



Einführung in die Statistik

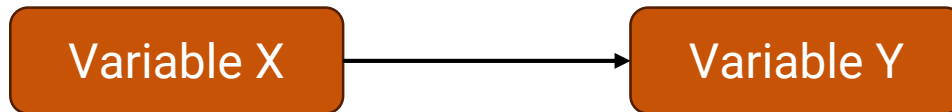
Moderations- und Mediationsanalyse

Überblick

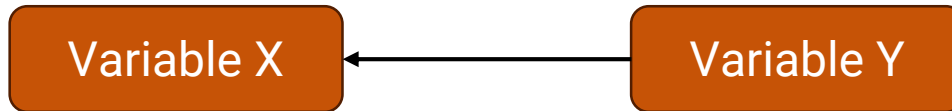
- Beziehungen zwischen Variablen
- Moderation
 - Statistisches Modell
 - Zentrierung
 - Regressionsgleichungen für kontinuierliche Moderatorvariable
 - Regressionsgleichungen für kategoriale Moderatorvariable
 - Signifikanzprüfung
 - Inferenzstatistische Voraussetzungen
- Mediation
 - Bootstrapping
 - Arten von Mediationen
 - Mediation als Regressionsmodell
 - Komplexe Modelle

Beziehungen zwischen Variablen (z. B. Döring, 2023)

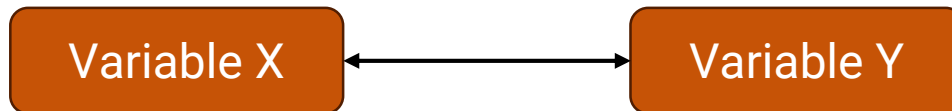
- **einfache Beziehungen zwischen zwei Variablen:** Richtung bzw. Kausalität der Beziehung ist relevant



X beeinflusst Y



Y beeinflusst X



X und Y beeinflussen
sich wechselseitig

- **Zur Erinnerung: Regressionen** setzen immer eine eindeutig Kausalrichtung zwischen Prädiktor (X) und Kriterium (Y) voraus

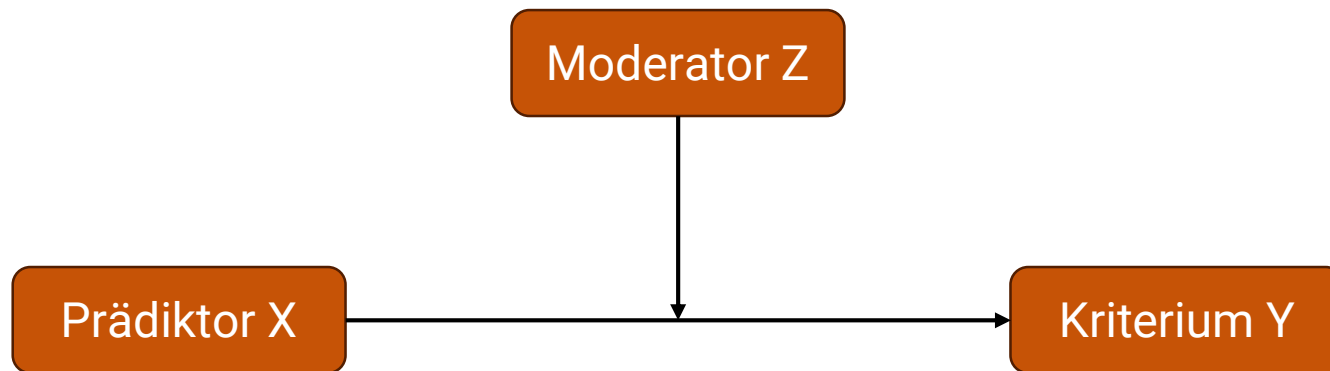
Beziehungen zwischen Variablen (z. B. Döring, 2023)

- **Einfluss dritter Variablen:** Beziehung zwischen zwei Variablen kann durch weitere Variablen (im einfachsten Fall: eine dritte) beeinflusst werden
- **wichtigste Varianten:** bei nur einer oder wenigen Einflussvariablen lassen sich einfache Modelle erstellen als
 - **Moderation**
 - **Mediation**
- **komplexere Modelle** mit mehreren Variablen und unterschiedlichen gegenseitigen Einflüssen werden typischerweise als Strukturgleichungsmodelle (Structural Equation Modeling, SEM) umgesetzt → kein Teil Ihrer Statistik-Ausbildung

Moderation

(z. B. Leonhart, 2009; Rudolf & Müller, 2012)

- **Moderation:** bezeichnet den Einfluss einer Moderatorvariable (kurz: Moderator), welche Stärke und/oder Richtung des Zusammenhangs zwischen zwei Variablen beeinflusst



- **Hypothesen** über Moderationseffekte spezifizieren den Einfluss der Beziehung zwischen Prädiktor und Kriterium bei unterschiedlichen Ausprägungen der Moderatorvariable bzw. Teilgruppen

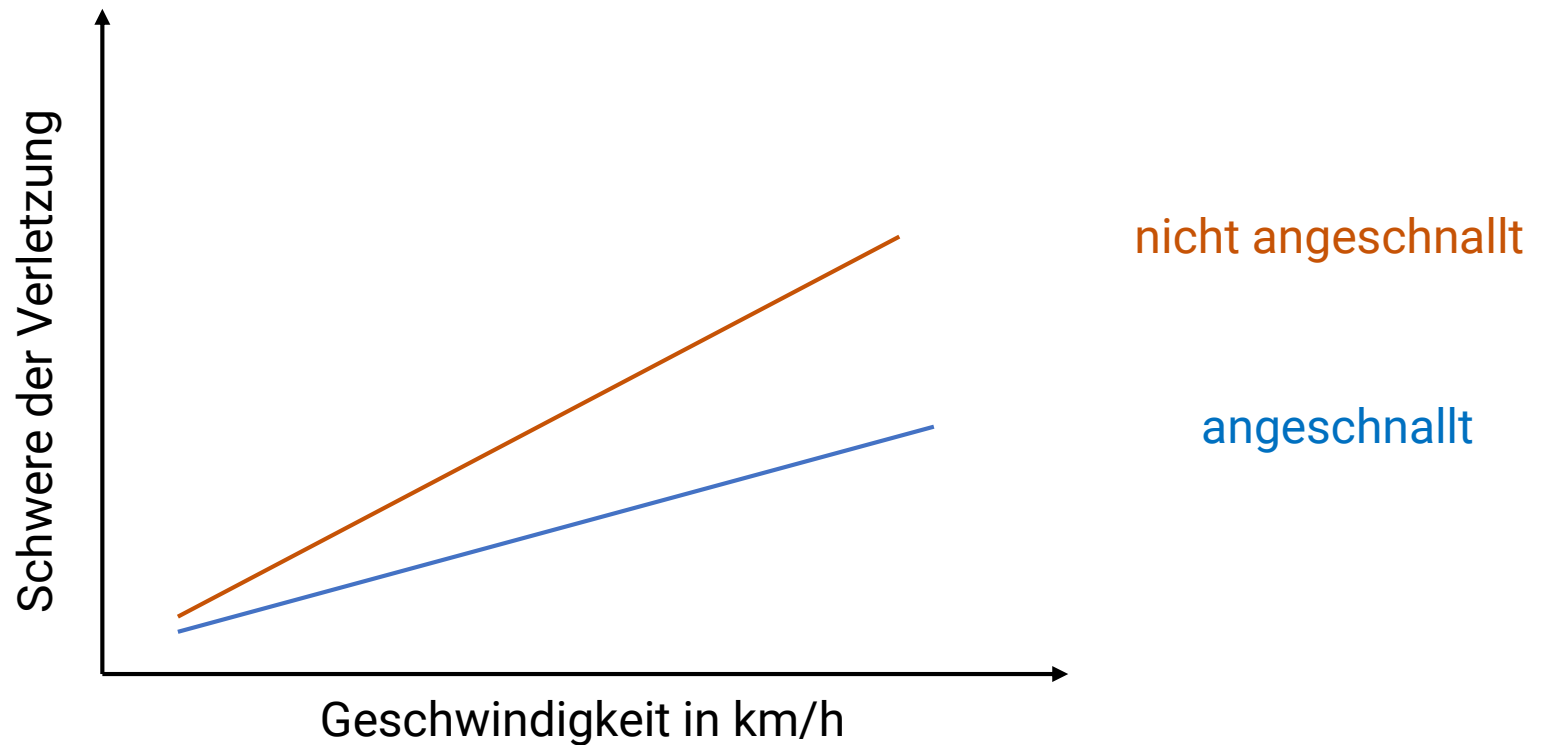
Moderation

(z. B. Leonhart, 2009; Rudolf & Müller, 2012)

- **Kategoriale Moderatorvariable:** unterscheidet zwischen Gruppen (mindestens zwei), bei denen der Zusammenhang zwischen Prädiktor und Kriterium je nach Gruppenzugehörigkeit (Moderator) jeweils anders ausgeprägt ist
- Beispiel: (in Anlehnung an Miles & Shevlin, 2001)
 - Prädiktorvariable: Geschwindigkeit des Fahrzeugs
 - Kriterium: Verletzungsgrad nach Unfall
 - Moderator: Anschnallverhalten mit den Ausprägungen angeschnallt vs. nicht angeschnallt
- signifikante Moderation liegt vor, wenn der Zusammenhang zwischen Kriterium und Prädiktor in der Gruppe „angeschnallt“ hinreichend von dem in der Gruppe „nicht angeschnallt“ abweicht

Moderation

- schematische Visualisierung



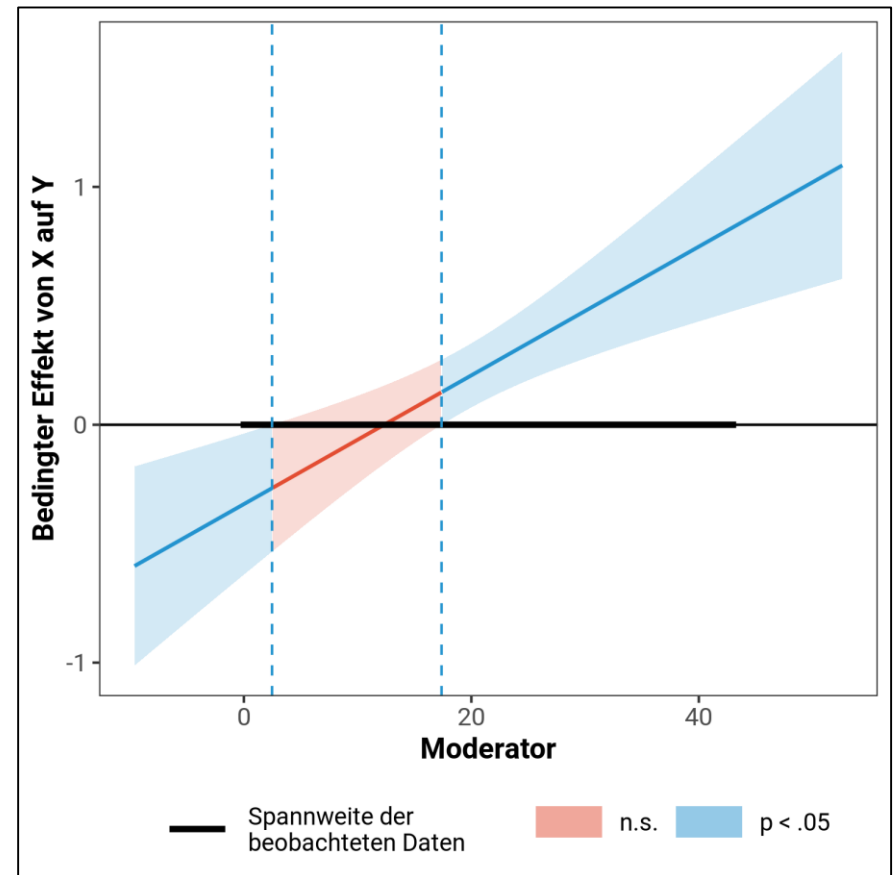
Moderation

(z. B. Leonhart, 2009; Rudolf & Müller, 2012)

- **kontinuierliche Moderatorvariable:** unterscheidet zwischen Wertebereichen (mindestens drei, häufig feiner aufgelöst) in der Moderatorvariable, bei denen der Zusammenhang zwischen Prädiktor und Kriterium jeweils anders ausgeprägt ist
- Beispiel:
 - Prädiktorvariable: tägliche Freizeit (in Minuten)
 - Kriterium: tägliche Social Media-Nutzung (in Minuten)
 - Moderator: Alter (in Jahren)
- signifikante Moderation (bzw. Interaktion) liegt vor, wenn der Zusammenhang zwischen Kriterium und Prädiktor in unterschiedlichen Bereichen der Moderatorvariable hinreichend verschieden aussieht

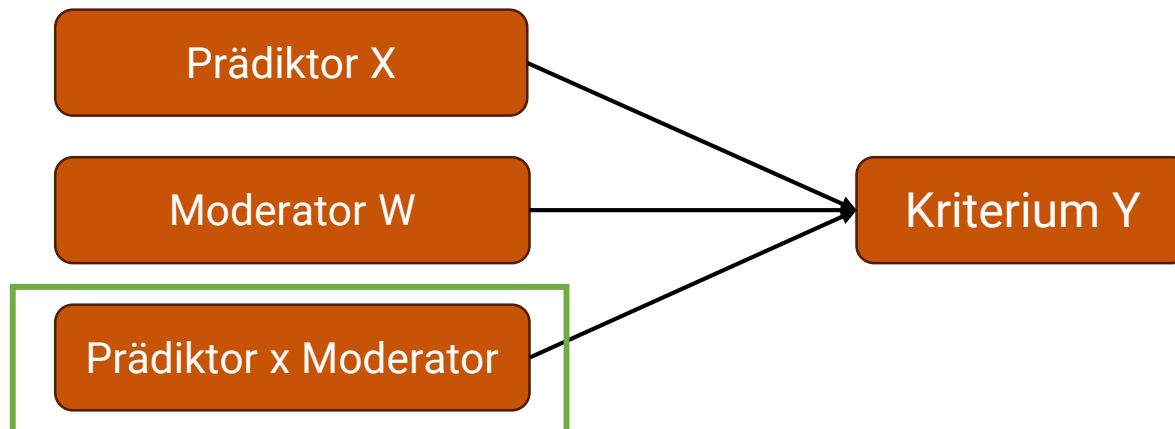
Moderation

- schematische Visualisierung: **Johnson-Neyman-Plot**
- zeigt die Wertebereiche, bei denen sich der Zusammenhang von Prädiktor und Kriterium signifikant durch den Moderator unterscheidet
- Visualisierung nicht in SPSS möglich, aber hier:
Hemmerich, W. (2020). StatistikGuru: Johnson-Neyman-Plots erstellen.
Retrieved from <https://statistikguru.de/rechner/johnson-neyman-plots-erstellen.html>
- SPSS-Notlösung: Vergleich von drei Slopes (für relevante Werte des Moderators) mit Mittelwert und +1 SD bzw. -1 SD



Moderation: Statistisches Modell (z. B. Field, 2018)

- **Modellierung** erfordert den Einbezug des Prädiktors und Moderators, auch wenn wir tatsächlich nur an deren Interaktion interessiert sind
 - Merke: mathematisch betrachtet ist eine Moderation eine Interaktion zwischen Variablen



- obwohl uns nur die Moderation (grüner Rahmen) interessiert, müssen wir die Haupteffekte mit in die Regressionsgleichung aufnehmen

Moderation: Statistisches Modell (z. B. Field, 2018)

- **Regressionsgleichung** der Moderation (kontinuierlicher Moderator):

$$\hat{y} = b_0 + b_1 \cdot X + b_2 \cdot W + b_3 \cdot (X \cdot W)$$

- Koeffizient b_3 (**Moderationseffekt**) wird nur dann korrekt geschätzt, wenn die Haupteffekte (b_1 und b_2) im Modell enthalten sind
- b_3 repräsentiert den inkrementellen Beitrag zur Erklärung der Varianz im Kriterium, nachdem der isolierte Einfluss von X und W bereits berücksichtigt wurde
- Aufnahme der Haupteffekte stellt sicher, dass man weiß, was die Haupteffekte bedeuten, und dass der Interaktionseffekt **rein als inkrementeller Beitrag** interpretiert werden kann.

\hat{y}	Vorhergesagte Kriteriumsvariable y
b_0	Achsenabschnitt
b_1	Steigung Prädiktor
X	Prädiktor
b_2	Steigung Moderator
W	Moderator
b_3	Steigung Moderation
XW	Moderationseffekt

Moderation: Zentrierung (z. B. Field, 2018)

- vor der eigentlichen Moderationsanalyse erfolgt typischerweise eine **Zentrierung um den Gesamtmittelwert** (Grand Mean Centring) bei **kontinuierlichen** Prädiktor und Moderator
- Berechnung: $Zentrierter\ Wert_i = Wert_i - \text{Arithmetisches Mittel}$
- Begründung für die Zentrierung:
 - **Interpretierbarkeit** der Messwerte: Wenn der Wert „0“ in der ursprünglichen Skala keine realistische oder sinnvolle Ausprägung darstellt (z. B. Alter, Puls), sind die Koeffizienten b_1 und b_2 nicht interpretierbar → bei Zentrierung wird z. B. b_1 interpretiert als der Effekt von X auf Y, wenn der Moderator W seinen durchschnittlichen Wert annimmt
 - Reduzierung der **Multikollinearität**: nicht der Hauptgrund, aber positiver Nebeneffekt
- typischerweise erledigen **Statistikprogramme** die Zentrierung automatisch (z. B. das in der Übung verwendete PROCESS-Makro)

Moderation: Statistisches Modell

- **Regressionsgleichung** der Moderation (kategorialer Moderator, zwei Ausprägungen z. B. Placebo vs. Wirkstoff):

$$\hat{y} = b_0 + b_1 \cdot X + b_2 \cdot W + b_3 \cdot (X \cdot W)$$

- Moderator:
 - $W = 0$ (Placebo) | $W = 1$ (Wirkstoff)
- Interpretation je Gruppe: für $W = 0$
 - $\hat{y}_{W=0} = b_0 + b_1 \cdot X$
- Interpretation je Gruppe: für $W = 1$
 - $\hat{y}_{W=1} = (b_0 + b_2) + (b_1 + b_3) \cdot X$
- Interpretation
 - b_2 ist der Mittelwertsunterschied der Gruppen bei $X=0$
 - b_3 ist der Unterschied der Steigungen zwischen den Gruppen

b_0	Achsenabschnitt
b_1	Steigung

$b_0 + b_2$	Achsenabschnitt
$b_1 + b_3$	Steigung

Moderation: Statistisches Modell

- **Regressionsgleichung** der Moderation (kategorialer Moderator, drei Ausprägungen z. B. Vergleich von drei Studiengängen) erfordert eine ...
- **Dummy-Codierung** (auch: Indikator-Codierung) der kategorialen Variable zur statistischen Verarbeitung (Umwandlung in binäre Form)
- Anzahl der benötigten Dummy-Variablen: $k - 1$, bei uns also $3 - 1 = 2$
- Codierung der Dummy-Variablen D_1 und D_2 bei Gruppen A, B, C:

Gruppe	D_1	D_2
A	0	0
B	1	0
C	0	1

- Gruppe A ist dabei die Referenzgruppe, mit der B und C verglichen werden (Vergleich von B und C erfordert entsprechende Umcodierung)

Moderation: Statistisches Modell

- **Regressionsgleichung** der Moderation (kategorialer Moderator, drei Ausprägungen z. B. Vergleich von drei Studiengängen) mit Dummy-Var.

$$\hat{y} = b_0 + b_1 \cdot X + b_2 \cdot D_1 + b_3 \cdot D_2 + b_4 \cdot (X \cdot D_1) + b_5 \cdot (X \cdot D_2)$$

- Gruppenspezifische Gleichung A: ($D_1 = 0; D_2 = 0$)
 - A: $\hat{y}_A = b_0 + b_1 \cdot X$
- Gruppenspezifische Gleichung B: ($D_1 = 1; D_2 = 0$)
 - $\hat{y}_B = (b_0 + b_2) + (b_1 + b_4) \cdot X$
- Gruppenspezifische Gleichung C: ($D_1 = 0; D_2 = 1$)
 - $\hat{y}_C = (b_0 + b_3) + (b_1 + b_5) \cdot X$

b_2	Unterschied des Achsenabschnitts von Gruppe B gegenüber Gruppe A
b_3	Unterschied des Achsenabschnitts von Gruppe C gegenüber Gruppe A
b_4	Unterschied der Steigung von Gruppe B im Vergleich zu Gruppe A
b_5	Unterschied der Steigung von Gruppe C im Vergleich zu Gruppe A

Signifikanzprüfung des Moderationseffekts

- generelles Vorgehen der Signifikanzprüfung ist **unabhängig** davon, ob Moderator kontinuierlich oder kategorial ist
- **Logik der Signifikanzprüfung** auf Modellebene oder Parameterebene möglich → identisches Ergebnis bei einfacher Moderation
- **Modellebene**: Prüfung, ob der Interaktionsterm der Moderation die Modellgüte signifikant verbessert
 - Berechnung ΔR^2 (Vergleich Modell 1: nur X und W mit Modell 2: X, W und $X*W$) und Prüfung via inkrementellen F-Test
 - wenn ΔR^2 signifikant ist, liegt eine Moderation vor
- **Parameterebene**: Prüfung, ob der Interaktionsterm selbst signifikant ist (also hinreichend unterschiedlich von 0)
 - t-Test des Regressionskoeffizienten der Moderation
 - wenn das Ergebnis signifikant ist, liegt eine Moderation vor
- beide Verfahren inhaltlich identisch, geläufiger ist Test auf **Modellebene**

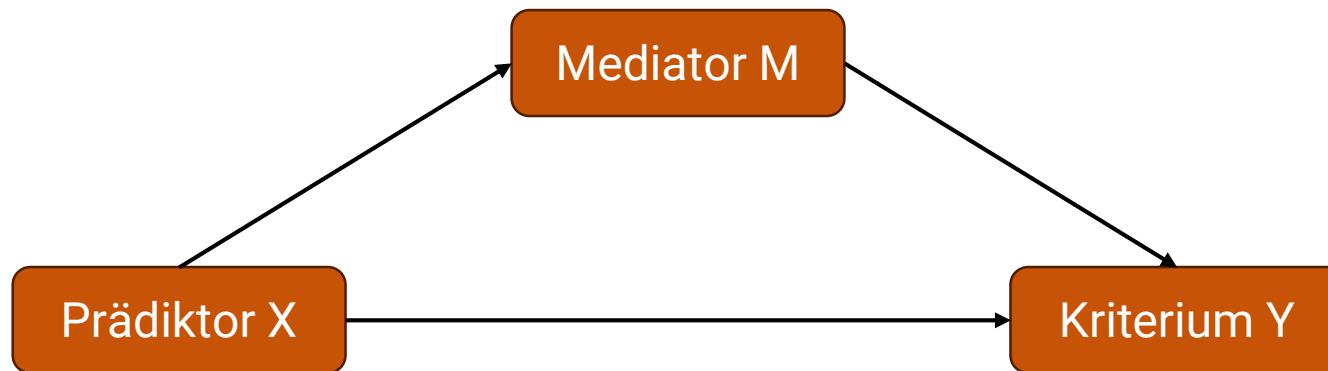
Inferenzstatistische Voraussetzungen

- es gelten exakt die gleichen Voraussetzungen wie bei der **Multiplen Linearen Regression** (siehe vergangene Sitzung)
- zu beachten sind folgende Besonderheiten
 - Achtung bei der VIF-Diagnostik: Moderationsmodelle haben immer erhöhte VIF-Werte → Verbesserung durch Zentrierung (siehe oben)
 - Interaktionsterme sind besonders empfindlich bei Ausreißerwerten bzw. Fällen mit großen Einfluss auf das Modell, daher Kennwerte prüfen für:
 - Cook-Distanz
 - Hebelwerte
 - Mahalanobis-Distanz
 - diese drei Kennwerte gibt das PROCESS-Makro optional mit aus (siehe Übung)

Mediation

(z. B. Leonhart, 2009; Rudolf & Müller, 2012)

- bei einer Mediation gibt es einen indirekten Einfluss des Prädiktors (X) auf das Kriterium (Y) in der Form, dass der Prädiktor auf die **Mediatorvariable** (M) wirkt und diese wiederum auf das Kriterium

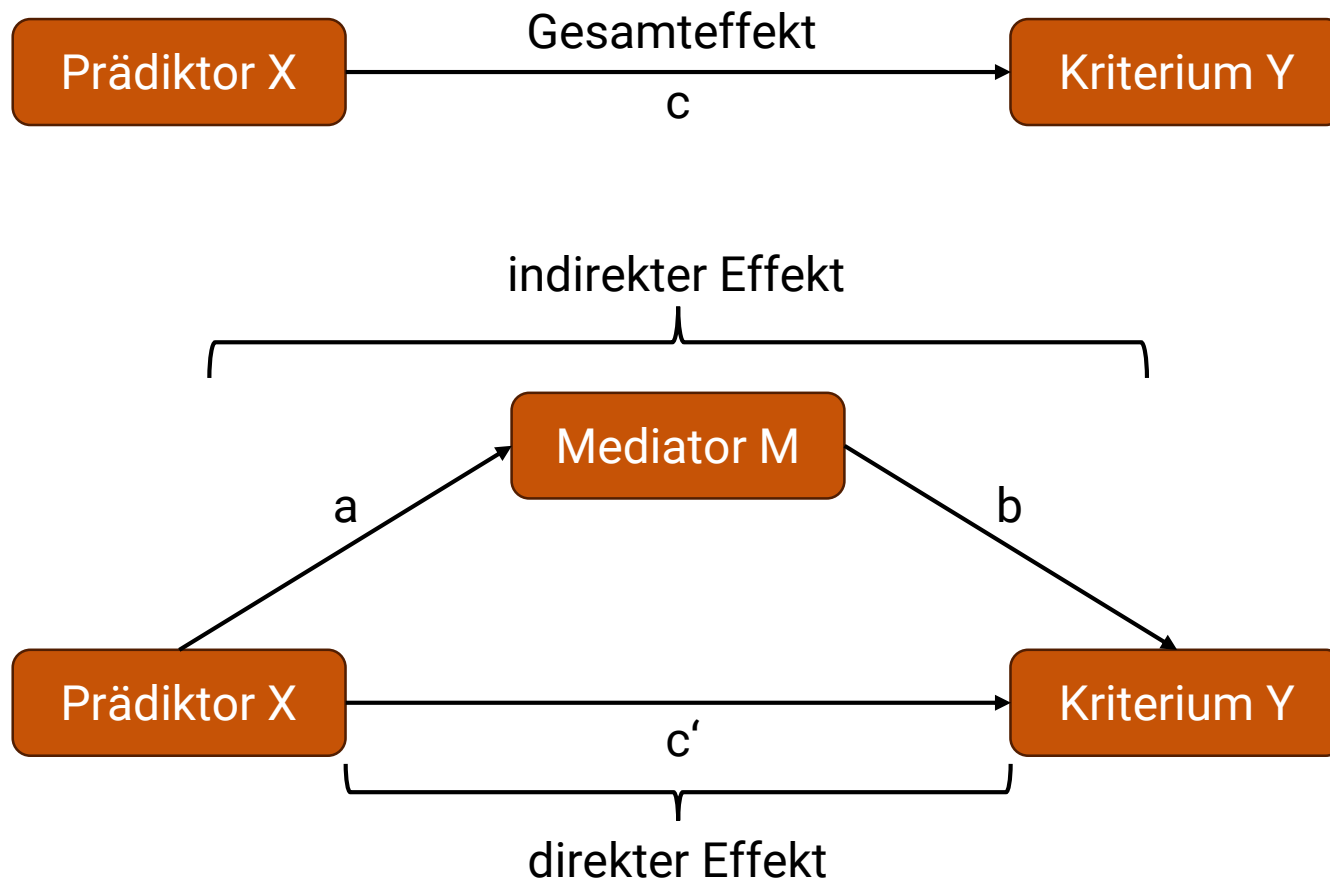


- **Hypothesen** über Mediationseffekte spezifizieren Mediatorvariablen als relevante Bindeglieder in der Kausalkette zwischen Ursache (Prädiktor) und Wirkung (Kriterium)

Mediation

(z. B. Leonhart, 2009; Rudolf & Müller, 2012)

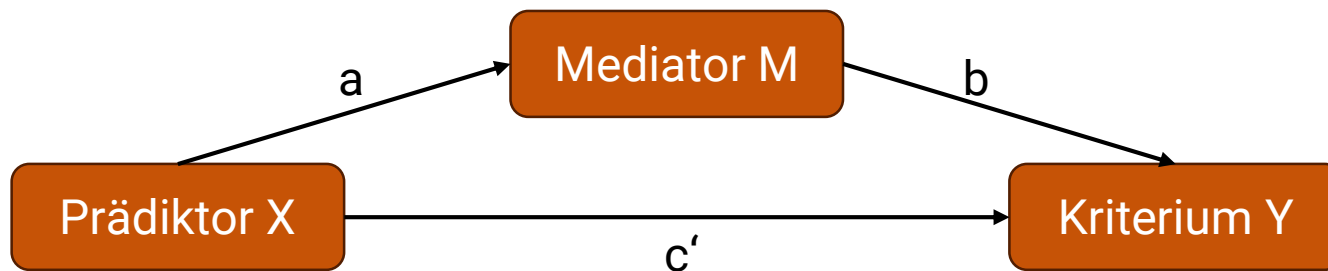
- Unterscheidung der verschiedenen **Wirkpfade** im Mediationsmodell



Mediation

(Baron & Kenny 1986)

- ursprünglicher Ansatz der Mediation nach Baron und Kenny (1986)
 - Test hat geringe Power und gilt inzwischen als veraltet
- Wenn der Gesamteffekt (c) signifikant ist, dann wird überprüft, ob ...
- ... der Prädiktor signifikant auf den Mediator wirkt (a)
- ... ob der Mediator signifikant auf das Kriterium wirkt (b)
- ... ob sich der Effekt von Prädiktor auf Kriterium (c') signifikant verkleinert hat im Vergleich zu vorher (c).
- Wenn diese Kriterien alle erfüllt sind, dann wurde der signifikante Einfluss der Mediatorvariable als nachgewiesen betrachtet.



Mediation

- moderner Ansatz der Mediation: Fokussierung auf den **indirekten Effekt** (Pfade a und b insgesamt, also $a \cdot b$)
- **Mediationseffekt** ($a \cdot b$) wird mittels Bootstrapping (siehe nächste Folie) ermittelt und gilt als signifikant, wenn das **Konfidenzintervall (CI)** des indirekten Effekts den **Wert 0** nicht enthält (meist 95%-Bootstrapping-CI)
- **Einzelpfade** a und b werden berechnet, müssen aber nicht signifikant sein → hauptsächlich Verwendung zur Prüfung der Richtung der Kausalkette (konsistente vs. inkonsistente Mediation, siehe unten)
- Signifikanz des **direkten Effekts** (c') ist irrelevant für die Frage, ob eine Mediation vorliegt → drückt aus, wie viel des Zusammenhangs zwischen X und Y nicht durch die Mediation bzw. Mediatorvariable erklärt wird

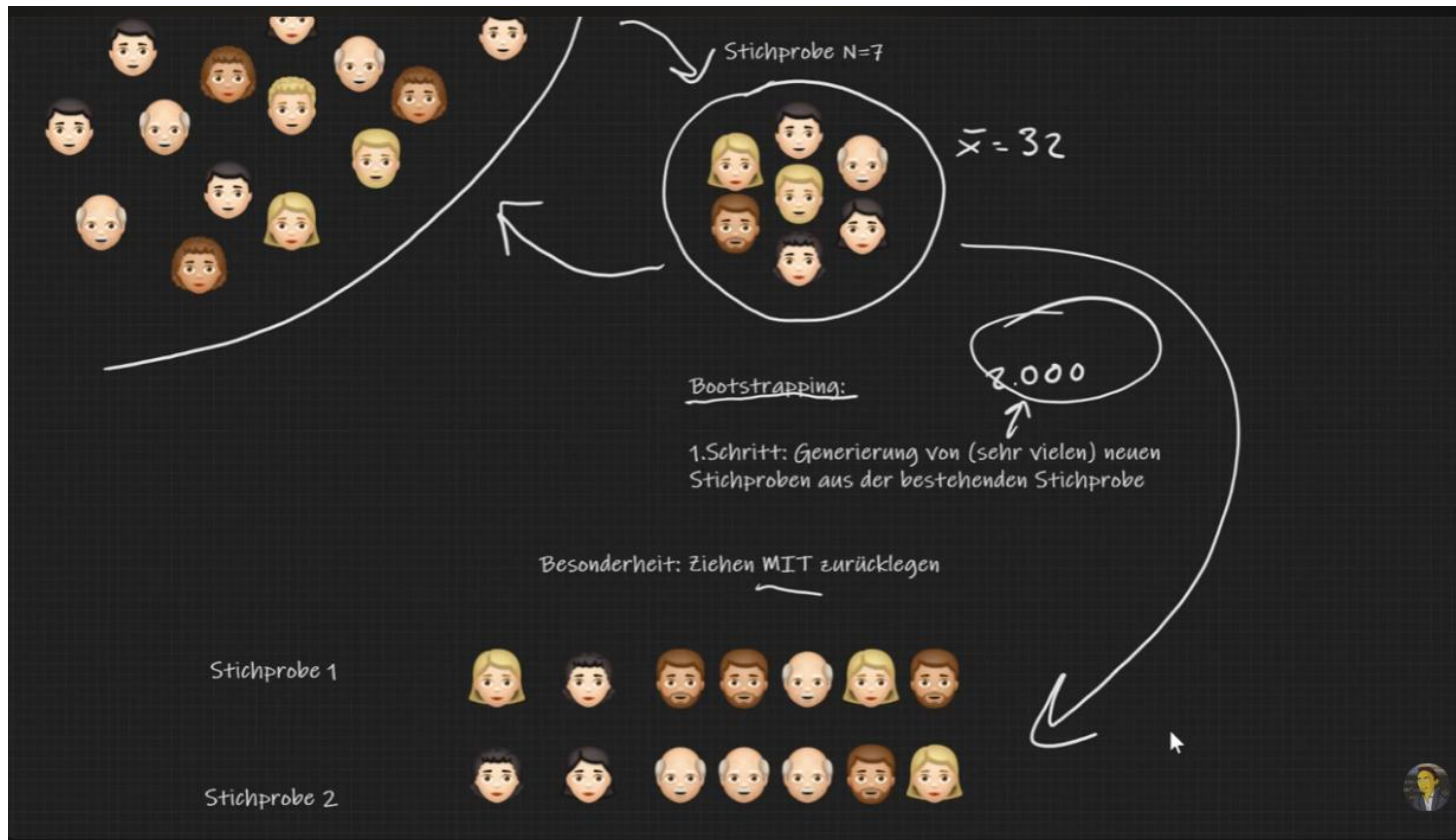
Mediation: Bootstrapping

(z. B. Bortz & Schuster, 2010)

- **Bootstrapping:** computergestützte Methode, um die Genauigkeit einer Statistik zu schätzen, in dem man den Wert durch simulierte Stichprobenziehungen auf Basis der Originalstichprobe vielfach ermittelt
- Vorgehen:
 - zufälliges Ermitteln einer Bootstrap-Stichprobe durch Ziehen (mit Zurücklegen) von Werten aus dem Originaldatensatz
 - Berechnung des gesuchten Wertes (z. B. $a \cdot b$) in der Bootstrap-Stichprobe
 - Wiederholung dieser Prozedur (meist 5.000 oder 10.000 Mal)
 - Ermittlung des Konfidenzintervalls auf Basis der ermittelten Werte
- Generelle Anwendung
 - Der Sinn des Bootstrappings ist es, die Verteilung einer Statistik zu schätzen, wenn man keine theoretischen Annahmen über deren Form machen kann oder die Schätzung unabhängig von einer theoretischen Annahme erfolgen soll.

Mediation: Bootstrapping

- **Bootstrapping:** Video zur Erklärung des generellen Vorgehens

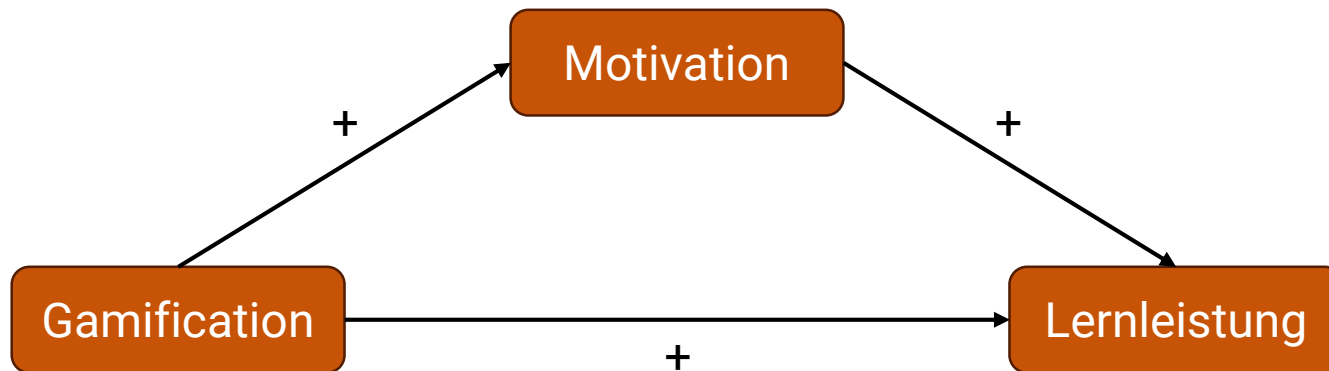


<https://www.youtube.com/watch?v=JU5m-aPrNJw>

Mediation: Arten von Mediationseffekten

Konsistente Mediation

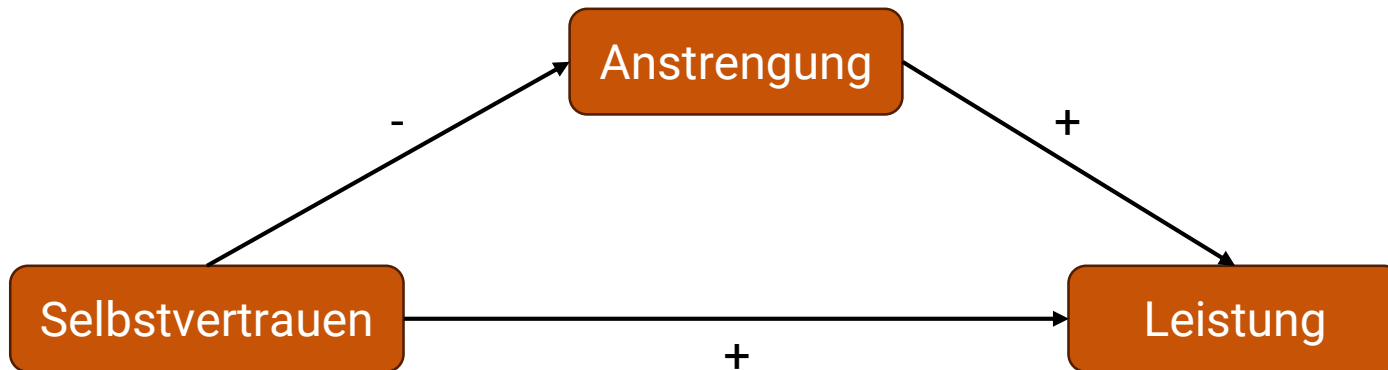
- Die Pfade a ($X \rightarrow M$) und b ($M \rightarrow Y$) haben dieselbe Richtung (z. B. beide positiv oder beide negativ).
- Ergebnis: Der indirekte Effekt ($a \cdot b$) und der direkte Effekt (c') verlaufen in dieselbe Richtung \rightarrow Mediator verstärkt oder erklärt den Gesamteffekt



Mediation: Arten von Mediationseffekten

Inkonsistente Mediation

- Die Pfade a ($X \rightarrow M$) und b ($M \rightarrow Y$) haben gegensätzliche Richtung (z. B. a positiv und b negativ).
- Ergebnis: Der indirekte Effekt ($a \cdot b$) und der direkte Effekt (c') verlaufen in entgegengesetzte Richtungen \rightarrow Mediator unterdrückt oder maskiert einen Teil des Gesamteffekts



Mediation: Größe des Mediationseffekts (z. B. Field, 2018)

- Größe der Mediation $a \cdot b$ ist abhängig vom Modell und den Skalen → keine allgemein hin anerkannten Effektgrößen

Übliche Varianten

- **Proportion der Mediation** (P_M):
$$P_M = \frac{\text{Indirekter Effekt } (a \cdot b)}{\text{Gesamteffekt } (c)}$$
 - bei konsistenter Mediation liegt Wert zwischen 0 und 1, bei $P_M = 0,40$ heißt das, dass 40% des Gesamteffekts von X auf Y über den Mediator läuft
 - **Achtung:** Bei inkonsistenten Mediationen (Suppressor-Effekt) kann der Wert größer als 1 werden → hier ist die Berechnung nicht sinnvoll
- **z-Standardisierung** der Variablen vor der Berechnung
 - ermöglicht eine Vergleichbarkeit der Effektgrößen mit den Ergebnissen anderer Untersuchungen und dadurch Bewertung
 - Interpretierbarkeit der z-standardisierten ist aber geringer, daher werden typischerweise beide Werte berichtet

Mediation als Regressionmodell (z. B. Field, 2018)

- Mediationsmodell besteht aus **zwei Regressionsgleichungen**

- Modell für den Mediator M:

$$M = i_M + aX$$

i_M Achsenabschnitt in der Mediatorregression

a Effekt von X auf M

- Modell für das Kriterium Y:

$$Y = i_Y + c'X + bM$$

i_Y Achsenabschnitt in der Y-Regression

c' direkter Einfluss von X auf Y

b Effekt des Mediators auf Y

- Y berechnet sich also als Multiple Regression mit den Prädiktoren X und M, wobei der Pfad a (Effekt von X auf M) durch den Mediator abgedeckt ist
- Substitution von M in Gleichung 2 durch Gleichung 1 ermöglicht Darstellung als Gesamtmodell (siehe nächste Folie)

Mediation als Regressionsmodell

$$Y = i_Y + c'X + bM \quad \xrightarrow{\quad M = i_M + aX \quad} \quad \text{Einsetzen}$$

$$Y = i_Y + c'X + b(i_M + aX) \quad \text{Ausmultiplizieren und Zusammenfassen}$$

$$Y = (i_Y + bi_M) + (c' + ab)X$$

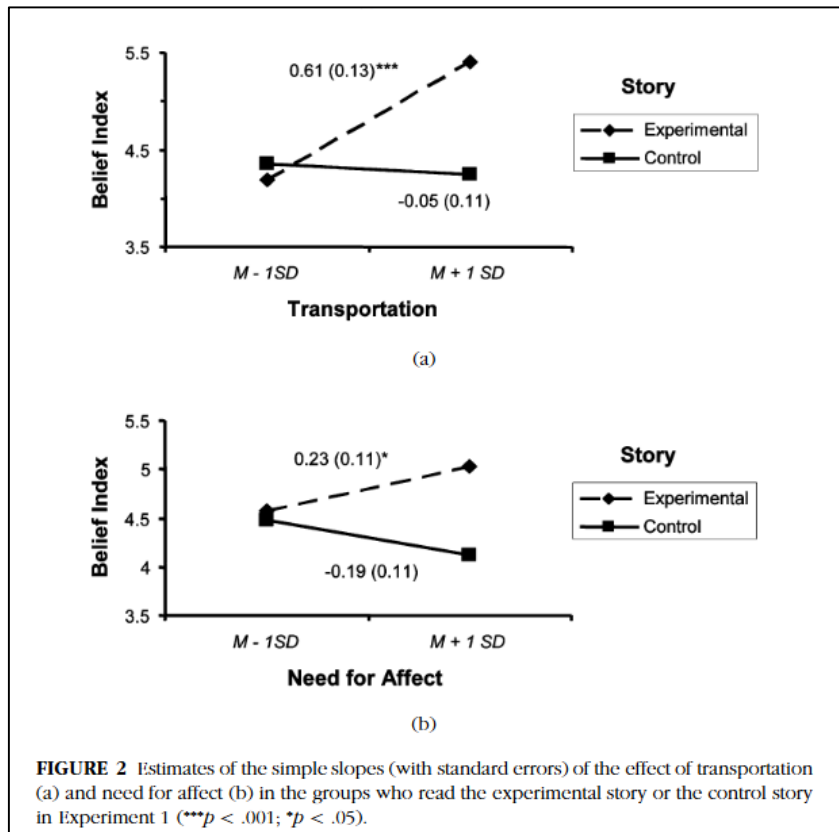
- Gesamtgleichung als Kombination der beiden Einzelregressionen
- gut erkennbar: Gesamteffekt entspricht der Summe aus direktem Effekt (c') und indirektem Effekt (ab)
- Struktur entspricht (wenig überraschend) unserer bekannten Form der allgemeinen Gleichung für lineare Modelle:

$$Y = \text{Achsenabschnitt} + \text{Koeffizient} \cdot X$$

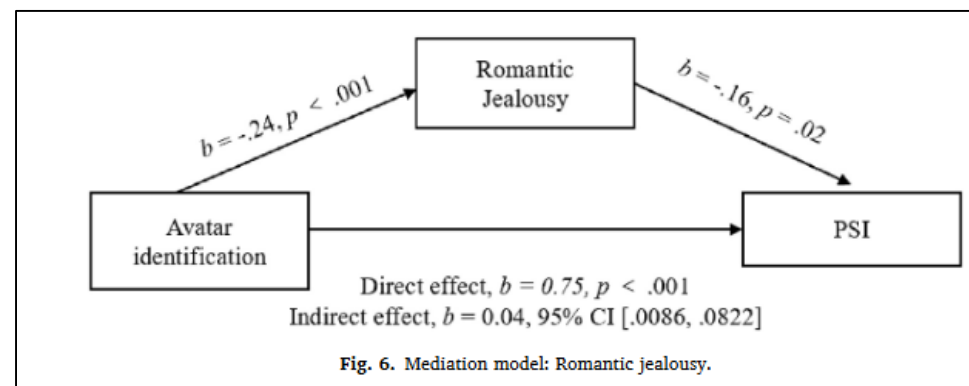
Komplexere Modelle (z. B. Bortz & Schuster 2010)

- Beziehungen zwischen Variablen sind in der Praxis teils deutlich komplexer → unterschiedliche **Varianten/Kombinationen** von Mediationen und Moderationen
- Beispiele:
 - mehrere Mediator-Variablen, die unterschiedlich wirken können (z. B. parallel oder sequenziell)
 - Mehrere Moderatorvariablen
 - Mediationen, bei denen Moderatorvariablen auf die Mediationspfade wirken (einzeln oder beide)
- einfache Modelle dieser Art lassen sich mithilfe des PROCESS-Makro in SPSS berechnen (siehe Übung)
- komplexe Variablenstrukturen erfordern andere Verfahren, typischerweise Strukturgleichungsmodelle und entsprechende Software wie die SPSS-Erweiterung AMOS

Beispiele für Moderation bzw. Mediation Fachzeitschriften



Beispiel für den Einfluss eines kategorialen Moderators | Quelle: Appel & Richter (2010)



Hypothesis 1b predicted that romantic jealousy would mediate the relationship between avatar identification and PSI. To test **H1b**, the mediating paths for PSI were analyzed (Fig. 6). There was a direct association between avatar identification and jealousy, ($\beta = -.24, p < .001$), as well as between avatar identification and PSI ($\beta = 0.79, p < .001$). However, when romantic jealousy was included in the model, the association between identification and PSI was reduced ($\beta = 0.75, p < .001$), and the indirect effect was 0.0392. It met our expectation that if players generated jealousy towards the female avatar, it might undermine PSI. Therefore, **H1b** was supported.

Beispiel für eine Mediation
Quelle: Hua & Xiao (2023)

Zusammenfassung I

- **Moderation:** Dient der Ermittlung des Einflusses (Richtung/Stärke) einer Moderatorvariable auf die Beziehung zwischen Prädiktor und Kriterium
- **Ausprägungen der Moderatorvariable** kann kategorial sein oder kontinuierlich
- **Effekt der Moderation** als Interaktionsterm zwischen Prädiktor und Moderator
- **Vorliegen einer Moderation** wenn inkrementeller Beitrag des Moderationsterms signifikante Modellverbesserung bewirkt
- **Follow-Up-Betrachtung** durch den Vergleich der Anstiege der Regressionsgeraden (Slopes) für Gruppen oder der Johnson-Neyman-Bereiche
- **Zentrierung** der kontinuierlichen Prädiktor-/Moderatorvariablen

Zusammenfassung II

- **Mediation** meint den indirekten Einfluss der Prädiktorvariable auf die Kriteriumsvariable über eine Mediatorvariable
- **Moderner Ansatz** betrachtet den indirekten Pfad über den Mediator ($a \cdot b$) unabhängig vom direkten Pfad (c')
- **Vorliegen einer Mediation**, wenn das Konfidenzintervall des indirekten Effekts nicht den Wert 0 enthält, also von 0 abweicht
- **Bootstrapping** dient der Berechnung des Konfidenzintervalls
- **Stärke der Mediation** meist ausgedrückt über die Proportion der Mediation (Verhältnis des indirekten zum Gesamteffekt)
- **Mediationsmodell** setzt sich aus zwei einzelnen Regressionen zusammen: Mediationsregression und Y-Regression
- **Strukturgleichungsmodelle** zur Berechnung komplexer Variablenkonfigurationen

Prüfungsliteratur

- Leonhart, R. (2009). *Lehrbuch Statistik. Einstieg und Vertiefung* (2. Auflage). Bern: Huber.
 - Mediatoranalyse und Moderatoranalyse (S. 321–325)

Weiterführende Literatur I

- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
<https://doi.org/10.1037/0022-3514.51.6.1173>
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer.
 - Partielle Korrelation und multiple lineare Regression (S. 339–361)
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6. Aufl.). Springer.
 - Moderator- und Mediatoreffekte (S. 683)
- Hayes, A.F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (3rd Edition). Guilford Press.

Weiterführende Literatur II

- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th Edition). Sage.
 - Moderation, Mediation and Multicategory Predictors (S. 481–518)
- Miles & Shevlin (2002). *Applying Regression & Correlation. A Guide for Students and Researchers*. Sage.
 - Moderator & Mediator Analysis (S. 165–191)
- Rudolf, M, & Müller, J. (2012). *Multivariate Verfahren* (2. Auflage). Hogrefe.
 - Moderator- und Mediatoranalyse (S. 60–68)