Datenanalyse mit R

Fortgeschrittene Verfahren

Markus Burkhardt
Johannes Titz
Peter Sedlmeier

Pearson

EXTRAS ONLINE

Datenanalyse mit R

Fortgeschrittene Verfahren

Markus Burkhardt
Johannes Titz
Peter Sedlmeier

Pearson

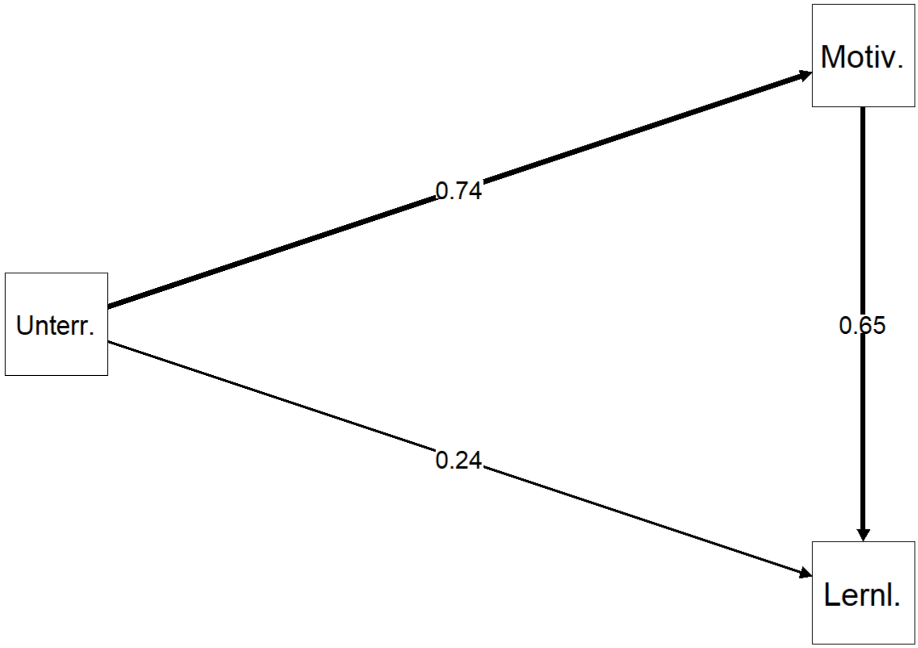

Abbildung 4.3: Pfadmodell mit standardisierten Pfadkoeffizienten einer Mediationsanalyse.

4.3 Signifikanztest für Pfadkoeffizienten

Das Paket *lavaan* bietet für die Koeffizienten als Teststatistik den z-Wert (z-*value*) des Wald-Tests auf Basis der Standardnormalverteilung. Der z-Wert ergibt sich aus dem Quotienten von standardisierten Pfadkoeffizienten und Standardfehler. Die Nullhypothese des Tests lautet, dass der Pfadkoeffizient den Wert 0 besitzt. Bei der herkömmlichen Regressionsrechnung werden die Koeffizienten mit einem t-Test (bei gleicher Nullhypothese) geprüft. Bei großen Stichproben geht der t-Test auch in den z-Test über.

Die Ausgabe der Pfadanalyse aus *Abschnitt 4.2* zeigt (bei einer Irrtumswahrscheinlichkeit von 5 %), dass der Zusammenhang zwischen Unterrichtsgüte und Lernleistung nicht von null abweicht. Die anderen Pfadkoeffizienten hingegen zeigen eine signifikante Abweichung von null, was auf eine Mediation schließen lässt. Allerdings besitzt der Wald-Test (multivariate) Normalverteilungsannahmen, die nur bei sehr großen Stichproben überprüft werden können. Daher ist das Ergebnis der Signifikanztests oft mit besonderer Vorsicht zu interpretieren. Im *Kapitel 5* (*Abschnitt 5.5.3*) diskutieren wir zwei Möglichkeiten zur Überprüfung dieser Voraussetzung.

Im *Kapitel 3* haben Sie mit dem Bootstrapping bereits eine Alternative zur „herkömmlichen“ Bestimmung von Standardfehlern kennengelernt. Auch diese Variante ist in *lavaan* implementiert. Wir nutzen in der Funktion `sem` das Argument `se = "boot"`, um den Standardfehler mit Bootstrapping zu bestimmen. Der anschließende Wald-

Test nutzt diesen Standardfehler für die Berechnung des z-Werts. Mit dem Argument `bootstrap = 1000` geben wir die Anzahl der Wiederholungen an.

```
set.seed(42)

LL_med2 <- sem(LL_med_model, data = Lehr_Lern, se = "boot", bootstrap = 1000)

summary(LL_med2, std = T)
```

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  Lernleistung ~
    Motivation        0.242    0.055    4.428    0.000    0.242    0.647
    Unterrichtsgut    0.456    0.285    1.604    0.109    0.456    0.242
  Motivation ~
    Unterrichtsgut    3.741    0.515    7.262    0.000    3.741    0.744
```

Im Unterschied zu den Standardeinstellungen erhalten wir nun veränderte Standardfehler und dementsprechend ändern sich auch die p-Werte. Allerdings sind die Unterschiede bei unserer Stichprobe relativ klein, was auch daran liegt, dass die Voraussetzungen in dem Beispiel erfüllt sind. Der Vorteil ist aber, dass mit dem Bootstrapping die Verteilungsannahmen nicht notwendigerweise erfüllt sein müssen. Wenn Sie diese Analyse mehrfach durchführen, werden Sie immer leicht unterschiedliche p-Werte erhalten, schließlich spielt der Zufall eine Rolle. Sollte man nun so lange probieren, bis der gewünschte p-Wert herauskommt? Natürlich nicht! Der Signifikanztest sollte für die Interpretation nur eine untergeordnete Rolle spielen, denn die Pfadkoeffizienten sind nach wie vor unverändert, und diese sind schließlich das relevante Ergebnis Ihrer Analyse[4].

Alternativ zum Signifikanztest empfehlen wir, die Konfidenzintervalle der Pfadkoeffizienten zu bestimmen. So können wir uns auf die Effekte konzentrieren, erhalten einen Eindruck von der Genauigkeit der Schätzung und können nebenbei auch die Signifikanz testen. Auch hier empfehlen wir, den Standardfehler mit `bootstrap` zu ermitteln, wie wir es bereits oben beschrieben haben. Anschließend können wir mit der Funktion `parameterestimates` die Konfidenzintervalle mit Bootstrapping und der Perzentilmethode (`boot.ci.type = "perc"`) ermitteln. Außerdem geben wir mit dem Argument `rsquare = TRUE` die Güte (Ausmaß der aufgeklärten Varianz) der zwei (Teil-) Regressionsgleichungen an (siehe die letzten beiden Zeilen der Ausgabe). Um die Ausgabe übersichtlicher zu gestalten, verzichten wir auf die Ausgabe der Signifikanz mit `zstat = FALSE`.

```
parameterEstimates(LL_med2, boot.ci.type = "perc", rsquare = TRUE,
                   zstat = FALSE)
```

```
           lhs op             rhs   est    se ci.lower ci.upper
1 Lernleistung  ~      Motivation 0.242 0.055    0.132    0.357
2 Lernleistung  ~ Unterrichtsguete 0.456 0.285   -0.139    0.995
3   Motivation  ~ Unterrichtsguete 3.741 0.515    2.623    4.681
```

4 Ein weiteres Problem ist das multiple Testen und die damit verbundene α-Fehler-Kumulierung. Wir haben dies Thematik in Sedlmeier und Burkhardt (2021) besprochen. Bei sehr komplexen Modellen stellt diese Problematik eine Einschränkung für die Interpretation des Signifikanztests dar.

```
4     Lernleistung ~~     Lernleistung  2.307 0.533    1.153    3.262
5       Motivation ~~       Motivation 25.286 7.427   11.049   39.154
6 Unterrichtsguete ~~ Unterrichtsguete  2.245 0.000    2.245    2.245
7     Lernleistung r2     Lernleistung  0.710    NA       NA       NA
8       Motivation r2       Motivation  0.554    NA       NA       NA
```

Wir erhalten in der Ausgabe die unstandardisierten Koeffizienten (*est*). Dabei wird in jeder Zeile der Pfadkoeffizient dargestellt. Die 4. bis 6. Zeile enthält die Varianzen der Residualwerte (beim jeweiligen Kriterium) und die Zeilen 7 und 8 den Determinationskoeffizienten (R^2). In den Spalten *ci.lower* und *ci.upper* sehen Sie die Grenzen des 95 %-Konfidenzintervalls. Hierbei lässt sich leicht ablesen, ob es sich um einen signifikanten Effekt handelt. Dazu prüfen wir, ob der Wert 0 von dem Intervall überdeckt wird.[5] Wenn Sie eine gerichtete Hypothese testen wollen, können Sie das Konfidenzintervall mit dem Argument `level` anpassen.

In manchen Fällen wird man an den standardisierten Konfidenzintervallen interessiert sein. Eine pragmatische Möglichkeit stellt die *z*-Standardisierung aller Variablen des Datensatzes vor der Pfadanalyse dar. Anschließend kann die von uns beschriebene Prozedur genutzt werden.

```
set.seed(43)

z_Lehr_Lern <- as.data.frame(scale(Lehr_Lern))

zLL_med_sem <- sem(LL_med_model, data = z_Lehr_Lern, se = "boot",
                   bootstrap = 1000)

parameterEstimates(zLL_med_sem, boot.ci.type = "perc", zstat = FALSE)
```

```
                lhs op              rhs   est    se ci.lower ci.upper
1     Lernleistung  ~       Motivation 0.647 0.142    0.365    0.927
2     Lernleistung  ~ Unterrichtsguete 0.242 0.147   -0.072    0.503
3       Motivation  ~ Unterrichtsguete 0.744 0.106    0.513    0.936
4     Lernleistung ~~     Lernleistung 0.280 0.065    0.139    0.391
5       Motivation ~~       Motivation 0.431 0.131    0.184    0.682
6 Unterrichtsguete ~~ Unterrichtsguete 0.967 0.000    0.967    0.967
```

Die direkte Wirkung der Lernleistung auf die Unterrichtsgüte ist vergleichsweise gering, hingegen zeigt die Mediation deutliche Effekte. Steigt die Unterrichtsgüte um eine (standardisierte) Einheit, dann steigt die Lernleistung um 0.72 (standardisierte Einheiten; $a * b + c = 0.74 * 0.65 + 0.24$). Davon gehen 0.48 Einheiten ($a * b$) auf den indirekten Effekt durch die Motivation und 0.24 Einheiten auf den direkten Pfad (die Unterrichtsgüte) zurück.[6]

5 Mit den Argumenten `zstat = TRUE` und `pvalue = TRUE` können zwar die *p*-Werte berechnet werden, diese folgen allerdings der Logik und den Voraussetzungen des „herkömmlichen“ Signifikanztest. Daher ist die Bewertung der Konfidenzintervalle die beste Methode, da keinerlei Verteilungsvoraussetzungen nötig sind.

6 Die Berechnung der Konfidenzintervalle ist streng genommen nicht korrekt, da wir zu Beginn der Analyse die *z*-Standardisierung durchführen. Tatsächlich wären zunächst die Koeffizienten zu ermitteln und anschließend zu standardisieren. Der Unterschied ist allerdings vernachlässigbar.

4.4 Güte von Pfadmodellen

Eine der häufigsten Fragen im Zusammenhang mit der Pfadanalyse ist, ob eine Mediation vorliegt und wie diese beschrieben werden kann. Oft besteht der Wunsch, eine qualitative Aussage wie: „Es liegt keine / eine partielle / eine vollständige Mediation vor“. Leider können wir keine Regeln formulieren, welche die eine oder andere Aussage mit Sicherheit rechtfertigt. In der Literatur existiert eine Kontroverse darüber, was die geeignete Effektgröße sei, um eine Mediation zu bewerten, und das Vorhandensein von Kontroversen deutet schon darauf hin, dass es keinen Konsens zu einem Thema gibt. Eine Übersicht über Effektstärken für Mediatoranalysen geben Preacher und Kelly (2011) sowie Wen und Fan (2015). Einige Autoren empfehlen die Betrachtung des direkten Pfadkoeffizienten. Dieser ist in unserem Beispiel mit $\lambda_c = 0.24$ nicht signifikant und somit nicht verschieden von 0, was zum Teil als vollständige Mediation gedeutet wird, wenn gleichzeitig der Mediatorpfad signifikant wird. Alternativ kann auch der indirekte Pfad $\lambda_b = 0.70$ multipliziert mit $\lambda_a = 0.74$ auf Signifikanz getestet werden (Sobel-Test). Dafür nutzen wir die Funktion `mediation.test` aus dem Paket *bda* (Bin, 2021).

```
library(bda)

mediation.test(mv = Lehr_Lern$Motivation, iv = Lehr_Lern$Unterrichtsguete,
               dv = Lehr_Lern$Lernleistung)
```

```
            Sobel   Aroian  Goodman
z.value 3.405327 3.373165 3.438427
p.value 0.000661 0.000743 0.000585
```

Die Ausgabe zeigt drei verschiedene Versionen des Sobel-Tests (Watson, 2016). Alle drei Varianten zeigen ein signifikantes Ergebnis für den indirekten Pfad ($p < 0.05$), was als Vorhandensein einer Mediation gedeutet wird. Diese Art des Tests wird aufgrund der strengen Voraussetzungen kritisiert (Hayes, 2009). Bei sehr kleinen Stichproben ist dieses Vorgehen daher oft nicht empfehlenswert.

Eine Alternative bieten relative Maße, z. B. das Verhältnis zwischen indirektem und direktem Pfad $(0.74 * 0.65) / 0.24 = 2$. Der Effekt des indirekten Pfades ist also doppelt so groß wie der des direkten. Relative Maße haben allerdings den Nachteil, dass gleiche relative Effektstärken bei sehr unterschiedlich starken Einzel-Effekten auftreten können. Aus diesem Grund sind sie immer im Zusammenhang mit den vorliegenden Effekten zu bewerten.

Schließlich kann auch das gesamte Pfadmodell bewertet werden. Dazu wird das R^2 der separaten Regressionen herangezogen. Von der Lernleistung können insgesamt 71 % der Varianz durch die Prädiktoren Motivation und Unterrichtsgüte erklärt werden. Von der Motivation immerhin noch 55 %. Die Erklärungskraft des Beispiels ist damit sehr hoch. In der Forschungspraxis wird man kaum so hohe Determinationskoeffizienten finden. Grundsätzlich gilt aber immer, dass die Interpretation inhaltlich erfolgen sollte.

4.5 Ein komplexeres Datenbeispiel

Am Beispiel eines psychologischen Experiments zum Power-Posing demonstrieren wir eine komplexere Pfadanalyse. Der Datensatz *Power_Pose* enthält die entsprechenden Variablen. Power-Posing bezeichnet kurzgesagt Gesten oder Haltungen, die Erfolge (z. B. zum Himmel gestreckte Arme) oder Misserfolge (z. B. herunterhängende Schultern) ausdrücken. Laut Carney et al. (2010) beeinflusst die vorangegangene Körperhaltung einer Person deren nachfolgende Bewertung. In dem hier vorgestellten Experiment wurden Teams aus jeweils zwei Personen (gleichen Geschlechts) gebildet.[7] Diese Dyade sollte nun, eine gemeinsam zu lösende Aufgabe, erfüllen. Es sollte eine Brücke mit einer bestimmten Stabilität aus Bausteinen gebaut werden. Eine Person des Teams wurde zufällig ausgewählt und nahm vor der kooperativen Aufgabe entweder eine Siegerpose oder eine Misserfolgspose ein (*Pose*). Die andere Person hingegen nahm eine neutrale Pose ein, wobei sich die Personen nicht sehen konnten. Anschließend wurde die Brückenaufgabe gelöst und die Person in der neutralen Pose sollte die Führungsstärke (*Fuehrung*) der anderen Person beurteilen. In der Literatur wurde die körperliche Aktivität (*Aktivitaet*) als Wirkmechanismus postuliert (Personen mit Siegerpose sollten sich nachfolgend mehr bewegen, was wiederum die Führungsstärke ausstrahlen sollte). Die körperliche Aktivität wurde mit Aktivitätssensoren (in einem Sensorgurt) erfasst. Als Kontrollvariable wurde auch die selbst eingeschätzte Dominanz (*Dominanz*) im Vorfeld des Experiments erhoben. Diese könnte als Prädiktor und auch als Moderator auf die Aktivität wirken. Die Moderatorbeziehung bedeutet, dass bei Menschen mit ausgeprägter Dominanz die Wirkung der Pose auf die Aktivität verstärkt sein sollte. ▶*Abbildung 4.4* zeigt das dazugehörige Pfadmodell. Der Pfeil *e* kennzeichnet die moderierende Beziehung (zur Durchführung von einfachen Moderatoranalysen mit R siehe Sedlmeier und Burkhardt, 2021).

7 Das Experiment wurde im Rahmen einer Masterarbeit von Müller (2018) durchgeführt.

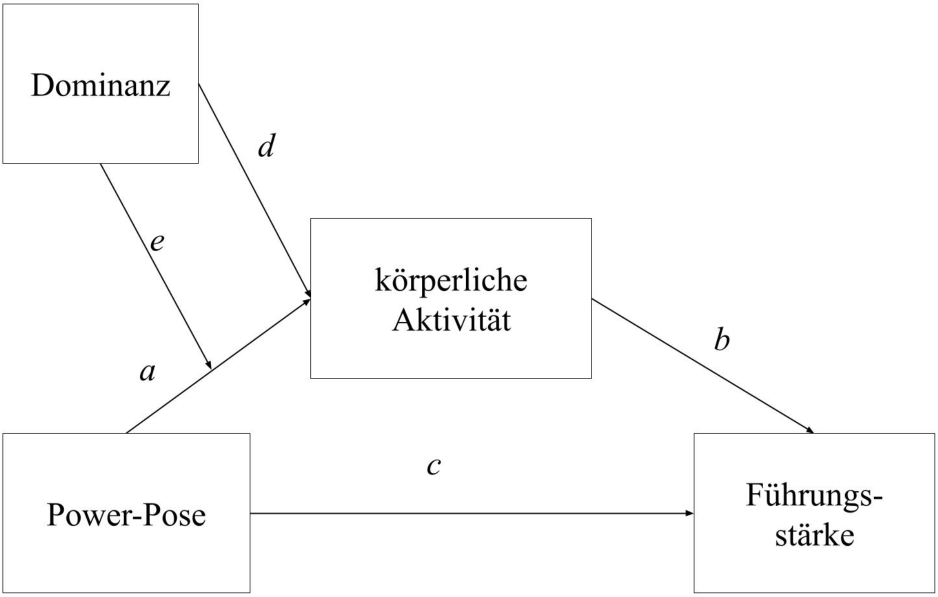

Abbildung 4.4: Pfadmodell zur Wirkung von Power-Posing auf die wahrgenommene Führungsstärke, vermittelt über die körperliche Aktivität (Mediator) unter Kontrolle der Dominanz und ihrer moderierenden Wirkung auf den Zusammenhang zwischen Power-Pose und Aktivität.

Da die Skalen in der ursprünglichen Skalierung mit extrem unterschiedlichen Varianzen vorliegen und die Statistik für Moderatoren sehr sensibel darauf reagiert, haben wir alle Variablen *z*-standardisiert (zur *z*-Standardisierung und Zentrierung siehe Kapitel 9 in Sedlmeier und Burkhardt, 2021). Der Moderator ist bereits im Datensatz vorhanden, kann aber auch durch die Multiplikation von *Pose* und *Dominanz* und anschließender Standardisierung erstellt werden.

Als Erstes verschaffen wir uns einen Überblick über die vorliegenden Daten und prüfen die Linearität (auch hier verweisen wir auf Sedlmeier und Burkhardt, 2021). Dabei fällt auf, dass hier lediglich eine sehr kleine Stichprobe vorliegt ($N = 16$), was an dem aufwendigen Studiendesign lag. Wir konzentrieren uns in der Analyse auf die Größe der standardisierten Effekte, wenngleich wir der Vollständigkeit halber die Signifikanztests ausgeben lassen. Das Pfadmodell aus ▶*Abbildung 4.4* kann entweder durch *Moderator* als die Variablen des Datensatzes oder alternativ als `Pose:Dominanz` in die Syntax eingefügt werden[8].

```
Model_PP <- " Fuehrung ~ Pose + Aktivitaet
             Aktivitaet ~ Pose + Dominanz + Moderator"

set.seed(44)

PA_PP <- sem(Model_PP, data = Power_Pose, se = "boot", test = "boot",
             bootstrap = 1000)
```

8 Tatsächlich erhalten Sie dabei leicht unterschiedliche Ergebnisse. Während unser Moderator im Datensatz nochmals *z*-standardisiert wurde, wird bei der Syntax mit `:` keine Standardisierung vorgenommen.

```
parameterEstimates(PA_PP, boot.ci.type = "perc")
```

```
         lhs op        rhs    est    se      z pvalue ci.lower ci.upper
1    Fuehrung  ~       Pose  0.095 0.526  0.180  0.857   -0.961    1.116
2    Fuehrung  ~ Aktivitaet  0.084 0.267  0.314  0.754   -0.473    0.597
3  Aktivitaet  ~       Pose -0.306 0.626 -0.489  0.625   -1.478    0.717
4  Aktivitaet  ~   Dominanz  0.440 0.489  0.899  0.369   -0.383    1.417
5  Aktivitaet  ~  Moderator -0.010 0.515 -0.019  0.985   -0.687    1.083
```

Wir sehen, dass weder die Pose noch die Aktivität einen Einfluss auf die Führungswahrnehmung haben. Auch die moderierende Wirkung der Dominanz ist so gut wie 0. Für die Pose lässt sich ein negativer Effekt (–0.31) finden. Wer zu Beginn eine Power-Pose durchführte, zeigte nachfolgend weniger Aktivität (was der Hypothese widerspricht). Einzig der Einfluss von Dominanz auf Aktivität mit einem Pfadgewicht von 0.44 zeigt einen hypothesenkonformen Koeffizienten. Die Konfidenzintervalle zeigen eine entsprechend große Breite, die der kleinen Stichprobe geschuldet ist. Vielleicht sind Sie darüber verwundert, dass einige Konfidenzintervalle Grenzwerte aufweisen, bei denen der standardisierte Pfadkoeffizient größer bzw. kleiner als eins ist. Tatsächlich können solche extremen Werte bei einer hohen Multikollinearität von Prädiktoren auftreten. Bei besonders kleinen Stichproben wie in unserem Fall sorgt der Zufall beim Bootstrapping dafür, dass auch Datensätze mit hoher Multikollinearität auftreten. In erster Linie ist dies ein Hinweis auf eine (zu) kleine Stichprobe. Sind nicht die Konfidenzintervallgrenzen, sondern die standardisierten Pfadgewichte größer bzw. kleiner als eins, ist gegebenenfalls eine der hoch korrelierenden Variablen aus dem Modell auszuschließen. Inhaltlich reiht sich dieses Ergebnis in die heterogene Befundlage zum Power-Posing ein (Elkjær et al., 2020). Betrachten wir schließlich noch die Güte der beiden Teilmodelle.

```
summary(PA_PP, rsquare = TRUE)
```

```
R-Square:
                   Estimate
    FW_fremd          0.008
    Aktiv             0.232
```

Die Führungswahrnehmung lässt sich in diesem Modell überhaupt nicht erklären. Die Varianz in der körperlichen Aktivität kann zumindest mit 23 % durch die verwendeten Prädiktoren erklärt werden.

Zusammenfassung

Pfadanalysen sind eine Generalisierung der Regressionsanalyse, bei der nicht nur direkte Beziehungen zwischen Kriterium und Prädiktoren, sondern auch beliebig komplexe und indirekte Beziehungen (über Mediatorvariablen) analysiert werden können. Solche komplexen Modelle kann man als Zusammenfassung mehrerer (Teil-) Regressionsanalysen zu einem einzigen Modell betrachten. Dies lässt sich gut in der Darstellung pfadanalytischer Gleichungen in dem von uns verwendeten Paket *lavaan* erkennen. Die Pfadanalyse ist ein konfirmatorisches Verfahren. Das heißt, dass die Kausalrichtungen theoretisch begründet werden müssen. Der entscheidende Vorteil gegenüber multiplen Regressionen ist die Modellierung indirekter Beziehungen zwischen den Variablen. Ähnlich wie bei der Regression lassen sich die Pfadkoeffizienten interpretieren und testen. Der Komplexität von Pfadanalysen sind dabei keine Grenzen gesetzt. Es können auch Moderatorbeziehungen und im Grunde beliebig viele Prädiktoren und Mediatoren einbezogen werden. Wenn Verteilungsannahmen verletzt sind, bietet es sich an, Bootstrapping einzusetzen. Typischerweise werden die Ergebnisse von Pfadanalysen grafisch dargestellt und sind dadurch gut kommunizierbar.

Die R-Funktionen in diesem Kapitel

In diesem Kasten stellen wir Ihnen die Funktionen, die wir in diesem Kapitel benutzt haben, noch einmal zusammen. Wenn Funktionen nicht schon vorinstalliert sind, werden sie zusammen mit dem Paket, aus dem sie stammen, aufgelistet (`Paket::Funktion`). Funktionen, die Sie bereits aus früheren Kapiteln kennen, sind hier weggelassen.

Funktion	Kurze Beschreibung
`base:scale`	*z*-Standardisierung
`bda::mediation.test`	Sobel-Test des indirekten Pfades
`lm.beta::lm.beta`	Erzeugt die standardisierten Regressionskoeffizienten β aus *lm*-Objekt
`lavaan::sem`	Berechnet alle Pfadkoeffizienten
`lavaan::parameterEstimates`	Tabellarische Ausgabe von Pfadkoeffizient, Signifikanztest und Konfidenzintervall
`semPlot::semPaths`	Zeichnet das Pfadmodell/Strukturgleichungsmodell mit Pfadgewichten

Ergänzende und weiterführende Literatur

Sedlmeier, P., & Burkhardt, M. (2021). *Datenanalyse mit R: Beschreiben, Explorieren, Schätzen und Testen.* Hallbergmoos: Pearson.

Eine Einführung in die Moderatoranalyse mit R

Sedlmeier, P. & Renkewitz, F. (2018). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (3. überarbeitete und erweiterte Auflage). Hallbergmoos: Pearson.

Eine leicht verständliche Einführung in die Pfadanalyse