

Tabellenverzeichnis

1.1	Das griechische Alphabet von A bis Ω	5
2.1	Religiösität, Rassismus und Lebensalter	12
2.2	Amtszeit von Kabinetten in Italien 1945-2008	29
2.3	Probabilistische Entscheidung über die Wahlbeteiligung	31
2.4	Einige typische Konstellationen von latenten Variablen, Indikatoren und Modellierungsmöglichkeiten	32
3.1	Zwei Dimensionen der Ausländerfeindlichkeit	67
3.2	Eine Dimension der Ausländerfeindlichkeit	69
3.3	Dimensionen der Ausländerfeindlichkeit: Modellvergleich	69
4.1	Struktur der Panelbefragung zur Bundestagswahl 2013	103

Glossar

χ^2 -Verteilung Eine eingipflige, asymmetrische Verteilung; dient als theoretisches Modell für die Verteilung von Zufallsvariablen und bildet damit die Grundlage für statistische Tests von Hypothesen. Form und Lage werden durch einen einzigen Parameter bestimmt, der der Zahl der Freiheitsgrade entspricht. 54, 55, 57, 60, 61

diagonale Matrix Quadratische Matrix, bei der alle Elemente außerhalb der Hauptdiagonalen den Wert 0 haben. 15, *siehe* Matrix

ECVI Expected Cross Validation Index. Mit den Informationskriterien verwandte Maßzahl. 58, *siehe* AIC, BIC, CAIC & nesting

Effekt Erwartete Reaktion einer abhängigen Variablen auf Veränderungen der unabhängige(n) Variable(n). Ob der Zusammenhang als kausal bzw. gerichtet interpretiert werden kann, hängt vom Design ab. 26, 42, 44, *siehe* Ex-post-facto-Design & Experimentaldesign

direkter Effekt Unmittelbarer Effekt einer unabhängigen Variablen auf eine abhängige Variable. 44

indirekter Effekt Effekt einer unabhängigen Variablen auf eine abhängige Variable, der über eine zwischengeschaltete Variable vermittelt ist. 44

totaler Effekt Summe des direkten und aller indirekten Effekte. 44

Einheitsmatrix (I) Diagonale Matrix, bei der die Elemente auf der Hauptdiagonalen den Wert 1 und alle anderen Elemente den Wert 0 haben. Auch: Identitätsmatrix. 15, 18, *siehe* diagonale Matrix & Matrix

Endogenität Ein Variable ist endogen, wenn sie von anderen Variablen innerhalb des Modells beeinflusst wird. Dies ist *per se* unproblematisch. Probleme ergeben sich aber im Falle von Rückkopplungsschleifen zwischen zwei Variablen. Aus diesen können verzerrte Schätzwerte resultieren. Dieses Problem wird als endogeneity bias bezeichnet (*siehe* Fußnote Abschnitt 2.4.1, Seite 39). 39, 41, *siehe* Exogenität

Exogenität Eine Variable ist exogen, wenn sie nicht auf andere Variablen innerhalb des Strukturgleichungsmodells zurückgeführt wird. 41, *siehe* Endogenität

Experimentaldesign Beim echten Experiment (mit Randomisierung) teilt die Forscherin die Versuchsobjekte zufällig auf Experimental- und Kontrollgruppe auf. Nur die Experimentalgruppe erhält den Stimulus. Beide Gruppen werden davor und danach untersucht. Durch die Randomisierung sind sich Experimental- und Kontrollgruppe so ähnlich wie nur möglich. Veränderungen in der Experimentalgruppe, die über die Veränderungen in der Kontrollgruppe hinausgehen, können deshalb mit großer Sicherheit auf einen kausalen Effekt des Stimulus zurückgeführt werden. 37, 38, *siehe* Ex-post-facto-Design

Ex-post-facto-Design Häufigstes Forschungsdesign in der Politikwissenschaft. Die Forscherin hat keine Kontrolle über den Stimulus, der durch die (soziale) Natur oder die Selbstselektion der Akteure gesetzt wird. Die Beobachtung erfolgt erst, nachdem der Stimulus gesetzt wurde (*ex post facto*). Kausale Schlüsse sind mit großen Problemen behaftet. 38, 39, 41, *siehe* Experimentaldesign

Faktorenanalyse Auch: Faktoranalyse. Oberbegriff für eine Reihe von unterschiedlicher statistischer Verfahren die versuchen, aus den empirischen Zusammenhängen zwischen beobachtbaren Variablen Informationen über eine kleinere Zahl von hypothetischen latenten Variablen zu extrahieren, die für diese Muster verantwortlich sein könnten. Explorative Faktorenanalysen treffen keine bzw. minimale Annahmen über die Struktur der latenten Variablen. Konfirmatorische Faktorenanalysen hingegen

testen, ob eine theoretisch plausible Struktur mit den Daten vereinbar ist. 3, 35, 43, 48, 62

CFA Konfirmatorische Faktorenanalyse. Faktorenanalyse bei der die Zahl der Faktoren und die Struktur der Ladungen auf Grundlage theoretischer Überlegungen vorgegeben werden. 35, 36, 54, 72, 98, 99, *siehe* MGCFA

MGCFA Konfirmatorische Mehr-Gruppen-Faktorenanalyse. CFA über mehrere Gruppen hinweg. 73–84, *siehe* CFA

FIML Full Information Maximum-Likelihood-Verfahren. Variante des Maximum-Likelihood-Verfahrens, das mit unvollständigen Daten umgehen kann, aber Zugriff auf die Rohdaten (nicht nur auf die Kovarianzen und Mittelwerte) voraussetzt. 88, 116, 123, *siehe* ML

Fit Qualität der Anpassung des Modells an die Daten. Kann durch diverse Maßzahlen beschrieben werden. 54–58, 70, 77, *siehe* Fit-Indizes, overfitting, AGFI, CFI, GFI, RMSEA & TLI

Fit-Indizes Maße, die die Qualität der Modellanpassung in einer einzelnen Größe zusammenfassen sollen. 55, 55, 56, 70, 77, 90, *siehe* Fit & overfitting

AGFI Adjusted Goodness of Fit-Index. Variante des GFI, die eine Korrektur für die Zahl der Freiheitsgrade enthält. Von der Nutzung wird in der neueren Literatur abgeraten. 55, *siehe* CFI, GFI, TLI, RMSEA & Freiheitsgrade

CFI Comparative Fit-Index. Vergleicht das aktuelle Modell mit einem Null-Modell ohne Parameter. Die Komplexität des Modells wird zusätzlich berücksichtigt, aber weniger stark gewichtet als beim TLI. Empirisch sind beide Werte eng korreliert, deshalb genügt es in der Regel, den TLI auszuweisen. 56, *siehe* AGFI, GFI, TLI & RMSEA

GFI Goodness of Fit-Index. Einer der ältesten Indizes für die Güte der Modellanpassung. Kann Werte zwischen 0 und 1 annehmen. Entspricht dem Anteil der beobachteten Kovarianzen, den die vom Modell implizierten Kovarianzen erklären können. Alte Faustregel: Werte über 0.9 (manchmal auch: 0.95) gelten als „gut“. Von der Nutzung wird in der neueren Literatur abgeraten. 55, *siehe* AGFI, CFI, TLI & RMSEA

NNFI Alternative Bezeichnung für den Tucker-Lewis-Index. 56, *siehe* TLI

RMSEA Root Mean Square Error of Approximation. Derzeit populärster Index für die Güte der Modellanpassung, unabhängig vom Stichprobenumfang. Basiert auf der Überlegung, dass Modelle die Population immer nur näherungsweise beschreiben. Hat eine bekannte Verteilung, so dass ein Konfidenzintervall für den Index berechnet werden kann. Faustregel: Werte kleiner als 0.1 (manchmal auch: 0.08) sind „akzeptabel“, Werte kleiner als 0.05 sind „gut“. 55–57, 70, 79, 91, *siehe* AGFI, TLI, CFI & GFI

TLI Tucker-Lewis-Index. Vergleicht das aktuelle Modell mit einem Null-Modell ohne Parameter, berücksichtigt zusätzlich die Komplexität des Modells. Verwendung wird empfohlen, wenn standardisierte Faktorladungen ≥ 0.5 . Ein TLI ≥ 0.95

gilt als „gut“. 56, 70, 91, *siehe* AGFI, CFI, GFI & RMSEA

Freiheitsgrade Zahl der unabhängigen Informationen, die für Signifikanztests genutzt werden können. 47, 53–55, 57, 60, 68–70, 77

Generalisiertes Lineares Modell Verallgemeinerung des linearen Regressionsmodells, die viele bekannte Modelle (lineares Regressionsmodell, binäre, ordinale und multinomiale logistische Regression, Poisson-Regression etc.) in einen gemeinsamen konzeptuellen Rahmen einordnet. 96, 97

Identifikation Ein Modell ist identifiziert, wenn für die Schätzung der Parameter eine eindeutige Lösung existiert. Dies setzt u. a. voraus, dass mindestens so viele unabhängige Informationen vorhanden sind, wie Parameter geschätzt werden sollen. 52, 62, 113

Informationskriterien Familie von Maßzahlen, die den Vergleich von Modellen auf der Grundlage von Fit und Komplexität ermöglichen und auf informationstheoretischen Ansätzen basieren. Diese Maßzahlen dienen *ausschließlich* dem Vergleich, ihr absoluter Wert ist nicht interpretierbar. Sie setzen nicht voraus, dass die Modelle verschachtelt sind. 58, *siehe* Fit

AIC Akaike Information Criterion. 58, 79, *siehe* BIC, CAIC, ECVI & nesting

BIC Bayesian Information Criterion. Alternative informationstheoretisch begründete Maßzahl. 58, 70, 79, *siehe* AIC, CAIC, ECVI & nesting

CAIC Consistent Akaike Information Criterion. Alternative informationstheoretisch begründete Maßzahl. 58, *siehe* AIC, BIC, ECVI & nesting

Inverse (A^{-1}) Die Inverse ist der Kehrwert einer quadratischen Matrix. Multipliziert man eine Matrix mit ihrer Inversen, so ergibt sich eine Einheitsmatrix: $AA' = I$. Matrizen, die nicht invertierbar sind, heißen singulär. 17–19, 53, *siehe* Einheitsmatrix & Matrix

Kollinearität (Fast) perfekte lineare Abhängigkeit innerhalb einer Reihe von Vektoren bzw. Variablen, führt zu instabilen und unpräzisen Schätzungen oder macht eine Schätzung unmöglich. Manchmal auch als Multi-Kollinearität bezeichnet. 2, 53, *siehe* Konvergenz

Konvergenz Parameter von Strukturgleichungsmodellen werden iterativ (schrittweise) geschätzt. Wenn sich die Schätzungen soweit an den wahren Wert annähern, dass von Schritt zu Schritt kaum noch Veränderungen der Schätzungen auftreten, spricht man von Konvergenz. Diese wird in den meisten Fällen schon nach wenigen Iterationen erreicht. Wenn das Modell komplex und/oder empirisch unzureichend identifiziert ist, wird das Stadium der Konvergenz nicht oder nur nach vielen Iterationen erreicht. 51, *siehe* Identifikation

Kovarianz ($\text{cov}(x, y)$) Maß für die gemeinsame Abweichung zweier Variablen von ihrem jeweiligen Mittelwert. 19

LCA Verfahren, die Untersuchungsobjekte genau einer latenten Klasse, d. h. einer Ausprägung einer kategorialen latenten Variable zuordnen. 124, 125, *siehe* LTA

LM-Test Lagrange-Multiplier-Test. Statistischer Test für die Aufnahme zusätzlicher Parameter in ein Modell. Manchmal auch als Score-Test bezeichnet, weil er auf dem Score-Vektor basiert, der die durch die zusätzlichen Parameter zu erwartenden Veränderungen Log-Likelihood enthält. 60, 79, *siehe* W-Test & LR-Test

LR-Test Likelihood-Ratio-Test, manchmal auch als χ^2 -Differenztest bezeichnet. Statistischer Test zum Vergleich zweier verschachtelter Modelle. 59–61, 78, 79, *siehe* LM-Test, nesting & W-Test

LTA Dynamische Erweiterung der LCA, die im Zeitverlauf den Übergang zwischen latenten Klassen zulässt. 124, 125, *siehe* LCA

Matrix (A, B etc.) Rechteckige Anordnung von reellen Zahlen mit m Zeilen und n Spalten. 12–19

Matrixmultiplikation Elementweise Multiplikation zweier Matrizen nach besonderen Regeln. 15, 16, 16, *siehe* Matrix

Mean Structure(s) Erweiterung des Strukturgleichungsmodells um Achsenabschnitte für die beobachteten Variablen, evtl. auch um Mittelwerte und gegebenenfalls Achsenabschnitte für die latenten Variablen. 47, 62

Messfehler Abweichung zwischen dem wahren Wert einer Variable und dem Ergebnis einer sozialwissenschaftlichen Messung. Idealerweise sollten beide Größen über viele Messungen hinweg nur zufällig voneinander abweichen. Kommt es jedoch zu systematischen Messfehlern, so spricht man von Verzerrung (bias). 2, 30, 33–35, 38, 43, 70, 86, 100, 102

MGSEM

Mehr-Gruppen-Strukturgleichungsmodell. Vollständiges Strukturgleichungsmodell, das über mehrere Gruppen hinweg geschätzt wird
. 84, *siehe* Faktorenanalyse

Missing Data Bezeichnet das Fehlen von Messwerten für einzelne Variablen (item non-response), im Unterschied zu Objekten, die überhaupt nicht untersucht bzw. befragt werden können (unit non-response). Missing Data reduzieren den Umfang des Datensatzes und damit die Effizienz der Schätzung. Systematische Ausfälle können die Schätzungen darüber hinaus verzerren. Treten Ausfälle in größerem Umfang auf, so müssen diese untersucht und ggf. durch den Einsatz besonderer Verfahren kompensiert werden. 87, 87, 88, 121–124, 126

MAR Missing At Random. Ausfallmechanismus, bei dem die Ausfallwahrscheinlichkeit nicht vom wahren Wert der Variablen selbst, aber vom Wert anderer Variablen abhängt. 122, 123, *siehe* MCAR & MNAR/NI

MCAR Missing Completely At Random. Ausfallmechanismus, bei dem die Ausfallwahrscheinlichkeit weder vom wahren Wert der Variablen selbst noch vom Wert anderer Variablen abhängt. 121, 122, *siehe* MAR & MNAR/NI

MNAR/NI Missing Not At Random oder Nonignorable. Ausfallmechanismus, bei dem die Ausfallwahrscheinlichkeit vom wahren Wert der Variablen selbst abhängt. Ausfälle sind damit nicht zufällig. 122, 123, *siehe* MCAR & MAR

ML Maximum-Likelihood-Verfahren. Schätzverfahren, das die Parameter findet, die am ehesten die Stichprobe hervorgebracht haben könnten. Setzt multivariate Normalverteilung der Daten voraus. 51, 52, 55, 66, 71, 78, 89, 102, 103, 112, 118, *siehe* multivariate Normalverteilung

Modifikationsindex Eine Maßzahl, die einen Hinweis darauf gibt, wie sehr sich der Fit durch die Aufnahme eines bestimmten Pfades in das Modell verbessern würde. 57, 60, *siehe* Fit

Multiple Imputation Verfahren, bei dem fehlende Werte durch mehrere plausible Schätzungen ersetzt werden. 123, *siehe* Missing Data

multivariate Normalverteilung Modell für die gemeinsame Normalverteilung mehrerer Variablen. Setzt u. a. voraus, dass bereits jede einzelne Variable normalverteilt ist. 51

nesting Zwei Modelle M_1, M_2 sind „nested“ (verschachtelt), wenn M_2 ein Spezialfall von M_1 ist. Dies bedeutet: (1) M_2 enthält ausschließlich Pfade, die auch in M_1 enthalten sind und (2) einige der Pfadkoeffizienten in M_2 werden nicht geschätzt, sondern sind auf einen bestimmten Wert (z. B. 0) fixiert. 57, 59, 77

OLS Ordinary Least Squares (auch: Verfahren der kleinsten Quadrate, Gauß-Verfahren). Schätzverfahren für lineare Regressionsmodelle, das die Summe der quadrierten Abweichungen zwischen den vom Modell implizierten Schätzungen für die abhängige Variable und den tatsächlich beobachteten Werten minimiert. Erfordert keine Iterationen, da eine analytische Lösung existiert. 28, 118

overfitting Im Forschungsprozess wird das Ausgangsmodell häufig modifiziert, um eine bessere Anpassung an die Daten zu erzielen. Wird ein Modell zu stark an die konkrete Stichprobe bzw. deren Idiosynkrasien angepasst, spricht man von overfitting. 55, 57

Pearsonscher Korrelationskoeffizient (r) Maß für den linearen Zusammenhang zwischen zwei Variablen, Wertebereich $[-1;1]$, basiert auf der Kovarianz. 23, *siehe* Kovarianz

Rang (rank) Der Rang ist ein Maß für den Informationsgehalt einer Matrix. 17, *siehe* Matrix

Skalarmultiplikation (Prä)Multiplikation einer Matrix mit einer reellen Zahl (Skalar). 15

Skalarprodukt Elementweise Multiplikation zweier Vektoren analog zur Matrixmultiplikation, manchmal auch als inneres Produkt oder Punktprodukt bezeichnet. 16, *siehe* Matrixmultiplikation

Skalenniveau Messen bedeutet einen empirischen Sachverhalt in Zahlen zu überführen. Der Informationsgehalt der Messung wird durch das Skalenniveau (nominal, ordinal, intervall- oder ratioskaliert) beschrieben. 11

Standardabweichung (s) Streuungsmaß; Quadratwurzel der Varianz. 19, 23, 24

Transposition ($A' = A^T$) Operation, die Zeilen und Spalten einer Matrix vertauscht. 17, *siehe* Matrix

Varianz (s^2) Streuungsmaß; Summe der quadrierten Abweichungen vom Mittelwert geteilt durch die Zahl der Beobachtungen. 19

Varianz-Kovarianz-Matrix (**S**) quadratische und symmetrische $m \times m$ Matrix, die Varianzen und Kovarianzen zwischen m Variablen enthält. 21, 22, *siehe* Kovarianz & Matrix

W-Test Wald-Test. Statistischer Test für die Herausnahme von Parametern aus einem Modell. 60, 78, 79, *siehe* LM-Test & LR-Test

WLS/ADF Weighted Least Squares/Asymptotically Distribution Free. Alternatives Schätzverfahren, das keine multivariate Normalverteilung der Daten voraussetzt. 51, 52, 55, 71, 89, 102, 103, *siehe* ML & multivariate Normalverteilung