

Die multiple lineare Regression im psychotherapiewissenschaftlichen Kontext

-Empfehlungen für Anwendung und Interpretation-

The multiple linear regression in psychotherapy science

-Recommendations for application and interpretation-

Oliver Wegenberger & Jan Aden

Kurzzusammenfassung

In diesem Beitrag der Serie Statistik in der Psychotherapiewissenschaft wird die Anwendung von Verfahren zur Analyse linearer Beziehungen im Sinne eines Best-Practice Ansatzes vorgestellt. Neben bivariaten Korrelationen wird vor allem das Verfahren der multiplen linearen Regression vorgestellt. Es werden Empfehlungen für (1) den Umgang mit Voraussetzungen, (2) der adäquaten Interpretation der Ergebnisse sowie (3) Reportkonventionen für die Ergebnisdarstellung gegeben. Darüber hinaus wird das Thema der Kausalität und Kausalinterpretationen empirischer Sachverhalte behandelt.

Schlüsselwörter

Korrelationen; Multivariate Verfahren; Multiple lineare Regressionsanalyse; Multikollinearität; Kausalität

Abstract

This contribution to the series Statistics in Psychotherapy Science aims at presenting methods for the analysis of linear relationships in a best-practice approach. Bivariate correlations are presented as well as the multiple linear regression, which is the focus of the present article. Recommendations are provided for (1) dealing with specific assumptions, (2) the adequate interpretation of results, and (3) statistical reporting. Furthermore, causality and causal interpretation of empirical facts are discussed.

Keywords

correlations; multivariate statistics; multiple linear regression; multicollinearity; causality

Einleitung

Einsatz-
bereiche
und Frage-
stellungen

Die psychotherapiewissenschaftliche Forschung befasst sich mit zahlreichen Problemstellungen und Fragen, die sich auf Zusammenhänge zwischen zwei oder mehreren Sachverhalten beziehen: Etwa ob die Qualität der therapeutischen Allianz/Arbeitsbeziehung (z.B. gemessen mit dem WAI-SR; Wilmers et al., 2008) mit der Reduktion depressiver Symptomatik nach Beendigung einer Therapie zusammenhängt; oder ob die Therapiemotivation (gemessen mit dem FPM; Nübling et al., 2005) der Klient*innen mit der Dauer der Therapie (z.B. Anzahl der Sitzungen) zusammenhängt. Die Bandbreite möglicher Fragestellungen ist groß und entsprechend häufig kommen statistische Verfahren zur Analyse solcher Sachverhalte zum Einsatz. Für eine empirische Untersuchung bivariater Wechselbeziehungen wird auf korrelative oder, im Falle bivariater Vorhersagen, (einfache) regressionsanalytische Verfahren zurückgegriffen. Im ersten Teil des vorliegenden Beitrags werden ausgewählte **korrelative Verfahren** sowie die **einfache lineare Regression** behandelt. Diese bilden gleichsam die Basis für das Verständnis multivariater Verfahren zur Analyse wechselseitiger Abhängigkeiten. Der Schwerpunkt des aktuellen Artikels liegt nämlich auf der Vorstellung des multivariaten Verfahrens der **Multiplen Linearen Regression**, welches für komplexe Vorhersagen von Sachverhalten durch mehr als eine andere Variable eingesetzt wird. Beispielsweise, wenn die Dauer der Therapie (Anzahl der Sitzungen) nicht mehr allein durch die Therapiemotivation (FPM) sondern gleichzeitig auch durch den Schweregrad der Symptomatik vorhergesagt werden soll, handelt es sich um eine Vorhersage, die der Komplexität des Sachverhalts adäquater Rechnung trägt, als es bei der Untersuchung eines Zusammenhangs von lediglich zwei Variablen der Fall wäre. In solchen Fällen ist der Einsatz der multiplen linearen Regression gefordert.

Eine Zielsetzung des Artikels besteht auch darin, auf die möglichen Fehlerquellen bei der Anwendung und Interpretation von multiplen linearen Regressionen zu sensibilisieren und gleichzeitig Informationen bereitzustellen, um in der Forschungspraxis eben solche Fehler zu vermeiden. Dazu werden einleitend die Grundlagen regressionsanalytischer Verfahren geschildert sowie der Umgang mit spezifischen Voraussetzungen, der Einsatz, die Interpretation und Reportierung von Ergebnissen multipler linearer Regressionsanalysen dargelegt. In einem gesonderten Exkurs befassen wir uns abschließend mit dem Spannungsfeld von Kausalität und Kovariation, das im Kontext regressionsanalytischer Verfahren von besonderer Bedeutung für die Ergebnisinterpretation ist.

Zusammenhangsmaße: Von Kovarianz und Korrelation zur Regression

Zusammenhangsmaße sind ein zentraler Bestandteil der Statistik. Dies reicht von der nicht-standardisierten Kovarianz über die standardisierte Korrelation bis hin zu Regressionsanalysen. Ein erster (klassischer) Schritt, um in die Welt der Zusammenhangsmaße einzusteigen, ist die Kovarianz. Sie drückt aus, ob zwei Variablen (x , y) in Beziehung stehen (Field, 2017). Wörtlich meint der Begriff der Kovarianz, dass Variablen ko-variiieren. Das bedeutet, dass, wenn eine Variable x variiert (also von ihrem Mittelwert abweicht) dies eine zweite Variable y ebenso tut. Ändert sich also ein Wert von Variable x , so ändert sich auch der Wert von Variable y (Bühner & Ziegler, 2009). Dabei kann diese Abweichung in die gleiche (positive Kovarianz) oder die entgegengesetzte Richtung (negative

Kovarianz) stattfinden (auf dies werden wir im Rahmen der Korrelationen gleich nochmal zu sprechen kommen). Die Kovarianz stellt ein nicht-standardisiertes Zusammenhangsmaß dar (Bortz & Schuster, 2010). Dies erschwert die Interpretation und stellt damit, aus Anwendungssicht, einen entscheidenden Nachteil des Maßes der Kovarianz dar.

Durch Standardisierung der Kovarianz (was typischerweise mithilfe der Standardabweichung geschieht) gelangt man zur Korrelation; genauer gesagt, erhält man den Korrelationskoeffizienten, r (Field, 2017). Um noch spezifischer zu sein, erhält man durch Standardisierung der Kovarianz den Korrelationskoeffizienten nach Pearson (auch Produkt-Moment Korrelationskoeffizient oder auch gelegentlich Bravais-Pearson-Korrelation, da die Korrelationsrechnung in einem Artikel von Bravais (1846) seinen Ursprung genommen hat (Bortz & Schuster, 2010)). Korrelationen, also, sind standardisierte Zusammenhangsmaße. Standardisiert bedeutet in diesem Zusammenhang, dass sie fixe Werte annehmen. Im Fall der Korrelation sind die Werte zwischen -1 und $+1$. Werte um 0 deuten auf einen Nullzusammenhang zwischen zwei Variablen, x und y , hin (obwohl eine Änderung in Variable x beobachtbar ist, bleibt Variable y gleich). Ein positiver Wert deutet auf einen positiven Zusammenhang hin („je mehr, desto mehr“). In anderen Worten heißt das, dass sich der Zusammenhang zwischen zwei Variablen – x und y – derart gestaltet, dass ein Anstieg in x , auch einen Anstieg in y bedeutet. Gleichermaßen bedeutet ein positiver Zusammenhang, dass ein Anstieg in y einen Anstieg in x bedeutet. Ein negativer Wert deutet auf einen negativen Zusammenhang hin („je mehr, desto weniger“). Also ein Anstieg in x geht mit einer Verringerung in y einher; vice versa (siehe positive und negative Kovarianz). Je näher der Wert des Korrelationskoeffizienten bei ± 1 liegt, desto stärker die positive/negative Korrelation ($r = 1$ bedeutet eine perfekt positive Korrelation; $r = -1$ bedeutet eine perfekt negative Korrelation). In Effektstärken bedeutet dies, dass Werte zwischen $r = \pm 0.10$ und $r = \pm 0.30$ auf einen kleinen Effekt, Werte zwischen $r = \pm 0.30$ und $r = \pm 0.50$ auf einen mittleren Effekt und Werte ab $r = \pm 0.50$ auf einen großen Effekt hinweisen.

In Abhängigkeit der Erfüllung spezifischer Voraussetzungen (Skalenniveau oder – wie im z.B. im Fall der Pearson Korrelation – Normalverteilung und Nicht-Vorhandensein von statistischen Ausreißern), finden verschiedene Verfahren zur Korrelationsberechnung ihren Einsatz. Unter anderem sind dies: Pearson Korrelation (Produkt Moment Korrelation), Spearman Korrelation, Punktbiseriale Korrelation etc. (siehe dazu z.B. auch: Bortz & Schuster (2010), Bühner & Ziegler (2009) oder Field (2017)).

Wie in Abb. 1 ersichtlich, stehen im Rahmen der klassischen bivariaten Korrelation die Variablen in einem wechselseitigen Verhältnis zueinander. Graphisch gesehen bedeutet das, dass der Richtungspfeil in beide Richtungen zeigt (siehe Abb. 1). Hier wird der bidirektionale Charakter der Korrelation deutlich. Dies stellt einen Unterschied zu den Regressionsanalysen dar, auf die wir im Folgenden eingehen werden.

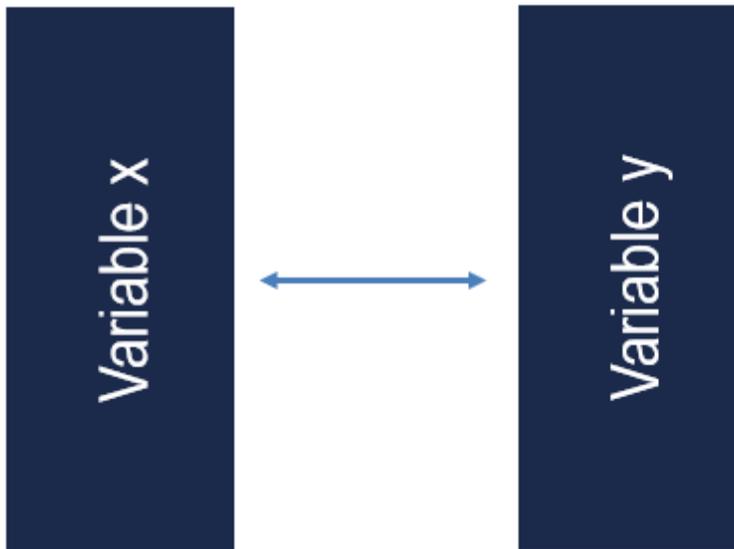


Abb. 1 Bivariate Korrelation

Regression¹

Im Rahmen der Regressionsanalyse wird die Beziehung zwischen einer abhängigen Variable und einer (im Fall der einfachen linearen Regression) oder mehreren (im Fall der multiplen linearen Regression) unabhängigen Variablen untersucht (Backhaus et al., 2018). Im Fall der Regression, wird, im Unterschied zur Korrelation, ein unidirektionales Verhältnis untersucht. Wie in Abb. 2 (Einfache lineare Regression) dargestellt, zeigt der Pfeil nicht in zwei (wie im Fall der Korrelation), sondern bloß in eine Richtung. Im Fall der einfachen linearen Regression soll also eine abhängige Variable y mithilfe einer anderen, unabhängigen Variable x hervorgesagt werden.

Korrelation
vs. Regression



Abb. 2 Einfache lineare Regression

¹ Im Rahmen dieses Artikels gehen wir auf die lineare Regression ein, für eine Auseinandersetzung mit z.B. logistischen Regressionen oder der Nichtlinearen Regression verweisen wir z.B. auf Field (2017) oder Backhaus et al. (2018) bzw. Backhaus et al. (2015).

Für die abhängige / unabhängige Variable lassen sich in der Literatur unterschiedliche Bezeichnungen finden; in Tab. 1 sind die gängigsten Bezeichnungen aufgelistet. Im Rahmen dieses Artikels werden wir vornehmlich die Begriffe Kriterium (abhängige Variable) und Prädiktor (unabhängige Variable) verwenden.

Tab. 1 Bezeichnungen für Abhängige und Unabhängige Variable(n) im Zuge der Regression

<i>Abhängige Variable (y)</i>	<i>Unabhängige Variable(n) (x)</i>
Kriteriumsvariable	Prädiktorvariable
Erklärte Variable	Erklärende Variable
Regressand	Regressor
Endogene Variable	Exogene Variable
Prognosevariable	

Die einfache lineare Regression

Im Rahmen der einfachen linearen Regression wird also eine Kriteriumsvariable y durch eine Prädiktorvariable x vorhergesagt. Die Formel dazu lautet: $\hat{y}_i = bx_i + a$. In dieser Formel stecken zwei entscheidende Details. Zum einen erinnert die Formel an die Geradengleichung, welche lautet: $y = kx + d$. Zum anderen fällt das „Dach“ über y auf, welches bedeutet, dass Werte von y anhand des Regressionsmodells geschätzt werden. In welcher Form die geschätzten Werte von y bei der Gesamtbeurteilung eines Regressionsmodells verwendet werden, behandeln wir in späterer Folge [im Zuge Modellbeurteilung]. Dieses ist im Fall der einfachen linearen Regressionen simpel gehalten: die Werte von y ergeben sich aus den konkreten Werten (i) des Prädiktors x multipliziert mit der Steigung b plus einer Konstanten a .

Was hat es nun mit der frappierenden Ähnlichkeit der Regressionsgleichung mit der Geradengleichung auf sich? Wir wollen uns dieser Frage graphisch nähern. Nehmen wir ein Streudiagramm zwischen Kriterium und Prädiktor. Dies würde z.B. so aussehen wie in Abb. 3.

Geraden-
gleichung
Regressions-
funktion

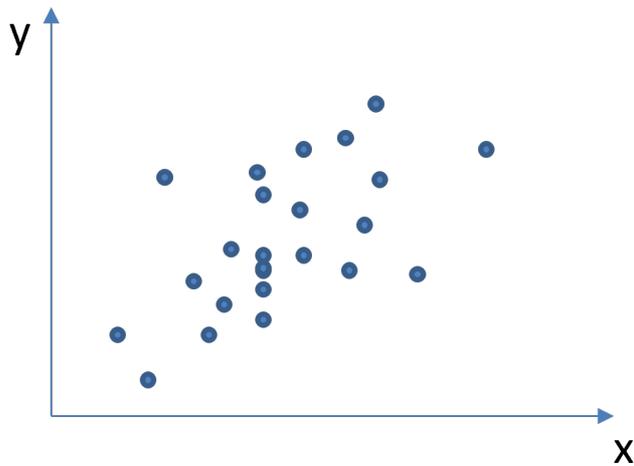


Abb. 3 Streudiagramm zwischen Kriterium y und Prädiktor x

Im Fall der einfachen linearen Regression wird durch die Funktion, $\hat{y}_i = bx_i + a$, eine Gerade erzeugt (Backhaus et al., 2018). Diese Gerade soll den Punktschwarm möglichst gut repräsentieren. Möglichst gut repräsentieren bedeutet, dass der Abstand (die Residuen bzw. Fehler) zwischen den tatsächlichen Werten (dargestellt durch die Punkte) und den geschätzten Werten (der Geraden) möglichst klein ausfällt (siehe Abb. 4). Die Residuen können, wie in Abb. 4 ersichtlich, sowohl positiv als auch negativ sein, wodurch das Problem entsteht, dass sich Fehler ausgleichen können. Aus diesem Grund werden die Residuen zuvor quadriert, sodass auch negative Werte, ein positives Vorzeichen erhalten. Für die Gerade bedeutet das, dass die Summe der quadrierten Residuen (Abweichungen), welche durch die Quadrierung ausschließlich positive Werte aufweisen, so klein wie möglich ausfallen soll. In anderen Worten werden die Abweichungen, durch das Finden der idealen linearen Gleichung minimiert (Backhaus et al., 2018; Wentura & Pospeschill, 2015). Im besten Fall liegt die Summe der Abweichungen bei 0 bzw. möglichst nahe bei 0 (Bortz & Schuster, 2010). Dieser Vorgang nennt sich die Methode der kleinsten Quadrate. Die Methode der kleinsten Quadrate (auch Kleinst-Quadrate- oder KQ-Schätzung) zählt zu den bedeutendsten Schätzverfahren in der Statistik (Backhaus et al., 2018). Im Rahmen der Regression stellt sie, wie beschrieben, die Methode für die Optimierung der Schätzung des Kriteriums dar (Eid et al., 2011).

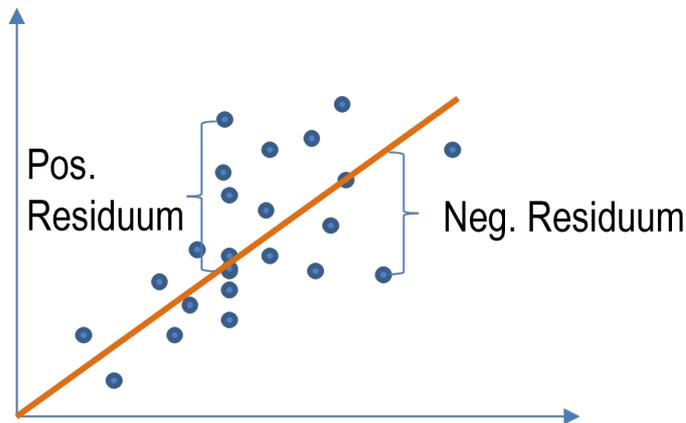


Abb. 4 Regressionsgerade mit positiven und negativen Residuen (Abweichungen)

Die multiple lineare Regression

Der Schritt von der einfachen linearen Regression zur multiplen linearen Regression ist ein simpler (Wentura & Pospeschill, 2015). Der Unterschied ist, dass die Kriteriumsvariable y im Rahmen der multiplen linearen Regression nicht auf Basis einer Prädiktorvariable, sondern basierend auf mehreren Prädiktoren vorhergesagt wird (siehe Abb. 5 – Regressionsmodell mit 4 Prädiktorvariablen). Warum wird das gemacht? Speziell in den Sozialwissenschaften werden oftmals multidimensionale Fragestellungen untersucht. Nehmen wir ein einfaches Beispiel: Es soll der Studienerfolg (Kriterium y) vorhergesagt werden. Dies kann beispielsweise anhand von Fähigkeit als Prädiktor geschehen. Möglicherweise stellt aber auch die Studienmotivation einen sinnvollen Prädiktor dar; darüber hinaus vielleicht auch die Interessenslage. Im Rahmen der einfachen linearen Regression steht man vor der Entscheidung „entweder/oder“. Diese „entweder/oder“ Entscheidung führt jedoch möglicherweise zu (zu) simplifizierten Aussagen. Die multiple lineare Regression erlaubt, all diese Prädiktoren (im obigen Bsp. sind das Fähigkeiten, Motivation, Interessen) zugleich zu betrachten und damit die Fragestellung nach beeinflussenden Variablen auf das Kriterium „Studienerfolg“ umfassender zu beleuchten. Mit dem Schritt von der einfachen zur multiplen linearen Regression kommt man als von der „entweder/oder“ Entscheidung zu einer „sowohl/als-auch“ Situation.

Auch im Zusammenhang empirischer Forschung zu psychischen Krankheitsbildern werden die Vorteile multipler/multivariater Regressionen deutlich. Beispielsweise lässt sich die Ausprägung der Schwere depressiver Symptomatik nicht allein durch einen (Risiko-) Faktor vorhersagen, sondern es benötigt die gleichzeitige/simultane Berücksichtigung weiterer, parallel existierender Einflussfaktoren. So sollte die Mehrdimensionalität eines (klinischen) Sachverhalts in einem diesbezügliches Regressionsmodell entsprechend abgebildet werden, um eine möglichst adäquate Vorhersage treffen zu können.

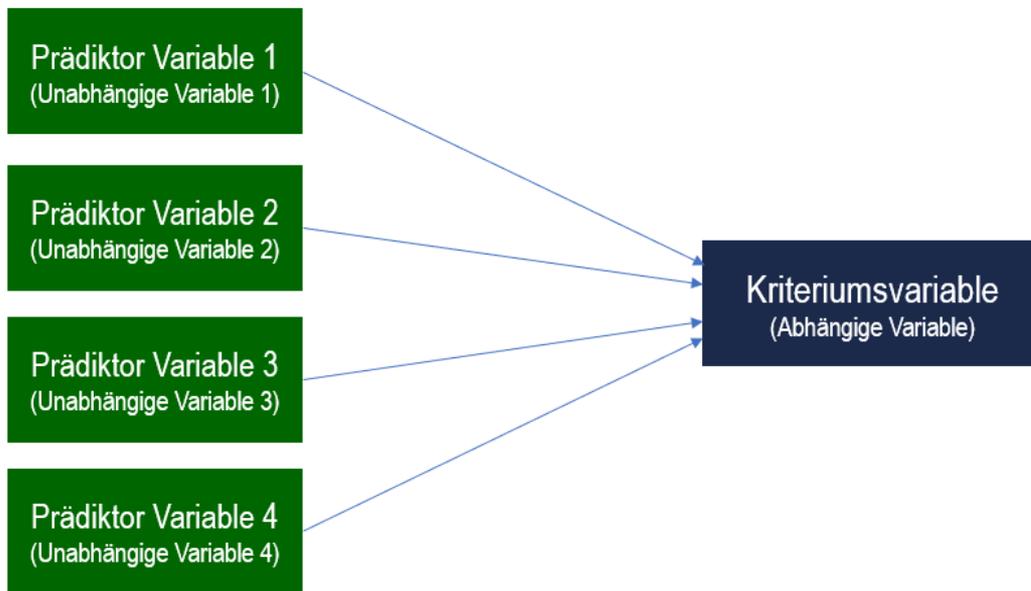


Abb. 5 Multiples lineares Regressionsmodell mit 4 Prädiktorvariablen

Der Umstand, dass ein Kriterium umfassender, also anhand mehrerer Prädiktoren erklärt wird, spiegelt sich auch in der Gleichung der multiplen linearen Regression wider. Wie oben dargestellt, sind es im Fall der einfachen linearen Regression ein Prädiktor x und dessen Steigung b (plus eine Konstante a), welche ein Kriterium y erklären. Im Fall der multiplen linearen Regression sind es mehrere Prädiktoren x_k und deren Steigungen b_k (plus einer Konstanten b_0 sowie der Fehlervarianz e), welche ein Kriterium y erklären. Die Gleichung einer multiplen linearen Regression lautet demgemäß: $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + e$. Wie auch im Fall der einfachen linearen Regression, werden die Regressionsparameter b (also die Steigungen) so gewählt, dass die Abweichung zwischen geschätzten Werten und den tatsächlichen Werten des Kriteriums y minimal ausfallen (Backhaus et al., 2018). Dies geschieht ebenso mithilfe der oben beschriebenen Methode der kleinsten Quadrate.

Im Zuge der multiplen linearen Regression spricht man von der Bildung eines Modells zur Erklärung bzw. Vorhersage einer Kriteriumsvariable. An dieser Stelle sei ein Wort der Vorsicht ausgesprochen. Die Bildung eines Regressionsmodells sollte theoriebasiert erfolgen und nicht in einer „Versuch/Irrtum“ Vorgehensweise. Die Auswahl der einzelnen Prädiktoren sollte theoretisch begründet sein, sodass ein Modell die Erkenntnisse vorangegangener empirischer Forschung und theoretischer Überlegungen bestmöglich abbildet. Für stärker explorativ ausgerichtete Fragestellungen, etwa zu bislang weniger untersuchten Sachverhalten, stellen die sog. *Schrittweisen Regressionen* (z.B. mit Vorwärts- und Rückwärtsalgorithmen) die Verfahren der Wahl dar; für vertiefte Informationen dazu siehe z.B. Bortz (2006).

Wenn aber ein theoretisch fundiertes Modell aufgestellt wurde, müssen die Prädiktoren sowie die Kriteriumsvariable messtheoretische Voraussetzungen erfüllen. Bezogen auf die Prädiktoren lassen sich diese wie folgt benennen: Prädiktoren müssen metrisch oder dichotom (max. zwei Ausprägungen) sein. Sollte(n) eine oder mehrere ausgewählte Prädiktorvariablen jedoch kateogreal sein und mehr als zwei Ausprägung aufweisen (z.B. höchste abgeschlossene Ausbildung: 1= Pflichtschule, 2= Lehre, 3= Matura/Abitur, 4= Studium), dann müssten solche Variablen in dichotome Variablen umcodiert werden (z.B. 1= ohne Matura/Abitur, 2= mit Matura/Abitur). Eine weitere Möglichkeit, kateogreal polytome Variablen in ein dichotomes Ausprägungsschema zu transformieren, stellt die Technik der Dummy-Codierung da. Details zur Anwendung dieser Technik sind u.a. bei Bortz (2006) zu finden und werden im Rahmen dieses Artikels nicht weiter ausgeführt.

Das Skalenniveau der Kriteriumsvariable ist das entscheidende Kriterium, anhand dessen das korrekte regressionsanalytische Verfahren ausgewählt wird: Gilt es eine kateogreale (dichotome oder polytom) Kriteriumsvariable vorherzusagen wird die Verfahrensgruppe der **Logistischen Regressionen** angewandt (siehe Tab. 2).

Soll ein metrisches Kriterium aus mehreren Prädiktorvariablen vorhergesagt werden, stellt die bereits beschriebene **Multiple Lineare Regression** das indizierte Verfahren dar. Der vorliegende Artikel wird sich in weiterer Folge ausschließlich dem Einsatz sowie der korrekten Interpretation eben der multiplen linearen Regression im psychotherapiewissenschaftlichen Kontext widmen. Für eine eingehende Auseinandersetzung mit Formen logistischer Regressionen sei exemplarisch auf Backhaus et al. (2018) verwiesen.

Tab. 2 Übersicht zu Arten multipler Regressionen

Prädiktoren	Kriterium	Verfahren
metrisch und/oder dichotom	metrisch	Multiple lineare Regression*
metrisch und/oder dichotom	kateogreal (dichotom)	Binär logistische Regression
metrisch und/oder dichotom	kateogreal (polytom)	Multinomiale logistische Regression

* bei explorativen Fragestellungen kommen schrittweise Verfahren zum Einsatz

Im Rahmen der multiplen linearen Regression können grob zwei Interpretationsebenen unterschieden werden, für welche wir im vorliegenden Artikel die folgenden Begriffe verwenden: (1) Modellebene und (2) Prädiktorebene.

Modellebene

Die Modellebene bezieht sich auf die Güte eines Modells insgesamt und gibt Aufschluss darüber, ob und wie gut sich das Modell – bestehend aus dem Set gewählter Prädiktoren – überhaupt für die Vorhersage des Kriteriums eignet. Zur Beurteilung der Modellgültigkeit wird für die inferenzstatistische Verallgemeinerung auf den F- und p-Wert einer ANOVA zurückgegriffen, die im Zuge regressionsanalytischer Berechnung ebenfalls kalkuliert wird. Bei einem signifikanten Ergebnis (p-Wert ≤ 0.05) besitzt das Modell eine zumindest inferenzstatistische Gültigkeit. Achtung: Mit einem signifikanten Ergebnis in der ANOVA ist die Güte eines Modells bei Weitem noch nicht ausreichend belegt. Ein weiterer zentraler Kennwert zur Beurteilung des Gesamtmodells ist das **adjustierte R^2** . Dieser Kennwert stellt das korrigierte Bestimmtheitsmaß der multiplen Korrelation R dar. Doch worum handelt es sich genau bei R und welche Informationen lassen sich zur Beurteilung der Modellgüte ableiten? Bei R handelt es sich schlicht um die bivariate Korrelationen zwischen dem auf Basis des Modells vorhergesagten Wertes \hat{y}_i (siehe Abbildung 6) und dem Wert, den eine Person tatsächlich aufweist, y_i . Es sei an dieser Stelle nochmals daran erinnert, dass die Berechnung von Regressionsmodellen auf Stichproben zurückgreift, von denen das vorherzusagende Kriterium tatsächlich gemessen wurde und für jede Person vorliegt. Im Zuge der Analyse wird nun eigentlich untersucht, inwiefern die tatsächlichen Werte y_i mit denjenigen, die auf Basis des Modells vorhergesagt wurden (\hat{y}_i), übereinstimmen. Die Methode der kleinsten Quadrate „sorgt“ dafür, dass die Diskrepanz so gering wie möglich ausfällt. Wie „ähnlich“ die tatsächlichen und vorhergesagten Werte aber wirklich sind, wird durch eine bivariate Korrelation der beiden Werte untersucht. R ist also der Korrelationskoeffizient zwischen \hat{y}_i und y_i . Die eingangs erwähnte ANOVA bzw. F-Test bezieht sich konkret darauf, ob eben diese (multiple) Korrelation $R^{(2)}$ signifikant von Null abweicht.

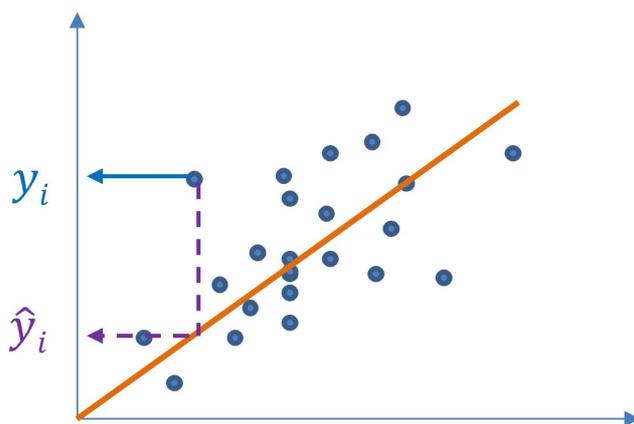


Abb. 6 Ermittlung der tatsächlichen (y_i) und auf Basis des Modells vorhergesagten Werte (\hat{y}_i)

Wird der Korrelationskoeffizient R quadriert und darüber hinaus mit 100 multipliziert, erhält man das Bestimmtheitsmaß in Form wechselseitig erklärter Varianz in Prozent. Alltagssprachlich formuliert, gibt R^2 multipliziert mit 100 an, wieviel Prozent der Varianz des Kriteriums durch die Prädiktoren des Modells erklärt werden können. Unter anderem um dem Ökonomiegebot multipler

Regressionsmodelle Rechnung zu tragen, wird das R^2 zuvor noch korrigiert, indem auch die Anzahl eingespielter Prädiktoren berücksichtigt wird (n = Stichprobengröße; k = Anzahl der Prädiktorvariablen):

$$R^2_{adj.} = R^2 - \frac{k * (1 - R^2)}{n - k - 1}$$

Das $R^2_{adj.}$ dient in weiterer Folge als Berechnungsgrundlage für die wechselseitig erklärte Varianz ($R^2_{adj.} * 100$).

Zusammenfassend lässt sich das Gesamtmodell also anhand des p-Wertes sowie des $R^2_{adj.}$ beurteilen. Im statistischen Report werden daher die Kennwerte der F-Statistik der ANOVA inklusive p-Wert ($F(df_1, df_2) = , p =$) sowie das $R^2_{adj.}$ angegeben:

„Das Modell zur Vorhersage des Kriteriums X, bestehend aus den Prädiktoren A, B, C und D erklärt 90.30% der Varianz ($R^2_{adj.} = .903$) des Kriteriums und ist mit $F(4, 296) = 63.63, p < .001$, signifikant.“

Prädiktorebene

Eine weitere Analyseebene bezieht sich – wie bereits erwähnt – auf die Beurteilung der einzelnen Prädiktoren. Hierbei steht die Frage im Vordergrund, wie hoch die Relevanz der einzelnen Prädiktoren für die Vorhersage des Kriteriums im Kontext des Modells ist. Die wichtigsten Kennwerte stellen in diesem Zusammenhang der unstandardisierte Regressionskoeffizient b sowie der standardisierte Regressionskoeffizient β (*Standardpartialregressionskoeffizienten*) dar.

Beim unstandardisierten Regressionskoeffizient b handelt es sich um die Steigung (vergleiche Geradengleichung bei einfacher linearer Regression). Der Regressionskoeffizient gibt also an, welche Auswirkung eine Änderung in einem Prädiktor auf das Kriterium hat (Backhaus et al., 2018). Darüber hinaus gibt b einen Aufschluss über die „Richtung“ dieser Auswirkung (positiv oder negativ). Die Höhe der Wertausprägung bei b lässt sich unter den Prädiktorvariablen schwer vergleichen, da die Steigung für eine spezifische Prädiktorvariable mit der Skalierung der Variable zusammenhängt. Eine Variable deren Ausprägungsbereich von 0-100 reicht, weist im b -Wert voraussichtlich einen anderen Wert auf als eine Variable, die Werte zwischen 0 und 10 annehmen kann. Hier ist also die Skalierung der Prädiktoren zu beachten. Um einen Vergleich der Prädiktionsleistung zwischen den Prädiktorvariablen anstellen zu können, werden die Regressionskoeffizienten b standardisiert. Die standardisierte Form ist der Koeffizient β . Diese Koeffizienten können Werte zwischen -1 und +1 annehmen und sind wie Korrelationskoeffizienten zu interpretieren (siehe Abschnitt zu Zusammenhangsmaßen). Die β -Koeffizienten geben Aufschluss darüber, in welcher Weise (positiv- oder negativ-linear) und wie stark (± 0.1 klein, ± 0.3 mittel, ± 0.5 groß (Cohen, 1988)) der jeweilige Prädiktor im Modellkontext zur Vorhersage des Kriteriums beiträgt.

Die Signifikanzprüfung der Regressionsgewichte b und β erfolgt in einer t-Verteilung. Im statistischen Report werden pro Prädiktor b und β sowie die t-Statistik inklusive p-Wert und Konfidenzintervall angegeben ($b = ; Kl_{95} = ; \beta = ; t(df) = ; p =$):

„...Prädiktor A konnte als relevanter Faktor zur Prädiktion des Kriteriums X identifiziert werden ($b=0.69$, $KI_{95}= 0.42 - 0.97$, $\beta= .67$, $t(296)= 5.23$; $p < .001$) und trägt von allen Prädiktorvariablen am stärksten zur Vorhersage bei. Prädiktor C weist die zweithöchste Prädiktionsleistung auf ($b=7.98$, $KI_{95}= 4.97 - 10.99$, $\beta= 0.36$, $t(296)= 5.49$; $p < .001$) und steht ebenfalls in einem positiv linearen Verhältnis mit dem Kriterium X. Die Prädiktoren B ($b=0.46$, $KI_{95}= -0.09 - 0.53$, $\beta= .09$, $t(296)= 1.09$; $p = .249$) und D ($b=0.69$, $KI_{95}= -0.18 - 1.12$, $\beta= .09$, $t(296)= 1.07$; $p = .253$) leisten keinen signifikanten Beitrag zur Vorhersage.“

Gerade bei umfangreicheren, d.h. eine große Zahl an Prädiktorvariablen umfassenden Modellen, bietet sich die Darstellung der Koeffizienten in tabellarischer Form an. Darüber hinaus besteht ebenso die Möglichkeit, weitere Kennwerte wie die Nützlichkeit (U_i) oder den Strukturkoeffizienten (C) anzuführen (siehe nähere Informationen dazu bei Backhaus et al. (2018)). Außerdem sei an dieser Stelle ein Wort der Warnung ausgesprochen. Der Regressionskoeffizient b bzw. β einer Prädiktorvariable ist in seiner jeweiligen Ausprägung durch die übrigen Prädiktoren beeinflusst. D.h., dass der Regressionskoeffizient einer Variable nur im Kontext des jeweils zur Prüfung stehenden Modells zu verstehen ist. Das b und das β -Gewicht eines Prädiktors können im Kontext anderer Prädiktorvariablen bei der Vorhersage desselben Kriteriums erheblich schwanken. Der Grund dafür besteht in der wechselseitigen Verbundenheit/Abhängigkeit (geteilte Varianz) der Prädiktoren untereinander. Ist die Abhängigkeit der Prädiktorvariablen untereinander zu hoch, können die b - und β -Gewichte stark verzerrt sein und sind dann nicht mehr adäquat zu interpretieren.

Voraussetzungen im Rahmen der multiplen linearen Regression

Daraus folgt, dass eine multiple lineare Regression nur dann durchgeführt und sachgerecht interpretiert werden sollte, wenn einige empirische Voraussetzungen erfüllt sind und dieser Umstand mit wiederum eigenen Verfahren überprüft worden ist. Dabei ist die Verletzung der einzelnen Voraussetzung mit unterschiedlich starken Konsequenzen für die Interpretierbarkeit des jeweiligen Regressionsmodells verbunden. Im Extremfall spiegeln die Ergebnisse der Regressionsanalyse die Realität nicht wider (Haupt et al., 2014). Folgend wird eine Zusammenschau der wichtigsten Voraussetzungen gegeben, vertiefende sowie umfänglichere Beschreibungen sind beispielsweise bei Backhaus et al. (2018) oder Field (2017) zu finden.

Die bereits erwähnte Unabhängigkeit der Prädiktoren stellt die, bei Verletzung, mit den stärksten Konsequenzen verbundene Voraussetzung dar. Mit Unabhängigkeit ist in diesem Zusammenhang gemeint, dass die Prädiktoren untereinander nicht zu hoch korrelieren dürfen. Vereinfacht ausgedrückt bedeutet dies, dass sich die Prädiktoren nicht zu ähnlich sein dürfen, da ansonsten u.a. nicht mehr eindeutig rückführbar ist, durch welchen von z.B. zwei sehr ähnlichen (hoch korrelierenden) Prädiktoren bestimmte (Varianz-)Anteile der Kriteriumsvariable erklärt werden. Zusammengefasst kommt in der Voraussetzung zum Ausdruck, dass Prädiktoren untereinander nicht (zu hoch) korrelieren dürfen. Ein moderates Ausmaß an Korreliertheit ist akzeptabel; eine zu starke Verbundenheit führt jedoch zu den benannten Problemen. Daher wird diese Voraussetzung auch als **Abwesenheit perfekter/extremer Multikollinearität** bezeichnet. Im Falle mangelhafter Berücksichtigung dieser Voraussetzung kann es zu schwerwiegenden Fehlern bei der Interpretation des Modells v.a. aber der Beurteilung einzelner Prädiktoren kommen. Daher ist eine eingehende Prüfung der Multikollinearität eine notwendige Bedingung für die Interpretation der Ergebnisse einer

multiplen linearen Regression. Die Prüfung kann anhand mehrerer Verfahren und deren zentralen Kennwerte erfolgen. In diesem Artikel empfehlen wir die Betrachtung der sogenannten **VIF-Werte** (Variance Inflation Factor) oder der **Toleranzstatistik** ($1/VIF$). Für jeden Prädiktor wird ein Toleranz- und VIF-Wert berechnet, die – alltagssprachlich formuliert – angeben, inwiefern die jeweilige Prädiktorvariable „von Multikollinearität betroffen ist“. In der Fachliteratur (z.B. Field (2017)) werden zumeist folgende Grenzwerte vorgeschlagen: Toleranzwerte von < 0.2 (entspricht $VIF= 5$) bzw. < 0.1 (entspricht $VIF= 10$) zeigen eine Verletzung der Voraussetzung an. Auf die Verletzung dieser Voraussetzung kann auf unterschiedliche Weise umgegangen werden. Es muss nämlich nicht gleich das gesamte Modell verworfen werden; in manchen Fällen könnte beispielsweise der betroffene Prädiktor einfach aus dem Modell entfernt werden. Der konkrete Umgang mit dem Vorliegen extremer Multikollinearität ist jedoch abhängig von der jeweiligen Fragestellung und dem Anspruch des geprüften Modells. Sicher ist jedoch, dass ein Modell, das von extremer Multikollinearität betroffen ist, nicht sinnvoll interpretiert werden kann und mindestens modifiziert werden muss. Neben den VIF-Werten und der Toleranzstatistik können zur Überprüfung ebenso bivariate Korrelationen zwischen den Prädiktoren interpretiert werden. Im statistischen Report müssen Angaben zur Multikollinearitätsprüfung enthalten sein, was standardmäßig über die Angaben der VIF- oder Toleranzwerte pro Prädiktor erfolgt.

Darüber hinaus gibt es weitere Voraussetzungen, die im Rahmen einer multiplen linearen Regression überprüft werden müssen. Die Voraussetzung der **Homoskedastizität** bildet die Forderung nach Gleichheit der Streuung der Fehler in der Beobachtungsreihe ab. Deren Verletzung hat Auswirkungen auf die Schätzung der Standardfehler der einzelnen Regressionskoeffizienten und in weiterer Folge auf deren inferenzstatistische Überprüfung. Die Homoskedastizität kann graphisch mittels Streudiagramm von (standardisierten) vorhergesagten Werten und den (standardisierten) Fehlern sowie mittels **Breusch-Pagan-Test** (darf nicht signifikant sein) überprüft werden. Im selben Streudiagramm lässt sich auch die Bedingung der **Linearität** graphisch beurteilen. Eine weitere Voraussetzung stellt die **Unkorreliertheit der Fehler/Residuen** dar. Ist diese Voraussetzung nicht erfüllt, kommt es zur sogenannten Autokorrelation der Fehler, welche ebenso die Schätzung der Standardfehler der einzelnen Regressionskoeffizienten beeinflusst. Zur Überprüfung wird die **Durbin-Watson-Statistik** herangezogen, deren Wert idealerweise bei 2 liegt. Ausprägungen, zwischen 1 und 3 sind allerdings noch akzeptabel. Eine weitere Voraussetzung, die sich auf die Fehler bezieht, ist die **Normalverteilung der Residuen**. Diese kann über ein **Normalverteilungsdiagramm** oder **Goodness-of-Fit Test** (z.B. Kolmogorov-Smirnov-Test oder Shapiro-Wilk-Test – dürfen nicht signifikant ausfallen) überprüft werden. Weiterführende Informationen zu den Konsequenzen bei Verletzung der jeweiligen Voraussetzung siehe z.B. Backhaus et al. (2018).

Weitere
prüfbare
Voraus-
setzungen

Exkurs: Kausalität & Kovariation

Was ist der Grund für ...? Warum ist ...? Fragen, die aus dem Alltag bestens bekannt sind und auf welche möglicherweise schnell (vermeintliche) Antworten gefunden sind. Seit (zehn-)tausenden Jahren,

beschäftigen sich Menschen mit derartigen Fragen (Kossakowski et al., 2021). Am befriedigendsten ist es, auf obige Fragen nach dem „Warum ist...?“ eine kausale Antwort geben zu können; schließlich war (ist) das Stellen der „Warum-Frage“ für die evolutionäre Entwicklung des Menschen von herausragender Bedeutung (Pearl & Mackenzie, 2019; Kossakowski et al., 2021). Viele Studien in den Sozialwissenschaften haben zum Ziel Fragen kausaler Natur zu beantworten (Pearl, 2009, 2010), jedoch sind nicht alle Zusammenhänge notwendigerweise kausal (Black, 1999). Doch nähern wir uns der Frage der Kausalität im Zusammenhang mit Korrelationen und Regressionen. Auch wenn kausale Beziehungen zwischen Variablen oft mithilfe von Zusammenhangsmaßen beschrieben werden (Black, 1999), sagt ein hoher Zusammenhang, auch wenn er noch so stark ist, für sich genommen noch nichts über eine etwaige Kausalität aus. Korrelation ist kein Beweis für Kausalität (Shadish et al., 2002). Ein Zusammenhang ist zwar eine Voraussetzung, liefert aber nicht ausreichend Informationen, um eine Aussage über einen kausalen Zusammenhang treffen zu können. In anderen Worten, „eine Korrelation zwischen zwei Variablen ist eine notwendige, aber keine hinreichende Voraussetzung für kausale Abhängigkeiten“ (Bortz & Schuster, 2010, S. 160). Im Rahmen der Korrelation kann die Frage nach der Kausalität zwischen zwei Variablen x und y folgendermaßen beantwortet werden: a) x wirkt kausal auf y ; b) y wirkt kausal auf x ; c) eine oder mehrere weitere Variablen wirken kausal auf x und y ; d) x und y wirken wechselseitig kausal aufeinander (Bortz & Schuster, 2010). Im Fall der Regression sind die Möglichkeiten beschränkter. Per Design wäre es der Prädiktor x , welcher auf y kausal wirkt. Aber auch im Fall der Regression ist die Frage der Kausalität ohne zusätzliche Informationen nicht hinreichend zu beantworten. Nehmen wir ein einfaches Beispiel, des Verhältnisses zwischen dem Krähen eines Hahnes und des Sonnenaufgangs (Beispiel in Anlehnung an Pearl & Mackenzie (2019)). Würde man eine Korrelation zwischen dem Krähen des Hahnes am Morgen und dem Sonnenaufgang rechnen, bekäme man einen ein perfekten (positiven) Zusammenhang. Auch in einer Regressionsrechnung würde das Krähen einen perfekten Prädiktor für das Kriterium des Sonnenaufgangs darstellen. Dennoch ist, plausibler Weise, das Krähen nicht ursächlich für den Sonnenaufgang.

Wann kann also von einer kausalen Beziehung zwischen Ursache und Wirkung gesprochen werden? Um es klassisch nach John Stuart Mill zu formulieren, ist dies der Fall, wenn: a) die Ursache vor der Wirkung auftritt; b) die Ursache mit der Wirkung zusammenhängt; c) sich keine andere – plausible – Erklärung für die Wirkung außer der Ursache finden lässt (Shadish et al., 2002). Die Klärung dieser Fragen geschieht in einem Experiment; es wird: a) die Realität so gestaltet, dass die Ursache vor der Wirkung auftritt; b) untersucht ob sich Änderungen in der Ursache in Veränderungen in der Wirkung niederschlagen; c) versucht die Plausibilität anderer Erklärungen für die Wirkung zu reduzieren (Shadish et al., 2002).

Kommen wir auf das Beispiel des Hahnkrähens und des Sonnenaufgangs zurück. Nehmen wir an man würde ein Experiment durchführen, in welchem Hähne zufällig in eine Versuchs- und eine Kontrollgruppe zugeordnet werden: die Versuchsgruppe wird dem Sonnenaufgang ausgesetzt, die Kontrollgruppe befindet sich in einem abgedunkelten Raum. Anhand dieses Designs könnte festgestellt werden, dass der Sonnenaufgang ursächlich für das Krähen am Morgen ist: a) die Ursache (Sonnenaufgang) tritt vor der Wirkung (Krähen) auf; b) die Änderungen in der Ursache (Sonnenaufgang / kein Sonnenaufgang) schlagen sich in der Wirkung (Krähen / kein Krähen) nieder; c) durch das Design

des Experiments (randomisierte Gruppen) lassen sich andere plausible Erklärungen – weitestgehend – ausschließen.

Konklusion

Abschließend ist zu zusammenzufassen, dass der Einsatz multivariater Verfahren und insbesondere der multiplen linearen Regression eine Kenntnis der Funktionsweise des Verfahrens verlangt, da die Gefahr falscher Rückschlüsse recht hoch ist. Außerdem ist eine intensive Planungsphase vorausgesetzt, die der theoriegestützten Formulierung eines Modells, sowie dessen Operationalisierung gewidmet ist. Der vorliegende Artikel soll einen Beitrag dazu leisten, auf die Wichtigkeit des sorgsamen und bewussten Einsatzes hinzuweisen und für die möglichen Fehler in der Ergebnisinterpretation zu sensibilisieren. Allerdings ist damit keineswegs Abschreckung intendiert, dieses Verfahren einzusetzen, vielmehr soll ein Grundverständnis des Verfahrens vermittelt werden, um die multiple lineare Regression sachgerecht anzuwenden.

Literaturverzeichnis

- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2018). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (15. Aufl.). Wiesbaden: Springer.
- Backhaus, K., Erichson, B., & Weiber, R. (2015). *Fortgeschrittene Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (3. Aufl.). Berlin, Heidelberg: Springer Gabler.
- Black, T. R. (1999). *Doing Quantitative Research in the Social Sciences. An Integrated Approach to Research Design, Measurement and Statistics*. London: Sage.
- Bortz, J. (2006). *Statistik: Für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- Bortz, J., & Schuster, C. (2010). *Statistik: für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- Bravais, A. (1846). Analyse mathématique sur les probabilités des erreurs de situation de point. *Mémoires présentés par divers savants à l'Académie des Sciences de l'Institut de France*, 9, 255–332.
- Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York: Academic Press.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden* (2. Aufl.). Weinheim Basel: Beltz.
- Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.) London: Sage.
- Haupt, H., Lösel, F., & Stemmler, M. (2014). Quantile Regression Analysis and Other Alternatives to Ordinary Least Squares Regression. *Methodology*, 10(3), 81–91. <https://doi.org/10.1027/1614-2241/a000077>
- Kossakowski, J. J., Waldorp, L. J., & van der Maas, H. L. J. (2021). The search for causality: A comparison of different techniques for causal inference graphs. *Psychological Methods*, 26(6), 719–742. <https://doi.org/10.1037/met0000390>
- Nübling, R., Schulz, H., Schmidt, J., Koch, U. & Wittmann, W. W. (2005). Fragebogen zur Psychotherapiemotivation (FPTM) – Testkonstruktion und Gütekriterien. In R. Nübling, F. A. Muthny & J. Bengel (Hrsg.), *Reha-Motivation und Behandlungserwartung* (S. 252-270). Bern: Huber.
- Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys* 3, 96–146. <https://doi.org/10.1214/09-SS057>

- Pearl, J. (2010). An Introduction to Causal Inference. *The International Journal of Biostatistics*, 6(2).
<https://doi.org/10.2202/1557-4679.1203>
- Pearl, J., & Mackenzie, D. (2019). *The Book of Why: The New Science of Cause and Effect*. New York: Penguin Books.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton, Mifflin and Company.
- Wentura, D., & Pospeschill, M. (2015). *Multivariate Datenanalyse: Eine kompakte Einführung*. Wiesbaden: Springer.
- Wilmers, F., Munder, T., Leonhart, R., Herzog, T., Plassmann, R., Barth, J., & Linster, H. W. (2008). Die deutschsprachige Version des Working Alliance Inventory-short revised (WAI-SR)-Ein schulübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz. *Klinische Diagnostik und Evaluation*, 1(3), 343–358.

Angaben zu den Autor*innen

Oliver Wegenberger, Jan Aden

Institut für Statistik
Fakultät Psychologie
Sigmund Freud PrivatUniversität Wien
Freudplatz 1, Raum 6011, 6. Stock
1020 Wien

E-Mail: oliver.wegenberger@mail.sfu.ac.at, jan.aden@sfu.ac.at