahierten Faktoren nicht erklärt wurde.

Faktorenextraktion

- Zur Faktorenextraktion stehen verschiedene Berechnungsverfahren zur Verfügung. Die beiden bedeutendsten sind die Hauptkomponenten-analyse und die Hauptachsenanalyse.
- Beide Verfahren unterscheiden sich vor allem hinsichtlich der Annahmen bezüglich der durch die Faktoren in den Variablen erklärbaren Varianz.
- *Im Rahmen der Lehrveranstaltung Testkonstruktion werden wir im praktischen Teil nur die Hauptkomponentenanalyse verwenden!*

Faktorenextraktion: Hauptkomponentenanalyse

- Bei der **Hauptkomponentenanalyse** wird für alle Variablen von Kommunalitäten gleich Eins ausgegangen und es werden so viele Faktoren extrahiert, wie Variablen untersucht werden.
- Es wird also angenommen, dass alle Varianz in den beobachteten Variablen durch gemeinsame Faktoren erklärt werden kann.
- Technisch bedeutet dies, dass bei der Faktorenextraktion die Korrelationsmatrix der untersuchten Variablen mit Einsen in der Hauptdiagonalen verwendet wird.

Faktorenextraktion: Hauptachsenanalyse

- Bei der **Hauptachsenanalyse** wird davon ausgegangen, dass nicht alle Varianz der beobachteten Variablen durch zugrundeliegende gemeinsame Faktoren erklärt werden kann.
- Die Kommunalitäten der Variablen sind also kleiner eins und müssen vielmehr für jede Variable geschätzt werden.
- Gebräuchlich hierfür ist z.B. die Reliabilität der einzelnen Variablen oder die quadrierte multiple Korrelation jeder Variable mit allen anderen Variablen.

Anzahl der Faktoren

- Die erste Frage, die mittels der exploratorischen Faktorenanalyse beantwortet werden soll, ist wie viele Faktoren den beobachteten Variablen zugrunde liegen.
- Um dies zu entscheiden, werden die Eigenwerte aller Faktoren herangezogen, nachdem so viele Faktoren extrahiert wurden wie Variablen untersucht wurden.

Bestimmung der Anzahl der Faktoren

- Es gibt mehrere Methoden zur Bestimmung der Faktorenanzahl, daher auch selten ein völlig eindeutiges Ergebnis.
- Die gebräuchlichsten Methoden zur Bestimmung der Anzahl relevanter Faktoren sind
  - das Kaiser- oder „Eigenwerte größer eins“-Kriterium,
  - der Scree-Test und
  - die Parallelanalyse.

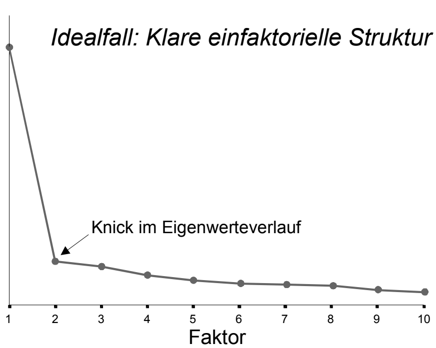
Bestimmung der Anzahl der Faktoren: Kaiser-Kriterium

- Beim Kaiser- oder „Eigenwerte größer eins"-Kriterium werden alle Faktoren als relevant betrachtet, die einen Eigenwert größer eins haben.
- Die Logik hierbei ist, dass eins der Varianz einer einzelnen Variablen entspricht.
- Das Kaiser-Kriterium führt i.d.R. zu einer zu großen Faktorenzahl und damit zu kaum interpretierbaren Ladungsmustern!
- Das Kaiser-Kriterium ist bei vielen Statistik-Programmen (z.B. SPSS) die Standardeinstellung → Die mit dieser Einstellung resultierenden Ergebnisse sollten nicht unreflektiert interpretiert werden!

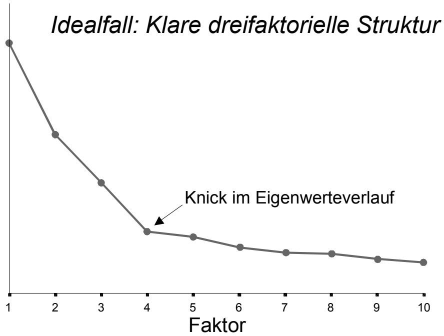
Bestimmung der Anzahl der Faktoren: Scree-Test

- Für den Scree-Test wird der Eigenwerteverlauf grafisch dargestellt („Scree-Plot").
- Anhand dieses Verlaufs wird nach dem Punkt gesucht, ab dem die Faktoren nur noch unbedeutende Restvarianz erklären.
- Dieser Punkt stellt im Idealfall einen deutlichen Knick im Eigenwerteverlauf dar.
- Alle Faktoren vor diesem Knick werden als bedeutsam betrachtet.
- Der Scree-Test ist in den meisten Fällen zuverlässig, stellt jedoch ein subjektives Verfahren dar, welches nicht immer eine eindeutige Entscheidung ermöglicht.

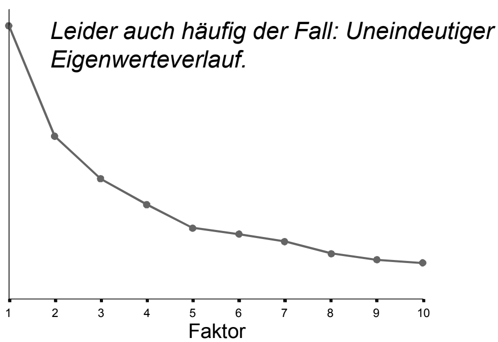
Bestimmung der Anzahl der Faktoren: Scree-Test



Bestimmung der Anzahl der Faktoren: Scree-Test



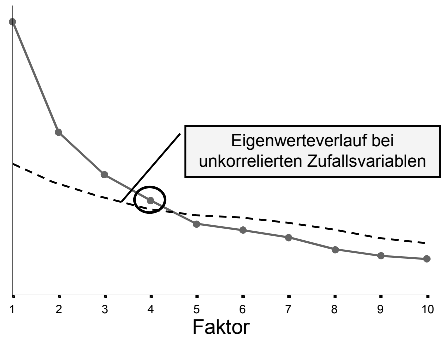
Bestimmung der Anzahl der Faktoren: Scree-Test



Bestimmung der Anzahl der Faktoren: Parallelanalyse

- Die Parallelanalyse trägt ausdrücklich dem Umstand Rechnung, dass einige Eigenwerte aus Stichprobendaten auch bei in der Population unkorrelierten Variablen größer Eins werden.
- Für die Parallelanalyse werden zunächst größere Mengen Zufallsdatensätze mit unkorrelierten Variablen generiert, wobei Variablenanzahl und Stichprobenumfang dem untersuchten empirischen Datensatz entsprechen.
- Mit diesen Zufallsdatensätzen werden Faktorenanalysen berechnet und jeweils die Eigenwerteverläufe bestimmt.
- Die Eigenwerte der Faktoren aus den Zufallsdatensätzen werden gemittelt.

Bestimmung der Anzahl der Faktoren: Parallelanalyse



Bestimmung der Anzahl der Faktoren: Parallelanalyse

- Als relevant werden alle Faktoren betrachtet, deren empirischer Eigenwert größer ausfällt als die mittels Parallelanalyse festgestellten Eigenwerte aus unkorrelierten Zufallsvariablen.
- Die Parallelanalyse stellt das zuverlässigste Verfahren zur Bestimmung der Faktorenanzahl im Rahmen der exploratorischen Faktorenanalyse dar.
- Die Parallelanalyse ist in Standard-Statistikprogrammen nicht implementiert, ihre Durchführung ist daher mit einem nicht unerheblichen Aufwand verbunden.

Faktorenrotation

- Die Faktorenextraktion werden zunächst nach dem Kriterium extrahiert, sukzessive so viel Varianz wie möglich zu erklären.
- Die Ladungsmuster, die aus der Faktorenextraktion resultieren, sind i.d.R. inhaltlich schwer zu erklären:
- Auf dem ersten Faktor finden sich über alle Variablen hinweg die höchsten Ladungen, auf dem zweiten die zweithöchsten usf.

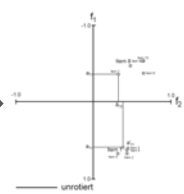
Faktorenrotation

- Um die Faktoren inhaltlich zu interpretieren, wird eine Faktorenrotation vorgenommen.
- Ziel ist es, dass auf jedem Faktor einige Variablen hoch und die übrigen Variablen möglichst niedrig laden.
- Umgekehrt sollte jede Variable nur auf einem einzelnen Faktor hoch laden, auf den übrigen niedrig.
- Ein solches Ladungsmuster bezeichnet man als Einfachstruktur.

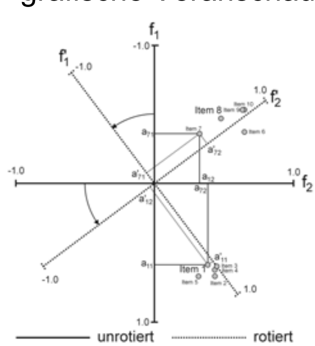
Faktorenrotation

- Grafisch kann man sich die Faktorenrotation veranschaulichen, indem man die Faktorladungen a_{mj} in der Ladungsmatrix A als Koordinaten der m Variablen im j-dimensionalen Faktorenraum betrachtet.

Variablen	Ladungen auf Faktor 1	Ladungen auf Faktor 2
x_1	a_{11}	a_{12}
x_2	a_{21}	a_{22}
x_3	a_{31}	a_{32}
x_4	a_{41}	a_{42}
Eigenwerte	$\sum_{m=1}^{k} a_{m1}^2 = \lambda_1$	$\sum_{m=1}^{k} a_{m2}^2 = \lambda_2$



Faktorenrotation: grafische Veranschaulichung



Faktorenrotation: Beispiel rotiertes und unrotiertes Ladungsmuster

	unrotiert		rotiert	
Variable	**Faktor 1**	**Faktor 2**	**Faktor 1'**	**Faktor 2'**
item 1	.590	.381	**.701**	-.044
item 2	.671	.430	**.795**	-.052
item 3	.597	.452	**.749**	.010
item 4	.622	.431	**.757**	-.022
item 5	.674	.314	**.729**	-.147
item 6	-.371	.644	.084	**.738**
item 7	-.358	.323	-.097	**.473**
item 8	-.471	.477	-.096	**.663**
item 9	-.530	.642	-.046	**.831**
item 10	-.533	.646	-.046	**.836**

Orthogonale vs. Oblique Faktorenrotation

- Die Faktoren werden so extrahiert, dass sie untereinander unkorreliert (orthogonal, d.h. „rechwinklig“ zueinander) sind.
- Bei der Faktorenrotation kann diese Unkorreliertheit aufgegeben werden, man spricht dann von einer obliquen („schiefwinkligen“) Rotation.
- Mit einer obliquen Rotation u.U. leichter eine Einfachstruktur des Ladungsmusters erreicht werden.

Orthogonale vs. Oblique Faktorenrotation

- Eine orthogonale Rotation ist dann angemessen, wenn das Ziel der Faktorenanalyse primär eine reine Datenreduktion ist.
- Wenn die Faktoren im Sinne latenter Variablen inhaltlich interpretiert werden sollen, ist eine Null-korrelation aller Dimensionen untereinander eine sehr strenge Annahme.
- In diesem Fall kann eine oblique Rotation inhaltlich angemessener sein.
- Die Ergebnisse einer obliquen Rotation sind etwas schwieriger zu interpretieren, z.B. summieren sich die quadrierten Ladungen nicht mehr zu den Eigenwerten der Faktoren und den Kommunalitäten der Variablen.

Gebräuchliche *orthogonale* Rotationstechniken

- Varimax: Maximiert die Varianz der quadrierten Faktorladungen innerhalb der Faktoren. Varimax ist der am häufigsten verwendete orthogonale Rotationsalgorithmus.
- Quartimax: Maximiert die Summe der vierten Potenz aller Faktorladungen. Es wird angestrebt, einen Faktor mit vielen hohen und mittleren Ladungen sowie verbleibende Faktoren mit wenigen hohen und sonst niedrigen Ladungen zu erhalten. Dieses Verfahren kann angezeigt sein, wenn ein gemeinsamer Generalfaktor erwartet wird.

Gebräuchliche *orthogonale* Rotationstechniken

- Promax: Dieser Algorithmus startet mit einer orthogonalen Rotation und transformiert die Faktoren anschließend in eine oblique Lösung mit dem Ziel, die absoluten Werte der Primärladungen zu maximieren und die Sekundärladungen gegen Null gehen zu lassen
- Direct Oblimin: Ein Rotationsalgorithmus, der die simultane Optimierung eines orthogonalen und eines obliquen Rotationskriteriums anstrebt.
- Bei beiden obliquen Rotationsverfahren kann die Höhe der Faktorinterkorrelationen vom Benutzer beeinflusst werden.

Praktisches Vorgehen bei der exploratorischen Faktorenanalyse

- Im konkreten empirischen Fall sieht das Vorgehen bei einer exploratorischen Faktorenanalyse wie folgt aus:
  - Faktorenextraktion (z.B. mittels Hauptkomponentenanalyse) und Ermittlung der Eigenwerte.
  - Entscheidung über die Anzahl relevanter Faktoren (z.B. mittels Scree-Test).
  - Entscheidung für einen Rotationsalgorithmus und Rotation der als relevant Betrachteten Faktoren (d.h. erneute Berechnung der Faktorenanalyse *mit Vorgabe der gewünschten Faktorenzahl*).
  - Interpretation des Ladungsmusters.