

# planung & analyse

Zeitschrift für Marktforschung und Marketing [www.planung-analyse.de](http://www.planung-analyse.de)

6/2013 D11700F

## Special

Für gute Benutzbarkeit:  
Usability-Forschung

## Report

Basis jeder Erhebung:  
Feldarbeit

## Forschung

Service-Siegel: Wirkung  
auf die Kaufbereitschaft

## Schwerpunkt

# Segmentierung



# 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

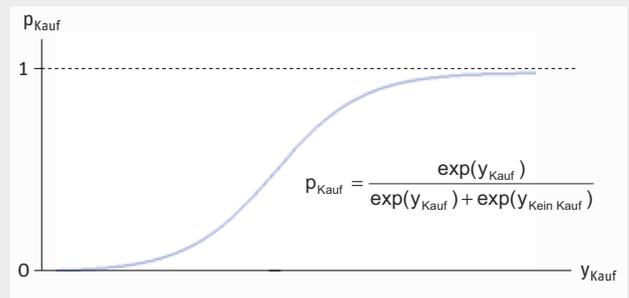


Abbildung 2: Logistische Funktion der Auswahlwahrscheinlichkeit

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. ◀

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

In Ausgabe 1/2014: Maximum-Likelihood-Schätzung

#### ► 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.