

Kolumnentitel: INTERAKTIONEN VON METRISCHEN UND KATEGORIALEN
PRÄDIKTOREN

Wie analysiert man Interaktionen von metrischen und kategorialen Prädiktoren? Nicht mit

Median-Splits!

(How to analyze interactions of metric and categorical predictors: Not with median splits!)

Tobias Richter

Florida State University

zur Veröffentlichung in der *Zeitschrift für Medienpsychologie* angenommen

Zusammenfassung

In der Medienpsychologie werden Interaktionen metrischer und kategorialer Prädiktorvariablen häufig mit inadäquaten Verfahren wie Median-Splits oder der Bestimmung separater Korrelationen analysiert. Dieser Beitrag zeigt, wie sich solche Datensituationen methodisch angemessen mit Hilfe von linearen Modellen analysieren lassen. Das Grundprinzip der Schätzung, Prüfung und Interpretation von Interaktionen metrischer und kategorialer Prädiktoren in linearen Modellen wird erläutert und mit einem Datenbeispiel aus der Medienpsychologie illustriert. Abschließend werden weiterführende Hinweise gegeben, die für eine Anwendung des Verfahrens relevant sind.

Schlagwörter: Allgemeines Lineares Modell, Interaktion, kategoriale Variable, metrische Variable, Regressionsanalyse

Abstract

In media psychology, interactions of metric and categorical predictor variables are frequently analyzed with inadequate techniques such as median-splits or separate correlations. This paper demonstrates that linear models provide ways to analyze these interactions in an adequate manner. The basic principles of estimating, testing, and interpreting interactions of metric and categorical predictors in linear models are explained and illustrated by sample data. Finally, further methodological recommendations are given that are relevant for applications of this method.

Key words: Categorical variable, General Linear Model, interaction, metric variable, regression analysis

Wie analysiert man Interaktionen von metrischen und kategorialen Prädiktoren? Nicht mit
Median-Splits!

Interaktionshypothesen bilden den Kern einer Vielzahl von Erklärungsmodellen zu zentralen Themenbereichen der Medienpsychologie. Der Theorie des Stimmungsmanagements (mood management, Zillmann, 1988) zufolge moderiert beispielsweise die aktuelle Stimmung von Rezipient/inn/en die Auswirkung von Eigenschaften der konsumierten Unterhaltungsmedien auf das Rezeptionsverhalten und – erleben. In der Psychologie des Textverstehens herrscht Konsens darüber, dass das Textverständnis wesentlich auf Interaktionen von Textmerkmalen mit Lesefähigkeiten, dem bereichsspezifischen Vorwissen, Interessen oder Verarbeitungszielen von Rezipient/inn/en beruht (z.B. Graesser, Singer & Trabasso, 1994; Kintsch, 1988; Richter & Christmann, 2006). Einige der wichtigsten Theorien zur Wirkung von Mediengewalt beinhalten Interaktionsannahmen, um nur ein weiteres Beispiel zu nennen. So besagt die Kultivierungshypothese, dass die Wirkung von Gewaltdarstellungen im Fernsehen auf die Gewaltbereitschaft durch das Ausmaß des individuellen Fernsehkonsums moderiert wird (Gerbner, Gross, Morgan & Signorielli, 1994), und die Stimulationshypothese postuliert eine Interaktion eines affektiven Zustands (Aggression) mit einer medialen Gewaltdarstellung als Auslöser (Berkowitz & Green, 1966).

Entsprechend dem zentralen Stellenwert von Interaktionshypothesen in medienpsychologischen Theorien stellt sich in vielen medienpsychologischen Untersuchungen das Auswertungsproblem, dass das Zusammenwirken von kontinuierlichen (metrischen) und kategorialen (qualitativen) Prädiktorvariablen auf eine metrische abhängige Variable untersucht werden soll. Das verbreitetste Beispiel sind Experimente, in denen eine oder mehrere unabhängige Variablen experimentell manipuliert wurden, aber zugleich der Effekt einer oder mehrerer personenbezogener Variablen einschließlich ihrer Interaktion(en)

mit der oder den experimentellen unabhängigen Variablen bestimmt werden sollen. Diese personenbezogenen Variablen haben in den meisten Fällen ein metrisches Skalenniveau (z.B. Persönlichkeitseigenschaften, Einstellungen oder Fähigkeiten), während die experimentell manipulierten Variablen über ein kategoriales Skalenniveau verfügen (z.B. experimentell variierte Merkmale eines Medienprodukts). Zur Analyse derartiger Daten wird häufig auf die Methode der Gruppenbildung zurückgegriffen, bei der die metrische Variable in zwei gleich große Gruppen mit niedrigen bzw. höheren Werten (*Median-Split* oder *Dichotomisierung*) aufgeteilt wird. Im Ergebnis enthält das Auswertungsdesign nur noch kategoriale Variablen, so dass sich zur Analyse von Interaktionseffekten das übliche Instrumentarium der Varianzanalyse anwenden lässt. Eine andere häufig eingesetzte Methode besteht darin, für die unterschiedlichen Ausprägungen der kategorialen Variablen getrennte Korrelationskoeffizienten der metrischen Prädiktor- und der Kriteriumsvariablen zu ermitteln. Eine Durchsicht der letzten Jahrgänge der *Zeitschrift für Medienpsychologie* (von 2000, Heft 1, bis 2006, Heft 2) illustriert, dass diese Vorgehensweisen auch in der Medienpsychologie üblich sind: Von insgesamt 45 empirischen Beiträgen in diesem Zeitraum beruhen zehn auf Datensätzen mit gemischt metrischen und kategorialen Prädiktoren, und neun dieser Beiträge arbeiten mit der Methode der Dichotomisierung metrischer Prädiktoren oder der Methode der Bestimmung separater Korrelationen.

Trotz ihrer offensichtlichen Popularität sprechen gegen beide Methoden aber gewichtige methodologische Gründe. Bei der Methode der Gruppenbildung gehen durch die Transformation einer metrischen in eine kategoriale Variable weite Teile der ursprünglich enthaltenen Information verloren. Die Messgenauigkeit der ursprünglichen Variablen wird künstlich reduziert, mit den Folgen einer Reduktion der Varianz, einer Unterschätzung statistischer Zusammenhänge und einer häufig geringeren Teststärke inferenzstatistischer Tests, die auf Analysen mit gruppierten Variablen beruhen (Cohen, 1983; Maxwell & Delaney, 1993). Darüber hinaus ist die in der metrischen Ausgangsvariablen enthaltene

Information verzerrt, da durch die notwendig willkürlich getroffene Gruppenbildung bestimmte Unterschiede zwischen Datenträgern übergewichtet, andere untergewichtet werden. Dadurch kann es insbesondere bei der Analyse von Interaktionseffekten zu verzerrten Schätzungen und falsch positiven Entscheidungen bei der Hypothesenprüfung kommen (MacCallum, Zhang, Preacher & Rucker, 2002; Whisman & McClelland, 2005). Zwar liegen bislang für den Fall von Interaktionen einer kategorialen mit einer metrischen Prädiktorvariablen noch keine systematischen Simulationsstudien darüber vor, wie robust Hypothesenentscheidungen gegen die Verzerrungen und den Informationsverlust sind, die durch das Verfahren der Gruppenbildung bedingt werden. In Analogie zu den Verhältnissen bei bivariaten Zusammenhängen und dem Fall der Interaktion zweier dichotomisierter metrischer Variablen (Maxwell & Delaney, 1993) lässt sich aber annehmen, dass die Robustheit eher gering ausfällt. So wird beispielsweise die bivariate Korrelation zweier bivariat normalverteilter Variablen auf etwa $4/5$ reduziert, wenn eine der beiden Variablen dichotomisiert wird (Peters & van Voorhis, 1940). Schließlich ist die Bedeutung der neu gebildeten kategorialen Variablen von der Verteilung der metrischen Ausgangsvariablen in den jeweils untersuchten Stichproben abhängig, was die Verallgemeinerbarkeit von Ergebnissen, die auf derartigen Variablen beruhen, praktisch unmöglich macht. All diese Probleme lassen nur den Schluss zu, dass die Anwendung der Dichotomisierung und anderer Methoden der Gruppenbildung anhand von metrischen Prädiktorvariablen grundsätzlich nicht zu rechtfertigen ist.

Die Methode der Bestimmung separater Korrelationen leidet unter dem Problem, dass unterschiedliche Korrelationen in den Subgruppen nicht zwangsläufig auf einen tatsächlichen Moderatoreffekt der kategorialen Variablen zurückgehen, sondern auch auf Unterschieden in der Varianz der metrischen Prädiktorvariablen oder der Kriteriumsvariablen beruhen kann (z.B. Whisman & McClelland, 2005). Daher eignet sich auch diese Methode nicht für die Überprüfung von Interaktionen von metrischen und kategorialen Prädiktoren.

Glücklicherweise stehen im Rahmen des Allgemeinen Linearen Modells alternative und einfach anwendbare Methoden zur Verfügung, mit denen sich Auswertungsdesigns mit beliebigen Kombinationen metrischer und kategorialer Prädiktoren einschließlich ihrer Interaktionen methodisch einwandfrei analysieren lassen. In der Literatur werden diese Methoden als Kovarianzanalysen mit nichtparallelen Regressionsgeraden (Werner, 1997), Regressionsanalysen mit Interaktionen von metrischen und kategorialen Prädiktoren (Jaccard & Turrisi, 2003) bzw. moderierte Regressionsanalysen mit metrischen und kategorialen Prädiktoren behandelt (Aiken & West, 1991; Cohen, Cohen, West & Aiken, 2003). Dieser Beitrag gibt einen anwendungsorientierten Überblick über die Analyse von Interaktionen von metrischen und kategorialen Prädiktoren in linearen Modellen. Im Folgenden werden zunächst das Grundprinzip und die wichtigsten Schritte bei der Anwendung dieser Methoden anhand eines Modells mit einer dichotomen kategorialen und einer metrischen Variablen erläutert und mit einem Datenbeispiel aus der Medienpsychologie illustriert. Im abschließenden Teil wird das Verfahren auf komplexere Auswertungssituationen verallgemeinert, und es werden weiterführende methodische Hinweise gegeben.

1 Grundprinzip von Regressionsanalysen mit Interaktionen metrischer und kategorialer Prädiktoren

Das Grundprinzip von Regressionsanalysen mit Interaktionen metrischer und kategorialer Prädiktoren soll anhand des einfachsten Falls mit einer zweifach gestuften kategorialen Prädiktorvariablen D_i und einer metrischen Prädiktorvariablen X_i erläutert werden. Gemäß der allgemeinen Logik der Modellierung von Interaktionen in linearen Modellen werden Interaktionen als Produkt von Prädiktorvariablen modelliert (vgl. Aiken & West, 1991). Die Regressionsgleichung lautet entsprechend

$$Y_i = b_0 + b_1 D_i + b_2 (X_i - \bar{X}) + b_3 D_i (X_i - \bar{X}) + e_i . \quad (1)$$

In Gleichung 1 repräsentiert b_1 den Haupteffekt des kategorialen Prädiktors, b_2 den Haupteffekt des metrischen Prädiktors und b_3 den Interaktionseffekt beider Prädiktoren. Zusätzlich enthält das Modell eine additive Konstante b_0 und einen Residualterm e_0 , der den individuellen Prädiktionsfehler wiedergibt. Durch die Einbeziehung des Interaktionseffekts werden die Effekte der beiden Prädiktorvariablen separiert in einen Anteil, der unabhängig von der Ausprägung des anderen Prädiktors ist (Haupteffektterme) und einen Anteil, der mit der Ausprägung des anderen Prädiktors variiert (Interaktionsterm). Dabei ist es wichtig, dass neben dem Interaktionsterm immer auch die Haupteffektterme aller betrachteten Prädiktoren im Modell berücksichtigt werden, da der Interaktionsterm ansonsten ein nicht interpretierbares Amalgam aus Haupt- und Interaktionseffekten widerspiegelt.

Die Schätzung und inferenzstatistische Testung der Parameter in Gleichung 1 unterscheidet sich nicht vom üblichen Vorgehen bei linearen Modellen. Auf der Basis von Gleichung 1 lässt sich demnach der Interaktionseffekt b_3 und sein Standardfehler mit Hilfe der Kleinst-Quadrate-Technik schätzen und über die t -Verteilung inferenzstatistisch gegen die Null-Hypothese prüfen, dass kein Interaktionseffekt vorliegt. Eine äquivalente Möglichkeit besteht darin, das Inkrement des Interaktionsterms zum Anteil der durch das Modell aufgeklärten Varianz zu bestimmen und anhand der F -Verteilung zu testen (vgl. z.B. Werner, 1997). Um einen signifikanten Interaktionseffekt und die Regressionskoeffizienten der beiden Haupteffektterme im Modell sinnvoll interpretieren zu können, müssen die Prädiktorvariablen allerdings in einer bestimmten Form vorliegen. Dafür kommt der Skalierung bzw. Kodierung von Prädiktorvariablen bei Regressionsanalysen mit Interaktionstermen eine entscheidende Bedeutung zu.

1.1 Skalierung und Kodierung von Prädiktorvariablen

Ganz generell ist bei Regressionsanalysen mit Interaktionstermen eine *Zentrierung* der metrischen Prädiktorvariablen zu empfehlen, wie sie beispielsweise in Gleichung 1 in Form durch die Verwendung von Abweichungswerten vom Gesamtmittelwert ($X_i - \bar{X}$) erfolgt ist. Für die Wahl dieser Skalierungstechnik sprechen zwei Gründe. Zum einen werden durch die Verwendung zentrierter Werte Probleme mit sogenannter nicht-essentieller Multikollinearität von Haupteffekt- und Interaktionstermen vermieden, die die Schätzung der Haupteffektterme und ihrer Standardfehler verfälschen könnte (Marquardt, 1980). Zum anderen wird durch die Verwendung zentrierter Prädiktorwerte, deren Stichprobenmittelwert Null ist, die Interpretation der im Modell enthaltenen Parameter vereinfacht. So lässt sich infolge der Zentrierung des metrischen Prädiktors in Gleichung 1 der Koeffizient b_1 als Haupteffekt des kategorialen Prädiktors bei einer mittleren Ausprägung des metrischen Prädiktors interpretieren.

Auch die Art und Weise, wie die kategoriale Prädiktorvariable kodiert wird, hat einen Einfluss auf die Schätzung sämtlicher im Modell enthaltenen Parameter. Bei Regressionsanalysen mit Interaktionen metrischer und kategorialer Prädiktoren sind zwei Kodierungsvarianten von besonderer Bedeutung (für eine ausführliche Diskussion verschiedener Kodierungstechniken, vgl. Cohen, Cohen, West & Aiken, 2003, Kap. 8):

(1) Um eine aussagekräftige Schätzung des Haupteffekts des metrischen Prädiktors zu erhalten, ist auch für den kategorialen Prädiktor die Wahl einer Kodierungsvariante anzuraten, die zu zentrierten Werten führt. Bei einem dichotomen Prädiktor, dessen Merkmalsausprägungen gleich häufig vertreten sind (d.h. bei gleich großen Gruppen), lässt sich eine Zentrierung durch die Verwendung einer Kontrastkodierung erzielen (den Datenträgern einer Gruppe wird der Wert -1, denen der anderen Gruppe der Wert 1 zugewiesen), die im Fall einer dichotomen Variablen identisch mit einer ungewichteten

Effektkodierung ist. Bei einem dichotomen Prädiktor, dessen Merkmalsausprägungen nicht gleich häufig vertreten sind (die Gruppengrößen n_1 und n_2 unterscheiden sich), ist eine gewichtete Effektkodierung zu verwenden (z.B. wird den Datenträgern der ersten Gruppe der Wert 1 zugewiesen, denen der zweiten Gruppe der Wert $-n_1/n_2$, also ein anhand des Verhältnisses der Gruppengrößen gewichteter negativer Wert, so dass die Summe der Kodiervariablen über alle Datenträger Null beträgt). Im Ergebnis spiegelt die Schätzung des Regressionskoeffizienten für den Haupteffektterm des metrischen Prädiktors, d.h. b_2 in Gleichung 1, den mittleren Effekt für diesen Prädiktor wider (ggf. korrigiert um ungleiche Gruppengrößen n_1 und n_2). Die Schätzung des Regressionskoeffizienten für den Haupteffektterm des kategorialen Prädiktors, d.h. b_1 in Gleichung 1, spiegelt bei Verwendung der Kontrastkodierung die Abweichung der mit dem positiven Wert kodierten Gruppe vom Gesamtmittelwert wider. Bei der gewichteten Effektkodierung handelt es sich um die Abweichung vom theoretischen Gesamtmittelwert unter der Annahme gleicher Gruppengrößen.

(2) Um die Interpretation des Interaktionseffekts zu erleichtern, verwendet man für den kategorialen Prädiktor eine einfache Dummykodierung. Dabei wird den Datenträgern einer Gruppe, der sogenannten Referenzgruppe, der Wert 0 zugewiesen, während den Datenträgern der anderen Gruppe der Wert 1 zugewiesen wird. Die Anwendung dieser Kodierungstechnik, die auch eine leicht zugängliche Interpretation des Haupteffekts des kategorialen Prädiktors ermöglicht, wird im folgenden Abschnitt erläutert.

1.2 Interpretation von Interaktionseffekten

Bei der Interpretation einer signifikanten Interaktion von metrischem und kategorialem Prädiktor können Forscher/innen je nach ihrem Erkenntnisinteresse zwei Frageperspektiven einnehmen. Erstens kann der kategoriale Prädiktor als Moderator des

Effekts des metrischen Prädiktors betrachtet werden. In diesem Fall stehen Unterschiede in den Steigungen der Regressionsgeraden des metrischen Prädiktors zwischen den Gruppen im Vordergrund, die durch den kategorialen Prädiktor gebildet werden. Zweitens kann der metrische Prädiktor als Moderator des Effekts des kategorialen Prädiktors betrachtet werden. Für diese Frageperspektive sind Unterschiede zwischen den durch den kategorialen Prädiktor definierten Gruppen in Abhängigkeit von verschiedenen Ausprägungen des metrischen Prädiktors relevant.

Kategorialer Prädiktor als Moderator. Zur Analyse der Moderatorfunktion des kategorialen Prädiktors für den Effekt des metrischen Prädiktors werden für den metrischen Prädiktor sogenannte einfache Regressionskoeffizienten (simple slopes) bestimmt, indem die Ausprägungen des kategorialen Prädiktors auf bestimmten Werten konstant gehalten und dann die Regressionsparameter neu geschätzt werden (vgl. Aiken & West, 1991). Für diesen Zweck eignet sich die einfache Dummykodierung besonders gut. Dies wird unmittelbar deutlich, wenn man die beiden Ausprägungen eines dummykodierte kategorialen Prädiktors in Gleichung 1 einsetzt. Für $D = 0$ resultiert daraus folgende Gleichung:

$$Y_i = b_0 + b_2(X_i - \bar{X}) + e_i , \quad (2)$$

und für $D = 1$ ergibt sich folgende Gleichung:

$$Y_i = b_0 + b_1 + (b_2 + b_3)(X_i - \bar{X}) + e_i . \quad (3)$$

Bei der Verwendung der Dummykodierung repräsentiert b_2 demnach den einfachen Regressionskoeffizienten für den metrischen Prädiktor in der mit Null kodierten Gruppe. Der einfache Regressionskoeffizient in der mit Eins kodierten Gruppe ergibt sich dagegen aus der Summe von b_2 und b_3 . Daraus folgt, dass bei Verwendung der Dummykodierung b_3 , also der Regressionskoeffizient des Interaktionsterms, den Unterschied der einfachen Steigungskoeffizienten in den beiden Gruppen repräsentiert. Liegt kein Interaktionseffekt vor, unterscheiden sich die beiden einfachen Regressionskoeffizienten demnach nicht. Auch der

Regressionskoeffizient b_1 , der dem Haupteffektterm des kategorialen Prädiktors zugeordnet ist, hat bei Verwendung von Dummykodierung und zentriertem metrischen Prädiktor eine leicht zugängliche Interpretation. Dieser Koeffizient repräsentiert den Unterschied zwischen den beiden durch den kategorialen Prädiktor definierten Gruppen bei einer mittleren Ausprägung des metrischen Prädiktors.

In der Praxis bestimmt man die einfachen Regressionskoeffizienten, indem die Regressionsparameter bei Dummykodierung des kategorialen Prädiktors zweimal geschätzt werden, wobei einmal die eine Gruppe und einmal die andere Gruppe als Referenzgruppe herangezogen wird. Auf diese Weise können die beiden einfachen Regressionskoeffizienten und ihre Standardfehler direkt geschätzt und auf Signifikanz geprüft werden. Wenn sich die Vorzeichen der einfachen Regressionskoeffizienten unterscheiden, kann übrigens auch der Fall auftreten, dass trotz eines signifikanten Interaktionseffekts keiner der beiden einfachen Regressionskoeffizienten signifikant von Null verschieden ist. Die Schätzungen der einfachen Regressionskoeffizienten entsprechen den Schätzungen, die man bei der Formulierung und Schätzung getrennter Regressionsmodelle für die beiden durch den kategorialen Prädiktor definierten Gruppen erhalten würde. Das hier vorgeschlagene Verfahren nutzt jedoch die in der Gesamtstichprobe enthaltene Information, wodurch für die inferenzstatistische Prüfung der einfachen Regressionskoeffizienten eine weitaus höhere Teststärke erzielt wird (vgl. Wishman & McClelland, 2005).

Zur graphischen Veranschaulichung der Interaktion können die einfachen Regressionsgeraden in einem Diagramm abgebildet werden, wobei für den metrischen Prädiktor ein Wertebereich von einer Standardabweichung unter dem Mittelwert bis eine Standardabweichung über dem Mittelwert üblich ist. Als Zusatzinformationen können noch weitere Informationen in das Diagramm aufgenommen werden, etwa Konfidenzintervalle der Parameterschätzungen oder die Datenpunkte selbst (vgl. Aiken & West, 1991).

Metrischer Prädiktor als Moderator. Zur Analyse der Moderatorfunktion des metrischen Prädiktors für den Effekt des kategorialen Prädiktors haben Aiken und West (1991) ein Verfahren vorgeschlagen, das auf einem ähnlichen Prinzip beruht wie die Bestimmung einfacher Regressionskoeffizienten. Dabei wird der Effekt des kategorialen Prädiktors für verschiedene Ausprägungen des metrischen Prädiktors geschätzt. Dies geschieht, indem der kategoriale Prädiktor in dummykodierter Form, der metrische Prädiktor dagegen als Abweichungswert $X_i - X_k$ von der jeweils interessierenden Prädiktorausprägung einbezogen wird. Im Ergebnis repräsentiert der Regressionskoeffizient b_1 , der dem Haupteffektterm des kategorialen Prädiktors zugeordnet ist, den Effekt des kategorialen Prädiktors unter der Bedingung, dass der metrische Prädiktor den Wert X_k annimmt. Anhand des ebenfalls geschätzten Standardfehlers kann auch geprüft werden, ob sich die beiden durch den kategorialen Prädiktor definierten Gruppen bei diesem Wert des metrischen Prädiktors unterscheiden.

2 Ein medienpsychologisches Datenbeispiel

Das Grundprinzip der Analyse von Interaktionen metrischer und kategorialer Prädiktoren und die wichtigsten Verfahrensschritte sollen anhand eines Beispieldatensatzes aus einer kleinen medienpsychologischen Untersuchung illustriert werden. In einem zweiten Schritt werden die dabei erzielten Ergebnisse mit Anwendungen der beiden suboptimalen Methoden der Gruppenbildung und der Bestimmung separater Korrelationen auf die Beispieldaten kontrastiert.

Das Datenbeispiel beruht auf einem Trainingsexperiment, in dem Versuchspersonen in der Anwendung von Internet-Suchstrategien trainiert wurden. Dabei kam in zwei Versuchsgruppen ($n_1 = 18$ und $n_2 = 17$) jeweils eine von zwei Trainingsmethoden (strukturiertes vs. exploratives Training) zum Einsatz (kategorialer Prädiktor D_i). Zusätzlich

wurde mit einem Subtest des Inventars zur Computerbildung (Richter, Naumann & Groeben, 2001) das praktische Computerwissen der Teilnehmer/innen gemessen (metrischer Prädiktor X_i). Als abhängige Variable wurde die Anzahl der Lösungsschritte erhoben, die zur Lösung einer im Anschluss an das Training vorgegebenen Web-Suchaufgabe benötigt wurden (Kriteriumsvariable Y_i).

Auf der Basis dieses Datensatzes sollte die Interaktionshypothese überprüft werden, dass die Wirksamkeit des explorativen Trainings durch praktisches Computerwissen verbessert wird, während dies für das strukturierte Training nicht der Fall ist. Erwartet wurde folglich ein negativer Zusammenhang von praktischem Computerwissen und der Anzahl benötigter Lösungsschritte in der Web-Suchaufgabe in der Gruppe, die das explorative Training absolviert hat, aber kein derartiger oder zumindest ein abgeschwächter Zusammenhang in der Gruppe, die das strukturierte Training erhalten hat. Umgekehrt betrachtet sollte das explorative Training dem strukturierten Training bei einer hohen Ausprägung des praktischen Computerwissens überlegen sein. Bei einer niedrigen Ausprägung sollte sich dagegen kein derartiger oder zumindest ein abgeschwächter Unterschied ergeben.

Diese Hypothese entspricht formal der Hypothese einer ordinalen Interaktion einer kategorialen und einer metrischen Variablen, wie sie für viele medienpsychologische Fragestellungen typisch ist (für aktuelle Beispiele vgl. z.B. Hoffner & Buchanan, 2005; Naumann, Richter, Flender, Christmann & Groeben, 2006; Tsfati & Cappella, 2005; vgl. auch die eingangs genannten Theoriebeispiele). Zur Überprüfung dieser Hypothese ist ein Regressionsmodell wie in Gleichung 1 erforderlich. Anhand unterschiedlicher Kodierungsvarianten lassen sich die einzelnen Effekte im Modell schätzen und interpretieren.

2.1 Interpretation des Haupteffekts der metrischen Variablen: Regressionsmodell mit zentrierten Prädiktoren

In einem ersten Auswertungsschritt wurde ein Regressionsmodell mit ausschließlich zentrierten Prädiktorvariablen formuliert, um eine Schätzung des Haupteffekts des metrischen Prädiktors praktisches Computerwissen zu ermöglichen. Für das praktische Computerwissen wurden dazu Abweichungswerte vom Stichprobenmittelwert gebildet. Für den kategorialen Prädiktor Trainingsmethode wurde aufgrund der leicht ungleichen Gruppengrößen eine gewichtete Effektkodierung verwendet, bei der das strukturierte Training mit -1 und das explorative Training mit 1.06 kodiert wurden. Insgesamt konnte das Modell mit den beiden Haupteffekt- und dem Interaktionsterm rund ein Drittel der Varianz in der Testleistungen in der Web-Suchaufgabe aufklären ($R^2 = .30$, $F(3,31) = 4.4$, $p < .05$). Die additive Konstante b_0 wurde auf 15.0 ($SE = 0.32$) geschätzt. Da das Modell ausschließlich zentrierte Prädiktoren enthielt, lässt sich dieser Wert als Schätzung der mittleren Anzahl benötigter Lösungsschritte in der Web-Suchaufgabe (gemittelt über beide Trainingsbedingungen und bei einer mittleren Ausprägung des praktischen Computerwissens) interpretieren. Der Regressionskoeffizient für den Haupteffektterm der Trainingsmethode (b_1) wurde auf -0.68 ($SE = 0.31$) geschätzt und war signifikant von Null verschieden ($t(34) = 2.1$, $p < .05$, $\Delta R^2 = .13$). Teilnehmer/innen, die das explorative Training absolviert hatten, benötigten demnach insgesamt weniger Lösungsschritte, um die Web-Suchaufgabe zu bewältigen. Der Regressionskoeffizient für den Haupteffektterm des praktischen Computerwissens (b_2) wurde auf -0.18 ($SE = 0.10$) geschätzt und war nicht signifikant von Null verschieden ($t(34) = -1.7$, $p = .10$). In der Gesamtstichprobe hatte das praktische Computerwissen demnach keinen signifikanten Effekt auf die in der Web-Suchaufgabe benötigten Lösungsschritte. Allerdings zeigte sich ein signifikanter und deutlicher Interaktionseffekt der Trainingsbedingung mit dem praktischen Computerwissen ($b_3 = -0.24$, $SE = 0.10$, $t(34) = -2.3$, $p < .05$, $\Delta R^2 = .15$).

2.2 Interpretation des Interaktionseffekts I: Bestimmung einfacher Regressionskoeffizienten

Aus den Ergebnissen für das im vorangegangenen Abschnitt formulierte Regressionsmodell lässt sich schließen, dass tatsächlich ein Interaktionseffekt des metrischen und des kategorialen Prädiktors vorliegt. Allerdings gibt die Schätzung der Regressionskoeffizienten keine Auskunft darüber, ob dieser Interaktionseffekt dem vorhergesagten Muster folgt. Um den Interaktionseffekt zu interpretieren, wurden daher zunächst die einfachen Regressionskoeffizienten des praktischen Computerwissens in den beiden Trainingsbedingungen bestimmt (Trainingsbedingung als Moderator des Effekts des praktischen Computerwissens). Dazu wurde der kategoriale Prädiktor Trainingsbedingung in dummykodierter Form in zwei komplementären Regressionsmodellen einbezogen und die Regressionsparameter wurden neu geschätzt. Der metrische Prädiktor praktisches Computerwissen wurde auch in diesem Modell in zentrierter Form berücksichtigt.

Im Modell zur Bestimmung des einfachen Regressionskoeffizienten in der Gruppe, die das strukturierte Training erhalten hatte, wurde das strukturierte Training mit Null kodiert, während das explorative Training mit Eins kodiert wurde. Auf die Modellgüte des Regressionsmodells hat die Wahl einer anderen Kodierung für den kategorialen Prädiktor keinen Einfluss, allerdings ändern sich die Schätzungen und die Bedeutung der Regressionsparameter erheblich. So wurde die additive Konstante b_0 auf 15.67 ($SE = 0.45$) geschätzt. Dieser Parameter repräsentiert im vorliegenden Modell eine Schätzung der mittleren Anzahl von Lösungsschritten in der Web-Suchaufgabe in der Gruppe mit strukturiertem Training bei einer mittleren Ausprägung des praktischen Computerwissens. Der Regressionskoeffizient für den Haupteffektterm der Trainingsmethode (b_1) spiegelt den Unterschied zwischen den beiden Trainingsbedingungen bei einer mittleren Ausprägung des praktischen Computerwissens wider. Dieser Koeffizient wurde auf -1.40 ($SE = 0.64$) geschätzt und war signifikant von Null verschieden ($t(34) = 2.1, p < .05, \Delta R^2 = .13$). Der Regressionskoeffizient für den Haupteffektterm des praktischen Computerwissens (b_2)

repräsentiert den einfachen Regressionskoeffizienten für diesen Prädiktor in der Gruppe, die das strukturierte Training erhalten hat. Dieser Koeffizient wurde auf 0.06 ($SE = 0.15$) geschätzt und war nicht signifikant ($t(34) = 0.43, p = .67$). Daraus lässt sich schließen, dass das praktische Computerwissen in der Gruppe mit strukturiertem Training keinen Effekt auf die Anzahl der in der Web-Suchaufgabe benötigten Lösungsschritte hatte. Schließlich war auch in diesem Modell der Regressionseffekt des Interaktionsterms signifikant ($b_3 = -0.49, SE = 0.21, t(34) = -2.3, p < .05, \Delta R^2 = .15$). Die unterschiedliche Kodierung gegenüber dem Modell mit zentrierten Koeffizienten hat grundsätzlich keinen Einfluss auf die inferenzstatistische Prüfung und Stärke des Interaktionseffekts.

Im nächsten Schritt wurde nach derselben Logik der einfache Regressionskoeffizient für die Gruppe ermittelt, die das explorative Training erhalten hat. Dafür wurde eine Dummykodierung verwendet, bei der das explorative Training mit Null und das strukturierte Training mit Eins kodiert wurden. Im Anschluss wurden die Regressionsparameter neu geschätzt. An der Modellgüte ändert dieser Schritt wiederum aus prinzipiellen Gründen nichts, sondern lediglich an den Schätzungen der Regressionsparameter. Die additive Konstante, durch die in diesem Modell die mittlere Anzahl von Lösungsschritten in der Gruppe mit explorativem Training bei einer mittleren Ausprägung des explorativen Trainings repräsentiert wird, wurde auf 14.27 ($SE = 0.46$) geschätzt. Die Schätzung des Regressionskoeffizienten für den Haupteffektterm der Trainingsmethode gibt in diesem Modell wiederum den Unterschied zwischen den beiden Trainingsbedingungen bei einer mittleren Ausprägung des praktischen Computerwissens wieder. Dieser Regressionskoeffizient war folglich –abgesehen vom Vorzeichen – identisch mit der Schätzung des vorherigen Modells ($b_1 = -1.40, SE = 0.64, t(34) = 2.1, p < .05, \Delta R^2 = .13$). Der Regressionskoeffizient für den Haupteffektterm des praktischen Computerwissens (b_2), der im vorliegenden Modell den einfachen Regressionskoeffizienten in der Gruppe mit explorativem Training repräsentiert, wurde auf -0.43 ($SE = 0.15$) geschätzt und war signifikant von Null

verschieden ($t(34) = -2.83, p < .01, \Delta R^2 = .20$). Anders als in der Gruppe, die das strukturierte Training erhalten hatte, hatte das praktische Computerwissen in der Gruppe mit explorativem Training also einen deutlich leistungsförderlichen Einfluss auf die Anzahl der benötigten Lösungsschritte in der Web-Suchaufgabe. Die Moderatorfunktion der Trainingsmethode entsprach also dem in der Interaktionshypothese vorhergesagten Ergebnismuster (Abbildung 1). Der Interaktionseffekt von metrischem und kategorialem Prädiktor war auch in diesem Modell signifikant, wobei die Teststatistiken und die Stärke des Effekts den Ergebnissen aus den zuvor geschätzten Modellen entsprachen.

Abbildung 1 hier einfügen.

2.3 Interpretation des Interaktionseffekts II: Effekt des kategorialen Prädiktors bei verschiedenen Ausprägungen des metrischen Prädiktors

Zusätzlich sollte der Interaktionseffekt auch aus einer umgekehrten Perspektive interpretiert werden (praktisches Computerwissen als Moderator des Effekts der Trainingsbedingung). Dazu wurden Unterschiede zwischen den beiden Trainingsbedingungen bei verschiedenen Ausprägungen der metrischen Variablen geschätzt und auf Signifikanz geprüft. Aus den Ergebnissen der zuvor geschätzten Modelle ließ sich bereits entnehmen, dass bei Teilnehmer/innen mit einer mittleren Ausprägung des praktischen Computerwissens das explorative Training zu einer besseren Testleistung führt als das strukturierte Training. Zusätzlich dazu sollte der Unterschied zwischen beiden Trainingsbedingungen bei einer niedrigen Ausprägung des praktischen Computerwissens (eine Standardabweichung unter dem Stichprobenmittelwert) und bei einer hohen Ausprägung des praktischen

Computerwissens (eine Standardabweichung über dem Stichprobenmittelwert) bestimmt werden.

Zu diesem Zweck wurde das praktische Computerwissen in Form von Abweichungswerten von der jeweils interessierenden Ausprägung skaliert, also einmal in Form von Abweichungswerten von dem Wert, der eine Standardabweichung über dem Stichprobenmittelwert lag $[X_i - (\bar{X} + SD(X_i))]$, und einmal in Form von Abweichungswerten von dem Wert, der eine Standardabweichung unter dem Stichprobenmittelwert lag $[X_i - (\bar{X} - SD(X_i))]$. Auf der Basis dieser beiden Skalierungsvarianten wurde je ein Regressionsmodell formuliert, in dem der kategoriale Prädiktor Trainingsmethode in dummykodierter Form einbezogen wurde (0 = strukturiertes Training und 1 = exploratives Training). In diesen Modellen repräsentiert der Haupteffektterm des kategorialen Prädiktors (b_1) eine Schätzung des Unterschieds zwischen beiden Trainingsmethoden bei derjenigen Ausprägung des praktischen Computerwissens, die zur Bildung der Abweichungswerte herangezogen wurde. Für das Modell mit Abweichungswerten von dem Wert, der eine Standardabweichung unter dem Mittelwert des praktischen lag, wurde dieser Koeffizient auf 0.12 ($SE = 0.92$) geschätzt und war nicht signifikant ($t(34) = 0.1, p = .90$). Für das Modell mit Abweichungswerten von dem Wert, der eine Standardabweichung über dem Mittelwert liegt, wurde derselbe Koeffizient dagegen auf -2.93 ($SE = 0.92$) geschätzt und war signifikant von Null verschieden ($t(34) = -3.2, p < .01, \Delta R^2 = .25$). Während sich bei Versuchspersonen mit einem (relativ zur Verteilung in der Stichprobe) unterdurchschnittlichen praktischen Computerwissen weder das strukturierte noch das explorative Training als besonders vorteilhaft erwies, profitierten übereinstimmend mit der skizzierten Hypothese Personen mit einem überdurchschnittlichen praktischen Computerwissen ganz besonders vom explorativen Training (vgl. Abbildung 1).

Diese Art von Analysen ließe sich im Prinzip mit beliebigen weiteren Abweichungswerten weiterführen, womit eine detaillierte und auf die jeweilige Fragestellung

flexibel zugeschnittene Analyse der Moderatorfunktion des metrischen Prädiktors ermöglicht wird. Dieses Vorgehen unterscheidet sich grundlegend von dem Verfahren der Gruppenbildung (über Median-Splits oder eine andere Variante). Während durch die Gruppenbildung Informationen verloren geht, indem Unterschiede zwischen Datenträgern künstlich beseitigt werden, wird bei dem hier skizzierten Verfahren die in der metrischen Variablen enthaltene Information vollständig genutzt.

2.4 Vergleich mit herkömmlichen Verfahren

In diesem Abschnitt sollen die Ergebnisse der Anwendung des hier vorgeschlagenen Verfahrens mit den Ergebnissen der beiden üblicherweise eingesetzten, suboptimalen Verfahren der Gruppenbildung und der Bestimmung separater Korrelationen verglichen werden. Dieser Vergleich soll zum einen deutlich machen, dass es sich bei den drei Verfahren keinesfalls um äquivalente Methoden handelt, und zum anderen die verbesserten Erkenntnismöglichkeiten illustrieren, die sich aus der Verwendung linearer Modelle mit Interaktionen kategorialer und metrischer Variablen ergeben.

Gruppenbildung (Dichotomisierung) der metrischen Variablen. Eine Dichotomisierung der metrischen Variablen praktisches Computerwissen anhand eines Mediansplits führt zu einer kategorialen Variablen mit zwei Gruppen, von denen eine über ein höheres und die andere um ein niedrigeres Computerwissen verfügt. In der Folge können die Daten mit einer zweifaktoriellen (Zwischengruppen-)Varianzanalyse mit dem Faktor Trainingsmethode (exploratives vs. strukturiertes Training) und Computerwissen (hoch vs. niedrig) und der abhängigen Variablen Anzahl von Lösungsschritten in der Web-Suchaufgabe ausgewertet werden. Die Anwendung dieses Verfahrens auf die Beispieldaten ergab für die übliche α -Fehler-Wahrscheinlichkeit von .05 keinerlei signifikante Effekte. Der Haupteffekt der Trainingsmethode verfehlte im Unterschied zu den Ergebnissen des hier vorgeschlagenen

Verfahrens knapp die Signifikanzgrenze, $F(1,31) = 4.0, p = .06, \eta^2 = .11$. Zudem waren weder der Haupteffekt des Faktors Computerwissen, $F(1,31) = 1.8, p = .20, \eta^2 = .05$, noch der hypothetisierte Interaktionseffekt von Trainingsmethode und Computerwissen signifikant, $F(1,31) = 2.1, p = .16, \eta^2 = .06$. Diese Ergebnisse veranschaulichen den gravierenden Verlust an Teststärke, der durch die Dichotomisierung der kontinuierlichen Prädiktorvariablen und den damit einhergehenden Informationsverlust auftreten kann. Davon ist im Beispieldatensatz in erster Linie die Prüfung des hypothetisierten Interaktionseffekts betroffen. Indirekt kann durch die Dichotomisierung aber auch die Schätzung und Prüfung des Haupteffekts der kategorialen Variablen verzerrt werden, wie sich am Beispiel des fehlenden Haupteffekts der Trainingsmethode ablesen lässt. Der Informationsverlust zeigt sich schließlich auch darin, dass die durch das varianzanalytische Modell insgesamt aufgeklärte Varianz in der Kriteriumsvariablen nur 20% (gegenüber 30% im linearen Modell mit kategorialem und metrischem Prädiktor) beträgt.

Bestimmung separater Korrelationen. Die Bestimmung separater Korrelationen führte für das Datenbeispiel in der Gruppe des explorativen Trainings zu einer Korrelation von $-.61 (p < .01)$ zwischen praktischem Computerwissen und der Anzahl von Lösungsschritten in der Web-Suchaufgabe, wohingegen sich in der Gruppe des strukturierten Trainings keine signifikante Korrelation ($r = .10, p = .09$) ergab. Ein inferenzstatistischer Vergleich ergab, dass sich beide Korrelationen signifikant voneinander unterscheiden ($z = 2.2, p < .05$). Dieses Ergebnismuster entspricht im vorliegenden Fall dem Ergebnismuster, das sich aus der Anwendung des linearen Modells mit kategorialem und metrischem Prädiktor ergeben hatte. Gegenüber dieser Auswertung lässt die Methode der Bestimmung separater Korrelationen im Datenbeispiel aber eine Reihe von Fragen offen. So lässt sich dem Korrelationsvergleich beispielsweise nichts über den Haupteffekt der Trainingsmethode oder Unterschiede zwischen den beiden verwendeten Trainingsmethoden bei verschiedenen

Ausprägungen des praktischen Computerwissens (praktisches Computerwissen als Moderator) entnehmen. Inhaltlich ausgedrückt bleibt unklar, für welche Nutzer/innen welche Trainingsform zu einer besseren Bewältigung der Web-Suchaufgabe führt. Derartige Fragen können mit linearen Modellen mit Interaktionstermen detailliert und im Rahmen eines einheitlichen Auswertungsmodells beantwortet werden.

3 Weiterführende Überlegungen

In diesem Abschnitt sollen einige ausgewählte Themen angesprochen werden, die für die Anwendung von Regressionsanalysen mit Interaktionen metrischer und kategorialer Prädiktoren in der medienpsychologischen Forschungspraxis von besonderer Bedeutung sind. Dabei handelt es sich um die Verallgemeinerung auf Interaktionen mit kategorialen Variablen, die mehr als zwei Abstufungen besitzen, sowie die Verallgemeinerung auf Modelle mit Interaktionen höherer Ordnung, Überlegungen zur Teststärke in linearen Modellen mit Interaktionstermen, Anmerkungen zu den Anwendungsvoraussetzungen des Verfahrens sowie Hinweise zur Verwendung standardisierter Regressionskoeffizienten.

3.1 Verallgemeinerungen auf komplexere Auswertungsmodelle

Das in den vorangegangenen Abschnitten dargestellte Grundprinzip lässt sich auf Fälle verallgemeinern, bei denen eine kategorialer Prädiktorvariable mehr als zwei Abstufungen hat oder Interaktionen höherer Ordnung analysiert werden sollen (für ausführliche Darstellungen vgl. Aiken & West, 1991; Jaccard & Turrisi, 2003). Bei kategorialen Prädiktoren mit mehr als zwei Abstufungen erfolgt die Interpretation von Haupt- und Interaktionseffekten analog zum Fall eines kategorialen Prädiktors mit zwei Abstufungen. Generell wird eine Prädiktorvariable mit k Abstufungen über $k-1$ Kodiervariablen in das

Regressionsmodell einbezogen. Dadurch ergeben sich schon im Fall von drei Abstufungen mehrere Möglichkeiten, wie die Kodiervariablen angelegt werden können, von denen dann die Schätzung und Interpretation der Regressionskoeffizienten abhängt. Bei Entscheidungen wie der Wahl der Referenzgruppe bei der Dummykodierung oder der Art und Weise, wie Kontrastvariablen anzulegen sind, spielen daher inhaltliche Erwägungen eine größere Rolle als im einfachsten Fall eines kategorialen Prädiktors mit zwei Abstufungen.

Interaktionen höherer Ordnung, etwa Interaktionen zweier kategorialer Prädiktoren mit einem metrischen Prädiktor, können ebenfalls anhand derselben Prinzipien analysiert werden wie der einfache Fall einer Interaktion eines metrischen mit einem kategorialen Prädiktor. Dreifach-Interaktionen z.B. lassen sich mit Hilfe der Dummykodierung im Hinblick auf die Moderatorfunktion eines kategorialen Prädiktors für die Interaktion eines anderen kategorialen Prädiktors mit einem metrischen Prädiktor interpretieren (vgl. z.B. Jaccard & Turrisi, 2003, Kap. 3). Insbesondere bei kategorialen Prädiktoren mit mehr als zwei Abstufungen, für die mehrere Kodiervariablen benötigt werden, kann allerdings die Anzahl der benötigten Terme im Regressionsmodell durch die Berücksichtigung von Interaktionen höherer Ordnung rasch stark ansteigen, was die Teststärke verringern kann. Deshalb kann es in Einzelfall sinnvoll sein, sich auf einfachere Modelle ohne Interaktionen höherer Ordnung zu beschränken.

3.2 Teststärke

Für lineare Modelle mit Interaktionstermen gelten im Prinzip dieselben Überlegungen für die Beziehungen zwischen Teststärke, Irrtumswahrscheinlichkeit für einen Fehler erster Art (α -Fehler-Wahrscheinlichkeit) und Stichprobenumfang wie bei additiven linearen Modellen (vgl. Cohen, 1988). Es gibt allerdings drei Besonderheiten, die im Einzelfall die zur Erzielung einer befriedigenden Teststärke benötigten Stichprobenumfänge

erhöhen können oder die Annahme einer höheren α -Fehler-Wahrscheinlichkeit nötig machen (McClelland & Judd, 1993):

(1) Für Haupteffekt- wie Interaktionsterme gilt gleichermaßen, dass eine mangelnde Reliabilität den theoretisch möglichen Effekt auf die Kriteriumsvariable verringert. So kann die Korrelation zweier Variablen prinzipiell nicht höher sein als das Produkt ihrer Reliabilitäten. Die Reliabilität eines Interaktionsterms bestimmt sich aus dem Produkt der Reliabilitäten der beteiligten Prädiktoren und kann daher im Falle messfehlerbehafteter Prädiktoren, die auch in korrelativen medienpsychologischen Untersuchungen die Regel sind, sehr geringe Werte annehmen. Diese Problematik kann sich bei Interaktionen von kategorialen und metrischen Prädiktoren vor allem dann bemerkbar machen, wenn neben den metrischen auch die kategorialen Prädiktoren gemessen (und nicht experimentell manipuliert) werden und wenn Interaktionen höherer Ordnung mit mehreren messfehlerbehafteten Variablen betrachtet werden. Aus diesen Überlegungen lässt sich als generelle Empfehlungen ableiten, möglichst reliable Messungen zu verwenden bzw. auf eine experimentelle Manipulation kategorialer Prädiktoren zurückzugreifen und sich bei kleinen Stichproben auf möglichst einfache Designs zu beschränken.

(2) Gerade im medienpsychologischen Kontext dürfte die Prüfung von Hypothesen über ordinale Interaktionen häufiger sein als die Prüfung von Hypothesen über disordinale Interaktionen. Im Fall einer ordinalen Interaktion ist es aus prinzipiellen mathematischen Gründen schwierig, ein Inkrement des Interaktionsterms gegenüber den Haupteffekttermen nachzuweisen. Wenn in der Population ein ordinaler Interaktionseffekt vorliegt, werden immer auch die Haupteffektterme einen Beitrag zur Aufklärung der Kriteriumsvarianz leisten (für eine Diskussion vgl. Rogers, 2002). Dadurch ergeben sich insbesondere für den Nachweis kleiner und mittlerer ordinaler Interaktionseffekte hohe Anforderungen an den Stichprobenumfang.

(3) Die Form der Verteilung der metrischen Prädiktorvariablen kann die Aufdeckung von Interaktionseffekten ebenfalls erschweren. Bei einer normalverteilten Variablen beispielsweise befinden sich die meisten Datenpunkte im mittleren Bereich und sind sich daher sehr ähnlich. Extremwerte, die für die Aufdeckung von Interaktionen besonders relevant sind, sind dagegen selten. Die Verwendung von Median-Splits oder einer anderen Form der Gruppierung stellt aus den genannten Gründen *keine* adäquate Lösung dieses Problems dar. Eine mögliche Lösung bietet dagegen das sogenannte Oversampling, bei der gezielt Datenträger mit extremen Werten erhoben werden (vgl. Cohen et al., 2003, Kap. 7.11; McClelland & Judd, 1993).

3.3 Anwendungsvoraussetzungen

Die Anwendungsvoraussetzungen von linearen Regressionsanalysen mit Interaktionen von metrischen und kategorialen Prädiktoren entsprechen den allgemeinen Anwendungsvoraussetzungen linearer Modelle. Dazu gehören insbesondere die Normalverteilung der Residualwerte, Unabhängigkeit und konstante Varianz der Residualwerten über verschiedene Werte der Prädiktorvariablen hinweg (Homoskedastizität), die (näherungsweise) Messfehlerfreiheit der Prädiktorvariablen, lineare Zusammenhänge zwischen Prädiktorvariablen und Kriteriumsvariablen sowie eine geringe Multikollinearität der Prädiktorvariablen untereinander (vgl. Cohen et al., 2003, Kap. 4 für Möglichkeiten der Prüfung dieser Anwendungsvoraussetzungen). Regressionsanalysen, die die herkömmliche Kleinst-Quadrate-Technik zur Schätzung der Regressionsparameter verwenden, sind nicht robust gegen eine Verletzung dieser Anwendungsvoraussetzungen. Vielmehr können Verletzungen dieser Voraussetzungen eine Miss-Spezifikation des Modells, eine Über- oder Unterschätzung von Regressionsparametern und Standardfehlern und in der Folge falsch positive oder falsch negative Hypothesenentscheidungen nach sich ziehen. Daher empfiehlt

sich die Einhaltung bestimmter Regeln bei der Modellkonstruktion, die die Gefahr einer Verletzung von Anwendungsvoraussetzungen minimieren. Dazu gehören eine möglichst vollständige Berücksichtigung theoretisch relevanter Prädiktorvariablen (einschließlich theoretisch relevanter Interaktionsterme) bei gleichzeitiger Vermeidung von redundanten Operationalisierungen desselben Konstrukts, die Verwendung möglichst zuverlässig gemessener oder experimentell manipulierter Prädiktorvariablen und die Eliminierung von Ausreißerwerten. Grundsätzlich müssen bei Berücksichtigung von Interaktionen in einem Regressionsmodell auch sämtliche Interaktionen niedrigerer Ordnung sowie alle Haupteffekte berücksichtigt werden, da das Modell ansonsten zwangsläufig fehlspezifiziert ist. Mittlerweile stehen auch Alternativen zum Kleinst-Quadrate-Schätzverfahren zur Verfügung, die robust gegen die Verletzung bestimmter Anwendungsvoraussetzungen sind (Wilcox, 2002).

Entgegen einem gelegentlich in der Literatur zu findenden Missverständnis beruht das Verfahren der linearen Regressionsanalyse nicht auf der Annahme intervallskalierter Prädiktorvariablen oder auf bestimmten Verteilungsannahmen bezüglich der Prädiktoren (auch wenn die Verwendung ordinalskalierter oder nicht-symmetrischer Prädiktoren eine Verletzung bestimmter Anwendungsvoraussetzungen nach sich ziehen kann). Insbesondere lassen sich – wie bei dem hier vorgestellten Verfahren – problemlos metrische und kategoriale Prädiktoren in ein gemeinsames Regressionsmodell einbeziehen.

3.4 Standardisierte Regressionskoeffizienten

Anders als bei rein additiven linearen Modellen ist bei linearen Modellen mit Interaktionstermen die herkömmliche Bestimmung und Interpretation standardisierter Regressionskoeffizienten nicht sinnvoll. In additiven linearen Modellen entsprechen die standardisierten Regressionskoeffizienten den Schätzungen, die sich bei ausschließlicher Einbeziehung z -standardisierter Prädiktorvariablen ergeben würden. Für lineare Modelle mit

Interaktionstermen würde diese Äquivalenzbeziehung nicht mehr gelten, da die z -Standardisierung eines Produktterms in der Regel nicht mit dem Produkt der z -standardisierten Ausgangsvariablen identisch ist (Jaccard & Turrisi, 2003). Aus diesem Grund können die standardisierten Regressionskoeffizienten, die von Statistikprogrammen üblicherweise ausgegeben werden, bei Regressionsanalysen mit Interaktionen von metrischen und kategorialen Prädiktoren nicht verwendet werden. Es empfiehlt sich stattdessen generell, unstandardisierte Regressionskoeffizienten zu berichten und zu interpretieren. Falls dennoch eine standardisierte Lösung benötigt wird, besteht der korrekte Weg darin, die beteiligten Prädiktoren einer z -Standardisierung zu unterziehen und als Interaktionsterm(e) die Produkte der z -standardisierten Prädiktoren zu verwenden (vgl. Aiken & West, 1991; Cohen et al., 2003, Kap. 7.5).

4 Fazit

In diesem Beitrag wurden Regressionsanalysen mit Interaktionstermen als Methode zur Analyse von Interaktionen metrischer und kategorialer Prädiktorvariablen vorgestellt. Obwohl diese Methode leicht zu handhaben ist, keine spezialisierte Auswertungssoftware erfordert und in einer Reihe guter einführender Texte dargestellt wird (Aiken & West, 1991; Cohen et al., 2003, Kap. 9; Jaccard & Turrisi, 2003), wird sie in der Medienpsychologie so gut wie nicht angewandt. Dies ist umso bedauernder, als dass theoretisch zentrale medienpsychologische Fragestellungen am Besten mit einer Analyse metrischer und kategorialer Prädiktoren zu klären sind. In der Forschung kommen bislang in aller Regel suboptimale Verfahren wie die Gruppenbildung auf Basis metrischer Prädiktoren oder die Bestimmung separater Korrelationen zum Einsatz. Beide Vorgehensweisen können einerseits zu einem gravierenden Verlust an Teststärke führen, so dass durch ihre Verwendung die Wahrscheinlichkeit erhöht wird, dass in der Population vorhandene Effekte in einer

Untersuchung nicht entdeckt werden. Sie führen andererseits zu verzerrten Schätzungen von Interaktionseffekten und deshalb unter Umständen auch zu falsch positiven Entscheidungen über Interaktionshypothesen. Medienpsychologen/innen sollten auf diese Verfahren verzichten und stattdessen Regressionsanalysen mit Interaktionstermen verwenden.

Literatur

Aiken, S. A. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Berkowitz, L. & Green, R.G. (1966). Film violence and the cue properties of available targets. *Journal of Personality and Social Psychology*, 3, 525-530.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249-253.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. & Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Gerbner, G., Gross, L., Morgan, M. & Signorielli, N. (1994). Growing up with television: The cultivation perspective. In J. Bryant & D. Zillmann (Eds.), *Perspectives on media effects* (pp. 17-41). Hillsdale, NJ: Erlbaum.

Graesser, A.C., Singer, M. & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.

Hoffner, C. & Buchanan, M. (2005). Young adults' wishful identification with television characters: The role of perceived similarity and character attributes. *Media Psychology*, 7, 325-351.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.

Jaccard, J. & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.

MacCallum, R.C., Zhang, S., Preacher, K.J. & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19-40.

Marquardt, D.W. (1980). You should standardize the predictor variables in your regression model. *Journal of the American Statistical Association*, 75, 87-91.

Maxwell, S.E. & Delaney, H.D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113, 181-190.

McClelland, G.H. & Judd, C.M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376-390.

Naumann, J., Richter, T., Flender, J., Christmann, U. & Groeben, N. (in press). Signaling in expository hypertexts compensates for deficits in reading skill. *Journal of Educational Psychology*.

Peters, C.C. & van Voorhis, W.R. (1940). *Statistical procedures and their mathematical bases*. New York: McGraw-Hill.

Richter, T. & Christmann, U. (2006). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (2. Aufl., S. 25-58). Weinheim: Juventa.

Richter, T., Naumann, J. & Groeben, N. (2001). Das Inventar zur Computerbildung (INCOBI): Ein Instrument zur Erfassung von Computer Literacy und computerbezogenen Einstellungen bei Studierenden der Geistes- und Sozialwissenschaften. *Psychologie in Erziehung und Unterricht*, 48, 1-13.

Rogers, W.M. (2002). Theoretical and mathematical constraints of interactive regression models. *Organizational Research Methods*, 5, 212-230.

Tsfati, Y. & Cappella, J. (2005). Why do people watch news they do not trust? The Need for Cognition as a moderator in the association between news media skepticism and exposure. *Media Psychology*, 7, 251-271.

Werner, J. (1997). *Lineare Statistik: Das Allgemeine Lineare Modell*. Weinheim: Psychologie Verlags Union.

Whisman, M.A. & McClelland, G.H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research. *Journal of Family Psychology, 19*, 111-120.

Wilcox, R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.

Wilcox, R. (2002). *Applying contemporary statistical techniques*. San Diego, CA: Academic Press.

Zillmann, D. (1988). Mood management: Using entertainment to full advantage. In L. Donohew, H.E. Sypher & E.T. Higgins (Eds.), *Communication, social cognition, and affect* (pp. 147-171). Hillsdale, NJ: Erlbaum.

Autorenhinweis

Tobias Richter, Florida State University, Tallahassee, USA.

Die Fertigstellung dieses Artikels wurde durch ein Forschungsstipendium der Deutschen Forschungsgemeinschaft (RI1100/3-1) gefördert.

Tobias Richter

Department of Psychology, Florida State University

Tallahassee, FL 32306-1270

USA

E-mail: richter@psy.fsu.edu

Bildunterschrift

Abbildung 1. Einfache Regressionsgeraden für praktisches Computerwissen (metrischer Prädiktor) in den beiden Trainingsbedingungen strukturiertes vs. exploratives Training (kategorialer Prädiktor).

