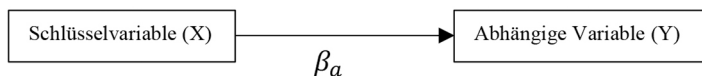


## Online-Anhang: Die Karlson-Holm-Breen-Methode (KHB-Methode)

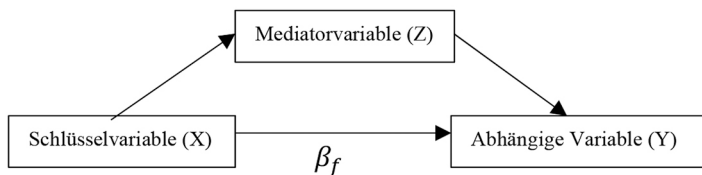
Das Ziel der Mediationsanalyse ist es, abzuschätzen, wie viel von der Stärke des Zusammenhangs zwischen einer abhängigen Variable und einer unabhängigen Variable (in diesem Zusammenhang *Schlüsselvariable* genannt) durch eine andere unabhängige Variable (in diesem Zusammenhang *Mediatorvariable* genannt) vermittelt wird bzw. durch diese im mathematischen Sinne erklärt werden kann. Genauer gesagt wird der vollständige Effekt der Schlüsselvariable auf die abhängige Variable zerlegt in einen indirekten Effekt (der Teil des vollständigen Effekts, der durch die Mediatorvariable vermittelt wird) und in einen direkten Effekt (der Teil der vollständigen Effekts, der nach Einbezug der Mediatorvariable übrig bleibt, also von dieser nicht vermittelt wird).

**Abbildung 1:** Logik der Mediationsanalyse

Ausgangsmodell (bzw. reduziertes Modell):



Vollständiges Modell:



(Quelle: Kohler 2011, leicht abgewandelt.)

Die Effekte werden auf folgende Weise ermittelt (die folgenden Ausführungen basieren auf Kohler 2011 und Kohler et al. 2011: 420-422):

- In einem ersten Schritt wird ein Regressionsmodell mit der Schlüsselvariable, aber ohne die Mediatorvariable geschätzt. Dieses Modell ist das *Ausgangsmodell* (siehe oberer Teil von Abbildung 1 und Gleichung A1).
- Danach wird ein weiteres Modell geschätzt, das zusätzlich die Mediatorvariable enthält. Dieses Modell wird als *vollständiges Modell* bezeichnet (siehe unterer Teil von Abbildung 1 und Gleichung A2).
- Im letzten Schritt wird die Differenz zwischen den Regressionsgewichten der Schlüsselvariable in beiden Regressionsmodellen ermittelt (siehe Gleichung A3). Diese Differenz ist der Effekt der Schlüsselvariable auf die abhängige Variable,

der durch die Mediatorvariable vermittelt wird (indirekter Effekt), während es sich bei dem Regressionsgewicht der Schlüsselvariable im vollständigen Modell um den direkten Effekt handelt. Bezieht man den indirekten Effekt auf das Regressionsgewicht der Schlüsselvariable im Ausgangsmodell erhält man den Anteil des Zusammenhangs zwischen Schlüsselvariable und abhängiger Variable, der durch die Mediatorvariable vermittelt wird bzw. durch diese im mathematischen Sinn erklärt werden kann.

$$Y = \alpha_a + \beta_a X + \epsilon \quad (\text{Gleichung A1})$$

$$Y = \alpha_f + \beta_f X + \gamma_f Z + \epsilon \quad (\text{Gleichung A2})$$

$$\beta_i = \beta_a - \beta_f \quad (\text{Gleichung A3})$$

Das eben beschriebene Verfahren der Mediationsanalyse gilt für lineare Regressionen. In der vorliegenden Studie wird die Mediationsanalyse jedoch nicht mit linearer Regression, sondern mit binärer logistischer Regression durchgeführt, da die abhängige Variable nicht metrisch, sondern dichotom ist. Bei logistischen Regressionen ändert sich das Verfahren zur Schätzung des indirekten Effekts, da sich durch das Hinzufügen einer (bedeutsamen) Mediatorvariable nicht nur der Effekt der Schlüsselvariable auf die abhängige Variable verändert, sondern auch die abhängige Variable neu skaliert wird (Kohler et al. 2011). Um dies besser zu verstehen, ist es notwendig, die Ausprägungen der abhängigen Variable eines binären logistischen Regressionsmodells als die Realisierung einer zugrundeliegenden latenten kontinuierlichen Variable ( $Y^*$ ) zu betrachten (vgl. Windzio 2013: 69-71.). Während in der linearen Regression das Hinzufügen einer (bedeutsamen) unabhängigen Variable die Varianz des Fehlerterms ( $\epsilon$ ) reduziert, ist die Situation in logistischen Regressionen anders. Da die Varianz des Fehlerterms in der logistischen Regression auf  $\pi^2/3$  fixiert ist, führt das Hinzufügen einer (bedeutsamen) unabhängigen Variable nicht zu einer Verringerung der Varianz des Fehlerterms, sondern zu einer Reskalierung der abhängigen Variable (Long/Freese 2014: 238; Windzio 2013: 69-71). Dieser Umstand macht es schwierig, mit der oben geschilderten Weise, die korrekte Größe des indirekten Effekts zu schätzen.

Eine Lösung für dieses Problem bietet die KHB-Methode (Karlson et al. 2012). Die Grundidee dieser Methode ist es, „aus [einer Mediatorvariable] die Informationen zu extrahieren, die nicht in [der Schlüsselvariablen] enthalten sind“ (Kohler et al. 2011: 423, eigene Übersetzung). Diese Information wird dann als (zusätzliche) unabhängige Variable im Ausgangsmodell verwendet. Dadurch erklären beide Modelle (Ausgangsmodell und volles Modell) den gleichen Anteil an der Variation der abhängigen Variable und forcieren die abhängige Variable auf die gleiche Skala. Im Detail geht man in drei Schritten vor (die folgenden Ausführungen basieren auf Kohler 2011 und Kohler et al. 2011: 423-424):

- In einem ersten Schritt wird eine lineare Regression mit der Mediatorvariable als abhängiger Variable und der Schlüsselvariable als unabhängige Variable geschätzt (siehe Gleichung B1)
- In einem zweiten Schritt werden die Residuen für jede Beobachtung geschätzt, indem die vorhergesagten Werte der Schlüsselvariablen von den beobachteten Werten der Schlüsselvariablen subtrahiert werden (siehe Gleichung B2).
- In einem dritten Schritt werden diese Residuen ( $r$ ) als weitere unabhängige Variable in das Ausgangsmodell eingesetzt (siehe Gleichung B3).

$$\widehat{Z} = \phi + \psi X \quad (\text{Gleichung B1})$$

$$r = Z - \widehat{Z} \quad (\text{Gleichung B2})$$

$$Y^* = \alpha_a + \beta_a X + \gamma_a r + \epsilon \quad (\text{Gleichung B3})$$

$$Y^* = \alpha_f + \beta_f X + \gamma_f Z + \epsilon \quad (\text{Gleichung B4})$$

$$\beta_i = \beta_a - \beta_f \quad (\text{Gleichung B5})$$

Nun unterscheiden sich  $r$  in Gleichung B3 und  $Z$  in Gleichung B4 „nur in der Komponente, die mit  $[X]$  korreliert ist“, was bedeutet, dass „das vollständige Modell von [Gleichung B4] [...] nicht mehr Vorhersagekraft besitzt als [das Modell von Gleichung B3]“ (Kohler et al. 2011: 423, eigene Übersetzung). Hierdurch haben die abhängigen Variablen beider logistischen Regressionsmodelle die gleiche Skala und es ist möglich, den indirekten Effekt gemäß Gleichung B5 genau zu schätzen. Mit diesem Effekt lässt sich dann auch leicht ermitteln, wieviel Prozent des vollen Effekts der Schlüsselvariable (auf die abhängige Variable) durch die Mediatorvariable vermittelt ist bzw. rechnerisch durch die Mediatorvariable erklärt werden kann (siehe Gleichung B6):

$$\text{conf}\% = 100 \times \frac{\beta_a - \beta_f}{\beta_a} \quad (\text{Gleichung B6})$$

Die KHB-Methode kam jüngst in zahlreichen Forschungsfeldern zum Einsatz, bspw. in Studien zum Wahlverhalten (Tutic/Hermann 2018, Hertel/Esche 2019), in Untersuchungen zu gesundheitlicher Ungleichheit (Arpino/Solé-Auró 2019, Kending et al. 2017), in Analysen zu Bildungsungleichheit (Ianelli 2013, Karlson/Holm 2011) und in Studien zum Rentenübergang (Eisman et al. 2018, Zwier et al. 2020).

## Literatur

- Arpino, Bruno & Solé-Auró, Aida (2019): Education inequalities in health among older European men and women: the role of active aging. *Journal of aging and health* 31(1): 185-208.
- Eismann, Michelle, Henkens, Kène & Kalmijn, Matthijs (2018): Why singles prefer to retire later: The role of retirement anxiety and spousal pull. *Innovation in Aging* 2(Suppl 1): 1011.

- Hertel, Florian R. & Esche, Frederike (2019): Die rechte Mitte? Zur Rolle objektiver Position und subjektiver Verunsicherung für die Identifikation mit rechten Parteien. In: Christiane Lübke & Jan Delhey (Hg.), *Diagnose Angstgesellschaft? Was wir wirklich über die Gefühlslage der Menschen wissen*. Bielefeld: transcript, S. 255-286.
- Iannelli, Cristina (2013): The role of the school curriculum in social mobility. *British Journal of Sociology of Education* 34(5-6): 907-928.
- Karlson, Kristian Bernt & Holm, Anders (2011): Decomposing primary and secondary effects: A new decomposition method. *Research in Social Stratification and mobility* 29(2): 221-237.
- Karlson, Kristian Bernt, Holm, Anders & Breen, Richard (2012): Comparing regression coefficients between same-sample nested models using logit and probit. A new method. *Sociological Methodology* 42(1): 286-313.
- Kendig, Hal, Gong, Cathy H., Yiengprugsawan, Vasoontara, Silverstein, Merril & Nazroo, James (2017): Life course influences on later life health in China: childhood health exposure and socioeconomic mediators during adulthood. *SSM-population health* 3: 795-802.
- Kohler, Ulrich (2011): *Comparing coefficients of nested nonlinear probability models using khb*. Paper presented at 9th German Stata Users Group Meeting, Bamberg, 1.7.2011.
- Kohler, Ulrich, Holm, Anders & Karlson, Kristian Bernt (2011): Comparing coefficients of nested nonlinear probability models. *The Stata Journal* 11(3): 420-438.
- Long, Scott J. & Freese, Jeremy (2014): *Regression models for categorical dependent variables using Stata*. College Station: Stata Press.
- Tutić, Andreas & von Hermann, Hagen (2018): Sozioökonomischer Status, Deprivation und die Affinität zur AfD – Eine Forschungsnotiz. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 70(2): 275-294.
- Windzio, Michael (2013): *Regressionsmodelle für Zustände und Ereignisse*. Wiesbaden: Springer VS.
- Zwier, Dieuwke, Damman, Marleen & Van den Heuvel, Swenne G. (2020): Explaining differences in retirement timing preferences between the solo self-employed and employees. *International Journal of Manpower* 42(2): 286-304.