

Inhaltsverzeichnis




1. GRUNDLAGEN VON SPSS®	1-5
1.1 DATEITYPEN	1-5
1.2 MENÜSTRUKTUR UND BEDIENUNG	1-5
1.3 GRUNDLEGENDE EINSTELLUNGEN	1-7
2. DATENDATEIEN	2-8
2.1 DATENSTRUKTUR	2-8
2.2 ERSTELLEN EINER DATENMASKE	2-8
2.2.1 <i>Name</i>	2-9
2.2.2 <i>Typ</i>	2-9
2.2.3 <i>Variablenlabels und Wertelabels</i>	2-10
2.2.4 <i>Fehlende Werte</i>	2-10
2.2.5 <i>Spalten und Ausrichtung</i>	2-11
2.2.6 <i>Messniveau</i>	2-11
2.2.7 <i>Rolle</i>	2-11
2.2.8 <i>Kopieren von Variablendefinitionen</i>	2-11
2.2.9 <i>Übersicht aller Variablenattribute</i>	2-11
2.2.10 <i>Variableneigenschaften definieren</i>	2-12
2.2.11 <i>Dateneigenschaften kopieren</i>	2-13
2.3 DATENEINGABE	2-13
2.3.1 <i>Eingabefunktionen</i>	2-14
2.3.2 <i>Editierfunktionen</i>	2-15
2.3.3 <i>Kopieren von Werten</i>	2-15
2.3.4 <i>Kopieren von Variablen und Fällen</i>	2-15
2.3.5 <i>Daten suchen und ersetzen</i>	2-16
2.3.6 <i>Sprung zu Fallnummer oder Variable</i>	2-16
2.4 DATENIMPORT UND -EXPORT	2-16
3. AUSGABEDATEIEN	3-18
3.1 CHARAKTER VON AUSGABEOBJEKTEN	3-18
3.2 ARBEITEN MIT OBJEKTEN IN AUSGABEDATEIEN	3-18
3.2.1 <i>Ausgabedateien öffnen und speichern</i>	3-18
3.2.2 <i>Arbeiten mit Ausgabeobjekten</i>	3-18
3.2.3 <i>Änderung der Darstellung</i>	3-19
3.2.4 <i>Ausgabe exportieren</i>	3-19
4. SYNTAX- UND MAKRODATEIEN	4-21
4.1 CHARAKTER VON SYNTAXDATEIEN	4-21
4.2 ARBEITEN MIT SYNTAXDATEIEN	4-21
4.2.1 <i>Syntaxdateien öffnen und speichern</i>	4-21
4.2.2 <i>Editieren einer Syntaxdatei</i>	4-22
4.2.3 <i>Grammatik von Syntaxdateien</i>	4-22
4.3 MAKRODATEIEN	4-23

5.	DATEN AUSWÄHLEN & TRANSFORMIEREN	5-25
5.1	SELEKTION VON FÄLLEN	5-25
5.1.1	<i>Fälle auswählen</i>	5-25
5.1.2	<i>Falls Bedingung formulieren</i>	5-25
5.1.3	<i>Nach Fällen aufteilen</i>	5-26
5.2	ANWENDUNG VON FUNKTIONEN AUF VARIABLEN	5-27
5.2.1	<i>Variable berechnen</i>	5-27
5.2.2	<i>Funktionsübersicht</i>	5-28
5.3	VARIABLENUMKODIERUNG	5-30
5.3.1	<i>Umkodieren in dieselben Variablen</i>	5-30
5.3.2	<i>Umkodieren in eine andere Variable</i>	5-31
5.3.3	<i>Automatisches Umkodieren von Variablen</i>	5-32
5.4	ZUSATZFUNKTIONEN BEI TRANSFORMATIONEN	5-33
5.4.1	<i>Zufallszahlengenerator</i>	5-33
5.4.2	<i>Häufigkeiten von Werten in Fällen auszählen</i>	5-33
5.4.3	<i>Transformation von Werten in Ränge</i>	5-34
5.5	ZUSATZFUNKTIONEN BEI DATENKODIERUNGEN	5-36
5.5.1	<i>Sortieren von Fällen oder Variablen</i>	5-36
5.5.2	<i>Transponieren von Daten</i>	5-37
5.5.3	<i>Verknüpfen von Datendateien</i>	5-37
5.5.4	<i>Aggregieren von Daten</i>	5-38
5.5.5	<i>Gewichten von Fällen</i>	5-41
5.5.6	<i>Umstrukturieren von Daten</i>	5-42
5.5.7	<i>Ermitteln doppelter Fälle</i>	5-44
5.5.8	<i>Visuelles Klassieren (Variablen kategorisieren)</i>	5-45
5.5.9	<i>Fehlende Werte ersetzen</i>	5-48
5.5.10	<i>Multiple Imputation</i>	5-49
5.5.11	<i>Assistent für Datum und Uhrzeit</i>	5-55
5.6	ZUSATZFUNKTIONEN ZUR AUFBEREITUNG UND VALIDIERUNG VON DATEN	5-56
5.6.1	<i>Optimales Klassieren</i>	5-56
5.6.2	<i>Ungewöhnliche Fälle identifizieren</i>	5-58
5.6.3	<i>Daten validieren</i>	5-61
5.6.4	<i>Daten hinsichtlich fehlender Werte analysieren</i>	5-64
6.	UNIVARIATE VERFAHREN	6-69
6.1	HÄUFIGKEITEN	6-69
6.2	UNIVARIATE STATISTIKEN	6-71
6.3	EXPLORATIVE DATENANALYSE	6-72
6.4	KREUZTABELLEN	6-76
6.5	VERGLEICH VON MITTELWERTEN	6-81
6.5.1	<i>Errechnen von Mittelwerten</i>	6-81
6.5.2	<i>t-Tests</i>	6-83
6.6	EINFAKTORIELLE VARIANZANALYSE (ANOVA)	6-86
6.7	KORRELATIONSKOEFFIZIENTEN	6-90
6.7.1	<i>Bivariate Korrelation</i>	6-90
6.7.2	<i>Partialkorrelation</i>	6-92
6.8	LINEARE REGRESSION	6-94
6.9	NICHPARAMETRISCHE TESTS	6-103
6.9.1	<i>Chi-Quadrat-Test et al.</i>	6-103
6.9.2	<i>Binomialtest</i>	6-104

6.9.3	Sequenzanalyse	6-105
6.9.4	Ein-Stichproben-Kolmogorov-Smirnov-Test.....	6-106
6.9.5	Tests für zwei unabhängige Stichproben.....	6-107
6.9.6	Tests für mehrere unabhängige Stichproben.....	6-108
6.9.7	Tests bei zwei abhängigen Stichproben.....	6-109
6.9.8	Tests für mehrere abhängige Stichproben.....	6-111
6.10	ANALYSE VON MEHRFACHANTWORTEN	6-112
6.10.1	Definieren von Mehrfachantworten.....	6-112
6.10.2	Analyse von Mehrfachantworten als Häufigkeiten.....	6-113
6.10.3	Analyse von Mehrfachantworten als Kreuztabellen.....	6-113
6.11	ANALYSE FEHLENDER WERTE	6-114
6.12	BOOTSTRAPPING	6-117
7.	MULTIVARIATE VERFAHREN	7-119
7.1	MEHRDIMENSIONALE VARIANZANALYSEN.....	7-119
7.1.1	Multiple Varianzanalyse	7-119
7.1.2	Multivariate Varianzanalyse.....	7-132
7.1.3	Varianzanalyse mit Messwiederholungen.....	7-139
7.2	FAKTORENANALYSE.....	7-148
7.2.1	Prinzip.....	7-149
7.2.2	Ergebnisausgabe.....	7-150
7.2.3	Durchführung.....	7-154
7.3	CLUSTERANALYSE	7-158
7.3.1	Distanz- und Ähnlichkeitsmaße	7-158
7.3.2	Hierarchische Clusteranalyse für mittlere Datenumfänge.....	7-165
7.3.3	Clusterzentrenanalyse für große Datenumfänge	7-171
7.3.4	Two-Step-Clusteranalyse	7-175
7.4	DISKRIMINANZANALYSE	7-181
7.4.1	Prinzip.....	7-181
7.4.2	Ergebnisausgabe bei Einschlussverfahren	7-183
7.4.3	Ergebnisausgabe bei schrittweiser Methode.....	7-189
7.4.4	Durchführung.....	7-190
7.5	MULTIDIMENSIONALE SKALIERUNG	7-194
7.5.1	Prinzip.....	7-195
7.5.2	Ergebnisausgabe bei symmetrischer Datenmatrix.....	7-196
7.5.3	Ergebnisausgabe bei asymmetrischer Datenmatrix.....	7-198
7.5.4	Ergebnisausgabe bei Rohvariablen.....	7-198
7.5.5	Durchführung.....	7-199
7.6	ITEM- UND SKALENANALYSE	7-202
7.6.1	Prinzip.....	7-202
7.6.2	Ergebnisausgabe.....	7-203
7.6.3	Durchführung.....	7-206
7.7	BINÄR LOGISTISCHE REGRESSION.....	7-207
7.7.1	Prinzip.....	7-208
7.7.2	Ergebnisausgabe.....	7-209
7.7.3	Durchführung.....	7-212
7.8	GEMISCHTE MODELLE	7-215
7.8.1	Prinzip.....	7-216
7.8.2	Ergebnisausgabe.....	7-218
7.8.3	Durchführung.....	7-225

7.9	WEITERE VERFAHREN IM ÜBERBLICK	7-228
7.9.1	<i>Multinomiale logistische Regressionsanalyse</i>	7-228
7.9.2	<i>Regression: Gewichtsschätzung</i>	7-228
7.9.3	<i>Regression: Kurvenanpassung</i>	7-228
7.9.4	<i>Ordinale Regression</i>	7-228
7.9.5	<i>Regression: Probit-Analyse</i>	7-229
7.9.6	<i>Nichtlineare Regression</i>	7-229
7.9.7	<i>Regression: Zweistufige kleinste Quadrate</i>	7-229
7.9.8	<i>Regression mit Optimaler Skalierung</i>	7-229
7.9.9	<i>Allgemeine loglineare Modelle</i>	7-230
7.9.10	<i>Logit-loglineare Modelle</i>	7-230
7.9.11	<i>Loglinear: Modellauswahl</i>	7-230
7.9.12	<i>Allgemeines Lineares Modell: Varianzkomponenten</i>	7-231
7.9.13	<i>Korrespondenzanalyse</i>	7-231
7.9.14	<i>Optimale Skalierung</i>	7-231
7.9.15	<i>OLAP-Würfel</i>	7-232
7.9.16	<i>Verhältnisstatistik</i>	7-232
7.9.17	<i>Multidimensionale Skalierung (PROXSCAL)</i>	7-232
7.9.18	<i>Verallgemeinerte lineare Modelle (GENLIN)</i>	7-233
7.9.19	<i>Verallgemeinerte Schätzungsgleichungen (GENLIN)</i>	7-233
7.9.20	<i>Mehrschichtiges Perzeptron (MLP)</i>	7-233
7.9.21	<i>Radiale Basisfunktion (RBF)</i>	7-233
8.	ANHANG	8-234
8.1	ÜBUNGSAUFGABEN	8-234
8.2	CONCEPT-MAPS.....	8-239
8.3	LITERATUR.....	8-242
8.3.1	<i>Einführende Literatur zur uni- und multivariaten Statistik</i>	8-242
8.3.2	<i>Primär- und Sekundärliteratur zu SPSS</i>	8-242
	STICHWORTREGISTER	8-243

Im Text wird folgende Symbolik verwendet:

-  ... kündigt eine Erläuterung bzw. einen Hinweis zu einem speziellen Thema an.
-  ... kündigt ein Rechenbeispiel (Erläuterung eines verwendeten Datensatzes) oder die Interpretation eines Rechenergebnisses zum verwendeten Datensatz an.
-  ... kündigt eine einleitende Kurzcharakterisierung zu einer statistischen Prozedur und/oder zur Syntax von SPSS an, mit Angaben zur Funktion, den erforderlichen Daten und notwendigen statistischen Annahmen zum Verfahren. Ein Doppelstern (**) hinter einem Unterbefehl der Syntax kennzeichnet eine Voreinstellung, die verwendet wird, wenn keine andere Angabe erfolgt.

Geklammerte Angaben wie z.B. (DESCRIPTIVES), (/SAVE) oder (MEAN) verweisen auf zugeordnete Syntax-Befehle (Befehle, Unterbefehle und Schlüsselwörter) aus der SPSS-Kommandosprache.

varname_ ist eine Bezeichnungsvariable, die bei der **ONEWAY**-Prozedur ohne Inhalt ist. Die unabhängige Variable (im Beispiel *grp*) befindet sich zwischen diesen beiden Variablen und die verbleibende letzte Variable ist die abhängige Variable; für beide Variablen werden die Werte-Labels aus der Datendatei übernommen (für die abhängige Variable ist dies allerdings nur noch bedingt sinnvoll). Über die Anweisung zum Einlesen ist es damit möglich, aufgrund der aggregierten Angaben in dieser Datei eine Varianzanalyse zu berechnen, ohne auf die Rohdaten zugreifen zu müssen. Allerdings funktioniert diese Option nur, wenn vorab die Matrix-Datei explizit (in den Daten-Editor) geladen wurde (`GET FILE=dateiname`); im Anschluss wird auf diese Datei durch `/MATRIX=IN(*)` zugegriffen. Werden deskriptive Statistiken angefordert, lässt sich kein Minimum und Maximum ermitteln; der Unterbefehl `/MISSING` wird ignoriert.

Ebenfalls nur über die Syntax erreichbar ist der Unterbefehl `/RANGES`, der erweiterte Möglichkeiten bei einigen Post-Hoc-Tests bietet. Bis zu 10 dieser Unterbefehle sind zulässig. So lassen sich unterschiedliche Alpha-Fehler für verschiedene Tests angeben. Das Schlüsselwort `MODLSD` ist dabei äquivalent mit `BONFERRONI` aus dem `POSTHOC` Unterbefehl.

Schließlich können über den Unterbefehl `/MISSING` benutzerdefinierte fehlende Werte wie gültige Werte im Verfahren berücksichtigt werden (`INCLUDE`).

6.7 Korrelationskoeffizienten

6.7.1 Bivariate Korrelation

Kurzcharakterisierung:		Bivariate Korrelation (<i>CORRELATIONS/NONPAR CORR</i>)	
Anwendung:	Prüfung von korrelativen Beziehungen zwischen Variablen nach Pearson oder deren Ränge anhand von Koeffizienten nach Spearman und Kendall.		
Datenniveau:	Für den Korrelationskoeffizienten nach Pearson müssen symmetrische quantitative Variablen vorliegen, für Spearman und Kendall mindestens ordinalskalierte Daten.		
Annahmen:	Für den Korrelationskoeffizienten nach Pearson wird angenommen, dass jedes Variablenpaar bivariat normalverteilt ist.		
Menüpfad:	Analysieren ► Korrelation ► Bivariat...		
Syntax: Pearson	<pre>CORRELATIONS [VARIABLES=] varlist [WITH varlist] [/varlist...] [/MISSING={PAIRWISE**} [INCLUDE]] {LISTWISE } [/PRINT={TWOTAIL**} {SIG**}] {ONETAIL } {NOSIG} [/MATRIX=OUT(file)] [/STATISTICS=[DESCRIPTIVES][XPROD][ALL]] .</pre>	Beispiel:	<pre>CORRELATIONS /VARIABLES=fpie fpin fpim /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE .</pre>
Syntax: Spearman/ Kendall	<pre>NONPAR CORR [VARIABLES=] varlist [WITH varlist] [/varlist...] [/PRINT={TWOTAIL**}{SIG**}{SPEARMAN**}] {ONETAIL }{NOSIG}{KENDALL } {BOTH } [/SAMPLE] [/MISSING={PAIRWISE**} [INCLUDE]] {LISTWISE } [/MATRIX=OUT(file)] .</pre>	Beispiel:	<pre>NONPAR CORR /VARIABLES=fpie fpin fpim /PRINT=SPEARMAN TWOTAIL NOSIG /MISSING=PAIRWISE .</pre>

Im Menü *Analysieren* steht zur Prüfung linearer Zusammenhänge zweier Variablen die Option *Korrelation ► Bivariat...* zur Verfügung. In der Ausgabe werden dabei alle Kombinationsmöglichkeiten der zu vergleichenden Variablen in eine quadratische Matrix geschrieben, die sich in der Hauptdiagonale spiegelt.

Die Ausgabe unterscheidet drei Angaben pro Variablenpaar: den Korrelationskoeffizienten (*Korrelation nach ...*), den Umfang der Stichprobe (*N*) und die Irrtumswahrscheinlichkeit (*Signifikanz – 2-seitig/1-seitig*), je nach Wahl aus dem Dialog als *zweiseitige* (`TWOTAIL`) oder *einseitige* (`ONETAIL`) Wahrscheinlichkeit (*Test auf Signifikanz*).

Korrelationen

		Extraversion	Emotionale Labilität	Maskulinität
Extraversion	Korrelation nach Pearson	1	-,270(*)	,523(**)
	Signifikanz (2-seitig)		,038	,000
	N	59	59	59

Emotionale Labilität	Korrelation nach Pearson	-,270(*)	1	-,701(**)
	Signifikanz (2-seitig)	,038		,000
	N	59	59	59
Maskulinität	Korrelation nach Pearson	,523(**)	-,701(**)	1
	Signifikanz (2-seitig)	,000	,000	
	N	59	59	59

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

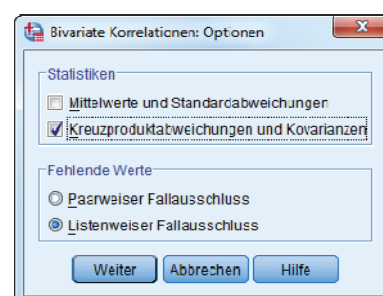
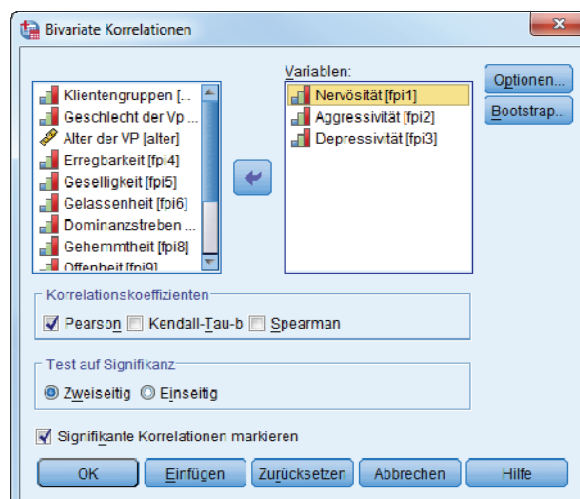
Für intervallskalierte Variablen oder die Berechnung einer punktbiseralen Korrelation (zwischen einer dichotomen und einer intervallskalierten Variable) kann auf die Korrelation nach *Pearson* zurückgegriffen werden; diese ist voreingestellt. Im Falle von ordinalskalierten Daten stehen alternativ zwei Rangkorrelationskoeffizienten zur Verfügung. Beide Koeffizienten, *Kendalls Tau-b* (KENDALL) und *Spearman's Rho* (SPEARMAN), liefern dabei Korrelationen anhand der in Ränge transformierten Variablen und bieten sich dann an, wenn ordinale und/oder nicht normalverteilte Daten vorliegen. Beide Maße basieren allerdings auf unterschiedlichen mathematischen Kalkülen und können daher in bestimmten Fällen

voneinander abweichen. Spearman ist die direkte Entsprechung zur Produkt-Moment-Korrelation für Rangdaten. Da in die Berechnung die Differenzen der Daten – die Ränge werden hier wie Messdaten aufgefasst – eingehen, müssen die aufeinander folgenden Ränge äquidistante Positionen (gleichabständige Ränge) auf dem untersuchten Merkmal abbilden. Fehlt die Voraussetzung, überschätzt der Koeffizient den wahren Zusammenhang; ist diese Voraussetzung gegeben, besitzt Spearman die größere Teststärke. Kendalls Tau⁸ basiert hingegen nicht auf dem Konzept der Produkt-Moment-Korrelation. Bei Tau geht lediglich die Anzahl der Größer- (Proversionen) und Kleinerrelationen (Inversionen) – also reine ordinale Information – in die Berechnung ein; Tau stellt geringere Anforderungen an die Daten und besitzt eine geringere Teststärke.

Die letzte Option *Signifikante Korrelationen markieren* entscheidet, ob das Signifikanzniveau zusätzlich als Symbol (* für $p < 0,05$ und ** für $p < 0,01$) ausgegeben werden soll.

Über die Schaltfläche [Optionen...](#) lassen sich zusätzlich *Mittelwerte und Standardabweichungen* (DESCRIPTIVES) mit den zugehörigen Stichprobenumfängen und die *Kreuzproduktabweichungen und Kovarianzen* (XPROD) anfordern. Das Kreuzprodukt der Abweichungen ist gleich der Summe der Produkte mittelwertkorrigierter Variablen; dies ist der Zähler des Korrelationskoeffizienten nach Pearson. Die Kovarianz ist ein nicht standardisiertes Maß für den Zusammenhang zwischen zwei Variablen und ist gleich der Kreuzproduktabweichung dividiert durch $N - 1$.

Die Funktion der neuen (seit Version 18) Schaltfläche [Bootstrap...](#) (BOOTSTRAP) wird im Kapitel 6.12 gesondert behandelt.



⁸ Begrifflich wird unterschieden zwischen der Variante *Tau-b*, die Rangbindungen im Nenner der Division berücksichtigt und *Tau-c*, die im Nenner den Stichprobenumfang und das Minimum der Zeilen- und Spaltenzahl berücksichtigt, aber Rangbindungen ignoriert; liegen keine Rangbindungen vor, führen beide Koeffizienten zu identischen Ergebnissen.

Beispiel:

```
GLM M1 M2 BY A B
/WSFACTOR ZEIT (2)
/MMATRIX = M1 1 M2 1; M1 1 M2 -1
/DESIGN A B .
```

Die Angabe von /WSFACTOR spezifiziert ein Messwiederholungsmodell mit zwei Zwischensubjekt-Faktoren A und B und dem Innersubjekt-Faktor ZEIT. Die **M**-Matrix ist:

```
1 1
1 -1
```

Da keine **L**- und **K**-Matrix spezifiziert wurde, gilt: **L** ist eine Matrix, die der schätzbaren Funktion des *Intercepts* des Zwischensubjekt-Modells entspricht und **K** ist ein zweizeiliger Vektor (0, 0).

7.2 Faktorenanalyse

Kurzcharakterisierung:		Faktorenanalyse (FACTOR)
Anwendung:	Bestimmung zugrunde liegender Faktoren, welche die Korrelationsmuster innerhalb eines Satzes beobachteter Variablen erklären. Anwendung erfolgt im Zuge einer Datenreduktion (zur Identifikation weniger Faktoren, welche den größten Teil der in den Variablen aufgetretenen Varianz erklären), zur Generierung von Hypothesen über kausale Mechanismen oder zum Sichten von Variablen für sich anschließende Analysen. Eine zusätzliche Ausgabe von univariaten Statistiken ist möglich. Faktorenanalyse wird explorativ und nicht konfirmativ durchgeführt.	
Datenniveau:	Variablen müssen quantitativ auf dem Intervall- oder Verhältnisskalenniveau sein. Kategoriale Daten sind nicht geeignet.	
Annahmen:	Daten müssen für jedes Variablenpaar einer bivariaten Normalverteilung entsprechen: Beobachtungen müssen unabhängig sein. Faktoren dürfen weder miteinander noch mit den gemeinsamen Faktoren korrelative Beziehungen aufweisen (keine Multikollinearität); zudem muss gelten: $n \geq \text{Var} + 2$	
Menüpfad:	Analysieren ► Dimensionsreduktion ► Faktorenanalyse	
Syntax:	<pre>FACTOR VARIABLES=varlist /[MISSING={LISTWISE**}] [INCLUDE] {PAIRWISE } {MEANSUB } {DEFAULT**} /[MATRIX=([IN({COR=file})] [OUT({COR=file}))) {COR=* } {COR=* } {COV=file} {COV=file} {COV=* } {COV=* } {FAC=file} {FAC=file} {FAC=* } {FAC=* } {FSC=file} {FSC=* } /[METHOD = {CORRELATION**}] {COVARIANCE } /[SELECT = varname(value)] /[ANALYSIS=varlist...] /[PRINT={DEFAULT**}] [INITIAL**] [EXTRACTION**] [ROTATION**] [UNIVARIATE] [CORRELATION] [COVARIANCE] [DET] [INV] [REPR] [AIC] [KMO] [FSCORE] [SIG] [ALL]] /[PLOT={EIGEN} [ROTATION [(n1,n2)]]] /[DIAGONAL={value list } {DEFAULT**} /[FORMAT={SORT} [BLANK(n)] [DEFAULT**]] /[CRITERIA={FACTORS(n)} [MINEIGEN({1.0**})] [ITERATE({25**})] {n } {n } [RCONVERGE({0.0001**})] [{KAISER**}] {n } {NOKAISER} [ECONVERGE({0.001**})] [DEFAULT**]] {n } /[EXTRACTION={PC** }]/[ROTATION={VARIMAX** }]</pre>	<pre>Beispiel: FACTOR /VARIABLES cmmt moe petwetw petgttw frkonz petzfgtw petoftw petsfgtw fewtw manntzw prozum /MISSING LISTWISE /ANALYSIS cmmt moe petwetw petgttw frkonz petzfgtw petoftw petsfgtw fewtw manntzw prozum /PRINT INITIAL EXTRACTION ROTATION /FORMAT SORT BLANK(.10) /CRITERIA MINEIGEN(1) ITERATE(25) /EXTRACTION PC /CRITERIA ITERATE(25) /ROTATION VARIMAX /SAVE REG(ALL) /METHOD= CORRELATION .</pre>

	{PA1** }	{EQUAMAX }	
	{PAF }	{QUARTIMAX }	
	{ALPHA }	{OBLIMIN({0}) }	
	{IMAGE }	{n }	
	{ULS }	{PROMAX({4}) }	
	{GLS }	{n }	
	{ML }	{NOROTATE }	
	{DEFAULT** }	{DEFAULT** }	
	[/SAVE={REG } ({ALL}[rootname]])]		
	{BART }	{n }	
	{AR }	{DEFAULT}.	

7.2.1 Prinzip

Die empirische Forschung versucht ihre hypothetischen Konstrukte vielfach durch einen homogenen Satz von Variablen zu erfassen. Die Faktorenanalyse bietet dazu ein Verfahren an, eine Vielzahl von Variablen, von denen nicht bekannt ist, ob und in welcher Weise sie etwas miteinander gemeinsam haben in Gruppen aufzuteilen. Hinter dieser Gruppierung wird dann vielfach (wenn sie eindeutig und interpretierbar ist) eine komplexe übergeordnete Variable vermutet, die ein spezifisches Konstrukt repräsentiert; diese übergeordneten Variablen werden im Rahmen dieser Analyse entsprechend *Faktoren* oder *Komponenten* genannt. Ziel der Faktorenanalyse ist es, den Grad an Komplexität, der durch eine Vielzahl von Variablen dargestellt wird, auf eine möglichst geringe und eindeutige Zahl von Faktoren zu reduzieren; entsprechend zählt die Faktorenanalyse zu den *dimensionsreduzierenden Verfahren*²³. Grundsätzlich sind bei Einsatz der Faktorenanalyse zwei Ausgangssituationen denkbar:

- Entweder sind zu Beginn weder die Art noch die Anzahl der Faktoren bekannt und sollen durch Anwendung der Faktorenanalyse bestimmt werden (in diesem Falle ist eine nachträgliche Interpretation und Benennung der Faktoren zu leisten),
- oder aber in der Konstruktion des Untersuchungsverfahrens (z.B. für bestimmte Gruppen von Items oder Fragen in einem psychodiagnostischen Test oder Fragebogen) steckt eine implizite Theorie über eingebettete hypothetische Konstrukte, die mittels der Faktorenanalyse überprüft bzw. bestätigt werden sollen.

Das Modell der Faktorenanalyse geht nun davon aus, dass sich jede Stichprobenvariable durch bestimmte Faktoren erklären lässt (ähnlich wie bei der multiplen Regression). Entsprechend muss eine Gleichung formuliert werden, die den Erklärungszusammenhang beschreibt. Dabei werden die Faktoren wie Prädiktoren (UV) betrachtet, die das Kriterium (AV) erklären; diese erklärenden Faktoren werden als *gemeinsame Faktoren* (*Common Factors*) bezeichnet. Um diesen Zusammenhang zu erklären, werden zusätzlich für jeden Faktor Regressionskoeffizienten ermittelt. Schließlich kommt ein weiterer Faktor hinzu, der dem Fehler (Residuum) der Regressionsschätzung entspricht; dieser nicht erklärte Rest wird als *Einzelrestfaktor* (*Unique Factor*) bezeichnet. Jede Variable in einem faktoranalytischen Modell besitzt einen solchen Einzelrestfaktor. Das Modell berücksichtigt dabei, dass einzelne Faktoren für bestimmte Variablen ggf. eine nur sehr geringe Bedeutung besitzen; dies würde einem kleinen Wert des Faktorkoeffizienten (*Faktorladung*) entsprechen.

Die Vorgehensweise ist nun die folgende: Zunächst werden lineare Kombinationen der Beobachtungsvariablen gebildet. Für Variablen, die miteinander hohe Korrelationen aufweisen, wird unterstellt, dass sie einen gemeinsamen Faktor besitzen; umgekehrt gilt, dass niedrige Korrelationen gegen

²³ Im Zusammenhang mit der Faktorenanalyse wird häufig auch die *kanonische Korrelation* diskutiert; damit lassen sich Assoziationen zwischen Gruppen von Variablen finden und quantifizieren. Dabei wird die Korrelation einer Linearkombination einer Menge von Variablen mit einer Linearkombination einer anderen Menge von Variablen verglichen; Anwendung ist meist die Vorhersage der Werte der einen Menge durch die Werte der anderen Menge. Eine kanonische Korrelationsanalyse kann in SPSS auf zweierlei Weise erfolgen: Entweder durch die Makro-Datei CANONICAL CORRELATION.SPS (gespeichert im SPSS-Ordner) oder mit der Option /DISCRIM=COR unter der Prozedur MANOVA, die sich nur (noch) über die SPSS-Syntax ausführen lässt. Weitere Hinweise zur Prozedur finden sich im SPSS *Command Syntax Reference Guide*.

einen gemeinsamen Faktor sprechen. Für die Ermittlung der Faktoren wird nun ein Schätzverfahren eingesetzt, das Koeffizienten c_i liefert, die den Zusammenhang mit dem Faktor oder den Faktoren liefert ($Factor^1 = c_1 \cdot var_1 + c_2 \cdot var_2 + \dots + c_i \cdot var_i$). Formell ist es dabei möglich, dass alle Variablen der Stichprobe durch einen Faktor erklärt werden; das Ziel besteht aber in der Regel darin, mehrere Faktoren zu finden, die nur einen spezifischen Teil der Variablen erklären. Auf der anderen Seite liegt aber der Erfolg einer Faktorenanalyse auch darin begründet, eine Vielzahl relevanter Variablen durch nur wenige Faktoren zu repräsentieren. Logischerweise liegt kein Informationsgewinn in einer Analyse, die n Variablen durch n Faktoren erklärt. Entscheidend ist aber (und dies ist gleichzeitig die schwierigste Aufgabe), dass die durch die Faktorenanalyse ermittelten Faktoren auch *sinnvoll* und *theoriegestützt* interpretiert bzw. benannt werden können.

Die Prozedur der Faktorenanalyse lässt sich grob in vier Schritten durchführen:

1. Mit der Berechnung der *Korrelationsmatrizen* lassen sich die Zusammenhänge der Variablen prüfen; Variablen mit geringer Korrelation bilden dabei häufig einen einzelnen Faktor und sollten ggf. aus der Analyse genommen werden. Liegen insgesamt nur geringe Korrelationen vor, sollte auf eine Faktorenanalyse verzichtet werden.
2. Danach folgt die eigentliche *Faktorextraktion*, die über verschiedene Extraktionsmethoden bestimmt werden kann. Statistische Maße können dabei Auskunft geben, ob das angenommene Faktorenmodell geeignet ist, die Variablen auf einfache Weise zu interpretieren.
3. Da die Faktoren häufig nicht eindeutig zu interpretieren sind, wird mittels der *Rotation* ein neues Koordinatensystem konstruiert, in dem die Verbindungen zu den Beobachtungsvariablen eindeutiger und damit besser interpretierbar werden. Die Rotation wird dabei durch Drehung der Koordinatenachsen erreicht.
4. Neben den Faktoren können schließlich für jede Variable auch *Faktorwerte* ermittelt werden.

7.2.2 Ergebnisausgabe

Im nachfolgenden Beispiel werden 11 Variablen aus einer Untersuchung an Frühgeborenen einer Faktorenanalyse unterzogen (Datensatz: **FRUEGEB.SAV**; vgl. Brosius, 2002). Dabei geht es um den Zusammenhang zwischen frühgeburtlichen Daten und neurologischen bzw. psychosozialen Faktoren. Differenziert werden soll hier die Intelligenzentwicklung von Frühgeborenen im Alter von sechs Jahren anhand 11 verschiedener Testverfahren. Als Verfahren für die Faktorextraktion wird die Voreinstellung der *Hauptkomponentenanalyse (Principal Component Analysis)* verwendet. Dabei werden Linearkombinationen der Variablen gebildet. Als erste Hauptkomponente (Faktor) wird diejenige ausgewiesen, die den größten Teil der Gesamtstreuung aller Variablen erklärt. Formal lassen sich so viele Hauptkomponenten berechnen wie Variablen im Modell vorhanden sind.

	cmmt	moe	petwetw	petgttw	frkonz	petzftw	petoftw	petsftw	fewtw	mannztw	prozum	intubier	o_21	o_40
1	27	6	33	0	0	46	0	0	0	0	31,00	4	3	3
2	50	10	53	57	29	27	48	53	46	46	59,00	1	2	1
3	59	10	63	67	30	50	56	51	55	50	58,00	1	1	1
4	24	10	31	20	0	0	0	0	0	0	30,00	2	2	2
5	54	9	44	29	31	46	39	46	49	39	68,00	2	2	1

Unter *Kommunalitäten (Communalities)* sind in der ersten Tabelle der Ausgabe die Variablen nach der Faktorextraktion ausgegeben und mit der Maßzahl der Kommunalität versehen (*PRINT*). Sie gibt an, welchen Betrag der Streuung *einer* Variablen *alle* Faktoren zusammengenommen erklären. Da alle Variablen zuvor z-transformiert wurden, besitzen sie eine Streuung von 1, d.h., die Kommunalität kann maximal den Wert 1 (minimal den Wert 0) annehmen. Dieser Zustand tritt ein, wenn die Streuung einer Variablen vollständig durch die Faktoren erklärt wird. Alle *anfänglichen* Werte (*INITIAL*) sind hier 1, da die Maßzahlen im Ausgangspunkt der Faktorextraktion angegeben werden; nach der *Extraktion (EXTRACTION)* verändern sich die Kommunalitäten entsprechend ihres Beitrags zur Varianzaufklärung aller Faktoren.

Kommunalitäten

	Anfänglich	Extraktion
Columbia Mental Maturity Test; transformiert in t-Werte	1,000	,573
Moering-Test	1,000	,399
Psychologischer Entwicklungstest: Wörter ergänzen; transformiert in t-Werte	1,000	,719

Psychologischer Entwicklungstest: Grammatik; transformiert in t-Werte	1,000	,684
Frankfurter Konzentrationstest für 5-jährige	1,000	,581
Psychologischer Entwicklungstest: Zahlenfolgedächtnis; transformiert in t-Werte	1,000	,562
Psychologischer Entwicklungstest: Objektfinden; transformiert in t-Werte	1,000	,720
Psychologischer Entwicklungstest: Symbolfolgedächtnis; transformiert in t-Werte	1,000	,536
Frostigs Entwicklungs-, Visualisierungs- und Wahrnehmungstest; transformiert in t-Werte	1,000	,584
Mannzeichen; transformiert in t-Werte	1,000	,606
Bildererkennung in einem zu einem gewissen Prozentsatz unvollständigen Mosaik; Summe aus verschiedenen Einzeltests	1,000	,302

Extraktionsmethode: Hauptkomponentenanalyse.

In der nachfolgenden Tabelle *Erklärte Gesamtvarianz* ist zu erkennen, dass im Anfangszustand nach der Methode der Hauptkomponentenanalyse genauso viele Faktoren extrahiert werden, wie Variablen vorhanden sind. Daher ist es auch nicht verwunderlich, dass die gesamte Streuung jeder Variablen restlos aufgeklärt wird. Weiter erfolgt in der Tabelle eine Angabe der *anfänglichen Eigenwerte (Eigenvalues)*, die angibt, welcher Betrag der Gesamtstreuung *aller* Variablen des Faktorenmodells durch *einen bestimmten* Faktor erklärt werden. Durch die vorherige z-Transformation der Werte ist die gesamte zu erklärende Streuung im Beispiel entsprechend 11. Von diesem Betrag erklärt der erste Faktor 4,879, das entspricht 44,355 % (*% der Varianz*) der Gesamtstreuung, der zweite Faktor 1,389, entsprechend 12,624 % der Gesamtstreuung usw. Die vierte Spalte (*Kumulierte %*) weist die kumulierten Prozentsätze der erklärten Gesamtstreuung aus.

Offensichtlich macht es aber keinen Sinn, im Modell so viele Faktoren zu berücksichtigen wie Variablen vorhanden sind. Vielmehr wird ein Anhaltspunkt gesucht, mit dem eine (sinnvolle) Begrenzung der Faktoren vorgenommen werden kann. Per Voreinstellung verwendet dazu SPSS folgende Regel: Wenn der Eigenwert eines Faktors kleiner als 1 ist, erklärt dieser Faktor an Gesamtstreuung weniger als jede einzelne Variable (denn die erklärt wenigstens sich selbst und damit eine Streuung von 1). Entsprechend verbleiben in der Ausgabe der *Erklärten Gesamtvarianz* unter der Überschrift *Summen von quadrierten Faktorladungen für Extraktion* nur die Faktoren, die dieses Kriterium erfüllen (im Beispiel 2).

Erklärte Gesamtvarianz

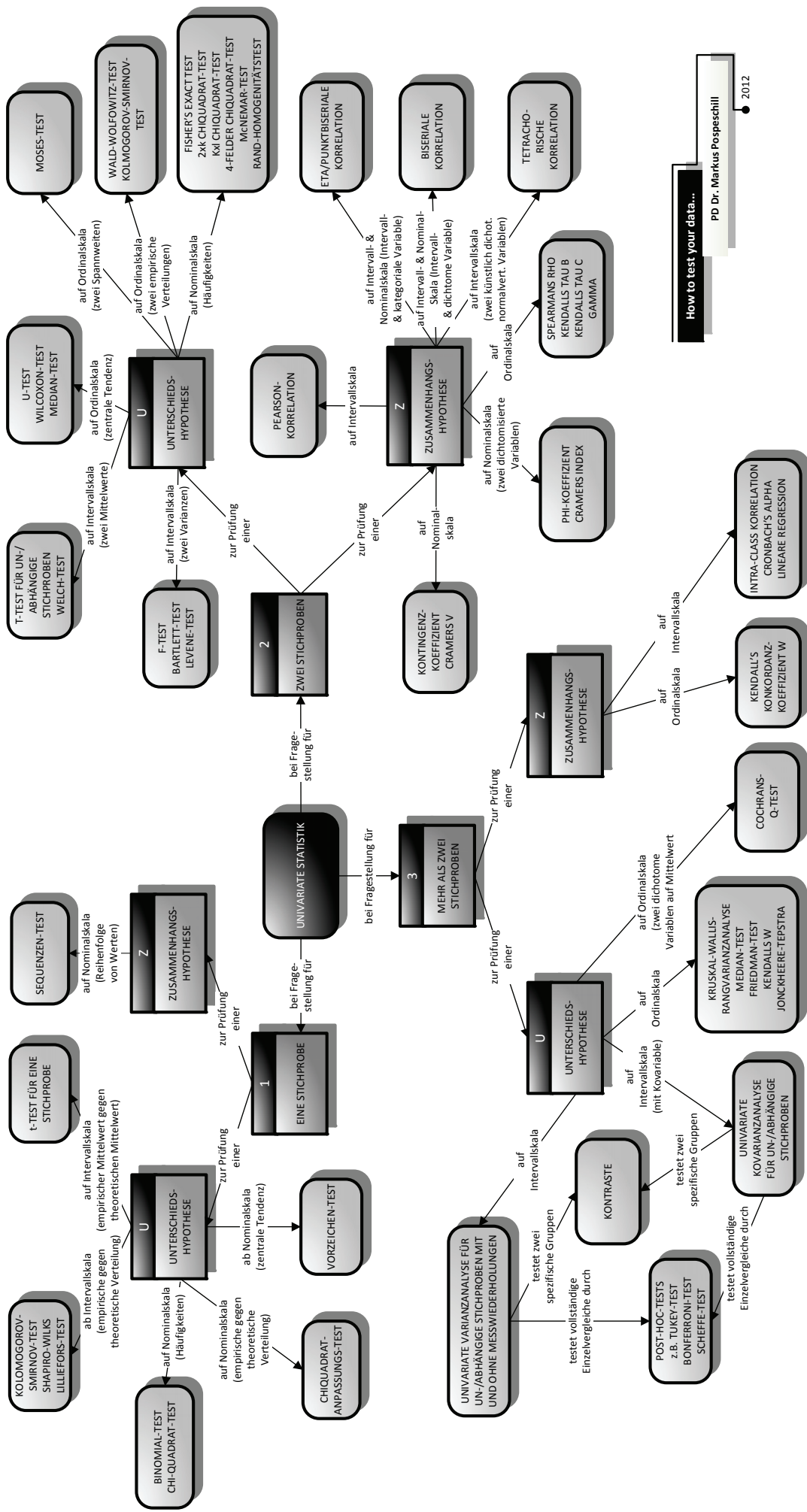
Komponente	Anfängliche Eigenwerte			Summen von quadrierten Ladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	4,879	44,355	44,355	4,879	44,355	44,355	3,702	33,657	33,657
2	1,389	12,624	56,979	1,389	12,624	56,979	2,565	23,323	56,979
3	,824	7,489	64,468						
4	,721	6,553	71,021						
5	,620	5,636	76,658						
6	,558	5,070	81,727						
7	,529	4,805	86,532						
8	,433	3,936	90,468						
9	,404	3,673	94,141						
10	,345	3,139	97,279						
11	,299	2,721	100,000						

Extraktionsmethode: Hauptkomponentenanalyse.

Ist zur besseren Interpretation der Faktoren ein Rotationsverfahren gewählt worden, erscheinen in der Spalte *Rotierte Summe der quadrierten Ladungen* die Veränderungen der Anteile der erklärten Varianz auf die extrahierten Faktoren. Anhand der kumulierten Prozentwerte kann aber abgelesen werden, dass sich durch das Rotationsverfahren die Güte des Gesamtmodells nicht verändert hat.

Zudem wird eine *Komponentenmatrix* (genauer Faktorladungsmatrix) angegeben, in der für jede Variable des Faktorenmodells die Koeffizienten der identifizierten Faktoren angegeben werden. Dabei gibt die *absolute* Größe der jeweiligen Faktorladung die Bedeutung des Faktors für die jeweilige Variable an²⁴.

²⁴ Bisher wurde davon ausgegangen, dass die Faktoren *orthogonal* sind, also keine Korrelationen aufweisen. Für diesen Fall gilt, dass ihre Koeffizienten (Ladungen) größengleich mit den Korrelationskoeffizienten zwischen dem Faktor und der Variablen sind. Diese Korrelationen zwischen Variablenpaaren aus den Faktorladungen lassen sich schätzen. Eine Abweichung dieses reproduzierten Korrelationskoeffizienten zum Koeffizienten der Originalvariablen wird Residuum genannt. Diese Annahme wird im Falle schiefwinkliger Faktoren aufgegeben.



How to test your data...
 PD Dr. Markus Pospeschill
 2012