Mikrogeographische Marktsegmentierung mit Mixture-Modellen

Stefanie Rankl

Mikrogeographische Marktsegmentierung mit Mixture-Modellen

Springer Gabler

Stefanie Rankl
Centrum für Marktforschung
Universität Passau
Neuburg, Deutschland

Dissertation Universität Passau, 2013

ISBN 978-3-658-02653-0
DOI 10.1007/978-3-658-02654-7

ISBN 978-3-658-02654-7 (eBook)

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Gabler

© Springer Fachmedien Wiesbaden 2013

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Gabler ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media.
www.springer-gabler.de

Nur wer das Ziel kennt, kann treffen.
(Griechisches Sprichwort)

Inhaltsverzeichnis	VII
Abbildungsverzeichnis	XI
Tabellenverzeichnis	XIII
Abkürzungsverzeichnis	XV
Symbolverzeichnis	XVII
I Einleitung	1
1. Problemstellung	1
2. Zielsetzung	3
3. Aufbau der Arbeit	4
II Mikrogeographische Marktsegmentierung	7
1. Grundlagen der Marktsegmentierung	7
1.1 Segmentierungsansätze	7
1.2 Segmentierungsmethoden	15
2. Grundlagen der mikrogeographischen Marktsegmentierung	28
2.1 Prinzipien der mikrogeographischen Marktsegmentierung	29
2.2 Entwicklung mikrogeographischer Systeme	30
2.3 Das System POINTplus®	33
2.4 Unterschiede zu traditionellen Segmentierungsverfahren	36
2.5 Beurteilung mikrogeographischer Segmentierungskriterien	40
2.6 Kritische Würdigung mikrogeographischer Segmentierung	42
III Mixture Modelle	45
1. Heterogenität von Konsumenten	45
1.1 Unbeobachtete Heterogenität	46
1.2 Modellierung von unbeobachteter Heterogenität	48

2. Grundlagen der Mixture Modelle	52
2.1 Einführung	52
2.2 Formale Darstellung des Grundmodells	53
2.3 Identifizierbarkeit	56
2.4 Maximum-Likelihood-Schätzung	57
3. Iterative Optimierungsverfahren	60
3.1 Das Newton-Raphson-Verfahren	60
3.2 Der EM-Algorithmus	62
3.3 Vergleich der Verfahren	66
4. Auswahl der optimalen Segmentzahl	68
4.1 Likelihood-Quotienten-Test (LQT)	68
4.2 Informationskriterien	70
4.3 Klassifikationskriterien	73
5. Erweiterungen des Grundmodells	75
6. Mixture Regression Modelle	76
6.1 GLIMMIX	78
6.2 Logit Mixture Regression	82
IV Empirische Studie	85
1. Datenbasis	85
2. Untersuchungsziel	89
3. Modellselektion	90
4. Verwendete Software	91
5. Ergebnisse der Segmentierungsanalyse	93
5.1 Untersuchung der Korrelationen	93
5.2 Auswahl der optimalen Segmentzahl: Gütekriterien	100
5.3 Parameter des Modells	103
5.4 Interpretation der Segmente	110

Inhaltsverzeichnis	IX
6. Weiterführende Schritte	115
V Abschließendes Fazit	118
Anhang	121
Anhang 1: Einige Verteilungen aus der Exponentialfamilie	122
Anhang 2: Der EM-Algorithmus für GLIMMIX	123
Anhang 3: Variablenübersicht	129
Anhang 4: GfK Euro-Socio-Styles®	134
Anhang 5: Bivariate Korrelationen	135
Anhang 6: Ergebnisse des gewählten 3-Segment-Modells	139
Anhang 7: Mittelwertunterschiede in den Segmenten	142
Literaturverzeichnis	143

Abbildungsverzeichnis

Abbildung II-2: Segmentierungsansätze	10
Abbildung II-3: Soziodemographische Segmentierungskriterien	10
Abbildung II-4: Psychographische Segmentierungskriterien	12
Abbildung II-5: Verhaltensorientierte Segmentierungskriterien	14
Abbildung II-6: Klassifikation der Clustermethoden	20
Abbildung II-7: Anbieter von mikrogeographischen Systemen	32
Abbildung II-8: Mikrogeographisches Datenangebot der GfK	33
Abbildung II-10: Räumliche Gliederungsebenen	37
Abbildung III-2: Modellierungsansätze unbeobachteter Heterogenität	48
Abbildung III-3: Erweiterungen des Grundmodells	75
Abbildung IV-2: Variablengruppen	87
Abbildung IV-3: Informationskriterien und AWE im gewählten Modell	102
Abbildung IV-4: Indexmittelwerte in den drei Segmenten	114
Abbildung IV-5: Kundenindex auf der Ebene der 5-stelligen Postleitzahlen	117

Tabellenverzeichnis

Tabelle II-1: Klassifikation ausgewählter Segmentierungsmethoden	16
Tabelle II-2: Zuordnungsmatrix verschiedener Clustermethoden	19
Tabelle III-1: Überblick über verschiedenen Informationskriterien	72
Tabelle III-2: Ausgewählte kanonische Link-Funktionen	80
Tabelle IV-1: Regressionsmodelle in Latent GOLD®	92
Tabelle IV-2: Korrelationsmatrix für ausgewählte Variablen	94
Tabelle IV-3: Kollinearitätsstatistik	97
Tabelle IV-4: Korrelationsmatrix für das gewählte Modell	99
Tabelle IV-5: Kollinearitätsstatistik für das gewählte Modell	99
Tabelle IV-6: Gütekriterien	101
Tabelle IV-7: Prozentsatz der richtig klassierten Fälle	101
Tabelle IV-8: Parameter des gewählten Modells	105
Tabelle IV-9: Regressionsparameter mit zugehörigen Odds Ratio	107
Tabelle IV-10: Veränderungen der Odds	108
Tabelle IV-11: Segmentgrößen und deren Mittelwerte	110
Tabelle IV-12: A-posteriori Wahrscheinlichkeiten ausgewählter Fälle	112
Tabelle IV-13: Mittelwerte in den Segmenten	112

Abkürzungsverzeichnis

Abb.	Abbildung
ACORN™	A Classification of Residential Neighborhoods
AIC	Akaike Information Criterion
AIC ₃	Modified Akaike Information Criterion
AID	Automatic Interaction Detection
AWE	Average Weight of Evidence
AZ DIAS	AZ Daten-, Informations- und Adress-System
Bd.	Band
BDSG	Bundesdatenschutzgesetz
BIC	Bayesian Information Criterion
BVSD	Bevölkerungsstrukturdaten
bzw.	beziehungsweise
ca.	circa
CAIC	Consistent Akaike Information Criterion
CART	Classification And Regression Trees
CAS	Clustertypologisches Analyse- und Selektionssystem
CHAID	Chi-square Automatic Interaction Detection
d.h.	das heißt
DINKs	Double Income No Kids
EM	Expectation-Maximization
ESS	Euro Socio Styles®
et al.	et alii
etc.	et cetera
f.	folgende
ff.	fortfolgende
GLIMMIX	Generalized Linear Model Mixture
GLM	Generalisiertes Lineares Modell
H.	Heft

Hrsg.	Herausgeber
IC	Information Criterion
Inc.	Incorporated
Infas	Institut für angewandte Sozialwissenschaft
Iss.	Issue
Jg.	Jahrgang
KNN	Künstliches Neuronales Netz
LL	Log-Likelihood
LQT	Likelihood-Quotienten-Test
MAIC	Modified Akaike Information Criterion
MDS	Multidimensionale Skalierung
Mio.	Millionen
MLP	Multilayer-Perceptron
NEC	Normalized Entropy Criterion
Nr.	Nummer
PRIZM®	Potential Rating Index for ZIP Markets
S.	Seite
SE	Societas Europaea
SOFM	Self-Organizing Feature Maps
sog.	sogenannt
Tab.	Tabelle
u.a.	unter anderem
usw.	und so weiter
vgl.	vergleiche
VIF	Variance Inflation Factor
Vol.	Volume
z.B.	zum Beispiel
ZIP	Zone Improvement Plan

Symbolverzeichnis

α	Signifikanzniveau
$\underline{\beta}_s$	Vektor der Regressionsparameter im Segment s
b_{js}	j-ter Regressionsparameter des s-ten Segments
$\underline{b}(\Phi)$	Vektor der ersten Ableitungen der LL-Funktion
c	Iterationsindex
$\frac{\partial f}{\partial x}$	Ableitung einer Funktion f nach der Variablen x
$\frac{df}{dx}$	Ableitung einer Funktion f nach der Variablen x
d	Strafkomponente
ε_i	Störterm des i-ten Objekts
η	linearer Prädiktor
E_s	Entropiekriterium
$\hat{f}(s \underline{x}_i)$	(geschätzte) a-posteriori Wahrscheinlichkeit der Segmentzugehörigkeit
$f_s(\underline{x}_i \theta_s)$	die bedingte Verteilungsfunktion im Segment Ω_s
$f(\underline{x}_i \Phi)$	unbedingte Verteilung der Segmentierungskriterien $\underline{x}_i$
g	Anzahl der Segmente / Teilgesamtheiten / Gruppen / Komponenten
$H(\Phi^c)$	Hesse-Matrix der zweiten Ableitungen der LL-Funktion
i	Objekt / Konsument / Haushalt mit $i=1, \dots, n$
j	Variable / Segmentierungskriterium mit $j=1, \dots, k$
k	Anzahl der Variablen / Segmentierungskriterien
λ	Lagrange Multiplikator
λ_s	Streuungsparameter

ℓ	Likelihood-Quotient
L	Likelihood-Funktion
L^*	Log-Likelihood-Funktion
L_c^*	Log-Likelihood-Funktion der vollständigen Daten
$L(\underline{x}_i \Phi)$	Likelihood-Funktion der Mischverteilung $f(\underline{x}_i \Phi)$
$L^*(\underline{x}_i \Phi)$	Log-Likelihood-Funktion der Mischverteilung
μ	Erwartungswert
n	Anzahl der Objekte / Konsumenten / Haushalte
N	Anzahl der Beobachtungswerte
Ω	Grundgesamtheit
Ω_s	s-tes Segment mit $s=1, \dots, g$
Φ	Vektor der unbekannten Parameter
$\Phi^{(0)}$	Startwert für den Vektor der unbekannten Parameter
Φ^c	Vektor der zu schätzenden Parameter im c-ten Iterationsschritt
π_s	Mischungsanteil des s-ten Segments (a-priori-Wahrscheinlichkeit der Segmentzugehörigkeit)
$\hat{\pi}_s$	Schätzung für π_s
P	Anzahl der zu schätzenden Parameter
$P(x)$	Wahrscheinlichkeit für die Variable x
p_{is}	(geschätzte) a-posteriori Wahrscheinlichkeit der Segmentzugehörigkeit
r^2	Bestimmtheitsmaß
s	Segment / Teilgesamtheit / Gruppe / Komponente mit $s=1, \dots, g$
$s_{b_{js}}$	Standardfehler von b_{js}

σ^2	Varianz einer Normalverteilung
θ_s	(unbekannte) Verteilungsparameter im Segment Ω_s
$\underline{\theta}_s$	Vektor der (unbekannten) Verteilungsparameter im s-ten Segment
T_j	Toleranz der Variablen x_j
V_{ijs}	Varianzfunktion
VIF_j	Variance Inflation Factor der Variablen x_j
w	Gewicht
$\underline{X}_c$	vollständige Datenmatrix $\underline{X}_c = (\underline{x}_i, \underline{z}_i)'$
$\underline{x}_i$	Vektor der Ausprägungen der Segmentierungskriterien beim i-ten Objekt
y_i	abhängige Variable des i-ten Objekts
$\underline{y}_i$	Vektor der abhängigen Variablen des i-ten Objekts
$\underline{z}_i$	Vektor der unbeobachteten Segmentzugehörigkeit des i-ten Objekts
z_{is}	Ausprägung des Vektors der unbeobachteten Segmentzugehörigkeit, dass Objekt i dem Segment s angehört