

planung & analyse

Zeitschrift für Marktforschung und Marketing
Eine Marke der dfv Mediengruppe

www.planung-analyse.de

Sonderdruck

Johannes Lüken
Heiko Schimmelpfennig

Statistik Kompakt

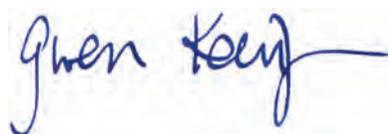


Zugang zur Statistik

Quantitative Marktforschung wird typischerweise untrennbar mit dem Begriff Statistik verbunden – und das wirkt oftmals abschreckend. Denn statistische Verfahren zu durchdringen und vollständig zu begreifen gilt gemeinhin als schwierig bis unlösbar. Häufig setzt sogar als Angstreflex eine Blockade ein, sobald es um weitreichendere quantitative Analysen als reine Häufigkeitsauszählungen geht. Oftmals komplexe und abstrakte Lehrbücher oder gar Vorlesungen haben ebenso ihren Anteil an dieser Abwehrhaltung, wie die verbreitete Ansicht, dass Statistik zu kompliziert zum Verstehen sei.

Das muss nicht sein und ist obendrein vergebenes Potenzial. Vielmehr bieten statistische Verfahren ein vielseitiges Werkzeugset, um den hinter Daten verborgenen Strukturen auf den Grund zu gehen und Zusammenhänge sichtbar zu machen. In der Marktforschung überaus wertvoll, um Meinungen und Einstellungen aufzudecken. Moderne Software macht es zudem häufig einfach, diese Werkzeuge anzuwenden. Doch um dies sinnvoll und zielgerichtet tun zu können, muss zuerst einmal bekannt sein, wozu sich welches statistische Werkzeug eignet. Ebenso schwierig, wie es ist, mit einer Säge ein Loch in die Wand zu bohren, da es nicht das geeignete Werkzeug ist, sind statistische Verfahren bei falscher Anwendung nutzlos. Eine einfache und auch für Ungeübte verständliche Gebrauchsanweisung fehlt jedoch oftmals, um Potenzial und Einsatzfelder der Werkzeuge zu verstehen.

Um die Verwendungszwecke statistischer Verfahren einfach und kompakt darzustellen und somit solche Gebrauchsanweisungen zu liefern, ist die Rubrik *Statistik Kompakt* in planung & analyse angetreten und greift seit Anfang 2012 in jeder Ausgabe ein neues Themenfeld auf. Die Serie umfasst Beiträge, die von verschiedenen Conjoint-Varianten bis zu Kausalanalysen statistische Verfahren aus unterschiedlichen Bereichen erläutern. Die Autoren **Johannes Lüken** und **Heiko Schimmelpfennig** verstehen es, in besonders anschaulicher und leicht nachvollziehbarer Weise, Voraussetzungen und Sachverhalte so einfach und zugänglich zu beschreiben, dass auch unerfahrene Leser die vorgestellten Methodiken begreifen können. Die kompakte Vorstellung auf jeweils einer Seite pro Thema ist als Einstieg gedacht, Hinweise auf weiterführende Literatur geben konkrete Anregung zur Vertiefung jedes Themas. Dieser Sonderdruck umfasst die bisher erschienenen 12 Beiträge der Serie Statistik Kompakt aus einem breiten Themenspektrum – wir wünschen eine erkenntnisreiche Lektüre!



Dr. Gwen Kaufmann
Chefredakteurin planung & analyse

Inhalt

Conjoint-Verfahren im Überblick	3
Varianten des Choice Based Conjoint	4
Auswahl geeigneter Conjoint-Verfahren	5
Marktsimulationen mit Conjoint-Analysen	6
Einführung in Signifikanztests	7
Signifikanz und Stichprobenumfang	8
Einführung in Kausalanalysen	9
Identifizierung von Scheinkorrelationen	10
Strukturgleichungsmodellierung	11
Interaktionseffekte in Kausalanalysen	12
Kategoriale Variablen in Regressionsmodellen	13
Discrete Choice Modelle	14
Autoreninfos	15

Statistik KOMPAKT

Conjoint-Verfahren im Überblick

Die Conjoint-Analyse wird eingesetzt, wenn Produkte oder Dienstleistungen marktgerecht und auf den Kundennutzen ausgerichtet gestaltet werden sollen. Analysiert werden die Präferenzen der Nachfrager. Die Verfahren gehen davon aus, dass sich der Gesamtnutzen eines Konzeptes (einer konkreten, mehrdimensionalen Produktidee) additiv aus den Teilnutzen seiner Merkmale zusammensetzt. Aus der ganzheitlichen Beurteilung (*Considered Jointly*) mehrerer Kombinationen von Merkmalsausprägungen wird auf deren Teilnutzen geschlossen (dekompositioneller Ansatz). Dazu braucht man nur einen Teil aller möglichen Kombinationen beurteilen zu lassen.

Reine Conjoint-Verfahren

In der klassischen Conjoint-Analyse (KCA) sind die durch das Erhebungsdesign bestimmten Konzepte von den Befragten entweder in eine Rangreihe zu bringen oder auf einer Rating-Skala zu bewerten. Darüber hinaus haben die Befragten in der Limit Conjoint-Analyse (LCA) eine Limit-Card zu setzen, das heißt zu bestimmen, bis zu welchem Rang oder Rating sie ein Konzept kaufen würden. Dadurch kann ein Schwellenwert berechnet werden, der zeigt, ab welchem Gesamtnutzen ein Konzept gekauft wird.

In der Choice-Based Conjoint-Analyse (CBC) haben die Befragten wiederholt aus (meistens 3 oder 4) Konzepten eines auszuwählen. Daneben kann eine None-Option angeboten werden, die es ihnen ermöglicht, sich für keines der Konzepte zu entscheiden. Zumeist werden jedem Befragten 10 bis 20 verschiedene solcher Choice-Sets vorgelegt.

Hybrid-Verfahren

In Hybrid-Verfahren werden zusätzlich Fragen zu den einzelnen Merkmalen gestellt, so dass im dekompositionellen Teil eine Konzentration auf die relevanten Merkmale bzw. Ausprägungen möglich ist. Im Vergleich zu reinen Conjoint-Verfahren erlauben sie damit die Berücksichtigung von deutlich mehr Merkmalen und Ausprägungen.

Das am weitesten verbreitete Hybrid-Verfahren ist die Adaptive Conjoint-Analyse (ACA). Zunächst haben die Befragten alle Ausprägungen jedes Merkmals in eine Rangfolge zu bringen bzw. auf einer Rating-Skala zu beurteilen. Anschließend ist für jedes Merkmal anzugeben, wie bedeutend der Unterschied zwischen der besten und schlechtesten Ausprägung ist. Im Weiteren werden mehrfach zwei Konzepte gegenübergestellt und nach der Stärke der Präferenz für das eine oder das andere Konzept gefragt.

Bereits während der Datenerhebung aktualisiert das Verfahren die Teilnutzen stetig. Um einen möglichst großen Informationsgewinn zu erzielen, werden in den Paarvergleichen Konzepte vorgelegt, die einen ähnlichen Gesamtnutzen aufweisen.

Mit dem Adaptive Choice-Based Conjoint (ACBC) wird die Idee des ACA, Antworten der Befragten im weiteren Verlauf zu berücksichtigen, auf das CBC übertragen. Das ACBC beginnt mit der Built Your Own-Frage, bei der sich die Befragten ihr Wunsch-Konzept zusammenstellen dürfen. In der folgenden Screening-Phase wird eine Reihe von Konzepten vorgelegt, die dieser bevorzugten Variante ähneln. Die Befragten geben für jedes Konzept an, ob es in Frage kommt oder nicht. Zwischendurch besteht die Möglichkeit, Ausprägungen als „must haves“ oder „unacceptables“ zu definieren, die bei der Konstruktion der weiteren Konzepte berücksichtigt werden. Im anschließenden Choice Tasks Tournament bildet das Verfahren aus den in Frage kommenden Konzepten Sets, aus denen jeweils das beste Konzept auszuwählen ist. Die Gewinner treten in weiteren Runden in neuen Choice-Sets gegeneinander an, bis der endgültige Sieger feststeht.

Das Hybrid Individualized Two-Level CBC (HIT-CBC) versucht, einen möglichen Number-of-Levels-Effekt, das heißt den positiven Zusammenhang zwischen der relativen Wichtigkeit eines Merkmals und der Anzahl seiner Ausprägungen, zu vermeiden. Die Befragten haben zunächst für jedes Merkmal die beste und die schlechteste Ausprägung zu bestimmen. Darauf basierend wird ein CBC durchgeführt, dessen Konzepte sich je Merkmal ausschließlich aus diesen beiden Ausprägungen zusammensetzen, und damit die Schätzung deren Teilnutzen ermöglicht. Die Teilnutzen der übrigen Ausprägungen werden auf Basis eines Ratings in Relation zur jeweils besten und schlechtesten Ausprägung bestimmt.

Falls eine thematische Gruppierung der Merkmale möglich ist, können für alle Verfahren hierarchische bzw. Multistage-Designs entwickelt werden. Für die Merkmale jeder Gruppe wird ein eigenes (Sub-)Design erstellt sowie ein Merkmal definiert, das Teil eines übergeordneten Hauptdesigns ist. Die Teilnutzen werden innerhalb jedes Designs geschätzt und anschließend mathematisch verknüpft.

Während bei diesem Vorgehen die Gruppierung für alle Befragten gleich ist, werden in der Hierarchisch Individualisierten Limit Conjoint-Analyse (HILCA) die Merkmale nach ihrer zuvor erhobenen individuellen Bedeutung gruppiert. Die Ausprägungen der relevanten Merkmale werden auf einer Rating-Skala beurteilt. Mit den fünf wichtigsten Merkmalen wird dann eine LCA durchgeführt. Für die übrigen relevanten Merkmale werden die Ratings auf die Skala der Teilnutzen transformiert. ◀

In Ausgabe 2/2012 werden Varianten der CBC vorgestellt.

► Literatur

Backhaus, K.; Erichson, B.; Weiber, R.: Auswahlbasierte Conjoint-Analyse. In: Fortgeschrittene Multivariate Analysemethoden. Berlin, Heidelberg: Springer, 2011, S. 317-380.

Buyer, D.; Cleff, T.; Frank, D.: Von Präferenzen zu Marktanteilen – Bestandsaufnahme zum Einsatz von Conjoint-Analysen zur Absatzprognose in der Marktforschungspraxis. In: *planung & analyse*, Jg. 37/2010, Nr. 4, S. 45-49.

Eggers, F.; Sattler, H.: Hybrid Individualized Two-Level Choice-Based Conjoint (HIT-CBC): A New Method for Measuring Preference Structures with

Many Attribute Levels. In: *International Journal of Research in Marketing*, Jg. 26/2009, Nr. 2, S. 108-118.

Johnson, R. M.; Orme, B. K.: A New Approach to Adaptive CBC. Sawtooth Software Research Paper Series. Sequim, 2007.

Louviere, J. J.; Gaeth, G. J.: Decomposing the Determinants of Retail Facility Choice using the Method of Hierarchical Information Integration: A Supermarket Illustration. In: *Journal of Retailing*, Jg. 63/1987, Nr. 1, S. 25-48.

Sawtooth Software: The ACA/Web v6.0. Sawtooth Software Technical Paper Series. Sequim, 2007.

Voeth, M.; Hahn, C.: Limit Conjoint-Analyse. In: *Marketing ZFP*, Jg. 20/1998, Nr. 2, S. 119-132.

Voeth, M.: Nutzenmessung in der Kaufverhaltensforschung: Die Hierarchisch Individualisierte Limit Conjoint-Analyse (HILCA). Wiesbaden: Gabler, 2000.

Wildner, R.; Dietrich, H.; Hölscher, A.: HILCA: Ein neues Conjoint-Verfahren zur verbesserten Abbildung von Kaufentscheidungen komplexer Produkte. In: *Jahrbuch der Absatz- und Verbrauchsforschung*, Jg. 52/2006, Nr. 4, S. 332-348.

Varianten des Choice Based Conjoint (CBC)

Unter den Conjoint-Verfahren ist das CBC am weitesten verbreitet und kommt aktuell in über 80 Prozent aller Conjoint-Analysen zur Anwendung. Der Vorteil dieser Methode liegt in ihrer realitätsnahen und einfachen Datenerhebung. Der Befragte wird mehrfach vor die Aufgabe gestellt, aus einem Set von Angeboten jenes auszuwählen, das ihm am meisten zusagt. Eine None-Option ermöglicht, sich gegen alle Angebote zu entscheiden. Die Abbildung zeigt ein Beispiel für ein Choice-Set mit fünf Merkmalen. Realitätsnah ist eine solche Entscheidungssituation allerdings nur, wenn der reale Entscheidungsprozess weitgehend simuliert werden kann. Um Einschränkungen durch die künstliche Testsituation in Grenzen zu halten, können mithilfe von Erweiterungen des CBC Besonderheiten berücksichtigt werden. Daneben ist es wichtig, den Befragten nicht zu überfordern. Je nach Komplexität der Produktbeschreibungen ist eine Bewertung von bis zu etwa acht Merkmalen zumutbar. Darüber hinaus nimmt die Differenzierungsfähigkeit so stark ab, dass die Wahlentscheidungen häufig nicht mehr ausreichend valide sind.

Welchen der hier angebotenen PKWs würden Sie wählen?

Herkunft			
Fahrzeugtyp	Limousine	Cabrio	Kombi
Hubraum	2.000 ccm	1.500 ccm	2.000 ccm
Verbrauch	8 l / 100 km	9 l / 100 km	7 l / 100 km
Preis	15.000 €	25.000 €	20.000 €

Ich würde keinen dieser PKWs wählen

Partial Profiles

Das Beispiel der Abbildung zeigt volle Profile, das heißt, dass alle fünf im Modell definierten Merkmale dem Befragten gleichzeitig zur Bewertung vorgelegt werden. Falls sich die Angebote aus mehr Merkmalen zusammensetzen, als einem Befragten zuzumuten sind, steht die Partial Profile-Variante zur Verfügung. Dabei wird in jedem Choice-Set nur eine zufällige Auswahl aus den definierten Merkmalen gezeigt. Alle Angebote eines Choice-Sets werden durch dieselbe Teilmenge von Merkmalen beschrieben. Diese Teilmenge wechselt von Aufgabe zu Aufgabe. Häufig ist es sinnvoll, ausgesuchte Merkmale wie Marke oder Preis immer anzuzeigen und nur aus den verbleibenden Merkmalen zu selektieren. Der reduzierte Informationsgewinn je Choice-Set wird durch eine höhere Anzahl Befragter und/oder Aufgaben je Befragtem kompensiert.

Alternative Specific Designs

Nicht immer lassen sich alle Angebote in einer Entscheidungssituation durch dieselben Merkmale beschreiben. Wenn beispielsweise auch die Bewertung der Funktionalität des Verdeckes beim Cabrio interessiert, können sich Variationen dieses Merkmals nur auf diesen Fahrzeugtyp beziehen. Bei Angeboten anderer Fahrzeugtypen bleibt das Merkmal außen vor. Das Auftreten eines Merkmals hängt also von den Ausprägungen eines oder mehrerer anderer Merkmale ab.

Conditional Pricing

Werden die Merkmalausprägungen völlig frei miteinander kombiniert, können unrealistische Angebote entstehen. Um das zu verhindern, lassen sich bestimmte Kombinationen von vornherein ausschließen, was jedoch die Qualität des Erhebungsdesigns reduziert. Erscheinen Angebote aufgrund zu hoher oder zu niedriger Preise unrealistisch, lässt sich der Qualitätsverlust durch ein Conditional Pricing vermeiden. Die Preislevel werden in Abhängigkeit der Ausprägungen eines oder mehrerer Merkmale dargestellt, beispielsweise in Abhängigkeit vom Hubraum. Die Abstände zwischen den Preisstufen bei höher oder geringer motorisierten Varianten bleiben jeweils gleich. Die Unterschiede in der durchschnittlichen Höhe der Preise je Variante spiegeln sich dann in den Teilnutzenwerten der Motorisierungen wider. Höhere Preise stärkerer Motoren mindern voraussichtlich deren Teilnutzenwerte. Eine nachträgliche Bereinigung um den Einfluss der Preislagen ist möglich.

Constant Sum

Charakteristisch am CBC ist die wahlbasierte Entscheidung (discrete choice). Allerdings kann nicht immer von einem solchen Entscheidungsprozess ausgegangen werden. Während dieser für einen privaten Autokäufer adäquat ist, gelten für einen Einkäufer eines Fuhrparks andere Mechanismen. Die Anforderungen an den Fuhrpark sind heterogen und so kommen für ihn unterschiedliche Modelle in Frage. Je nach Bedarf und Präferenz entscheidet er sich zum Beispiel beim Kauf der nächsten zehn Fahrzeuge fünf Mal für ein Modell X, drei Mal für ein Modell Y und zwei Mal für ein Modell Z. Ein solcher Entscheidungsprozess kann durch die Aufteilung der nächsten zehn Käufe auf die gezeigten Angebote abgebildet werden. Der Befragte hat je Option innerhalb eines Choice-Sets eine Anzahl zwischen 0 und 10 anzugeben, wobei die Summe der einzelnen Werte der vorgegebenen Gesamtzahl 10 entsprechen muss – die Summe also konstant ist.

Dual Response None

Die Verwendung einer None-Option erlaubt es dem Befragten kundzutun, dass keines der gezeigten Angebote für ihn in Frage kommt. Dies ist ein entscheidendes Kriterium, um realistische Rahmenbedingungen zu schaffen. Die None-Option liefert wertvolle Informationen über die Einstiegsschwelle der Käufer und bildet somit die Grundlage für eine realistische Marktmodellierung und Potenzialschätzungen. Allerdings liefern Choice-Sets, in denen die None-Option gewählt wurde, keinerlei Informationen über die Relationen der Merkmale zueinander und reduzieren die Genauigkeit der Schätzungen der Teilnutzenwerte. Beide Informationen lassen sich gewinnen, wenn anstelle der None-Option zusätzlich gefragt wird, ob das gewählte Angebot (1. Antwort) tatsächlich gekauft werden würde (2. Antwort). ◀

In Ausgabe 3/2012 wird die Wahl des geeigneten Conjoint-Verfahrens diskutiert.

► Literatur

Chrzan, K.; Elrod, T.: Choice-Based Approach for Large Numbers of Attributes. In: Marketing News, Jg. 29/1995; Nr. 1, S. 20.

Orme, B.: Three Ways to Treat Overall Price in Conjoint Analysis. Sawtooth Research Paper Series. Sequim, 2007.

Sawtooth Software: The CBC Advanced Design Module (ADM) Technical Paper. Sawtooth Software Technical Paper Series. Sequim, 2008.

Auswahl geeigneter Conjoint-Verfahren

Jedes Verfahren der Conjoint-Analyse besitzt Stärken und Schwächen. Insofern lässt sich keine generelle Empfehlung geben, welches vorzugsweise eingesetzt werden sollte. Die Entscheidung für ein Verfahren ist vor dem Hintergrund der Charakteristika der jeweiligen Fragestellung und Erhebung zu treffen. Relevante Kriterien werden nachfolgend vorgestellt.

Anzahl der Merkmale

Menschen sind in der Lage, bis zu acht Merkmale gegeneinander abzuwägen. Soll eine größere Anzahl berücksichtigt werden, muss eine Auswahl derjenigen Merkmale erfolgen, die die im dekompositionellen Teil ganzheitlich zu beurteilenden Konzepte beschreiben. Dies leisten Partial-Profile-CBC, ACA und HILCA. Multistage Designs ermöglichen durch die Gruppierung von Merkmalen komplexe Beschreibungen in mehrere kleinere aufzuteilen und somit insgesamt eine Vielzahl an Merkmalen einzubeziehen.

Interaktionseffekte

Kommt für einen Befragten beispielsweise ein rotes Auto ausschließlich in Kombination mit einer bestimmten Marke in Frage, liegt eine Interaktion zwischen beiden Ausprägungen vor. Die Stärke von Interaktionseffekten kann nur in CBC-Ansätzen bestimmt werden (im HIT-CBC jedoch allein zwischen der besten bzw. schlechtesten Ausprägung von zwei Merkmalen). Liegen Interaktionseffekte vor, erhöht ihre Berücksichtigung im Allgemeinen die Güte der Prognose der Entscheidungen, reduziert jedoch die Genauigkeit der Schätzung der Haupteffekte.

Preisanalyse

Steht der Preis im Fokus einer Untersuchung, sind ACA und Partial-Profile-CBC zu meiden, da sie die relative Bedeutung des Preises häufig unterschätzen.

Stichprobengröße

Mit der KCA/LCA und dem ACBC gelingt es, von einem Befragten am meisten Informationen über seine Präferenzen zu gewinnen, so dass diese mit kleineren Stichproben auskommen. Verfahren, die bei der Konzeptbeurteilung jeweils nur auf einen Teil aller Merkmale zurückgreifen, können den resultierenden Informationsnachteil durch eine größere Stichprobe kompensieren.

Flexibilität bezüglich des Befragungsmodus

In Paper-Pencil-Befragungen sind nur KCA/LCA oder ein (Partial-Profile-) CBC möglich. ACA, ACBC, HIT-CBC und HILCA benötigen eine Computerunterstützung. In reinen CATI-Befragungen kann am ehesten die ACA mit wenigen Merkmalen im dekompositionellen Teil Verwendung finden. Bei Verknüpfung von telefonischen mit einer Online-Befragung lassen sich alle Verfahren nutzen.

Interviewdauer

Adaptive, hybride und hierarchische Ansätze führen zu durchschnittlich längeren Interviews als die einphasigen Verfahren KCA/LCA und CBC.

Realitätsnähe

Am besten gelingt es CBC-Ansätzen, den realen Entscheidungsprozess abzubilden. Partial-Profile-CBC und HIT-CBC verlieren durch die Darstellung nur unvollständig beschriebener Konzepte bzw. die Verwendung ausschließlich extremer Merkmalsausprägungen allerdings an Realitätsnähe. KCA und LCA verlangen vom Befragten eine Ordnung aller Konzepte, das heißt auch derjenigen, die er in der Realität völlig außer Acht lassen würde.

Joy of Use

Erhebungen, die auf gegebene Antworten reagieren und abwechslungsreich sind, motivieren den Befragten, ein Interview sorgfältig zu Ende zu führen. In diesem Sinne fördert das ACBC in Verbindung mit seiner Realitätsnähe am ehesten Interesse und Spaß an der Befragung. ◀

In Ausgabe 4/2012: Marktsimulationen mit Conjoint-Analysen.

► Literatur

Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: Conjoint-Analyse. In: Multivariate Analysemethoden, 13. Auflage. Berlin, Heidelberg: Springer, 2011, S. 457-505.

Hartmann, A.; Sattler, H.: Wie robust sind Methoden zur Präferenzmessung? In: zfbf, Jg. 56/2004, Nr. 2, S. 3-22.

Karger, M.: Prüfung bestehender Varianten der Conjoint Analyse zur Validierung des Customer Perceived Value Accountings. In: Zahlungsbereitschaftsmessung für industrielle Hybride Leistungsbündel. Wiesbaden, 2011, S. 70-102.

Melles, T.: Welche Conjoint-Methode ist die beste? Workshoppräsentation zur Marktforschungsmesse. München, 2011.

Orme, B.: Which Conjoint Method Should I Use? Sawtooth Research Paper Series. Sequim, 2009.

	Klassische/ Limit Conjoint- Analyse (KCA/LCA)	Choice-Based Conjoint (CBC)	Partial- Profile- CBC	Adaptive Conjoint- Analyse (ACA)	Adaptive Choice-Based Conjoint (ACBC)	Hybrid Individualized Two-Level CBC (HIT-CBC)	Hierarchische Individualisierte LCA (HILCA)	Multistage/ Hierarchische Designs
Anzahl der Merkmale	●	●	●	●	●	●	●	●
Interaktionseffekte	●	●	●	●	●	●	●	●
Preisanalyse	●	●	●	●	●	●	●	●
Stichprobengröße	●	●	●	●	●	●	●	*
Flexibilität bez. des Befragungsmodus	●	●	●	●	●	●	●	*
Interviewdauer	●	●	●	●	●	●	●	●
Realitätsnähe	●	●	●	●	●	●	●	●
Joy of Use	●	●	●	●	●	●	●	*

● Stärke ● weder/noch ● Schwäche * abhängig von den verknüpften Verfahren

© Lüken/Schimmelpfennig; planung & analyse 3/12

Abbildung: Stärken und Schwächen der Conjoint-Verfahren

Marktsimulationen mit Conjoint-Analysen

Die Conjoint-Analyse liefert mit individuellen Nutzenwerten eine gute Datenbasis für Marktsimulationen der Auswirkungen von Produkt- und Preisvariationen auf Marktanteile und andere ökonomische Größen.

Grundlegendes Vorgehen

Im Rahmen der im Conjoint-Modell enthaltenen Merkmale und Ausprägungen werden Produkte definiert und zu einem Set zusammengestellt. Metrische Merkmale sind dabei nicht auf die abgefragten Ausprägungen beschränkt. Durch eine Interpolation können auch Ausprägungen zwischen den betrachteten Stufen berücksichtigt werden. Eine Extrapolation über das kleinste bzw. größte abgefragte Level hinaus sollte aber allenfalls bis zu einer halben Breite des Intervalls zwischen zwei Ausprägungen erfolgen.

Für jedes der Produkte wird durch die Summe der individuellen Teilnutzenwerte der jeweiligen Merkmalsausprägungen ein Gesamtnutzenwert für jeden Befragten bestimmt. Anschließend wird überprüft, wie jeder Einzelne reagiert, wenn genau dieses Set an Alternativen zur Auswahl steht. Dazu bedarf es einer Annahme über das Entscheidungsverhalten der Befragten. Gleichgültig welche Auswahlregel zugrunde gelegt wird, ergeben sich die „Marktanteile“ durch die Mittelwerte der individuellen Auswahlwahrscheinlichkeiten für die Alternativen.

Entscheidungsregeln

Die *First Choice-Regel* nimmt an, dass ein Befragter sich für das Angebot entscheidet, das für ihn den höchsten Gesamtnutzenwert erzielt. Für diese Alternative ist die individuelle Auswahlwahrscheinlichkeit gleich eins, für alle anderen gleich null – auch wenn deren Gesamtnutzenwerte nur geringfügig schlechter sind. Nur bei Alternativen mit exakt demselben Gesamtnutzenwert verteilen sich die Wahrscheinlichkeiten gleichmäßig auf diese Alternativen. Bei extensiven und habituellen Kaufentscheidungen ist diese Regel grundsätzlich angemessen, widerspricht bei impulsiven oder limitierten Kaufentscheidungen, in denen die Auswahl nicht eindeutig determiniert ist, aber dem Verhalten.

Im Gegensatz zur First Choice-Regel geht die *Share of Preference-Regel* davon aus, dass für einen Befragten alle Alternativen eines Sets in Frage kommen. Die Auswahlwahrscheinlichkeiten verteilen sich auf die konkurrierenden Angebote gemäß der Höhe der Gesamtnutzenwerte. Das Verfahren weist jedoch Schwächen auf, wenn Produkte in einem Set ähnlich sind: Die Anteile solcher Produkte sind dann unrealistisch hoch. Die *Randomized First Choice-Methode* berücksichtigt, dass es zufallsbedingte Einflüsse auf die Auswahl geben kann. Für jeden Befragten werden mehrere First Choice-Berechnungen durchgeführt. Bei jeder dieser Iterationen werden Zufallswerte zu den Nutzenwerten hinzugefügt, so dass auch Alternativen mit ursprünglich geringerem Gesamtnutzen die beste Alternative werden können. Die individuelle Auswahlwahrscheinlichkeit einer Alternative ist dann bestimmt durch das Verhältnis der Anzahl der Iterationen, in denen sie die beste war, zu der Anzahl aller Iterationen. Wie die Share of Preference-Regel geht die Randomized First Choice-Methode von keiner eindeutigen Wahl des Befragten aus, vermeidet aber bei sinnvoller Spezifikation der Modellparameter die Überschätzung der Anteile ähnlicher Produkte.

Typische Anwendungen

Marktsimulationen ermöglichen vor allem „was wäre, wenn ...“-Betrachtungen. Ausgehend von einem Anfangsszenario lassen sich die Auswirkungen von Veränderungen des Preises oder anderer Merkmale eines oder mehrerer Produkte auf die Anteile oder Umsätze aller Produkte bestimmen. Bei Berücksichtigung von Kosten für die Merkmalsausprägungen können auch Gewinne prognostiziert werden.

Ferner sind Marktsimulationen hilfreich zur Untersuchung der Auswirkungen der Einführung neuer Produkte, um zum Beispiel Kannibalisierungseffekte zu identifizieren, bzw. die grundsätzliche Akzeptanz eines neuen Produktes zu überprüfen.

Daneben lassen sich auch Analysen mit dem Ziel durchführen, die durch die Nachfrager wahrgenommenen geldwerten Vor- oder Nachteile von Produktfeatures zu erkennen. Typischerweise werden hierfür zwei identische Produkte definiert. Eines wird in der zu untersuchenden Komponente verändert und im Preis kontinuierlich variiert. Bei einer Produktverbesserung wird der berechnete Wahlanteil bei gleichem Preis zunächst höher sein als der des Originalproduktes. Eine Erhöhung des Preises führt dann zu einer Verringerung des Anteils. Der Preispunkt, an dem das Original- und das modifizierte Produkt den gleichen Marktanteil erzielen, zeigt die subjektive geldwerte Differenz seitens der Nachfrager.

Optimierung von Produkten

Eine mögliche Aufgabenstellung besteht darin, ein einzelnes Produkt oder ein Portfolio aus mehreren Produkten im Hinblick auf maximalen Absatz, Umsatz oder Gewinn zu optimieren. Aufgrund der schnell ansteigenden Anzahl möglicher Kombinationen aus Merkmalsausprägungen und Preisstellungen sind manuelle Vorgehensweisen hierbei häufig unpraktikabel. Simulationstools bieten daher die Möglichkeit, solche Optimierungen zu automatisieren.

Justierung des Modells

Je weniger berechnete „Marktanteile“ eines Szenarios von den entsprechenden realen Marktanteilen abweichen, desto valider ist das Conjoint-Modell. Allerdings wird es eine exakte Übereinstimmung kaum geben können, da im Allgemeinen ein Modell die Realität nicht vollständig abbildet. Bei der Share of Preference-Regel oder der Randomized First Choice-Methode können jedoch in den meisten Fällen durch eine Justierung der Parameter die „Marktanteile“ sehr weit an die reale Marktsituation angepasst werden. Somit gelingt es, eine Ausgangsbasis für realitätsnahe Simulationen zu schaffen, ohne die Datenbasis zu verändern. ◀

In Ausgabe 5/2012: Einführung in Signifikanztests

► Literatur

Green, P. E.; Krieger, A. M.: Choice Rules and Sensitivity Analysis in Conjoint Simulators. In: Journal of the Academy of Marketing Science, Jg. 16/1988, Nr. 1, S. 114-127.

Huber, J.; Orme, B.; Miller, R.: Dealing with Product Similarity in Conjoint Simulations. In: Sawtooth Software Conference Proceedings, Sequim, 1999, S. 253-266.

Orme, B.: Market Simulators for Conjoint Analysis. In: Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research, 2. Auflage, Madison, 2010, S. 89-103.

Skiera, B.; Gensler, S.: Berechnung von Nutzenfunktionen und Marktsimulationen mit Hilfe der Conjoint-Analyse (Teil 2). In: Wirtschaftswissenschaftliches Studium, Jg. 31/2002, Nr. 5, S. 258-263.

Einführung in Signifikanztests

Signifikanztests untersuchen, inwieweit die auf Grundlage einer Stichprobe erzielten Ergebnisse auch in der dahinter liegenden Grundgesamtheit gelten. Da die wahren Werte von Parametern der Grundgesamtheit wie Mittelwert, Anteil, Korrelations- oder Regressionskoeffizient nicht bekannt sind, werden hypothetische Werte angenommen und überprüft, ob diese mit den Werten in der Stichprobe im Widerspruch stehen.

Ein Beispiel wäre die Untersuchung, ob sich nach dem erstmaligen Verbrauch eines Produktes die Kaufbereitschaft für das Produkt verändert. Dazu wird vorher und nachher dieselbe zufällig ausgewählte Stichprobe der Zielgruppe befragt. Nimmt man an, dass sich im Durchschnitt die Kaufbereitschaft aller Personen der Zielgruppe nicht verändert, so würde als Hypothese folgen: „Die wahren Mittelwerte der Kaufbereitschaft sind vor und nach dem Verbrauch gleich“.

Intuitives Vorgehen

Man berechnet für jeden Befragten der Stichprobe die Differenz zwischen seiner Kaufbereitschaft vor und nach dem Verbrauch und bildet dann den Mittelwert über diese Differenzen. Gegen die aufgestellte Hypothese würde sprechen, wenn der Mittelwert deutlich von null abweicht. Mithilfe eines Signifikanztests wird nun präzisiert, ab wann Abweichungen des Mittelwertes von null in der Stichprobe deutlich oder bloß zufällig sind.

Stichprobenverteilung

Eine vorliegende Stichprobe ist nur eine von vielen möglichen. Geht man von der Richtigkeit der Hypothese aus und würde sämtliche Stichproben des gleichen Umfangs ziehen, so verteilen sich deren Mittelwerte der Differenzen wie im roten Histogramm der Abbildung skizziert. Für sehr viele Stichproben ergibt sich ein Mittelwert, der in der Nähe von null liegt, für sehr wenige Stichproben ein Mittelwert, der sehr weit davon entfernt ist. Aber auch ohne alle möglichen Stichproben gezogen zu haben, lässt sich unter der Bedingung, dass die Hypothese richtig ist, die für einen Signifikanztest maßgebliche Verteilung der Teststatistik ermitteln. Im beschriebenen Beispiel ist der Mittelwert der Differenzen für Stichproben mit einem Umfang größer als 30 annähernd normalverteilt, wie die schwarze Kurve in der Abbildung veranschaulicht.

Signifikanzniveau

Ein sehr großer Mittelwert der Differenzen in einer Stichprobe kann in dem Beispiel zwei Gründe haben: Entweder ist in der Grundgesamtheit der Mittelwert tatsächlich nicht null, oder der Mittelwert von null ist richtig, jedoch hat man eine „ungünstige“ Stichprobe erwischt. Im ersten Fall wird mit Ablehnung der Hypothese die richtige Entscheidung getroffen, im zweiten Fall jedoch eine falsche (Fehler 1. Art).

Da der Mittelwert der Differenzen sowohl positiv als auch negativ sein kann, ist im Hinblick auf die Entscheidung über die Hypothese ein Bereich für positive wie ein Bereich für negative Mittelwerte festzulegen. Diese beiden Bereiche sollen alle diejenigen Mittelwerte umfassen, für die die Hypothese abgelehnt wird, weil sie zu weit von null entfernt sind. In der Abbildung sind diese grau unterlegt. Die Wahrscheinlichkeit des Fehlers 1. Art ist damit unter Annahme der Richtigkeit der Hypothese gleich der Wahrscheinlichkeit, eine Stichprobe zu ziehen, deren Mittelwert der Differenzen in einen dieser beiden Bereiche fällt.

Damit die Wahrscheinlichkeit dieses Fehlers sehr gering ist, könnte man die Bereiche sehr schmal halten. Ist die Hypothese aber tatsächlich falsch, wird dies dann leicht nicht mehr erkannt, und man macht ebenfalls einen Fehler (Fehler 2. Art). Die Wahrscheinlichkeit beider Fehler kann somit nicht gleichzeitig minimiert werden. Insofern ist es üblich, die Wahrscheinlichkeit für den Fehler 1. Art, die gerade noch akzeptiert wird – das Signifikanzniveau α – vorzugeben und einen Test zu verwenden, bei dem die Wahrscheinlichkeit für den Fehler 2. Art möglichst klein wird. Durch die Vorgabe von α können mithilfe von Tabellen der ermittelten Verteilung die Grenzen bestimmt werden, innerhalb derer Abweichungen der Mittelwerte der Differenzen von null noch zufällig sind und außerhalb derer sie zu groß sind, um die Hypothese „Die wahren Mittelwerte sind vor und nach dem Verbrauch gleich“ beizubehalten.

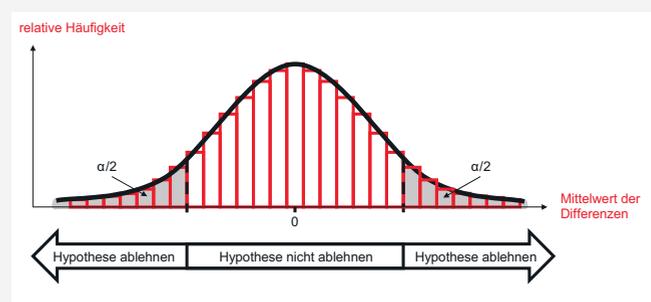
Entscheidung anhand des p-Wertes

Mit dem Einsatz von Statistik-Software erübrigt sich eine Bestimmung der Bereiche, um über eine Hypothese zu entscheiden. Die Software weist mit dem p-Wert die Wahrscheinlichkeit aus, unter der Annahme der Richtigkeit der Hypothese in diesem Beispiel einen Mittelwert der Differenzen (allgemein: eine Teststatistik) zu erhalten, der vom Betrag genauso groß wie oder größer als der in der vorliegenden Stichprobe beobachtete Mittelwert ist. Damit gibt p die konkrete Wahrscheinlichkeit für den Fehler 1. Art an. Eine Entscheidung über die Hypothese kann dann unmittelbar auf der Grundlage des p-Wertes getroffen werden.

Ist die Wahrscheinlichkeit p, mit der Ablehnung der Hypothese einen Fehler zu begehen, kleiner als die maximal akzeptierte Fehler-Wahrscheinlichkeit α wird die Hypothese folglich abgelehnt. Entsprechend der Höhe von p kann dann differenziert werden, ob ein Ergebnis schwach signifikant ($p \leq 0,1$), signifikant ($p \leq 0,05$) oder hoch signifikant ($p \leq 0,01$) ist.

Ist zum Beispiel $p = 0,03$, so ist die Veränderung der Kaufbereitschaft signifikant, das heißt in der Stichprobe so hoch, dass voraussichtlich auch in der Grundgesamtheit ein Unterschied vorliegt – ohne aber eine Aussage über dessen Höhe zu machen. Dazu wäre es nötig, eine andere Hypothese wie beispielsweise „Der wahre Mittelwert der Kaufwahrscheinlichkeit ist nach dem Verbrauch um mindestens 10 Prozentpunkte höher als vor dem Verbrauch“ zu testen. ◀

In Ausgabe 6/2012: Signifikanz und Stichprobenumfang



► Literatur

Bleymüller, J.: Statistik für Wirtschaftswissenschaftler, 16. Auflage, München, 2012.

Eckey, H.-F.; Türck, M.: Statistische Signifikanz (p-Wert). In: Wirtschaftswissenschaftliches Studium, Jg. 35/2006, Nr. 7, S. 415-418.

Siegel, S.: Nichtparametrische statistische Methoden, 5. Auflage, Eschborn, 2001.

Statistik KOMPAKT

Signifikanz und Stichprobenumfang

Nicht selten kommt es bei der Tabellierung von Marktforschungsdaten vor, dass ein Unterschied zwischen zwei Teilgruppen als signifikant und ein Unterschied zwischen zwei anderen Gruppen nicht als signifikant ausgewiesen wird, obwohl dieser größer ist. Eine mögliche Erklärung dafür ist, dass die ersten beiden Gruppen stärker besetzt sind als die anderen beiden.

Signifikanztest

Ein in einer Stichprobe beobachteter Effekt ist signifikant, wenn dieser wahrscheinlich nicht zufällig aufgetreten ist. Man kann dann davon ausgehen, dass ein Effekt auch in der entsprechenden Grundgesamtheit vorhanden ist. Die Prüfung auf Signifikanz erfolgt mit einem statistischen Test. Diesem liegt die Hypothese zugrunde, dass kein Effekt vorliegt. Es wird die Wahrscheinlichkeit p bestimmt, mit dem Verwerfen der Hypothese einen Fehler zu begehen. Ist p kleiner als ein vorgegebenes Signifikanzniveau α (zumeist ist $\alpha = 0,05$), so wird die Hypothese abgelehnt, das heißt der Effekt ist signifikant.

Abhängigkeit der Signifikanz vom Stichprobenumfang

Soll beispielsweise überprüft werden, ob sich das Ausprobieren eines Produkts positiv auf die Kaufbereitschaft auswirkt, wird die Kaufbereitschaft vor und nach dem Ausprobieren bei denselben zufällig ausgewählten Befragten auf einer Ratingskala erhoben. Im Mittel ergibt sich in der Stichprobe eine um 0,2 Skalenpunkte höhere Kaufbereitschaft nach dem Ausprobieren. Ist diese Erhöhung signifikant? Hierfür wird die einseitige Hypothese getestet, dass sich die Kaufbereitschaft nach dem Ausprobieren nicht erhöht. In Abbildung 1 sind die für mögliche Stichprobenumfänge und verschiedenen große Grundgesamtheiten resultierenden p -Werte dargestellt. Obwohl allen derselbe Mittelwert der Differenzen zwischen der Kaufbereitschaft vor und nach dem Ausprobieren sowie dieselbe Standardabweichung der Differenzen zugrunde liegt, unterscheiden sich die p -Werte. Im Fall einer sehr großen Grundgesamtheit ist die Verbesserung der durchschnittlichen Kaufbereitschaft für $n = 100$ und $n = 150$ nicht signifikant, da p größer als 0,05 ist. Für $n = 200$ ist sie dagegen signifikant.

Stichprobenumfang	Grundgesamtheit sehr groß	Grundgesamtheit N = 1000	Grundgesamtheit N = 500
n = 100	0,11	0,10	0,08
n = 150	0,06	0,05	0,03
n = 200	0,04	0,02	0,01

Abbildung 1: p -Werte für verschiedene Umfänge

Das heißt bei einem größeren ist es wahrscheinlicher als bei einem kleineren Stichprobenumfang, dass die Hypothese abgelehnt wird und ein Effekt signifikant ist. Insofern spricht das Verwerfen einer Hypothese bei einer kleinen Stichprobe für einen stärkeren Effekt in der Grundgesamtheit als bei einer großen Stichprobe.

Statistischer Hintergrund

Ob ein Effekt signifikant ist, hängt vom Ergebnis in der Stichprobe ebenso ab wie von der Stichprobenverteilung. Die Stichprobenverteilung

beschreibt die Verteilung beispielsweise des Mittelwerts für alle denkbaren Stichproben eines bestimmten Umfangs. Die Abbildung 2 zeigt die Stichprobenverteilungen für zwei unterschiedliche Stichprobenumfänge. Bei einem großen Stichprobenumfang ist die Verteilung viel schmäler, das heißt die Standardabweichung des Mittelwerts – der Standardfehler – kleiner. Die hellblaue Fläche unter der Kurve entspricht dem p -Wert beim großen Stichprobenumfang. Beim kleinen Stichprobenumfang ist der p -Wert trotz des gleichen Ergebnisses in der Stichprobe aufgrund des höheren Standardfehlers um die graue Fläche größer.

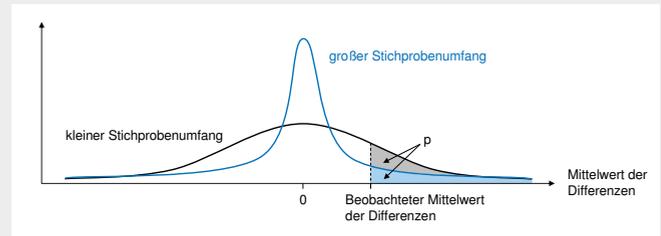


Abbildung 2: Stichprobenverteilung

Konsequenz für gewichtete Daten

Eine Fallgewichtung verändert nicht nur das Ergebnis in der Stichprobe wie beispielsweise den Mittelwert. Der gewichtete Stichprobenumfang besitzt ebenfalls Einfluss auf den Ausgang des Signifikanztests. Um die Signifikanz eines Effekts nicht ungerechtfertigt zu begünstigen, sollte unbedingt vermieden werden, dass die Summe der Gewichte größer als der ursprüngliche Stichprobenumfang ist. Um die aufgrund der Gewichtung erhöhte Unsicherheit bei der Schätzung des Mittelwerts zu berücksichtigen, wird vielmehr empfohlen, einen reduzierten Stichprobenumfang – die so genannte effective base – anzusetzen, die umso kleiner ist, je höher die Varianz der Fallgewichtungen ist.

Kleine Grundgesamtheiten und Vollerhebungen

Je kleiner die Grundgesamtheit im Verhältnis zum Stichprobenumfang ist, desto geringer ist der Standardfehler, wodurch die Stichprobenverteilung schmalgipfliger und der p -Wert damit kleiner wird. Der Vergleich der p -Werte in Abbildung 1 veranschaulicht diesen Zusammenhang.

Auf Basis einer Vollerhebung lassen sich die wahren Werte bestimmen. Das Formulieren von Hypothesen über die Werte in der Grundgesamtheit und Signifikanztests erübrigen sich dann. In dem Beispiel würde jede noch so kleine positive Abweichung des Mittelwerts von null einen positiven Effekt des Ausprobierens auf die Kaufbereitschaft bedeuten. Sinnvoll sind Signifikanztests dann nur, um zu überprüfen, ob ein beobachteter Effekt aufgrund von Messfehlern zustande gekommen ist. ◀

In Ausgabe 1/2013: Einführung in Kausalanalysen

► Literatur

Behnke, J.: Lassen sich Signifikanztests auf Vollerhebungen anwenden? Einige essayistische Anmerkungen. In: Politische Vierteljahresschrift, Jg. 46/2005, Heft 1, S. 0-1-0-15.

Kish, L.: Survey Sampling, New York et al., 1965, S. 427.

Lipovetzky, S.: Post-Stratification with Optimized Effective Base. In: Proceedings of the Survey Research Methods Section, American Statistical Association, 2007, S. 2313-2320.

Einführung in Kausalanalysen

Kausale Zusammenhänge beschreiben Beziehungen zwischen Variablen in Form von Wenn-dann- bzw. Je-desto-Sätzen. Die Wenn- oder Je-Komponente enthält die Ursache bzw. unabhängige Variable, die Dann- oder Desto-Komponente die Wirkung bzw. abhängige Variable. Mit Kausalanalysen werden Vermutungen über kausale Zusammenhänge getestet, um diejenigen Treiber zu identifizieren, die tatsächlich einen Einfluss auf eine Zielgröße haben. Dazu können verschiedene Methoden eingesetzt werden, die für die Prüfung der Anforderungen an einen kausalen Zusammenhang unterschiedlich gut geeignet sind.

Anforderungen an einen kausalen Zusammenhang

Von einem kausalen Zusammenhang zwischen einer unabhängigen und einer abhängigen Variable kann man ausgehen, wenn

- zwischen der unabhängigen Variable und der abhängigen Variable ein statistischer Zusammenhang besteht,
- die Veränderung der unabhängigen Variable der der abhängigen Variable zeitlich vorausgeht, und
- alternative Erklärungen für den statistischen Zusammenhang ausgeschlossen werden können.

Wird zum Beispiel vermutet, dass die wahrgenommene Produktqualität ein Treiber der Kundenzufriedenheit ist, so

- müssen Produktqualität und Zufriedenheit miteinander korrelieren,
- muss sich die Produktqualität verändert haben, bevor sich die Zufriedenheit verändert, und
- dürfen Veränderungen von Produktqualität und Zufriedenheit nicht auf andere (dritte) Variablen zurückzuführen sein,

damit der kausale Zusammenhang zwischen den beiden Variablen bestätigt wird.

Überprüfung mittels randomisierten Experimenten

Als ideal für den Test eines kausalen Zusammenhangs gilt das randomisierte Experiment. Die Befragten werden zufällig in so viele Gruppen aufgeteilt, wie Ausprägungen der unabhängigen Variable untersucht werden sollen. Jede andere Variable weist dann bei einer großen Stichprobe in den Gruppen abgesehen von Zufallsschwankungen die gleiche Verteilung auf. Anschließend werden die Ausprägungen der unabhängigen Variable zufällig den Gruppen zugewiesen. Beispielsweise werden zur Überprüfung des Einflusses einer Werbung auf die Kaufbereitschaft die Befragten zufällig in zwei Gruppen aufgeteilt. Einer Gruppe wird ein Werbespot gezeigt und einer zweiten (Kontroll-)Gruppe nicht. Unterschiede zwischen den Gruppen in der abhängigen Variable (Kaufbereitschaft) können dann eindeutig auf die unabhängige Variable (Werbespot gesehen oder nicht) zurückgeführt werden, da sich die Gruppen in den Variablen vor dem Anschauen des Spots nicht unterscheiden haben. Damit lässt sich der durchschnittliche kausale Effekt durch einen Vergleich der Mittelwerte der abhängigen Variable zwischen den Gruppen oder bei vielen Ausprägungen der unabhängigen Variable durch die Korrelation der Werte der unabhängigen und der abhängigen Variable bestimmen. Zudem stellt ein experimentelles Design die zeitliche Asymmetrie von Ursache und Wirkung sicher.

Überprüfung bei nicht-experimentellen Designs

In nicht-experimentellen Designs werden Ausprägungen der unabhängigen Variable bei den Befragten nicht künstlich herbeigeführt, sondern liegen bereits vor und diktieren damit die Einteilung in Gruppen mit jeweils gleichen Ausprägungen der unabhängigen Variable. Allerdings werden sich auch die Werte anderer Variablen zwischen den Gruppen unterscheiden. Beispielsweise könnten in einer Gruppe, die die Produktqualität als gut beurteilt, sowohl die Kundenzufriedenheit als auch das Markenimage besser sein als in einer anderen Gruppe. Der statistische Zusammenhang zwischen der wahrgenommenen Qualität und der Zufriedenheit kann dann ebenso auf einen Einfluss des Images auf diese beiden Variablen zurückzuführen sein (siehe Abbildung).

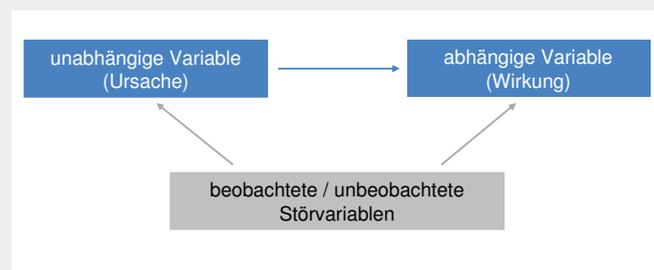


Abbildung: Einfluss von Störgrößen bei kausalen Zusammenhängen

Um die Gruppen vergleichbar zu machen, sind solche störenden Variablen zu kontrollieren. Wenn Daten zu einem einzigen Zeitpunkt erhoben werden (Querschnittsdaten), ist die Regressionsanalyse eine mögliche Methode. Werden alle Störvariablen, die die Ursache und die Wirkung beeinflussen, mit in den Regressionsansatz aufgenommen, wird deren Einfluss auf die beiden interessierenden Variablen herausgerechnet. Strukturgleichungsmodelle (häufig als Kausalanalysen bezeichnet) stellen eine Erweiterung der Regressionsanalyse für die Analyse komplexerer kausaler Zusammenhänge mit mehreren abhängigen Variablen dar.

Bei Querschnittsdaten kann die zeitliche Asymmetrie von Ursache und Wirkung nur durch eine entsprechende Annahme als erfüllt angesehen werden. Des Weiteren ist die Kontrolle unbeobachteter Störvariablen ausgeschlossen. Liegen Längsschnittdaten aus einem Panel vor, können diese Einschränkungen unter der Annahme der Konstanz der unbeobachteten Variablen und ihrer Einflüsse mit Fixed-Effect-Modellen überwunden werden. ◀

In Ausgabe 2/2013: Identifizierung von Scheinkorrelationen

► Literatur

Agresti, A.; Finlay, B.: Introduction to Multivariate Relationships. In: Statistical Methods for the Social Sciences, Upper Saddle River, 2009, S. 301-315.

Legewie, J.: Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jg. 64/2012, Heft 1, S. 123-153.

Opp, K.-D.: Kausalität als Gegenstand der Sozialwissenschaften und der multivariaten Statistik. In: Wolf, C.; Best, H. (Hrsg.): Handbuch der sozialwissenschaftlichen Datenanalyse, Wiesbaden, 2010, S. 9-38.

Identifizierung von Scheinkorrelationen

Kontrolle von Störvariablen mit der Regressionsanalyse

Ein statistischer Zusammenhang wie beispielsweise eine positive Korrelation zwischen Produktqualität und Kundenzufriedenheit genügt allein nicht, um die Produktqualität als Treiber der Zufriedenheit auszumachen. Es muss zudem ausgeschlossen sein, dass dieser statistische Zusammenhang auf Störvariable(n) zurückzuführen ist. So könnte etwa das Markenimage beide Größen beeinflussen und damit die Korrelation zumindest zu einem Teil hervorrufen. Mithilfe der Regressionsanalyse kann der Einfluss einer Störvariable auf Ursache und Wirkung eines vermuteten kausalen Zusammenhangs kontrolliert und somit eine Scheinkorrelation aufgedeckt werden.

Lineare Regressionsanalyse

Mit der Regressionsanalyse wird die Abhängigkeit einer metrischen Variable (Regressand) von einer oder mehreren metrischen und/oder dichotomen Variablen (Regressoren) untersucht. Für jeden Regressor bestimmt sie einen Koeffizienten, der dessen Einflussstärke auf den Regressanden misst. Damit lässt sich eine Regressionsfunktion aufstellen, mittels der für jeden Fall der Stichprobe auf Basis der beobachteten Werte der Regressoren ein Wert des Regressanden berechnet (prognostiziert) werden kann. Die Differenz zwischen dem prognostizierten und dem beobachteten Wert des Regressanden ist das Residuum. Die Regressionskoeffizienten werden so berechnet, dass die Summe der quadrierten Residuen über alle Befragten minimiert wird (Ordinary Least Squares (OLS)-Schätzung).

Beispiel zur Kontrolle einer Störvariable

Vermutet wird, dass ein kausaler Zusammenhang zwischen der Ursache x und der Wirkung y besteht. Betrachten wir dies in einem Beispiel, bei dem die Korrelation zwischen x und y 0,44 beträgt. Für die Korrelationen einer möglichen Störvariable mit der Ursache und Wirkung werden im Weiteren drei Situationen unterschieden*.

Situation 1: z_1 korreliert mit Ursache und Wirkung.

Mit Hilfe jeweils einer einfachen Regression wird der Teil der Ursache beziehungsweise der Wirkung berechnet, der allein durch die Störvariable determiniert ist. Die Residuen für x und y spiegeln dann die Werte der Ursache beziehungsweise der Wirkung wider, die nach Herausrechnen des Einflusses von z_1 übrig bleiben. Ist die Korrelation dieser Residuen 0 ist der positive statistische Zusammenhang zwischen Ursache und Wirkung vollständig auf den Einfluss der Störvariable auf die beiden Variablen zurückzuführen. Die Korrelation von 0,44 zwischen x und y ist eine *Scheinkorrelation*.

Der Umweg über die Berechnung der Residuen ist jedoch nicht notwendig, um den Einfluss der Störvariable auszuschalten. In einer multiplen Regression gibt ein Regressionskoeffizient die Veränderung des Regressanden an, wenn sich der Wert des Regressors um eine Einheit verändert – bei Konstanz der übrigen Regressoren. Um Regressoren mit unterschiedlichen Skalierungen vergleichbar zu machen, wird der standardisierte Regressionskoeffizient berechnet. Eine multiple Regression mit der Ursache und der Störvariable als Regressoren sowie der Wirkung als Regressand führt zu einem standardisierten Regressions-

koeffizienten der Ursache von 0 (siehe rote Werte in der Abbildung). Das heißt bei konstanter Störvariable verändert sich y nicht, wenn sich x ändert. Damit besteht zwischen Ursache und Wirkung offensichtlich kein kausaler Zusammenhang.

Situation 2: z_2 korreliert nur mit der Wirkung.

Der standardisierte Regressionskoeffizient der Ursache in der multiplen Regression ist gleich 0,44 (siehe violette Werte in der Abbildung). Das heißt er ist genauso hoch wie die Korrelation zwischen Ursache und Wirkung. Trotz ihres Einflusses auf y ist z_2 somit keine Störvariable, die den statistischen Zusammenhang zwischen Ursache und Wirkung beeinträchtigt.

Situation 3: Im Vergleich zu Situation 1 ist die Korrelation zwischen Störvariable (z_3) und Ursache gleich, die Korrelation zwischen Störvariable und Wirkung aber geringer. Der standardisierte Regressionskoeffizient der Ursache ist in der multiplen Regression mit 0,28 positiv (siehe grüne Werte in der Abbildung). Insofern ist nicht der gesamte statistische Zusammenhang zwischen Ursache und Wirkung auf die Störvariable zurückzuführen. In dieser Situation ist der durch die Störvariable erklärte Teil der Varianz der Wirkung geringer als in Situation 1, so dass Varianz übrig bleibt, die durch x erklärt wird.

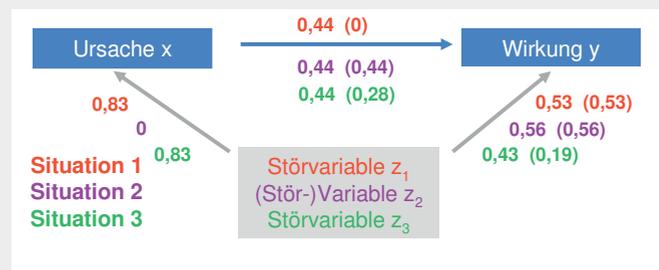


Abbildung: Korrelationen (Standardisierte Regressionskoeffizienten)

Fazit

Eine Regression, die im Regressionsansatz gleichzeitig Ursache und Störvariable enthält, berechnet mit dem (standardisierten) Regressionskoeffizienten der Ursache den statistischen Zusammenhang zwischen Ursache und Wirkung, der um den Einfluss dieser Störvariable bereinigt ist. Dieser Koeffizient ist – neben der Korrelation zwischen Ursache und Wirkung – bestimmt durch die Korrelation zwischen Ursache und Störvariable sowie durch die Korrelation der Störvariablen mit der Wirkung, falls Ursache und Störvariable korrelieren. Bei mehreren Störvariablen kann analog vorgegangen werden, indem diese gemeinsam mit der Ursache in den Regressionsansatz aufgenommen werden. ◀

* Der zugrunde liegende Datensatz kann bei den Autoren (hschimmel-pfennig@ifad.de) angefordert werden.

In Ausgabe 3/2013: Strukturgleichungsmodellierung

► Literatur

Bortz, J.; Schuster, C.: Partielle Korrelation und multiple lineare Regression. In: Statistik für Human- und Sozialwissenschaftler, 7. Auflage, Berlin, Heidelberg, 2010, S. 339-361.

Firebaugh, G.: The Fifth Rule: Compare Like With Like. In: Seven Rules for Social Research, Princeton, Oxford, 2008, S. 120-164.

Strukturgleichungsmodellierung

Die Strukturgleichungsmodellierung zählt zu den Methoden der Kausalanalyse. Sie erweitert die multiple Regressionsanalyse und ermöglicht, nicht nur kausale Zusammenhänge mehrerer unabhängiger und einer abhängigen Variable, sondern auch komplexere Zusammenhänge zwischen Variablen zu überprüfen. Auch nicht unmittelbar beobachtbare latente Variablen wie Einstellung, Involvement oder Loyalität können berücksichtigt werden.

Bestandteile eines Strukturgleichungsmodells

Üblicherweise werden die vermuteten Zusammenhänge zwischen den Variablen in einem Pfaddiagramm grafisch veranschaulicht (siehe Abbildung). Latente Variablen (LV) werden durch Ellipsen gekennzeichnet. Gemeinsam mit den Pfeilen, die zeigen, welche latenten Variablen sich

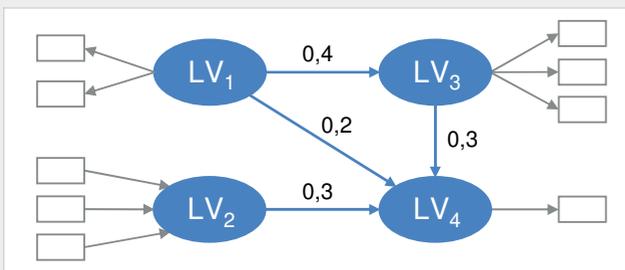


Abbildung: Beispiel eines Strukturgleichungsmodells

beeinflussen, bilden sie das Strukturmodell. Da eine LV sowohl andere bestimmen als auch von anderen bestimmt werden kann, wird nicht zwischen unabhängigen und abhängigen, sondern zwischen exogenen und endogenen LV differenziert. Diejenigen, die von keiner LV beeinflusst werden, sind exogen (in der Abbildung LV₁ und LV₂). Die übrigen sind endogen (in der Abbildung LV₃ und LV₄).

Zur Messung einer LV bedarf es eines oder mehrerer Indikatoren, die direkt beobachtbar sind. Diese werden im Pfaddiagramm als Rechtecke dargestellt. Die jeweils zu einer latenten Variable gehörenden Indikatoren bilden ihr Messmodell. Grundlegend ist die Unterscheidung zwischen reflektiven und formativen Messmodellen. Einem reflektiven Messmodell liegt die Annahme zugrunde, dass die LV verantwortlich für die Ausprägungen der Indikatoren ist (in der Abbildung zum Beispiel LV₁). In einem formativen Messmodell machen die Indikatoren zusammen die LV aus (in der Abbildung LV₂).

Ein Beispiel ist die Messung von Trunkenheit: Diese kann reflektiv z. B. durch den Atemalkohol und die Fähigkeit, auf einer geraden Linie gehen zu können, oder aber formativ durch die konsumierten Mengen an Bier, Wein etc. gemessen werden. Unterschieden werden zwei Arten formativer Messmodelle: die beobachteten Indikatoren bestimmen die LV entweder vollständig oder nur zu einem Teil. Wird angenommen, dass Trunkenheit durch die Mengen der verschiedenen Alkoholika vollständig erfasst ist, wird die LV durch diese Indikatoren auch inhaltlich eindeutig definiert: die Gesamtmenge des Alkoholkonsums. Geht die inhaltliche Bedeutung einer Definition des Begriffs *Trunkenheit* aber darüber hinaus, so würde sie durch die konsumierten Mengen nur unvollständig gemessen.

Schätzverfahren

Zur Schätzung der Pfadkoeffizienten, das heißt der Stärke der Einflüsse der LV aufeinander, sowie der Beziehungen zwischen den LV und ihren Indikatoren, haben sich zwei Ansätze etabliert. Die Kovarianzstrukturanalyse, die auch unter der Bezeichnung Lisrel (Linear Structural Relationships) geläufig ist, betrachtet ein Strukturgleichungsmodell als Ganzes. Sie bestimmt die Koeffizienten so, dass die Kovarianzen der Indikatoren möglichst gut mit den aufgrund der geschätzten Koeffizienten reproduzierten Kovarianzen übereinstimmen. Im Gegensatz dazu zerlegt das Partial-Least-Squares (PLS-)Verfahren das Gesamtmodell in Teilmodelle von unmittelbar zusammenhängenden Variablen und berechnet Fallwerte für die LV. Die Pfadkoeffizienten werden durch multiple Regressionen bestimmt, in denen jeweils eine endogene latente Variable die abhängige Größe ist. Eng verwandt mit PLS ist RALV (Relationships Among Latent Variables), womit insbesondere Verzerrungen der geschätzten Koeffizienten aufgrund von Multikollinearität vermieden werden können.

Beide Verfahren weisen Einschränkungen bei der Berücksichtigung formativer Messmodelle auf. Die Kovarianzstrukturanalyse kann diese nur für exogene, PLS dagegen für exogene und endogene latente Variablen valide abbilden. Allerdings gehen formative Messmodelle in PLS grundsätzlich davon aus, dass die LV durch die beobachteten Indikatoren vollständig bestimmt ist. Dagegen kann in der Kovarianzstrukturanalyse durch eine Residualgröße berücksichtigt werden, dass die Bedeutung der LV über die erhobenen Indikatoren hinausgeht.

Insbesondere bei kleineren Stichproben zeigt sich ein Nachteil der Kovarianzstrukturanalyse. Sie liefert oft kein brauchbares Ergebnis. Entweder ist die Lösung unzulässig, zum Beispiel aufgrund geschätzter negativer Varianzen, oder man erhält gar keine Lösung, weil das iterative Verfahren zur Berechnung der Koeffizienten nicht konvergiert. Nur durch zusätzliche Restriktionen, beispielsweise durch das Fixieren eines Pfadkoeffizienten auf einen bestimmten Wert, oder das Weglassen von Variablen, ergibt sich dann möglicherweise ein zulässiges Ergebnis. In PLS und verwandten Ansätzen treten diese Probleme nicht auf.

Indirekte und totale Effekte

Typisch für Strukturgleichungsmodelle sind indirekte Beziehungen zwischen LV. In dem gezeigten Modell bestimmt LV₁ die Variable LV₄ sowohl direkt als auch indirekt über LV₃. Ergeben sich beispielsweise die in der Abbildung dargestellten Pfadkoeffizienten, so ist die Stärke des indirekten Effekts gleich dem Produkt der beiden Koeffizienten $0,4 \times 0,3 = 0,12$. Zusammen mit dem direkten Effekt von $0,2$ beträgt der totale Effekt somit $0,2 + 0,12 = 0,32$. Auch wenn der direkte Effekt kleiner ist, besitzt LV₁ insofern einen stärkeren Einfluss auf LV₄ als LV₂. ◀

In Ausgabe 4/2013: Interaktionseffekte

► Literatur

Homburg, C.; Pflesser, C.; Klarmann, M.: Strukturgleichungsmodelle mit latenten Variablen: Kausalanalyse. In: Hermann, A.; Homburg, C.; Klarmann, M. (Hrsg.): Handbuch Marktforschung, 3. Auflage, Wiesbaden, 2008, S. 547-577.

Lüken, J.; Schimmelpfennig, H.: Erfolgswirkung von Marken-Touchpoints. Neue Wege der Strukturgleichungsmodellierung für Treiberanalysen. In: planung & analyse, Jg. 38/2011, Nr. 4, S. 61-64.

Scholderer, J.; Balderjahn, I.: Was unterscheidet harte und weiche Strukturgleichungsmodelle nun wirklich? Ein Klärungsversuch zur Lisrel-PLS-Frage. In: Marketing ZFP, Jg. 28/2006, Nr. 1, S. 57-70.

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:

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

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.

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. ◀

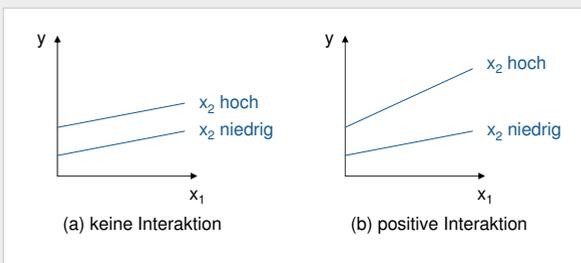


Abbildung: Regressionsgeraden für verschiedene Werte von x_2

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.

Kategoriale Variablen in Regressionsmodellen

Regressionsmodelle sind nicht beschränkt auf metrische unabhängige Variablen. Kategoriale Variablen wie Geschlecht, Beruf etc. können Berücksichtigung finden, wenn ihre Ausprägungen als Zahlen dargestellt werden. Eine gängige Vorgehensweise ist die Dummy-Codierung.

Dummy-Codierung unabhängiger dichotomer Variablen

Es soll untersucht werden, welchen Einfluss neben dem Preis das Schalten einer Werbung auf den monatlichen Absatz besitzt. Die lineare Regressionsfunktion ist somit

$$\text{Absatzmenge} = b_0 + b_1 \cdot \text{Preis} + b_2 \cdot \text{Werbung}$$

Während der Preis eine metrische Variable ist, weist die Werbung nur zwei Kategorien auf: es wurde Werbung (zu Beginn eines Monats) geschaltet oder nicht. Um diese Einflussgröße im Regressionsmodell zu berücksichtigen, sind beiden Ausprägungen Zahlen zuzuordnen. Folgt man der Dummy-Codierung, ist einer Referenzkategorie der Wert 0 und der anderen Kategorie der Wert 1 zu geben. In diesem Beispiel bietet es sich an, als Referenzkategorie den Verzicht auf Werbung festzulegen. Der Regressionskoeffizient b_2 gibt dann genau die Menge an, um die sich der Absatz durch das Schalten einer Werbung gegenüber der Referenzkategorie „keine Werbung“ bei konstantem Preis verändert.

Dummy-Codierung unabhängiger Variablen mit mehr als zwei Kategorien

Es wird zusätzlich differenziert, ob eine TV- oder Print-Werbung geschaltet wurde. Insofern sind drei Kategorien zu unterscheiden. Damit

	W ₁	W ₂	W ₃
TV-Werbung	1	0	0
Print-Werbung	0	1	0
keine Werbung	0	0	1

Abbildung 1: Dummy-Codierung

bedarf es zur Codierung der zwei Variablen $W(\text{erbung})_1$ und $W(\text{erbung})_2$ (siehe Abbildung 1).

Die Kombination der Variablen mit den Ausprägungen $W_1 = 1, W_2 = 0$ repräsentiert somit TV-Werbung, $W_1 = 0, W_2 = 1$ Print-Werbung und $W_1 = 0, W_2 = 0$ keine Werbung. Durch diese Kombinationen sind alle drei Kategorien eindeutig definiert. Eine dritte Variable W_3 wäre nicht nur redundant, sondern würde zu exakter Multikollinearität führen, so dass das Regressionsmodell nicht schätzbar wäre. „Keine Werbung“ ist auch hier die Referenzkategorie, da für diese beide Codiervariablen gleich 0 sind. In der entsprechenden Regressionsfunktion

$$\text{Absatzmenge} = b_0 + b_1 \cdot \text{Preis} + b_2 \cdot W_1 + b_3 \cdot W_2$$

quantifizieren b_2 und b_3 die Wirkungen der TV- bzw. Print-Werbung auf die Absatzmenge im Vergleich zur Referenzkategorie. Die Differenz zwischen b_2 und b_3 gibt an, um wie viel sich die Wirkung einer Werbung zwischen den beiden Medien unterscheidet.

Interaktionseffekte mit kategorialen Variablen

Die Interpretation der Regressionskoeffizienten geht davon aus, dass keine Mehrfachnennungen für die kategoriale Variable vorliegen. Das heißt, es darf im selben Monat nicht in TV und Print geworben worden sein. Um auch den Effekt gemeinsamer Werbung in beiden Medien zu bestimmen, ist eine eigene zusätzliche Kategorie „TV & Print“ zu berücksichtigen (siehe Abbildung 2).

	TV & Print als eigene Kategorie			TV & Print als Interaktionseffekt		
	W ₁	W ₂	W ₃	TV	Print	TV&Print
TV-Werbung	1	0	0	1	0	0
Print-Werbung	0	1	0	0	1	0
TV & Print-Werbung	0	0	1	1	1	1
keine Werbung	0	0	0	0	0	0

Abbildung 2: Dummy-Codierung bei Mehrfachnennungen

Alternativ lassen sich TV-Werbung (ja/nein) und Print-Werbung (ja/nein) als zwei eigenständige dichotome Variablen auffassen. Geht man davon aus, dass gleichzeitige Werbung in beiden Medien nicht additiv wirkt, besteht zwischen diesen ein Interaktionseffekt. Dieser kann durch Aufnahme des Produkts der beiden Variablen im Modell abgebildet werden:

$$\text{Absatzmenge} = b_0 + b_1 \cdot \text{Preis} + b_2 \cdot \text{TV} + b_3 \cdot \text{Print} + b_4 \cdot \text{TV} \cdot \text{Print}$$

Die Wirkung gemeinsamer Werbung ist dann gleich der Summe der Einzeleffekte und des Interaktionseffekts ($b_2 + b_3 + b_4$). Diese entspricht dem Regressionskoeffizienten der Codiervariable W_3 aus Abbildung 2, falls TV & Print-Werbung als eigene Kategorie dargestellt wird.

Ebenso kann ein Interaktionseffekt zwischen Werbung und Preis, das heißt zwischen einer kategorialen und einer metrischen Variable, im Modell berücksichtigt werden. Im einführenden Beispiel der Dummy-Codierung der dichotomen Variable Werbung wird ihr Produkt mit dem Preis in die Regressionsfunktion aufgenommen:

$$\text{Absatzmenge} = b_0 + b_1 \cdot \text{Preis} + b_2 \cdot \text{Werbung} + b_3 \cdot \text{Preis} \cdot \text{Werbung}$$

Angenommen, der Zusammenhang zwischen Preis und Absatzmenge ist negativ ($b_1 < 0$), so bedeutet ein negativer Koeffizient b_3 , dass die Wirkung des Preises auf den Absatz bei Schalten einer Werbung stärker ist als ohne Werbung. Ein positiver Koeffizient deutet dagegen auf geringere Preissensibilität hin.

Neben der Dummy-Codierung sind die Effekt- und die Kontrast-Codierung übliche Vorgehensweisen. Die Art der Codierung beeinflusst zwar die Regressionskoeffizienten und deren Interpretation. Das Bestimmtheitsmaß und damit Ergebnisse der Prüfungen der Signifikanz von Verbesserungen des Bestimmtheitsmaßes infolge der Berücksichtigung weiterer Variablen oder von Interaktionseffekten sind davon jedoch unabhängig. ◀

In Ausgabe 6/2013: Discrete Choice Modelle

► Literatur

Cohen, J.; Cohen, P.; West, S. G.; Aiken, L. S.: Interactions With Categorical Variables, In: Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3. Auflage, Mahwah, New Jersey, 2003, S. 354-389.

Eid, M.; Gollwitzer, M.; Schmitt, M.: Multiple Regressionsanalyse, In: Statistik und Forschungsmethoden, 2. Auflage, Weinheim, Basel, 2011, S. 648-677.

**Statistik
KOMPAKT** *Discrete Choice
Modelle*

Eine bedeutende Aufgabenstellung der Marktforschung ist die Bestimmung der maßgeblichen Einflüsse auf Kaufentscheidungen. Die zu erklärende Variable bildet die Entscheidung ab und ist somit kategorial. Zur Analyse derartiger diskreter Auswahl-situationen kommen häufig Logit-Modelle zum Einsatz.

Binäres Logit-Modell

Im einfachsten Fall stehen zwei Alternativen zur Auswahl, zum Beispiel was die Entscheidung für oder gegen den Kauf eines Produktes bestimmt. Abbildung 1 zeigt einen Ausschnitt aus einer möglichen Datenbasis, um die Einflüsse von Alter, Einkommen und Haushaltsgröße auf die Kaufentscheidung zu untersuchen.

#	Kauf	Alter	Einkommen	HH-Größe
1	Ja	45	2700 €	2
2	Nein	28	3600 €	1
3	Nein	61	4500 €	4

Abbildung 1: Daten für ein binäres Logit-Modell

Die Grundidee eines Logit-Modells ist, auf Basis der unabhängigen Variablen, die metrisch und/oder kategorial sein können, die Wahrscheinlichkeit zu schätzen, dass eine Alternative ausgewählt wird. Dazu werden für jede der beiden Alternativen „Kauf“ und „Kein Kauf“ die Werte der unabhängigen Variablen wie in einem linearen Regressionsansatz mit Regressionskoeffizienten gewichtet und zu einem Wert zusammengefasst. Zu jeder Alternative gehört ein eigenes Set von Koeffizienten. Für die Alternative „Kauf“ ist somit

$$y_{\text{Kauf}} = b_1^{\text{Kauf}} \cdot \text{Alter} + b_2^{\text{Kauf}} \cdot \text{Einkommen} + b_3^{\text{Kauf}} \cdot \text{Haushaltsgröße}$$

Die Wahrscheinlichkeit, dass das Produkt gekauft wird, ist dann

$$p_{\text{Kauf}} = \frac{\exp(y_{\text{Kauf}})}{\exp(y_{\text{Kauf}}) + \exp(y_{\text{Kein Kauf}})}$$

und dementsprechend die Wahrscheinlichkeit, dass es nicht gekauft wird, $1 - p_{\text{Kauf}}$. Zur Schätzung des Modells werden alle Regressionskoeffizienten einer Alternative – zum Beispiel die für „Kein Kauf“ – auf 0 gesetzt und die Koeffizienten der Alternative „Kauf“ mittels Maximum-Likelihood-Schätzung bestimmt.

Abbildung 2 veranschaulicht den sich aus der Berechnung ergebenden typischen logistischen Funktionsverlauf der Auswahl- bzw. Kaufwahrscheinlichkeit. Im Gegensatz zu einer linearen Regression ist für die Veränderung der abhängigen Variable – der Wahrscheinlichkeit – infolge der Änderung einer unabhängigen Variable entscheidend, wo man sich auf der Kurve befindet, das heißt welche konkreten Werte die unabhängigen Variablen aufweisen. Geeignete Tests wie der Likelihood-Ratio-Test ermöglichen die Bestimmung der Signifikanz der Einflussgrößen.

Multinomiales Logit-Modell

Ein Multinomiales Logit-Modell (MNL-Modell) stellt die Verallgemeinerung des binären Logit-Modells auf mehr als zwei zur Auswahl stehende Alternativen dar, um zum Beispiel die Auswahlentscheidung zwischen mehreren Produkten abzubilden. Wie im binären Logit-Modell sind die

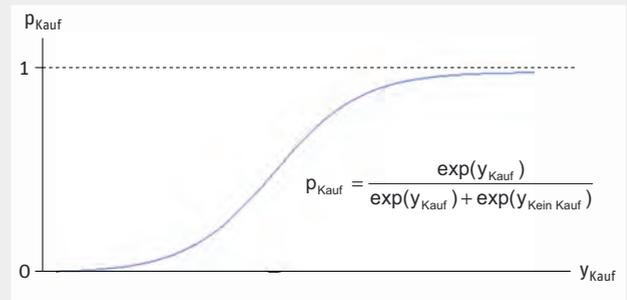


Abbildung 2: Logistische Funktion der Auswahlwahrscheinlichkeit

unabhängigen Variablen Merkmale der Befragten wie Einkommen oder Alter, die über die Befragten hinweg variieren, aber für alle Alternativen gleich sind. Für jede Alternative i werden eigene Regressionskoeffizienten geschätzt, auf Basis derer wiederum Auswahlwahrscheinlichkeiten berechnet werden können:

$$p_i = \frac{\exp(y_i)}{\sum_j \exp(y_j)}$$

Conditional Logit-Modell

Soll nicht nur der Einfluss von Merkmalen der Befragten, sondern auch der von Merkmalen der Alternativen auf die Entscheidung untersucht werden, ist das Conditional Logit-Modell zu verwenden. Abbildung 3 zeigt einen Ausschnitt aus einer Datenbasis zur Analyse des Einflusses von Leistung, Verbrauch und Entfernung zum Händler auf die Wahl eines Autos. Die Merkmale variieren entweder nur über die Alternativen (Leistung, Verbrauch) oder über Alternativen und Befragte (Entfernung). Entsprechend der Anzahl der Alternativen liegen für jeden Befragten

#	Alternative (Auto)	Gewählte Alternative	Leistung	Verbrauch auf 100 km	Entfernung zum Händler
1	1	0	100 PS	6,8 Liter	3 km
1	2	1	75 PS	5,4 Liter	2 km
1	3	0	100 PS	7,1 Liter	5 km
2	1	0	100 PS	6,8 Liter	4 km
2	2	0	75 PS	5,4 Liter	5 km
2	3	1	100 PS	7,1 Liter	8 km

Abbildung 3: Daten für ein Conditional Logit-Modell

mehrere – in diesem Beispiel drei – Datensätze vor. Eine Dummy-Variablen (Gewählte Alternative) gibt dann die Entscheidung des Befragten an. Beispielsweise hat der Befragte 1 das Auto 2 ausgewählt. Im Gegensatz zum MNL-Modell wird je Variable nicht für alle Alternativen ein eigener, sondern ein für alle Alternativen gleicher Regressionskoeffizient geschätzt. Die Berechnung der Auswahlwahrscheinlichkeiten erfolgt analog zum MNL-Modell. Darüber hinaus erlaubt das Conditional Logit-Modell auch, Merkmale der Befragten wie das Einkommen, deren Ausprägungen für alle Alternativen gleich sind, zu berücksichtigen. ◀

► **Literatur**

Cameron, A. C.; Trivedi, P. K.: Multinomial Models, In: Microeconomics, Cambridge et al., 2005, S. 490-503.

Maddala, G. S.: Discrete Regression Models, In: Limited-Dependent and Qualitative Variables in Econometrics, Cambridge et al., 1983, S. 13-46.

IfaD – Institut für angewandte Datenanalyse GmbH

IfaD wurde 1976 in Hamburg gegründet. Die Aufgabe bis heute: Marktforscher in Instituten und Unternehmen bei ihren vielfältigen Aufgaben mit maßgeschneiderten Dienstleistungen und Softwareprodukten professionell zu unterstützen. IfaD fasst seine fünf Produktbereiche als „Research Support“ zusammen.

- ▶ Data Sciences: Beratung, Statistik, Analysen
- ▶ Data Support: Daten erfassen, Tabellieren, Codieren
- ▶ CIS Software: Befragung, Codierung, Panel
- ▶ CIS Fragebogen-Services: Scripting, Hosting, Hardware
- ▶ reportbook: Analyzer, Dashboard, Presenter, Network

Eine grundlegende Expertise von IfaD ist die Datenanalyse. Der Institutsname ist hier Verpflichtung. Wo Standardverfahren nicht mehr die gewünschten Möglichkeiten bieten, passt IfaD laufend die vorhandenen Methoden an, erweitert oder optimiert sie in ihrer Darstellungsform. Für die konkrete Kundenaufgabe entwickelt der Bereich Data Sciences Varianten bestehender Verfahren oder neue Methoden.

▶ Die Autoren



Johannes Lüken, Diplom Psychologe, ist der Leiter des Bereichs Data Sciences bei IfaD, Institut für angewandte Datenanalyse, Hamburg. Schwerpunkte seiner Tätigkeit sind die Entwicklung neuer Methoden, deren Implementierung in Analysetools, sowie die Anwendung, Schulung und Beratung im Hinblick auf diese Verfahren.



Prof. Dr. Heiko Schimmelpfennig, Diplom-Kaufmann, ist Projektleiter für Data Sciences bei IfaD, Institut für angewandte Datenanalyse, Hamburg. Er ist schwerpunktmäßig für die methodische Beratung, Anwendung der Analyseverfahren und Schulungen verantwortlich. Dr. Schimmelpfennig hat eine Professur an der Business and Information Technology School (BITS) am Standort Hamburg für quantitative Methoden der Betriebswirtschaftslehre.

IfaD

Institut für angewandte
Datenanalyse GmbH

Uhlandstraße 68
22087 Hamburg

Telefon: +49 40 25 17 13 0
Telefax: +49 40 25 17 13 33

info@ifad.de
www.ifad.de