

planung & analyse

Zeitschrift für Marktforschung und Marketing
Eine Marke der dfv Mediengruppe

www.planung-analyse.de

4/2014 D11700F

Schwerpunkt

Fishing the River of Information



Special

Pharmamarktforschung –
Der Gesundheitsbranche
auf der Spur

Report

Alles nur Spiel –
Gamification in der
Marktforschung

Kongressbericht

15. M-Motion-Tag:
Markenführung nicht
ohne Marktforschung

Maximum Difference Scaling (MaxDiff)

Maximum Difference Scaling (MaxDiff) bzw. Best-Worst-Scaling ist ein Verfahren zur Messung der Wichtigkeit oder Präferenz beispielsweise von Produkteigenschaften, Marken oder Werbeslogans. Im Allgemeinen trennt es besser zwischen den zu beurteilenden Items als entsprechende Abfragen auf einer Ratingskala bzw. vereinfacht die Erhebung im Vergleich zu einer Konstantsummenskala.

Klassisches MaxDiff

Aus einer vorgegebenen Menge von Items wird jedem Befragten mehrfach eine Teilmenge aus zumeist drei bis fünf Items vorgelegt (siehe Abbildung). Soweit möglich wird jedem Befragten jedes Item durchschnittlich drei Mal vorgelegt und über alle Befragten hinweg annähernd gleich häufig mit jedem anderen kombiniert.

Welche der folgenden vier Eigenschaften ist für Sie persönlich die wichtigste und welche die unwichtigste bei der Entscheidung für eine Kreditkarte?

Am wichtigsten		Am unwichtigsten
<input type="checkbox"/>	Kundenzeitschrift mit Rabatt-Tipps	<input type="checkbox"/>
<input type="checkbox"/>	Abrechnung erst nach 3 Monaten	<input type="checkbox"/>
<input type="checkbox"/>	Kostenloses Bargeld am Automaten	<input type="checkbox"/>
<input type="checkbox"/>	Hohe Guthabenverzinsung	<input type="checkbox"/>

Abbildung: Beispiel einer typischen MaxDiff-Aufgabe

Der Befragte hat jeweils das aus seiner Sicht beste/wichtigste und schlechteste/unwichtigste Item anzugeben. Damit liegt eine Reihe von Informationen über die Items einer einzelnen Aufgabe vor: Bei drei Items je Aufgabe lässt sich eine vollständige und eindeutige Rangfolge dieser drei Items bilden. Bei vier Items liefern die Antworten Informationen zu fünf der sechs möglichen Paarvergleiche. Ist beispielsweise A das wichtigste Item und D das unwichtigste, so ist folglich $A > B$, $A > C$, $A > D$, $B > D$ und $C > D$, wobei „>“ für „wichtiger als“ steht. Bei fünf Items je Aufgabe sind noch die Präferenzen von sieben der zehn möglichen Paare definiert.

Ziel der Auswertung der so gewonnenen Daten ist es, für jedes Item i einen Wert u_i zu bestimmen, der dessen Nutzen widerspiegelt. Dazu wird ein Logit-Modell zugrunde gelegt, das die Wahrscheinlichkeit liefert, mit der ein Item aus einer Menge von Items

$$\text{als bestes } P_i = \frac{\exp(u_i)}{\sum_j \exp(u_j)} \quad \text{bzw. schlechtestes } P_i = \frac{\exp(-u_i)}{\sum_j \exp(-u_j)}$$

ausgewählt wird. Mittels Hierarchical-Bayes-Schätzung werden damit für jeden Befragten individuelle Nutzenwerte bestimmt. Dabei fließen nicht nur die eigenen Antworten ein, sondern es werden zur Stabilisierung der Schätzungen auch die Angaben der übrigen Befragten genutzt. Zur besseren Interpretation werden die geschätzten Nutzenwerte anschließend wiederum gemäß Logit-Modell in die erwartete Wahrscheinlichkeit transformiert, mit der ein Item in einer beliebigen Aufgabe als bestes ausgewählt wird. Die sich ergebenden Werte werden schließlich so normiert, dass für jeden Befragten die Summe über alle Items gleich 100 ist.

Anchored MaxDiff

Die (transformierten und normierten) Nutzenwerte eines herkömmlichen MaxDiff können nur in Relation zueinander interpretiert werden. Es kann nicht ausgeschlossen werden, dass beispielsweise alle betrachteten Items eigentlich unwichtig sind. Um einen Rückschluss auf die absolute Wichtigkeit bzw. Präferenz zu ermöglichen, ist die Erhebung zusätzlicher Informationen notwendig. Entweder wird nach jeder einzelnen Aufgabe gefragt, ob alle, einige oder keins der soeben gezeigten Items sehr wichtig sind. Oder am Ende aller Aufgaben wird der Befragte gebeten, aus allen Items diejenigen auszuwählen, die für ihn sehr wichtig sind. Werden diese Angaben bei der Schätzung der Nutzenwerte mit berücksichtigt, repräsentieren positive Nutzenwerte wichtige und negative Nutzenwerte unwichtige Items.

MaxDiffs mit sehr vielen Items

Bei insgesamt 30 zu beurteilenden Items, die jeweils drei Mal vorgelegt werden, und fünf Items je Aufgabe hat ein Befragter 18 ($= 30 \cdot 3 / 5$) Aufgaben zu bewältigen. Verdoppelt sich die Anzahl an Items, verdoppelt sich auch die Anzahl der notwendigen Aufgaben. Um ein MaxDiff mit sehr vielen Items zweckmäßig einsetzen zu können, ist das klassische Vorgehen somit zu modifizieren.

Es wird vorgeschlagen, entweder jedem Befragten (a) nur eine individuelle Auswahl der Items oder (b) jedes Item weniger als durchschnittlich drei Mal vorzulegen. Dadurch verringert sich die Informationsdichte, die für jeden Befragten aufgrund seiner eigenen Antworten vorliegt. Da es das Prinzip der Hierarchical-Bayes-Schätzung ist, fehlende Information aus der Gesamtstichprobe zu gewinnen, ist es insbesondere bei diesen Vorgehensweisen wichtig, dass die Stichprobe bezüglich der Präferenzen weitgehend homogen ist. Falls die Menge an Items nur so groß ist, dass jedes Item jedem Befragten bei einer akzeptablen Menge an Aufgaben mindestens ein Mal vorgelegt werden kann, ist (b) der Option (a) vorzuziehen.

Um die Validität der geschätzten Nutzenwerte zu erhöhen, kann ein Befragter vorab gebeten werden, die Items hinsichtlich seiner Wichtigkeit/Präferenz zu gruppieren. Zum Beispiel hat er aus 80 Items zunächst die wichtigsten 20, aus den verbleibenden 60 wiederum die (nächst)wichtigsten 20 etc. auszuwählen. Erst anschließend wird dann ein MaxDiff in der Variante (a) oder (b) durchgeführt. Die Information aus der Gruppierung fließt mit in die Auswertung und in (b) auch mit in die Auswahl der Items ein. Allerdings bleibt die Erhebung ähnlich lang wie im klassischen MaxDiff, vermeidet aber eine Monotonie infolge mehrerer Dutzend MaxDiff-Aufgaben.

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

In Ausgabe 5/2014: Total Unduplicated Reach and Frequency (TURF)

► Literatur

Hendrix, P.; Drucker, S.: Alternative Approaches to MaxDiff with Large Sets of Disparate Items – Augmented and Tailored MaxDiff, Proceedings of the Sawtooth Software Conference, Santa Rosa, 2007, 169-187.

Lattery, K.: Anchoring Maximum Difference Scaling Against a Threshold – Dual Response and Direct Binary Responses, Sawtooth Research Paper Series. Sequim, 2011.

Sawtooth Software: The MaxDiff/Web v6.0 Technical Paper. Sawtooth Software Technical Paper Series. Sequim, 2005-2007.

Wirth, R.; Wolfrath, A.: Using MaxDiff for Evaluating Very Large Sets of Items, Proceedings of the Sawtooth Software Conference, Orlando, 2012, 59-78.