

4. Skalenbildung: Faktorenanalyse, Aggregation von Items & Normierung

Über die interne Struktur eines Tests bestehen präzise theoretische Vorstellungen

--> ja: Konfirmatorische Faktorenanalyse CFA

--> nein: Explorative Faktorenanalyse EFA

Die Information aus der Faktorenanalyse wird evtl. Genutzt um Modifikationen des Testinhalts durchzuführen.

Die verbliebenen Items müssen zu Skalen aggregiert werden und die entsprechenden Skalenrohwert an Normgruppen relativiert (ggf.. Transformiert) werden.

4.1 Explorative Faktorenanalyse

4.1.1 Grundlagen der Faktorenanalyse

Ursprung	Theoretische Intelligenzforschung
Ableitung	Aus allgemeinem linearen Modell; mit anderen multivariaten Verfahren rechnerisch verwandt
Anwendungsgebiet	- Analyse der internen Struktur eines Tests & Skalenbildung - Analyse der Kovariationsmuster von Variablen - Gruppierung von Variablen zu Faktoren
Ziel	Datenstrukturierung & Zusammenfassung vieler manifester V. zu wenigen latenten V

Verallgemeinerung der EFA nach Catell:

Unterscheidung von drei Varianzquellen (Variablen, Vpn, Situationen) und 6 mgl. Arten von Korrelationsmatrizen.

--> R-Technik

Zentrale Zwecke der EFA nach Thompson:

Validitätsprüfung	Überprüfung der Gültigkeit theoretischer Modellannahmen über die Struktur eines Variablenatzes/ Messinstrumentes
Theorieentwicklung	Über die interne Struktur psychologischer Konstrukte/ deren messbaren Manifestationen
Datenreduktion	Vereinfachende Beschreibung des Datensatzes durch Zusammenfassung von Variablen zu übergeordneten Faktoren

* EFA setzt im Vergleich zur CFA keine theoretischen Modellvorstellungen voraus.

* EFA eignet sich dazu Modellvorstellungen im Rahmen eines theoriebildendes Verfahrens zu entdecken

* CFA eignet sich zur Modellprüfung besser als EFA

* CFA eignet sich um vorliegenden Datensatz zu vereinfachen

Grundgedanke der Faktorenanalyse:

Die Ausprägung eines Individuums hinsichtlich der beobachteten Variablen wird ursächlich durch dahinter stehende, latente Konstrukte beeinflusst.

Latente Variablen	Faktoren (streng zu unterscheiden von UV in Varianzanalyse)
Mittelpunkt der FA	Simultaner Einfluss jedes Faktors auf mehrere beobachtete Variablen
Faktorenanalyse	Effektindikatoren-Modell
Regressionsanalyse	Kausalindikatoren-Modell

Faktorenanalyse unterstellt Kovarianz/ Korrelation der Items: Werden Items von demselben Faktor/ latenten Konstrukt beeinflusst, sollten diese Items untereinander korreliert sein.

--> Ausprägungen auf den Items als lineare Funktion dahinter stehender Faktoren

--> Aus dem Muster unterschiedlich hoher/ geringer Korrelationen/ Kovarianzen auf manifester Ebene wird in der FA darauf geschlossen wie viele latente Konstrukte hinter welchen manifesten Variablen stehen. (ggf.. Auch wie die latenten Konstrukte inhaltlich zu interpretieren sind.

4.1.2 Varianten & Vorgehensweise bei der Explorativen Faktorenanalyse

Häufigste Verfahren nach Thompson

PAF	principal axis factor analysis : Hauptachsenanalyse --> Aufdeckung latenter Strukturen im Datensatz
PCA	principal components analysis : Hauptkomponentenanalyse --> Datenreduktion & Beschreibung

Prinzip der FA:

Faktoren-Extraktion	Untersuchung einer Kovarianz-/ Korrelationsmatrix durch Extraktion einer kleineren Anzahl latenter Variablen, welche sich jeweils aus einer gewichteten Summe der beobachteten Variablen zusammensetzen
Faktoren-Rotation	Die vorläufig neuen Faktoren werden im geometrischen Raum gedreht, dass sie die enthaltene Information möglichst gut wiedergeben & möglichst eindeutig interpretierbar sind
Zuweisung der Faktorenwerte	Den so entstandenen endgültigen Faktoren können nun Vpn Messwerte zugewiesen werden, welche ihre Ausprägung auf den latenten Merkmalen beschreiben

Schritte der FA:

1. Prüfung der Voraussetzungen	<ul style="list-style-type: none"> - Korrelation der Items - Stichprobengröße - Uni-/ multivariate Verteilungseigenschaften der Items/ manifesten Variablen
---------------------------------------	--

Korrelation	<ul style="list-style-type: none"> - Existenz latenter Faktoren äußert sich in Interkorrelation der Items --> Items sollten substantiell korrelieren - <i>Mindestvoraussetzung</i>: Beobachtete Korrelationsmatrix sollte signifikant von der vollständigen Unabhängigkeit der Variablen in der Population abweichen
Bartlett-Test auf Sphärizität	<ul style="list-style-type: none"> - In SPSS implementierter Signifikanztest; alternativen Prüfgrößen unterlegen - Sollte signifikant werden, damit Hypothese verworfen werden kann
Kaiser-Meyer-Olkin-Koeffizient	<ul style="list-style-type: none"> - KMO-Koeffizient: sollte möglichst hoch sein >0.50 - Schätzt Ausmaß der Interkorreliertheit ein
Kommunalitäten	Indiz für Eignung einzelner Items zur Faktorisierung; >0.20
MSA-Koeffizient	Measure of sample adequacy: ähnlich KMO, jedoch auf einzelne Items bezogen
Stichproben	<ul style="list-style-type: none"> - Anforderung an Stichprobengröße steigen <ul style="list-style-type: none"> - Je mehr Items in Analyse eingehen - Je weniger Items je Faktor zusammengefasst werden - Je weniger reliabel die einzelnen Items sind
Untergrenze nach MacCallum	N=60, aber nur unter sonst günstigen Voraussetzungen (Anzahl der Variablen darf durch Stichprobengröße nicht unterschritten werden)
Verteilungseigenschaft (VE)	<ul style="list-style-type: none"> - FA unterstellt <i>Intervallskalenniveau</i> sowie <i>Normalverteilung</i> - Post hoc - Überprüfung der VE durch Inspektion der deskriptiven Itemstatistiken: Finden sich innerhalb desselben Faktors viele Items mit ähnlichen VE, sich die VE aber zwischen den Faktoren über alle Items je Faktor auffällig unterscheiden, deutet das darauf hin, dass die Faktoren lediglich Methodenartefakte abbilden und nicht inhaltlich überprüft werden sollten
Transformierung	Items vor Analyse transformieren um Normalverteilung anzunähern

Parcels	Zusammenfassung von Items zu Päckchen--> Verbesserung der Reliabilität (Empfehlung für CFA)
Faktorenanalyse höherer Ordnung	(Empfehlung für EFA)
Korrelationsanalyse	- tetrachorisch bei dichotomen Items - Polychorisch bei Ratingskalen - Anstelle der Produkt-Moment-Korrelation

2. Faktorenextraktion	<ul style="list-style-type: none"> - Ausprägung einer Person auf einer beobachteten Variable setzt sich aus einer gewichteten Kombination von Ausprägung auf latenten Variablen plus Fehlerterm zusammen - Standardisierten Gewichte a je Items und Faktor werden <i>Ladungen</i> genannt - Quadrierte Ladungen sind der Anteil gemeinsamer Varianz an der Gesamtvarianz der beteiligten Variablen - Ladungsmatrix
------------------------------	---

Ladungsmatrix:

	Faktoren/Komponenten			Zeilensumme
	F1	F2	F3	h^2
Item1	a11	a12	a13	a1j
Item2	a21	a22	a23	a2j
Item3	a31	a32	a33	a3j
Spaltensumme	am1	am2	am3	Varianzaufklärung

Zeilensumme: Kommunalität h^2 = Anteil der Varianz eines Items, der durch alle extrahierten Faktoren gemeinsam aufgeklärt werden kann

Grundproblem der FA: Kommunalitäten müssen vor Durchführung der FA (bevor Anzahl der Faktoren und Ladungen bekannt ist) geschätzt werden

--> PCA: in der Hauptdiagonalen der Korrelationsmatrix werden Einsen eingetragen = Die erste Schätzung der Kommunität geht von einer vollständigen Varianzaufklärung aus

--> PAF: Schätzungen der Kommunalitäten sind meist quadrierten multiple Korrelationskoeffizient R^2

Prinzip der PCA	- Aus einem Variablensatz wird ein neuer Variablensatz extrahiert, wobei die Variablen nach ihrer Bedeutung geordnet sind indem jede Komponente ein Maximum der verbleibenden Gesamtvarianz aller beteiligten Variablen aufklärt
------------------------	--

Prinzip der PAF	<ul style="list-style-type: none"> - Suche nach einer Gerade (1. Hauptkomponente), die von allen Punkten im Raum die geringste Entfernung hat - Extraktion 2. Komponente (von 1. HK unabhängig): klärt Maximum der Varianz auf, welche von der 1HK noch nicht aufgeklärt werden konnte - Es wird solange fortgefahren, bis Varianz vollständig aufgeklärt ist - Der neue Variablensatz hat soviel Komponenten wie der ursprüngliche, ist aber nach Varianzaufklärung geordnet - rechnerisches Vorgehen entspricht zunächst PCA - Werden dabei weniger als der vollständige Satz an Komponenten extrahiert entstehen neue Schätzungen für Kommunitäten - Kommunalitäten werden in Hauptdiagonale der Korrelationsmatrix eingetragen - Neue PCA --> Ergebnis geht in neue Kommunalitätenschätzung ein - Wiederholung des Prozesses bis Schätzungen nach einem festgelegten Abbruchkriterium konvergieren
------------------------	--

3. Festlegung der Anzahl extrahierter Faktoren	<ul style="list-style-type: none"> - Abhängig vom Ziel der FA - Kein allgemeingültiges Kriterium für Abbruch der Extraktion - Evtl. Inhaltliches Kriterium kann Existenz theoretischer Vorstellungen sein
---	--

Eigenwerte	<ul style="list-style-type: none"> - Ergeben sich (wie Kommunitäten) aus quadrierten Ladungen a^2 über die Ladungen aller Items auf demselben Faktor - Varianz des Faktors gemessen in Einheiten der Varianz je Item - --> bei PCA liegt Varianz des Faktors definitionsgemäß bei 1 - Beispiel: Komponente mit Eigenwert 3 klärt bei einem Test aus 10 Items 30% der Gesamtvarianz auf
Spur der Matrix	Summe der Eigenwerte unkorrelierter Faktoren = Summe der Elemente in der Hauptdiagonalen der Korrelationsmatrix
Abbruchkriterien	
Faustregel	Kaiser-Gutmann-Kriterium: Extraktion aller Eigenwerte >1
Faktorenzahl nach Scree-Test	<ul style="list-style-type: none"> - grafische Analyse des Eigenwertverlaufs über alle Faktoren - Auffälliger Knick gilt als Hinweis, alle Faktoren vor dem Knick zu extrahieren
Bartlett-Test	Signifikanztest: Anwendung nach Extraktion auf Residualmatrix um zu prüfen, ob diese immer noch signifikant von Identitätsmatrix abweicht

Parallelanalyse nach Horn	<ul style="list-style-type: none"> - Simulation eines Datensatzes aus der Grundgesamtheit - Produktion vieler Matrizen welche den gleichen Rang (p) wie die empirische Matrix besitzen, aber auf Zufallswerten beruhen --> Komponentenextraktion & Ordnung nach Eigenwerten
MAP-Test	Minimum-Average-Partial-Test:

4. Rotation	<ul style="list-style-type: none"> - Kriterium der <i>Einfachstruktur</i> wird angestrebt: Ladungen je Item sollen sich möglichst nur auf einen Faktoren (hoch) verteilen - Achsen (Komponenten/ Faktoren) werden so gedreht, dass sie sich einigen Punkten (Items, beobachtete Variablen) annähern und sich von anderen Punkten entfernen. So können die Punkte den Achsen eindeutig zugeordnet werden
--------------------	---

Orthogonale Rotation	<ul style="list-style-type: none"> - die beiden Achsen stehen im rechten Winkel zueinander - Maximiert theoretische Sparsamkeit
Varimax	<ul style="list-style-type: none"> Minimierung Faktorkomplexität, indem Unterschiede der Ladungen innerhalb eines Faktors maximiert werden --> Erhöht Interpretierbarkeit der Faktoren
Quartimax	Ziel: Eindeutigkeit der Zuordnung der Items zu einem Faktor
Equamax	Kompromiss zwischen Varimax und Quartimax
Oblique Rotation	<ul style="list-style-type: none"> - Winkel zwischen den Achsen wird verändert, sodass eine Korrelation zwischen den Achsen zugelassen wird - Korrelierte Faktoren sind oft einfacher interpretierbar, weil sich Mehrfachladungen der Items besser verteilen
Promax	Ladungen aus einer orthogonalen Lösungen werden mit einem Exponenten potenziert, wodurch sich Ladungen verkleinern & geringe Ladungen ganz verschwinden
Direkte Oblimin-Rot.	Lässt über Einstellung des Delta-Wertes eine Variation zwischen max. Korreliertheit (0) und Orthogonalität (4) zu
Kriteriums-/ Prokrustes-Rotation	Versuch eine erwartete/ früher gefundene Faktorstruktur anhand der empirischen Daten zu reproduzieren

5. Interpretation	<ul style="list-style-type: none"> - Durch Rotation entsteht Strukturmatrix mit Strukturkoeffizienten - Zusätzlich entsteht Mustermatrix in der die partiellen standardisierten Regressionsgewichte des Items mit den rotierten Faktoren stehen --> enthält Gewichte $a =$ Faktormusterkoeffizienten - Muster der Ladungskoeffizienten ist bedeutsam
--------------------------	---

Ladungen	<ul style="list-style-type: none"> - Unbedeutend $<.30$: werden unterdrückt - Bedeutend: $>.40 - .60$
Eindeutige Zuordnung	<p>Wenn Item auf einen Faktor hoch lädt und auf alle anderen niedrig</p> <p>--> Items die das besonders gut tun, heißen Markiertvariablen</p>
Überdeterminiert heit	Ein Faktor sollte durch möglichst viele Variablen mit jeweils möglichst hohen Ladungen definiert werden

6. Faktorwerte	<ul style="list-style-type: none"> - Ergeben sich - nach Umstellung der faktoranalytischen Grundgleichung - als eine gewichtete Kombination der Items die zum jeweiligen Faktor beitragen - SPSS: Koeffizientenmatrix der Faktorwerte
-----------------------	---

Voreinstellung SPSS	Regressionsrechnung (führt zu z-standardisierten Variablen)
Möglichkeiten	<ul style="list-style-type: none"> - Vergleiche zwischen Teilstichproben - Oblique Rotation: Faktorenanalyse zweiter Ordnung

4.1.3 Beispieldurchführung der EFA in SPSS

Variablenauswahl	ANALYSIEREN --> DIMENSIONSREDUZIERUNG --> FAKTORENANALYSE
Beurteilung der Voraussetzung	<p>Button: DESKRIPTIVE STATISTIK</p> <p>--> KMO (mgl. hoch) & Bartlett-Test (mgl. Signifikant) auf Sphärizität</p> <p>--> Anti-Image: MSA-Koeffizient</p>
PCA	Button: EXTRAKTION --> „Hauptkomponenten“ voreingestellt
PAF	METHODE --> HAUPTACHSENFAKTORENANALYSE
Auswahl der Optionen	<ul style="list-style-type: none"> - Voreinstellung: rotierte Faktorenlösung - Screeplot - KG-Kriterium Eigenwerte >1
SPSS-Ausgabe	<ul style="list-style-type: none"> - <i>Kommunalitäten</i> (bei PCA generell höher): Interpretation nach Entscheidung über Faktorenanzahl - Nach extrahierten Komponenten/ Faktoren aufgeschlüsselte <i>Gesamtvarianz</i> - Screeplots - Unrotierte Komponenten-/ Faktorenmatrix - Parallelanalyse: zusätzliche Klarheit über Faktorenanzahl

Festlegung auf Faktorenanzahl	<ul style="list-style-type: none"> - <i>KG-Kriterium</i>: Extraktion aller Faktoren >1 - <i>Screeplot</i>: Extraktion aller Faktoren vor dem Knick - <i>Parallelanalyse</i>: Extraktion alle Faktoren deren empirischer Eigenwert den Mittelwert als auch das 95. Perzentil überschreitet <p>FAKTORENANALYSE: EXTRAKTION --> Eigenwerte auf FESTE ANZAHL VON FAKTOREN</p>
Rotation	<p>FAKTORENANALYSE: ROTATION</p> <p>--> OPTIONEN: Koeffizienten $<.25$ unterdrücken um nur einigermaßen bedeutsame Ladungen und Nebenladungen zu erhalten</p> <p>--> Auswahl der gewünschten Rotation</p>
FA 2. Ordnung	<ul style="list-style-type: none"> - Berechnung Faktorwerte: FAKTORENANALYSE --> WERTE

4.1.4 Bewertung der Explorativen Faktorenanalyse

VORTEILE	NACHTEILE
<ul style="list-style-type: none"> - große Popularität durch einfache Anwendbarkeit - Grundlage zahlreicher theoretischer Modelle - Vereinfachende Beschreibung komplexer Datensätze 	<ul style="list-style-type: none"> - es müssen viele nicht-objektive Entscheidungen getroffen werden - Um stabile Ergebnisse zu erzeugen bedarf es <ul style="list-style-type: none"> - Großer Stichproben - Relativ vieler Indikatoren pro Faktor - Einzelindikatoren mit hoher Messqualität - Substantielle Interkorrelationen - Angemessene Verteilungseigenschaften - Zur Theorieprüfung wenig geeignet

--> Beschränkung auf Zwecke, für die sich das Verfahren eignet!