planung &analyse

SEIT 30 JAHREN

Zeitschrift für Marktforschung und Marketing www.planung-analyse.de 4/2013 D11700F

Schwerpunkt

Advanced Analytics

Kongress

14. M-Motion-Tag: Aufschlag Marktforschung

Special

Immer abgesichert: Finanzen & Versicherungen

Report

Überall dabei: Mobile Research

Statistik KOMPAKT

Interaktionseffekte in Kausalanalysen

Interaktionen bzw. Wechselwirkungen zwischen unabhängigen Variablen liegen in einer Kausalanalyse vor, wenn sich Variablen in ihrer Wirkung auf eine abhängige Variable gegenseitig verstärken oder abschwächen. Häufig wird beispielsweise berichtet, dass reiche Menschen gesünder sind als arme und sich dieser Unterschied mit zunehmendem Alter vergrößert. Das Alter verstärkt demnach den Einfluss des Einkommens auf die Gesundheit. Eine Berücksichtigung von Interaktionen kann Vorhersagen somit deutlich verbessern.

Interaktion zwischen zwei Variablen

Ein lineares Regressionsmodell zur Untersuchung der Zusammenhänge zwischen der abhängigen Variable y und zwei unabhängigen Variablen x_1 und x_2, das keine Interaktion berücksichtigt, lässt sich formal schreiben als

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2$$

Die Regressionskoeffizienten b_1 und b_2 quantifizieren die Einflussstärken der Variablen. Die Stärke des Einflusses von x_1 ist immer gleich b_1 – gleichgültig welche Ausprägung x_2 aufweist. Liegt dagegen ein Interaktionseffekt vor, verändert sich die Einflussstärke von x_1 (im Beispiel das Einkommen), wenn sich x_2 (das Alter) ändert. Um diese Abhängigkeit zu erfassen, wird die Einflussstärke von x_1 zusätzlich zu b_1 durch den Term $b_3 \cdot x_2$ gemessen:

$$y = b_0 + (b_1 + b_3 \cdot x_2) \cdot x_1 + b_2 \cdot x_2 \qquad (*)$$

Grafisch lassen sich die beiden Situationen anhand von Regressionsgeraden veranschaulichen, die die Beziehung zwischen x_1 und der abhängigen Variable jeweils für zwei unterschiedliche Ausprägungen von x_2 zeigen (siehe Abbildung).

Wird keine Interaktion berücksichtigt bzw. liegt keine Interaktion vor, verlaufen beide Regressionsgeraden parallel zueinander. Die Steigungen respektive Einflussstärken sind gleich – egal ob der Wert von x_2 hoch oder niedrig ist. Bei einer positiven Interaktion ist die Regressionsgerade für den größeren Wert von x_2 dagegen steiler. Allgemein ist die Einflussstärke von x_1 umso höher, je größer x_2 ist.

Interaktionseffekte sind nicht einseitig, sondern immer wechselseitig. Die Möglichkeit der Transformation der Gleichung (*) zu

$$y = b_0 + b_1 \cdot x_1 + (b_2 + b_3 \cdot x_1) \cdot x_2$$

verdeutlicht, dass $(b_2 + b_3 \cdot x_1)$ die Einflussstärke von x_2 misst. Die Einflussstärken können folglich nicht generell, sondern nur unter Vorgabe von Werten für x_1 respektive x_2 angegeben werden. Insofern kann ein Test der Signifikanz beispielsweise des Einflusses von x_1 unter Berücksichtigung eines Interaktionseffekts immer nur für einen bestimmten Wert von x_2 erfolgen. Auf diese Weise lassen sich Bereiche für x_2 ermitteln innerhalb derer der Einfluss von x_1 signifikant ist.

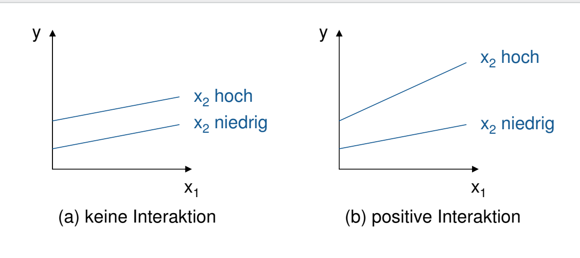

Abbildung: Regressionsgeraden für verschiedene Werte von x_2

Schätzung des Modells

Das Ausmultiplizieren der Klammern in Gleichung (*) führt zu

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_1 \cdot x_2$$

Um eine mögliche Interaktion in einem Regressionsmodell zu berücksichtigen, wird dieses ergänzt um das Produkt der beiden Variablen. Das heißt, das Modell enthält eine weitere Variable, deren Werte sich durch Multiplikation von x_1 und x_2 ergeben. Die Koeffizienten können dann mithilfe der üblichen Ordinary Least Squares (OLS)-Schätzung bestimmt werden. Ob der Interaktionseffekt signifikant ist, lässt sich durch Prüfung der Signifikanz des Koeffizienten b_3 oder der Erhöhung des Bestimmtheitsmaßes im Vergleich zum Modell ohne Interaktion feststellen.

Zentrieren unabhängiger metrischer Variablen

Die Einflussstärke einer Variable mit Interaktion entspricht nur dann genau ihrem Regressionskoeffizienten, wenn die andere Variable gleich null ist. Die Interpretation ist insofern häufig wenig aussagekräftig. Repräsentiert x_2 zum Beispiel das Alter, so wäre b_1 die Einflussstärke von x_1 für diejenigen, die 0 Jahre alt sind. Erleichtert wird die Interpretation, wenn die unabhängigen Variablen vorab zentriert werden. Das heißt von jedem Wert einer Variable wird der Mittelwert dieser Variable subtrahiert. Eine zentrierte Variable ist somit immer gleich null, wenn ihr ursprünglicher Wert dem Mittelwert entspricht. Dann gibt b_1 die Einflussstärke von x_1 für das durchschnittliche Alter in der Stichprobe an.

Das Zentrieren besitzt noch einen weiteren Vorteil: In einem Modell mit Interaktion korrelieren x_1 und x_2 hoch mit der Variable, die aus dem Produkt beider Variablen gebildet wird. Insofern sind die geschätzten Koeffizienten aufgrund hoher Multikollinearität verzerrt. Das Zentrieren der Variablen vor ihrer Multiplikation vermindert das Ausmaß an Multikollinearität und damit Verzerrungen erheblich.

Interaktionen bei mehr als zwei Variablen

Enthält ein Modell mehr als zwei unabhängige Variablen, kann zum einen die Interaktion zwischen drei oder mehr Variablen untersucht werden. Analog zu einer paarweisen Interaktion wird das Produkt der entsprechenden Variablen in das Regressionsmodell aufgenommen. Zum anderen lassen sich verschiedene paarweise Interaktionen betrachten. Allerdings kann mit der Berücksichtigung mehrerer Produktterme im Modell die Multikollinearität trotz Zentrierens wieder zu einem größeren Problem werden. Dann ist es zwar möglich, das Gesamtmodell zur Vorhersage zu nutzen, die einzelnen Effekte können aber nicht valide bestimmt werden. ◄

Johannes Lüken und **Dr. Heiko Schimmelpfennig**, Experten für Multivariate Analyse bei IfaD, Institut für angewandte Datenanalyse GmbH.

In Ausgabe 5/2013: Kategoriale Variablen in Regressionsmodellen

► Literatur

Cohen, J.; Cohen, P.; West, S. G.; Aiken, L. S.: Interactions Among Continuous Variables, In: Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3. Auflage, Mahwah, New Jersey, 2003, S. 255-291.